

Índice General

Parte I Metodología	1
Capítulo I: Programación Estocástica	3
1 Introducción	3
2 Definiciones básicas	4
3 Programación con restricciones probabilísticas	5
4 Función objetivo aleatoria	12
5 Bibliografía	19
Capítulo II: Aproximación metodológica a la optimización en la incertidumbre	23
1 El latir de una sociedad compleja	23
2 De los principios a la lógica	25
3 Una propuesta de matemática de la incertidumbre	29
4 Breve referencia a la matemática numérica para la optimización . .	32
5 Elementos no numéricos para la optimización	37
6 Consideraciones finales	45
7 Bibliografía	48
Capítulo III: Programación estocástica multiobjetivo	51
1 Introducción	51
2 Enfoque Multiobjetivo	55
3 Enfoque Estocástico	64
4 Conclusiones	68
5 Bibliografía	69
Capítulo IV: Métodos y Modelos de Programación Lineal Borrosa	71
1 Introducción	71
2 Conceptos básicos	74
3 Programación Lineal Borrosa	79
4 Ejemplos	87
5 Conclusión y epílogo	90

6	Bibliografía	92
Capítulo V: Branch-and-fix coordinado, un esquema de resolución de problemas estocásticos multietápicos 0–1 mixtos.		95
1	Introducción	95
2	Minimización del valor esperado	97
3	Minimización de las funciones objetivo <i>valor esperado–riesgo</i> y CaR	100
4	Acotación del Branch-and-Bound	102
5	Branch-and-Fix Coordinado	102
6	Bibliografía	107
Capítulo VI: Modelado de algoritmos de descomposición con GAMS		111
1	Introducción	111
2	Descomposición de Benders	112
3	Relajación Lagrangiana	129
4	Implantación en grandes modelos	144
5	Bibliografía	145
Parte II Aplicaciones		147
Capítulo VII: Optimización estocástica aplicada a plantas petroquímica y de producción de energía		149
1	Introducción	149
2	Las ideas	152
3	Metodología	152
4	Modelización de unidades: Diseños y modos de funcionamiento .	160
5	Nomenclatura	169
6	Los métodos de optimización	173
7	Conclusiones	173
8	Agradecimiento	173
9	Bibliografía	174
Capítulo VIII: Long-term electric power planning in liberalized markets using the Bloom and Gallant formulation		179
1	Introduction and Motivation	179
2	The load-duration curve	180
3	Thermal Units	181
4	Matching the load-duration curve	181
5	Bloom & Gallant's model for matching the load-duration curve when there are non-load-matching constraints	183
6	Long-term maximization of profit in a “competitive” market	186
7	Coding the load-matching constraints	190

8	The Ford-Fulkerson column-generation method applied to the multi-interval problem	190
9	Murtagh and Saunders algorithm using a Column Generation procedure	191
10	Computational results	198
11	Conclusions	203
12	Bibliografía	203
Capítulo IX: Aplicaciones en sistemas de energía eléctrica		207
1	Introducción	207
2	Planificación de la expansión de la generación	208
3	Programación semanal	217
4	Programación semanal en mercados eléctricos	225
5	Bibliografía	235
Capítulo X: Algunos problemas estocásticos de localización discreta: un enfoque unificador		239
1	Introducción	239
2	Modelos de recurso para Problemas Estocásticos de Localización Discreta	242
3	Función de recurso para Problemas Estocásticos de localización-asignación (ELA)	245
4	Función de recurso para Problemas Estocásticos de localización-rutas (ELR)	247
5	Algoritmos para ELA	249
6	Algoritmos para ELR	253
7	Conclusiones	260
Capítulo XI: Localización Minimax con Incertidumbre		263
1	Introducción	263
2	El modelo	265
3	El criterio del valor esperado	269
4	El criterio de máximo cubrimiento en probabilidad	271
5	El criterio de la restricción de incertidumbre	275
6	Conclusiones y líneas futuras de investigación	279
7	Bibliografía	279
Capítulo XII: Cálculo de medianas robustas con incertidumbre en las demandas		285
1	Introducción	285
2	Cálculo de los escenarios	287
3	Planteamiento del modelo coordinado	291
4	Cuando la demanda depende de un único parámetro	296

5	Conclusiones	301
6	Bibliografía	301
Capítulo XIII: Solución numérica de problemas de control estocástico en economía		303
1	Introducción	303
2	Un modelo de crecimiento sencillo	306
3	Condiciones de estabilidad: caso determinista	309
4	Condiciones de estabilidad: el caso estocástico	313
5	Solución del modelo de crecimiento estocástico básico	314
6	Resumen	321
7	Bibliografía	322
Capítulo XIV: Learning by Doing e incertidumbre aditiva: solución analítica		325
1	Introducción	325
2	Revisión de la literatura	326
3	Discusión sobre el problema abordado	329
4	Modelo	331
5	Solución	332
6	Implicaciones Económicas	334
7	Futuras investigaciones	335
8	Bibliografía	338
Capítulo XV: Programación estocástica por metas		343
1	Introducción	343
2	Programación por metas determinista	344
3	Programación por metas estocástica	346
4	Una aplicación al diseño de sistemas Bonus-Malus	354
5	Definición de un sistema Bonus-Malus	355
6	Bibliografía	360
Capítulo XVI: Planificación óptima de la actividad quirúrgica en hospitales públicos mediante un modelo de Programación Compromiso Posibilística		363
1	Introducción	363
2	Metodología	364
3	Presentación del problema.	370
4	Conclusiones	382
5	Bibliografía	383

Capítulo XVII: Análisis envolvente de datos bajo condiciones de incertidumbre: Aplicación a la liga de fútbol profesional española e italiana	387
1 Introducción	387
2 Análisis envolvente de datos clásico	388
3 Análisis envolvente de datos con incertidumbre o datos imprecisos	390
4 Modelos de eficiencia con incertidumbre	394
5 Aplicación a las ligas de fútbol profesionales	399
6 Resultados computacionales	401
7 Conclusiones	406
8 Bibliografía	407

Presentación

En la primavera de 2002, ASEPUMA editó su primer volumen monográfico sobre programación con objetivo múltiples que fue coordinado por los profesores Rafael Caballero y Gabriela Fernández. Con él se iniciaba una nueva etapa dentro del programa de publicaciones de la Asociación Española de Profesores Universitarios de Matemática Aplicada a la Economía y la Empresa.

El éxito de esta publicación animó a la Junta Directiva a proponer la edición de un segundo volumen. En la reunión de la Junta Directiva de ASEPUMA celebrada el ? se acordó encargar la coordinación de este volumen sobre Optimización en Incertidumbre a los profesores Emilio Cerdá, Laureano Escudero y Ramón Sala.

En la primera toma de contacto para diseñar el contenido de este número se consideró conveniente la incorporación de Antonio Alonso-Ayuso, y así quedó constituido el grupo responsable de llevar a cabo esta edición.

Los coordinadores de este monográfico contactaron con diversos profesores especialistas en este campo y aceptaron realizar un trabajo para este número, cada uno desde una visión particular.

El libro se estructura en dos partes: una parte *metodológica* y otra de *aplicaciones*, formando un total de 17 capítulos de contenido y orientación diferentes, pero a la vez complementarios.

La primera parte contiene seis trabajos, mientras que son nueve las aplicaciones que se comentan.

El primer Capítulo elaborado por Emilio Cerdá y Julio Moreno, presenta una Introducción a la Programación Estocástica. Se hace un recorrido por sus métodos fundamentales, haciendo especial énfasis en los problemas de recursos, restricciones probabilísticas y modos de transformar un objetivo estocástico en su equivalente determinista.

Se trata, en definitiva, de presentar una panorámica general de la **programación estocástica**, que proporcione una buena introducción a los recién llegados a este campo de la metodología fundamental.

El profesor Jaime Gil-Aluja de la Universitat de Barcelona, plantea un Capítulo sobre la visión fuzzy de la matemática de la incertidumbre. Esta introducción, junto con las referencias bibliográficas aportadas suponen un punto de partida

adecuado a la iniciación a las técnicas de la teoría de los subconjunto difusos o borrosos.

Como extensión del primer Capítulo, Caballero, Cerdá, Muñoz y Rey, presentan la **Programación estocástica multiobjetivo**. Este trabajo se centra en el estudio de problemas de decisión en los que el número de objetivos es múltiple y algunos o todos los parámetros del problema son variables aleatorias con distribución conocida. De esta forma, se relaja la hipótesis frecuente cuando se plantea un modelo de optimización referente a que el objetivo del proceso de decisión puede representarse a través de una única función a optimizar. Se acompañan de una exposición amplia de los diferentes procedimientos para resolver este tipo de problemas cuando aparecen diferentes objetivos a coordinar.

Jose Miguel Cadenas y Jose Luis Verdegay, complementan el Capítulo del profesor Gil-Aluja con la descripción de la metodología de la programación lineal fuzzy, con el trabajo titulado: **Métodos y modelos de programación lineal borrosa**.

En el trabajo de Alonso-Ayuso, Clement, Escudero, Gil y Ortuno, escriben el trabajo **Branch and fixcoordinado, un esquema de resolución de problemas estocásticos multietapicos 0-1 mixtos**. En muchos problemas de optimización, en especial en aquellos que evolucionan en el tiempo a lo largo de un horizonte de planificación dado, es habitual que algunos de los coeficientes de la función objetivo y del vector del término independiente e, incluso, de la matriz de restricciones, no se conozcan con exactitud en el momento de tomar las decisiones correspondientes, aunque se disponga de alguna información sobre ellos. En este capítulo se estudia la utilización del procedimiento llamado *Branch-and-Fix Coordinado (BFC)* para obtener la solución óptima 0-1 mixta del problema estocástico original. Se puede utilizar tanto la descomposición Lagrangiana como la descomposición de Benders, entre otras metodologías, para aprovechar la estructura del *DEM*. Estos tipos de descomposiciones permiten obtener soluciones factibles continuas una vez obtenidas soluciones factibles 0-1, de forma que todas ellas satisfagan las llamadas condiciones de *no anticipación* en la solución óptima.

No podía acabar esta primera parte con una referencia a los procedimientos informáticos de resolución de los problemas estocásticos. Santiago Cerisola, Andrés Ramos y Álvaro Baillo, han escrito el trabajo **Modelado de algoritmos de descomposición con GAMS** describen los métodos fundamentales de resolución de problemas estocásticos, la descomposición de Benders y la relajación Lagrangiana que han sido citado en los capítulos anteriores. Además aportan la estructura del código de resolución en lenguaje GAMS de estos dos procedimientos, así como su aplicabilidad a la resolución de unos ejemplos sencillos pero muy ilustrativos.

La segunda parte está dedicada a presentar algunas aplicaciones en las que se emplean los métodos de resolución de la optimización estocástica, borrosa, robusta, etc. La secuencia de los capítulos va desde las aplicaciones .

Con el Capítulo 7 escrito por Francisco Javier Quintana con el título Optimización estocástica aplicada al diseño de procesos, se inicia la parte de aplicaciones. En este capítulo se describe la optimización estocástica aplicada al diseño de procesos, entre los que encontramos las plantas petroquímicas y de producción de energía. Exponemos algunas características y ventajas de este sistema con ejemplos que pueden aclarar ideas.

Los dos capítulos siguientes presentan aplicaciones de la optimización estocástica al sector eléctrico. El primero de ellos de Narcis Nabona y Adela Pagés presenta el trabajo: **Long-term electric power planning in liberalized markets using the Bloom and Gallant formulation**, que como el título indica esta redactado en inglés. El Capítulo plantea la planificación a largo plazo es el elemento clave para la generación de energía por parte de las empresas productoras, pero esta planificación tiene su concreción a en la planificación de las decisiones a corto plazo.

En el Capítulo de Santiago Cerisola, Andrés Ramos y Álvaro Baíllo, sobre **Aplicaciones en sistemas de energía eléctrica** en él se presentan varios ejemplos característicos de planificación y operación de sistemas de energía eléctrica para cuya resolución se utilizan frecuentemente técnicas de descomposición.

El Capítulo siguiente escrito por María Albareda-Sambola y Elena Fernández con título: **Algunos problemas estocásticos de localización discreta: un enfoque unificador**, analiza los problemas discretos de localización tratando de seleccionar las ubicaciones óptimas para un conjunto de centros de servicios (plantas) entre un conjunto de ubicaciones potenciales que es conocido a priori. Algunos de los elementos del problema son datos aleatorios, y con ello obliga a introducir la metodología de la programación estocástica como método de resolución de este tipo de problemas.

Los capítulos 11 y 12, también presentan dos aplicaciones a los problemas de localización, aunque difieren en el procedimiento de solución, ya que mientras que el primero utiliza la metodología de la programación estocástica, el segundo utiliza la optimización robusta como enfoque para la resolución.

El Capítulo 11, escrito por Blas Pelegrín lleva por título: **Localización Minimax con Incertidumbre**. El trabajo presenta un modelo general en el plano, donde la distancia viene medida por cualquier norma y se presenta incertidumbre en los coeficientes de la distancia, que vienen dados por variables aleatorias con distribuciones de probabilidad arbitrarias. Se consideran tres criterios de decisión, se analizan las propiedades básicas de los correspondientes modelos de optimización y se plantean procedimientos para su resolución.

En Capítulo siguiente **Cálculo de medianas robustas con incertidumbre en las demandas**, escrito por María José Canós, Marisa Martínez y Manuel Mocholí, describe uno de los problemas a los que se enfrentan las empresas es decidir donde ubicar sus instalaciones de modo que sus costes de aprovisionamiento y distribución sean mínimos. La optimización robusta no necesita que la incerti-

dumbre esté provocada por un solo parámetro (el tiempo), como la optimización dinámica, ni tampoco que exista una distribución de probabilidad asociada, como la optimización estocástica, requisito no trivial en problemas que, como los de localización, estudian fenómenos únicos con poca o ninguna información histórica.

El Capítulo **Solución Numérica de Problemas de Control Estocástico en Economía** ha sido escrito por Domínguez, Novales, Pérez y Ruiz. En el trabajo se analiza el comportamiento de un consumidor que trata de maximizar su nivel de utilidad agregada en el tiempo, que deriva del consumo de los distintos bienes, así como del nivel de ocio de que disfruta en cada período. Las restricciones a que se enfrenta este decisor, una para cada instante de tiempo, especificarán que la cantidad que paga por los bienes que consume no puede exceder de la renta de que dispone. En un contexto más amplio, los mercados de capitales existen para que el consumidor pueda reservar parte de su renta cada período en la forma de ahorro. De ese modo, un período puede decidir llevar a cabo un consumo cuyo valor de mercado es inferior a la renta del período, ahorrando la renta no gastada. En otro período, podría suceder lo contrario, utilizando la renta de dicho instante, junto con parte del ahorro que arrastra de períodos anteriores, para financiar su nivel de consumo.

El Capítulo 14 lleva por título: **Learning by Doing e incertidumbre aditiva: solución analítica**, cuyo autor es Francisco Álvarez. En este trabajo se presenta la solución analítica al fenómeno de que algunas empresas reducen sus costes de producción a lo largo del tiempo como consecuencia de la acumulación de experiencia. Esto se denomina learning by doing. El problema de elección por parte de una empresa de su senda temporal de niveles de producción cuando ésta tiene learning by doing puede plantearse matemáticamente como un problema de optimización dinámica. Además se analiza la presencia de shocks aleatorios.

En el Capítulo 15 Antonio Heras Martínez y Ana García Aguado realizan una aplicación de la **Programación estocástica por metas** al campo actuarial, y en particular a los sistemas de tarificación a posterior como el diseño de Sistemas de Tarificación Bonus-Malus, un problema clásico de tarificación en el seguro del automóvil. La aplicación de técnicas de Programación por Metas Estocástica permite obtener algunas características deseables de las soluciones que no son tenidas en cuenta por los métodos clásicos de resolución de tales problemas.

Los dos capítulos finales están dedicados a las aplicaciones de la metodología de los conjuntos borrosos. Así el Capítulo 16 escrito por Antomil, Arenas, Bilbao, Pérez y Rodríguez Uría con el título: **Planificación óptima de la actividad quirúrgica en hospitales públicos mediante un modelo de Programación Compromiso Probabilística**. Este Capítulo analiza el fenómeno de las problemáticas de las listas de espera que es un problema de toma de decisiones racionales con presencia de criterios múltiples en un entorno de incertidumbre e imprecisión. Se trata por tanto de un problema de Programación Multiobjetivo Lineal con datos vagos/impresos. El trabajo propone un instrumento de gestión de listas

de espera quirúrgicas basándonos en los datos de un hospital público.

El último Capítulo realizado por José Emilio Boscá, Vicente Liern, Aurelio Martínez y Ramón Sala, presenta un modelo de Análisis Envolvente de Datos (DEA) en presencia de incertidumbre en los datos. El modelo propuesta presenta una aplicación a las ligas profesionales de fútbol de España e Italia, en donde se sustituye el score único de los modelos DEA tradicionales por un intervalo de eficiencia que permite analizar en qué casos un equipo (DMU) puede llegar a ser eficiente mejorando determinados aspectos de sus inputs y outputs.

No quisiéramos acabar esta presentación del volumen sin dedicar una pocas líneas a las personas que han hecho posible que este volumen vea la luz. En primer lugar a los miembros de la Junta Directiva de ASEPUMA, sin los cuales no hubiera sido posible la realización de este volumen.

A todos los autores que han colaborado de forma de desinteresada en este volumen.

Finalmente, aunque no los últimos, a Vicente Liern y Carlos Ivorra, miembros del Consejo de Redacción de Rect@, por su trabajo de conversión de los trabajos de Word a LaTeX, por la composición de los capítulos y por su dedicación a que este libro fuera una realidad en los plazo fijados.

Valencia, Junio de 2004.

ANTONIO ALONSO-AYUSO, EMILIO CERDÁ,
LAUREANO ESCUDERO Y RAMÓN SALA

Coordiadores

Primera parte

Metodología

Programación Estocástica

E. Cerdá^a, J. Moreno^b

^a Departamento Análisis Económico. UCM.

^b Departamento de Estadística. UCM.

1 Introducción

Tal como su nombre indica, la Programación Estocástica trata problemas de Programación Matemática en cuya formulación aparece algún elemento estocástico. Por tanto, mientras que en un problema determinístico de Programación Matemática, ya sea de Programación Lineal, Programación No Lineal, Programación Entera, Programación Mixta Lineal Entera o Programación Mixta No Lineal Entera, todos los datos (coeficientes) que aparecen en su formulación son números conocidos, en Programación Estocástica dichos datos (o al menos alguno de ellos) son desconocidos, aunque para ellos se conoce o se puede estimar su distribución de probabilidad. Para precisar más, veamos las dos definiciones que propone Prekopa [29]:

Primera definición: “Programación Estocástica es la ciencia que ofrece soluciones para problemas formulados en conexión con sistemas estocásticos, en los que el problema numérico resultante a resolver es un problema de Programación Matemática de tamaño no trivial”.

Segunda definición: “La Programación Estocástica trata problemas de Programación Matemática en los que algunos de los parámetros son variables aleatorias, bien estudiando las propiedades estadísticas del valor óptimo aleatorio o de otras variables aleatorias presentes en el problema o bien reformulando el problema en otro de decisión en el que se tiene en cuenta la distribución de probabilidad conjunta de los parámetros aleatorios”.

Los problemas resultantes de ambas definiciones son llamados problemas de Programación Estocástica.

La aleatoriedad en coeficientes en unos casos se deberá a la falta de fiabilidad en los datos recogidos, en otros casos a errores de medida, en otros a eventos futuros aún no conocidos, etc.

Tal como indica Dantzig [11], la Programación Estocástica comenzó en 1955 con los trabajos de Dantzig [10] y Beale [2]. y ya en la misma década alcanzó con Markowitz [23] una aplicación muy destacada al problema de selección de carteras que le llevaría a la consecución del Premio Nóbel. En [34] se recogen unas 800 referencias sobre trabajos publicados entre 1955 y 1975, clasificadas en función de su contenido.

En 1974 se celebró en Oxford (Inglaterra) la primera conferencia internacional en Programación Estocástica, organizada por Michael Dempster. En 1981 se celebró en Keszeg (Hungría) la segunda conferencia, organizada por Andra Prekopa. En dicho encuentro se puso en marcha el Committee on Stochastic Programming (COSP), como una rama de la Mathematical Programming Society. Dicho comité ha sido el responsable de organizar las sucesivas conferencias que se han ido celebrando. La novena conferencia internacional se celebró en Berlín (Alemania) en 2001 y la décima se celebrará los días 9 a 12 de Octubre de 2004 en Tucson, Arizona (USA).

El COSP ha puesto en funcionamiento la página web <http://stoprog.org> en la que se puede encontrar mucha información y documentación sobre Programación Estocástica.

2 Definiciones básicas

Se considera el siguiente problema de Programación Estocástica:

$$\begin{aligned} & \min_x \tilde{g}_0(x, \tilde{\xi}), \\ & \text{sujeto a :} \\ & \tilde{g}_i(x, \tilde{\xi}) \leq 0, i = 1, 2, \dots, m, \\ & x \in D, \end{aligned} \tag{1.1}$$

donde el conjunto $D \subset \mathbb{R}^n$, $\tilde{\xi}$ es un vector aleatorio definido sobre un conjunto $E \subset \mathbb{R}^s$. Suponemos que están dados una familia de eventos F , formada por subconjuntos de E y una distribución de probabilidad P , definida sobre F . Por tanto, para cada $A \subset E$, es $A \in F$, y la probabilidad $P(A)$ es conocida. Además suponemos que las funciones $\tilde{g}_i(x, \cdot) : E \rightarrow \mathbb{R}, \forall x, i$ son variables aleatorias y que la distribución de probabilidad P es independiente del vector de variables de decisión x .

Obsérvese que en el problema formulado (PE) para cada realización ξ del vector aleatorio $\tilde{\xi}$ se tiene un problema determinístico. Un vector $x \in D$ puede ser factible para una realización del vector aleatorio y no serlo para otra realización.

Asimismo puede ocurrir que para una realización ξ^1 sea $g_0(x^1, \xi^1) < g_0(x^2, \xi^1)$ y en cambio para otra realización ξ^2 del vector aleatorio $\tilde{\xi}$ sea $g_0(x^2, \xi^2) < g_0(x^1, \xi^2)$.

Un caso particular del problema (PE) es el siguiente problema de Programación Lineal Estocástica:

$$\begin{array}{ll} \min_x & c^T(\tilde{\xi})x, \\ \text{sujeto a:} & Ax = b, \\ & T(\tilde{\xi})x \geq h(\tilde{\xi}), \\ & x \geq 0, \end{array} \quad (1.2)$$

donde la matriz A y el vector b son determinísticos. La matriz $T(\cdot)$ y los vectores $c(\cdot)$ y $h(\cdot)$ dependen del vector aleatorio $\tilde{\xi}$ y por tanto son estocásticos.

Normalmente el problema estocástico se reemplaza por un problema determinístico, que se llama determinista equivalente cuya solución óptima pasa a considerarse la solución óptima del problema estocástico.

Fundamentalmente existen dos tipos de modelos en Programación Estocástica:

- Modelos “esperar y ver” (“wait and see”) o modelos de programación estocástica pasiva, basados en la suposición de que el decisor es capaz de esperar a que se produzca la realización de las variables aleatorias y hacer su decisión con información completa de dicha realización, con lo que el problema se convierte en determinístico y es posible encontrar el valor óptimo de las variables de decisión con las técnicas habituales de programación matemática determinística. En ocasiones puede tener interés el conocer la distribución de probabilidad del valor objetivo óptimo o algunos de sus momentos (valor esperado o varianza) antes de conocer la realización de sus variables aleatorias. Tales problemas se llaman **problemas de distribución**. Estos problemas se estudian en [4], [33], [29].
- Modelos “aquí y ahora” (“here and now”) o modelos de programación estocástica activa. En estos modelos el decisor toma la decisión sin el conocimiento de la realización de las variables aleatorias, sin que por ello queden afectadas las distribuciones de probabilidad de las mismas. En los siguientes apartados veremos diferentes enfoques para resolver el problema.

3 Programación con restricciones probabilísticas

Se considera el problema (1.1) en el que se supone que la función objetivo no contiene ninguna variable aleatoria:

$$\begin{aligned} & \min_x \quad g_0(x), \\ & \text{sujeto a : } \tilde{g}_i(x, \tilde{\xi}) \leq 0, \quad i = 1, 2, \dots, m, \\ & \quad x \in D, \end{aligned} \tag{1.3}$$

El método de restricciones de azar (chance constrained) fue introducido por Charnes, Cooper y Symonds en 1958. Véanse [7], [8]. La idea consiste en transformar el problema dado en un determinista equivalente en el que se verifiquen las restricciones con, al menos, una determinada probabilidad fijada de antemano. Hay que distinguir dos casos según se fije la probabilidad para el conjunto de las restricciones o para cada una de ellas por separado.

Restricciones de azar conjuntas:

Se considera el problema (3.1). Sea $p \in [0, 1]$ dado. Se define el determinista equivalente:

$$\begin{aligned} & \min_x \quad g_0(x), \\ & \text{sujeto a : } P\left(\tilde{g}_1(x, \tilde{\xi}) \leq 0, \tilde{g}_2(x, \tilde{\xi}) \leq 0, \dots, \tilde{g}_m(x, \tilde{\xi}) \leq 0\right) \geq p, \\ & \quad x \in D. \end{aligned} \tag{1.4}$$

Para este problema, $1 - p$ es el riesgo admisible para el decisor de que la solución del problema sea no factible.

En el caso particular de que para cada $x \in D$ las variables aleatorias

$$\tilde{g}_1(x, \tilde{\xi}), \tilde{g}_2(x, \tilde{\xi}), \dots, \tilde{g}_m(x, \tilde{\xi})$$

sean mutuamente estadísticamente independientes, el problema equivalente determinista anterior se puede expresar de la siguiente forma.

$$\begin{aligned} & \min_x \quad g_0(x) \\ & \text{sujeto a : } P\left(\tilde{g}_1(x, \tilde{\xi}) \leq 0\right) P\left(\tilde{g}_2(x, \tilde{\xi}) \leq 0\right) \dots P\left(\tilde{g}_m(x, \tilde{\xi}) \leq 0\right) \geq p, \\ & \quad x \in D. \end{aligned} \tag{1.5}$$

Restricciones de azar separadas o individuales:

Se considera el problema (1.3). Para cada restricción $i \in \{1, 2, \dots, m\}$ sea $p_i \in [0, 1]$ dado. Se define el determinista equivalente:

$$\begin{aligned} & \min_x \quad g_0(x), \\ & \text{sujeto a : } P\left(\tilde{g}_i(x, \tilde{\xi}) \leq 0\right) \geq p_i, \quad \text{para } i = 1, 2, \dots, m, \\ & \quad x \in D. \end{aligned} \tag{1.6}$$

La siguiente Proposición recoge la relación entre los dos casos:

Proposición 3.1. *Supongamos que \hat{x} es una solución factible del problema (1.6) para los valores p_1, p_2, \dots, p_m . Entonces para $p = 1 - m + \sum_{i=1}^m p_i$, se verifica que \hat{x} es factible para el problema (1.4).*

Demuestra: Sea \hat{x} solución factible del problema (1.4). Ello quiere decir que se verifica: $P(\xi | g_i(\hat{x}, \xi) \leq 0) \geq p_i$, para todo $i = 1, 2, \dots, m$. Definimos los eventos A_i de la siguiente forma: $A_i = \{\xi | g_i(\hat{x}, \xi) \leq 0\}$, para $i = 1, 2, \dots, m$.

Se verifica que $P(A_i) \geq p_i$, $P(A_i^C) \leq 1 - p_i$. Veamos que se verifica que

$$P\left(\bigcap_{i=1}^m A_i\right) \geq p,$$

lo cual quiere decir que \hat{x} es factible para el problema (1.4). En efecto:

Teniendo en cuenta la desigualdad de Boole: $P\left(\bigcup_k S_k\right) \leq \sum_k P(S_k)$, se tiene que

$$\begin{aligned} P\left(\bigcap_{i=1}^m A_i\right) &= 1 - P\left(\left(\bigcap_{i=1}^m A_i\right)^C\right) = 1 - P\left(\bigcup_{i=1}^m (A_i)^C\right) \geq \\ &\geq 1 - \sum_{i=1}^m P((A_i)^C) \geq 1 - \sum_{i=1}^m (1 - p_i) = p. \end{aligned}$$

■

Sean:

$$\begin{aligned} q(x) &= P(\xi | g_1(x, \xi) \leq 0, g_2(x, \xi) \leq 0, \dots, g_m(x, \xi) \leq 0), \\ q_i(x) &= P(\xi | g_i(x, \xi) \leq 0), \quad i = 1, 2, \dots, m. \end{aligned}$$

El conjunto factible del problema (1.4) lo podemos representar de la siguiente forma: $C(p) = \{x \in D | q(x) \geq p\}$.

Sea: $C_i(p_i) = \{x \in D | q_i(x) \geq p_i\}$, $i \in \{1, 2, \dots, m\}$.

El conjunto factible del Problema (1.6) lo podemos representar como

$$\hat{C}(p_1, p_2, \dots, p_m) = \bigcap_{i=1}^m C_i(p_i).$$

Sería deseable que los conjuntos $C(p)$ y $\hat{C}(p_1, p_2, \dots, p_m)$, que son los conjuntos de soluciones factibles de los deterministas equivalentes que estamos estudiando, fueran no vacíos, cerrados y convexos. Las siguientes proposiciones tratan sobre dichas cuestiones.

Proposición 3.2 Sea $C(p)$ el conjunto de soluciones factibles del Problema (1.4). En dicho conjunto se verifican las siguientes propiedades:

- 1) Si $p^1 \leq p^2$, entonces $C(p^1) \supset C(p^2)$.
- 2) $C(0) = D$.
- 3) $C(p)$ es no vacío para todo $p \in [0, 1] \iff C(1) \neq \emptyset$.

Demostración:

- 1) Sea $p^1 \leq p^2$. Si $x \in C(p^2)$, es $q(x) = P(\xi | g_1(x, \xi) \leq 0, g_2(x, \xi) \leq 0, \dots, g_m(x, \xi) \leq 0) \geq p^2 \geq p^1 \implies x \in C(p^1)$.
- 2) $C(0) = \{x \in D | q(x) \geq 0\} = D$, ya que $q(x)$ es una probabilidad y por tanto es mayor o igual que cero.
- 3) Si $C(p) \neq \emptyset, \forall p \in [0, 1] \implies C(1) \neq \emptyset$. Por otra parte, si $C(1) \neq \emptyset \implies \forall p \leq 1$, por 1) es $C(p) \supset C(1) \neq \emptyset$. ■

Obsérvese que si $C(p) \neq \emptyset, \forall p \in [0, 1]$, entonces $\hat{C}(p_1, p_2, \dots, p_m) \neq \emptyset$, para todo p_1, p_2, \dots, p_m en $[0, 1]$.

La siguiente proposición, cuya demostración se encuentra en [17, ?] da condiciones que aseguran que los conjuntos que estamos considerando son cerrados.

Proposición 3.3 Si las funciones $g_i : R^n \times E \rightarrow R$ son continuas, entonces los conjuntos factibles $C(p)$ y $\hat{C}(p_1, p_2, \dots, p_m)$ son cerrados.

A continuación se aborda el problema de la convexidad de los conjuntos $C(p)$ y $\hat{C}(p_1, p_2, \dots, p_m)$. Estos conjuntos en general no son convexos. Veamos condiciones en que sí lo son. Las demostraciones de las proposiciones siguientes se encuentran en [17, ?]. Véase también [29].

Definición 3.1 Una medida de probabilidad $P : F \rightarrow [0, 1]$ se dice que es cuasicónica si $\forall S_1, S_2 \in F$, siendo S_1 y S_2 conjuntos convexos, y $\forall \lambda \in [0, 1]$, se verifica que $P(\lambda S_1 + (1 - \lambda) S_2) \geq \min\{P(S_1), P(S_2)\}$.

Definición 3.2 Una medida de probabilidad $P : F \rightarrow [0, 1]$ se dice que es log-cónica si $\forall S_1, S_2 \in F$, siendo S_1 y S_2 conjuntos convexos, y $\forall \lambda \in [0, 1]$, se verifica que $P(\lambda S_1 + (1 - \lambda) S_2) \geq [P(S_1)]^\lambda [P(S_2)]^{1-\lambda}$.

Las dos proposiciones siguientes dan condiciones para que una medida de probabilidad sea cuasi-cónica.

Proposición 3.3 Si P es una medida de probabilidad log-cónica en F , entonces P es cuasicónica.

Proposición 3.4 Sea P una medida de probabilidad en R^s , de tipo continuo con función de densidad asociada f . Entonces se verifica:

- P es log-cóncava si y sólo si el logaritmo de f es una función cóncava.
- P es cuasi-cóncava si y sólo si $f^{-1/s}$ es convexa.

La siguiente proposición da condiciones suficientes para que los conjuntos que estamos estudiando sean convexos.

Proposición 3.5 Si $g_i(\cdot, \cdot)$ es conjuntamente convexa en (x, ξ) , para cada $i = 1, 2, \dots, m$ y P es cuasi-cóncava, entonces $C(p)$ es convexo para todo $p \in [0, 1]$ y $\hat{C}(p_1, p_2, \dots, p_m)$ es convexo, $\forall p_1, p_2, \dots, p_m$ en $[0, 1]$.

Algunas medidas de probabilidad cuasi-cóncavas son: La uniforme k -dimensional, sobre un conjunto convexo $S \subset \mathbb{R}^k$, la distribución exponencial en \mathbb{R} , la normal multivariante en \mathbb{R}^k , la distribución de Dirichlet, la beta, la distribución de Wishart, la gamma para ciertos valores del parámetro, la distribución de Cauchy, la distribución de Pareto para determinados valores etc.

El caso lineal:

Se considera el problema lineal estocástico (1.2), en el cual se supone que la función objetivo no contiene ninguna variable aleatoria:

$$\begin{array}{ll} \min_x & c^T x, \\ \text{sujeto a :} & Ax = b, \\ & T(\tilde{\xi})x \geq h(\tilde{\xi}), \\ & x \geq 0, \end{array} \quad (1.7)$$

Para el Problema (1.7), dado el valor $p \in [0, 1]$, el programa determinista equivalente correspondiente al método de restricciones de azar tomadas en conjunto será:

$$\begin{array}{ll} \min_x & c^T x, \\ \text{sujeto a :} & Ax = b, \\ & P(T(\tilde{\xi})x \geq h(\tilde{\xi})) \geq p, \\ & x \geq 0, \end{array} \quad (1.8)$$

Para el mismo Problema (1.7), dados los valores p_1, p_2, \dots, p_m , pertenecientes al intervalo $[0, 1]$, el programa determinista equivalente correspondiente al método de restricciones de azar tomadas de manera separada será:

$$\begin{array}{ll} \min_x & c^T x, \\ \text{sujeto a :} & Ax = b, \\ & P\left(T_i(\tilde{\xi})x \geq h_i(\tilde{\xi})\right) \geq p_i, \quad i = 1, 2, \dots, m, \\ & x \geq 0, \end{array} \quad (1.9)$$

Sean $C(p)$ el conjunto factible del programa (1.8) y $\hat{C}(p_1, p_2, \dots, p_m)$ el conjunto factible de (1.9). Aunque el programa estocástico inicial (1.7) es lineal, los conjuntos de soluciones factibles $C(p)$, $\hat{C}(p_1, p_2, \dots, p_m)$ no tienen por qué ser convexos, como se puede observar en el siguiente ejemplo.

Se considera el siguiente programa estocástico con una sola variable de decisión x :

$$\begin{aligned} & \min_x g_0(x), \\ & \text{sujeto a : } Tx \geq h(\tilde{\xi}), \end{aligned}$$

en donde $T = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$,

$h(\tilde{\xi})$ toma los valores:

$$\begin{pmatrix} -4 \\ 0 \end{pmatrix}, \quad \text{con probabilidad } 1/2, \text{ y } \begin{pmatrix} -10 \\ 3 \end{pmatrix} \quad \text{con probabilidad } 1/2.$$

Para este programa estocástico se tiene que, para todo $p \in [0, 1/2]$ es $C(p) = \hat{C}(p) = [0, 2] \cup [3, 5]$, que no es convexo ni conexo.

Las siguientes proposiciones recogen los principales resultados conocidos para el tipo de problema que estamos considerando.

Proposición 3.6 *Se considera el programa estocástico (1.7). Supongamos que $\tilde{\xi}$ es un vector aleatorio cuya distribución de probabilidad es discreta y finita. Sea $P(\xi = \xi^k) = \alpha_k$, para $k = 1, 2, \dots, K$. Entonces para $p > 1 - \min_{k \in \{1, 2, \dots, K\}} \alpha_k$ se verifica que el conjunto factible $C(p)$ es convexo.*

La demostración se encuentra en [17, ?]

A la vista de la proposición anterior, se comprueba inmediatamente que si $p_j > 1 - \min_{k \in \{1, 2, \dots, K\}} \alpha_k$ para cada $j = 1, 2, \dots, m$, el conjunto $\hat{C}(p_1, p_2, \dots, p_m)$ es convexo.

Proposición 3.7 *Se considera el programa estocástico (1.7). Supongamos que $T(\tilde{\xi}) = T$ y que la probabilidad P correspondiente a $h(\tilde{\xi}) = \tilde{h}$ es cuasi-cónica. Entonces los conjuntos $C(p)$ y $\hat{C}(p_1, p_2, \dots, p_m)$ son cerrados y convexos.*

La demostración se puede ver en [5]

Proposición 3.8 *Se considera el programa estocástico (1.7). Sean $\tilde{T}_{1\cdot}, \tilde{T}_{2\cdot}, \dots, \tilde{T}_{m\cdot}$ las filas respectivas de la matriz $T(\tilde{\xi})$, $h(\tilde{\xi}) = \tilde{h}$. Supongamos que $\tilde{T}_{1\cdot}, \tilde{T}_{2\cdot}, \dots, \tilde{T}_{m\cdot}$, \tilde{h} tienen distribución normal con*

$$E \left[\left(\tilde{T}_{i\cdot} - E(\tilde{T}_{i\cdot}) \right) \left(\tilde{T}_{j\cdot} - E(\tilde{T}_{j\cdot})^T \right) \right] = r_{ij} C, \quad \text{para } i, j = 1, 2, \dots, m,$$

$$E \left[\left(\tilde{T}_{i\cdot} - E(\tilde{T}_{i\cdot}) \right) \left(\tilde{h} - E(\tilde{h}) \right) \right] = s_i C, \text{ para } i, j = 1, 2, \dots, m,$$

donde r_{ij} y s_i son constantes para todo i, j . Entonces, $C(p)$ es convexo para $p \geq 0, 5$.

La demostración se puede ver en [5]

Ejemplos:

- 1) Se considera el programa estocástico con conjunto factible

$$g(x) \geq \tilde{\xi}, \quad (1.10)$$

en donde $x \in \mathbb{R}^n$, $g(x) = (g_1(x), g_2(x), \dots, g_m(x))^T$ no contiene ningún elemneto aleatorio y $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2, \dots, \tilde{\xi}_m)$ es un vector aleatorio de dimensión m .

En este caso para $p \in [0, 1]$ se tiene que el conjunto factible del determinista equivalente para restricciones de azar conjuntas es

$$\begin{aligned} C(p) &= \{x \in \mathbb{R}^n \mid P(\xi \mid g(x) \geq \tilde{\xi}) \geq p\} = \\ &= \left\{ x \in \mathbb{R}^n \mid F_{\tilde{\xi}}(g(x)) \geq p \right\}, \end{aligned}$$

en donde $F_{\tilde{\xi}}$ es la función de distribución del vectora aleatorio $\tilde{\xi}$.

Para $p_i \in [0, 1]$, considerando restricciones de azar individuales se tiene que

$$\begin{aligned} C_i(p_i) &= \{x \in \mathbb{R}^n \mid P(\xi_i \mid g_i(x) \geq \xi_i) \geq p_i\} = \\ &= \left\{ x \in \mathbb{R}^n \mid F_{\xi_i}(g_i(x)) \geq p_i \right\} = \{x \in \mathbb{R}^n \mid g_i(x) \geq \gamma_i\}, \end{aligned}$$

en donde $\gamma_i = F_{\xi_i}^{-1}(p_i)$.

- 2) Se considera el programa estocástico lineal (1.7) y su determinista equivalente (1.9) para restricciones de azar separadas. Sea la restriccción estocástica $T_i(\tilde{\xi})x \geq h_i(\tilde{\xi})$ de la forma $\tilde{t}^T x \geq \tilde{h}$, siendo $(\tilde{t}^T, \tilde{h})^T$ un vector aleatorio con distribución conjunta normal de media $\mu \in \mathbb{R}^{n+1}$, y matriz de varianzas y covarianzas V , de dimensión $(n+1) \times (n+1)$. Calculemos su correspondiente restriccción en el determinista equivalente (para restricciones de azar separadas). $P((t^T, h)^T \mid t^T x \geq h) = P((t^T, h)^T \mid x^T t - h \geq 0) = P(\eta \mid \eta(x) \geq 0)$, en donde $\tilde{\eta}(x) = x^T \tilde{t} - \tilde{h}$. La variable aleatoria $\tilde{\eta}$ es normal (unidimensional), por ser combinación lineal de variables conjuntamente normales. Su media es $m_{\tilde{\eta}}(x) = \sum_{j=1}^n \mu_j x_j - \mu_{n+1}$, y su varianza es $\sigma_{\tilde{\eta}}^2(x) = z(x)^T V z(x)$, donde $z(x) =$

$$(x_1, x_2, \dots, x_n, -1)^T.$$

$$\begin{aligned} P(\tilde{\eta}(x) \geq 0) \geq p_i &\iff P\left(\frac{\tilde{\eta}(x) - m_{\tilde{\eta}}(x)}{\sigma_{\tilde{\eta}}(x)} \geq \frac{-m_{\tilde{\eta}}(x)}{\sigma_{\tilde{\eta}}(x)}\right) \geq p_i. \iff \\ 1 - P\left(\frac{\tilde{\eta}(x) - m_{\tilde{\eta}}(x)}{\sigma_{\tilde{\eta}}(x)} < \frac{-m_{\tilde{\eta}}(x)}{\sigma_{\tilde{\eta}}(x)}\right) &\geq p_i \iff 1 - \Phi\left(\frac{-m_{\tilde{\eta}}(x)}{\sigma_{\tilde{\eta}}(x)}\right) \geq p_i, \end{aligned}$$

donde Φ es la función de distribución de la normal de media cero y varianza 1.

Por tanto, la restricción de azar correspondiente queda como:

$$\begin{aligned} 1 - \Phi\left(\frac{-m_{\tilde{\eta}}(x)}{\sigma_{\tilde{\eta}}(x)}\right) \geq p_i &\iff \Phi\left(\frac{-m_{\tilde{\eta}}(x)}{\sigma_{\tilde{\eta}}(x)}\right) \leq 1 - p_i \iff \\ \iff \frac{-m_{\tilde{\eta}}(x)}{\sigma_{\tilde{\eta}}(x)} &\leq \Phi^{-1}(1 - p_i) \iff -\Phi^{-1}(1 - p_i)\sigma_{\tilde{\eta}}(x) - m_{\tilde{\eta}}(x) \leq 0. \end{aligned}$$

El conjunto de los $x \in R^n$ que verifican esa condición es convexo si y sólo si $\Phi^{-1}(1 - p_i) \leq 0$, lo cual se verifica si y sólo si $p_i \geq 0,5$.

Pueden encontrarse más ejemplos en [14], [27], [28], [30], [33], [35]. En [26] se presenta una aplicación muy interesante.

4 Función objetivo aleatoria

Consideremos el siguiente problema estocástico, en el que todas las restricciones son determinísticas y la función objetivo es aleatoria.

$$\begin{array}{ll} \min_x & \tilde{g}_0(x, \tilde{\xi}), \\ \text{sujeto a :} & x \in X \end{array} \quad (1.11)$$

El conjunto factible $X \subset R^n$ está compuesto por restricciones determinísticas, bien porque lo sean de manera natural, bien porque se haya obtenido el determinista equivalente utilizando el método de restricciones de azar.

Se trata de transformar el objetivo estocástico en su determinista equivalente. Ello puede hacerse utilizando distintos criterios, que vamos a ver a continuación, siguiendo el enfoque de los trabajos [6] y [27].

4.1 Algunos conceptos de solución

a) Criterio del valor esperado.

Se convierte la variable aleatoria $\tilde{g}_0(x, \tilde{\xi})$ en una función determinística tomando la esperanza matemática

$$E[\tilde{g}_0(x, \tilde{\xi})].$$

El determinista equivalente del problema estocástico (1.11) será

$$\begin{aligned} \min_x & \quad E[\tilde{g}_0(x, \tilde{\xi})], \\ \text{sujeto a: } & \quad x \in X \end{aligned} \quad (1.12)$$

Para resolver el problema de programación estocástica siguiendo este criterio, basta con conocer el valor esperado de la función objetivo estocástica y , por tanto, es aplicable aún en el caso en el que se desconozca la distribución de probabilidad de la variable aleatoria $\tilde{g}_0(x, \tilde{\xi})$.

En [29] se señala que para que este criterio sea considerado apropiado se deben cumplir dos condiciones:

- 1) El sistema debe repetir su realización de manera independiente un gran número de veces, para asegurar que la media de los resultados sea bastante próxima al valor esperado.
- 2) La magnitud de la variación del resultado no debe ser grande. En otro caso nuestra política óptima puede llevar al sistema a la bancarrota antes de que la deseada media a largo plazo pueda ser alcanzada.

En muchas situaciones prácticas estas condiciones no se cumplen y, por tanto, este criterio no debería ser utilizado en tales casos.

b) Criterio de mínima varianza.

Se convierte la variable aleatoria $\tilde{g}_0(x, \tilde{\xi})$ en una función determinística tomando su varianza: $Var[\tilde{g}_0(x, \tilde{\xi})] = E[(\tilde{g}_0(x, \tilde{\xi}))^2] - \{E[\tilde{g}_0(x, \tilde{\xi})]\}^2$.

La utilización de este criterio da lugar a la elección de aquel vector x para el que la variable aleatoria $\tilde{g}_0(x, \tilde{\xi})$ está más concentrada alrededor de su valor esperado, de manera que el determinista equivalente según el criterio de mínima varianza puede interpretarse como una medida de error cuadrático.

El criterio de optimización es el de mínima varianza tanto si se trata de minimizar la función objetivo (como estamos suponiendo en este trabajo) como si se trata de maximizarlo.

Para poder utilizar este criterio es suficiente con que se conozca la varianza de la variable aleatoria $\tilde{g}_0(x, \tilde{\xi})$. No hace falta que se conozca su distribución de probabilidad.

El determinista equivalente del problema estocástico (1.11), según el criterio de mínima varianza será

$$\begin{aligned} \min_x & \quad Var[\tilde{g}_0(x, \tilde{\xi})] \\ \text{sujeto a: } & \quad x \in X \end{aligned} \quad (1.13)$$

c) Criterio de eficiencia valor esperado desviación estándar.

Este concepto de eficiencia fue introducido por Markowitz en 1952 para resolver problemas de selección de carteras en el campo de las finanzas. Véase [22] y también [23] y [24].

Se trata de elegir una política x^0 que sea eficiente en el sentido de Markowitz. Expliquemos su significado:

$$\text{Sean } \mu(x) = E[\tilde{g}_0(x, \tilde{\xi})], \quad \sigma^2(x) = \text{Var}[\tilde{g}_0(x, \tilde{\xi})].$$

Se tiene que verificar que no existe ningún $x \in X$ para el cual se tenga que $\mu(x) = \mu(x^0)$ y $\sigma(x) < \sigma(x^0)$, o bien $\sigma(x) = \sigma(x^0)$ y $\mu(x) < \mu(x^0)$.

El conjunto de puntos eficientes normalmente tiene infinitos elementos. Por tanto, normalmente este criterio no especifica un único punto como solución óptima. Si se quiere llegar a “una” solución óptima habrá que añadir otras consideraciones al conjunto obtenido de puntos eficientes.

El cálculo de soluciones eficientes valor esperado desviación estándar se traduce en el cálculo de soluciones eficientes del siguiente problema biobjetivo determinista equivalente:

$$\begin{aligned} \min_x & \quad (E[\tilde{g}_0(x, \tilde{\xi})], \text{Var}[\tilde{g}_0(x, \tilde{\xi})]), \\ \text{sujeto a: } & \quad x \in X \end{aligned} \tag{1.14}$$

d) Criterio de mínimo riesgo.

Este criterio fue introducido por Bereanu [3] con el nombre de criterio de mínimo riesgo y por Charnes y Cooper [9] con el nombre de P-modelo.

Se trata de maximizar la probabilidad de que la función objetivo sea menor o igual que cierto valor previamente establecido. Por tanto, para resolver el problema hay que fijar un nivel para la función objetivo estocástica, $\lambda \in R$, al que se denomina nivel de aspiración, y maximizar la probabilidad de que el objetivo sea menor o igual que ese nivel: $P\{\tilde{g}_0(x, \tilde{\xi}) \leq \lambda\}$.

La idea del nivel de aspiración es que “como mucho el valor objetivo sea λ ”.

El determinista equivalente del problema estocástico (1.11), según el criterio de mínimo riesgo será

$$\begin{aligned} \max_x & \quad P\{\tilde{g}_0(x, \tilde{\xi}) \leq \lambda\}, \\ \text{sujeto a: } & \quad x \in X \end{aligned} \tag{1.15}$$

Teniendo en cuenta que $\max_x P\{\tilde{g}_0(x, \tilde{\xi}) \leq \lambda\} = \max_x \{1 - P\{\tilde{g}_0(x, \tilde{\xi}) > \lambda\}\} = 1 - \min_x P\{\tilde{g}_0(x, \tilde{\xi}) > \lambda\}$, el problema (1.15) es equivalente a

$$\begin{aligned} \min_x & \quad P\{\tilde{g}_0(x, \tilde{\xi}) > \lambda\}, \\ \text{sujeto a: } & \quad x \in X, \end{aligned}$$

y el problema puede interpretarse como la minimización del riesgo de que la función objetivo sobrepase el nivel de aspiración λ .

Si el problema a resolver consistiera en maximizar la función objetivo (en lugar de minimizar como estamos considerando), es decir, si el problema original fuera

$$\begin{aligned} & \max_x \quad \tilde{g}_0(x, \tilde{\xi}), \\ & \text{sujeto a : } x \in X \end{aligned}$$

el problema de mínimo riesgo determinista equivalente sería

$$\begin{aligned} & \max_x \quad P \left\{ \tilde{g}_0(x, \tilde{\xi}) \geq \lambda \right\}, \\ & \text{sujeto a : } x \in X \end{aligned}$$

En este caso la idea del nivel de aspiración es que “el valor objetivo al menos sea λ ”.

e) Criterio de Kataoka o criterio β -fractil

El criterio fue introducido por Kataoka [20].

Se comienza fijando por el decisor una probabilidad $\beta \in (0, 1)$ para la función objetivo y se determina el menor nivel que puede alcanzar la función objetivo con esa probabilidad. En concreto, el determinista equivalente del problema estocástico (1.11), según el criterio de Kataoka¹ será:

$$\begin{aligned} & \min_{(x^T, \lambda)} \quad \lambda \\ & \text{sujeto a : } P \left\{ \tilde{g}_0 \left(x, \tilde{\xi} \right) \leq \lambda \right\} = \beta, \\ & \quad x \in X \end{aligned}$$

Si el problema a resolver consistiera en maximizar la función objetivo (en lugar de minimizar como estamos considerando), es decir, si el problema original fuera

$$\begin{aligned} & \min_x \quad \tilde{g}_0 \left(x, \tilde{\xi} \right), \\ & \text{sujeto a : } x \in X \end{aligned}$$

el problema de Kataoka determinista equivalente sería

$$\begin{aligned} & \min_{(x^T, \lambda)} \quad \lambda \\ & \text{sujeto a : } P \left\{ \tilde{g}_0 \left(x, \tilde{\xi} \right) \geq \lambda \right\} = \beta, \\ & \quad x \in X \end{aligned}$$

¹En trabajos posteriores al de Kataoka otros autores como Stancu-Minasian [33] plantean el problema con restricción probabilística de desigualdad: $P \left\{ \tilde{g}_0 \left(x, \tilde{\xi} \right) \leq \lambda \right\} \geq \beta$. Se demuestra que si la variable aleatoria $\tilde{g}_0 \left(x, \tilde{\xi} \right)$ es continua el resultado del problema es el mismo en ambos casos.

Comparando los tres primeros criterios con los dos últimos (que se llaman de máxima probabilidad) aparecen algunas diferencias:

- En el criterio de mínimo riesgo se fija el nivel de aspiración y en el criterio de Kataoka se fija la probabilidad, luego ambos dependen de los valores que se asignen a estos parámetros, mientras que en los tres primeros casos no hay que fijar ningún parámetro.
- En los criterios valor esperado, mínima varianza y eficiencia valor esperado desviación estándar sólo necesita conocerse la esperanza t/o la varianza , no haciendo falta la distribución de probabilidad.

La elección de un criterio u otro deberá realizarse en base a las características del problema y a las preferencias del decisor. De todas formas, los cinco criterios están relacionados entre sí, dado que cada uno de ellos utiliza diferentes características de la función objetivo.

4.2 Relaciones entre las soluciones según los distintos criterios

En [6] se obtienen algunos resultados para problemas estocásticos como (1.11) que cumplen algunas condiciones adicionales. Veamos algunos de dichos resultados.

Consideremos el problema estocástico (1.11) en el que suponemos ahora que el conjunto de soluciones factibles $X \subset \mathbb{R}^n$ es no vacío, cerrado, acotado y convexo. Suponemos también que ξ es un vector aleatorio definido sobre un conjunto $E \subset \mathbb{R}^s$ cuyas componentes son variables aleatorias continuas y cuya distribución de probabilidad es independiente de las variables de decisión del problema x_1, x_2, \dots, x_n .

Las demostraciones de todas las proposiciones que presentamos a continuación se encuentran en [6].

Proposición 4.1 *Se considera el problema estocástico (1.11) con las hipótesis adicionales introducidas en este subapartado.*

a) *Si la solución óptima del problema según el criterio del valor esperado es única, entonces es una solución eficiente valor esperado desviación estándar. Si no es única sólo se puede asegurar que las soluciones óptimas valor esperado son soluciones débilmente eficientes valor esperado desviación estándar, pero no tienen por qué ser eficientes valor esperado desviación estándar.*

b) *Si la varianza de la función objetivo es una función estrictamente convexa, el problema de varianza mínima tiene solución única que es una solución eficiente valor esperado desviación estándar. Si no es única sólo se puede asegurar que las*

soluciones óptimas de mínima varianza son soluciones débilmente eficientes valor esperado desviación estándar, pero no tienen por qué ser eficientes valor esperado desviación estándar.

La siguiente proposición establece relación entre las soluciones óptimas según los criterios de mínimo riesgo y de Kataoka.

Proposición 4.2 *Se considera el problema estocástico (1.11) con las hipótesis adicionales introducidas en este subapartado. Supongamos que la función de distribución de la variable aleatoria $\tilde{g}_0(x, \xi)$ es estrictamente creciente. Entonces x^* es la solución de mínimo riesgo para el nivel de aspiración λ^* si y sólo si $(x^{*T}, \lambda^*)^T$ es la solución de Kataoka con probabilidad β^* , con λ^* y β^* verificando $P(\tilde{g}_0(x, \xi) \leq \lambda^*) = \beta^*$.*

A la vista de la proposición anterior se puede asegurar que en las condiciones que estamos considerando en este subapartado:

- Para cada nivel de aspiración λ , la solución de mínimo riesgo es también la solución de Kataoka con una probabilidad β igual a la máxima probabilidad obtenida en el problema de mínimo riesgo.
- Para cada valor β fijado, la solución de Kataoka es también solución de mínimo riesgo para un nivel de aspiración igual al valor óptimo del problema de Kataoka.

En [6] se establecen también relaciones entre soluciones de Kataoka y soluciones eficientes valor esperado desviación estándar para algunos tipos de programas estocásticos lineales.

4.3 Ejemplo

Como ejemplo vamos a considerar el caso de función objetivo lineal con distribución de probabilidad normal.

Sea el programa estocástico lineal

$$\begin{array}{ll} \min & \tilde{\xi}^T x, \\ x & \text{sujeto a : } x \in X \end{array} \quad (1.16)$$

El conjunto factible $X \subset R^n$ está compuesto por restricciones determinísticas, bien porque lo sean de manera natural, bien porque se haya obtenido el determinista equivalente utilizando el método de restricciones de azar. Se supone que el vector aleatorio $\tilde{\xi}$ sigue una distribución de probabilidad normal multivariante, con valor esperado ξ y matriz de varianzas y covarianzas S definida positiva.

En estas condiciones la variable aleatoria $\tilde{\xi}^T x$ es normal con valor esperado $\bar{\xi}^T x$ y varianza $x^T S x$. Por tanto se tiene que $\frac{\tilde{\xi}^T x - \bar{\xi}^T x}{\sqrt{x^T S x}}$ es una variable aleatoria $N(0, 1)$ (normal con valor esperado 0 y desviación típica 1).

A continuación se calcula el determinista equivalente del programa estocástico (1.16) para cada uno de los criterios considerados en este apartado.

a) Criterio del valor esperado

$$\begin{aligned} & \min_x \quad \tilde{\xi}^T x, \\ & \text{sujeto a: } x \in X \end{aligned}$$

b) Criterio de mínima varianza

$$\begin{aligned} & \min_x \quad x^T S x, \\ & \text{sujeto a: } x \in X \end{aligned}$$

c) Criterio de eficiencia valor esperado desviación estándar

$$\begin{aligned} & \min_x \quad \left(\tilde{\xi}^T x, \sqrt{x^T S x} \right), \\ & \text{sujeto a: } x \in X \end{aligned}$$

d) Criterio de mínimo riesgo de nivel λ

$$\begin{aligned} & \max_x \quad P \left\{ \tilde{\xi}^T x \leq \lambda \right\}, \\ & \text{sujeto a: } x \in X \end{aligned} \tag{1.17}$$

Pero

$$P \left\{ \tilde{\xi}^T x \leq \lambda \right\} = P \left(\frac{\tilde{\xi}^T x - \bar{\xi}^T x}{\sqrt{x^T S x}} \leq \frac{\lambda - \bar{\xi}^T x}{\sqrt{x^T S x}} \right) = \Phi \left(\frac{\lambda - \bar{\xi}^T x}{\sqrt{x^T S x}} \right), \tag{1.18}$$

donde Φ es la función de distribución de la $N(0, 1)$, que es estrictamente creciente, por lo que $\max_x P \left\{ \tilde{\xi}^T x \leq \lambda \right\} = \max_x \Phi \left(\frac{\lambda - \bar{\xi}^T x}{\sqrt{x^T S x}} \right) = \Phi \left(\max_x \frac{\lambda - \bar{\xi}^T x}{\sqrt{x^T S x}} \right)$, y el problema (1.17) es equivalente a

$$\begin{aligned} & \max_x \quad \frac{\lambda - \bar{\xi}^T x}{\sqrt{x^T S x}} \\ & \text{sujeto a: } x \in X. \end{aligned} \tag{1.19}$$

Una vez resuelto este problema, la probabilidad máxima para la que se puede asegurar que la función objetivo estocástica es menor o igual que el nivel de aspiración fijado λ , es: $\Phi \left(\max_x \frac{\lambda - \bar{\xi}^T x}{\sqrt{x^T S x}} \right)$.

e) **Criterio de Kataoka o criterio β -fractil**

$$\begin{array}{ll} \min_{(x^T, \lambda)} & \lambda \\ \text{sujeto a :} & P \left\{ \xi^T x \leq \lambda \right\} = \beta, \\ & x \in X \end{array}$$

Teniendo en cuenta (1.18), y que la función de distribución Φ es estrictamente creciente, se tiene que $\Phi \left(\frac{\lambda - \xi^T x}{\sqrt{x^T S x}} \right) = \beta \iff \frac{\lambda - \xi^T x}{\sqrt{x^T S x}} = \Phi^{-1}(\beta) \iff \lambda = \Phi^{-1}(\beta) \sqrt{x^T S x} + \xi^T x$, por lo que el problema (4.11) se puede expresar:

$$\begin{array}{ll} \min_{(x^T, \lambda)} & \lambda \\ \text{sujeto a :} & \lambda = \Phi^{-1}(\beta) \sqrt{x^T S x} + \xi^T x, \\ & x \in X \end{array}$$

y este problema es equivalente al problema con n variables de decisión:

$$\begin{array}{ll} \min_x & \Phi^{-1}(\beta) \sqrt{x^T S x} + \xi^T x, \\ \text{sujeto a :} & x \in X. \end{array}$$

Este problema es convexo para $\beta \geq 0,5$. Una vez resuelto este problema, el menor nivel λ para el que podemos afirmar que la función objetivo no supera ese nivel con probabilidad β es $\lambda = \min_x \Phi^{-1}(\beta) \sqrt{x^T S x} + \xi^T x$.

5 Bibliografía

- [1] Álvarez, F., Cerdá, E. "A Solution Method for a Class of Learning by Doing Models with Multiplicative Uncertainty". *Top*, 7, 1, 1-23, 1999.
- [2] Beale, E.M.L., "On Minimizing a Convex Function Subject to Linear Inequalities". *Journal of the Royal Statistical Society, B* 17, 173-184, 1955.
- [3] Bereanu, B. "Programme de Risque Minimal en Programmation Linéaire Stochastique". *C. R. Acad. Sci. Paris*, 259, 981-983, 1964.
- [4] Bereanu, B. *The Generalized Distribution Problem of Stochastic Linear Programming*. Symposia Mathematica. Academic Press, 1976.
- [5] Birge, J.R., Louveaux, F.V. *Introduction to Stochastic Programming*. Springer, 1997.
- [6] Caballero, R., Cerdá, E., Muñoz, M.M., Rey, L. "Analysis and Comparisons of Some Solution Concepts for Stochastic Programming Problems". *Top*, 10(1), 101-124, 2002.

- [7] Charnes, A., Cooper, W.W. "Chance-Constrained Programming". *Management Science*, 5, 73-79, 1959.
- [8] Charnes, A., Cooper, W.W., Symonds, G.H. "Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil". *Management Science*, 4, 235-263, 1958.
- [9] Charnes, A., Cooper, W.W. "Deterministic Equivalents for Optimizing and Satisfying under Chance Constraints". *Operations Research*, 11, 1, 18-39, 1963.
- [10] Dantzig, G.B. "Linear Programming under Uncertainty". *Management Science*, 1, 197-206, 1955.
- [11] Dantzig, G.B. "Planning under Uncertainty". *Annals of Operations Research*, 85, 1999.
- [12] Dempster, M.A.H. *Stochastic Programming*. Academic Press, 1980.
- [13] Diwelar, U. "Optimization under Uncertainty: An Overview". *SIAG/OPT Views -and-News*, 13,11-8, 2002
- [14] Goicoechea, A., Hansen, D.R., Duckstein, L. *Multiobjective Decision Analysis with Engineering and Business Applications*. John Wiley and Sons, 1982.
- [15] Hammer, P.L. *Stochastic Programming. State Of The Art*, 1998. Annals of Operations Research, Vol 85, 1999.
- [16] Higle, J.L., Sen, S. *Stochastic Decomposition*. Kluwer Academic Publishers, 1996.
- [17] Kall, P. *Stochastic Linear Programming*. Springer-Verlag, 1976.
- [18] Kall, P. "Stochastic Programming". *European Journal of Operational Research*, 10, 125-130, 1982.
- [19] Kall, P., Wallace, S.W. *Stochastic Programming*. John Wiley, 1994.
- [20] Kataoka, S. "A Stochastic Programming Model". *Econometrica*, 31, 1-2, 186-196, 1963.
- [21] Kibzun, A.I., Kan, I.S. *Stochastic Programming Problems with Probability and Quantile Functions*. John Wiley, 1996.
- [22] Markowitz, H. "Portfolio Selection". *The Journal of Finance*, 7, 77-91, 1952.
- [23] Markowitz, H. *Portfolio Selection: Efficient Diversification of Investment*. Cowles Commission Monograph 16, John Wiley and Sons, 1959.

- [24] Markowitz, H. *Mean Variance Analysis in Portfolio Choice and Capital Markets*. Blackwell, 1987.
- [25] Marti K. (ed). *Stochastic Optimization. Numerical Methods and Technical Applications*. Lecture Notes in Economics and Mathematical Systems 379. Springer-Verlag, 1992.
- [26] Medova, E. "Chance Constrained Stochastic Programming for Integrated Services Network Management". *Annals of Operations Research*, 81, 213-229, 1998.
- [27] Muñoz M.M. *Programación Estocástica: Algunas Aportaciones Teóricas y Computacionales*. Tesis Doctoral. Universidad Complutense de Madrid, 1998.
- [28] Prawda, J. *Métodos y Modelos de Investigación de Operaciones*. Vol. 2. Modelos Estocásticos. Ed. Limusa, 1980.
- [29] Prekopa, A. *Stochastic Programming*. Kluwer Academic Publishers, 1995.
- [30] Ríos Insua, S. *Investigación Operativa. Optimización*. Editorial Centro de Estudios Ramón Areces, S.A. Segunda Edición, 1993.
- [31] Ruszcynski, A., Shapiro, A. (ed). *Stochastic Programming*. Handbooks in Operations Research and Management Science, Vol 10. North-Holland, 2003.
- [32] Sen, S., Higle, J.L. "An Introductory Tutorial on Stochastic Linear Programming Models". *Interfaces*, 29:2, 33-61, 1999.
- [33] Stancu-Minasian, I.M. *Stochastic Programming with Multiple Objective Functions*. D. Reidel Publishing Company, 1984.
- [34] Stancu-Minasian, I.M., Wets, M.J. "A Research Bibliography in Stochastic Programming 1955-1975". *Operations Research*, 24, 1078-1119, 1976.
- [35] Vajda, S. *Probabilistic Programming*. Academic Press, 1972.
- [36] Wets, R.J.B. "Stochastic Programming ". En NEmhauser, G.L., Rinnooy Kan, A.H.G., Todd, M.J. (ed). *Optimization*. Vol. 1 North-Holland, 1989.

Aproximación metodológica a la optimización en la incertidumbre

Jaime Gil Aluja

Catedrático de la Universitat de Barcelona
e-mail: gilaluja@fuzzyeconomics.com

1 El latir de una sociedad compleja

La sociedad en la que se desenvuelve la actividad económica está soportando, cada vez más, el efecto de cambios profundos para los que no resulta fácil estimar las consecuencias sobre los hombres, las instituciones y las relaciones entre unos y otros. De una manera muy resumiva, acostumbramos a hablar de las nuevas perspectivas utilizando palabras tales como revoluciones, convulsiones,... que dan lugar a comportamientos no lineales que conducen hacia un futuro cargado de incertidumbre.

Conocer, explicar y tratar este nuevo mundo constituye el principal objetivo de quienes deseamos una sociedad mejor al servicio del hombre. Pero para ello, será necesario vencer no pocos obstáculos colocados por quienes son remisos a la apertura de su pensamiento a la recepción del aire fresco que exigen los nuevos tiempos.

Somos conscientes de las dificultades que entraña traspasar el umbral de una nueva era, sobre todo, si se tiene en cuenta que todos cuantos nos hemos formado a través del estudio de la ciencia económica sabemos del gran arraigo de la matemática determinista como elemento capaz de describir las interrelaciones existentes entre los fenómenos económicos. Somos conscientes de la enorme atracción ejercida por los razonamientos sobre ella fundamentados, avalados por la seguridad que otorga tantos siglos de experiencia. Pero pretender traducir un mundo tan cambiante en el lenguaje numérico tradicional resulta hoy, misión casi

imposible. Bien es cierto que, afortunadamente, aún se dan en determinadas ocasiones circunstancias aptas para la utilización de la aritmética de la certeza y del azar, pero también lo es que cada vez resulta más difícil incluso acotar convenientemente los fenómenos, lo que constituye la condición mínima para el empleo del más rudimentario de los números inciertos.

Desde hace unas pocas décadas los estudiosos de la economía y gestión de empresas están intentando canalizar sus inquietudes a través de un buen número de propuestas que, en diferentes sentidos, pretenden dar un nuevo tratamiento tanto a viejos problemas como a los que van surgiendo del complejo entramado de relaciones económico-financieras. Sin embargo en muchas de ellas se constatan, bajo formas diferentes, viejas rutinas que no han conseguido generar horizontes capaces de ofrecer luz a las obscuras profecías de brujos y adivinadores de inquietantes futuros.

Creemos que la principal causa de tanto inmovilismo viene dada por la influencia del “principio del tercio excluso”, que ha constituido norma a la vez que guía para los científicos durante más de 2000 años. Hemos pensado llegado el momento de enunciar un nuevo principio, el de la “simultaneidad gradual”, que puede cobijar un elevado número de razonamientos lógicos capaces de crear conceptos, establecer métodos y elaborar modelos y algoritmos, aptos para proporcionar, por lo menos, algunas de las respuestas esperadas.

Hemos repetido en muchas ocasiones, que el saber científico no debe servir para explicar y tratar el universo que nos gustaría vivir sino el que realmente vivimos¹. Para ello es necesario revisar en profundidad aquellos conocimientos “sagrados” hasta ahora, que describen un mundo estable y crear una nueva estructura del pensamiento capaz de convivir con los desequilibrios y equilibrios inestables que conducen a la incertidumbre, aunque para ello sea necesario reflexionar sobre la idoneidad del comportamiento científico generalmente aceptado.

Una de las ideas recurrentes en la ciencia occidental durante muchos siglos ha sido la idea de las leyes de la naturaleza². Según ella, la naturaleza sigue ciertas reglas estructuradas en torno a la **certeza**. Parece llegado el momento de poner de manifiesto que cada vez resulta más patente la contradicción de esos planteamientos con el aspecto cambiante del universo en general y con los cambios que se producen en el entorno del ser humano en particular. Creemos llegado el momento de impulsar una idea darwiniana de la evolución de los sistemas sociales y económicos, colocando dentro de sus justos límites la idea geométrica. Es necesario saber “explicar” los fenómenos que van apareciendo, día a día, dando razón de los cambios inductores de incertidumbres. Afortunadamente, somos hoy capaces de extraer también de la incertidumbre ciertos comportamientos expresables, la

¹Gil Aluja, J.: “La incertidumbre en la economía y gestión de empresas”. Actas del IV Congreso de la Asociación Española sobre Tecnología y Logica Fuzzy.Blanes, 14 de Septiembre de 1994. Págs: 9-14

²Gil Aluja, J.: “Investment un uncertainty”.Kluwer academica Publishers. Dordrecht 1998. Págs.:19-20

mayor parte de ellos mediante **posibilidades**, algunos a través de **probabilidades** y muy pocos por la **certeza**.

Ha sido, sobre todo, a lo largo de los últimos decenios cuando más importantes han sido las mutaciones, no sólo en cuanto a la fenomenología socio-económica, sino también en los comportamientos y en el pensamiento individual. Ante este contexto, parece lícito preguntarse cómo se puede concebir una actividad científica cuando el pensamiento humano, cargado de un alto grado de subjetividad, intenta encontrar, entre tanto cambio, lo objetivo.

No resulta, pues, honesto cerrar los ojos y decir que se cree en las certezas, aunque éstas no residan en nuestro mundo, sobre todo teniendo en cuenta que la incertidumbre puede dar lugar a una nueva manera de representar las instabilidades y a través de ella comprender mejor el papel que juegan en unas nuevas reglas que explican la naturaleza. Parece abierta, así, una puerta que quizá sea angosta, pero es suficiente para salir del universo determinista que resulta alienante ya que en él todo se halla predeterminado e inscrito en el Big Bang. Al traspasar esta puerta aparecerá un universo incierto que permitirá explicar los nuevos hábitos, las nuevas normas de conducta, los nuevos fenómenos.

La matemática del determinismo ha tenido un gran predicamento y ha imperado y continúa imperado aún hoy en muchos ámbitos de la actividad científica en economía y gestión de empresas. Pero al iniciarse en la sociedad importantes cambios, que tienen cada vez mayor presencia en el campo económico, se levantan ciertas voces clamando por una nueva manera de enfocar los problemas. Se subraya, así, la insuficiencia del mecanicismo para explicar los fenómenos y comportamientos de la nueva sociedad en emergencia.

En el ámbito de la matemática aplicada, se están intentando crear elementos capaces de llegar a un adecuado tratamiento de los fenómenos que tienen lugar en el seno de los estados y de las empresas, cuando su conocimiento se produce de manera tan poco precisa que no somos capaces de representar numéricamente las magnitudes resultantes de la actividad económica. No podemos, hoy, ante la imposibilidad de recoger las complejas e inciertas realidades, recurrir a una simplificación inicial para realizar los desarrollos posteriores en base a estos elementos simplificadores.

2 De los principios a la lógica

La revisión del complejo entramado científico existente, aceptado sin fisuras significativas hasta hace poco, comporta hurgar en los principios mismos que forman la **base** a partir de la cual se construyen los elementos teóricos necesarios para, finalmente, llegar a un adecuado tratamiento de los problemas susceptibles de ser planteados. **A partir de ellos** se levantan cuatro niveles o estratos: lógica, matemática pura, matemática aplicada y estudios de optimización.

En la configuración de la actividad investigadora ³, la ciencia occidental se ha ido construyendo a lo largo de los siglos tomando como uno de los indiscutibles soportes el principio del tercio excluso, a partir del cual, una **arquitectura lógica** va tomando forma hasta consolidarse con los trabajos de George Boole. Un único operador lógico permite el encadenamiento de proposiciones a través de dos modos: hacia adelante con el “modus ponens” y hacia atrás con el “modus tollens”. El paso del razonamiento lógico, expresado en palabras, a su presentación mediante símbolos ha dado lugar al desarrollo de la **matemática en su estado más puro** que, con el tiempo, ha adquirido “formas” diferentes, pero siempre dentro de un mecanismo, tanto en cuanto se ha seguido el camino de la certeza como en cuanto se ha emprendido la senda del azar. La formulación de la matemática binaria resulta esclarecedora de cuanto acabamos de exponer. En base a estos razonamientos lógicos, expresados mediante símbolos, se han elaborado técnicas específicas en forma casi siempre de modelos y algoritmos, cuyo objetivo ha sido y continúa siendo el suministro de instrumentos capaces de prestar una ayuda en la adopción de decisiones. La **matemática aplicada** alcanza, así, su mayor nivel de operatividad. La utilización de estos elementos en los diferentes ámbitos en los cuales aparece el problema decisional constituye el último peldaño de esta escalera que los investigadores han subido para suministrar la necesaria ayuda a quienes tienen responsabilidades en el gobierno de las naciones y en la gestión de empresas e instituciones.

En los momentos actuales y, como consecuencia de las razones reiteradamente expuestas y otras que podríamos añadir, no es posible asentar toda la actividad investigadora sobre el principio del tercio excluso, el cual ha resultado estrecho para albergar las “explicaciones” de las complejas realidades y fenómenos propios de nuestra época. La formulación de un nuevo principio se ha convertido, así, en cuestión fundamental si se desean abrir nuevas puertas a las aventuras que el futuro propone.

Hace casi un siglo, en 1910, Lukaciewicz ⁴ expuso su “principio de valencia”, señalando que entre la verdad y la falsedad había una tercera posición. Entre el 1 y el 0 colocaba, así, el 0.5 que representaba la indeterminación. Retomaba, de esta manera, la idea de los epicúreos.

Los amplios desarrollos habidos en los últimos decenios en el ámbito de la matemática no determinista y las fructíferas aplicaciones sobre todo en el área de la ingeniería y de la economía nos ha llevado al intento de definir un principio capaz de cobijar los distintos operadores lógicos que habían ido surgiendo, los cuales difícilmente tenían cabida dentro del principio del tercio excluso ya que su encastre sólo era posible a través de medios artificiosos. Pretendíamos, además, que el nuevo principio pudiera generalizar el de tercio excluso. De estas reflexiones

³Gil Aluja, J.: “Elements for a theory of decision in uncertainty”. Kluwer Academic Publ. Boston. London, Dordrecht 1999. Pág.: 15

⁴Lukaviewicz, J.: “O zasadzie wylaczonego srodkę”. Przegl'd Filozficzny. 13. 1910

nació el principio de la simultaneidad gradual. Este principio puede ser enunciado de la siguiente manera: “Una proposición puede ser a la vez verdadera y falsa. a condición de asignar un grado a su verdad y un grado a su falsoedad”.

Pasemos, mediante unos ejemplos, a poner en evidencia el contenido y significación de este enunciado, al tiempo que mostramos la posición relativa del mismo con respecto al principio tradicional.⁵

Existen ciertas proposiciones para las cuales normalmente no se plantean problemas mayores para aceptar el cumplimiento del principio del tercio excluso. Así, cuando se anuncia “Pedro pertenece al sexo masculino”, concluimos en la **verdad** de esta proposición y en la falsoedad de la negación. Pero se pueden aportar otras proposiciones para las cuales el cumplimiento no resulta tan claro. En efecto, proposiciones tales como “Pedro es alto”, “Pedro es moreno”, “Pedro es inteligente”, plantean ciertos problemas dada la relatividad del calificativo “alto”, “moreno”, “inteligente”. Para solucionarlos, los estudios tradicionales acostumbran a establecer un umbral (evidentemente subjetivo y arbitrario) a partir del cual se asume la verdad de la proposición. Si se acepte en el caso de la altura que el umbral es 1 m. 80 cms., quienes lleguen o sobrepasen esta altura serán consideradas personas altas y para los que no alcanzan esta talla se les asignará la falsoedad de la proposición. Llegamos, de esta manera, a admitir que una persona con una altura de 1 m. 80 cms. es **alta** mientras que otra de 1 m. 79 cms. es **no alta**. Creemos que la aceptación del principio de la simultaneidad gradual permite una buena solución convirtiendo el principio del tercio excluso en un caso particular. Veámoslo.

Para una mayor comodidad continuaremos con uno de nuestros ejemplos. En cuanto a la proposición concerniente a la altura se buscarán aquellas medidas para las cuales se cumple plenamente la verdad y la falsoedad de esta proposición. El ser humano más alto del mundo mide 2 m 38 cms y el más bajo (proposición negativa de la anterior) 0 m. 56 cms. A partir de estas cotas, en nuestro caso numéricas (aunque no tienen porque ser siempre así) se establece un orden desde la verdad (alto) hasta la falsoedad (bajo). Este escalado puede venir enmarcado por un intervalo, por ejemplo [0, 1] o por cualquier otro concepto apto para permitir la ordenación. Si, a efectos de una mayor sencillez y para no separarnos del ámbito numérico se acepta el intervalo [0, 1], asignaremos un 1 a la verdad y un 0 a la falsoedad y, entonces, el gigante de 2 m. 38 cms. será alto en un grado 1 y el enano de 0 m. 56 cms, alto en grado 0. Una persona con altura de 1 m. 70 cms. podría ser considerada alta en un grado 0.3, por ejemplo. Se establece, así, que cuando más nos acercamos a la verdad de la proposición el grado asignado se halla más próximo de 1 y cuanto más nos alejamos, el grado adscrito se encuentra más cerca de 0.

⁵El posterior desarrollo ha sido recogido de la obra de Gil-Aluja, J.: "Elements for a theory of decision in uncertainty". Kluwer Academic Publ. Boston, Londres, Dordrecht 1999, Págs.: 16-18

Pero cuando planteamos la falsedad de la proposición, convirtiendo el **alto** en **no alto**, o bajo si se quiere, sucede que los mismos protagonistas juegan un papel diferente, al serles asignado un grado distinto. Así, como bajo, el enano lo es en grado sumo por lo que le será asignado una valuación de 1, mientras que el gigante no es bajo en absoluto y su grado será 0. Quien con altura de 1 m 70 cms. era alto en un grado 0.3, será bajo en un grado 0.7, pero ejemplo. De esta manera, cualquier persona es alta y baja a la vez, si tomamos la precaución de acompañar al calificado de un grado expresable numéricamente (como en nuestro ejemplo) o no numéricamente, si así procediera.

Pero es que, además, a todas aquellas proposiciones a las cuales se acomoda confortablemente el principio del tercio excluso también les son aplicables el principio de simultaneidad gradual. Lo que sucede, entonces, es que de todos los grados posibles sólo le son asignados los extremos. De esta manera, con las reservas derivadas del caso, a la proposición “Pedro pertenece al sexo masculino” se le asigna un 1 en su grado de verdad y a la proposición “Pedro no pertenece al sexo masculino” un grado 0 de verdad. Por tanto, pertenece y no pertenece al sexo masculino, pero en grado distinto.

Esta breve exposición permite, así lo esperamos, poner de manifiesto la generalidad del principio de simultaneidad gradual, capaz de albergar una gran variedad de desarrollos lógicos. En efecto, frente a las limitaciones existentes en la lógica formal surgidas del rigor derivado del principio del tercio excluso, se ha comprobado la enorme flexibilidad y adaptabilidad de las llamadas lógicas multivalentes. Un claro ejemplo de cuanto acabamos de señalar lo hallamos en los operadores lógicos de inferencias. Así, en la inferencia binaria el único operador es:

$$c = \bar{a} \vee b$$

en donde a y b son los predicados y c la inferencia, mientras que en el ámbito de la multivalencia son posibles una práctica infinidad de operadores. Se acostumbran a citar como más conocidos, la llamada inferencia de Lee, cuyo operador es el mismo que el del campo binario ya presentado; la inferencia de Lukaciewicz, cuya formulación habitual es la siguiente:

$$c = 1 \wedge (\bar{a} + b)$$

la inferencia de Goguen, cuya expresión, que reproducimos, permite la utilización de valuaciones o probabilidades:

$$c = 1 \wedge \frac{b}{a}$$

así como las surgidas de cualquier T-conorma⁶. A ellas es posible añadir un

⁶Como es conocido de toda T-conorma se puede obtener un operador de inferencia con sólo sustituir la proposición o predicado de a por \bar{a} .

sinnúmero de operadores que cumplen los requisitos necesarios para ser consideradas como inferencias. No vamos a extendernos más en estos planteamientos y sí, en cambio, hacer referencia al tránsito de la lógica a la matemática fundamental.

3 Una propuesta de matemática de la incertidumbre

Cuando se encadenan palabras y/o proposiciones de una manera “razonada” llegamos normalmente a ciertos **resultados**. Estos resultados se amalgaman a su vez entre sí y siempre a través de palabras, expresadas de forma verbal o escrita, se consiguen nuevos resultados. Todo este proceso puede ser considerado propio del ámbito de la lógica. Pero tanto las palabras como las proposiciones son susceptibles de representación a través de símbolos, de la misma manera que los lazos entre ellas, pueden ser expresados mediante operadores. Cuando se actúa así, y se traspasa la frontera de la palabra para llegar al reino de los símbolos, se transita de la lógica a la matemática.

Desde hace casi cincuenta años, han aparecido algunos intentos de llevar a cabo la construcción formal de una matemática a partir de las incipientes lógicas multivalentes. Poco a poco los resquicios de estos ensayos irían formando un caldo de cultivo del cual nacería una idea capaz de aglutinar, con el tiempo, a centenares de miles de investigadores cuyos trabajos abrirían el camino a una **matemática de la incertidumbre**. Tiene lugar, en primer término, el desarrollo de los elementos numéricos, dando paso a una nueva aritmética de la incertidumbre. A los operadores llamados “duros”, propios del mecanicismo y aptos para la manipulación de magnitudes objetivas, se van incorporando otros operadores, considerados “blandos”, los cuales ejercen una función central cuando se trata de amalgamar, de la mejor manera posible, elementos con una alta carga de **subjetividad**. Si el operador más representativo de los primeros es el de composición suma-producto, el que posee una mayor significación entre los segundos es el de composición o convolución maxmin.

Durante el último tercio del siglo XX, se observa que matemáticos e ingenieros principalmente, van pasando desde el campo de la investigación tradicional y secular a esta nueva manera de formalizar los razonamientos. Los investigadores se afanan en el análisis y desarrollo de nuevos conceptos, aportando una simbología capaz de identificar los **elementos numéricos** que conforman una emergente matemática, diferenciada de la sustentada sobre razonamientos mecanicistas. Paralelamente, quizás con un cierto “gap”, se van abriendo paso algunos conceptos, surgidos casi siempre de planteamientos combinatorios, los cuales permiten vislumbrar la formación de un núcleo de conocimientos básicos con un contenido suficientemente unitario para poder hablar de una **matemática no numérica de la incertidumbre**. Así se está consiguiendo, cada vez más, expresar con

mayor fidelidad los razonamientos surgidos de las lógicas multivalentes mediante la matemática numérica de la incertidumbre, por una parte, y no numérica de la incertidumbre, por otra.

Todos cuantos, de alguna manera, nos hallamos en el ámbito de las investigaciones sociales, económicas o de gestión, somos conscientes de la necesidad de disponer de un “instrumental” apto para ser utilizado en el tratamiento de los problemas complejos, característicos de nuestros tiempos y que, difícilmente, la matemática en su estado puro puede suministrar, por lo menos de manera inmediata. De ahí la importancia de una **matemática aplicada** que, en los diferentes campos del conocimiento ha tomado nombres distintos. Recordemos a este respecto la **investigación operativa** de los estudios clásicos o los trabajos recogidos con la denominación de **técnicas operativas de gestión**, en época más reciente. La incorporación al acervo científico de trabajos realizados en el ámbito de la matemática pura han facilitado (y están facilitando) la construcción de herramientas de indudable eficacia. Estas herramientas son presentadas, generalmente mediante modelos o algoritmos. Estos elementos técnicos no son siempre novedosos, aparecido de manera instantánea y súbita, sino que, en buena medida, constituye total o parcialmente, reformulaciones de conceptos conocidos o bien adaptaciones de esquemas utilizados con éxito en otros momentos.

Quizás uno de los aspectos que permite vislumbrar mejor la línea que separa las matemáticas del determinismo y del azar con los de la incertidumbre viene dado por la naturaliza de las asignaciones numéricas. En efecto, sabemos que el concepto de **medida**, utilizando en teoría de conjuntos, significa un dato que es aceptado con carácter general porque se le supone *objetivo*. Teóricamente debe satisfacer determinadas propiedades entre las que se encuentran la “aditividad”. Así cuando se consideran dos subconjuntos booleanos A y B que son disjuntos (no poseen ningún elemento comán) se puede escribir:

$$m(A \cup B) = m(A) + m(B)$$

En el campo de las probabilidades, la noción de medida adquiere toda su significación, ya que el “evento” constituye, en si mismo, un concepto objetivo.

Sin embargo cuando se hace referencia a una “sensación” o “percepción” de tipo subjetivo que no es posible o no se sabe medir, se recurre a otro concepto: el de **valuación**, utilizado, entre otros por la teoría de los subconjuntos borrosos.

Así, dados \mathcal{A} y \mathcal{B} , si se supone que \mathcal{A} está incluida en \mathcal{B} , es decir, que para cada elemento el nivel de pertenencia es siempre igual o mayor en \mathcal{B} que en \mathcal{A} , se podrá escribir que, dado que \mathcal{A} se halla incluido en \mathcal{B} :

$$v(\mathcal{A}) \leq v(\mathcal{B})$$

Esta propiedad se sustenta en el concepto subjetivo de *sensación*.

En el ámbito de la economía y la gestión de empresas, se han realizado intentos, creemos que de manera parcial logrados, de crear unos elementos capaces de

llegar a un adecuado tratamiento de los fenómenos que tienen lugar en el seno de los estados y de las empresas, cuando su conocimiento tiene lugar de manera poco precisa. Para ello se han utilizado la teoría de los errores, la de los intervalos de confianza, la de los números borrosos, de los subconjuntos borrosos y todas las generalizaciones propuestas, que ya hemos empleado. La diferencia con el tratamiento realizado en los esquemas tradicionales, es importante. Ante la imposibilidad de recoger con precisión la compleja e incierta realidad económica, se recurrió a una *simplificación inicial* para realizar los desarrollos posteriores en base a estos elementos simplificadores. Las posibles desviaciones iniciales se iban acumulando y ampliando a medida que el proceso operativo avanzaba. Se perdía además, una información desde el principio que ya no era posible recuperar.

Por nuestra parte preferimos recoger los fenómenos económicos y de gestión con su incertidumbre, para realizar los pertinentes desarrollos *conservando la imprecisión* (y también toda la información) para hacerla “caer” lo más tarde posible, dado que siempre es posible (perdiendo información) reducir la incertidumbre.

Cuanto acabamos de señalar, tiene como finalidad plantear ciertas reflexiones sobre los profundos cambios que se están produciendo en el ámbito de estudio de los problemas en las empresas e instituciones de nuestros días. Creemos que, de una manera casi imperceptible pero continuada, los investigadores van aceptando las nuevas bases sobre las que se asientan las modernas técnicas de optimización.

Ahora bien, conviene poner en evidencia una vez más que no todos los conceptos, métodos y técnicas que se están utilizando han nacido de manera espontánea y súbita. Algunos de ellos, incluso, han sido empleados hace ya varias décadas, aunque en otros contextos clásicos. Quizá haya sido la irrupción de la teoría de los subconjuntos borrosos⁷ la espoleta que ha elevado el nivel de su interés y utilidad, hasta convertirlas en el eje de los nuevos movimientos.

Todo cuanto ha sido expuesto no debe inducir a la falsa creencia de la inutilidad de los modelos basados en los instrumentos surgidos al amparo de los paradigmas clásicos. Es más, los indudables avances que en su aplicación se han producido en los últimos años han permitido un perfeccionamiento de los estudios cuantitativos dirigidos al tratamiento de los fenómenos complejos, resultado, así, válidos aquellos esquemas, en muchas ocasiones afortunadamente. Sólo cuando no es posible, honestamente, considerar estimaciones “objetivas”, se debe recurrir a principios y modos de actuación diferentes. Pero en un mundo convulsionado como el de nuestros días, no parece que la incertidumbre vaya a remitir y únicamente conviviendo con ella resultará fácil la aceptación de nuevas reglas.

La noción de subconjunto borroso forma parte, ya hoy, del conglomerado de elementos conocidos con el nombre de **matemáticas**. Se adapta muy bien tanto al tratamiento de lo subjetivo como de lo incierto. La nueva concepción de la incertidumbre, a partir de los conceptos borrosos, ha dado lugar a una

⁷Zadeh, L.: “Fuzzy Sets”. Information and Control. 8 de Junio de 1965. Págs.: 338-353

distinta manera de pensar que reúne el rigor del razonamiento con la riqueza de la imaginación, asociando, así, las posibilidades secuenciales de la máquina a las posibilidades de las partes menos utilizadas del cerebro humano.

Los elementos instrumentales surgidas de los desarrollos de la matemática borrosa no son mucho más complicados que los utilizados normalmente, e incluso resultan más simples y mucho más cercanos a la manera habitual de pensar del hombre. La teoría de los subconjuntos borrosos es un intento, por el momento parcialmente logrado, de rehabilitar científicamente la subjetividad y la imprecisión.

La utilización de los esquemas borrosos tiene lugar actualmente, en la práctica totalidad de los campos de estudio de las ciencias. Se encuentra en la gestión de las empresas, en biología, en medicina, en geología, en sociología, en fonética y hasta en música, por sólo citar algunos. Todo problema situado en el ámbito de la incertidumbre es susceptible de ser tratado a través de la teoría de los subconjuntos borrosos. A medida que transcurre el tiempo, cada vez se van incorporando en los esquemas formales más mecanismos del pensamiento tales como las sensaciones y las opiniones numéricas y no numéricas.

Desde hace 70 años en elevado número de matemáticos se ha interesado por las lógicas multivalentes, entre ellos cabe destacar a B. Rusell, Lukaciewicz, Post, etc., pero es en 1965 en que Lofti Zadeh publica su primer artículo⁸ donde se inicia el verdadero avance en el estudio de esta nueva rama de las matemáticas. Hoy se estima que existen más de 600.000 investigadores dedicados a su estudio y desarrollo.

Así, pues, se concibe hoy una matemática borrosa de la misma manera que no existió problema en su momento en concebir una matemática del azar. Es más, si para las relaciones hombre-ordenador en la situación actual resulta imprescindible recurrir a teorías mecanicistas, para la relación entre el hombre y sus semejantes parece más adecuada la utilización de la teoría de la incertidumbre en la que la borrosidad juega un papel esencial.

4 Breve referencia a la matemática numérica para la optimización

Los modelos numéricos establecidos para el tratamiento de los problemas en un ambiente de incertidumbre se asientan, fundamentalmente, en la definición de un amplio abanico de números en cuya esencia se halla ausente el concepto de precisión. Se considera, así, inicialmente una magnitud para la que no se conoce de manera precisa su valor o, si se prefiere, que es “incierta”. Supongamos que la magnitud pertenece a \mathbb{R} .

⁸Zadeh, L.: “Fuzzy Sets”. Information and Control. 8 de Junio de 1965. Págs.: 338-353

Existen muchas situaciones en las que se puede afirmar, sin embargo, que una magnitud x es superior o igual a $a_1 \in \mathbb{R}$ e inferior o igual a $a_2 \in \mathbb{R}$, es decir que x pertenece al segmento $[a_1, a_2] \subset \mathbb{R}$. Por otra parte, se supone que no existe ley de probabilidad alguna, incluida la ley uniforme o equiprobable, que pueda ser afectado a los elementos de este segmento. Se dirá entonces que:

$$A = [a_1, a_2]$$

es un “intervalo de confianza” en \mathbb{R} .

Hay que señalar que:

$$(a_2 = a_1) \quad (A = [a_1, a_1] = a_1)$$

Un intervalo de confianza se puede generalizar a través de un “dominio de confianza” en \mathbb{R}^n . Se incorpora, normalmente, la hipótesis según la cual un “dominio de confianza” (es decir, un intervalo de confianza si se considera $n = 1$) es siempre convexo.

La definición de la convexidad de un dominio de confianza es la siguiente:

Sea A un dominio de confianza en \mathbb{R}^n , $n = 1, 2, 3, \dots$. Si se consideran dos puntos (x^1) y (x^2) que pertenecen a A , entonces todo punto (x^*) situado sobre el segmento que une x^1 y x^2 pertenece a A .

También se puede utilizar otra definición que, evidentemente equivale a la anterior:

$$(A \subset \mathbb{R}^n \text{ convexa}) \iff (\forall(x^1) \in A, \forall(x^2) \in A) :$$

$$(\mu_A(\lambda(x^1) + (1 - \lambda)(x^2))) \geq (\text{Min } [\mu_A(x^1), \mu_A(x^2)]), \quad \forall \lambda \in [0, 1])$$

en donde:

$$(\mu_A(x) = 0) \iff (x \notin A)$$

$$(\mu_A(x) = 1) \iff (x \in A)$$

Cuando el dominio de confianza D no es convexo resulta preferible utilizar las palabras “subconjunto de confianza”. Se trata aquí pura y simplemente del concepto de subconjunto ordinario tal que al introducir el concepto de “función característica” se pude hacer:

Para todo $x \in E$:

$$\mu_D(x) = \begin{cases} 1, & x \in D \\ 0, & x \notin D \end{cases}$$

El concepto de subconjunto de confianza añade al concepto de subconjunto una idea suplementaria. Se hace referencia a uno o varios elementos del referencial y se establece la hipótesis que este (o estos) elementos se hallan en el dominio y no pueden encontrarse fuera del dominio. A la definición de subconjunto se añade una idea de decisión, de elección, de confianza.

Vamos ahora a avanzar e imaginar que el subconjunto de confianza puede ser considerado a diferentes niveles de presunción⁹. Generalmente, cuando el nivel de presunción se eleva el número de elementos afectados, disminuye. Esto va a justificar la utilización de la palabra presunción. En el ámbito de la matemática se acostumbra a expresar, partiendo de una referencial E y de subconjuntos de confianza A_a que dependen de a , de la siguiente manera:

$$(a' > a) \iff (A_{a'} \subset A_a)$$

Dicho de otra manera, estos subconjuntos de confianza se encajan de manera monótona los unos dentro de los otros cuando crece.

El valor de α se llama “nivel de presunción”. Generalmente la determinación de A_a para cada valor es un “dato subjetivo” aunque en ciertas ocasiones puede ser una medida. Este dato subjetivo se halla, de esta manera, asociado a una apreciación de la incertidumbre. El nivel 0 corresponde siempre al referencial. A medida que aumenta el nivel de presunción los subconjuntos obtenidos no pueden aumentar (ampliarse). Es posible que al nivel 1 el subconjunto sea vacío.

De esta manera, mediante un encaje, se ha construido un nuevo concepto que se llama “subconjunto borroso” o de manera impropia para simplificar “conjunto borroso”. Se le puede, así, presentar como una generalización del concepto de subconjunto ordinario.

L. A. Zadeh presenta los subconjuntos borrosos a partir de la noción de conjunto. Se puede representar un subconjunto vulgar a partir de un par:

$$(E, \mu_A(x))$$

en donde E es el referencial y $A \subset E$ se halla definido por su función característica:

$$\mu_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

de esta manera, en los subconjuntos ordinarios o vulgares, la pertenencia de un elemento de subconjunto es de todo o nada.

En el supuesto de un subconjunto borroso se escribirá:

$$(E, \mu_A(x))$$

en donde E es el referencial y se escribe también $A \subset E$ que se halla definido por:

$$\mu_A(x) = a, \quad x \in A_a$$

imponiéndose la anteriormente enunciada propiedad de encaje.

Se ha podido observar la utilización del símbolo \sim (tilde en lengua española) para indicar que se trata del concepto de subconjunto borroso. Se ha colocado

⁹Del latín “praesumptio”: coyuntura, confianza, exceso de confianza....

este símbolo debajo de la letra mayúscula que indica que se trata de un conjunto o subconjunto; algunos autores colocan el \sim debajo de la letra mientras que otros no utilizan este símbolo con objeto de simplificar la elaboración tipográfica.

Los intervalos de confianza constituyen un medio de tratamiento de la incertidumbre en \mathbb{R} y en \mathbb{Z} cuando se dispone, para informaciones aceptadas como ciertas, de los extremos inferior y superior. Pues bien si se asocia la noción de subjetividad a la de incertidumbre, se llega, para \mathbb{R} o \mathbb{Z} , a la noción de número borroso. Para ello, vamos a considerar dos propiedades concretas de los subconjuntos borrosos.

Subconjunto borroso normal. Un subconjunto borroso $\tilde{A} \subset \mathbb{R}$ es “normal” cuando:

$$\vee_x \mu_{\tilde{A}}(x) = 1$$

Subconjunto borroso convexo. Se ha podido observar, que todo subconjunto borroso está constituido por un encaje de subconjuntos vulgares en función del nivel de presunción considerado. Supongamos ahora, que todos los subconjuntos vulgares de nivel α sean segmentos o intervalos de confianza $A_\alpha = [a_1^{(\alpha)}, a_2^{(\alpha)}]$ entonces la anterior propiedad ($\alpha' > \alpha$) $\iff (A_{\alpha'} \subset A_\alpha)$ se escribirá:

$$(\alpha' > \alpha) \iff \left([a_1^{(\alpha')}, a_2^{(\alpha')}] \subset [a_1^{(\alpha)}, a_2^{(\alpha)}] \right)$$

Se dirá entonces que el subconjunto borroso A es convexo.

Una definición equivalente de la convexidad adaptada al caso \mathbb{R} sería:

$$(\tilde{A} \subset \mathbb{R}^n \text{ convexa}) \iff (\forall x_1 \in A_\alpha, \forall x_2 \in A_\alpha, \forall \alpha \in [0, 1], \lambda \in [0, 1] :)$$

$$\mu_{A_\alpha}(\lambda x_1 + (1 - \lambda)x_2) \geq \mu_{A_\alpha}(x_1) \wedge \mu_{A_\alpha}(x_2)$$

Se llega, así, a la noción de número borroso.

Se define un número borroso como un subconjunto borroso del referencial de los reales, que tiene una función de pertenencia normal (debe existir una x_i para la que $\mu(x)$ toma el valor uno) y convexa (cualquier desplazamiento a la derecha e izquierda de x este valor $\mu(x)$ va disminuyendo).

Un número borroso puede ser representado a través de los segmentos formados al “cortar” (asignar un valor) la función de pertenencia a unos determinados niveles.

La teoría de los números borrosos puede considerarse como una ampliación de la teoría de los intervalos de confianza, cuando se consideran estos intervalos a todos los niveles desde 0 hasta 1, en lugar de considerar un solo nivel. También es posible, a partir de un número borroso, conocer los intervalos de confianza, para cada uno de los niveles desde 0.1 hasta 1 para un determinado número borroso.

De todo ello se deduce que un número borroso se caracteriza por los pares “nivel de presunción” “intervalo de confianza”, ya que a cada nivel de presunción se le adscribe un intervalo de confianza.

Para realizar operaciones con números borrosos se actúa de la misma manera que con los números reales ordinarios, operando nivel a nivel tal como se hace con los intervalos de confianza.

De entre todos los números borrosos aparece, por su facilidad de utilización, el *número borroso triangular* cuya singularidad consiste en que se halla determinado por tres cantidades: una por debajo de la cual no va a descenderse, otra en la que por encima no será posible llegar, y finalmente, aquella que representa el máximo nivel de presunción. La representación gráfica de un número borroso triangular (a_1, a_2, a_3) queda reflejado, en un sistema de coordenadas, por un triángulo. De ahí su denominación.

El número borroso triangular permite formalizar de manera muy fidedigna gran cantidad de situaciones económicas en la que se estiman magnitudes localizadas en el futuro. Así, en la estimación del coste de un producto a elaborar, es frecuente pensar que su precio no va a ser inferior a 40 ni superior a 70, siendo el precio que tiene la máxima posibilidad 55 unidades monetarias: se ha definido, entonces, en el campo de la incertidumbre, un número borroso triangular.

Dado que en el ámbito de la economía y gestión de las empresas se estudian problemas cuyas magnitudes se proyectan hacia el futuro, no exigen, frecuentemente, una extrema precisión sino la mayor adaptación posible a la realidad. Un presupuesto no precisa de una exactitud al céntimo sino que refleja lo que va a suceder en la realidad con una “buena aproximación”. Una estimación de ventas para un período no puede realizarse de una manera totalmente rígida, pues hay demasiados elementos que influyen en ella. Los ejemplos surgen a millares. En actividades repetitivas, en cambio, la probabilidad resulta altamente fructífera, pero en la gestión de las empresas, la repetitividad constituye la excepción. De ahí el interés en la utilización de los números borrosos en general y de los triangulares en particular.

Conviene señalar, en última instancia, que la transformación de los modelos tradicionales de carácter numérico al campo de la incertidumbre, basada en la sustitución de números precisos por números inciertos, los generalizan e incluso los hacen más aptos para el tratamiento de la realidad. Ahora bien, en sí misma, esta transformación no resulta suficiente para abordar la compleja realidad de nuestros días. Son los modelos inciertos, desarrollados a partir de conceptos emergidos de la matemática no numérica, quienes son capaces de llenar el vacío, cada vez más profundo, en el estudio de unos fenómenos que escapan, no sólo a la medición sino también a la valuación¹⁰, aun cuando ésta sea realizada a través del más simple de los números inciertos.

La incorporación de estos instrumentos significa una clara ruptura en relación a situaciones precedentes. Y ello, aun cuando en su desarrollo se utilicen elemen-

¹⁰Recordemos que el concepto de valuación se asimila a una asignación numérica realizada subjetivamente. La subjetividad es la característica que la diferencia de la noción de medida, de carácter eminentemente objetivo

tos ya empleados en los estudios clásicos. La actual situación de equilibrio en la cohabitación de modelos numéricos y no numéricos en la incertidumbre, está dando paso a una mayor supremacía de estos últimos, como consecuencia de las crecientes dificultades de acotar siquiera los fenómenos sociales, económicos y de gestión. Por ello, aquellos conceptos que exigen inevitablemente ser **expresados numéricamente** (en la certeza o en la incertidumbre) han ido dejando protagonismo por las dificultades de expresarlas objetiva y hasta subjetivamente, habida cuenta del contexto cada vez más incierto en el cual se inscriben.

5 Elementos no numéricos para la optimización

La optimización implica, frecuentemente, la necesidad de tomar partido por una alternativa frente a otra u otras. Los estudios económicos y de gestión se han desarrollado, en gran parte, con la búsqueda de elementos capaces de dar pautas, ayudar en definitiva, a aquellos en quienes recae la tarea de pasar de los planteamientos a su ejecución. En los estudios clásicos las ayudas toman muchas veces la forma de **criterios** basados en conceptos con **apoyo numérico**. Así, se hallan presentes en prácticamente todas las áreas del conocimiento económico las nociones de economicidad, rentabilidad, productividad, ... las cuales comparten informaciones siempre expresadas mediante números precisos, aleatorios o, de manera más reciente, inciertos. Ahora bien, mantenerse en esta línea exige plantear la pregunta relativa a cómo poder suministrar criterios cuando honestamente **no** es posible establecer unas asignaciones numéricas, ni siquiera subjetivas. El camino parece, pues, cortado y se hace precisa la búsqueda de nuevas vías.

Nos hemos acostumbrado a disponer, cuando se inicia un proceso de optimización, de una cifra representativa del nivel de apreciación de cada objeto físico o mental que en él interviene. El mayor (o en su caso el menor) de estos **números** determina muchas veces el resultado buscado. Tan asociados se han hallado número y nivel de apreciación que es difícil concebir el uno sin el otro. Sin embargo, esta inveterada costumbre no debe impedir el intento de separar este matrimonio, no siempre bien avenido. Creemos llegado el momento de dar una solución amistosa a las incompatibilidades, haciendo durar la convivencia en todos aquellos casos en los cuales no sea necesario el engaño y dejando que el número y el nivel de apreciación puedan estar presentes de manera independiente en los diferentes planteamientos de la ciencia económica, cuando las circunstancias lo exijan.

Sabemos muy bien, es evidente, como utilizar el análisis numérico, principalmente en la certeza y el azar, pero también en la incertidumbre. El nivel de apreciación en cambio, carece de la costumbre de actuar por sí solo y se ve necesitado de una estructura suficientemente sólida para poder enfrentarse a los planteamientos más generales.

Puesto de manifiesto la existencia de un nuevo contenido de la matemática

para la optimización, pasamos ahora, a la descripción de aquellos elementos que constituyen su soporte. En otras palabras, vamos a ver cuál es la tipología básica sobre la que se asienta la optimización. Nos referimos a los conceptos de **relación, asignación, agrupación y ordenación**. El estudio de la optimización en las ciencias sociales pone de manifiesto que en la práctica totalidad de los casos la optimización tiene lugar bien para establecer una relación, bien para afectar una “cosa” a otra “cosa”, bien para realizar agrupaciones casi siempre homogéneas, (las cuales sirven también para separar los grupos formados) o bien para establecer un orden de prioridades, unas veces de mejor a peor (de más a menos) , otras veces en sentido inverso.

Vamos a proceder, a continuación a la anunciada descripción, muy somera¹¹ como procede en este contexto, de estos cuatro elementos:

a) Relación

El fenómeno de la **relación** forma parte consustancial de la vida social, económica y de la actividad de las empresas. Afecta tanto a los seres inteligentes como a los animales inferiores y objetos inanimados. Se trata de lo que para generalizar llamamos “objetos” físicos o mentales.

Quizás el aspecto de las relaciones que resulta más familiar, se sitúa en el ámbito de las conexiones entre las personas integrantes de nuestra sociedad. Los individuos que viven dentro de un contorno social forman **círculos** dentro de los cuales realizan la totalidad de sus actividades vitales. La característica de **cada uno** de estos vínculos es que un sujeto cualquiera es capaz de relacionarse con todo otro sujeto, bien de manera directa o a través de otros sujetos.

La vida se desarrolla en el interior de cada círculo unas veces aumentando el **nivel** de la relación, otras veces “enfriándolo”, es decir, reduciendo la fuerza de la conexión.

El paso de un círculo o “casta” a otro resulta, la mayor parte de las veces, muy difícil. Casi siempre es necesario algún tipo de revolución. Ahora bien, cuanto se acaba de señalar no impide la existencia de algunas relaciones entre uno o varios miembros de un “círculo” con uno o varios miembros de otros “círculos”. Sin embargo, esta relación **no es simétrica** en el sentido de que una cierta relación en un sentido no es correspondida en el sentido inverso (prestar un servicio, conceder una dádiva, ... sin contrapartidas). A pesar de ello, es posible imaginar una sociedad ideal, en la cual todos sus miembros se hallan relacionados entre sí, de manera directa o indirecta.

La fuerza de la costumbre nos ha llevado a considerar el “conjunto producto” como la noción capaz de representar la totalidad de relaciones existentes entre los elementos de un conjunto referencial o las relaciones entre dos o más referenciales. La incorporación del concepto de “grafo” permite desarrollar la idea de relaciones entre una parte de los elementos del conjunto producto, con fructíferos resultados.

¹¹Gil Aluja, J.: “Elements for a theory of decision in uncertainty”. Kluwer Academic Publisher. Boston. Londres. Dordrecht. Págs: 20-28

Los estudios clásicos, basados en la lógica booleana, consideran la alternativa de existencia o no de relación. La necesidad de establecer un grado o nivel de relación, aconseja pasar al ámbito de las lógicas multivalentes. La **relación borrosa** o **matriz borrosa** adquiere, entonces, un papel relevante.

En efecto, a partir de un grafo borroso, presentado en forma matricial o satigada, es posible, mediante adecuados algoritmos, conocer si existe **un solo círculo** de relaciones (grafo fuertemente conexo) o **varios** círculos de relaciones (grafo no fuertemente conexo). En este último caso parece la noción de clase de equivalencia o subgrafo fuertemente conexo.

Establecidos los nexos de relación y el nivel de los mismos, cabe preguntar cuál será la evolución de los mismos a través de sucesivos períodos de tiempo o etapas, en su caso. En otro orden de ideas, también se puede centrar el nivel de interés en el conocimiento de las variaciones en el grado o intensidad de la relación a medida que van interviniendo más elementos que actúan como intermediarios. En otras palabras, cuando la cadena de relaciones se va alargando.

Existe un operador capaz de dar cumplida respuesta a tales cuestiones. Se trata del **operador convolución maxmin**. La utilización de este interesante agente conector permite llegar a la conclusión de que el tiempo, el espacio, o las conexiones, según los casos, desemboca en tres situaciones a largo plazo, tipificadas por la convergencia en el límite, la periodicidad y las situaciones caóticas.

Una amplia tipología de relaciones permite establecer ciertas características y propiedades poseídas por cada tipo de relación. De esta manera se abre un abanico de formas teóricas aptas para representar diversas realidades. A partir de ellas el encadenamiento de relaciones entre objetos pertenecientes a varios conjuntos conduce a conexiones directas e indirectas.

Finalmente, otros aspectos merecen especial atención. A medida que transcurre el tiempo o se suceden etapas, se pueden generar nuevas relaciones dentro de cada subgrafo o entre elementos pertenecientes a subgrafos distintos. En este caso la red de relaciones va siendo cada vez más tupida. Pero no tiene porque ser siempre así. En ciertos supuestos desaparecen lazos, antes existentes.

Pero es quizás al analizar la variación en las intensidades, niveles o fuerza de las relaciones, cuando surgen particularidades que es necesario tener en cuenta a la hora de representar fielmente las realidades sociales, económicas y de gestión. Así, en multitud de supuestos tiene lugar la “degradación” en el nivel de una o varias relaciones, mientras que en otros, no pocos también, el transcurso del tiempo o la sucesión de etapas “fortalece” los vínculos.

El estudio de este tipo de relaciones presenta unas particularidades tales que no pueden ser tratadas de la misma manera que las no acumulativas. Resulta entonces necesario variar, en cierto modo, el camino seguido, haciendo intervenir además de la relación o relaciones primarias entre los objetos de un conjunto con los de otro u otros, la relación de objetos de cada conjunto consigo mismo. Así, cuando se trata de establecer la relación acumulada directa o indirectamente

entre los elementos de dos conjuntos, además de presentar una relación en forma matricial o sagitada entre los objetos de un conjunto y los del otro, se deben definir las relaciones en forma matricial o sagitada entre los objetos de un conjunto entre sí y los del otro conjunto, también entre sí. Se tiene, entonces, que el proceso formalizador parte bien sea de tres matrices o de tres grafos sagitados.

Un caso particular de cuanto acabamos de exponer viene dado por las llamadas relaciones de incidencia o causalidad ¹². En ellas un conjunto de causas está “conectado” con un conjunto de efectos. A medida que se van obteniendo los efectos acumulados de primera y segunda generación se observa un aumento del nivel de la relación. En este caso, las relaciones entre objetos de un mismo conjunto deben poseer la propiedad reflexiva, en el sentido de que la incidencia de un objeto consigo mismo es total. Es así que sus representaciones matriciales se caracterizan por poseer el valor unidad en todos los elementos de su diagonal principal. Lo mismo sucede en la forma sagitada en la cual cada arco que une un vértice consigo mismo está valuado con un uno. Nadie puede dudar de la importancia ejercida por las relaciones de causalidad a lo largo del desarrollo del pensamiento científico. El proceso propuesto permite, además, obtener “todas” las relaciones directas e indirectas, sin posibilidad de error u omisión, recuperando lo que se ha venido denominando “efectos olvidados”.

b) Asignación:

Una manera muy especial de establecer relaciones tiene lugar mediante un proceso conocido con el nombre de “asignación”. También se emplean, como sinónimos, otros términos tales como afectación o adscripción. Por nuestra parte utilizaremos habitualmente el primero de estos vocablos en el bien entendido que no hacemos bandera de esta elección.

La calificación de relación “especial” creemos tiene su plena justificación en un aspecto consustancial con la asignación, tan diferenciable que lo ha ido apartando de los estudios relationales clásicos. Nos referimos al hecho de que el objeto susceptible de afectación y objeto al cual éste debe ser afectado, no pueden jugar papeles reversibles, en caso alguno. El sentido, pues, de la adscripción es siempre el mismo. De ahí, la consideración fundamental de un conjunto de elementos a asignar y de otro conjunto de elementos que **siempre** recibirán la asignación. Y todo ello con independencia del propio fenómenos de la asignación, cuyas características específicas lo hacen merecedor de especial atención.

El planteamiento del problema de la asignación parte de la existencia de tres conjuntos, normalmente finitos, de objetos físicos o mentales. El primero recoge los elementos a asignar, el segundo los elementos que deben recibir la asignación y el tercero los elementos en los cuales se basa el proceso asignador (cualidades, características, singularidades, ...), en definitiva lo que podríamos denominar criterios de asignación. Cómo organizar el papel que juega cada uno de estos

¹²Kaufmann, A. y Gil Aluja, J.: ” Modelos para la investigación de efectos olvidados”. Ed. Milladoiro. Santiago de Compostela, 1988

conjuntos, constituye el punto de arranque a partir del cual se desarrollan las distintas técnicas que se han elaborado. Consideramos, por nuestra parte, que una buena manera de representar estos aspectos viene dada por la construcción de un subconjunto borroso para cada uno de los objetos a asignar tomando como referencial el conjunto de “criterios” de asignación. Se tienen, entonces, unos **descriptores**, tantos como elementos posee el primero de los conjuntos (elementos a asignar). De igual manera, se elaboran el mismo número de subconjuntos borrosos, con idéntico referencial, del conjunto de los “criterios”, como objetos receptores de la asignación. Se trata, también en este caso, de **descriptores**, pero ahora de los elementos del segundo de los conjuntos (objetos receptores).

Con objeto de obtener las relaciones, a partir de las cuales iniciar el proceso para la asignación, se puede recurrir a alguno de los índices capaces de expresar el “alejamiento” o “acercamiento”, en su caso, entre los objetos a afectar y los objetos a los cuales debe realizarse la afectación. Entre los más conocidos caben citar los que surgen de la **noción de distancia** y los que parten de la **noción de adecuación**. A partir de ellos, es posible utilizar una amplia gama de variantes.

La tarea de asignar convenientemente un objeto a otro objeto, tiene en sí misma, un carácter combinatorio. Por este motivo las técnicas precisas para esta labor han tenido que ser buscadas en este campo de la matemática. No es de extrañar, entonces, que se haya hurgado en aquellos algoritmos capaces de acotar el número de operaciones necesarias para encontrar la o las soluciones óptimas. La justificación de tales algoritmos proporciona una buena base teórica sobre la cual es posible sustentar una teoría de la asignación, quizás aun hoy incompleta.

Elementos de la programación matemática, flujos en redes, acotaciones en arborescencias, ... confluyen para encauzar los estudios de asignación hacia procedimientos de cálculo capaces de dar amplia respuesta a los problemas planteados. Estos procedimientos se concretan en algoritmos. Entre los más utilizados citaremos el **algoritmo por eliminación de filas y columnas**, al que es necesario añadir otros dos aptos para la optimización: el **algoritmo húngaro** y **algoritmo branch and bound**.

Las posibilidades de utilizar estos algoritmos en las realidades sociales, económicas y de gestión, son muchas. De hecho, disponemos de una buena experiencia, en este sentido, en campos tales como los recursos humanos, finanzas e inversiones, marketing, ... e incluso conocemos de un estudio realizado para la asignación de un jugador de fútbol a una posición del equipo. Un amplio ventanal se ha abierto para las optimizaciones basadas en la afectación, adscripción o, si se quiere, asignación.

c) Agrupación

El problema de la agrupación homogénea de objetos físicos o mentales constituye una constante para la optimización. En efecto, son frecuentes las situaciones en las cuales es necesario reunir, en bloques, “objetos” muchas veces con apariencia diferente, bien para una selección entre los componentes del mismo grupo bien para elegir un grupo entre varios de ellos. En la actividad social, económica y de

gestión, son múltiples los casos en que se presenta este planteamiento. Así, en el ámbito productivo, la separación por grupos homogéneos de materias primas, productos semielaborados o productos acabados¹³, constituye una constante en el día a día de los procesos de fabricación. En la actividad financiera la agrupación de productos o instrumentos de financiación¹⁴ o la de objetos de la inversión¹⁵ es algo que de manera expresa o tácita se halla en la mente de los agentes activos de este campo. La actividad comercial presenta una variada gama de planteamientos decisarios en cuyo origen se halla la necesidad de agrupar¹⁶. Y la gestión de los recursos humanos no es ajena a este problema¹⁷. Nos hemos limitado a señalar algunos de los muchos casos que aparecen en los estudios de optimización.

La formalización de estos planteamientos ha tenido lugar, a lo largo del tiempo, a partir de ciertas técnicas cuyo desarrollo no ha proporcionado, en nuestra opinión, los resultados deseados para dar cumplida respuesta a los retos planteados. La posterior dispersión de los caminos emprendidos ha evitado la formación de un esquema unitario con la suficiente generalidad para abarcar el amplio espectro de la compleja realidad social de nuestra época.

En efecto, vamos a tomar como referencia el concepto de semejanza. Este término adquirió en su momento, gracias al desarrollo de la matemática de la incertidumbre, una importancia especial, sobre todo para procesos de agrupación o separación. Con el tiempo se ha ido constatando que este concepto no ha resultado lo suficientemente útil para una eficaz agrupación, como consecuencia de la **no** existencia de la propiedad transitiva. En efecto, si se dispone de un grupo de objetos concretos o abstractos, *A*, *B*, *C*, y se cumple que *A* y *B* son semejantes a un determinado nivel, así como *B* y *C* son semejantes, también a este nivel, no tiene porque cumplirse que al nivel especificado lo sean *A* y *C*. Para que esto se cumpla es necesaria la intervención de la citada propiedad transitiva.

Es suficientemente conocida la manera de obtener los subgrafos transitivos a partir de un grafo de semejanzas (reflexivo y simétrico). Estos subgrafos expresan “relaciones de similitud” entre algunos elementos del referencial, formando el mayor “grupo” posible con características similares: Son las llamadas subrelaciones máximas de similitud. Para su obtención se han elaborado algunos algoritmos en-

¹³Gil Aluja, J.:MAPCLAN. Modelo for assembling products by means of clans. Proceedings de la Third International Conference on modelling and simulation MS'97. Melbourne 29-31 Octubre 1997. Págs 496-504

¹⁴Gil Lafuente, A.M.:Fundamentos de Análisis Financiero. Ed Ariel. Barcelona. 1993. Págs: 285-300

¹⁵Gil Aluja, J.: Invertir en el incertidumbre. Ed. Piramide. Madrid 1997. Págs: 115-139

¹⁶Gil Lafuente,J.:Marketing para el nuevo milenio. Ed. Piramide. Madrid. 1997. Págs:145-161

¹⁷Gil Aluja, J.: La gestión interactiva de los recursos humanos en la incertidumbre. Ed. CEURA. Madrid. Págs: 145-158

tre los cuales caben citar el de Pichat¹⁸ y el de Lafosse-Marin Kaufmann¹⁹. Los subconjuntos que se forman no son disjuntos. Y aunque esta circunstancia carece de importancia en muchos casos, cabe plantearse el problema de la formación de subrelaciones máximas de similitud disjuntas. Cuando esta propiedad se convierte en una exigencia, habrá que recurrir a la transformación del grafo borroso de semejanza en un grafo borroso de similitud a través del cierre transitivo.

La importancia que ha representado disponer de estos elementos teóricos capaces de obtener agrupaciones por semejanzas o similitudes, ha sido enorme. A pesar de ello, estos elementos no han resultado suficientes para resolver toda la gama de problemas de optimización que las realidades económicas y de gestión plantean. No se olvide que el punto de partida es una **matriz cuadrada**, en la cual los elementos de las filas coinciden en cantidad y en esencia con los elementos de las columnas. Y esto es un caso particular de otro más general, en el que no coinciden, ni en número ni en concepto, las filas con las columnas. Para llegar a él se puede partir de una matriz rectangular, ni simétrica ni reflexiva.

Conscientes de esta necesidad, intentamos en su día abordar la construcción de un conglomerado de conocimientos a partir de la noción de afinidad, ya utilizada en trabajos anteriormente, realizados junto al añorado profesor Kaufmann²⁰. Los resultados han permitido poner en evidencia el carácter general de este concepto, del cual la importante noción de similitud constituye un caso particular.

Para lograr este objetivo, tomamos como punto de arranque el concepto de relación, en su aspecto más amplio, en el sentido que pueden existir conexiones a distintos niveles entre elementos de dos conjuntos referenciales finitos. La presentación de estas relaciones mediante una relación borrosa, normalmente rectangular, permite un amplio juego del que se derivan interesantes reflexiones.

Una de ellas, y no la menos importante, hace referencia a la flexibilidad derivada de la posibilidad de obtener, a partir del tratamiento mediante α -cortes, un abanico de matrices booleanas capaz de permitir la adaptabilidad necesaria en la formación de agrupaciones, considerando los niveles deseados de homogeneidad.

Al disponer de matrices booleanas, se consigue un entronque con los estudios clásicos de la lógica binaria, lo que permite aprovechar relevantes aspectos, hasta ahora creemos poco utilizados en los procesos de optimización. Cabe destacar a este respecto, el conglomerado de conocimientos elaborados a partir del concepto de “familia de Moore”, los cuales permiten la obtención de adecuadas agrupaciones. La presentación de estas agrupaciones mediante estructuras reticulares pone

¹⁸Pichat, E.: Algorithm for finding the maximal elements of a finite universal algebra. Inform Proceeding 68. Publ. North Holland. 1969

¹⁹Kaufmann, A.: Modèles mathématiques pour la stimulation inventive. Ed. Albin Michel, 1979. Pág.: 62

²⁰Señalemos a título de ejemplo las obras de Kaufmann, A. y Gil Aluja, j.: Técnicas de gestión de empresa. Previsiones, decisiones y estrategias. Ed. Pirámide. Madrid. 1972, Cap. 10 y Técnicas especiales para la gestión de expertos. Ed. Milladoiro. Santiago de Compostela, 1993. Cap 13, de los mismos autores

en evidencia las afinidades. Para ello se ha recurrido a los retículos de Galois, a cuya belleza formal se añade su gran capacidad de adaptación.

Las posibilidades de utilizar este conjunto de elementos teóricos y técnicos a la optimización en la incertidumbre, se ven facilitadas por la puesta a punto de algunos algoritmos de empleo alternativo según las circunstancias concretas de cada momento. En efecto los algoritmos de la “correspondencia inversa máxima” y de las “submatrices completas máximas” constituyen eficaces instrumentos de solución rápida y eficaz a los problemas planteados. A ellos, hemos podido añadir un tercero, cuya base se halla en la llamada teoría de clanes. Creemos que, con esta trilogía, se puede cerrar, aunque sea provisionalmente, la esfera de la “agrupación” como concepto fundamental de la optimización en la incertidumbre.

d) Ordenación

Se llega finalmente, a la noción de orden. Si en los estudios tradicionales la ubicación de los elementos numéricos adquiere categoría de protagonismo, cuando las realidades sociales sufren convulsiones que incapacitan la adopción de técnicas lineales, su desplazamiento resulta inevitable. Es entonces cuando el proceso de ordenación asume la mayor responsabilidad.

En efecto, cada vez resulta más difícil circunscribir la fenomenología actual al estricto reducto de un número e incluso acotarla entre dos o más números. Ante estas circunstancias, el recurso a una comparación basada en formas no manifestadas cuantitativamente, puede proporcionar excelentes soluciones. Resulta evidente que sería preferible, para todos, disponer de la información suficiente, capaz de permitir una asignación numérica a cada fenómeno susceptible de optimización. Pero a falta de soportes aptos para el suministro de tales elementos, creemos más honesto conformarnos con menos, aunque este menos impida precisiones muchas veces superficiales.

Con objeto de obtener estructuras capaces de conducirnos hacia algoritmos aptos para la ordenación se han creado ciertos desarrollos que permiten una suficiente gama de caminos capaces de cubrir un amplio abanico de problemas hasta hace poco sin solución satisfactoria.

El primero parte de la noción de **función ordinal de un grafo**. La definición de este concepto a partir de la teoría de grafos resulta altamente fructífera al proporcionar una estructura básica a partir de la cual, siguiendo cada una de las formas señaladas, se elaboran algoritmos de fácil asimilación y eficacia probada para su utilización en los planteamientos económicos y de gestión.

La teoría de grafos suministra, así, interesantes esquemas para establecer una relación de orden. Si se exceptúan los grafos “fuertemente conexos” los cuales, representados por matrices con una sola clase de equivalencia, no permiten hallar un orden entre sus vértices, la posibilidad de descomponer un grafo **no** fuertemente conexo en subgrafos **sí** fuertemente conexos, ha abierto las puertas a la ordenación cuando no de vértices, por lo menos de conjunto de vértices (los que componen cada clase de equivalencia o subgrafo fuertemente conexo).

Una vez obtenidas todas las clases de equivalencia o subgrafos fuertemente conexos, se ha vencido el escollo más importante que podía evitar la ordenación. Ahora, cuando no se pueden ordenar vértices por la existencia de circuitos, es posible la ordenación de clases de equivalencia, es decir, de grupos de objetos físicos o mentales.

Sin abandonar la misma metodología general, se puede abordar un segundo camino que tiene como punto de arranque la denominada **matriz latina**. Su andadura, pues, tiene lugar, por lo menos inicialmente, a partir de una forma matricial. La utilización de adecuados operadores, principalmente el de convolución maxmin, conduce al método de la composición P-latina. La exigencia de una propiedad específica (la del camino elemental) permite establecer la enumeración de los caminos elementales de un grafo, que son, en sí mismos, fuente inmediata de ordenación. El algoritmo que se desprende hace el resto.

El último sendero susceptible de utilización ofrece unas características diferenciales de los dos anteriores. Su fundamento se halla en conocidas propiedades del cálculo matricial, las cuales fueron sagazmente utilizadas primero por Saaty²¹ y luego por Dhin²²

En un sistema social y económico marcado por la incertidumbre el concepto de orden ocupa un puesto de privilegio para la optimización. Ordenar inversiones, ordenar fuentes de financiación, ordenar recursos, ... es la antesala de la optimización. Cuando no es posible obtener un cuadro “valorado” de objetos, apelar a un “orden no cuantificado” de los mismo puede ser suficiente para una optimización aceptable.

Relación, asignación, agrupación, ordenación, ... y tantos otros vocablos que en un futuro inevitablemente van a aparecer, son conceptos que van emergiendo, en este intento de buscar nuevos caminos para dar solución a los complejos problemas que la optimización plantea.

6 Consideraciones finales

Tradicionalmente, la ciencia económica ha tenido como uno de sus objetivos fundamentales, la búsqueda de la optimización. No puede extrañar, entonces, que estemos viviendo unos momentos de desconcierto, cuando una realidad, llena de convulsiones que hacen la vida inestable, quiere ser tratada como se había hecho en situaciones de equilibrio, envueltas en estabilidades. Nos hemos de acostumbrar a pensar que la sociedad, la economía y la actividad de las empresas, tal como las hemos conocido hasta ahora, no tienen ninguna posibilidad de sobrevivir en un futuro muy inmediato y que muchos, pero que muchos cambios serán inevitables.

²¹Saaty, T.L.:Exploring the interface between hierarchies, multiples objectives and fuzzy sets. Fuzzy Sets and Systems. 1978. Vol. 1 N. 1. Págs: 57-68

²²Dinh, Xuan Bá.: A method for estimating the membership function of a fuzzy set. Revue Buseval, L.S.I. University Paul Sabatier. Toulouse 1984, N. 19 Págs.:68-82

Creemos que consenso e inestabilidad son conceptos que van cada vez más unidos y todavía lo irán más en el futuro. Tanto es así, que en las nuevas tendencias de la investigación científica en el ámbito de la economía y la gestión de las empresas, están tomando una posición cada vez más fundamental. Inestabilidad y volatilidad de los mercados financieros, inestabilidad y fluctuación de los precios, inestabilidad y precariedad de las ofertas de trabajo, son claros ejemplos.

Pueda resultar útil, en este contexto, una breve reflexión en torno a las posibilidades que ofrecen las propuestas que llegan de los laboratorios en donde se ensayan los nuevos hallazgos científicos. En ellos, podemos comprobar que la actividad científica se halla en una encrucijada en la que está en juego el futuro de la humanidad. Por un lado, la concepción geométrica del universo, por el otro, la concepción darwiniana. De una parte la imposición de unas creencias pre establecidas desde el esplendoroso amanecer newtoniano, en el que se soñaba con reducir el funcionamiento del mundo a la predictibilidad de un mecano. De otra parte, el vacío de lo desconocido. La atracción de la aventura, sólo guiados por la esperanza de abrir nuevos horizontes. La respuesta a la llamada de Bertran Russell, de Lukasiewicz, de Zadeh, de Lorenz, de Prigogine, de Kaufmann. El rechazo al yugo de la predestinación y la proclamación de la libertad de decisión que una y otra vez choca con el muro de la incertidumbre.

Incertidumbre. Vocablo casi proscrito hace sólo unos pocos decenios. Sujeto a maltrato y objeto de confusión por quienes más debían ser celosos guardianes del “grail” de la Ciencia. Hoy, gracias a los originales trabajos de Zadeh, resurge acaparando un protagonismo que facilita nuevos cauces a tantos investigadores deseosos de proporcionar respuestas a los interrogantes que la sociedad convulsa de hoy plantea.

Aunque resulta difícil buscar las raíces más profundas de las nuevas miradas con que los investigadores escrutan el mundo, no existe la menor duda de que algo importante estaba pugnando por emerger a la superficie de la actividad científica cuando destilaban las primeras esencias del evolucionismo, rica herencia del siglo XIX.

En efecto, en su fundamental obra “**El origen de las especies**” publicada en 1859, **Darwin** considera que las **fluctuaciones** en las especies biológicas, gracias a la selección del medio, dan lugar a una evolución biológica **irreversible**. De la asociación entre **fluctuaciones** (que asimila a la idea de azar, diríamos nosotros incertidumbre) e **irreversibilidad** tiene lugar una **autoorganización** de sistemas con una creciente complejidad.

Por su parte, **Clausius** formula, en 1865, la “**ley de aumento de la entropía**”, con la correspondiente división entre procesos **reversibles** e **irreversibles**. Esta distinción se hace explícita en la **segunda ley** que postula la **existencia de una función, la entropía**, la cual, en un sistema aislado, aumenta cuando existen procesos irreversibles y se mantiene constante en presencia de procesos reversibles. Por lo tanto, la entropía alcanza un valor máximo cuando el

sistema está llegando al equilibrio y acaba el proceso irreversible.

Tanto en el caso de **Darwin** como en el de **Boltzmann**, azar (o si quiere incertidumbre) y evolución se hallan estrechamente relacionados, pero el resultado de sus respectivas investigaciones conducen a conclusiones contrapuestas. En **Boltzmann**, la probabilidad llega a su **máximo** cuando se está alcanzando la **uniformidad**, mientras que en **Darwin** la evolución conduce a **nuevas estructuras autoorganizadas**.

En contraposición con estas perspectivas, el prototipo de la **física clásica** y como consecuencia de la teoría económica marginalista, es la **mecánica del movimiento**, la descripción de procesos de carácter **reversible y determinista**, en donde la dirección del tiempo no juega papel alguno, en la cual no existe un lugar ni para la incertidumbre ni para la irreversibilidad. En definitiva, los sistemas económicos y de gestión constituyen **inmensos autómatas**.

Es bien cierto que algunos fenómenos que surgen en la vida de los estados, instituciones y empresas se pueden perfectamente describir mediante **ecuaciones deterministas** pero, en cambio, otros comportan **procesos inciertos** o, en todo caso, **estocásticos**. Podría suceder que nuestra propia existencia, con toda su complejidad, se hallara, también inscrita en las **leyes generales** desde el momento primigenio del Big-Bang. Pero la ciencia, de tanto **buscar las generalidades, las simetrías y las leyes**, ha encontrado lo **mutable, lo temporal y lo complejo**.

Nos encontramos, pues, en una encrucijada que podría ser bautizada como geometrismo – darwinismo, en cuyo epicentro se halla una querella que data de más de dos mil años. En efecto, Aristóteles (384 - 322 a.C.) señalaba que: “respecto de las cosas presentes o pasadas, las proposiciones, sean positivas o negativas, son por necesidad verdaderas o falsas. Y de las proposiciones que se oponen contradictoriamente debe ser una verdadera y una falsa”. En esta misma línea se situaba el pensamiento de los **estoicos** a una de cuyas figuras centrales, Crisipo de Soli (\simeq 281 -208 a.C.), se le atribuye la formulación del llamado “principio del tercio excluso”. Los **epicúreos**, con su fundador Epicuro (341 - 270 a.C.) al frente, contestaron con vigor este principio, señalado que sólo es aceptable si no se da una tercera posibilidad “tertium non datur” (tercio excluso).

Tienen que transcurrir veintidós siglos para que Lukaciewicz, retomando la idea de los epicúreos, señalara que existen proposiciones que no son ni verdaderas ni falsas, sino **indeterminadas**. Esto le permite enunciar su “principio de valencia” (cada proposición tiene un valor de verdad). Se inicia, así, el camino para las llamadas lógicas multivalentes.

Con ocasión del Congreso Internacional SIGEF de Buenos Aires, intentamos asentar la posición epicárea en las nuevas coordenadas surgidas de la importante obra de Zadeh, enunciando el “principio de la simultaneidad gradual” (toda proposición puede ser a la vez verdadera y falsa, a condición de asignar un grado a su verdad y un grado a su falsoedad). Antes y después, un buen número de

científicos han ido colocando, piedra tras piedra, los cimientos de lo que puede ser un nuevo edificio del saber económico. Desde nuestro ámbito de estudio, muchos nombres jalanan este ya fructífero camino y numerosos grupos de investigación pertenecientes a universidades de los cinco continentes han aceptado el testigo y están trabajando en las distintas ramas del árbol de la ciencia, para proporcionar a quienes se hallan inmersos en el intrincado mundo de las optimizaciones, unos elementos teóricos y técnicos susceptibles de permitir una lucha más equilibrada frente a las incertidumbres de hoy y de mañana.

7 Bibliografía

- [1] Dihm, Xuân Bá.: A method for estimating the membership function of a fuzzy set. Revue Busefal, L. S. I. University Paul Sabatier. Toulouse 1984.
- [2] Gil Aluja, J.: "La incertidumbre en la economía y gestión de empresas". Actas del IV Congreso de la Asociación Española sobre Tecnología y Lógica Fuzzy. Blanes, 14 de Septiembre 1994.
- [3] Gil Aluja, J.: La gestión interactiva de los recursos humanos en la incertidumbre. Ed. CEURA. Madrid, 1996.
- [4] Gil Aluja, J.: MAPCLAN. Model for assembling products by means of clans. Proceedings de la Third International Conference on modelling and simulation MS'97. Melbourne, 29 – 31 Octubre 1997.
- [5] Gil Aluja, J.: Invertir en la incertidumbre. Ed. Pirámide. Madrid, 1997.
- [6] Gil Aluja, J.: "Investment in uncertainty". Kluwer Academic Publishers. Dordrecht 1998.
- [7] Gil Aluja, J.: "Elements for a theory of decision in uncertainty". Kluwer Academic Publ. Boston, London, Dordrecht 1999.
- [8] Gil Lafuente, A. M.: Fundamentos de Análisis Financiero. Ed. Ariel. Barcelona, 1993.
- [9] Gil Lafuente, J.: Marketing para el nuevo milenio. Ed. Pirámide. Madrid, 1997.
- [10] Kaufmann, A.: Modèles mathématiques pour la stimulation inventive. Ed. Albin Michel, 1979.
- [11] Kaufmann, A .y Gil Aluja, J.: "Modelos para la investigación de efectos olvidados". Ed. Milladoiro. Santiago de Compostela, 1988

- [12] Kaufmann, A. y Gil Aluja, J.: Técnicas de gestión de empresa. Previsiones, decisiones y estrategias. Ed. Pirámide. Madrid, 1992.
- [13] Kaufmann, A. y Gil Aluja, J.: Técnicas especiales para la gestión de expertos. Ed. Milladoiro. Santiago de Compostela, 1993.
- [14] Lukaciewicz, J.: “O zasadzie wylaczonego srodka”. Przegl'd Filozficzny. 13, 1910.
- [15] Pichat, E.: Algorithm for findig the maximal elements of a finite universal algebra. Inform Proceeding 68. Publ North Holland, 1969
- [16] Saaty, T. L.: Exploring the interface between hierarchies, multiple objectives and fuzzy sets. Fuzzy Sets and Systems. 1978.
- [17] Zadeh, L.: “Fuzzy Sets”. Information and Control, 8 de Junio 1965.

Programación estocástica multiobjetivo

R. Caballero¹, E. Cerdá², M. M. Muñoz¹, L Rey¹

¹ Departamento de Economía Aplicada (Matemáticas). Universidad de Málaga

² Departamento de Fundamentos de Análisis Económico I. Universidad Complutense de Madrid.

1 Introducción

En este trabajo nos centramos en el estudio de problemas de decisión en los que el número de objetivos es múltiple y algunos o todos los parámetros del problema son variables aleatorias con distribución conocida. De esta forma, se relaja la hipótesis frecuente cuando se plantea un modelo de optimización referente a que el objetivo del proceso de decisión puede representarse a través de una única función a optimizar. En general, un problema de programación estocástica multiobjetivo se puede formular como:

$$\begin{aligned} \text{Opt}_x \quad & \tilde{z}(x, \tilde{\xi}) = ((\tilde{z}_1(x, \tilde{\xi}), (\tilde{z}_2(x, \tilde{\xi}), \dots, (\tilde{z}_q(x, \tilde{\xi})) \\ \text{s.a.:} \quad & \tilde{g}_i(x, \tilde{\xi}) \leq 0, \quad i = 1, 2, \dots, m \\ & x \in D \subset \mathbb{R}^n \end{aligned}$$

donde $x \in \mathbb{R}^n$ es el vector de variables de decisión del problema y $\tilde{\xi}$ es un vector aleatorio definido sobre un conjunto $E \subset \mathbb{R}^n$. Suponemos dada la familia F de eventos, es decir, subconjuntos de E , y la distribución de probabilidad P definida sobre F , de manera que para cualquier subconjunto de E , $A \subset E$, $A \subset F$, la probabilidad de A , $P(A)$, es conocida. Además, se mantiene la hipótesis de que la distribución de probabilidad, P , es independiente de las variables de

decisión, x_1, \dots, x_n . Suponemos que las funciones $\tilde{z}_1(x, \tilde{\xi}), \tilde{z}_2(x, \tilde{\xi}), \dots, \tilde{z}_q(x, \tilde{\xi})$ y $\tilde{g}_1(x, \tilde{\xi}), \tilde{g}_2(x, \tilde{\xi}), \dots, \tilde{g}_m(x, \tilde{\xi})$, están definidas en todo el espacio $\mathbb{R}^n \times E$.

A partir de los análisis realizados en artículos previos, sabemos que la resolución de problemas de programación estocástica pasa por la elección de criterios para obtener, a partir del problema estocástico, un problema determinista equivalente cuya solución es considerada solución óptima del problema original. Esto da lugar a que para un mismo problema estocástico se puedan obtener distintas soluciones, una por cada uno de los criterios de obtención del determinista equivalente, y que se puedan considerar todas ellas soluciones óptimas del problema estocástico de partida. La elección de un criterio u otro dependerá de las características del proceso de decisión a partir del cual se genera el problema.

Esta diversidad de criterios es trasladable al caso multiobjetivo, con la dificultad añadida de que el número de objetivos del problema es mayor que uno.

Los estudios realizados hasta ahora en programación estocástica multiobjetivo abordan la resolución del problema planteado desde distintos enfoques. Para clasificar los trabajos realizados distinguiremos entre problemas con variables aleatorias discretas y problemas con variables aleatorias continuas.

Dentro del caso discreto, cabe destacar los trabajos de Ben Abdelaziz (1992) y Ben Abdelaziz, Lang y Nadeau (1994 y 1995) en los que se analiza la obtención de soluciones eficientes de problemas de programación estocástica multiobjetivo con variables aleatorias discretas. En estos trabajos se dan distintos conceptos de eficiencia de problemas de programación estocástica multiobjetivo y el estudio que se realiza es paralelo al análisis de eficiencia en programación estocástica (con un solo objetivo). Por otro lado existen también estudios en los que se aborda la resolución del problema planteado bajo la filosofía de la programación recurso de la programación estocástica. Así, Teghem, Dufrane, Thauvoye y Kunsch (1986) plantean la resolución de problemas de programación estocástica multiobjetivo con parámetros que son variables aleatorias discretas con distribución conocida. Estos autores proponen el método STRANGE, un método interactivo de resolución de problemas con las características descritas. En este método, basado en el método STEM de la programación multiobjetivo, se define una función que penaliza la violación de las restricciones, que se añade al resto de objetivos del problema determinista equivalente. Posteriormente, Urli y Nadeau (1990) proponen el método PROMISE para resolver problemas de programación estocástica multiobjetivo con información incompleta, esto es, para el caso en el que sólo se conoce una medida de tendencia central y una de dispersión de cada uno de los parámetros aleatorios que intervienen en el problema. Al igual que el anterior, este método es interactivo, basado en el método STEM, y se utiliza una función que penaliza la posible violación de las restricciones del problema, que se incorpora como un objetivo más del problema determinista equivalente.

En cuanto al estudio de problemas de programación estocástica multiobjetivo con parámetros que son variables aleatorias continuas existen en la literatura

múltiples trabajos que lo abordan. De entre ellos cabe destacar los trabajos de Stancu-Minasian (1984), Stancu-Minasian y Tigan (1988), Leclercq (1982), Szidarovszky, Gershon y Duckstein (1986), Goicoechea, Hansen y Duckstein (1982) y el libro de Slowinski y Teghem (1990). La mayor parte de los estudios realizados hasta ahora resuelven el problema estocástico multiobjetivo transformando el problema en uno de optimización determinista equivalente y consideran solución del problema estocástico multiobjetivo a la solución del problema transformado. En general, esta transformación consta de dos etapas que constituyen una doble transformación. En una de ellas se transforma el problema estocástico en uno determinista equivalente y en la otra se transforma el problema multiobjetivo en uno de optimización. El orden en que se lleven a cabo estas transformaciones depende del método que se siga para resolver el problema. Así, Stancu-Minasian (1984) plantea que para la resolución de estos problemas se pueden considerar dos posibles formas de abordarlo:

- Transformar el problema de programación estocástica multiobjetivo en un problema de programación estocástica con una única función objetivo y resolver éste mediante alguno de los criterios que se utilizan en programación estocástica con un solo objetivo.
- Transformar el problema de programación estocástica multiobjetivo en un problema multiobjetivo determinista equivalente, fijando un criterio de transformación para cada objetivo estocástico y, posteriormente, buscar soluciones eficientes del problema multiobjetivo determinista obtenido.

En este sentido, Ben Abdelaziz (1992) clasifica los métodos de resolución de los problemas en tres grupos, que denomina enfoques o aproximaciones, para la resolución de problemas de programación estocástica multiobjetivo. Esta clasificación se realiza en función del orden en el que se lleven a cabo las transformaciones anteriormente mencionadas y distingue:

- Enfoque multiobjetivo: En él se transforma el problema en uno multiobjetivo determinista equivalente y posteriormente en uno de optimización.
- Enfoque estocástico: Consiste en reducir el problema a un problema de programación estocástica con una función objetivo y, posteriormente, a uno un problema de optimización determinista equivalente.
- Enfoque interactivo: Se combinan ambos enfoques para obtener, en interacción con el decisor, una solución de compromiso. La siguiente figura muestra las dos transformaciones básicas que generalmente se siguen para resolver problemas de programación estocástica multiobjetivo.

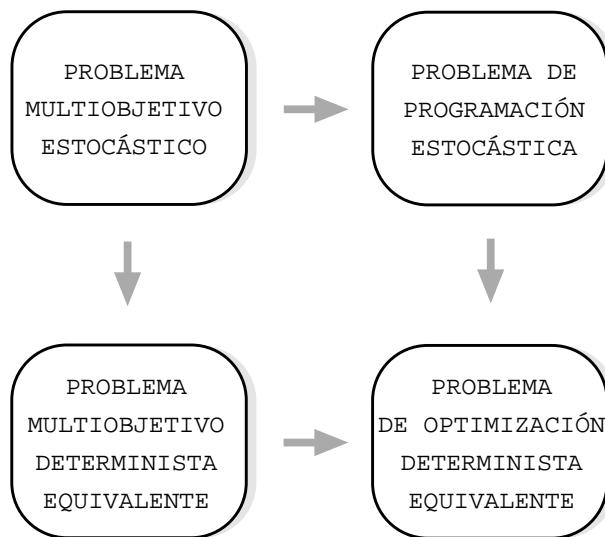


Fig. 1 Transformaciones usuales de un problema multiobjetivo estocástico

A esta clasificación habría que añadir los trabajos realizados para resolver el problema mediante programación estocástica por metas. El primero de ellos se debe a Contini (1968) y posteriormente aparecen otros trabajos debidos a Stancu-Minasian (1984), Stancu-Minasian y Tigan (1988) y Ballester (2001). El análisis de estos trabajos nos ha llevado a considerarlos como un bloque aparte de la clasificación anterior, si bien, algunos de los problemas planteados mediante este enfoque se podrían encuadrar dentro de alguno de los enfoques que distingue Ben Abdelaziz.

En cuanto a los métodos interactivos, hemos de señalar la importancia del método PROTRADE, (Goicoechea, Hansen y Duckstein (1982)), basado en el método STEM, y del método MULT propuesto por Leclerq (1982), y en el que se abordan problemas de programación estocástica con restricciones estocásticas. Para resolver este problema, Leclerq transforma las restricciones estocásticas en restricciones de azar y las considera objetivos adicionales del problema. El problema resultante se resuelve mediante un método interactivo, que propone el autor.

En adelante nos centraremos en el análisis de la resolución de problemas mediante el enfoque multiobjetivo y el enfoque estocástico, básicamente nos centraremos en los distintos conceptos de solución asociados y ofreciendo referencias de las posibles relaciones que pueden existir en la resolución de problemas mediante los mismos. El problema que abordaremos es:

$$\begin{aligned} \text{Min } & \tilde{z}(x, \tilde{\xi}) = ((\tilde{z}_1(x, \tilde{\xi}), (\tilde{z}_2(x, \tilde{\xi}), \dots, (\tilde{z}_q(x, \tilde{\xi})) \\ \text{s.a.: } & x \in D \end{aligned} \tag{PEM}$$

en el que se supone que el conjunto de oportunidades es determinista o bien ha sido transformado en su determinista equivalente siguiendo el criterio de restricciones de azar separadas (obsérvese que la transformación del conjunto de oportunidades es igual en programación estocástica y en programación estocástica multiobjetivo) y se mantiene como criterio de optimización el de mínimo.

2 Enfoque Multiobjetivo

Abordamos la resolución de problemas de programación estocástica multiobjetivo mediante el enfoque multiobjetivo, es decir, siguiendo las siguientes etapas:

Etapa 1: Transformación del problema estocástico multiobjetivo en uno multiobjetivo determinista equivalente, siguiendo algún criterio que se considera apropiado.

Etapa 2: Resolución del problema multiobjetivo determinista obtenido en la etapa anterior, sin considerar el carácter estocástico del problema de partida, salvo por la transformación realizada en la primera etapa.

Evidentemente, existen muchos posibles criterios para llevar a cabo la transformación del problema estocástico multiobjetivo siguiendo las dos etapas descritas. Así, por ejemplo, en un problema de dos objetivos estocásticos es posible que se considere adecuado el criterio valor esperado para transformar el primer objetivo estocástico y el criterio mínimo riesgo para el segundo y, una vez obtenido el problema biobjetivo determinista equivalente, obtener soluciones eficientes, satisfacientes o de compromiso del mismo.

Sin embargo, existen en la literatura conceptos de solución eficiente de problemas de programación estocástica multiobjetivo que se encuadran dentro de este enfoque. Estos conceptos son generalizaciones de los criterios básicos de resolución de problemas de programación estocástica: valor esperado, mínima varianza, valor esperado desviación estándar, mínimo riesgo y Kataoka. La idea básica en todos ellos es elegir un criterio de transformación de los objetivos estocásticos, aplicarlo a cada uno de ellos y construir un problema multiobjetivo determinista equivalente con las transformaciones obtenidas.

La mayor crítica que se realiza a esta forma de resolver problemas de programación estocástica multiobjetivo es que, al aplicar el criterio de obtención del problema determinista equivalente a cada objetivo por separado, puede ocurrir que no se tenga en cuenta la posible dependencia estocástica entre unos objetivos estocásticos y otros, de manera que, en cierto modo, se prima la naturaleza multiobjetivo del problema sobre la naturaleza estocástica. Por otro lado, la mayor ventaja del enfoque multiobjetivo es que es fácilmente aplicable para la resolución de problemas estocásticos multiobjetivo.

En este epígrafe se recogen algunos conceptos de solución eficiente de problemas de programación estocástica multiobjetivo y se analizan las relaciones entre ellos. Como veremos a continuación, estos conceptos de eficiencia están relacionados entre sí.

En Caballero, Cerdá, Muñoz y Rey (2002) se analizan con detenimiento los procesos de obtención del determinista equivalente de un objetivo estocástico mediante algunos de los criterios más utilizados y se realizan comparaciones entre ellos, por lo que en lo que sigue no nos detendremos en ese aspecto concreto del problema que nos ocupa. Tampoco entraremos en la segunda de las etapas anteriormente señaladas, esto es, la resolución del problema multiobjetivo determinista equivalente resultante de aplicar estos criterios.

2.1 Conceptos de solución eficiente

Consideremos el problema de programación estocástica multiobjetivo:

$$\begin{aligned} \text{Min } & \tilde{z}(x, \xi) = (\tilde{z}_1(x, \xi), (\tilde{z}_2(x, \xi), \dots, (\tilde{z}_q(x, \xi)) \\ \text{s.a.: } & x \in D \end{aligned} \tag{PEM}$$

Se supone que el conjunto D de soluciones factibles es no vacío, compacto y convexo.

Existen en la literatura distintos conceptos de solución eficiente de este problema multiobjetivo. Los conceptos que vamos a ver a continuación son aquellos en los que se transforma el problema estocástico multiobjetivo en uno multiobjetivo determinista equivalente, fijando algún criterio de transformación de los existentes en programación estocástica para transformar el objetivo estocástico en determinista. Todos los conceptos que analizamos transforman cada una de las funciones objetivo del problema en determinista mediante un mismo criterio de transformación, que se aplica por separado a cada una de las funciones objetivo del problema.

Eficiencia en Esperanza

La definición de solución eficiente en esperanza no es más que una consecuencia de uno de los métodos más utilizados para resolver el problema (PEM), que consiste en la obtención del determinista equivalente del problema estocástico tomando el valor esperado de cada una de las funciones objetivo del problema, es decir:

$$\begin{aligned} \text{Min}_x & \quad \left(E\{\tilde{z}_1(x, \tilde{\xi})\}, E\{\tilde{z}_2(x, \tilde{\xi})\}, \dots, E\{\tilde{z}_q(x, \tilde{\xi})\} \right) \\ \text{s.a.: } & \quad x \in D \end{aligned} \quad (\text{E})$$

Una vez planteado este problema podemos definir el concepto de solución eficiente valor esperado como sigue:

Definición 1: Solución eficiente valor esperado

Sea $x \in D$. x es eficiente valor esperado del problema de programación estocástica multiobjetivo (PEM) si es solución eficiente en el sentido de Pareto del problema (E).

Denotamos por ε_E al conjunto de soluciones eficientes valor esperado.

A partir del concepto que acabamos de definir, dado un problema de programación estocástica multiobjetivo, podemos obtener soluciones eficientes del mismo sin más que considerar el problema planteado anteriormente. Los objetivos del problema determinista equivalente serán lineales si lo son las funciones objetivo estocásticas y cuadráticos y convexos en el caso de que las funciones estocásticas lo sean. En cuanto a la bondad de este criterio para resolver problemas de programación estocástica multiobjetivo, mantenemos las mismas consideraciones que se hacen en programación estocástica, esto es, consideraremos que el valor esperado no es más que una medida de tendencia central de la variable aleatoria y, en este sentido, la elección de este criterio puede no ser adecuada en determinados casos, puesto que sólo se recogen determinados aspectos del objetivo estocástico.

Al igual que en programación estocástica, la obtención de soluciones eficientes valor esperado del problema (PEM) es posible siempre que se conozca el valor esperado de cada una de las funciones objetivo del problema, aún si se desconoce la distribución de probabilidad de alguna de las funciones objetivo estocásticas. Sin embargo, al igual que en programación estocástica, este criterio recoge sólo el valor esperado de las funciones objetivo estocásticas del problema y, por tanto, recoge sólo una de las características estocásticas de los objetivos aleatorios del problema.

Eficiencia Mínima Varianza

Al igual que hemos definido el concepto de solución eficiente en esperanza, cabe definir el concepto de solución eficiente mínima varianza sin más que plantear el

problema de programación multiobjetivo de minimizar la varianza de cada una de las funciones del problema de programación estocástica multiobjetivo (PEM), es decir:

$$\begin{array}{ll} \text{Min}_x & \left(\text{Var}\{\tilde{z}_1(x, \xi)\}, \text{Var}\{\tilde{z}_2(x, \xi)\}, \dots, \text{Var}\{\tilde{z}_q(x, \xi)\} \right) \\ \text{s.a :} & x \in D \end{array} \quad (\sigma^2)$$

Una vez planteado este problema, que será siempre de mínimo, independientemente de que el problema de partida sea de mínimo (como suponemos en este trabajo) o de máximo, podemos definir el conjunto de soluciones eficientes mínima varianza del problema de programación estocástica multiobjetivo (PEM) de la siguiente forma:

Definición 2. Solución eficiente mínima varianza

$x \in D$ es solución eficiente mínima varianza del problema de programación estocástica multiobjetivo (PEM) si es solución eficiente en el sentido de Pareto del problema (σ^2) .

Denotamos por E_{σ^2} al conjunto de soluciones eficientes del problema σ^2 . Así pues, para la obtención de soluciones eficientes mínima varianza del problema (PEM) hemos de construir un problema de q funciones objetivo, formado por cada una de las varianzas del problema multiobjetivo estocástico de partida. Estas funciones son cuadráticas y convexas tanto en el caso de que los objetivos estocásticos del problema sean lineales como en el caso de que sean cuadráticos. Al igual que con el criterio valor esperado, podremos obtener soluciones eficientes mínima varianza si conocemos la varianza de cada una de las funciones objetivo estocásticas del problema, independientemente de que se conozca o no la distribución de probabilidad de cada una de las funciones objetivo del problema estocástico multiobjetivo de partida. La elección del criterio mínima varianza supone buscar soluciones que acerquen el valor de cada una de las funciones objetivo estocásticas a su valor esperado y, en este sentido se puede considerar un criterio poco arriesgado.

Eficiencia Valor Esperado Desviación Estándar

Al igual que en el caso de optimización estocástica monobjetivo consideramos la transformación del problema estocástico en uno determinista biobjetivo, con objetivos el valor esperado y la desviación estándar de la función objetivo estocástica, planteamos ahora la posibilidad de establecer estos dos criterios de transformación de cada función objetivo estocástica y consideramos el siguiente problema:

$$\begin{array}{ll} \text{Min}_x & \left(E\{\tilde{z}_1(x, \xi)\}, \dots, E\{\tilde{z}_q(x, \xi)\}, \sqrt{\text{Var}\{\tilde{z}_1(x, \xi)\}}, \dots, \sqrt{\text{Var}\{\tilde{z}_q(x, \xi)\}} \right) \\ \text{s.a :} & x \in D \end{array} \quad (E_\sigma)$$

Una vez planteado este problema podemos definir el concepto de solución eficiente valor esperado desviación estándar de la siguiente forma:

Definición 3

$x \in D$ es solución eficiente valor esperado desviación estándar del problema multiobjetivo estocástico (PEM) si es solución eficiente en el sentido de Pareto del problema (E_σ) .

Sea ε_{E_σ} el conjunto de soluciones eficientes valor esperado desviación estándar del problema de programación estocástica multiobjetivo (PEM).

Este concepto ha sido ampliamente utilizado en la literatura para resolver problemas de programación estocástica. En algunos trabajos se ha considerado el criterio de eficiencia valor esperado varianza. Para ello se plantea un problema con $2q$ objetivos que recoge el valor esperado de cada función objetivo y la varianza de cada uno de ellos en lugar de su desviación estándar. En este trabajo hemos preferido definirlo mediante la desviación estándar, dado que de esta forma conseguimos establecer relaciones entre este concepto y otros conceptos de solución eficiente de problemas de programación estocástica multiobjetivo que se definen más adelante. En cualquier caso, es fácil demostrar que el conjunto de soluciones eficientes valor esperado desviación estándar coincide con el de soluciones eficientes valor esperado varianza.

Las funciones objetivo del problema que se genera para la obtención de soluciones eficientes valor esperado desviación estándar son funciones convexas, para problemas estocásticos multiobjetivo lineales.

Eficiencia Mínimo Riesgo de niveles u_1, u_2, \dots, u_q .

Definido por Stancu-Minasian y Tigan (1984), este concepto de solución considera soluciones eficientes del problema de programación estocástica multiobjetivo (PEM) a las soluciones eficientes del problema multiobjetivo determinista que se obtiene al aplicar a cada una de las funciones objetivo del problema el criterio mínimo riesgo. Para aplicar este criterio hemos de fijar un nivel de aspiración a alcanzar para cada uno de los objetivos estocásticos, $u_1, u_2, \dots, u_q, u_k \in \mathbb{R}$, $k = 1, 2, \dots, q$. Una vez fijados estos valores, el problema mínimo riesgo, equivalente determinista del problema (PEM) consiste en maximizar la probabilidad de que cada uno de los objetivos estocásticos no supere el nivel de aspiración fijado, es decir:

$$\begin{aligned} \text{Min}_x & \quad \left(P(\tilde{z}_1(x, \xi) \leq u_1), \dots, P(\tilde{z}_q(x, \xi) \leq u_q) \right) \\ \text{s.a : } & \quad x \in D \end{aligned} \quad (\text{MR}(\mathbf{u}))$$

Una vez planteado este problema, Stancu-Minasian y Tigan (1984) definen el concepto de solución vectorial mínimo riesgo de nivel u del problema (PEM) de la siguiente forma:

Definición 4. Solución eficiente mínimo riesgo de niveles u_1, u_2, \dots, u_q

$x \in D$ es solución vectorial mínimo riesgo de nivel \mathbf{u} si es solución eficiente en el sentido de Pareto del problema (MR(\mathbf{u})).

En adelante nos referiremos a estas soluciones como soluciones eficientes mínimo riesgo de niveles u_1, u_2, \dots, u_q . Denotamos por $\varepsilon_{mr}(\mathbf{u})$ al conjunto de soluciones eficientes del problema (MR(\mathbf{u})).

Obsérvese que el problema multiobjetivo determinista que se obtiene al aplicar este criterio, (MR(\mathbf{u})), depende, en general, del vector de niveles de aspiración fijado, \mathbf{u} , de tal forma que, podemos afirmar que en general, dados $\mathbf{u}, \mathbf{u}' \in \mathbb{R}^q$, si $\mathbf{u}=\mathbf{u}'$ entonces los conjuntos de soluciones eficientes mínimo riesgo de niveles \mathbf{u} y \mathbf{u}' serán distintos: $\varepsilon_{mr}(\mathbf{u}) = \varepsilon_{mr}(\mathbf{u}')$.

Puesto que para aplicar este criterio de eficiencia, hemos de fijar un nivel de aspiración, $u_k \in \mathbb{R}$, para cada uno de los objetivos estocásticos del problema, es necesaria la intervención del decisor para generar soluciones eficientes mínimo riesgo. Además, dado que en el problema que se genera interviene la función de distribución de cada una de las funciones objetivo estocásticas, la aplicación de este criterio se centra en los casos ya conocidos en programación estocástica: problemas estocásticos multiobjetivo lineales con hipótesis de normalidad o de aleatoriedad simple. En ambos casos, los objetivos del problema determinista equivalentes son fraccionales. Puesto que esto sale del objetivo de este trabajo, no plantearemos los problemas que se generan. Por otro lado, en los casos en los que desconocemos la función de distribución de los objetivos estocásticos del problema, podemos aplicar la desigualdad de Cantelli a la función de distribución y obtener una cota inferior para la misma. Así, si sustituimos las funciones de distribución por estas cotas en el problema (MR(\mathbf{u})) obtenemos el problema:

$$\begin{aligned} \text{Max}_x & \left(\frac{(u_1 - E\{\tilde{z}_1(x, \xi)\})^2}{Var(\tilde{z}_1(x, \xi)) + (u_1 - E\{\tilde{z}_1(x, \xi)\})^2}, \dots, \frac{(u_q - E\{\tilde{z}_q(x, \xi)\})^2}{Var(\tilde{z}_q(x, \xi)) + (u_q - E\{\tilde{z}_q(x, \xi)\})^2} \right) \\ \text{s.a.: } & E\{\tilde{z}_k(x, \xi)\} \leq u_k, \quad k = 1, 2, \dots, q \end{aligned} \quad (\text{AMR}(\mathbf{u}))$$

$$x \in D$$

Evidentemente, el conjunto de soluciones eficientes del problema (AMR(\mathbf{u})), que denotamos por $\varepsilon_{AMR}(\mathbf{u})$, no coincide, en general, con el conjunto de soluciones eficientes del problema (MR(\mathbf{u})), $\varepsilon_{AMR}(\mathbf{u}) = \varepsilon_{MR}(\mathbf{u})$, y sólo puede tomarse como aproximación del mismo.

Eficiencia con Probabilidades $\beta_1, \beta_2, \dots, \beta_q$

Finalmente, consideramos un concepto de solución que se basa en una idea expresada en Goicoechea, Hansen y Duckstein (1982), el concepto de solución estocástica no dominada de nivel β .

Definición 5: Solución estocástica no dominada de nivel β

Sea $z_k(x)$ un valor perteneciente al rango o soporte de la variable aleatoria $\tilde{z}_k(x)$, $k = 1, 2, \dots, q$. Se dice que $x \in D$ es solución estocástica no dominada de nivel $\beta \in (0, 1)$ si:

- (i) $P\{\tilde{z}_k(x, \tilde{\xi}) \leq z_k(x)\} = \beta$ para todo $k \in 1, 2, \dots, q$
- (ii) no existe ningún vector $y \in D$ tal que: $P\{\tilde{z}_k(x, \tilde{\xi}) \leq z_k(x)\} = \beta$ para todo $k \in 1, 2, \dots, q$ existe $l \in 1, 2, \dots, q$ tal que $z_1(y) < z_1(x), z_k(y) \leq z_k(x)$, para todo $k \in 1, 2, \dots, q, k \neq 1$

A partir de esta definición, dado el problema de programación estocástica multiobjetivo (PEM), si aplicamos el criterio de Kataoka a cada una de las funciones objetivo estocásticas del problema para una probabilidad β , el problema que se genera es:

$$\begin{aligned} \text{Min}_{x,u} \quad & \mathbf{u} = (u_1, u_2, \dots, u_q) \\ \text{s.a :} \quad & P\{\tilde{z}_k(x, \tilde{\xi}) \leq u_k\} = \beta, \quad k = 1, 2, \dots, q \\ & x \in D \end{aligned}$$

y tenemos que el conjunto de soluciones eficientes de este problema es el conjunto de soluciones no dominadas de nivel β que se ha definido anteriormente, puesto que para cada $k \in 1, 2, \dots, q$ la variable u_k será una función $z_k(x)$ que se obtiene a partir de la igualdad $P\{\tilde{z}_k(x, \tilde{\xi}) \leq u_k\} = \beta$. Así, tenemos que el conjunto de soluciones no dominadas de nivel β se obtiene a partir de aplicar el criterio de Kataoka a cada una de las funciones objetivo del problema multiobjetivo estocástico, fijando el mismo nivel de probabilidad para todas las funciones estocásticas. A partir de aquí, cabe la posibilidad de generalizar este concepto considerando distintos niveles de probabilidad para las funciones objetivo del problema, sin más que plantearnos el problema:

$$\begin{aligned} \text{Min}_{x,u} \quad & \mathbf{u} = (u_1, u_2, \dots, u_q) \\ \text{s.a :} \quad & P\{\tilde{z}_k(x, \tilde{\xi}) \leq u_k\} = \beta_k, \quad k = 1, 2, \dots, q \\ & x \in D \end{aligned} \tag{K(\beta)}$$

Una vez planteado el problema, definimos el concepto de solución eficiente con probabilidades $\beta_1, \beta_2, \dots, \beta_q$ de la siguiente forma:

Definición 6. Solución eficiente con probabilidades $\beta_1, \beta_2, \dots, \beta_q$

Sea $x \in D$. Se dice que x es solución eficiente con probabilidades $\beta_1, \beta_2, \dots, \beta_q$ si existe un $u \in \mathbb{R}^n$ tal que $(\mathbf{x}^t, \mathbf{u}^t)^t$ es solución eficiente del problema $(K(\beta))$.

Denotamos por $\varepsilon_k(\beta) \subset \mathbb{R}^n$ al conjunto de soluciones eficientes con probabilidades $\beta_1, \beta_2, \dots, \beta_q$.

Obsérvese que el concepto de solución eficiente con probabilidades $\beta_1, \beta_2, \dots, \beta_q$ se define para los vectores \mathbf{x} , aunque las soluciones del problema $(K(\beta))$ sean vectores $(\mathbf{x}^t, \mathbf{u}^t)^t \in \mathbb{R}^{n+q}$.

Al igual que en el caso mínimo riesgo, este concepto de eficiencia va asociado a unos niveles de probabilidad fijados a priori, con lo cual, el problema multiobjetivo determinista mediante el que se obtienen soluciones eficientes con probabilidades $\beta_1, \beta_2, \dots, \beta_q$, $(K(\beta))$, depende, en general, del vector de probabilidades fijado, $\beta = (\beta_1, \beta_2, \dots, \beta_q)^t$, de tal forma que, podemos afirmar que en general, dados, $\beta, \beta' \in \mathbb{R}^q$, si $\beta = \beta'$ entonces el conjunto de soluciones eficientes para β es distinto del que se obtiene para $\beta' : \varepsilon_k(\beta) \neq \varepsilon_k(\beta')$.

De nuevo, para poder aplicar este criterio de eficiencia, hemos de fijar una probabilidad $\beta \in (0, 1)$, para cada uno de los objetivos estocásticos del problema. Esto hace necesaria la intervención del decisor para generar estas soluciones eficientes. Además, en el problema interviene la función de distribución de cada una de las funciones objetivo estocásticas. Consideramos los mismos casos anteriores citados: funciones objetivo lineales bajo la hipótesis de normalidad o bajo la hipótesis de aleatoriedad simple. En ambos casos es posible obtener soluciones eficientes del problema de programación estocástica multiobjetivo y los problemas resultantes son, bajo determinadas hipótesis, problemas convexos.

Por otro lado, en los casos en los que desconocemos la función de distribución de los objetivos estocásticos del problema, podemos aplicar la desigualdad de Cantelli a la función de distribución y obtener una cota inferior para la misma. Así, si sustituimos las funciones de distribución por estas cotas en el problema $(K(\beta))$ obtenemos el problema:

$$\begin{aligned} \text{Min}_{x,u} \quad & u = (u_1, u_2, \dots, u_q) \\ \text{s.a. :} \quad & E\{\tilde{z}_k(x, \tilde{\xi})\} + \sqrt{\frac{\beta_k}{1-\beta_k}} \sqrt{\text{Var}\{\tilde{z}_k(x, \tilde{\xi})\}}, \quad k = 1, \dots, q \\ & x \in D \subseteq \mathbb{R}^n \end{aligned} \quad (\text{AK}(\beta))$$

Al igual que en el caso mínimo riesgo, el conjunto de soluciones eficientes del problema $(\text{AK}(\beta))$, que denotamos por $(\varepsilon_{\text{AK}}(\beta))$ será distinto del conjunto de soluciones eficientes del problema $K(\beta)$, $(\varepsilon_{\text{AK}}(\beta)) \neq (\varepsilon_K(\beta))$, pero, podemos tomar este conjunto como aproximación.

2.2 Relaciones entre los diferentes conceptos de solución eficiente definidos

En Caballero, Cerdá, Muñoz y Rey (2000) se obtiene una relación entre las soluciones eficientes mínimo riesgo de niveles u_1, u_2, \dots, u_q y las soluciones eficientes con probabilidades $\beta_1, \beta_2, \dots, \beta_q$: Consideremos los problemas $MR(\mathbf{u})$ y $K(\beta)$ definidos anteriormente. Supongamos que los conjuntos factibles de ambos

problemas $D \subset \mathbb{R}^n$ y

$$(x^t, u^t) \in D \times \mathbb{R}^q / P\{\tilde{z}_k(x, \tilde{\xi}) \leq u_k\} = \beta_k, \quad k = 1, 2, \dots, q$$

son cerrados, acotados y no vacíos, y por tanto ambos problemas tienen soluciones eficientes. También suponemos que para cada $k \in 1, 2, \dots, q$, y cada $x \in D$, la función de distribución de la variable aleatoria $\tilde{z}_k(x, \tilde{\xi})$ es continua y estrictamente creciente. Estas hipótesis implican que para cada probabilidad β_k , existe un único número real u_k , tal que $P\{\tilde{z}_k(x, \tilde{\xi}) \leq u_k\} = \beta_k$.

El siguiente Teorema, cuya demostración se puede ver en la referencia dada, relaciona los conjuntos de soluciones eficientes $\varepsilon_{mr}(\mathbf{u})$ (del problema $MR(\mathbf{u})$) y $\varepsilon_k(\beta)$ (del problema $K(\beta)$).

Teorema 1: Supongamos que la función de distribución de la variable aleatoria $\tilde{z}_k(x, \tilde{\xi})$ es continua y estrictamente creciente. Entonces \mathbf{x} es una solución eficiente del Problema $MR(\mathbf{u})$ si y sólo si $(\mathbf{x}^t, \mathbf{u}^t)^t$ es una solución eficiente del Problema $K(\beta)$, con \mathbf{u} y β tales que

$$P\{\tilde{z}_k(x, \tilde{\xi}) \leq u_k\} = \beta_k, \quad \forall k \in 1, 2, \dots, q$$

Corolario 1:

$$\bigcup_{u \in \mathbb{R}^q} (E_{MR}(u)) = \bigcup_{\beta \in B} (E_k(\beta))$$

con $B = \{(\beta_1, \beta_2, \dots, \beta_q)^t \in \mathbb{R}^q \mid \beta_k \in (0, 1), k = 1, 2, \dots, q\}$.

De los resultados obtenidos se ve que las uniones de los conjuntos eficientes de ambos problemas coinciden. Además, si $x \in D$ es una solución eficiente del Problema $K(\beta)$, para ciertas probabilidades dadas $\beta = (\beta_1, \beta_2, \dots, \beta_q)^t$, por el Teorema 1 sabemos que es también una solución eficiente mínimo riesgo de niveles u_1, u_2, \dots, u_q , manteniendo para los niveles de satisfacción y probabilidades la relación que aparece en el enunciado del Teorema, y viceversa. Este resultado permite realizar el análisis de estas soluciones eficientes por uno cualquiera de los dos conceptos y, por el Teorema 1, obtener el nivel o probabilidad para el cual es eficiente de acuerdo con el otro.

En Caballero, Cerdá, Muñoz, Rey y Stancu-Minasian (2001) se obtienen las siguientes relaciones entre soluciones eficientes valor esperado, soluciones eficientes mínima varianza y soluciones eficientes valor esperado desviación estándar:

Teorema 2:

Se considera el problema (PEM). Sean los conjuntos de soluciones eficientes valor esperado (E_E), mínima varianza (E_{σ^2}), y valor esperado desviación estándar ($E_{E\sigma}$). Sean $E_E^d, E_{\sigma^2}^d, E_{\sigma}^d$, los conjuntos de soluciones débilmente eficientes asociados a los problemas $(E_E^d), (E_{\sigma^2}^d), (E_{\sigma}^d)$ respectivamente.

Entonces,

1. $E_E \cap E_{\sigma^2} \subset E_{E_\sigma}$. Cada solución que es a la vez valor esperado eficiente y mínima varianza eficiente es también solución valor esperado desviación estándar.
2. $E_E \cup E_{\sigma^2} \subset E_{E_\sigma}^d$. Cada solución eficiente valor esperado o mínima varianza es solución débilmente eficiente valor esperado desviación estándar.
3. $E_E^d \cap E_{\sigma^2}^d \subset E_{E_\sigma}^d$. El conjunto de soluciones débilmente eficientes valor esperado desviación estándar incluye a la unión de los conjuntos de soluciones débilmente eficientes valor esperado y mínima varianza.—

Asimismo, en el mismo artículo se obtienen relaciones entre soluciones eficientes valor esperado desviación estándar y soluciones eficientes con probabilidades $\beta_1, \beta_2, \dots, \beta_q$ cuando las funciones objetivo son lineales y el vector de parámetros aleatorios sigue una distribución normal o bien verifica la hipótesis de aleatoriedad simple. Estas relaciones y la demostración del Teorema 2 se encuentran en Caballero, Cerdá, Muñoz, Rey y Stancu-Minasian (2001).

3 Enfoque Estocástico

Nos centramos ahora en la resolución de problemas de programación estocástica multiobjetivo mediante el enfoque estocástico. Como ya se ha comentado anteriormente, en la clasificación que realiza Ben Abdelaziz (1992) de los métodos de resolución de problemas de programación estocástica multiobjetivo agrupa, en lo que denomina enfoque estocástico, a todos los métodos de resolución de estos problemas en los que se siguen las dos etapas siguientes:

Etapa 1: Transformación del problema estocástico multiobjetivo en un problema de programación estocástica con una única función objetivo siguiendo alguno de los criterios existentes para ello en programación multiobjetivo.

- **Etapa 2:** Resolución del problema de programación estocástica obtenido en la etapa uno mediante algún método de resolución de programación estocástica. La solución obtenida en esta etapa es considerada, en estos métodos, solución al problema estocástico multiobjetivo de partida. Abordaremos ahora la resolución de problemas de programación estocástica con objetivos múltiples siguiendo estas dos etapas, tal y como se describe en la Figura 2.

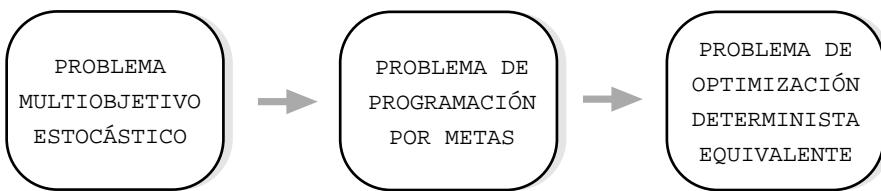


Fig. 2 Transformación de un problema multiobjetivo estocástico mediante el enfoque estocástico

Para resolver problemas de programación estocástica multiobjetivo siguiendo los pasos descritos, pueden aplicarse distintos criterios de transformación de los q objetivos estocásticos del problema en la etapa uno. Hecho ésto se obtiene un problema de programación estocástica con una única función objetivo estocástica y, para su resolución, es posible aplicar también distintos criterios. Por tanto, el proceso de resolución de problemas de programación estocástica multiobjetivo mediante este enfoque pasa por un proceso de decisión previo en el que se ha de elegir la manera de transformar los objetivos del problema para obtener un problema de programación estocástica y en la elección de un criterio de resolución de este problema.

De todo lo anterior se desprende que la diversidad de criterios de transformación del problema en las dos etapas de resolución, dará lugar, en general, a todo un conjunto de posibles soluciones al problema de programación estocástica multiobjetivo que, de acuerdo con lo descrito, no son comparables.

En este trabajo hemos optado por fijar un único criterio de transformación en la etapa uno del proceso y, al llegar a la etapa dos, resolver el problema obtenido mediante algunos de los criterios para la programación estocástica. Así, dado el problema de programación estocástica multiobjetivo, hemos considerado la aplicación del método de la ponderación al mismo. Hecho ésto, nos hemos planteado su resolución mediante los criterios valor esperado, mínima varianza, mínimo riesgo y Kataoka. No hemos considerado la resolución del problema ponderado mediante el criterio de eficiencia valor esperado desviación estándar, puesto que consideramos poco lógico transformar el problema multiobjetivo en un problema ponderado y, posteriormente, considerar un problema bicriterio para resolverlo. Para los criterios considerados analizaremos el problema de optimización determinista equivalente que se obtiene tras aplicar las dos etapas e intentamos establecer relaciones entre la solución obtenida mediante este proceso y los conceptos de solución eficiente de problemas de programación estocástica multiobjetivo definidos en la sección anterior.

Antes de abordar la resolución de este problema mediante el método de las ponderaciones consideramos importante señalar que este mismo estudio ha sido abordado por Stancu-Minasian (1984) para problemas con funciones objetivo estocásticas lineales que verifican la hipótesis de aleatoriedad simple. En ese trabajo, Stancu-Minasian aplica el criterio de Chebychev al problema multiobjetivo

estocástico y, posteriormente, aplica al problema resultante el criterio mínimo riesgo (véase Stancu-Minasian (1984), pág. 223-225).

Consideremos el problema de programación estocástica multiobjetivo (PEM). Si aplicamos al problema anterior el método de las ponderaciones de la programación multiobjetivo, fijando unos pesos no negativos, $\mu_1, \mu_2, \dots, \mu_q, \mu_k \geq 0$, para todo $k \in 1, 2, \dots, q$, obtenemos el problema de programación estocástica:

$$\begin{aligned} \text{Min}_x \quad & \tilde{f}(x, \tilde{\xi}) = \sum_{k=1}^q \mu_k^2 \tilde{z}(x, \tilde{\xi}) \\ \text{s.a.:} \quad & x \in D \end{aligned} \quad (\text{AE})$$

Para resolver este problema consideramos a continuación los criterios valor esperado, mínima varianza, mínimo riesgo y de Kataoka. Dividimos nuestro estudio en función del criterio que apliquemos.

Criterio Valor Esperado

Sea el problema (AE) y consideremos la resolución del mismo mediante el criterio valor esperado, que nos lleva a obtener el siguiente problema determinista equivalente:

$$\begin{aligned} \text{Min}_x \quad & \bar{f}(x) = \sum_{k=1}^q \mu_k^2 E\{\tilde{z}(x, \tilde{\xi})\} = \sum_{k=1}^q \mu_k \bar{z}_k(x) \\ \text{s.a.:} \quad & x \in D \end{aligned} \quad (\text{AEE})$$

Por tanto, si se aplica el criterio valor esperado al problema (AE), el problema resultante minimiza una combinación lineal de los valores esperados de las funciones objetivo estocásticas del problema original, y los coeficientes de tal combinación lineal no son más que los pesos asignados a los objetivos estocásticos en la primera etapa de la resolución del problema. En otras palabras, el problema que se obtiene es el mismo que resulta trasformando el problema original multiobjetivo en otro multiobjetivo en que cada objetivo es el determinista equivalente que se obtiene al aplicar el criterio valor esperado a cada uno de los objetivos estocásticos del problema (PEM) y aplicando entonces el método de la ponderación para obtener las soluciones eficientes valor esperado.

Criterio mínima varianza

Sea el problema (AE) y consideremos la resolución del mismo mediante el criterio mínima varianza, que nos lleva a obtener el siguiente problema determinista equivalente:

$$\begin{aligned} \text{Min}_x \quad & \sigma^2(x) = \sum_{k=1}^q \mu_k \sigma^2(x) + 2 \sum_{k,s=1, k < s}^q \mu_k \mu_s \sigma_{ks}(x) \\ \text{s.a.:} \quad & x \in D \end{aligned} \quad (\text{AEV})$$

La relación entre el problema estocástico obtenido al utilizar el método de ponderación y el problema determinista equivalente multiobjetivo que se obtiene al aplicar a cada objetivo estocástico el criterio de mínima varianza no es tan directa como en el caso anterior. Distingamos dos casos:

- (i) Las covarianzas de las funciones objetivo son cero. Es decir:

$$\sigma_{ks}(x) = 0 \text{ para cada } k, s \in \{1, 2, \dots, q\} \text{ con } k = s \text{ y para cada } x \in D.$$

Entonces el problema (AEV) resultante de aplicar el criterio de mínima varianza al problema ponderado (AE) es:

$$\begin{aligned} \text{Min}_x \quad & \sigma^2(x) = \sum_{k=1}^q \mu_k \sigma^2(x) \\ \text{s.a.:} \quad & x \in D \end{aligned}$$

y este problema es el mismo que se habría obtenido resolviendo el problema estocástico multiobjetivo mediante el enfoque multiobjetivo, es decir, aplicando a cada uno de los objetivos estocásticos el criterio de mínima varianza y a continuación el método de ponderación al determinista equivalente multiobjetivo para obtener sus soluciones eficientes.

- (ii) Las covarianzas de las funciones objetivo no son todas cero. En este caso la solución obtenida al resolver el problema (AEV) no tiene por qué ser eficiente mínima varianza al aplicar el enfoque multiobjetivo, como se ve en un ejemplo que se presenta en Caballero, Cerdá, Muñoz y Rey (2004).

Criterios mínimo riesgo y Kataoka.

Para aplicar el criterio de mínimo riesgo al problema (AE) debemos fijar un valor u (nivel de aspiración de la función objetivo del problema) y resolver:

$$\begin{aligned} \text{Max}_x \quad & P \left\{ \sum_{k=1}^q \mu_k \tilde{z}(x, \tilde{\xi}) \leq u \right\} \\ \text{s.a.:} \quad & x \in D \end{aligned}$$

Dado que $\bar{f}(x, \tilde{\xi}) = \sum_{k=1}^q \mu_k \tilde{z}(x, \tilde{\xi})$ es una función de las q funciones objetivo para el problema, la elección del valor u no es nada sencilla. Nótese que este nivel debe ser fijado para la variable aleatoria $\bar{f}(x, \tilde{\xi}) = \sum_{k=1}^q \mu_k \tilde{z}(x, \tilde{\xi})$ que ha sido construida a partir del problema original para resolverlo. Por consiguiente, el valor u , que en programación estocástica es el nivel de aspiración del decisor para el objetivo estocástico, no tiene ahora tal significado. Se puede determinar el nivel de aspiración u_k para cada función objetivo y tomar $u = \sum_{k=1}^q \mu_k u_k$ y entonces el problema determinístico equivalente es:

$$\begin{aligned} \text{Max}_x \quad & P \left\{ \sum_{k=1}^q \mu_k \tilde{z}(x, \tilde{\xi}) \leq \sum_{k=1}^q \mu_k u_k \right\} \\ \text{s.a.:} \quad & x \in D \end{aligned} \tag{AEMR(u)}$$

Por otra parte, si aplicamos el criterio de Kataoka al problema (AE), el problema determinista equivalente generado es el siguiente:

$$\begin{aligned} \text{Min}_{x,u} \quad & u \\ \text{s.a.:} \quad & P \left\{ \sum_{k=1}^q \mu_k \tilde{z}(x, \tilde{\xi}) \leq u \right\} = \beta \\ & x \in D \end{aligned} \tag{AEK(\beta)}$$

Al igual que hemos hecho en el criterio de mínimo riesgo, podemos tomar $\beta = \sum_{k=1}^q \mu_k \beta_k$ considerando la posibilidad de pedir al decisor que fije una probabilidad β_k para cada uno de los objetivos estocásticos.

Una vez establecida una posible forma de fijar los niveles de aspiración y las probabilidades en los dos criterios de máxima probabilidad, para plantear las posibles relaciones que se obtienen en los dos enfoques, señalemos que se puede aplicar el Teorema 1, tomando $k = 1$, obteniendo relaciones entre los problemas deterministas equivalentes (AEMR(u)) y (AEK(\beta)).

Para profundizar en el enfoque estocástico y en las relaciones entre las soluciones por los enfoques estocástico y multiobjetivo véase Caballero, Cerdá, Muñoz y Rey (2004).

4 Conclusiones

Este trabajo se centra en el análisis de los distintos conceptos de solución eficiente que aparecen en la programación estocástica multiobjetivo y en la transformación del problema estocástico en uno determinista equivalente. A la hora de abordar esta transformación, se observa que, ahora, ésta es "doble", y consiste, básicamente, en pasar del problema de programación estocástica multiobjetivo a uno determinista de un único objetivo. Basándonos en estudios previos, denominamos *enfoque multiobjetivo* a la resolución del problema transformando, en una primera etapa el problema multiobjetivo estocástico en uno determinista equivalente y, posteriormente, centrándonos en la obtención de soluciones eficientes de este último, y *enfoque estocástico* al conjunto de técnicas que, en una primera etapa, transforman el problema en uno de programación estocástica con un solo objetivo y, posteriormente, resuelven el problema estocástico obtenido mediante cualquier técnica.

Evidentemente, la existencia de distintos criterios de transformación del problema en cada una de las etapas descritas da lugar a que se pueda obtener todo un abanico de posibles problemas equivalentes del problema de partida. Una vez

definidos estos conceptos, se analizan en este trabajo las posibles relaciones entre los mismos.

5 Bibliografía

- [1] Ballester, E. (2001). "Stochastic Goal Programming: A Mean-Variance Approach". European Journal of Operational Research, 131, 476-481.
- [2] Ben Abdelaziz, F., (1992). L'efficacité en programmation multi-objectifs stochastique. Ph. D. Thesis, Université de Laval, Québec.
- [3] Ben Abdelaziz, F., Lang P. and Nadeau, R., (1994) "Pointwise Efficiency in Multiobjective Stochastic Linear Programming". Journal of the Operational Research Society, 45, 11, 1324-1334.
- [4] Ben Abdelaziz, F., Lang P. and Nadeau, R., (1995) "Distributional Efficiency in Multiobjective Linear Programming". European J. Operational Research, 85, 399-415.
- [5] Caballero, R., Cerdá, E., Muñoz, M. M. y Rey, L., (2000). "Relations among Several Efficiency Concepts in Stochastic Multiple Objective Programming". En: Y. Y. Haimes and R. Steuer (Eds), Research and Practice in Multiple Criteria Decision Making, Lectures Notes in Economics and Mathematical Systems, 487, 57-68, Springer-Verlag, Berlin, Germany.
- [6] Caballero, R., Cerdá, E., Muñoz, M. M. y Rey, L., (2002). "Analysis and Comparisons of Some Solution Concepts for Stochastic Programming Problems". Top , 10(1) 101-124.
- [7] Caballero, R., Cerdá, E., Muñoz, M. M. y Rey, L., (2004). "Stochastic Approach Versus Multiobjective Approach for Obtaining Efficient Solutions in Stochastic Multiobjective Programming Problems". European Journal of Operational Research. Por aparecer.
- [8] Caballero, R., Cerdá, E., Muñoz, M.M., Rey, L., Stancu Minasian, I. M., (2001). "Efficient solution concepts and their relations in stochastic multiobjective programming". Journal of Optimization, Theory and Applications, 110 (1) 53-74.
- [9] Contini B. (1968). "A Stochastic Approach to Goal Programming". Operations Research, 16, 3, 576-586.
- [10] Goicoechea, A., Hansen, D. R., Duckstein, L., (1982). Multiobjective decision analysis with engineering and business applications. John Wiley and Sons, New York.

- [11] Leclercq J.P. (1982). "Stochastic Programming: An Interactive Multicriteria Approach". European Journal of Operational Research, 10, 33-41.
- [12] Slowinski, R., Teghem, J. (Eds.), (1990). Stochastic versus fuzzy approaches to multiobjective mathematical programming under uncertainty. Kluwer Academic Publishers, Dordrecht.
- [13] Stancu-Minasian, I. M., (1984). Stochastic programming with multiple objective functions. D. Reidel Publishing Company, Dordrecht.
- [14] Stancu-Minasian, I., Tigan, S., (1984). "The vectorial minimum risk problem". Proceedings of the Colloquium on Approximation and Optimization, Cluj-Napoca, 321-328.
- [15] Stancu-Minasian, I. y Tigan, S. (1988). "A Stochastic Approach to Some Linear Fractional Goal Programming Problems". Kybernetika, 24, 2, 139-149.
- [16] Szidarovszky, F., Gershon, M. y Duckstein, L. (1986). Techniques for Multiobjective Decision Making in Systems Management. Elsevier. Amsterdam.
- [17] Teghem, J., Dufrane, D., Thauvoye, M. And Kunsch, P, (1986). "STRANGE: An Interactive Method for Multi-Objective Linear Programming Under Uncertainty". European J. Operational Research, 26, 1, 65-82.
- [18] Urli, B. and Nadeau, R., (1990) "Stochastic MOLP with Incomplete Information: An Interactive Approach with Recourse". Journal of the Operational Research Society, 41, 12, 1143-1152.

Métodos y Modelos de Programación Lineal Borrosa

J. M. Cadenas^a, VJ. L. Verdegay^b

^a Dpto. Ingeniería de la Información y las Comunicaciones. Facultad de Informática. Universidad de Murcia. Campus de Espinardo. 30071-Espinardo. Murcia. e-mail: jcadenas@dif.um.es

^b Dpto. Ciencias de la Computación e Inteligencia Artificial. ETSII. Universidad de Granada. C/Periodista Daniel Saucedo Aranda s/n, 18071. Granada. e-mail: verdegay@decsai.ugr.es

1 Introducción

Desde que en 1965, [26], el profesor estadounidense, aunque de origen iraní, Lotfi A. Zadeh introdujera el concepto de conjunto borroso permitiendo la pertenencia de un elemento a un conjunto de forma gradual, y no de manera absoluta como establece la teoría conjuntista clásica, es decir, admitiendo pertenencias valoradas en el intervalo $[0,1]$ en lugar de en el conjunto $\{0,1\}$, las aplicaciones y desarrollos basados en este sencillo concepto han evolucionado de tal modo que, hoy en día, es prácticamente imposible calcular el volumen de negocio que generan en todo el mundo, pudiendo encontrar productos cuyo funcionamiento está directamente basado en dicho concepto desde los más usuales electrodomésticos, lavadoras, microondas, cámaras fotográficas, ..., hasta los más sofisticados sistemas, frenado de trenes, control de hornos, navegación, ...

La necesidad de encontrar la solución óptima, o la mejor solución entre las disponibles, en un problema correctamente planteado es por lo que se estudian las teorías, y se proponen metodologías adecuadas al campo científico en el que surge la cuestión que se ha de resolver. Desde un punto de vista más concreto, pero

muy general, una importante clase de problemas son los conocidos con el nombre de problemas de optimización, habitualmente asociados a tener que encontrar el máximo, o el mínimo, valor que una determinada función puede alcanzar en un cierto conjunto previamente especificado. Todo lo relativo a estos problemas se enmarca dentro del cuerpo doctrinal denominado Programación Matemática, que incluye una enorme variedad de situaciones, según que se consideren casos lineales, no lineales, aleatoriedad, un solo decisor o varios decisores, etc.

Entre todos los modelos que se incluyen en la Programación Matemática, el más y mejor estudiado, así como el que ha probado tener unas repercusiones prácticas más importantes, es el correspondiente al caso lineal uni-objetivo, tema del que se ocupa la Programación Lineal. Los métodos y modelos de la Programación Lineal tienen relevantes aplicaciones en las diferentes áreas de las Ingenierías, la Economía, las Matemáticas, la Investigación Operativa, la Inteligencia Artificial, y demás disciplinas más o menos relacionadas con la optimización, y constituyen un sustrato teórico más que adecuado para abordar de un modo elegante y eficiente situaciones muy complejas.

Aunque como se ha dicho, los modelos y técnicas de la Programación Lineal son los más y mejor estudiados, es justamente por ese motivo, junto con la mencionada eficiencia y elegancia que los caracteriza, por lo que son fácilmente adaptables a nuevos contextos tecnológicos, lo que impulsa a su vez el que sean protagonistas en los más recientes desarrollos científicos, como es el caso de su incorporación e implementación en los sistemas generadores de modelos de los Sistemas de Ayuda a la Decisión. De este modo la Programación Lineal aparece entronizada en una de las más prometedoras líneas de desarrollo en el ámbito de la Inteligencia Artificial, y consiguientemente, y a pesar de sus más de cincuenta años de vida, a la vanguardia del avance científico.

En ese contexto de Sistemas de Ayuda a la Decisión, dentro del marco de la Inteligencia Artificial, lo que principalmente se persigue es disponer de sistemas automáticos que, desde implementaciones que faciliten actuaciones lo más cercanas a la realidad de la inteligencia humana, sean capaces de actuar como lo haría en cada ocasión concreta una persona. Esto significa que los modelos de Programación Lineal que vayamos a usar en esas condiciones no van a poder ser, en general, los conocidos y bien desarrollados hasta ahora, porque van a tener que ser redefinidos para adecuarlos a ese nuevo contexto.

Es de sobra conocido que habitualmente el planteamiento de un problema real se hace en términos que, siendo perfectamente comprensibles, son difícilmente representables de forma eficaz: "el costo del transporte será de unas 750 pesetas", "el beneficio es de un 30%", ... Cuando hemos de manejar datos de esa naturaleza, que obviamente no tiene porqué ser probabilística, generalmente se actúa forzando los datos a tomar aquellos valores que entendemos son los más representativos de los verdaderos, por ejemplo 750 y 30%, planteando de esta manera problemas que podríamos denominar deformados, y que pueden llevarnos a ob-

tener soluciones que, siendo optimales para el problema planteado, están muy alejadas de la verdadera solución que correspondería al problema original, si este se hubiera planteado sobre sus auténticos valores, que podrían haber sido 742 y 28.5%.

Es por todo esto que en el contexto de los Sistemas de Ayuda a la Decisión, y de la Inteligencia Artificial entre otras disciplinas, la representación adecuada de la información es una tarea primordial, como garantía de corrección de las soluciones que se persiguen y porque, además, según la versión que adoptemos de imprecisión, podemos encontrarnos con diferentes conceptos de óptimo y, por tanto, de optimización.

En todo lo que sigue, por imprecisión entenderemos lo que habitualmente se conoce por borrosidad (fuzziness), es decir, esa vaguedad lingüística que tiene perfecto sentido para los seres humanos, a pesar de la falta de información exacta que muestren ("no sé que edad tiene, pero es joven"). Escogemos esta versión de imprecisión por entender que en general es la más cercana y apropiada para los desarrollos que nos interesan en el ámbito de la Inteligencia Artificial, ya que lo que habitualmente hacemos a la hora de definir algo, es extraer objetos de la realidad como conceptos lingüísticamente etiquetados en el dominio referencial que se considere, reflejando la borrosidad de cualquier etiqueta lingüística cierta distancia entre los objetos etiquetados y algún punto referencial, que en cada caso depende del contexto, por lo que modeliza bastante adecuadamente el modo de razonamiento y comunicación humano.

En este marco, y desde un punto de vista mucho más concreto, en general un problema de Programación Lineal (PL) se formula como

$$\text{Max}\{z = cx / Ax \leq b, x \geq 0\}$$

siendo A una matriz de números reales de dimensión $m \times n$, b un vector en \mathbb{R}^m y c un vector de costos en \mathbb{R}^n .

Sobre este planteamiento podemos suponer, a tenor de lo comentado con anterioridad, que el decisor se expresa, conoce o formula los datos del problema de forma imprecisa, pero perfectamente clara para él: "el rendimiento será superior al del año pasado", "se trabajará un número elevado de horas", "el salario bruto es de unos tres millones", etc. En este ambiente de optimización con tal tipo de datos, es en el que nace la Programación Lineal Borrosa (PLB).

Aunque la PLB tiene su antecedente teórico en 1970 en el magistral trabajo sobre Teoría de la Decisión de R. Bellman y L.A. Zadeh ya clásico en la literatura científica [1], los problemas de PLB nacen formalmente en 1974, año en que separadamente en dos trabajos, [21] y [28], se propuso el mismo modelo para tratar los problemas de PL en los que el conjunto de restricciones estaba dado por un conjunto borroso. A pesar de la coincidencia, enfocaron su resolución desde puntos de vista, y por tanto con métodos, diferentes que proporcionaban una solución constituida por un único punto, y que por tanto puede considerarse

ajena al contexto borroso en el que se calculaba. Más adelante se demostró que tales métodos eran casos particulares de uno más general que permitía obtener una solución borrosa, contexto-dependiente, y que englobaba las que se habían propuesto previamente en los referidos trabajos, [23].

Concretando, el problema central en PLB consiste en resolver un problema de PL en el que el conjunto de restricciones es borrosa,

$$\begin{aligned} \text{Max } z &= cx \\ \text{s.a : } & \\ Ax &\leq_f b \\ x &\geq 0 \end{aligned}$$

es decir, donde se supone que el decisor puede aceptar violaciones moderadas sobre el cumplimiento de las restricciones, evaluándose el grado con que se efectúan estas violaciones mediante ciertas funciones de pertenencia,

$$\mu_i : \mathbb{R} \rightarrow [0, 1], \quad i = 1, \dots, m$$

que el mismo decisor establece.

Aunque desde este planteamiento inicial, las líneas de trabajo que se han seguido en este tema han sido muchas. De forma resumida pueden sintetizarse en los siguientes apartados:

- a) Extensiones del modelo anterior a problemas más complejos. Particularmente en el campo donde más se ha incidido ha sido en el de los problemas multiobjetivo, aunque también es de destacar el trabajo realizado en otras parcelas, como es el caso de la Programación Estocástica o la Fraccional.
- b) Métodos de resolución de los diferentes problemas.
- c) Aplicaciones en dominios concretos (transporte, juegos, política hidráulica, agricultura, razonamiento a partir de conocimiento proposicional, ...).

Inicialmente abordaremos aquellos elementos teóricos básicos que son necesarios para el desarrollo del artículo. En este sentido, y sin profundizar hasta conceptos que puedan resultar triviales, se presentan las ideas más elementales relativas a conjuntos y números fuzzy. A continuación, pero dentro del contexto borroso, introducimos los problemas más típicos de Programación Lineal Borrosa.

2 Conceptos básicos

Un concepto que se muestra básico es el de número borroso. Desde el punto de vista de que un número borroso es un conjunto borroso en \mathbb{R} , podemos decir

que la noción de número borroso se introduce en 1965 en el célebre trabajo de L.A. Zadeh, [26].

Sin embargo, los números borrosos no toman carta de naturaleza hasta aproximadamente 1978 con los trabajos de S. Nahmias sobre variables borrosos y de D. Dubois y H. Prade sobre el manejo de cantidades imprecisas. Desde entonces, el estudio de las posibles definiciones de número borroso y, sobre todo, el cómo manipularlos y compararlos, es un tema de gran interés en el área de los conjuntos borrosos, [25].

En esta sección se introducen las nociones y operaciones elementales entre conjuntos borrosos para llegar al referido concepto de número. Establecida esta noción, se dedica el resto de la sección al problema de comparar dos números borrosos. Este es un problema complejo porque, dado el carácter impreciso de las cantidades que se consideran, por ejemplo A y B, de antemano no puede garantizarse el que $A \leq B$, o el que $B \leq A$, sino que, ambas propiedades van a verificarse simultáneamente con ciertos grados de cumplimiento. Esto hace que existan múltiples métodos de comparar dos números borrosos, lo que en la literatura especializada se ha desarrollado mediante los llamados índices de comparación.

2.1 Introducción al Concepto de Conjunto Borroso

Sea X un conjunto cuyos elementos notaremos por x , y sea A un subconjunto de X . La pertenencia de un elemento x de X al conjunto A viene dada por la función característica

$$\mu_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

donde $\{0, 1\}$ es el llamado conjunto valoración.

Si el conjunto valoración es el intervalo real $[0, 1]$, A se denomina un conjunto borroso ([26]) y μ_A mide el grado de pertenencia del elemento x a A . A se caracteriza por el conjunto de pares $\{(x, \mu_A(x)), x \in X\}$.

Dos conjuntos borrosos A y B se consideran iguales ($A = B$) si y sólo si: $\forall x \in X, \mu_A(x) = \mu_B(x)$.

Definición 1 ([26]) *Dado un conjunto borroso $A = \{(x, \mu_A(x))\}$, se define su soporte como el conjunto ordinario $Sop(A) = \{x \in X / \mu_A(x) > 0\}$.*

Definición 2 ([26]) *Dado un conjunto borroso A , llamamos α -corte de dicho conjunto, al conjunto ordinario $A_\alpha = \{x \in X / \mu_A(x) \geq \alpha\}$ con $\alpha \in [0, 1]$.*

Claramente se ve cómo los conjuntos A_α , $\alpha \in [0, 1]$ constituyen una sucesión decreciente. Si $\alpha_1 \geq \alpha_2 \Leftrightarrow A_{\alpha_1} \subseteq A_{\alpha_2}$, $\alpha_1, \alpha_2 \in [0, 1]$.

Teorema 1 (*Teorema de Representación*) Si A es un conjunto borroso y A_α sus α -cortes, $\alpha \in [0, 1]$, se verifica que:

$$A = \bigcup_{\alpha \in [0, 1]} \alpha A_\alpha$$

entendiendo esta notación formal como la igualdad entre las funciones de pertenencia de ambos conjuntos. Si $\mu_{A_\alpha}(x)$ nota la función característica de A_α , caso particular de la función de pertenencia,

$$\mu_{A_\alpha}(x) = \begin{cases} 1 & \text{si y sólo si } x \in A_\alpha \\ 0 & \text{en otro caso} \end{cases}$$

la función de pertenencia del conjunto borroso A puede expresarse en términos de las funciones características de sus α -cortes según la fórmula

$$\mu_A(x) = \sup_{\alpha \in [0, 1]} \min(\alpha, \mu_{A_\alpha}(x))$$

Definición 3 ([26]) Un conjunto borroso es convexo si y sólo si sus α -cortes son convexos.

Una definición equivalente a la convexidad es que A es convexo si y sólo si $\forall x_1, x_2 \in X, \forall \lambda \in [0, 1], \mu_A(\lambda x_1 + (1 - \lambda)x_2) \geq \min(\mu_A(x_1), \mu_A(x_2))$.

Definición 4 Se define la altura de un conjunto borroso $Alt(A) = \sup_{x \in X} \mu_A(x)$.

Definición 5 Un conjunto borroso se dice normalizado si y sólo si $\exists x \in X$ en el que $\mu_A(x) = 1$.

2.2 Números Borrosos

Definición 6 [10] Un número borroso A , es un conjunto μ_A de la recta real, convexo, normalizado y tal que

- a) $\exists x_0 \in \mathbb{R} / \mu_A(x_0) = 1$, que suele llamarse moda, y
- b) μ_A es continua a trozos.

Todo número borroso está pues caracterizado por una función de pertenencia $\mu_A : \mathbb{R} \rightarrow [0, 1]$ y toda función como la anterior engendra un número borroso donde, $\forall x \in \mathbb{R}, \mu_A(x)$ es el grado de pertenencia de x al número borroso A .

Notaremos por $F(\mathbb{R})$ al conjunto de las funciones de pertenencia sobre \mathbb{R} , por tanto nos podemos referir al hablar de número borroso tanto al elemento $A \in F(\mathbb{R})$ como a $\mu_A \in F(\mathbb{R})$.

Un número borroso A , se dice que es del tipo $\mathcal{L} - \mathcal{R}$, si y sólo si su función de pertenencia μ_A es de la forma

$$\mu_A(x) = \begin{cases} \mathcal{L}\left[\frac{(m-x)}{\alpha}\right] & \text{para } x \leq m, \alpha > 0 \\ \mathcal{R}\left[\frac{(x-m)}{\beta}\right] & \text{para } x \geq m, \beta > 0 \end{cases}$$

donde m es la moda de A y α (β) la amplitud por la izquierda (derecha), \mathcal{L} y \mathcal{R} representan una función a la izquierda y derecha de m respectivamente, \mathcal{L} no decreciente y \mathcal{R} no creciente. El número borroso A lo notaremos abreviadamente por $A = (m - \alpha, m, m + \beta)_{\mathcal{LR}}$.

Definición 7 [11] *Un número borroso plano es un número borroso A tal que*

$$\exists(m_1, m_2) \in \mathbb{R}, \quad m_1 < m_2 \text{ y } \mu_A(x) = 1, \quad \forall x \in [m_1, m_2]$$

Un número borroso plano puede modelar un intervalo borroso. Un número borroso plano A de tipo $\mathcal{L} - \mathcal{R}$ está definido como

$$\mu_A(x) = \begin{cases} \mathcal{L}\left[\frac{(m_1-x)}{\alpha}\right] & \text{para } x \leq m_1, \alpha > 0 \\ \mathcal{R}\left[\frac{(x-m_2)}{\beta}\right] & \text{para } x \geq m_2, \beta > 0 \\ 1 & \text{para } m_1 \leq x \leq m_2 \\ 0 & \text{otro caso} \end{cases} \quad (4.1)$$

Más brevemente, se va a notar por $(m_1 - \alpha, m_1, m_2, m_2 + \beta)_{\mathcal{LR}}$.

Como es evidente, según sean las funciones \mathcal{L} y \mathcal{R} , obtendremos distintos tipos de números borrosos.

Vamos a considerar números borrosos planos, lineales y normalizados, cuya función de pertenencia analítica es la siguiente.

Un número borroso plano, que notaremos por $u_j^f = (r_j, \underline{u}_j, \bar{u}_j, R_j)$, tendrá como función de pertenencia

$$\forall v \in \mathbb{R}, \quad \mu_{u_j^f}(v) = \begin{cases} \frac{(v-r_j)}{(\underline{u}_j-r_j)} & r_j \leq v \leq \underline{u}_j \\ \frac{(R_j-v)}{(R_j-\bar{u}_j)} & \bar{u}_j \leq v \leq R_j \\ 1 & \underline{u}_j \leq v \leq \bar{u}_j \\ 0 & \text{otro caso} \end{cases} \quad (4.2)$$

En adelante, se utilizarán con bastante frecuencia números borrosos expresados como combinaciones lineales $y^f = \sum_j u_j^f x_j$ con $x_j \in \mathbb{R}$, $j = 1, \dots, n$.

En [22] podemos encontrar la función de pertenencia de dichos números que expresamos a continuación.

Proposición 1 Si $y^f = \sum_j u_j^f x_j = u^f x$ es una expresión lineal, en la que los u_j^f , $j = 1, \dots, n$, son números borrosos con funciones de pertenencia lineales, dados por $u_j^f = (r_j, \underline{u}_j, \bar{u}_j, R_j)$ y $x_j \geq 0$, $j = 1, \dots, n$, entonces la función de pertenencia de y^f es

$$\mu(z) = \begin{cases} \frac{(z-rx)}{(\underline{u}x-rx)} & \text{si } x > 0 \text{ y } rx \leq z \leq \underline{u}x \\ \frac{(Rx-z)}{(Rx-\bar{u}x)} & \text{si } x > 0 \text{ y } \bar{u}x \leq z \leq Rx \\ 1 & \text{si } \underline{u}x \leq z \leq \bar{u}x \\ 0 & \text{otro caso} \end{cases}$$

donde $r = (r_1, \dots, r_n)$, $\underline{u} = (\underline{u}_1, \dots, \underline{u}_n)$, $\bar{u} = (\bar{u}_1, \dots, \bar{u}_n)$ y $R = (R_1, \dots, R_n)$.

2.3 Formas de Comparar Números Borrosos

Un problema de constante actualidad a lo largo de los últimos 15 años ha sido el de la ordenación de cantidades imprecisas, y por tanto el de la comparación de números borrosos. Los diversos y múltiples enfoques con los que se puede plantear el problema, justifican la existencia de un amplio catálogo de métodos que permiten realizar la mencionada comparación. Al respecto, una excelente recopilación de técnicas, métodos y enfoques puede encontrarse en [31], [25].

Utilizaremos las formas de comparar números sólo como un medio, para analizar la repercusión que, en un problema de Programación Lineal Borrosa, tiene el empleo de diferentes métodos de comparación. Desde este punto de vista, no es nuestro objetivo recoger aquí todas las formas posibles que hay para la comparación.

La resolución del problema puede abreviarse por alguna de las 2 siguientes vías, según se emplee un método basado en la definición de una función ordenadora o basado en la comparación de alternativas.

Métodos Basados en la Definición de una Función Ordenadora

Consideramos $A, B \in F(\mathbb{R})$. Un método simple de comparación entre ellos consiste en la definición de una función $g : F(\mathbb{R}) \rightarrow \mathbb{R}$. Si se conoce esta función

$g(\cdot)$, entonces

$$\begin{aligned} g(A) < g(B) &\Leftrightarrow A \text{ es menor que } B \\ g(A) > g(B) &\Leftrightarrow A \text{ es mayor que } B \\ g(A) = g(B) &\Leftrightarrow A \text{ es igual que } B \end{aligned}$$

Usualmente, g se llama función de ordenación lineal si

- 1) $\forall A, B \in F(\mathbb{R}), g(A + B) = g(A) + g(B)$
- 2) $\forall r \in \mathbb{R}, r > 0, g(rA) = rg(A), \forall A \in F(\mathbb{R})$

En este caso los índices pueden clasificarse conforme a que la función ordenadora sea lineal o no.

Métodos Basados en la Comparación de Alternativas

Estos métodos, consisten en obtener el conjunto borroso de las alternativas optimales:

$$O^f = \{i, \mu_{O^f}(i)\}, \mu_{O^f}(i) = \mu_{O^f}(A^i), A^i \in F(\mathbb{R})$$

donde $\mu_{O^f}(i)$ representa el grado con el cual la alternativa i -ésima puede ser considerada la mejor.

Destacaremos finalmente que a pesar de la extraordinaria abundancia de métodos para comparar números borrosos, aún son pocos los índices que se han estudiado, puesto que es perfectamente justificable el que cada decisor humano que se considera tenga su propio método de comparación, que no tendrá porque ajustarse a ninguno de los métodos teóricos descritos en la literatura. Un estudio detallado sobre este aspecto puede encontrarse en [19] donde se emplea una red neuronal artificial que aprende el criterio de ordenación de cada decisor que se considere.

3 Programación Lineal Borrosa

En general un problema de PL se plantea como

$$\text{Max}\{z = cx / Ax \leq b, x \geq 0\}$$

donde A es una matriz $m \times n$ de números reales, $b \in \mathbb{R}^m$ y $c \in \mathbb{R}^n$.

Como es obvio, se está asumiendo que el decisor dispone de una información exacta sobre los elementos que intervienen en el problema. Sin embargo, aunque esto fuera así, es usual que dicho decisor se encuentre más comodo expresando su conocimiento en términos lingüísticos, es decir, mediante etiquetas lingüísticas convencionales [27], que utilizando datos numéricos de precisión exacta. Por tanto tiene perfecto sentido que hablemos de problemas de optimización planteados a partir de ese tipo de predicados vagos, entendiendo que esa vaguedad está

producida por la forma de expresar el conocimiento que posee el decisor, y no como causa de una cierta naturaleza aleatoria que, para nada, será considerada. En definitiva se trata de suponer que la imprecisión de los datos que definen al problema es borroso.

El primer antecedente sobre problemas de optimización con planteamiento borroso que existe en la literatura se remonta a más de tres décadas, [1]. De ese artículo provienen los conceptos clave de restricción, objetivo y decisión óptima borrosa que ya son clásicos.

Igual que ocurre con la PL en el contexto de la optimización convencional, los métodos de PLB han sido uno de los temas más estudiados en el ámbito borroso. Como ya se dijo, sin ser exhaustivos hay tres tipos más importantes de problemas de PLB, según que la imprecisión se establezca en las restricciones, en los coeficientes de la matriz tecnológica o en los costos que definen la función objetivo. Dedicamos el resto de este apartado a estudiar cada uno de ellos.

Existen en la literatura muchos modelos y métodos para la resolución de estos problemas, fundamentalmente para el caso en que f y g_i , $i \in M$, son funciones lineales [7], [8], [14], [30]. Se obtienen soluciones puntuales en algunos casos y soluciones borrosas en otros, más acordes con el planteamiento impreciso de los problemas. Estas últimas recogen un conjunto de buenas alternativas, que engloban a las soluciones puntuales obtenidas con otros métodos. En cualquier caso, el decisor debe tomar la decisión final para su problema.

Aunque en la literatura podemos encontrar muchos modelos de PLB, [12], [4], [6], [15], [7], [18], [17], ..., la mayoría de ellos suponen vaguedad sólo en algunos de los elementos descritos en el modelo. Al final de esta sección se presenta un modelo general de PLB en el que todos los elementos que intervienen son borrosos. Para ello se presentarán los tipos de modelos más importantes en programación lineal borrosa y el modelo general de PLB, [3]. De este modelo, se deriva cada caso particular del problema de PLB fácilmente y resulta conforme a sus características.

3.1 Programación Lineal con Restricciones Borrosas

Consideraremos el caso en el que un decisor asume que puede haber cierta tolerancia en el cumplimiento de las restricciones, en el sentido de estar dispuesto a tolerar cierto margen de violación que él mismo establece, [3]. Para cada restricción esta suposición se puede representar de la forma

$$a_i x \leq_f b_i, \quad i \in M = \{1, 2, \dots, m\},$$

y modelizarla por medio de una función de pertenencia

$$\mu_i : \mathbb{R} \rightarrow [0, 1] / \mu_i(a_i x) = \begin{cases} 1 & \text{si } a_i x \leq b_i \\ f_i(a_i x) & \text{si } b_i \leq a_i x \leq b_i + t_i \\ 0 & \text{si } a_i x \geq b_i + t_i \end{cases}$$

Estas funciones expresan que el decisor tolera violaciones en cada restricción hasta un valor $b_i + t_i$, $i \in M$. Por otra parte, las funciones f_i se asumen no decrecientes y continuas para estas restricciones.

La función μ_i se define para cada $x \in X$ y da el grado de cumplimiento de la i -ésima restricción para $x \in X$.

El problema asociado se representa de la siguiente forma

$$\begin{aligned} \text{Max } z &= cx \\ \text{s.a : } & \begin{aligned} Ax &\leq_f b \\ x &\geq 0 \end{aligned} \end{aligned} \tag{4.3}$$

donde $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, A es una matriz $m \times n$ de números reales.

Aunque el origen de (4.3) se encuentra en [1], este problema fue desarrollado en [21] y [29], donde se consideraron algunas hipótesis adicionales sobre la naturaleza de la función objetivo, que aquí no son relevantes.

Para resolver (4.3), se pueden considerar tres aproximaciones diferentes [21], [29] y [23]. En particular, y haciendo uso del Teorema de Representación para conjuntos borrosos, en [23] se demuestra cómo encontrar una solución borrosa a (4.3) por medio del problema auxiliar de PL paramétrica.

$$\begin{aligned} \text{Max } z &= cx \\ \text{s.a : } & \begin{aligned} Ax &\leq g(\alpha) \\ x &\geq 0, \alpha \in [0, 1] \end{aligned} \end{aligned}$$

donde $g(\alpha) \in \mathbb{R}^m$ es un vector columna definido por las funciones inversas de las f_i , $i \in M$. Obviamente, la linealidad y la dimensión de (4.3) se conserva en este último modelo.

Las soluciones propuestas en [21] y [29] aparecen de la solución borrosa propuesta en [23] para valores particulares del parámetro $\alpha \in [0, 1]$.

Suponiendo la linealidad de las f_i , tenemos que el modelo auxiliar que resuelve (4.3) es

$$\begin{aligned} \text{Max } z &= cx \\ \text{s.a : } & \begin{aligned} Ax &\leq b + t(1 - \alpha) \\ x &\geq 0, \alpha \in [0, 1] \end{aligned} \end{aligned} \tag{4.4}$$

con $t = (t_1, \dots, t_m) \in \mathbb{R}^m$.

Si notamos $S(\alpha) = \{x \in \mathbb{R}^n / z(x) = \max cx, x \in X(\alpha)\}$ con $X(\alpha) = \{x \in \mathbb{R}^n / Ax \leq b + t(1 - \alpha), \alpha \in [0, 1]\}$, definimos una solución borrosa a (4.3) como,

Definición 8 La solución borrosa a (4.3) es un conjunto borroso con función de pertenencia

$$\lambda(x) = \begin{cases} \sup_{x \in S(\alpha)} \alpha & \text{si } x \in \bigcup_{\alpha} S(\alpha) \\ 0 & \text{otro caso} \end{cases}$$

Consideramos (4.3) y la imprecisión de las restricciones representada por funciones de pertenencia no lineales

$$\mu_i'(x) = \begin{cases} 1 & \text{si } a_i x \leq b_i \\ f'_i(a_i x) & \text{si } b_i \leq a_i x \leq b_i + t_i \\ 0 & \text{si } a_i x \geq b_i + t_i \end{cases}$$

donde la función $f'_i(\cdot)$ se supone estrictamente decreciente y continua, $f'_i(b_i) = 1$ y $f'_i(b_i + t_i) = 0$.

Si usamos la aproximación propuesta en [23] y con una discusión similar a la anterior para el caso lineal, la solución óptima borrosa para (4.3) puede obtenerse de la solución paramétrica óptima del problema

$$\begin{aligned} \text{Max } & z = cx \\ \text{s.a : } & Ax \leq g'(\alpha) \\ & x \geq 0, \alpha \in [0, 1] \end{aligned}$$

donde $g'(\alpha) = f'^{-1}(\alpha)$, $\forall \alpha \in [0, 1]$.

En [9], se muestra una relación entre las soluciones obtenidas del caso lineal y el no lineal. Los siguientes resultados se demuestran en [9].

Proposición 2 Sea $[a, b]$ un intervalo real y $f : [a, b] \rightarrow [0, 1]$ continua, lineal y estrictamente decreciente con $f(a) = 1$ y $f(b) = 0$. Para cualquier otra función continua, estrictamente decreciente $f' : [a, b] \rightarrow [0, 1]$, tal que $f'(a) = 1$ y $f'(b) = 0$, existe una función $r : [0, 1] \rightarrow [0, 1]$ tal que $r(\cdot) \circ f(\cdot) = f'(\cdot)$.

Proposición 3 Consideramos el problema de PLB (4.3). Notamos $x(\cdot)$ y $x'(\cdot)$ a las soluciones óptimas borrosas para este problema, usando funciones de pertenencia lineales y no lineales, respectivamente, para la imprecisión de las restricciones. Entonces $x'(\alpha) = x(r^{-1}(\alpha))$, donde $r(\cdot)$ es la obtenida en la proposición anterior.

Con estos resultados, el valor de la función objetivo será $z'(\alpha) = cx'(\alpha) = cx(r^{-1}(\alpha))$.

Esto demuestra que resolver un problema de PL con restricciones borrosas modelizadas por funciones de pertenencia lineales pueden obtener la solución borrosa del mismo problema modelizado por funciones de pertenencia no lineales. Para el caso en que el problema de PL tiene restricciones borrosas modelizadas por funciones de pertenencia definidas a trozos obtenemos un resultado similar, [9]. Así, no se pierde generalidad si siempre suponemos el problema de restricciones borrosas con funciones de pertenencia lineales.

3.2 Programación Lineal con Costos Borrosos

En este caso, el decisor no conoce con exactitud los valores de los coeficientes c , representando esta situación por el problema de PLB siguiente, [3].

$$\begin{aligned} \text{Max } z &= c^f x \\ \text{s.a : } & \begin{aligned} Ax &\leq b \\ x &\geq 0 \end{aligned} \end{aligned} \tag{4.5}$$

con $c^f \in (F(\mathbb{R}))^n$ y suponiendo funciones de pertenencia de la forma

$$\mu_j : \mathbb{R} \rightarrow [0, 1], \quad j \in N \quad / \quad \mu_j(v) = \begin{cases} 0 & \text{si } v \leq r_j \text{ ó } v \geq R_j \\ h_j(v) & \text{si } r_j \leq v \leq \underline{c}_j \\ g_j(v) & \text{si } \bar{c}_j \leq v \leq R_j \\ 1 & \text{si } \underline{c}_j \leq v \leq \bar{c}_j \end{cases} \tag{4.6}$$

donde $h_j(\cdot)$ y $g_j(\cdot)$ son funciones continuas estrictamente crecientes y decrecientes, respectivamente, tales que, $h_j(\underline{c}_j) = g_j(\bar{c}_j) = 1, \forall j \in N$.

Aunque existe una gran gama de funciones h_j y g_j (lineales, exponenciales, logarítmicas, parabólicas cóncavas y convexas, etc.) se suelen considerar los costos borrosos como números borrosos planos con funciones $h_j(\cdot)$ y $g_j(\cdot)$ lineales. Así para el número $(r_j, \underline{c}_j, \bar{c}_j, R_j)$ estas funciones vendrán dadas de la forma:

$$h_j(v) = \begin{cases} \frac{v-r_j}{\underline{c}_j-r_j} & \text{si } r_j \leq v \leq \underline{c}_j \\ 0 & \text{otro caso} \end{cases}, \quad g_j(u) = \begin{cases} \frac{R_j-u}{\bar{c}_j-R_j} & \text{si } \bar{c}_j \leq u \leq R_j \\ 0 & \text{otro caso} \end{cases}$$

Para resolver (4.5) hay diferentes aproximaciones [6], [20], [22]. En [8] se demuestra que el método propuesto en [6] da un contexto formal para encontrar la solución de (4.5) englobando las soluciones de los métodos propuestos por [20], [22], [18].

La solución borrosa propuesta en [6] para este problema se puede obtener de

la solución del siguiente problema paramétrico multiobjetivo

$$\begin{aligned}
 \text{Max } z &= [c^1 x, c^2 x, \dots, c^{2^n} x] \\
 \text{s.a.:} \\
 Ax &\leq b, \quad x \leq 0 \\
 c_j^k &\in \{h_j^{-1}(1-\alpha), g_j^{-1}(1-\alpha)\} \\
 \alpha &\in [0, 1], \quad k = 1, \dots, 2^n, \quad j \in N
 \end{aligned}$$

3.3 Programación Lineal con Números Borrosos en la Matriz Tecnológica

Ahora, consideramos que los coeficientes en la matriz tecnológica y los coeficientes de la parte derecha se representan por números borrosos, siendo números reales los costos que definen la función objetivo, [3].

Este tipo de problema de PLB se plantea en los siguientes términos,

$$\begin{aligned}
 \text{Max } z &= cx \\
 \text{s.a.:} \\
 a_i^f x &\leq_f b_i^f, \quad i \in M \\
 x &\geq 0
 \end{aligned} \tag{4.7}$$

donde para cada $i \in M$, $a_i^f = (a_{i1}^f, \dots, a_{in}^f)$, $a_{ij}^f \in F(\mathbb{R})$, $j \in N$, $b_i^f \in F(\mathbb{R})$, $x \in X = \{x \in \mathbb{R}^n / a_i^f x \leq_f b_i^f, i \in M, x \geq 0\}$ y $c \in \mathbb{R}^n$.

Una versión casi idéntica de (4.7) fue el punto de partida de este tipo de modelos (en [22] se presentó un problema similar a éste, pero suponiendo también imprecisión en el objetivo). Ahora bien, para resolver (4.7), con la intención de obtener una solución borrosa y no una puntual, como se hacía en (4.3) podemos suponer que se admiten violaciones en el acoplamiento de sus restricciones, hasta una amplitud máxima de valor t_i^f , $i \in M$. [7]. Nótese que, a diferencia de (4.3), t_i^f ha de ser un número borroso debido a la naturaleza de los coeficientes que toman parte en cada restricción.

Desde este punto de vista, en [7] se propone un método de resolución para el modelo general (4.7). La aproximación consiste en la sustitución del conjunto de restricciones de (4.7) por el siguiente conjunto borroso convexo:

$$a_i^f x \leq_g b_i^f + t_i^f(1-\alpha), \quad i \in M, \quad \alpha \in [0, 1]$$

donde $a_i^f = (a_{i1}^f, \dots, a_{in}^f)$, t_i^f es un número fijado por el decisor que nos da la violación tolerada en la restricción, y \leq_g es una relación entre tales números.

Así el problema (4.7) queda de la siguiente forma

$$\begin{aligned} \text{Max } z &= cx \\ \text{s.a.:} \\ a_i^f x &\leq_g b_i^f + t_i^f(1 - \alpha), \quad i \in M \\ x &\geq 0, \quad \alpha \in [0, 1] \end{aligned}$$

En [5], la solución al problema originalmente planteado se obtiene por particularización, en el problema auxiliar, de la relación para cada diferente método de comparación de números borrosos.

3.4 Un Modelo General de Programación Lineal Borrosa

Un modelo general de PLB, [3], en el que consideramos todos los casos anteriores es un problema del tipo:

$$\begin{aligned} \text{Max } z &= \sum_{j=1}^n c_j^f x_j \\ \text{s.a.:} \\ \sum_{j=1}^n a_{ij}^f x_j &\leq_f b_i^f \\ x_j &\geq 0, \quad i \in M, \quad j \in N \end{aligned} \tag{4.8}$$

donde los elementos borrosos se consideran dados por:

- a) Para cada costo $\exists \mu_j \in F(\mathbb{R})$ tal que $\mu_j : \mathbb{R} \rightarrow [0, 1]$, $j \in N$ las cuales definen el vector de costos borrosos.
- b) Para cada fila $\exists \mu_i \in F(\mathbb{R})$ tal que $\mu_i : \mathbb{R} \rightarrow [0, 1]$, $i \in M$ las cuales definen el número borroso en la parte derecha.
- c) Para cada $i \in M$ y $j \in N$ $\exists \mu_{ij} \in F(\mathbb{R})$ tal que $\mu_{ij} : \mathbb{R} \rightarrow [0, 1]$ las cuales definen los números borrosos en la matriz tecnológica.
- d) Para cada fila $\exists \mu^i \in F[F(\mathbb{R})]$ tal que $\mu^i : F(\mathbb{R}) \rightarrow [0, 1]$, $i \in M$ que nos da para cada $x \in \mathbb{R}^n$, el grado de acoplamiento del número borroso $a_{i1}^f x_1 + a_{i2}^f x_2 + \dots + a_{in}^f x_n$, $i \in M$ con respecto a la i -ésima restricción, es decir, la adecuación entre estos números borrosos y el correspondiente b_i^f con respecto a la i -ésima restricción.

Un método de resolución, [2], para el modelo general (4.8) consiste en la sustitución del conjunto restricción de (4.8) por un conjunto borroso convexo.

Sea g una función ordenadora de números borrosos y sea la función $\psi : F(\mathbb{R}) \times F(\mathbb{R}) \rightarrow F(\mathbb{R})$ tal que

$$\psi(a_i^f x, b_i^f) = \begin{cases} t_i^f & \text{si } a_i^f x \leq_g b_i^f \\ t_i^f \ominus a_i^f x \oplus b_i^f & \text{si } b_i^f \leq_g a_i^f x \leq_g b_i^f \oplus t_i^f \\ 0 & \text{si } a_i^f x \geq_g b_i^f \oplus t_i^f \end{cases}$$

con $t_i^f \in F(\mathbb{R})$ tal que su soporte esté incluido en \mathbb{R}^+ y \leq_g una relación que mide el que $A \leq_g B$, $\forall A, B \in F(\mathbb{R})$, y \ominus y \oplus las operaciones usuales entre números borrosos.

Definición 9 La función de pertenencia asociada a la restricción borrosa $a_i^f x \leq_f b_i^f$, con t_i^f un número borroso, dando la violación máxima en la verificación de la i -ésima restricción, es la siguiente:

$$\mu^i : F(\mathbb{R}) \rightarrow [0, 1] \quad / \quad \mu^i(a_i^f x, b_i^f) = \frac{g(\psi(a_i^f x, b_i^f))}{g(t_i^f)} \quad (4.9)$$

donde g es una función ordenadora de números borrosos.

Si consideramos el problema (4.8), \leq_f con funciones de pertenencia (4.9) y utilizando el teorema de representación para números borrosos, obtenemos que

$$\mu^i(a_i^f x, b_i^f) \geq \alpha \Leftrightarrow \frac{g(f(a_i^f x, b_i^f))}{g(t_i^f)} \geq \alpha \Leftrightarrow \frac{g(t_i^f \ominus a_i^f x \oplus b_i^f))}{g(t_i^f)} \geq \alpha \Leftrightarrow$$

$$g(t_i^f) - g(a_i^f x) + g(b_i^f) \geq g(t_i^f)\alpha \Leftrightarrow g(a_i^f x) \leq g(b_i^f \oplus t_i^f(1 - \alpha)) \Leftrightarrow$$

$$a_i^f x \leq_g b_i^f \oplus t_i^f(1 - \alpha)$$

donde \leq_g es la relación correspondiente a g .

Con lo cual, un problema auxiliar para resolver (4.8) es el siguiente:

$$\begin{aligned} \text{Max} \quad & \sum_{j=1}^n c_j^f x_j \\ \text{s.a :} \quad & \sum_{j=1}^n a_{ij}^f x_j \leq_g b_i^f \oplus t_i^f(1 - \alpha), \quad i \in M \\ & x \geq 0, \quad \alpha \in [0, 1], \quad j \in N \end{aligned} \quad (4.10)$$

Si en el problema (4.8) no hubiera números borrosos en su formulación, sino sólo restricciones borrosas, este enfoque coincide con el correspondiente modelo

$$\begin{aligned} \text{Max } z &= cx \\ \text{s.a :} \\ Ax &\leq b + t(1 - \alpha) \\ x &\geq 0, \alpha \in [0, 1] \end{aligned}$$

Es decir, en el caso de una restricción borrosa $a_i x \leq_f b_i$ la función de pertenencia asociada a esta restricción será de la forma:

$$\mu^i : F(\mathbb{R}) \rightarrow [0, 1] / \mu^i(a_i x, b_i) = \frac{g(f(a_i x, b_i))}{g(t_i)}$$

donde g es el orden clásico en \mathbb{R} y t_i es la violación máxima en el acoplamiento de la i -ésima restricción ($t_i \in \mathbb{R}$).

$$\mu^i(a_i x, b_i) = \begin{cases} 1 & a_i x \leq b_i \\ (1 - \frac{a_i x - b_i}{t_i}) & b_i \leq a_i x \leq b_i + t_i \\ 0 & a_i x \geq b_i + t_i \end{cases}$$

Se pueden utilizar, para resolver dicho problema, las distintas relaciones de comparación de números borrosos, tanto en las restricciones como en el objetivo, o relaciones de comparación en las restricciones y α -cortes en el objetivo, que nos llevarán a obtener distintos modelos convencionales, lo que permite la obtención de una solución propiamente borrosa.

4 Ejemplos

Ejemplo 1 Un país productor de un cierto mineral se ve obligado a exportar anualmente una cantidad del producto no inferior a 2000 toneladas aproximadamente ni superior a 4000 toneladas aproximadamente. La venta del producto se puede hacer en el mercado internacional a 2500 unidades monetarias la tonelada o bien a un país vecino a un precio de 2000 unidades monetarias por tonelada. El gobierno desea saber qué parte del mineral producido (x_2) debe de vender en el mercado internacional y qué parte (x_1) al país vecino si su objetivo es maximizar los ingresos.

Si x_1 y x_2 son las cantidades de mineral vendidas al país vecino y en el mercado internacional, respectivamente, y teniendo en cuenta que las ventas en general no pueden ser inferiores a 2000 toneladas aproximadamente ni superiores a 4000 aproximadamente, tendremos que la suma de las cantidades que nos reflejará el total exportado deberá verificar

$$x_1 + x_2 \geq_f 2000$$

$$x_1 + x_2 \leq_f 4000$$

donde estamos suponiendo que se permite vender un poco menos de 2000 toneladas (25 toneladas) y un poco más de 4000 toneladas (100 toneladas). Estas restricciones borrosas podrían modelarse como

$$\mu_1(x_1, x_2) = \begin{cases} 0 & \text{si } x_1 + x_2 \leq 1975 \\ \frac{x_1+x_2-1975}{25} & \text{si } 1975 \leq x_1 + x_2 \leq 2000 \\ 1 & \text{otro caso} \end{cases}$$

$$\mu_2(x_1, x_2) = \begin{cases} 1 & \text{si } x_1 + x_2 \leq 4000 \\ \frac{4100-x_1-x_2}{100} & \text{si } 4000 \leq x_1 + x_2 \leq 4100 \\ 0 & \text{otro caso} \end{cases}$$

Como el gobierno tiene que decidir sobre las cantidades que vende en cada mercado, y debe vender como mínimo 100 toneladas al país vecino y 200 toneladas al mercado internacional, las variables x_i , $i = 1, 2$ deberán ser mayores o iguales a 100 y 200, respectivamente. Puesto que el gobierno quiere calcular x_1 y x_2 con objeto de maximizar sus ingresos, deberá de resolver el problema

$$\begin{aligned} \text{Max} \quad & 2000x_1 + 2500x_2 \\ \text{s.a :} \quad & \\ & x_1 + x_2 \geq_f 2000 \\ & x_1 + x_2 \leq_f 4000 \\ & x_1 \geq 100 \\ & x_2 \geq 200 \end{aligned}$$

El modelo auxiliar, utilizando [23], para resolver el problema es el siguiente problema paramétrico:

$$\begin{aligned} \text{Max} \quad & 2000x_1 + 2500x_2 \\ \text{s.a :} \quad & \\ & x_1 + x_2 \geq 2000 - 25(1 - \alpha) \\ & x_1 + x_2 \leq 4000 + 100(1 - \alpha) \\ & x_1 \geq 100 \\ & x_2 \geq 200 \\ & \alpha \in [0, 1] \end{aligned}$$

donde 25 y 100 son los márgenes de la violaciones permitidas en el verificación de las restricciones.

La solución óptima es

$$x_\alpha^* = (100, 4000 - 100\alpha)$$

y el valor óptimo de

$$10200000 - 250000\alpha \text{ unidades monetarias}$$

para $\alpha \in [0, 1]$.

Ejemplo 2 Dos productos, A y B, para la exportación deben producirse utilizando tres procesos diferentes (corte, doblado y empaquetado). La producción de una unidad del producto A (B) requiere 10 (6) minutos de tiempo de transformación en el departamento del corte, 5 (10) minutos en el departamento de doblaje, y sobre 7 (7) minutos en el departamento de empaquetado. El tiempo total disponible para cada proceso de producción es 2500 minutos para el corte, como máximo 2000 minutos para doblar (aunque se permiten violaciones hasta 2064 minutos) y alrededor de 2050 minutos para empaquetar (dependiendo de la urgencia del envío), pero este tiempo de empaquetado total nunca puede excederse de 2124 minutos. Cuando se vende al exterior, el producto A (B) tiene un beneficio alrededor de 23 euros (32 euros) por unidad, dependiendo del cambio actual. Si los encargados desean maximizar el beneficio, podrán resolver el problema:

$$\begin{aligned} \text{Max } & 23^f x_1 + 32^f x_2 \\ \text{s.a : } & \\ & 10x_1 + 6x_2 \leq 2500 \\ & 5x_1 + 10x_2 \leq_f 2000 \\ & 7^f x_1 + 7^f x_2 \leq 2050^f \\ & x_j \geq 0, j = 1, 2 \end{aligned}$$

La segunda restricción puede modelarse como

$$\mu(x_1, x_2) = \begin{cases} 1 & \text{si } 5x_1 + 10x_2 \leq 2000 \\ \frac{2064 - 5x_1 - 10x_2}{64} & \text{si } 2000 \leq 5x_1 + 10x_2 \leq 2064 \\ 0 & \text{otro caso} \end{cases}$$

y la tercera restricción nunca puede exceder del número borroso $2124^f = (2094, 2124, 2154)$.

Suponemos que los números borrosos están definidos por funciones de pertenencia lineales: $23^f = (22, 23, 25)$, $32^f = (31, 32, 34)$, $7^f = (6, 7, 8.5)$ y $2050^f = (2025, 2050, 2075)$.

El primero modelo auxiliar para resolver el problema es el siguiente problema paramétrico:

$$\begin{aligned} \text{Max } & 23^f x_1 + 32^f x_2 \\ \text{s.a : } & \\ & 10x_1 + 6x_2 \leq 2500 \\ & 5x_1 + 10x_2 \leq 2000 + 64(1 - \alpha) \\ & 7^f x_1 + 7^f x_2 \leq_g 2050^f + 74^f(1 - \alpha) \\ & x_j \geq 0, j = 1, 2, \alpha \in [0, 1] \end{aligned}$$

donde 64 y 74^f son los márgenes de la violaciones permitidas en el verificación de la segunda y tercera restricción, respectivamente, y \leq_g es la relación asociada con la función de ordenación lineal g .

Puesto que hay diferentes aproximaciones y métodos para resolverlo, podemos obtener diferentes soluciones. En cualquier caso, y siendo coherente con la naturaleza imprecisa del problema, todas las soluciones serán borrosas.

Por ejemplo, si elegimos como función de ordenación lineal g para la restricción tercera el primer índice de Yager, [31], y como función de ordenación para números borrosos de la función objetivo el segundo índice de Yager, [31], obtenemos el siguiente problema auxiliar

$$\begin{aligned} \text{Max } & \frac{24x_1+34x_2}{2x_1+2x_2+1} \\ \text{s.a : } & \\ & 10x_1 + 6x_2 \leq 2500 \\ & 5x_1 + 10x_2 \leq 2000 + 64(1 - \alpha) \\ & 7.16x_1 + 7.16x_2 \leq 2050 + 74(1 - \alpha) \\ & x_j \geq 0, j = 1, 2, \alpha \in [0, 1] \end{aligned}$$

cuya solución óptima es

$$x_\alpha^* = (0, 206.4 - 6.4\alpha)$$

y el valor óptimo alrededor de

$$6604.8 - 204.8\alpha \text{ euros}$$

(definido por el número borroso $(6398.4 - 198.4\alpha, 6604.8 - 204.8\alpha, 7017.6 - 217.6\alpha)$) para un α fijado.

5 Conclusión y epílogo

Los métodos de optimización basados en la lógica borrosa no terminan en la PLB. En efecto, la facilidad de resolver problemas reales de dimensión cada vez mayor, gracias a la mayor potencia y el menor costo de los computadores, la imposibilidad de conocer en todos los casos las soluciones exactas que les corresponden a esos problemas, y la necesidad de dar respuestas a las situaciones prácticas contempladas en multitud de casos (problemas de secuenciación, de diseño de rutas, de localización, etc.), han motivado que los algoritmos de tipo heurístico sean empleados cada vez más, como valiosas herramientas capaces de proporcionar soluciones donde los algoritmos exactos no son capaces de encontrarlas. Así en los últimos años ha aparecido un enorme catálogo de técnicas diversas, animadas por el principio de que es mejor satisfacer que optimizar, o lo que es lo mismo que, antes que no poder dar la solución óptima a un problema, es mejor dar una solución que satisfaga al usuario en algún sentido que previamente habrá especificado, y que se han demostrado extraordinariamente efectivas. Ejemplos de esas técnicas pueden ser los algoritmos de Búsqueda Tabú, Enfriamiento Simulado, GRASP ("Greedy Randomized Adaptive Search Procedure"), Genéticos, o los

mas recientes: Meméticos, VNS (Búsqueda por Entornos Variables), Colonias de Hormigas, Búsqueda Dispersa, Programación por Restricciones, que en definitiva, demuestran el gran interés de este campo, y la falta de un mínimo marco teórico en el que encuadrar, relacionar y poder comparar estos algoritmos.

Se puede decir que en la mayoría de los casos, estas heurísticas se han inspirado en algún modelo real de la naturaleza, la sociedad, la física, ... para producir modelos teóricos que se ajustan a las circunstancias consideradas. Desde esta perspectiva se ha conseguido proporcionar solución a casos que, hace muy poco tiempo, eran intratables con las técnicas convencionales. Sin embargo, las soluciones conseguidas no han sido en la inmensa mayoría de los casos las óptimas. Han sido soluciones "cercañas a las óptimas", que frecuentemente se han obtenido con cargo a criterios distintos del clásico "conseguir el mejor valor de la función objetivo", al considerar características subjetivamente establecidas por el decisor.

Como a lo largo de este trabajo ha quedado patente, cuando hablamos de subjetividad asociada a personas humanas, o incluso de cercanía a un valor óptimo, la forma mejor contrastada de modelizar ese tipo de situaciones es mediante los conjuntos borrosos. Sin embargo esa forma de modelización de la subjetividad, tan desarrollada en otros ámbitos, prácticamente no ha sido aplicada al caso del diseño de algoritmos heurísticos, [24], a pesar de todo apunta que este puede ser un enfoque muy prometedor porque, aparte de proporcionar soluciones tan cercanas al óptimo como las otras heurísticas convencionales ya conocidas,

- a) encuentran la solución del problema en cuestión con un menor costo que los demás métodos, y
- b) como es habitual en el marco de las metodologías borrosas, generalizan las heurísticas ya conocidas (que deben ser casos particulares de las aquí presentadas, en los valores 0 o 1 del grado de cumplimiento que se considere).

Estos aspectos describen de forma general el contexto por el que parece que discurrirán las investigaciones en el futuro mas inmediato, para poder abordar diferentes tareas:

1. De generalización: mediante el estudio, diseño e implementación de nuevos algoritmos heurísticos basados en la lógica borrosa, que se demuestren operativos sobre problemas reales. A su vez, estos nuevos algoritmos podrán obtenerse desde diferentes perspectivas: Considerando reglas de parada borrosas, suponiendo que los conjuntos solución están definidos de forma vaga, midiendo el rendimiento de los objetivos de forma cualitativa, etc.
2. De sistematización: desarrollando estructuras comunes ("templates", esqueletos, esquemas, ...) para incluir tantas heurísticas como sea posible, y que permitan diseñar metaheurísticas híbridas o ajustar los parámetros de cada una de ellas. Con estos se podrá realizar la implementación de procedimientos que contemplen los aspectos efectivos de las diversas heurísticas, y

que permitan al usuario o el propio sistema decidir que componentes incluir y que valores asignar a los parámetros. Estas tareas de sistematización, a su vez, deberían contemplar los casos uni y multi-objetivo, así como la viabilidad de parallelizar las componentes del "template" para facilitar la labor del usuario a la hora de elegir una vía de solución.

3. De implementación: para obtener Sistemas de Ayuda a la Decisión que, incorporando en sus Sistemas Gestores de Modelos esos u otros templates integrados por Algoritmos Heurísticos basados en metodologías borrosas, resuelvan efectivamente problemas de tanta trascendencia hoy día como los de Diseño de Redes, de Planificación Logística o de Bioinformática.

Agradecimientos

Este trabajo ha sido realizado parcialmente al amparo de los proyectos de investigación nacionales TIC2002-4242-CO3-02 y TIC2002-04021-C02-01 , financiados por el Ministerio de Ciencia y Tecnología.

6 Bibliografía

- [1] Bellman RE , Zadeh LA (1970) Decisión Making in a Fuzzy Environment. Management Science 17 (B) 4: 141–164
- [2] Cadenas JM, Verdegay JL (1997) Using Fuzzy Numbers in Linear Programming. IEEE Transactions on Systems, Man, and Cybernetics 27 (B) 6: 1017–1022
- [3] Cadenas JM, Verdegay JL (1999) Modelos de Optimización con Datos Imprecisos. Univesidad de Murcia. Servicio de Publicaciones
- [4] Campos L (1986) Modelos de la PLD para la resolución de Juegos Matriciales imprecisos. Tesis doctoral, Universidad de Granada
- [5] Campos L, Verdegay JL (1989) Linear Programming Problems and Ranking of Fuzzy Numbers. Fuzzy Sets and Systems 32: 1–11
- [6] Delgado M, Verdegay JL, Vila MA (1987) Imprecise costs in mathematical programming problems. Control and Cybernet 16 (2): 113–121
- [7] Delgado M, Verdegay JL, Vila MA (1989) A general model for fuzzy linear programming. Fuzzy Sets and systems 29: 21–29
- [8] Delgado M, Verdegay JL, Vila MA (1990) Relating different approaches to solve linear programming problems with imprecise costs. Fuzzy Sets and Systems 37: 33–42

- [9] Delgado M, Herrera F, Verdegay JL, Vila MA (1993) Post-optimality analysis on the membership function of a fuzzy linear programming problema. *Fuzzy Sets and Systems* 53: 289–297
- [10] Dubois D, Prade H (1978) Operations on Fuzzy Numbers. *International Journal Systems Science* 9: 613–626
- [11] Dubois D, Prade H (1980) *Fuzzy Sets and Systems. Theory and Applications*. Academic Press
- [12] Dubois D, Prade H (1983) Ranking of Fuzzy Numbers in the setting of Possibility Theory. *Information Science* 30: 183–244
- [13] Dubois D., Prade H (1987) The mean value of a fuzzy number. *Fuzzy Sets and Systems* 24: 279–300
- [14] Fedrizzi M, Kacprzyk J, Verdegay JL (1991) A Survey of Fuzzy Optimization and Fuzzy Mathematical Programming. In: Fedrizzi M, Kacprzyk J, Roubens M (eds) *Interactive Fuzzy Optimization*. Springer Verlag, Berlin
- [15] González A (1988) *Métodos subjetivos para la comparación de Números difusos*. Tesis doctoral, Universidad de Granada
- [16] González A (1990) A studing of the ranking function approach through mean values. *Fuzzy Sets and Systems* 35: 29–41
- [17] Herrera F, Kovacs M, Verdegay JL (1992) Fuzzy Linear programming problems with homogeneous linear fuzzy functions. Presented to IIPMU'92
- [18] Ishibuchi H, Tanaka H (1990) Multipleobjective programming in optimization of the interval objective function. *EJOR* 48: 219–225
- [19] Requena I (1992) *Redes Neuronales en problemas de decisión con ambiente difuso*. Tesis doctoral, Universidad de Granada
- [20] Rommelfanger H, Hanuscheck R, Wolf J (1989) Linear programming with fuzzy objectives. *Fuzzy Sets and Systems* 29: 31–48
- [21] Tanaka H, Okuda T, Asai K (1974) On Fuzzy Mathematical Programming. *Journal of Cybernetics* 3, 4: 37–46
- [22] Tanaka H, Ichihashi H, Asai F (1984) A formulation of fuzzy linear programming problems based a comparison of fuzzy numbers. *Control and Cybernet* 13: 185–194
- [23] Verdegay JL (1982) Fuzzy Mathematical Programming. In: Gupta MM , Sanchez E (eds) *Fuzzy Information and Decisión Processes*

- [24] Verdegay JL (2003) Fuzzy Sets based Heuristics for Optimization. Studies in Fuzziness and Soft Computing. Springer Verlag
- [25] Wang X, Kerre E (1996) On the classification and the Dependencies of the Ordering Methods. In: Ruan D (ed) Fuzzy Logic Foundation and Industrial Applications. International Series in Intelligent Technologies. Kluwer
- [26] Zadeh LA (1965) Fuzzy Sets. *Information and Control* 8: 338–353
- [27] Zadeh LA (1975) The Concept of a Linguistic Variable and its Applications to Approximate Reasoning. Part I, *Information Sciences* 8: 199–249, Part II, *Information Sciences* 8: 301–357, Part III, *Information Sciences* 9: 43–80
- [28] Zimmermann HJ (1974) Optimization in fuzzy environments. Presentado al XXI International TIMS and 46th ORSA Conference, San Juan, Puerto Rico
- [29] Zimmermann HJ (1976) Description and Optimization of fuzzy Systems. *International Journal of General Systems* 2: 209–215
- [30] Zimmermann HJ (1987) Fuzzy Sets, Decision Making, and Expert Systems. Kluwer Academic Publishers, Boston
- [31] Zhu Q, Lee ES (1992) Comparison and Ranking of Fuzzy Numbers. In: Kacprzyk J, Fedrizzi M (eds) Fuzzy Regression Analysis. Onmitech Press Warsan and Physica-Verlag

Branch-and-fix coordinado, un esquema de resolución de problemas estocásticos multietápicos 0–1 mixtos.

**A. Alonso-Ayuso¹, M. F. Clement², L. F. Escudero², M. L. Gil²
y M. T. Ortúñoz³**

¹Escuela de CC. Experimentales y Tecnología,
Universidad Rey Juan Carlos, Móstoles(Madrid), Spain
e-mail: a.alonso@escet.urjc.es

²Centro de Investigación Operativa
Universidad Miguel Hernández, Elche (Alicante), Spain
e-mail: {mf.clement, escudero, ml.gil} @umh.es

³Dpto. de Estadística e Investigación Operativa
Universidad Complutense de Madrid, Spain
e-mail: mteresa@mat.ucm.es

1 Introducción

En muchos problemas de optimización, en especial en aquellos que evolucionan en el tiempo a lo largo de un horizonte de planificación dado, es habitual que algunos de los coeficientes de la función objetivo y del vector del término independiente e, incluso, de la matriz de restricciones, no se conozcan con exactitud en el momento de tomar las decisiones correspondientes, aunque se disponga de alguna información sobre ellos.

Supongamos que es posible representar la incertidumbre de estos coeficientes mediante un árbol de escenarios, ver más abajo. En este caso, una de las disciplinas más apropiadas para tratar el problema es la Programación Estocástica (*Stochastic Programming, SP*) vía análisis de escenarios. La gran ventaja de este

enfoque es que, con él, se puede considerar el riesgo de una mala decisión en el propio modelo. La contraposición a la *SP* es una mala utilización de la Programación Determinista en cualquiera de sus modalidades (programación lineal, no lineal, entera, combinatoria, etc.). Esta mala utilización se pone de relieve en el caso muy frecuente en el que se reemplazan los parámetros inciertos por su propio promedio, y se optimiza el nuevo modelo para obtener la solución óptima a proponer al decisor. Se puede observar que dicha solución sólo es la solución óptima del escenario promedio que, incluso, puede no existir como tal escenario. La implantación de dicha solución en los diversos escenarios puede dar un resultado muy malo en la función objetivo e, incluso, la solución puede ser infactible para algunos de ellos. En cambio, la *SP* contempla todos los escenarios, sin subordinarse a ninguno, de forma que el riesgo de una mala decisión viene reflejado en el mismo modelo. El campo de aplicación de la *SP* es, pues, el mismo campo de la Programación Determinista, pero con incertidumbre en algunos parámetros. Para una panorámica general sobre los modelos y técnicas de *SP* pueden consultarse los libros [6, 10, 11, 17].

El tratamiento de las variables 0-1 en *SP* es más reciente que el tratamiento del caso continuo, pero su campo de aplicación es enorme, como lo es el campo de aplicación de la programación mixta 0-1, sobre todo en los entornos con horizonte temporal. La mayoría de los enfoques hoy día existentes también se basan en el análisis de escenarios para aprovechar la estructura del *Modelo Determinista Equivalente (Deterministic Equivalent Model, DEM)* del problema estocástico 0-1 mixto de recurso completo. Un punto clave será la descomposición del *DEM* en lo que llamaremos *clusters de escenarios*.

Es interesante destacar la aparición reciente de trabajos en la literatura en los que la función a optimizar (sea, minimizar) no es sólo el valor esperado de la función objetivo, sino que se contempla una función compuesta, de forma ponderada, por dicho valor esperado y la probabilidad de que la solución a implantar permita valores no deseables en la función objetivo. Esta alternativa permite considerar la variabilidad del valor de la función objetivo al obtener la solución que minimiza su valor esperado.

En este capítulo se estudia la utilización del procedimiento llamado *Branch-and-Fix Coordinado (BFC)* para obtener la solución óptima 0-1 mixta del problema estocástico original. Se puede utilizar tanto la descomposición Lagrangiana como la descomposición de Benders, entre otras metodologías, para aprovechar la estructura del *DEM*. Estos tipos de descomposiciones permiten obtener soluciones factibles continuas una vez obtenidas soluciones factibles 0-1, de forma que todas ellas satisfagan las llamadas condiciones de *no anticipación* en la solución óptima.

El resto del capítulo se estructura de la forma siguiente: La sección 2 revisa algunos conceptos importantes en programación estocástica 0-1, analizando la minimización del valor esperado de la función objetivo. La sección 3 trata algunas

medidas de riesgo que permitan valorar la bondad de la solución obtenida. La sección 4 revisa la descomposición Lagrangiana como medio de acotación en una metodología *Branch-and-Bound*. Por último, la sección 5 estudia el esquema *Branch-and-Fix Coordinado*.

2 Minimización del valor esperado

Sea el siguiente modelo determinista

$$\begin{aligned} & \min cx + ay \\ & \text{s.a. } Ax + By = b \\ & x \in \{0,1\}^{nc}, y \geq 0, \end{aligned} \tag{5.1}$$

donde c y a son los vectores de coeficientes de la función objetivo, b es el vector (de dimensión m) del término independiente, A y B son las matrices de restricciones (de dimensiones $m \times nc$ y $m \times n$ respectivamente), x e y son los vectores (de dimensiones nc y n , respectivamente) de las variables 0–1 y continuas, respectivamente, a optimizar sobre un conjunto de etapas \mathcal{T} , y m , n y nc son el número de restricciones, variables 0–1 y variables continuas, respectivamente. El modelo debe extenderse para recoger la incertidumbre en los valores de algunos parámetros; por tanto, se precisa un procedimiento para modelizar dicha incertidumbre.

Como se ha mencionado antes, la Programación Estocástica 0–1 (*S01P*) es más reciente que la *SP* continua, pudiéndose encontrar numerosas aplicaciones prácticas, en campos como la planificación de la producción [1, 2, 3, 14], la planificación energética [8, 12, 15, 21] y finanzas [7, 22], entre otros muchos.

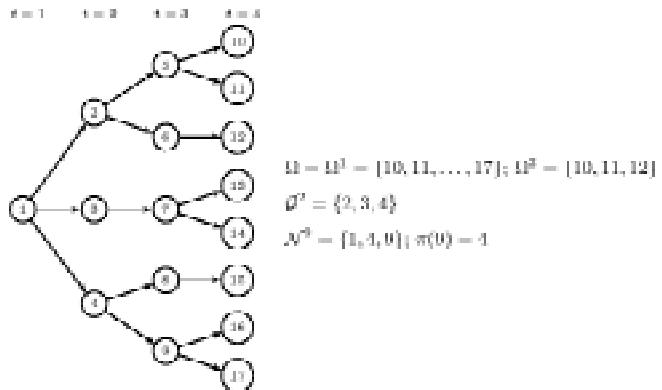


Fig. 1 Scenario tree

Sean las siguientes definiciones básicas:

Definición 1. Dado un horizonte temporal, se llama *etapa* a un conjunto de periodos de tiempo en el que los parámetros inciertos toman un determinado valor.

Al inicio de cada etapa se suponen conocidos los valores que toman todos los parámetros del problema para todos los períodos de tiempo incluidos en dicha etapa.

Definición 2. Un *escenario* es una realización de los parámetros inciertos, junto con los parámetros determinísticos, a lo largo de las etapas del horizonte temporal.

Definición 3. Un *grupo de escenarios* para una etapa dada es el conjunto de escenarios para los que los parámetros inciertos toman el mismo valor hasta la etapa considerada.

Muchos de los enfoques actuales en *SP* y, desde luego, en *S01P*, gestionan la incertidumbre a partir del análisis de escenarios. Para ilustrar este concepto, sea la Figura 1: cada nodo representa un punto en el tiempo donde puede tomarse una decisión. Una vez que la decisión está tomada, pueden ocurrir distintas eventualidades (e.g., en este ejemplo tenemos tres posibles eventualidades en el periodo $t = 2$), y la información relativa a dichas eventualidades está disponible al principio de la etapa siguiente (en este caso, periodo de tiempo). Esta información se suele representar por medio de un árbol, donde cada camino de la raíz a las hojas representa un escenario, y se corresponde con una realización de todo el conjunto de parámetros inciertos. Cada nodo en el árbol ha de asociarse con un grupo de escenarios, de forma que dos escenarios pertenecen al mismo grupo en una etapa dada si tienen las mismas realizaciones de los parámetros inciertos hasta dicha etapa. De acuerdo con el principio de *no-anticipación*, ver [18], las variables con índices de tiempo correspondientes a las etapas anteriores hasta la etapa considerada inclusive deben tomar el mismo valor para ambos escenarios.

Sea la siguiente notación relacionada con el árbol de escenarios:

\mathcal{T} , conjunto de etapas (en este caso, períodos de tiempo) del horizonte temporal dado. $\mathcal{T}_1 \equiv \mathcal{T} - \{|\mathcal{T}|\}$.

Ω , conjunto de escenarios.

\mathcal{G} , conjunto de grupos de escenarios.

\mathcal{G}^t , conjunto de grupos de escenarios en la etapa t , para $t \in \mathcal{T}$ ($\mathcal{G}^t \subseteq \mathcal{G}$).

Ω^g ,] conjunto de escenarios en el grupo g , para $g \in \mathcal{G}$ ($\Omega^g \subseteq \Omega$).

$\pi(g)$, nodo predecesor inmediato del nodo g , para $g \in \mathcal{G}$.

\mathcal{N}^g , conjunto de grupos de escenarios $\{k\}$ tales que $\Omega^g \subseteq \Omega^k$, para $g \in \mathcal{G}$ ($\mathcal{N}^g \subset \mathcal{G}$). Es decir, conjunto de grupos de escenarios (uno por cada etapa) cuyos conjuntos de escenarios contienen el subconjunto de escenarios del

grupo g . Nótese que el (único) camino desde el nodo asociado con el grupo de escenarios g hasta el nodo raíz en el árbol de escenarios correspondiente atraviesa todos los nodos asociados con grupos de escenarios en \mathcal{N}^g . Por motivos técnicos, se supone que $g \in \mathcal{N}^g$.

w_g , peso asociado al grupo de escenarios g , para $g \in \mathcal{G}$. Nota: $w_g = \sum_{\omega \in \Omega^g} w^\omega$, donde w^ω representa el peso que el modelizador asocia al escenario ω , $\omega \in \Omega$, y $\sum_{\omega \in \Omega} w^\omega = 1$ y $\sum_{g \in \mathcal{G}^t} w_g = 1 \forall t \in \mathcal{T}$.

Dependiendo del tipo de recurso que se considere, simple, parcial o completo, pueden construirse distintos tipos de modelos. En este trabajo sólo se considera el recurso completo, y, en primer lugar, se trata la minimización del valor esperado. En ese caso, la versión estocástica del programa (5.1) resulta

$$\begin{aligned} \min Q_E &= \sum_{\omega \in \Omega} w^\omega (c^\omega x^\omega + a^\omega y^\omega) \\ \text{s.a. } Ax^\omega + By^\omega &= b^\omega \quad \forall \omega \in \Omega \\ v &\in \mathcal{N} \\ x^\omega &\in \{0, 1\}^n, y^\omega \geq 0 \quad \forall \omega \in \Omega, \end{aligned} \tag{5.2}$$

donde c^ω y a^ω son los vectores de coeficientes de la función objetivo y b^ω es el vector del término independiente para el escenario ω , x^ω e y^ω son las variables correspondientes, $v = (x, y)$ y \mathcal{N} es el espacio factible correspondiente a las condiciones de *no anticipación* para las variables x e y , tal que

$$v \in \mathcal{N} = \{v_t^\omega | v_t^\omega = v_t^{\omega'} \quad \forall \omega \in \Omega^g : g \in \mathcal{G}^t, t \in \mathcal{T}_1\}, \tag{5.3}$$

donde v_t^ω es tal que $v^\omega = (v_t^\omega, \forall t \in \mathcal{T})$ y $\omega' \in \Omega^g$.

Nota: Los modelos de recurso completo son aquellos que satisfacen las condiciones de *no anticipación* para todas las variables, exigiendo una variable única para cada grupo de escenarios, sin subordinarse a ninguno de ellos.

Para modelizar las condiciones (5.3) pueden utilizarse dos enfoques distintos: la representación con *variables divididas* y la representación *compacta*. En el primer caso se pueden utilizar dos tipos de formulaciones. Una de ellas es la conocida como representación *asociada a los nodos* (o asociada a los grupos de escenarios). Dicha formulación requiere utilizar copias de las variables con elementos no negativos en las restricciones que pertenecen a etapas distintas. La otra formulación, conocida como representación *asociada a los escenarios*, requiere utilizar copias de todas las variables del modelo. En ambos casos, las condiciones de *no anticipación* deben incorporarse explícitamente al modelo, pero la segunda formulación preserva la estructura del modelo de una forma más adecuada para el enfoque

considerado en este trabajo; el modelo asociado es el siguiente,

$$\begin{aligned} \min Q_E &= \sum_{\omega \in \Omega} w^\omega (c^\omega x^\omega + a^\omega y^\omega) \\ \text{s.a. } Ax^\omega + By^\omega &= b^\omega \quad \forall \omega \in \Omega \\ v_t^\omega - v_t^{\omega'} &= 0, \text{ donde } \omega' \in \Omega^g, \quad \forall \omega \in \Omega^g : g \in \mathcal{G}^t, t \in \mathcal{T}_1 \\ x^\omega &\in \{0, 1\}^n, y^\omega \geq 0 \quad \forall \omega \in \Omega. \end{aligned} \tag{5.4}$$

La representación *compacta* requiere modelizar con más detalle las relaciones entre las variables. Como ilustración, se supone que el vector de variables v_t^ω tiene coeficientes no negativos en las restricciones correspondientes a las etapas t y $t+1$, de forma que el modelo determinista puede escribirse de la forma siguiente,

$$\begin{aligned} \min cx + ay \\ \text{s.a. } A_t^- x_{t-1} + A_t x_t + B_t^- y_{t-1} + B_t y_t &= b_t \quad \forall t \in \mathcal{T} \\ x_t &\in \{0, 1\}^{n'}, y_t \geq 0 \quad \forall t \in \mathcal{T}, \end{aligned} \tag{5.5}$$

donde x_t e y_t son los vectores de variables para la etapa t tal que $x = (x_t, \forall t \in \mathcal{T})$ e $y = (y_t, \forall t \in \mathcal{T})$, n' es la dimensión de los vectores x_t , y A_t^- , A_t , B_t^- y B_t son las correspondientes matrices de restricciones. Abusando ligeramente de la notación, la versión estocástica del modelo puede escribirse de la siguiente forma,

$$\begin{aligned} \min Q_E &= \sum_{g \in \mathcal{G}} w_g (c_g x_g + a_g y_g) \\ \text{s.a. } A_t^- x_{\pi(g)} + A_t x_g + B_t^- y_{\pi(g)} + B_t y_g &= b_g \quad \forall g \in \mathcal{G}^t, t \in \mathcal{T} \\ x_g &\in \{0, 1\}^{n'}, y_g \geq 0 \quad \forall g \in \mathcal{G}, \end{aligned} \tag{5.6}$$

donde c_g y a_g son los vectores de coeficientes de la función objetivo, b_g es el vector del término independiente, y x_g e y_g son los vectores de las variables del grupo de escenarios g , con $c_g = c_t^\omega$, $a_g = a_t^\omega$ y $b_g = b_t^\omega$ donde, en general, $d^\omega = (d_t^\omega, \forall t \in \mathcal{T})$, para $\omega \in \Omega^g : g \in \mathcal{G}^t, t \in \mathcal{T}$.

3 Minimización de las funciones objetivo *valor esperado–riesgo* y CaR

Los modelos contemplados en la sección anterior buscan la minimización del valor esperado de la función objetivo. Sin embargo, existen otros enfoques que se ocupan también de medidas de riesgo, por medio de semi-desviaciones [16] y *probabilidades de exceso* [20]. Estos enfoques son más adecuados que los clásicos de media-varianza, sobre todo cuando aparecen variables 0–1. Ver también [19].

Sea ϕ un umbral pre establecido para la *probabilidad de exceso*, Q_P , tal que

$$Q_P = P(\omega \in \Omega : c^\omega x^\omega + a^\omega y^\omega > \phi). \quad (5.1)$$

Así, como alternativa a la minimización de Q_E (5.2), donde

$$Q_E = \sum_{\omega \in \Omega} w^\omega (c^\omega x^\omega + a^\omega y^\omega), \quad (5.2)$$

la función de *valor esperado–riesgo* a minimizar es,

$$Q_E + \beta Q_P, \quad (5.3)$$

donde β es un parámetro de peso no negativo.

Una expresión más manejable que (5.3), a efectos computacionales al menos, puede ser

$$\begin{aligned} & \min Q_E + \beta \sum_{\omega \in \Omega} w^\omega \nu^\omega \\ & \text{s.a. } c^\omega x^\omega + a^\omega y^\omega \leq \phi + M \nu^\omega \quad \forall \omega \in \Omega \\ & \nu^\omega \in \{0, 1\} \quad \forall \omega \in \Omega, \end{aligned} \quad (5.4)$$

donde ν^ω es una variable 0–1, que toma el valor 1 si el valor en la función objetivo para el escenario ω es mayor que el umbral ϕ , y toma el valor 0 en otro caso, y M es el valor más pequeño que no elimina ninguna solución factible bajo ningún escenario en el problema estocástico.

Como alternativa a $\min Q_E$ y $\min Q_E + \beta Q_P$, la función de *Cost-at-Risk* (*CaR*) a optimizar para un nivel de riesgo α dado puede expresarse como

$$\begin{aligned} & \min CaR \\ & \text{s.a. } c^\omega x^\omega + a^\omega y^\omega \leq CaR + M \nu^\omega \quad \forall \omega \in \Omega \\ & \sum_{\omega \in \Omega} w^\omega \nu^\omega \leq \alpha, \end{aligned} \quad (5.5)$$

con $0 \leq \alpha < 1$.

Nótese que la sustitución de la función de valor esperado Q_E (5.2) por el sistema de valor esperado–riesgo (5.4) no modifica la estructura del modelo. Por el contrario, la minimización de *CaR* (5.5) sí destruye dicha estructura, ya que la restricción $\sum_{\omega \in \Omega} w^\omega \nu^\omega \leq \alpha$ incluye variables de todos los escenarios y no es separable. Sin embargo, puede advertirse que la relajación Lagrangiana que resulta de su dualización recupera la separabilidad del modelo y permite el uso del esquema *BFC*, ver sección 5. Nuestra conjetura es que su solución puede proporcionar un buen *CaR*.

4 Acotación del Branch-and-Bound

Los casos del *DEM* 0–1 mixto (5.4) pueden ser de dimensiones tan grandes que la utilización de las herramientas estándar de optimización no puedan resolverlos. Puede utilizarse el esquema de la descomposición de Benders [5], al menos para casos de tamaño mediano; ver [6, 13] para problemas enteros, entre otros.

Como alternativa, se considera un esquema de tipo *Branch-and-Bound* (*BB*) para optimizar el *DEM* en el caso de variables enteras, de forma que se pueda utilizar la *Descomposición Lagrangiana* en cada nodo *BB*, dualizando las condiciones de *no anticipación*

$$v_t^\omega - v_t^{\omega'} = 0, \text{ donde } \omega' \in \Omega^g, \quad \forall \omega \in \Omega^g : g \in \mathcal{G}^t, t \in \mathcal{T}_1, \quad (5.1)$$

ver [8, 12, 15, 19, 21], entre otros. En cualquier caso, sería preciso utilizar heurísticas Lagrangianas.

El modelo Lagrangiano es el siguiente,

$$\begin{aligned} & \min \sum_{\omega \in \Omega} w^\omega (c^\omega x^\omega + a^\omega y^\omega + \beta \nu^\omega) + \sum_{t \in \mathcal{T}_1, \omega \in \Omega^g : g \in \mathcal{G}^t} \mu_t^\omega (v_t^\omega - v_t^{\omega'}) \\ & \text{s.a. } c^\omega x^\omega + a^\omega y^\omega \leq \phi + M \nu^\omega \quad \forall \omega \in \Omega \\ & \quad Ax^\omega + By^\omega = b^\omega \quad \forall \omega \in \Omega \\ & \quad 0 \leq x^\omega \leq 1, y^\omega \geq 0 \quad \forall \omega \in \Omega, \end{aligned} \quad (5.2)$$

donde μ_t^ω , $\forall \omega \in \Omega^g : g \in \mathcal{G}^t, t \in \mathcal{T}_1$ es el vector de multiplicadores de Lagrange asociados a las condiciones de *no anticipación* (5.1).

Nótese que el número de multiplicadores de Lagrange depende del número de variables del vector v y del número de grupos de escenarios, $|\mathcal{G}|$, a lo largo del horizonte temporal \mathcal{T} .

5 Branch-and-Fix Coordinado

5.1 Enfoque general

Como alternativa al enfoque de tipo *Branch-and-Bound*, consideremos el enfoque denominado *Branch-and-Fix Coordinado* (*BFC*), que trata de forma coordinada los $|\Omega|$ modelos independientes

$$\begin{aligned} & \min c^\omega x^\omega + a^\omega y^\omega + \beta \nu^\omega \\ & \text{s.a. } c^\omega x^\omega + a^\omega y^\omega \leq \phi + M \nu^\omega \\ & \quad Ax^\omega + By^\omega = b^\omega \\ & \quad x^\omega \in \{0, 1\}^n, \nu^\omega \in \{0, 1\}, y^\omega \geq 0, \end{aligned} \quad (5.1)$$

que aparecen tras la relajación de las condiciones (5.1). En cualquier caso, pueden incluirse técnicas Lagrangianas en el desarrollo de la metodología. *BFC* está diseñado especialmente para coordinar la selección de la variable y el nodo sobre el que ramificar en el árbol *Branch-and-Fix* (*BF*) correspondiente a cada escenario, de forma que las condiciones (5.1), que han sido relajadas, se satisfagan al fijar las variables adecuadas a cero o a uno. El procedimiento también coordina y refuerza la fase de eliminación de nodos activos, así como la fijación de variables y la acotación de la función objetivo del subproblema correspondiente a cada nodo.

En [4] puede encontrarse el desarrollo de los resultados en los que se asienta *BFC*. En [3] pueden encontrarse aplicaciones a problemas bietápicos 0–1 mixtos donde en la primera etapa sólo aparecen variables 0–1, y en [2] aplicaciones a problemas multietápicos 0–1 puros. En [7] se presenta un enfoque para la resolución de un problema bietápico con variables 0–1 y variables continuas en la primera etapa y variables continuas en la segunda etapa; se utiliza el esquema de la descomposición de Benders para resolver los subproblemas *LP* correspondientes a ciertos nodos en los árboles *BF*, aquellos llamados *Familias de Nodos Gemelos Enteros*, que definiremos más adelante. En la actualidad, estamos trabajando en algunos desarrollos de problemas multietápicos con variables 0–1 y continuas en todas las etapas; para resolverlos se está desarrollando un algoritmo híbrido que utiliza tanto el esquema *BFC* como algunos esquemas de descomposición tipo Benders y Lagrangianos.

Para presentar el esquema *BFC*, sea \mathbb{R}^ω el árbol *BF* asociado al escenario ω , \mathcal{E}^ω el conjunto de nodos activos en \mathbb{R}^ω para $\omega \in \Omega$, \mathcal{I} el conjunto de índices de las variables x en cualquier grupo de escenarios, y $(x_t^\omega)_i$ la variable i -ésima del vector x_t^ω , para $t \in \mathcal{T}, i \in \mathcal{I}, \omega \in \Omega$. Nota: $n' \equiv |\mathcal{I}|$.

Definición 4. Dos variables $(x_t^\omega)_i$ y $(x_t^{\omega'})_i$, se llamarán variables *comunes* para los escenarios ω y ω' , si $\omega, \omega' \in \Omega^g : g \in \mathcal{G}^t$, para $\omega \neq \omega', t \in \mathcal{T}_1, i \in \mathcal{I}$. Nótese que dos variables *comunes* tienen elementos no nulos en la condición de *no anticipación* correspondiente a un grupo de escenarios dado.

Definición 5. Diremos que dos nodos activos, $e \in \mathcal{E}^\omega$ y $e' \in \mathcal{E}^{\omega'}$ son nodos *gemelos* respecto a un grupo de escenarios dado si los caminos desde sus nodos raíz hasta ellos en sus árboles *BF* \mathbb{R}^ω y $\mathbb{R}^{\omega'}$, respectivamente, o bien todavía no han ramificado en sus variables *comunes* o bien tienen los mismos valores 0–1 en las variables *comunes* ya ramificadas/fijadas $(x_t^\omega)_i$ y $(x_t^{\omega'})_i$ para $\omega, \omega' \in \Omega, t \in \mathcal{T}_1, i \in \mathcal{I}$.

Definición 6. Llamamos *Familia de Nodos Gemelos* (*Twin Node Family*, *TNF*), \mathcal{J}_f , a un conjunto de nodos tales que cualquiera de ellos es *gemelo* de todos los demás nodos de la familia, para $f \in \mathcal{F}$, donde \mathcal{F} es el conjunto de familias. Nota 1: Por motivos prácticos, se considerará que todos los nodos *BF* pertenecen al menos a una *TNF*, aunque sea de cardinal uno. Nota 2: Puede haber nodos *gemelos* en el mismo árbol *BF*. Sería el caso de nodos cuyos caminos

desde la raíz hasta ellos en su propio árbol *BF* han ramificado en alguna variable *común* cuyo grupo de escenarios pertenece a una etapa posterior al conjunto de etapas a las que pertenecen las otras variables ramificadas.

Definición 7. Una *TNF candidata* es una *TNF* cuyos miembros todavía no han ramificado/fijado todas sus variables *comunes* relativas a un grupo de escenarios dado.

Definición 8. Una *TNF entera* es una *TNF* en la que todas las variables x y ν toman valores enteros, y se satisfacen las condiciones de *no anticipación* $(x_t^\omega)_i - (x_t^{\omega'})_i = 0 \forall \omega, \omega' \in \Omega^g : g \in \mathcal{G}^t, \omega \neq \omega', t \in \mathcal{T}_1, i \in \mathcal{I}$.

Sea el árbol de escenarios y los árboles *BF* que se muestran en la Figura 2, donde x_h^ω es una variable subindicada h bajo el escenario ω , y x_h es la notación genérica de dicha variable. Para simplificar la exposición, se supone que se está optimizando sólo el valor esperado de la función objetivo (i.e., no hay variables ν). A título ilustrativo, sea el siguiente orden de ramificación es x_1, x_2, x_3, x_4 . Las *TNFs* son: $\mathcal{J}_1 = \{1, 2, 3\}$, $\mathcal{J}_2 = \{4, 6, 8\}$, $\mathcal{J}_3 = \{5, 7, 9\}$, $\mathcal{J}_4 = \{4, 6\}$, $\mathcal{J}_5 = \{8\}$, $\mathcal{J}_6 = \{10, 12\}$, $\mathcal{J}_7 = \{11, 13\}$, $\mathcal{J}_8 = \{14\}$, $\mathcal{J}_9 = \{15\}$, $\mathcal{J}_{10} = \{16, 18\}$, $\mathcal{J}_{11} = \{17, 19\}$, $\mathcal{J}_{12} = \{20\}$, $\mathcal{J}_{13} = \{21\}$, $\mathcal{J}_{14} = \{22, 26, 30\}$, ..., $\mathcal{J}_{16} = \{22, 26, 32\}$, $\mathcal{J}_{17} = \{22, 27, 30\}$, ..., $\mathcal{J}_{19} = \{22, 27, 32\}$, and $\mathcal{J}_{20} = \{24, 28, 30\}$, ..., $\mathcal{J}_{22} = \{24, 28, 32\}$.

Se puede observar que la primera *TNF candidata* es J_1 , ya que la variable de la etapa 1 es una variable *común* a todos los nodos raíz. Además, J_2 es una familia que ha ramificado en el mismo valor de la variable *común* x_1 . Nótese que el nodo 7 es no factible, y, por tanto, también lo es la *TNF* J_3 . Como sólo hay una variable en el nodo raíz del árbol de escenarios, J_2 no es una *TNF candidata*. Adviértase que los nodos 4 y 6 también forman una *TNF*, ya que aún no se ha ramificado en sus variables comunes, x_2 y x_3 , ni éstas han sido fijadas. Por otro lado, el nodo 8 es el único nodo en la *TNF* J_5 . Ramificando en la variable x_2 resultan las *TNFs* J_6, \dots, J_9 . Se toma la decisión de ramificar independientemente en las *TNFs* J_7 y J_9 considerando x_3 la variable de ramificación, y, de esta forma, se construyen las *TNFs* J_{10}, J_{11}, J_{12} y J_{13} . Finalmente, se ramifica en la variable x_4 . Como resultado, supóngase que los nodos 23, 25, 29 y 33 son no factibles. Por otro lado, surgen las siguientes *TNFs enteras*: J_{14}, \dots, J_{22} , y la *TNF* J_6 se elimina (se supone que la cota dada por la *TNF* es peor que el valor de la mejor *TNF entera*). Finalmente, a título ilustrativo, nótese que los nodos 22 y 27 son *gemelos* para el grupo de escenarios b (i.e., han ramificado en los mismos valores de sus variables *comunes* x_1, x_2, x_3), y los nodos 22, 27, 32 son *gemelos* para el grupo de escenarios a (i.e., han ramificado en el mismo valor de su variable *común* x_1).

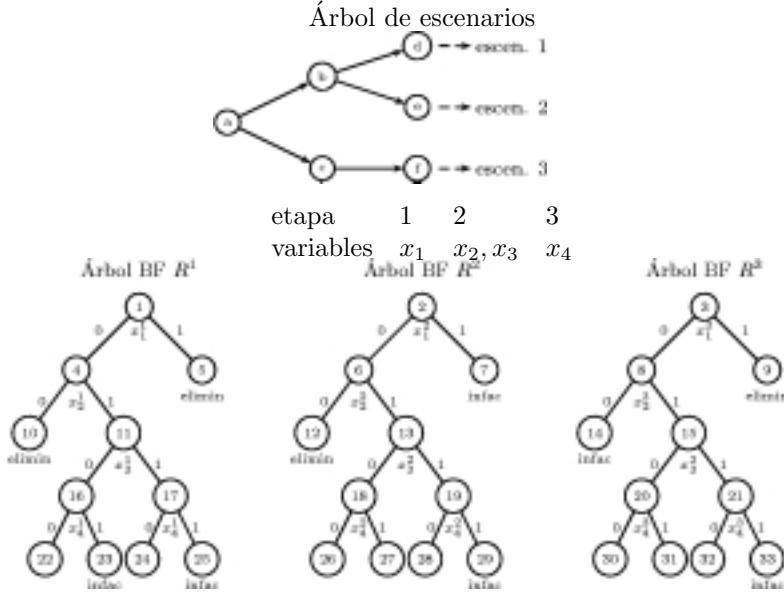


Fig. 2 Esquema del Branch-and-Fix Coordinado

Es claro que no es necesario relajar las condiciones de *no anticipación* (5.1) para todos los pares de escenarios al objeto de resolver el problema original. El número de escenarios a considerar conjuntamente en un modelo dado depende básicamente de las dimensiones del modelo correspondiente a los escenarios (5.1).

Definición 9. Un *cluster de escenarios* es un conjunto de escenarios cuyas condiciones de *no anticipación* están consideradas explícitamente en el modelo.

El criterio para la formación de los clusters $\Omega_1, \dots, \Omega_q$, donde q es el número de *clusters de escenarios*, depende de cada caso. De cualquier forma, nótese que $\Omega_p \cap \Omega_{p'} = \emptyset$, $p, p' = 1, \dots, q$: $p \neq p'$ y $\Omega = \sum_{p=1, \dots, q} \Omega_p$.

El modelo a considerar para cada *cluster de escenarios* $p = 1, \dots, q$ puede escribirse utilizando la representación *compacta* (5.2), donde ω para $d \in G^{|\mathcal{T}|}$ es el único escenario tal que $\omega \in \Omega^d$ y, por otro lado, $\mathcal{G}_p = \{g \in \mathcal{G} : \Omega^g \cap \Omega_p \neq \emptyset\}$.

$$\begin{aligned}
 & \min \sum_{d \in \mathcal{G}^{|\mathcal{T}|} \cap \mathcal{G}_p} w^\omega \sum_{g \in \mathcal{N}^d} (c_g x_g + a_g y_g) + \beta \sum_{\omega \in \Omega_p} w^\omega \nu^\omega \\
 & \text{s.a. } \sum_{g \in \mathcal{N}^d} (c_g x_g + a_g y_g) \leq \phi + M \nu^\omega \quad \forall d \in \mathcal{G}^{|\mathcal{T}|} \cap \mathcal{G}_p \\
 & A_t^- x_{\pi(g)} + A_t x_g + B_t^- y_{\pi(g)} + B_t y_g = b_g \quad \forall g \in \mathcal{G}^t \cap \mathcal{G}_p, t \in \mathcal{T} \\
 & x_g \in \{0, 1\}^{n'}, y_g \geq 0, \quad \forall g \in \mathcal{G}_p \\
 & \nu^\omega \in \{0, 1\} \quad \forall \omega \in \Omega_p.
 \end{aligned} \tag{5.2}$$

Adviértase que \mathcal{N}^d es el conjunto de nodos en el camino predecesor desde el nodo hoja d hasta el nodo raíz en el árbol de escenarios.

Los modelos correspondientes a los *cluster de escenarios* (5.2) están relacionados mediante las condiciones de *no anticipación*:

$$x_{gp} - x_{gp'} = 0 \quad (5.3)$$

$$y_{gp} - y_{gp'} = 0, \quad (5.4)$$

$\forall p, p' = 1, \dots, q : p \neq p'$, donde $g^p \in \mathcal{G}_p, g^{p'} \in \mathcal{G}_{p'}$ y $g^p = g^{p'}$.

En [2] se presenta un procedimiento heurístico para obtener (posibles) buenas soluciones para el modelo (5.2)–(5.4), mediante un procedimiento que modifica el esquema *BFC* explorando de forma selectiva sólo algunas *TNFs*. Este esquema es útil especialmente para resolver problemas dinámicos de gran tamaño (i.e., casos con docenas de miles de restricciones y variables en un horizonte temporal).

5.2 Acotación del Branch-and-Fix Coordinado

Para mayor simplicidad en la exposición, se considera el caso sin *clusters de escenarios*, ni *probabilidad de exceso*. Se puede obtener la acotación de una *TNF* dada, $\mathcal{J}_f, f \in \mathcal{F}$, mediante la resolución de $|\mathcal{J}_f|$ modelos *LP* independientes asociados con los nodos en la familia. Sin embargo, puede obtenerse una acotación mejor utilizando la *Descomposición Lagrangiana* (*Lagrangian Decomposition*, *LD*). Abusando ligeramente de la notación sea el modelo *LD*,

$$\begin{aligned} Z_D(\mu) = \min & \sum_{j \in \mathcal{J}_f} w^j (c^j x^j + a^j y^j) + \sum_{j \in \mathcal{J}_f} \mu^j (v^j - v^{j+1}) \\ \text{s.t. } & Ax^j + By^j = b^j \quad \forall j \in \mathcal{J}_f \\ & 0 \leq x^j \leq 1, y^j \geq 0 \quad \forall j \in \mathcal{J}_f, \end{aligned} \quad (5.5)$$

donde w^j, c^j y a^j , y b^j son el peso, los vectores de coeficientes de la función objetivo y el vector del término independiente para el escenario correspondiente al nodo *gemelo* j , respectivamente, A y B son las matrices de restricciones, x^j y y^j son los vectores de variables x e y , respectivamente, $v = x$ para una *TNF candidata* y $v = y$ para una *TNF entera*, y μ^j denota el vector de los multiplicadores de Lagrange asociados a las condiciones de *no anticipación* $v^j - v^{j+1} = 0 \forall j \in \mathcal{J}_f$, tal que $j+1 \in \mathcal{J}_f$ para $j = |\mathcal{J}_f|$. El modelo se puede descomponer en modelos *LP* independientes. Nótese que ya se han ramificado/fijado algunas variables del vector x^j en los caminos desde los nodos raíz en los árboles *BF* hasta los nodos miembros de la *TNF*. Nótese también que el número de multiplicadores de Lagrange es el número de variables no ramificadas/fijadas en el vector v^j multiplicado por el

número de nodos, $|\mathcal{J}_f|$, en la familia. Este número es menor (y puede ser mucho menor) que el número de multiplicadores en un nodo de BB , véase la sección 4.

La acotación LD puede expresarse como

$$Z_D(\mu^*), \quad (5.6)$$

donde

$$\mu^* = \operatorname{argmax}\{Z_D(\mu)\}. \quad (5.7)$$

Alternativamente, puede obtenerse otra acotación usando una estrategia de *sustitución Lagrangiana*. En nuestro caso, consiste en agregar [9] las condiciones de *no anticipación*, de forma que el nuevo término Lagrangiano puede expresarse

$$\lambda \sum_{j \in \mathcal{J}_f} (w^j - P/r)v^j, \quad (5.8)$$

donde λ es el vector de los nuevos multiplicadores de Lagrange, $r \equiv |J_f|$ y $P = \sum_{j \in \mathcal{J}_f} w^j$. Nótese que la dimensión de los multiplicadores de Lagrange es n' , la dimensión de los vectores v_t . Y, finalmente, la nueva cota no es peor que la cota proporcionada por la relajación lineal del modelo. La nueva acotación, alternativa al modelo (5.6), se puede expresar como

$$Z_D(\lambda^*), \quad (5.9)$$

donde

$$\lambda^* = \operatorname{argmax}\{Z_D(\lambda)\}. \quad (5.10)$$

6 Bibliografía

- [1] S. Ahmed, A.J. King y G. Parija. A multi-stage stochastic integer programming approach for capacity expansion under uncertainty. *Journal of Global Optimization*, 26:3–24, 2003.
- [2] A. Alonso-Ayuso, M.F. Clement, L.F. Escudero, M.L. Gil y M.T. Ortúñoz. FRC-S3, On dealing with the uncertainty for stochastic sequencing and scheduling problem solving. Centro de Investigación-Operativa, Universidad Miguel Hernández, Elche, 2004.
- [3] A. Alonso-Ayuso, L.F. Escudero, A. Garín, M.T. Ortúñoz y G. Pérez. An approach for strategic supply chain planning based on stochastic 0–1 programming. *Journal of Global Optimization*, 26:97–124, 2003.
- [4] A. Alonso-Ayuso, L.F. Escudero y M.T. Ortúñoz. BFC, a Branch-and-Fix Coordination algorithmic framework for solving some types of stochastic pure and mixed 0–1 programs. *European Journal of Operational Research*, 151:503–519, 2003.

- [5] J.F. Benders. Partitioning procedures for solving mixed variables programming problems. *Numerische Mathematik*, 4:238–252, 1962.
- [6] J.R. Birge y F.V. Louveaux. *Introduction to Stochastic Programming*. Springer, 1997.
- [7] L.F. Escudero, A. Garín, M. Merino y G. Pérez. On structuring Mortgage-Backed Securities portfolios under uncertainty. Report I-2003-04. Centro de Investigación-Operativa, Universidad Miguel Hernández, Elche, 2003.
- [8] N. Groewe-Kuska, K. Kiwiel, M.P. Nowak, W. Römisch e I. Wegner. Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation, En C. Greengard y A. Ruszcynski, editores. *Decision making under uncertainty: Energy and Power*, 39–70. 2001.
- [9] M. Guignard, Lagrangean Decomposition and Lagrangean Substitution for Stochastic Integer Programming. Technical note. OPIM Dept., Wharton School, University of Pennsylvania, Philadelphia, USA, 2003.
- [10] J.L. Higle y S. Sen. *Stochastic Decomposition*. Kluwer Academic Publishers, 1996.
- [11] P. Kall y S.W. Wallace. *Stochastic Programming*. John Wiley, 1994.
- [12] W.K. Klein Haneveld y M.H. van der Vlerk. Optimizing electricity distribution using integer recourse models. En S. Uryasev y P.M. Pardalos, editores. *Stochastic Optimization: Algorithms and Applications*. Kluwer Academic Publishers, 137–154, 2001.
- [13] G. Laporte y F.V. Louveaux. An integer L-shaped algorithm for the capacitated vehicle routing problem with stochastic demands. *Operations Research*, 50:415–423, 2002.
- [14] S.A. MirHassani, C. Lucas, G. Mitra y C.A. Poojari. Computational solution of a capacity planning model under uncertainty. *Parallel Computing Journal*, 26:511–538, 2000.
- [15] M.P. Novak, R. Schultz y M. Westphalen. Optimization of simultaneous power production and trading by stochastic integer programming. Stochastic Programming E-Print Series, <http://dochost.rz.hu-berlin.de/speps>, 2002.
- [16] W. Ogryczak y A. Ruszcynski. From stochastic dominance to mean-risk models: semi-deviations as risk measures. *European Journal of Operational Research*, 116:33–50, 1999.
- [17] A. Prekopa. *Stochastic Programming*. Kluwer Academic Publishers, 1995.

- [18] R.T. Rockafellar y R.J-B Wets. Scenario and policy aggregation in optimisation under uncertainty. *Mathematics of Operations Research*, 16:119–147, 1991.
- [19] R. Schultz. Stochastic programming with integer variables. *Mathematical Programming*, Ser. B 97:285–309, 2003.
- [20] R. Schultz y S. Tiedemann. Risk aversion via excess probabilities in stochastic programs with mixed-integer recourse. *SIAM Journal on Optimization*, 14:115–138, 2004.
- [21] S. Takriti y J.R. Birge. Lagrangean solution techniques and bounds for loosely coupled mixed-integer stochastic programs. *Operations Research*, 48:91–98, 2000.
- [22] W.T. Ziemba y J.M. Mulvey, editores. *Worldwide Asset and Liability Modeling*. Cambridge University Press, 1998.

Modelado de algoritmos de descomposición con GAMS

Santiago Cerisola^a, Andrés Ramos^a
Álvaro Baíllo^a

^a Escuela Técnica Superior de Ingeniería - ICAI,
e-mails: {Santiago.Cerisola, Andrés.Ramos, Alvaro.Baillo}@iit.icai.upco.es

1 Introducción

Las técnicas de descomposición tiene sentido aplicarlas a un problema de optimización cuya estructura específica permite identificar partes del mismo que son fácilmente resolubles de modo individual. Los problemas multietapa y los problemas estocásticos de gran tamaño son ejemplos de problemas cuya resolución se puede abordar mediante técnicas de descomposición. Los problemas enteros mixtos tales que la relajación de un conjunto de restricciones reduce la dificultad del mismo son también ejemplos en los que el uso de técnicas de descomposición puede ser preferible a la resolución del problema de un modo directo.

Este documento describe una implantación de la Descomposición de Benders [2, 14] y de la Relajación Lagrangiana [8, 9] en el lenguaje de modelado algebraico GAMS [4]. La descripción se centra en problemas lineales deterministas aunque su extensión para problemas estocásticos [11, 7] (en el caso de Benders) o para problemas mixtos (en el caso de Relajación Lagrangiana) es inmediata. A lo largo del documento se describe brevemente cada una de las dos descomposiciones y su implantación en GAMS para la resolución de problemas académicos concretos.

2 Descomposición de Benders

EL método de descomposición de Benders [2, 14] se centra en la interpretación de un problema de optimización como un problema bietapa PL-2. Para este tipo de problemas, las variables de decisión aparecen claramente separadas en dos bloques: las variables de la primera etapa y las variables de la segunda etapa. Esta división induce de modo natural un algoritmo iterativo de resolución en el que los problemas asociados a las variables de la primera etapa (problema maestro) y los asociados a las variables de la segunda etapa (subproblema) son resueltos consecutivamente. Este método de descomposición recibe también el nombre de *descomposición primal* (porque el problema maestro fija variables del primal), *descomposición en L* (porque se aplica a problemas con matriz de restricciones con dicha forma) y *descomposición por recursos* (porque el maestro asigna directamente las decisiones sobre los recursos al subproblema).

Un problema lineal bietapa PL-2 se representa matemáticamente de la forma siguiente,

$$\begin{aligned} & \min_{x,y} c^T x + q^T y \\ & Ax = b \\ & Tx + Wy = h \\ & x, y \geq 0 \end{aligned} \tag{6.1}$$

donde x representa el conjunto de variables de la primera etapa e y representa el conjunto de variables de la segunda etapa. Supondremos que $A \in \mathbb{R}^{m_1 \times n_1}$ y $W \in \mathbb{R}^{m_2 \times n_2}$ y que las restantes dimensiones son conformes con éstas¹. El tamaño del problema completo es $(m_1 + m_2) \times (n_1 + n_2)$. La estructura de la matriz de restricciones del problema (denominada triangular inferior por bloques) se presenta en la figura 6.1. Las restricciones $Ax = b$ afectan únicamente a las variables de la primera etapa x , mientras que las restricciones $Tx + Wy = h$ afectan a ambos conjuntos de variables x e y .

El problema lineal bietapa PL-2 (6.1) se puede interpretar también de esta manera:

$$\begin{aligned} & \min_x c^T x + \theta(x) \\ & Ax = b \\ & x \geq 0 \end{aligned} \tag{6.2}$$

donde la función de recurso, $\theta(x)$, representa la función objetivo de la segunda etapa como función de las decisiones de la primera etapa y tiene la siguiente expresión:

¹Por convención en la formulación los vectores son columna, su transposición se representa por un superíndice T , las variables se ubican a la izquierda de las expresiones y los coeficientes de las variables preceden a éstas.

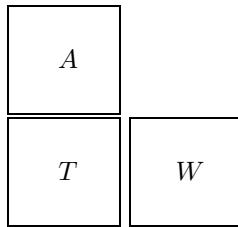


Figura 6.1: Estructura de la matriz de coeficientes de las restricciones en problemas lineales bietapa.

$$\begin{aligned} \theta(x) = & \min_y q^T y \\ & Wy = h - Tx \quad : \pi \\ & y \geq 0 \end{aligned} \tag{6.3}$$

donde π son las variables duales de las restricciones.

El problema (6.2) se conoce en la literatura como problema maestro y el problema (6.3), que evalúa la función de recurso para las decisiones de la primera etapa, como subproblema de la descomposición de Benders. El problema maestro puede ser reformulado de modo lineal utilizando la representación dual de subproblema. Con esta reformulación, el algoritmo de descomposición resuelve en cada iteración un problema maestro lineal.

2.1 Representación lineal del problema maestro

El subproblema, expresado en su forma dual es:

$$\begin{aligned} \theta(x) = & \max_{\pi} (h - Tx)^T \pi \\ & W^T \pi \leq q \end{aligned} \tag{6.4}$$

Sea $\Pi = \{\pi_1^1, \pi_2^1, \dots, \pi_2^\nu\}$ el conjunto finito de vértices del poliedro convexo definido por la región factible $W^T \pi \leq q$. Obsérvese que la región factible del problema dual no depende del valor de x . Dado que la solución óptima de un problema lineal reside en un vértice el problema se puede resolver por enumeración de todos ellos:

$$\theta(x) = \max \{(h - Tx)^T \pi^l\} \quad l = 1, \dots, \nu \tag{6.5}$$

De la ecuación anterior se deriva que la función de recurso $\theta(x)$ es una función poligonal convexa de las variables x . El subproblema de Benders puede ser reformulado como:

$$\begin{aligned}\theta(x) = & \min_{\theta} \theta \\ & \theta \geq (h - Tx)^T \pi^1 \\ & \vdots \\ & \theta \geq (h - Tx)^T \pi^\nu\end{aligned}\tag{6.6}$$

con $\theta \in \mathbb{R}$. Obsérvese que la variable θ es libre. Las restricciones de esta formulación se denominan cortes de Benders y constituyen una aproximación exterior de la función de recurso. Como el número de vértices (soluciones duales del subproblema de Benders) es finito, esta aproximación exterior es una función convexa a tramos. El problema original PL-2 se puede expresar como:

$$\begin{aligned}\min_{x,\theta} & c^T x + \theta \\ Ax &= b \\ \theta &\geq (h - Tx)^T \pi^1 \\ &\vdots \\ \theta &\geq (h - Tx)^T \pi^\nu \\ x &\geq 0\end{aligned}\tag{6.7}$$

Esta formulación se denomina *problema maestro completo*, ya que contiene todos los cortes posibles. Presenta todas las restricciones de la primera etapa más todas las condiciones necesarias derivadas de la segunda etapa. Desde el punto de vista práctico, la resolución del problema maestro completo implica disponer de forma explícita de todos los cortes de Benders, lo cual es prácticamente imposible en problemas de tamaño realista. Por esta razón, la resolución del problema original PL-2 es reemplazada por la resolución iterativa del *problema maestro relajado*² definidos como:

$$\begin{aligned}\min_{x,\theta} & c^T x + \theta \\ Ax &= b \\ \theta &\geq (h - Tx)^T \pi^l & l = 1, \dots, j \\ x &\geq 0\end{aligned}\tag{6.8}$$

donde $\theta \in \mathbb{R}$ y l representa el índice de iteraciones³. De cara a una implantación eficiente del algoritmo de descomposición, los cortes de Benders aceptan la siguiente formulación como linealización de la función de recurso en torno a los valores de las variables de la primera etapa y de los valores de la variable θ obtenidos en cada iteración:

²Por simplicidad de la exposición, se denomina como problema maestro al problema maestro relajado.

³Se entiende por iteración un ciclo maestro-subproblema, en este orden.

$$\begin{aligned}\theta &\geq \pi^{jT}(h - Tx) = \pi^{jT}(h - Tx + Tx^j - Tx^j) = \\ &= \pi^{jT} [h - Tx^j - T(x - x^j)] = \pi^{jT}(h - Tx^j) - \pi^{jT}T(x - x^j) = \\ &= f^j - \pi^{jT}T(x - x^j)\end{aligned}\quad (6.9)$$

siendo x^j y $f^j = \pi^{jT}(h - Tx^j)$ los valores de las variables de la primera etapa y el de la función de recurso para la iteración j . De este modo el corte para la iteración j también se expresa como

$$\theta - f^j \geq \pi^{jT}T(x^j - x) \quad (6.10)$$

o de modo equivalente como

$$\theta + \pi^{jT}Tx \geq f^j + \pi^{jT}Tx^j \quad (6.11)$$

Esta expresión indica que $\pi^{jT}T$ es un subgradiente del valor de la función de recurso $\theta(x)$ para la propuesta x^j del maestro.

Con esta formulación el *problema maestro relajado de Benders* tiene ahora esta expresión:

$$\begin{aligned}\min_{x,\theta} & c^T x + \theta \\ Ax &= b \\ \theta + \pi^{lT}Tx &\geq f^l + \pi^{lT}Tx^l \quad l = 1, \dots, j \\ x &\geq 0\end{aligned}\quad (6.12)$$

y el *subproblema de Benders* para cada iteración j se formula como:

$$\begin{aligned}f^j &= \min_y q^T y \\ Wy &= h - Tx^j \quad : \pi^j \\ y &\geq 0\end{aligned}\quad (6.13)$$

2.2 Cortes de infactibilidad

La descripción anterior del algoritmo de Benders ha supuesto que el subproblema de Benders es factible y acotado para cualquier propuesta del problema maestro. Esta hipótesis, conocida en la literatura como *recurso parcialmente completo*⁴, no suele satisfacerse en la práctica y el algoritmo de descomposición es modificado cuando esto ocurre. La modificación del algoritmo consiste en la construcción de otro tipo de corte, *corte de infactibilidad*, que elimina la solución propuesta en el problema maestro. La construcción de este corte se comenta a continuación.

⁴en un problema bietapa por *recurso completo* se entiende que el subproblema de la segunda etapa es siempre factible para cualquier valor de las variables de la primera etapa.

Si el subproblema es factible para un valor de x los valores duales de las restricciones forman un corte en el problema maestro denominado *corte de optimalidad* tal como se ha presentado anteriormente. El caso en el que un subproblema es no acotado carece de interés algorítmico, puesto que esta situación implica que el problema bietapa PL-2 es no acotado. La situación que debe ser destacada es aquélla en la que el subproblema es infactible para la propuesta del maestro. En ese caso se puede generar un *corte de infactibilidad* derivado de la Fase I del simplex [14]. El subproblema de minimización de infactibilidades se formula como:

$$\begin{aligned} \theta^*(x) = & \min_{v^+, v^-} e^T v^+ + e^T v^- \\ & Wy + Iv^+ - Iv^- = h - Tx^j \quad : \pi \\ & y, v^+, v^- \geq 0 \end{aligned} \quad (6.14)$$

siendo $e^T = (1 \ \dots \ 1)$, I matriz identidad ($m_2 \times m_2$) y π las variables duales de las restricciones para la solución óptima. $\theta^*(x)$ representa la función de recurso asociada a la fase I del subproblema de Benders.

El corte de optimalidad que el subproblema de minimización de infactibilidades obtiene (siguiendo la metodología descrita hasta el momento) viene dado como:

$$\theta^* \geq \pi^{jT}(h - Wx) \quad (6.15)$$

Reemplazando la variable de recurso θ^* por 0 se obtiene una condición necesaria para aquellas soluciones del problema maestro que son factibles en el subproblema de Benders. Esta condición indica que la suma de infactibilidades para esas soluciones debe ser menor que 0. Linealizando en torno a la solución obtenida en el problema maestro se obtiene la siguiente expresión para el corte de infactibilidad, similar a la del corte de optimalidad:

$$\pi^{jT}Tx \geq f^j + \pi^{jT}Tx^j \quad (6.16)$$

donde f^j representa el valor del subproblema de minimización de infactibilidades para la solución propuesta.

El problema maestro relajado considerando ambos tipos de cortes se formula de la siguiente manera:

$$\begin{aligned} & \min_{x, \theta} c^T x + \theta \\ & Ax = b \\ & \delta^l \theta + \pi^{lT}Tx \geq f^l + \pi^{lT}Tx^l \quad l = 1, \dots, j \\ & x \geq 0 \end{aligned} \quad (6.17)$$

siendo $\delta^l = 1$ para los cortes de optimalidad y $\delta^l = 0$ para los de infactibilidad.

2.3 Algoritmo de descomposición de Benders

El algoritmo de descomposición de Benders resuelve en cada iteración el problema maestro relajado y pasa el valor de las variables de la primera etapa x^j al subproblema. La resolución del subproblema de Benders con el término de la derecha (RHS) ($h - Tx^j$) genera unas variables duales π^j que se usan para generar un nuevo corte de Benders que se añade al problema maestro relajado. En cada iteración, el valor obtenido por la función objetivo del problema maestro relajado \underline{z} es una *cota inferior* del problema completo PL-2. La sucesión de estas cotas inferiores es monótona creciente dado que en cada iteración en problema maestro relajado contiene mayor número de restricciones. Por otra parte, una *cota superior* \bar{z} del valor óptimo de la función objetivo del problema original PL-2 viene dada en cada iteración por $(c^T x^j + q^T y^j)$ siendo x^j e y^j soluciones factibles en maestro y subproblema en esa iteración. El criterio de parada del algoritmo es la coincidencia de ambas cotas con una tolerancia relativa ϵ (por ejemplo, 10^{-4}).

$$\frac{|\bar{z} - \underline{z}|}{|\bar{z}|} = \frac{|q^T y^j - \theta^j|}{|c^T x^j + q^T y^j|} \leq \epsilon \quad (6.18)$$

En cada iteración del algoritmo de Benders, la variable dual generada en el subproblema es distinta del conjunto de variables duales generadas con anterioridad por el algoritmo [14]. Dado que el conjunto de posibles valores duales es finito, el algoritmo de descomposición de Benders converge en un número finito de iteraciones.

Para obtener el valor x^0 de las variables de la primera etapa en la primera iteración del algoritmo de descomposición, se suele proceder a la resolución del problema maestro relajado sin cortes, lo que equivale a fijar el valor de la variable de recurso a cero, $\theta = 0$. Otra alternativa consiste en estimar razonablemente este valor en caso de que la naturaleza del problema sea conocida.

Para el caso en que las variables de la primera etapa sean continuas, en cada iteración se dispone de una base del problema maestro relajado que es infactible sólo por una variable básica, la variable de holgura del nuevo corte. Cuando se soluciona el dual del problema maestro relajado los cortes aparecen como columnas. Al añadir una nueva columna la solución previa sigue siendo factible y la nueva solución óptima puede obtenerse en pocas iteraciones del simplex. Por ello es teóricamente conveniente resolver el maestro mediante el simplex dual. Por otra parte, cada subproblema sólo cambia las cotas de las restricciones en cada iteración. Por esta razón, suponiendo que ninguna solución del maestro ocasiona infactibilidad en el subproblema, es conveniente resolverlo mediante el método simplex primal (siempre que su tamaño lo aconseje).

Esquemáticamente el algoritmo se formula a continuación:

1. Inicialización: $j = 0$, $\bar{z} = \infty$, $\underline{z} = -\infty$, $\epsilon = 10^{-4}$

2. Resolución del problema maestro

$$\begin{aligned} & \min_{x, \theta} c^T x + \theta \\ & Ax = b \\ & \delta^l \theta + \pi^{lT} T x \geq f^l + \pi^{lT} T x^l \quad l = 1, \dots, j \\ & x \geq 0 \end{aligned} \tag{6.19}$$

Obtener la solución x^{j+1}, θ^{j+1} y evaluar la cota inferior $\underline{z} = (c^T x^{j+1} + \theta^{j+1})$

3. Resolución del subproblema de suma de infactibilidades

$$\begin{aligned} f^{j+1} = & \min_{v^+, v^-} e^T v^+ + e^T v^- \\ & Wy + Iv^+ - Iv^- = h - Tx^{j+1} \quad : \pi \\ & y, v^+, v^- \geq 0 \end{aligned} \tag{6.20}$$

Si $f^{j+1} \geq 0$, obtener π^{j+1} , formar un corte de infactibilidad y añadirlo al problema maestro, incrementar el numero de iteraciones $j = j + 1$ e ir al paso 2.

Si $f^{j+1} = 0$, ir al paso 4.

4. Resolución del subproblema de Benders

$$\begin{aligned} f^{j+1} = & \min_y q^T y \\ & Wy = h - Tx^{j+1} \quad : \pi \\ & y \geq 0 \end{aligned} \tag{6.21}$$

Obtener y^{j+1} y actualizar cota superior $\bar{z} = (c^T x^{j+1} + q^T y^{j+1})$

5. Regla de parada Si $\frac{|\bar{z} - \underline{z}|}{|\bar{z}|} \leq \epsilon$ detener el algoritmo. En otro caso, obtener π^{j+1} , formar un corte de optimalidad y añadirlo al problema maestro, incrementar el numero de iteraciones $j = j + 1$ e ir al paso 1.

Mientras no se haya generado ningún corte de optimalidad, se fija el valor de la variable de recurso θ a cero, pues en otro caso el problema maestro es no acotado. Una vez obtenido algún corte de infactibilidad, esta variable pasa a ser libre.

La siguiente sección presenta con un ejemplo los pasos del algoritmo de descomposición de Benders.

2.4 Ejemplo

Supongamos que se desea resolver el siguiente problema de programación lineal:

$$\begin{aligned} \min_{x_1, x_2, y_1, y_2} \quad & -x_1 - 2x_2 - 2y_1 - 3y_2 \\ & x_1 + x_2 \leq 600 \\ & x_1 - 2x_2 \leq 0 \\ & x_1 + x_2 + y_1 + y_2 \leq 1000 \\ & x_1 + y_1 \leq 500 \\ & -2y_1 + y_2 \leq 0 \\ & x_1, x_2, y_1, y_2 \geq 0 \end{aligned}$$

entonces,

$$c = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \quad q = \begin{pmatrix} -2 \\ -3 \end{pmatrix}, \quad b = \begin{pmatrix} 600 \\ 0 \end{pmatrix}, \quad h = \begin{pmatrix} 1000 \\ 500 \\ 0 \end{pmatrix},$$

$$A = \begin{pmatrix} 1 & 1 \\ 1 & -2 \\ 0 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{y} \quad W = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ -2 & 1 \end{pmatrix}$$

El algoritmo comienza con un problema maestro que no contiene ningún corte de Benders, por lo que el valor de la variable de recurso θ está fijado a cero. La resolución del primer problema maestro obtiene la siguiente solución para las variables de la primera etapa:

$$x = \begin{pmatrix} 0 \\ 600 \end{pmatrix}$$

Para estos valores de la primera etapa, el subproblema es factible. Como solución se obtiene

$$y = \begin{pmatrix} 133.3 \\ 266.7 \end{pmatrix} \quad \text{y} \quad \pi = \begin{pmatrix} -2.667 \\ 0.000 \\ -0.333 \end{pmatrix}$$

y un valor para la cota superior $\bar{z} = -2266.7$. Como la diferencia entre ambas cotas es elevada (la cota superior es este momento es infinito) se continúa iterando. Con los valores duales se forma un corte para el problema maestro dado por:

$$\theta - 2.667x_1 - 2.667x_2 \geq -2666.7$$

La resolución del problema maestro obtiene una nueva solución

$$x = \begin{pmatrix} 0.000 \\ 0.000 \end{pmatrix}$$

y un valor para la cota inferior $\underline{z} = -2666.7$. Al introducir en el subproblema estos valores se obtiene como solución

$$y = \begin{pmatrix} 333.3 \\ 666.7 \end{pmatrix} \quad \text{y} \quad \pi = \begin{pmatrix} -2.667 \\ 0.000 \\ -0.333 \end{pmatrix}$$

Se calcula la cota superior en esta iteración, siendo $\bar{z} = -2666.7$. La diferencia entre ambas cotas es ahora nula y, por lo tanto, acaba el algoritmo. La función objetivo es $z^* = -2666.7$ y la solución óptima es:

$$x^* = \begin{pmatrix} 0.000 \\ 0.000 \end{pmatrix} \quad \text{e} \quad y^* = \begin{pmatrix} 333.3 \\ 666.7 \end{pmatrix}$$

2.5 Descomposición de Benders en GAMS

Dentro del lenguaje GAMS la descomposición de Benders puede implantarse creando un modelo para el problema maestro y un modelo para el subproblema. La adición de cortes para el problema maestro dentro del proceso iterativo se consigue declarando la ecuación de cortes de Benders sobre un índice estático sobre el que se construye un conjunto dinámico que representa el conjunto de cortes activos en cada iteración. Posteriormente, la ecuación de cortes se construye sobre este conjunto dinámico, de modo que durante el proceso iterativo la activación incremental del conjunto dinámico implica la inclusión incremental de estas restricciones adicionales en el problema maestro. Estas ideas se presentan a continuación en el siguiente código que resuelve el problema académico de la anterior sección. Se ha seguido en la manera de lo posible la notación y la formulación presentada anteriormente.

```
$TITLE Descomposición de Benders (Bd)

SETS
  I     índice de variables      de la primera etapa / i1 * i2 /
  L     índice de variables      de la segunda etapa / l1 * l2 /
  M     índice de restricciones de la primera etapa / m1 * m2 /
  N     índice de restricciones de la segunda etapa / n1 * n3 /

  J     índice de iteraciones          / j1 * j20 /
  JJ(j) subconjunto de iteraciones
```

* Datos del problema

```

PARAMETERS
  C(i)      coeficientes función objetivo primera etapa
  / i1      -1
    i2      -2 /
  Q(l)      coeficientes función objetivo segunda etapa
  / l1      -2
    l2      -3 /
  B(m)      cotas restricciones primera etapa
  / m1      600
    m2      0 /
  H(n)      cotas restricciones segunda etapa
  / n1      1000
    n2      500
    n3      0 /

```

```

TABLE A(m,i)  matriz de restricciones primera etapa
              i1      i2
  m1      1      1
  m2      1      -2

```

```

TABLE T(n,i)  matriz de restricciones segunda etapa
              i1      i2
  n1      1      1
  n2      1      0
  n3      0      0

```

```

TABLE W(n,l)  matriz de restricciones segunda etapa
              l1      l2
  n1      1      1
  n2      1      0
  n3      -2     1

```

* Fin datos del problema

```

POSITIVE VARIABLES
  X(i)      variables primera etapa
  Y(l)      variables segunda etapa

```

```

VARIABLES
  Z1      función objetivo primera etapa
  Z2      función objetivo segunda etapa
  theta   función de recurso

```

```

SCALARS
  TOL      tolerancia relativa / 1e-6 /
  Z_INF    cota inferior      / -INF /
  Z_SUP    cota superior       / INF /

```

```

PARAMETERS
  PI(n,j)  variables duales restricciones segunda etapa en la iteración j
  DELTA(j)  tipo de corte (infactible 0 óptimo 1)      de la iteración j
  X_J(i,j)  valores de las variables de la primera etapa en la iteración j

```

```
Z2_J(j)      valor de la función objetivo del subproblema en la iteración j;
```

* Declaración de las ecuaciones

```
EQUATIONS
  EQ_OBJ      función objetivo problema completo
  EQ_Z1       función objetivo primera etapa
  EQ_Z2       función objetivo segunda etapa
  EQ_R1(m)    restricciones primera etapa
  EQ_R2(n)    restricciones segunda etapa
  CORTES(j)   cortes de Benders ;
```

* Construcción de las ecuaciones

```
EQ_OBJ        .. Z1 =E= SUM(i, C(i)*X(i)) + SUM(l, Q(l)*Y(l)) ;

EQ_Z1         .. Z1 =E= SUM(i, C(i)*X(i)) + theta ;

EQ_Z2         .. Z2 =E= SUM(l, Q(l)*Y(l)) ;

EQ_R1(m)     .. SUM(i, A(m,i)*X(i)) =L= B(m) ;

EQ_R2(n)     .. SUM(i, T(n,i)*X(i)) + SUM(l, W(n,l)*Y(l)) =L=
H(n) ;

CORTES(jj)   .. DELTA(jj) * theta =G= Z2_J(jj) +
SUM(n $(ORD(n) < 3), PI(n,jj)*SUM(i, T(n,i)*(X_J(i,jj) - X(i)))) ;

MODEL MAESTRO / EQ_R1, CORTES, EQ_Z1 / ; MODEL SUBPROBLEMA /
EQ_R2, EQ_Z2 / ; MODEL COMPLETO / EQ_R1, EQ_R2, EQ_OBJ /
;
```

```
FILE COPT / cplex.opt / ;
```

```
PUT COPT PUT 'scaind -1'/'lpmethod 1' / 'preind 0' / 'epopt
1.1e-9' / 'eprhs 1.1e-9' PUTCLOSE COPT ;
```

```
MAESTRO.OPTFILE = 1 ; SUBPROBLEMA.OPTFILE = 1 ;
```

* Algoritmo de Descomposición de Benders

```
* Inicialización de parámetros del problema
JJ(j) = NO ;
theta.FX = 0 ;
DELTa(j) = 0 ;
Z2_J(j) = 0 ;
X_J(i,j) = 0 ;
PI(n,j) = 0 ;
```

```

* Iteraciones del algoritmo

LOOP(j $(ABS(1-Z_INF/Z_SUP) > TOL),

*      Resolución del problema maestro
      SOLVE MAESTRO USING LP MINIMIZING Z1 ;
*      Adquisición de la solución
      X_J(i,j) = X.L(i) ;

*      Fijación de la variable de la primera etapa y
*      Resolución del subproblema
      X.FX(i) = X.L(i) ;
      SOLVE SUBPROBLEMA USING LP MINIMIZING Z2;

*      Adquisición de los parámetros para formar un nuevo corte
      IF(SUBPROBLEMA.MODELSTAT = 4,
*         Subproblema infactible
         DELTA(j) = 0 ;
         Z2_J(j) = SUBPROBLEMA.SUMINFES;
      ELSE
*         Actualización de la cota inferior
         Z_INF = Z1.L ;
         DELTA(j) = 1 ;
         theta.LO = -INF ; theta.UP = INF ;
*         Actualización de la cota superior
         Z_SUP = SUM(i, C(i)*X.L(i)) + SUM(l, Q(l)*Y.L(l)) ;
      ) ;

      Z2_J(j) = Z2.L;
      PI(n,j) = EQ.R2.M(n) ;
      X.L0(i) = 0 ; X.UP(i) = INF;

*      Incremento del conjunto de cortes
      JJ(j) = YES;
) ;

```

Los lenguajes de modelado algebraico como GAMS permiten el uso de diferentes optimizadores para resolver los problemas. En el caso anterior, el optimizador escogido ha sido CPLEX [10]. Para que las variables duales devueltas por el subproblema de Benders sean correctas para generar un corte de Benders tanto en el caso de optimalidad como en el de infactibilidad, para este optimizador deben ser desactivadas las opciones de *preproceso* (*preind* 0) y de *escalado* (*scaind* -1) y el subproblema debe ser resuelto mediante el algoritmo simplex primal (*lpmethod* 1). Con el uso de otros optimizadores se debe comprobar la corrección de las variables duales tanto para optimalidad como para infactibilidad. Si las variables duales no son las adecuadas para generar un corte de infactibilidad, el subproblema de minimización de infactibilidades debe de ser planteado explícitamente

y resuelto en el código GAMS.

2.6 Problema de transporte con coste fijo mediante descomposición de Benders

Veamos a continuación un caso ejemplo característico de la aplicación del método de descomposición de Benders. Se trata del problema de transporte donde algunos o todos los arcos tienen un coste fijo asociado a la decisión de su instalación o a su uso. El problema consiste en la minimización de los costes fijos y variables sujeto a las restricciones de respetar las ofertas máximas de los orígenes y las demandas en los destinos. El problema se formula de la siguiente manera

$$\begin{aligned} & \min_{x_{ij}} \sum_{ij} (c_{ij}x_{ij} + f_{ij}y_{ij}) \\ & \sum_j x_{ij} \leq a_i \\ & \sum_i x_{ij} \geq b_j \\ & x_{ij} \leq M_{ij}y_{ij} \\ & x_{ij} \geq 0, y_{ij} \in \{0, 1\} \end{aligned} \tag{6.22}$$

siendo c_{ij} el coste variable unitario de transporte, f_{ij} el coste fijo asociado a la decisión de inversión en el arco ij , a_i la oferta máxima de producto en el origen i , b_j la demanda del destino j , x_{ij} la variable que indica el flujo que recorre el arco ij , y_{ij} la variable que representa la decisión de inversión en el arco ij y M_{ij} una cota superior de cualquier flujo en dicho arco ij (por ejemplo, $M_{ij} = \min\{a_i, b_j\}$).

Las variables y_{ij} son binarias. Una vez conocidas el problema anterior es un problema clásico de transporte. Las variables y_{ij} son las variables que complican la resolución y, por consiguiente, son asignadas al problema maestro en un entorno de descomposición de Benders. El subproblema se formula de la siguiente manera

$$\begin{aligned} & \min_{x_{ij}} \sum_{ij} c_{ij}x_{ij} \\ & \sum_j x_{ij} \leq a_i \\ & \sum_i x_{ij} \geq b_j \\ & x_{ij} \leq M_{ij}y_{ij}^j \quad : \pi_{ij} \\ & x_{ij} \geq 0 \end{aligned} \tag{6.23}$$

y el problema maestro como

$$\begin{aligned} & \min_{y_{ij}, \theta} \sum_{ij} (\theta + f_{ij}y_{ij}) \\ & \theta + \sum_{ij} (\pi_{ij}^l M_{ij}y_{ij}) \geq f^l + \sum_{ij} (\pi_{ij}^l M_{ij}y_{ij}^l) \quad l = 1, \dots, k \\ & y_{ij} \in \{0, 1\} \end{aligned} \tag{6.24}$$

A continuación se expresa en GAMS este problema para un caso ejemplo en el que se suponen cuatro orígenes del producto y tres puntos de demanda. El

problema debe decidir la combinación óptima de arcos de entre todos los posibles, dados en la figura 2.

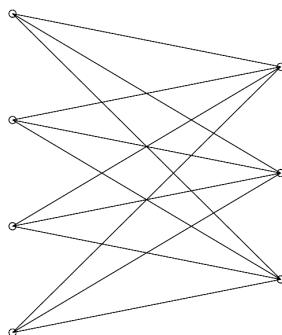


Fig. 2: Arcos posibles.

```
$TITLE Problema de transporte con coste fijo (Bd)
```

```
OPTION OPTCR = 0
```

```
SETS
  J      indice de iteraciones / j1 * j20 /
  JJ(j) subconjunto de iteraciones
  I      orígenes           / i1 * i4 /
  L      destinos          / i1 * i3 /
```

```
* Datos del problema
```

```
PARAMETERS
  A(i)    ofertas de producto
           / i1 10, i2 30, i3 40, i4 20 /
  B(l)    demandas de producto
           / l1 20, l2 50, l3 30 /
```

```
TABLE C(i,l) coste variable unitario de transporte
      11 12 13
  i1  1  2  3
  i2  3  2  1
  i3  2  3  4
  i4  4  3  2
```

```
TABLE F(i,l) coste fijo de transporte
      11 12 13
  i1  10 20 30
  i2  20 30 40
  i3  30 40 50
  i4  40 50 60
```

```

* Fin datos del problema

ABORT $(SUM[i, A(i)] < SUM[l, B(l)]) 'Problema infactible'

POSITIVE VARIABLES
  X(i,l)    flujo por los arcos

BINARY VARIABLE
  Y(i,l)    decisiones de inversión en los arcos

VARIABLES
  Z1        función objetivo primera etapa
  Z2        función objetivo segunda etapa
  theta     función de recurso

SCALARS
  TOL      tolerancia relativa / 1e-6 /
  Z_INF    cota inferior      / -INF /
  Z_SUP    cota superior      / INF /

PARAMETERS
  Y_J(i,l,j)  valores de las variables de la primera etapa en la iteración j
  PI(i,l,j)   variables duales restricciones segunda etapa en la iteración j
  DELTA(j)    tipo de corte (infactible 0 óptimo 1) de la iteración j
  Z2_J(j)    valor de la función objetivo del subproblema en la iteración j;

* Declaración de las ecuaciones

EQUATIONS
  EQ_Z1      función objetivo primera etapa
  EQ_Z2      función objetivo segunda etapa
  EQ_OBJ     función objetivo problema completo
  OFERTA(i)  ofertas de los orígenes
  DEMANDA(l) demanda de los destinos
  LIMITE(i,l) límite de uso del arco
  CORTES(j)  cortes de Benders ;

EQ_Z1      .. Z1 =E= SUM[(i,l), F(i,l)*Y(i,l)] + theta ;
EQ_Z2      .. Z2 =E= SUM[(i,l), C(i,l)*X(i,l)] ;
EQ_OBJ     .. Z1 =E= SUM[(i,l), F(i,l)*Y(i,l)] + SUM[(i,l),
C(i,l)*X(i,l)];
OFERTA(i)  .. SUM[l, X(i,l)] =L= A(i) ;
DEMANDA(l) .. SUM[i, X(i,l)] =G= B(l) ;

```

```

LIMITE(i,1) .. X(i,1) =L= MIN[A(i),B(1)] * Y(i,1) ;

CORTES(jj) .. DELTA(jj) * theta - Z2_J(jj) =G=
- SUM[(i,1), PI(i,1,jj) * MIN[A(i),B(1)] * (Y_J(i,1,jj) - Y(i,1))] ;

MODEL MAESTRO / EQ_Z1, CORTES / ;

MODEL SUBPROBLEMA / EQ_Z2, OFERTA, DEMANDA, LIMITE / ;

MODEL COMPLETO / EQ_OBJ, OFERTA, DEMANDA, LIMITE / ;

X.UP(i,1) = MIN[A(i),B(1)] ;

FILE COPT / cplex.opt / ;

* para que los resultados de un problema infactible sean los correctos con CPLEX sólo
* se puede utilizar el método simplex y sin preproceso ni escalado

PUT COPT PUT 'scaind -1' / 'lpmethod 1' / 'preind 0' / 'epopt
1.1e-9' / 'eprhs 1.1e-9' PUTCLOSE COPT ;

MAESTRO.OPTFILE = 1 ; SUBPROBLEMA.OPTFILE = 1 ;

* inicialización de parámetros del problema

JJ(j)      = NO ; theta.FX    = 0 ; DELTA(j)   = 0 ; Y_J(i,1,j) = 0
; Z2_J(j)   = 0 ; PI(i,1,j) = 0 ;

* Iteraciones del algoritmo
LOOP(j $(ABS(1-Z_INF/Z_SUP) > TOL),

*     Resolución del problema maestro
      SOLVE MAESTRO USING MIP MINIMIZING Z1 ;
*     Adquisición de la solución
      Y_J(i,1,j) = Y.L(i,1) ;

*     Fijación de la variable de la primera etapa y
*     Resolución del subproblema
      Y.FX(i,1) = Y.L(i,1) ;
      SOLVE SUBPROBLEMA USING RMIP MINIMIZING Z2;

*     Adquisición de los parámetros para formar un nuevo corte
      IF(SUBPROBLEMA.MODELSTAT = 4,
        DELTA(j) = 0 ;
        Z2_J(j) = SUBPROBLEMA.SUMINFES;
      ELSE
*       Actualización de la cota inferior

```

```

Z_INF = Z1.L;
DELTA(j) = 1 ;
theta.L0 = -INF; theta.UP = INF;
*   Actualización de la cota superior
Z_SUP = SUM[(i,1), F(i,1)*Y.L(i,1)] + SUM[(i,1), C(i,1)*X.L(i,1)];
Z2_J(j) = Z2.L;
) ;

PI(i,1,j) = LIMITE.M(i,1) ;
Y.L0(i,1) = 0 ; Y.UP(i,1) = 1 ;

*   Incremento del conjunto de cortes
JJ(j) = YES;
) ;

```

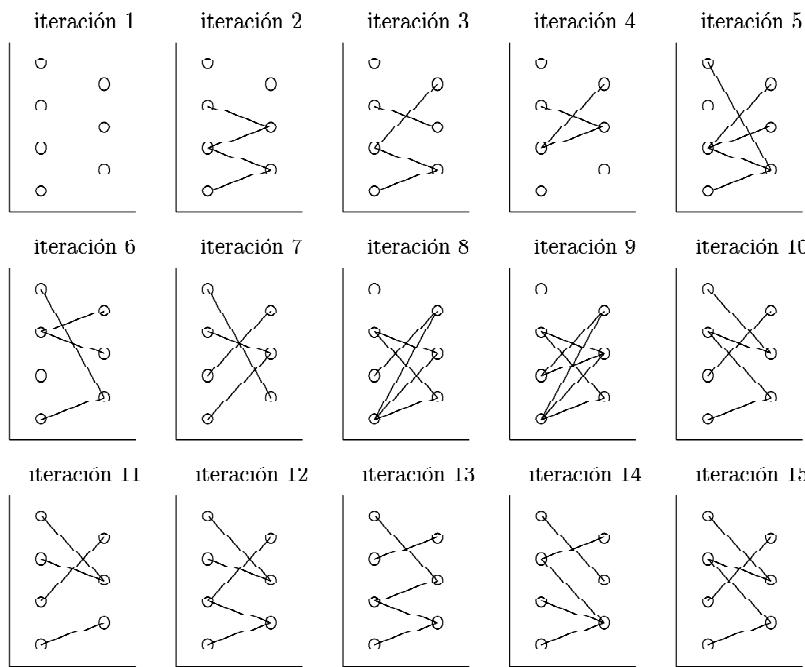


Fig. 3: Evolución de arcos propuestos por el algoritmo de Benders.

La solución óptima es $y_{11} = y_{23} = y_{31} = y_{32} = y_{42} = 1$ y se alcanza en 15 iteraciones con un coste total fijo más variable de 380. La evolución de las combinaciones de arcos propuestas por el algoritmo de descomposición se presenta en la figura 3. Debe destacarse que solamente las propuestas de las iteraciones 7, 10 y 15 son factibles para el subproblema, de modo que la mayoría de cortes que el algoritmo genera son de infactibilidad y sólo dos cortes son de optimalidad. Por

Iteración	Cota Inferior	Cota Superior
1	$-\infty$	∞
2	$-\infty$	∞
3	$-\infty$	∞
4	$-\infty$	∞
5	$-\infty$	∞
6	$-\infty$	∞
7	140	390
8	140	390
9	120	390
10	360	380
11	360	380
12	360	380
13	360	380
14	360	380
15	380	380

Tabla 6.1: Evolución de cotas en el algoritmo de Benders

último, el cuadro 6.1 presenta la evolución de la cota superior y la cota inferior del problema durante el algoritmo. En este ejemplo, el algoritmo converge cuando la cota inferior es exactamente la misma que la cota superior. La convergencia se alcanza también cuando aparece una propuesta repetida. En este ejemplo, la propuesta de arcos de la última iteración es de hecho la propuesta de la iteración 10.

3 Relajación Lagrangiana

La Relajación Lagrangiana [9] es una de las técnicas más extendidas en optimización discreta. Se emplea principalmente cuando en un problema de programación matemática aparece un conjunto de ecuaciones que *complica* la resolución del problema. Esto es, la resolución del problema sin esas ecuaciones tiene una estructura cuya resolución es más sencilla. Esta técnica se basa en la *dualización* de esas ecuaciones y en la formulación de un *problema dual* cuya resolución proporciona una aproximación del valor óptimo del problema original mejor que la resolución de su relajación lineal⁵. En caso de que el problema sea lineal y que el conjunto de ecuaciones de complicación sean el conjunto de restricciones del problema, este problema dual es el problema dual del problema original.

Se considera el siguiente problema de optimización lineal:

⁵Por relajación lineal de un problema P se entiende el problema P en el que las variables enteras son sustituidas por variables continuas.

$$\begin{aligned}
 & \min_x c^T x \\
 & Ax = b \\
 & Dx \leq d \\
 & x \geq 0
 \end{aligned} \tag{6.25}$$

y supongamos que $Ax = b$ son el conjunto de restricciones de complicación. Supondremos que $A \in \mathbb{R}^{m_1 \times n}$, $D \in \mathbb{R}^{m_2 \times n}$ y que el resto de dimensiones son conformes con éstas.

Dado $\lambda \in \mathbb{R}^{m_1}$ se formula la función de Lagrange $L(x, \lambda)$ como:

$$L(x, \lambda) = c^T x + \lambda^T (Ax - b) \tag{6.26}$$

y la función dual $\omega(\lambda)$ como evaluación del siguiente problema, *subproblema de Lagrange*:

$$\begin{aligned}
 \omega(\lambda) = & \min_x c^T x + \lambda^T (Ax - b) \\
 & Dx \leq d \\
 & x \geq 0
 \end{aligned} \tag{6.27}$$

Para todo $\lambda \in \mathbb{R}^{m_1}$ es inmediato observar que $\omega(\lambda) \leq c^T x^*$, siendo x^* la solución óptima del problema (25). El problema dual consiste en buscar el vector de multiplicadores λ para el cual esta cota inferior dada por la función dual es máxima:

$$\max_{\lambda \in \mathbb{R}^{m_1}} \omega(\lambda) \tag{6.28}$$

La función dual es cóncava independientemente de que el problema original sea lineal o lineal entero mixto. Por esta razón, las técnicas basadas en subgradientes aparecen de modo natural a la hora de maximizar la función dual [13, 12]. Estas técnicas actualizan los valores de los multiplicadores siguiendo la dirección del subgradiente, utilizando diferentes longitudes de paso en cada iteración [1]. Nótese que si x^j es la solución óptima del subproblema de Lagrange para un valor λ^j del multiplicador, $Ax^j - b$ es un subgradiente de la función dual. Una alternativa a las técnicas basadas en el subgradiente es la representación lineal del problema dual, lo que induce de modo natural un algoritmo de aproximación exterior similar al algoritmo de descomposición de Benders. La próxima sección presenta este algoritmo que posteriormente es formulado en GAMS.

La Relajación Lagrangiana recibe también el nombre de *descomposición dual* o *descomposición por precios* porque se basa en la evaluación del subproblema para distintos precios o variables duales.

Se describe a continuación la formulación del algoritmo de aproximación exterior de la función dual, que denominaremos algoritmo de descomposición basado en Relajación Lagrangiana.

3.1 Representación lineal del problema dual

Consideremos de nuevo el subproblema de Lagrange que evalúa la función dual $\omega(\lambda)$:

$$\begin{aligned}\omega(\lambda) = \min_x & c^T x + \lambda^T (Ax - b) \\ & Dx \leq d \\ & x \geq 0\end{aligned}\tag{6.29}$$

En caso de disponer de forma explícita del conjunto de soluciones factibles de la región $\{Dx \leq d, x \geq 0\}$, el anterior problema podría ser resuelto mediante enumeración de todas ellas como:

$$\omega(\lambda) = \min_l c^T x^l + \lambda^T (Ax^l - b) \quad l = 1, \dots, \nu\tag{6.30}$$

La expresión anterior indica que la función dual es cóncava y que el problema dual puede reformularse como el siguiente problema lineal, denominado *problema maestro de la Relajación Lagrangiana*:

$$\begin{aligned}\max & \omega \\ \omega \leq & c^T x^1 + \lambda^T (Ax^1 - b) \\ \vdots \\ \omega \leq & c^T x^\nu + \lambda^T (Ax^\nu - b)\end{aligned}\tag{6.31}$$

Cada restricción del problema anterior se denomina *Corte de Lagrange*. Dado que para problemas realistas disponer del conjunto de soluciones es prácticamente imposible, la optimización del problema dual se reemplaza por la resolución iterativa de problemas maestros relajados⁶, cuyo número de cortes de Lagrange aumenta con cada iteración. Cada resolución de un problema maestro relajado propone un nuevo valor del multiplicador λ que, una vez evaluado en la función dual a través del subproblema de Lagrange, propone un nuevo corte de Lagrange que aumenta el problema maestro relajado⁷.

De modo similar a la posibilidad de subproblemas infactibles en el algoritmo de descomposición de Benders, el algoritmo de descomposición debe afrontar la posibilidad de encontrar subproblemas de Lagrange no acotados. Cuando esto ocurre, el multiplicador propuesto no es válido y se debe ser introducir una restricción en el problema maestro que lo elimine del conjunto de multiplicadores posibles. Este tipo de corte se denota *Corte de Acotamiento* y su construcción se comenta en la siguiente sección.

⁶Relajado hace referencia a que no se dispone de la totalidad de los cortes, sino solamente de un conjunto de ellos

⁷Por simplicidad de la exposición, se denomina como problema maestro al problema maestro relajado.

3.2 Cortes de acotamiento

Consideremos un valor del multiplicador λ y consideremos de nuevo el subproblema de Lagrange

$$\begin{aligned}\omega(\lambda) = \min_x & c^T x + \lambda^T (Ax - b) \\ & Dx \leq d \\ & x \geq 0\end{aligned}\tag{6.32}$$

Supongamos que para este valor λ del multiplicador existe $v \geq 0$ y $Dv \leq 0$ ⁸ tal que $c^T - \lambda^T A < 0$. Entonces, si x_0 es una solución del subproblema de Lagrange, observando que $tv \geq 0$ y $Dtv \leq 0$ para cualquier valor positivo de t se tiene que

$$c^T(x_0 + tv) + \lambda^T(A(x_0 + tv) - b) = c^T x_0 + \lambda^T A(x_0 - b) + t(c^T - \lambda^T A)v \tag{6.33}$$

con lo que el subproblema no está acotado dado que toma valores muy pequeños cuando t tiende a infinito.

Del razonamiento anterior se deduce que el conjunto de multiplicadores aceptables debe verificar que

$$(c^T + \lambda^T A)x \geq 0 \quad \forall x \in \{x \geq 0, Dx \leq 0\} \tag{6.34}$$

La condición anterior, una vez propuesto un valor para el multiplicador, puede ser comprobada resolviendo el siguiente problema, que denotamos *subproblema de acotamiento*:

$$\begin{aligned}\omega^*(\lambda) = \min_x & c^T x + \lambda^T Ax \\ & Dx \leq 0 \\ & 0 \leq x \leq 1\end{aligned}\tag{6.35}$$

y en caso de que tenga un valor negativo, se debe ser introducir un corte de acotamiento en el problema maestro de la Relajación Lagrangiana de la forma:

$$0 \leq c^T x^j + \lambda^T A x^j \tag{6.36}$$

De este modo, el algoritmo de la Relajación Lagrangiana itera entre un problema maestro, formado por cortes de Lagrange y cortes de acotamiento, y un subproblema de Lagrange que evalúa los multiplicadores propuestos por el maestro. El problema maestro, considerando ambos tipo de corte, puede ser formulado de la siguiente forma

$$\begin{aligned}\max & \omega \\ \delta^j \omega \leq & c^T x^j + \lambda^T (Ax^j - \delta^j b) \quad j = 1, \dots, k\end{aligned}\tag{6.37}$$

⁸Es decir $v \in \{x \geq 0, Dx \leq 0\}$ que es el sistema homogéneo asociado a $\{x \geq 0, Dx \leq d\}$

considerando $\delta^j = 1$ para los cortes de Lagrange y $\delta^j = 0$ para los cortes de acotamiento. Con esta formulación, el algoritmo de Relajación Lagrangiana se describe esquemáticamente en la siguiente sección.

3.3 Algoritmo de Relajación Lagrangiana

El algoritmo de la Relajación Lagrangiana, considerando la interpretación lineal del problema dual anterior, es descrito en los siguientes pasos.

1. Inicialización: $j = 0, \epsilon = 10^{-4}$
2. Resolución del problema maestro de la Relajación Lagrangiana

$$\begin{array}{ll} \max_{\lambda, \omega} & \omega \\ \text{s.t.} & \delta^j \omega \leq c^T x^j + \lambda^T (Ax^j - \delta^j b) \quad j = 1, \dots, k \end{array} \quad (6.38)$$

Obtener valor de λ e ir al paso 3

3. Resolución del problema de acotamiento

$$\begin{array}{ll} \omega^*(\lambda) = & \min_x c^T x + \lambda^T Ax \\ & Dx \leq 0 \\ & 0 \leq x \leq 1 \end{array} \quad (6.39)$$

Si $\omega^*(\lambda) \geq 0$ ir al paso 4. En otro caso obtener la solución x^j y formar corte de acotamiento

$$0 \leq c^T x^j + \lambda^T Ax^j \quad (6.40)$$

Ir al paso 2.

4. Resolución del subproblema de Lagrange

$$\begin{array}{ll} \omega(\lambda) = & \min_x c^T x + \lambda^T (Ax - b) \\ & Dx \leq d \\ & x \geq 0 \end{array} \quad (6.41)$$

Obtener la solución x^j , y formar corte de Lagrange

$$\omega \leq c^T x^j + \lambda^T (Ax^j - b) \quad (6.42)$$

Ir al paso 5.

5. Regla de parada

Cálculo de la diferencia entre los multiplicadores propuestos en la iteración j y la anterior $j - 1$.

Si $d(\lambda^j - \lambda^{j-1}) < \epsilon$ detener⁹.

En otro caso, ir al paso 2.

En el algoritmo de la Relajación Lagrangiana, los multiplicadores λ en el problema maestro están acotados por valores suficientemente amplios que impiden que éste sea no acotado. El problema lineal resuelto por Benders es ahora resuelto mediante el algoritmo de Relajación Lagrangiana para ver en detalle los pasos del algoritmo.

3.4 Ejemplo

Se desea resolver el siguiente problema lineal:

$$\begin{aligned} \min_{x_1, x_2, y_1, y_2} \quad & -x_1 - 2x_2 - 2y_1 - 3y_2 \\ & x_1 + x_2 \leq 600 \\ & x_1 - 2x_2 \leq 0 \\ & x_1 + x_2 + y_1 + y_2 \leq 1000 \\ & x_1 + y_1 \leq 500 \\ & -2y_1 + y_2 \leq 0 \\ & x_1, x_2, y_1, y_2 \geq 0 \end{aligned}$$

entonces,

$$c = \begin{pmatrix} -1 \\ -2 \\ -2 \\ -3 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -2 & 0 & 0 \\ 0 & 0 & -2 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 600 \\ 0 \\ 0 \end{pmatrix},$$

$$D = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \text{ y } d = \begin{pmatrix} 1000 \\ 500 \end{pmatrix}$$

esto es, se está considerando que las ecuaciones $x_1 + x_2 + x_3 + x_4 \leq 1000$ y $x_1 + x_4 \leq 500$ son de complicación.

Para este ejemplo académico, codificado en GAMS tal como se presenta en la próxima sección, las variables λ que representan los multiplicadores han sido acotadas entre 0 y 10. El algoritmo comienza resolviendo un problema maestro

⁹ $d(\lambda^j - \lambda^{j-1})$ representa distancia entre λ^j y λ^{j-1}

de Relajación Lagrangiana sin ningún corte, proponiendo un primer valor para los multiplicadores dado por

$$\lambda = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Este valor de los multiplicadores no es adecuado y el subproblema de acotamiento lo rechaza introduciendo un corte en el problema maestro dado por

$$0 \leq -5 + 2\lambda_1 + 2\lambda_2$$

Una nueva resolución del problema maestro propone

$$\lambda = \begin{pmatrix} 2.5 \\ 0 \end{pmatrix}$$

y el subproblema de acotamiento

$$0 \leq -4 + 1.5\lambda_1 + 0.5\lambda_2$$

con lo que una nueva resolución del maestro obtiene

$$\lambda = \begin{pmatrix} 2.667 \\ 0 \end{pmatrix}$$

Este valor de multiplicadores es aceptado por el problema de acotamiento y el algoritmo pasa a resolver el subproblema de Lagrange. Una vez resuelto, éste propone un corte de Lagrange para el problema maestro que es

$$\omega \leq -1000\lambda_1 - 500\lambda_2$$

La nueva resolución del maestro propone el mismo multiplicador de Lagrange, por lo que el algoritmo termina.

Para problemas lineales, el algoritmo de Relajación Lagrangiana termina con una solución que es óptima para el dual del problema lineal. Esta solución viene dada como el opuesto de este multiplicador óptimo. Este hecho puede apreciarse en este ejemplo. Los valores de las variables duales de las ecuaciones relajadas son (cuando se resuelve el problema completo):

$$\pi = \begin{pmatrix} -2.667 \\ 0 \end{pmatrix}$$

Una desventaja de este método de descomposición es la pérdida de factibilidad de la solución primal durante las sucesivas iteraciones del método. En este ejemplo, la última solución primal, aunque factible ($x_1 \ x_2 \ x_3 \ x_4$) = (0 0 0 0) no

es óptima. Sin embargo, puede comprobarse que la solución óptima, ($x_1 \ x_2 \ x_3 \ x_4$) = (0 0 333.333 666.667) es una solución alternativa del último subproblema de Lagrange resuelto.

Cuando la Relajación Lagrangiana se aplica a problemas lineales, la sucesión de los valores óptimos del problema maestro (que siempre es decreciente) converge al valor óptimo del problema lineal. Por el contrario, esto no se satisface cuando el problema es entero mixto. Este valor al que la sucesión decreciente converge es simplemente una cota inferior del valor óptimo del problema. Esta diferencia entre el valor óptimo del problema y el valor obtenido por la Relajación Lagrangiana es lo que se conoce en la literatura como *intervalo de dualidad (duality gap)*. La aparición del intervalo de dualidad origina que la solución primal propuesta por la Relajación Lagrangiana pueda ser infactible. Una alternativa para evitar esta situación consiste normalmente en realizar un postprocesado de las soluciones obtenidas para encontrar la solución factible. Este postprocesado depende del problema concreto que se resuelva y suele estar basado en el conocimiento específico de dicho problema. Otra alternativa para obtener una solución factible consiste en aumentar la función objetivo del subproblema de Lagrange con un término cuadrático que penalize las infactibilidades asociadas a las ecuaciones relajadas. Esta técnica es conocida como *Relajación Lagrangiana Aumentada* y existen numerosas aplicaciones en la literatura.

Las próximas secciones presentan la implantación en GAMS del algoritmo de descomposición basado en Relajación Lagrangiana para el ejemplo académico anterior y para el problema de coste fijo que anteriormente ha sido resuelto mediante la descomposición de Benders.

3.5 Relajación Lagrangiana en GAMS

La implantación en GAMS de un código de Relajación Lagrangiana es similar a la implantación de la descomposición de Benders. Se formula un modelo para el problema maestro y otro para el subproblema, y éstos son resueltos sucesivamente en un bucle de iteraciones. Las ecuaciones que representan los cortes de Lagrange (una única ecuación que engloba a los dos tipos de corte) es declarada sobre un conjunto estático de índices y construida sobre un conjunto dinámico. Durante el proceso algorítmico, este conjunto dinámico se va actualizado, aumentando por tanto el número de cortes de Lagrange que contiene el problema maestro.

```
$TITLE Relajación Lagrangiana en GAMS (RL)

OPTION OPTCR = 0

SETS
  J           índice de iteraciones /j1 * j100 /
```

```

JJ(j)      subconjunto de índices
I          índice de variables /i1 * i4 /
M          índice de restricciones /m1 * m2/
N          índice de restricciones /n1 * n3/

* Datos del problema

TABLE A(m,i) matriz A
   i1 i2 i3 i4
m1  1  1  1  1 m2  1  0  1  0

TABLE D(n,i) matriz D
   i1 i2 i3 i4
n1  1  1 n2  1 -2 n3      -2  1

PARAMETERS
  C(i)      coeficientes objetivo
            / i1 -1, i2 -2, i3 -2, i4 -3 /
  B(m)      término B
            / m1 1000, m2 500 /
  Dd(n)     matriz d (notación matemática)
            / n1 600, n2 0, n3 0 /

* Fin datos del problema

SCALARS
  TOL       tolerancia relativa / 1e-6 /
  Z_INF     cota inferior      / -INF /
  Z_SUP     cota superior      / INF /
  DELTA     0 para subproblema de acotamiento 1 para Lagrange
  DIF       diferencia en iteraciones

PARAMETERS
  W_J(j)    función objetivo del problema dual en la iteración j
  X_J(i,j)  valores de las variables de flujo en la iteración j
  DELTAJ(j) tipo de corte (acotación 0 óptimo 1) de la iteración j
  LAMBDA_J(m,j) multiplicadores en la iteración j

POSITIVE VARIABLES
  X(i)      variables
  LAMBDA(m) multiplicador

VARIABLES
  Z          variable objetivo primal (subproblema)
  W          variable dual

* Declaración de las ecuaciones
EQUATIONS
  EQ_OBJ    función objetivo problema
  EQ_R1(m)  restricciones que se relajan
  EQ_R2(n)  resto de restricciones
  CORTES_LR(j) cortes de Relajación Lagrangiana

```

```

CORTE_RE      corte redundante;

EQ_OBJ .. Z =E= SUM(i, C(i)*X(i)) + SUM(m,
LAMBDA.L(m)*(SUM(i,A(m,i)*X(i))-DELTA*B(m))) ;

EQ_R1(m) .. SUM(i,A(m,i)*X(i)) =L= B(m) ;

EQ_R2(n) .. SUM(i,D(n,i)*X(i)) =L= DELTA*Dd(n) ;

CORTES_LR(jj).. DELTAJ(jj)*W =L= SUM(i, C(i)*X_J(i,jj)) +
SUM(m, LAMBDA(m)*(SUM(i,A(m,i)*X_J(i,jj))-DELTAJ(jj)*B(m))) ;

CORTE_RE .. W =L= 10000 ;

MODEL MAESTRO_LR / CORTES_LR , CORTE_RE / ;

MODEL SUB_LR      / EQ_OBJ, EQ_R2 / ;

MODEL COMPLETO    / EQ_OBJ, EQ_R1, EQ_R2 / ;

FILE COPT / cplex.opt / ;

PUT COPT PUT 'scaind -1' / 'lpmethod 1' / 'preind 0' / 'epopt
1.1e-9' / 'eprhs 1.1e-9' PUTCLOSE COPT ;

MAESTRO_LR.OPTFILE = 1 ; SUB_LR.OPTFILE = 1 ; COMPLETO.OPTFILE = 1
;

* inicialización de parámetros del problema

JJ(j)      = NO ; DELTAJ(j)     = 0 ; DELTA        = 0 ;
X_J(i,j)   = 0 ; LAMBDA.L0(m) = 0 ; LAMBDA.UP(m) = 10 ; DIF =
INF;

* Iteraciones del algoritmo
LOOP(j $(DIF>TOL),

* Resolución del problema maestro de Relajación Lagrangiana
IF(ORD(j) > 1,
  SOLVE MAESTRO_LR USING LP MAXIMIZING W;
  W_J(j) = W.L;
ELSE
  LAMBDA.L(m) = 0 ;
) ;
LAMBDA_J(m,j) = LAMBDA.L(m) ;

* Resolución del subproblema de acotamiento

```

```

* Normalización de las cotas no nulas de las variables
X.UP(i) = 1 ;
DELT A = 0 ;
SOLVE SUB_LR USING RMIP MINIMIZING Z;

* Si la solución es negativa, generar un corte de acotamiento
IF(Z.L < -TOL,
  X_J(i,j) = X.L(i) ;
  DELTAJ(j) = 0 ;

* En caso contrario resolver el subproblema de Relajación Lagrangiana
ELSE
  X.UP(i) = INF ;
  DELTA = 1 ;
  SOLVE SUB_LR USING RMIP MINIMIZING Z;
  X_J(i,j) = X.L(i) ;
  DELTAJ(j) = 1 ;
) ;

* Actualización del conjunto de cortes de Lagrange
JJ(j) = YES;

* Actualización de la diferencia de multiplicadores
IF(ORD(j)>1,
  DIF = SUM(m, MAX(LAMBDA_J(m,j)-LAMBDA_J(m,j-1),
    LAMBDA_J(m,j-1)-LAMBDA_J(m,j))) ;
) ;
) ;

X.UP(i) = INF ; LAMBDA.FX(m) = 0 ; DELTA = 1 ; SOLVE COMPLETO
USING RMIP MINIMIZING Z ;

```

Debe destacarse la necesidad de resolver independientemente el subproblema de acotamiento y el subproblema de Lagrange. En el código presentado esto se ha simplificado introduciendo un scalar (**DELT A**) que, declarado como 0 o bien como 1, genera el subproblema de acotamiento o el subproblema de Lagrange respectivamente. Otra característica de este código es la introducción de una ecuación redundante para el problema maestro $\omega \leq 10000$. Esta ecuación es introducida en el maestro para que la variable objetivo que es maximizada, ω , no desaparezca del problema. La desaparición de la variable objetivo de un problema es interpretado por este lenguaje como un error.

En la siguiente sección se presenta un código en GAMS orientado a resolver el problema de coste fijo anteriormente descrito mediante Relajación Lagrangiana.

3.6 Problema de coste fijo mediante Relajación Lagrangiana

Consideremos el problema de transporte con coste fijo de la sección 2 y supongamos que son las restricciones de capacidad límite las que son relajadas e introducidas en la función objetivo a través del lagrangiano. Para un valor λ_{ij} el subproblema de Lagrange presenta la forma:

$$\begin{aligned} & \min_{x_{ij}} \sum_{ij} (c_{ij}x_{ij} + f_{ij}y_{ij}) + \lambda_{ij}(x_{ij} - M_{ij}y_{ij}) \\ & \sum_j x_{ij} \leq a_i \\ & \sum_i x_{ij} \geq b_j \\ & x_{ij} \geq 0, y_{ij} \in \{0, 1\} \end{aligned} \tag{6.43}$$

que reformulado queda como

$$\begin{aligned} & \min_{x_{ij}} \sum_{ij} (c_{ij} + \lambda_{ij})x_{ij} + (f_{ij} - \lambda_{ij}M_{ij})y_{ij} \\ & \sum_j x_{ij} \leq a_i \\ & \sum_i x_{ij} \geq b_j \\ & x_{ij} \geq 0, y_{ij} \in \{0, 1\} \end{aligned} \tag{6.44}$$

Esta reformulación permite observar que el subproblema es separable en dos problemas. Un problema de transporte en el que el coste variable es ligeramente modificado por el multiplicador y un segundo problema que es entero puro y cuya solución puede obtenerse de modo inmediato.

Problema de transporte

$$\begin{aligned} & \min_{x_{ij}} \sum_{ij} (c_{ij} + \lambda_{ij})x_{ij} \\ & \sum_j x_{ij} \leq a_i \\ & \sum_i x_{ij} \geq b_j \\ & x_{ij} \geq 0 \end{aligned} \tag{6.45}$$

Problema entero puro

$$\begin{aligned} & \min_{y_{ij}} \sum_{ij} (f_{ij} - \lambda_{ij}M_{ij})y_{ij} \\ & y_{ij} \in \{0, 1\} \end{aligned} \tag{6.46}$$

Esta formulación de la Relajación Lagrangiana para el problema del transporte con coste fijo puede implantarse en GAMS con el siguiente código.

```
$TITLE Problema de transporte con coste fijo (RL)
```

```
OPTION OPTCR = 0
```

```

SETS
  J      índice de iteraciones / j1 * j100 /
  JJ(j)  subconjunto de índices
  I      orígenes          / i1 * i4 /
  L      destinos         / l1 * l3 /

* Datos del problema

PARAMETERS
  A(i)    ofertas de producto
          / i1 10, i2 30, i3 40, i4 20 /
  B(l)    demandas de producto
          / l1 20, l2 50, l3 30 /

TABLE C(i,l) coste variable unitario de transporte
      11 12 13
  i1  1  2  3
  i2  3  2  1
  i3  2  3  4
  i4  4  3  2

TABLE F(i,l) coste fijo de transporte
      11 12 13
  i1 10 20 30
  i2 20 30 40
  i3 30 40 50
  i4 40 50 60

* Fin datos del problema

ABORT ${SUM[i, A(i)] < SUM[l, B(l)]} 'Problema infactible'

SCALARS
  TOL      tolerancia relativa / 1e-6 /
  Z_INF    cota inferior      / -INF /
  Z_SUP    cota superior      / INF /
  DIF      diferencia en iteraciones

PARAMETERS
  X_J(i,l,j)  valores de las variables de flujo   en la iteración j
  Y_J(i,l,j)  valores de las variables de inversión en la iteración j
  DELTA(j)    tipo de corte (acotación O óptimo 1) de la iteración j

POSITIVE VARIABLES
  X(i,l)    flujo por los arcos
  LAMBDA(i,l) multiplicador

BINARY VARIABLE
  Y(i,l)    decisiones de inversión en los arcos

VARIABLES
  Z        variable objetivo primal (subproblema)

```

```

W           variable dual

* Declaración de las ecuaciones
EQUATIONS
EQ_OBJ      función objetivo problema completo
OFERTA(i)   ofertas de los orígenes
DEMANDA(l)  demanda de los destinos
LIMITE(i,l) límite de uso del arco
CORTES_LR(j) cortes de Relajación Lagrangiana ;

EQ_OBJ .. Z =E= SUM[(i,l), F(i,l)*Y(i,l)] + SUM[(i,l),
C(i,l)*X(i,l)]
          + SUM[ (i,l), LAMBDA.L(i,l)*(X(i,l)-MIN[A(i),B(l)] * Y(i,l))];

OFERTA(i) .. SUM[l, X(i,l)] =L= A(i) ;

DEMANDA(l) .. SUM[i, X(i,l)] =G= B(l) ;

LIMITE(i,l) .. X(i,l) =L= MIN[A(i),B(l)] * Y(i,l) ;

CORTES_LR(jj).. DELTA(jj)*W =L= SUM[(i,l), F(i,l)*Y_J(i,l,jj)] +
SUM[(i,l), C(i,l)*X_J(i,l,jj)] +
SUM[(i,l), LAMBDA(i,l)*(X_J(i,l,jj)-MIN[A(i),B(l)] * Y_J(i,l,jj))];

MODEL MAESTRO_LR / CORTES_LR /;

MODEL SUB_LR      / EQ_OBJ, OFERTA, DEMANDA /;

MODEL COMPLETO    / EQ_OBJ, OFERTA, DEMANDA, LIMITE /;

X.UP(i,l) = MIN[A(i),B(l)] ;

FILE COPT / cplex.opt /;

* para que los resultados de un problema infactible sean los correctos con CPLEX sólo
* se puede utilizar el método simplex y sin preprocess

PUT COPT PUT 'scaind -1' / 'lpmethod 1' / 'preind 0' / 'epopt
1.1e-9' / 'eprhs 1.1e-9' PUTCLOSE COPT;

MAESTRO_LR.OPTFILE = 1 ; SUB_LR.OPTFILE = 1 ; COMPLETO.OPTFILE = 1
;

* inicialización de parámetros del problema

JJ(j)      = NO ; DELTA(j)      = 0 ; Y_J(i,l,j)      = 0 ;
X_J(i,l,j) = 0 ; LAMBDA.L0(i,l) = 0 ; LAMBDA.UP(i,l) = 10 ;

```

```

DIF = INF;

* Iteraciones del algoritmo
LOOP(j $(DIF>TOL),

      * Resolución del problema maestro de Relajación Lagrangiana
      IF(ORD(j) > 1,
         SOLVE MAESTRO_LR USING LP MAXIMIZING W;
      ELSE
         LAMBDA.L(i,1) = 0 ;
      )
      LAMBDA_J(i,1,j) = LAMBDA.L(i,1) ;

      * Resolución del subproblema de acotamiento
      * Normalización de las cotas no nulas de las variables
      X_UP(i,1) = 1 ;
      SOLVE SUB_LR USING RMIP MINIMIZING Z ;

      * Si la solución es negativa, generar un corte de acotamiento
      IF(Z.L < 0,
         X_J(i,1,j) = X.L(i,1) ;
         Y_J(i,1,j) = Y.L(i,1) ;
         DELTA(j) = 0 ;
      )

      * En caso contrario resolver el subproblema de Relajación Lagrangiana
      ELSE
         X_UP(i,1) = MIN[A(i),B(1)] ;
         SOLVE SUB_LR USING MIP MINIMIZING Z;
         X_J(i,1,j) = X.L(i,1) ;
         Y_J(i,1,j) = Y.L(i,1) ;
         DELTA(j) = 1 ;
      )

      * Actualización del conjunto de cortes de Lagrange
      JJ(j) = YES;

      * Actualización de la diferencia de multiplicadores
      IF(ORD(j)>1,
         DIF = SUM((i,1), MAX(LAMBDA_J(i,1,j)-LAMBDA_J(i,1,j-1),
                               LAMBDA_J(i,1,j-1)-LAMBDA_J(i,1,j))) ;
      )
   )

X_UP(i,1) = MIN[A(i),B(1)] ; LAMBDA.FX(i,1) = 0 ; SOLVE COMPLETO
USING MIP MINIMIZING Z ;

```

En este ejemplo, la sucesión de valores del problema maestro y la sucesión de óptimos de los subproblemas convergen a la solución dada por la relajación lineal del problema, véase la figura. En este caso, la cota inferior del valor óptimo del problema dada por la relajación lineal del mismo no puede ser superada por la aplicación de la Relajación Lagrangiana¹⁰. Por otra parte, los valores obtenidos

¹⁰Cuando esto ocurre se dice que el problema satisface la *propiedad de integralidad* (*integrality*

por las variables de decisión no son factibles en ninguna de las iteraciones del método. Se necesita un método de postprocesado de estas soluciones para obtener una solución. En este problema, una posibilidad consiste en considerar la solución dada para las variables continuas x_{ij} por el problema de transporte (lineal), y adecuar las variables binarias a esta solución, es decir

$$u_{ij} = 1 \text{ si } x_{ij} > 0$$

Con este procesado de las soluciones puede comprobar que en la iteración 51 (el algoritmo converge en la iteración 55) la solución obtenida es en este caso la solución óptima del problema entero mixto:

$$y_{11} = y_{23} = y_{31} = y_{32} = y_{42} = 1$$

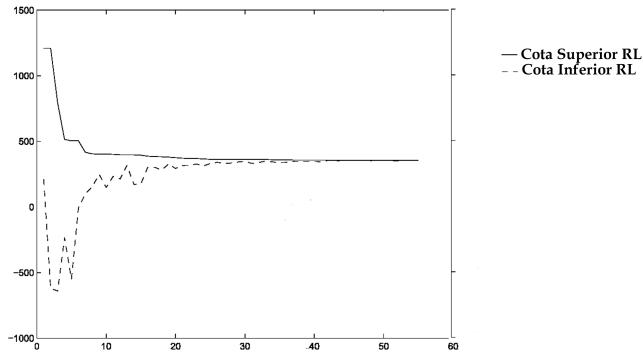


Fig. 4: Evolución de cotas en Relajación Lagrangiana.

4 Implantación en grandes modelos

GAMS fue desarrollado para permitir la implantación rápida de modelos de optimización concentrando la labor del modelador en su formulación. Además, dada la potencia del lenguaje permite la creación de modelos complejos con un número reducido de instrucciones. Esto hace que el mantenimiento o la modificación de un modelo sea muy sencilla. De la misma manera estas ventajas se extienden a la escritura de métodos de descomposición. Sin embargo, el tiempo de resolución de un problema en GAMS por métodos de descomposición es muy elevado. Las razones son las siguientes:

property) y tiene poco sentido el uso de la Relajación Lagrangiana como técnica de aproximación del valor óptimo del problema.

- el paso de un problema definido en GAMS al optimizador encargado de su resolución se hace escribiéndolo en disco en lugar de a través de memoria principal. Esto hace que se consuma mucho tiempo cuando esta tarea se debe hacer repetidas veces para cada maestro o subproblema. Este inconveniente se puede soslayar parcialmente recurriendo a la creación de un disco virtual en memoria principal mediante un controlador RAMDISK.
- cada maestro o subproblema se crea de nuevo cada vez aunque las modificaciones entre una iteración y la siguiente para cada problema son menores. Precisamente este idea de resolución iterativa de problemas muy similares es el núcleo central de la descomposición y es la que se debería poder aprovechar computacionalmente tanto en la fase de creación del problema como en el algoritmo de resolución (mediante el simplex dual). GAMS genera desde cero cada problema aunque sí permite guardar la base previa.

En conclusión, GAMS puede ser utilizado para el desarrollo de prototipos de modelos de optimización y de métodos de descomposición. Sin embargo, una vez validados estos prototipos se deben utilizar lenguajes alternativos de menores tiempos de computación.

5 Bibliografía

- [1] Bazaraa, M.S. "On the choice of step sizes in subgradient optimization" *Journal of Operational Research*. Vol 7, pp 380-388.
- [2] Benders, J.F. "Partitioning Procedures for Solving Mixed-Variable Programming Problems" *Numerische Mathematik*. Vol 4, pp 238-252. 1962.
- [3] Birge, J.R.. and Louveaux, F.V. "A Multicut Algorithm for Two-Stage Stochastic Linear Programs". *European Journal of Operational Research*. Vol 34, No 3, pp 384-392. March 1988.
- [4] Brooke, A., Kendrick, D. and Meeraus, A. *GAMS A User's Guide*. The Scientific Press. South San Francisco, USA. 1988.
- [5] Dantzig, G.B. and Thapa, M.N. *Linear Programming, Extensions & Software*. Unpublised Book. September 1991.
- [6] Dantzig, G.B. and Wolfe, P. "Decomposition Principle for Linear Programs" *Operations Research*. Vol 8, No 1, pp 101-111. January-February 1960.
- [7] Ermoliev, Y. and Wets, R.J-B. (eds.) *Numerical Techniques for Stochastic Optimization*. Springer-Verlag. Berlin, Germany. 1988.
- [8] Geoffrion, A.M. "Elements of Large-Scale Mathematical Programming" *Management Science*. Vol 16, No 11, pp 652-691. July 1970.

- [9] Geoffrion, A.M. "Lagrangean relaxation for integer programming" *Mathematical Programming Study*. Vol. 2, pp. 82-114.
- [10] ILOG, "<http://www.ilog.com>." 2003.
- [11] Infanger, G. "Monte Carlo (Importance) Sampling within a Benders Decomposition Algorithm for Stochastic Linear Programs. Extended Version: Including Results of Large-Scale Problems" *Systems Optimization Laboratory. Department of Operations Research. Stanford University*. SOL 91-6. March 1991.
- [12] Jiménez, N. "Coordinación Hidrotérmica en el Corto Plazo Mediante Técnicas de Relajación Lagrangiana" *Ph.D. Thesis, Universidad de Málaga* Málaga, 1999.
- [13] Kiwiel, K. C. "Proximity control in bundle methods for convex nondifferentiable minimization" *Mathematical Programming*. Vol. 46, pp. 105-122. 1990.
- [14] Van Slyke, R.M. and Wets, R.J.-B. "L-Shaped Linear Programs with Application to Optimal Control and Stochastic Programming" *SIAM Journal on Applied Mathematics*. Vol 17, No 4, pp 638-663. July 1969.

Segunda parte

Aplicaciones

Optimización estocástica aplicada a plantas petroquímica y de producción de energía

Francisco Javier Quintana Martín

Departamento de Ingeniería Química Industrial y Medio Ambiente
Escuela Técnica Superior de Ingenieros Industriales
Universidad Politécnica de Madrid

1 Introducción

La vida de una planta petroquímica o de producción de energía comienza con su estudio de viabilidad, dimensionamiento y localización. Siguen el diseño, la construcción, después la puesta en marcha que es algo que por su problemática podemos separar del funcionamiento normal, pasamos a la fase de producción o funcionamiento normal, durante la que puede tener lugar alguna modificación de la misma con vistas a su posible ampliación de capacidad o mejora de rendimiento y finalmente su cierre por obsolescencia económica o tecnológica.

Nos centraremos en la fase de diseño de una nueva planta o en la modificación de una existente, si bien las ideas son aplicables también al funcionamiento.

Cuando abordamos el diseño de una planta debemos determinar que unidades vamos a construir, que forma y dimensiones tendrá cada unidad y de qué materiales estará construida, como estarán interconectadas, en qué lugares físicos hemos de situarlas, como será el sistema de control, y un sin número de características fáciles de imaginar. Los modelos que habitualmente aplicamos, sean más o menos sencillos o complejos, son deterministas. Solemos analizar: una hipótesis básica

que es la recomendada, una pesimista y otra optimista. Así nos cubrimos algo en salud pero pasamos "el muerto" de la decisión última a nuestro jefe.

Generalmente diseñamos la planta para unas "condiciones nominales" de funcionamiento. Fijamos factores tales como la capacidad de producción del producto o productos principales y secundarios, las materias primas que vamos a consumir. En el cálculo de las unidades de que consta suponemos conocidos parámetros tales como velocidades de reacción y/o coeficientes de transferencia de materia, funcionamiento de hornos, calderas, turbinas, perdidas de presión en tuberías, cargas que afectarán a la estructura y otros similares. Y "por si acaso" tomamos un coeficiente de seguridad con el que nos cubrimos en salud. Ésto era aceptable en una época en la que los algoritmos de cálculo y la capacidad de los ordenadores, cuando les había, no permitían otra cosa. Pero la realidad no es esa, y en la era de la globalización y de la competencia creciente no queda mas remedio que afinar todo lo que podamos. Con ello no solo salvaremos a nuestra empresa o entidad en la que trabajamos, sino que estaremos beneficiando a toda la colectividad.

Pero, ¿cómo podemos lograrlo?, ¿qué debemos hacer para mejorar nuestros diseños?

Está claro que acercarnos mas a la realidad. Lo difícil no es decir ésto sino saber llevarlo a la práctica. Trataremos de daros algunas ideas que os ayuden a hacerlo.

En primer lugar, por el método clásico damos valores fijos a los parámetros que intervienen en el diseño, pero la realidad es que solo podemos indicar la frecuencia con la que esperamos que tomen uno u otro valor. Disponemos de la distribución de probabilidades de que esto suceda. Si esta distribución de probabilidades tiene una desviación típica reducida respecto del valor que toma su media, el adoptar ésta como valor del parámetro en cuestión no origina una desviación apreciable en los resultados, pero si no es éste el caso, estamos ante la posibilidad de mejorar el diseño teniendo en cuenta en los cálculos la incertidumbre existente.

Al llegar a este punto pensamos en la incertidumbre existente en las previsiones de demanda y precios de nuestros productos principales y subproductos, que pueden alcanzar uno u otro valor, en la probabilidad de que la competencia decida llevar a efecto un proyecto complementario o competitivo con el nuestro, de la probabilidad de que dispongamos de ciertas materias primas y de los precios que puedan alcanzar, o de que en un futuro no lejano aparezca un producto o proceso ventajoso que origine la obsolescencia económica del nuestro. Todo ello afectará al diseño de nuestra planta.

Pero hay mas incertidumbres. ¿Porqué sobredimensionamos tanto las unidades de extracción líquido-líquido o los cristalizadores? ¿Y ciertos reactores?

La respuesta es sencilla: porque los modelos que representan su funcionamiento son poco exactos. Cuando determinamos los coeficientes de velocidad de reacción a partir de resultados experimentales, lo que realmente obtenemos -aunque no lo sepamos- es el valor mas probable y la desviación típica con la que

evaluamos dicho valor, que suele seguir una distribución normal. Si la relación entre el valor mas probable y la desviación típica es grande, el valor aludido es altamente probable que coincida con el real y podemos usar un modelo determinista en el cálculo de nuestro reactor. Si la relación aludida es pequeña la incertidumbre en la velocidad de reacción es alta y deberemos utilizar en nuestros cálculos un modelo que tenga en cuenta esta incertidumbre, esto es, un modelo estocástico. Pero incluso en el primer caso, tampoco está claro que el modelo determinista sea suficiente, pues seguramente exista incertidumbre en los caudales y composiciones de las corrientes que alimentan el reactor, por no hablar de costes, flujos internos y muchos otros factores.

Hasta con los valores de los parámetros exactos, podemos tener la incertidumbre en la bondad del propio modelo. El reactor antes aludido puede no comportarse como reactor de mezcla perfecta ni de flujo de pistón perfecto, sino como algo intermedio, acercándose mas o menos a uno u otro tipo de flujo, en función de una pluralidad de factores de imposible control durante el funcionamiento.

Los factores inciertos no solo son de los tipos hasta ahora expuestos. ¿Podemos estar seguros de que no va a salir un producto con mejores propiedades que el que nosotros vamos a fabricar y mas barato que el nuestro o de que no sale un proceso que desplaza al nuestro por fabricar lo mismo que nosotros con menores costes? Y ¿como protegernos contra esto al diseñar nuestra planta? ¿Quizá diseñando nuestra planta para que puede fabricar también otros productos diferentes?

Ni que decir tiene que nosotros estamos del lado de los modelos que contemplen la incertidumbre existente, esto es, de los modelos estocásticos por una razón muy sencilla: el mundo es así, la predestinación no existe. Pero para poder aplicar en este momento las técnicas de optimización estocástica a problemas reales medianos o grandes, tales como procesos completos o partes importantes de los mismos no queda mas remedio que transformar nuestro modelo en uno lineal entero mixto.

La solución que proporciona un modelo de programación matemática estocástica (que es el nombre de la optimización estocástica) debemos denominarla *solución inmunizada contra la incertidumbre* o *solución robusta* en vez de solución óptima.

Cuando transcurre el tiempo se suceden los acontecimientos. Si habíamos hecho un modelo determinista y la casualidad hace que las cosas ocurran de acuerdo con nuestros supuestos, estamos de enhorabuena. Ahora bien, lo mas probable es que las cosas no sucedan así. Lo normal es que no podamos celebrarlo.

En cambio, si habíamos utilizado un modelo en el que contemplábamos las probabilidades de que las cosas sucedan de una u otra forma, es fácil que lo que ocurra no nos pille desprevenidos. Quizá nuestra solución no sea la óptima pero si suficientemente buena. Es casi seguro que lo celebraremos.

La optimización determinista ha sido estudiada por muchos autores. La estocástica está de moda pero es mucho mas compleja y los éxitos son menores. En

relación con la síntesis de procesos citaremos [3], [9], [11], [15], [18], [27], [33], [14]. Exponen aspectos de la síntesis de procesos aplicando técnicas no lineales [5], [6], [7], [28], [29]. Un buen libro de síntesis de procesos es [1]. Otro que aborda la optimización de procesos desde la óptica no lineal, con revisión de algunos algoritmos es [12].

En [21], [22] mostramos nuestros primeros trabajos encaminados a la optimización determinista del diseño de procesos. Nuestros planteamientos estocásticos aparecen en [26], [24], [25], [23], y [10].

2 Las ideas

En la optimización del diseño, mejora y/o funcionamiento de un proceso encontramos elementos cuyo valor es incierto tanto en factores externos -precios, demandas de productos, disponibilidades de materias primas- como internos relacionados con su funcionamiento. Además, en ocasiones, mediante el proceso resultante deberemos ser capaces de fabricar mas de un producto.

Planteando nuestro modelo de una forma que pudiéramos llamar exacta, llegamos a un sistema que es: no lineal, tanto en la función objetivo (a veces mas de una) como en las restricciones; las mas de las veces fuertemente no convexo; con variables reales (que pueden tomar cualquier valor comprendido entre dos límites) y binarias (que toman solo los valores 0 ó 1); y algunos de los que normalmente serían parámetros han sido sustituidos por funciones de distribución de probabilidades. Este tipo de modelos solo funcionará en casos de muy reducidas dimensiones. Como línea de investigación básica es muy interesante, pero para resolver problemas reales no.

Para poder tratar casos medianos y grandes, nosotros consideraremos la incertidumbre vía escenarios y transformaremos el modelo no lineal en otro lineal entero mixto, en el que intervengan variables reales -continuas que pueden tomar cualquier valor comprendido entre dos límites- y binarias -que solo toman los valores 0 y 1-. El modelo resultante será generalmente de muy grandes dimensiones. El tratar con muchas variables binarias no será algo fácil pero si factible. Es un sistema que hemos aplicado con éxito a plantas petroquímicas reales y es el método que vamos a aplicar aquí.

3 Metodología

En nuestro sistema distinguiremos las:

- *Variables de diseño*, que determinan todo lo que interviene en el diseño del proceso. Son las que debemos determinar en el momento diseñarle. Tales son en un intercambiador de calor el diámetro, longitud, y disposición de los tubos, el número de pasos por la carcasa y la disposición de las pantallas

deflectoras, los materiales empleados, los rangos máximos de temperatura y presión de funcionamiento, las unidades a las que conectamos cada una de sus entradas y salidas de fluido y otros detalles constructivos. Existirá un único diseño, común a todos los escenarios.

- *Variables de funcionamiento*, que indican como funcionará nuestro proceso en cada escenario. Para cada escenario el proceso funcionará de una manera diferente, y lo único que conocemos es la probabilidad de que esto suceda. Debemos prever en nuestro modelo la posibilidad de que en algunos escenarios ciertas restricciones sean flexibles. Así, el modelo debe poder decidir para una planta de etileno una capacidad de producción de 500.000 t/año, en tanto que la demanda prevista para algunos escenarios sea de 700.000 t/año, teniendo en cuenta los posibles efectos que puede originar el que no suministremos 200.000 t/año. En el intercambiador antes aludido, variables de funcionamiento serían para cada escenario los fluidos y caudales que circulan por los tubos y por la carcasa y sus temperaturas de entrada y salida.

Nuestro modelo utilizará:

- *Variables reales* que pueden tomar cualquier valor comprendido entre dos límites que por defecto son 0 e ∞ . A este tipo pertenecen variables tales como caudales, temperaturas, presiones, concentraciones.
- *Variables binarias* que toman únicamente los valores 0 ó 1, cuyo significado es "si" o "no" instalamos, conectamos o colocamos algo, y en general las utilizaremos para imponer condiciones lógicas.

Nuestro modelo estocástico, tendría la forma compacta siguiente:

- Función objetivo a maximizar o minimizar, según el caso:

$$\max \widehat{\mathbf{c}}^X \cdot \mathbf{x}^D + \widehat{\mathbf{c}}^Y \cdot \mathbf{y}^D + \sum_{g \in \mathcal{G}} \widehat{w}_g \cdot \widehat{\mathbf{c}}_g^X \cdot \mathbf{x}_g^F + \sum_{g \in \mathcal{G}} \widehat{w}_g \cdot \widehat{\mathbf{c}}_g^Y \cdot \mathbf{y}_g^F \quad (7.1)$$

- sujeta a:

$$\underline{\mathbf{b}}_g \leq \widehat{\mathbf{A}}_g^D \cdot \mathbf{x}^D + \widehat{\mathbf{B}}_g^D \cdot \mathbf{y}^D + \widehat{\mathbf{A}}_g^F \cdot \mathbf{x}_g^F + \widehat{\mathbf{B}}_g^F \cdot \mathbf{y}_g^F \leq \bar{\mathbf{b}}_g \quad (7.2) \\ \forall g \in \mathcal{G}$$

- con los límites:

$$\underline{\mathbf{x}}^D \leq \mathbf{x}^D \leq \bar{\mathbf{x}}^D \quad (7.3)$$

$$\mathbf{y}^D \in \{0, 1\} \quad (7.4)$$

$$\underline{\mathbf{x}}_g^F \leq \mathbf{x}_g^F \leq \bar{\mathbf{x}}_g^F \quad (7.5)$$

$$\mathbf{y}_g^F \in \{0, 1\} \quad (7.6)$$

Recomendamos seguir los siguientes pasos:

1. *Definición del problema.*

No es sencillo definir el problema en el caso del diseño de un proceso o de una parte del mismo. Tenemos que decidir el conjunto de materias primas de posible utilización, el conjunto de productos de entre los que el sistema decidirá cuales produce, si deseamos que el proceso sea multiproducto en cuyo caso es como si tuviéramos "varios procesos en uno". El objetivo perseguido parece claro que es el beneficio máximo, pero generalmente no es el único. Aspectos tales como el impacto medioambiental, los riesgos en el abastecimiento de ciertas materias primas, la diversificación de riesgos y otros de política de empresa la complementan.

2. *Elaboración de un primer esbozo de superestructura.*

La superestructura es un diagrama de flujo que contemplará muchas si no todas las alternativas. Veamos un ejemplo sencillo que nos aclare las ideas.

Supongamos que disponemos de una corriente A que es una mezcla de los compuestos $C1$, $C2$ y $C3$ en concentración y caudal variable entre ciertos límites. Deseamos obtener tres corrientes cuyas concentraciones mínimas limitativas son: 98% de $C1$ la B , 97% de $C2$ la C y 99% de $C3$ la D . Supongamos también que utilizamos para ello una red de columnas de destilación.

Una columna de destilación, como bien sabemos, dispone de varios platos, un condensador en la parte superior en el que extraemos calor y un hervidor en la inferior mediante el que proporcionamos calor. El condensador y el hervidor son intercambiadores de calor. Para una columna de destilación dada, la temperatura de la corriente a calentar en el hervidor es mayor que la temperatura de la corriente a enfriar en el condensador. En un plato pasan los productos mas volátiles desde la fase líquida a la fase vapor y los menos volátiles desde la fase vapor a la líquida. La corriente de vapor asciende por la columna y se enriquece progresivamente en compuestos volátiles, empobreciéndose en compuestos menos volátiles. A la corriente líquida, que circula en sentido inverso, le sucede lo contrario. El vapor que sale por la parte superior de la columna se condensa en el condensador. Una parte es reciclado a la columna y otra es la corriente que sale por cabeza. Por el fondo sale una corriente líquida, parte de la cual se evapora en un hervidor y se devuelve a la columna, constituyendo el resto la corriente de fondo que sale de la misma. En ocasiones existen salidas laterales por las que salen corrientes de composiciones intermedias. El resultado es que por la parte superior salen los productos mas volátiles y por la inferior los menos volátiles. Si hay salidas intermedias, por ellas salen corrientes cuya volatilidad es menor a medida que la salida está mas próxima al fondo.

Un intercambiador de calor es un dispositivo formado por tubos por el que circulan dos corrientes, una de las cuales -llamada fría- se calienta y otra -denominada caliente y cuya temperatura es mayor que la de la anterior- se enfriá. El calor que pierde la corriente caliente es igual al calor que pasa a la corriente fría.

La figura 7.1 muestra la superestructura del ejemplo expuesto. Por simplicidad hemos dibujado solamente las columnas de destilación y sus interconexiones pero no las de la red de intercambiadores de calor (de la que si aparecen los condensadores y hervidores de las columnas) y otros elementos necesarios para un buen aprovechamiento energético. Columnas de destilación con su condensador y hervidor asociados estarán representadas en los lugares $U - 1, U - 2$ y $U - 3$. Tuberías de conexión son las líneas $L1, L2, \dots, L11$.

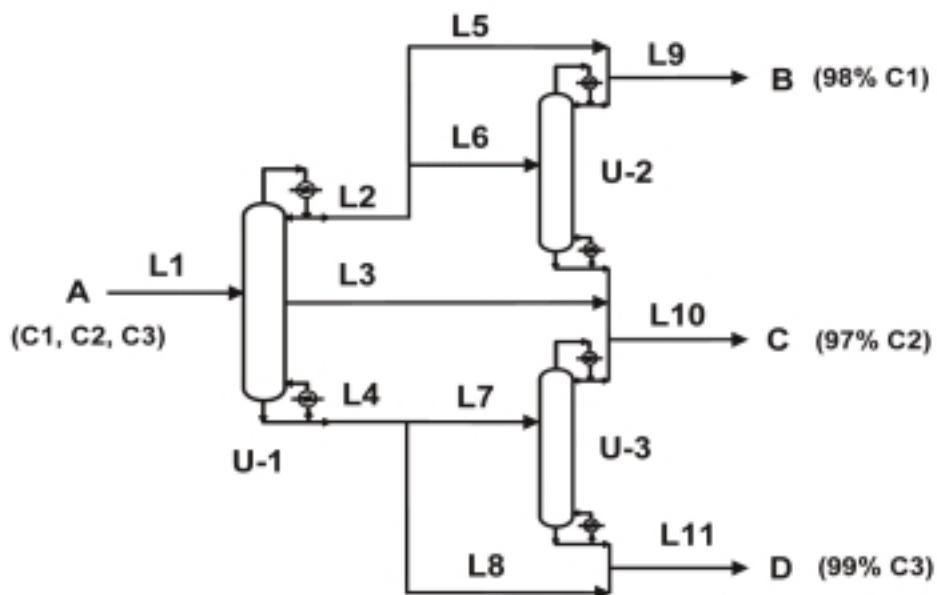


Figura 7.1: Ejemplo de superestructura

Nuestro modelo deberá decidir de que elementos constará nuestro proceso de entre los mostrados en la superestructura: columnas de destilación; intercambiadores de calor; fuentes de vapor y de agua de refrigeración; si

debemos o no emplear turbinas de vapor de contrapresión para mejorar el rendimiento energético. Deberá asimismo decidir que diseño e interconexiones deberá tener cada una de estas piezas.

3. *Elaboración de los escenarios y de sus probabilidades asociadas.*

La elaboración de escenarios conlleva hacer previamente una lista de parámetros inciertos. Para cada uno de ellos debemos determinar la distribución de frecuencias, o mas exactamente, de probabilidades de que tomen uno u otro valor.

En el caso de las demandas y precios de productos que posiblemente fabriquemos y de las materias primas que consumamos y sus restricciones los escenarios, podemos emplear para su predicción modelos estadísticos tales como los econométricos o los de Box-Jenkins. La distribución de probabilidades deseada nos la proporcionan directamente. Lo mismo sucede con parámetros de tipo técnico obtenidos experimentalmente, como coeficientes de velocidad de reacción. El uso de técnicas estadísticas como son sencillas regresiones.

Pero en otras ocasiones no queda mas remedio que recurrir a probabilidades de carácter subjetivo. Tel sucede por ejemplo en la inversión de un equipo con un diseño dado. Podemos solicitar valores de la inversión: mas probable, y los límites razonables superior e inferior (con probabilidades respectivas del 10% de ser o no superados), y ajustar una distribución Beta de probabilidades con estos datos.

Hemos de tener presente que hay parámetros interrelacionados, cuyos valores, si bien inciertos, no son independientes.

Supongamos que únicamente existe incertidumbre significativa en: la cantidad de producto q_1 , q_2 y q_3 a ser producida por nuestro proceso, cuyas respectivas probabilidades son w_{11} , w_{21} y w_{31} ; y en las probabilidades de que los precios de venta sean p_1 y p_2 , siendo con probabilidades w_{12} y w_{22} . Sean independientes ambos parámetros -cantidades y precios-. Resultarán los 6 escenarios mostrados en la figura 7.2. Así: el escenario 1, de probabilidad $w_1 = w_{11} \cdot w_{12}$, conllevará una producción q_1 y un precio de venta p_1 ; el escenario 2, de probabilidad $w_2 = w_{11} \cdot w_{22}$, conllevará una producción q_1 y un precio de venta p_2 ; y el escenario 6, de probabilidad $w_6 = w_{31} \cdot w_{22}$, conllevará una producción q_3 y un precio de venta p_2 .

El problema surge cuando del caso académico pasamos al real. La primera gran sorpresa es que el número de escenarios es infinito, o mas exactamente, varios millones, para no exagerar. La manera de resolver tal problema es generar escenarios equiprobables y extraer aleatoriamente un número razonable de ellos. Si somos un poco hábiles no necesitaremos generarles

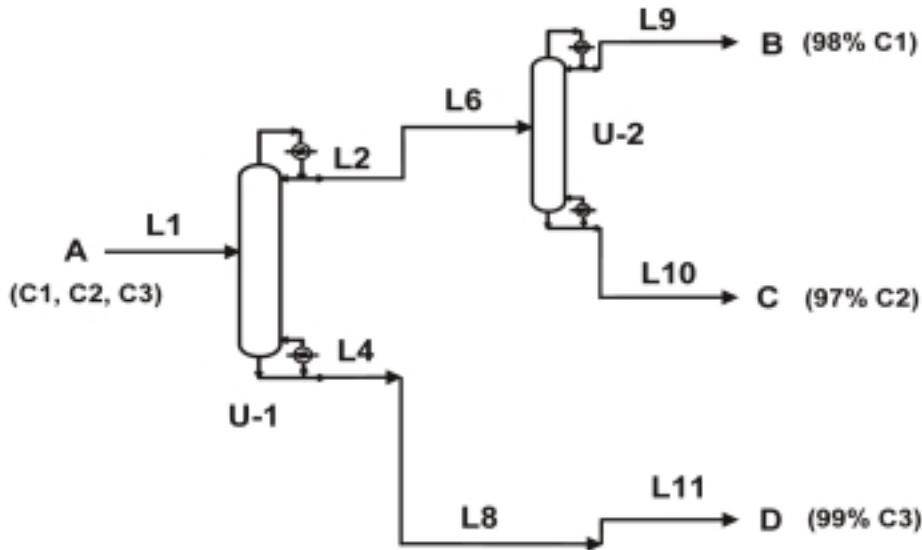


Figura 7.2: Ejemplo de escenarios

físicamente. Elaboraremos aleatoriamente solo los que necesitemos. Usaremos por tanto en nuestro modelo una muestra del colectivo total de escenarios. Esto implicará que nuestros resultados tendrán un cierto nivel de significación, pero por bajo que sea este, nuestra solución siempre será mejor que la obtenida por métodos deterministas.

4. *Elaboración de la superestructura.*

A la vista del resultado y consideraciones hechas en el paso 3 puede ser necesario retocar la superestructura elaborada en el paso 2. Llegamos a la superestructura que vamos a utilizar en nuestro modelo.

5. *Determinación de los límites de funcionamiento de cada unidad.*

En la superestructura de nuestro ejemplo (figura 7.1) encontramos que en los lugares $U - 1$, $U - 2$ y $U - 3$ habrá unidades importantes -columnas de destilación- cuya existencia y características deberá decidir el modelo. Habida cuenta de la superestructura y de los escenarios, determinaremos el conjunto de límites que englobarán las distintas maneras de funcionamiento

de cada unidad, en caso de que exista. Es fácil realizar modelos basados en flujos en redes con diferentes funciones objetivo, que nos den dichos límites. Así en nuestro ejemplo, la unidad $U-1$, cuando exista, podrá separar, según los casos por cabeza $C1$ o bien $C1$ y $C2$, por el centro $C2$ o nada, y por el fondo $C3$ o $C2$ y $C3$.

6. Determinación del catálogo de diseños de las unidades.

Esta es una de las fases críticas de nuestro método. Conocemos la superestructura y los límites de funcionamiento de cada una de las unidades de que consta.

Una forma de abordar el problema es incluir en nuestro modelo los "modelos" de diseño y funcionamiento de cada una de las unidades. El problema es que son modelos generalmente no lineales fuertemente no convexos. Ni siquiera un modelo determinista de una planta real medianamente compleja que los incluyera funcionaría. Si el lector ha usado modelos comerciales de simulación de, por ejemplo, columnas de destilación, recordará sin esfuerzo lo que le ha costado obtener resultados, ¡y lo único que hacen es resolver un sistema no lineal de ecuaciones!.

Pero cuando, por ejemplo, calculamos una tubería o una estructura metálica lo que hacemos es decidir que tubería o viga a usamos en cada lugar, seleccionándola de entre un catálogo de tuberías o vigas disponibles normalizadas. Ahora bien, ¿porqué las tuberías y vigas están normalizadas y las columnas de destilación o reactores químicos no? La respuesta es sencilla: porque las tuberías y vigas se consumen mucho y las columnas de destilación no.

Nuestra propuesta es aplicar el principio de *davit el vincit*. Elaboramos a nuestra medida un catálogo o conjunto de diseños de cada una de las unidades de nuestra superestructura. Para cada unidad y diseño determinamos un conjunto de modos de funcionamiento. Un diseño podrá funcionar de distintos modos. Cada modo será tal que pueda ser representado con suficiente precisión mediante un modelo lineal. Lo que hacemos con esto es separar el problema de diseño y funcionamiento de las unidades por un lado y la parte combinatoria estocástica de optimización, que es la realmente compleja, por otro.

Para determinar el diseño y funcionamiento de las unidades puedes usar los simuladores existentes y en caso de que fuera necesario resultados experimentales. Si usamos simuladores recomendamos hacer un sencillo programa que los maneje al objeto de elaborar de forma automática el conjunto de diseños de cada unidad.

El conjunto de diseños de cada unidad debe ser tal que cubran los límites de funcionamiento marcados en el paso 5.

7. Determinación de los modos de funcionamiento de cada diseño.

Para cada diseño de cada unidad deberemos determinar un conjunto de modos de funcionamiento. Cada modo tendrá unos límites tales que el funcionamiento pueda ser representado mediante un modelo lineal con la suficiente precisión. Es como si ajustáramos mediante facetas la función que representa el funcionamiento. Esto conlleva el uso de técnicas estadísticas, que en los casos sencillos se reduce a regresiones. Debemos de terminar simultáneamente las funciones lineales y los límites.

Dedicaremos un apartado a aclarar este punto mas adelante.

8. Elaboración del modelo determinista.

Elaboraremos un modelo determinista que contemple todos los diseños y modos de funcionamiento de las distintas unidades, su posible existencia o no, sus interconexiones y todos los demás aspectos necesarios. Tendrá una o mas funciones objetivo.

El modelo determinista estará constituido por la función objetivo (7.1), las restricciones (7.2), los límites de variables (7.5) y (7.6) cuando sustituimos el conjunto de escenarios \mathcal{G} por el escenario $g \in \mathcal{G}$ objeto de estudio, y además añadimos los límites de variables (7.3) y (7.4).

Pasaremos este modelo determinista para las condiciones específicas de cada escenario. Esto nos permitirá:

- Detectar los errores en el modelo. Es siempre mas fácil en este modelo que en el estocástico mas complejo.
- Analizar la solución proporcionada para cada escenario. Obtenemos una primera idea de la situación del problema y de la filosofía subyacente.
- Servir de paso previo al modelo estocástico. Las soluciones de cada escenario nos ayudan a elaborar una buena solución estocástica inicial o al menos una cota inicial que nos permita reducir el tiempo de cálculo.

9. Elaboración del modelo estocástico.

El modelo estocástico, como ya dijimos, está constituido por la función objetivo (7.1), las restricciones (7.2), y los límites de variables (7.3), (7.4), (7.5) y (7.6). Es generalmente un modelo de muy grandes dimensiones. Una manera de resolverlo en menor tiempo de cálculo es aplicar la descomposición de Benders, conservando en el problema maestro las variables binarias \mathbf{y}^D e \mathbf{y}_g^F , $\forall g \in \mathcal{G}$. El problema maestro le resolvemos en el ordenador mas rápido de que dispongamos y los problemas derivados en paralelo en todos los que podamos.

Sobre el modelo estocástico trataremos mas detenidamente en siguientes apartados.

10. *Análisis de los resultados del modelo estocástico.*

Llamamos solución robusta o solución inmunizada contra incertidumbre a la solución óptima del modelo estocástico. Es una solución que quizá para ningún caso o escenario sea la óptima, pero casi siempre suficientemente buena.

En nuestro ejemplo, la solución robusta podría ser la mostrada en la figura 7.3. Lógicamente, la solución robusta contiene mucha mas información que la incluida en esa figura. Parte de la solución es: el diseño de las unidades $U - 1$ y $U - 2$ (la $U - 3$ no se construye) que será el mismo sea cual fuere el o los escenarios que después ocurran, y para cada escenario temperaturas, presiones, flujos y composiciones de todas las corrientes que circulan dentro de las citadas unidades y por las tuberías que las interconectan, entre las que se encuentran las que abastecen de materias primas el proceso (L_1 en nuestro caso) y sacan de la misma los productos obtenidos (L_9 , L_{10} y L_{11}). Y a la hora de analizar la solución robusta no olvidemos tener en cuenta uno de los datos: la probabilidad de que suceda cada escenario.

4 Modelización de unidades: Diseños y modos de funcionamiento

Hemos comentado que debemos modelar las unidades de forma que el modelo resultado sea lineal. Veamos unas ideas o metodología de como hacerlo. Nos valdremos de un sencillo ejemplo de un intercambiador de calor.

El problema

Sea un intercambiador de calor cuyo diagrama y esquema de funcionamiento indicamos en la figura 7.4. Supongamos que es un intercambiador de calor en contracorriente de un paso por la tubería y otro por la carcasa. Las corrientes entrada y salida del fluido caliente son CE y CS respectivamente. Las corrientes de entrada y salida del fluido frío son FE y FS respectivamente. Denominemos $CC1$ al intercambiador.

Suponemos que cuando no exista el intercambiador conectamos la corriente fría de entrada con la de salida y la caliente de entrada con la de salida.

Por simplicidad suponemos que únicamente hay transferencia de calor sensible, si bien incluir el calor latente no ofrece ningún problema.

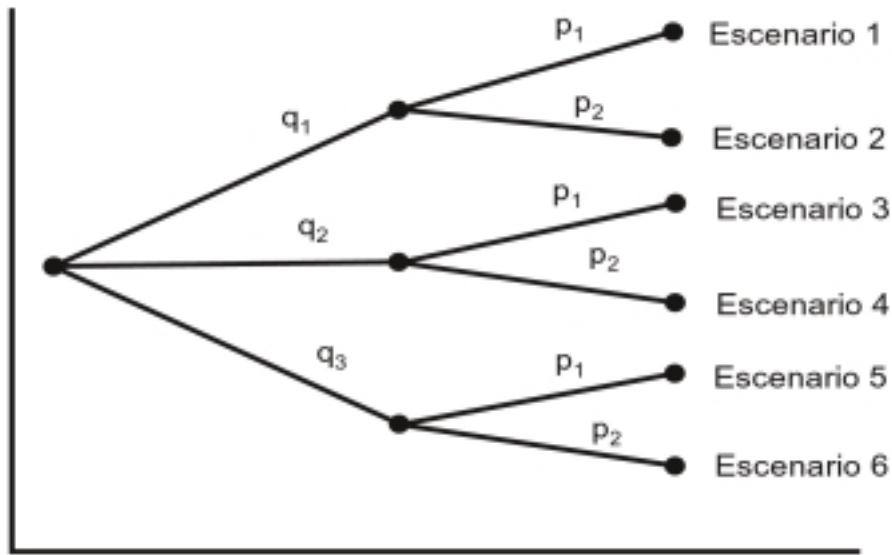


Figura 7.3: Ejemplo de solución robusta

La modelización exacta o casi exacta

Supondremos que: $\hat{c}_{CE} = \hat{c}_{CS}$ y $\hat{c}_{FE} = \hat{c}_{FS}$

Señalaremos que las *variables que definen el diseño* son: la existencia o no de cada unidad Y_u , el área A_u , y la inversión I_u que depende del área.

Las *variables que definen el funcionamiento* dependen del escenario que tenga lugar, y son: el calor intercambiado Q_{ug} , los caudales máicos de las corrientes F_{cg} , y sus temperaturas T_{cg} de las corrientes.

Empezamos por elaborar un modelo sencillo del intercambiador $CC1$ de nuestro ejemplo.

- En todos los casos:

$$F_{CE,g} = F_{CS,g} \quad (7.7)$$

$$F_{FE,g} = F_{FS,g} \quad (7.8)$$

- Si existe el intercambiador $CC1$ será $Y_{CC1} = 1$, y se cumplirá:

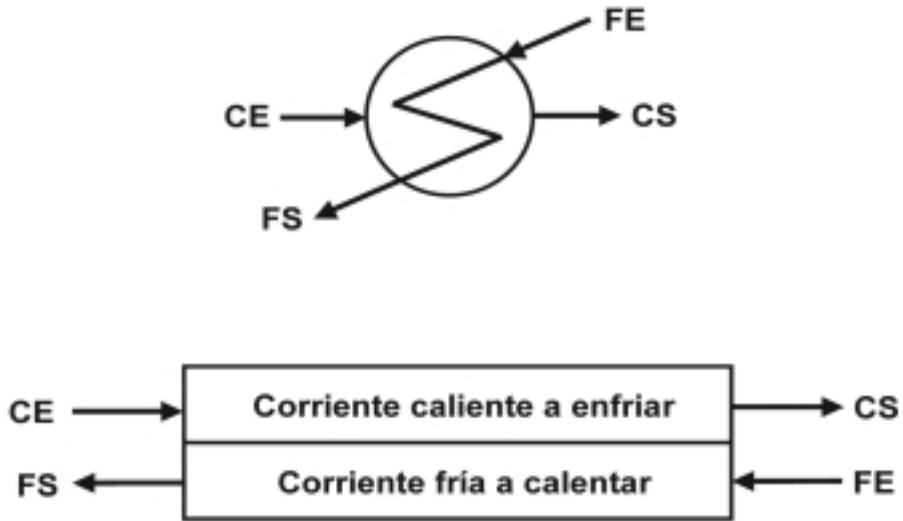


Figura 7.4: Ejemplo de intercambiador de calor: Diagrama y esquema

- Calor cedido por la corriente caliente:

$$Q_{CC1,g} = F_{CE,g} \cdot \hat{c}_{CE} (T_{CE,g} - T_{CS,g}) \quad (7.9)$$

$$g \in \mathcal{G}$$

- Calor tomado por la corriente fría:

$$Q_{CC1,g} = F_{FE,g} \cdot \hat{c}_{FE} (T_{FS,g} - T_{FE,g}) \quad (7.10)$$

$$g \in \mathcal{G}$$

- Calor transferido en el intercambiador:

$$Q_{CC1,g} = \hat{U}_{CC1} \cdot A_{CC1} \frac{T_{CE,g} - T_{FS,g} - T_{CS,g} + T_{FE,g}}{\ln \left(\frac{T_{CE,g} - T_{FS,g}}{T_{CS,g} - T_{FE,g}} \right)} \quad (7.11)$$

$$g \in \mathcal{G}$$

- Inversión:

$$I_{CC1} = \hat{\alpha}_1 (A_{CC1})^{\hat{\alpha}_2} \quad (7.12)$$

- Relaciones entre temperaturas:

$$\begin{aligned} T_{FS,g} &\geq T_{FE,g} \\ g \in \mathcal{G} \end{aligned} \tag{7.13}$$

$$\begin{aligned} T_{CE,g} &\geq T_{CS,g} \\ g \in \mathcal{G} \end{aligned} \tag{7.14}$$

$$\begin{aligned} T_{CE,g} &\geq T_{FS,g} \\ g \in \mathcal{G} \end{aligned} \tag{7.15}$$

$$\begin{aligned} T_{CS,g} &\geq T_{FE,g} \\ g \in \mathcal{G} \end{aligned} \tag{7.16}$$

- Si no existe el intercambiador CC1 será $Y_{CC1} = 0$. Se cumplirá entonces que:

- El área del intercambiador es nula:

$$A_{CC1} = 0 \tag{7.17}$$

- El calor intercambiado es nulo:

$$\begin{aligned} Q_{CC1,g} &= 0 \\ g \in \mathcal{G} \end{aligned} \tag{7.18}$$

- En las corrientes implicadas no hay variaciones de temperatura, ni de estado, ni de entalpia:

$$\begin{aligned} T_{FE,g} &= T_{FS,g} \\ g \in \mathcal{G} \end{aligned} \tag{7.19}$$

$$\begin{aligned} T_{CE,g} &= T_{CS,g} \\ g \in \mathcal{G} \end{aligned} \tag{7.20}$$

En nuestro modelo observamos no linealidades en las ecuaciones (7.9), (7.10), (7.11) y (7.12). Estas no linealidades nos van a complicar el proceso de optimización, ¡y el modelo exacto es mucho mas complejo que este!

En efecto. El modelo exacto debe proporcionar además: la longitud, diámetro, distancia y disposición de las tuberías; la distancia entre pantallas deflectoras; el

diseño de los cabezales y de las conexiones de entradas y salidas de los fluidos; los soportes; los aislantes térmicos; y los espesores y tipo de material de todos los elementos de que consta.

Incluso esto no es todo, pues durante el funcionamiento "se mancha o ensucia" con mayor o menor rapidez y a una velocidad que oscila entre ciertos límites. De éstas resistencias de ensuciamiento no es difícil encontrar datos en la bibliografía, aunque puede que en muchos casos no tengamos mas remedio que basarnos en resultados de ensayos en planta real o piloto, en cuyo caso el tipo de modelos "exacto o casi exacto" del tipo del expuesto en este apartado no es el mas adecuado.

La modelización suficientemente aproximada pero resoluble

Una manera de abordar el problema de las no linealidades es linealizar por facetas con la ayuda de variables binarias las ecuaciones en las que se encuentran.

Otra forma es el método que exponemos a continuación. Consiste en linealizar la unidad en su conjunto. Es mas ventajoso cuanto mas compleja es la unidad.

Continuemos con nuestro intercambiador de calor (unidad $u = CC1$). Intervienen las variables: $Y_u, A_u, I_u, Q_{ug}, \{F_{cg}, T_{cg}, \forall c \in \mathcal{C}_u\}$, y los parámetros: $\hat{c}_{CE}, \hat{c}_{FE}, \hat{U}_{ud}$.

Las restricciones que en nuestro nuevo modelo debemos conservar por cumplirse siempre y ser lineales son:

- las ecuaciones (7.7) y (7.8)
- las ecuaciones limitativas (7.13), (7.14), (7.15) y (7.16)
- las ecuaciones (7.17), (7.18), (7.19) y (7.20) aplicables cuando el intercambiador no existe.

En la figura 7.5 mostramos las variables que intervienen en una unidad de proceso, concretamente en un intercambiador de calor.

Las ecuaciones no lineales: (7.9), (7.10), (7.11) y (7.12) son las que debemos sustituir por facetas. Los siguientes pasos pueden servirnos de guía:

1. *Definir y discretizar las variables de diseño.*

Tratamos aquí de definir un conjunto de diseños que cubran lo comprendido entre los límites bajo los que sea posible que en algún momento tenga que funcionar la unidad en cuestión. Esto forma parte del paso 5 (página 157) de la metodología. Es recomendable que incluso haya solapamientos entre las zonas cubiertas por cada uno.

En nuestro caso elaboraremos un subconjunto \mathcal{D}_{CC1} de diseños tales que cada uno de ellos $d \in \mathcal{D}_{CC1}$ vendrá definido por su área \hat{A}_{ud} , su inversión \hat{I}_{ud} , e incluso elementos de detalle como longitud, disposición y materiales

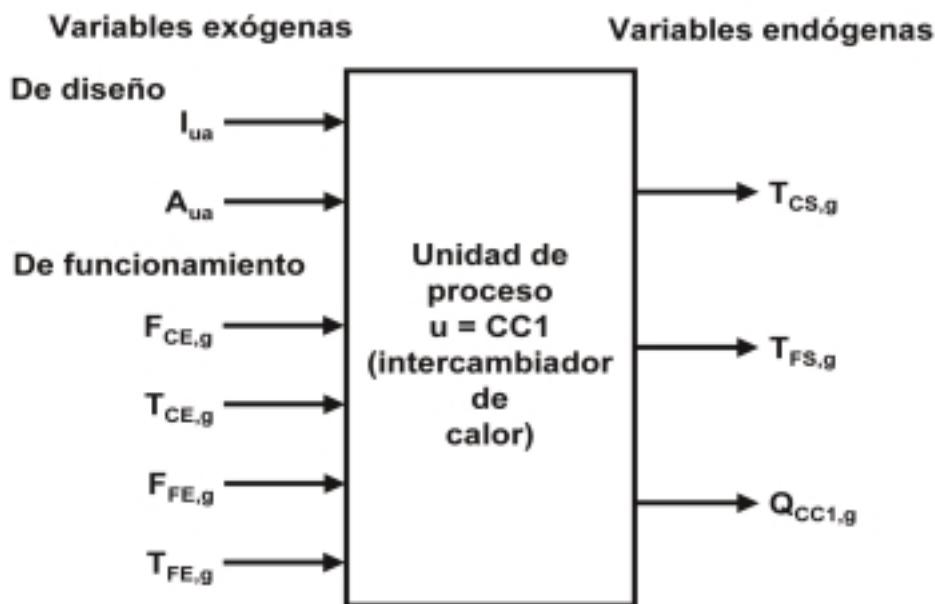


Figura 7.5: Ejemplo de intercambiador de calor: Clasificación de variables

de los tubos y deflectores, y todos los detalles constructivos que creamos necesarios. Podemos fijar las áreas en progresión geométrica e incluso realizar varios diseños diferentes para el mismo área.

Por tanto, en el momento de establecer un diseño fijamos todas las variables de diseño, como son en nuestro caso $\hat{A}_{ud}, \hat{I}_{ud}, \forall u, d \in \mathcal{U}^Q \times \mathcal{D}_u$.

Observemos que con éste truco hemos eliminado de un plumazo restricciones no lineales como la expresada por la ecuación (7.12).

2. *Elaborar la tabla de datos básicos.*

Denominamos *ensayo* a los resultados obtenidos mediante: un ensayo realizado en laboratorio o en planta piloto; las medidas efectuadas en una unidad de una planta industrial; un modelo de simulación mas o menos complejo; un análisis de tipo estadístico; o una mezcla de todo lo anterior.

En nuestro caso podemos realizar los cálculos mediante las ecuaciones antes mostradas o bien con otro modelo mas exacto y complejo, complementado si acaso con medidas en planta de los incrementos de las resistencias (de

Tabla 7.1: Resultados de ensayos con diseño d de unidad u

e	$F_{CE,e}$	$T_{CE,e}$	$T_{CS,e}$	$F_{FE,e}$	$T_{FE,e}$	$T_{FS,e}$	$Q_{CC1,e}$
$J7$	554	230	87	987	48	180	126.442
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$B5$	320	150	74	477	56	140	338.871

transmisión de calor) debidas al progresivo ensuciamiento a lo largo del tiempo.

Para cada diseño d de cada unidad u (tales que $u, d \in \mathcal{U}^Q \times \mathcal{D}_u$) elaboramos un subconjunto \mathcal{E}_{ud} de ensayos cuyos resultados dispondremos en forma de tabla que nos permita linealizar por facetas el funcionamiento de la unidad. Cada faceta será un *modo de funcionamiento*, representado en nuestro nuevo modelo por un conjunto de ecuaciones lineales. Cada fila de la tabla corresponderá a un ensayo $e|e \in \mathcal{E}_{ud}$. Las columnas de la tabla serán el ensayo e y las variables $F_{FE,e}, F_{CE,e}, T_{FE,e}, T_{CE,e}, T_{FS,e}, T_{CS,e}, Q_{CC1,e}$. También formarán parte de los resultados del ensayo otras variables y parámetros como $\widehat{U}_{ud}, \widehat{c}_{CE}, \widehat{c}_{FE}$, de gran interés, que no mencionamos aquí explícitamente por no intervenir directamente en nuestro modelo estocástico de diseño del proceso. En cada ensayo debemos fijar 4 de estas variables, pues tenemos un total de 7 y un número de ecuaciones de restricción de 3. Decidimos fijar como exógenas o independientes las variables $F_{FE,e}, F_{CE,e}, T_{FE,e}, T_{CE,e}$, y en función de ellas determinar las variables endógenas o dependientes: $T_{FS,e}, T_{CS,e}, Q_{CC1,e}$. Para cada $u, d|u, d \in \mathcal{U}^Q \times \mathcal{D}_u$ llegamos a una tabla como la 7.1 de nuestro ejemplo.

3. Análisis estadístico de los datos básicos y determinación de las facetas.

A partir del contenido de la tabla 7.1 determinaremos los coeficientes \widehat{a}_{udmij} de nuestro modelo lineal, que sustituirá al expresado mediante las (7.9), (7.10), (7.11) y (7.12). La forma mas sencilla de obtenerlos es:

- (a) Definir el subconjunto de *modos de funcionamiento* \mathcal{M}_{ud} de forma que no se solapen y cubran el espacio completo de funcionamiento del diseño d de la unidad m . Lo mas sencillo es establecer una malla dividiendo cada variable explicativa en intervalos que no tienen porqué estar uniformemente distribuidos.
- (b) Agrupar los ensayos en subconjuntos \mathcal{E}_{udm} , de acuerdo con la definición de modos.

(c) Determinar los coeficientes \hat{a}_{udmij} mediante:

$$\min \sum_{e \in \mathcal{E}_{udm}} \left[(\varepsilon_{umde1})^2 + (\varepsilon_{umde2})^2 + (\varepsilon_{umde3})^2 \right] \quad (7.21)$$

Sujeto a:

$$\begin{aligned} T_{FS,e} = & \hat{a}_{udm10} + \hat{a}_{udm11} \cdot F_{FE,e} + \hat{a}_{udm12} \cdot F_{CE,e} \\ & + \hat{a}_{udm13} \cdot T_{FE,e} + \hat{a}_{udm14} \cdot T_{CE,e} + \varepsilon_{umde1} \end{aligned} \quad (7.22)$$

$$\forall e \in \mathcal{E}_{udm}$$

$$\begin{aligned} T_{CS,g} = & \hat{a}_{udm20} + \hat{a}_{udm21} \cdot F_{FE,g} + \hat{a}_{udm22} \cdot F_{CE,g} \\ & + \hat{a}_{udm23} \cdot T_{FE,g} + \hat{a}_{udm24} \cdot T_{CE,g} + \varepsilon_{umde2} \end{aligned} \quad (7.23)$$

$$\forall g \in \mathcal{G}$$

$$\begin{aligned} Q_{CC1,g} = & \hat{a}_{udm30} + \hat{a}_{udm31} \cdot F_{FE,g} + \hat{a}_{udm32} \cdot F_{CE,g} \\ & + \hat{a}_{udm33} \cdot T_{FE,g} + \hat{a}_{udm34} \cdot T_{CE,g} + \varepsilon_{umde3} \end{aligned} \quad (7.24)$$

$$\forall g \in \mathcal{G}$$

(d) Si las desviaciones ε_{umde1} , ε_{umde2} y ε_{umde3} son aceptables y pensamos que no es posible reducir el número de modos de funcionamiento sin que ello las deteriore, terminamos el ajuste.

En caso contrario volvemos al punto 3a, modificando el subconjunto de modos de funcionamiento \mathcal{M}_{ud} .

En muchas ocasiones no te será tan fácil distinguir las variables significativas de las que no lo son ni las explicativas de las explicadas. Para ello tendrás que aplicar técnicas estadísticas que te permitan:

- reducir al máximo el número de variables que explique de manera significativa el funcionamiento del diseño de la unidad,
- diferenciar las exógenas de las endógenas, y
- estimar de que variables exógenas depende cada una de las endógenas,

En casos complejos, las técnicas de cluster, componentes principales, análisis discriminante nos pueden ayudar en el intento. En [19] hay una buena descripción de ellas.

4. Ecuaciones del modelo.

Nuestro proceso estará compuesto por varias unidades, una de las cuales es el intercambiador de calor que venimos poniendo como ejemplo. Mostramos a continuación la parte del modelo del proceso correspondiente a esta unidad.

- Función objetivo:

$$\min z = \dots + \sum_{d \in \mathcal{D}_{CC1}} \widehat{c}_{CC1,d}^F \cdot Y_{CC1,d} \quad (7.25)$$

Si en vez costes representara beneficios, cambiaríamos el signo de este término y maximizaríamos esta función.

- Ecuaciones de restricción:

Balances de materia:

$$F_{CE,g} - F_{CS,g} = 0 \quad (7.26) \\ \forall g \in \mathcal{G}$$

$$F_{FE,g} - F_{FS,g} = 0 \quad (7.27) \\ \forall g \in \mathcal{G}$$

Ecuaciones lineales de la unidad que sustituirán a las no lineales (7.9), (7.10), (7.11) y (7.12)::

$$T_{FS,g} - \sum_{d \in \mathcal{D}_u} \sum_{m \in \mathcal{M}_{ud}} [\widehat{a}_{udm10} + \widehat{a}_{udm11} \cdot F_{FE,g} \\ + \widehat{a}_{udm12} \cdot F_{CE,g} + \widehat{a}_{udm13} \cdot T_{FE,g} \\ + \widehat{a}_{udm14} \cdot T_{CE,g}] \cdot Y_{CC1,d,m,g} = 0 \quad (7.28) \\ \forall u, g \in \mathcal{U} \times \mathcal{G}$$

$$T_{CS,g} - \sum_{d \in \mathcal{D}_u} \sum_{m \in \mathcal{M}_{ud}} [\widehat{a}_{udm20} + \widehat{a}_{udm21} \cdot F_{FE,g} \\ + \widehat{a}_{udm22} \cdot F_{CE,g} + \widehat{a}_{udm23} \cdot T_{FE,g} \\ + \widehat{a}_{udm24} \cdot T_{CE,g}] \cdot Y_{CC1,d,m,g} = 0 \quad (7.29) \\ \forall u, g \in \mathcal{U} \times \mathcal{G}$$

$$Q_{CC1,g} - \sum_{d \in \mathcal{D}_u} \sum_{m \in \mathcal{M}_{ud}} [\widehat{a}_{udm30} + \widehat{a}_{udm31} \cdot F_{FE,g} \\ + \widehat{a}_{udm32} \cdot F_{CE,g} + \widehat{a}_{udm33} \cdot T_{FE,g} \\ + \widehat{a}_{udm34} \cdot T_{CE,g}] \cdot Y_{CC1,d,m,g} = 0 \quad (7.30) \\ \forall u, g \in \mathcal{U} \times \mathcal{G}$$

Recordemos que los términos cuadráticos formados por el producto de una variable real por una binaria que aparecen en estas ecuaciones se transforman sin problema en lineales añadiendo una variable real adicional.

- Transformamos las ecuaciones (7.13) y (7.19) en:

$$T_{FS,g} - T_{FE,g} - u_{FS,FE,g} = 0 \quad (7.31)$$

$$u_{FS,FE,g} - \bar{M} \cdot Y_{CC1} = 0 \quad (7.32)$$

$$\forall g \in \mathcal{G}$$

y las ecuaciones (7.14) y (7.20) en:

$$T_{CE,g} - T_{CS,g} - u_{CE,CS,g} = 0 \quad (7.33)$$

$$u_{CE,CS,g} - \bar{M} \cdot Y_{CC1} = 0 \quad (7.34)$$

$$\forall g \in \mathcal{G}$$

- Las ecuaciones (7.15) y (7.16) las escribimos como:

$$T_{FS,g} - T_{CE,g} + \bar{M} \cdot Y_{CC1} \leq \bar{M} \quad (7.35)$$

$$\forall g \in \mathcal{G}$$

$$T_{FE,g} - T_{CS,g} + \bar{M} \cdot Y_{CC1} \leq \bar{M} \quad (7.36)$$

$$\forall g \in \mathcal{G}$$

Señalaremos que en la modelización de una unidad hay ecuaciones que dependen de ciertas hipótesis. En nuestro ejemplo podíamos haber supuesto que si no existía el intercambiador las corrientes afectadas tendrían flujo nulo. Teníamos entonces que sustituir algunos de las restricciones impuestas.

5 Nomenclatura

Utilizamos la siguiente nomenclatura:

Conjuntos e (índices):

\mathcal{C}	Conjunto de corrientes, (c) Una corriente puede ser una tubería por la que circulan fluidos, un transportador de sólidos, un conducto que transporta energía eléctrica, un cable de fibra óptica por el que circulan señales. Empiezan en una unidad y terminan en otra.
\mathcal{C}_u	$\mathcal{C}_u \subseteq \mathcal{C}$, conjunto de corrientes que llegan o salen de la unidad u , para $u \in \mathcal{U}$ En nuestro ejemplo es: $\mathcal{C} = \{CE, CS, FE, FS\}$
\mathcal{D}	Conjunto de diseños de las unidades del proceso, (d) En tuberías, bombas y otros elementos los diseños están normalizados. Aquellos tipos de unidades que no lo estén, podemos "normalizarlos" nosotros de acuerdo con la conveniencia de nuestro caso.

\mathcal{D}_u	$\mathcal{D}_u \subseteq \mathcal{U}$, conjunto de diseños del intercambiador u , para $u \in \mathcal{U}$
\mathcal{E}	Conjunto de ensayos y/o resultados de cálculos realizados con los modelos de simulación, (e)
\mathcal{E}_u	$\mathcal{E}_u \subseteq \mathcal{E}$, subconjunto de ensayos y/o resultados de cálculos realizados con los modelos de simulación para la unidad u , para $u \in \mathcal{U}$
\mathcal{E}_{ud}	$\mathcal{E}_{ud} \subseteq \mathcal{E}_u$, subconjunto de ensayos y/o resultados de cálculos realizados con los modelos de simulación para el diseño d de la unidad u , para $u, d \in \mathcal{U}^Q \times \mathcal{D}_u$
\mathcal{E}_{udm}	$\mathcal{E}_{udm} \subseteq \mathcal{E}_{ud}$, subconjunto de ensayos y/o resultados de cálculos realizados con los modelos de simulación para el diseño d de la unidad u en el modo o forma de funcionamiento m , para $u, d, m \in \mathcal{U} \times \mathcal{D}_u \times \mathcal{M}_{ud}$
\mathcal{G}	Conjunto de escenarios contemplados, (g)
\mathcal{M}	Conjunto de modos o formas de funcionar las diversas unidades, (m)
\mathcal{M}_{ud}	$\mathcal{M}_{ud} \subseteq \mathcal{M}$, subconjunto de modos o formas de funcionar el diseño d de la unidad u , para $u, d \in \mathcal{U} \times \mathcal{D}_u$
\mathcal{U}	Conjunto de unidades del proceso, (u) Una unidad de proceso puede ser una operación básica, un reactor, una caldera, un tanque, un almacén, un nodo en el que confluyen varias tuberías, un controlador, e incluso el "exterior" de la planta, proceso o parte de los mismos que estamos analizando. En nuestro proceso, por el motivo que fuere, podemos desglosar en unidades independientes lo que habitualmente constituye una unidad. Así, nos puede interesar considerar como una unidad independiente cada plato de una columna de destilación, si el hecho de que existe además en él reacción química con o sin la ayuda de un catalizador es importante.
\mathcal{U}^Q	$\mathcal{U}^Q \subseteq \mathcal{U}$, subconjunto de unidades del proceso que son intercambiadores de calor En nuestro ejemplo es: $\mathcal{U}^Q = \{CC1\}$

Subíndices especiales:

\widehat{d}_e	$\widehat{d}_e \subseteq \mathcal{D}$, diseño de una unidad de proceso u utilizado en el ensayo e , para $e \in \mathcal{E}$ Para un $u \in \mathcal{U}$, debe ser: $\widehat{d}_e \subseteq \mathcal{D}_u$ y $e \in \mathcal{E}_u$
-----------------	--

Parámetros:

$\hat{\alpha}_1, \hat{\alpha}_2$	Coeficientes usados para el cálculo de la inversión de un intercambiador de calor en función de su área
\hat{a}_{udmij}	Coeficiente de la variable j en la ecuación lineal que determina la variable i en el modelo del diseño d de la unidad u cuando el punto de funcionamiento se halla en el modo de funcionamiento m , para $u, d, m \in \mathcal{U} \times \mathcal{D}_u \times \mathcal{M}_{ud}$ Los coeficientes \hat{a}_{udmij} aparecen en modelos como el representado por las ecuaciones 7.25, 7.26 y 7.27
$\hat{\mathbf{A}}_g^D, \hat{\mathbf{B}}_g^D$	Matrices de coeficientes de las restricciones del escenario g correspondientes a las variables de diseño reales y binarias, para $g \in \mathcal{G}$
\hat{A}_{ud}	Área del intercambiador u para el diseño d , para $u, d \in \mathcal{U}^Q \times \mathcal{D}_u$
$\bar{\mathbf{b}}_g, \underline{\mathbf{b}}_g$	Vectores de límites superior e inferior de las restricciones correspondientes al escenario g , para $g \in \mathcal{G}$
$\hat{\mathbf{c}}^X, \hat{\mathbf{c}}^Y$	Vectores de coeficientes de la función objetivo de las variables de diseño reales y binarias
\hat{c}_c	Calor específico de la corriente c , para $c \in \mathcal{C}$
$\hat{\mathbf{c}}_g^X, \hat{\mathbf{c}}_g^Y$	Vectores de coeficientes de la función objetivo de las variables de funcionamiento reales y binarias del escenario g , para $g \in \mathcal{G}$
\hat{c}_{ud}^F	Coste fijo anual del diseño d del intercambiador u , para $u, d \in \mathcal{U}^Q \times \mathcal{D}_u$
	Está compuesto por las cargas de capital mas el coste anual de mantenimiento, calculándose ambos como un porcentaje de la inversión
\hat{I}_{ud}	Inversión del diseño d del intercambiador u , para $u, d \in \mathcal{U}^Q \times \mathcal{D}_u$
\overline{M}	Valor positivo que excede a la mayor diferencia de temperatura entre los diversos pares de corrientes del intercambiador de que se trate
\hat{U}_u	Coeficiente global de transferencia de calor del intercambiador u , para $u \in \mathcal{U}^Q$
\hat{U}_{ud}	Coeficiente global de transferencia de calor del intercambiador u para el diseño d , para $u, d \in \mathcal{U}^Q \times \mathcal{D}_u$
\hat{w}_g	Probabilidad de que suceda el escenario g , para $g \in \mathcal{G}$ Observemos que: $\sum_{g \in \mathcal{G}} \hat{w}_g = 1$

$\bar{\mathbf{x}}^D, \underline{\mathbf{x}}^D$	Vector de límites superior e inferior de las variables de diseño reales
$\bar{\mathbf{x}}_g^F, \underline{\mathbf{x}}_g^F$	Vector de límites superior e inferior de las variables de funcionamiento reales, para $g \in \mathcal{G}$

Variables:

ε_{udmei}	Desviación en la ecuación de ajuste de la variable endógena i del ensayo e con el diseño d de la unidad de proceso u , para $u, d, m, e \in \mathcal{U}^Q \times \mathcal{D}_u \times \mathcal{M}_{ud} \times \mathcal{E}_{udm}$
A_u	Área del intercambiador u , para $u \in \mathcal{U}^Q$
I_u	Inversión del intercambiador u , para $u \in \mathcal{U}^Q$
F_{ce}	Caudal másico de la corriente c en el ensayo e , para $u, d, c, e \in \mathcal{U}^Q \times \mathcal{D}_u \times \mathcal{C} \times \mathcal{E}_{ud}$
F_{cg}	Caudal másico de la corriente c en el escenario g , para $c, g \in \mathcal{C} \times \mathcal{G}$
Q_{ude}	Calor transferido en el ensayo e del diseño d de la unidad (intercambiador) u , para $u, d, e \in \mathcal{U}^Q \times \mathcal{D}_u \times \mathcal{E}_{ud}$
Q_{ug}	Calor transferido en la unidad (intercambiador) u en el escenario g , para $u, g \in \mathcal{U}^Q \times \mathcal{G}$
T_{ce}	Temperatura de la corriente c en el ensayo e , para $u, d, e \in \mathcal{U}^Q \times \mathcal{D}_u \times \mathcal{E}_{ud}$
T_{cg}	Temperatura de la corriente c en el escenario g , para $c, g \in \mathcal{C} \times \mathcal{G}$
$u_{cc'g}$	$0 \leq u_{cc'g} \leq \infty$, variable auxiliar utilizada en una ecuación en la que intervienen las corrientes c y c' en el escenario g , para $c, c', g \in \mathcal{C} - \{c\} \times \mathcal{G}$
\mathbf{x}^D	Vector de variables de diseño reales (que toman valores continuos)
\mathbf{x}_g^F	Vector de variables de funcionamiento reales (que toman valores continuos) del escenario g , para $g \in \mathcal{G}$
\mathbf{y}^D	Vector de variables de diseño binarias (que toman valores 0 ó 1)
\mathbf{y}_g^F	Vector de variables de funcionamiento binarias (que toman valores 0 ó 1) del escenario g , para $g \in \mathcal{G}$
Y_u	$Y_u \in \{0, 1\}$, variable binaria que indica que la unidad u existe ($Y_u = 1$) o no ($Y_u = 0$) en la solución del modelo, para $u \in \mathcal{U}^Q$
Y_{ud}	$Y_{ud} \in \{0, 1\}$, variable binaria que indica que la unidad u (si, $Y_{ud} = 1$) o (no, $Y_{ud} = 0$) existe y tiene el diseño d en la solución del modelo, para $u, d \in \mathcal{U}^Q \times \mathcal{D}_u$

Y_{udmg}	$Y_{udmg} \in \{0, 1\}$, variable binaria que indica que la unidad u (si, $Y_{ud} = 1$) o (no, $Y_{ud} = 0$) existe, tiene el diseño d y funciona en el modo m en el escenario g en la solución robusta del modelo, para $u, d, m, g \in \mathcal{U}^Q \times \mathcal{D}_u \times \mathcal{M}_{ud} \times \mathcal{G}$
------------	--

6 Los métodos de optimización

La fuerte no linealidad que en ocasiones tienen las unidades del proceso hace que hoy día no sea viable la optimización estocástica no lineal para problemas que excedan las dimensiones académicas. En el mejor de los casos nunca tendremos la certeza de que el óptimo hallado sea global y no local. Sí es posible, sin embargo, transformar el modelo no lineal en uno lineal entero mixto, no sencillo de resolver por el gran número de variables binarias que aparecen. Nos pueden prestar gran ayuda en esta tarea aplicar, además de las técnicas que están habitualmente disponibles como la ramificación y limitación, la de planos secantes, y las heurísticas de generación de columnas, y la de relajar y fijar. Técnicas de descomposición como la de Benders y la lagrangeana aumentada nos permiten paralelizar los cálculos acortando tiempos de proceso.

En [13], [8], [4], [17], [30], [32], [31] encontramos amplia información sobre técnicas de programación lineal entera mixta.

En [16], [2], [20] hallamos técnicas de optimización estocástica.

7 Conclusiones

Tener en cuenta la incertidumbre existente a la hora de diseñar un proceso nos proporciona una solución con mas garantías de éxito durante su futuro funcionamiento que la proporcionada por los métodos deterministas de corte tradicional, en los que consideramos como máximo el escenario base y uno o dos pesimistas y optimistas.

Basar el modelo en técnicas de programación matemática estocástica lineal entera mixta es un camino que nos permite hoy día llegar a una buena solución en casos reales de grandes dimensiones.

8 Agradecimiento

El desarrollo de esta línea de investigación no hubiera sido posible sin los consejos de los profesores Escudero y M. Sanchez, que además han colaborado en algunos de estos trabajos, del profesor V. S. Vassiliadis, con quien ha estado alguno de los doctorandos, la ayuda del profesor Turet de la Fundación Repsol, la cual ha soportado económicamente algunas de estas investigaciones, la de empresas como Cepsa y Repsol que nos han permitido aplicar nuestros modelos a

plantas reales y han analizado y comentado nuestros resultados, del trabajo en estos temas en sus Proyectos de Fin de Carrera y Tesis Doctorales de Ingenieros como J. Colombás, I. Pereira, J. C. Gómez Gascó que fueron con nosotros los iniciadores de estos temas, y J. Torres, J. I. Marín, P. Penzol, S. Benito, D. Gómez, I. Domarco, J. C. del Río, P. Rincón y F. Pinilla que mantienen viva la llama hasta hoy.

9 Bibliografía

- [1] Biegler, T, Grossmann, IE, Westerberg, AW, (1996) Systematic Methods of Chemical Process Design, Prentice Hall International
- [2] Birge, JR, and Louveaux, F, (1997) Introduction to Stochastic Programming, Springer-Verlag, NY.
- [3] Bruno, J.C., F. Fernandez, F. Castells and I.E. Grossmann, (1998) MINLP Model for Optimal Synthesis and Operation of Utility Plants, Transaction of the Institution of Chemical Engineers , 76, pp.246-258
- [4] Ceria, S, Cordier, C, Marchand, H, and Wolsey, LA, (1998), Cutting Planes for Integer Programs with General Integer Variables, Mathematical Programming 81, 201-214.
- [5] Chen, T.W.C., and V.S. Vassiliadis, (2003), Solution of general nonlinear optimization problems using the penalty/modified barrier method with the use of exact Hessians, Comp. chem. Engng., 27 (4), 501-525.
- [6] Conejeros R. and V. S. Vassiliadis, (1998), Analysis and Optimization of Biochemical Process Reaction Pathways. Part I: Pathway Sensitivities and Identification of Limiting Steps, Industrial & Engineering Chemistry Research, 37 (12), 4699-4708.
- [7] Conejeros R. and V. S. Vassiliadis, (1998), Analysis and Optimization of Biochemical Process Reaction Pathways. Part II: Selection of Reaction Steps for Modification, Industrial & Engineering Chemistry Research, 37 (12), 4709-4714.
- [8] Conforti, M, Rinaldi, G, Wolsey, LA, (1999), On the cut polyhedron, R. 500, 4/, <http://www.iasi.rm.cnr.it/iasi/Publications/1999.html>
- [9] Duran, M.A. and I.E. Grossmann, (1986) Simultaneous Optimization and Heat Integration of Chemical Processes, AIChE J. 32, 123.
- [10] Escudero, LF, Quintana, FJ, Salmerón, J, (1999) CORO, a modeling and an algorithmic framework for oil supply, transformation and distribution

- optimization under uncertainty, European Journal of Operational Research, 114:638–656, Volume 114, Number 3, May 1.
- [11] Floudas, C.A., A.R. Ceric and I.E. Grossmann, (1986) Automatic Synthesis of Optimal Heat Exchanger Network Configurations, AIChE J. 32, 276.
 - [12] Floudas, C.A., (1996) Nonlinear and Mixed-Integer Optimization : Fundamentals and Applications, Oxford University Press.
 - [13] Felici, G, Gentile, C, Rinaldi, G, (2000), Solving Large MIP Models in Supply Chain Management by Branch & Cut, Rapporto di Ricerca IASI numero 522, Gennaio (ISSN 1128-3378).
 - [14] Grossmann, I.E. and C.A. Floudas, (1987), Active Constraint Strategy for Flexibility Analysis in Chemical Processes, Computers and Chemical Engineering 11, 675.
 - [15] Grossmann, I.E., J.A. Caballero and H. Yeomans, (1999), Advances in Mathematica Programming for Automated Design, Integration and Operation of Chemical Processes, Proceedings of the International Conference on Process Integration (PI'99), Copenhagen, Denmark.
 - [16] Higle, JL, and Sen S, (1996) Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming, Kluver Academic Publishers, London.
 - [17] Nemhauser, GL, and Wolsey, LA, (1988), Integer and Combinatorial Optimization, New York, Wiley.
 - [18] Papoulias, S.A. and I.E. Grossmann, (1983), A Structural Optimization Approach in Process Synthesis. Part I: Utility Systems,” Part II: Heat Recovery Networks, Part III: Total Processing Systems, Computers and Chemical Engineering 7, 695
 - [19] Peña, D, (2002), Análisis de Datos multivariantes, Mc Graw Hill, Madrid.
 - [20] Prékopa, A, (1995) Stochastic Programming, Kluver Academic Publisers.
 - [21] Quintana, FJ, Escudero, LF, Colombas, J, (1985) Diseño de Plantas Químicas Asistido por Ordenador: Optimización de Redes de Intercambio de Calor mediante Programación Matemática, XV Reunión Nacional de Estadística e Investigación Operativa. Gijón, Septiembre.
 - [22] Quintana, FJ, Escudero, LF, Pereira, I, Gómez, JC, (1985) Diseño de Plantas Químicas Asistido por Ordenador: Optimización de Sistemas de Separación de Componentes con Columnas de Destilación mediante Programación Matemática, XV Reunión Nacional de Estadística e Investigación Operativa. Gijón, Septiembre.

- [23] Quintana, FJ, Escudero, LF, Marín, JI, (1998) Optimización Estocástica del Diseño de Plantas Químicas, XXIV Congreso Nacional de Estadística e Investigación Operativa, Sociedad de Estadística e Investigación Operativa, 20 a 23 octubre.
- [24] Quintana, FJ, Gómez, D, Marín, JI, (2000) Optimización Estocástica del Diseño de una Planta de Termohidrodesalquilación Mediante la Técnica de Escenarios, XXV Congreso Nacional de Estadística e Investigación Operativa, Sociedad de Estadística e Investigación Operativa, 4 a 7 abril.
- [25] Quintana, FJ, Benito, S, Marín, JI, Hernández, A, (2000) Diseño de una Planta de Separación de Naftas de FCC mediante Optimización Estocástica, XXV Congreso Nacional de Estadística e Investigación Operativa, Sociedad de Estadística e Investigación Operativa, 4 a 7 abril.
- [26] Quintana, FJ, Escudero, L. F., Marín, JI, (2000) Una metodología para la optimización del diseño de procesos químicos, Congreso sobre Técnicas de Ayuda a la Decisión en la Defensa (II RIMO y I REM: II Reunión de Investigación Militar Operativa y I Reunión de Estadística Militar), Ministerio de Defensa, 12 a 15 de diciembre.
- [27] Turkay, M. and I.E. Grossmann, (1998), Structural Flowsheet Optimization with Complex Investment Cost Functions, Computers and Chemical Engineering 22, 673-686.
- [28] Vassiliadis, V. S., R. W. H. Sargent, and C. C. Pantelides, (1994), Solution of a Class of Multistage Dynamic Optimization Problems. Part I - Problems Without Path Constraints, Industrial & Engineering Chemistry Research, 33 (9), 2111-2122.
- [29] Vassiliadis, V. S., R. W. H. Sargent, and C. C. Pantelides, (1994), Solution of a Class of Multistage Dynamic Optimization Problems. Part II - Problems With Path Constraints, Industrial & Engineering Chemistry Research, 33 (9), 2123-2133.
- [30] Wolsey, LA, (1998) Integer Programming, New York, Wiley.
- [31] Wolsey, LA, (2002) Solving multi-item lot-sizing problems with an MIP solver using classification and reformulation, Management Science 48, 1587-1602.
- [32] Wolsey, LA, (2003) Strong formulations for mixed integer programs: valid inequalities and extended formulations, Mathematical Programming B 97, 423-447.

- [33] Yeomans, H. and I.E. Grossmann, (1999), A Systematic Modeling Framework of Superstructure Optimization in Process Synthesis, *Computers and Chemical Engineering* , 23, 709-731.

Long-term electric power planning in liberalized markets using the Bloom and Gallant formulation

Narcís Nabona, Adela Pagès

Dept. Statistics and Operations Research
Univ. Politècnica de Catalunya, 08028 Barcelona, Spain
e-mails: narcis.nabona@upc.es, adela.pages@upc.es

1 Introduction and Motivation

Long-term generation planning is a key issue in the operation of an electricity generation company. Its results are used both for budgeting and planning fuel acquisitions and to provide a framework for short-term generation planning.

The long-term problem is a well-known stochastic optimisation problem because several of its parameters are only known as probability distributions (for example: load, availability of thermal units, hydrogeneration and generations from renewable sources in general).

A long-term planning *period* (e.g., a natural year) is normally subdivided into shorter *intervals* (e.g., a week or a month), for which parameters (e.g., the load-duration curve) are known or predicted, and optimized variables (e.g., the expected energy productions of each generating unit) must be found.

Predicted load-duration curves (LDC's) — equivalent to cumulative probability load distributions — for each interval are used as data for the problem, which is appropriate since load uncertainty can be suitably described through the LDC. The probability of failure for each thermal unit is assumed to be known.

Bloom and Gallant [3] proposed a linear model (with an exponential number of

inequality constraints) and used an *active set* methodology [18] to find the optimal way of matching the LDC of a single interval with thermal units only, when there are load-matching and other operational non-load-matching constraints. These could be, for example, limits on the availability of certain fuels, or environmental maximum emission limits. The optimal *loading order* obtained with Bloom and Gallant's method may include permutations with respect to the *merit order* and *splittings* in the loading of units [3, 14]. In this way the energies generated satisfy the limitations imposed by the non-load-matching constraints while having the best possible placement, with respect to generation cost, in the matching of the LDC.

When the long-term planning power problem is to be solved for a generation company operating in a competitive market, the company has not a load of its own to satisfy, but it bids the energies of its units to a *market operator*, who selects the lowest-price among bidding companies to match the load. In this case, the scope of the problem is no longer that of the generation units of a single generation company but that of all units of all companies bidding in the same competitive market, matching the load of the whole system. This makes planning problems much larger than before and is a reason for developing more efficient codes to solve them.

The Bloom and Gallant model has been successfully extended to multi-interval long-term planning problems using either the active-set method [18], the Dantzig-Wolfe column-generation method [8, 23], or the Ford-Fulkerson column-generation (FFcg) method [9, 21]. The FFcg and the Dantzig-Wolfe procedures have many common steps. The model has also been coded using the modeling language *AMPL* [10] and has been solved with a linear/quadratic programming package *Cplex 7.5* [5] as carried out in [20] for a single interval.

A quadratic model is put forward here to formulate the long term profit maximization of generation companies in a liberalized electricity market [19] and the performance of several solution procedures for solving this problem is compared [22].

2 The load-duration curve

The LDC is the most sensible way to represent the load of a future interval. The main features of an LDC (corresponding to the i^{th} interval) can be described through 5 characteristics: the duration T^i , the peak load power \hat{P}^i , the base load power \underline{P}^i , the total energy \hat{E}^i and the shape, which is not a single parameter and is usually described through a table of durations and powers, or through a function.

The LDC for future intervals must be predicted. For a past interval, for which the hourly load record is available, the LDC is equivalent to the load over time

curve sorted in order of decreasing power. It should be noted that in a *predicted* LDC, random events such as weather, shifts in consumption timing, etc., that cause modifications of different signs in the load tend to cancel out, and that the LDC keeps all the power variability of the load.

3 Thermal Units

As far as loading an LDC is concerned, the relevant parameters of a thermal unit are:

- * *power capacity*: (C_j for the j^{th} unit) maximum power output (MW) that the unit can generate
- * *outage probability*: (q_j for the j^{th} unit) probability of a unit not being available when it is required to generate
- * *linear generation cost*: (\tilde{f}_j for the j^{th} unit) production cost in €/MWh

Other associated concepts are:

- * *merit order*: units are ordered according to their efficiency in generating electric power (€/MWh); all units will work at their maximum capacity since no unit should start to generate until the previous unit in the merit order is generating at its maximum capacity,
- * *loading order*: units will have load allocated to them in a given order; loading order and merit order may not coincide when there are other constraints to be satisfied.

4 Matching the load-duration curve

Due to the outages of thermal units (whose probability is >0), the LDC does not coincide with the estimated production of thermal units. It is usual for the installed capacity to be higher than the peak load: $\sum_{j=1}^{n_u} C_j > \hat{P}$.

The *generation-duration curve* is the expected production of the thermal units over the time interval to which the LDC refers. The energy generated by each unit is the slice of area under the generation-duration curve which corresponds to the capacity of the thermal unit.

The probability that there are time lapses within the time interval under consideration, where, due to outages, there is not enough generation capacity to cover the current load, is not null. Therefore, *external energy* (from other interconnected utilities) will have to be imported and paid for at a higher price

than the most expensive unit in ownership. The peak power of the generation-duration curve is $\sum_{j=1}^{n_u} C_j + \hat{P}$ and the area above power $\sum_{j=1}^{n_u} C_j$ is the external energy.

4.1 Convolution method of finding the generation-duration curve

The loading of thermal units in an LDC was first formulated in [1] and practical procedures to compute the expected generation can be found in [26]. Analytically, given the probability density function of load $p(x)$, the cumulative load distribution function $L_0(x)$ (see Fig. ??) is calculated as follows:

$$L_0(x) = 1 - \int_0^x p(y) dy$$

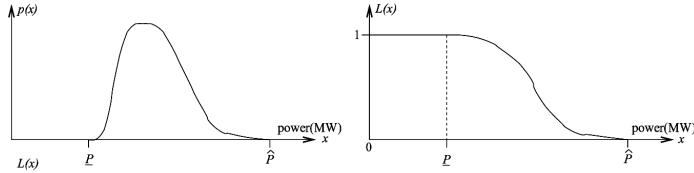


Fig. 1 Probability density function of load $p(x)$ (left), and cumulative load distribution function $L_0(x)$ (right).

The method calculates the production of each thermal unit, given a loading order. The load is modeled through its distribution $L_0(x)$, which is the probability of requiring x MW, or more. Let:

- C_j : maximum power capacity in MW of unit j
- q_j : outage probability of unit j
- $1 - q_j$: in-service probability of unit j
- U_j : set of unit indices $1, 2, \dots, j$
- $L_{U_{j-1}}(x)$: probability distribution of uncovered load after loading units $1, 2, \dots, j-1$
- $L_{U_j}(x)$: probability distribution of load still to be matched after loading units $1, 2, \dots, j-1, j$
- x : load in MW

the convolution computes $L_{U_j}(x)$ from $L_{U_{j-1}}(x)$ as [1, 26]:

$$L_{U_j}(x) = q_j L_{U_{j-1}}(x) + (1 - q_j) L_{U_{j-1}}(x + C_j) \quad (8.1)$$

Recalling that $E=P \cdot T$, the expected energy generated by unit j is [1]:

$$E_j = (1 - q_j) T \int_0^{C_j} L_{U_{j-1}}(x) dx . \quad (8.2)$$

4.2 Unsupplied load after a set of thermal units is loaded

Let $L_0(x)$ be the cumulative probability distribution of the power load corresponding to the LDC. It is not difficult to derive that, given a set of units whose indices 1,2, etc. are the elements of the set of indices Ω , the unsupplied load after loading all the units in Ω will have a cumulative probability distribution $L_\Omega(x)$

$$L_\Omega(x) = L_0(x) \prod_{m \in \Omega} q_m + \sum_{U \subseteq \Omega} \left(L_0(x + \sum_{i \in U} C_i) \prod_{i \in U} (1 - q_i) \prod_{i \in \Omega \setminus U} q_i \right) \quad (8.3)$$

We can thus say that the cumulative probability distribution $L_\Omega(x)$ of the unsupplied load is the same no matter the order in which the units in Ω have been loaded.

The unsupplied energy $W(\Omega)$ is computed as:

$$W(\Omega) = T \int_0^{\hat{P}} L_\Omega(x) dx \quad (8.4)$$

The integration in (8.4) is to be carried out numerically.

5 Bloom & Gallant's model for matching the load-duration curve when there are non-load-matching constraints

Let the Bloom & Gallant formulation (for a single interval) [3] be given by:

$$\min_{E_j} \quad \sum_{j=1}^{n_u+1} \tilde{f}_j E_j \quad (8.5)$$

$$\text{subject to} \quad \sum_{j \in U} E_j \leq \hat{E} - W(U) \quad \forall U \subset \Omega = \{1, \dots, n_u\} \quad (8.6)$$

$$A_{\geq} E \geq R_{\geq} \quad (8.7)$$

$$A_{=} E = R_{=} \quad (8.8)$$

$$\sum_{j=1}^{n_u+1} E_j = \hat{E} \quad (8.9)$$

$$E_j \geq 0 \quad j = 1, \dots, n_u, n_u + 1 \quad (8.10)$$

where:

n_u+1	: index representing the external energy
n_{\geq}	: total number of non-load-matching inequality constraints
A_{\geq}	: $\in \mathbb{R}^{n_{\geq} \times n_u}$ matrix of non-load-matching inequality constraints
R_{\geq}	: <i>rhs</i> of non-load-matching inequality constraints
$A_=$: $\in \mathbb{R}^{n_= \times n_u}$ matrix of non-load-matching equality constraints
$R_=$: <i>rhs</i> of non-load-matching equality constraints
U	: subset of Ω
$W(U)$: unsupplied energy after loading all units $j \in U \subset \Omega$

The objective function (8.5) can be simplified using (8.9), which leads to:

$$\sum_{j=1}^{n_u} f_j E_j + \tilde{f}_{n_u+1} \hat{E} \quad \text{where } f_j = \tilde{f}_j - \tilde{f}_{n_u+1}$$

with $\tilde{f}_{n_u+1} \hat{E}$ being a constant.

5.1 The case where no constraint (8.7) is active

Constraints (8.7) and (8.8) are the non-load-matching constraints. The Appendix of [14] contains a proof that the merit-order loading energies correspond to a minimum of the formulation (8.5–8.10) when there are no active constraints (8.7) and in case that there should be no non-load-matching equality constraints (8.8).

Assuming that units are ordered in order of merit, the active constraints at the minimizer of the set of inequalities (8.6) would be:

$$\begin{aligned} E_1 &= \hat{E} - W(1) \\ E_1 + E_2 &= \hat{E} - W(1, 2) \\ E_1 + E_2 + E_3 &= \hat{E} - W(1, 2, 3) \\ &\dots \\ E_1 + E_2 + E_3 + \dots + E_{n_u} &= \hat{E} - W(1, 2, \dots, n_u) \end{aligned} \tag{8.11}$$

5.2 The case with equalities (8.8) or where a constraint (8.7) or nonnegativity bound (8.10) is active

In this case, the equalities (8.8) or at least one of the constraints in (8.7) or nonnegativity bound (8.10) will be active, which means that at least one of the active constraints in (8.11) will not be satisfied as an equality.

5.3 The multi-interval Bloom and Gallant model

As power planning for a long time period cannot take into account changes over time of some parameters, the time period is subdivided into shorter intervals

in which all parameters can be assumed to be constant. We will use superscript i to indicate that variables and parameters refer to the i^{th} interval.

Therefore some constraints refer only to variables of a single interval, while others may refer to variables in several intervals. E.g., constraints on the minimum consumption of gas may affect several or all the intervals, while emission limit constraints, or the constraint associated with the units composing a combined-cycle unit refer to each single interval.

The overhauling of thermal units must be taken into account. Therefore, there will be intervals where some units must remain idle. The set of available units in each interval may be different. Let Ω^i be the set of available units in the i^{th} interval, and let n_u^i be $n_u^i = |\Omega^i|$ (the cardinality of this set).

The Bloom and Gallant linear optimization model extended to n_i intervals, with inequality and equality non-load-matching constraints, can thus be expressed as:

$$\min_{E_j^i} \quad \sum_{i=1}^{n_i} \sum_{j=1}^{n_u} f_j E_j^i \quad (8.12)$$

$$\text{subject to: } \sum_{j \in U} E_j^i \leq \hat{E}^i - W^i(U) \quad \forall U \subset \Omega^i \quad i = 1, \dots, n_i \quad (8.13)$$

$$A_{\geq}^i E^i \geq R_{\geq}^i \quad i = 1, \dots, n_i \quad (8.14)$$

$$\sum_i A_{\geq}^{0i} E^i \geq R_{\geq}^0 \quad (8.15)$$

$$A_{=}^i E^i = R_{=}^i \quad i = 1, \dots, n_i \quad (8.16)$$

$$\sum_i A_{=}^{0i} E^i = R_{=}^0 \quad (8.17)$$

$$E_j^i \geq 0 \quad j = 1, \dots, n_u, \quad i = 1, \dots, n_i \quad (8.18)$$

where:

A_{\geq}^i	: $\in \mathbb{R}^{n_{\geq}^i \times n_u}$ matrix of inequalities that refer only to interval i
A_{\geq}^{0i}	: $\in \mathbb{R}^{n_{\geq}^0 \times n_u}$ matrix of inequalities that refer to more than one interval
R_{\geq}^i	: $\in \mathbb{R}^{n_{\geq}^i}$ rhs of inequalities that refer only to energies of interval i
R_{\geq}^0	: $\in \mathbb{R}^{n_{\geq}^0}$ rhs of inequalities that refer to more than one interval
$A_{=}^i$: $\in \mathbb{R}^{n_{=}^i \times n_u}$ matrix of equalities that refer only to energies of interval i
$A_{=}^{0i}$: $\in \mathbb{R}^{n_{=}^0 \times n_u}$ matrix of equalities that refer to more than one interval
$R_{=}^i$: $\in \mathbb{R}^{n_{=}^i}$ rhs of equalities that refer only to energies of interval i
$R_{=}^0$: $\in \mathbb{R}^{n_{=}^0}$ rhs of equalities that refer to energies of more than one interval

The number of variables is now $\sum_i^{n_i} n_u^i$ and there are $\sum_i^{n_i} (2^{n_u^i} - 1)$ load-matching constraints plus $n_- = n_-^0 + \sum_i n_-^i$ non-load-matching equalities, and $n_\geq = n_\geq^0 + \sum_i n_\geq^i$ non-load-matching inequalities. Note that supraindices 0 indicate constraints which affect variables of more than one interval.

Should constraint sets (8.15) and (8.17), which are the multi-interval constraints, be empty, the problem would be separable into n_i subproblems, one for each interval. Otherwise a joint solution must be found.

5.4 Approximate model of long-term hydrogeneration

The long term model described is appropriate for thermal generation units but not for hydrogeneration, which requires additional variables to represent the variability of water storage in reservoirs and discharges necessary for the calculation of the hydroenergy generated.

A coarse model of hydrogeneration, which does not consider any of the reservoir dynamics, can be employed. All or a part of the reservoir systems of one or several basins are considered as a single pseudo-thermal unit H with cost $\tilde{f}_H = 0$, outage probability $q_H = 0$ and capacity C_H (normally lower than the maximum installed hydropower capacity), with a constraint binding the intervals' hydrogenerations over the successive intervals so that they add up to a total expected hydrogeneration R_H^0 for the whole period:

$$\sum_i^{n_i} E_H^i = R_H^0, \quad (8.19)$$

6 Long-term maximization of profit in a “competitive” market

In the classical electricity markets, utility companies have both generation and distribution of power. These companies have their own load to supply, corresponding to their clients plus other contracts, and try to minimize their generation cost. In “competitive” electricity markets, generation companies have no distribution, and therefore no load of their own. Generation companies must bid their generation to the market operator and a market price is determined for each hour by matching the demand with the generation of the lowest bids. Generation companies are no longer interested in generating at the lowest cost but in obtaining the maximum profit, which is the difference between market price and generation cost for all accepted generation bids. In long-term operation all accepted bids in a time interval (a week, or a month) must match the LDC of this interval.

There is no specific load to be matched by a specific generation company (SGC). The only known loads are the predicted LDC's for the whole market in

each interval. As all generation companies pursue their maximum profit, it is natural to attempt to maximize the profit of all generation companies combined.

The SGC must thus solve the problem of the maximization of profit of all generation companies, taking into account the total market load. The SGC should introduce its own operation constraints (fuel and emission limits, contracts, etc.) and may also introduce a market-share constraint for its units in one or several intervals. (The Lagrange multiplier value of this constraint will tell whether the market share imposed, though feasible, is reasonable or not.) The long-term results will indicate how the SGC should program its units so that its profit be maximized while meeting all its operation constraints.

6.1 Long-term market price function of a given interval

From the records of past market-price and load series (see Fig. ??) it is possible to compute a market-price function for a given interval. This function is to be used with expected generations that match the LDC of the interval, so market prices should correspond in duration with the duration of loads, from peak to base load in the interval.

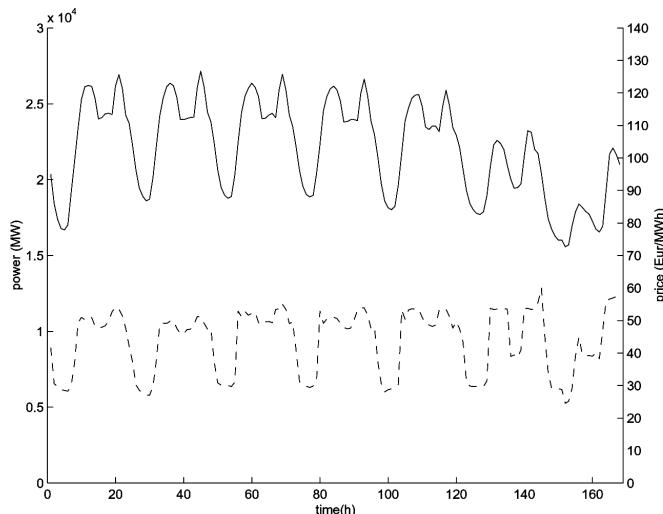


Fig. 2 Hourly loads (continuous curve) and market prices (dashed).

Both the load and the market price series should be reordered in decreasing load order obtaining a LDC and a price-duration curve that corresponds to the loads in the LDC. The price-duration-curve obtained will be nonsmooth and may

even be nondecreasing (see Fig. ??). However, fitting a straight line or a low order polynomial to it, a decreasing line or function will generally be obtained. Given the variability of the price-duration curve, it seems reasonable to fit a straight line to it. Let b^i and l^i be the basic and linear coefficient of such line for the i^{th} interval. (Predictions of b^i and l^i could be obtained taking into account both the series corresponding to the same interval in several successive years and that of successive intervals.)

6.2 Maximum profit objective function

In order to determine the maximum-profit objective function, a simplifying assumption is convenient regarding the shape of the unit contributions in the generation-duration curve. Instead of having some units (particulary those with the lowest loading order) with an irregular shape in its right side, it will be assumed that the contribution of all units will have a rectangular shape with height C_j (for unit j) and base length E_j^i/C_j as in Fig. 4.

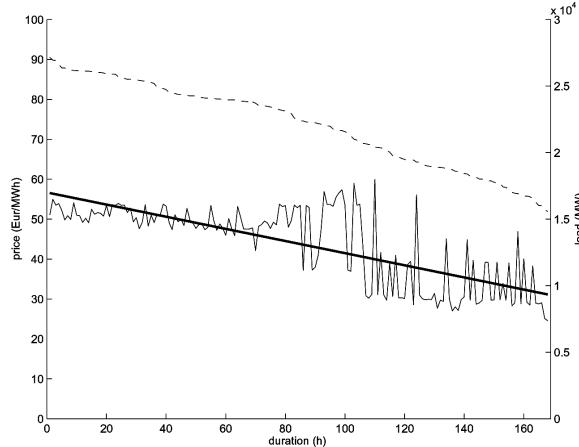


Fig. 3 Market prices ordered by decreasing load power (thin continuous curve) in weekly interval, market-price linear function (thick line), and LDC (dashed).

The profit (price minus cost) of unit j in interval i will be:

$$\int_0^{E_j^i/C_j} C_j \{b^i + l^i t - \tilde{f}_j\} dt = (b^i - \tilde{f}_j) E_j^i + \frac{l^i}{2C_j} E_j^i {}^2$$

and adding for all intervals and units, taking into account the external energy and using (8.9) we get the profit function to be maximized:

$$\sum_i^{n_i} \left[\sum_j^{n_u} \left\{ (b^i - f_j) E_j^i + \frac{l^i}{2C_j} E_j^i {}^2 \right\} - \tilde{f}_{n_u+1} \hat{E}^i \right] \quad (8.20)$$

with $f_j = \tilde{f}_j - \tilde{f}_{n_u+1}$, which is quadratic in the generated energies. Given that

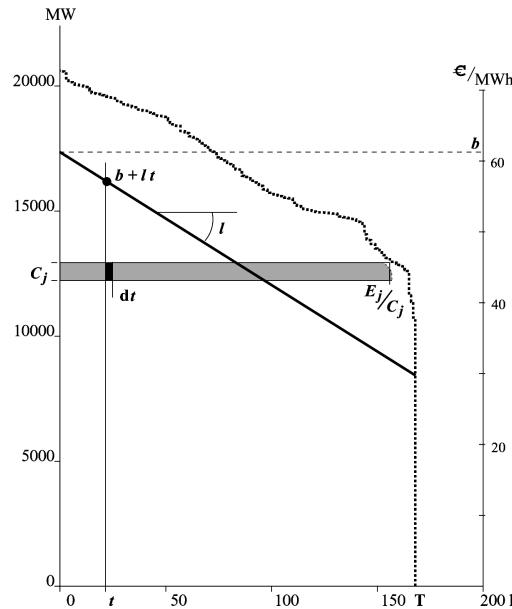


Fig. 4 Long-term price function for a time interval and contribution of j^{th} unit.

$\tilde{f}_{n_u+1} \hat{E}^i$ is a constant, the problem to be solved is:

$$\min_{E_j^i} \quad \sum_i^{n_i} \sum_j^{n_u} \left\{ (f_j - b^i) E_j^i - \frac{l^i}{2C_j} E_j^i {}^2 \right\} \quad (8.21)$$

$$\text{subject to: } \sum_{j \in U} E_j^i \leq \hat{E}^i - W^i(U) \quad \forall U \subset \Omega^i \quad i = 1, \dots, n_i \quad (8.22)$$

$$A_{\geq}^i E^i \geq R_{\geq}^i \quad i = 1, \dots, n_i \quad (8.23)$$

$$\sum_i A_{\geq}^{0i} E^i \geq R_{\geq}^0 \quad (8.24)$$

$$A_{\underline{=}}^i E^i = R_{\underline{=}}^i \quad i = 1, \dots, n_i \quad (8.25)$$

$$\sum_i A_{\underline{=}}^{0i} E^i = R_{\underline{=}}^0 \quad (8.26)$$

$$E_j^i \geq \underline{0} \quad j = 1, \dots, n_u, \quad i = 1, \dots, n_i \quad (8.27)$$

It should be noted that, should all b^i and l^i be zero, the solution of the maximum profit problem (8.24) would be the same as that of the minimum cost problem (8.12-8.18). Otherwise, the cost of the maximum profit solution is higher than that of the minimum cost solution.

Given that $l^i < 0$, the quadratic of the objective function of (8.24) is positive definite, thus problem (8.24) has a unique global minimizer.

7 Coding the load-matching constraints

The main difficulty of the direct solution of the Bloom and Gallant model is the exponential number of load-matching inequality constraints (8.13). These constraints are avoided in the application of the Ford-Fulkerson [22] or the Dantzig-Wolfe column generation method [23, 21, 22], or are generated as they are required in the active set method [18]. In a direct solution by linear or quadratic programming all $n_i \times (2^{n_u} - 1)$ constraints must be explicitly created.

Leaving aside the storage and processing time for these many load-matching inequality constraints, their creation has two parts: the linear coefficients, which is fast [19], and the *rhs*'s, which is very time consuming as it requires lots of calculation.

For each interval i and for the units of each subset U of the set Ω^i we must first calculate $L_U^i(x)$ starting from $L_0^i(x)$ by successive convolution for all units j in U using (8.1), and then compute

$$\widehat{E}^i - W^i(U) = \widehat{E}^i - T^i \int_0^{\widehat{P}^i} L_U^i(x) dx$$

using numerical integration. This means a lot of arithmetic operations.

8 The Ford-Fulkerson column-generation method applied to the multi-interval problem

Constraints (8.22) and (8.27) define, for each interval, a convex polyhedron whose vertices can be easily calculated. To apply the Ford-Fulkerson procedure, energies $E^i \in \mathbb{R}^{n_u}$ must be expressed as convex combinations of all vertices V_k^i of the i^{th} interval polyhedron:

$$E^i = V^i \Lambda^i, \quad V^i \in \mathbb{R}^{n_u \times n_V^i} \quad \Lambda^i \geq \underline{0}, \quad \mathbb{I}' \Lambda^i = 1 \quad \forall i$$

$\mathbb{I}' = [1 \ 1 \ \dots \ 1]$ being the all one vector.

The number n_V^i of vertices of one such polyhedron is very high as the number of constraints (8.22) that define it, jointly with the nonnegativity bounds (8.27), is exponential: 2^{n_u} (which is over a million for $n_u=20$). Note that no account is made of extreme-rays as the nature of the constraints and nonnegativity bounds prevents these.

Subtracting surpluses $S^i \in \mathbb{R}^{n_{\geq}^i}$, $i=0, 1, \dots, n_i$ in the inequalities, problem (8.21-8.27) can be rewritten as:

$$\min_{S^0, S^i, \Lambda^i} \sum_{i=1}^{n_i} \left\{ (f - b^i)' V^i \Lambda^i + \frac{1}{2} \Lambda^i' V^i Q^i V^i \Lambda^i \right\} \quad (8.28)$$

$$\text{subject to: } \mathbb{I}' \Lambda^i = 1 \quad i = 1, \dots, n_i \quad (8.29)$$

$$\begin{cases} A_{\leq}^i V^i \Lambda^i = R_{\leq}^i \\ A_{\geq}^i V^i \Lambda^i - S^i = R_{\geq}^i \end{cases} \quad i = 1, \dots, n_i \quad (8.30)$$

$$\sum_{i=1}^{n_i} A_{\leq}^{0i} V^i \Lambda^i = R_{\leq}^0 \quad (8.31)$$

$$\sum_{i=1}^{n_i} A_{\geq}^{0i} V^i \Lambda^i - S^0 = R_{\geq}^0 \quad (8.32)$$

$$S^0 \geq \underline{0} \quad S^i \geq \underline{0}, \quad \Lambda^i \geq \underline{0} \quad i = 1, \dots, n_i. \quad (8.33)$$

which is quadratic in Λ^i and lends itself to being solved by the column-generating method of Ford-Fulkerson [9].

The convex coefficients $\Lambda^i \in \mathbb{R}^{n_V^i}$, $i=1, \dots, n_i$ and the surpluses $S^i \in \mathbb{R}^{n_{\geq}^i}$, $i=0, 1, \dots, n_i$ of the inequalities are the variables in the problem. In (8.29-8.33) there are linear equality constraints and non-negativity bounds only.

9 Murtagh and Saunders algorithm using a Column Generation procedure

Given a problem such as (8.28-8.33) we apply the Murtagh and Saunders algorithm [12] using the column generation procedure. The outline of the method is:

- 0.- $k := 0$; Given an initial feasible point Λ_0 , S_0 , classify each variable as basic, superbasic or nonbasic. Let n_U be the number of superbasic variables.
- 1.- Compute the projected gradient, $\|Z'G\|$
- 2.- If $\|Z'G\| \leq \epsilon$

- Compute the Lagrange multipliers Σ of the active non-negativity bounds.
 - Look for a constraint c_l , having a negative multiplier, $\sigma_l < 0$
 - If there is any $\sigma_l < 0$ then
 - $n_U := n_U + 1$
 - Update $\|Z'G\|$
 - else
 - END
- 3.- If $\|Z'G\| > \epsilon$
- Compute a descent direction for the basic and superbasic variables, d_k
 - Determine the step length, α_k
 - Update the variables:
 - $\Lambda_{k+1} := \Lambda_k + \alpha_k d_{\Lambda_k}$
 - $S_{k+1} := S_k + \alpha_k d_{S_k}$
 - Update the basic, superbasic and nonbasic sets
 - $k := k + 1$; go to step 1.-

9.1 Obtaining an initial feasible point

Obtaining a feasible point is not trivial when there are non-load-matching constraints.

As with the active set methodology [18], the feasible point is obtained from a point satisfying only the load-matching constraints of all intervals and adding one constraint at a time, plus either a non-zero surplus for the constraint added or a new vertex, until all constraints are satisfied. The details of this process can be found in [21].

9.2 Variable classification

In an active set methodology (such as Murtagh and Saunders is), the active constraints at a feasible point Λ_k , S_k are either general linear constraints or simple bounds.

At a typical iteration, the matrix of active constraints \hat{A} will contain all the general linear constraints and an additional set of rows of the identity matrix that corresponds to variables at zero.

It is important to mention that in this problem there are only lower (non-negativity) bounds because the upper bound 1 for the convex coefficients λ_k^i is implicit in the convexity constraints ($\mathbf{I}'\Lambda^i = 1 \forall i$).

As we are in quadratic programming, there is no *a priori* number of fixed variables. Let n_N denote the number of fixed ($=0$) variables at the current iteration. Then the constraints matrix is (conceptually) partitioned as follows

$$\mathbf{B} = \begin{array}{|c|c|c|} \hline & \mathbb{I} & \mathbb{I} \\ \hline & \mathbf{A}_\leq \mathbf{V}_B & \emptyset \\ \hline & \mathbf{A}_> \mathbf{V}_B & -\hat{\mathbb{1}}_B \\ \hline \end{array} \quad \mathbf{U} = \begin{array}{|c|c|c|} \hline & \mathbb{I} & \mathbb{I} \\ \hline & \mathbf{A}_\leq \mathbf{V}_U & \emptyset \\ \hline & \mathbf{A}_> \mathbf{V}_U & -\hat{\mathbb{1}}_U \\ \hline \end{array} \quad \mathbf{N} = \begin{array}{|c|c|c|} \hline & \mathbb{I} & \mathbb{I} \\ \hline & \mathbf{A}_\leq \mathbf{V}_N & \emptyset \\ \hline & \mathbf{A}_> \mathbf{V}_N & -\hat{\mathbb{1}}_N \\ \hline \end{array}$$

Fig. 5 General constraint matrix partitioned into basic, superbasic and nonbasic matrices.

The $n_B \times n_B$ (where $n_B = n_+ + n_\geq + n_i$) “basis” matrix B is square and non-singular, and its columns correspond to the *basic* variables. The n_N columns of N correspond to the *nonbasic* variables (those fixed at 0). The $n_U = (n_V + n_\geq) - n_B - n_N$ columns of the U matrix correspond to the remaining variables, which will be termed *superbasic*.

9.3 The projected gradient

A necessary (but not sufficient) condition to be at the optimizer is that the projected gradient vanishes:

$$\|Z'G\| \leq \epsilon$$

We define the matrix of the null space Z , such that $\hat{A}Z = 0$, as

$$Z = \begin{bmatrix} -B^{-1}U \\ \mathbb{1} \\ \mathbf{0} \end{bmatrix}$$

$\mathbb{1}$ is the identity matrix of size n_U , number of superbasics.

G is the gradient of the objective function. As we are dealing with a quadratic function, the gradient at the point Λ_k , S_k for each group of variables is:

$$\begin{aligned} \cdot G_{\Lambda_B} &= (f - b)'V_B + V'_B Q V_B \Lambda_B + V'_B Q V_U \Lambda_U + V'_B Q V_N \Lambda_N & G_{S_B} &= 0 \\ \cdot G_{\Lambda_U} &= (f - b)'V_U + V'_U Q V_B \Lambda_B + V'_U Q V_U \Lambda_U + V'_U Q V_N \Lambda_N & G_{S_U} &= 0 \\ \cdot G_{\Lambda_N} &= (f - b)'V_N + V'_N Q V_B \Lambda_B + V'_N Q V_U \Lambda_U + V'_N Q V_N \Lambda_N & G_{S_N} &= 0 \end{aligned}$$

The terms where Λ_N appears, vanish because $\Lambda_N = 0$. The final expression of the projected gradient is as follows:

$$Z'G = G_U - U'B^{-1}'G_B = G_U - U'\Pi \tag{8.34}$$

where G_B and G_U refer to the gradient with respect to the basic and superbasic, G_N to that of the nonbasic, and Π comes from solving system $B'\Pi = G_B$.

9.4 Computation of the multipliers and generation of new vertices

The overdetermined system $\widehat{A}' \begin{bmatrix} \Pi \\ \Sigma \end{bmatrix} = G$ is compatible when $Z'G = 0$. The detailed subsystem $\begin{bmatrix} B' & \\ N' & \mathbb{1} \end{bmatrix} \begin{bmatrix} \Pi \\ \Sigma \end{bmatrix} = \begin{bmatrix} G_B \\ G_N \end{bmatrix}$ is:

The solution procedure solves first

$$B' \begin{bmatrix} \Pi_\Lambda \\ \Pi_A \end{bmatrix} = G_B$$

for $\Pi_\Lambda \in \mathbb{R}^{n_i}$, which are the multipliers of the convexity constraints, and $\Pi_A \in \mathbb{R}^{n=+n_{\geq}}$, which are the multipliers of the non-load-matching constraints. This calculation is already performed in the projected gradient computation (8.34).

From the equations that yield the multipliers Σ two possible types of equation follow. Either:

$$\pi_\lambda^i + v_{Nk}^{i'}(A'\Pi_A)^i + \sigma_k^i = G_{Nk}^i \quad i = 1, \dots, n_i \quad (8.36)$$

hence (recall that $G_{Nk}^i = (f - b^i)' v_{Nk}^i + v_{Nk}^{i'} Q^i V_B^i \Lambda_B^i + v_{Nk}^{i'} Q^i V_U^i \Lambda_U^i$):

$$\sigma_k^i = ((f - b^i) + Q^i V_B^i \Lambda_B^i + Q^i V_U^i \Lambda_U^i - (A' \Pi_A)^i)' v_{Nk}^i - \pi_\lambda^i \quad (8.37)$$

i.e., there is a nonbasic vertex of interval i if $\sigma_k^i < 0$, and this will be so if for the modified costs $\widehat{f}^i = (f - b^i) + Q^i V_B^i \Lambda_B^i + Q^i V_U^i \Lambda_U^i - (A' \Pi_A)^i$ the vector of energies $v_{N_k}^i$ yields a cost lower than π_λ^i .

The other equation we obtain from the Σ equations is:

$$-\pi_{\geq k}^i + \sigma_k^i = 0 \quad \rightarrow \quad \sigma_k^i = \pi_{\geq k}^i \quad i = 0, 1, \dots, n_i \quad (8.38)$$

which tells that the surplus $s_{\geq k}^i$ ($i=1, 2, \dots, n_i, 0$) will become superbasic (relaxing the active constraint $A_{\geq k}^i$) whenever $\pi_{A \geq k}^i < 0$.

The problem of finding a (nonbasic) vertex v_{Nk}^i loading order, and compute the elements of v_{Nk}^i by successive convolution (8.1) and integration (8.2). (And checking whether $\hat{f}^i' v_{Nk}^i < \pi_\lambda^i$ for some interval i .)

It is in the calculation of vertices that the nonavailability of units — by programmed overhauling during the interval — is taken into account.

9.5 Finding a descent direction

If we have not reached the optimizer, we must find another feasible point that decreases the objective function value. As we are dealing with a constrained problem, a feasible direction is $d = Zp_z$, for any p_z : $d = \begin{bmatrix} -B^{-1}U \\ \mathbf{1} \\ \mathbf{0} \end{bmatrix} p_z =$

$\begin{bmatrix} -B^{-1}Up_z \\ p_z \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} d_B \\ d_U \\ d_N \end{bmatrix}$ where the nonbasic variables do not change their value. The projected gradient direction, $p_z = -Z'G'$, can be employed or Newton's method:

$$Z'HZp_z = -Z'G' \quad (8.39)$$

where $H = V'QV$ is the Hessian matrix.

Our computational experience is that the projected gradient direction has a poor convergence. Newton's direction is computationally harder to obtain but is much more efficient. When the step length applied is 1, only one iteration is required to achieve $\|Z'G'\| < \epsilon$. However, the computational experience shows that, when applying a step length of 1 using Newton's direction, sometimes rounding errors make necessary more than one iteration.

9.6 Computation of the step length

Given a feasible point Λ_k , S_k and a direction d_k , we choose a new point as $\Lambda_{k+1} := \Lambda_k + \alpha_k d_{\Lambda_k}$ and $S_{k+1} := S_k + \alpha_k d_{S_k}$ where d_{Λ_k} and d_{S_k} are the components of d_k related to Λ_k and S_k respectively. The optimal step length

$$\alpha_k^* = \frac{-G'_{\Lambda_k} d_{\Lambda_k}}{d'_{\Lambda_k} V' Q V d_{\Lambda_k}}$$

should be $\alpha_k^* = 1$ if we use Newton's direction. α_k^* may lay beyond the upper limits due to the basic and superbasic variable change.

The variables must be nonnegative, thus:

$$\begin{aligned}\bar{\alpha}_{Bk} &= \min \left\{ \frac{\lambda_{Bk}^j}{|d_{\Lambda_{Bk}}^j|} \mid \forall j \mid d_{\Lambda_{Bk}}^j < 0, \frac{S_{Bk}^j}{|d_{S_{Bk}}^j|} \mid \forall j \mid d_{S_{Bk}}^j < 0 \right\} \\ \bar{\alpha}_{Uk} &= \min \left\{ \frac{\lambda_{Uk}^j}{|d_{\Lambda_{Uk}}^j|} \mid \forall j \mid d_{\Lambda_{Uk}}^j < 0, \frac{S_{Uk}^j}{|d_{S_{Uk}}^j|} \mid \forall j \mid d_{S_{Uk}}^j < 0 \right\}\end{aligned}$$

The step length is $\alpha_k = \min\{\bar{\alpha}_{Bk}, \bar{\alpha}_{Uk}, \alpha_k^*\}$. Depending on which one gives α_k , changes in the basic, superbasic and nonbasic sets may occur.

9.7 Changes in the variable sets

In this section we will use variable both to refer to Λ or to S) reaches its bound. In the first case, where all step length can be done, any change in the basic, superbasic and nonbasic sets is needed. In the second case, when some variable limits the step length, there are some changes in the groups.

Also, each variable is upper bounded by 1, but we do not have to take care about this because there are the coefficient convexity constraints ($\sum \Lambda^i = 1 \forall i$) explicitly in the model.

nonbasic set. No more changes happen. basic set is defined as a set of variables related to columns of the constraints matrix which form a base. For so, a superbasic has to be chosen in order to belong to the basic set.

Should α_k be $\alpha_k = \alpha_k^*$, no changes occur in the working set. Should α_k be $\alpha_k = \bar{\alpha}_{Uk}$, a superbasic variables becomes zero and changes to nonbasic.

The case of $\alpha_k = \bar{\alpha}_{Bk}$ is more complicated because a basic variable becomes zero, and changes to nonbasic while a superbasic variable changes to basic to substitute it. The new basis has to be refactorized.

In theory, the superbasic chosen to be basic has only to be linearly independent from the remaining basics. In practice, sometimes we can get stuck without any apparent reason.

9.8 Choosing a superbasic variable to enter the basis

choosing one of them but not all of them work properly. Not only we have to get a new basis (so the column related to the new basic must be linearly independent from the older ones) but also the condition number of the new B should be low enough.

the direction and multipliers computations, both crucial for the suitable ongoing of the algorithm).

The choice of superbasic to enter the basis is important for the solution accuracy and convergence. It is convenient to keep the condition number of B as low as possible in order to get accurate calculations of Π and d_B .

Once known the basic variable l that leaves the basis, which of the superbasic ones will perform better?

The new basis \tilde{B} will be as the former one B except the leaving column l . The change using an η matrix, can be expressed

$$\tilde{B} = B\eta$$

different from the unit matrix is in position l .

$$\eta = \begin{pmatrix} 1 & w_1 \\ & w_2 \\ 1 & \ddots & \vdots \\ & w_l \\ \ddots & & \vdots \\ w_{n_B} & & 1 \end{pmatrix}$$

with the components w_i obtained from the vector w that solves $Bw = U_e$. U_e being the entering column of the superbasic set. It is easy to find an upper bound to the condition number of the new basis:

$$\text{cond}(\tilde{B}) \leq \text{cond}(B) \cdot \text{cond}(\eta)$$

The eigenvalues of η are all ones except w_l , thus its condition number is

$$\text{cond}(\eta) = \begin{cases} \text{if } w_l \leq 1 & \rightarrow 1/w_l \\ \text{if } w_l > 1 & \rightarrow w_l \end{cases}$$

The calculation of the l th row of $B^{-1}U$ is at no cost if using Newton's direction, since the explicit calculation of the upper part of Z is required.

9.9 Management of the nonbasic set. Differences between Ford-Fulkerson algorithm and Dantzig-Wolfe algorithm

The main advantage of the column-generation procedure is that vertices (columns) are only generated when they are required. The basic and superbasic vertices have to be generated and stored properly. At the beginning, there is no nonbasic vertex but as the procedure evolves some nonbasic vertices are known. We can do two things with them: get rid of them or store them (and in a next iteration any known nonbasic vertex can become superbasic again).

The version in which the known nonbasic vertexs are deleted is called the Ford-Fulkerson algorithm (FF) and the one that keeps them is called the Dantzig-Wolfe algorithm (DW). In the DW algorithm, before generating a new vertex, the multipliers of the known nonbasic vertices are computed and if there is any negative, it is reentered as a superbasic. In the results section both methods are compared.

10 Computational results

10.1 Test cases

The characteristics of the test cases employed are summarized in Table 8.1. The fourth column, $\sum_i n_u^i$, is the number of variables and the last but two column contains $\sum_i (2^{n_u^i} - 1)$, which is the number of load-matching inequality constraints. All cases except ltp06 correspond to a certain Spanish generation company together with the rest of the Spanish power pool with a different degree of disaggregation of the generation units; the loads satisfied are those of the Spanish power pool. Case ltp06 refers to the planning of a single German generation company considering only its own load. One or more pseudo-units represent, in all cases, the hydrogenation of one or several basins using the approximate hydromodel of section 5.4.

Tabla 8.1: Test cases for long-term electric power planning

c”se	n_i	n_u	$\sum_i n_u^i$	$\sum_i n_{\leq}^i$	n_{\leq}^0	$\sum_i n_{\geq}^i$	n_{\geq}^0	$\sum_i (2^{n_u^i} - 1)$	$\sum_i \sum_j^{n_u+1} f_j E_j^i$ solver (€)
ltp01a	11	13	140	0	2	0	2	79861	Cplex 4837512292
ltp01b	11	13	140	0	2	1	4	79861	Cplex 4854704625
ltp02a	11	15	162	0	2	33	3	319477	Cplex 3587429530
ltp02b	11	15	162	0	2	34	5	319477	Cplex 3622023526
ltp03a	11	17	183	0	2	54	5	1245173	Cplex 3580260681
ltp03b	11	17	183	0	2	55	7	1245173	Cplex 3624657306
ltp04a	11	18	193	0	2	64	6	2457589	Cplex 3579624419
ltp04b	11	18	193	0	2	65	8	2457589	Cplex 3624160513
ltp06	15	29	416	0	1	15	3	3758096369	ac.set 1070527267

Market-share constraints can be imposed. Cases whose name ends with “a” do not have any market-share constraint imposed. Cases ending with “b” have market-share constraints associated to the units of the SGC imposed and active.

As mentioned earlier, the purpose of these problems and computational tests is twofold:

- to test the models developed, described in this work, and to observe the influence of several parameters associated with the models, and
- to have reliable results (obtained with a reliable code for linear and quadratic programming: *Cplex 7.5*) for the problems posed with which to check alternative specialised algorithms to solve the same problems, specifically

the Dantzig-Wolfe and Ford-Fulkerson column generation algorithms, the active set algorithm, and other algorithms to be developed.

10.2 Performance of the Ford-Fulkerson procedure and comparison with the active set method

Tabla 8.2: Comparison of the active set, and the Ford-Fulkerson column generation method

case	active set method				Ford-Fulkerson cg						Cplex 7.5		
	feas. iters.	total iters.	time (s)	dig. ag.	feas. iters.	total iters.	time (s)	ver. gen.	ver. opt.	dig. ag.	total iters.	time (s)	rhs (h)
ltp01a	193	246	6.6	10	21	79	7.2	147	15	10	781	1.3	0.44
ltp01b	239	312	9.0	9	21	224	16.4	396	18	9	2354	2.35	0.44
ltp02a	450	642	62.5	10	128	357	14.4	254	20	10	3285	11.0	2.28
ltp02b	513	734	80.1	9	128	516	16.1	293	24	10	7646	16.9	2.28
ltp03a	672	964	197.1	10	310	831	20.5	348	23	10	12622	56.8	9.52
ltp03b	781	1096	348.0	9	310	1213	21.6	354	33	9	23213	86.2	9.52
ltp04a	938	1233	508.2	10	400	796	23.7	383	25	9	17447	115.1	19.27
ltp04b	1075	1404	756.6	10	400	1768	38.5	603	45	9	42785	212.0	19.27
ltp06	1803	2646	24.3	—	51	585	5.0	466	31	10	n.a.	n.a.	n.a.

Both the active set and the FFcg methods require a considerable number of iterations to reach a feasible solution. Their numbers appear under the heading “feas. iters.” (feasibility iterations) in Table 8.2; the number of iterations to achieve the optimizer is shown next. After that, the required CPU time, and the number of figures of agreement of the objective function value with that obtained with a different solver are shown, as indicated in the last two columns of Table 8.1. The last three columns in Table 8.2 show the results obtained using an AMPL plus Cplex 7.5 solution [19], the last column giving the long computation times required, in hours(!), to have the rhs’s of the $\sum_i(2^{n_u^i}-1)$ load-matching constraints (8.13).

Several conclusions can be drawn from the results of Table 8.2. The first is that the FFcg method is quicker to get to the solution and that the rate of increase of the time required with problem size is lower in the case of FFcf than with the active set or the direct linear programming solution.

The next issue is precision. Direct linear programming, the active set method and the FFcg procedure reach practically the same optimizer (the number of agreement digits of these methods’ solution is 9 or more for all cases). Four agreement digits would be fairly acceptable from an engineering view-point, given

that many data in this problem are approximations or predictions. Therefore it could be thought that the optimization process could be stopped when the objective function does not change in the first five or six figures over a number of iterations. It must be borne in mind that the active set method for a linear program behaves like linear programming, and obtaining the right set of active constraints produces exactly the same optimizer. However, the FFcg procedure generates the optimizer as the convex combination of vertices of the polyhedrons of feasible points (one for each interval in long-term power planning). Thus the calculation of the optimizer, and its objective function value, requires many more arithmetic operations. The column with header “ver. opt.” contains the number of vertices at the optimizer. On average, we have 2 vertices for each interval.

Through the `-pg` option in the Fortran compilation of the programs and the standard Unix program `gprof` (profiling), it is possible to analyze where the CPU time is spent during execution. It was found that most of the execution time of the active set implementation (over 90%) went to calculating the *rhs*'s of the new active constraints tried and, on average, about 20 new constraints are tried per iteration. With the FFcg implementation almost as much computation (about 80%) is due to calculating new vertices, which involve the same routines of convolution and integration as the calculation of the *rhs*'s. However in the FFcg the number of vertices generated per iteration is less than one, as in many iterations a slack variable is made active, and the number of iterations required has been always below that of the active set procedure.

It is not surprising that case ltp06, though bigger than cases ltp02 and ltp03, and requiring more iterations than former cases, takes less time to convergence. This is because the convolutions are much shorter in ltp06 than in the other cases because the load to be matched (of a single company in Germany) is much lower than that of the Spanish power pool, and a uniform 1 MW step is taken for storing the probability distributions of load still to be supplied, and for integration.

All test cases have been solved with two different objective functions: the linear minimum cost (8.12) and the quadratic of maximum profit (8.20). The linear cost problems have been solved using the linear programming code in *Cplex* 7.5 package [5], while for the quadratic profit problem the barrier separable QP solver [27] in *Cplex* 7.5 package is employed, both through an *AMPL* [10] model and data files. Prior to the solution, the *rhs*'s of the load-matching inequality constraints (8.13) have been calculated using an separate program, whose required CPU time is reported in the last column of Table 8.1. The calculated *rhs*'s are a part of the *AMPL* data files used.

The solutions obtained with the Ford-Fulkerson column generation, which is the most efficient [22], are compared with those obtained through *AMPL* plus *Cplex* 7.5 quadratic programming and with Dantzig-Wolfe column generation..

The second column of Table 8.3 has the input times required by the *AMPL* data files. These times are important because the data files, due to the *rhs*'s of

Tabla 8.3: Comparison of AMPL plus Cplex, and the Dantzig-Wolfe and Ford-Fulkerson column generation methods

case	AMPL plus Cplex 7.5				Dantz-Wl.		Ford-Fulk. column gen.		
	input (s)	b qp iters.	b. qp (s)	obj. fun. (8.21) (€)	D.W. ites.	(s)	F.F. ites.	(s)	obj. fun. (8.21) (€)
ltp01a	1.3	34	97.56	9552335013	289	12.1	262	11.6	9552335013
ltp01b		46	55.09	9536489728	258	9.1	240	9.5	9536489725
ltp02a	5.69	59	183.5	10986157177	842	41.4	629	36.5	10986157163
ltp02b		56	176.9	10961049191	1248	51.8	893	44.1	10961049198
ltp03a	24.47	78	1020.3	11004938184	1321	82.3	957	60.4	11004938185
ltp03b		75	977.7	10977720297	1934	97.4	1341	77.9	10977720295
ltp04a	46.88	87	4393.2	11006374461	1423	91.8	1132	79.1	11006374462
ltp04b		116	5787.0	10979064726	2063	109.0	1545	92.0	10979064723
ltp06					1103	558.3	838	423.4	936301399

the load-matching constraints, are very large, e.g., the data file for case ltp04a is over 100Mbyte.

It can be observed that the Ford-Fulkerson column generation proves to be systematically more efficient in iterations and CPU time than Dantzig-Wolfe's. In the table, the enormous time required to calculate the *rhs* terms of the load-matching constraints when using *AMPL* plus *Cplex 7.5* is not included.

10.3 Solutions of long-term maximum profit planning and comparison with the minimum cost solution

Tabla 8.4: Minimum cost and maximum profit solutions with an approximate and linearized full hydromodel

It is clear from the results in Table 8.4 that the maximization of profit with respect to the minimum cost solution brings about a greater increase in generation cost than an increase in profits.

10.4 Effect of market-share constraints

Three market-share constraints have been introduced in cases whose name ends with “b”: one for the first interval, one for the intervals corresponding to the rest of the first year (intervals 2 to 7), and a third for the intervals of the second year (8 to 11). These three sets of successive intervals will be referred to

with the supraindices I , II and III associated to the variables. The market-share constraints refer to the units of the SGC, and force their generation to add up to over a given percentage of the load in the corresponding intervals.

$$\sum_{i \in I_k} \sum_{j \in SGC} E_j^i \geq \mu^{I_k} \sum_{i \in I_k} \hat{E}^i \quad I_k : I, II, III, \quad (8.40)$$

which are of type (8.15), except set I (a single interval) which is of type (8.14).

The criterion employed to fix a market-share μ^{I_k} for the units in the set SGC is based on the Lagrange multiplier values of the market-share constraints $\lambda_{m-s}^{I_k}$ and the expected profit rate in the power pool r^{I_k} : total profit over total load. The Lagrange multipliers $\lambda_{m-s}^{I_k}$ express the rate of change in pool profit due to a market-share increase by the SGC. The reaction of competitor generating companies to a market-share increase by the SGC would be proportional to the resulting $\lambda_{m-s}^{I_k}/r^{I_k}$. Therefore, attainable market-shares are those that produce a small enough value $\lambda_{m-s}^{I_k}/r^{I_k}$. In the cases reported in Table 8.5 the market-shares μ^{I_k} of the SGC have been pushed up until the ratio $\lambda_{m-s}^{I_k}/r^{I_k}$ was close to but did not exceed $\frac{1}{3}$.

Tabla 8.5: Effect of market share constraints on the profit of the SGC

case	μ^I %	λ_{m-s}^I	r^I	μ^{II} %	λ_{m-s}^{II}	r^{II}	μ^{III} %	λ_{m-s}^{III}	r^{III}	total profit (€)	SGC profit (€)
ltp01a	3.75	0.0		3.36	0.0		3.44	0.0		9552335013	263380937
ltp01b	4.2	8.25	26.12	4.2	8.93	27.39	4.2	8.24	25.10	9538257985	268453956
ltp02a	1.85	0.0		1.94	0.0		2.2	0.0		10986157177	174506646
ltp02b	3.4	9.64	29.31	3.3	10.24	31.12	3.4	9.45	29.14	10963147542	206313710
ltp03a	2.08	0.0		2.25	0.0		2.57	0.0		11004938184	205156894
ltp03b	3.6	9.58	29.35	3.6	10.14	31.17	3.8	9.64	29.20	10981583153	235282378
ltp04a	2.08	0.0		2.25	0.0		2.59	0.0		11006374461	205051615
ltp04b	3.6	9.35	29.36	3.6	9.84	31.18	3.7	8.94	29.21	10985461774	232575618

There are also cases whose name ends with “a” in Table 8.5. These cases are the same as those ending in “b” but without the market-share constraints. They are thus equivalent to having imposed a nonactive market share, lower than the share the SGC gets in the solution. It should be noted that the market-share constraints imposed slightly decrease the overall profit, but they noticeably increase the SGC profit.

11 Conclusions

- The long-term hydrothermal planning of the electricity generation problem has been presented and an extension of the Bloom and Gallant model has been put forward in order to solve it. multicommodity network flow model, has been described and a comparative analysis with classical scenario based stochastic programming has been presented. renewable energy sources) in long-term planning has been described.
- A new way of formulating the long-term profit maximization of generating companies in a competitive market has been described.
- An implementation of the Ford-Fulkerson and of Dantzig-Wolfe column generation procedures for solving a quadratic or a linear problem has been presented.
- Implementation details of the solution with *AMPL* of the minimum cost and the maximum profit long-term planning problems have been given.
- The computational experience with the Ford-Fulkerson and of Dantzig-Wolfe column generation procedures and with *AMPL* plus *Cplex 7.5* linear programming and barrier quadratic programming has been reported. This includes:
 - The calculation of the *rhs*'s of the load-matching constraints for the data files required by *AMPL*, which is extremely time-consuming, and which is fairly time-consuming to be read in the solution process. This lengthy calculation, requiring extremely long files to store the results, makes this procedure impractical to use for real cases (where the number of units to consider may be well above one hundred).
 - The solution of the minimum cost and the maximum profit long-term problems.
 - The comparison of the three procedures implemented for a set of real cases, using the approximate hydrogeneration representation, showing that the Ford-Fulkerson column generation is the most efficient, and that *AMPL* plus *Cplex 7.5* is not practical for big cases.
 - The analysis of the effect of market-share constraints for a SGC in the maximum profit solution.

12 Bibliografía

- [1] Balériaux, H. and Jamoule, E. and Linard de Gertechin, F., (1967). Simulation de l'exploitation d'un parc de machines thermiques de production

- d'électricité couplé à des stations de pompage, *Revue E de la Soc. Royale Belge des Electriciens*, 7, 225–245.
- [2] J.R. Birge and F. Louveaux, (1997). Introduction to stochastic programming, *Springer-Verlag*, New York, USA.
 - [3] Bloom, J.A. and Gallant, L., (1994). Modeling dispatch constraints in production cost simulations based on the equivalent load method, *ieeetrps*, 9, 2, 598–611.
 - [4] Conejo, A.J. and Prieto, (2001). Mathematical programming and electricity markets, *TOP*, 9, 1, 1–53.
 - [5] Using the CPLEX Callable Library. User's Manual, version 4.0, (1985). CPLEX Optimization Inc..
 - [6] Harwell Subroutine Library. A catalogue of Subroutines, (1988). *Computer Science and Systems Division, Harwell Laboratory, AERE – R 9185*.
 - [7] Escudero, L.F. and Salmerón, J. and Paradinas, I. and Sánchez, M., (1998). SEGEM: a Simulation Approach for Electric Generation Management, *ieeetrps*, 13, 3, 738–748.
 - [8] Dantzig, G.B. and Wolfe, P., (1960). title=Decomposition principle for linear programmes, *Operations Research*, 8, 101–110.
 - [9] Ford, L.R. and Fulkerson, D.R., (1958). A suggested computation for maximal multicommodity network flows, *Management Science*, 5, 97 - 101.
 - [10] R. Fourer and D.M. Gay and B.W. Kernighan, (1993). AMPL a modeling language for mathematical programming, *Boyd and Fraser Publishing Company*.
 - [11] Gondzio, J., (1998). Warm start of the primal-dual method applied in the cutting-plane scheme, *Mathematical Programming*, 83, 125–143.
 - [12] P.E. Gill and W. Murray and M.H. Wright, (1981). Practical Optimization, *Academic Press*, London, G.B.
 - [13] Nabona, N., (1993). Multicommodity network flow model for long-term hydro-generation optimization, *ieeetrps*, 8, 2, 395 – 404.
 - [14] Nabona, N. and Díez, M.A., (1998). Long-term thermal multi-interval thermal generation scheduling using Bloom and Gallant's linear model, *eioupc* DR9803.
 - [15] Nabona, N., (1998). Probabilistic long-term hydrothermal coordination using expected hydrogeneration contribution functions, *eioupc*, DR9801.

- [16] Nabona, N. and Gil, C., (1999). The MIBG code of multi-interval long-term thermal generation scheduling using Bloom and Gallant's model, *eioupc* DR9901.
- [17] Nabona, N. and Castro, J. and González, J.A, (1995). title=Optimum Long-Term Hydrothermal Coordination with Fuel Limits, *ieeetrps*, 10, 2, 1054–1062.
- [18] Nabona, N. and Gil, C. and Albrecht, J., (2001). title=Long-term thermal power planning at VEW ENERGIE using a multi-interval Bloom and Gallant method, *ieeetrps*, 16, 1, 69–77.
- [19] Nabona, N. and Pagès, A., (2002). Long-term electric power planning in a competitive market using the Bloom and Gallant procedure and a modeling language, *eioupc*, DR2002/24.
- [20] Pagès, A., (2002). Modelització amb AMPL de la coordinació tèrmica a llarg termini de la generació elèctrica segons la formulació de Bloom i Gallant, *Supervised Assignment*, Facultat de Matemàtiques i Estadística, Univ. Politècnica de Catalunya.
- [21] Pagès, A., (2002). Modelització amb AMPL de la coordinació tèrmica a llarg termini de la generació elèctrica segons la formulació de Bloom i Gallant.*Final-year project*, Facultat de Matemàtiques i Estadística, Univ. Politècnica de Catalunya.
- [22] Pagès, A. and Nabona, N., (2002). Long-term electric power planning using the Bloom and Gallant's linear model. A comparison of solution techniques, *eioupc*, DR2002/25.
- [23] Pérez-Ruiz, J. and Conejo, A.J., (2000). Multi-period probabilistic production cost model including dispatch constraints, *ieeetrps*, 15, 2, 502–507.
- [24] Prieto, F.J. and Conejo, A.J., Competitive markets, *TOP*, 15, 2, 502–507.
- [25] Rosenthal, R.E., (1981). A Nonlinear Network Flow Algorithm for Maximization of Benefits in a Hydroelectric Power System, *Operations Research*, 29, 4, 763–784.
- [26] A.J. Wood and B.F. Wollenberg, (1984). Power generation, operation and control, John Wiley & Sons, New York, USA.
- [27] S.J. Wright, Primal-dual interior-point methods, *SIAM*,

Aplicaciones en sistemas de energía eléctrica

Santiago Cerisola, Andrés Ramos
Álvaro Baillo

Escuela Técnica Superior de Ingeniería - ICAI,
e-mails: {Santiago.Cerisola, Andrés.Ramos, Alvaro.Baillo}@iit.icai.upco.es

1 Introducción

Los sistemas de energía eléctrica siempre han sido un campo habitual de aplicaciones de optimización en general (véanse los artículos [5, 8, 12, 20] o el libro [17]) y, en particular, de optimización estocástica [7]. Después de la desregulación que está ocurriendo en numerosos sistemas eléctricos surgen de nuevo oportunidades de modelado resueltas mediante optimización estocástica.

En este capítulo se van a presentar varios ejemplos característicos de planificación y operación de sistemas de energía eléctrica para cuya resolución se utilizan frecuentemente técnicas de descomposición. Existen numerosas referencias de aplicaciones específicas, pero la presentación que se realiza en este capítulo está orientada a mostrar los problemas de una forma didáctica eliminando o simplificando algunas de las complejidades que pueden encontrarse cuando se modelan sistemas eléctricos reales específicos. Las dos primeras aplicaciones, planificación de la expansión de la generación y programación semanal, son clásicas en la literatura de sistemas eléctricos. La tercera, generación de ofertas, corresponde a un modelo específico de programación semanal para mercados eléctricos.

Por otra parte, la introducción de la incertidumbre en el modelado de los problemas de optimización, incrementa el tamaño de los mismos así como la dificultad de su resolución. En este capítulo, se sugieren estrategias de descomposición de los problemas propuestos para abordar su resolución numérica.

2 Planificación de la expansión de la generación

Un modelo de planificación de la expansión de la generación o modelo de inversión tiene por objetivo determinar la composición óptima de las nuevas inversiones de generación (denominada *política óptima*) para satisfacer la demanda de electricidad con el mínimo coste total de inversión en las nuevas instalaciones más el coste de operación de todo el sistema. Otras funciones objetivo alternativas o complementarias pueden ser: medidas de fiabilidad del sistema, emisiones contaminantes, requisitos financieros [15, 19], etc. que pueden ser combinadas bajo técnicas de decisión multicriterio. Las nuevas inversiones pueden estar originadas por varios factores: la retirada de equipos al alcanzar su vida útil, el crecimiento de la demanda de electricidad, la aparición de nuevas tecnologías competitivas que reemplazan a las actuales, la evolución de los costes de combustibles que influyen en los costes de operación de los equipos actuales, etc. La influencia de la demanda se produce no sólo por la energía total solicitada sino también por la potencia máxima asociada. El modelo de planificación que se presenta aquí se orienta a entornos regulatorios tradicionales, aunque ya están surgiendo modelos de expansión de la generación para entornos desregulados [13].

Un modelo de planificación de expansión de la generación minimiza los costes totales (fijos y variables) de expansión del equipo generador para un alcance de varios años t^1 , $t = 1, \dots, T$. Un alcance razonable puede ser de 5 a 15 años. Se trata de un modelo de planificación dinámica, donde se considera explícitamente la cronología de las decisiones de inversión a lo largo del tiempo. Las decisiones principales que se toman con ayuda de este modelo son la potencia a instalar de cada tipo de generación o generador en cada año del alcance del modelo.

La demanda de electricidad tiene un comportamiento estacionario cuyo valor varía a lo largo del tiempo. El modelado habitual en problemas de planificación utiliza la curva denominada duración-carga o monótona de carga. La monótona de demanda de un año t se representa dividida en períodos p , $p = 1, \dots, P$. Sea D_{tp}^s la demanda de potencia de cada período p para cada escenario s^2 , $s = 1, \dots, S$, siendo $Prob^s$ la probabilidad del escenario y sea Dur_{tp} la duración de dicho período. La demanda se supone constante en dicho período. Esto hace que el problema de optimización sea lineal.

Por simplicidad, sólo se considera generación térmica. El sistema de generación se caracteriza mediante un coste fijo, un coste variable y una tasa de disponibilidad, que indica la proporción del tiempo en que el grupo se encuentra disponible. Sea F_{ti} el coste fijo anualizado de inversión de cada generador candidato i , $i = 1, \dots, I$, a ser instalado, depende del año de instalación t ; V_{pi}^s el coste variable de producción del generador en cada período y escenario y A_i la tasa de

¹Se utiliza el índice t para la unidad temporal más relevante en este problema, típicamente el año.

²Los escenarios pueden contemplar entre otros parámetros (como se verá en la siguiente sección) las variaciones de la demanda o de los costes de combustibles a lo largo del tiempo.

disponibilidad de cada grupo. El coste anualizado de inversión se calcula como el coste total de inversión del equipo repartido a lo largo de su vida económica teniendo en cuenta la tasa de descuento del dinero. Utilizando este coste anualizado se simplifica el tratamiento de los efectos finales del modelo. Las tecnologías o grupos a instalar dependen de los costes fijos y variables de cada uno de ellos, de los costes de los combustibles y de las estrategias de abastecimiento energético del país. Los costes variables incluyen los costes de combustible más los costes variables de operación y mantenimiento.

Las variables del problema serán la potencia a instalar de cada generador en cada año del alcance del modelo x_{ti} (son variables enteras) y la potencia de operación de cada generador en cada periodo y escenario p_{tpi}^s . Obsérvese que la potencia a instalar no depende del escenario, mientras que las potencias de operación dependen del escenario. La potencia acumulada instalada al comienzo del alcance del modelo se considera conocida y_{0i} .

Habitualmente se tienen en cuenta estas restricciones en las decisiones de expansión: potencia instalada inicial conocida, máxima (y/o mínima) potencia instalable, inversión máxima (y/o mínima), número máximo (y/o mínimo) de generadores instalables en cada año. Además, también se consideran entre otras estas restricciones de operación: el balance generación demanda en cada año.

2.1 Formulación determinista

Se considera un sistema generador puramente térmico. En la formulación del problema se consideran los siguientes conjuntos de índices, parámetros, variables y ecuaciones.

<i>Conjuntos</i>	
T	Conjunto de años
P	Conjunto de periodos
I	Conjunto de grupos térmicos
S	Conjunto de escenarios

<i>Índices</i>	
t	Índice de años
p	Índice de periodos
i	Índice de grupos térmicos
s	Índice de escenarios

<i>Parámetros</i>		
D_{tp}^s	Demanda de potencia en el periodo p del año t	[MW]
Dur_{tp}	Duración en el periodo p del año t	[h]
F_{ti}	Coste fijo anualizado del grupo i en el año t	[€ /MWaño]
V_{pi}^s	Coste variable del grupo i en el periodo p para el escenario s	[€ /MWh]
A_i	Tasa de disponibilidad del grupo i	[p.u.]
$Prob^s$	Probabilidad del escenario s	[p.u.]
R_t	Coeficiente de margen de reserva en el año t	[p.u.]
\bar{M}_t	Presupuesto máximo anual en el año t	[€]
X_{ti}	Potencia mínima a invertir en el año t para el grupo i	[MW]

<i>Variables</i>		
x_{ti}	Potencia a instalar del grupo i en el año t (variable entera)	[MW]
y_{ti}	Potencia acumulada instalada del grupo i en el año t	[MW]
p_{tpi}^s	Potencia producida por el grupo i en el periodo p del año t para el escenario s	[MW]

Para facilitar la comprensión del modelo vamos a ver en primer lugar un escenario determinista cualquiera. Es decir, s' es un escenario conocido de demanda, por ejemplo el de demanda media o cualquier otro. En este caso, la función objetivo será minimizar la suma de costes fijos de inversión más costes variables de operación para dicho escenario para el alcance del modelo.

$$\min_{x_{ti}, y_{ti}, p_{tpi}^{s'}} \sum_{t=1}^T \sum_{i=1}^I \left(F_{ti} y_{ti} + \sum_{p=1}^P V_{pi}^{s'} Dur_{tp} p_{tpi}^{s'} \right) \quad (9.1)$$

siendo y_{ti} la potencia acumulada instalada cuyo cálculo se formula como

$$y_{ti} = \sum_{t' \leq t} x_{t'i} \quad \forall t, i \quad (9.2)$$

o bien expresándolo con ecuaciones de continuidad para cada año y tecnología

$$y_{ti} = y_{t-1} + x_{ti} \quad \forall t, i \quad (9.3)$$

siendo la potencia instalada al comienzo del alcance del modelo y_{0i} conocida.

Las decisiones de instalación de los generadores son enteras. Luego el problema de optimización es lineal entero mixto.

Las restricciones que condicionan este problema las podemos separar en dos grupos. Uno, las *restricciones de inversión* que sólo afectan a las variables de inversión. Dos, las *restricciones de operación*, que representan la operación del sistema.

Restricciones de inversión

Entre las primeras conviene mencionar, por ejemplo, una limitación en la potencia mínima a invertir en cada año del alcance del modelo \underline{X}_{ti} y un presupuesto máximo anual disponible \overline{M}_t .

$$x_{ti} \geq \underline{X}_{ti} \quad \forall t, i \quad (9.4)$$

$$\sum_{i=1}^I F_{ti} x_{ti} \leq \overline{M}_t \quad \forall t \quad (9.5)$$

Éstas ecuaciones deben incluir también, aunque no está puesta en ellas, la retirada de potencia de generadores al final de su vida útil.

También se puede incluir una restricción que indique que la potencia instalada total en cada año debe ser superior a la demanda máxima más un cierto margen de reserva

$$\sum_{i=1}^I (y_{0i} + y_{ti}) \geq R_t \max_p (d_{tp}^{s'}) \quad \forall t \quad (9.6)$$

siendo R_t un coeficiente de margen de reserva anual.

Restricciones de operación

Entre las restricciones de operación están la cobertura de la demanda y la que relaciona la potencia instalada en cada año con la potencia de operación utilizable para satisfacer la demanda.

La restricción de cobertura de la demanda se formula para cada periodo p de cada año t

$$\sum_{i=1}^I p_{tpi}^{s'} = D_{tp}^{s'} \quad \forall t, p \quad (9.7)$$

Se supone que se dispone de potencia de operación suficiente para cubrir la demanda en cualquier periodo y año. Para ello se asume que siempre existe un generador con potencia suficiente y coste de operación muy elevado (denominado potencia no suministrada) al que se puede recurrir. De esta manera el modelo es

de *recurso completo*, es decir, cualquier decisión de inversión es factible desde el punto de vista de la operación del sistema.

La potencia de operación en cualquier periodo de un año debe ser inferior a la potencia instalada acumulada en ese año reducida por su disponibilidad A_i

$$p_{tpi}^{s'} \leq A_i (y_{0i} + y_{ti}) \quad \forall t, p, i \quad (9.8)$$

Esta ecuación liga las decisiones de inversión con las de operación.

Además de estas restricciones principales están el conjunto detallado de condiciones de funcionamiento de los grupos entre límites, entre otras.

2.2 Formulación estocástica

Entre los parámetros que pueden considerarse estocásticos en un modelo de expansión de la generación podemos citar: los costes fijos y variables de las tecnologías, la demanda, la fecha de disponibilidad de las nuevas tecnologías, incluso la vida útil de los equipos y por consiguiente su fecha de retirada, ver [10, 11, 14]. Los dos más importantes suelen ser la demanda y los costes variables de los generadores. La evolución de la demanda está influenciada por diferentes factores como son: la actividad industrial, el crecimiento económico, las políticas de ahorro energético o las tarifas. Los costes variables están influidos principalmente por la evolución de los costes de los combustibles. Existe otro tipo de incertidumbre que resulta más difícilmente representable en modelos de optimización estocástica como es la incertidumbre regulatoria pero que puede afectar las decisiones de inversión en sistemas desregulados.

En la formulación de este problema estocástico se suponen los escenarios independientes entre sí. Las decisiones de inversión deben ser únicas para el conjunto de escenarios, luego es un problema de *planificación bietapa* siendo las decisiones de inversión las de la primera etapa y las de operación las de la segunda.

La función objetivo recoge los costes de inversión más los costes esperados de operación del conjunto de todos los escenarios

$$\min_{x_{ti}, y_{ti}, p_{tpi}^s} \sum_{t=1}^T \sum_{i=1}^I \left(F_{ti} y_{ti} + \sum_{s=1}^S \sum_{p=1}^P Prob^s V_{pi}^s Dur_{tp} p_{tpi}^s \right) \quad (9.9)$$

Las ecuaciones que afectan únicamente a las decisiones de inversión permanecen idénticas.

$$x_{ti} \geq \underline{X}_{ti} \quad \forall t, i \quad (9.10)$$

$$\sum_{i=1}^I F_{ti} x_{ti} \leq \overline{M}_t \quad \forall t \quad (9.11)$$

$$y_{ti} = y_{t-1,i} + x_{ti} \quad \forall t, i \quad (9.12)$$

$$\sum_{i=1}^I (y_{0i} + y_{ti}) \geq R_t \max_p(D_{tp}^{s'}) \quad \forall t \quad (9.13)$$

mientras que las ecuaciones de operación ahora se repiten para cada escenario

$$\sum_{i=1}^I p_{tpi}^s = D_{tp}^s \quad \forall s, t, p \quad (9.14)$$

$$p_{tpi}^s \leq A_i (y_{0i} + y_{ti}) \quad \forall s, t, p, i \quad (9.15)$$

Este problema de expansión de la generación es de *recurso fijo* puesto que los coeficientes de las variables de inversión y de operación en las restricciones no dependen del escenario.

Descomposición de Benders

Aparte de la formulación directa del problema de optimización anterior cuando las dimensiones lo requieren se puede utilizar el método de *descomposición de Benders* [3, 4, 16]. El lector interesado puede encontrar la descripción del método de descomposición de Benders en el capítulo de este libro titulado *Modelado de algoritmos de descomposición con GAMS*. En este método el problema maestro o problema de inversión envía propuestas de inversión al subproblema o problema de operación y éste devuelve las variables duales de las restricciones donde éstas aparecen. El problema maestro es un problema lineal entero mixto mientras que el subproblema es lineal. El subproblema resulta separable por escenarios y años³ y resoluble independientemente.

El problema maestro se formula como

$$\begin{aligned} & \min_{x_{ti}, y_{ti}, p_{tpi}^s} \sum_{t=1}^T \sum_{i=1}^I F_{ti} y_{ti} + \sum_{s=1}^S \sum_{t=1}^T Prob^s \theta^{st}(x) \\ & x_{ti} \geq \underline{X}_{ti} \quad \forall t, i \\ & \sum_{i=1}^I F_{ti} x_{ti} \leq \bar{M}_t \quad \forall t \\ & y_{ti} = y_{t-1,i} + x_{ti} \quad \forall t, i \\ & \sum_{i=1}^I (y_{0i} + y_{ti}) \geq R_t \max_p(D_{tp}^{s'}) \quad \forall t \end{aligned} \quad (9.16)$$

y el subproblema para cada escenario s y año t

³Según la formulación matemática presentada el problema también resulta separable por periodo. Sin embargo, en formulaciones más realistas existen restricciones de acoplamiento entre periodos que aquí, por ser poco detallada no se han presentado.

$$\begin{aligned}\theta^{st}(x) = & \min_{p_{tpi}^s} \sum_{p=1}^P \sum_{i=1}^I V_{pi}^s D_{tp} p_{tpi}^s \\ & \sum_{i=1}^I p_{tpi}^s = D_{tp}^s \quad \forall p \\ & p_{tpi}^s \leq A_i (y_{0i} + y_{ti}^j) \quad \forall p, i\end{aligned}\tag{9.17}$$

siendo $\theta^{st}(x)$ la función de recurso del subproblema para el escenario s y y_{ti}^j la propuesta de decisión de inversión acumulada para la iteración j .

2.3 Caso ejemplo

Veamos a continuación un caso de estudio muy sencillo escrito en GAMS con las ecuaciones presentadas en las formulaciones determinista y estocástica anteriores. Sus resultados se utilizan de forma pedagógica para mostrar la validez de la solución estocástica frente a las deterministas. El alcance del modelo es de 1 año dividido en 3 períodos. El sistema de generación está compuesto de 4 generadores y se consideran 3 escenarios de demanda con probabilidades 0.2, 0.5 y 0.3.

```
$TITLE Planificación óptima de la expansión de la generación

SETS
  J periodos           / per-1 * per-3 /
  I generadores        / gen-1 * gen-4 /
  S escenarios de demanda / s-1   * s-3   /

PARAMETERS
  F(i) coste fijo de inversión [euro]
  / gen-1 10
    gen-2  7
    gen-3 16
    gen-4  6 /

  PROB(s) probabilidad de cada escenario [p.u.]
  / s-1  0.2
    s-2  0.5
    s-3  0.3 /

  DEM(j) demanda para un escenario [MW]
  TABLE V(i,j) coste variable de operación [euro por MW]

  gen-1  per-1  per-2  per-3
gen-1   40     24     4
gen-2   45     27     4.5
gen-3   32     19.2   3.2
gen-4   55     33     5.5

  TABLE DEMS(s,j) demanda estocástica [MW]
  s-1    per-1  per-2  per-3
  s-1    3      3      2
  s-2    5      3      2
  s-3    7      3      2
```

```

SCALARS
  POTMIN potencia mínima a instalar [MW] / 12 /
  PRSPTO límite presupuestario [euro] / 120 /

VARIABLES
  X(i)      potencia a instalar          [MW]
  Y(j,i)    potencia en operación        [MW]
  YS(s,j,i) potencia en operación estocástica [MW]
  COSTE    coste total

POSITIVE VARIABLES X, Y, YS

EQUATIONS
  COST     coste total           [euro]
  COSTS   coste total estocástico [euro]
  PRESUP  limitación presupuestaria [euro]
  INSMIN  potencia mínima instalada      [MW]
  BALPOT  potencia en operación menor que instalada [MW]
  BALPOTS potencia en operación menor que instalada estocástica [MW]
  BALDEM  balance de demanda          [MW]
  BALDEMS balance de demanda estocástico [MW] ;

COST .. COSTE =E= SUM(i, F(i) * X(i))
           + SUM((j,i), V(i,j) * Y(j,i)) ;
COSTS .. COSTE =E= SUM(i, F(i) * X(i))
           + SUM((s,j,i), PROB(s) * V(i,j) * YS(s,j,i)) ;

PRESUP .. SUM(i, F(i) * X(i)) =L= PRSPTO ;
INSMIN .. SUM(i, X(i)) =G= POTMIN ;
BALPOT(j,i) .. Y(j,i) =L= X(i) ;
BALPOTS(s,j,i) .. YS(s,j,i) =L= X(i) ;

BALDEM(j) .. SUM(i, Y(j,i)) =G= DEM(j) ;
BALDEMS(s,j) .. SUM(i, YS(s,j,i)) =G= DEMS(s,j) ;

MODEL DETERM / COST, INSMIN, PRESUP, BALPOT, BALDEM / ;
MODEL ESTOCA / COSTS, INSMIN, PRESUP, BALPOTS, BALDEMS / ;

* este bucle resuelve cada escenario determinista por separado

LOOP (s,
  DEM(j) = DEMS(s,j) ;
  SOLVE DETERM MINIMIZING COSTE USING LP ;
) ;

* ahora se resuelve el escenario de demanda media

DEM(j) = SUM(s, PROB(s) * DEMS(s,j)) ;
SOLVE DETERM MINIMIZING COSTE USING LP ;

* ahora se resuelve el problema estocástico

SOLVE ESTOCA MINIMIZING COSTE USING LP ;

```

El problema se resuelve en primer lugar de forma determinista para cada escenario de demanda por separado. Los resultados aparecen en las tres primeras columnas de la siguiente tabla. A continuación se resuelve el problema para

	Escenario Determinista 1	Escenario Determinista 2	Escenario Determinista 3	Escenario Medio	Problema Estocástico
Generador 1	.	0.33	3.67	0.67	0.67
Generador 2	2
Generador 3	3	4.67	3.33	4.53	4.33
Generador 4	9	7	5	6.8	5
Coste total	262	346.67	437.33	355.73	362.47

el valor medio ponderado de la demanda (siguiente columna) y finalmente el problema estocástico (en la última columna). Obsérvese que en este ejemplo sencillo no hay coincidencia entre la solución de ningún escenario y la del problema estocástico.

La función objetivo del problema estocástico es el coste fijo de las decisiones de la primera etapa más el valor esperado de la operación asociada a las decisiones de la segunda etapa (362.47 en el ejemplo). A continuación se calculan algunos parámetros característicos que ayudan a valorar la utilidad del problema estocástico frente a soluciones de escenarios deterministas. El *valor de la solución estocástica* (value of stochastic solution VSS) es la diferencia entre la función objetivo del problema estocástico y la del problema determinista para el valor medio de los parámetros ($362.47-355.73=6.73$). Se denomina *valor esperado con información perfecta* (expected value with perfect information EVWPI) a la suma ponderada para cada escenario de la función objetivo total sabiendo que dicho escenario va a ocurrir con certeza (356.93 para el ejemplo). Es decir, el valor esperado si se revela la incertidumbre antes de tomar las decisiones de la primera etapa. Este valor siempre será menor o igual, en un problema de minimización, que la función objetivo del problema estocástico. Para cada escenario, la solución del problema estocástico es siempre peor o igual que la solución con información perfecta (la función es 280, 349.33 y 439.33 respectivamente). Se denomina *valor esperado de la información perfecta* (expected value of perfect information EVPI) o *arrepentimiento* a la diferencia entre ambas ($280-262=18$, $349.33-346.67=2.66$, $439.33-437.33=2$).

2.4 Caso español

Vamos a estimar a continuación las dimensiones del problema estocástico presentado previamente para un sistema eléctrico de tamaño como el español. Supongamos un alcance del estudio de 10 años, $T = 10$. Cada año la monótona de carga considera períodos de punta, llano y valle en laborable y festivo para cada mes, siendo un total de 72 períodos en cada año, $P = 72$. Suponemos 10 escenarios de demanda, $S = 10$, y un sistema de generación con 80 grupos térmicos, $I = 80$.

Las restricciones de la primera etapa (potencia mínima a invertir en cada año y cada generador, presupuesto máximo anual disponible, margen de potencia instalada con respecto a la demanda, más las restricciones de cálculo de la potencia instalada acumulada) suman un total de $TI + T + TI + T = 2(TI + T) = 1620$. Las restricciones que acoplan decisiones de la primera y segunda etapa son $STPI = 576000$. Las restricciones de operación de la segunda etapa son $STP = 7200$.

Las decisiones de inversión de la primera etapa son $2TI = 1600$ variables enteras mientras que las de la segunda etapa son $STPI = 576000$ variables continuas.

Este tamaño está muy al límite o excede las capacidades actuales de los optimizadores lineales enteros mixtos, luego se necesitaría recurrir a métodos de descomposición para la resolución del problema de planificación de la expansión del sistema eléctrico peninsular español.

3 Programación semanal

El problema de la programación semanal (*unit commitment*) consiste en determinar el conjunto de grupos de generación que se deben conectar y sus niveles de operación para minimizar el coste total de operación sujeto a la satisfacción de la demanda y de una cierta reserva de generación. Para la resolución de este problema, tradicionalmente se han utilizado técnicas de optimización ordinal así como técnicas basadas en programación dinámica. Recientemente, debido a la evolución de los optimizadores para problemas lineales enteros mixtos, este problema es formulado y resuelto como un problema lineal entero mixto de gran tamaño. Sin embargo, esta mejora posibilita la exigencia de modelos de programación semanal más complejos que incluyan incertidumbre en los parámetros modelados. Por otra parte, los nuevos entornos regulatorios modifican el planteamiento de los problemas de programación semanal como problemas de minimización de costes. Ahora se plantea como un problema de determinación de la oferta óptima formulando un problema de maximización del beneficio obtenido.

Estas nuevas características de los problemas de programación semanal complican el modelado tradicional de los mismos así como su resolución algorítmica. En esta sección se presenta el modelado tradicional de un problema de programación semanal y el planteamiento de la Relajación Lagrangiana como método de resolución. Posteriormente se describe un problema estocástico en el que la estocasticidad en la demanda se introduce mediante un árbol de escenarios. La siguiente sección de este capítulo está dedicada a los problemas de programación semanal en mercados eléctricos.

3.1 Formulación determinista

Se considera un sistema generador puramente térmico. En la formulación del problema se consideran los siguientes conjuntos de índices, parámetros, variables

y ecuaciones.

<i>Conjuntos</i>		
<i>T</i>	Conjunto de periodos	
<i>I</i>	Conjunto de grupos térmicos	

<i>Índices</i>		
<i>t</i>	Índice de periodos	
<i>h</i>	Índice auxiliar de periodos	
<i>i</i>	Índice de grupos térmicos	

<i>Parámetros</i>		
<i>D_t</i>	Demanda del periodo <i>t</i>	[MW]
<i>R_t</i>	Coeficiente de margen de reserva térmica del periodo <i>t</i>	[p.u.]
<i>Dur_t</i>	Duración del periodo <i>t</i>	[h]
<i>P_i^{max}</i>	Potencia máxima del grupo <i>i</i>	[MW]
<i>P_i^{min}</i>	Potencia mínima del grupo <i>i</i>	[MW]
<i>L_i^{up}</i>	Límite rampa subida del grupo <i>i</i>	[MW/h]
<i>L_i^{down}</i>	Límite rampa bajada del grupo <i>i</i>	[MW/h]
<i>F_i</i>	Coste fijo del grupo <i>i</i>	[€/h]
<i>V_i</i>	Coste variable de combustible del grupo <i>i</i>	[€/MWh]
<i>C_i^{up}</i>	Coste de arranque del grupo <i>i</i>	[€]
<i>C_i^{down}</i>	Coste de parada del grupo <i>i</i>	[€]
<i>τ_i</i>	Tiempo mínimo de parada del grupo <i>i</i>	[h]
<i>κ_i</i>	Tiempo mínimo de acoplamiento del grupo <i>i</i>	[h]

<i>Variables</i>		
<i>p_{ti}</i>	Potencia producida por el grupo <i>i</i> en el periodo <i>t</i>	[MW]
<i>u_{ti}</i>	Acoplamiento del grupo <i>i</i> en el periodo <i>t</i>	{0,1}
<i>s_{ti}^{up}</i>	Arranque del grupo <i>i</i> en el periodo <i>t</i>	[0,1]
<i>s_{ti}^{down}</i>	Parada del grupo <i>i</i> en el periodo <i>t</i>	[0,1]

El problema de programación semanal debe satisfacer la demanda de potencia en cada uno de los períodos⁴ del alcance temporal

$$\sum_{i=1}^I p_{ti} = D_t \quad \forall t \quad (9.18)$$

imponiendo un margen de reserva que es modelado como

⁴Obsérvese que se utiliza el índice *t* para la unidad temporal más relevante en este problema, típicamente una hora.

$$\sum_{i=1}^I (P_i^{max} u_{ti} - p_{ti}) \geq R_t D_t \quad \forall t \quad (9.19)$$

Para cada grupo térmico, los valores de potencia producida están acotados entre su mínimo técnico y su máxima capacidad disponible

$$P_{ti}^{min} u_{ti} \leq p_{ti} \leq P_{ti}^{max} u_{ti} \quad \forall t, i \quad (9.20)$$

Los cambios en la potencia producida por los grupos térmicos vienen limitados por las ecuaciones de rampa⁵

$$L_i^{down} Dur_t \leq p_{ti} - p_{t-1,i} \leq L_i^{up} Dur_t \quad \forall t, i \quad (9.21)$$

Las decisiones de arranque y parada de los grupos son gestionadas mediante el siguiente grupo de ecuaciones

$$u_{ti} - u_{t-1,i} = s_{ti}^{up} - s_{ti}^{down} \quad \forall t, i \quad (9.22)$$

Algunos modelos de programación semanal más avanzados incluyen requerimientos de tiempo mínimo para unidades que han sido acopladas o desacopladas. En caso de que hayan sido acopladas, se exige un tiempo mínimo de acoplamiento y por el contrario, en caso de que hayan sido desacopladas, debe transcurrir un cierto número de horas hasta que puedan ser acopladas de nuevo. Estas restricciones de tiempo mínimo de parada son modeladas mediante el siguiente conjunto de ecuaciones

$$u_{t+h_t,i} \leq 1 + u_{ti} - u_{t-1,i} \quad \forall t, h_t, i \quad (9.23)$$

donde el conjunto de periodos desplazados, controlados por el índice h_t , puede ser restringido para valores de $h_t \geq 1$ tales que

$$\tau_i \leq \sum_{l=0}^{h_t-1} Dur_{t+l} \quad (9.24)$$

De modo similar, las restricciones de mínimo tiempo de funcionamiento son modeladas como

$$u_{t+h_t,i} \geq u_{ti} - u_{t-1,i} \quad \forall t, h_t, i \quad (9.25)$$

limitando, en este caso, el conjunto de periodos desplazados a aquellos $h_t \geq 1$ tales que

⁵La ecuación siguiente aparece simplificada y supone que las duraciones de los periodos son iguales. La ecuación de rampa se debería aplicar únicamente a la potencia producida por encima del mínimo técnico, aunque en la ecuación 9.21 por simplicidad se aplica a la potencia producida total.

$$\kappa_i \leq \sum_{l=0}^{h_t-1} Dur_{t+l} \quad (9.26)$$

Dado el anterior conjunto de restricciones, el problema de programación semanal minimiza el coste total de explotación que viene dado como

$$\sum_{t=1}^T \sum_{i=1}^I \{Dur_t F_i u_{ti} + Dur_t V_i p_{ti} + C_i^{up} s_{ti}^{up} + C_i^{down} s_{ti}^{down}\} \quad (9.27)$$

El problema de programación semanal, formulado anteriormente, puede ser resuelto mediante técnicas de programación entera mixta como las técnicas de ramificación y acotamiento (*branch and bound*) o técnicas de ramificación y corte (*branch and cut*). En la literatura, también es común encontrar el problema de la programación semanal resuelto mediante la técnica de la Relajación Lagrangiana. Esta técnica es revisada en la siguiente sección, particularizada para el problema de programación semanal anteriormente descrito.

Relajación Lagrangiana

La Relajación Lagrangiana explota la estructura matricial de un problema de optimización en el que un conjunto de ecuaciones complican la resolución del mismo. Para el problema de la programación semanal, este conjunto de restricciones son las ecuaciones de balance de potencia para cada uno de los períodos del alcance temporal (tradicionalmente una semana).

La Relajación Lagrangiana maximiza la función dual obtenida al minimizar la función objetivo dada por el *lagrangiano* sobre el conjunto de restricciones que definen el problema de optimización.

Considérese el problema de programación semanal descrito en la sección anterior. Por simplicidad en la exposición supongamos también que no se incluyen las restricciones asociadas a la reserva de potencia ni las restricciones de mínimo tiempo de acoplamiento o parada. En este caso, el lagrangiano es formulado como

$$\begin{aligned} L(p_{ti}, u_{ti}, s_{ti}^{up}, s_{ti}^{down}, \lambda_t) = \\ \sum_{t=1}^T \sum_{i=1}^I \{Dur_t F_i u_{ti} + Dur_t V_i p_{ti} + C_i^{up} s_{ti}^{up} + C_i^{down} s_{ti}^{down}\} + \\ \sum_{t=1}^T \lambda_t (\sum_{i=1}^I p_{ti} - D_t) \end{aligned} \quad (9.28)$$

siendo $\lambda = \lambda_t$, $t = 1, \dots, T$ el conjunto de multiplicadores de Lagrange. Esta expresión, una vez reordenados los términos, se plantea como

$$\begin{aligned} L(p_{ti}, u_{ti}, s_{ti}^{up}, s_{ti}^{down}, \lambda_t) = \\ \sum_{t=1}^T \sum_{i=1}^I \{Dur_t F_i u_{ti} + (Dur_t V_i + \lambda_t) p_{ti} + C_i^{up} s_{ti}^{up} + C_i^{down} s_{ti}^{down}\} - \\ \sum_{t=1}^T \lambda_t D_t \end{aligned} \quad (9.29)$$

La función dual $\omega(\lambda)$ se obtiene al evaluar el lagrangiano para un valor específico λ del multiplicador sobre el conjunto de restricciones que definen el problema salvo la ecuación de complicación.

$$\begin{aligned} \omega(\lambda) = & \min_{p_{ti}, u_{ti}, s_{ti}^{up}, s_{ti}^{down}} L(p_{ti}, u_{ti}, s_{ti}^{up}, s_{ti}^{down}, \lambda_t) \\ & P_{ti}^{\min} u_{ti} \leq p_{ti} \leq P_{ti}^{\max} u_{ti} \quad \forall t, i \\ & L_i^{\text{down}} \text{Dur}_t \leq p_{ti} - p_{t-1,i} \leq L_i^{\text{up}} \text{Dur}_t \quad \forall t, i \\ & u_{ti} - u_{t-1,i} = s_{ti}^{up} - s_{ti}^{down} \quad \forall t, i \end{aligned} \quad (9.30)$$

El anterior problema de optimización, llamado *subproblema de Lagrange*, es separable en I problemas de optimización, uno por cada grupo térmico considerado, dados como

$$\begin{aligned} \min_{p_{ti}, u_{ti}, s_{ti}^{up}, s_{ti}^{down}} & \sum_{t=1}^T \{ \text{Dur}_t F_i u_{ti} + (\text{Dur}_t V_i + \lambda_t) p_{ti} + C_i^{\text{up}} s_{ti}^{up} + C_i^{\text{down}} s_{ti}^{down} \} \\ & P_{ti}^{\min} u_{ti} \leq p_{ti} \leq P_{ti}^{\max} u_{ti} \quad \forall t \\ & L_i^{\text{down}} \text{Dur}_t \leq p_{ti} - p_{t-1,i} \leq L_i^{\text{up}} \text{Dur}_t \quad \forall t \\ & u_{ti} - u_{t-1,i} = s_{ti}^{up} - s_{ti}^{down} \quad \forall t \end{aligned} \quad (9.31)$$

Existen numerosas técnicas en la literatura orientadas a obtener el máximo de la función dual. Se destacan las técnicas basadas en el subgradiente así como aquellas de aproximación exterior. No es el objetivo de este capítulo presentar una revisión extensa de estos métodos. El lector interesado puede encontrar la descripción del método de aproximación exterior de la función dual en el capítulo de este libro titulado *Modelado de algoritmos de descomposición con GAMS*.

El modelado presentado del problema de programación semanal no incorpora estocasticidad en ninguno de los parámetros del problema. En la siguiente sección se presenta un modelado estocástico del problema de la programación semanal en el cual la estocasticidad ha sido introducida en el parámetro de la demanda a través de una distribución de probabilidad discreta dada como un árbol de escenarios.

3.2 Formulación estocástica

Consideremos el problema de programación semanal en el que la variable de demanda es introducida como un parámetro aleatorio. Se considera que la distribución de dicho parámetro tiene soporte finito (*distribución discreta*) y viene dada a través de un árbol de escenarios. En la figura 9.1 se muestra un ejemplo de un árbol de cuatro escenarios para un caso realista en el que se considera una ramificación del árbol de escenarios en la hora 25 y otra en la hora 49.

Cuando se introducen parámetros aleatorios en un problema de optimización, normalmente se formula y resuelve el *problema determinista equivalente*. Este problema introduce copias de las variables para cada uno de los posibles escenarios

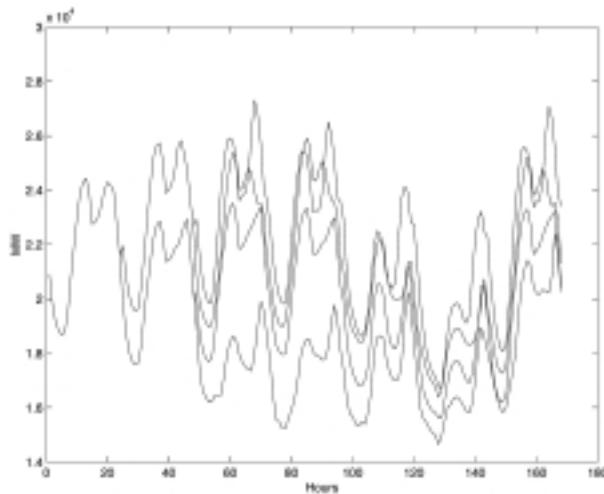


Figura 9.1: Árbol de 4 escenarios

que la realización del parámetro estocástico produce. Esto aumenta drásticamente el tamaño del problema complicando, por tanto, su resolución. Esta representación del problema equivalente puede formularse de modo *compacto* o mediante la introducción de *restricciones de no anticipatividad*. En el primero de los casos, sólo se generan variables para las ramas diferentes del árbol del escenarios. Para el ejemplo de la figura, se genera una copia para cada una de las variables de las primeras 24 horas, dos copias para cada una de las variables de las segundas 24 horas y cuatro copias para el resto de las variables. Por el contrario, una formulación no compacta genera una copia para cada uno de los escenarios⁶ posibles. Las restricciones de no anticipatividad se introducen para igualar los valores de aquellas variables que comparten parte común del árbol.

Para mostrar el modelado de un problema de optimización introduciendo restricciones de no anticipatividad se utiliza el problema de programación semanal anteriormente presentado. El conjunto de índices y parámetros es prácticamente el mismo que en el caso determinista y el conjunto de variables incorpora el superíndice escenario.

<i>Conjuntos</i>	
<i>T</i>	Conjunto de periodos
<i>I</i>	Conjunto de grupos térmicos
<i>S</i>	Conjunto de escenarios

⁶En optimización estocástica un escenario se define como cualquiera de los caminos que van desde el periodo inicial al final (de la raíz a las hojas).

Índices

<i>t</i>	Índice de periodos
<i>h</i>	Índice auxiliar de periodos
<i>i</i>	Índice de grupos térmicos
<i>s</i>	Índice de escenarios

Parámetros

D_t^s	Demanda del periodo t en escenario s	[MW]
R_t^s	Reserva térmica del periodo t en escenario s	[MW]
Dur_t	Duración del periodo t	[h]
P_i^{max}	Potencia máxima del grupo i	[MW]
P_i^{min}	Potencia mínima del grupo i	[MW]
L_i^{up}	Límite rampa subida del grupo i	[MW/h]
L_i^{down}	Límite rampa bajada del grupo i	[MW/h]
F_i	Coste fijo del grupo i	[€/h]
V_i	Coste variable de combustible del grupo i	[€/MWh]
C_i^{up}	Coste de arranque del grupo i	[€]
C_i^{down}	Coste de parada del grupo i	[€]
τ_i	Tiempo mínimo de parada del grupo i	[h]
κ_i	Tiempo mínimo de acoplamiento del grupo i	[h]
$Prob^s$	Probabilidad del escenario s	[p.u.]

Variables para cada escenario s

p_{ti}^s	Potencia producida por el grupo i en el periodo t	[MW]
u_{ti}^s	Acoplamiento del grupo i en el periodo t	{0,1}
$s_{ti}^{up\ s}$	Arranque del grupo i en el periodo t	[0,1]
$s_{ti}^{down\ s}$	Parada del grupo i en el periodo t	[0,1]

El conjunto de restricciones del problema estocástico de programación semanal coincide con el del planteamiento determinista. Debe tenerse en cuenta que cada ecuación es duplicada tantas veces como número de escenarios tiene el problema. Juntamente con estas restricciones, las de no anticipatividad son modeladas como

$$\begin{aligned} p_{ti}^s &= p_{ti}^{s'} && \text{si } s, s' \in N(t) \\ u_{ti}^s &= u_{ti}^{s'} && \text{si } s, s' \in N(t) \\ s_{ti}^{up\ s} &= s_{ti}^{up\ s'} && \text{si } s, s' \in N(t) \\ s_{ti}^{down\ s} &= s_{ti}^{down\ s'} && \text{si } s, s' \in N(t) \end{aligned} \quad (9.32)$$

donde $N(t)$ representa los diferentes conjuntos de escenarios que comparten la misma parte del árbol para el periodo t . En las ecuaciones anteriores es conveniente considerar s como el escenario de índice menor que pertenece a $N(t)$ y $s < s'$ de modo que el número de restricciones generadas sea el menor posible.

La función objetivo minimiza el coste esperado de explotación y es formulada como

$$\sum_{s=1}^S \text{Prob}^s \sum_{t=1}^T \sum_{i=1}^I \{ \text{Dur}_t F_i u_{ti} + \text{Dur}_t V_i p_{ti} + C_i^{up} s_{ti}^{up} + C_i^{down} s_{ti}^{down} \} \quad (9.33)$$

Para la resolución del problema con restricciones de no anticipatividad una alternativa muy popular en la literatura consiste en la relajación de estas restricciones, obteniendo para cada escenario un problema de programación semanal (modificando ligeramente su función objetivo). Esta idea, conocida como descomposición por escenarios (*scenario decomposition*), es presentada brevemente en la próxima sección.

Descomposición por escenarios del problema estocástico

Considérese el problema anteriormente planteado y sean $\lambda_{ti}^s, \mu_{ti}^s, \gamma_{ti}^{up\ s}, \gamma_{ti}^{down\ s}$ las variables duales de las restricciones de no anticipatividad presentadas en la sección anterior. La formulación del lagrangiano para este problema tiene la siguiente expresión

$$\begin{aligned} \sum_{s=1}^S \text{Prob}^s & \sum_{t=1}^T \sum_{i=1}^I \{ \text{Dur}_t F_i u_{ti}^s + \text{Dur}_t V_i p_{ti}^s + C_i^{up} s_{ti}^{up\ s} + C_i^{down} s_{ti}^{down\ s} \} \\ & + \lambda_{ti}^s (p_{ti}^s - p_{ti}^{s'}) \\ & + \mu_{ti}^s (u_{ti}^s - u_{ti}^{s'}) \\ & + \gamma_{ti}^{up\ s} (s_{ti}^{up\ s} - s_{ti}^{up\ s'}) \\ & + \gamma_{ti}^{down\ s} (s_{ti}^{down\ s} - s_{ti}^{down\ s'}) \end{aligned} \quad (9.34)$$

La reordenación de los términos de la función objetivo anterior obtiene una expresión del lagrangiano similar a la función objetivo del problema estocástico de programación semanal. La diferencia aparece en la modificación de los coeficientes que acompañan a las variables de decisión. Esta modificación queda reflejada en la siguiente expresión, donde la notación ha sido simplificada para facilitar su comprensión⁷

$$\sum_{s=1}^S \text{Prob}^s \sum_{t=1}^T \sum_{i=1}^I \{ (\text{Dur}_t F_i + \bar{\mu}_{ti}^s) u_{ti}^s + (\text{Dur}_t V_i + \bar{\lambda}_{ti}^s) p_{ti}^s + (C_i^{up} + \bar{\gamma}_{ti}^{up\ s}) s_{ti}^{up\ s} + (C_i^{down} + \bar{\gamma}_{ti}^{down\ s}) s_{ti}^{down\ s} \} \quad (9.35)$$

El problema de minimización del lagrangiano sobre el conjunto de restricciones del problema estocástico semanal, es claramente separable en S problemas independientes, una vez que se han relajado las restricciones de no anticipatividad. Cada uno de estos problemas puede ser formulado como

⁷Los coeficientes $\bar{\mu}, \bar{\lambda}, \bar{\gamma}, \bar{\gamma}$ reflejan las manipulaciones realizadas para simplificar la expresión de la función objetivo del subproblema lagrangiano.

$$\begin{aligned}
& \min_{p_{ti}^s, u_{ti}^s, s_{ti}^{up}, s_{ti}^{down}} \sum_{t=1}^T \sum_{i=1}^I \{(Dur_t F_i + \bar{\mu}_{ti}^s) u_{ti}^s + (Dur_t V_i + \bar{\lambda}_{ti}^s) p_{ti}^s + \\
& (C_i^{up} + \bar{\gamma}_{ti}^{up}) s_{ti}^{up} + (C_i^{down} + \bar{\gamma}_{ti}^{down}) s_{ti}^{down}\} \\
& \sum_{i=1}^I p_{ti}^s = D_t^s \quad \forall t \\
& P_{ti}^{min} u_{ti}^s \leq p_{ti}^s \leq P_{ti}^{max} u_{ti}^s \quad \forall t \\
& L_i^{down} Dur_t \leq p_{ti}^s - p_{t-1,i}^s \leq L_i^{up} Dur_t \quad \forall t \\
& u_{ti} - u_{t-1,i} = s_{ti}^{up} - s_{ti}^{down} \quad \forall t
\end{aligned} \tag{9.36}$$

con lo que se obtiene un problema similar al problema de programación semanal. Para cada valor del multiplicador propuesto, la resolución del anterior problema proporciona el valor de la función dual. Tal como se ha comentado anteriormente, esta función dual es maximizada utilizando técnicas de optimización basadas en el subgradiente o técnicas de aproximación exterior. Para la posterior búsqueda de una solución factible (i.e., una que satisfaga las condiciones de no anticipatividad relajadas), es necesario el uso de heurísticos basados en el conocimiento del problema. Recientemente, se han utilizado técnicas basadas en ramificación y acotamiento para tal fin [18].

4 Programación semanal en mercados eléctricos

La diferencia principal en un mercado eléctrico de cara a la planificación semanal reside en que las empresas de generación son responsables de su propia producción total. La demanda de potencia, que en los modelos tradicionales aparece como un parámetro, debe ser introducida como una variable de decisión que un problema de maximización del beneficio debe optimizar. El beneficio obtenido $B(p)$ viene dado por el nivel de ingresos, que depende del precio de mercado π , y por el coste de generación de la producción total determinada $c(p)$.

$$B(p) = \pi p - c(p) \tag{9.37}$$

En un entorno de mercado, el precio viene determinado como función de la demanda. Para una empresa estratégica (no tomadora de precio), la parte de demanda que es capaz de cubrir viene dada por la curva de demanda residual⁸, $p = R(\pi)$, de modo que la función objetivo aparece como una función no lineal de la cantidad de potencia producida.

$$B(p) = R^{-1}(p)p - c(p) \tag{9.38}$$

La función de demanda residual es una función escalonada que puede ser aproximada como una función poligonal. Por otra parte, la función de ingresos es

⁸La función de demanda residual se define como el resultado de substraer de la función de demanda agregada la producción del resto de las empresas.

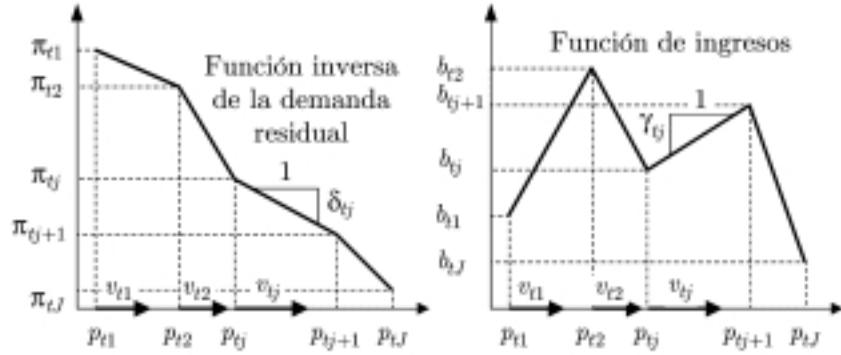


Figura 9.2: Representación poligonal de la curva de demanda residual y de la curva de ingresos

en general una función no cóncava [9] que puede ser modelada como una función poligonal (figura 9.2). Para ello se utiliza una conjunto de variables binarias que representa la cantidad producida como suma de las cantidades asociadas a los tramos individuales. El precio, dado por la función de demanda residual así como la función de ingresos siguen el mismo patrón. Este modelado de la producción, precio y curva de ingresos, utilizando dicho conjunto de variables binarias es presentado en la siguiente sección.

En caso de que supongamos conocida la curva de demanda residual, el problema de determinación de la cantidad de producción óptima es un problema determinista similar al problema de programación semanal. Por el contrario, en caso de que esta curva de demanda residual no sea conocida, su estimación a través de una distribución de probabilidad discreta conduce a la formulación de un problema de optimización que lleva a la determinación, en cada periodo, de una curva de producción dependiente del precio. Esta curva es exactamente la curva óptima de oferta en el mercado. Estas dos situaciones, la situación determinista y la situación estocástica, son descritas en las siguientes secciones.

4.1 Situación determinista: determinación de la producción óptima

Supongamos un conjunto de periodos dado por el conjunto T y una curva conocida de demanda residual para cada periodo, $t \in T$. El siguiente conjunto de índices, parámetros y variables extiende la formulación determinista del problema de la programación semanal para el caso de maximización del beneficio.

<i>Conjuntos</i>	
J	Conjunto de tramos para representar la curva de demanda residual
<i>Índices</i>	
j	Índice de tramos
<i>Parámetros</i>	
δ_{tj}	Pendiente del tramo j de la curva de demanda residual en el periodo t
δ_{tj}	[€ /MW]
π_{tj}	Pendiente del tramo j de la curva de beneficio en el periodo t
π_{tj}	[€ /MW]
p_{tj}	Precio asociado al tramo j de la curva de demanda residual en el periodo t
p_{tj}	[€ /MWh]
\bar{p}_{tj}	Cantidad asociada al tramo j de la curva de demanda residual en el periodo t
\bar{p}_{tj}	[MW]
\bar{b}_{tj}	Beneficio asociado al tramo j de la curva de beneficio en el periodo t
\bar{b}_{tj}	[€]
<i>Variables</i>	
v_{tj}	Variable binaria asociada al tramo j en el periodo t
v_{tj}	{0,1}
p_{tj}	Producción total asociada al tramo j en el periodo t
p_{tj}	[MW]
π_t	Precio en el periodo t
π_t	[€]
b_t	Beneficio en el periodo t
b_t	[€]

El problema de maximización del beneficio para un conjunto T de periodos considera una función objetivo que viene dada como

$$\max \sum_{t=1}^T b_t - c(p_t) \quad (9.39)$$

considerando que el beneficio está modelado como

$$b_t = \bar{b}_{t0} + \sum_{j=1}^{J-1} \gamma_{tj} p_{tj} \quad (9.40)$$

y que la producción viene dada como

$$p_t = p_{t0} + \sum_{j=1}^{J-1} p_{tj} \quad (9.41)$$

Las producciones totales asociadas a cada uno de los tramos j está acotada de modo natural por

$$(\bar{p}_{tj} - \bar{p}_{tj-1})v_{t+1,j} \leq p_{tj} \leq (\bar{p}_{tj} - \bar{p}_{tj-1})v_{tj} \quad (9.42)$$

imponiendo una condición de monotonía al conjunto de variables representativas de los tramos que puede ser modelada como

$$v_{tj} \geq v_{tj+1} \quad j = 1, \dots, J-1 \quad (9.43)$$

En esta formulación, la ecuación de balance de potencia en cada periodo impone que la suma de las producciones de los grupos sea la producción total. La introducción de la variable de producción total reemplaza el parámetro D_t del modelo tradicional de programación semanal (o programación multiperiodo en un caso general).

$$\sum_{i=1}^I p_{ti} = p_t \quad \forall t \quad (9.44)$$

El resto de restricciones del problema de maximización, como las restricciones de potencia máxima y mínima, ecuaciones de rampa y ecuaciones de gestión de arranques y paradas son aquéllas del problema de programación semanal (presentadas en la sección 3 de este capítulo) son introducidas en el problema de maximización del beneficio.

4.2 Situación estocástica: determinación de la curva de oferta

En caso de que la curva de demanda residual sea desconocida, o exista una cierta incertidumbre en torno a su realización, el modelado determinista de la sección anterior no es suficientemente válido. La empresa generadora debe tener en cuenta este abanico de posibilidades y su decisión de generación se transforma en una curva de producción dependiente del precio resultante en el mercado. En cada periodo, la decisión de producción de una empresa no se limita a un valor fijo de cantidad producida. Por cada realización de la curva de demanda residual (que ahora se considera como una variable aleatoria), existe una variable de decisión (producción en dicho periodo) que determina un precio óptimo para dicha decisión de producción. Este conjunto de parejas de cantidades y precios constituye la curva de oferta, (figura 9.3).

Al introducir estocasticidad en la representación de la curva de demanda residual, se deben tener en cuenta dos características de modelado. Por una parte, la función objetivo pasa a ser considerada como la maximización del beneficio esperado. El modo de interpretar la incertidumbre así como la manera de modelar la función objetivo se comentará más adelante. Por otra parte, la curva de oferta debe ser una función monótona creciente, y se debe introducir específicamente en el modelo de maximización tal cualidad.

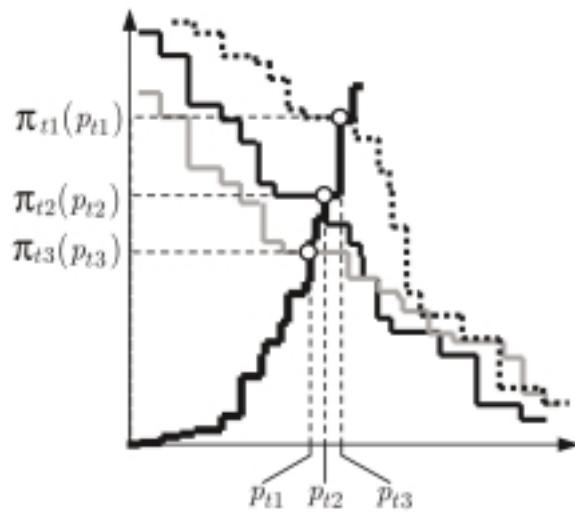


Figura 9.3: Demanda residual estocástica y curva de oferta para una hora particular

En los problemas de generación en mercado eléctricos, es posible identificar *días tipo* en función de la forma de las funciones de demanda residual de las 24 horas del día [2]. Esto es, al identificar que la curva de demanda residual para una hora particular t se comporta según un determinado patrón, la colección de demandas residuales para el resto de las horas se comporta de modo similar al patrón de día en que se encuentra incluida esa hora t . Esto lleva a considerar una representación de la incertidumbre identificando escenarios compuestos por *días tipo*.

En las figuras 9.4 y 9.5 está representada la interpretación del modelado de la incertidumbre. Un escenario o situación determinista consiste en un conjunto de funciones de demanda residual desde el primero de los períodos hasta el último. Por el contrario, en una situación estocástica son considerados una colección de escenarios que no comparten parte común en ningún momento. Debe observarse el diferente tratamiento de la incertidumbre en este problema con respecto al tratamiento de la incertidumbre dada en modo de árbol de escenarios, más adecuada para otros modelos. Esta diferencia es trasladada al modelado utilizando el subíndice k para representar cada una de las realizaciones de la incertidumbre asociada a la demanda residual (en lugar de utilizar el subíndice s).

Dada esta representación, consideremos el siguiente conjunto de índices, parámetros

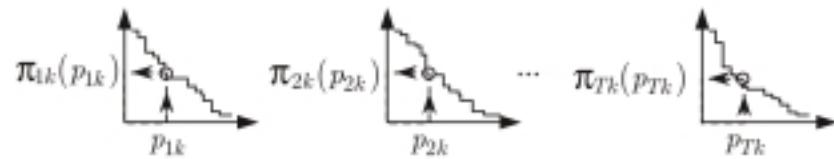


Figura 9.4: 1 escenario de curvas de demanda residual

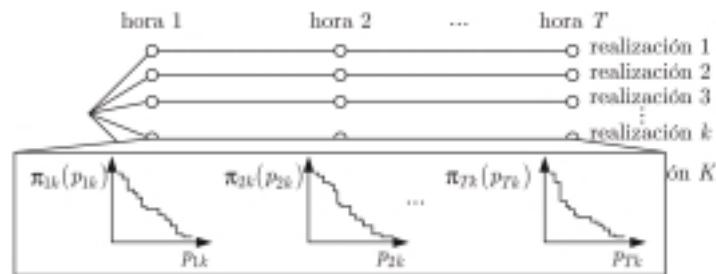


Figura 9.5: K escenarios de curvas de demanda residual

y variables que es posteriormente utilizado para la representación estocástica del problema de maximización.

Conjuntos

T	Conjunto de periodos
I	Conjunto de grupos térmicos
J	Conjunto de tramos para representar la curva de demanda residual
K	Conjunto de escenarios

Índices

t	Índice de periodos
i	Índice de grupos térmicos
j	Índice de tramos
k	Índice de escenarios

<i>Parámetros</i>	<i>deterministas</i>
Dur_t	Duración del periodo t [h]
P_i^{max}	Potencia máxima del grupo i [MW]
P_i^{min}	Potencia mínima del grupo i [MW]
L_i^{up}	Límite rampa subida del grupo i [MW/h]
L_i^{down}	Límite rampa bajada del grupo i [MW/h]
F_i	Coste fijo del grupo i [€/h]
V_i	Coste variable de combustible del grupo i [€/MWh]
C_i^{up}	Coste de arranque del grupo i [€]
C_i^{down}	Coste de parada del grupo i [€]

<i>Parámetros</i>	<i>estocásticos</i>
δ_{tj}^k	Pendiente del tramo j de la curva de demanda residual en el periodo t y escenario k [€/MW]
δ_{tj}^k	Pendiente del tramo j de la curva de beneficio en el periodo t y escenario k [€/MW]
π_{tj}^k	Precio asociado al tramo j de la curva de demanda residual en el periodo t y escenario k [€]
\bar{p}_{tj}^k	Cantidad asociada al tramo j de la curva de demanda residual en el periodo t y escenario k [MW]
\bar{b}_{tj}^k	Beneficio asociado al tramo j de la curva de beneficio en el periodo t y escenario k [€]
$Prob^k$	Probabilidad del escenario k

<i>Variables</i>	
v_{tj}^k	Variable binaria asociada al tramo j en el periodo t y escenario k {0,1}
p_t^k	Producción total en el periodo t y escenario k [MW]
p_{tj}^k	Producción total asociada al tramo j en el periodo t y escenario k [MW]
p_{ti}^k	Potencia producida por el grupo i en el periodo t y escenario k [MW]
π_t^k	Precio en el periodo t y escenario k [€]
b_t^k	Beneficio en el periodo t y escenario k [€]
u_{ti}^k	Acoplamiento del grupo i en el periodo t y escenario k {0,1}
$s_{ti}^{up\ k}$	Arranque del grupo i en el periodo t y escenario k [0,1]
$s_{ti}^{down\ k}$	Parada del grupo i en el periodo t y escenario k [0,1]
$x_t^{kk'}$	Variable binaria asociada a la monotonía de la oferta en el periodo t y escenarios k y k' {0,1}

La ecuación de balance de potencia es pues modelada como

$$\sum_{i=1}^I p_{ti}^k = p_t^k \quad \forall t, k \quad (9.45)$$

siendo modelada la cantidad total p_t^k producida en un periodo t para una realización k de la incertidumbre como

$$p_t^k = \bar{p}_{t0}^k + \sum_{j=1}^{J-1} p_{tj}^k \quad (9.46)$$

La curva de ingresos es modelada a tramos de igual modo que en la situación determinista

$$b_t^k = \bar{b}_{t0}^k + \sum_{j=1}^{J-1} \gamma_{tj}^k p_{tj}^k \quad (9.47)$$

extendiendo las cotas de las variables de la producción por tramo t a la situación estocástica

$$(\bar{p}_{tj}^k - \bar{p}_{tj-1}^k) v_{t+1,j}^k \leq p_{tj}^k \leq (\bar{p}_{tj}^k - \bar{p}_{tj-1}^k) v_{tj}^k \quad (9.48)$$

e imponiendo la condición de monotonía del conjunto de variables binarias representativas de los tramos, que ahora en la situación estocástica depende del escenario k

$$v_{tj}^k \geq v_{tj+1}^k \quad j = 1, \dots, J-1 \quad (9.49)$$

Dada la representación de la incertidumbre considerada, las restricciones de potencia máxima y mínima, las restricciones de rampa y las restricciones de gestión de arranques y paradas son introducidas de modo independiente para cada escenario k .

$$P_{ti}^{min} u_{ti}^k \leq p_{ti}^k \leq P_{ti}^{max} u_{ti}^k \quad \forall t, k \quad (9.50)$$

$$L_i^{down} Dur_t \leq p_{ti}^k - p_{t-1,i}^k \leq L_i^{up} Dur_t \quad \forall t, k \quad (9.51)$$

$$u_{ti}^k - u_{t-1,i}^k = s_{ti}^{up,k} - s_{ti}^{down,k} \quad \forall t, k \quad (9.52)$$

El conjunto de restricciones anterior forman el bloque principal de ecuaciones del modelo de determinación de la producción óptima bajo incertidumbre (curva de oferta). La curva de oferta debe ser monótona creciente y por ello el siguiente conjunto de ecuaciones es introducido en el modelo con tal propósito.

El precio obtenido al considerar la producción óptima p_t^k de un periodo t y una realización k de la incertidumbre dada por la demanda residual, es modelada mediante tramos utilizando la siguiente representación.

$$\pi_t^k = \bar{\pi}_{t0}^k + \sum_{j=1}^{J-1} \delta_{tj}^k p_{tj}^k \quad (9.53)$$

Con esta formulación, debe imponerse la condición de que el conjunto de parejas $(p_t^k, \pi_t^k), k \in K$, formen una curva monótona creciente. Esta condición, que puede ser modelada mediante una restricción no lineal, admite la siguiente formulación lineal entera mixta, mediante la introducción de un conjunto auxiliar de variables binarias [1]

$$p_t^k - p_t^{k'} \geq -x_t^{kk'} M^p \quad \forall t, k, k' \quad k' > k \quad (9.54)$$

$$\pi_t^k - \pi_t^{k'} \geq -x_t^{kk'} M^\pi \quad \forall t, k, k' \quad k' > k \quad (9.55)$$

$$p_t^k - p_t^{k'} \geq -(1 - x_t^{kk'}) M^p \quad \forall t, k, k' \quad k' > k \quad (9.56)$$

$$\pi_t^k - \pi_t^{k'} \geq -(1 - x_t^{kk'}) M^\pi \quad \forall t, k, k' \quad k' > k \quad (9.57)$$

Una vez determinado el conjunto de restricciones que dan forma a la curva de oferta y determinan el sistema generador, resta introducir la función objetivo que el problema estocástico de maximización del beneficio considera. Para el problema presentado, se considera la función objetivo que maximiza el beneficio esperado, modelada como

$$\max \sum_{t=1}^T \sum_{k=1}^K \text{Prob}^k [b_t^k - c(p_t^k)] \quad (9.58)$$

Este conjunto de restricciones modela la toma de decisión de un problema de programación multiperiodo cuando se considera incertidumbre en la curva de demanda residual. El conjunto final de decisión es una curva de oferta para cada uno de los periodos considerados en el alcance temporal. Este problema ha sido modelado como un problema de programación entera mixta de gran tamaño, pudiendo ser resuelto por optimizadores comerciales para tamaños intermedios del mismo. Por simplicidad en la exposición ha sido considerado un parque generador compuesto únicamente por grupos térmicos, si bien la formulación anterior es extensible a sistemas generadores más complejos que incluyan grupos hidráulicos así como contratos de medio y largo plazo y opciones de compra [2].

Esta sección finaliza con una breve indicación sobre la posibilidad de descomponer el problema anteriormente propuesto. Entre las posibilidades de descomposición del problema anterior figuran la Relajación Lagrangiana (mediante la relajación de la ecuación de balance de potencia) o descomposición bietapa en un problema maestro que genera la oferta y un subproblema que determina la factibilidad de la oferta propuesta dado un sistema generador existente.

Descomposición bietapa del problema estocástico

El modelado anterior presenta dos bloques bien diferenciados de restricciones. Aquéllas dirigidas a la construcción de la curva de oferta y las representativas del sistema generador. Entre el primer conjunto de restricciones figuran aquéllas de representación a tramos de la variable de producción total por periodo, de la variable de ingresos y de la variable precio. También figuran aquellas restricciones de monotonía de las variables binarias de representación de los tramos y monotonía de la curva de oferta. Un segundo grupo de ecuaciones son aquéllas que, dada una producción p_t^k por periodo t y escenario k , determinan el conjunto de grupos y niveles de operación para satisfacer ese requerimiento de potencia. Esta diferenciación entre los dos tipos de restricciones da lugar a la descomposición del problema en un problema maestro que determina y propone una curva de oferta para cada periodo t y un subproblema que evalúa la factibilidad del conjunto de soluciones propuestas.

De este modo, se considera un problema maestro dado como

$$\begin{aligned}
 & \max \sum_{t=1}^T \sum_{k=1}^K \text{Prob}^k(b_t^k) + \theta(p_t^k) \\
 p_t^k &= \bar{p}_{t0} + \sum_{j=1}^{J-1} p_{tj}^k \\
 b_t^k &= \bar{b}_{t0} + \sum_{j=1}^{J-1} \gamma_{tj}^k p_{tj}^k \\
 \pi_t^k &= \bar{\pi}_{t0}^k + \sum_{j=1}^{J-1} \delta_{tj}^k p_{tj}^k \\
 (\bar{p}_{tj}^k - \bar{p}_{tj-1}^k) v_{t+1,j}^k &\leq p_{tj}^k \leq (\bar{p}_{tj}^k - \bar{p}_{tj-1}^k) v_{tj}^k \\
 v_{tj}^k &\geq v_{tj+1}^k \quad j = 1, \dots, J-1 \\
 p_t^k - p_t^{k'} &\geq -x_t^{kk'} M^p \quad \forall t, k, k' \quad k' > k \\
 \pi_t^k - \pi_t^{k'} &\geq -x_t^{kk'} M^\pi \quad \forall t, k, k' \quad k' > k \\
 p_t^k - p_t^{k'} &\geq -(1 - x_t^{kk'}) M^p \quad \forall t, k, k' \quad k' > k \\
 \pi_t^k - \pi_t^{k'} &\geq -(1 - x_t^{kk'}) M^\pi \quad \forall t, k, k' \quad k' > k
 \end{aligned} \tag{9.59}$$

y un subproblema que evalúa el coste de esa decisión

$$\begin{aligned}
 \theta(p_t^k) = & \max - \sum_{k=1}^K \text{Prob}^k c(\sum_{i=1}^I p_{ti}^k) \\
 \sum_{i=1}^I p_{ti}^k &= p_t^k \quad \forall t, k \\
 P_{ti}^{\min} u_{ti}^k &\leq p_{ti}^k \leq P_{ti}^{\max} u_{ti}^k \quad \forall t, k \\
 L_i^{\downarrow} Dur_t &\leq p_{ti}^k - p_{t-1,i}^k \leq L_i^{\uparrow} Dur_t \quad \forall t, k \\
 u_{ti}^k - u_{t-1,i}^k &= s_{ti}^{up,k} - s_{ti}^{down,k} \quad \forall t, k
 \end{aligned} \tag{9.60}$$

La función de coste, $c(p)$, sintetiza de modo general la suma de los costes asociados a cada grupo: coste variable de combustible, coste de acoplamiento y coste de arranque. Es inmediato observar que el problema anterior es separable en K subproblemas independientes, uno por cada escenario de demanda residual considerado para representar la incertidumbre, dados como

$$\begin{aligned} \theta(p_t^k) = & \max -\text{Prob}^k c(\sum_{i=1}^I p_{ti}^k) \\ & \sum_{i=1}^I p_{ti}^k = p_t^k \quad \forall t \\ & P_{ti}^{\min} u_{ti}^k \leq p_{ti}^k \leq P_{ti}^{\max} u_{ti}^k \quad \forall t \\ & L_i^{down} Dur_t \leq p_{ti}^k - p_{t-1,i}^k \leq L_i^{up} Dur_t \quad \forall t \\ & u_{ti}^k - u_{t-1,i}^k = s_{ti}^{up,k} - s_{ti}^{down,k} \quad \forall t \end{aligned} \quad (9.61)$$

debido a que todas las restricciones que ligan los escenarios han sido incorporadas en el problema maestro.

Esta descomposición en dos etapas es característica de los algoritmos de descomposición de Benders. La diferencia esencial radica en que en este caso en el carácter entero de alguna de las variables que aparecen en el subproblema (o en los subproblemas). Bajo ciertas hipótesis de modelado es posible relajar esta condición de integralidad impuesta sobre las variables de acoplamiento de los grupos térmicos. Otra posibilidad consiste en la generalización de los algoritmos de descomposición de dos etapas (para problemas lineales) a algoritmos enteros mixtos. Se están realizando avances en este campo, como puede observarse en [6].

5 Bibliografía

- [1] A. Baíllo *et al.* “Strategic Unit Commitment for Generation in Deregulated Electricity Markets”. *The Next Generation of Electric Power Unit Commitment Models*. International Series in Operations Research and Management Science, B. F. Hobbs, M. H. Rothkopf, R. P. O’Neill and H.-P. Chao, Eds., 1st ed., Kluwer Academic Publishers, Boston, 2001, pp. 227-248.
- [2] A. Baíllo, “A methodology to develop optimal schedules and offering strategies for a generation company operating in a short-term electricity market,” Ph.D. Thesis, Department of Industrial Organization, Universidad Pontificia Comillas, Madrid, 2002.
- [3] J. A. Bloom “Solving an electricity generation capacity expansion planning problem by generalized Benders’ decomposition”. *Operations Research*, 31:84-100, 1983.
- [4] A. B. Borison, P. A. Morris, and S. S. Oren. “A state-of-the-world decomposition approach to dynamics and uncertainty in electric utility generation expansion planning”. *Operations Research*, 32:1052-1068, 1984.

- [5] Carpentier, J., Merlin, A. "Optimization methods in planning and operation". *International Journal of Electrical Power and Energy Systems*, vol.4, no.1, Jan. 1982. p. 11-18.
- [6] S. Cerisola, "Benders decomposition for mixed integer problems: application to a medium term hydrothermal coordination model," Ph.D. Thesis, Instituto de Investigación Tecnológica, Universidad Pontificia Comillas, Madrid, 2002.
- [7] G. Dantzig and G. Infanger. "Approaches to stochastic programming with application to electric power systems". In K. Frauendorfer, H. Glavitsch, and R. Bacher, editors, *Optimization in Planning and Operation of Electric Power Systems*, pages 125-138. Physica-Verlag, Heidelberg, 1993.
- [8] Delson, J.K.; Shahidehpour, S.M. "Linear programming applications to power system economics, planning and operations". *IEEE Transactions on Power Systems*, vol.7, no.3, Aug. 1992. p. 1155-63.
- [9] J. García-González, J. Barquín and J. Román. "Building supply functions under uncertainty for a day-ahead electricity market". *6th International Conference on Probabilistic Methods Applied to Power Systems, Funchal, Madeira (Portugal)*, September 25-28,2000.
- [10] D. T. Gardner and J. S. Rogers. "Planning electric power systems under demand uncertainty with different technology lead times". *Management Science*, 45:1289-1306, 1999.
- [11] Gorenstein, B.G., Campodonico, N.M., Costa, J.P., Pereira, M.V.F. "Power system expansion planning under uncertainty". *IEEE Transactions on Power Systems*, vol.8, no.1, Feb. 1993. p. 129-36.
- [12] Irving, M.R., Yong-Hua Song "Optimisation techniques for electrical power systems. 1. Mathematical optimisation methods" *Power Engineering Journal*, vol.14, no.5, Oct. 2000. p. 245-54.
- [13] Kagiannas, A. G., Askounis, D. Th., Psarras, J. "Power generation planning: a survey from monopoly to competition". *International Journal of Electrical Power and Energy Systems*, Jul 2004, Vol. 26 Issue 6, p413, 9p
- [14] F.V. Louveaux and Y. Smeers. "Optimal investments for electricity generation: A stochastic model and a test-problem". In *Numerical techniques for stochastic optimization*, Springer Ser. Comput. Math. 10, 445-453, 1988.
- [15] Majumdar, S.; Chattopadhyay, D.; "A model for integrated analysis of generation capacity expansion and financial planning" *IEEE Transactions on Power Systems* vol.14, no.2, May 1999. p. 466 - 471.
- [16] Marin, A. Salmeron, J. "Electric capacity expansion under uncertain demand: decomposition approaches" *IEEE Transactions on Power Systems* vol.13, no.2, May 1998. p. 333 - 339.

- [17] J. A. Momoh *Electric Power System Applications of Optimization* Marcel Dekker Inc. New York. 2001.
- [18] M. P. Nowak, "Stochastic Lagrangian Relaxation in Power Scheduling of a Hydro-Thermal System under Uncertainty," Ph.D. Thesis, Mathematisch-Naturwissenschaftlichen Fakultät II, Humboldt-Universität zu Berlin, Berlin, 1999.
- [19] H.D. Sherali, A.L. Soyster, F.H. Murphy, and S. Sen. "Allocation of capital costs in electric utility capacity expansion planning under uncertainty". *Management Science* 30:1-19, 1984.
- [20] Yong-Hua Song, Irving, M.R. "Optimisation techniques for electrical power systems. II. Heuristic optimisation methods". *Power Engineering Journal*, vol.15, no.3, June 2001. p. 151-60.

Algunos problemas estocásticos de localización discreta: un enfoque unificador

Maria Albareda-Sambola^a, Elena Fernández^b

^a Departamento de Estadística, Universidad Carlos III de Madrid.
e-mail: malbared@est-econ.uc3m.es

^b Departament d'Estadística i Investigació Operativa, Universitat
Politècnica de Catalunya. e-mail: e.fernandez@upc.es

1 Introducción

Los problemas discretos de localización tratan de seleccionar las ubicaciones óptimas para un conjunto de centros de servicios (plantas) entre un conjunto de ubicaciones potenciales que es conocido a priori. Desde las plantas seleccionadas deberá satisfacerse la demanda de un conjunto de clientes conocido. Por tanto, la decisión a tomar es doble y requiere identificar cuál es el conjunto de plantas que se debe abrir y cómo se debe satisfacer la demanda de los clientes.

Hay distintos factores que pueden intervenir en un problema discreto de localización y que dan lugar a una gran variedad de problemas de este tipo. Los más relevantes están relacionados con la forma en la que se dotará el servicio a los clientes y con la existencia o no de una capacidad limitada en los centros de servicio. En algunas situaciones, el servicio se realiza en las propias plantas, de manera que los clientes deben acudir a ellas. Esto ocurre cuando los centros de servicio son hospitales, escuelas, oficinas de correo, etc. En estos casos la decisión sobre el servicio a clientes da lugar a problemas de localización-asignación (LA) que se han estudiado en numerosos contextos (ver, por ejemplo, [23]). Sin embargo, en otras situaciones el servicio se realiza donde se sitúan los clientes. En estos

casos, cuando la naturaleza de los servicios admite que se realice una secuencia de ellos sin volver a la planta, la decisión sobre el servicio a los clientes da lugar a problemas combinados de localización-rutas (LR). Hay que señalar que en el caso de los problemas LR también es necesario realizar una asignación de clientes a plantas abiertas previa al propio diseño de la ruta, para saber desde qué planta se realizará la ruta que atienda a cada cliente. Tradicionalmente los problemas LR han sido menos estudiados que los problemas LA, aunque recientemente han recibido mayor atención (ver p.ej. [15, 21]).

Dada la naturaleza discreta tanto de los problemas LR como de los problemas LA normalmente éstos se estudian en el ámbito de la Programación Entera. En ambos casos a menudo se supone que los problemas son deterministas. Es decir, que todos los datos que intervienen son conocidos a priori. Sin embargo, con frecuencia estos problemas son en la práctica de naturaleza estocástica y en el momento de tomar las decisiones hay elementos de incertidumbre. Una alternativa clásica es utilizar estimaciones de los datos aleatorios y resolver el problema estocástico como en el caso determinista. Otra alternativa consiste en incorporar explícitamente la incertidumbre a los modelos dando lugar a problemas de Programación Estocástica. El aumento en la literatura de Programación Estocástica ([13, 25, 6]) ilustra el interés de tales modelos.

Dentro de los elementos aleatorios que suelen presentarse en problemas LA y LR cabe resaltar la demanda de los clientes, ya que ésta no suele ser conocida en el momento de decidir las ubicaciones de las plantas y la forma de servicio de los clientes. En estos casos podemos modelar la demanda mediante variables aleatorias y considerar los problemas en el contexto de la Programación Estocástica Entera ([27, 14, 17, 19]).

En este capítulo abordamos el estudio de algunos problemas discretos de localización en los que supondremos que las plantas en caso de abrirse tendrán una capacidad limitada conocida. Para los problemas LR supondremos además que desde cada planta abierta se realiza una única ruta para atender a los clientes asignados a la planta. Finalmente, supondremos que la demanda de cada cliente se ajusta a una variable aleatoria Bernouilli, que tiene el mismo parámetro (p) para todos los clientes. Es decir, supondremos que los clientes requieren o no servicio con una determinada probabilidad p , pero que cuando existe demanda la *cantidad* de servicio requerido es irrelevante para el problema (o, equivalentemente, que todos los clientes que solicitan servicio tienen la misma cantidad de demanda). En este contexto la capacidad de una planta representa una limitación sobre el número máximo de clientes que pueden servirse desde ella en caso de abrirse. Este tipo de demanda ha sido considerada en problemas estocásticos de rutas [5, 16] así como en problemas estocásticos de asignación generalizada [22, 1, 3]. Conocemos un único trabajo donde se haya abordado un problema estocástico LA (ELA) [24], donde se usa un problema ELA sin capacidades como

ejemplo ilustrativo y muy pocos trabajos sobre problemas estocásticos LR (ELR) (algunos ejemplos son [7, 10, 18]). Que nosotras sepamos, el único trabajo sobre SLR en el que se considera el tipo de demanda descrito anteriormente es [2].

En el tratamiento de problemas estocásticos de programación matemática existen distintas alternativas: el uso de restricciones probabilísticas, el estudio de escenarios, y la programación con recurso. En este trabajo, trataremos tanto los problemas ELA como los ELR siguiendo este último enfoque. En la programación con recurso los problemas se estructuran según un esquema binivel que distingue las decisiones que se toman antes de conocer los valores de los elementos estocásticos del problema (*solución a priori*) de aquellas que se toman *a posteriori* para adaptar la solución *a priori* al escenario que tiene lugar. En los dos casos, la solución *a priori* consistirá en un conjunto de plantas a abrir junto con una asignación de los clientes a las plantas abiertas. Una vez conocidas las demandas, si para la instancia en cuestión la demanda total asignada a alguna planta abierta excede su capacidad la solución *a priori* no será factible para esa instancia concreta y aplicaremos la acción de recurso.

Para los problemas ELA la acción de recurso consiste en reasignar algunos clientes asignados a una planta violada a otra planta abierta con capacidad disponible a un coste prefijado. Para una instancia dada puede ocurrir que la demanda total supere la capacidad total de las plantas abiertas. En ese caso parte de los clientes se perderán (no serán servidos) y se generará un coste adicional, que puede entenderse como la penalización por perder al cliente o como el coste de adquisición de recursos adicionales para poder proporcionar el servicio.

Por el contrario, la asignación de los clientes no cambia en la acción de recurso para los problemas ELR, aunque la capacidad de alguna planta esté violada por la solución *a priori*. Para definir la acción de recurso, suponemos que cuando el número de clientes asignados a una planta que solicitan servicio supera su capacidad, alguno de ellos se perderá (no será servido) generándose un coste adicional que se interpreta de forma análoga al caso de los problemas ELA. Los clientes no servidos se eligen aleatoriamente para no priorizar aquellos clientes con una posición privilegiada en la ruta *a priori*. Una vez se haya establecido el conjunto de clientes que recibirán servicio, éstos son visitados en el orden definido por las rutas a *a priori* como se muestra en la Figura 1(b).

En ambos casos el objetivo es minimizar el coste total, que se define como la suma de los costes fijos de apertura de las plantas más el coste esperado de la acción de recurso. A su vez, el coste esperado de la acción de recurso consta de dos términos: el coste esperado de la solución *a posteriori* más el coste esperado de las penalizaciones debidas a los clientes no atendidos.

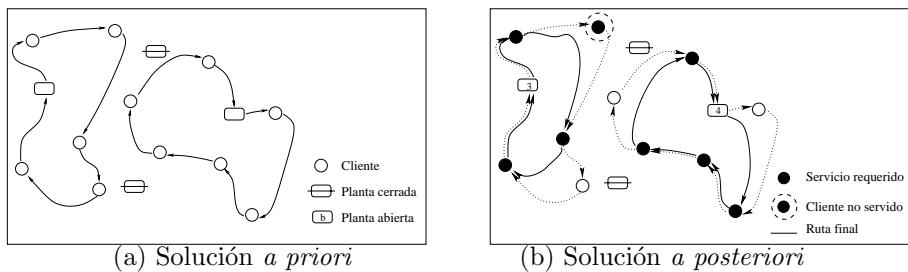


Fig 1: Acción de recurso

En la siguiente sección presentamos la notación que usaremos a lo largo del capítulo, y los elementos comunes de los modelos ELD. El modelado y las propiedades de las distintas funciones de recurso para problemas ELA y probemas ELR se tratan en las Secciones 3 y 4, respectivamente, mientras que en las Secciones 5 y 6 describimos los algoritmos que se han desarrollado para los dos tipos de problemas. Para los problemas ELA (Sección 5) describimos dos algoritmos de solución, uno heurístico y uno exacto, mientras que para problemas ELR (Sección 6) proponemos una heurística y una cota inferior. Finalizamos el capítulo con una sección de conclusiones donde, además, se presentan algunas líneas de investigación futura.

2 Modelos de recurso para Problemas Estocásticos de Localización Discreta

En esta sección formulamos un Problema Estocástico de Localización Discreta (PELD) mediante un modelo de recurso. El modelo considerado permite dar un tratamiento unificado para problemas ELA y ELR. Una instancia de un PELD viene dada por los siguientes datos:

- Los conjuntos de índices de plantas y clientes: I y J , respectivamente. Sean $m = |I|$ y $n = |J|$.
 - El coste no negativo de viaje entre pares de clientes o entre clientes y plantas: c_{ij} . Para los problemas ELA supondremos que c_{ij} representa el coste de asignación del cliente j a la planta i ;
 - El coste fijo de apertura para la planta $i \in I$: f_i ;
 - Capacidad de la planta $i \in I$ en caso de ser abierta: b_i ;
 - La probabilidad de que un cliente dado solicite servicio y su complementario: p y $q = 1 - p$, respectivamente. Suponemos que las peticiones de servicio de los clientes son independientes y todos con la misma probabilidad p ;

- El vector que representa la demanda de los clientes: $(\xi_j, j \in J)$. Puede modelarse mediante variables aleatorias Bernouilli de parámetro común p ;
- El coste de reasignación del cliente $j \in J$: $r_j \geq 0$;
- La penalización en la que se incurre cuando se rechaza la petición de servicio de un cliente: $g \geq 0$. Suponemos que esta penalización es la misma para todos los clientes.

Las soluciones *a priori* para un PELD se caracterizan por:

- Unas variables binarias $y_i, i \in I$ que representan si la planta i está o no abierta;
- Para cada $i \in I$ con $y_i = 1$:
 - $J(i) \subseteq J$, conjunto de clientes asignados a la planta i ;
 - $d_i = |J(i)|$, número de clientes asignados a la planta i ;
 - x_i , vector de incidencias que representa la asignación *a priori* para la planta i .

Dada una realización ξ del vector de demandas, para cada $i \in I$ con $y_i = 1$, sea $\eta_i = \sum_{j \in J(i)} \xi_j$, la demanda total asignada a la planta i . Puesto que las peticiones de servicio son variables aleatorias Bernouilli *i.i.d.*, tenemos que

$$\eta_i \sim \text{Binomial}(d_i, p). \quad (10.1)$$

El objetivo que se plantea en un PELD es encontrar el conjunto de plantas a abrir y la asignación de los clientes a las plantas abiertas que minimice el coste total esperado. Con la notación anterior podemos modelar la función objetivo como:

$$\sum_{i \in I} f_i y_i + \Omega(y, x),$$

donde la función de recurso $\Omega(y, x)$ está definida por

$$\Omega(y, x) = \mathbb{E}_\xi [v(y, x, \xi)] \quad (10.2)$$

siendo $\mathbb{E}_\xi [v(y, x, \xi)]$ el valor esperado de la solución *a posteriori* para la solución *a priori* (y, x) y el vector de demandas ξ .

Por tanto, un problema ELD puede formularse utilizando el siguiente modelo de recurso en dos etapas:

$$(PDE1) \text{ Min } \sum_{i \in I} f_i y_i + Q(y, x) \quad (10.3)$$

$$\sum_{i \in I} x_{ij} = 1 \quad j \in J \quad (10.4)$$

$$x_{ij} \leq y_i \quad i \in I, j \in J \quad (10.5)$$

$$x_{ij} \in \{0, 1\} \quad i \in I, j \in J \quad (10.6)$$

$$y_i \in \{0, 1\} \quad i \in I \quad (10.7)$$

Las ecuaciones (10.4) garantizan que cada cliente es asignado a una única planta, mientras que las restricciones (10.5) aseguran que ningún cliente está asignado a una planta que no esté abierta. Nótese que la solución *a priori* permite que haya más clientes asignados a una determinada planta que su capacidad. Si esto no fuese así, todas las soluciones de la primera etapa serían factibles para cualquier realización del vector de demandas con lo que nunca se realizaría la acción de recurso.

La estructura de (10.3) permite abordarlo mediante algoritmos del tipo L-shaped en los que aprovechando las buenas propiedades de la función de recurso, ésta se aproxima inferiormente de forma lineal mediante cortes de optimalidad. Este método, que está basado en la descomposición de Benders ([4]), fue inicialmente propuesto por Van Slyke y R. Wets [26] para problemas estocásticos con variables continuas y es también válido para problemas en los que únicamente las variables de la primera etapa tengan limitaciones de integridad. Sin embargo, cuando la función de recurso está definida por problemas enteros, sus buenas propiedades se pierden. A pesar de ello, en problemas como los que nos ocupan, es posible extender la función de recurso a una función convexa de forma que la aproximación inferior por una función lineal a trozos es de nuevo un enfoque viable. En esta dirección, Laporte y Louveaux ([17]) desarrollaron una nueva familia de cortes para problemas con variables binarias en el primer nivel. Las primeras aplicaciones para el caso de variables binarias tanto en la primera como en segunda etapa se deben a Wollmer [27] y Laporte y Louveaux ([17]). En particular, sea $F = \{(y, x) : \text{que satisfacen (10.4) - (10.7)}\}$ el conjunto de soluciones posibles para el modelo (PDE1) que suponemos indexadas en un con junto K . Entonces, (PDE1) es equivalente al modelo

$$(PDE2) \text{ Min } \sum_{i \in I} f_i y_i + \theta \quad (10.8)$$

$$\sum_{i \in I} x_{ij} = 1 \quad j \in J \quad (10.9)$$

$$x_{ij} \leq y_i \quad i \in I, j \in J \quad (10.10)$$

$$\theta \geq Q(y^k, x^k) + \langle \nabla Q(y^k, x^k), (y - y^k, x - x^k) \rangle \quad k \in K \quad (10.11)$$

$$x_{ij} \in \{0, 1\} \quad i \in I, j \in J \quad (10.12)$$

$$y_i \in \{0, 1\} \quad i \in I, \quad (10.13)$$

donde $\langle \cdot, \cdot \rangle$ denota el producto escalar y $\nabla Q(y^k, x^k)$ es un subgradiente de la función de recurso en el punto (y^k, x^k) .

Hay que resaltar que dado el gran número de restricciones (10.11) es inviable incorporarlas inicialmente a la formulación de (10.8). Por el contrario, los métodos L-shaped son métodos iterativos en los que en cada iteración se identifica una nueva solución *a priori* $(y^k, x^k) \in F$ a partir de la cual se genera una nueva desigualdad (10.11) que se incorpora a la formulación del problema. La eficiencia de estos métodos depende de la capacidad que se tenga de generar desigualdades (10.11) que acotan de manera eficaz el espacio de búsqueda. La expresión concreta de tales desigualdades depende de la función de recurso que se haya definido. En las próximas secciones veremos la expresión específica para los problemas ELA y ELR.

3 Función de recurso para Problemas Estocásticos de localización-asignación (ELA)

Dada una solución *a priori* $(y, x) \in F$, la acción de recurso para un problema ELA consiste en reasignar algunos clientes asignados a una planta violada a otra planta abierta con capacidad disponible a un coste prefijado, $r_j, j \in J$. Adicionalmente, para una instancia dada puede ocurrir que la demanda agregada supere la capacidad total de las plantas abiertas. En ese caso parte de los clientes se perderán (no serán servidos) y se incurrirá en una penalización g por cada cliente no atendido. Por tanto, en la función de recurso $Q^{LA}(y, x) := \xi [v^{LA}(y, x, \xi)]$, el valor de $v^{LA}(y, x, \xi)$ viene dado por:

$$v^{LA}(y, x, \xi) = \text{Min} \quad \sum_{i \in I} \sum_{j \in J} c_{ij} w_{ij} + \sum_{j \in J} r_j z_j \quad (10.14)$$

$$w_{ij} + z_j \geq \xi_j x_{ij} \quad i \in I, j \in J \quad (10.15)$$

$$\sum_{i \in I} w_{ij} \geq \xi_j \quad j \in J \quad (10.16)$$

$$\sum_{j \in J} w_{ij} \leq b_i y_i \quad i \in I \quad (10.17)$$

$$w_{ij} \in \{0, 1\} \quad i \in I, j \in J \quad (10.18)$$

$$z_j \in \{0, 1\} \quad j \in J \quad (10.19)$$

El problema de la segunda etapa (10.14)-(10.19) establece la solución *a posteriori* una vez conocidas las demandas. Las variables binarias w_{ij} ($(i, j) \in I \times J$) determinan las asignaciones finales, es decir, $w_{ij} = 1$ si el cliente j se asigna a

la planta i . Por otro lado, las variables z_j ($j \in J$) indican aquellos clientes con demanda no nula que se han reasignado. El primer grupo de restricciones (10.15) fijan z_j a 1 si el cliente j tiene demanda no nula y se reasigna a una planta diferente de la que estaba asignado. Las restricciones (10.16) imponen que todos los clientes con demanda no nula estén asignados a alguna planta abierta, mientras que las restricciones (10.17) garantizan que no se violen las capacidades de las plantas.

Observaciones

- En general, para el programa (PDE1) de un problema ELA no podrá garantizarse *recurso relativamente completo*, puesto que puede haber soluciones *a priori* $(y, x) \in F$ tales que

$$n > \sum_{i \in I} b_i y_i. \quad (10.20)$$

Sin embargo, es posible obtener recurso relativamente completo, incluyendo una planta ficticia $0 \in I$ con suficiente capacidad como para absorber cualquier exceso de demanda con un coste fijo de apertura $f_0 = 0$ y con un coste unitario de asignación que refleje la penalización por cliente no atendido g . Es decir, dado que cualquier asignación *a posteriori* a esta planta también producirá el correspondiente coste de reasignación r_j , los costes de asignación de la planta ficticia serán $c_{0j} = g - r_j$, $j \in J$, para no pagar ambas penalizaciones simultáneamente.

- Para una solución *a priori* (y, x) la evaluación de la función de recurso requiere la resolución de muchos problemas de segunda etapa, que en este caso son problemas con variables binarias. Por tanto, desde el punto de vista computacional la evaluación de la función de recurso es costosa, puesto que estos problemas no son sencillos de resolver. En este contexto resulta especialmente importante estudiar propiedades de $v^{LA}(y, x, \xi)$ que permitan su evaluación de forma más eficiente.

En este punto debemos notar que la función de recurso aquí definida para un problema ELA es similar a la función de recurso tratada en [3] para un problema estocástico de asignación generalizada (PEAG) en el que, como en el problema ELA que ahora estudiamos, las demandas de los clientes son variables aleatorias de Bernouilli, con un parámetro común p . Teniendo en cuenta que una vez que se selecciona un conjunto de plantas a abrir en un problema ELA, el subproblema de asignación correspondiente es un PEAG, podemos observar que la función de recurso $\Omega^{LA}(y, x)$ coincide con la estudiada en [3] puesto que la estructura del problema de segunda etapa $v^{LA}(y, x, \xi)$ es la misma en ambos casos. Por tanto, podemos aplicar los resultados de [3] y sustituir (10.14)-(10.19) por su relajación lineal, puesto que para todos los vectores de términos independientes correspondientes a

soluciones factibles de la primera etapa, ambos problemas tienen la misma solución.

Por tanto, para una solución a priori (y, x) el valor de $\mathcal{Q}^{LA}(y, x)$ puede obtenerse mediante la resolución de una serie de problemas continuos de segunda etapa de la forma

$$v^{LA}(y, x, \xi) = \text{Min} \quad \sum_{i \in I} \sum_{j \in J} c_{ij} w_{ij} + \sum_{j \in J} r_j z_j \quad (10.21)$$

$$w_{ij} + z_j \geq \xi_j x_{ij} \quad i \in I, j \in J \quad (10.22)$$

$$\sum_{i \in I} w_{ij} \geq \xi_j \quad j \in J \quad (10.23)$$

$$\sum_{j \in J} w_{ij} \leq b_i y_i \quad i \in I \quad (10.24)$$

$$0 \leq w_{ij} \leq 1 \quad i \in I, j \in J \quad (10.25)$$

$$z_j \geq 0 \quad j \in J \quad (10.26)$$

Proposición 4 Sea S el conjunto de índices de realizaciones para la demanda. Para $s \in S$, sean ξ^s el correspondiente vector de demandas y p^s la probabilidad de tal realización de forma que

$$\mathcal{Q}^{LA}(y, x) = \sum_{s \in S} p^s v^{LA}(y, x, \xi^s), \quad (y, x) \in \mathbb{R}^m \times \mathbb{R}^{m \cdot n}. \quad (10.27)$$

Sea $\lambda(y, x, \xi^s)$ un vector de variables duales asociadas a las restricciones (10.24) y a las restricciones (10.22) para el vector (y, x, ξ^s) . Entonces $u(y, x)$,

$$u(y, x) = \sum_{s \in S} p^s \lambda(y, x, \xi^s) \text{diag}(b, \bar{\xi}_1^s, \dots, \bar{\xi}_n^s), \quad (10.28)$$

es un subgradiente de \mathcal{Q} en (y, x) . Aquí, $\bar{\xi}_j^s$ es un vector de m componentes, todas iguales a la demanda del cliente j en la realización s y b es el vector de capacidades.

4 Función de recurso para Problemas Estocásticos de localización-rutas (ELR)

Para los problemas ELR una solución *a priori* $(y, x) \in F$ define no sólo un conjunto de plantas a abrir y una asignación de clientes a plantas abiertas, sino también un conjunto de rutas desde las plantas abiertas que visitan cada cliente exactamente una vez. Estas rutas están bien definidas puesto que son las rutas de coste esperado mínimo que visitan todos los clientes asignados a una determinada

planta y pueden identificarse resolviendo para cada una de las plantas abiertas un Problema Probabilista del Viajante de Comercio (PPVC) (ver, p.ej. [11]) sobre el conjunto de nodos $V^i = \{i\} \cup J(i)$. Por tanto, para cada planta abierta en la solución *a priori*, su *ruta a priori* es la solución del correspondiente PPVC. Sea z^i el vector de incidencias de la ruta *a priori* para una planta i .

En la acción de recurso que consideramos para los problemas ELR la asignación de los clientes no cambia aunque la capacidad de alguna planta esté violada por la solución *a priori* (haya más clientes con demanda asignados a una determinada planta que su capacidad). Para definir la acción de recurso, suponemos que cuando el número de clientes asignados a una planta que solicitan servicio supera su capacidad, alguno de ellos se perderá (no será servido) generándose un coste adicional g por cliente no atendido. Los clientes no servidos se eligen aleatoriamente para no priorizar aquellos clientes con una posición privilegiada en la ruta *a priori*. Una vez establecido el conjunto de clientes que recibirán servicio, éstos son visitados en el orden definido por las rutas a *a priori* z^i , como se muestra en la Figura 1(b).

Para una solución *a priori* $(y, x) \in F$, y una realización del vector de demandas ξ , sea $J'(i) \subseteq J(i)$ el conjunto de clientes atendidos en la ruta *a posteriori* de la planta i . Podemos distinguir dos casos: si $\eta_i \leq b_i$, entonces $J'(i) = \{j \in J(i) : \xi_j = 1\}$; en otro caso, si $\eta_i > b_i$, entonces $J'(i)$ es un subconjunto aleatorio de $\{j \in J(i) : \xi_j = 1\}$, de cardinalidad b_i . Sea \hat{z}^i , el vector de incidencias de los arcos de la ruta obtenida a partir de z^i pero que sólo visita los clientes que están en $J'(i)$. Utilizando la notación anterior, la función de recurso $\mathcal{Q}^{LR}(y, x)$ para un problema ELR puede expresarse como:

$$\mathcal{Q}^{LR}(y, x) = \mathbb{E}_\xi \left[\sum_{i | y_i=1} g \cdot (\eta_i - b_i)^+ + \mathbb{E}_{J'(i)} \left[\sum_{i | y_i=1} c \hat{z}^i \right] \right], \quad (10.29)$$

con $(\cdot)^+ = \max\{\cdot, 0\}$. Puesto que la función esperanza es lineal, (10.29) puede expresarse como

$$\mathcal{Q}^{LR}(y, x) = \mathcal{S}(y, x) + \sum_{i | y_i=1} \mathcal{T}^i(y, x), \quad (10.30)$$

donde $\mathcal{S}(y, x) = \mathbb{E}_\xi \left[\sum_{i | y_i=1} g \cdot (\eta_i - b_i)^+ \right]$ es la penalización esperada y $\mathcal{T}^i(y, x) = \mathbb{E}_\xi [\mathbb{E}_{J'(i)} [c \hat{z}^i]]$ es el coste esperado de la ruta asociada a la planta $i \in I$.

Hay que notar que $\mathcal{S}(y, x)$ tiene la estructura de una función de recurso asociada a un problema estocástico con recurso simple, que ha sido ampliamente tratada en la literatura (ver p.ej. [6]). Por el contrario, las funciones $\mathcal{T}^i(y, x)$ no tienen una estructura sencilla. En [2] se obtiene la siguiente expresión exacta para $\mathcal{T}^i(y, x)$:

$$\mathcal{T}^i(y, x) = \sum_{h=1}^{d_i} R_{h-1} c_{ij_h} + \sum_{h=1}^{d_i-1} \sum_{t=h+1}^{d_i} P_{t-h-1} c_{j_h j_t} + \sum_{h=1}^{d_i} R_{d_i-h} c_{j_h i}. \quad (10.31)$$

donde

$$P_\ell = p^2 \left[\sum_{k=2}^{k_1} \binom{d_i - \ell - 2}{k-2} p^{k-2} q^{d_i-k} + \sum_{k=b_i+1}^{d_i} p^{k-2} q^{d_i-k} \sum_{t=t_0}^{t_1} \binom{\ell}{t} \binom{d_i - \ell - 2}{k-t-2} \binom{k-t-2}{b_i-2} / \binom{k}{b_i} \right], \quad (10.32)$$

y

$$R_\ell = p \left[\sum_{k=1}^{k_1} \binom{d_i - \ell - 1}{k-1} p^{k-1} q^{d_i-k} + \sum_{k=b_i+1}^{d_i} p^{k-1} q^{d_i-k} \sum_{t=t_0}^{t_1} \binom{\ell}{t} \binom{d_i - \ell - 1}{k-t-1} \binom{k-t-1}{b_i-1} / \binom{k}{b_i} \right]. \quad (10.33)$$

Tanto en (10.32) como en (10.33), $k_1 = \min\{b_i, d_i - \ell\}$, $t_0 = \max\{k - d_i + \ell, 0\}$, y $t_1 = \min\{k - b_i, \ell\}$.

5 Algoritmos para ELA

En esta sección presentamos dos algoritmos para resolver problemas ELA. El primero de ellos es una heurística, mientras que el segundo es un método exacto.

5.1 Heurística para ELA

La heurística que presentamos a continuación está basada en una de las heurísticas propuestas en [1] para PEAG. Resolveremos dos problemas auxiliares deterministas cuya solución nos proporcione una aproximación de (PDE1). Inicialmente, para elegir un conjunto adecuado de plantas para abrir resolvemos el siguiente problema de la mochila:

$$(PM) \quad \text{Min} \quad \sum_{i \in I} f_i y_i \quad (10.34)$$

$$\sum_{i \in I} b_i y_i \geq (1 + \alpha)pn \quad (10.35)$$

$$y_i \in \{0, 1\} \quad i \in I \quad (10.36)$$

En las restricciones (10.35), $\alpha \in [0, \frac{q}{p}]$ es un parámetro prefijado. Por tanto, las restricciones (10.35) garantizan que el conjunto de plantas seleccionado tiene

una capacidad total mayor que un valor que oscila entre la demanda esperada ($\alpha = 0$) y la demanda conjunta de todos los clientes ($\alpha = \frac{q}{p}$).

Sea $I^* = \{i \in I : y_i = 1\}$ en la solución óptima de (10.34) – (10.36) el conjunto de plantas a abrir. Ahora obtenemos una asignación *a priori* resolviendo el siguiente problema de transporte restringido al conjunto de plantas I^* :

$$(PT) \quad \text{Min} \quad \sum_{i \in I^*} \sum_{j \in J} c_{ij} x_{ij} \quad (10.37)$$

$$\sum_{i \in I^*} x_{ix} \geq 1 \quad j \in J \quad (10.38)$$

$$\sum_{j \in J} x_{ix} \leq \tilde{b}_i \quad i \in I^* \quad (10.39)$$

$$x_{ij} \in \{0, 1\} \quad i \in I^*, j \in J \quad (10.40)$$

donde $\tilde{b}_i = \left\lceil b_i n \left(\sum_{i \in I} b_i \right)^{-1} \right\rceil$. Es decir, escalamos las capacidades de las plantas abiertas de forma que en el problema auxiliar pueda satisfacerse la demanda de todos los clientes.

Cabe resaltar que dada la estructura de (PT), éste puede resolverse de forma óptima como un problema continuo, eliminando las condiciones de integridad sobre las variables.

5.2 Algoritmo exacto para ELA

El algoritmo que describimos a continuación es del tipo *branch-and-cut*. Utiliza dos tipos de planos secantes que son cortes de optimalidad y que se basan, respectivamente, en la convexidad de la función de recurso Q^{LA} y en el hecho que las variables son binarias.

Cortes de optimalidad

La primera familia de cortes que utilizamos son los cortes de optimalidad (10.11) definidos en la Sección 2, que se utilizan en los algoritmos del tipo L-shaped [26] para problemas continuos con recurso en dos etapas. En nuestro caso, podemos utilizarlos considerando la aproximación convexa de la función de recurso estudiada en la Sección 3. Para obtenerlos basta con aplicar la Proposición 4. Nos referiremos a estos cortes como de *∂ -optimalidad* para distinguirlos de los de L-L-optimalidad que describimos a continuación.

Los cortes de L-L-optimalidad fueron introducidos por Laporte and Louveaux [17] y son válidos para todas las soluciones binarias de primera etapa, para variables de segunda etapa generales. Dado un vector binario (\bar{y}, \bar{x}) , la estructura de estos cortes en nuestro caso es

$$\theta \geq (\mathcal{Q}^{LA}((\bar{y}, \bar{x})) - L) \left(\sum_{\bar{y}_i=1} y_i + \sum_{\bar{x}_{ij}=1} x_{ij} - \sum_{\bar{y}_i=0} y_i - \sum_{\bar{x}_{ij}=0} x_{ij} \right) - (\mathcal{Q}^{LA}(\bar{y}, \bar{x}) - L)(\sum_i \bar{y}_i + \sum_{i,j} \bar{x}_{ij} - 1) + L, \quad (10.41)$$

donde L es una cota inferior del valor de \mathcal{Q}^{LA} . Este corte se cumple como igualdad en (\bar{y}, \bar{x}) , mientras que está dominado por $\mathcal{Q}^{LA}(y, x) \geq L$ para otras soluciones binarias.

Algoritmo de *branch-and-cut* (BC)

El algoritmo trata la integridad de las variables de la primera etapa mediante el esquema enumerativo y aproxima la función de recurso incorporando cortes de optimalidad de forma sucesiva. Por tanto, dado un *problema actual* definido por una serie de cortes de optimalidad y un subconjunto de variables de decisión fijadas a 0 ó a 1, el algoritmo BC procede *i*) resolviendo la relajación lineal del problema actual, *ii*) añadiendo cuando los haya cortes de optimalidad violados, y *iii*) ramificando cuando la solución actual sea no entera.

Dada una colección de cortes de optimalidad $k = 1, \dots, K$, y dos pares de subconjuntos disjuntos de $I \times I$ e $I \times J$, respectivamente, $Y = (Y^0, Y^1)$, y $X = (X^0, X^1)$ definimos el problema $P_{K,Y,X}$ como:

$$(P_{K,Y,X}) \text{Min} \quad \theta \quad (10.42)$$

$$\sum_{i \in I} x_{ij} = 1 \quad i \in I \quad (10.43)$$

$$\alpha^k + \beta^k \cdot x \leq \theta \quad k = 1, \dots, K \quad (10.44)$$

$$y_i = 0 \quad i \in Y^0, \quad (10.45)$$

$$y_i = 1 \quad i \in Y^1, \quad (10.46)$$

$$x_{ij} = 0 \quad (i, j) \in X^0, \quad (10.47)$$

$$x_{ij} = 1 \quad (i, j) \in X^1, \quad (10.48)$$

$$y_i \in [0, 1], x_{ij} \in [0, 1] \quad \text{en otro caso.} \quad (10.49)$$

El algoritmo BC es esencialmente una combinación de técnicas estándar de *branch-and-bound* y planos secantes. Sin embargo aprovecha la propiedades específicas de ELA ya que utiliza la aproximación convexa de la función de recurso Q^{LA} de la Sección 3 que permite tratar ELA como un problema binario con función objetivo convexa. Esto permite generar cortes de ∂ -optimalidad en las soluciones fraccionales de cada problema actual, que pueden utilizarse en un algoritmo estándar de *branch-and-bound*. Hay que resaltar que tales cortes son

en general muy eficientes para obtener una mejor aproximación de Q^{LA} en un entorno de la solución óptima *fraccional* del problema relajado, pero que esto no suele ser cierto para la solución óptima binaria del problema actual. Para

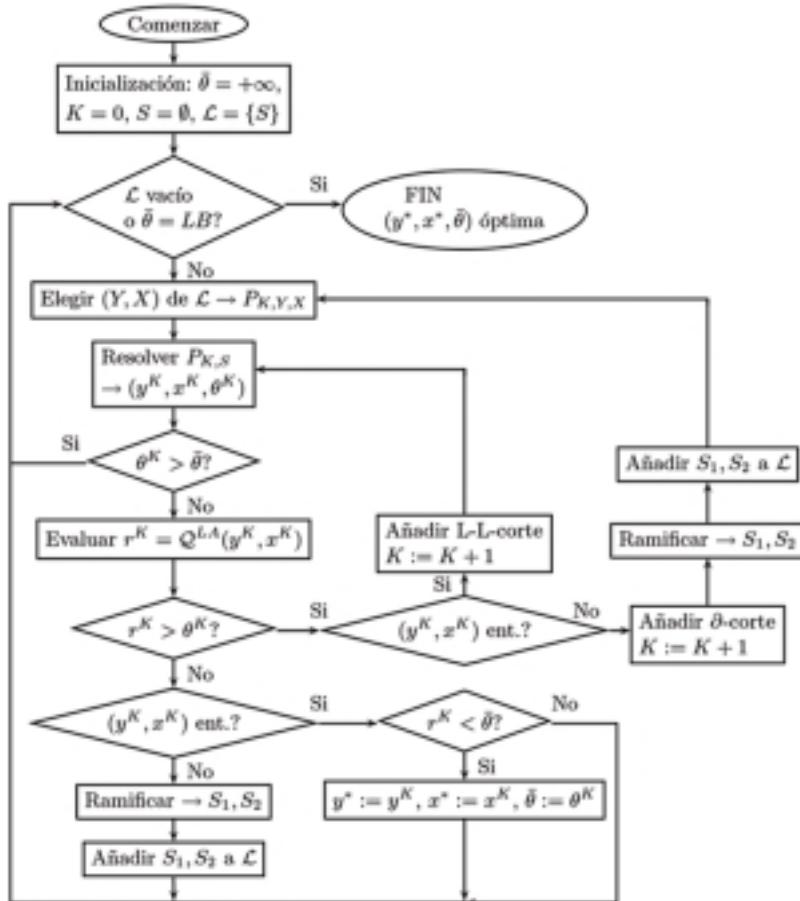


Fig 2: Algoritmo BC.

reducir este inconveniente, cuando la solución del problema actual sea binaria nuestro algoritmo BC utiliza la interpretación alternativa de ELA como un problema con recurso en dos etapas con variables binarias en la primera etapa y siguiendo el algoritmo entero L-shaped de [17] diseñado para esta clase de problemas, generamos un corte de L-L-optimalidad (en vez de generar un corte de

∂ -optimalidad). En definitiva, el algoritmo BC puede verse como una variante del algoritmo entero L-shaped, del que también se utilizan las reglas de ramificación y eliminación, combinado con una cota inferior y cortes de ∂ -optimalidad que se generan en soluciones fraccionales de los problemas actuales. Los detalles del algoritmo se presentan en la Figura 2. El algoritmo selecciona el problema actual de la lista de subproblemas pendientes de evaluación (que se denota \mathcal{L}), con una política *last-in-first-out*. Puesto que los cortes de optimalidad $k = 1, \dots, K$ son válidos para todos los subproblemas, basta con indicar los conjuntos de índices $S = (Y, X)$, con $Y = (Y^0, Y^1)$ y $X = (X^0, X^1)$ para identificar los subproblemas de la lista $P_{K,S}$.

El algoritmo BC converge puesto que se trata de una modificación del algoritmo entero L-shaped [17], que encuentra una solución óptima en un número finito de iteraciones, en la que se añaden cortes violados de ∂ -optimalidad en la soluciones fraccionales de los problemas actuales, de los que hay un número finito.

6 Algoritmos para ELR

En esta sección describimos una heurística para obtener soluciones posibles para ELR así como una cota inferior válida, que permite contrastar la calidad de las soluciones obtenidas con la heurística, que han sido propuestas en [2]. La heurística utiliza una aproximación de la función de recurso Q^{LR} definida en la Sección 4. Esto se debe a que la evaluación de la función de recurso Q^{LR} cuya expresión analítica viene dada por (10.30) es muy costosa puesto que requiere la evaluación de la expresión (10.31) para conocer el valor de la función \mathcal{T}^i lo cual, a su vez, requiere el cálculo de las probabilidades (10.32) y (10.33), que implican el cálculo de números combinatorios. Por tanto, utilizaremos una aproximación de la función \mathcal{T}^i que puede evaluarse de forma mucho más rápida. Existen algunos algoritmos para problemas estocásticos que han utilizado de forma satisfactoria aproximaciones de la función objetivo, cuando su evaluación requiere un esfuerzo de cálculo importante (ver p.ej. [9]). En nuestro caso, la dificultad de la evaluación del coste de la función de recurso se debe en gran medida al hecho de que las visitas a distintos clientes en la misma ruta *a priori* no son independientes debido a la restricción de capacidad de la planta. Para superar esta dificultad construimos una aproximación $\tilde{\mathcal{T}}^i$ en la que las probabilidades (10.31) y (10.33) se sustituyen por las que resultarían si los servicios a los clientes fuesen independientes. De hecho este es el caso cuando $d_i \leq b_i$. Sea \bar{p} la probabilidad de visitar un cliente dado, calculada como p veces la probabilidad condicional de visitar el cliente dado que tiene demanda. Si en la solución *a priori* el cliente está asignado a planta i , esta probabilidad condicional puede expresarse como (ver [2]):

$$\bar{p} = p \left(\sum_{k=1}^{b_i} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} + \sum_{k=b_i+1}^{d_i} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} b_i/k \right).$$

Sustituyendo en (10.31) las probabilidades P_ℓ y R_ℓ por

$$\tilde{P}_\ell = \bar{p}^2 (1-\bar{p})^\ell \quad \text{and} \quad \tilde{R}_\ell = \bar{p} (1-\bar{p})^\ell$$

obtenemos la siguiente aproximación de \mathcal{T}^i :

$$\begin{aligned} \tilde{\mathcal{T}}^i(y, x) = & \sum_{k=1}^{d_i} \bar{p} (1-\bar{p})^{k-1} c_{ij_k} + \sum_{k=1}^{d_i-1} \sum_{t=k+1}^{d_i} \bar{p}^2 (1-\bar{p})^{t-k-1} c_{j_k j_t} \\ & + \sum_{k=1}^{d_i} \bar{p} (1-\bar{p})^{d_i-k} c_{j_k i}. \end{aligned}$$

Esta expresión coincide con la función objetivo propuesta en [12] para el PPVC en el caso particular en que la probabilidad de presencia de todos los clientes sea la misma. En adelante nos referimos a esta aproximación de la función de recurso como $\tilde{\mathcal{Q}}^{LR}(y, x)$.

6.1 Heurística para ELR

En [2] se propone una heurística en dos fases para ELR. En la fase constructiva se obtiene una solución inicial que se mejora de forma iterativa en la fase de búsqueda local.

Fase constructiva

Para obtener la solución inicial, se descompone el problema ELR en una secuencia de tres subproblemas: 1) selección del conjunto de plantas a abrir, 2) asignación de clientes a plantas abiertas, y 3) diseño de una ruta para cada planta abierta.

Selección del conjunto de plantas abiertas.

El conjunto de plantas abiertas O se elige de forma que la capacidad total sea suficiente como para atender todas las solicitudes de demanda con una determinada probabilidad α , es decir:

$$\mathbb{P} \left[\sum_{j \in J} \xi_j \leq \sum_{i \in O} b_i \right] \geq \alpha. \quad (10.50)$$

La menor capacidad conjunta b_α que satisface (10.50) puede obtenerse fácilmente de la distribución de probabilidad binomial del número total de solicitudes de servicio. Una vez calculado b_α , el conjunto O se determina resolviendo el siguiente Problema de la Mochila:

$$(PM_\alpha) \quad \min \sum_{i \in I} f_i y_i \quad (10.51)$$

$$\sum_{i \in I} b_i y_i \geq b_\alpha \quad (10.52)$$

$$y_i \in \{0, 1\} \quad \forall i \in I. \quad (10.53)$$

Cabe observar que, por construcción de b_α , las restricciones (10.50) y (10.52) son equivalentes. Por tanto, el conjunto $O = \{i \in I : y_i = 1 \text{ en la solución de } PM_\alpha\}$ tiene coste mínimo entre los que satisfacen (10.50).

Asignación de clientes a plantas abiertas

Una vez que el conjunto de plantas abiertas está fijo, se identifica una asignación de clientes a las plantas abiertas. Para ello, resolvemos un PEAG en el que la función de coste viene dada por las distancias entre las plantas y los clientes. Como en el caso de la heurística para ELA, en esta fase podemos utilizar una de las heurísticas de [1] y resolvemos un problema de transporte similar a (10.37)-(10.40). Ahora, debemos tener en cuenta que la función que deseamos optimizar es el valor de la función de recurso $Q^{LR}(y, x)$ (10.30) y que en ELR no permitimos reasignaciones de clientes. Por tanto, en la función objetivo del problema de transporte, establecemos todos los costes de reasignación a un valor superior al de la penalización por no atenderlos cuando tienen demanda.

Construcción del conjunto de rutas a priori

Comenzando con las rutas vacías, los clientes se insertan sucesivamente en la ruta de la planta a la que ha sido asignado, utilizando el criterio de *inserción del más próximo* respecto a la función objetivo del problema determinista. Una vez que cada cliente se ha insertado en su correspondiente ruta, se realizan intercambios sucesivos de dos arcos en la misma ruta, hasta que todas las rutas sean *2-opt* respecto a la función objetivo \tilde{T}^i .

Fase de mejora

Una vez obtenida una solución inicial, ésta se mejora utilizando búsqueda local. Todos los movimientos se realizan utilizando el criterio de *inserción del más próximo* respecto a \tilde{T} . Se exploran cuatro entornos:

- $N_1(x)$: Reasignación de un cliente. $N_1(x)$ contiene todas las soluciones que pueden obtenerse a partir de x quitando un cliente de su ruta e insertándolo en otra diferente.

- $N_2(x)$: *Intercambio de dos clientes.* $N_2(x)$ contiene todas las soluciones que pueden obtenerse a partir de x seleccionando un par de clientes j_1 y j_2 que estén en rutas diferentes i_1 e i_2 , respectivamente, eliminándolos de sus correspondientes rutas, e insertando j_1 en la ruta de la planta i_2 y j_2 en la ruta de la planta i_1 .
- $N_3(x)$: *Intercambio de plantas.* Las soluciones de $N_3(x)$ son las que pueden obtenerse a partir de x cerrando una planta abierta i_1 , abriendo una planta cerrada i_2 , y asignando la ruta de i_1 a i_2 .
- $N_{2\text{-opt}}(x)$: *Intercambio de arcos.* Las soluciones de $N_{2\text{-opt}}(x)$ se obtienen a partir de x intercambiando pares de arcos en la misma ruta.

La Figura 3 muestra un esquema de la heurística, en el que se indica el orden de exploración de los distintos entornos.

Heurística para ELR

Fase constructiva

Seleccionar α , determinar b_α y resolver $(KP_\alpha) \rightarrow O$.

Asignar clientes a O utilizando heurística $\rightarrow \{J(i) : i \in O\}$.

Para ($i \in O$) hacer

Para ($j \in J(i)$) Insertar j en la ruta de la planta i .

Explorar $N_{2\text{-opt}}(x)$.

FinPara

Fase de mejora

CriterioTerminación \leftarrow falso.

Mientras(*CriterioTerminación* falso) hacer

Repetir

Explorar $N_2(x)$ y actualizar x .

Explorar $N_1(x)$ y actualizar x .

Hasta que (no se encuentren movimientos de mejora)

Repetir

Explorar $N_{2\text{-opt}}(x)$ y actualizar x .

Hasta que (no se encuentren movimientos de mejora)

Explorar $N_3(x)$ y actualizar x .

Actualizar *CriterioTerminación*.

FinMientras

Fig. 3. Esquema de la heurística para ELR

6.2 Cota inferior para ELR

Para obtener cotas inferiores válidas para ELR podemos descomponer la función objetivo en dos partes y obtener una cota separada para cada una de ellas. Estas dos partes son 1) los costes asociados con las plantas (costes fijos de apertura más penalizaciones), y 2) los costes de las propias rutas.

Costes fijos de apertura de las plantas más penalizaciones esperadas por clientes no servidos: z_{SKP}

Es importante acotar conjuntamente la penalización esperada más los costes fijos de apertura puesto que estos dos términos son contradictorios entre ellos. Abrir una planta aumenta los costes fijos pero disminuye las penalizaciones. La penalización esperada asociada a una solución con un conjunto de plantas abiertas O , no será menor que

$$\mathbb{E}_\xi \left[g \cdot \left(\sum_{j \in J} \xi_j - \sum_{i \in O} b_i \right)^+ \right],$$

independientemente de la asignación de los clientes a las plantas. Por tanto, una cota inferior de la suma de los costes fijos más las penalizaciones esperadas viene dada por la solución del siguiente Problema Estocástico de la Mochila.

Este problema es un problema en dos etapas con recurso simple. Puede verse que un modelo determinista equivalente para este problema es

$$\begin{aligned} \min_i \quad & \sum_{i \in I} f_i y_i + \sum_{s=0}^{n+1} \delta^s u^s \\ & \sum_{i \in I} b_i y_i - \sum_{s=0}^{n+1} u^s = np \\ & u^0 \leq 1 - np \\ & 0 \leq u^{l+1} \leq 1 \quad l = 0, \dots, n \\ & y_i \in \{0, 1\} \quad i \in I, \end{aligned}$$

(OJO) donde $\delta^0 = -g$, y $\delta^{l+1} = g \cdot \left(-1 + \mathbb{P} \left[\sum_{j \in J} \xi_j \leq l \right] \right)$, $l = 0, \dots, n$.

Se trata de un problema de programación entera mixta que puede resolverse de forma eficiente utilizando cualquier software de propósito general.

Coste esperado de las rutas *a posteriori*

Para calcular una cota inferior del coste de las rutas *a posteriori* utilizamos un procedimiento similar al propuesto en [16] para el PPVC. Es decir, calculamos una cota inferior del coste \bar{z}_{RS} de las rutas *a priori* y restamos una cota superior de los ahorros esperados por *saltarse* los clientes.

Cota inferior \underline{z}_{PVC} , del coste de una ruta *a priori*

Para acotar el coste de una ruta *a priori* utilizamos la solución de un PVC definido como sigue. Consideramos un grafo dirigido completo con conjunto de vértices $V = I \cup I' \cup J$, donde I' es una copia del conjunto de plantas. El vértice i representa la planta como punto inicial de la ruta, e i' representa la misma planta, como punto final de la ruta. Definimos los siguientes costes para los arcos:

$$\tilde{c}_e = \begin{cases} c_e & \text{si } e \in J \times J, e \in I \times J \text{ o } e \in J \times I' \\ 0 & \text{si } e \in I' \times I \text{ o } e = (i, i'), i \in I \\ \infty & \text{en otro caso.} \end{cases}$$

Con estos costes, en la solución óptima de PVC los vértices que representan las plantas estarán conectados solamente en dos casos: 1) un arco de un planta-terminal a un nodo planta-inicial, que se corresponde con una ruta que termina en una planta seguida de otra ruta que comienza en la otra planta, y 2) un arco de un nodo planta-inicial a su correspondiente nodo planta-terminal, que representa una ruta vacía. La Figura ?? ilustra cómo puede obtenerse una solución factible para este PVC a partir de cualquier conjunto de rutas *a priori* para ELR.

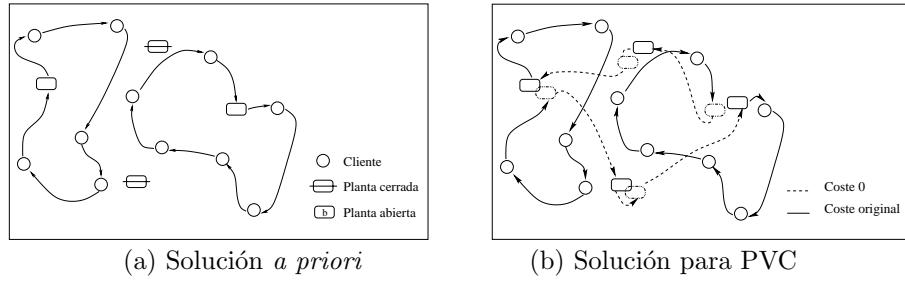


Fig 4: Cota del coste de la ruta *a priori*, \underline{z}_{PVC}

Para resolver este PVC puede utilizarse el modelo propuesto en [20] y reforzado posteriormente en [8]:

$$\underline{z}_{PVC} = \min \sum_{e \in A} c_e x_e \quad (10.54)$$

$$\sum_{e \in \delta^-(j)} x_e = 1 \quad i \in V \quad (10.55)$$

$$\sum_{e \in \delta^+(j)} x_e = 1 \quad i \in V \quad (10.56)$$

$$u_i - u_j + (\bar{n} - 1)x_{ij} + (\bar{n} - 3)x_{ji} \leq \bar{n} - 2 \quad \begin{matrix} i, j > 1 \\ i \neq j \end{matrix} \quad (10.57)$$

$$x_e \in \{0, 1\} \quad e \in A \quad (10.58)$$

$$1 \leq u_i \leq \bar{n} - 1 \quad i > 1, \quad (10.59)$$

donde $\bar{n} = |V| = 2m + n$, A es el conjunto de arcos del grafo y, dado un vértice j , $\delta^-(j)$ (resp. $\delta^+(j)$) es el conjunto de arcos que entran en (resp. salen de) j . Las variables x_e indican si el arco e se utiliza en la solución, y cada variable u_i da la posición del vértice i en el circuito. En [2] se mejora el valor de la cota inferior \underline{z}_{PCV} , reforzando el modelo anterior (10.54)-(10.59) mediante la incorporación de algunas restricciones adicionales.

Cota superior \bar{z}_{RS} , de los ahorros esperados por “saltar” clientes

Para acotar los costes esperados de las rutas *a posteriori* en [2] se utiliza una extensión del modelo propuesto en [16] para el PTSP. La idea del modelo es encontrar una familia adecuada de atajos con un valor total máximo. Un atajo representa la sustitución de un camino (i, k, j) en la solución *a priori* por el arco (i, j) (ver Figura 1(b)). En nuestro caso, para cada cliente k se definen conjuntos de tripletas para todos los posibles atajos como sigue:

- 1) $A(k) = \{(i, k, i) : i \in I\};$
- 2) $B(k) = \{(i, k, j) : i \in I, j \in J, j \neq k\};$
- 3) $C(k) = \{(j_1, k, j_2) : j_1, j_2 \in J, j_1, j_2 \neq k, j_1 \neq j_2\},$ y
- 4) $D(k) = A(k) \cup B(k) \cup C(k).$

Para cada tripleta (i, k, j) se define una variable y_{ikj} . Adicionalmente se incluyen variables binarias w_i , $i \in I$, que indican si la planta i está o no abierta. Los coeficientes de la función objetivo se corresponden con los valores de los atajos. Dada una tripleta (i, k, j) , el correspondiente coeficiente viene dado por $d_{ikj} = q(c_{ik} + c_{kj} - c_{ij})$, donde los índices i, j , se refieren a clientes o plantas. Nótese que si $i = j \in I$ (variable de tipo A), entonces el último término c_{ij} es cero. Con esta notación el modelo resultante es:

$$\max \sum_{k \in J} \sum_{(i, k, j) \in D(k)} d_{ikj} y_{ikj} \quad (10.60)$$

$$\text{s. a.} \quad \sum_{(i,k,j) \in D(k)} y_{ikj} = 1, \quad k \in J \quad (10.61)$$

$$\sum_{k \in J} y_{iki} \leq 1, \quad i \in I \quad (10.62)$$

$$\sum_{k \in J \setminus \{j\}} y_{ikj} \leq 2i \in I, j \in J \quad (10.63)$$

$$\sum_{k \in J \setminus \{j_1, j_2\}} y_{j_1 k j_2} \leq 1 \quad j_1 \neq j_2 \in J \quad (10.64)$$

$$\sum_{k \in J \setminus \{j\}} \sum_{i \in I} y_{ikj} + \sum_{\substack{k, j_2 \in J \setminus \{j\} \\ k \neq j_2}} y_{jkj_2} \leq 2 \quad j \in J \quad (10.65)$$

$$2 \sum_{k \in J} y_{iki} + \sum_{k, j \in J} y_{ikj} \leq 2w_i \quad i \in I \quad (10.66)$$

$$\sum_{i \in I} f_i w_i \leq npg \quad (10.67)$$

$$2 \sum_{i \in I} y_{iji} + \sum_{i \in I} \sum_{\substack{k \in J \\ k \neq j}} y_{ikj} + \sum_{\substack{k, j_2 \in J \\ j \neq k \neq j_2 \neq j}} y_{jkj_2} \leq 2 \quad j \in J \quad (10.68)$$

$$y_{ikj}, w_i \in \{0, 1\}, \quad k \in J, (i, k, j) \in D(k), i \in I. \quad (10.69)$$

Las restricciones (10.61) garantizan que para cada cliente se elige exactamente un atajo y las restricciones (10.62)-(10.65) sirven para controlar la utilización de cualquier vértice como nodo terminal de un atajo. En general, cualquier vértice puede ser nodo terminal de dos atajos diferentes, excepto las plantas con atajos de tipo A, que pueden serlo de un único atajo. Las restricciones (10.66) impiden realizar atajos en plantas que no estén abiertas. La restricción (10.67) establece una cota superior para los costes fijos totales de las plantas abiertas. Dado que el término independiente npg es el coste de una solución que no abre ninguna planta y esa decisión se corresponde con una solución posible para ELR, se trata de una cota superior del coste de cualquier solución óptima y, en particular, de los costes fijos de apertura asociados con ella. Finalmente, las restricciones (10.68) evitan que clientes asociados con un atajo de tipo A sean nodos terminales de cualquier otro atajo. La solución óptima de este modelo proporciona una cota superior \bar{z}_{RS} de los ahorros esperados por “saltarse” clientes en cualquier conjunto factible de rutas.

7 Conclusiones

En este capítulo hemos estudiado algunos problemas estocásticos de localización discreta. En concreto, hemos considerado problemas con un factor aleatorio

rio determinado por el hecho de que la demanda de los clientes no es conocida *a priori*. Sin embargo, hemos asumido que conocemos la probabilidad con la que cada cliente tiene demanda, que es la misma para todos los clientes. En este contexto hemos planteado dos problemas discretos que difieren entre ellos en la forma en la que se proporciona el servicio a los clientes. En el primer caso, hemos supuesto que los clientes se desplazan hasta la planta para recibir servicio, dando lugar a problemas estocásticos de localización-asignación (ELA). En el segundo caso hemos supuesto que el servicio a los clientes se realiza donde se sitúan los clientes y que la naturaleza de los servicios permite que se realice una secuencia de ellos sin volver a la planta, dando lugar a problemas estocásticos de localización-rutas (ELR). Hemos abordado ambos problemas desde una perspectiva unificadora de programación con recurso. En cada caso hemos definido la solución *a priori* y la función de recurso para obtener la solución *a posteriori* que resulta al adaptar la solución *a priori* a cada posible escenario. Para el caso de los problemas ELA hemos presentado una heurística para obtener soluciones posibles, así como un algoritmo exacto de tipo L-shaped que aprovecha las propiedades de la función de recurso. Para los problemas de localización-rutas hemos obtenido la expresión analítica de la función de recurso, pero hemos sugerido una aproximación de la misma, dado que computacionalmente la evaluación de esta función resulta costosa. También hemos presentado una cota superior obtenida mediante una heurística y una cota inferior que consta de dos términos, uno asociado a las plantas abiertas y otro a las propias rutas. El estudio realizado permite concluir que las técnicas utilizadas son adecuadas para el tratamiento de estos problemas. Sin embargo, cabe resaltar que dada la complejidad de los problemas estudiados, desde el punto de vista computacional es necesario realizar un esfuerzo adicional que permita abordar problemas de mayores dimensiones. Desde el punto de vista teórico, pueden plantearse diversas extensiones del problema, entre las que cabe resaltar el estudio de otro tipo de distribuciones de probabilidad para las demandas de los clientes.

Agradecimientos

Este trabajo ha sido parcialmente financiado por las ayudas CICYT TIC2003-05982-C05-04 y CICYT BFM2002-10418-E.

Localización Minimax con Incertidumbre

Blas Pelegrín Pelegrín

Departamento de Estadística e Investigación Operativa.
Universidad de Murcia, e-mail: pelegrin@um.es

1 Introducción

Una decisión importante que hay que tomar al establecer un nuevo centro de servicio o actividad económica, tanto en el sector público como en el privado, es dónde localizarlo. Esto sucede por ejemplo con un hospital, un colegio, un restaurante, un almacén para servir a consumidores o detallistas, una máquina en una planta de producción, o un punto de venta de una franquicia. Una buena localización está ligada, entre otros factores, con una mayor eficacia, una mayor utilización y una disminución en los costes operativos del centro correspondiente. Existen una gran cantidad de modelos matemáticos que permiten estudiar los aspectos cuantitativos de estos problemas, conocidos con el nombre de *Modelos de Localización*. En situación de monopolio, el objetivo más frecuente es optimizar alguna función de los costes de instalación y de transporte, mientras que en un oligopolio el objetivo suele ser maximizar el beneficio del centro, o bien su cuota de mercado. Excelentes estudios y revisiones sobre estos modelos pueden consultarse en [3, 4, 5, 7, 45, 1, 8, 2].

En monopolio, cada modelo se suele formular como un problema de optimización, donde la función objetivo depende de la distancia entre los puntos de demanda y los centros que los sirven. Ejemplos de tales funciones son el coste total de transporte, la distancia máxima entre puntos de demanda y centros, y la distancia mínima entre dichos puntos. En particular, hablaremos de *Modelos de Localización Mínimax* si el problema consiste en minimizar el máximo de una serie de funciones, cada una de las cuales es a su vez una función de la distancia entre un punto de demanda y un centro. Los modelos difieren unos de otros de-

pendiendo del espacio de localización , de las funciones que se utilicen para medir la distancia, y de la función objetivo. Si el espacio es el *Plano*, cada centro y cada punto de demanda vienen representados por dos componentes (sus coordenadas geográficas) y la distancia entre dos puntos se mide mediante alguna función de sus coordenadas, normalmente usando una norma $\ell_p, 1 < p < \infty$, o $\ell_{2b}, b > 0$ [34, 42]). Si el espacio es una *Red de Transporte*, los puntos de demanda suelen estar en los nodos. Los centros están en los nodos, o en los tramos, midiéndose la distancia entre pares de puntos por la longitud del camino más corto que los une [11].

Cuando las funciones de la distancia son lineales, este tipo de problema se conoce con el nombre de *Problema del p-centro*, donde p hace referencia al número de centros a localizar, que ha sido ampliamente estudiado en ambiente de certidumbre. En el plano, para $p = 1$, tiene su origen en Silvester [19], quién en 1987 planteó el problema de encontrar el círculo de menor radio que contuviera a un conjunto de puntos dado (lo que equivale a encontrar un punto que minimice su distancia máxima a tales puntos). Sin embargo, este problema no se resolvió de forma eficiente hasta 1972, fecha que se publicó el algoritmo de Elzinga y Hearn [27], que más tarde fue generalizado al caso de funciones lineales de la distancia [31, 32]. En una red lo encontramos en 1964 con Hakimi [43], quién planteó y resolvió el problema cuando el candidato es cualquier punto. En ambos espacios, plano y red, el problema fue estudiado más adelante para funciones no lineales de la distancia [44, 38]. Para $p > 1$ ha sido extensamente estudiado, utilizándose distintas medidas de la distancia en el caso del plano. Debido a que es un problema NP-duro, sólo se han utilizado algoritmos de tipo heurístico para encontrar soluciones a problemas con un elevado número de puntos de demanda [30, 20, 41, 39]. No obstante, existen algoritmos exactos que permiten resolver problemas con hasta una o dos centenas de puntos [29, 33, 20, 18, 25, 24, 14, 23]. Su extensión al espacio euclídeo de dimensión $n > 2$ puede verse en

Los modelos minimax han sido aplicados a problemas de localización de servicios de emergencia, como bomberos y ambulancias; de centros multimedia, como emisoras y antenas de telefonía; y de servicios de mensajería urgente [46, 35]. Su estudio en el espacio euclídeo con dimensión $n > 2$ es de interés en Clasificación y Análisis de Datos [9, 13], donde se utilizan para la determinación de representantes de grupos. Como cabe esperar, en numerosas situaciones se presenta incertidumbre. Esto sucede en particular en la localización de servicios de emergencia, donde se deberá atender cualquier incidente en una determinada región, o en conjunto de puntos de riesgo. Si el criterio de decisión es minimizar el tiempo máximo de desplazamiento desde el centro a los puntos de demanda, el tiempo de viaje se puede considerar como una variable aleatoria. Ello conduce a considerar que el factor que pondera la distancia es *una variable aleatoria*. En esta dirección, los primeros estudios en una red fueron realizados por Frank [49, 50], y en el plano (con los puntos de demanda alineados) por Wesolowsky [48], ambos considerando

una distribución de probabilidad *Normal* para las variables aleatorias. Este tipo de incertidumbre ha sido extendido a otros modelos de localización [16, 51, 47, 17], pero sólo muy recientemente ha sido estudiado para el problema del 1-centro con otras distribuciones de probabilidad, como la distribución de *Bernouilli* ([52] y la *Uniforme* ([53, 54], y ha sido reconsiderado en una red usando la distribución *Normal* en [55].

En este trabajo presentamos un modelo general en el plano, donde la distancia viene medida por cualquier norma y se presenta incertidumbre en los coeficientes de la distancia, que vienen dados por variables aleatorias con distribuciones de probabilidad arbitrarias. Se consideran tres criterios de decisión, se analizan las propiedades básicas de los correspondientes modelos de optimización y se plantean procedimientos para su resolución.

2 El modelo

Sean $p_i = (p_{i1}, p_{i2})$, $i = 1, \dots, n$, un conjunto de puntos del plano cuyas coordenadas son conocidas. Cada punto p_i tiene asociada una cantidad $w_i \geq 0$ y va a ser servido desde un punto $x = (x_1, x_2)$, cuyas coordenadas se desconocen. El objetivo es encontrar las coordenadas del punto x , con objeto de minimizar la distancia máxima de dicho punto a los n puntos dados, ponderada por las cantidades indicadas. Es decir, se pretende resolver el siguiente problema:

$$(P) \min_{x \in \mathbb{R}^2} F(x) = \max_{1 \leq i \leq n} \{w_i d_i(x)\}. \quad (11.1)$$

Dependiendo de la situación que se estudie, los puntos p_i representan lugares donde tiene que llegar un determinado servicio, como una señal de TV, telefonía móvil,etc.; pueden ser puntos de riesgo de un determinado incidente, como un incendio, una explosión, una colisión de vehículos,etc.; puntos de venta a los que hay que entregar un producto perecedero , etc. La cantidad w_i asociada a cada punto mide su importancia en relación a los demás puntos, debido al tamaño de la población existente, grado de riesgo del posible incidente, cantidad de demanda,etc.; o bien representa el inverso de la velocidad media en desplazarse desde x al punto p_i , si se trata de un servicio de emergencia. La distancia $d_i(x)$ se suele medir con la norma euclídea, o con la norma rectangular, según se consideren grandes zonas geográficas , o zonas urbanas. Como ya se ha indicado en la sección 1, existen otras normas que proporcionan una mejor estimación de la distancia recorrida en los desplazamientos, por lo que vamos a considerar $d_i(x) = \|x - p_i\|$, donde $\|\cdot\|$ es una norma arbitraria, que especificaremos cuando sea necesario.

Si no se presenta incertidumbre, es bien conocido que el problema (P) tiene siempre solución óptima y al menos existe una solución óptima en la envolvente convexa de los puntos p_i . Su resolución puede hacerse por diferentes procedimientos, que dependen principalmente de la norma elegida y de si las ponderaciones

son iguales o distintas. Denotaremos por f^* al valor óptimo y por x^* a una solución óptima. A continuación vamos a describir los algoritmos de resolución que se utilizan con mayor frecuencia.

2.1 Puntos colineales

Si los puntos p_i están alineados, podemos representarlos de la forma $p_i = p_0 + \lambda_i v$, donde p_0 es un punto de la línea elegido de forma arbitraria y v es la dirección de alineación. Cualquiera que sea la norma, una solución óptima del problema viene determinada por el algoritmo siguiente.

Algoritmo 1

1. Calcular la cantidad:

$$F_\lambda = \max_{i \neq j} \left\{ \frac{w_i w_j |\lambda_i - \lambda_j|}{w_i + w_j} \right\}$$

2. Determinar dos índices t y s para los cuales se alcanza el valor máximo anterior.
3. Calcular el siguiente valor:

$$\lambda^* = \frac{w_t \lambda_t + w_s \lambda_s}{w_t + w_s}$$

4. La solución óptima y el valor óptimo son:

$$x^* = p_0 + \lambda^* v$$

,

$$f^* = F_\lambda$$

2.2 Puntos no colineales y norma ℓ_1

Si los puntos se encuentran en una zona urbana y los desplazamientos se realizan a través de sus calles, puede resultar apropiado estimar la distancia por la norma ℓ_1 , conocida también por norma rectangular o métrica de Manhattan. En este caso tenemos que :

$$d_i(x) = |x_1 - p_{i1}| + |x_2 - p_{i2}|$$

Para resolver el problema con esta métrica se usa la transformación $u_1 = x_1 + x_2$, $u_2 = x_1 - x_2$. Los puntos $p_i = (p_{i1}, p_{i2})$ se transforman en los puntos $u_i = (u_{i1}, u_{i2})$ y la distancia viene dada por:

$$d_i(x) = \max\{|u_1 - u_{i1}|, |u_2 - u_{i2}|\}$$

Reemplazando la expresión anterior en la función objetivo se obtiene que el problema es separable en las nuevas variables. Cada uno de los dos subproblemas resultantes es del tipo visto para puntos colineales. Entonces se obtiene el siguiente procedimiento de resolución [7].

Algoritmo 2

1. Calcular las cantidades:

$$F_{u1} = \max_{i \neq j} \left\{ \frac{w_i w_j |u_{i1} - u_{j1}|}{w_i + w_j} \right\}$$

$$F_{u2} = \max_{i \neq j} \left\{ \frac{w_i w_j |u_{i2} - u_{j2}|}{w_i + w_j} \right\}$$

$$F_u = \max \{ F_{u1}, F_{u2} \}$$

2. Determinar dos índices t y s para los cuales se alcanza el valor máximo F_{u1} .

Determinar dos índices h y k para los cuales se alcanza el valor máximo F_{u2} . Calcular los siguientes valores:

$$u_1^* = \frac{w_t u_{t1} + w_s u_{s1}}{w_t + w_s}$$

$$u_2^* = \frac{w_h u_{h2} + w_k u_{k2}}{w_h + w_k}$$

3. i) Si $F_{u1} = F_{u2}$ sólo hay una solución óptima x^* cuyas coordenadas son:

$$x_1^* = \frac{u_1^* + u_2^*}{2}$$

$$x_2^* = \frac{u_1^* - u_2^*}{2}$$

ii) Si $F_{u1} < F_{u2}$ las soluciones óptimas son los puntos $x = (x_1, x_2)$ del segmento:

$$\frac{1}{2}(u_2^* + \max_{1 \leq i \leq n} \{u_{i1} - \frac{F_u}{w_i}\}) \leq x_1 \leq \frac{1}{2}(u_2^* + \min_{1 \leq i \leq n} \{u_{i1} + \frac{F_u}{w_i}\})$$

$$x_1 - x_2 = u_2^*$$

iii) Si $F_{u1} > F_{u2}$ las soluciones óptimas son los puntos $x = (x_1, x_2)$ del segmento:

$$\frac{1}{2}(u_1^* + \max_{1 \leq i \leq n} \{u_{i2} - \frac{F_u}{w_i}\}) \leq x_1 \leq \frac{1}{2}(u_1^* + \min_{1 \leq i \leq n} \{u_{i2} + \frac{F_u}{w_i}\})$$

$$x_1 + x_2 = u_1^*$$

4. El valor óptimo es: $f^* = F_u$

2.3 Puntos no colineales y norma ℓ_2

En ocasiones los puntos de demanda están dispersos en grandes áreas y resulta más conveniente usar la norma euclídea. Entonces se tiene:

$$d_i(x) = \sqrt{(x_1 - p_{i1})^2 + (x_2 - p_{i2})^2}$$

Como esta métrica es una S-norma, el problema (P) tiene solución única y viene dada por la solución óptima de un problema similar con sólo dos o tres puntos de demanda [38]. Si tenemos en cuenta que dos puntos p_i, p_j , con $w_i \neq w_j$, determinan una circunferencia formada por todos los puntos que cumplen $w_i d_i(x) = w_j d_j(x)$, vamos a ver como se obtendría la solución óptima.

Para $n = 2$ la solución es el punto $x^* = \frac{w_1 p_1 + w_2 p_2}{w_1 + w_2}$.

Para $n = 3$ la solución está determinada por dos de los tres puntos, o por los tres. En el primer caso, hay que obtener las soluciones óptimas para los tres pares de puntos, según acabamos de indicar. En el segundo caso, la solución óptima se encuentra de la siguiente forma: Si $w_1 = w_2 = w_3$ es la intersección de las mediatrices del triángulo determinado por los tres puntos. Si $w_1 = w_2 \neq w_3$ es la intersección de la mediatriz del segmento $[p_1, p_2]$ con las circunferencias determinadas por los pares de puntos p_1, p_3 y p_2, p_3 . Los casos $w_1 = w_3 \neq w_2$ y $w_2 = w_3 \neq w_1$ son similares al anterior. Si $w_1 \neq w_2 \neq w_3$ es la intersección de las tres circunferencias determinadas por los pares p_1, p_2 , p_1, p_3 y p_2, p_3 . Un algoritmo muy eficiente para resolver el problema es el siguiente [31].

Algoritmo 3

- 1 Elegir dos puntos cualesquiera, que notamos por p_1 y p_2 . Obtener la solución óptima x^* para estos dos puntos. Hacer $N = 2$ e ir al paso 2.
- 2 Si $w_1 d_1(x^*) = F(x^*)$, PARAR. De lo contrario ir al paso 3 si $N = 2$ y al paso 4 si $N = 3$.
- 3 Denotemos por p_3 a un punto tal que $w_3 d_3(x^*) = F(x^*)$. Encontrar la solución x^* para p_1, p_2 y p_3 . Tomar N igual al número de puntos que determinan x^* . Si $N = 2$ denominar a los puntos correspondientes por p_1 y p_2 . Ir al paso 2.
- 4 Denotemos por p_4 a un punto tal que $w_4 d_4(x^*) = F(x^*)$. Encontrar la solución x^* para estos cuatro puntos explorando los pares $p_1, p_4 ; p_2, p_4 ; p_3, p_4$ y los tríos $p_1, p_2, p_4 ; p_1, p_3, p_4 ; p_2, p_3, p_4$. Tomar N igual al número de puntos que determinan x^* . Si $N = 2$ denominar a los puntos correspondientes por p_1 y p_2 . Si $N = 3$ denominarlos por p_1, p_2 y p_3 . Ir al paso 2.

En las secciones siguientes vamos a analizar este modelo cuando se presenta incertidumbre en los coeficientes de la distancia. Esto suele pasar cuando el producto $w_i d_i(x)$ representa el tiempo de viaje entre los puntos p_i y x . En este caso

w_i es el inverso de la velocidad media y puede considerarse una variable aleatoria, ya que la velocidad media es mayor o menor, dependiendo de la circulación existente. También puede haber incertidumbre si el riesgo de incidente en cada punto no es fijo, sino que varía dependiendo de las condiciones atmosféricas, del tráfico, o cualquier otro factor, según el problema objeto de estudio. Vamos a suponer entonces que los coeficientes de la distancia en el modelo son *variables aleatorias independientes con distribuciones de probabilidad arbitrarias*. Al vector n -dimensional aleatorio determinado por estas variables lo denotaremos por \mathbf{w} , de manera que $w = (w_1, w_2, \dots, w_n)$ representará un posible valor de \mathbf{w} . A la función objetivo la denotaremos por $F(x, \mathbf{w})$, que es una variable aleatoria que tomará el valor $\max_{1 \leq i \leq n} \{w_i d_i(x)\}$ si $\mathbf{w} = w$. Supondremos que la distribución de probabilidad de \mathbf{w} es *discreta*, o *absolutamente continua*, y que cada una de las variables aleatorias w_i toma valores dentro de un intervalo $[a_i, b_i]$, donde $a_i \geq 0$.

3 El criterio del valor esperado

Si aplicamos el criterio del valor esperado en el modelo anterior, tendremos que resolver el problema:

$$(E) \min_{x \in \mathbb{R}^2} E(x) = E_{\mathbf{w}}[F(x, \mathbf{w})] \quad (11.2)$$

donde $E_{\mathbf{w}}$ denota el valor esperado utilizando la correspondiente distribución de probabilidad del vector aleatorio \mathbf{w} . Vamos a ver que (E) es un problema de *Programación Convexa* con algunas características particulares. Sea H la envolvente convexa de los puntos p_i .

Teorema 2 *La función $E(x)$ es convexa en \mathbb{R}^2 y siempre se puede encontrar una solución óptima en H .*

Demostración: Para cada valor de la variable \mathbf{w} , la función $F(x, \mathbf{w})$ es convexa en la variable x , ya que está definida como el máximo de n funciones convexas (la distancia viene dada por una norma y puede comprobarse fácilmente que es una función convexa). Por consiguiente, si $x = \lambda y + (1 - \lambda)z$, $0 \leq \lambda \leq 1$, resulta que:

$$\begin{aligned} E(x) &= E_{\mathbf{w}}[F(x, \mathbf{w})] \leq E_{\mathbf{w}}[\lambda F(y, \mathbf{w}) + (1 - \lambda)F(z, \mathbf{w})] \\ &= \lambda E_{\mathbf{w}}[F(y, \mathbf{w})] + (1 - \lambda)E_{\mathbf{w}}[F(z, \mathbf{w})] = \lambda E(y) + (1 - \lambda)E(z). \end{aligned}$$

Como $E(x)$ es convexa en \mathbb{R}^2 , tiene que ser continua. Si $x' \notin H$, cualquiera que sea la norma que se utilice debe existir un punto $x \in H$ tal que $d_i(x) \leq d_i(x')$ (véase [56]). Por lo tanto:

$$E(x) = E_{\mathbf{w}}[F(x, \mathbf{w})] \leq E_{\mathbf{w}}[F(x', \mathbf{w})] = E(x')$$

Por ser $E(x)$ continua y la envolvente convexa un conjunto compacto, de la desigualdad anterior se desprende que su valor mínimo se alcanza en algún punto de H .

A los valores de toma la variable $F(x, \mathbf{w})$ los denotaremos por la letra f y a su función de distribución de probabilidad por $G_x(f)$. Esta función viene dada por:

$$G_x(f) = P_r[F(x, \mathbf{w}) \leq f] = \prod_{i=1}^n P_r[\mathbf{w}_i \leq \frac{f}{d_i(x)}] = \prod_{i=1}^n G_i\left(\frac{f}{d_i(x)}\right)$$

donde G_i es la función de distribución de probabilidad de la variable \mathbf{w}_i . Una vez determinada $G_x(f)$ se obtendrá su valor esperado $E(x)$.

Del teorema anterior se desprende que cualquier mínimo local de la función $E(x)$ es un mínimo global, por lo que se pueden utilizar los algoritmos usuales de descenso de la *Programación no Lineal* para encontrar una solución óptima del problema (E) . Como punto inicial para comenzar las iteraciones puede tomarse $x^0 = \frac{1}{n} \sum_{1 \leq i \leq n} p_i$, que pertenece a H . Una gran dificultad en el uso de estos algoritmos reside en la evaluación de la función $E(x^k)$ en cada una de las iteraciones, ya que la función de distribución de $F(x, \mathbf{w})$, aunque conocida, es en general complicada. El uso eficiente de un algoritmo determinado va a depender de cada situación, es decir de cuales sean las funciones G_i . El caso más sencillo es cuando las variables \mathbf{w}_i son binarias y ha sido recientemente estudiado en [52]. Dada la gran dificultad computacional para resolver este problema, se describen a continuación algunas alternativas para obtener una localización adecuada del centro en la situación de incertidumbre que estamos considerando.

Si los valores de cada variable \mathbf{w}_i están muy agrupados entorno a su valor esperado, que denotaremos por μ_i , una buena localización sería la solución óptima del problema (P) , tomando el valor μ_i como coeficiente de $d_i(x)$. Si el problema es localizar un servicio de emergencia, ello supondría minimizar el *máximo de los tiempos de viaje esperados*, en lugar del valor esperado del tiempo máximo de viaje. Con este criterio alternativo habría entonces que resolver el problema:

$$(P_\mu) \min_{x \in \mathbb{R}^2} F_\mu(x) = \max_{1 \leq i \leq n} \{\mu_i d_i(x)\} \quad (11.3)$$

Su solución óptima la denotaremos por x_μ^* y su valor óptimo por f_μ^* , los cuales pueden obtenerse según se ha indicado en la sección 2. La bondad de esta propuesta se debe a que la solución óptima de (P_μ) está determinada por a lo sumo tres de los puntos de demanda, los cuales están a una distancia (ponderada) del punto x_μ^* igual a f_μ^* (véase [38]). Para los otros puntos p_i , lo más normal es que $\mu_i d_i(x_\mu^*)$ sea bastante menor que f_μ^* . Por lo que para esos puntos $w_i d_i(x)$ no va a superar nunca f_μ^* , cualquiera que sea el valor que tome la variable \mathbf{w}_i , siempre que haya suficiente concentración entorno a su media.

Otra alternativa consiste en elegir como localización del centro una solución del problema (P) tomando el valor b_i como coeficiente de $d_i(x)$. Con este criterio

habría que resolver el problema:

$$(P_b) \min_{x \in \mathbb{R}^2} F_b(x) = \max_{1 \leq i \leq n} \{b_i d_i(x)\} \quad (11.4)$$

Su solución óptima la denotaremos por x_b^* y su valor óptimo por f_b^* . Con esta localización se garantiza que para cualquier realización de las variables aleatorias la función objetivo no sobrepase el valor f_b^* , ya que cada variable \mathbf{w}_i está acotada superiormente por b_i .

Sea $v(E)$ el valor óptimo del problema (E) . Con las dos alternativas anteriores se obtienen cotas para este valor, según se indica a continuación.

Teorema 3 *Se verifica que:*

$$f_\mu^* \leq v(E) \leq f_b^*$$

Demuestra: Como $\mathbf{w}_i d_i(x) \leq F(x, \mathbf{w})$, resulta que $\mu_i d_i(x) \leq E_{\mathbf{W}}[F(x, \mathbf{w})]$. Por lo tanto $F_\mu(x) \leq E(x)$, de donde se desprende que $F_\mu^* \leq E(x), \forall x$. Por otra parte $F(x, \mathbf{w}) \leq \max_{1 \leq i \leq n} \{b_i d_i(x)\}, \forall x$, de donde resulta que

$$E_{\mathbf{W}}[F(x, \mathbf{w})] \leq \max_{1 \leq i \leq n} \{b_i d_i(x)\}, \forall x.$$

Al minimizar en las expresiones anteriores se obtienen las desigualdades propuestas.

4 El criterio de máximo cubrimiento en probabilidad

En ocasiones, fijar un valor f y localizar el centro en un punto con probabilidad máxima de que la variable $F(x, \mathbf{w})$ no sobrepase dicho valor sea , puede resultar más interesante que usar el criterio del valor esperado. Para encontrar ese punto habría que resolver el problema:

$$(C_f) \max_{x \in \mathbb{R}^2} P(x) = P_r[F(x, \mathbf{w}) \leq f] \quad (11.5)$$

La función objetivo de este problema se puede evaluar muy fácilmente, al contrario de lo que sucedía con la del problema (E) , ya que viene dada por:

$$P(x) = G_x(f) = \prod_{i=1}^n G_i\left(\frac{f}{d_i(x)}\right)$$

Como ya se ha indicado, G_i es la función de distribución de la variable \mathbf{w}_i , que suponemos conocida. Sin embargo, esta función no es cóncava (los máximos

locales no son globales), por lo que para resolver el problema (C_f) no se puede usar el mismo tipo de algoritmos que para el problema (E) . Denotemos por $v(C_f)$ y $X(C_f)$ respectivamente al valor óptimo y al conjunto de soluciones óptimas del problema (C_f) . Vamos a ver en primer lugar sus propiedades más notables.

Teorema 4 *Para cualquier valor $f > 0$, el problema (C_f) tiene solución y siempre se puede encontrar una solución óptima en H .*

Demostración: Según se ha visto en el teorema 2, si $x' \notin H$, entonces existe un punto $x \in H$ tal que $d_i(x) \leq d_i(x')$. Por lo tanto:

$$P(x) = \prod_{i=1}^n G_i\left(\frac{f}{d_i(x)}\right) \geq \prod_{i=1}^n G_i\left(\frac{f}{d_i(x')}\right) = P(x')$$

Entonces al menos una solución óptima se encuentra en H , si existe alguna. Veamos que el valor óptimo se alcanza y por consiguiente el problema tiene solución. Si \mathbf{w} es de tipo discreto, es evidente que la función objetivo $P(x)$ toma un número finito de valores y en consecuencia se alcanza un valor máximo. Si \mathbf{w} fuese absolutamente continua, las funciones G_i serían continuas y por lo tanto se alcanzaría un valor máximo en H , que es un conjunto compacto.

Sea Ω el espacio muestral de la variable aleatoria \mathbf{w} . Por A denotaremos un suceso en dicho espacio. A cada punto $x \in \mathbb{R}^2$ le asociaremos el suceso $A(x, f) = \{w = (w_1, w_2, \dots, w_n) \in \Omega : w_i d_i(x) \leq f, i = 1, 2, \dots, n\}$ (ver figura ??). Si definimos la familia de sucesos $\Omega_f = \{A(x, f) : x \in \mathbb{R}^2\}$, el problema (C_f) equivale a encontrar un suceso en dicha familia que tenga probabilidad máxima. Para buscar tal suceso, podemos descartar aquellos que no sean maximales respecto de la relación de inclusión.

Sea $A \in \Omega_f$, definimos los valores $w_i(A) = \sup\{w_i : w = (w_1, \dots, w_i, \dots, w_n) \in A\}$, $i = 1, 2, \dots, n$, y el conjunto $X(A, f) = \{x \in \mathbb{R}^2 : A(x, f) = A\}$. Denotemos por $B(p_i, r)$ la bola de centro p_i y radio r , es decir $B(p_i, r) = \{y \in \mathbb{R}^2 : d_i(y) \leq r\}$. Entonces, tenemos la siguiente caracterización de suceso maximal.

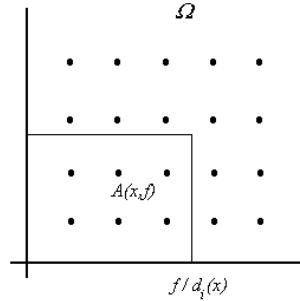


Fig 1: Sucesos $A(x, f)$ en Ω

Teorema 5 Sea $A \in \Omega_f$, entonces:

- i) $X(A, f) \subset \bigcap_{i=1,n} B(p_i, \frac{f}{w_i(A)})$
- ii) A es maximal si y sólo si $X(A, f) = \bigcap_{i=1,n} B(p_i, \frac{f}{w_i(A)})$

Demostración:

i) Si $x \in X(A, f)$, entonces $w_i d_i(x) \leq f, i = 1, 2, \dots, n, \forall w \in A$, por lo tanto $w_i(A) d_i(x) \leq f, i = 1, \dots, n$, lo que implica que $x \in \bigcap_{i=1,n} B(p_i, \frac{f}{w_i(A)})$. Por consiguiente : $X(A, f) \subset \bigcap_{i=1,n} B(p_i, \frac{f}{w_i(A)})$.

ii) Sea A maximal. Si $x \in \bigcap_{i=1,n} B(p_i, \frac{f}{w_i(A)})$, entonces $w_i(A) d_i(x) \leq f, i = 1, 2, \dots, n$, de donde se deduce que $w_i d_i(x) \leq f, i = 1, 2, \dots, n, \forall w \in A$ y por consiguiente $A \subset X(x, f)$. Como A es maximal, resulta que $A = X(x, f)$, por lo tanto $x \in X(A, f)$. En consecuencia, $\bigcap_{i=1,n} B(p_i, \frac{f}{w_i(A)}) \subset X(A, f)$. Si tenemos en cuenta i), resulta que la inclusión anterior es una igualdad. Recíprocamente, supongamos que se da la igualdad. Si $A \subset A' \in \Omega_f$, tomemos un punto x tal que $A(x, f) = A'$, entonces tiene que cumplirse que $w_i d_i(x) \leq f, i = 1, 2, \dots, n, \forall w \in A'$, y en particular $\forall w \in A$. Por lo tanto $x \in \bigcap_{i=1,n} B(p_i, \frac{f}{w_i(A)}) = X(A, f)$, lo que significa que $A(x, f) = A$. Por consiguiente $A = A'$, es decir A es maximal.

Teorema 6 Si la variable aleatoria \mathbf{w} es discreta, $X(C_f)$ es la unión de un número finito de conjuntos que son intersecciones de bolas centradas en los puntos de demanda.

Demostración:

Si \mathbf{w} es discreta, habrá un número finito de sucesos maximales con probabilidad máxima. Las localizaciones correspondientes a dichos sucesos son las óptimas. Por el teorema anterior, dichas localizaciones son los puntos intersección de bolas centradas en los puntos de demanda.

Como vemos, los conjuntos de soluciones óptimas del problema (C_f) son de la forma $X(A, f) = \bigcap_{i=1,n} B(p_i, \frac{f}{w_i(A)})$, donde A es algún suceso maximal. Los sucesos maximales en Ω_f vienen determinados por puntos de Ω que son maximales en el orden natural de \mathbb{R}^n (los puntos $w(A) = (w_1(A), \dots, w_n(A))$). El elemento maximal correspondiente al suceso Ω es $w(\Omega) = b = (b_1, b_2, \dots, b_n)$. Resolviendo el problema (P_b) de la sección 3, se obtiene que si $f_b^* \leq f$ entonces $v(C_f) = 1$ y $X(C_f) = \bigcap_{i=1,n} B(p_i, \frac{f}{b_i})$. Para valores de f menores que f_b^* , el valor óptimo de (C_f) es menor que la unidad, siempre que la variable \mathbf{w} tenga probabilidad positiva en un entorno del punto b . Por otro lado, si tomamos el elemento minimal del espacio muestral $a = (a_1, a_2, \dots, a_n)$, podemos resolver el siguiente problema:

$$(P_a) \min_{x \in \mathbb{R}^2} F_a(x) = \max_{1 \leq i \leq n} \{a_i d_i(x)\} \quad (11.6)$$

Su solución óptima la denotaremos por x_a^* y su valor óptimo por f_a^* . El valor óptimo de (C_f) es positivo si $f_a^* \leq f$, mientras que $v(C_f) = 0$ si $f < f_a^*$, en cuyo

caso todos los puntos del plano son soluciones óptimas. Así pues, si descartamos los dos casos anteriores, queda por resolver (C_f) para $f_a^* \leq f < f_b^*$.

Si la variable \mathbf{w} es discreta, para resolverlo se puede realizar un proceso de exploración de sucesos maximales según se indica a continuación. Para valores de f próximos a f_b^* , los sucesos maximales de Ω_f corresponden a los elementos de Ω que preceden en el orden natural de \mathbb{R}^n al punto b . Estos elementos son los puntos $pb(i) = (b_1, \dots, pb_i, \dots, b_n), i = 1, \dots, n$, donde pb_i es el valor inmediatamente anterior a b_i que toma la variable \mathbf{w}_i . Para cada uno de ellos se resuelve el problema (P) tomando $w = pb_i$. Sea $f_{pb_i}^*$ su valor óptimo. Si $f_{pb_i}^* \leq f$, entonces el suceso $A_i = \{w \in \Omega : w \leq pb_i \text{ in } \Omega_f\}$ y es maximal. Entre los sucesos maximales de Ω_f seleccionaremos los de probabilidad máxima, a los que nos referiremos por A_k . Entonces el conjunto de soluciones óptimas de (C_f) vendrá dado por la unión de los conjuntos $X(A_k, f)$. Si $f < f_{pb_i}^*$, habrá que seguir el proceso de búsqueda con los siguientes elementos de Ω que preceden a pb_i . Mediante este proceso de inspección, podemos encontrar todas las soluciones óptimas, como se ilustra con el siguiente ejemplo.

Ejemplo 1: Sean $p_1 = (2, 1)$, $p_2 = (3 - \frac{1}{\sqrt{2}}, 3 + \frac{3}{\sqrt{2}})$ y $p_3 = (5, 2)$. Supongamos que las variables aleatorias $\mathbf{w}_1, \mathbf{w}_2$ y \mathbf{w}_3 tienen una distribución de probabilidad de Bernoulli, con parámetros $\frac{1}{2}, \frac{3}{5}$ y $\frac{2}{5}$ respectivamente. La distancia viene medida por la norma ell_2 . El espacio muestral es $\Omega = \{w^1 = (0, 0, 0), w^2 = (1, 0, 0), w^3 = (0, 1, 0), w^4 = (0, 0, 1), w^5 = (1, 1, 0), w^6 = (1, 0, 1), w^7 = (0, 1, 1), w^8 = (1, 1, 1)\}$.

Comenzamos resolviendo el problema (P_b) , donde ahora b es w^8 . La solución se obtiene mediante el *Algoritmo 3* y viene dada por $x_{w^8}^* = (3, 3)$, que corresponde al circuncentro del triángulo que determinan los puntos p_i . El valor óptimo es $f_{w^8}^* = 2,2361$. Por consiguiente, para $f \geq 2,2361$, el conjunto de soluciones óptimas es $X(C_f) = \bigcap_{i=1,2,3} B(p_i, f)$ y $v(C_f) = 1$.

Para $f < f_{w^8}^*$, los sucesos maximales de Ω_f corresponden a los puntos: $pb_1 = w^7$, $pb_2 = w^6$ y $pb_3 = w^5$. Las soluciones óptimas de los problemas (P_{pb_i}) , $i = 1, 2, 3$, son respectivamente:

$$x_{w^7}^* = \left(4 - \frac{1}{2\sqrt{2}}, \frac{5}{2} + \frac{3}{2\sqrt{2}}\right), x_{w^6}^* = \left(\frac{7}{2}, \frac{3}{2}\right), x_{w^5}^* = \left(\frac{5}{2} - \frac{1}{2\sqrt{2}}, 2 + \frac{3}{2\sqrt{2}}\right)$$

y sus valores óptimos son:

$$f_{w^7}^* = 2,0659, f_{w^6}^* = 1.5811, f_{w^5}^* = 2,0659$$

Las probabilidades de los correspondientes sucesos maximales son:

$$P_r(A_1) = P_r(\mathbf{w}_1 = 0) = 0,5, P_r(A_2) = P_r(\mathbf{w}_2 = 0) = 0,4, P_r(A_3) = P_r(\mathbf{w}_3 = 0) = 0,6$$

. Si $f_{w^5}^* \leq f < f_{w^8}^*$, el suceso de máxima probabilidad es A_3 , por lo tanto el conjunto de soluciones óptimas es $X(C_f) = \bigcap_{i=1,2} B(p_i, f)$ y $v(C_f) = 0,6$.

Para $f < f_{w^5}^*$, los sucesos maximales de Ω_f corresponden a los puntos w^6 y w^3 , que son los que dominan a los otros predecesores de w^5 y w^7 en el orden natural de \mathbb{R}^3 . El suceso que corresponde a w^6 es A_2 , que ya ha sido considerado. El que corresponde a w^3 lo denotamos por A_4 y su probabilidad es $P_r(A_4) = P_r(\mathbf{w}_1 = 0, \mathbf{w}_2 = 0) = 0,3$. La solución óptima de (P_{w^3}) es $x_{w^3}^* = p_2$ y su valor óptimo es $f_{w^3}^* = 0$. Si $f_{w^6}^* \leq f < f_{w^5}^*$, el suceso de máxima probabilidad es A_2 , por lo que el conjunto de soluciones óptimas es $X(C_f) = \bigcap_{i=1,3} B(p_i, f)$ y $v(C_f) = 0,4$.

Para $f < f_{w^6}^*$ los sucesos maximales de Ω_f corresponden a los puntos w^4 , w^3 y w^2 . La solución óptima de (P_{w^4}) es $x_{w^4}^* = p_3$ y su valor óptimo es $f_{w^4}^* = 0$. El correspondiente suceso maximal es A_5 y su probabilidad es $P_r(A_5) = P_r(\mathbf{w}_1 = 0, \mathbf{w}_2 = 0) = 0,2$. Para w^2 se obtiene $x_{w^2}^* = p_1$ y $f_{w^2}^* = 0$. El correspondiente suceso maximal es A_6 y su probabilidad es $P_r(A_6) = P_r(\mathbf{w}_2 = 0, \mathbf{w}_3 = 0) = 0,24$. Como vemos el suceso de máxima probabilidad es A_4 , por lo tanto si $0 \leq f < f_{w^6}^*$ entonces $X(C_f) = B(p_2, f)$ y $v(C_f) = 0,3$.

Si la variable \mathbf{w} es absolutamente continua, no se puede aplicar el proceso de enumeración antes descrito. La función objetivo $P(x)$ es continua, pero no es cóncava (ni convexa), por lo que en este caso habrá que utilizar *éTécnicas de Optimización Global para resolver (C_f)* . Algunos algoritmos basados en estas técnicas que pueden usarse para su resolución pueden verse en [21, 22, 6, 10, 12]. Ilustraremos este caso con un ejemplo sencillo.

Ejemplo 2 : Sean $p_1 = (0,0)$, $p_2 = (7,0)$. Supongamos que las variables aleatorias \mathbf{w}_1 y \mathbf{w}_2 toman valores en el intervalo $(0, 1)$ con las funciones de distribución : $G_1(w_1) = w_1^2$ y $G_2(w_2) = w_2$. Por el teorema 4, una solución óptima se encuentra en $H = [p_1, p_2]$. Los puntos de este segmento son de la forma $x = (a, 0)$, $0 \leq a \leq 7$, por lo que la función objetivo de (C_f) en tales puntos podemos expresarla tomando a como argumento. Dicha función es:

$$P(a) = \begin{cases} (f/a)^2(f/(7-a)) & \text{si } f < a < 7-f \\ (f/a)^2 & \text{si } f < a, 7-f \leq a \\ (f/(7-a)) & \text{si } a \leq f, a < 7-f \\ 1 & \text{si } 7-f \leq a \leq f \end{cases}$$

Está claro que $P(a)$ puede tomar el valor 1 si $7-f \leq f$. Esto sucede si $\frac{7}{2} \leq f$, en cuyo caso el subintervalo de $[p_1, p_2]$ correspondiente a los valores de a en $[7-f, f]$ es un conjunto de localizaciones óptimas. Si $f < \frac{7}{2}$, resulta sencillo comprobar que el valor máximo de $P(a)$ es $\frac{f}{7-f}$ y se alcanza para $a = f$.

5 El criterio de la restricción de incertidumbre

Cuando se trata de servicios de emergencia, es importante que el valor f , que se fija para usar el criterio de la sección anterior, sea lo más pequeño posible. En tal caso, como la probabilidad máxima de cubrimiento es no decreciente con

f , podría ocurrir que dicha probabilidad fuese muy pequeña. Ello nos lleva a fijar una probabilidad mínima α para cubrimiento y buscar una localización que minimice el valor f condicionado a la misma. Es decir, habría que resolver el siguiente problema:

$$(R_\alpha) \min\{f : P_r[F(x, \mathbf{w}) \leq f] \geq \alpha, x \in \mathbb{R}^2, f \geq 0\} \quad (11.7)$$

Denotemos por $v(R_\alpha)$ y $X(R_\alpha)$ respectivamente al valor óptimo y al conjunto de localizaciones óptimas del problema (R_α)

Teorema 7 $v(R_\alpha) = f_\alpha^*$ si y sólo si $v(C_f) < \alpha, \forall f < f_\alpha^*$ y $v(C_f) \geq \alpha, \forall f > f_\alpha^*$.

Demostración:

Sea $v(R_\alpha) = f_\alpha^*$. Si $f < f_\alpha^*$, entonces $P_r[F(x, \mathbf{w}) \leq f] < \alpha, \forall x \in \mathbb{R}^2$. Como el valor máximo del problema (C_f) se alcanza, resulta que $v(C_f) < \alpha$. Si $f > f_\alpha^*$, debe haber algún punto \hat{x} tal que $P_r[F(\hat{x}, \mathbf{w}) \leq f] \geq \alpha$, por consiguiente $v(C_f) \geq \alpha$.

Recíprocamente, supongamos que para un valor f_α^* se verifican las desigualdades. Entonces no existe ningún $x \in \mathbb{R}^2$ tal que $P_r[F(x, \mathbf{w}) \leq f] \geq \alpha$ si $f < f_\alpha^*$, mientras que ocurre lo contrario si $f > f_\alpha^*$. por lo tanto $v(E_\alpha) = f_\alpha^*$.

Si definamos la función $\varphi(f) = v(C_f)$, del teorema anterior resulta que el problema (R_α) lo podemos expresar de la siguiente forma:

$$(R_\alpha) \min\{f : \varphi(f) \geq \alpha, f \geq 0\} \quad (11.8)$$

Este es un problema de optimización unidimensional sencillo de resolver si se conociese la función $\varphi(f)$. En efecto, el valor óptimo f_α^* es el menor valor de f para el que $\varphi(f) \geq \alpha$ y el conjunto de localizaciones óptimas viene dado por $X(R_\alpha) = x \in \mathbb{R}^2 : P_r[A(x, f_\alpha^*)] \geq \alpha$. Vamos a ver que propiedades tiene la función $\varphi(f)$ y la relación entre las soluciones óptimas de (R_α) y (C_f) .

Teorema 8 Si \mathbf{w} es discreta, entonces:

i) Existe un número finito de valores $f_0 = 0 < f_1 < f_2 < \dots < f_k$ tales que:

$$\varphi(f) = \varphi(f_h), \text{ si } f_h \leq f < f_{h+1}$$

donde $\varphi(f_k) = 1$ y $f_{k+1} = \infty$.

ii) $X(R_\alpha) \supset X(C_{f_h})$ si $\varphi(f_{h-1}) < \alpha \leq \varphi(f_h)$ y $X(R_\alpha) = X(C_{f_h})$ si y estos conjuntos son iguales si $\alpha = \varphi(f_h)$.

Demostración:

i) Para cada valor f , denotemos por Ω_f^* a la familia formada por los sucesos de Ω_f que tienen probabilidad máxima. Si $A \in \Omega_f^*$, entonces $X(A, f)$ es un conjunto de soluciones óptimas del problema (C_f) . Sea $\Omega^* = \{A : A \in \Omega_f^*, f \geq 0\}$, este

conjunto está formado por un número finito de sucesos. Sean $\alpha_0 < \alpha_1 < \alpha_2 < \dots < \alpha_k$ las diferentes probabilidades de realización de tales sucesos. Si $A \in \Omega^*$, el conjunto $X(A, f)$ es no vacío si $f \geq f_A$, donde por simplicidad denotamos por f_A al valor óptimo del problema (P) tomando como coeficientes de las distancias los valores $w_i(A)$, $i = 1, \dots, n$. Sea $f_h = \min\{f_A : A \in \Omega^*, P_r(A) = \alpha_h\}, h = 0, 1, 2, \dots, k$, donde $P_r(A)$ es la probabilidad del suceso A . Como la probabilidad es no decreciente con f se verifica que $f_0 = 0 < f_1 < f_2 < \dots < f_k$. Si $f_h \leq f < f_{h+1}$, tendremos que $X(A, f)$ es no vacío para algún suceso A tal que $P_r(A) = \alpha_h$ y $X(A, f)$ es vacío para cualquier suceso con probabilidad mayor, por lo que $\varphi(f) = \alpha_h$. Como Ω es maximal para valores de f suficientemente grandes resulta que $\varphi(f) = 1$ si $f \geq f_k$.

ii) Si $x \in X(C_{f_h})$ entonces $P_r[F(x, \mathbf{w}) \leq f_h] = \varphi(f_h) \geq \alpha$, por lo tanto $x \in X(R_\alpha)$. Si $\alpha = \varphi(f_h)$, resulta que $x \in X(R_\alpha)$ si $P_r[F(x, \mathbf{w}) \leq f_h] \geq \varphi(f_h)$. Como esta desigualdad no puede ser estricta, el suceso $A(x, f_h)$ tiene probabilidad máxima y $x \in X(C_{f_h})$.

Un procedimiento de resolver R_α , cualquiera que sea el valor de α , consiste en resolver previamente los problemas (C_f) , con objeto de determinar la función $\varphi(f) = v(C_f)$. Una vez obtenida esta función, podemos encontrar el valor óptimo f_α^* según se ha indicado. Las localizaciones óptimas las encontraremos a partir de los sucesos $A(x, f_\alpha^*)$ con probabilidad mayor o igual que α . Vamos a resolver R_α para el ejemplo 1.

Según se vio en la sección 3, la función $\varphi(f)$ viene dada por la siguiente expresión:

$$\varphi(f) = \begin{cases} 0,3 & \text{si } 0 \leq f < 1,5811 \\ 0,4 & \text{si } 1,5811 \leq f < 2,0659 \\ 0,6 & \text{si } 2,0659 \leq f < 2,2361 \\ 1 & \text{si } 2,2361 \leq f \end{cases}$$

Los valores óptimos de los problemas (R_α) vienen reflejados en la Tabla 11.1. Las localizaciones óptimas para los diferentes valores de α , las obtendremos siguiendo un orden de probabilidad descendente. Para $0,6 < \alpha \leq 1$, únicamente el suceso $A = \Omega$ tiene una probabilidad superior a igual a α , por lo que la solución es el punto $x_{w^8}^*$. Para $0,5 < \alpha \leq 0,6$, sólo el suceso A tal que $w(A) = w^5$ tiene probabilidad superior, o igual, a igual a α , por lo que la solución es $x_{w^5}^*$. Para $0,4 < \alpha \leq 0,5$, los sucesos correspondientes a $w(A) = w^5$ y $w(A) = w^7$ son los únicos con probabilidad superior o igual a α , por lo que hay dos soluciones óptimas $x_{w^5}^*$ y $x_{w^7}^*$. Siguiendo este proceso se obtiene la Tabla 11.2 que muestra los resultados para todos los valores de α . Obsérvese que para resolver (R_α) no es necesario resolver (C_f) para todos los valores de f , sino sólo para aquellos que verifican $\varphi(f) \geq \alpha$.

Teorema 9 Si \mathbf{w} es absolutamente continua, entonces:

Tabla 11.1: Valores óptimos para (R_α) en el ejemplo 1

α	$v(R_\alpha)$
$0,6 < \alpha \leq 1$	2,2361
$0,4 < \alpha \leq 0,6$	2,0659
$0,3 < \alpha \leq 0,4$	1.5811
$0 < \alpha \leq 0,3$	0

Tabla 11.2: Localizaciones óptimas para (R_α) en el ejemplo 1

α	$X(R_\alpha)$
$0,6 < \alpha \leq 1$	$x_{w^8}^*$
$0,5 < \alpha \leq 0,6$	$x_{w^5}^*$
$0,4 < \alpha \leq 0,5$	$x_{w^5}^*, x_{w^7}^*$
$0,3 < \alpha \leq 0,4$	$x_{w^6}^*$
$0,24 < \alpha \leq 0,3$	p_2
$0,2 < \alpha \leq 0,24$	p_1, p_2
$0,12 < \alpha \leq 0,2$	p_1, p_2, p_3
$0 < \alpha \leq 0,12$	\mathbb{R}^2

- i) $\varphi(f)$ es continua.
- ii) $X(R_\alpha) = X(C_{f_\alpha}^*)$.

Demostración:

i) Sea $f_h \rightarrow f$. Por el teorema 4, para cada f_h existirá un punto x_h en H que es solución óptima de C_{f_h} . Como H es compacto, la sucesión x_h contendrá una subsucesión x_k convergente a un punto $x \in H$. Entonces:

$$\lim_{f_k \rightarrow f} \varphi(f_k) = \lim_{f_k \rightarrow f} G_{x_k}(f_k) = G_x(f) \leq \varphi(f)$$

Por otra parte, si x^* es una solución óptima de C_f se verifica que $G_{x^*}(f_k) \leq \varphi(f_k)$, por lo que:

$$\lim_{f_k \rightarrow f} \varphi(f_k) \geq \lim_{f_k \rightarrow f} G_{x^*}(f_k) = G_{x^*}(f) = \varphi(f)$$

Por consiguiente $\lim_{f_k \rightarrow f} \varphi(f_k) = \varphi(f)$, y como la función φ es monótona también se verifica que $\lim_{f_h \rightarrow f} \varphi(f_h) = \varphi(f)$.

ii) Como $\varphi(f)$ es continua se verifica que $\varphi(f_\alpha^*) = \alpha$. Entonces $x \in X(C_{f_\alpha}^*)$ equivale a que $P_r[F(x, \mathbf{w}) \leq f_\alpha^*] \geq \alpha$ y esto a su vez equivale a que $x \in X(R_\alpha)$.

Para el ejemplo 2 la función $\varphi(f)$ viene dada por:

$$\varphi(f) = \begin{cases} \frac{f}{7-f} & \text{si } 0 \leq f < 3,5 \\ 1 & \text{si } 3,5 \leq f \end{cases}$$

Es fácil comprobar que $v(R_\alpha) = \frac{7\alpha}{1+\alpha}$ y que este valor se alcanza en $x^* = (\frac{7\alpha}{1+\alpha}, 0)$.

6 Conclusiones y líneas futuras de investigación

Se ha presentado y analizado un modelo general para la localización de un centro con criterio minimax cuando se presenta incertidumbre en los coeficientes de las distancias. En lugar de tratar algún caso concreto, como en la mayoría de las referencias citadas, se han considerado variables aleatorias con distribuciones de probabilidad arbitrarias, lo que permite abordar un gran número de situaciones.

El modelo se ha analizado para tres criterios de decisión en situación de incertidumbre. Para cada uno de ellos se han obtenido propiedades fundamentales, en base a las cuales es posible desarrollar distintos procedimientos de resolución. Para el criterio del valor esperado, su resolución es posible mediante técnicas de *Programación Convexa*, si bien el cálculo de la función objetivo requiere una gran cantidad de recursos computacionales. Para el criterio del máximo cubrimiento en probabilidad, en general es necesario usar técnicas de *Optimización Global*, pero también se pueden desarrollar procedimientos de enumeración si las variables aleatorias son discretas. El criterio de la restricción de incertidumbre se basa en el anterior y precisa por tanto de los mismos procedimientos. El desarrollo de algoritmos eficientes que permitan resolver problemas de tamaño real constituye la principal línea de investigación en el estudio de cada uno de los problemas que se han presentado. Esto conduce a tener que especificar las distribuciones de probabilidad de las mencionadas variables, de manera que tales algoritmos habrá que diseñarlos de forma adecuada para cada caso concreto.

7 Bibliografía

- [1] Handler GY and Mirchandani PB (1979), Location on Networks: Theory and Algorithms. MIT Press.
- [2] Mirchandani PB and Francis RL (1990) Discrete Location Theory. John Wiley & Sons.
- [3] Daskin M.S. (1995) Network and Discrete Location: Models, Algorithms and Applications. John Wiley & Sons, New York.
- [4] Drezner Z (1995) Facility Location: A Survey of Applications and Methods. Springer Verlag.
- [5] Drezner Z and Hamacher HW (2002) Facility Location: Applications and Theory. Springer Verlag.

- [6] Hansen E. (1992) Global Optimization Using Interval Analysis. Marcel Dekker.
- [7] Francis RL, McGinnis LF and White JA (1992) Facility Layout and Location: An Analytical Approach. Prentice Hall.
- [8] Love RF, Morris JG and Wesolowsky GO (1988) Facilities Location: Models and Methods. North Holland.
- [9] Kaufman L, Rousseeuw PJ (1990) Finding groups in data. Wiley.
- [10] Kearfott R.B. (1996) Rigorous Global Search: Continuous Problems. Kluwer, Dordrecht.
- [11] Pelegrín B, Cánovas L y Fernández (1992) Algoritmos en Grafos y Redes. PPU.
- [12] Ratschek H. and Rokne J. (1988) New Computer Methods for Global Optimization. Ellis Horwood, Chichester.
- [13] Späth H (1985) Cluster Analysis Algorithms for data reduction and classification of objects. Ellis Horwood limited.
- [14] Pelegrín B (1996) The p -center problem under bidirectional polyhedral norms. In : Pelegrín (ed) Proceedings EWGLA III 151-169. Publicaciones de la Universidad de Sevilla.
- [15] Pelegrín B (1996) El problema general del p -centro en \mathbb{R}^n . In : Puerto (ed) Lecturas en Teoría de Localización 149-166. Publicaciones de la Universidad de Sevilla.
- [16] Berman O and Krass D (2002), Facility location problems with stochastic demands and congestion. In: Drezner Z and Hamacher H (eds) Facility Location: Applications and Theory 329-371. Springer Verlag.
- [17] Marianov V and ReVelle C (1995) Siting emergency services. In : Drezner Z (ed) Facility Location: A Survey of Applications and Methods. Springer Verlag.
- [18] Drezner Z (1987) On the rectangular p -center problem. Naval Research Logistics 34:229-234.
- [19] Sylvester JJ (1857) A question in geometry of situation Quart. J. Pure Applied Math 1:79.
- [20] Drezner Z (1984) The p -center problem: Heuristics and optimal algorithms. Journal of the Operational Research Society 35:741-748.

- [21] Fernández J, Fernández P and Pelegrín B. (2000) A continuous location model for siting a non-noxious undesirable facility within a geographical region, European Journal of Operational Research 121: 259–274.
- [22] Fernández J and Pelegrín B (2001) Using interval analysis for solving planar single-facility location problems: new discarding tests. Journal of Global Optimization 19: 61–81.
- [23] Vijay J (1985) An algorithm for the p -center problem in the plane, Transportation Science 19:235-245.
- [24] Kariv O and Hakimi SL (1979) An algorithmic approach to network location problems. Part I :The p -centers. SIAM Journal on Applied Mathematics 37: 539-560.
- [25] Halpern J and Maimon O (1982) Algorithms for the m -center problem: A survey. European Journal of Operations Research 10:90-99.
- [26] Pelegrín B (1991) The p -center problem in \mathbb{R}^n with weighted Tchebychell norms , JORBEL 31:49-62.
- [27] Elzinga J and Hearn DW (1972) Geometric solutions for some minimax location problems, Transportation Science 6:379–394.
- [28] Elzinga J and Hearn DW (1972) The minimum covering sphere problem. Management Science 19:96-104.
- [29] Aneja YT, Chandrasekaran R and Nir KPK (1988) A note on the m -center problem with rectilinear distances. European Journal of Operational Research 35:118-123.
- [30] Chandrasekaran R and Tamir A (1982) Polynomially bounded algorithms for locating p -center pproblems on a tree. Mathematical Programming 22:304-315.
- [31] Charalambous C (1982) Extension of the Elzinga and Hearn algorith to the weigthed case. Operations Research 30:591-594.
- [32] Hearn DW and Vijay J (1982) Efficient algorithms for the weigthed minimum circle problem,Operations Research 30:777–795.
- [33] Caruso C, Colomni A and Alois L (2003) Dominant, an algorithm for the p -center problem. European Journal of Operational Research 149: 53-64.
- [34] Berens W, Koerling FJ (1988) On Estimating Road Distances by Mathematical Functions- a Rejoinder , European Journal of Operational Research 36, 254-255.

- [35] Toregas CR, Swan C, Revelle C, and Bergman L (1971) The location of emergency service facilities, *Operations Research* 19, 1366–1373.
- [36] Minieka E (1981) , A polynomial time algorithm for finding the absolute center of a network , *Networks*, 11 , 351-355.
- [37] Pelegrín B. (1986), A general approach to the 1-center problem, *Cahiers du CERO* 28, 293-301.
- [38] Pelegrín B, Michelot C and Plastria F (1985) On the uniqueness of optimal solutions in continuous location theory, *European Journal of Operations Research* 29:98-110.
- [39] Pelegrín B (1991) Heuristics methods for the p -center problem, *RAIRO Recherche Opérationnelle* 25:65-72.
- [40] Pelegrín, B. and Cánovas, L.(1998) The minimum l_{pb} -hypersphere problem. *Computational Optimization and Applications* 9: 85–97.
- [41] Dyer ME and Frieze AM (1985)A simple heuristic for the p -center problem. *Operations Research Letter* 3:285-288.
- [42] Fernández J, Fernández P and Pelegrín, B (2002) Estimating actual distances by norm functions: A comparison between the $l_{k,p,\theta}$ -norm and the $l_{b_1,b_2,\theta}$ -norm and a study about the selection of the data set . *Computers and Operations Research* 29:609-623.
- [43] Hakimi S.L. (1964), Optimal Location of Switching Centers and the Absolute Centers and Medians of a Graph, *Operations Research* 12, 450–459.
- [44] Hakimi S.L. and Maheshwari S.N. (1972), Optimum Locations of Centers in Networks, *Operations Research* 20, 967–973.
- [45] Hamacher, H.W. and Nickel, S. (1998), Classification of location models, *Location Science*, 6 , 229–242 .
- [46] Bräuer ML and Chiu SS (1989)An overview of representative problems in location research. *Management Science* 35:645-674.
- [47] Louveaux F (1993) Stochastic Location Analysis. *Location Science* 1:127-154.
- [48] Wesolowsky GO (1977) Probabilistic weights in the one-dimensional location problem. *Management Science* 24:224-229.
- [49] Frank H (1966)Optimal locations on a graph with probabilistic demands. *Operations Research* 14:409-421.

- [50] Frank H (1966) Optimal locations on a graph with correlated normal demands. *Operations Research* 15:552-557.
- [51] Daskin MS (1983) A maximum expected covering location model: Formulations, properties and heuristic solution. *Transportation Science* 17:48-70.
- [52] Berman O, Drezner Z and Wesolowsky GO (2003) The expected maximum distance objective in facility location. *Journal of Regional Science* 43:735-748.
- [53] Berman O, Wang H, Drezner Z and Wesolowsky GO (2003) A probabilistic minimax location problem on the plane. *Annals of Operations Research* 122:59-70.
- [54] Berman O, Wang H, Drezner Z and Wesolowsky GO (2003) A probabilistic minimax location problem on the plane. *Annals of Operations Research* 122:59-70.
- [55] Berman O and Drezner Z (2003) A probabilistic one-center location problem on a network. *Journal of the Operational Research Society* 54:871-877.
- [56] Wendell RE and Hurter AP (1973) Location theory, dominance, and convexity. *Operations Research* 21:314-320.
- [57] Ross DW (1977) Lysosomes and storage diseases. MA Thesis, Columbia University, New York

Cálculo de medianas robustas con incertidumbre en las demandas

M. J. Canós Darós^{a1}, M. L. Martínez Romero^b,
M. Mocholí Arce^a

^a Dep. Matemática Económico-Empresarial, Universitat de València,

^c Departamento de Ingeniería y Matemáticas, Florida Universitaria

1 Introducción

Uno de los problemas a los que se enfrentan las empresas es decidir donde ubicar sus instalaciones de modo que sus costes de aprovisionamiento y distribución sean mínimos. Existen numerosos modelos que tratan de resolver estos problemas de localización. Los más básicos sólo necesitan la demanda y una función de distancia que permita determinar el coste de satisfacer dicha demanda.

El problema de la p -mediana es un modelo básico de localización cuando la estructura topológica subyacente del problema es una red en la que las longitudes de las aristas (distancias) y los pesos de los vértices (demandas) son conocidos. Consiste en encontrar p puntos de la red de modo que se minimice la distancia total (o media) ponderada entre estos puntos y los vértices. Fue introducido por Hakimi [9, 10] quien demostró la propiedad de optimalidad en los vértices, a saber, que siempre existe una p -mediana en los vértices de la red. Suponer que los datos son deterministas y estáticos, sin embargo, es poco realista, pues la mayoría proviene de estimaciones sobre hechos que todavía no se han producido y de los cuales, en muchos casos, no hay información histórica. Además, estas

¹Este trabajo ha sido parcialmente subvencionado por TIC 2002-04242-C03

aproximaciones inexactas deben proyectarse hacia el futuro para un largo periodo de tiempo en el cual los cambios en el entorno son muy difíciles, sino imposibles, de predecir. Así, surgen de forma natural los modelos y métodos que aceptan incertidumbre en los datos, como el análisis de sensibilidad, análisis paramétrico, modelos dinámicos y modelos estocásticos. Un resumen de ellos aparece en el artículo de Owen y Daskin [14]. Más recientemente se han aplicado también técnicas borrosas [3, 4, 5] y técnicas de optimización robusta [12, 6, 15, 2, 1].

Las técnicas de optimización robusta, al contrario que el análisis de sensibilidad o el análisis paramétrico, consideran que la incertidumbre es una característica inherente al sistema y que, en lugar de eliminarla, es mucho más provechoso hacer un esfuerzo para estructurarla hasta donde sea posible, entenderla y manejarla. La optimización robusta no necesita que la incertidumbre esté provocada por un solo parámetro (el tiempo), como la optimización dinámica, ni tampoco que exista una distribución de probabilidad asociada, como la optimización estocástica, requisito no trivial en problemas que, como los de localización, estudian fenómenos únicos con poca o ninguna información histórica. Cualquier técnica robusta ha de seguir tres pasos [12]:

1. Planificación de los escenarios. Un escenario es una realización potencial de los datos inciertos del problema. La filosofía de la optimización robusta es la de estar preparados para enfrentarse a (casi) cualquier suceso futuro. Por tanto, del buen diseño de los escenarios depende el éxito o el fracaso de todo el proceso posterior.
2. Elección de un criterio de robustez. Puesto que es imposible que sepamos que va a ocurrir en el futuro, el criterio de robustez debe llevarnos a una solución del problema que se porte bien bajo cualquier escenario.
3. Planteamiento de un modelo coordinado. El modelo coordinado recoge el criterio de robustez, la información proporcionada por todos los escenarios y, una vez planteado, puede resolverse por técnicas conocidas.

Cuando el decisor obtiene una solución para el modelo coordinado desea que esté lo más cerca posible de la factibilidad y de la optimalidad para todos y cada uno de los escenarios. Para ello, se admiten desviaciones respecto de la factibilidad y la optimalidad y se intenta minimizar ambas [13]; o bien se supone que los escenarios tienen soluciones factibles en común y sólo se permiten desviaciones respecto a la optimalidad [12].

Todos los trabajos que aplican técnicas robustas al problema de la p -mediana, excepto el de Canós y Mocholí [2], trabajan bajo la última hipótesis; es decir, suponen que todas las restricciones de todos los escenarios deben cumplirse. En particular, suponen que toda la demanda debe ser atendida. Sin embargo, en algunos casos podría interesar dejar una pequeña parte de la demanda insatisfecha a cambio de otras compensaciones. Este último caso es el que planteamos. En

este trabajo consideramos el problema de la mediana sobre un árbol en el que la incertidumbre viene reflejada por el hecho de que cada demanda puede variar en un rango de valores. El cálculo de estos intervalos puede realizarse por el método propuesto por Canós y Mocholí [2]. Nuestro objetivo es conseguir una solución que, cuando los datos tomen algún valor en el futuro, minimice la demanda insatisfecha y cuyo coste no esté muy lejos del mínimo coste que hubiésemos obtenido de haber conocido con antelación dicho valor. Puesto que la forma de plantear matemáticamente este objetivo no es única, presentamos y comparamos varias opciones entre las que puede elegir el decisor.

2 Cálculo de los escenarios

Consideremos un árbol $T = (V, E)$, donde $V = \{v_1, \dots, v_n\}$ es el conjunto de vértices y E el de aristas. Cada arista tiene asociada una longitud positiva conocida. Supongamos que cada vértice tiene una demanda incierta asociada representada por un intervalo. Así, $w_j = [\underline{\omega}_j, \bar{\omega}_j]$ es la demanda del vértice v_j , para $j = 1, \dots, n$. Es evidente que tenemos infinitos escenarios. Al conjunto de todos los escenarios lo llamaremos S . Dado un escenario $s \in S$, denotamos por ω_j^s el valor de la demanda de v_j en dicho escenario. Obviamente, se cumplirá que $\underline{\omega}_j \leq \omega_j^s \leq \bar{\omega}_j$. Para todo par de puntos x e y de T , $d(x, y)$ es la distancia entre x e y , calculada como la longitud del único camino entre x e y .

Definición 10 *Diremos que un punto m^s es una mediana de T bajo el escenario s si para todo x de T se cumple que*

$$\sum_{j=1}^n \omega_j^s d(m^s, v_j) \leq \sum_{j=1}^n \omega_j^s d(x, v_j)$$

Definición 11 *Diremos que un punto x de T es una mediana de escenario si existe algún s de S tal que x es una mediana de T bajo el escenario s .*

Puesto que en cada uno de los escenarios tenemos un problema con datos conocidos y constantes, se puede aplicar la propiedad de optimalidad en los vértices. Por tanto, podemos suponer que las medianas de escenario son vértices de T . Además, una mediana de escenario cumplirá cualquier propiedad conocida para las medianas. En particular cumplirá la propiedad de la semisuma.

Teorema 10 (Propiedad de la semisuma. Kariv y Hakimi [11]) *Un vértice v_j es una mediana de un árbol si, y sólo si, cada subárbol con raíz en v_j tiene un peso total (excluyendo v_j) menor o igual que la mitad de la suma de todos los pesos del árbol.*

Como consecuencia, los siguientes teoremas, cuya demostración se basa en la anterior propiedad y cuya redacción hemos modificado ligeramente para adaptarlos a nuestro problema particular, son ciertos independientemente del criterio de robustez elegido.

Teorema 11 (Chen y Lin [6]) *Sean v_i , v_j y v_k tres vértices de T tales que v_k está en el único camino que une v_i con v_j . Si v_i y v_j son dos medianas de escenario entonces v_k también es una mediana de escenario.*

Como corolario, todas las medianas de escenario están conectadas y podemos hablar del camino de las medianas, cuyos vértices serán medianas bajo algún escenario.

Definición 12 Consideremos $v_j \in V$. Sean T_1, T_2, \dots, T_k los k subárboles cuya raíz es v_j . El árbol resultante de convertir cada uno de estos subárboles en un vértice cuya peso es la suma, con la aritmética habitual de intervalos, de los pesos de todos los vértices del subárbol se denomina *árbol agregado de T con raíz en v_j* .

Es evidente, por la propiedad de la semisuma, que v_j es una mediana de T si, y sólo si, es una mediana del árbol agregado de T con raíz en v_j .

Teorema 12 (Chen y Lin [6]) Consideremos el árbol agregado con raíz en v_j . Entonces v_j es una mediana de escenario si, y sólo si, es una mediana del escenario agregado c_j , cuyos pesos vienen dados por:

$$\omega_j^{cj} = \bar{\omega}_j, \quad \omega_i^{cj} = \min\{\underline{\omega}_k, \bar{\omega}_i\} \quad \forall i \neq j$$

siendo

$$\underline{\omega}_k = \max_{i \neq j} \underline{\omega}_i$$

Con el teorema anterior podemos calcular si un vértice determinado es o no una mediana de escenario. Si lo es, conocemos un escenario agregado asociado. Para calcular un escenario de S , basta asignar a cada vértice de T un valor factible y que cumpla la propiedad de que al sumar los pesos de los vértices de cada subárbol el resultado sea el peso correspondiente del escenario agregado. A cada escenario que cumpla esta propiedad lo llamaremos desagregado.

Algoritmo ESCEGEN.

INPUT. Introducir como datos el árbol T y las demandas w_j , $j = 1, \dots, n$.

INICIALIZACIÓN. Hacer $i = 1$.

PASO 1. Tomar el vértice v_i y crear el árbol agregado de T con raíz en v_i .

PASO 2. Comprobar si v_i es una mediana del árbol agregado bajo el escenario ci . En caso afirmativo, es una mediana de escenario en T . Calcular un escenario desagregado dci . En caso negativo, no existe ningún escenario en T para el que v_i sea óptimo.

PASO 3. Comprobar si $i = n$. En caso afirmativo, parar. En caso negativo, hacer $i = i + 1$ e ir al paso 1.

OUTPUT. Obtener como resultado $M = \{v_{(i)} : (i) = 1, \dots, k\}$ el conjunto de las medianas de escenario, y $SD = \{dc(i) : (i) = 1, \dots, k\}$, el conjunto de los escenarios desagregados calculados.

Este algoritmo permite calcular todas las medianas del problema y un escenario, al menos, para cada una de ellas.

Ejemplo 2.1. Consideremos el siguiente árbol T con 12 vértices.

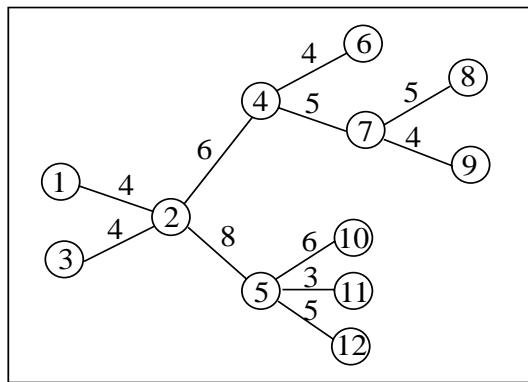


Figura 12.1: Árbol T

Las demandas de cada vértice son

$$\begin{array}{llll} \omega_1 = [0, 9] & \omega_2 = [15, 18] & \omega_3 = [16, 30] & \omega_4 = [5, 20] \\ \omega_5 = [225, 250] & \omega_6 = [50, 85] & \omega_7 = [65, 90] & \omega_8 = [150, 350] \\ \omega_9 = [45, 70] & \omega_{10} = [15, 20] & \omega_{11} = [31, 40] & \omega_{12} = [23, 26] \end{array}$$

Vamos a ilustrar los pasos del algoritmo tomando como v_i el vértice 2.

El árbol agregado para el vértice 2 es
cuyas demandas serán:

$$\omega_{1a} = [0, 9] \quad \omega_2 = [15, 18] \quad \omega_{3a} = [16, 30] \quad \omega_{4a} = [315, 615] \quad \omega_{5a} = [294, 336]$$

El escenario $c2$ tiene los pesos

$$(\omega_{1a}^{c2}, \quad \omega_2^{c2}, \quad \omega_{3a}^{c2}, \quad \omega_{4a}^{c2}, \quad \omega_{5a}^{c2}) = (9, \quad 18, \quad 30, \quad 315, \quad 315)$$

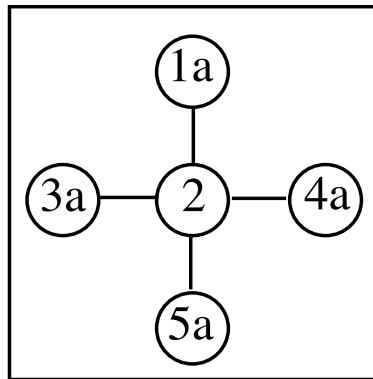


Figura 12.2: Árbol agregado

Puesto que se cumple que cada uno de ellos es menor que la semisuma total, deducimos que el vértice 2 es una mediana bajo el escenario $c2$ en el árbol agregado. Por tanto, el vértice 2 es una mediana de escenario para el árbol T .

Ahora debemos calcular un escenario desagregado al que llamaremos $dc2$. Nuestro procedimiento ha sido el siguiente:

1. Si el vértice agregado coincide con un único vértice del árbol T , a dicho vértice le asignamos el peso que le corresponde en el árbol agregado. Así, puesto que el vértice $1a$ coincide con el vértice 1, habrá un escenario $dc2$ en T de forma que el vértice 2 será una mediana bajo $dc2$ y $\omega_1^{dc2} = \omega_{1a}^{c2} = 9$. En el caso particular del vértice que hemos tomado como raíz, vértice 2 en nuestro ejemplo, esto siempre será cierto.
2. Si el vértice agregado ha sido creado a partir de un subárbol de T con más de un vértice, hemos asignado a cada uno de ellos su cota inferior. Si la suma coincide con el peso del vértice agregado bajo el escenario agregado, estos valores son los pesos del escenario desagregado asociado. Así ocurre con el vértice $4a$. En caso contrario, empezamos añadiendo la diferencia al vértice del subárbol con intervalo de mayor longitud, hasta su cota superior; continuamos por el de mayor longitud de los restantes, y así sucesivamente hasta que la suma coincide con el peso del escenario agregado. Esto ocurre en el vértice $5a$ donde hemos necesitado aumentar la cota inferior del vértice 5 en 21 unidades.

Aplicando este procedimiento al vértice 2 obtenemos un escenario desagregado donde los pesos de los vértices son

$$(9, 18, 30, 5, 246, 50, 65, 150, 45, 15, 31, 23).$$

Al ejecutar el algoritmo ESCEGEN para el árbol T , obtenemos el conjunto

$$M = \{2, 4, 7\}$$

y el conjunto SD compuesto por los escenarios dados por los siguientes pesos:

	ω_1^s	ω_2^s	ω_3^s	ω_4^s	ω_5^s	ω_6^s	ω_7^s	ω_8^s	ω_9^s	ω_{10}^s	ω_{11}^s	ω_{12}^s
$s = dc2$	9	18	30	5	246	50	65	150	45	15	31	23
$s = dc4$	0	15	16	20	225	85	65	215	45	15	31	23
$s = dc7$	0	15	16	5	225	50	90	350	70	15	31	23

y su correspondiente demanda total atendida y coste total

	demandada total atendida	coste total
$s = dc2$	687	7294
$s = dc4$	755	7884
$s = dc7$	890	8794

El algoritmo ESCEGEN nos da una caracterización de los óptimos bajo cualquier escenario, pero no de todos los escenarios asociados a cada óptimo. Veremos en la sección 4 que, en el caso particular en que todas las demandas recorran su intervalo dependiendo de un único parámetro (por ejemplo, si suponemos que la fuente de la incertidumbre es el tiempo), podemos refinar el algoritmo obteniendo una caracterización de todos los escenarios.

3 Planteamiento del modelo coordinado

Nuestro criterio de robustez está perfectamente definido. Es un criterio en esencia biobjetivo puesto que trata de minimizar la demanda no atendida por la solución robusta al mismo tiempo que obtiene un coste cercano al coste mínimo que se hubiera obtenido de saber con antelación qué escenario se iba a realizar. La forma de plantear matemáticamente este criterio en un modelo coordinado no es única.

Bajo la hipótesis de que las distancias son conocidas, planteamos tres modelos coordinados:

Modelo 1: su objetivo consiste en la minimización de las desviaciones respecto de la demanda no atendida en cada uno de los escenarios desagregados

calculados y el coste asociado a la solución robusta.

$$\begin{aligned} \text{Min} \quad & C \sum_{(i)=1}^k DD_{(i)}^2 + \sum_{i,j=1}^n d_{ij}x_{ij} \\ \text{s.a.} \quad & \sum_{i=1}^n y_i = 1 \end{aligned} \quad (1.1)$$

$$\underline{\omega}_j y_i \leq x_{ij} \leq \bar{\omega}_j y_i \quad i, j = 1, \dots, n \quad (1.2)$$

$$\sum_{j=1}^n \omega_j^{dc(i)} + DD_{(i)} = \sum_{i,j=1}^n x_{ij} \quad (i) = 1, \dots, k \quad (1.3)$$

$$y_i \in \{0, 1\}$$

En este modelo, además de los parámetros definidos anteriormente, aparecen $d_{ij} = d(v_i, v_j)$ y C que es una constante de escala. Puesto que las desviaciones vienen expresadas en unidades de producto y el coste en unidades monetarias se hace necesario utilizar un coeficiente de ponderación. En el ejemplo 3.1 hemos utilizado

$$C = \frac{\sum_{i,j=1}^n d_{ij}}{n^2}$$

Las variables de localización y_i son variables binarias que valen 1 si localizamos un centro de servicio en el vértice i y 0 en caso contrario. Las variables de asignación x_{ij} representan la demanda del vértice j atendida desde el vértice i . Las variables $DD_{(i)}$ son las desviaciones respecto de la demanda atendida en el escenario desagregado para los que el óptimo es el vértice (i) .

Por último, la ecuación (1.1) indica el número de centros de servicio que hay que localizar, las ecuaciones (1.2) indican que la demanda total del vértice j que puede ser atendida desde el centro de servicio debe estar comprendida entre el valor de su demanda mínima y máxima. Además, aseguran que el vértice i no atenderá demanda si no hay ningún centro de servicio ubicado en él así como la no negatividad de las variables de asignación. Por último, la ecuación (1.3) recoge la desviación $DD_{(i)}$ de la demanda total atendida por el modelo coordinado con respecto a cada escenario desagregado calculado en el que el óptimo es el vértice (i) .

Modelo 2: su objetivo es la minimización de las desviaciones del coste de la solución robusta respecto de los costes de cada escenario.

$$\begin{aligned} \text{Min} \quad & \sum_{(i)=1}^k DC_{(i)}^2 \\ \text{s.a.} \quad & \sum_{i=1}^n y_i = 1 \end{aligned} \tag{2.1}$$

$$\underline{\omega}_j y_i \leq x_{ij} \leq \bar{\omega}_j y_i \quad i, j = 1, \dots, n \tag{2.2}$$

$$\sum_{j=1}^n \omega_j^{dc(i)} d_{(i)j} + DC_{(i)} = \sum_{i,j=1}^n d_{ij} x_{ij} \quad (i) = 1, \dots, k \tag{2.3}$$

$$y_i \in \{0, 1\}$$

En este modelo, las variables $DC_{(i)}$ representan la desviación del coste total del modelo coordinado respecto del coste del escenario desagregado cuyo óptimo es el vértice (i) y las ecuaciones (2.3)recogen las desviaciones respecto del coste.

Modelo 3: su objetivo es una combinación de los objetivos de los dos modelos anteriores.

$$\begin{aligned} \text{Min } & C \sum_{(i)=1}^k DD_i^2 + \sum_{(i)=1}^k DC_{(i)}^2 \\ \text{s.a. } & \sum_{i=1}^n y_i = 1 \end{aligned} \quad (3.1)$$

$$\frac{\underline{\omega}_j}{\bar{\omega}_j} y_i \leq x_{ij} \leq \bar{\omega}_j y_i \quad i, j = 1, \dots, n \quad (3.2)$$

$$\sum_{j=1}^n \omega_j^{dc(i)} + DD_{(i)} = \sum_{i,j=1}^n x_{ij} \quad (i) = 1, \dots, k \quad (3.3)$$

$$\sum_{j=1}^n \omega_j^{dc(i)} d_{(i)j} + DC_{(i)} = \sum_{i,j=1}^n d_{ij} x_{ij} \quad (i) = 1, \dots, k \quad (3.4)$$

$$y_i \in \{0, 1\}$$

Ejemplo 3.1. Consideremos el árbol T del ejemplo 2.1 cuya matriz de distancias es la siguiente:

$$D = \left(\begin{array}{cccccccccccc} 0 & 4 & 8 & 10 & 12 & 14 & 15 & 20 & 19 & 18 & 15 & 17 \\ 4 & 0 & 4 & 6 & 8 & 10 & 11 & 16 & 15 & 14 & 11 & 13 \\ 8 & 4 & 0 & 10 & 12 & 14 & 15 & 20 & 19 & 18 & 15 & 17 \\ 10 & 6 & 10 & 0 & 14 & 4 & 5 & 10 & 9 & 20 & 17 & 19 \\ 12 & 8 & 12 & 14 & 0 & 18 & 19 & 24 & 23 & 6 & 3 & 5 \\ 14 & 10 & 14 & 4 & 18 & 0 & 9 & 14 & 13 & 24 & 21 & 23 \\ 15 & 11 & 15 & 5 & 19 & 9 & 0 & 5 & 4 & 25 & 22 & 24 \\ 20 & 16 & 20 & 10 & 24 & 14 & 5 & 0 & 9 & 30 & 27 & 29 \\ 19 & 15 & 19 & 9 & 23 & 13 & 4 & 9 & 0 & 29 & 26 & 28 \\ 18 & 14 & 18 & 20 & 6 & 24 & 25 & 30 & 29 & 0 & 9 & 11 \\ 15 & 11 & 15 & 17 & 3 & 21 & 22 & 27 & 26 & 9 & 0 & 8 \\ 17 & 13 & 17 & 19 & 5 & 23 & 24 & 29 & 28 & 11 & 8 & 0 \end{array} \right)$$

Con esta tabla de distancias, las soluciones obtenidas para cada uno de los modelos anteriores están recogidas en las tablas 12.1, 12.2 y 12.3, las cuales demuestran que cada uno de los modelos corresponde a una actitud diferente del decisor. El modelo 1 trata de ajustar la demanda atendida respecto de cualquier realización futura obteniendo como resultado la menor desviación posible respecto de cada uno de los escenarios, teniendo en cuenta además el coste para que éste no sea excesivo, aunque las desviaciones en coste respecto a realizaciones futuras puede llegar a ser grande.

El modelo 2 es conservador respecto al coste, minimizando las desviaciones respecto de cualquier realización futura pero a costa de dejar mucha demanda sin atender, en las desviaciones de la tabla 12.3 se aprecia que no llega a cubrir

	Demanda Mínima	Demanda atendida Modelo 1	Demanda atendida Modelo 2	Demanda atendida Modelo 3	Demanda Máxima
Vértice 1	0	9	9	0	9
Vértice 2	15	18	15	18	18
Vértice 3	16	30	16	16	30
Vértice 4	5	20	5	20	20
Vértice 5	225	225	225	230,083	250
Vértice 6	50	85	50	85	85
Vértice 7	65	90	65	90	90
Vértice 8	150	161,212	168,028	204,25	350
Vértice 9	45	70	45	45	70
Vértice 10	15	15	15	15	20
Vértice 11	31	31	31	31	40
Vértice 12	23	23	23	23	26

Tabla 12.1: Tabla de demanda atendida

	Modelo 1	Modelo 2	Modelo 3
Localización óptima	4	5	4
Coste óptimo	7944,116	7990,667	7990,667
Demandas total atendida	777,212	667,028	777,333

Tabla 12.2: Tabla de soluciones

	Modelo 1	Modelo 2	Modelo 3
DD(2)	90,212	-19,972	90,233
DD(4)	22,212	-879,72	22,333
DD(7)	-112,788	-222,972	-112,667
DC(2)	700,116	696,667	696,667
DC(4)	110,116	106,667	106,667
DC(7)	-799,884	-803,333	-803,333

Tabla 12.3: Tabla de desviaciones

la demanda de ninguno de los escenarios. En este sentido, podemos afirmar que el modelo 1 resultará más adecuado para la localización de centros de servicio públicos, mientras que el modelo 2 lo será para centros de servicio privados.

Por último, el modelo 3 intenta cubrir ambas posibilidades simultáneamente. Desde este punto de vista, es el mejor de los modelos. Sin embargo, el esfuerzo computacional que requiere es mucho mayor que el de los modelos 1 y 2. Por ello, en cada caso en concreto, el decisor deberá valorar el binomio eficiencia-coste computacional, eligiendo el modelo más adecuado para su problema real.

Puesto que la función de demanda total y la función de coste son crecientes en la demanda, es lógico que cada uno de los modelos intente ajustar las desviaciones a la desviación “central” correspondiente al vértice central del camino de las medianas óptimas, en este caso, el vértice 4. Ello no implica, como se demuestra en el modelo 2, que dicho vértice central sea siempre la ubicación óptima para el modelo coordinado.

4 Cuando la demanda depende de un único parámetro

Supongamos ahora que todos los pesos pueden ser escritos como una función lineal de un único parámetro $t \in [L, U]$ y denotemos por m_{t_0} a la mediana correspondiente al escenario obtenido para el valor t_0 . Vamos a ver que, en este caso, podemos obtener una caracterización completa de los infinitos escenarios asociados con cada mediana de escenario.

Teorema 13 (Erkut y Tansel [7]) . Para todo $t \in [L, U]$ existe una mediana m_t en el camino que une m_L y m_U .

Además, Erkut y Tansel (1992) demostraron que se podían calcular los subintervalos de $[L, U]$ para los que era óptima cada una de las medianas de escenario. Aplicando su método a nuestro problema, obtenemos el siguiente algoritmo:

Algoritmo ESCEPAR

INPUT. Introducir como datos el árbol T y las demandas ω_j , $j = 1, \dots, n$.

INICIALIZACIÓN. Parametrizar las demandas, haciendo

$$\omega_j(t) = a_j + b_j t, \text{ con } t \in [0, 1], \quad a_j = \underline{\omega}_j \text{ y } b_j = \overline{\omega}_j - \underline{\omega}_j.$$

PASO 1. Calcular el vértice inicial del camino de las medianas óptimas, m_0 , correspondiente al escenario con pesos $\omega_j(0)$ para $j = 1, \dots, n$, y el vértice final, m_1 correspondiente al escenario con pesos $\omega_j(1)$ para $j = 1, \dots, n$, mediante algún algoritmo conocido.

PASO 2. Calcular el camino desde m_0 hasta m_1 , renombrar los vértices del camino como $\{m_0 = v_{(1)}, v_{(2)}, \dots, m_1 = v_{(k)}\}$ y asignar a cada vértice del camino el peso del subárbol del cual es raíz $wc_{(i)}(t) = ac_{(i)} + bc_{(i)}t$, para $(i) = 1, \dots, k$.

PASO 3. Calcular los puntos de ruptura del parámetro mediante la siguiente expresión:

$$t_{(i)(i+1)} = \frac{-(ac_{(1)} + \dots + ac_{(i)}) + (ac_{(i+1)} + \dots + ac_{(k)})}{(bc_{(1)} + \dots + bc_{(i)}) - (bc_{(i+1)} + \dots + bc_{(k)})},$$

para $(i) = 1, \dots, k - 1$, donde $t_{(i)(i+1)}$ representa el valor de t para el cual toda la arista $(v_{(i)}, v_{(i+1)})$ son medianas.

OUTPUT. Obtener como resultado M , el conjunto de las medianas de escenario; ST , el conjunto de los subintervalos del parámetro t y SP , el conjunto de los escenarios definidos por subintervalos de las demandas.

Ejemplo 4.1. Consideremos el árbol T del ejemplo 2.1.

Para calcular los vértices inicial (vértice 2) y final (vértice 7) del camino de las medianas hemos utilizado el algoritmo de Goldman (1971) cuya complejidad algorítmica es $O(n)$. El camino de las medianas es $M = \{(1), (2), (3)\} = \{2, 4, 7\}$ con pesos

$$wc_2(t) = 325 + 68t, \quad wc_4(t) = 55 + 50t, \quad wc_7(t) = 260 + 250t$$

y el conjunto ST viene dado por:

Mediana	Valores de t
Vértice 2	$[0, 5/116[$
Arista (2, 4)	$5/116$
Vértice 4	$]5/116, 10/11[$
Arista (4, 7)	$10/11$
Vértice 7	$]10/11, 1]$

A partir de ST, el cálculo de SP es inmediato.

En este caso, los modelos coordinados pueden afinar más la solución que en los ejemplos anteriores puesto que no conocemos sólo un escenario para cada óptimo, sino el conjunto de los infinitos asociados a cada mediana. Con esto, y con los mismos objetivos que en la sección 3, proponemos:

Modelo 4: su objetivo consiste en la minimización de las desviaciones respecto de la demanda no atendida en cada uno de los escenarios y el coste asociado

a la solución robusta.

$$\begin{aligned} \text{Min} \quad & C \sum_{\substack{(i)=1 \\ (i)=1}}^k (t_{(i)(i+1)} - t_{(i-1)(i)}) DD_{(i)}^2 + \sum_{i,j=1}^n d_{ij} x_{ij} \\ \text{s.a.} \quad & \sum_{i=1}^n y_i = 1 \end{aligned} \quad (4.1)$$

$$x_{ij} \leq \bar{\omega}_j y_i \quad i, j = 1, \dots, n \quad (4.2)$$

$$\sum_{\substack{i=1 \\ 0 \leq t \leq 1}}^n x_{ij} = \underline{\omega}_j + (\bar{\omega}_j - \underline{\omega}_j) t \quad j = 1, \dots, n \quad (4.3)$$

$$0 \leq t \leq 1 \quad (4.4)$$

$$\sum_{j=1}^n [\underline{\omega}_j + (\bar{\omega}_j - \underline{\omega}_j) t_{(i)}] + DD_{(i)} = \sum_{i,j=1}^n x_{ij} \quad (i) = 1, \dots, k \quad (4.5)$$

$$t_{(i-1)(i)} \leq t_{(i)} \leq t_{(i)(i+1)} \quad i = 1, \dots, k \quad (4.6)$$

$$x_{ij} \geq 0, y_i \in \{0, 1\}$$

Con la misma notación que en los modelos de la sección 3, la ecuación (4.1) indica el número de centros de servicio que hay que localizar, las ecuaciones (4.2) representan que el vértice i no atenderá demanda si no hay ningún centro de servicio ubicado en él, la ecuación (4.3) indica que la demanda total del vértice j que puede ser atendida desde todos los centros de servicio debe estar comprendida entre el valor de su demanda mínima y máxima y la ecuación (4.5) recoge la desviación $DD_{(i)}$ de la demanda total atendida por el modelo coordinado con respecto a cada escenario particular en el que el óptimo es el vértice (i) .

Modelo 5: su objetivo es la minimización de las desviaciones del coste de la solución robusta respecto de los costes de cada escenario.

$$\begin{aligned} \text{Min} \quad & \sum_{\substack{(i)=1 \\ (i)=1}}^k (t_{(i)(i+1)} - t_{(i-1)(i)}) DC_{(i)}^2 \\ \text{s.a.} \quad & \sum_{i=1}^n y_i = 1 \end{aligned} \quad (5.1)$$

$$x_{ij} \leq \bar{\omega}_j y_i \quad i, j = 1, \dots, n \quad (5.2)$$

$$\sum_{\substack{i=1 \\ 0 \leq t \leq 1}}^n x_{ij} = \underline{\omega}_j + (\bar{\omega}_j - \underline{\omega}_j) t \quad j = 1, \dots, n \quad (5.3)$$

$$0 \leq t \leq 1 \quad (5.4)$$

$$\sum_{j=1}^n [d_{(i)j} (\underline{\omega}_j + (\bar{\omega}_j - \underline{\omega}_j) t_{(i)})] + DC_{(i)} = \sum_{i,j=1}^n d_{ij} x_{ij} \quad (i) = 1, \dots, k \quad (5.5)$$

$$t_{(i-1)(i)} \leq t_{(i)} \leq t_{(i)(i+1)} \quad i = 1, \dots, k \quad (5.6)$$

$$x_{ij} \geq 0, y_i \in \{0, 1\}$$

En este modelo, las variables $DC_{(i)}$ representan la desviación del coste total del modelo coordinado respecto del coste de los escenarios cuyo óptimo es el vértice (i) y la ecuación (5.5) recoge las desviaciones respecto del coste.

Modelo 6: su objetivo es una combinación de los objetivos de los dos modelos anteriores.

$$\begin{aligned} \text{Min } & C \sum_{\substack{(i)=1 \\ (i)=1}}^k (t_{(i)(i+1)} - t_{(i-1)(i)}) DD_{(i)}^2 + \\ & \sum_{\substack{(i)=1 \\ (i)=1}}^k (t_{(i)(i+1)} - t_{(i-1)(i)}) DC_{(i)}^2 \\ \text{s.a. } & \sum_{i=1}^n y_i = 1 \end{aligned} \quad (6.1)$$

$$x_{ij} \leq \bar{\omega}_j y_i \quad i, j = 1, \dots, n \quad (6.2)$$

$$\sum_{i=1}^n x_{ij} = \underline{\omega}_j + (\bar{\omega}_j - \underline{\omega}_j)t \quad j = 1, \dots, n \quad (6.3)$$

$$0 \leq t \leq 1 \quad (6.4)$$

$$\sum_{j=1}^n [d_{(i)j}(\underline{\omega}_j + (\bar{\omega}_j - \underline{\omega}_j)t_{(i)})] + DC_{(i)} = \sum_{i,j=1}^n d_{ij}x_{ij} \quad (i) = 1, \dots, k \quad (6.5)$$

$$\sum_{j=1}^n [\underline{\omega}_j + (\bar{\omega}_j - \underline{\omega}_j)t_{(i)}] + DD_{(i)} = \sum_{i,j=1}^n x_{ij} \quad (i) = 1, \dots, k \quad (6.6)$$

$$t_{(i-1)(i)} \leq t_{(i)} \leq t_{(i)(i+1)} \quad i = 1, \dots, k \quad (6.7)$$

$$x_{ij} \geq 0, y_i \in \{0, 1\}$$

Ejemplo 4.2.

Consideremos el árbol T del ejemplo 2.1 cuya matriz de distancias viene dada en el ejemplo 3.1, y las soluciones obtenidas están consignadas en las tablas 12.4, 12.5 y 12.6. Vemos que el comportamiento de los modelos 4, 5 y 6 responde a la misma actitud del decisor que los modelos 1, 2 y 3 respectivamente. No obstante, la información que el decisor tiene en este caso es mucho mayor que en el caso anterior y puede utilizarse para elegir adecuadamente uno de los modelos.

	Modelo 4	Modelo 5	Modelo 6
Localización óptima	Vértice 4	Vértice 7	Vértice 4
Coste óptimo	9213,403	9236,632	9236,632
Demanda total atendida	869,529	847,337	872,045

Tabla 12.4: Tabla de demanda atendida

	Demanda Mínima	Demanda atendida Modelo 4	Demanda atendida Modelo 5	Demanda atendida Modelo 6	Demanda Máxima
Vértice 1	0	5,613	5,071	5,675	9
Vértice 2	15	16,871	16,690	16,892	18
Vértice 3	16	24,732	23,888	24,828	30
Vértice 4	5	14,356	13,451	14,458	20
Vértice 5	225	240,593	239,085	240,764	250
Vértice 6	50	71,830	69,720	72,069	85
Vértice 7	65	80,593	79,085	80,764	90
Vértice 8	150	274,744	262,683	276,111	350
Vértice 9	45	60,593	59,085	60,764	70
Vértice 10	15	18,119	17,817	18,153	20
Vértice 11	31	36,613	36,071	36,675	40
Vértice 12	23	24,871	24,690	24,892	26

Tabla 12.5: Tabla soluciones

	Modelo 4	Modelo 5	Modelo 6
DD(2)	213,667	-	216,183
DD(4)	0	-	0
DD(7)	-105,016	-	-102,501
DC(2)	-	1996,167	1996,167
DC(4)	-	0	0
DC(7)	-	-946,458	-946,458

Tabla 12.6: Tabla de desviaciones

5 Conclusiones

En este trabajo hemos presentado el problema de la mediana sobre un árbol con incertidumbre en las demandas resuelto mediante técnicas de optimización robusta. Esta incertidumbre viene dada por intervalos, esto es, hemos establecido la hipótesis de que la demanda de cada vértice varía en un rango conocido de valores. Presentamos dos procedimientos para manejar los infinitos escenarios que aparecen. Cuando no conocemos cual es el comportamiento de cada demanda dentro de su rango posible de valores, el algoritmo ESCEGEN nos permite calcular todos los óptimos posibles y al menos un escenario asociado a cada uno. Si sabemos que el comportamiento de todas las demandas viene determinado por un único parámetro, podemos utilizar el algoritmo ESCEPAR para calcular todas las medianas y todos los escenarios asociados a cada una de ellas. En ambos casos los cálculos no dependen de las distancias.

Con estos datos hemos planteado tres modelos coordinados para cada caso. Nuestro criterio de robustez está perfectamente definido. Es un criterio en esencia biobjetivo puesto que trata de minimizar la demanda no atendida por la solución robusta al mismo tiempo que obtiene un coste cercano al coste mínimo que se hubiera obtenido de saber con antelación que escenario se iba a realizar. La forma de plantear matemáticamente este criterio en un modelo coordinado no es única. Los objetivos de cada modelo son, respectivamente, cubrirse respecto a todas las posibles demandas futuras, cubrirse respecto a todos los posibles costes futuros o cubrirse respecto a ambas situaciones. Por supuesto, el esfuerzo computacional asociado a cada uno de los modelos no es el mismo. Por ello, el decisor debe valorar su objetivo preferente frente al coste computacional para elegir el mejor modelo en cada caso concreto.

6 Bibliografía

- [1] Averbakh, I. y Berman, O. *Minmax Regret Median Location on a Network under Uncertainty*. *INFORMS Journal on Computing*, Vol. n 12, pp. 104-110, 2000.
- [2] Canós, M.J. y Mocholí, M. *Técnicas de optimización robusta aplicadas al problema de la p-mediana en condiciones de incertidumbre*. *Actas de las VI Jornadas de ASEPUA. Santiago de Compostela*, pp. 149-156, 1998.
- [3] Canós, M.J., Ivorra, C. y Liern, V. *An Exact Algorithm for the Fuzzy p-Median Problem*. *European Journal of Operational Research*, Vol. n 116, pp. 80-86, 1999.
- [4] Canós, M.J., Ivorra, C. y Liern, V. *The Fuzzy p-Median Problem: A Global*

- Analysis of the Solutions. European Journal of Operational Research, Vol. n 130*, pp. 430-436, 2001.
- [5] Canós, M.J., Ivorra, C. y Liern, V. *Finding Satisfactory Near-Optimal Solutions in Location Problems.* Verdegay, J. L. (editor), Fuzzy Sets Based Heuristics for Optimization, Springer, 2003.
 - [6] Chen, B. y Lin, C.S. *Minmax-Regret Robust 1-Median Location on a Tree. Networks.* Vol. n 31, pp. 93-103, 1998.
 - [7] Erkut, E. y Tansel, B.C. *On Parametric Medians of Trees. Transportation Science.* Vol. n 26, pp. 149-156, 1992.
 - [8] Goldman, A. J. *Optimal Center Location in Simple Networks. Transportation Science.* Vol. n 5, pp. 212-221, 1971.
 - [9] Hakimi, S. L. *Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph. Operations Research.* Vol. n 12, pp. 450-459, 1964.
 - [10] Hakimi, S. L. *Optimum Distribution of Switching Centers in a Communication Network and Some Related Graph Theoretic Problems. Operations Research.* Vol. n 13, pp. 462-475, 1965.
 - [11] Kariv, O. y Hakimi, S.L. *An Algorithmic Approach to Network Location Problems. Part 2: The p-Medians. SIAM Journal on Applied Mathematics* Vol. n 37, pp. 539-560, 1979.
 - [12] Kouvelis, P. y Yu, G.W. *Robust Discrete Optimization and its Applications.* Kluwer Academic Publishers, 1997.
 - [13] Mulvey, J.M., Vanderbei, R.J. y Zenios, S.A. *Robust Optimization of Large-Scale Systems. Operations Research.* Vol. n 43, pp. 264-281, 1995.
 - [14] Owen, S.H. y Daskin, M.S. *Strategic Facility Location: A Review. European Journal of Operational Research.* Vol. n 111, pp. 423-447, 1998.
 - [15] Serra, D. y Marianov, V. *The p-Median Problem in a Changing Network: the Case of Barcelona. Location Science.* Vol. n 6, pp. 383-394, 1998.

Solución numérica de problemas de control estocástico en economía

E. Domínguez^a, A. Novales^b, J. Pérez^c,
J. Ruiz^{b1}

^a Dep. de Fundamentos de Análisis Económico, Universidad Pública de Navarra,

^b Dep. Economía Cuantitativa, Universidad Complutense

^c Fundación CentrA y Universidad Pablo Olavide, Sevilla

e-mails anovales@ccee.ucm.es

1 Introducción

En los últimos años se ha venido adoptando como norma la especificación de modelos económicos en los que los distintos agentes económicos derivan sus reglas de comportamiento como solución a la optimización de una determinada función objetivo sujeta a unas restricciones. Tanto las funciones objetivo como el conjunto de restricciones son específicas de cada uno de los distintos agentes. Esta corriente trata de aportar mayor rigor formal a un concepto que siempre ha estado presente en la especificación de modelos de economías agregadas. Las distintas ecuaciones que integran estos modelos, denominadas función de consumo, inversión, etc., determinan los niveles de dichas variables como función de los valores numéricos de sus determinantes, que pueden ser variables como los precios, tipos de interés, etc.. Supuestamente, esto era consecuencia de un comportamiento optimizador por parte de los agentes económicos, aunque dicho comportamiento no se hacía explícito en el modelo. En el lenguaje de la teoría de optimización, los modelos

¹Los autores agradecen la financiación recibida del Ministerio de Ciencia y Tecnología a través del proyecto BEC2003-03965.

económicos constaban de ecuaciones que no eran sino el resultado de agregar las reglas de decisión de agentes análogos, como consumidores, empresas, etc., pero dejando implícita la especificación de sus funciones objetivo o de las restricciones a que estos agentes se enfrentan.

Cuando, por el contrario, se hace explícito el problema de optimización que resuelve cada unidad de decisión que constituye una economía, se distingue entre distintos tipos de agentes: consumidores/trabajadores, empresas y gobierno, cada uno de ellos con una función objetivo diferente, y sujeto a distintas restricciones. Cada agente decide los valores numéricos de sus variables de control, tomando sus variables de estado como dadas. Las variables de estado de un agente son variables que toma como dadas cuando toma sus decisiones. Las variables de decisión o de control de un agente pueden ser variables de estado para otro agente distinto: el gobierno puede decidir sobre la tasa de crecimiento de la oferta monetaria de manera que minimice una función objetivo que penaliza un nivel elevado de la tasa de inflación. Ese crecimiento monetario será una variable de estado para un consumidor o una empresa en la misma economía. Así, las variables de estado de un agente pueden ser variables exógenas para toda la economía, como sería en el caso español el nivel de los tipos de interés fijado por la Reserva Federal de EEUU. Pueden ser también variables de decisión de otros agentes de la misma economía, como ocurriría con la producción de un determinado bien, cuya cuantía es decidida por la empresa productora, y los consumidores toman como dada, siendo, por tanto, una variable de estado para ellos. Finalmente, las variables de estado pueden ser también sus propias decisiones pasadas.

A lo largo de este artículo, y exclusivamente por razones de simplicidad, vamos a centrarnos en una economía sencilla, en la que ignoraremos el efecto que sobre la economía puede tener la actividad del gobierno, a través de su política fiscal: tipos impositivos, emisión de deuda, etc., su política monetaria: tipos de interés, crecimiento de la cantidad de dinero en circulación, etc., y a través de su actividad como agente económico: gasto público, inversión pública, etc.. Además, suponemos, como es habitual, que los consumidores son propietarios de las empresas, de modo que las decisiones de consumo y las decisiones de inversión y producción se toman simultáneamente por parte del mismo agente. Todos los consumidores son idénticos, por lo que nos basta con considerar las decisiones que toma un consumidor representativo.

Dicho consumidor tratará de maximizar su nivel de utilidad agregada en el tiempo, que deriva del consumo de los distintos bienes, así como del nivel de ocio de que disfruta en cada período. Las restricciones a que se enfrenta este decisor, una para cada instante de tiempo, especificarán que la cantidad que paga por los bienes que consume no puede exceder de la renta de que dispone. En un contexto más amplio, los mercados de capitales existen para que el consumidor pueda reservar parte de su renta cada período en la forma de ahorro. De ese modo, un período puede decidir llevar a cabo un consumo cuyo valor de mercado

es inferior a la renta del período, ahorrando la renta no gastada. En otro período, podría suceder lo contrario, utilizando la renta de dicho instante, junto con parte del ahorro que arrastra de períodos anteriores, para financiar su nivel de consumo.

Problemas como los descritos tienen una estructura que no es lineal-cuadrática. Si lo fuese, la aplicación del principio de equivalencia cierta, que permite separar el problema de estimación del problema de control, facilitaría el cálculo de la solución de dicho problema, en sus versiones estocásticas. Al ser su estructura más compleja,² el problema carece de solución analítica incluso en su versión determinista. Es decir, es imposible encontrar reglas de decisión que resuelvan el problema, especificando el valor óptimo en cada período de cada una de las variables de decisión, en función de los valores que en dicho período toman las variables de estado (variables exógenas, y variables de control de períodos anteriores).

Al no existir solución analítica, en los últimos años se ha propuesto una variedad de procedimientos alternativos para obtener soluciones numéricas a este tipo de problemas de control. En este trabajo revisamos, a modo de introducción, las principales características de uno de dichos métodos,³ que se basa en la caracterización de las condiciones de estabilidad del sistema, es decir, de las ecuaciones que definen el subespacio de convergencia a un equilibrio estacionario o punto de reposo del mismo. El método, introducido por Sims (2001) basándose en trabajo previo de Blanchard y Kahn (1980) y extendido al caso no lineal cuadrático por Novales et al. (1999), caracteriza las condiciones de estabilidad de la aproximación lineal del modelo, puesto que se carece de procedimientos generales para caracterizar dichas condiciones en modelos no lineales. Posteriormente, la solución al modelo, en la forma de una serie temporal para cada una de las variables del mismo, se obtiene combinando las condiciones de estabilidad estimadas con la estructura analítica no lineal del modelo original. La imposición de condiciones de estabilidad para obtener soluciones numéricas no explosivas es precisa incluso en problemas deterministas. En problemas estocásticos es aún más necesaria, para garantizar que las perturbaciones estocásticas que afectan cada período al sistema no lo sitúen en una trayectoria divergente.

Por último, hay que recordar que un modelo de control estocástico no está completamente definido hasta que se establece un supuesto acerca del mecanismo de formación de expectativas utilizado por los agentes. Distintos supuestos acerca de estos mecanismos conducen a problemas de control diferentes, con soluciones que pueden tener características muy distintas. Aunque la tradición acerca de la

²Generalmente las funciones objetivo son más complejas que una función cuadrática. Sin embargo, el problema más importante es que, en los problemas más interesantes, en los que tanto el comportamiento de las variables cantidad como de los precios es determinado simultáneamente, las restricciones del problema no son lineales.

³El lector interesado puede consultar el volumen 8 de la revista *Journal of Business and Economic Statistics* (1990) dedicado especialmente a la descripción de estos métodos, o el libro *Computational Methods for the Study of Dynamic Economies*, de Oxford University Press, editado por A. Marimón y A. Scott.

manera en que los agentes económicos procesan la información disponible para formar sus expectativas es muy extensa, consideramos aquí únicamente el supuesto de racionalidad de expectativas, en el sentido de que los agentes utilizan eficientemente en todo momento la información de que disponen al formar sus expectativas. Ante la ambigüedad que implica suponer lo contrario, el supuesto de racionalidad de expectativas se ha convertido asimismo en el estándar a utilizar en el análisis de modelos económicos en los que se hace explícito el comportamiento optimizador de agentes que se enfrentan a situaciones de incertidumbre.

En la sección 2 describimos con todo detalle el modelo de crecimiento neoclásico más simple. En la sección 3 detallamos la caracterización de las condiciones de estabilidad en un contexto determinista, dejando para la sección 4 la caracterización de las condiciones de estabilidad en un contexto estocástico y bajo el supuesto de formación de expectativas racional. El procedimiento de solución es ilustrado en la sección 5 utilizando el modelo de crecimiento estocástico básico. Por último, el trabajo se cierra con un resumen.

2 Un modelo de crecimiento sencillo

Consideremos un consumidor en una economía simple, con un único bien. Las unidades de dicho bien disponibles en cada instante pueden consumirse o pueden ahorrarse. Las unidades que se ahorren contribuyen a incrementar el stock de capital de cada período. Éste experimenta una depreciación a una tasa igual a δ por lo que, incluso si el ahorro de cada período es positivo, el stock de capital puede disminuir si el nivel de ahorro no es suficientemente elevado como para compensar la pérdida por depreciación. El consumidor resuelve el problema de maximización de su utilidad agregada a través del tiempo,

$$\max_{c_t, s_t} \sum_{t=1}^{\infty} \beta^t U(c_t) \quad (13.1)$$

sujeto a la sucesión de restricciones,

$$\begin{aligned} c_t + s_t &= y_t, \quad t = 1, 2, \dots \\ y_t &= f(k_{t-1}), \quad t = 1, 2, \dots \\ k_t &= (1 - \delta)k_{t-1} + s_t, \quad t = 1, 2, \dots \\ c_t, s_t &\geq 0, \quad t = 1, 2, \dots \end{aligned}$$

y dado k_0

siendo c_t, s_t los niveles de consumo y ahorro de cada instante, y $f(\cdot)$ la función de producción que representa la tecnología disponible en cada instante, que genera en el período t unidades del bien de consumo a partir del stock de capital disponible

al inicio de dicho período, k_{t-1} . Aunque la decisión sobre el valor numérico de k_{t-1} se toma en el período $t-1$, a través de la decisión de ahorro de dicho período, s_{t-1} , el stock de capital k_{t-1} no es productivo hasta el período siguiente, t . Este supuesto es crucial, pues genera toda la dinámica del modelo. Este modelo, es conocido como el modelo de Cass-Koopmans, y en su versión estocástica, como modelo de Brock-Mirman.

Suponemos que la función de utilidad es monótona creciente y cóncava, con: $U' > 0, U'' < 0, U'(0) = \infty, \lim_{c \rightarrow \infty} U'(c) = 0$. En cuanto a la función de producción, suponemos asimismo monotonía y concavidad, con $f' > 0, f'' < 0, \lim_{k \rightarrow 0} f(k) = \infty$ y $\lim_{k \rightarrow \infty} f(k) = 0$. Un supuesto habitual sobre la forma funcional, que adoptamos en lo sucesivo es, $f(k_t) = Ak_t^\alpha$, $0 < \alpha < 1$, que satisface las condiciones descritas, donde A es un parámetro que representa el nivel de la tecnología y α es la elasticidad de la producción respecto del stock de capital. Puesto que $\alpha < 1$, un incremento en k_t en un factor μ genera un incremento en la producción en un factor inferior a μ , por lo que se dice que este tipo de funciones de producción presenta rendimientos decrecientes.

Sustituyendo la segunda restricción en la primera, tenemos, el Lagrangiano

$$L = \sum_{t=1}^{\infty} \beta^t U(c_t) + \lambda_t [f(k_{t-1}) - c_t - k_t + (1 - \delta) k_{t-1}],$$

cuyas condiciones de optimalidad son,

$$\begin{aligned} \beta^t U'(c_t) &= \lambda_t, \quad t = 1, 2, \dots, \\ \lambda_{t+1} f'(k_t) &= \lambda_t - (1 - \delta) \lambda_{t+1}, \quad t = 1, 2, \dots, \end{aligned}$$

de las que, eliminando el multiplicador de Lagrange, tenemos la conocida *regla de Keynes-Ramsey*,

$$\frac{U'(c_t)}{\beta U'(c_{t+1})} = f'(k_t) + (1 - \delta), \quad t = 1, 2, \dots,$$

que, junto con la secuencia de restricciones presupuestarias,

$$f(k_{t-1}) - c_t - k_t + (1 - \delta) k_{t-1}, \quad t = 1, 2, \dots$$

debe darnos las sucesiones $\{c_t, s_t\}_{t=1}^{\infty}$ que resuelven el problema de maximización de utilidad a lo largo del tiempo, a partir del stock de capital inicial, k_0 .

Una condición adicional es la llamada condición de transversalidad, que en este problema adopta la forma,

$$\lim_{\tau \rightarrow \infty} \beta^{t+\tau} U'(c_{t+\tau}) k_{t+\tau} = 0. \quad (13.2)$$

Esta condición se obtiene considerando la versión en tiempo finito (T períodos) del problema de optimización (13.1) y maximizando con respecto al último valor observado de la variable de estado, k_T . Al ser una condición del tipo Kuhn-Tucker, especifica que el producto del valor la variable de estado en dicho instante terminal, multiplicado por el valor de la derivada parcial correspondiente del Lagrangiano en dicho instante debe ser igual a cero,

$$\frac{\partial L_T}{\partial k_T} k_T = 0,$$

donde L_T denota el Lagrangiano de dicho problema en horizonte finito, T . Puesto que $\frac{\partial L_T}{\partial k_T} = \lambda_T$, utilizando la primera condición de optimalidad y pasando al límite, se obtiene (13.2). En un problema más general, habría tantas condiciones de transversalidad como variables de estado.

2.1 Estado estacionario y estabilidad

La solución al problema anterior consiste en un conjunto de trayectorias temporales que se reflejarían en un conjunto de series temporales, una para cada variable relevante del modelo, que evolucionan a partir de unas condiciones iniciales. En realidad, existe una condición inicial para cada variable de estado, que en el problema anterior es únicamente el stock de capital. Como tendremos ocasión de discutir posteriormente, lo que resulta verdaderamente importante es determinar los valores óptimos iniciales de las variables de control.

Dichas trayectorias temporales pueden converger a un punto estable, diverger hacia infinito, con signo positivo o negativo, o bien oscilar alrededor de un punto que nunca se alcanza. Para caracterizar las propiedades de estabilidad de un sistema, es conveniente introducir antes la noción de estado estacionario. Definimos estado estacionario como un conjunto de valores numéricos para las tasas de crecimiento de las variables del modelo en términos per capita tales que, si alguna vez se alcanzase, entonces el sistema ya nunca lo abandonaría. Puede probarse que, bajo el supuesto de rendimientos decrecientes en cada input de la función de producción, las únicas tasas de crecimiento que pueden mantenerse indefinidamente en una economía como la descrita son iguales a cero. En este caso particular, el estado estacionario se convierte en unos niveles para las variables per cápita del modelo tales que, una vez alcanzados, ya nunca se abandonarían.

Las cuestiones relevantes se refieren entonces al número de estados estacionarios posibles, y a sus características de estabilidad. De acuerdo con la definición, los estados estacionarios del modelo anterior se caracterizarían re-escribiendo el modelo sin subíndices temporales, es decir, $c_t = c_{t-1} = c^*$ y $k_t = k_{t-1} = k^*$,

$$\frac{1}{\beta} = f'(k^*) + (1 - \delta), \quad (13.3)$$

$$f(k^*) - c^* - k^* + (1 - \delta)k^*.$$

La primera ecuación, que proviene de la igualdad entre la relación marginal de sustitución y la productividad marginal del capital, nos proporciona el nivel de estado estacionario del stock de capital y , una vez conocido éste, la segunda ecuación, que proviene de la restricción de recursos, nos proporciona el nivel de estado estacionario del consumo. Pueden existir múltiples estados estacionarios (*indeterminación global del equilibrio*), dependiendo de las propiedades de la función de producción. Sin embargo, es fácil ver que las condiciones antes descritas acerca de la función f garantizan que la solución a (13.3) existe siempre y es única. Por tanto, bajo tales supuestos existe un único estado estacionario. Pero, por supuesto, podría darse el caso de que existe un estado estacionario y sin embargo, la economía evoluciona en el tiempo sin dirigirse hacia él o que exista un estado estacionario y que se converja hacia él a través de infinitas sendas de equilibrio (*indeterminación local del equilibrio*). Que ocurra alguna de estas circunstancias o no dependerá de las condiciones de estabilidad.

En problemas localmente estables, las condiciones de transversalidad ayudan a seleccionar, de entre todas las trayectorias temporales que satisfacen las condiciones de optimalidad, aquella que constituye el punto de silla. La medida en que las condiciones de trasversalidad son restrictivas determina las características de estabilidad del problema. En problemas localmente estables, las condiciones de transversalidad no eliminan trayectorias, por lo que existen múltiples, quizás infinitas, trayectorias, que satisfacen las condiciones de optimalidad. En problemas localmente inestables, las condiciones de transversalidad eliminan todas las trayectorias que satisfacen las condiciones de optimalidad, por lo que no existe ninguna trayectoria óptima; en particular, no se produce la convergencia a ningún estado estacionario.

3 Condiciones de estabilidad: caso determinista

Contrariamente a lo que suele pensarse frecuentemente, incluso en modelos deterministas es preciso imponer condiciones de estabilidad adecuadas para garantizar que las condiciones de transversalidad se cumplan, como ilustramos en esta sección. Al igual que ocurre en los problemas de control estocástico, dichas condiciones están determinadas por los autovectores por la izquierda asociados a autovalores inestables de la matriz de transición en la aproximación lineal del modelo.

El problema de crecimiento económico de Brock-Mirman antes descrito tiene estructura de punto de silla. Como ya hemos visto, presenta un único estado estacionario (c^*, k^*) , y hay un subespacio bien determinado en el plano (c_t, k_t)

desde el que la economía converge a (c^*, k^*) , divergiendo hacia valores negativos (no admisibles) de alguno de los inputs en caso contrario.⁴ Ello significa que, dado k_0 , una elección inicial de c_0 distinta de la que corresponde a k_0 sobre el subespacio estable, hará que la economía se aleje del estado estacionario. Pero, además, como muestra la condición de estabilidad, la optimalidad de la trayectoria requiere que la economía converja a su estado estacionario, por lo que estabilidad y optimalidad son dos caras de la misma moneda en este tipo de problemas.

En el caso del modelo más sencillo, sin considerar el ocio como un argumento en la función de utilidad, supongamos que ésta adopta la forma,⁵ $U(c_t) = \frac{c_t^{1-\sigma}-1}{1-\sigma}$, $\sigma > 0$, por lo que las condiciones de optimalidad son:

$$\frac{c_t^\sigma}{c_{t+1}^\sigma} = \beta [f'(k_t) + (1 - \delta)], \quad (13.4)$$

$$k_t = (1 - \delta)k_{t-1} + f(k_{t-1}) - c_t, \quad (13.5)$$

$$\lim_{\tau \rightarrow \infty} \beta^{t+\tau} c_{t+\tau}^{-\sigma} k_{t+\tau} = 0.$$

Las dos primeras ecuaciones pueden aproximarse linealmente alrededor de los valores de estado estacionario del consumo y el stock de capital, c^* , k^* , obteniendo el siguiente sistema de ecuaciones en diferencias lineal:

$$\begin{pmatrix} k_t - k^* \\ c_t - c^* \end{pmatrix} = \begin{pmatrix} \frac{1}{\beta} & -1 \\ -\frac{1}{\beta}M & 1+M \end{pmatrix} \begin{pmatrix} k_{t-1} - k^* \\ c_{t-1} - c^* \end{pmatrix} = A \begin{pmatrix} k_{t-1} - k^* \\ c_{t-1} - c^* \end{pmatrix}, \quad (13.6)$$

donde

$$M = -\frac{1}{\sigma} \beta A \alpha (\alpha - 1) c^* k^{*\alpha-2} > 0,$$

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} \frac{1}{\beta} & -1 \\ -\frac{1}{\beta}M & 1+M \end{pmatrix}.$$

Utilizando la descomposición espectral de la matriz A , $A = \Gamma \Lambda \Gamma^{-1}$, donde Γ^{-1} tiene por filas los autovectores por la izquierda de A ,⁶ podemos representar

⁴La discusión de este punto, que queda lejos del ámbito de este artículo puede verse en cualquier libro de crecimiento, como Novales y Sebastián (1999).

⁵Esta es la función de utilidad de elasticidad de sustitución intertemporal del consumo constante ($1/\sigma$).

⁶Los autovectores por la derecha son: $(x_1, x_2) = (1, \frac{\lambda_1 - a_{11}}{a_{12}})$ y $(y_1, y_2) = (1, \frac{\lambda_2 - a_{11}}{a_{12}})$, y la matriz inversa:

$$\begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix} = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}^{-1} = \frac{1}{x_1 y_2 - x_2 y_1} \begin{pmatrix} y_2 & -y_1 \\ -x_2 & x_1 \end{pmatrix}$$

la dinámica de la solución al problema de control a partir de valores iniciales c_0, k_0 :

$$\begin{aligned} \begin{pmatrix} k_t - k^* \\ c_t - c^* \end{pmatrix} &= \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix} \begin{pmatrix} k_{t-1} - k^* \\ c_{t-1} - c^* \end{pmatrix} \\ &= \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} \begin{pmatrix} \lambda_1^t & 0 \\ 0 & \lambda_2^t \end{pmatrix} \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix} \begin{pmatrix} k_0 - k^* \\ c_0 - c^* \end{pmatrix} \end{aligned} \quad (13.7)$$

Que este modelo tiene estructura de punto de silla es consecuencia de que una de las raíces de la ecuación, digamos que λ_1 , es mayor que 1 en valor absoluto, mientras que la otra raíz, λ_2 , es inferior⁷ a 1⁸.

La matriz producto en la expresión anterior es:

$$\begin{aligned} \begin{pmatrix} k_t - k^* \\ c_t - c^* \end{pmatrix} &= \\ &= \begin{pmatrix} x_1 \lambda_1^t [u_1(k_0 - k^*) + v_1(c_0 - c^*)] + y_1 \lambda_2^t [u_2(k_0 - k^*) + v_2(c_0 - c^*)] \\ x_2 \lambda_1^t [u_1(k_0 - k^*) + v_1(c_0 - c^*)] + y_2 \lambda_2^t [u_2(k_0 - k^*) + v_2(c_0 - c^*)] \end{pmatrix}, \end{aligned}$$

por lo que la condición de transversalidad en el stock de capital se cumplirá sólo si el coeficiente asociado a λ_1 es igual a cero. Pero x_1 depende de los valores de los parámetros estructurales, de modo que es el término en paréntesis que acompaña a λ_1^t quien ha de ser igual a cero. Dicha condición es la misma para la ecuación del stock de capital que para la ecuación del consumo: $u_1(k_0 - k^*) + v_1(c_0 - c^*) = 0$, de modo que estabilidad requiere que el consumo inicial se escoja de modo que:

⁷Un sistema 2×2 con ecuación característica,

$$\lambda^2 - \left(1 + \frac{1}{\beta} + M\right)\lambda + \frac{1}{\beta} = 0$$

tiene estructura de punto de silla puesto que esta ecuación tiene un valor propio mayor que $1/\beta$, y otro menor que 1. Dichas raíces han de satisfacer,

$$\lambda_1 + \lambda_2 = 1 + \frac{1}{\beta} + M, \quad \lambda_1 \lambda_2 = \frac{1}{\beta}$$

de modo que,

$$\lambda_1 + \frac{1}{\beta \lambda_1} = 1 + \frac{1}{\beta} + M$$

La función $f(\lambda_1) = \lambda_1 + \frac{1}{\beta \lambda_1}$ tiene un mínimo en $\lambda_1 = \sqrt{1/\beta}$, y toma valor $f(\lambda_1) = 1 + \frac{1}{\beta}$ en $\lambda_1 = 1$ y en $\lambda_1 = \frac{1}{\beta}$. Como $1 + \frac{1}{\beta} + M > 1 + \frac{1}{\beta}$, entonces los dos posibles valores para λ_1 están no por debajo de 1, y el otro por encima de $\frac{1}{\beta}$.

⁸La tasa de crecimiento crítica, por debajo de la cual la solución es estable, es específica de cada modelo. El requisito para la existencia de una solución bien definida es que la función objetivo sea acotada, lo que requiere imponer cotas superiores sobre la velocidad de crecimiento de sus argumentos. Dichas cotas dependerán de la forma funcional de la función objetivo.

$$c_0 - c^* = -(k_0 - k^*) \frac{u_1}{v_1} = (k_0 - k^*) \frac{y_2}{y_1} = (k_0 - k^*) \frac{\lambda_2 - a_{11}}{a_{12}},$$

lo que implica, a partir de dicho instante:

$$\begin{aligned} k_t - k^* &= y_1 \lambda_2^t (u_2(k_0 - k^*) + v_2(c_0 - c^*)), \\ c_t - c^* &= y_2 \lambda_2^t (u_2(k_0 - k^*) + v_2(c_0 - c^*)) = \frac{y_2}{y_1} (k_t - k^*) = \frac{\lambda_2 - a_{11}}{a_{12}} (k_t - k^*), \end{aligned}$$

de modo que en cada período a partir de $t = 0$ se cumplirá la misma condición entre las desviaciones respecto de los valores de estado estacionario del stock de capital y el consumo. Esta es la representación lineal aproximada del subespacio estable del problema propuesto. Es precisamente el hecho de que la condición de estabilidad se cumplirá en cada período t , lo que permite que el modelo pueda resolverse utilizando dicha condición y una de las condiciones de primer orden del problema de optimización (13.4), (13.5).

3.1 Elección inicial de variables de control

La condición de estabilidad anterior puede escribirse como el producto interior: $(y_2, -y_1) (k_0 - k^*, c_0 - c^*)' = 0$, donde $(y_2, -y_1)$ es el autovector por la izquierda asociado a la raíz inestable, λ_1 , de A . Por tanto, en modelos deterministas, las condiciones de estabilidad seleccionan los valores iniciales de las variables de decisión, dados los valores iniciales de las variables de estado. Un número menor de condiciones de estabilidad que de variables de decisión⁹ llevará a resolver numéricamente el modelo como función de un valor inicial arbitrario de una o más variables de decisión, obteniendo así un continuo de trayectorias todas convergiendo al estado estacionario, por lo que la solución es localmente indeterminada, en el sentido de Benhabib y Perli (1994) y Xie (1994).

El sistema carece de solución cuando hay más condiciones de estabilidad independientes que variables de control. El subespacio estable se reduce entonces al estado estacionario, por lo que la economía será globalmente inestable, iniciando trayectorias divergentes tan pronto como experimente cualquier mínima desviación del estado estacionario. Finalmente, la solución será única cuando el conjunto de condiciones de estabilidad pueda utilizarse para representar las variables de control de forma única como función de las variables de estado.

El análisis de las condiciones de estabilidad en el caso estocástico es muy similar, como vamos a ver en la sección siguiente.

⁹Después de utilizar las ecuaciones que contienen sólo variables de control contemporáneas para eliminar algunas variables de control del problema.

4 Condiciones de estabilidad: el caso estocástico

Sims (2001) generalizó el trabajo de Blanchard and Kahn (1980) en varias direcciones, proponiendo una discusión general acerca del cálculo de la solución a un modelo lineal, estocástico bajo expectativas racionales. Dicho modelo puede expresarse:

$$\Gamma_0 y_t = \Gamma_1 y_{t-1} + C + \Psi z_t + \Pi \eta_t, \quad (13.8)$$

donde C es un vector de constantes. y_t es un vector que incluye:

- las variables determinadas por el modelo, denominadas variables endógenas, como el consumo, la producción, el stock de capital, etc.; incluye las esperanzas condicionales que aparecen en el modelo, redefinidas como nuevas variables, pero no los errores de expectativas asociados. Se incluyen en y_t las variables con subíndice más adelantado, mientras que las que tengan subíndices anteriores se incluirán en y_{t-1} ,
- las variables que son exógenas a los agentes pero siguen leyes de movimiento conocidas; ejemplos de este último tipo serían algunas variables de política, como el crecimiento monetario o los tipos impositivos, o también algunas perturbaciones exógenas, como la perturbación que introduciremos en la sección siguiente en la productividad de los factores.¹⁰

El vector z_t contiene las innovaciones que aparecen en las leyes de evolución temporal de las variables exógenas. Por ejemplo, el supuesto habitual acerca de la ley de movimiento de una posible perturbación en productividad: $\log(\theta_t) = \rho \log(\theta_{t-1}) + \epsilon_t$, conducirá a que un componente de y_t sea $\log(\theta_t)$, mientras que ϵ_t sea un componente de z_t . Finalmente, el vector η_t contiene los errores de expectativas racionales, que satisfacen $E_t(\eta_{t+1}) = 0$, y cuyo valor numérico se obtendrá, junto con el de las demás variables, como parte de la solución del modelo.

Modelos que incluyen más de un retardo temporal pueden incluirse en la formulación anterior añadiendo como nuevas variables dentro del vector y algunos retardos de variables ya incluidas en el modelo. También pueden incorporarse variables de expectativas adicionales de modo que todos los errores de expectativas que resulten correspondan a previsiones un período hacia el futuro. Lo mismo puede hacerse si el modelo incluye expectativas a horizontes superiores a un período.

Tomando condiciones iniciales arbitrarias y_0 y realizaciones muestrales para las variables z_t en (13.8), es inmediato obtener trayectorias para y_t . Sin embargo,

¹⁰Las variables en z_t son independientes: si dos perturbaciones exógenas están correlacionadas, la aproximación lineal a su relación ha de añadirse al sistema. Una de las variables formaría parte del vector z_t , mientras que la otra entraría a formar parte del vector y_t .

tales trayectorias incumplirán las condiciones de transversalidad, a menos de que se incorporen al sistema las condiciones de estabilidad apropiadas, que están definidas mediante los autovectores asociados a los autovalores inestables de las matrices en la representación (13.8). Cuando Γ_0 es invertible, son los autovalores inestables de la matriz $\Gamma_0^{-1}\Gamma_1$ quienes caracterizan las condiciones de estabilidad, mientras que cuando Γ_0 es singular, hemos de caracterizar los autovalores generalizados inestables del par (Γ_0, Γ_1) .

Novales et al. (1999) extienden el procedimiento anterior a la solución de problemas de control no lineales. Para ello, proponen calcular la aproximación lineal alrededor del estado estacionario del conjunto de ecuaciones no lineales que componen el modelo. Como consecuencia, el vector de constantes C resulta ser cero. Después de redefinir variables de modo adecuado, las matrices Γ_0 y Γ_1 que aparecen en la aproximación lineal al modelo original, contienen: i) las derivadas parciales de cada ecuación del sistema con respecto a cada una de las variables en y_t , evaluadas en estado estacionario, y ii) filas de unos y ceros, correspondiendo a variables intermedias que se han añadido al sistema para dotarle de una estructura autorregresiva de primer orden en presencia de retardos de orden superior o expectativas a más de un período hacia el futuro, o expectativas formadas en períodos distintos de t . En definitiva, (13.8) es la aproximación lineal al sistema formado por las condiciones de optimalidad del problema de control, las restricciones del mismo, las reglas de política económica exógenas al modelo, así como las leyes de movimiento de las perturbaciones también exógenas al modelo, estando todas las variables en desviaciones respecto de sus valores de estado estacionario.

Las condiciones de estabilidad se obtienen a partir de esta aproximación lineal, pero es la estructura original, no lineal del modelo la que es finalmente utilizada en el cálculo de la solución numérica. Con ello, se obtiene una realización de serie temporal para cada una de las variables del modelo, incluidas cada una de las expectativas condicionales que en él aparecen, y cada uno de los errores de expectativas asociados, por lo que es posible estimar cualquier propiedad estadística de este amplio vector de variables.

Conviene puntualizar que aunque estamos describiendo el procedimiento de solución dentro del marco de un modelo económico en que los agentes resuelven problemas de control explícitos, conviene poner énfasis en que el mismo procedimiento puede utilizarse para obtener soluciones aproximadas a cualquier conjunto de ecuaciones en diferencia estocásticas no lineales.

5 Solución del modelo de crecimiento estocástico básico

Ilustramos en esta sección el procedimiento general descrito en la sección previa mediante una aplicación al modelo estocástico más básico de crecimiento,

propuesto por Brock y Mirman (1972). La matriz Γ_0 que aparece en la aproximación lineal a este modelo resulta ser invertible, por lo que el cálculo de la solución numérica es bastante sencilla. El modelo supone que el consumidor deriva utilidad únicamente del nivel de consumo que realiza. Incorpora asimismo una perturbación aleatoria en la función de producción,

$$y_t = f(\theta_t, k_{t-1}) = \theta_t A k_{t-1}^\alpha, \quad 0 < \alpha < 1,$$

donde la perturbación aleatoria θ_t toma valores de acuerdo con una distribución de probabilidad centrada alrededor de 1. Valores por encima de dicha referencia hacen que la cantidad de producto exceda de lo que cabría esperar, en ausencia de incertidumbre, de las cantidades utilizadas de los inputs, mientras que valores de θ_t por debajo de 1 hacen que la cantidad de producto sea inferior a lo esperado. Se dice que θ_t afecta a la productividad de los factores productivos. De hecho, la productividad es la contribución al nivel de producto de una unidad adicional del input, k_t , por lo que se mide por la derivada parcial de la función f en la que θ_t aparece como una factor de escala. Nótese que, si bien k_t es conocido en t , $f(k_t) = \theta_{t+1} k_t^\alpha$ no se conoce hasta el período $t + 1$.

Dado k_0 , el consumidor representativo escoge sucesiones $\{k_t, c_t\}_{t=1}^\infty$ que resuelvan el problema,

$$\max_{\{k_t, c_t\}_{t=1}^\infty} E_0 \sum_{t=1}^{\infty} \beta^{t-1} \frac{c_t^{1-\sigma} - 1}{1 - \sigma}, \quad \sigma > 0 \quad (13.9)$$

sujeto a las restricciones

$$-c_t - k_t + (1 - \delta)k_{t-1} + \theta_t k_{t-1}^\alpha = 0, \quad (13.10)$$

$$-\log(\theta_t) + \rho \log(\theta_{t-1}) + \epsilon_t = 0, \quad \epsilon_t \text{ i.i.d.} \sim N(0, \sigma_\epsilon), \quad |\rho| < 1, \quad (13.11)$$

y dadas las condiciones iniciales, k_0, θ_0 . La primera es la restricción global de recursos en la economía, que estipula que la suma del consumo y la inversión bruta es igual a los recursos disponibles mediante la actividad de producción del período. La segunda especifica una evolución temporal para la perturbación en la función habitualmente aceptada, y es que la perturbación de productividad evoluciona suavemente, es decir, con alta autocorrelación positiva, a lo largo del tiempo.

Después de formar el Lagrangiano, tomar derivadas parciales respecto a las variables de decisión y los multiplicadores, y eliminar estos últimos, las condiciones de optimalidad resultan,

$$c_t = \theta_t k_{t-1}^\alpha - k_t + (1 - \delta)k_{t-1}, \quad (13.12)$$

$$c_t^{-\sigma} = \beta E_t [c_{t+1}^{-\sigma} ((1 - \delta) + \alpha \theta_{t+1} k_t^{\alpha-1})], \quad (13.13)$$

junto con la ley de movimiento de la perturbación estocástica del modelo $\log(\theta_t) = \rho \log(\theta_{t-1}) + \epsilon_t$, ϵ_t i.i.d. $\sim N(0, \sigma_\epsilon^2)$, la condición de transversalidad

$$\lim_{\tau \rightarrow \infty} E_t [c_{t+\tau}^{-\sigma} k_{t+\tau} \beta^\tau] = 0,$$

y las condiciones iniciales k_0, θ_0 .

En este punto es donde se hace preciso incorporar al modelo algún supuesto específico acerca del modo en que los agentes forman sus expectativas. Vamos a considerar en lo sucesivo que dichas expectativas son racionales, lo que equivale a afirmar que, al formar expectativas, los agentes hacen uso eficiente de la información de que disponen. Ello significa que los errores de expectativas no tienen nada en común con información que era disponible en el momento de construir la expectativa, lo que en términos estadísticos, se traduce en ausencia de correlación entre el error de expectativas y cualquier variable conocida al formar la expectativa. Si denotamos por W_t la esperanza condicional que aparece en (13.13), $W_t = E_t [c_{t+1}^{-\sigma} (1 - \delta) + \alpha \theta_{t+1} k_t^{\alpha-1}]$, e introducimos el correspondiente error de expectativas, η_t , $\eta_t = c_t^{-\sigma} [(1 - \delta) + \alpha \theta_t k_{t-1}^{\alpha-1}] - W_{t-1}$, deberemos tener, una vez resuelto numéricamente el modelo, $E(\eta_t X_t) = 0$, siendo X_t cualquier variable medible respecto al sigma-álgebra sobre el cual se define el operador esperanza condicional E_t .¹¹

Con esta notación podemos escribir el sistema anterior como:

$$0 = -W_t + E_t [c_{t+1}^{-\sigma} (1 - \delta) + \alpha \theta_{t+1} k_t^{\alpha-1}], \quad (13.14)$$

$$0 = -c_t^{-\sigma} + \beta W_t, \quad (13.15)$$

$$0 = -W_{t-1} + c_t^{-\sigma} [(1 - \delta) + \alpha \theta_t k_{t-1}^{\alpha-1}] - \eta_t, \quad (13.16)$$

donde $E_t[\eta_{t+1}] = 0$, junto con (13.12).

Las condiciones que caracterizan el estado estacionario son:

$$\begin{aligned} c^* &= \theta^* k^{*\alpha} - k^* + (1 - \delta)k^*, \quad W^* = c^{*-\sigma} \left[(1 - \delta) + \alpha \theta^* (k^*)^{\alpha-1} \right], \\ 0 &= -c^{*-\sigma} + \beta W^*. \end{aligned}$$

donde $\theta^* = 1$, lo que conduce a, $k^* = \left(\frac{\alpha \beta}{1 - (1 - \delta) \beta} \right)^{\frac{1}{1-\alpha}}$, $c^* = k^{*\alpha} - \delta k^*$, $W^* = \frac{1}{\beta} (c^*)^{-\sigma}$, $y^* = k^{*\alpha}$.

El sistema que hemos de aproximar linealmente consiste en las ecuaciones (13.11), (13.15), (13.12) y (13.16). Las variables de estado son: k_{t-1} y $\log(\theta_t)$, mientras que las variables de decisión son: c_t y k_t . Para linealizar, consideramos

¹¹Nótese que nuestra notación $E_t(z_{t+1})$ es equivalente a $E(z_{t+1}/\Omega_t)$ siendo Ω_t , $t = 1, 2, \dots$ una sucesión de sigma-álgebras con $\Omega_t \subseteq \Omega_{t+1} \forall t$. Puede interpretarse que cada sigma-álgebra Ω_t es la generada por el espacio de sucesos observables hasta el instante t .

cada ecuación como una función: $f(c_t, k_t, W_t, \log(\theta_t), \eta_t, \epsilon_t) = 0$ y, en línea con la notación de la sección anterior, definimos los vectores $y_t = (c_t - c^*, k_t - k^*, W_t - W^*, \log(\theta_t))'$, $\eta_t = (\eta_t)$, $z_t = (\epsilon_t)$, siendo estos dos últimos de dimensión 1×1 . La aproximación de primer orden alrededor del estado estacionario es:

$$\frac{\partial f}{\partial y_t}|_{ss} y_t + \frac{\partial f}{\partial y_{t-1}}|_{ss} y_{t-1} + \frac{\partial f}{\partial \eta_t}|_{ss} \eta_t + \frac{\partial f}{\partial \epsilon_t}|_{ss} \epsilon_t = 0,$$

donde los valores de estado estacionario de η_t y ϵ_t son cero. Alineando verticalmente estas aproximaciones, podemos escribir el sistema linearizado como:

$$\Gamma_0 y_t = \Gamma_1 y_{t-1} + \Psi z_t + \Pi \eta_t, \quad (13.17)$$

donde:

$$\begin{aligned} \Gamma_0 &= \begin{pmatrix} -\sigma(c^*)^{-\sigma-1} & 0 & -\beta & 0 \\ \sigma(c^*)^{-\sigma-1}(\alpha(k^*)^{\alpha-1} + 1 - \delta) & 0 & 0 & -\alpha c^{*\sigma} (k^*)^{\alpha-1} \\ 1 & 1 & 0 & -k^{*\alpha} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \\ \Gamma_1 &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \alpha(\alpha-1)c^{*\sigma}(k^*)^{\alpha-2} & -1 & 0 \\ 0 & \alpha(k^*)^{\alpha-1} + 1 - \delta & 0 & 0 \\ 0 & 0 & 0 & \rho \end{pmatrix}, \Psi = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \Pi = \begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \end{pmatrix}. \end{aligned}$$

5.1 Caracterización de las condiciones de estabilidad

El término constante en (13.17) es cero, puesto que las variables recogidas en y_t están expresadas en desviaciones respecto de sus valores de estado estacionario. Además, para valores paramétricos razonables, Γ_0 resulta ser invertible. Multiplicando por la izquierda por la inversa de Γ_0 , obtenemos un sistema transformado en el que la matriz de coeficientes y_t es la identidad y, después de redefinir adecuadamente las matrices ($\tilde{\Gamma}_1 = \Gamma_0^{-1}\Gamma_1$; $\tilde{\Psi} = \Gamma_0^{-1}\Psi$; $\tilde{\Pi} = \Gamma_0^{-1}\Pi$), tenemos:

$$y_t = \Gamma_0^{-1}\Gamma_1 y_{t-1} + \Gamma_0^{-1}\Psi z_t + \Gamma_0^{-1}\Pi \eta_t = \tilde{\Gamma}_1 y_{t-1} + \tilde{\Psi} z_t + \tilde{\Pi} \eta_t. \quad (13.18)$$

La matriz $\tilde{\Gamma}_1$ admite una descomposición de Jordan¹²: $\tilde{\Gamma}_1 = P \Lambda P^{-1}$, donde P es la matriz que tiene por columnas los autovectores por la derecha de $\tilde{\Gamma}_1$, y Λ tiene los valores propios de $\tilde{\Gamma}_1$ en su diagonal principal, y ceros en el resto de coordenadas¹³. La matriz inversa de P , P^{-1} , tiene por filas los vectores propios

¹²Para cuyo cálculo existe una función o rutina en casi todo programa de cálculo numérico.

¹³Consideramos únicamente el caso en que todos los valores propios son distintos entre sí. Sims (2001) discute la extensión al caso de autovalores múltiples, en modelos lineales.

por la izquierda de $\tilde{\Gamma}_1$. Multiplicando este sistema por P^{-1} y definiendo $w_t = P^{-1}y_t$, tenemos:

$$w_t = \Lambda w_{t-1} + P^{-1} (\tilde{\Psi} z_t + \tilde{\Pi} \eta_t), \quad (13.19)$$

que es un sistema en combinaciones lineales de las variables que componen el vector original y_t . Tendremos una ecuación correspondiente a cada valor propio λ_j de $\tilde{\Gamma}_1$:

$$w_{jt} = \lambda_{jj} w_{j,t-1} + P^{j\bullet} (\tilde{\Psi} z_t + \tilde{\Pi} \eta_t), \quad (13.20)$$

donde $P^{j\bullet}$ denota la fila j -ésima de P^{-1} .

Es fácil ver que autovalores λ_{jj} mayores que uno en valor absoluto generarán un comportamiento explosivo en la variable w_{jt} excepto si:

$$w_{jt} = P^{j\bullet} y_t = 0, \quad \forall t, \quad (13.21)$$

lo que genera una condición de estabilidad en la forma de una condición de ortogonalidad entre el autovalor por la izquierda asociado a un autovalor inestable de la matriz $\tilde{\Gamma}_1 = \Gamma_0^{-1}\Gamma_1$ y el vector de variables y_t , en desviaciones respecto de los valores de estado estacionario.

5.2 Cálculo de la solución numérica

Para valores paramétricos como los habitualmente utilizados en modelos económicos, $\sigma = 1.5$, $\delta = 0.025$, $\alpha = 0.36$, $\beta = 0.99$, $\rho = 0.95$, tenemos el estado estacionario:

$$y^* = 3.7041; c^* = 2.7543; k^* = 37.9893$$

y las matrices:

$$\Gamma_0 = \begin{pmatrix} -0.1191 & 0 & -0.99 & 0 \\ 0.1203 & 0 & 0 & -0.0077 \\ 1 & 1 & 0 & -3.7041 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \Gamma_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -0.0001 & -1 & 0 \\ 0 & 1.0101 & 0 & 0 \\ 0 & 0 & 0 & 0.95 \end{pmatrix}.$$

La matriz $\tilde{\Gamma}_1 = \Gamma_0^{-1}\Gamma_1$ admite una descomposición $\tilde{\Gamma}_1 = P\Lambda P^{-1}$, con matrices P, Λ, P^{-1} :

$$P = \begin{pmatrix} 1 & -0.0277 & -0.0388 & -0.0318 \\ 0 & 0.9996 & -0.9992 & -0.9995 \\ 0 & 0.0033 & 0.0047 & 0.0038 \\ 0 & 0 & 0 & 0.0085 \end{pmatrix}, \Lambda = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1.0388 & 0 & 0 \\ 0 & 0 & 0.9723 & 0 \\ 0 & 0 & 0 & 0.95 \end{pmatrix},$$

$$P^{-1} = \begin{pmatrix} 1 & 0 & 8.3097 & 0 \\ 0 & 0.5842 & 124.9696 & 12.4795 \\ 0 & -0.4163 & 125.0167 & -105.3230 \\ 0 & 0 & 0 & 117.7817 \end{pmatrix}.$$

La condición de estabilidad viene dada por la fila de la matriz P^{-1} , asociada al único autovalor mayor que 1:

$$w_{2t} = P^{2\bullet} y_t = 0 \quad \forall t \Rightarrow W_t - W^* + 0.00467(k_t - k^*) + 0.09986 \log(\theta_t) = 0. \quad (13.22)$$

Una única condición de estabilidad es, de hecho, lo que cabía esperar de la discusión teórica que llevamos a cabo en la sección 3 al haber una única variable de control en el modelo. Además, hay una única esperanza condicional en el modelo, por lo que una condición de estabilidad es suficiente para identificarla.

La diferencia con el caso determinista es que la condición de estabilidad no garantiza que la condición estocástica de optimalidad en la que aparece dicha expectativa, (13.13) se satisfaga en todos los períodos, debido a la existencia del error de expectativas. El papel de esta ecuación, una vez que la escribimos en la forma (13.15-13.16), es precisamente proporcionar la realización del error de expectativas¹⁴. La condición de estabilidad impone una relación exacta entre el error de expectativas racionales η_t y la innovación en la perturbación de productividad, ϵ_t . La única razón por la que los agentes cometan errores de predicción es porque el sistema está sujeto a dichas perturbaciones, y la relación entre ambos tipos de variables es una característica importante del modelo que no se obtiene con otros procedimientos de solución numérica.

Para calcular soluciones estables, se añade la condición de estabilidad estimada (13.22) al modelo no lineal original, lo que permite resolver valores numéricos para las variables endógenas, así como para los errores de expectativas. El procedimiento de solución comienza utilizando la ley de movimiento de la perturbación en productividad (13.11) para generar una serie temporal para dicha variable θ_t , de la longitud muestral deseada, a partir de realizaciones independientes de la innovación ϵ_t y de un valor inicial θ_0 . Este puede tomarse igual a su esperanza matemática, cero, o puede escogerse arbitrariamente. Al ser esta variable exógena respecto a todas las demás variables del modelo, puede simularse su realización temporal sin necesidad de conocer la evolución temporal de ninguna otra variable. A partir de la condición inicial sobre k_0 , utilizamos la restricción presupuestaria para expresar k_1 como función de c_1 . Llevando esta función a la condición de estabilidad, junto con (13.15) y el valor numérico obtenido previamente para θ_1 , tenemos una única ecuación no lineal en c_1 . Una vez calculado c_1 , invertimos la función anterior para obtener k_1 . Se repite el mismo proceso en cada período,

¹⁴Cuyo valor numérico no será sino una aproximación al verdadero error de expectativas, puesto que incorporará el error numérico derivado de haber estimado el subespacio estable en la aproximación lineal del modelo.

hasta cubrir la longitud muestral deseada. Una vez que tengamos series temporales para el consumo y el stock de capital, podemos generar la serie temporal de W_t de (13.15), y la correspondiente al error de expectativas η_t , a partir de (13.16).

A partir de (13.20), es claro que imponer la condición de estabilidad: $w_{jt} = 0 \quad \forall t$ es equivalente a:

$$P^{j\bullet} (\tilde{\Psi} z_t + \tilde{\Pi} \eta_t) = 0 \quad \forall t, \quad (13.23)$$

que ilustra cómo la estabilidad de la solución requiere que los errores de expectativas fluctúen en respuesta a los valores observados de la perturbación estructural en productividad. Con la parametrización descrita, se tiene:

$$P^{2\bullet} (\tilde{\Psi} z_t + \tilde{\Pi} \eta_t) = 0 \quad \forall t \Rightarrow 0.1051\epsilon_t + \eta_t = 0 \quad \forall t, \quad (13.24)$$

que es una relación exacta entre el error de expectativas del modelo y la innovación en la perturbación de productividad. El error de expectativas que calculemos a partir de (13.16) depende de modo no lineal de las variables de decisión y de estado y, por tanto, de las perturbaciones exógenas del modelo. En consecuencia, no satisfará la relación (13.24) exactamente, que es una aproximación lineal a la verdadera relación no lineal que existe entre estas variables.

Una vez generada una realización de la solución numérica, en la forma de una serie temporal para cada una de las variables del modelo, volvemos a generar por simulación una serie temporal diferente para la perturbación de productividad θ_t y, a partir de ella, obtendremos otra realización de series temporales para todas las variables del modelo. De este modo, obtenemos un conjunto numeroso de realizaciones para cada variable, que nos permitirá calcular asimismo una amplio número de realizaciones de cualquier estadístico: volatilidad de la producción y_t , coeficiente de correlación entre ésta y el consumo c_t , funciones de autocorrelación simple y parcial de la productividad $f'(k_{t-1})$, etc.. Las estimaciones de las matrices P , Λ , así como la condición de estabilidad (13.21) o la relación (13.24) son comunes a todas las simulaciones, al ser función únicamente de los parámetros del modelo, pero no de los valores numéricos de las perturbaciones θ_t .

5.3 Existencia de solución

Por supuesto que nada garantiza que un problema de control estocástico tenga solución. Siguiendo la metodología de solución numérica que hemos descrito, la existencia de un número mayor de condiciones de estabilidad que de expectativas condicionales que aparecen en el modelo, haría imposible que el vector de esperanzas condicionales pudiera ajustarse en cada período de manera que compensara los valores tomados por las perturbaciones exógenas del modelo, que en nuestro caso es la perturbación de productividad. En consecuencia, no existiría

una relación bien definida entre los errores de expectativas y las innovaciones estructurales, no existiendo una solución al modelo.¹⁵

Un número de condiciones de estabilidad igual al de expectativas condicionales del modelo, sin ser ninguna de ellas redundante hará, generalmente, que el modelo tenga una solución única que puede obtenerse combinando condiciones de estabilidad con las ecuaciones no lineales del modelo, obteniendo un conjunto de series temporales, una para cada una de las variables del modelo, más las variables que hemos definido como expectativas condicionales, más los errores de expectativas.

Un número de condiciones de estabilidad inferior al de expectativas condicionales del modelo generará lo que conocemos habitualmente como equilibrios *sunspot*, en los que podemos fijar arbitrariamente los valores numéricos de alguna de las expectativas y resolver el sistema de modo que se cumplan todas las ecuaciones del modelo. Por tanto, tenemos en general todo un continuo de soluciones (*indeterminación local estocástica*).

Otro sentido importante en que debe valorarse la existencia de soluciones es desde el punto de vista de la racionalidad de expectativas. Como hemos explicado anteriormente, el método propuesto incorpora el supuesto de racionalidad, lo que hace que los errores de expectativas tengan determinadas propiedades: media cero, ausencia de correlación serial, y correlación nula con las variables que eran conocidas en el momento de formar las expectativas. Todas estas características son contrastables estadísticamente¹⁶ y es, de hecho, una importante laguna en la investigación actual en este campo, el que habitualmente no se contrasten una vez generada la solución numérica. El incumplimiento de las propiedades estadísticas citadas implicaría rechazar el supuesto de racionalidad de expectativas que supuestamente se ha utilizado en la generación de la solución numérica. Ello podría deberse a los errores numéricos generados por utilizar una aproximación lineal al modelo para generar las condiciones de estabilidad, parte integral en la solución del mismo. Dichos errores numéricos se acumulan a los errores de expectativas, por lo que estos pudieran incumplir las propiedades mencionadas, invalidando la solución obtenida.

6 Resumen

Hemos descrito un método para la solución de problemas de control estocástico como los habitualmente propuestos en economía. El procedimiento incorpora y utiliza en su diseño el supuesto de racionalidad de expectativas, consistente en

¹⁵Salvo en el caso, muy infrecuente en aplicaciones de interés, en que alguna condición de estabilidad fuese redundante con el resto del sistema de tal modo que (13.23) pudiera cumplirse incluso si el número de filas en $P^{j\bullet}$ excede de la dimensión del vector η_t .

¹⁶Véase Den Haan y Marcer (1994) para un contraste acerca de la incorrelación entre errores de expectativas y el conjunto de información utilizado en la formación de las mismas.

cada agente decisor utiliza eficientemente toda la información de que dispone cuando toma sus decisiones. Las condiciones de estabilidad del sistema de ecuaciones de primer orden del problema de optimización juega un papel fundamental en el método de solución, informando acerca de la posible multiplicidad de soluciones, así como de la posible inexistencia de las mismas. La solución numérica toma la forma de una serie temporal, de la longitud deseada, para cada una de las variables del modelo. Mediante un experimento Monte Carlo, puede obtenerse un alto número de tales soluciones y caracterizar la densidad empírica de frecuencias de cualquier estadístico, a la luz de la cual puede evaluarse la verosimilitud del estadístico análogo calculado a partir de datos de economías reales. De este modo, el modelo económico puede entenderse como una función que convierte la distribución de probabilidad de las innovaciones estocásticas del modelo, en una distribución de probabilidad para cada uno de los estadísticos de interés relativo a una o varias variables del modelo: varianzas, coeficientes de correlación, etc.. La solución numérica precisa de una determinada parametrización del modelo, por lo que puede diseñarse un procedimiento de estimación consistente en hallar los valores de los parámetros que minimizan el valor numérico de una determinada función distancia entre los valores numéricos obtenidos por simulación para un determinado vector de estadísticos, y sus valores numéricos en una economía real.

7 Bibliografía

- [1] Benhabib, J. y R. Perli (1994), "Uniqueness and Indeterminacy: On the Dynamics of Endogenous Growth", *Journal of Economic Theory*, nº63, págs. 113-142.
- [2] Blanchard, O.J. y C.M. Kahn (1980), "The Solution of Linear Difference Models under Rational Expectations", *Econometrica*, vol. 48(5), págs. 1305-1311.
- [3] Brock, W.A. y L.J. Mirman (1972), "Optimal Economic Growth and Uncertainty: The Discounted Case", *Journal of Economic Theory*, vol. 4(3), págs. 497-513.
- [4] Den Haan, W.J. y A. Marcet (1994), "Accuracy in Simulations", *Review of Economic Studies*, vol. 61, págs. 3-17.
- [5] Marimón, R. y A. Scott (1999), *Computational Methods for the Study of Dynamic Economies*, R. Marimón y A. Scott (editores), Oxford University Press.
- [6] Novales, A., E. Domínguez, J. Pérez y J. Ruiz (1999), "Solving Nonlinear Rational Expectations Models by Eigenvalue-Eigenvector Decompositions", en *Computational Methods for the Study of Dynamic Economies*, R. Marimón y A. Scott (editores), Oxford University Press, págs. 62-92.

- [7] Novales, A., y C. Sebastián (1999), *Análisis Macroeconómico, vol.II*, Marcial Pons, Madrid.
- [8] Sims, C.A. (2001), "Solving Linear Rational Expectations Models", *Journal of Computational Economics*, vol. 20(1-2), págs.1-20.
- [9] Xie, D. (1994), "Divergence in Economic Performance: Transitional Dynamics with Multiple Equilibria", *Journal of Economic Theory*, nº63, págs. 97 112.

Learning by Doing e incertidumbre aditiva: solución analítica

Francisco Alvarez¹

Dto. Fundamentos del Análisis Económico II, Universidad Complutense de
Madrid. Email: fralvare@ccee.ucm.es

1 Introducción

En un trabajo pionero, Wright (1936) encontró evidencia empírica de que el tiempo requerido para construir un avión decrecía a medida que aumentaba el número de aviones que ya se habían fabricado. Este fenómeno, que ha recibido el nombre de *learning by doing*, ha sido -a partir del trabajo de Wright- encontrado en diferentes industrias y en diferentes períodos de tiempo.

La literatura que analiza las implicaciones de dicho efecto en un marco teórico comenzó algunos años mas tarde, con el trabajo de Arrow (1962). Al igual que ocurrió con su vertiente empírica, el trabajo de Arrow ha dado paso a una enorme cantidad de trabajos teóricos posteriores.

El learning by doing es un fenómeno naturalmente dinámico: las decisión de hoy de la empresa (*¿cuántos aviones producir hoy?*) afecta a sus posibilidades de mañana (horas de trabajo que serán necesarias mañana para producir un avión). Ello da lugar a modelizar el problema de decisión de la empresa desde un punto de vista teórico como un problema de optimización dinámica. Si además

¹Agradezco los comentarios de E. Cerdá. Cualquier error es responsabilidad mía.

suponemos, como parece natural, que en el momento de decidir hoy, la empresa no conoce con exactitud los efectos que sus decisiones presentes tendrán sobre sus posibilidades futuras, el problema anterior es esencialmente un problema de programación estocástica. En dicho problema, la incertidumbre aparece recogida por la presencia de shocks aleatorios.

En este trabajo se aborda el problema teórico de una empresa que tiene learning by doing e incertidumbre sobre como sus decisiones presentes afectan a sus posibilidades futuras. El principal resultado es la obtención analítica de la solución de dicho problema, para una modelización específica, mediante técnicas de programación dinámica. Esta obtención analítica de la solución permite estudiar propiedades de la misma que de otra forma involucran un análisis mucho mas complejo. Como ejemplo de propiedad a estudiar a partir de la solución analítica, en este trabajo comparamos la solución del problema abordado con la de el problema análogo determinístico.

La estructura general del trabajo es la siguiente. Tras esta sección introducatoria, la sección 2 contiene una breve revisión de la literatura. La sección 3 sitúa el problema abordado en el marco de la literatura previa. Las tres siguientes secciones contienen, respectivamente, la presentación formal del modelo (sección 4), la solución (sección 5) y algunas implicaciones económicas (sección 6). La sección 7 presenta algunas ideas para futuras investigaciones.

2 Revisión de la literatura

Hay una enorme literatura dentro de teoría económica y la literatura de investigación operativa posterior a los trabajos de Wright y Arrow. Esta revisión no pretende -ni mucho menos- ser exhaustiva, si no solamente destacar algunos trabajos que ha su vez han generado una literatura posterior que se basa en ellos. Dividimos los trabajos en empíricos y teóricos.

2.1 Evidencia empírica

Como hemos indicado, Wright (1936) comprobó empíricamente que el tiempo requerido para la construcción de un avión decrecía a medida que aumentaba el número de aviones fabricados. A partir de este trabajo los estudios se han extendido a diferentes sectores industriales, períodos de tiempo y formas de modelización.

De manera general, los trabajos empíricos usan datos de una industria o del sector industrial en su conjunto para ajustar lo que se conoce como *curva de aprendizaje*. Una curva de aprendizaje es una representación gráfica que relaciona

la cantidad producida acumulada, variable independiente, con alguna medida de la productividad (por ejemplo horas-trabajador necesarias para construir un avión), que es la variable independiente. La forma funcional más utilizada es la función de potencia:

$$z = aQ^{-b}$$

donde z mide las horas-trabajo de la última unidad producida, Q es la producción acumulada (a lo largo del tiempo) y a y b son los parámetros que se estiman a partir de los datos. Es importante notar que hemos definido learning by doing como una reducción de costes, que no tiene por qué ser equivalente a un aumento de productividad, si bien todos los trabajos empíricos toman ambos conceptos como equivalentes.

Los trabajos empíricos extienden el esquema anterior en alguna de las siguientes líneas: (a) estiman la curva de aprendizaje para diferentes industrias, (b) estudian posibles factores adicionales para explicar la mejora de productividad; (c) sugieren formas funcionales alternativas. Comentamos brevemente cada una de estas extensiones.

Se ha encontrado evidencia empírica de existencia de una curva de aprendizaje en la producción de energía eléctrica en centrales nucleares de EE.UU. (Joskov et al. (1979)) y de Francia (Lester et al. (1993)), en la producción de microconductores en Japón (Dick (1991)), en las industrias químicas de EE.UU. (Lieberman (1984)), en la producción de rayón² en EE.UU. (Jarmain (1994)), en la industria farmaceútica (Lieberman (1987)) y construcción de barcos (Argote et al. (1990)).

A un nivel de agregación del sector industrial mayor que el anterior, Jefferson (1988) encuentra evidencia de curvas aprendizaje en el sector manufacturero de EE.UU., y Backus et al. (1992) extienden posteriormente este estudio a más de sesenta países. Mitra (1992) encuentra también evidencia de una curva de aprendizaje en el sector manufacturero de la India. Webb et al. (1993) señalan que las exportaciones de determinados sectores dentro de las manufacturas de Costa Rica (maquinaria y manufacturas intermedias) generan mejoras de productividad en los mismos sectores de otros países que participan en el mercado común de América Central.

Pueden existir diversos factores que explican las mejoras de productividad, además de la producción acumulada. En este sentido, Fellner (1969) propone la inclusión del tiempo como variable relevante en un modelo en el que estudia la mejora en el rendimiento de los atletas. Moreh (1985) señala como variable

²Fibra sintética compuesta fundamentalmente por celulosa.

relevante el esfuerzo invertido en formación dentro de la empresa. Resultados similares aparecen en Shaw (1989). En el trabajo antes citado de Lieberman (1984) se señala la influencia de los gastos en I+D. El tamaño de la empresa y el reparto de mercado (cuántas empresas y que poder de mercado tiene cada una) es señalado también en Lieberman (1987). Argote et al. (1990) señalan un conjunto de factores que pueden dar lugar a que el aprendizaje se deprecie, por lo que deberían entrar en la curva de aprendizaje afectando negativamente a la productividad. Estos factores son, por ejemplo: diferentes turnos de empleados o interrupciones de la producción debidas a huelgas,...

Básicamente, se han propuesto tres alternativas a la forma más habitual de curva de aprendizaje, indicada anteriormente, y que como puede verse es lineal en logartimos.

Conway et al. (1959) sugieren que, en una escala logarítmica, la relación debe ser una linea recta con pendiente decreciente para valores iniciales de la producción acumulada y una recta horizontal (por tanto, sin mejoras adicionales de productividad) a partir de cierto valor crítico de ésta. Este modelo se conoce como el modelo *plateau*. Bahk et al. (1993) señalan tambien la importancia de permitir que el aprendizaje tenga una duración finita en el tiempo.

Como segunda alternativa está el modelo S, introducido por Cochran (1960), que es una versión *suavizada* del anterior: la curva de aprendizaje no llega a ser absolutamente horizontal, aunque si disminuye el valor absoluto de su pendiente con el paso del tiempo.

Finalmente el modelo Standford B, introducido por Garg et al. (1961), propone exactamente lo contrario: a partir de una etapa de aprendizaje relativamente lento, este se acelera. Este último modelo parece representar adecuadamente la evolución de la productividad en sectores con mucho cambio técnico en períodos muestrales relativamente cortos.

2.2 Estudios teóricos

En este apartado se presentan, casi a modo de enumeración, diferentes líneas de investigación teóricas que se han seguido a partir del trabajo pionero de Arrow (1962).

Una de las cuestiones que más atención ha recibido en la literatura teórica de learning by doing es la que relaciona dicho fenómeno con el crecimiento económico. Esencialmente, se ha introducido este efecto en las funciones de producción de modelos de crecimiento neoclásico para probar como el learning by doing puede generar crecimiento endógeno, o como puede explicar diferencias persistentes entre

las tasas de crecimiento de diferentes países. Además, un conjunto de trabajos estudia cuestiones relativas a inefficiencias dinámicas en la solución de mercado cuando existe learning by doing. Sin entrar en los aspectos concretos que aborda cada trabajo, en esta línea tenemos: Cigno (1984), Chamley (1993), d'Autume et al. (1993), Ishikawa (1992), King et al. (1993), Lucas (1988) y (1993), Matsuyama (1992), Parente (1994), Simon et al. (1984), Torvik (1993) y Young (1991).

Otro conjunto de trabajos analiza el fenómeno a un nivel industrial. Por ejemplo, ¿debe protegerse a una industria que actualmente es inefficiente pero que tiene potencial para mejorar su eficiencia porque se espera que tenga un efecto learning by doing? De forma algo sorprendente, el interés por esta cuestión ha descendido en los últimos años (al menos lo han hecho el número de publicaciones) a pesar de que la literatura disponible está lejos de agotar la cuestión. Artículos clásicos dentro de esta línea son: Clemhout et al. (1970), Chen et al. (1994), Feder et al. (1976), Kohn et al. (1992) y Succar (1987).

Otro tipo de trabajos dentro de economía industrial son aquellos que relacionan learning by doing y temas clásicos dentro de esta literatura. Ejemplo de ello es learning by doing en un modelo de diferenciación vertical de producto (Goe-ring (1993), Grubber (1992) y Stokey (1986)), learning by doing en un modelo de bienes duraderos (Olsen (1992)) o como el learning by doing puede determinar la propia estructura industrial (Agliardi (1990), Jovanovic et al. (1989), Mookherjee et al. (1991) y Spence (1981)).

3 Discusión sobre el problema abordado

La pregunta que abordamos en este trabajo es teórica. Desde un punto de vista matemático, la decisión de cuánto debe producir una empresa que en su estructura productiva tiene learning by doing puede representarse mediante un problema de optimización dinámica, lo que puede a su vez verse como un problema de programación matemática. Por otra parte, parece natural pensar que la empresa no tiene capacidad para predecir sin error alguno la evolución que va a tener en el futuro su estructura de costes, lo que añadido a lo anterior da lugar a un problema de programación (dinámica) estocástica.

La aportación fundamental de este trabajo es encontrar de manera analítica la política óptima de producción de una empresa que en su estructura productiva tiene learning by doing y dentro del cual existen elementos aleatorios que le impiden a la empresa predecir sin error su estructura de costes futura. La obtención de dicha solución analítica permite obtener propiedades de la misma que de otra manera son difíciles de conocer. Como ilustración de este punto, en este trabajo presentamos una comparación de la solución obtenida con la del

problema análogo determinístico, de modo que podemos medir cuál es el impacto de introducir incertidumbre en el proceso de aprendizaje.

Hemos elegido un modelo con una estructura industrial sencilla: un monopolio sin amenazas de entrada. Por otra parte, el modelo recoge la idea de que el learning by doing tiene una duración esencialmente finita en el tiempo. Concretamente, utilizamos una extensión natural al caso estocástico del modelo determinístico que se presenta en Dasgupta et al. (1988). Es un modelo formulado en tiempo discreto, con un horizonte temporal de dos períodos. El coste en el segundo período es una función de la producción del primer período. Dicha función es lineal definida en dos tramos: el primer tramo es decreciente y el segundo es horizontal (recogiendo la idea esencial del modelo de plateau, comentado anteriormente).

Considerar un modelo de dos períodos implica concentrar la fase de aprendizaje (de learning by doing) en un solo período y la fase de *madurez* de la industria (en la que ya no hay learning by doing) en otro. En una generalización a un modelo de T períodos la fase de aprendizaje dura varios períodos, por lo que la política óptima de la empresa indicaría como se reparte la producción a dentro de los períodos de aprendizaje.

Una cuestión de interés es cómo introducir incertidumbre en un modelo determinístico. En particular, en un modelo de learning by doing puede haber varias fuentes de incertidumbre.

Una primera distinción clara es si hay incertidumbre que disminuye en el tiempo o no. Por ejemplo, si la incertidumbre radica en que algunos parámetros del proceso de aprendizaje son desconocidos, cabe esperar que la empresa mejore las estimaciones de dichos parámetros -y por tanto la incertidumbre disminuya- a medida que pase el tiempo y haya datos disponibles. Como caso contrario, si hay una perturbación aleatoria de varianza estable en el tiempo, dicha incertidumbre no va disminuir: la empresa tendrá la misma capacidad predictiva al principio que la final de su vida activa. En este artículo nos centramos en este segundo tipo de incertidumbre. El primero ha sido estudiado en Alvarez et al. (1999a).

Aún dentro de la incertidumbre que no disminuye en el tiempo, podemos considerar dos tipos, que, usando una notación propia del control óptimo, denominaremos incertidumbre *multiplicativa* y *aditiva*, respectivamente. Imaginemos que la perturbación aleatoria entra en la curva de aprendizaje multiplicada por la producción presente (caso multiplicativo). Ello quiere decir que si la producción presente es cero no hay incertidumbre alguna. Un efecto diferente se tiene si la perturbación aparece sumada a la producción presente (caso aditivo). A un nivel más intuitivo, el grado de incertidumbre sobre los costes futuros puede depender o no de la producción presente (multiplicativo y aditivo, respectivamente). Como

su título indica, en este trabajo nos centramos en el caso aditivo. El multiplicativo ha sido estudiado en Alvarez et al. (1999b).

4 Modelo

Consideramos un monopolio neutral ante el riesgo sin amenazas de entrada y que tiene efecto de *learning by doing* en su estructura productiva. El horizonte temporal es dos períodos, 0 y 1, genéricamente denotados por t . El monopolista elige las cantidades (no negativas) a producir en cada período, siendo $q(t)$ la cantidad elegida en el período t . La función inversa de demanda es estacionaria en el tiempo, en el período t su inversa es:

$$p(t) = a - bq(t) \quad (14.1)$$

donde a y b son parámetros estrictamente positivos. En el período t , la función de costes es lineal, siendo $c(t)$ el coste unitario en dicho período, de modo que el coste total de producir $q(t)$ es $c(t)q(t)$. El efecto de learning by doing está en que el coste unitario cambia en el tiempo. En concreto, suponemos que:

$$c(1) = \max\{\tau, c(0) - \beta q(0) + \varepsilon(0)\} \quad (14.2)$$

en donde τ y β son parámetros estrictamente positivos y $\varepsilon(0)$ es una variable aleatoria cuya distribución de probabilidad es $\Pr(\varepsilon(0) = -\theta) = \Pr(\varepsilon(0) = \theta) = 1/2$, con $\theta \geq 0$. Todos los parámetros son conocidos y $c(0)$ está dado. Interpretamos τ como el menor coste unitario posible y β como el efecto que la producción presente tiene sobre coste unitario futuro. En nuestro modelo, el monopolista conoce ambas magnitudes, pero ello -junto con su decisión de producción presente- no le permite conocer con exactitud el coste futuro, eventualmente ni siquiera estará seguro de que vaya a haber reducción de coste unitario.

El objetivo del monopolista es maximizar el flujo esperado y descontado de beneficios, siendo $\lambda \in (0, 1)$ su factor de descuento³. A partir de (14.1) y de la expresión para los costes totales, es sencillo ver que los beneficios del monopolista en el período t son $(a - bq(t) - c(t))q(t)$, por lo que su objetivo es:

$$\max_{q(0), q(1)} E \left\{ \sum_{t=0}^1 \lambda^t (a - bq(t) - c(t)) q(t) \right\} \quad (14.3)$$

donde la esperanza es condicional en $c(0)$ y además $c(1)$ viene dado por (14.2). Otros supuestos adicionales del modelo son:

³El supuesto de que el factor de descuento está en $(0, 1)$ es habitual en la literatura económica y juega un papel esencial en problemas de optimización dinámica de horizonte temporal infinito (Kamien et al. (1991)). En nuestro modelo mantenemos dicho supuesto por analogía con la literatura.

$$c(0) > \tau a > c(0) + \theta \quad (14.4)$$

La primera de las desigualdades en (14.4) indica que en nuestro modelo puede haber reducción de costes. La segunda desigualdad asegura que la producción óptima es estrictamente positiva en cada período.

Cuatro comentarios finales sobre el modelo pueden ser de interés. En primer lugar, este modelo es una generalización a un caso de incertidumbre del propuesto por Dasgupta et al. (1988), que es determinístico ($\theta = 0$). En segundo lugar, es sencillo generalizar el método de solución que presentaremos posteriormente al caso en que $\varepsilon(0)$ tiene un soporte discreto y finito arbitrario, sin que dicha generalización modifique las principales implicaciones económicas de introducir incertidumbre. En tercer lugar, hemos de recordar que en nuestra modelización, el coste unitario puede aumentar de un período a otro. En cuarto lugar, debemos indicar que la generalización a un caso de T períodos es más compleja.

5 Solución

El problema planteado es de optimización dinámica estocástica con observación completa, tiene un horizonte temporal finito y está formulado en tiempo discreto. La variable de estado es el coste unitario y la variable de control es la cantidad producida. Lo resolveremos usando la metodología de programación dinámica propuesta por Bellman. La dificultad en resolver de modo analítico la ecuación funcional de Bellman asociada al problema planteado es identificar la probabilidad con que se alcanza τ en el período 1.

Para la presentación formal del método de solución introducimos la siguiente notación. Sea $\pi = \{q(0), q(1)\}$ en donde $q(t) \geq 0$ para $t \in \{1, 2\}$, decimos que π es una *política factible*. El conjunto de las políticas factibles lo denotamos por S . Dado $\pi \in S$ y $c(0)$, el flujo esperado y descontado de beneficios asociados lo denotamos por $V(\pi, c(0))$. Decimos que $\pi^* = \{q^*(0), q^*(1)\} \in S$ soluciona el problema planteado si $V(\pi^*, c(0)) \geq V(\pi, c(0))$ para todo $\pi \in S$. Además definimos $V^*(c(0)) = V(\pi^*, c(0))$ y lo denominamos *función valor*.

Hemos de notar que la segunda desigualdad en (14.4) garantiza que $a > c(1)$, a partir de lo cual es sencillo ver que $q^*(1) = (2b)^{-1}(a - c(1))$, de modo que en lo sucesivo nos centramos en la obtención de $q^*(0)$ y $V^*(c(0))$. A lo largo de los resultados que siguen definimos un conjunto de valores umbrales, sea $\{R_1, R_2, R_3, R_4\}$, que define una partición en subintervalos del intervalo de posibles valores de $c(0)$, a partir de (14.4) es sencillo ver que dicho intervalo es $(\tau, a - \theta)$ y los umbrales satisfacen $a - \theta > R_1 > R_2 > R_3 > R_4 > \tau$, de modo que

$q^*(0)$ y $V^*(c(0))$ son funciones de $c(0)$ definidas por tramos, correspondiendo los tramos con los subintervalos de dicha partición. Además, los diferentes resultados muestran que, como es natural, la probabilidad de alcanzar τ bajo π^* disminuye con $c(0)$. Las demostraciones están en el apéndice. Previamente a cada resultado se define la notación nueva utilizada en el mismo.

Notación para el teorema 1. Sean: $K = (4b)^{-1}$, $\phi = 2K$, $\phi_1 = (1 + 2\lambda\beta K)(2b - 2\lambda\beta^2 K)^{-1}$, $K_{1,2} = \lambda K + \frac{1}{2}(1 + 2\lambda\beta K)\phi_1$, $K_{1,0} = \lambda\theta^2 K$, $R_1 = (\tau + \theta + \beta\phi_1 a)(1 + \beta\phi_1)^{-1}$.

Teorema 1. Si se verifica: (a) $c(0) > R_1$, (b) $b > \lambda\beta^2 K$, entonces: (i) $q^*(0) = \phi_1(a - c(0))$ y $V^*(c(0)) = K_{1,0} + K_{1,2}(a - c(0))^2$, (ii) bajo π^* es $\Pr(c(1) > \tau) = 1$.

Notación para el teorema 2. Sean: $\phi_2 = (1 + \lambda\beta K)(2b - \lambda\beta^2 K)^{-1}$, $\alpha_2 = \lambda\beta\theta K(2b - \lambda\beta^2 K)^{-1}$, $K_{2,2} = \lambda K + \frac{1}{2}(1 + 2\lambda\beta K)\phi_2$, $K_{2,1} = -2aK_{2,2} + \lambda\theta(1 + \beta\phi_2)K$, $K_{2,0} = \frac{1}{2}\lambda K((a - \tau)^2 + (a - \theta)^2) + \frac{1}{2}((1 + \lambda\beta K)a - \lambda\beta\theta K)(\phi_2 a - \alpha_2)$, $R_2 = (\tau + \theta + \beta(\phi_2 a - \alpha_2))(1 + \beta\phi_2)^{-1}$, $R_3 = (\tau - \theta + \beta(\phi_2 a - \alpha_2))(1 + \beta\phi_2)^{-1}$.

Teorema 2. Si se verifica: (a) $R_2 \geq c(0) > R_3$, (b) $2b > \lambda\beta^2 K$, entonces: (i) $q^*(0) = \phi_2(a - c(0)) - \alpha_2$ y $V^*(c(0)) = K_{2,0} + K_{2,1}c(0) + K_{2,2}c(0)^2$, (ii) bajo π^* es $\Pr(c(1) = \tau) = 1/2$.

Notación para el teorema 3. Sean: $K_{3,0} = \lambda K(a - \tau)^2$, $R_4 = (\tau - \theta + \beta\phi a)(1 + \beta\phi)^{-1}$.

Teorema 3. Si se verifica: (a) $R_4 \geq c(0)$, entonces: (i) $q^*(0) = \phi(a - c(0))$ y $V^*(c(0)) = K_{3,0} + K(a - c(0))^2$, (ii) bajo π^* es $\Pr(c(1) = \tau) = 1$.

Proposición 1. Se verifica: $a - \theta > R_1 > R_2 > R_3 > R_4 > \tau$.

La prueba de todos los teoremas se basa en definir un problema que bajo las hipótesis (y sólo bajo éstas) es equivalente al original. Las hipótesis (b) de los teoremas 1 y 2 son requerimientos de concavidad de la función objetivo de cada problema equivalente. Es sencillo ver que la condición del teorema 1 es más restrictiva que la del teorema 2. Por otra parte, en virtud de la proposición 1, hay dos subintervalos de posibles valores de $c(0)$, (R_2, R_1) y (R_4, R_3) , para los que los teoremas anteriores no dan la solución. Dicha solución se presenta en los dos próximos teoremas. La diferencia entre los teoremas anteriores y los dos siguientes es que en los primeros el problema equivalente tiene solución anterior, mientras que en los últimos la solución es de esquina.

Notación para el teorema 4. Sean: $K_{4,0} = -\beta^{-1} (a + b\beta^{-1}(\tau + \theta)) (\tau + \theta) + \frac{1}{2}K\lambda((a - \tau)^2 + (a - \tau - 2\theta)^2)$, $K_{4,1} = \beta^{-1} (a + b\beta^{-1}(\tau + \theta)) + \beta^{-1} (1 + b\beta^{-1}) (\tau + \theta)$, $K_{4,2} = -\beta^{-1} (1 + b\beta^{-1})$.

Teorema 4. Si se verifica: (a) $R_1 \geq c(0) > R_2$, (b) $2b > \lambda\beta^2 K$, entonces: (i) $q^*(0) = \beta^{-1}(c(0) - \tau - \theta)$ y $V^*(c(0)) = K_{4,0} + K_{4,1}c(0) + K_{4,2}c(0)^2$, (ii) bajo π^* es $\Pr(c(1) = \tau) = 1/2$.

Notación para el teorema 5. Sean: $K_{5,0} = \lambda K (a - \tau)^2 - \beta^{-1} (a + b\beta^{-1}(\tau - \theta)) (\tau - \theta)$, $K_{5,1} = \beta^{-1} (a + b\beta^{-1}(\tau - \theta)) + \beta^{-1} (1 + b\beta^{-1}) (\tau - \theta)$, $K_{5,2} = K_{4,2}$.

Teorema 5. Si se verifica: (a) $R_3 \geq c(0) > R_4$, (b) $2b > \lambda\beta^2 K$, entonces: (i) $q^*(0) = \beta^{-1}(c(0) - \tau + \theta)$ y $V^*(c(0)) = K_{5,0} + K_{5,1}c(0) + K_{5,2}c(0)^2$, (ii) bajo π^* es $\Pr(c(1) = \tau) = 1$.

Esto conluye la presentación de la solución del problema planteado. Como caso particular, los resultados anteriores presentan también la solución del problema determinístico ($\theta = 0$).

6 Implicaciones Económicas

En esta sección analizamos las implicaciones sobre la política óptima del período 0, $q^*(0)$, de introducir incertidumbre en el proceso de reducción de costes. Hay dos formas de hacer este análisis. La primera, que llamaremos *impacto marginal*, es considerar pequeñas variaciones en la incertidumbre una vez que ésta existe, lo que en nuestro modelo puede hacerse estudiando cómo varía la solución ante pequeñas variaciones del parámetro θ . La segunda, que llamaremos *impacto general*, es estudiar la diferencia entre que haya o no incertidumbre, lo que en nuestro modelo equivale a estudiar la diferencia entre la política óptima del período 0 para el caso determinístico ($\theta = 0$) y estocástico ($\theta > 0$).

El análisis del impacto marginal es directo a partir de la solución presentada en la sección anterior, por lo que omitimos aquí una presentación rigurosa del mismo. Basicamente, ante pequeños aumentos de θ , si $q^*(0)$ varía, lo hace de manera continua, aumentando o disminuyendo dependiendo del valor de $c(0)$. Hemos de notar que aumentar θ implica hacer *mean-preserving spread* de la distribución de $\varepsilon(0)$, es decir, aumentar la varianza sin alterar la media. En un contexto como el considerado aquí, en el que el tomador de decisiones (monopolista), es neutral ante el riesgo, es de esperar que, tal como indicamos, el efecto que dicha variación sea ambigüo.

El análisis del impacto general está en línea con el marginal, que acabamos de comentar. Queda recogido en la siguiente proposición.

Proposición 1. Sea $q_d^*(0)$ la política óptima en el período 0 del problema determinístico. Entonces: (i) si $c(0) > R_1$ o bien $c(0) < R_4$, $q_d^*(0) = q^*(0)$; (ii) existen valores c_a y c_b satisfaciendo $R_1 > c_b > c_a > R_4$ tales que $c(0) = c_a$ implica que $q_d^*(0) < q^*(0)$ y $c(0) = c_b$ implica que $q_d^*(0) > q^*(0)$.

7 Futuras investigaciones

En este trabajo hemos obtenido, usando programación dinámica, la solución para un modelo de learning by doing en el que hay incertidumbre aditiva. A pesar de la simplicidad del modelo elegido, su solución analítica pone de manifiesto que algunas de las propiedades de la política óptima de la empresa bajo learning by doing en el caso determinístico no se mantienen al introducir shocks aleatorios. Profundizar en el estudio del efecto de estos shocks es, a nuestro juicio, el campo de trabajo más prometedor.

A modo puramente de ejemplo, en trabajos previos de la literatura se ha estudiado en un contexto determinístico como el learning by doing puede determinar la estructura de la industria. La intuición subyacente es sencilla: empresas con una ventaja relativa inicial (digamos en términos de poder de mercado) *agrandan* esa ventaja a lo largo del tiempo debido a que ellas aprenden a reducir costes más rápido que sus rivales (dado que producen más). Ello implica que la existencia de learning by doing puede inducir a la disminución del número de empresas a lo largo del tiempo, lo que tiene conocidos efectos sobre la eficiencia social de la asignación de mercado. Ahora bien, el razonamiento anterior podría ser completamente distinto si las empresas con una ventaja inicial tienen que enfrentarse *antes* que las demás (y quizás por hacerlo antes lo hagan en mayor medida) a la incertidumbre inherente a sus procesos de aprendizaje.

Apendice: demostraciones.

Demostración del Teorema 1. Sea el problema auxiliar siguiente: la función objetivo es la dada por (14.3) s.a.: $c(1) = c(0) + \beta q(0) + \varepsilon(0)$ con $c(0)$ dado. Si la política óptima del problema auxiliar, sea $\pi_a^* = \{q_a^*(0), q_a^*(1)\}$, satisface la restricción adicional siguiente $c(1) \geq \tau$, entonces es también la política óptima del problema original. La demostración del teorema se hace en dos pasos.

en el primer paso se demuestra que π_a^* es la que aparece en el enunciado del teorema. En el segundo paso se muestra que bajo π_a^* se verifica $\Pr(c(1) > \tau) = 1$. Primer paso. Para $t = 0, 1$, denotamos por $V_a^*(c(t), t)$ la función valor del problema auxiliar en el período t , donde $t \in \{0, 1, 2\}$. La ecuación funcional de Bellman asociada al problema auxiliar es:

$$V_a^*(c(t), t) = \max_{q(t) \geq 0} \{(a - bq(t) - c(t))q(t) + \lambda E\{V_a^*(c(t+1), t+1)\}\} \text{ para } t = 0, 1$$

con $V_a^*(c(2), 2) = 0$ dado. Resolvemos recursivamente la anterior ecuación, primero para $t = 1$ y después para $t = 0$. Para $t = 1$ se obtiene $V_a^*(c(1), 1) = K(a - c(1))^2$, mientras que para $t = 0$, bajo (b), se obtiene $q_a^*(0) = q^*(0)$ y $V_a^*(c(0)) = V^*(c(0))$ donde $q^*(0)$ y $V^*(c(0))$ están dados en el enunciado del teorema. Segundo paso. Bajo π_a^* , se tiene que $c(1) > \tau$ c.p.1. $\Leftrightarrow c(0) - \beta q_a^*(0) - \theta > \tau \Leftrightarrow c(0) > R_1$. Q.E.D.

A partir de la demostración anterior, el siguiente Corolario, que será utilizado posteriormente, es trivial.

Corolario 1. Si se verifica: (a) $c(0) \leq R_1$, (b) $2b > \lambda\beta^2 K$, entonces: bajo π^* es $\Pr(c(1) = \tau) > 0$.

La demostración del Teorema 2 es totalmente análoga a la del Teorema 1 utilizando el siguiente problema auxiliar: la función objetivo es la dada por (14.3) s.a.: $c(1)$ condicionado a $c(0)$ y $q(0)$ tiene la siguiente distribución de probabilidad: $\Pr(c(1) = \tau) = \Pr(c(1) = c(0) - \beta q(0) + \theta) = \frac{1}{2}$, con $c(0)$ dado, y donde la política óptima del problema adicional debe satisfacer la restricción adicional siguiente: $c(0) - \beta q(0) - \theta \leq \tau < c(0) - \beta q(0) + \theta$. A partir de éste análisis, se obtiene además el siguiente Corolario, que será utilizado posteriormente.

Corolario 2. Si bajo π^* es $\Pr(c(1) = \tau) > 0$ y además (a) $c(0) \leq R_3$ y (b) $2b > \lambda\beta^2 K$, entonces bajo π^* es $\Pr(c(1) = \tau) > 1/2$.

La demostración del Teorema 3 es totalmente análoga a la del Teorema 1 utilizando el siguiente problema auxiliar: la función objetivo es la dada por (14.3) s.a.: $c(1) = \tau$, y donde la política óptima del problema adicional debe satisfacer la restricción adicional siguiente: $c(0) - \beta q(0) + \theta \leq \tau$.

La demostración de la proposición 1 es sencilla mediante álgebra.

Demostración del teorema 4. Teniendo en cuenta el Corolario 1, $R_1 \geq c(0)$ y (b), ha de ser $\Pr(c(1) = \tau) > 0$ bajo π^* . Pero $\Pr(c(1) = \tau) > 0 \Leftrightarrow q(0) \geq \beta^{-1}(c(0) - \tau - \theta)$, por tanto en cualquier problema cuya solución sea la óptima del problema inicial está debe satisfacer la restricción anterior. En el problema

auxiliar de la demostración del Teorema 2 se toma el problema inicial considerando la menor de las probabilidades posibles de que $c(1) = \tau$ sin imponer la restricción anterior, y se obtiene que bajo la política óptima de dicho problema auxiliar restricción se satisface si y solo si $c(0) \leq R_2$, lo cual no sucede bajo las hipótesis de este teorema. Además, por verificarce (b), la función objetivo del citado problema auxiliar es globalmente cóncava, por lo que el mejor $q(0)$ es el que satisface exactamente la restricción, es decir, $q^*(0) = \beta^{-1}(c(0) - \tau - \theta)$. De sustituir en dicha función objetivo este valor de $q(0)$ se obtiene $V^*(c(0))$. *Q.E.D.*

Demostración del teorema 5. Teniendo en cuenta el Corolario 2, $c(0) \leq R_3$ y (b), ha de ser $\Pr(c(1) = \tau) > 1/2$ bajo π^* , y la siguiente probabilidad posible mayor a $1/2$ es 1, verificándose además $\Pr(c(1) = \tau) = 1 \Leftrightarrow q(0) \geq \beta^{-1}(c(0) - \tau + \theta)$, por tanto en cualquier problema cuya solución sea la óptima del problema inicial está debe satisfacer la restricción anterior. En el problema auxiliar de la demostración del Teorema 3 se toma el problema inicial considerando $c(1) = \tau$ sin imponer la restricción anterior, y se obtiene que bajo la política óptima de dicho problema auxiliar restricción se satisface si y solo si $c(0) \leq R_4$, lo cual no sucede bajo las hipótesis de este teorema. Dado que la función objetivo del citado problema auxiliar es globalmente cóncava, el mejor $q(0)$ es el que satisface exactamente la restricción, es decir, $q^*(0) = \beta^{-1}(c(0) - \tau + \theta)$. De sustituir en dicha función objetivo este valor de $q(0)$ se obtiene $V^*(c(0))$. *Q.E.D.*

Demostración de la proposición 1. La política óptima del problema determinístico se obtiene al particularizar los resultados de la sección 3 para el caso $\theta = 0$. A partir de ahí, la prueba de (i) es trivial. Veamos (ii). Sea la función auxiliar $g(x, y, z) = (\tau + x + \beta(ya - z))(1 + \beta y)^{-1}$. O bien ocurre (a) $g(-\theta, \phi_2, \alpha_2) < g(0, \phi, 0)$ o bien ocurre (b) $g(-\theta, \phi_2, \alpha_2) \geq g(0, \phi, 0)$. Si ocurre (a), existe $c_a \in (g(-\theta, \phi_2, \alpha_2), g(0, \phi, 0))$ tal que si $c(0) = c_a$ es $q_d^*(0) = \phi(a - c_a)$ y $q^*(0) = \phi_2(a - c_a) - \alpha_2$, por lo que $q_d^*(0) < q^*(0) \Leftrightarrow a > 2b\theta(2b + \beta)^{-1}$, y la última desigualdad se verifica bajo (14.4). Si ocurre (b), existe $c_a \in [g(0, \phi, 0), g(-\theta, \phi_2, \alpha_2)]$ tal que si $c(0) = c_a$ es $q_d^*(0) = \beta^{-1}(c_a - \tau)$ y $q^*(0) = \beta^{-1}(c_a - \tau + \theta)$, por lo que $q_d^*(0) < q^*(0)$ se verifica. Además, o bien ocurre (c) $g(\theta, \phi_2, \alpha_2) < g(0, \phi_1, 0)$ o bien ocurre (d) $g(\theta, \phi_2, \alpha_2) \geq g(0, \phi_1, 0)$. Si ocurre (c), existe $c_b \in (g(\theta, \phi_2, \alpha_2), g(0, \phi_1, 0))$ tal que c_b es mayor que cualquiera de los c_a seleccionados anteriormente y además si $c(0) = c_b$ es $q_d^*(0) = \beta^{-1}(c_b - \tau)$ y $q^*(0) = \beta^{-1}(c_b - \tau - \theta)$ por lo que $q_d^*(0) > q^*(0)$ se verifica. Si ocurre (d), existe $c_b \in [g(0, \phi_1, 0), g(\theta, \phi_2, \alpha_2)]$ tal que c_b es mayor que cualquiera de los c_a seleccionados anteriormente y además si $c(0) = c_b$ es $q_d^*(0) = \phi_1(a - c_b)$ y $q^*(0) = \phi_2(a - c_b) - \alpha_2$, por lo que $q_d^*(0) > q^*(0)$ se verifica. Para ver que cualesquiera de los c_a y c_b seleccionados pertenecen a (R_4, R_1) basta notar que: $c_a \geq \min\{g(-\theta, \phi_2, \alpha_2), g(0, \phi, 0)\} > g(-\theta, \phi, 0)$ y $c_b \leq \min\{g(0, \phi_1, 0), g(\theta, \phi_2, \alpha_2)\}$. *Q.E.D.*

8 Bibliografía

- [1] Agliardi, E. (1990) Learning by doing and the emergence of monopoly, *Economics letters*, 32, pp. 353-357.
- [2] Alvarez, F. y Amman, H. M. (1999a) Learning by Doing under uncertainty, *Computational Economics*, 14, pp. 255-262.
- [3] Alvarez, F. y Cerdá, E. (1999b) A solution method for a class of Learning by Doing models with multiplicative uncertainty, *TOP of Operations Research*, 7, 1, pp 1-23.
- [4] Argote, L. Beckman, S. L. y Apple, D. (1990) Learning curves in manufacturing, *Science*, 247, pp. 920-924.
- [5] Arrow, K. J. (1962) The economic implications of learning by doing, *Review of Economic Studies*, 29, pp. 155-173.
- [6] Backus, D. K. Kehoe, P. J. y Kehoe, T. (1992) In the search of scale effects in trade and growth, *Journal of Economic Theory*, 58, pp. 377-409.
- [7] Bahk, B. y Gort, M. (1993) Decomposing learning by doing in new plants, *Journal of Political Economy*, 101, 4, pp. 561-583.
- [8] Cigno, A. (1984) Further implications of learning by doing: the effect of population on per-capita income, *Bulletin of Economic Research*, 36, 2, pp. 98-108.
- [9] Chamley, C. (1993) Externalities and dynamics in models of learning by doing, *International Economic Review*, 34, 3, pp. 583-609.
- [10] Chen, L. L. y Deveraux, J. (1994) Protection, export promotion and development, *Journal of Development Economics*, 40, pp. 387-395.
- [11] Clemhout, S. y Wan, H. Y. Jr. (1970) Learning by doing and infant industry protection, *Review of Economic Studies*, 37, pp. 33-56.
- [12] Conway, R. W. y Schultz, A. (1959) The manufacturing progress function, *Journal of Industrial Engineering*, 10.
- [13] Dasgupta, P. y Stiglitz, J. (1988) Learning by doing, market structure and industrial trade policies, *Oxford Economic Papers*, 40, pp. 246-268.
- [14] d'Autume, A. y Michel, P. (1993) Endogenous growth in Arrow's learning by doing model, *European Economic Review*, 37, pp. 1175-1184.
- [15] Dick, A. R. (1991) Learning by doing and dumping in the semiconductor industry, *Journal of Law & Economics*, XXXIV, pp. 133-159.

- [16] Feder, G. y Schmitz, A. (1976) Learning by doing and infant industry protection: a partial equilibrium approach, *Review of Economic Studies*, 43(1), pp. 175-178.
- [17] Fellner, W. (1969) Specific interpretations of learning by doing, *Journal of Economic Theory*, 1, pp. 119-140.
- [18] Garg, A. y Milliman, P. (1961) The aircraft progress curve modified for design changes, *Journal of industrial Engineering*, 12, 23.
- [19] Goering, G. E. (1993) Learning by doing and product durability, *Economica*, 60, pp. 311-326.
- [20] Grubber, H. (1992) Persistence of leadership in product innovation, *The Journal of Industrial Economics*, XL, 4, pp. 359-375.
- [21] Ishikawa, J. (1992) Learning by doing, changes in industrial structure and trade patterns, and economic growth in a small open economy, *Journal of International Economics*, 33, pp. 221-244.
- [22] Jarmin, R. S. (1994) Learning by doing and competition in the early rayon industry, *RAND Journal of Economics*, 25, 4, pp. 441-454.
- [23] Jefferson, G. H. (1988) The aggregate production function and productivity growth: Verdoorn's law revisited, *Oxford Economic Papers*, 40, pp. 671-691.
- [24] Joskow, P. L. y Rozanski, G. A. (1979) The effect of learning by doing on nuclear plant operating reability, *The review of Economic and Statistics*, LXI, 2, pp. 161-168.
- [25] Jovanovic, B. y Lach, S. (1989) entry, exit, and diffusion with learning by doing, *American Economic Review*, 79, 4, pp. 690-699.
- [26] Kamien, M. I. y N. L. Schwartz (1991) Dynamic Optimization, Elsevier, The Netherlands.
- [27] King, I. y Ferguson, D. (1993) Dynamic inefficiency, endogenous growth and ponzi games, *Journal of Monetary Economics*, 32, pp. 79-104.
- [28] Kohn, M. y Marrion, N. (1992) The implications of knowledge-based growth for the optimality of open capital markets, *Canadian Journal of Economics*, XXV, 4, pp. 865-883.
- [29] Lester, r. K. y McCabe, M. J. (1993) The effect of industrial Structure on learning by doing in nuclear power plant operation, *RAND Journal of Economics*, 24, 3, pp. 418-438.

- [30] Lieberman, M. B. (1984) The learning curve and pricing in the chemical processing industries, *RAND Journal of Economics*, 15, 2, pp. 213-228.
- [31] Lieberman, M. B. (1987) Patents, learning by doing, and market structure in the chemical processing industries, *International Journal of Industrial Organization*, 5, pp. 257-276.
- [32] Lucas, R. E. Jr. (1988) On the mechanics of economic development, *Journal of Monetary Economics*, 22, pp. 3-42.
- [33] Lucas, R. E. Jr. (1993) Making a miracle, *Econometrica*, 61, 2, pp. 251-272.
- [34] Matsuyama, K. (1992) Agricultural productivity, comparative advantage, and economic growth, *Journal of Economic Theory*, 58, pp. 317-334.
- [35] Mitra, P. K. (1992) Tariff design and reform in a revenue-constrained economy: theory and illustration from India, *Journal of Public Economics*, 47, pp. 227-251.
- [36] Mookherjee, D. y Ray, D. (1991) Collusive market structure under learning by doing and increasing returns, *Review of Economic Studies*, 58, pp. 991-1009.
- [37] Moreh, J. (1985) Human capital and learning by doing, *Metroeconomica*, 37, 3, pp. 307-329.
- [38] Olsen, T. E. (1992) Durable goods monopoly, learning by doing and the Coase conjecture, *European Economic Review*, 36, pp. 157-177.
- [39] Parente, S. L. (1994) Technology adoption, learning by doing and economic growth, *Journal of Economic Theory*, 63, pp. 346-369.
- [40] Shaw, K. L. (1989) Life-cycle labor supply with human capital accumulation, *International Economic Review*, 30, 2, pp. 431-456.
- [41] Simon, J. L. y Steinmann, G. (1984) The economic implication of learning by doing for population size and growth, *European Economic Review*, 26, pp. 167-185.
- [42] Spence, A. M. (1981) The learning curve and competition, *Bell Journal of Economics*, 12, pp. 49-79.
- [43] Stokey, N. L. (1986) The dynamics of the industrywide learning, Ch. 3. Equilibrium analysis: essays in honour of Kenneth J. Arrow, vol II, W. P. Heller, R. M. Starr y D. A. starrett, Cambridge University Press.

- [44] Succar, P. (1987) The need for industrial policy in LDC's a re-statement of the infant industry argument, *International Economic Review*, 28,2, pp. 521-534.
- [45] Torvik, R. (1993) Talent, growth and income distribution, *Scandinavian Journal of Economics*, 95(4), pp. 581-593.
- [46] Young, A. (1991) Learning by doing and the dynamic effects of international trade, *Quarterly Journal of Economics*, 106, pp. 369-405.
- [47] Wright, T. P. (1936) Factors affecting the costs of airplanes, *Journal of Aeronautical Science*, 3, pp. 122-128.

Programación estocástica por metas

Antonio Heras Martínez^a, Ana García Aguado^b

^a Dep. de Economía Financiera y Contabilidad, Universidad Complutense,

^b Dep. de Estadística e Investigación Operativa II (Métodos de decisión), Universidad Complutense,
e-mails: @ccee.ucm.es, a.g.aguado@emp.ucm.es

1 Introducción

Como es bien conocido, los problemas de decisión en ambiente de certeza a menudo se convierten en problemas de optimización matemática. La Programación Multiobjetivo se ocupa de resolver algunos de dichos problemas, a saber, aquellos en los cuales se trata de optimizar a la vez varias funciones objetivo. Dentro de los métodos de resolución de problemas multiobjetivo, la denominada Programación por Metas ocupa un lugar destacado. Posiblemente se deba, al menos en parte, a que los decisores, en palabras de Herbert Simon (1957), frecuentemente no actúan conforme a un paradigma "optimizador" sino que lo hacen de acuerdo con un paradigma "satisfaciente". En efecto, en los problemas multiobjetivo reales es muy común que los decisores no aspiren a la optimización (en el sentido paretiano) de todas las funciones objetivo involucradas, sino que se conformen con la consecución de un conjunto de objetivos o metas definidas previamente y que se conocen como niveles de aspiración. Cuando los costes de no consecución de los niveles de aspiración (bien por defecto o por exceso) se pueden agregar, por ejemplo, cuando tales costes se pueden agregar monetariamente, las técnicas de Programación por Metas son de gran utilidad en los procesos de ayuda a la toma de decisiones (véanse, por ejemplo, Romero (1991, 1993 y 2002)).

Por otra parte, a menudo algunas de las variables que intervienen en los problemas reales no toman valores de forma determinista sino que tienen naturaleza estocástica. Esto ha motivado el desarrollo de una rama de la Programación Matemática, conocida como Programación Estocástica, que se ocupa de la resolución de programas matemáticos en los cuales algunos parámetros se modelizan matemáticamente como variables aleatorias (véanse, por ejemplo, Kall y Wallace (1994), Kibzun y Kan (1996), Birge y Louveaux (1997); para el caso multiobjetivo, véase Stancu-Minasian (1984)). Un caso especialmente interesante de la Programación Estocástica Multiobjetivo es el de la Programación por Metas Estocástica (Heras y G. Aguado (1998 y 1999)), que estudia aquellos problemas de Programación por Metas en los que los niveles de aspiración no son conocidos de antemano con certidumbre, aunque se pueden asignar probabilidades a sus distintos valores (en otras palabras, los niveles de aspiración constituyen variables aleatorias). Numerosos problemas de la vida real se pueden en principio modelizar matemáticamente mediante programas en los que se trata de encontrar los valores de las variables de decisión que hacen que ciertas funciones objetivo tomen valores cercanos a unas metas que no son conocidas con precisión. Pensemos, por ejemplo, en una empresa cuyo objetivo fundamental consiste en satisfacer la demanda futura de los bienes que produce, demanda cuyo valor exacto evidentemente desconoce. Muchos problemas de la Matemática Actuarial podrían, asimismo, encuadrarse dentro de esta categoría. Así, por ejemplo, en un problema de tarificación se pretende calcular, a partir de las características de cada asegurado, una prima que sea lo más cercana posible al verdadero valor de su siniestralidad, el cual es evidentemente desconocido de antemano. En el presente artículo expondremos algunos resultados importantes sobre la Programación por Metas Estocástica. En el apartado segundo recordaremos los conceptos más importantes referentes a la Programación por Metas determinista. En el apartado tercero expondremos la formulación y algunas propiedades importantes de los programas estocásticos por metas. En los apartados siguientes estudiaremos con cierto detalle una aplicación de lo anterior al diseño de Sistemas de Tarificación Bonus-Malus, un problema clásico de tarificación en el seguro del automóvil. Veremos que la aplicación de técnicas de Programación por Metas Estocástica permite obtener algunas características deseables de las soluciones que no eran tenidas en cuenta por los métodos clásicos de resolución de tales problemas.

2 Programación por metas determinista

El propósito del decisor en un problema de Programación por Metas es hallar una solución factible lo más cercana posible a una meta dada que, habitualmente, es inalcanzable. Los problemas de Programación por Metas se pueden resolver por varios métodos, entre los que destacan los métodos gráfico, secuencial, multifase y de Arthur y Ravindran para la resolución de programas lexicográficos, y el

método de programación por metas ponderadas. Este último es probablemente el más utilizado en la literatura del tema, y se formula mediante el siguiente programa:

$$\begin{aligned} \min & \sum_{i=1}^p (w_i^+ y_i^+ + w_i^- y_i^-) \\ s.a. & f(\bar{x}) + \bar{y}^+ - \bar{y}^- = \bar{m} \\ & \bar{y}^+, \bar{y}^- \geq \bar{0} \\ & \bar{x} \in X \\ & w_i^+, w_i^- \geq 0 \end{aligned} \quad (1)$$

donde f la función objetivo, es decir $f(\bar{x}) = (f_1(\bar{x}), \dots, f_p(\bar{x}))$, $\bar{m} = (m_1, \dots, m_p)$ el vector de los “niveles de aspiración”, $\bar{x} = (x_1, \dots, x_n)$ es el vector de variables de decisión, $X \subset \Re^n$ el conjunto factible y los pesos w_i^+, w_i^- representan las tasas de intercambio (trade-offs) o coeficientes que se asocian a las variables de desviación, y_i^+, y_i^- , por exceso y por defecto, respectivamente. Cuando estas ponderaciones representan los costes monetarios de sobrepasar y de no alcanzar las metas, la solución óptima del programa (1) minimiza el coste global de todas las desviaciones. El programa (1) es un programa lineal siempre que las funciones objetivo sean lineales, ($f(\bar{x}) = A\bar{x}$ donde A es una matriz de orden $p \times n$), y el conjunto factible, X , esté definido mediante restricciones lineales. Otra formulación equivalente, aunque menos utilizada, de este problema es la siguiente: Para cada $\bar{x} \in X$ definimos

$$\begin{aligned} Q(\bar{x}) = & \min_{\bar{y}^+, \bar{y}^-} \sum_{i=1}^p (w_i^+ y_i^+ + w_i^- y_i^-) \\ \text{s. a.} & A\bar{x} + \bar{y}^+ - \bar{y}^- = \bar{m}, \bar{y}^+, \bar{y}^- \geq \bar{0} \end{aligned} \quad (2)$$

Los valores óptimos de las variables y_i^+, y_i^- verifican que $y_i^+ \cdot y_i^- = 0$ para todo i y representan el exceso y el defecto del i -ésimo objetivo, respectivamente, asociado con la solución factible \bar{x} .

La solución óptima del problema de Programación por Metas es la solución factible que minimiza la desviación global de los p objetivos, es decir, la solución del programa

$$\begin{aligned} \min & Q(\bar{x}) \\ s.a. & \bar{x} \in X \end{aligned} \quad (3)$$

Esta no es la formulación usual para la Programación por Metas porque, en lugar del programa lineal (1), se tienen dos programas, el (2), que también es un programa lineal, y el (3), que es un programa convexo. Pero incluimos aquí esta formulación alternativa porque es similar a la solución que propondremos, en el apartado siguiente, para resolver el problema de Programación Estocástica por Metas.

El origen de la metodología de la Programación por Metas se debe a Charnes, Cooper y Ferguson (1955). A partir de entonces son numerosísimos los trabajos publicados desarrollando aspectos teóricos, aplicaciones prácticas y posibles extensiones de la Programación por Metas. Como ya hemos comentado en la Introducción, debido a que los problemas de decisión reales frecuentemente dependen de parámetros desconocidos de naturaleza estocástica, una de las extensiones naturales de la Programación por Metas es la denominada Programación Estocástica por Metas. En la sección siguiente estudiamos este problema cuando la aleatoriedad está únicamente en los "niveles de aspiración", es decir, cuando el vector de metas, \bar{m} , es un vector aleatorio que representaremos por $\bar{\xi} = (\bar{\xi}_1, \dots, \bar{\xi}_p)$.

3 Programación por metas estocástica

Por analogía con la Programación por Metas Determinista, los problemas de Programación Estocástica por Metas se pueden representar así:

$$\begin{aligned} \text{mín} \quad & \sum_{i=1}^p (w_i^+ y_i^+ + w_i^- y_i^-) \\ \text{s.a.} \quad & f(\bar{x}) + \bar{y}^+ - \bar{y}^- = \bar{\xi} \\ & \bar{y}^+, \bar{y}^- \geq \bar{0} \\ & \bar{x} \in X \\ & w_i^+, w_i^- \geq 0 \end{aligned} \tag{4}$$

cuyos términos tienen la misma interpretación que en el problema (1).

Supongamos que el programa (4) es un programa lineal, es decir, que las funciones objetivo son lineales, ($f(\bar{x}) = A\bar{x}$ donde A es una matriz de orden $p \times n$) y el conjunto factible, X , está definido mediante restricciones lineales. Bajo estas condiciones, (4) es la expresión matemática de un problema de *Programación Lineal Estocástica*, ya que se trata de un programa lineal que contiene algunos parámetros que son variables aleatorias. Por lo tanto, se podría intentar resolver mediante alguno de los procedimientos utilizados en la resolución de este tipo de problemas. Por otro lado, el problema también se puede considerar como un caso de decisión en ambiente de riesgo, por lo que para su resolución también se podría recurrir a técnicas de la *Teoría Bayesiana de la Decisión*.

Comencemos por esta última. Como es bien sabido, si las preferencias del decisor sobre las posibles consecuencias de sus decisiones son consistentes con ciertos axiomas de comportamiento racional (véase, por ejemplo, DeGroot (1970), capítulo 7), entonces es posible definir una función sobre dichas consecuencias (*la función de utilidad*) que representa numéricamente dichas preferencias, en el sentido de que una decisión es preferida a otra si, y solamente si, la utilidad esperada de las consecuencias es mayor para la primera decisión que para la segunda. En muchos problemas de decisión es habitual considerar la función opuesta de la utilidad, denominada función de *desutilidad* o de pérdida. En tales

casos, el decisor escoge como decisión óptima aquella que minimiza la pérdida esperada de sus consecuencias. Más detalladamente, el decisor debe proceder de la siguiente manera: en primer lugar, debe definir una consecuencia para cada decisión factible y cada posible realización del parámetro aleatorio; en segundo lugar, debe especificar la función de pérdida utilizada; finalmente, debe calcular la pérdida esperada de cada decisión factible, y elegir como óptima aquella decisión que tenga asociada una pérdida esperada mínima.

Si aplicamos este procedimiento a nuestro problema, deberíamos comenzar preguntándonos por la apropiada definición de las consecuencias de la elección de una decisión factible \bar{x} y una realización del vector aleatorio $\bar{\xi}$. Es evidente que la Programación por Metas estocástica debe ser una generalización de la determinista, y que para cada realización del vector aleatorio tenemos un problema determinista. Por tanto, puede parecer razonable definir la consecuencia asociada con \bar{x} y $\bar{\xi}$ como el resultado óptimo del programa (2) cuando $\bar{m} = \bar{\xi}$, al que denotaremos como $Q(\bar{x}, \bar{\xi})$. La solución del programa (4) que buscamos consiste, de acuerdo con la metodología bayesiana, en minimizar la pérdida esperada de las consecuencias $Q(\bar{x}, \bar{\xi})$. Debemos, pues, definir una función de pérdida. Si el rango de variación de la expresión anterior no es demasiado grande, o si la decisión debe repetirse un gran número de veces, parece razonable asumir la indiferencia ante el riesgo del decisor y por tanto elegir una función de pérdida lineal. La solución del problema (4) que proponemos en estas condiciones es, pues, el resultado de resolver el programa

$$\begin{aligned} \min & \quad E_{\bar{\xi}}[Q(\bar{x}, \bar{\xi})] \\ \text{s. a.} & \quad \bar{x} \in X \end{aligned} \quad (5)$$

donde

$$Q(\bar{x}, \bar{\xi}) = \min_{\bar{y}^+, \bar{y}^-} \left\{ \sum_{i=1}^p (w_i^+ y_i^+ + w_i^- y_i^-), \mid f(\bar{x}) + \bar{y}^+ - \bar{y}^- = \bar{\xi}, \bar{y}^+, \bar{y}^- \geq \bar{0} \right\} \quad (6)$$

Por otro lado, los expertos en Programación Lineal Estocástica no tendrán dificultad en reconocer a los programas (5) y (6) como los resultantes de la aplicación de la metodología de la Programación Estocástica con Recursos al programa (4). De hecho, se trata del caso particular de la Programación Estocástica con Recursos Simples (véanse Kall y Wallace (1994), Birge y Louveaux (1997)).

A continuación estudiaremos algunas propiedades de interés de los programas (5) y (6).

3. 1. El programa (6)

$$Q(\bar{x}, \bar{\xi}) = \min_{\bar{y}^+, \bar{y}^-} \left\{ (w_i^+)' y_i^+ + (w_i^-)' y_i^- \mid \bar{y}^+ - \bar{y}^- = \bar{\xi} - f(\bar{x}), \bar{y}^+ \geq 0, \bar{y}^- \geq 0 \right\}$$

al que, por analogía con el Problema de Recursos Simples, podemos denominar *Programa de Segunda Fase*, es evidentemente factible $\forall \bar{x} \in X$ y $\forall \bar{\xi} \in \Xi$ fijos. Además es acotado inferiormente, por lo tanto siempre tiene solución.

Además la solución óptima verifica que $y_i^+ \cdot y_i^- = 0$, $i = 1, \dots, p$, por lo que al menos uno de estos dos términos $y_i^+, y_i^- = 0$, es cero $i = 1, \dots, p$. En efecto: $\bar{x}^*, \bar{y}^{+*}, \bar{y}^{-*}$ sean los vectores óptimos del programa (6). Supongamos que $\exists j \in \{1, 2, \dots, p\} / y_j^{+*} > 0, y_j^{-*} > 0$. Definamos $\tilde{y}_j^* = y_j^{+*} - \delta > 0$, $\tilde{y}_j^{-*} = y_j^{-*} - \delta > 0$, para un δ suficientemente pequeño.

Sean

$$\tilde{y}_i^+ = \begin{cases} y_i^{+*} & \forall i \neq j \\ \tilde{y}_j^+ & \text{para } i = j \end{cases} \quad \tilde{y}_i^- = \begin{cases} y_i^{-*} & \forall i \neq j \\ \tilde{y}_j^- & \text{para } i = j \end{cases}$$

entonces, puesto que $\bar{w} > \bar{0}$, es decir, $w_i^+, w_i^- > 0$, $\forall i = 1, \dots, p$, siendo al menos una de las desigualdades estricta, se verifica que

$$\bar{w}'(\hat{y}^+ + \hat{y}^-) < \bar{w}'(\bar{y}^{+*} + \bar{y}^{-*}) \quad y \quad f(\bar{x}^*) - (\hat{y}^+ + \hat{y}^-) = \bar{\xi}$$

en contradicción con que $\bar{y}^{+*}, \bar{y}^{-*}$ son óptimos.

3.2. Si X es un conjunto convexo, para cada $\bar{\xi} \in \Xi$ fijo, la función $Q(\circ, \bar{\xi})$ es convexa. En efecto: sean $\bar{x}^1, \bar{x}^2 \in X$, $\bar{x}^3 = \lambda \bar{x}^1 + (1 - \lambda) \bar{x}^2$ con $\lambda \in [0, 1]$. Si las soluciones del problema de segunda fase son $(\bar{y}^{1+}, \bar{y}^{1-})$ para $\bar{x} = \bar{x}^1$ y $(\bar{y}^{2+}, \bar{y}^{2-})$ para $\bar{x} = \bar{x}^2$, tenemos que:

$$\begin{aligned} Q(\bar{x}^1, \bar{\xi}) &= (\bar{w}^+)' \bar{y}^{1+} + (\bar{w}^-)' \bar{y}^{1-} \\ Q(\bar{x}^2, \bar{\xi}) &= (\bar{w}^+)' \bar{y}^{2+} + (\bar{w}^-)' \bar{y}^{2-} \end{aligned}$$

Entonces

$$(\lambda \bar{y}^{1+} + (1 - \lambda) \bar{y}^{2+}, \lambda \bar{y}^{1-} + (1 - \lambda) \bar{y}^{2-})$$

es factible para $\bar{x} = \bar{x}^3$. En efecto, es obvio que

$$\begin{aligned} \lambda \bar{y}^{1+} + (1 - \lambda) \bar{y}^{2+} &\geq \bar{0} \\ \lambda \bar{y}^{1-} + (1 - \lambda) \bar{y}^{2-} &\geq \bar{0}. \end{aligned}$$

Además

$$\begin{aligned} f(\bar{x}^3) + \lambda \bar{y}^{1-} + (1 - \lambda) \bar{y}^{2-} - \lambda \bar{y}^{1+} + (1 - \lambda) \bar{y}^{2+} \\ &= f(\lambda \bar{x}^1 + (1 - \lambda) \bar{x}^2) + \lambda \bar{y}^{1-} + (1 - \lambda) \bar{y}^{2-} - \lambda \bar{y}^{1+} + (1 - \lambda) \bar{y}^{2+} \\ &= \lambda[f(\bar{x}^1 + \bar{y}^{1-} - \bar{y}^{1+})] + (1 - \lambda)[f(\bar{x}^2 + \bar{y}^{2-} - \bar{y}^{2+})] \\ &= \lambda \bar{\xi} + (1 - \lambda) \bar{\xi} = \bar{\xi} \end{aligned}$$

Por otro lado

$$\begin{aligned} Q(\bar{x}^3, \bar{\xi}) &= Q(\lambda \bar{x}^1 + (1 - \lambda) \bar{x}^2, \bar{\xi}) \leq \\ &\leq (w^+)' [\lambda \bar{y}^{1+} + (1 - \lambda) \bar{y}^{2+}] + (w^-)' [\lambda \bar{y}^{1-} + (1 - \lambda) \bar{y}^{2-}] = \\ &= \lambda [(w^+)' \bar{y}^{1+} + (w^-)' \bar{y}^{1-}] + (1 - \lambda) [(w^+)' \bar{y}^{2+} + (w^-)' \bar{y}^{2-}] = \\ &= \lambda Q(\bar{x}^1, \bar{\xi}) + (1 - \lambda) Q(\bar{x}^2, \bar{\xi}) \end{aligned}$$

3.3. Si el dominio Ξ de $\bar{\xi}$ es un conjunto convexo, entonces para cada $\bar{x} \in X$ fijo, la función $Q(\bar{x}, \circ)$, es convexa.

En efecto, sean $\bar{\xi}^1, \bar{\xi}^2 \in \Xi$, con $\bar{\xi}^3 = \lambda\bar{\xi}^1 + (1 - \lambda)\bar{\xi}^2$ con $\lambda \in [0, 1]$. Si las soluciones del problema de segunda fase son $(\bar{y}^{1+}, \bar{y}^{1-})$ para $\bar{\xi} = \bar{\xi}^1$ y $(\bar{y}^{2+}, \bar{y}^{2-})$ para $\bar{\xi} = \bar{\xi}^2$, se debe verificar:

$$Q(\bar{x}, \bar{\xi}^1) = (\bar{w}^+)' \bar{y}^{1+} + (\bar{w}^-)' \bar{y}^{1-}, \bar{y}^{1-} - \bar{y}^{1+} = \bar{\xi}^1 - f(\bar{x})$$

$$Q(\bar{x}, \bar{\xi}^2) = (\bar{w}^+)' \bar{y}^{2+} + (\bar{w}^-)' \bar{y}^{2-}, \bar{y}^{2-} - \bar{y}^{2+} = \bar{\xi}^2 - f(\bar{x})$$

Entonces $(\lambda\bar{y}^{1+} + (1 - \lambda)\bar{y}^{2+}, \lambda\bar{y}^{1-} + (1 - \lambda)\bar{y}^{2-})$, es factible para $\bar{\xi} = \bar{\xi}^3$. En efecto, es evidente que $\lambda\bar{y}^{1+} + (1 - \lambda)\bar{y}^{2+} \geq \bar{0}$ y $\lambda\bar{y}^{1-} + (1 - \lambda)\bar{y}^{2-} \geq \bar{0}$. Además

$$\begin{aligned} & \lambda\bar{y}^{1-} + (1 - \lambda)\bar{y}^{2-} - \lambda\bar{y}^{1+} - (1 - \lambda)\bar{y}^{2+} = \\ & \lambda(\bar{y}^{1-} - \bar{y}^{1+}) + (1 - \lambda)(\bar{y}^{2-} - \bar{y}^{2+}) = \\ & \lambda(\bar{\xi}^1 - f(\bar{x})) + (1 - \lambda)(\bar{\xi}^2 - f(\bar{x})) = \\ & \lambda\bar{\xi}^1 + (1 - \lambda)\bar{\xi}^2 - f(\lambda\bar{x} + (1 - \lambda)\bar{x}) = \\ & \bar{\xi}^3 - f(\bar{x}) \end{aligned}$$

Finalmente,

$$\begin{aligned} Q(\bar{x}, \bar{\xi}^3) &= Q(\bar{x}, \lambda\bar{\xi}^1 + (1 - \lambda)\bar{\xi}^2) \leq \\ &\leq (w^+)'(\lambda\bar{y}^{1+} + (1 - \lambda)\bar{y}^{2+}) + (w^-)'(\lambda\bar{y}^{1-} + (1 - \lambda)\bar{y}^{2-}) = \\ &= \lambda[(w^+)' \bar{y}^{1+} + (w^-)' \bar{y}^{1-}] + (1 - \lambda)[(w^+)' \bar{y}^{2+} + (w^-)' \bar{y}^{2-}] = \\ &= \lambda Q(\bar{x}, \bar{\xi}^1) + (1 - \lambda)Q(\bar{x}, \bar{\xi}^2) \end{aligned}$$

3.4. Si X es un conjunto convexo, el programa (5):

$$\begin{array}{ll} \min & E_{\bar{\xi}}[Q(\bar{x}, \bar{\xi})] \\ s.a. & \bar{x} \in X \end{array}$$

también es convexo.

En efecto, X es un conjunto convexo por hipótesis. Además para cada $\bar{\xi} \in \Xi$, $\bar{\xi}$ fijo, se verifica que $Q(\lambda\bar{x}^1 + (1 - \lambda)\bar{x}^2, \bar{\xi}) = \lambda Q(\bar{x}^1, \bar{\xi}) + (1 - \lambda)Q(\bar{x}^2, \bar{\xi})$. Luego

$$\begin{aligned} E[Q(\lambda\bar{x}^1 + (1 - \lambda)\bar{x}^2), \bar{\xi}] &\leq \\ &\leq E[\lambda Q(\bar{x}^1, \bar{\xi}) + (1 - \lambda)Q(\bar{x}^2, \bar{\xi})] = \\ &= \lambda E[Q(\bar{x}^1, \bar{\xi})] + (1 - \lambda)E[Q(\bar{x}^2, \bar{\xi})] \end{aligned}$$

Como consecuencia, un modo de resolver el problema (5) es utilizar los algoritmos de programación convexa.

3.5. Es evidente que la función

$$\begin{aligned} Q(\bar{x}, \bar{\xi}) = & \min_{\bar{y}^+, \bar{y}^-} \{ \bar{w}^+ \bar{y}^+ + \bar{w}^- \bar{y}^- / \bar{y}^- - \bar{y}^+ = \bar{\xi} - f(\bar{x}), \bar{y}^+ \geq \bar{0}, \bar{y}^- \geq \bar{0} \} = \\ & \min_{\bar{y}^+, \bar{y}^-} \sum_{i=1}^p (w^{+i} \bar{y}^+ + w^{-i} \bar{y}^-) / \bar{y}^- - \bar{y}^+ = \xi_i - f_i(\bar{x}), \bar{y}^+ \geq 0, \bar{y}^- \geq 0, i = 1, \dots, p \end{aligned}$$

se puede obtener a partir de las soluciones de los p programas asociados en los que solamente se considera un objetivo en cada programa

$$\begin{array}{ll} \min & w^+ \bar{y}^+ + w^- \bar{y}^- \\ \text{s. a.} & \bar{y}^- - \bar{y}^+ = \xi_i - f_i(\bar{x}), \quad i = 1, \dots, p \\ & \bar{y}^+ \geq 0, \bar{y}^- \geq 0 \end{array} \quad (7)$$

Veamos qué forma tiene dicha solución. Por la teoría de la dualidad de la programación lineal (véase Balbás y Gil (1990)), sabemos que, dado un programa en la forma estándar

$$\begin{array}{ll} \min & \bar{c}' \bar{x} \\ \text{s. a.} & A \bar{x} = \bar{b} \\ & \bar{x} \geq \bar{0} \end{array}$$

su dual es

$$\begin{array}{ll} \max & \bar{\lambda}' \bar{b} \\ \text{s. a.} & \bar{\lambda}' A \leq \bar{c} \end{array}$$

Nuestro programa es

$$\begin{array}{ll} \min_{\bar{y}^+, \bar{y}^-} & w_1^+ y_1^+ + w_1^- y_1^- + \dots + w_p^+ y_p^+ + w_p^- y_p^- \\ \text{s. a.} & y_1^- - y_1^+ = \xi_1 - f_1(\bar{x}) \\ & \dots \\ & y_p^- - y_p^+ = \xi_p - f_p(\bar{x}) \\ & \bar{y}^+, \bar{y}^-, \bar{x} \geq 0 \end{array}$$

Por lo tanto su dual es

$$\begin{array}{ll} \max & \lambda_1 (\xi_1 - f_1(\bar{x})) + \dots + \lambda_p (\xi_p - f_p(\bar{x})) \\ \text{s.a.} & (\lambda_1, \dots, \lambda_p) \begin{pmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix} \leq \begin{pmatrix} w_1^+ \\ w_1^- \\ \vdots \\ w_p^+ \\ w_p^- \end{pmatrix} \end{array}$$

siendo las restricciones equivalentes a

$$\begin{cases} w_1^- \leq \lambda_1 \leq w_1^+ \\ \vdots \\ w_p^- \leq \lambda_p \leq w_p^+ \end{cases}$$

Sea

$$Q_i(f_i(\bar{x}), \xi_i) = \max \left\{ \lambda_i (\xi_i - f_i(\bar{x}) / \bar{w}_i^- \leq \lambda_i \leq \bar{w}_i^+ \right\}, i = 1, 2, \dots, p \quad (8)$$

La solución $\lambda_i^*, i = 1, 2, \dots, p$, de estos p programas (8) es, obviamente

- $(\xi_i - f_i(\bar{x}) < 0, \lambda_i^*(\xi_i, f_i(\bar{x})) = -w_i^-$
- $(\xi_i - f_i(\bar{x}) > 0, \lambda_i^*(\xi_i, f_i(\bar{x})) = w_i^+$
- $(\xi_i - f_i(\bar{x}) = 0, \lambda_i^*(\xi_i, f_i(\bar{x}))$

toma cualquier valor en el intervalo $[-w_i^-, -w_i^+]$.

Por lo tanto las soluciones de (8) son para $i = 1, 2, \dots, p$

$$\lambda_i^* = \begin{cases} w_i^+ & \text{si } \xi_i - f_i(\bar{x}) \geq 0 \\ -w_i^- & \text{si } \xi_i - f_i(\bar{x}) \leq 0 \end{cases}$$

Es decir, para $i = 1, 2, \dots, p$,

$$Q_i(f_i(\bar{x}), \xi_i) = \begin{cases} w_i^+(\xi_i - f_i(\bar{x})) & \text{si } \xi_i \geq f_i(\bar{x}) \\ -w_i^-(\xi_i - f_i(\bar{x})) & \text{si } \xi_i \leq f_i(\bar{x}) \end{cases}$$

Finalmente, como consecuencia de las propiedades de la dualidad, podemos concluir que la solución buscada es

$$Q(\bar{x}, \bar{\xi}) = \sum_{i=1}^p Q_i(f_i(\bar{x}), \xi_i)$$

Como habíamos observado anteriormente, la función $Q(\bar{x}, \bar{\xi})$ resulta ser la suma de las soluciones de los p programas asociados (8) en los que únicamente se considera un objetivo cada vez. Así, si denominamos

$$Q(f(\bar{x}), \bar{\xi}) = \sum_{i=1}^p Q_i(f_i(\bar{x}), \xi_i), Q(\bar{x}) = E_{\bar{\xi}}[Q(\bar{x}, \bar{\xi})],$$

$$Q(f(\bar{x})) = E_{\bar{\xi}}[Q(f(\bar{x}), \bar{\xi})],$$

podemos escribir que

$$E_{\bar{\xi}}[Q(\bar{x}, \bar{\xi})] = Q(\bar{x}) = Q(f(\bar{x})) = E_{\bar{\xi}}[Q(f(\bar{x}), \bar{\xi})]$$

Consecuentemente la función objetivo del problema (5)

$$\begin{array}{ll} \min & E_{\bar{\xi}}[Q(\bar{x}, \bar{\xi})] \\ \text{s. a.} & \bar{x} \in X \end{array}$$

se puede escribir con cualquiera de las expresiones de la última igualdad. (Observación: no se deben confundir $Q(\bar{x})$ y $Q(f(\bar{x}))$ ya que sus dominios son \mathbb{R}^n y \mathbb{R}^p respectivamente).

Hemos demostrado anteriormente que la función objetivo del problema (5) es convexa. A continuación vamos a comprobar que, cuando la distribución de probabilidad de la variable aleatoria $\bar{\xi}$ es discreta, el problema (5) es equivalente a un problema de programación lineal. Lo haremos en la siguiente propiedad.

3.6. Si la distribución de probabilidad conjunta de la variable aleatoria $\bar{\xi}$ es una distribución discreta finita, es decir, $\bar{\xi}$ toma los valores $\bar{\xi}_1, \dots, \bar{\xi}_p$, con probabilidades p_1, \dots, p_p , entonces el problema (5)

$$\begin{array}{ll} \min & E_{\bar{\xi}}[Q(\bar{x}, \bar{\xi})] \\ \text{s.a.} & \bar{x} \in X \end{array}$$

donde

$$Q(\bar{x}, \bar{\xi}) = \min_{\bar{y}^+, \bar{y}^-} \left\{ \sum_{i=1}^p (w_i^+ \bar{y}_i^+ + w_i^- \bar{y}_i^-), \text{s. a. } f(\bar{x}) + \bar{y}^- - \bar{y}^+ = \bar{\xi}, \bar{y}^+ \geq \bar{0}, \bar{y}^- \geq \bar{0} \right\}$$

tiene la forma:

$$\begin{array}{ll} \min & E_{\bar{\xi}}[Q(\bar{x}, \bar{\xi})] = p_1 Q(\bar{x}, \bar{\xi}_1) + \dots + p_p Q(\bar{x}, \bar{\xi}_p) \\ \text{s. a.} & \bar{x} \in X \end{array}$$

donde, para $i = 1, \dots, p$,

$$Q(\bar{x}, \bar{\xi}_i) = \min_{\bar{y}_i^+, \bar{y}_i^-} \left\{ (w_i^+ \bar{y}_i^+ + w_i^- \bar{y}_i^-) / f(\bar{x}) + \bar{y}_i^- - \bar{y}_i^+ = \bar{\xi}_i, \bar{y}^+, \bar{y}^- \geq \bar{0} \right\}.$$

Problema que denotaremos como (9).

Este problema es equivalente al programa lineal

$$\begin{array}{ll} \min_{\bar{x}, \bar{y}_i^+, \bar{y}_i^-} & \sum_{i=1}^p p_i (\bar{w}_i^+ \bar{y}_i^+ + \bar{w}_i^- \bar{y}_i^-) \\ \text{s. a.} & f(\bar{x}) + \bar{y}_i^- - \bar{y}_i^+ = \bar{\xi}_i \\ & \bar{y}_i^+, \bar{y}_i^- \geq \bar{0} \\ & \bar{x} \in X \end{array} \quad (9)$$

Demostración

a) Toda solución de (10) lo es de (9). En efecto, supongamos que $p_i > 0, \forall i = 1, \dots, p$. Sean $\tilde{x}_i, \tilde{\bar{y}}_i^+, \tilde{\bar{y}}_i^-, \forall i = 1, \dots, p$ soluciones de (10). Entonces, $\forall i = 1, \dots, p Q(\bar{x}_i, \bar{\xi}_i) = \bar{w}^+ \tilde{\bar{y}}_i^+ + \bar{w}^- \tilde{\bar{y}}_i^-$, es decir, no existen $\tilde{\bar{y}}_j^+, \tilde{\bar{y}}_j^- \geq 0$ para algún tales que

$$f(\bar{x}_1) + \tilde{\bar{y}}_j^+ - \tilde{\bar{y}}_j^- = \bar{\xi}_i$$

$$\bar{w}^+ \tilde{\bar{y}}_j^+ + \bar{w}^- \tilde{\bar{y}}_j^- < \bar{w}^+ \tilde{\bar{y}}_j^+ + \bar{w}^- \tilde{\bar{y}}_j^-$$

porque si para algún j existieran tales $\tilde{\bar{y}}_j^+, \tilde{\bar{y}}_j^-$, entonces la solución de (10) no sería

$$\bar{x}_1, \tilde{\bar{y}}_1^+, \dots, \tilde{\bar{y}}_p^+, \tilde{\bar{y}}_1^-, \dots, \tilde{\bar{y}}_p^-$$

sino

$$\bar{x}_1, \tilde{\bar{y}}_1^+, \dots, \tilde{\bar{y}}_{j-1}^+, \tilde{\bar{y}}_j^+, \tilde{\bar{y}}_{j+1}^+, \dots, \tilde{\bar{y}}_p^+, \tilde{\bar{y}}_1^-, \dots, \tilde{\bar{y}}_{j-1}^-, \tilde{\bar{y}}_j^-, \tilde{\bar{y}}_{j+1}^-, \dots, \tilde{\bar{y}}_p^-,$$

Luego si $, \bar{x}_1, \tilde{\bar{y}}_1^+, \tilde{\bar{y}}_1^-, i = 1, \dots, p$ es la solución de (10), entonces

$$\sum_{i=1}^p p_i (\bar{w}^+ \tilde{\bar{y}}_i^+ + \bar{w}^- \tilde{\bar{y}}_i^-) = \sum_{i=1}^p p_i Q(\bar{x}_1, \bar{\xi}_i)$$

y \bar{x}_1 resulta ser solución de (9), ya que si existe $\bar{x}_0 \in X$, tal que

$$p_1 Q(\bar{x}_0, \bar{\xi}_1) + \dots + p_9 Q(\bar{x}_0, \bar{\xi}_p) <$$

$$p_1 Q(\bar{x}_1, \bar{\xi}_1) + \dots + p_9 Q(\bar{x}_1, \bar{\xi}_p)$$

entonces, existirán $\hat{\bar{y}}_i^+, \hat{\bar{y}}_i^- \geq 0, i = 1, \dots, p$, tales que

$$f(\bar{x}_0) + \hat{\bar{y}}_i^- - \hat{\bar{y}}_i^+ = \bar{\xi}_i, \quad i = 1, \dots, p$$

$$\sum_{i=1}^p p_i (\bar{w}^+ \hat{\bar{y}}_i^+ + \bar{w}^- \hat{\bar{y}}_i^-) \leq \sum_{i=1}^p p_i (\bar{w}^+ \tilde{\bar{y}}_i^+ + \bar{w}^- \tilde{\bar{y}}_i^-)$$

y por lo tanto $\bar{x}_1, \hat{\bar{y}}_i^+, \hat{\bar{y}}_i^-, i = 1, \dots, p$, no sería solución de (10).

b) Toda solución de (9) lo es de (10). En efecto, sea \bar{x}_0 solución de (9) y sea

$$Q(\bar{x}_0, \bar{\xi}_i) = \bar{w}^+ \hat{\bar{y}}_i^+ + \bar{w}^- \hat{\bar{y}}_i^-.$$

Entonces,

$$\sum_{i=1}^p p_i Q(\bar{x}_0, \bar{\xi}_i) \leq \sum_{i=1}^p p_i Q(\bar{x}, \bar{\xi}_i), \quad \forall \bar{x} \in X,$$

lo que significa que se verifica que

$$\sum_{i=1}^p p_i (\bar{w}^+ \hat{\bar{y}}_i^+ + \bar{w}^- \hat{\bar{y}}_i^-) \leq \sum_{i=1}^p p_i (\bar{w}^+ \bar{y}_i^+ + \bar{w}^- \bar{y}_i^-), \quad \forall \bar{x} \in X,$$

siendo $\bar{y}_i^+, \bar{y}_i^- \geq 0$ tales que $f(\bar{x}_0) + \hat{\bar{y}}_i^- - \hat{\bar{y}}_i^+ = \bar{\xi}_i, \forall i = 1, \dots, p$. Pero esto es equivalente a decir que $\bar{x}_0, \hat{\bar{y}}_i^+, \hat{\bar{y}}_i^-, \forall i = 1, \dots, p$, es la solución de (10).

4 Una aplicación al diseño de sistemas Bonus-Malus

En este apartado aplicaremos los resultados anteriores a la resolución de un problema de tarificación en el marco de la Matemática Actuarial, el problema del diseño de Sistemas Bonus-Malus óptimos. Como es bien sabido, las compañías de seguros a menudo incorporan en sus sistemas de tarificación algunas reglas que modifican las primas de acuerdo con la experiencia de siniestros de los asegurados, de forma que los asegurados que tienen un mayor número de siniestros paguen asimismo una prima mayor. Una forma muy habitual de llevar a cabo estas penalizaciones consiste en establecer un número finito de posibles tipos o clases de asegurados asociados con diferentes tarifas, de forma que los asegurados con más siniestros tiendan a permanecer en las clases con mayores primas, y viceversa. En tal caso, se dice que la compañía ha establecido un Sistema de tarificación Bonus-Malus. Evidentemente, los asegurados pueden cambiar de clase de acuerdo con ciertas reglas de transición. El diseño de un Sistema Bonus-Malus requiere establecer el número de clases, las reglas de transición entre ellas, la clase inicial y la escala de primas. En este artículo solamente nos referiremos a la elección de una escala de primas que sea óptima respecto de algún criterio. Este problema ha recibido mucha atención en la literatura, siendo Norberg (1976) quien estableció el método más importante para el cálculo de una escala de primas óptima, la denominada *escala de Bayes*. La importancia de la escala de Bayes se basa fundamentalmente en que garantiza que el sistema Bonus-Malus tenga la importante propiedad de *equilibrio financiero*, es decir, garantiza que la cuantía total de los ingresos por primas de la compañía sea suficiente para hacer frente a los siniestros, en términos de esperanza matemática, lo cual resulta esencial para la supervivencia de la empresa en el largo plazo. Se trata, sin duda, de una propiedad importantísima. Sin embargo, en la práctica pueden resultar interesantes otras propiedades adicionales que ya no están garantizadas en la escala de Bayes. Por ejemplo, la compañía de seguros puede desear suavizar el sistema de penalizaciones, en el caso de que el sistema original resulte excesivamente duro y muchos asegurados prefieran cambiar de compañía. En los siguientes apartados veremos

que la metodología de Programación Estocástica por Metas permite diseñar escalas de primas óptimas que pueden tener en cuenta propiedades interesantes como la que acabamos de comentar, siempre que se representen matemáticamente mediante restricciones lineales. Por supuesto, si el decisor impone un gran número de tales propiedades, el programa matemático resultante puede ser infactible, por lo cual el decisor deberá eliminar algunas de ellas o al menos suavizarlas. El proceso de diseño de un Sistema Bonus-Malus se convierte así en un proceso interactivo, con el objetivo de alcanzar un grado razonable de compromiso entre un gran número de propiedades que probablemente sean incompatibles entre sí.

5 Definición de un sistema Bonus-Malus

Consideremos una cartera de pólizas en la que existe heterogeneidad debido a ciertas características inobservables de las pólizas. Como es usual en la literatura actuaria, supondremos que las características respecto al riesgo de cada póliza están resumidas en el valor de cierto parámetro λ , que habitualmente se identifica con la frecuencia media de siniestros. Supondremos asimismo que este número medio de siniestros se mantiene constante a lo largo del tiempo, y que es independiente de la cuantía de los mismos. Si tomamos como unidad monetaria la cuantía media de un siniestro, nuestro objetivo consistirá en diseñar un sistema para calcular una prima (pura) para cada asegurado lo más cercana posible al verdadero valor de su parámetro (λ) (que es evidentemente desconocido). Intentaremos conseguir este objetivo mediante el diseño de un Sistema de Tarificación Bonus-Malus (en adelante, SBM). Tales sistemas son habitualmente utilizados por las compañías para la tarificación en el seguro del automóvil, dependiendo en la gran mayoría de los casos del número de siniestros de cada asegurado y no de su cuantía (Lemaire (1985 y 1995)). Supondremos, finalmente, que el parámetro de riesgo Λ es una variable aleatoria cuya función de distribución $U(\lambda)$ (denominada *función de estructura*) es conocida.

Siguiendo a Lemaire (1995, p. 6), diremos que una compañía de seguros utiliza un SBM cuando se verifican las siguientes condiciones:

- Existe un número finito de clases $\{C_1, \dots, C_n\}$ tales que cada póliza permanece en una clase durante todo el periodo asegurado (usualmente un año).
- La prima calculada para cada póliza depende únicamente de la clase en la que permanece.
- La clase para cada periodo está determinada por la clase en el periodo anterior y el número de siniestros durante dicho periodo.

Cada SBM está definido por tres elementos:

- *La clase inicial*, donde son asignadas las nuevas pólizas.

- *La escala de primas* (P_1, \dots, P_n), donde P_i es la prima asignada a las pólizas de la clase C_i .
- *Las reglas de transición*, es decir, las reglas que establecen las condiciones bajo las cuales las pólizas de la clase C_i se transfieren a la clase C_j durante el siguiente periodo.

Las reglas de transición se definen habitualmente mediante transformaciones T_k tales que $T_k(i) = j$ cuando las pólizas de la clase C_i que declaran k siniestros son transferidas a la clase C_j durante el siguiente periodo. Las transformaciones T_k se definen mediante matrices,

$$T_k = (t_{ij}^k)$$

tales que

$$t_{ij}^k = 1 \quad \text{si} \quad T_k(i) = j$$

$$t_{ij}^k = 0 \quad \text{si} \quad T_k(i) \neq j$$

La *probabilidad condicional de transición* de la clase C_i a la C_j en un periodo, supuesto que $\Lambda = \lambda$, se calcula como

$$p_{ij}(\lambda) = \sum_{k=0}^{\infty} p_k(\lambda) t_{ij}^k$$

donde $p_k(\lambda)$ es la probabilidad de declarar k siniestros en un periodo, condicionada a que $\Lambda = \lambda$. Es decir,

$$p_k(\lambda) = Pr[N = k / \Lambda = \lambda]$$

La matriz de probabilidades condicionadas, supuesto que, se define como

$$P(\lambda) = (p_{ij}(\lambda))$$

Estas definiciones permiten considerar al SBM como una *cadena de Markov*. La cadena será *homogénea*, ya que hemos supuesto que cada frecuencia media λ es estacionaria respecto al tiempo. La matriz de transición $P(\lambda)$ que acabamos de definir, será la matriz de transición de la correspondiente cadena de Markov.

Si suponemos además que la cadena es regular, entonces es posible asegurar (véase, por ejemplo, Kemeny y Snell (1976)) que existe una distribución estacionaria de probabilidades condicionales ($\pi_1(\lambda), \dots, \pi_n(\lambda)$), donde $\pi_i(\lambda)$ se define como el valor límite (cuando el número de periodos tiende a infinito) de la probabilidad de que una póliza pertenezca a la clase C_i , condicionada a que $\Lambda = \lambda$. Se puede demostrar que la distribución estacionaria de probabilidad coincide con el autovector por la izquierda asociado con el autovalor 1 de la correspondiente matriz de transición, cuyas componentes suman la unidad.

Se puede definir también la distribución estacionaria de probabilidades (no condicionadas) (π_1, \dots, π_n) para una póliza arbitraria, como el valor medio de las correspondientes distribuciones estacionarias condicionadas $(\pi_1(\lambda), \dots, \pi_n(\lambda))$. Es decir,

$$\pi_1 = \int \pi_i(\lambda) dU(\lambda)$$

Evidentemente, π_1 y $\pi_1(\lambda)$ pueden interpretarse como las probabilidades de que una póliza arbitraria y una póliza condicionada a que $\Lambda = \lambda$, respectivamente, pertenezcan a la clase C_i cuando se alcanza la estacionariedad. El conocimiento de estas distribuciones estacionarias resulta muy útil a la hora de diseñar un SBM, ya que nos informa del comportamiento a largo plazo de las pólizas. De esta forma, añadiremos a nuestras hipótesis anteriores el supuesto de que el SBM ha alcanzado, o al menos se ha aproximado, a su estado estacionario.

5.1 Cálculo de la escala de primas mediante criterios asintóticos

El método estándar para calcular la escala de primas de un SBM, para un número de clases y reglas de transición dadas, es conocido como la *escala de Bayes* (Norberg (1976)). Dicha escala surge como solución de un problema de decisión bayesiano, en el que las decisiones factibles coinciden con las posibles escalas de primas (P_1, \dots, P_n) , los parámetros aleatorios son el verdadero valor del parámetro Λ y la clase C_i a la que pertenece la póliza al alcanzar el estado estacionario, y la función de pérdida es una función cuadrática del error de tarificación, de la forma $(P_i - \lambda)^2$. La escala óptima o escala de Bayes minimiza entonces el llamado *error cuadrático de tarificación esperado*,

$$\int_0^\infty \sum_{i=1}^n (P_i - \lambda)^2 \pi_i(\lambda) dU(\lambda)$$

siendo la solución de la forma

$$P_i = \frac{1}{\pi_i} \int_0^\infty \lambda \pi_i(\lambda) dU(\lambda)$$

Como hemos comentado anteriormente, la escala de Bayes constituye el método básico para la construcción de un SBM óptimo: se trata de una escala de primas fundamentada teóricamente, fácilmente calculable y con la importante propiedad de equilibrio financiero. Sin embargo, puede dar lugar a sorpresas desagradables: por ejemplo, a veces las primas resultantes no forman una secuencia monótona, lo que impide totalmente su implementación práctica; o, aún siendo monótonas, las bonificaciones o penalizaciones entre clases consecutivas pueden variar bruscamente o tomar valores excesivamente grandes o pequeños, lo que dificulta asimismo su uso en un problema real. Por supuesto, todas estas propiedades interesantes (y otras que no mencionamos aquí: véase Heras, Vilar y Gil (2002))

podrían incluirse como restricciones adicionales del problema de minimización que da lugar a la escala de Bayes, pero esto exigiría la resolución de un programa cuadrático con un gran número de restricciones. Y, lo que es más importante, la introducción de nuevas restricciones impedirá en general que el SBM resultante esté financieramente equilibrado, lo que elimina la principal razón esgrimida en la literatura en defensa de la escala de Bayes.

Como exponen Heras, Vilar y Gil (2002), la mayor parte de las propiedades razonables que debería cumplir la escala de primas de un SBM (incluida la propiedad de equilibrio financiero) se pueden representar matemáticamente por medio de restricciones lineales. Sería interesante, por tanto, que el programa matemático que origina la escala de primas fuese asimismo un programa lineal. Como veremos a continuación, esto se puede conseguir mediante las técnicas de Programación Estocástica por Metas expuestas en los apartados anteriores.

En efecto, en última instancia el problema de tarificación que estamos estudiando consiste en definir un conjunto de primas para las clases del SBM que estén lo más cerca posible del verdadero valor del parámetro Λ de cada asegurado, lo cual no tendría ninguna dificultad si conociéramos exactamente el valor de dicho parámetro. El problema radica, obviamente, en que no conocemos su valor, debiendo contentarnos con una distribución de probabilidad. En otras palabras, debemos elegir el valor de ciertas variables de decisión (las primas) de forma que nos acerquemos lo más posible al valor de un objetivo que varía aleatoriamente (el parámetro). Esto es, evidentemente, un problema que puede modelizarse mediante Programación Estocástica por Metas, siendo los valores de las variables de decisión \bar{x} iguales a las posibles escalas de primas $\bar{P} = (P_1, \dots, P_n)$, y siendo asimismo los valores aleatorios de los objetivos $\bar{\xi}$ iguales a los posibles valores del parámetro Λ . La escala óptima debería, según lo que hemos visto en apartados anteriores, minimizar la función (11)

$$\min_{\bar{P}} \{E_\lambda(Q(\bar{P}, \lambda)\}$$

siendo la función $Q(\bar{P}, \lambda)$ el resultado de resolver el programa de segunda fase (12):

$$\begin{aligned} Q(\bar{P}, \lambda) = & \min_{y_i^+, y_i^- \geq 0} \sum_{i=1}^n (y_i^+ + y_i^-) \pi_i(\lambda) \\ & \text{sujeto a :} \\ & P_1 + y_1^+ - y_1^- = \lambda \\ & \dots \\ & P_n + y_n^+ - y_n^- = \lambda \end{aligned}$$

No es difícil darse cuenta que los programas (11) y (12) también podrían haberse obtenido como solución de un programa bayesiano de decisión totalmente análogo al que da lugar a la escala de Bayes, con la única diferencia de que la función de pérdida no estaría definida como el cuadrado del error de tarificación, sino como su valor absoluto $|P_i - \lambda|$. La escala de primas que es solución óptima

de (11) y (12) minimiza, pues, el *error absoluto de tarificación esperado*, definido como

$$\int_0^\infty \sum_{i=1}^n |P_i - \lambda| \pi_i(\lambda)$$

Como sabemos, la resolución de los programas (11) y (12) puede ser un proceso complicado. Ahora bien, en la literatura actuarial es habitual trabajar con aproximaciones discretas de la función de estructura $U(\lambda)$. En tal caso, sabemos por la propiedad 3.9 que los programas (11) y (12) son equivalentes a un programa lineal. En efecto, si el parámetro Λ puede tomar únicamente los valores $(\lambda_1, , \lambda_m)$ con probabilidades $(q_1, , q_m)$, entonces los programas (11) y (12) equivalen al programa lineal (13):

$$\min \sum_{j=1}^m \sum_{i=1}^n (y_{ij}^+ + y_{ij}^-) \pi_i(\lambda) q_j$$

sujeto a :

$$\begin{aligned} P_1 + y_{11}^- + y_{11}^+ &= \lambda_1 \\ &\dots \\ P_1 + y_{1m}^- + y_{1m}^+ &= \lambda_m \\ &\dots \\ P_n + y_{n1}^- + y_{n1}^+ &= \lambda_1 \\ &\dots \\ cP_n + y_{nm}^- + y_{nm}^+ &= \lambda_m \\ P_i, y_{ij}^+, y_{ij}^- &\geq 0 \end{aligned}$$

La resolución del programa (13) se puede llevar a cabo fácilmente en un ordenador personal, añadiendo todas las restricciones (lineales) que el decisor considere conveniente.

Veamos un ejemplo sencillo. Supongamos que el parámetro Λ puede tomar únicamente tres valores, $\lambda_1 = 0.5$, $\lambda_2 = 1$, $\lambda_3 = 1.5$, todos ellos con probabilidad $1/3$, y que solamente existen tres clases de bonus-malus.

Supongamos asimismo que las reglas de transición implican las siguientes distribuciones condicionales estacionarias:

$$\begin{aligned} \pi_1(\lambda_1) &= \frac{3}{5}; & \pi_2(\lambda_1) &= \frac{1}{5}; & \pi_3(\lambda_1) &= \frac{1}{5} \\ \pi_1(\lambda_2) &= \frac{1}{3}; & \pi_2(\lambda_2) &= \frac{1}{2}; & \pi_3(\lambda_2) &= \frac{1}{6} \\ \pi_1(\lambda_3) &= \frac{1}{4}; & \pi_2(\lambda_3) &= \frac{1}{4}; & \pi_3(\lambda_3) &= \frac{1}{2} \end{aligned}$$

La distribución estacionaria incondicional será:

$$\begin{aligned} \pi_1 &= \left(\frac{3}{5} + \frac{1}{3} + \frac{1}{4} \right) \frac{1}{3} = \frac{71}{60} \frac{1}{3} \\ \pi_2 &= \left(\frac{1}{5} + \frac{1}{2} + \frac{1}{4} \right) \frac{1}{3} = \frac{19}{20} \frac{1}{3} \\ \pi_3 &= \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{2} \right) \frac{1}{3} = \frac{13}{15} \frac{1}{3} \end{aligned}$$

Supongamos que, por razones comerciales, deseamos obtener una escala de primas que penalice a los malos conductores con una prima cuya cuantía sea el doble que la de los buenos. Es decir, tal que $P_3 = 2P_2$. En tal caso deberíamos resolver el programa lineal (13), añadiendo la restricción anterior y, por supuesto, la referente al equilibrio financiero:

$$P_1\pi_1 + P_2\pi_2 + P_3\pi_3 = E(\Lambda) = 1$$

La escala de primas óptima, obtenida de la resolución del programa estocástico por metas (13) junto con las dos nuevas restricciones adicionales, resulta ser

$$\begin{aligned} P_1 &= 0.7031 \\ P_2 &= 1.0 \\ P_3 &= 1.4062 \end{aligned}$$

La metodología expuesta permitiría la introducción de todas las restricciones lineales adicionales que sean necesarias, con la única limitación de que el programa lineal resultante siga siendo factible.

6 Bibliografía

- [1] Balbás, A. y Gil, J. A. (1990). “Programación Matemática”. Editorial A.C. Madrid.
- [2] Birge, J. R. & Louveaux, F. (1997). “Introduction to Stochastic Programming”, Springer, New York.
- [3] Charnes, A., Cooper, W. W. & Ferguson, R.O. (1955). “Optimal Estimation of Executive Compensation by Linear Programming”, Management Science 1 (2), 138-151.
- [4] Dantzig, G.B. (1963). “Linear Programming and Extensions”. Princeton University Press, Nueva Jersey.
- [5] De Groot, M. H. (1970). “Optimal Statistical Decisions”. McGraw-Hill, New York.
- [6] Heras, A. & García Aguado, A. (1998). “Stochastic Goal Programming with Recourse”, Revista de la Real Academia de CC. Físicas, Exactas y Naturales 92 (4), 409-414.
- [7] Heras, A. & García Aguado, A. (1999). “Stochastic Goal Programming”, Central European Journal of Operations Research 7 (3), 139-158.

- [8] Heras, A., Vilar, J. L. & Gil, J. A. (2002). "Asymptotic Fairness of Bonus-Malus Systems and Optimal Scales of Premiums", *The Geneva Papers on Risk and Insurance Theory* 27, 61-82.
- [9] Kall, P. (1976). "Stochastic Linear Programming". Springer-Verlag, New York. 10.
Kall, P. & Wallace, S. W. (1994). "Stochastic Programming", John Wiley & Sons, New York.
- [10] Kemeny, J. G. & Snell, J. L. (1976). "Finite Markov Chains", Springer-Verlag, Berlin.
- [11] Kibzun, A. & Kan, Y. S. (1996). "Stochastic Programming Problems", John Wiley and Sons, Chichester.
- [12] Lemaire, J. (1985). "Automobile Insurance. Actuarial Models", Kluwer-Nijhoff Publishing, Dordrecht.
- [13] Lemaire, J. (1995). "Bonus-Malus Systems in Automobile Insurance", Kluwer Academic Publishers, Dordrecht.
- [14] Norberg, R. (1976). "A Credibility Theory for Automobile Bonus Systems", Scandinavian Actuarial Journal, 92-107.
- [15] Romero, C. (1991). "Handbook of Critical Issues in Goal Programming", Pergamon Press, Oxford.
- [16] Romero, C. (1993). "Teoría de la Decisión Multicriterio: conceptos, técnicas y aplicaciones", Alianza Universidad Textos, Madrid.
- [17] Romero, C. (2002). "Programación por Metas (Goal Programming): pasado, presente y futuro" en Caballero, R. & Fernández, G. (eds) "Toma de decisiones con criterios múltiples", monográfico 1 de Rect@, ASEPUA, Valencia.
- [18] Simon, H. A. (1957). "Models of Man" Wiley. New York.
- [19] Stancu-Minasian, I. M. (1984). "Stochastic Programming with Multiple Objective Functions". Reidel Publishing Company.

Planificación óptima de la actividad quirúrgica en hospitales públicos mediante un modelo de Programación Compromiso Posibilística

**J. Antomil Ibias^a, M. Arenas Parra^b, A. Bilbao Terol^c,
B. Pérez Gladish^c, M. V. Rodríguez Uría^c**

^a Dep.

^b Dep.

^c Dep.

1 Introducción

Las listas de espera sanitarias representan un grave problema que ha sido desde los años setenta objeto de debate entre los profesionales de la salud. La existencia de listas de espera puso en duda la correcta gestión de la sanidad española y movilizó a las Administraciones Sanitarias para solventar la situación lo antes posible. En la actualidad, la situación es lo suficientemente grave como para justificar la búsqueda conjunta de soluciones en todos los ámbitos: el político, el de los profesionales de la salud y el académico.

La existencia de listas de espera en la atención sanitaria es un problema que afecta al derecho de los ciudadanos a la protección de la salud reflejando un desajuste entre la demanda y la oferta de asistencia sanitaria, que implica una necesidad sanitaria no cubierta en el momento en que se solicita. Esta situación

supone un deterioro de la calidad asistencial que percibe el paciente que siente angustia ante el desconocimiento del tiempo de espera en la lista; pudiendo llegar a suponer, incluso, un detrimento de su salud.

Los factores que pueden influir en la generación y persistencia de las listas de espera son diversos: los avances en medicina e investigación médica que generan demanda de atención, el incremento en la esperanza de vida y la mayor disponibilidad de información que tienen los pacientes que hace más difícil en la actualidad que acepten demoras.

Los gobiernos de diferentes países han llevado a cabo actuaciones dirigidas a paliar las implicaciones sanitarias y sociales que comportan las listas de espera. Inicialmente los esfuerzos se centraron en reducir el volumen de las listas, entendiendo por *volumen* el número de pacientes en espera; sin embargo, en la actualidad la atención se centra en la disminución del tiempo de espera más que en su volumen, que en sí mismo no posee significado alguno.

Aunque son diversas las líneas de actuación propuestas para abordar el problema de la existencia de listas de espera, todas parten de la idea común de que el problema es de asignación de recursos y de su correcta gestión. La literatura en el ámbito de la sanidad, encaminada a la propuesta de soluciones, es escasa en relación a la magnitud del problema planteado. No existen prácticamente trabajos que afronten el tema desde una perspectiva cuantitativa, por ello es necesario disponer de instrumentos que permitan conocer, ordenar, gestionar, analizar, decidir y evaluar con fiabilidad, en tiempo real y según los requerimientos de cada momento, las formas de organización y funcionamiento de los hospitales públicos.

El problema de la reducción de las listas de espera no es estrictamente de financiación económica, sino principalmente, como ya hemos señalado, de gestión eficiente, lo entendemos como un problema de decisión con criterios múltiples. En este trabajo exponemos un instrumento de gestión de listas de espera quirúrgicas desarrollado con metodología multicriterio difusa y con los datos de un hospital de agudos del tipo II del INSALUD, extensible a cualquier otro hospital de similares características.

2 Metodología

Es habitual que el ser humano deba tomar decisiones sobre problemas de cierta complejidad que incluyen varios objetivos, que pueden ser total o parcialmente conflictivos entre sí, de manera que la mejora en cualquiera de ellos puede empeorar el valor de otros objetivos que son evaluados de acuerdo a múltiples criterios y donde no es evidente la mejor u óptima alternativa.

Hoy en día, superando los procesos clásicos de análisis que atienden a un solo criterio económico, se consideran múltiples criterios u objetivos en los procesos de planificación de entidades o empresas de interés público. Este tipo de análisis se recomienda, y a veces se impone por ley en algunos países, como obligatorio para

analizar problemas de aprovechamiento de recursos naturales como el agua, las minas o cualquier otro, respecto del que la población tenga expectativas explícitas; entre estos problemas está la provisión de los servicios sanitarios que constituye el núcleo de este trabajo.

Son muchos los problemas de naturaleza económica que se caracterizan porque en la elección de la mejor decisión se han de tener en cuenta varios criterios y, por tanto, se desea alcanzar más de un objetivo. La Programación Multiobjetivo y, en general, la teoría de la decisión multicriterio, se encarga de la resolución de problemas de este tipo y, por tanto, su aplicación a problemas económicos es clara.

La formalización, que normalmente implica una simplificación de la realidad, obliga al investigador a “elegir entre realizar desde el inicio una selección de elementos a considerar, para poder operar después con un instrumental preciso, o bien captar la realidad con toda su imprecisión y operar con estas informaciones borrosas, aún sabiendo que los resultados vendrán dados de manera imprecisa” (Kaufmann y Gil Aluja, (1986)). Al final la duda que como modelizadores se nos plantea, es elegir, como señalan estos autores “entre un modelo preciso pero que no refleja la realidad y un modelo vago pero más adecuado a la realidad”.

La Teoría de los Subconjuntos Difusos, debida a por Lofti A. Zadeh (1965) y la Teoría de la Posibilidad asociada a ella, nacen como una solución matemática a la multitud de problemas y situaciones de la vida real a los que las teorías clásicas -conjuntista o probabilística- no podían dar solución. Los subconjuntos difusos tienen la capacidad de modelizar modos de razonamiento no preciso, que juegan un papel esencial en la toma de decisiones racionales en entornos de incertidumbre e imprecisión. La Teoría de los Subconjuntos Difusos proporciona herramientas adecuadas para representar un problema de Programación Multiobjetivo Lineal con datos vagos/ imprecisos.

Hasta hace pocos años, se han utilizado herramientas de la Teoría de la Probabilidad para modelizar la incertidumbre que surge en los procesos de decisión. Sin embargo, en muchas ocasiones el Decisor no puede comportarse de una forma que pueda ser descrita mediante los axiomas de las probabilidades subjetivas. Es necesario hacer una distinción entre la incertidumbre que corresponde a hechos frecuencialmente estables, y que por tanto están sujetos a leyes del azar (con lo que admiten un tratamiento probabilístico) y la incertidumbre inherente a aquellos fenómenos a los que tienen que enfrentarse las ciencias sociales, en los que además de los hechos inciertos de la naturaleza surgen los que introduce el ser humano producto de su libertad y poder de imaginación. Aparece entonces la necesidad de recurrir a nuevas herramientas para modelizar la imprecisión.

Mediante el presente trabajo proponemos un instrumento de gestión de listas de espera quirúrgicas basándonos en los datos de un hospital público. El problema de las listas de espera es como ya hemos dicho, un problema de toma de decisiones racionales con presencia de criterios múltiples en un entorno de incertidumbre e

imprecisión . Se trata por tanto de un problema de Programación Multiobjetivo Lineal con datos vagos/imprecisos.

Para su resolución aplicaremos una metodología basada en la Programación Compromiso y en la Teoría de los Subconjuntos Difusos, la *Programación Compromiso Posibilística* (Arenas et al. (2004)) que nos permitirá manejar la imprecisión de los datos del modelo y obtener un *compromiso* entre los logros de eficiencia y la equidad.

2.1 Programación Compromiso Posibilística

Sea el siguiente problema de Programación Multiobjetivo Lineal con parámetros difusos:

$$\begin{aligned} \min \quad & \tilde{z} = (\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_k) = (\tilde{c}_1 x, \tilde{c}_2 x, \dots, \tilde{c}_k x) \\ \text{s.a.} \quad & x \in \xi(\tilde{A}, \tilde{b}) = \left\{ \begin{array}{l} \tilde{a}_i x \leq \tilde{b}_i, \quad i = 1, \dots, m \\ x \geq 0 \end{array} \right\} \end{aligned} \quad (\text{FP-MOLP})$$

donde $x^t = (x_1, x_2, \dots, x_k)$ es el vector de variables decisión (no difusas), $\tilde{c}^t = (\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_k)$ son los parámetros difusos de los k objetivos considerados, $\tilde{A} = [\tilde{a}_{ij}]_{m \times n}$ es la matriz de coeficientes tecnológicos difusos y $\tilde{b}^t = (\tilde{b}_1, \tilde{b}_2, \dots, \tilde{b}_m)$ son los términos independientes de las restricciones, también parámetros difusos.

Supondremos que los parámetros anteriores están descritos por números difusos representados por sus distribuciones de posibilidad, que el analista matemático estima a partir de la información que le proporciona el Decisor (Tanaka, (1987)).

El desconocimiento o imprecisión de los parámetros de las restricciones del modelo implican la necesidad de definir en este contexto el concepto de factibilidad de un vector de decisión x . Esta cuestión conlleva la necesidad de comparar números difusos. En este trabajo utilizaremos el método de ordenación de números difusos desarrollado por Jiménez (1996) manipulando los números difusos a través de sus intervalos y valores esperados, definidos por Heilpern (1992)¹.

El intervalo esperado de un vector difuso $\tilde{a}_i = (\tilde{a}_{i1}, \tilde{a}_{i2}, \dots, \tilde{a}_{in})$ es un vector cuyas componentes son los intervalos esperados de cada número difuso del vector \tilde{a}_i , es decir: $EI(\tilde{a}_i) = (EI(\tilde{a}_{i1}), EI(\tilde{a}_{i2}), \dots, EI(\tilde{a}_{in}))$.

Definición 13 *Dados dos números difusos \tilde{a} y \tilde{b} , definimos la relación difusa de preferencia $\mu_M(\tilde{a}, \tilde{b})$, \tilde{a} es preferido a \tilde{b} (\tilde{a} es menor que \tilde{b}) mediante la siguiente*

¹Dado un número difuso triangular $\tilde{a} = (a^L, a^C, a^R)$, obtendremos su intervalo esperado ($EI(\tilde{a})$) y su valor esperado ($EV(\tilde{a})$) de la manera siguiente:

$$EI(\tilde{a}) = [E_1^{\tilde{a}}, E_2^{\tilde{a}}] = \left[\frac{a^L + a^C}{2}, \frac{a^C + a^R}{2} \right], \quad EV(\tilde{a}) = \frac{E_1^{\tilde{a}} + E_2^{\tilde{a}}}{2}.$$

función de pertenencia:

$$\mu_M(\tilde{a}, \tilde{b}) = \begin{cases} 0 & \text{si } E_1^{\tilde{a}} > E_2^{\tilde{b}} \\ \frac{E_2^{\tilde{b}} - E_1^{\tilde{a}}}{E_2^{\tilde{a}} - E_1^{\tilde{a}} + E_2^{\tilde{b}} - E_1^{\tilde{b}}} & \text{si } \in [E_1^{\tilde{b}} - E_2^{\tilde{a}}, E_2^{\tilde{b}} - E_1^{\tilde{a}}] \\ 1 & \text{si } E_2^{\tilde{a}} > E_1^{\tilde{b}} \end{cases} \quad (16.1)$$

Si $\mu_M(\tilde{a}, \tilde{b}) \leq \beta$, con $\beta \in [0, 1]$ diremos que \tilde{a} es preferido a \tilde{b} al menos en un grado igual a β denotándolo por $\tilde{a} \leq_{\beta} \tilde{b}$. Si $\beta = \frac{1}{2}$ los valores esperados de \tilde{a} y \tilde{b} son iguales y por ello diremos que \tilde{a} y \tilde{b} son indiferentes. Teniendo en cuenta la definición 1 tenemos que:

$$(1 - \beta)E_1^{\tilde{a}} + \beta E_2^{\tilde{a}} \leq E_1^{\tilde{b}} + (1 - \beta)E_2^{\tilde{b}} \quad (16.2)$$

La consideración de la relación de preferencia difusa de Jiménez (1996) da lugar a la siguiente definición:

Definición 14 Un vector de decisión $x \in \mathbb{R}^n$, diremos que es β -factible para el problema (FP-MOLP) si verifica las restricciones al menos en un grado β . Es decir,

$$\tilde{a}_i x \leq_{\beta} \tilde{b}_i, \quad i = 1, \dots, m. \quad (16.3)$$

Las definiciones anteriores nos permiten plantear el siguiente problema β -paramétrico mediante el que resolveremos el problema inicial FP-MOLP:

$$\begin{aligned} \min \quad & \tilde{z} = (\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_k) = (\tilde{c}_1 x, \tilde{c}_2 x, \dots, \tilde{c}_k x) \\ \text{s.a.} \quad & x \in \chi(\beta) = \left\{ \begin{array}{l} x \in \mathbb{R}^n \quad | \quad \tilde{a}_i x \leq_{\beta} \tilde{b}_i, \quad i = 1, \dots, m \\ x \geq 0 \end{array} \right\} \quad (\beta\text{-FP-MOLP}) \end{aligned}$$

Teniendo en cuenta la definición 1, este problema es equivalente a:

$$\begin{aligned} \min \quad & \tilde{z} = (\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_k) = (\tilde{c}_1 x, \tilde{c}_2 x, \dots, \tilde{c}_k x) \\ \text{s.a.} \quad & \left. \begin{array}{l} (1 - \beta)E_1^{\tilde{a}_i} + \beta E_2^{\tilde{a}_i} \leq E_1^{\tilde{b}_i} + (1 - \beta)E_2^{\tilde{b}_i}, \quad i = 1, \dots, m \\ x \geq 0 \end{array} \right\} = \chi(\beta) \end{aligned}$$

Para poder aplicar la Programación Compromiso a la resolución del problema, necesitamos obtener en primer lugar la solución ideal difusa del problema β -FP-MOLP. Nos basaremos para ello en el método de resolución de un problema mono-objetivo lineal con parámetros difusos (β -FLP), propuesto por Arenas *et al.* (1998):

$$\begin{aligned} \min \quad & \tilde{z} = \tilde{c}x \\ \text{s.a.} \quad & x \in \chi(\beta) \quad (\beta\text{-FLP}) \end{aligned}$$

Este método proporciona una solución difusa, en el espacio de objetivos definida por su distribución de posibilidad; está basado en el principio de extensión (Zadeh, (1975)) y en la distribución de posibilidad conjunta de los parámetros difusos del modelo y se basa en los α -cortes de la solución para la obtención de su distribución de posibilidad (Arenas et al. (1998), (1999a), (1999b)). En el citado trabajo se ha probado que la solución $\tilde{z}^*(\beta)$ del problema (β -FLP) es un número difuso. Basándonos en estos resultados definimos a continuación el concepto de solución -ideal difusa del problema FP-MOLP:

Definición 15 Una solución β -ideal difusa del problema FP-MOLP en el espacio de objetivos, $\tilde{z}^*(\beta) = (\tilde{z}_1^*(\beta), \tilde{z}_2^*(\beta), \dots, \tilde{z}_k^*(\beta))$, es un vector cuyas componentes son números difusos obtenidos resolviendo k β -FLP problemas.

Una vez obtenida la solución ideal difusa, trataremos de encontrar un vector de decisión x , con un grado de factibilidad fijado por el Decisor, que determine una solución difusa $\tilde{c}_r x$ lo más próxima posible a la solución ideal difusa $z_r^*(\beta)$. El problema a resolver será por lo tanto el siguiente:

Hallar un $x \in \chi(\beta)$ tal que:

$$EI(\tilde{c}_r x) \xrightarrow{\sim} EI(z_r^*(\beta)), \quad r = 1, \dots, k, \quad (16.4)$$

siendo $EI(\tilde{c}_r x) = [E_1^{\tilde{c}_r x}, E_2^{\tilde{c}_r x}]$ y $EI(z_r^*(\beta)) = [E_1^{z_r^*(\beta)}, E_2^{z_r^*(\beta)}]$.

Definición 16 (Arenas et al. (2004)). Dados dos números difusos $\tilde{c}_r x$ y $\tilde{z}_r^*(\beta)$ representados por sus intervalos esperados, llamaremos conjunto discrepancia entre ambos y lo denotaremos por $CD_r(\tilde{c}_r x, \tilde{z}_r^*(\beta))$, al siguiente conjunto de números reales no negativos:

$$CD_r(\tilde{c}_r x, \tilde{z}_r^*(\beta)) = \left\{ \left| E_1^{\tilde{z}_r^*(\beta)} - E_1^{\tilde{c}_r x} \right|, \left| E_2^{\tilde{z}_r^*(\beta)} - E_2^{\tilde{c}_r x} \right| \right\}, \quad r = 1, \dots, k \quad (16.5)$$

por tanto, los elementos del conjunto discrepancia son las diferencias en valor absoluto entre los extremos correspondientes de los intervalos esperados.

Definición 17 (Arenas et al. (2004)). Llamaremos discrepancia entre los números difusos $\tilde{c}_r x$ y $\tilde{z}_r^*(\beta)$ y lo denotaremos D_r , al elemento máximo del conjunto discrepancia:

$$D_r = \max \left\{ \left| E_1^{\tilde{z}_r^*(\beta)} - E_1^{\tilde{c}_r x} \right|, \left| E_2^{\tilde{z}_r^*(\beta)} - E_2^{\tilde{c}_r x} \right| \right\}, \quad r = 1, \dots, k. \quad (16.6)$$

Consideraremos preferible la obtención de soluciones difusas menos imprecisas (con menor amplitud) que sus correspondientes soluciones en el punto β -ideal difuso:

$$E_1^{\tilde{c}_r x} - E_1^{\tilde{c}_r x} \leq E_2^{\tilde{z}_r^*(\beta)} - E_1^{\tilde{z}_r^*(\beta)} \quad (16.7)$$

lo cual permite definir la discrepancia entre cada objetivo difuso y su correspondiente componente en la solución β -ideal difusa, $\tilde{z}_r^*(\beta)$, de la siguiente manera:

$$D_r = \max \left\{ E_1^{\tilde{c}_r x} - E_1^{\tilde{z}_r^*(\beta)}, E_2^{\tilde{z}_r^*(\beta)} - E_2^{\tilde{c}_r x} \right\}, \quad r = 1, \dots, k. \quad (16.8)$$

Con la definición de discrepancia y la condición de menor imprecisión, es posible afirmar que:

$$EI(\tilde{c}_r x) \xrightarrow{\sim} EI(z_r^*(\beta)), \quad r = 1, \dots, k, \quad \text{si y sólo si} \quad D_r \rightarrow 0 \quad (16.9)$$

Es decir, consideramos que si la discrepancia es nula, el intervalo esperado de cada objetivo difuso y el intervalo esperado de la correspondiente componente de la solución ideal difusa, son iguales. Por lo tanto, podemos concluir que alcanzar la solución ideal difusa es equivalente a alcanzar discrepancia nula para todos los objetivos:

$$(D_1, D_2, \dots, D_k) = (0, 0, \dots, 0) \quad (16.10)$$

Abordaremos esta tarea mediante un problema de Programación Compromiso no difusa, cuyo objetivo es minimizar la discrepancia entre la solución β -ideal difusa y los objetivos difusos, siendo la solución ideal de este nuevo problema no difuso el vector nulo y definiéndose las soluciones β -compromiso de la siguiente manera:

Definición 6. Un vector de decisión x^* es una solución β -compromiso del problema FP-MOLP, si es solución compromiso del siguiente problema:

$$\begin{aligned} & \min && (D_1, D_2, \dots, D_k) \\ & \text{s. a :} && x \in \chi(\beta) \end{aligned} \quad (16.11)$$

El problema (16.11) no es difuso y el enfoque de Programación Compromiso que utilizaremos para resolverlo estará basado en la siguiente familia de problemas basados en distancias L_p :

Problem L_P :

$$\begin{aligned} & \min && L_p = \min \left(\sum_{r=1}^k w_r^p D_r^p \right)^{\frac{1}{p}} \\ & \text{s. a :} && x \in \chi(\beta) \end{aligned} \quad (16.12)$$

donde $w_r \geq 0$ es un coeficiente normalizador y que pondera la importancia relativa que tiene la discrepancia entre cada función objetivo difusa y su correspondiente componente en la solución ideal difusa. Estos coeficiente son fijados por el Decisor mediante un proceso interactivo con el analista matemático.

Si consideramos la métrica $p = 1$ la correspondiente solución compromiso se obtiene resolviendo el siguiente problema de programación lineal:

Problem L_1 :

$$\begin{aligned} \min \quad & L_1 = \min \sum_{r=1}^k w_r D_r \\ \text{s. a :} \quad & x \in \chi(\beta) \end{aligned} \quad (16.13)$$

La consideración de la métrica L_∞ , conlleva la minimización de la máxima discrepancia de entre las discrepancias individuales. Consecuentemente, la solución compromiso correspondiente a esta métrica se obtiene resolviendo el problema:

Problem L_∞ :

$$\begin{aligned} \min \quad & L_\infty = \min \max_r w_r D_r \\ \text{s. a :} \quad & x \in \chi(\beta) \end{aligned} \quad (16.14)$$

Este problema es un problema MIN-MAX que puede reformularse de la siguiente manera:

$$\begin{aligned} \min \quad & DT \\ \text{s. a :} \quad & x \in \chi(\beta) \\ & DT \geq w_r D_r, \quad r = 1, \dots, k \end{aligned} \quad (16.15)$$

siendo $DT = \max w_r D_r, \quad r = 1, \dots, n$.

Las soluciones correspondientes a los problemas (16.13) y (16.14) son las soluciones compromiso que se obtienen con más frecuencia debido principalmente, a que para métricas distintas a $p = 1$ y $p = \infty$ los problemas planteados no son lineales y requieren para su resolución de algoritmos de resolución de programas matemáticos no lineales. Además, Blasco *et al.* (1999) demuestran como bajo ciertas condiciones, estas soluciones acotan el conjunto completo de soluciones compromiso.

3 Presentación del problema.

Descripción del modelo: datos y variables.

Hemos trabajado, como ya hemos señalado, con seis servicios quirúrgicos pertenecientes a un hospital tipo II del INSALUD. Sin embargo en este trabajo presentaremos los resultados correspondientes a un solo servicio, Cirugía General, que consideramos suficientemente representativo del problema.

Denotaremos las variables del problema según los servicios a los que pertenecen los procesos y a aquellos por su inicial; en nuestro caso Cirugía General que será denotado mediante una C . Si el proceso se realiza en forma horaria normal, la variable que define el servicio carecerá de prefijo; si se realiza en forma extraordinaria, sea interna o externa, el prefijo será una X . Cada proceso vendrá definido por dos subíndices. Un primer índice i , hará referencia al proceso y un segundo

índice j , nos informará del mes en que debe realizarse el proceso de referencia. Si se trata de nombrar la lista de cualquiera de los procesos, se precederá la variable representativa del servicio de la letra L .

Tabla 1. Formato de las variables de actividad o variables de decisión.

Modalidad	Servicio	Proceso	Mes
--/X	C	i	j

Tabla 2. Formato de las variables de lista de espera o variables de estado.

Lista	Servicio	Proceso	Mes
L	C	i	j

Consideraremos nueve procesos quirúrgicos que representan aproximadamente un 45% de la actividad total del servicio considerado:

Tabla 3. Servicio: Cirugía General.

CODIGO	Nombre Proceso	Variable
241	Bocio Multinodular	C01
278	Obesidad Mórbida	C02
454	Varices	C03
455	Hemorroides	C04
550	Hernias Inguinales	C05
553	Otras Hernias	C06
565	Fisura/Fístula anal	C07
574	Colelitiasis	C08
685	Quiste Pilonidal	C09

Trabajaremos con un período de planificación actualizado que comprende los meses desde abril a diciembre de 1999.

Restricciones.

Las restricciones del modelo que nos ocupa serán de cuatro tipos:

- Ecuaciones de estado.
- Horas de quirófanos por servicio.
- Límites superiores a la permanencia en lista de espera.
- Cotas superiores de actividad extraordinaria.

a) Ecuaciones de estado, $j = 4, \dots, 12$.

$$\widetilde{LC}_{i(j+1)} = \widetilde{LC}_{ij} + \widetilde{AN}_{ij} - XC_{ij} - C_{ij} \quad (16.16)$$

donde \widetilde{LC}_{ij} refleja el estado de la lista de espera para el proceso i al principio del mes j . $\widetilde{AN}_{ij} = \tilde{A}_{ij} - \tilde{S}_{ij}$ recoge el número de admisiones netas estimadas (admisiones, \tilde{A}_{ij} , menos exclusiones, \tilde{S}_{ij} , sin haber sido sometidos a intervención quirúrgica) para el proceso i del servicio C por mes j . Estas ecuaciones definen el estado de la lista de espera para cada proceso y en cada mes. Determinada la actividad quirúrgica a llevar a cabo durante el período de planificación, proporcionan información a cerca de la evolución mensual de las listas de espera de los procesos considerados.

Los datos sobre admisiones (\tilde{A}_{ij}) y exclusiones (\tilde{S}_{ij}) para el año 1999 serán tratados como números difusos triangulares dado su carácter incierto y tendrán la siguiente estructura:

$$\begin{aligned} \text{Admisiones } \tilde{A}_{ij} &= (ME_{ij}^A, PR_{ij}^A, PE_{ij}^A) \\ \text{Exclusiones } \tilde{S}_{ij} &= (ME_{ij}^S, PR_{ij}^S, PE_{ij}^S) \end{aligned}$$

donde PR_{ij}^A y PR_{ij}^S recogen los peores datos reales históricos para las admisiones y exclusiones respectivamente; PE_{ij}^A y PE_{ij}^S son las estimaciones pesimistas del Decisor y ME_{ij}^A y ME_{ij}^S son sus estimaciones optimistas.

Las tablas 4 y 5 recogen las estimaciones sobre admisiones y exclusiones que se elaboran, considerando dichas admisiones y exclusiones como datos imprecisos representados por números difusos triangulares.

Tabla 4. Admisiones estimadas.

COD.	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
241	(7, 8 10)	(5, 6 8)	(4, 4 5)	(5, 5 7)	(5, 5 7)	(14, 15 20)	(7, 8 10)	(5 5 7)	(8, 9 12)
278	(3, 3 4)	(4, 4 5)	(2, 2 3)	(1, 1 1)	(2, 2 3)	(5, 5 7)	(4, 4 5)	(2, 2 3)	(5, 5 7)
454	(22, 24 31)	(23, 25 33)	(20, 22 29)	(15, 17 22)	(11, 12 16)	(19, 21 27)	(22, 24 31)	(26, 29 38)	(14, 15 20)
455	(15, 17 22)	(14, 15 20)	(10, 11 14)	(11, 12 16)	(7, 8 10)	(7, 8 10)	(12, 13 17)	(16, 18 23)	(13, 14 18)
550	(49, 54 70)	(54, 60 78)	(51, 57 74)	(42, 47 61)	(27, 30 39)	(46, 51 66)	(56, 62 81)	(65, 72 94)	(37, 41 53)
553	(23, 26 34)	(19, 21 27)	(23, 25 33)	(23, 25 33)	(14, 15 20)	(26, 29 38)	(31, 34 44)	(32, 35 46)	(18, 20 26)
565	(16, 18 23)	(14, 15 20)	(14, 16 21)	(11, 12 16)	(9, 10 13)	(8, 9 12)	(19, 21 27)	(19, 21 27)	(14, 16 21)
574	(21, 23 30)	(27, 30 39)	(22, 24 31)	(15, 17 22)	(13, 14 18)	(19, 21 27)	(23, 25 33)	(21, 23 30)	(18, 20 26)
685	(24, 27 35)	(29, 32 42)	(23, 25 33)	(17, 19 25)	(13, 14 18)	(17, 19 25)	(35, 39 51)	(43, 48 62)	(30, 33 43)

Tabla 5. Exclusiones estimadas.

COD.	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
241	(0, 0 0)	(0, 0 0)	(1, 1 1)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(0, 0 0)
278	(1, 1 1)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(0, 0 0)
454	(2, 2 3)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(1, 1 1)	(1, 1 1)	(0, 0 0)	(0, 0 0)
455	(0, 0 0)	(1, 1 1)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(1, 1 1)	(0, 0 0)	(2, 2 3)	(1, 1 1)
550	(4, 4 5)	(1, 1 1)	(0, 0 0)	(1, 1 1)	(0, 0 0)	(5, 5 7)	(4, 4 5)	(5, 5 7)	(0, 0 0)
553	(1, 1 1)	(0, 0 0)	(1, 1 1)	(0, 0 0)	(2, 2 3)	(5, 5 7)	(3, 3 4)	(1, 1 1)	(1, 1 1)
565	(0, 0 0)	(1, 1 1)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(0, 0 0)	(1, 1 1)
574	(1, 1 1)	(0, 0 0)	(5, 6 8)	(1, 1 1)	(0, 0 0)	(4, 4 5)	(2, 2 3)	(2, 2 3)	(0, 0 0)
685	(5, 5 7)	(1, 1 1)	(1, 1 1)	(0, 0 0)	(1, 1 1)	(2, 2 3)	(2, 2 3)	(2, 2 3)	(1, 1 1)

b) Horas de quirófano por servicio y mes.

Estas restricciones sólo afectan a la planificación quirúrgica que se lleva a cabo en horario ordinario:

$$\sum_{i=1}^9 \tilde{t}_i C_{ij} \leq CQ_j \quad (16.17)$$

donde CQ_j indica en minutos el tiempo de quirófano del que dispone este servicio para el mes j y donde \tilde{t}_i es la estimación hecha sobre la duración de cada intervención quirúrgica.

Consideramos como dato difuso, la duración de cada intervención quirúrgica que representaremos mediante el siguiente número difuso triangular:

$$\tilde{t}_i = (ME_i^t, DR_i^t, PE_i^t) \quad (16.18)$$

donde DR_i^t es el tiempo medio de duración de una intervención, que ha sido determinado utilizando los datos procedentes de los partes de quirófano, más veinte minutos, que es el tiempo requerido para preparar los quirófanos para la siguiente intervención; ME_i^t , y PE_i^t son la mejor y peor estimación acerca de la duración de cada intervención, proporcionadas por el Centro Decisor.

Tabla 6. Estimación de la duración de las intervenciones quirúrgicas.

COD.	Nombre Proceso	Variable	Estimación Tiempo (minutos)		
			ME	DR	PE
241	Bocio Multinodular	C01	151	151	182
278	Obesidad Mórbida	C02	155	155	186
454	Varices	C03	135	135	162
455	Hemorroides	C04	74	74	89
550	Hernias Inguinales	C05	114	114	137
553	Otras Hernias Abdominales	C06	134	134	161
565	Fisura/ Fístula Anal	C07	62	62	75
574	Colelitiasis	C08	135	135	162
685	Quiste Pilonidal	C09	63	63	76

Los datos referentes a los tiempos de quirófano asignados, en este caso, al servicio de Cirugía General son los que se muestran en la tabla siguiente:

Tabla 7. Horas de quirófano asignadas al servicio de Cirugía.

	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Minutos	3955	4704	4976	3316	2997	2997	5183	4815	4385

c) Límites superiores a la permanencia en lista de espera: no más de seis meses.

Con las siguientes ecuaciones reflejamos el hecho de que a lo largo del año 1999 el tiempo máximo que un paciente puede permanecer en lista de espera debe ser de seis meses, para ello exigimos que la suma de las actividades ordinaria y extraordinaria realizadas entre enero del año en curso y el mes k -ésimo, supere al número de paciente que llevarían seis o más meses en lista de espera para cada proceso i en el momento k :

$$\sum_{j=04}^k [C_{ij} + XC_{ij}] \geq s_{ik} \quad (16.19)$$

El cálculo de las cotas mínimas, que serán la parte derecha de las ecuaciones, es sencillo, se realiza a partir del estado de la lista de espera al comienzo del periodo de planificación y a partir de ahí se van acumulando estos valores mes a mes. El estado de la lista de espera al comienzo del período de planificación (1 de abril de 1999) es el siguiente:

Tabla 8. Estado inicial de las listas de espera.

COD.	241	278	454	455	550	553	565	574	685
Lista	17	13	51	21	118	58	28	59	49

Los parámetros de los términos independientes, por proceso y mes, calculados de la forma descrita y para las previsiones llevadas a cabo, se recogen en la siguiente tabla:

Tabla 9. Parámetros s_{ik} correspondientes a Cirugía General.

COD.	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
241	5	8	11	21	23	28	32	40	40
278	2	3	6	7	9	19	22	26	28
454	13	13	16	27	39	67	91	102	139
455	9	10	10	11	19	31	46	52	67
550	21	39	48	62	103	166	199	258	346
553	15	25	29	46	64	85	125	143	183
565	5	8	9	11	19	37	53	69	93
574	4	13	26	42	57	85	95	123	141
685	13	14	16	16	27	65	90	102	122

d) Cotas al número de procesos realizables fuera de horario normal

Estas restricciones consistirán en desigualdades del tipo:

$$XC_{ij} \geq r_{ij} \quad (16.20)$$

El significado de estas ecuaciones no es otro que el de acotar la actividad global mínima que a priori se acuerda derivar en base a los datos históricos. Esta actividad mínima ha de indicarse por proceso y mes. Reflejan información de tipo cualitativo: el Decisor conoce de antemano que algunos procesos, imposibles de asumir, serán derivados en al menos ciertas cantidades, que a lo largo del año se irán concretando. Igualmente conoce que algunos otros procesos no se derivarán, salvo graves problemas. Este es el caso de las intervenciones de Hernias Inguinales que es el proceso quirúrgico que mayores listas de espera presenta en el servicio de Cirugía General; el Centro Hospitalario establece que para este proceso se derive a otros centros al menos una intervención al mes:

$$XC_{5j} \geq 1, \quad j = 4, \dots, 12. \quad (16.21)$$

También existen procesos quirúrgicos que debido a su complejidad clínica el Centro Hospitalario no desea derivar a otros centros. Este es el caso de tres procesos: Bocio Multinodular, Obesidad Mórbida y Colelitiasis:

$$\sum_{j=4}^{12} XC_{1j} = 0; \quad \sum_{j=4}^{12} XC_{2j} = 0; \quad \sum_{j=4}^{12} XC_{8j} = 0. \quad (16.22)$$

e) No impondremos que las variables de este problema sean enteras por dos razones: Una de ellas es que los tiempos estimados por proceso son tiempos

medios, lo que imposibilita la exactitud de resultados, y la otra deriva de la complejidad computacional asociada a un problema entero con tan gran número de variables.

Objetivos.

Este problema, como ya hemos dicho, pretende determinar la actividad quirúrgica ordinaria y extraordinaria óptimas que es necesario planificar por parte de la gerencia del Hospital para que, teniendo en cuenta el flujo de pacientes de las diversas patologías que acuden al centro hospitalario, aquellos puedan ser atendidos dentro del marco de excelencia que es exigible a un servicio público de este tipo.

La disponibilidad de tiempo de quirófano es la más fuerte de las restricciones a la actividad de cada servicio dado que, hoy en día, como ya se ha señalado, muchos procesos son susceptibles de ser realizados mediante Cirugía Mayor Ambulatoria y por ello el número de camas instaladas en el centro limita poco o nada la actividad.

La determinación del número de procesos, por mes y servicio, que es necesario realizar para mantener las listas de espera en los niveles deseados, nos lleva a considerar dos objetivos.

El primero proporcionará la *máxima capacidad operativa* del centro en su horario y configuración ordinarios y el segundo determinará la *actividad mínima* indispensable que ha de realizarse de modo *extraordinario*, todo ello para verificar el requisito marco de máxima permanencia en lista de espera. Estos objetivos se expresan mediante las funciones objetivo que a continuación detallamos:

La primera de las funciones objetivo se formulará simplemente como la maximización del total de la actividad a realizar de manera ordinaria por mes, proceso y servicio, a lo largo del año:

$$\max \sum_{i=1}^9 \sum_{j=04}^{12} [C_{ij}] \quad (16.23)$$

La segunda función objetivo minimiza la suma de la actividad externa que será necesario llevar a cabo:

$$\min \sum_{i=1}^9 \sum_{j=04}^{12} [XC_{ij}] \quad (16.24)$$

El Decisor hospitalario considera que los dos objetivos poseen la misma importancia, por lo que fija los siguientes pesos $W_1 = W_2 = 1$.

El problema que resolveremos será por tanto:

$$\begin{aligned}
 \max(\tilde{F}_1, -\tilde{F}_2) &= \left(\sum_{i=1}^9 \sum_{j=4}^{12} \tilde{t}_i C_{ij}, \sum_{i=1}^9 \sum_{j=4}^{12} \tilde{t}_i X C_{ij} \right) \\
 \text{sujeto a:} \\
 \widetilde{LC}_{i(j+1)} &= \widetilde{LC}_{ij} + \widetilde{AN}_{ij} - XC_{ij} - C_{ij}, \quad i = 1, \dots, 9, j = 4, \dots, 12 \\
 \sum_{i=1}^9 \tilde{t}_i C_{ij} &\leq CQ_j \\
 \sum_{i=1}^9 \sum_{j=4}^k [C_{ij} + XC_{ij}] &\geq s_{ik} \\
 XC_{5j} &\geq 1, \quad j = 4, \dots, 12 \\
 \sum_{j=4}^{12} XC_{1j} &= 0; \quad \sum_{j=4}^{12} XC_{2j} = 0; \quad \sum_{j=4}^{12} XC_{8j} = 0
 \end{aligned} \tag{P}$$

El problema (P) es un problema de Programación Multiobjetivo Lineal Posibilística.

Problemas resueltos

En ausencia de pesos subjetivos preferenciales asignaremos el mismo peso a ambos objetivos ($W_1 = W_2 = 1$). El problema Multiobjetivo Lineal β -Paramétrico mediante el que determinaremos las soluciones β -eficientes del problema (P) es el siguiente:

$$\begin{aligned}
 \max(-\tilde{F}_1, \tilde{F}_2) &= \left(-\sum_{i=1}^9 \sum_{j=4}^{12} \tilde{t}_i C_{ij}, \sum_{i=1}^9 \sum_{j=4}^{12} \tilde{t}_i X C_{ij} \right) \\
 \text{sujeto a:} \\
 \begin{aligned}
 LC_{i(j+1)} &= LC_{ij} + (1-\beta)E_2^{\widetilde{AN}_{ij}} + \beta E_1^{\widetilde{AN}_{ij}} - XC_{ij} - C_{ij} \\
 \sum_{i=1}^9 [(1-\beta)E_1^{\tilde{t}_i} + \beta E_2^{\tilde{t}_i}] C_{ik} &\leq CQ_k \\
 \sum_{i=1}^9 \sum_{j=4}^k [C_{ij} + XC_{ij}] &\geq s_{ik} \\
 XC_{5j} &\geq 1, \quad j = 4, \dots, 12 \\
 \sum_{j=4}^{12} XC_{1j} &= 0; \quad \sum_{j=4}^{12} XC_{2j} = 0; \quad \sum_{j=4}^{12} XC_{8j} = 0 \\
 i = 1, \dots, 9, j = 4, \dots, 12
 \end{aligned} \tag{16.25}
 \end{aligned}$$

Hemos resuelto el problema (16.25) de manera interactiva con el Decisor que fija sucesivamente los distintos niveles de factibilidad que está dispuesto a asumir. Fijado el nivel en que se han de satisfacer las restricciones, hemos obtenido la solución β -ideal difusa mediante sus α -cortes. En la Tabla 10 se incluyen las distribuciones de posibilidad del punto ideal $(\tilde{F}_1^*, \tilde{F}_2^*)$ formado por los valores óptimos de cada objetivo para cada nivel de factibilidad fijado por el Decisor.

Como se puede observar en la Tabla 10, la solución “empeora” a medida que el Decisor establece factibilidades, β , mayores, es decir, a mayor nivel de factibilidad menores niveles de actividad interna y mayores de actividad externa.

Tabla 10. Distribuciones de posibilidad de la solución β -ideal difusa.

$\beta = 0.6$		
α	\tilde{F}_1^*	\tilde{F}_2^*
0	[30854, 43544]	[80325, 113380]
0, 2	[31930, 42148]	[82895, 109478]
0, 4	[33001, 40631]	[85828, 105792]
0, 6	[34131, 39160]	[88572, 101899]
0, 8	[35160, 37657]	[91601, 98198]
1	[36244, 36244]	[94312, 94312]
$\beta = 0.8$		
α	\tilde{F}_1^*	\tilde{F}_2^*
0	[29838, 42107]	[81341, 114816]
0, 2	[30877, 40737]	[83945, 110901]
0, 4	[31897, 39288]	[86932, 107135]
0, 6	[32960, 37845]	[89732, 103211]
0, 8	[33982, 36403]	[92780, 99452]
1	[35034, 35034]	[95522, 95522]
$\beta = 1$		
α	\tilde{F}_1^*	\tilde{F}_2^*
0	[28775, 40617]	[82403, 116316]
0, 2	[29781, 39330]	[85041, 112294]
0, 4	[30776, 37901]	[88053, 108532]
0, 6	[31817, 36535]	[90874, 104521]
0, 8	[32785, 35109]	[93978, 100746]
1	[33811, 33811]	[96746, 96746]

La maximización de la actividad interna como único criterio, por ejemplo, implicaría niveles de actividad externa muy elevados para poder mantener las listas de espera en los niveles deseados (Tabla 11), por ello no parece aconsejable la aplicación de políticas que tengan en cuenta un solo criterio de decisión.

Tabla 11. Actividad externa correspondiente a la solución \tilde{F}_1^* con nivel de posibilidad $\alpha = 1$ y $\beta = 0.8$.

COD.	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dic
241	0	0	0	0	0	0	0	0	0
278	0	0	0	0	0	0	0	0	0
454	74	26	23	3	0	0	0	0	0
455	39	15	12	1	0	0	0	0	0
550	338	1	1	1	1	1	1	1	1
553	183	0	0	0	0	0	0	0	0
565	93	0	0	0	0	0	0	0	0
574	0	0	0	0	0	0	0	0	0
685	122	0	0	0	0	0	0	0	0

Una vez hallada la solución ideal, determinamos las soluciones β -compromiso difusas, L_1 y L_∞ . Utilizaremos para ello la siguiente notación:

$$\begin{aligned}
 D_r &= \max \left\{ E_1^{\tilde{c}_r x} - E_1^{\tilde{z}_r^*(\beta)}, E_2^{\tilde{z}_r^*(\beta)} - E_2^{\tilde{c}_r x} \right\}, \quad r = 1, 2, \quad DT = \max D_r \\
 EI(\tilde{c}_1 x) &= EI(\tilde{F}_1) = \left[E_1^{\tilde{F}_1}, E_2^{\tilde{F}_1} \right] = \left[E_1^{\tilde{t}_i}, E_2^{\tilde{t}_i} \right] \sum_{i=1}^9 \sum_{j=4}^{12} C_{ij} \\
 EI(\tilde{c}_2 x) &= EI(\tilde{F}_2) = \left[E_1^{\tilde{F}_2}, E_2^{\tilde{F}_2} \right] = \left[E_1^{\tilde{t}_i}, E_2^{\tilde{t}_i} \right] \sum_{i=1}^9 \sum_{j=4}^{12} XC_{ij} \\
 EI(\tilde{z}_r^*) &= EI(\tilde{F}_r^*) = \left[E_1^{\tilde{F}_r^*}, E_2^{\tilde{F}_r^*} \right]
 \end{aligned}$$

En la siguiente tabla aparecen recogidos los intervalos esperados de las soluciones compromiso para los distintos niveles de factibilidad fijados por el Decisor.

Tabla 12. Soluciones compromiso β -difusas.

$\beta = 0.6$				
L_1	\tilde{F}_1^* (30837, 36214, 43542)	\tilde{F}_2^* (80341, 94342, 113381)	$EI(\tilde{F}_1^*)$ (33526, 39878)	$EI(\tilde{F}_1^*)$ (87342, 103862)
	L_∞ (30840, 36230, 43540)	$(80339, 94326, 113384)$	$(33535, 39885)$	$(87333, 103855)$
$\beta = 0.8$				
L_1	(29784, 34976, 42053)	(81395, 95580, 114870)	(32380, 38515)	(88488, 105225)
	L_∞ (29820, 35018, 42107)	$(81358, 95539, 114816)$	$(32419, 38563)$	$(88449, 105178)$
$\beta = 1$				
L_1	(28771, 33806, 40602)	(82407, 96750, 116321)	(31289, 37204)	(89579, 106536)
	L_∞ (28775, 33810, 40604)	$(82403, 96746, 116319)$	$(31293, 37207)$	$(89575, 106533)$

El problema A1 (ver apéndice) proporciona una solución de máxima eficiencia al minimizar la suma ponderada de las discrepancias, es decir, al maximizar el nivel de logro agregado y ponderado de los objetivos. Esta solución, sin embargo, puede presentar un sesgo hacia el nivel de logro de algún objetivo en particular. Por el contrario, el problema A2 (ver apéndice) proporciona una solución de máximo equilibrio entre los niveles de logro de los objetivos, al tener en cuenta sólo la minimización de la mayor de las discrepancias individuales. Resulta, por lo tanto, interesante presentar al Decisor ambas soluciones y ofrecerle la posibilidad de elegir o bien una planificación quirúrgica que maximice la eficiencia o bien una planificación quirúrgica que implique máximo equilibrio entre todos los criterios considerados.

Si observamos las tablas 13, 14, 15 y 16 vemos como para ambas soluciones compromiso la actividad interna total es prácticamente la misma debido a las restricciones en la disponibilidad de quirófanos en el hospital; pese a ello, existen diferencias en la distribución de dicha actividad entre los distintos procesos a lo largo del período de planificación.

Tabla 13. Solución compromiso L_1 : Actividad Interna. $\beta = 0.8$.

COD.	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
241	5	3	13	0	7	0	12	0	0
278	2	1	5	0	11	0	6	1	2
454	0	0	0	0	0	0	0	0	3
455	0	0	0	0	0	0	0	0	0
550	0	0	0	0	0	0	0	0	0
553	0	0	0	0	0	0	0	0	0
565	8	0	3	0	0	0	18	22	16
574	16	28	13	23	0	17	8	22	18
685	0	0	0	0	0	13	0	0	0

Tabla 14. Solución compromiso L_∞ : Actividad Interna. $\beta = 0.8$.

COD.	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
241	5	16	0	0	2	5	12	0	0
278	2	12	0	0	11	9	0	4	2
454	0	0	0	0	0	0	0	0	0
455	0	0	0	0	0	0	0	0	0
550	0	0	0	0	0	0	0	0	0
553	0	0	0	0	0	0	0	0	0
565	2	3	17	0	16	0	12	18	22
574	19	0	27	23	11	5	17	21	18
685	0	0	0	0	0	0	0	0	0

Tabla 15. Solución compromiso L_1 : Actividad Externa. $\beta = 0.8$.

COD.	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
241	0	0	0	0	0	0	0	0	0
278	0	0	0	0	0	0	0	0	0
454	74	26	23	13	0	0	0	0	0
455	39	3	0	25	0	0	0	0	0
550	338	1	1	1	1	1	1	1	1
553	183	0	0	0	0	0	0	0	0
565	0	0	0	9	0	0	0	0	0
574	0	0	0	0	0	0	0	0	0
685	122	0	0	0	0	0	0	0	0

Tabla 16. Solución compromiso L_∞ : Actividad Externa. $\beta = 0.8$.

COD.	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
241	0	0	0	0	0	0	0	0	0
278	0	0	0	0	0	0	0	0	0
454	74	26	0	39	0	0	0	0	0
455	39	15	12	1	0	0	0	0	0
550	338	1	1	1	1	1	1	1	1
553	183	0	0	0	0	0	0	0	0
565	3	0	0	0	0	0	0	0	0
574	0	0	0	0	0	0	0	0	0
685	122	0	0	0	0	0	0	0	0

Finalmente, podemos completar la información ofrecida al Decisor con la evolución mensual esperada de las listas de espera quirúrgicas, determinada por la planificación quirúrgica elegida. (Figura 1 y Figura 2).

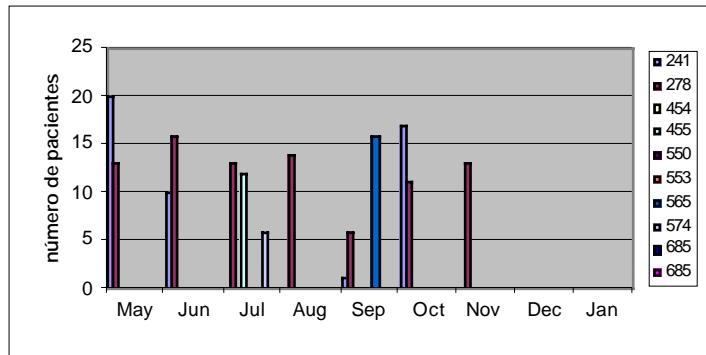
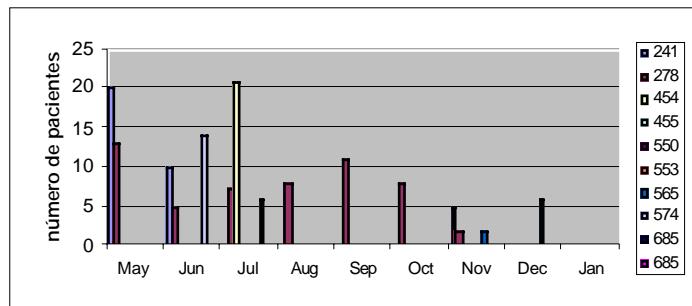
Figura 1. Evolución esperada de las listas de espera quirúrgicas. Solución Compromiso L_1 para $\beta = 0.8$.

Figura 2. Evolución esperada de las listas de espera quirúrgicas. Solución Compromiso L_∞ para $\beta = 0.8$.



La información a cerca de la evolución esperada de las listas de espera puede ser muy útil para el Decisor a la hora de determinar si es o no posible lograr los objetivos relativos a la máxima permanencia en lista de espera establecidos por el Ministerio de Sanidad al principio de cada año, por ello estos datos pueden ser utilizados por el Decisor hospitalario a la hora de negociar con el Ministerio de Sanidad las condiciones a firmar en los Contratos de Gestión.

4 Conclusiones

En este trabajo hemos presentado un modelo que permite planificar de manera óptima la actividad quirúrgica de un hospital público, con el propósito de proporcionar al Decisor hospitalario datos cuantitativos que le sirvieran de soporte para analizar la coherencia de los objetivos relativos a las listas de espera quirúrgicas, planteados por las autoridades sanitarias.

El desconocimiento y/o imprecisión de algunos datos del problema implicó la necesidad de considerarlos como números difusos. Las estimaciones sobre admissions y exclusiones de pacientes, así como las estimaciones sobre la duración de las intervenciones quirúrgicas han sido establecidas por el Decisor en términos de la lógica difusa debido a la imposibilidad de obtener series temporales para esos datos. Esta situación nos ha llevado a resolver el problema mediante un enfoque de Programación Multiobjetivo Lineal Posibilística. La aplicación de la Programación Compromiso a la resolución del problema nos proporcionó un conjunto de soluciones compromiso dependientes del nivel de cumplimiento de las restricciones establecido por el Decisor. De entre estas soluciones son de especial interés la solución L_1 y la solución L_∞ , que implican planificaciones quirúrgicas de máxima eficiencia y máximo equilibrio entre los objetivos, respectivamente.

Creemos que el modelo propuesto es un instrumento útil para la toma de decisiones en problemas de gestión de listas de espera en hospitales públicos, debido a la presencia de múltiples criterios de decisión y a la imprecisión y/o incertidumbre de los datos que se manejan en este tipo de problemas.

5 Bibliografía

- [1] Arenas Parra, M., Bilbao Terol, A., Jiménez López, M., Rodríguez Uría, M. V. (1998): “A Theory of Possibilistic Approach to the Solution of a Fuzzy Linear Programming”, Applied Decision Analysis, Girón, J. (Ed.), Kluwer Academic Publishers.
- [2] Arenas Parra, M.; Bilbao Terol, A.; Rodríguez Uría, M. V. (1999a): “Solving the multiobjective possibilistic linear programming problem”, European Journal of Operational Research, 117, 175-182.
- [3] Arenas Parra, M.; Bilbao Terol, A.; Rodríguez Uría, M. V. (1999b): “Solution of a possibilistic multiobjective linear programming problem”, European Journal of Operational Research, 119, 338-344.
- [4] Arenas Parra, M.; Bilbao Terol, A.; Cerdá Tena, E.; Rodríguez Uría, M. V. (2000): “Management of surgical waiting lists in public hospitals”. Research and Practice in Multiple Criteria Decision Making. Haimes Y.Y. Steuer. (Eds.) 473-482. Ed. Springer-Verlag
- [5] Arenas, M.; Bilbao, A.; Pérez Gladish, B.; Rodríguez Uría, M. V. (2004): “Solving a Multiobjective Possibilistic Problem through Compromise Programming”. European Journal of Operational Research (available online).
- [6] Blasco, E.; Cuchillo-Ibáñez, E.; Morón, M. A., Romero, C. (1999): “On the Monotonicity of the Compromise Set in Multicriteria Problems”, Journal of Optimization Theory and its Applications, Vol. 102, nº 1, 69-82. Plenum Publishing Corporation.
- [7] Heilpern, S. (1992): “The expected value of a fuzzy number”. Fuzzy Sets and Systems 47, 81-86. Jiménez, M. (1996): “Ranking fuzzy numbers through the comparison of its expected intervals”. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol.4, n. 4, 379-388.
- [8] Kaufmann, A., Gil Aluja, J. (1986): Introducción de la Teoría de Subconjuntos Borrosos a la Gestión de Empresas. Ed. Milladoiro.
- [9] Tanaka, H. (1987): “Fuzzy Data Analysis by Possibilistic Linear Models”, Fuzzy Sets and Systems 24, 363-375.

[10] Yu, P. L. (1973): “A class of solutions for group decision problems”. Management Science, 19, 936-946.

[11] Zadeh, L. A. (1965): “Fuzzy Sets”, Information and Control 8, 338-353.

[12] Zadeh, L. A. (1975): “The concept of a linguistic variable and its applications to approximate reasoning”, part 1: Information Sciences 8, 199-249; part 2: . Information Sciences 8, 301-357; part 3: Information Sciences 9, 43-80.

[13] Zeleny, M. (1973): “Compromise Programming” in Multiple Criteria Decision Making (Cochrane, J. L., Zeleny, M. Eds.). University of South Carolina Press, Columbia, 262-301.

Apéndice

Problema A1: determinación de la solución β -compromiso L_1 .

$$\begin{aligned}
 & \text{Min} \quad D_1 + D_2 + \delta_1 + \delta_2 \\
 & \text{sujeto a :} \\
 & \sum_{i=1}^9 \sum_{j=1}^{12} E_1^{\tilde{t}_i} C_{ij} - E_1^{\tilde{F}_1^*} - D_1 \leq M\delta_1 \\
 & E_1^{\tilde{F}_1^*} - \sum_{i=1}^9 \sum_{j=1}^{12} E_1^{\tilde{t}_i} C_{ij} - D_1 \leq M(1 - \delta_1) \\
 & \sum_{i=1}^9 \sum_{j=1}^{12} E_1^{\tilde{t}_i} X C_{ij} - E_1^{\tilde{F}_2^*} - D_2 \leq M\delta_2 \\
 & E_1^{\tilde{F}_2^*} - \sum_{i=1}^9 \sum_{j=1}^{12} E_1^{\tilde{t}_i} X C_{ij} - D_2 \leq M(1 - \delta_2) \\
 & E_1^{\tilde{F}_1^*} - \sum_{i=1}^9 \sum_{j=1}^{12} E_2^{\tilde{t}_i} X C_{ij} - D_1 \leq M\delta_1 \\
 & \sum_{i=1}^9 \sum_{j=1}^{12} E_2^{\tilde{t}_i} C_{ij} - E_2^{\tilde{F}_1^*} - D_1 \leq M(1 - \delta_1) \\
 & E_2^{\tilde{F}_1^*} - \sum_{i=1}^9 \sum_{j=1}^{12} E_2^{\tilde{t}_i} X C_{ij} - D_2 \leq M\delta_2 \\
 & \sum_{i=1}^9 \sum_{j=1}^{12} E_2^{\tilde{t}_i} C_{ij} - E_2^{\tilde{F}_1^*} - D_2 \leq M(1 - \delta_1) \\
 & [E_2^{\tilde{t}_i} - E_1^{\tilde{t}_i}] \sum_{i=1}^9 \sum_{j=1}^{12} C_{ij} \leq [E_2^{\tilde{F}_1^*} - E_1^{\tilde{F}_2^*}] + M\delta_1 \\
 & [E_2^{\tilde{t}_i} - E_1^{\tilde{t}_i}] \sum_{i=1}^9 \sum_{j=1}^{12} C_{ij} + M(1 - \delta_1) \geq [E_2^{\tilde{F}_1^*} - E_1^{\tilde{F}_2^*}] \\
 & [E_2^{\tilde{t}_i} - E_1^{\tilde{t}_i}] \sum_{i=1}^9 \sum_{j=1}^{12} X C_{ij} \leq [E_2^{\tilde{F}_1^*} - E_1^{\tilde{F}_2^*}] + M\delta_2 \\
 & [E_2^{\tilde{t}_i} - E_1^{\tilde{t}_i}] \sum_{i=1}^9 \sum_{j=1}^{12} X C_{ij} + M(1 - \delta_2) \geq [E_2^{\tilde{F}_1^*} - E_1^{\tilde{F}_2^*}] \\
 & \delta_p \in \{0, 1\}, \quad p = 1, 2. \\
 & \chi(\beta)
 \end{aligned}$$

Problema A1: determinación de la solución β -compromiso L_∞ .

$$\begin{aligned}
 & \text{Min} \quad DT + \delta_1 + \delta_2 \\
 & \text{sujeto a :} \\
 & \sum_{i=1}^9 \sum_{j=1}^{12} E_1^{\tilde{t}_i} C_{ij} - E_1^{\tilde{F}_1^*} - DT \leq M\delta_1 \\
 & E_1^{\tilde{F}_1^*} - \sum_{i=1}^9 \sum_{j=1}^{12} E_1^{\tilde{t}_i} C_{ij} - DT \leq M(1 - \delta_1) \\
 & \sum_{i=1}^9 \sum_{j=1}^{12} E_1^{\tilde{t}_i} XC_{ij} - E_1^{\tilde{F}_2^*} - DT \leq M\delta_2 \\
 & E_1^{\tilde{F}_2^*} - \sum_{i=1}^9 \sum_{j=1}^{12} E_1^{\tilde{t}_i} XC_{ij} - DT \leq M(1 - \delta_2) \\
 & E_1^{\tilde{F}_1^*} - \sum_{i=1}^9 \sum_{j=1}^{12} E_2^{\tilde{t}_i} XC_{ij} - DT \leq M\delta_1 \\
 & \sum_{i=1}^9 \sum_{j=1}^{12} E_2^{\tilde{t}_i} C_{ij} - E_2^{\tilde{F}_1^*} - DT \leq M(1 - \delta_1) \\
 & E_2^{\tilde{F}_2^*} - \sum_{i=1}^9 \sum_{j=1}^{12} E_2^{\tilde{t}_i} XC_{ij} - DT \leq M\delta_2 \\
 & \sum_{i=1}^9 \sum_{j=1}^{12} E_2^{\tilde{t}_i} C_{ij} - E_2^{\tilde{F}_1^*} - DT \leq M(1 - \delta_1) \\
 & [E_2^{\tilde{t}_i} - E_1^{\tilde{t}_i}] \sum_{i=1}^9 \sum_{j=1}^{12} C_{ij} \leq [E_2^{\tilde{F}_1^*} - E_1^{\tilde{F}_2^*}] + M\delta_1 \\
 & [E_2^{\tilde{t}_i} - E_1^{\tilde{t}_i}] \sum_{i=1}^9 \sum_{j=1}^{12} C_{ij} + M(1 - \delta_1) \geq [E_2^{\tilde{F}_1^*} - E_1^{\tilde{F}_2^*}] \\
 & [E_2^{\tilde{t}_i} - E_1^{\tilde{t}_i}] \sum_{i=1}^9 \sum_{j=1}^{12} XC_{ij} \leq [E_2^{\tilde{F}_1^*} - E_1^{\tilde{F}_2^*}] + M\delta_2 \\
 & [E_2^{\tilde{t}_i} - E_1^{\tilde{t}_i}] \sum_{i=1}^9 \sum_{j=1}^{12} XC_{ij} + M(1 - \delta_2) \geq [E_2^{\tilde{F}_1^*} - E_1^{\tilde{F}_2^*}] \\
 & \delta_p \in \{0, 1\}, \quad p = 1, 2. \\
 & \chi(\beta)
 \end{aligned}$$

Análisis envolvente de datos bajo condiciones de incertumbre: Aplicación a la liga de fútbol profesional española e italiana

José E. Boscá^a, Vicente Liern^{b1}, Aurelio Martínez^c,
Ramón Sala^b

^a Dep. Análisis Económico, Universitat de València,

^b Dep. Matemática Económico-Empresarial, Universitat de València,

^c Dep. Economía Aplicada II (Estructura Económica), Universitat de València,

e-mails: {Jose.E.Bosca, Vicente.Liern, Ramon.Sala, Aurelio.Martinez}@uv.es

1 Introducción

El Análisis Envolvente de Datos, DEA², es una metodología que permite analizar las fronteras de producción y proporciona una medida de la eficiencia productiva [11]. Aunque pueden encontrarse algunos antecedentes en la literatura económica [28], sus orígenes pueden establecerse en el trabajo de Charnes, Cooper y Rhodes [8] y los de Banker, Charnes y Cooper [1] o Charnes *et al.* [9].

En esencia, los métodos DEA definen la frontera de producción a partir de dos ideas básicas:

¹Este trabajo ha sido parcialmente subvencionado por TIC 2002-04242-C03

²Son las siglas de Data Envelopment Analysis.

- a) La noción de Pareto-Koopmans [12, 19] de eficiencia, dado que una unidad productiva, DMU³, es eficiente si y sólo si, ninguno de sus inputs o outputs puede mejorarse sin empeorar alguno de los restantes inputs o outputs.
- b) El concepto de eficiencia relativa, puesto que una unidad se evalúa como eficiente si, y sólo si, el comportamiento del resto de DMUs no muestra que alguno de sus inputs o outputs puede ser mejorado sin empeorar alguno de los restantes inputs o outputs.

Se trata por tanto de identificar las unidades productivas que se muestran mejores y éstas son las que definen la frontera de posibilidades de producción. A continuación la eficiencia del resto de DMUs se mide evaluando su desviación respecto de esta frontera. Precisamente esta medición relativa es la que marca la importancia de la homogeneidad y tecnología común que debe exigirse a las diferentes DMUs, puesto que de otro modo la comparación podría carecer de sentido [19].

Los modelos DEA suponen una alternativa a los enfoques paramétricos clásicos de análisis de fronteras de producción que precisan de una relación funcional explícita que relacione la variable dependiente con las variables independientes. Además, en estos métodos suelen ser necesarias hipótesis relativas a la distribución del término asociado con la ineficiencia y del término de error, por ejemplo que sean una muestra aleatoria de una distribución normal.

Si bien es innegable que las técnicas DEA, se han convertido en algo más de dos décadas en una herramienta imprescindible en muchos estudios de eficiencia, lo cierto es que en los últimos años ha habido autores [2, 4, 16, 17] que han advertido de las dificultades de utilizar directamente estas técnicas a problemas reales. En estos modelos se asume que los inputs y los outputs utilizados son datos precisos, sin embargo se sabe que a menudo los datos que se manejan están afectados de incertidumbre. Por lo tanto, es necesario extender estas técnicas de manera que pueda incorporarse la incertidumbre al propio modelo. En nuestro trabajo mostraremos algunas de las formas de conseguir este objetivo utilizando subconjuntos *borrosos* [33]. No obstante, no ha sido nuestra intención recoger de forma exhaustiva todas las propuestas que existen en la literatura actual, sino aquellas que, a nuestro juicio, tenían una aplicación más directa al análisis de la eficiencia en las ligas de fútbol profesional.

2 Análisis envolvente de datos clásico

El modelo matemático básico que formularon Charnes, Cooper y Rhodes (CCR) [7, 8], para n DMUs con m inputs y s outputs, cada una, fue el siguiente modelo de programación matemática fraccional:

³Las siglas corresponden a Decision Making Units.

$$\begin{aligned}
 \text{Max} \quad w_o &= \frac{\sum_{r=1}^s u_{ro} y_{ro}}{\sum_{i=1}^m v_{io} x_{io}} \\
 \text{s.a.} \quad &\frac{\sum_{r=1}^s u_{ro} y_{rj}}{\sum_{i=1}^m v_{io} x_{ij}} \leq 1, \quad 1 \leq j \leq n \\
 &u_{ro} \geq 0, \quad 1 \leq r \leq s \\
 &v_{io} \geq 0, \quad 1 \leq i \leq m
 \end{aligned} \tag{17.1}$$

donde el subíndice 0 representa la DMU que se va a evaluar, x_{ij} e y_{rj} representan las cantidades del input i y del output r de la DMU j -ésima, y u_{ro} y v_{io} son las ponderaciones asignadas al input i y al output r correspondientes a la DMU cuya eficiencia se está evaluando.

Posteriormente, Charnes, Cooper y Rhodes [9] exigieron que las ponderaciones fuesen estrictamente positivas, es decir, sustituyeron las restricciones de no negatividad por $u_{ro} \geq \epsilon$ y $v_{io} \geq \epsilon$ con $\epsilon > 0$. Por otro lado, como resolver el modelo fraccional supone dificultades computacionales, los mismos autores propusieron dos modelos lineales equivalentes conocidos como CCR lineal con *orientación input* (si minimizamos el denominador manteniendo fijo el numerador) y CCR lineal con *orientación output* (si maximizamos el numerador manteniendo fijo el denominador). Estos modelos son:

ORIENTACIÓN OUTPUT	ORIENTACIÓN INPUT
$\text{Max} \quad \phi_o = \sum_{i=1}^n u_{ro} y_{ro}$ sujeto a $\sum_{r=1}^s u_{ro} y_{rj} - \sum_{i=1}^m v_{io} x_{ij} \leq 0, \quad 1 \leq j \leq n$ $\sum_{i=1}^m v_{io} x_{ij} = 1$ $u_{ro} \geq \epsilon, \quad 1 \leq r \leq s$ $v_{io} \geq \epsilon, \quad 1 \leq i \leq m$	$\text{Min} \quad \varphi_o = \sum_{j=1}^R v_{jo} x_{jo}$ sujeto a $\sum_{r=1}^s u_{ro} y_{rj} - \sum_{i=1}^m v_{io} x_{ij} \leq 0, \quad 1 \leq j \leq n$ $\sum_{r=1}^s u_{ro} y_{rj} = 1$ $u_{ro} \geq \epsilon, \quad 1 \leq r \leq s$ $v_{io} \geq \epsilon, \quad 1 \leq i \leq m$

En nuestra aplicación, como ocurre en la mayoría de trabajos de análisis de eficiencia se trabaja con los modelos duales de los anteriores, que son los siguientes:

ORIENTACIÓN OUTPUT	ORIENTACIÓN INPUT
Min $\theta_o - \epsilon \left(\sum_{i=1}^m s_i^+ + \sum_{r=1}^s s_r^- \right)$	Max $\eta_o + \epsilon \left(\sum_{i=1}^m s_i^+ + \sum_{r=1}^s s_r^- \right)$
sujeto a	sujeto a
$\sum_{j=1}^n \lambda_j x_{ij} + s_i^+ = \theta_0 x_{io}, \quad 1 \leq i \leq m$	$\sum_{j=1}^n \lambda_j x_{ij} + s_i^+ = x_{io}, \quad 1 \leq i \leq m$
$\sum_{j=1}^n \lambda_j y_{rj} - s_r^- = y_{ro}, \quad 1 \leq r \leq s$	$\sum_{j=1}^n \lambda_j y_{rj} - s_r^- = \eta_0 y_{ro}, \quad 1 \leq r \leq s$
$\lambda_j \geq 0, \quad 1 \leq j \leq n$	$\lambda_j \geq 0, \quad 1 \leq j \leq n$
$s_i^+ \geq 0$	$s_i^+ \geq 0$
$s_r^- \geq 0$	$s_r^- \geq 0$

donde la variable θ_o (ó η_o) representa el ratio de eficiencia de la DMU que se está analizando, las variables λ_j son los parámetros con los que se obtiene el grupo de referencia de la DMU objeto de estudio y las variables s_i^+ y s_r^- son variables de holgura.

Suele preferirse el enfoque dual, en lugar del primal, porque el dual tiene $m+s$ restricciones mientras que el primal $n+1$. Cooper et al. [10] establecieron que para que el análisis tenga sentido debería verificarse $n+1 > m+s$. En este sentido, resulta más eficiente resolver el dual, y determinar de forma más directa los pesos asociados a cada uno de los inputs y outputs.

3 Análisis envolvente de datos con incertidumbre o datos imprecisos

En los procesos de toma de decisiones, cuando se utiliza un modelo matemático éste puede verse afectado por los valores numéricos que han sido introducidos. Conviene ser conscientes de que la validez de los resultados puede depender de la asignación numérica a parámetros desconocidos, para los que sólo podemos tener en cuenta estimaciones o conjjeturas. En ocasiones no hay realmente ninguna base fundada para suponer que el parámetro en cuestión va a seguir una distribución de probabilidad concreta. Por tanto, no se trata de una *incertidumbre estocástica*, donde es posible un tratamiento probabilístico, sino de otro tipo de incertidumbre para la que resulta muy útil el uso de la teoría de conjuntos borrosos.

La teoría de conjuntos borrosos, concebida por Zadeh en la década de los sesenta [33], incluye la incertidumbre en el propio formalismo. En esencia, consiste en sustituir los conjuntos tradicionales, a los cuales un elemento dado puede pertenecer o no, por las funciones de pertenencia, que son aplicaciones de un conjunto referencial dado X en el intervalo $[0, 1]$. Es decir, un conjunto (o subconjunto) borroso \tilde{A} de X es

$$\tilde{A} = \{ (x, \mu_{\tilde{A}}(x)), \quad x \in X \}$$

donde $\mu_{\tilde{A}}(x)$ es una función que indica el grado de pertenencia a \tilde{A} de los elementos de X llamada función de pertenencia .

Por ejemplo, un conjunto borroso $\tilde{5}$ podría representarse gráficamente de la forma siguiente:

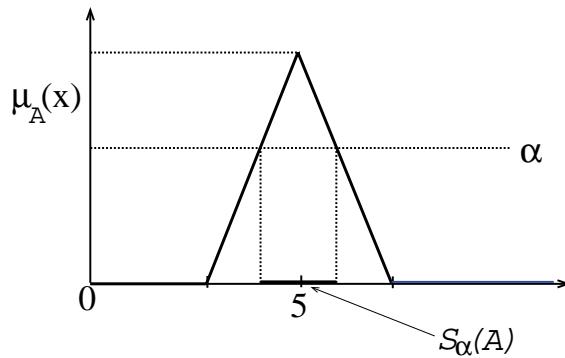


Fig 1: Función de pertenencia para el número borroso $\tilde{5}$.

Un grado de pertenencia nulo se interpreta como no pertenencia, el 1 como pertenencia en el sentido booleano y los números intermedios reflejan una pertenencia incierta, que será interpretada de diversos modos según cada aplicación. La potencia de esta teoría se debe a que a través de la pertenencia a un conjunto se puede modelizar cualquier situación.

A continuación recordamos algunas definiciones básicas que permiten formular esta teoría:

- Un conjunto borroso \tilde{A} es *normalizado* si, y sólo si, $\sup_{x \in X} \mu_{\tilde{A}}(x) = 1$
- Un α -corte de un conjunto \tilde{A} (ver Fig. 1) es el conjunto

$$S_\alpha(\tilde{A}) = \{x \in X : \mu_{\tilde{A}}(x) \geq \alpha\}.$$

- c) Un conjunto borroso es *convexo* si cada α -corte es convexo.
- d) Un *número borroso* es un conjunto borroso normal y convexo con función de pertenencia una función real continua a trozos.

Entre todas las posibilidades de números borrosos, los más utilizados son los que se conocen como números *LR*-borrosos:

Definición 18 *Un número borroso \tilde{M} se dice que es un número LR-borroso,*

$$\tilde{M} = (m^L, m^R, \alpha^L, \alpha^R)_{L,R},$$

si su función de pertenencia tiene la forma siguiente:

$$\mu_{\tilde{M}}(r) = \begin{cases} L\left(\frac{m^L - r}{\alpha^L}\right) & r \leq m^L \\ 1 & m^L \leq r \leq m^R \\ R\left(\frac{r - m^R}{\alpha^R}\right) & r \geq m^R \end{cases}$$

donde L y R son funciones referencia, es decir, $L, R: [0, +\infty[\rightarrow [0, 1]$ son estrictamente decrecientes en $\text{supp}(\tilde{M}) = \{r : \mu_{\tilde{M}}(r) > 0\}$ y semicontinuas superiormente de modo que

$$L(0) = R(0) = 1.$$

Cuando el soporte de $\mu_{\tilde{M}}$ está acotado, las funciones L y R se definen en $[0, 1]$ y se verifica $L(1) = R(1) = 0$. Gráficamente, sería lo siguiente:

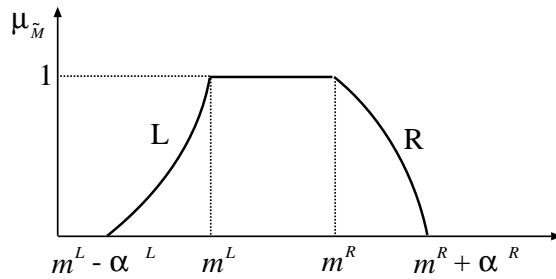


Fig 2: Función de pertenencia para el número LR -borroso.

Cuando las funciones L y R son lineales, es decir

$$L(z) = R(z) = \max\{0, 1 - z\},$$

el número borroso se denomina *trapezoidal* y si además $m^L = m^R$ entonces se tiene *número borroso triangular*

Aunque hay diversas formas de evaluar una combinación lineal de números borrosos, nosotros emplearemos una de las más utilizadas en programación lineal borrosa: la T -norma mínimo que, para números LR -borrosos consiste en lo siguiente: Dados n números LR -borrosos $\tilde{a}_j = (a_j^L, a_j^R, \alpha_j^L, \alpha_j^R)_{L,R}$, $j = 1, \dots, n$ y n escalares $x_j > 0$, $j = 1, \dots, n$, tenemos

$$\sum_{j=1}^n \tilde{a}_j x_j = \left(\sum_{j=1}^n a_j^L x_j, \sum_{j=1}^n a_j^R x_j, \sum_{j=1}^n \alpha_j^L x_j, \sum_{j=1}^n \alpha_j^R x_j \right)_{L,R}, \quad (17.2)$$

donde L y R son las funciones de referencia comunes a todos ellos.

Ordenación de números borrosos

La ordenación de números borrosos ha sido, y es, un tema que ha ocupado a muchos investigadores, puesto que en la práctica puede depender mucho de la situación modelizada y del entorno al que se aplica. En muchas ocasiones, una forma práctica de resolverlo ha sido comparar los números mediante la comparación de sus α -cortes. Aunque con este método no se maneja toda la información que recogen los números borrosos, lo cierto es que ha sido uno de los métodos más utilizados en programación matemática borrosa.

Para precisar las ordenaciones a la que nos estamos refiriendo, conviene precisar el concepto de máximo de dos números borrosos:

Definición 19 *Dados \tilde{M} y \tilde{N} dos números borrosos, $\tilde{M} \vee \tilde{N}$ representa el número borroso cuya función de pertenencia viene dada por*

$$\mu_{\tilde{M} \vee \tilde{N}}(r) = \sup_{r=s \vee t} \{\mu_{\tilde{M}}(s) \wedge \mu_{\tilde{N}}(t)\}. \quad (17.3)$$

Dubois and Prade propusieron ordenar dos números borrosos \tilde{M} y \tilde{N} de la forma siguiente [31]:

$$\tilde{M} \gtrsim \tilde{N} \iff \tilde{M} \vee \tilde{N} = \tilde{M}, \quad (17.4)$$

Posteriormente, Ramík y Římánek [31] establecieron una caracterización más operativa de este orden en términos de α -cortes:

Proposición 5 *Dados dos números borrosos \tilde{M} y \tilde{N} , entonces*

$$\tilde{M} \gtrsim \tilde{N} \iff \begin{cases} \inf\{s : \mu_{\tilde{M}}(s) \geq h\} \geq \inf\{t : \mu_{\tilde{N}}(t) \geq h\} \\ \text{y} \\ \sup\{s : \mu_{\tilde{M}}(s) \geq h\} \geq \sup\{t : \mu_{\tilde{N}}(t) \geq h\} \end{cases} \quad \forall h \in [0, 1] \quad (17.5)$$

En particular, para números borrosos trapezoidales $\tilde{M} = (m^L, m^R, \alpha^L, \alpha^R)$ y $\tilde{N} = (n^L, n^R, \beta^L, \beta^R)$, la expresión (17.5) es equivalente a

$$\tilde{M} \gtrsim \tilde{N} \iff \begin{cases} m^L \geq n^L, & m^L - \alpha^L \geq n^L - \beta^L \\ m^R \geq n^R, & m^R + \alpha^R \geq n^R + \beta^R \end{cases} \quad (17.6)$$

A pesar de que esta ordenación es muy clara desde el punto de vista teórico, lo cierto es que en la práctica suele conducir a situaciones de indecisión. Para evitar este inconveniente, Tanaka, Ichihasi y Asai propusieron una noción menos restrictiva de orden [31]:

Definición 20 *Sean \tilde{M} y \tilde{N} dos números borrosos y un número real $h \in [0, 1]$. Entonces,*

$$\tilde{M} \gtrsim {}^h \tilde{N} \iff \begin{cases} \inf\{s : \mu_{\tilde{M}}(s) \geq k\} \geq \inf\{t : \mu_{\tilde{N}}(t) \geq k\} \\ \text{y} \\ \sup\{s : \mu_{\tilde{M}}(s) \geq k\} \geq \sup\{t : \mu_{\tilde{N}}(t) \geq k\} \end{cases} \quad \forall k \in [h, 1] \quad (17.7)$$

Si los números que se comparan son *LR*-borrosos con soporte acotado, para un h dado, este método de ordenación resulta

$$\tilde{M} \gtrsim {}^h\tilde{N} \leftrightarrow \begin{cases} m^L - L^*(h) \alpha^L \geq n^L - L^*(h) \beta^L, \\ m^R + R^*(h) \alpha^R \geq n^R + R^*(h) \beta^R, \end{cases} \quad (17.8)$$

donde $L^*(k) = \sup \{z : L(z) \geq k\}$, $R^*(k) = \sup \{z : R(z) \geq k\}$.

Este orden, al ser menos restrictivo que (17.6) permite abordar para cada nivel h , muchas situaciones que en global no estaban representadas por números borrosos ordenados, es decir de las que no se podía asegurar ni $\tilde{M} \gtrsim \tilde{N}$ ni $\tilde{N} \gtrsim \tilde{M}$.

4 Modelos de eficiencia con incertidumbre

En los últimos años han aparecido muchos modelos que permiten calcular la eficiencia teniendo en cuenta la incertidumbre. Estos modelos no sólo difieren desde el punto de vista técnico, sino que abordan la eficiencia desde enfoques diferentes. En este trabajo presentamos ejemplos de dos enfoques que agrupan a gran número de ellos:

- a) Dado un conjunto $\{DMU_i\}_{i=1}^n$ de n DMUs, podemos definir el conjunto borroso

$$\widetilde{Ef} = \left\{ \left(DMU_i, \mu_{\widetilde{Ef}}(DMU_i) \right), \quad i = 1, \dots, n \right\},$$

de manera que será más eficiente aquella DMU que tenga un grado de pertenencia mayor a \widetilde{Ef} .

- b) Para cada DMU se obtiene la eficiencia como un conjunto borroso. Por lo tanto, disponemos de n conjuntos borrosos $\{\tilde{E}_i\}_{i=1}^n$. En este contexto, analizar la eficiencia consiste en ordenar los números borrosos \tilde{E}_i .

Para mostrar algunos modelos de estos enfoques partimos de que la incertidumbre en los inputs y los outputs está expresada mediante números borrosos trapezoidales⁴, es decir

$$\begin{aligned} \tilde{x}_{ij} &= (x_{ij}^L, x_{ij}^R, \alpha_{ij}^L, \alpha_{ij}^R), \quad i = 1, \dots, m, j = 1, \dots, n, \\ \tilde{y}_{rj} &= (y_{rj}^L, y_{rj}^R, \beta_{rj}^L, \beta_{rj}^R), \quad r = 1, \dots, s, j = 1, \dots, n. \end{aligned}$$

Por tanto, sus α -cortes se puede escribir como

$$x_{ij} = [x_{ij}^L + \alpha_{ij}^L(\alpha - 1), x_{ij}^R - \alpha_{ij}^R(\alpha - 1)], \quad \alpha \in [0, 1]$$

⁴En realidad los métodos que aquí se presentan son válidos para números *LR*-borrosos. Sin embargo no creemos conveniente complicar la notación y los cálculos puesto que en las aplicaciones que presentamos (y en la gran mayoría de las que aparecen en la literatura) suelen usarse números trapezoidales o triangulares.

$$y_{rj} = [y_{rj}^L + \beta_{rj}^L(\alpha - 1), y_{rj}^R - \beta_{rj}^R(\alpha - 1)], \quad \alpha \in [0, 1]$$

4.1 Cálculo de eficiencia con datos inciertos

Para ilustrar este método seguiremos la propuesta de León et al [17]. Partimos del modelo

$$\begin{aligned} & \min \theta_0 \\ \text{s.a. : } & \sum_{j=1}^n \lambda_j \tilde{x}_{ij} \lesssim \theta_0 \tilde{x}_{i0} \quad i = 1, \dots, m \\ & \sum_{j=1}^n \lambda_j \tilde{y}_{rj} \gtrsim \tilde{y}_{r0} \quad r = 1, \dots, s \\ & \lambda_j \geq 0 \quad j = 1, \dots, n \end{aligned} \tag{17.9}$$

Como los inputs y los outputs son números *LR*-borrosos, las restricciones pueden verse como desigualdades entre números *LR*. Si en particular \lesssim se interpreta como hemos dicho en (17.5) y las combinaciones lineales como (17.2), el problema puede escribirse como

$$\begin{aligned} (\mathbf{P}^h) \quad & \min \theta_0 \\ \text{sujeto a : } & \sum_{j=1}^n \lambda_j x_{ij}^L \leq \theta_0 x_{i0}^L \quad i = 1, \dots, m \\ & \sum_{j=1}^n \lambda_j x_{ij}^R \leq \theta_0 x_{i0}^R \quad i = 1, \dots, m \\ & \sum_{j=1}^n \lambda_j x_{ij}^L - (1-h) \sum_{j=1}^n \lambda_j \alpha_{ij}^L \leq \theta_0 x_{i0}^L - (1-h) \theta_0 \alpha_{i0}^L \quad i = 1, \dots, m \\ & \sum_{j=1}^n \lambda_j x_{ij}^R + (1-h) \sum_{j=1}^n \lambda_j \alpha_{ij}^R \leq \theta_0 x_{i0}^R + (1-h) \theta_0 \alpha_{i0}^R \quad i = 1, \dots, m \\ & \sum_{j=1}^n \lambda_j y_{rj}^L \geq y_{r0}^L \quad r = 1, \dots, s \\ & \sum_{j=1}^n \lambda_j y_{rj}^R \geq y_{r0}^R \quad r = 1, \dots, s \\ & \sum_{j=1}^n \lambda_j y_{rj}^L - (1-h) \sum_{j=1}^n \lambda_j \beta_{rj}^L \geq y_{r0}^L - (1-h) \beta_{r0}^L \quad r = 1, \dots, s \\ & \sum_{j=1}^n \lambda_j y_{rj}^R + (1-h) \sum_{j=1}^n \lambda_j \beta_{rj}^R \geq y_{r0}^R + (1-h) \beta_{r0}^R \quad r = 1, \dots, s \\ & \lambda_j \geq 0 \quad j = 1, \dots, n \end{aligned} \tag{17.10}$$

En este caso, para el conjunto $\{DMU_j\}_{j=1}^n$ de las n DMUs, podemos definir un conjunto borroso de eficiencia como

$$\widetilde{Ef} = \left\{ \left(DMU_j, \mu_{\widetilde{Ef}}(DMU_j) \right), \quad j = 1, \dots, n \right\}.$$

donde la función de pertenencia está dada por

$$\mu_{\widetilde{Ef}}(DMU_j) = \begin{cases} 0 & \text{si } \theta_j^*(h) \neq 1 \quad \forall h \in [0, 1], \\ \sup\{h : \theta_j^*(h) = 1\} & \text{si } \theta_j^*(h) = 1 \quad \text{para algún } h \in [0, 1]. \end{cases}$$

Y entonces, el grado de pertenencia $\mu_{\widetilde{Ef}}(DMU_j)$ proporciona una medida de la eficiencia de la j -ésima DMU.

4.2 Ordenación de DMUs por eficiencias

Para este método seguiremos el enfoque de Kao y Liu [16]. El objetivo es ordenar las DMUs por eficiencias suponiendo que tanto los inputs como los outputs son números LR -borrosos. Para calcular la eficiencia de cada DMU y calcular la eficiencia de cada DMU de la

- a) Calcula la eficiencia de la j -ésima DMU en el caso más desfavorable, es decir se suponen los mayores inputs para ella y los menores para el resto, mientras que se suponen los menores outputs para la unidad analizada y los mayores para las restantes. El cálculo se realiza a través de α -cortes, es decir, para cada $\alpha \in [0, 1]$ se calcula el valor $(E_j)_\alpha^L$ a partir del modelo siguiente:

$$(E_j)_\alpha^L = \begin{array}{ll} \max & \sum_{r=1}^s u_r(y_{rj}^L + \beta_{rj}^L(\alpha - 1))/v_i(x_{ij}^R - \alpha_{ij}^R(\alpha - 1)) \\ \text{s.a.} & \sum_{k=1}^t u_k(y_{kj}^L + \beta_{kj}^L(\alpha - 1))/v_j(x_{ij}^R - \alpha_{ij}^R(\alpha - 1)) \leq 1 \\ & \sum_{k=1}^t u_k(y_{kj}^R - \beta_{kj}^R(\alpha - 1))/v_j(x_{ij}^L + \alpha_{ij}^L(\alpha - 1)) \leq 1, \quad 1 \leq i \leq n \\ & u_k, v_j \geq \epsilon > 0, \end{array}$$

- b) Calcula la eficiencia de la j -ésima DMU en el caso más favorable, es decir se suponen los menores inputs para ella y los mayores para el resto y por el contrario se suponen los mayores outputs para la DMU que se está analizando y los menores para las restantes. Como en el caso anterior, para cada $\alpha \in [0, 1]$ el valor $(E_j)_\alpha^R$ se obtiene como

$$\begin{aligned}
 (E_j)_{\alpha}^R = & \max \sum_{r=1}^s u_r (y_{rj}^R + \beta_{rj}^R(\alpha - 1)) / v_i (x_{ij}^L - \alpha_{ij}^L(\alpha - 1)) \\
 \text{s.a. } & \sum_{k=1}^t u_k (y_{rj}^R + \beta_{rj}^R(\alpha - 1)) / v_j (x_{ij}^L - \alpha_{ij}^L(\alpha - 1)) \leq 1 \\
 & \sum_{k=1}^t u_k (y_{rj}^L - \beta_{rj}^L(\alpha - 1)) / v_j (x_{ij}^R + \alpha_{ij}^R(\alpha - 1)) \leq 1, \quad 1 \leq i \leq n \\
 & u_k, v_j \geq \epsilon > 0,
 \end{aligned}$$

Podemos considerar $(E_j)_{\alpha}^L$ y $(E_j)_{\alpha}^U$ como funciones de $\alpha \in [0, 1]$, así si ambas son invertibles tenemos

$$L(z) := [(E_j)_{\alpha}^L]^{-1}, \quad R(z) := [(E_j)_{\alpha}^U]^{-1}$$

Es decir, que para la j -ésima DMU tenemos un número LR -borroso \tilde{E}_j . Si repetimos esto con cada DMU tenemos n números borrosos $\{\tilde{E}_r\}_{r=1}^n$ que, una vez ordenados, nos proporcionan una ordenación por eficiencias.

4.3 DEA con tolerancias

Con los métodos anteriores, puede ocurrir que al tratar con problemas reales de grandes dimensiones, para cada α la eficiencia de cada DMU es un intervalo excesivamente grande, y con ello el análisis de eficiencia o la ordenación por eficiencias de las DMUs puede resultar complicado. Por otro lado, la propia modelización o el cálculo de la función de pertenencia para la eficiencia puede resultar complejo e incluso tener que recurrir a métodos heurísticos. A continuación presentamos un modelo que intenta paliar estas dos dificultades.

Suponemos que los inputs y los outputs vienen dados con tolerancias. Es decir, hay valores x_{ij}, y_{rj} que podemos considerar que son los que tienen mayores posibilidades de ser ciertos, pero sobre los que tenemos alguna incertidumbre que viene expresada por los intervalos siguientes⁵:

$$[x_{ij} - \alpha_{ij}^L, x_{ij} + \alpha_{ij}^R], \quad i = 1, \dots, m, j = 1, \dots, n, \quad (17.11)$$

$$[y_{rj} - \beta_{rj}^L, y_{rj} + \beta_{rj}^R], \quad r = 1, \dots, s, j = 1, \dots, n. \quad (17.12)$$

Nota: A los valores $\alpha_{ij}^L, \alpha_{ij}^R, \beta_{rj}^L, \beta_{rj}^R$ les llamaremos *tolerancias* cuando no sea necesario hacer distinciones específicas.

Nuestra propuesta, que recoge en parte el esquema de las dos que hemos presentado anteriormente, parte de un modelo de DEA clásico y consta de los tres pasos siguientes:

⁵Nótese que no es más que un caso particular de los inputs y outputs dados como números LR -borrosos. Estos intervalos corresponden al α -corte de números triangulares con $\alpha = 0$.

- a) Calculamos la eficiencia de cada DMU con los datos x_{ij}, y_{rj} , a la que denominaremos eficiencia original, y la denotaremos $\{E_j^o\}_{j=1}^n$. Para ello, utilizamos, por ejemplo, el modelo siguiente:

$$\begin{aligned} (\text{E}_j^o) \quad & \min \theta_0 \\ \text{sujeto a :} \quad & \sum_{j=1}^n \lambda_j x_{ij} + \leq \theta_0 x_{i0} \quad i = 1, \dots, m \\ & \sum_{j=1}^n \lambda_j y_{rj} \geq y_{r0} \quad r = 1, \dots, s \\ & \lambda_j \geq 0 \quad j = 1, \dots, n \end{aligned} \quad (17.13)$$

- b) Calculamos la eficiencia de la j -ésima DMU de forma desfavorable. Suponemos los mayores inputs para ella, $x_{ij} + \alpha_{ij}^R$, y el resto se suponen ciertos, $x_{ik}, \quad k \neq j$, mientras que se suponen los menores outputs para la unidad analizada, $y_{rj} - \beta_{rj}^L$, y los de las restantes se suponen ciertos $y_{rk}, \quad k \neq j$. Si ahora sustituimos estos valores de inputs y outputs en el modelo DEA elegido, obtenemos las “eficiencias peores” de las n DMUs, $\{E_j^P\}_{j=1}^n$:

$$\begin{aligned} (\text{E}_j^P) \quad & \min \theta_0 \\ \text{sujeto a :} \quad & \sum_{j=1}^n \lambda_j x_{ij} + \lambda_j \alpha_{i0}^R \leq \theta_0 (x_{i0} + \alpha_{i0}^R) \quad i = 1, \dots, m \\ & \sum_{j=1}^n \lambda_j y_{rj} - \lambda_j \beta_{r0}^L \geq y_{r0} - \beta_{r0}^L \quad r = 1, \dots, s \\ & \lambda_j \geq 0 \quad j = 1, \dots, n \end{aligned} \quad (17.14)$$

- c) Calculamos la eficiencia de la j -ésima DMU de forma favorable. Suponemos los menores inputs para ella, $x_{ij}^R - \alpha_{ij}^L$, y el resto se suponen ciertos, $x_{ik}, \quad k \neq j$, mientras que se suponen los mayores outputs para la unidad analizada, $y_{rj} + \beta_{rj}^R$, y los de las restantes se suponen ciertos $y_{rk}, \quad k \neq j$. Al sustituir en el modelo DEA obtenemos las “eficiencias mejores” de las n DMUs, $\{E_j^M\}_{j=1}^n$:

$$\begin{aligned} (\text{E}_j^M) \quad & \min \theta_0 \\ \text{sujeto a :} \quad & \sum_{j=1}^n \lambda_j x_{ij} - \lambda_j \alpha_{i0}^L \leq \theta_0 (x_{i0} - \alpha_{i0}^L) \quad i = 1, \dots, m \\ & \sum_{j=1}^n \lambda_j y_{rj} + \lambda_j \beta_{r0}^R \geq y_{r0} + \beta_{r0}^R \quad r = 1, \dots, s \\ & \lambda_j \geq 0 \quad j = 1, \dots, n \end{aligned} \quad (17.15)$$

Como veremos en la aplicación a las ligas de fútbol profesional, el resultado será que la eficiencia de DMU viene expresada por un intervalo

$$[E_j^P, E_j^M], \quad 1 \leq j \leq n,$$

que, nos permite analizar las posibilidades de variabilidad en la eficiencia de cada DMU. Además, por propia construcción, se verifica

$$E_j^o \in [E_j^P, E_j^M], \quad 1 \leq j \leq n.$$

5 Aplicación a las ligas de fútbol profesionales

Las características básicas de la competición futbolística hacen que ésta pueda ser analizada como cualquier actividad productiva. S. Rottenberg utilizó por primera vez el concepto de “función de producción deportiva” en 1956 para el béisbol [22], aunque la primera aportación empírica, en 1974, se debe a G. W. Scully [25]. A estos trabajos pioneros les sucedieron varios artículos centrándose en diferentes deportes: el baloncesto [34, 24], el cricket [23], el fútbol [29, 30], o bien centrándose en la función de producción y el estudio de la eficiencia: en el rugby [5] o en la liga de fútbol inglesa [6, 13].

Formalmente, el proceso productivo del fútbol puede describirse mediante una función de producción

$$Y: \mathbb{R}^k \implies \mathbb{R}^n,$$

cuyas funciones coordenadas $Y_i(X_j)$, $1 \leq i \leq n$, $1 \leq j \leq k$ representan la medida del output del fútbol para el equipo i -ésimo (normalmente el porcentaje de puntos o victorias sobre el total, o el ratio o la diferencia entre goles marcados y encajados) y X_j es un vector de inputs.

La práctica habitual es aproximar los inputs que entran en la función de producción con variables que miden las habilidades técnicas de los jugadores. De esta forma, en la literatura se han utilizado diferentes métodos de estimación: tomar como unidad cada partido de fútbol de una liga [5], o el agregado a lo largo de una o varias ligas [13], que tienen en común el tratar de cuantificar la importancia relativa de cada uno de los inputs en la obtención de los outputs y la intención de explicar la razón por la que difiere la productividad aparente de cada uno de los equipos.

Para abordar este tipo de cuestiones suele analizarse si algunas de estas empresas tienen acceso a tecnologías más avanzadas, y en este caso las estimaciones de funciones de producción suele realizarse a través de técnicas paramétricas que permiten recuperar elasticidades output homogéneas para todas las empresas y controlar las diferencias por las diferencias de nivel entre las distintas funciones de producción individuales, que se interpretan como indicadores del nivel de

desarrollo tecnológico de cada empresa. Sin embargo, cuando el acceso a la tecnologías es similar para todas las empresas, las diferencias en la productividad vienen determinadas por la gestión de los recursos o la eficiencia con la que estos son organizados. Esta posibilidad, conocida como niveles de eficiencia X, será el que más interese en el caso del fútbol, puesto que podemos suponer que todos los equipos de las ligas profesionales tienen acceso a una tecnología común [14]. Por esta razón se utilizará el análisis envolvente de datos para analizar la eficiencia de los equipos.

El análisis de la eficiencia que vamos a realizar tiene en cuenta explícitamente algunas características importantes del proceso productivo del fútbol. Uno de los problemas más relevantes en la interpretación de los resultados que se derivan de las estimaciones paramétricas de las funciones de producción es que las medidas del output utilizadas combinan la productividad ofensiva (producir goles) con la eficiencia defensiva (evitar recibirlos). Como consecuencia de ello los inputs utilizados en las estimaciones son una combinación de indicadores de capacidad ofensiva y defensiva de los equipos, por lo que los signos esperados son para algunos inputs positivos, para otros negativos y para otros indeterminados. En una estimación de funciones de producción estándar para empresas los inputs siempre deben presentar un signo positivo, ya que si añaden output a las empresas, con lo que su correlación con éste debe de ser positiva, o las empresas no los utilizarían. Con los modelos DEA, donde puede haber varios outputs, este problema queda resuelto y, de hecho, en el análisis de eficiencia se ha separado siempre la eficiencia ofensiva de la defensiva.

Como en cualquier problema económico en el que se analizan funciones o fronteras de producción, lo deseable es contar con indicadores lo más fiables posibles del flujo de servicios que proporcionan los distintos factores productivos. En el caso del fútbol el único factor productivo son los propios jugadores, que organizados por un entrenador disputan los partidos. En consecuencia, lo deseable sería contar con indicadores precisos que midieran las habilidades concretas de los distintos jugadores que forman las plantillas o, lo que es lo mismo con indicadores del capital humano específico para jugar al fútbol de cada jugador. Aunque en algunos trabajos se ha utilizado información referida al coste económico de los jugadores para aproximar sus habilidades, ésta información es difícil de conseguir, además de presentar serias dudas en su interpretación.

En este trabajo hemos optado por incluir como inputs, indicadores técnicos del fútbol que entendemos ofrecen una buena aproximación a la calidad de las plantillas, a la estructura del equipo o al sistema de juego empleado. Dado que ambos niveles (ofensivo/defensivo, casa/fuera) son relevantes se ha procedido a seleccionar, dentro de la base de datos, aquellos inputs que estuvieran disponibles para las dos competiciones ligueras y en todas las combinaciones de análisis de la eficiencia que pretendemos abordar: defensiva general (casa y fuera), defensiva en casa, defensiva fuera, ofensiva general (casa y fuera), ofensiva en casa y ofensiva

fueras. Los datos utilizados han sido facilitados de forma desinteresada por la empresa GECA Sport, que posee una de las bases de datos más extensas disponibles referidas al fútbol español e italiano. Aunque en esta base de datos se recoge una gran variedad de aspectos del juego, con amplia información desglosada en muy diferentes categorías, nosotros sólo hemos utilizado una parte pequeña que, a nuestro juicio, permite analizar la eficiencia técnica de los clubes.

De esta forma se han seleccionado cuatro inputs ofensivos y cuatro defensivos elegidos que presentan correlaciones positivas con los distintos outputs, si bien en algún caso concreto podría ser nula⁶. En cuanto a los outputs ofensivos y defensivos, hemos seleccionado dos: los puntos y el número de goles. De esta forma estamos recogiendo el hecho de que a los equipos se les exige ganar (maximizar los puntos conseguidos), pero valorando que se consigan muchos goles⁷.

En la tabla siguiente detallamos los inputs y outputs que emplearemos en nuestro estudio:

Tabla 1: Inputs y outputs seleccionados.

	Ofensivos	Defensivos
Inputs	Remates realizados	Remates recibidos
	Jugadas de ataque	Jugadas del contrario en área propia
	Centros al área realizados	Centros al área recibidos
	Minutos de posesión	Minutos de posesión
Outputs	Puntos	No puntos*
	Goles marcados	Goles encajados

* El número de no puntos se ha calculado como:

$$\text{No puntos} = 3 \times \text{núm. partidos} - \text{puntos}$$

6 Resultados computacionales

A pesar de que, hemos analizado las eficiencias defensivas (general, en casa y fuera) y ofensivas (general, en casa y fuera) de las ligas 2000/2001, 2001/2002 y 2002/2003 para la ligas profesionales española e italiana, pensamos que, para

⁶Téngase en cuenta, por ejemplo, que aunque la correlación entre los minutos de posesión de balón del adversario y los puntos dejados de obtener es claramente positiva en el total de la liga, la misma correlación no es distinta de cero en casa, aunque sí que es claramente positiva fuera de casa.

⁷Aunque en el mundo del fútbol lo más importante es ganar, los aficionados prefieren hacerlo por la máxima diferencia de goles posible por al menos dos razones. La primera es que un partido con muchos goles a favor es un espectáculo mejor valorado. La segunda es que en caso de empate a puntos en la clasificación entre dos equipos en ligas como la española queda por delante aquel que tiene la mayor diferencia de goles a favor.

evitar un exceso de resultados computacionales, en esta sección sólo presentaremos los datos ofensivos y defensivos de la liga 2002/2003.

Para utilizar la metodología propuesta en la sección 4.3, hemos dividido ambas ligas en tres grupos formados por los 5 equipos que encabezan la clasificación (Grupo 1), los del centro de la tabla (Grupo 2), 10 equipos en España y 8 en Italia, y los 5 equipos del final de la tabla (Grupo 3). Analizando las variaciones conjuntas de las dos ligas durante tres temporadas, hemos asignado las siguientes tolerancias a los diferentes grupos:

Tabla 2: Tolerancias para los inputs y outputs.

	Casa		Fuera		General	
	Inputs	Outputs	Inputs	Outputs	Inputs	Outputs
Grupo 1	3	1.5	5	2.5	4	2
Grupo 2	5	2.5	7	3.5	6	3
Grupo 3	7	3.5	9	4.5	8	4

Nota: Los datos de la tablan expresan porcentajes.

Esto significa que si, por ejemplo, estamos analizando la eficiencia *en casa*, todos los equipos que pertenecen al Grupo 1, verifican

$$\alpha_{ij}^L = \alpha_{ij}^R = 0.03x_{ij}, \quad \beta_{rj}^L = \beta_{rj}^R = 0.015y_{rj}$$

siendo α_{ij} y β_{rj} los expresados en (17.11) y (17.12) respectivamente. Entonces, según se propone en el modelo de la sección 4.3, para calcular las eficiencias mejor y peor de la k -ésima DMU (que pertenece al Grupo 1) utilizamos las tolerancias siguientes:

$$\alpha_{ij}^L = \alpha_{ij}^R = \begin{cases} 0.03x_{ij} & \text{si } j = k \\ 0 & \text{si } j \neq k \end{cases} \quad \beta_{rj}^L = \beta_{rj}^R = \begin{cases} 0.015y_{rj} & \text{si } j = k \\ 0 & \text{si } j \neq k \end{cases}$$

donde $i \in \{1, \dots, m\}$ y $r \in \{1, \dots, s\}$.

En las tablas 3 y 4 mostramos los resultados para los diferentes escenarios del análisis de la eficiencia ofensiva y en las tablas 5 y 6 los de la eficiencia defensiva. El orden en el que aparecen los equipos aparece

Tabla 3: Cálculo de eficiencia ofensiva en la liga italiana 2002-2003.

EQUIPO	Peor	Original	Mejor	Peor	Original	Mejor	Peor	Original	Mejor
	GENERAL			CASA			FUERA		
Juventus FC	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9407	1,0000	1,0000
Inter	1,0000	1,0000	1,0000	0,9528	0,9964	1,0000	1,0000	1,0000	1,0000
Milan AC	0,8214	0,8721	0,9249	0,9356	0,9790	1,0000	0,7601	0,8183	0,8830
SS Lazio	0,8805	0,9356	0,9931	0,8825	0,9225	0,9637	1,0000	1,0000	1,0000
AC Parma	0,9376	0,9961	1,0000	0,9271	0,9687	1,0000	0,8086	0,8724	0,9395
Udinese	0,9092	0,9970	1,0000	0,9357	1,0000	1,0000	0,7099	0,7920	0,8801
Chievo	0,9484	1,0000	1,0000	1,0000	1,0000	1,0000	0,7150	0,7957	0,8822
AC Roma	0,7987	0,8752	0,9563	0,8804	0,9499	1,0000	0,6339	0,7065	0,7844
Brescia	0,6519	0,7139	0,7797	0,6269	0,6750	0,7283	0,6162	0,6857	0,7603
Perugia	0,6772	0,7416	0,8100	0,8050	0,8670	0,9352	0,6185	0,6902	0,7660
Bologna	0,6624	0,7265	0,7946	0,7709	0,8317	0,8957	0,5435	0,6058	0,6727
Modena	0,6330	0,6940	0,7587	0,6444	0,6953	0,7487	0,5354	0,5936	0,6601
Empoli	0,6802	0,7457	0,8152	0,5903	0,6352	0,6857	0,7327	0,8169	0,9072
Atalanta	0,5759	0,6503	0,7308	0,5547	0,6171	0,6840	0,6644	0,7603	0,8706
Reggina	0,6069	0,6853	0,7702	0,6863	0,7631	0,8456	0,3993	0,4593	0,5250
Piacenza	0,6957	0,7861	0,8841	0,7758	0,8624	0,9553	0,5082	0,5815	0,6658
Como	0,5110	0,5773	0,6490	0,6258	0,6964	0,7722	0,3520	0,4036	0,4602
Torino	0,4544	0,5126	0,5757	0,4289	0,4767	0,5281	0,5685	0,6544	0,7484

Tabla 4: Cálculo de eficiencia ofensiva en la liga española 2002-2003.

EQUIPO	Peor	Original	Mejor	Peor	Original	Mejor	Peor	Original	Mejor
	GENERAL			CASA			FUERA		
Real Madrid	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Real Sociedad	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9600	1,0000	1,0000
Dep. A Coruña	0,9149	0,9721	1,0000	0,9429	0,9876	1,0000	0,9428	1,0000	1,0000
R. C. Celta Vigo	0,8686	0,9229	0,9794	0,8132	0,8514	0,8908	0,9345	1,0000	1,0000
Valencia C.F.	0,7589	0,8063	0,8557	0,7094	0,7425	0,7767	0,7773	0,8371	0,9030
F.C. Barcelona	0,7659	0,7985	0,8726	0,8435	0,9109	0,9821	0,5975	0,6659	0,7394
Ath. Bilbao	0,8102	0,8447	0,9231	0,8109	0,8749	0,9422	0,9884	1,0000	1,0000
Real Betis	0,7334	0,7646	0,8356	0,7798	0,8399	0,9058	0,7340	0,8155	0,9059
R.C.D. Mallorca	0,7128	0,7431	0,8120	0,6061	0,6526	0,7039	0,8164	0,9080	1,0000
Sevilla F.C.	0,7289	0,7600	0,8305	0,7883	0,8491	0,9159	0,6209	0,6898	0,7637
Olasauna	0,6256	0,6522	0,7127	0,5601	0,6032	0,6507	0,6369	0,7089	0,7861
Atl. de Madrid	0,5990	0,6245	0,6824	0,6127	0,6600	0,7117	0,6199	0,6898	0,7649
Málaga C.F.	0,6822	0,7112	0,7772	0,7665	0,8256	0,8904	0,5087	0,5642	0,6275
Real Valladolid	0,6932	0,7227	0,7898	0,7406	0,7977	0,8603	0,5420	0,6017	0,6692
Villarreal	0,5908	0,6160	0,6731	0,6311	0,6797	0,7332	0,6227	0,6930	0,7684
Racing Santander	0,6843	0,7587	0,8536	0,7545	0,8387	0,9290	0,7428	0,8509	0,9731
Espanyol	0,6590	0,7306	0,8219	0,7909	0,8803	0,9762	0,5742	0,6589	0,7515
Recreat. Huelva	0,5054	0,5603	0,6303	0,5412	0,6021	0,6674	0,4281	0,4929	0,5637
Dep. Alavés	0,5225	0,5953	0,6517	0,4822	0,5364	0,5945	0,6447	0,7404	0,8451
Rayo Vallecano	0,4585	0,5083	0,5719	0,4545	0,5057	0,5607	0,4772	0,5497	0,6290

Tabla 5: Cálculo de eficiencia defensiva en la liga italiana 2002-2003.

EQUIPO	Peor	Original	Mejor	Peor	Original	Mejor	Peor	Original	Mejor
	GENERAL			CASA			FUERA		
Juventus FC	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Inter	0,7059	0,7194	0,7347	0,7372	0,7480	0,7597	0,8012	0,8202	0,8423
Milan AC	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9992	1,0000	1,0000
SS Lazio	0,7898	0,8049	0,8218	0,6268	0,6360	0,6462	0,8060	0,8251	0,8474
AC Parma	0,7378	0,7519	0,7679	0,6498	0,6593	0,6696	0,8464	0,8471	0,8700
Udinese	0,7589	0,7803	0,8059	0,8334	0,8532	0,8762	0,6128	0,6327	0,6574
Chievo	0,6575	0,6760	0,6982	0,6647	0,6805	0,6989	0,7011	0,7239	0,7521
AC Roma	0,6716	0,6905	0,7132	0,5395	0,5523	0,5672	0,5773	0,5961	0,6194
Brescia	0,6634	0,6821	0,7045	0,6356	0,6507	0,6683	0,6737	0,6956	0,7227
Perugia	0,5332	0,5483	0,5663	0,7772	0,7957	0,8172	0,4393	0,4536	0,4713
Bologna	0,5170	0,5316	0,5490	0,4960	0,5078	0,5215	0,4662	0,4814	0,5002
Modena	0,5115	0,5259	0,5432	0,5187	0,5310	0,5453	0,4336	0,4477	0,4652
Empoli	0,5039	0,5181	0,5351	0,4416	0,4521	0,4643	0,4715	0,4868	0,5058
Atalanta	0,4946	0,5128	0,5359	0,4253	0,4392	0,4563	0,4646	0,4836	0,5085
Reggina	0,4931	0,5113	0,5343	0,4839	0,4997	0,5192	0,3871	0,4029	0,4237
Piacenza	0,3873	0,4016	0,4197	0,3185	0,3289	0,3417	0,3778	0,3933	0,4136
Como	0,4274	0,4432	0,4631	0,3708	0,3829	0,3978	0,3969	0,4132	0,4345
Torino	0,4100	0,4251	0,4443	0,3622	0,3740	0,3885	0,3860	0,4018	0,4225

Tabla 6: Cálculo de eficiencia defensiva en la liga española 2002-2003.

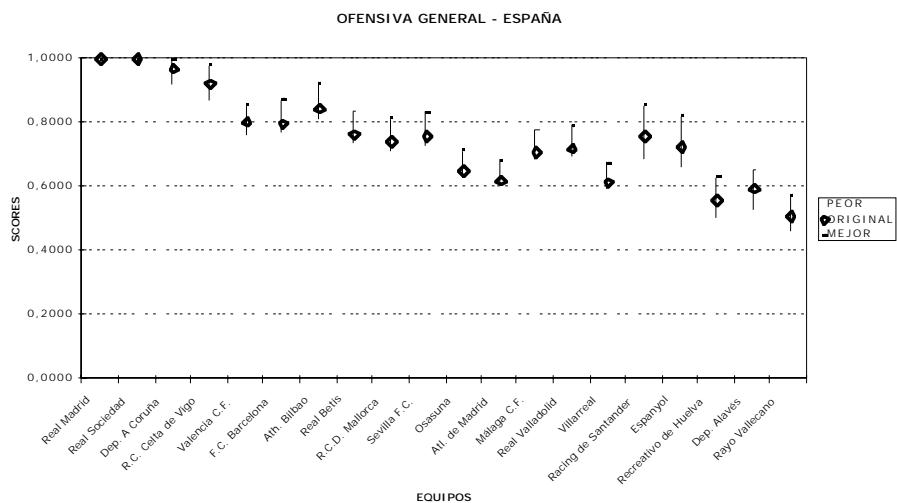
EQUIPO	Peor	Original	Mejor	Peor	Original	Mejor	Peor	Original	Mejor
	GENERAL			CASA			FUERA		
Real Madrid	1,0000	1,0000	1,0000	0,9031	0,9162	0,9306	1,0000	1,0000	1,0000
Real Sociedad	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Dep. A Coruña	0,8992	0,9165	0,9359	0,9187	0,9321	0,9467	0,9332	0,9554	0,9812
R. C. Celta Vigo	0,9697	0,9883	1,0000	0,9992	1,0000	1,0000	0,8871	0,9082	0,9326
Valencia C.F.	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
F.C. Barcelona	0,8035	0,8261	0,8533	0,7042	0,7209	0,7403	0,8620	0,8901	0,9247
Ath. Bilbao	0,6550	0,6735	0,6957	0,8282	0,8478	0,8706	0,6036	0,6233	0,6475
Real Betis	0,7423	0,7632	0,7883	0,9117	0,9333	0,9584	0,7295	0,7533	0,7826
R.C.D. Mallorca	0,6334	0,6514	0,6728	0,4858	0,4973	0,5107	0,7911	0,8169	0,8487
Sevilla F.C.	0,9058	0,9313	0,9619	0,7516	0,7694	0,7901	1,0000	1,0000	1,0000
Olasana	0,8748	0,8995	0,9291	0,6881	0,7044	0,7234	0,8803	0,9090	0,9443
Atl. de Madrid	0,6951	0,7147	0,7382	0,6474	0,6628	0,6806	0,6715	0,6934	0,7203
Málaga C.F.	0,7399	0,7607	0,7857	0,7382	0,7557	0,7760	0,6939	0,7165	0,7444
Real Valladolid	0,8715	0,8960	0,9255	0,8661	0,8866	0,9104	0,8173	0,8440	0,8769
Villarreal	0,6982	0,7180	0,7415	0,7816	0,8002	0,8218	0,6062	0,6259	0,6502
Racing Santander	0,6073	0,6296	0,6580	0,7766	0,8019	0,8331	0,6050	0,6298	0,6623
Espanyol	0,6454	0,6691	0,6993	0,6560	0,6774	0,7038	0,5920	0,6163	0,6481
Recreat. Huelva	0,5643	0,5851	0,6114	0,6129	0,6329	0,6575	0,5170	0,5382	0,5660
Dep. Alavés	0,5424	0,5623	0,5877	0,7371	0,7611	0,7907	0,5659	0,5891	0,6195
Rayo Vallecano	0,5807	0,6020	0,6292	0,6577	0,6791	0,7055	0,5200	0,5413	0,5693

Nota: Las columnas Peor, Original y Mejor representan, respectivamente, los modelos E_j^P , E_j^o y E_j^M descritos en la sección 4.3.

Además de conocer los scores de eficiencia de cada DMU, los intervalos expresados en las tablas anteriores informan a los directivos de los equipos acerca de la estabilidad del análisis de eficiencia. Piénsese que, en un análisis de eficiencia clásico, hay ocasiones en las que una DMU considerada eficiente dejaría de serlo con ligerísimas modificaciones en algún input u output [17], y en este caso las conclusiones que podrían extraerse en las ligas de fútbol podría estar muy distorsionada.

Por otro lado, el análisis de eficiencia por intervalos permite conocer si los equipos han aprovechado de forma adecuada sus posibilidades. Analicemos, por ejemplo, el gráfico de los intervalos de eficiencia ofensiva general en la liga española.

La posición del score original en el intervalo nos informa de si la utilización de los recursos ha sido adecuada. En concreto, cuanto más próximo esté del extremo superior del intervalo, podemos interpretar que más la utilización de recursos ha sido más adecuada.

Fig. 1 Intervalos de eficiencia ofensiva general en la liga española 2002/2003.

7 Conclusiones

El uso de modelos DEA en los que se puedan utilizar datos inciertos permite ampliar el espectro de aplicación del análisis envolvente de datos tradicional. En nuestro caso, hemos utilizado como DMUs los equipos de las ligas de fútbol profesional española e italiana. En la práctica, estos modelos presentan dos tipos de ventajas frente a los tradicionales: ofrecen mayor información de los resultados (desde el punto de vista interno y externo) y tienen en cuenta la estabilidad de los mismos.

Desde el punto de vista interno, el cálculo de eficiencias por intervalos permite conocer el campo de variación de su score de eficiencia frente a la mejora (emporamiento) de cada uno de los inputs y/o outputs. Este análisis, denominado “análisis de incidencia” [3], permite conocer el efecto que provocan las variaciones de cada uno de los diferentes elementos sobre el resultado final.

Este análisis de incidencia puede servir de guía u orientación para el director deportivo y/o entrenador del equipo para conocer qué acciones deben mejorar (aumentar o reducir) para conseguir un aumento en los outputs (goles y puntos).

En cuanto a la información externa, este análisis puede servir para estudiar el efecto de otros factores que afectan a los resultados del equipo. Ahora bien, en este caso necesitamos contar con un referente externo para los outputs (o inputs, en su caso).

Una de las aplicaciones en este sentido es analizar la influencia que las actuaciones arbitrales han podido tener sobre los resultados de los partidos, y por tanto en el total de la liga.

En una publicación periódica, con soporte electrónico, ha dedicado una sección: <http://www.as.com/lectorarbitra/> incluye la posibilidad de decantarse y votar por el resultado de una determinada acción que puede suscitar diferentes interpretaciones. Con estos resultados, que carecen de base científica para considerarse objetivos, puede reconstruirse la serie de outputs de los diferentes equipos. Con esta serie "externa" se calcula el score de eficiencia (ofensiva, por ejemplo) S^o con los inputs originales y se compara con el intervalo de eficiencia:

- a) Si el score está incluido en el intervalo, i. e. $S^o \in [E^P, E^M]$, entendemos que el equipo no ha sido influido por factores externos, y su eficiencia es aceptable.
- b) Si el score es mayor que la cota superior del intervalo, es decir $S^o > E^M$, interpretamos que ha obtenido mejores resultados que los realmente conseguidos, por tanto, se trataría de un equipo que ha sido "beneficiado" por decisiones externas.
- c) Si el score es menor que la cota inferior del intervalo, es decir $S^o < E^P$, interpretamos que el equipo ha sido "perjudicado" por decisiones externas.

Sin embargo, no hemos hecho referencia a los resultados computacionales en este sentido porque, a nuestro juicio, esto no contribuiría más que a suscitar polémicas innecesarias.

Por último, en cuanto a la *estabilidad* de los resultados de los modelos DEA, el uso de datos inciertos, que sustituyen un score único por un intervalo, nos informa acerca de la y seguridad a los resultados obtenidos, ya que un pequeño cambio en los datos puede afectar de forma importante al resultado global.

8 Bibliografía

- [1] Banker, R. D., Charnes, A. y Cooper, W.W. (1984), Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, Management Science, 30, 1078-1092.
- [2] Bonilla, M; Casasus, T., Medal, A. y Sala, R (2002), Traffic in Spanish ports: an efficiency analysis, International Journal of Transport Economics, 29, 215 - 230.
- [3] Boscá, J. E.; Liern, V., Martínez, A. y Sala, R (2002), Un modelo DEA para datos imprecisos: Tolerancias y análisis de incertidumbre. Actas X Jornadas de ASEPUA, Madrid.

- [4] Briec, W. (1997), A Graph-Type Extension of Farrell Technical Efficiency Measure, *Journal of Productivity Analysis*, 8, 95-110.
- [5] Carmichael, F. y Thomas, D. (1995), Production and efficiency in team sports: an investigation of rugby league football, *Applied Economics*, 27, 859-869.
- [6] Carmichael, F., Thomas, D. y Ward, R. (2000), Team Performance: The Case of English Premiership Football, *Managerial and Decision Economics*, 21, 31 - 45.
- [7] Charnes, A., Cooper, W.W. y Rhodes, E. (1978), Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research*, 2, 429-444.
- [8] Charnes, A., Cooper, W.W. y Rhodes, E. (1979), Short Communication: Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research*, 3, 339.
- [9] Charnes, A., Cooper, W.W. , Lewin, A.Y. Morey, R.C. y Rousseau, J. (1985): Sensitivity and stability in DEA . *Annals of Operations Research*. 2, 139-156.
- [10] Charnes, A., Cooper, W.W. y Rhodes, E. (1990), Data Envelopment Analysis (en H. E. Breda ed.). *Operational Research'90*. Pergamon Press, Oxford.
- [11] Coelli y T. (1999): An introduction to efficiency and productivity analysis. Kluwer. Boston.
- [12] Cooper, W.W., Seiford L.M. y Tone, K. (2000): Data Envelopment Analysis. Kluwer. Boston.
- [13] Dawson, P., Dobson, S. y Gerrard, B. (2000), Stochastic Frontiers and the Temporal Structure of Managerial Efficiency in English Soccer, *Journal of Sports Economics*, vol. 1, n^o 4, 341-362.
- [14] De la Fuente, Ángel (1998), Fronteras de producción eficiente y convergencia regional: un comentario al trabajo de Maudos, Pastor y Serrano, *Revista Española de Economía*, 15, 265-271.
- [15] Färe, R. y Lovell, C.A.K. (1978). Measuring the Technical Efficiency of Production, *Journal of Economic Theory*, 19, 150-162.
- [16] Kao, C, Liu, S. T. (2000), Fuzzy Efficiency Measures in Data Envelopment Analysis, *Fuzzy Sets and Systems*, 119, 149-160.

- [17] León, T, Liern, V. Ruiz, J. L., y Sirvent, I. (2002), A Possibilistic Programming Approach to the Assessment of Efficiency with DEA Models, *Fuzzy Sets and Systems*, 139, 407-419.
- [18] Maudos, J., Pastor, J.M. y Serrano, L (1998), Convergencia en las regiones españolas: cambio técnico, eficiencia y productividad, *Revista Española de Economía*, 15, 235-264.
- [19] Martínez Cabrera, M. (2003): La medición de la eficiencia en las instituciones de educación superior. Fundación BBVA. Bilbao.
- [20] Olesen, O.B. y Petersen, N.C. (1996): A presentation of GAMS for DEA. *Computers and Operations Research*, 23-4, 323-339.
- [21] Pastor, J.T., Ruiz, J.L. y Sirvent, I (1999): A statistical test for detecting influential observations in DEA. *European Journal of Operational Research*, 115, 542-554.
- [22] Rottenberg, S. (1956), The baseball player's labor-market, *Journal of Political Economy*, 64, 242-258.
- [23] Schofield, J.A. (1988), Production functions in the sports industry: An empirical analysis of professional cricket, *Applied Economics*, 20, 177-193.
- [24] Scott, F.A., Long, J.E. y Somppi, K. (1985), Salary vs marginal revenue product under monopsony and competition: The case of professional basketball, *Atlantic Economic Journal*, 13, 50-59.
- [25] Scully, G W. (1974), Pay and Performance in Major League Baseball, *American Economic Review*, 64, 915-930.
- [26] Sengupta, J. K. (1992), A Fuzzy Systems Approach in Data Envelopment Analysis, *Computers and Mathematics with Applications*, 24, 259-266.
- [27] Sengupta, J.K. (1987). Data Envelopment Analysis for Efficiency Measurement in the Stochastic Case, *Computers and Operations Research*, 14, 117-129.
- [28] Sirvent, I. (2001), Algunos aspectos del problema de la evaluación de la eficiencia con modelos DEA, PhD. Dissertation, Universidad Miguel Hernández, Elche.
- [29] Sloane, P.J. (1971), The economics of professional football, the football club as a utility maximiser, *Scottish Journal of Political Economy*, 8, 121-146.
- [30] Sloane, P.J. ed. (1997), The economic of sport, *Economic Affairs*, 17, Special issue.

- [31] Slowinski, J.K. (1998), Fuzzy Sets in Decision Analysis, Operations Research and Statistics, Kluwer Academic Publishers.
- [32] Triantis, K and Girod, O (1998), A Mathematical Programming Approach for Measuring Technical Efficiency in a Fuzzy Environment, Journal of Productivity Analysis, 10, 85-102.
- [33] Zadeh, L.A. (1965), Fuzzy Sets, *Information and Control*, 8, 338-353.
- [34] Zak, T. A., Huang, C.J. y Siegfried, J.J. (1979), Production efficiency: the case of professional basketball, Journal of Business, 52, 379-392.