



Universidad de Valladolid

MÁSTER EN INVESTIGACIÓN EN MATEMÁTICAS
Curso 2016-2017

**Relajación Lagrangiana, métodos heurísticos
y metaheurísticos en algunos modelos de
Localización e Inventarios**

Autor: Adrián ESTEBAN PÉREZ

Tutor: Jesús SÁEZ AGUADO

Índice general

Resumen (Abstract)	9
Introducción	11
1. Preliminares	13
1.1. Relajación Lagrangiana y dualidad	13
1.1.1. Conceptos básicos	13
1.1.2. Propiedades del problema dual Lagrangiano	15
1.1.3. Subgradientes y el problema dual	17
1.2. Optimización Subgradiente	21
1.3. Modelos de inventario con revisión continua	22
1.3.1. El modelo EOQ	23
1.3.2. El modelo (r, Q)	25
1.4. Métodos heurísticos y metaheurísticos	28
1.4.1. Métodos heurísticos	29
1.4.2. Métodos metaheurísticos	31
2. Un modelo de Localización-Inventario	35
2.1. Introducción	35
2.2. Formulación del modelo. Formulación como problema no lineal entero (INLP)	36

2.3.	Análisis poliédrico	40
2.4.	Propiedades básicas del problema	42
2.5.	Formulación como problema no lineal entero mixto (MINLP)	43
2.6.	Reformulación como problema no lineal entero mixto (MINLP)	44
2.7.	Relajación Lagrangiana	45
2.7.1.	Descomposición espacial del problema	46
2.7.2.	Subproblemas de la relajación Lagrangiana	47
2.7.3.	Obtención de soluciones factibles: Heurística Lagrangiana	49
2.7.4.	Algoritmo de resolución	50
2.8.	Otras formulaciones equivalentes	52
2.8.1.	Formulación cónica	52
2.8.2.	Formulación set covering	52
2.9.	Resultados computacionales	53
2.10.	Conclusiones	56
3.	El problema de la p-mediana con restricciones de distancia máxima	59
3.1.	Introducción	59
3.2.	Formulación	60
3.2.1.	Formulación densa	61
3.2.2.	Formulación dispersa	62
3.3.	Relajación Lagrangiana	63
3.4.	GRASP	66
3.5.	Resultados computacionales	67
3.5.1.	Intervalos de factibilidad de distancia máxima	68
3.5.2.	Análisis de los resultados computacionales	72
3.6.	Conclusiones	75

4. Programación Entera bi-objetivo. Aplicaciones a modelos de Localización	77
4.1. Introducción	77
4.2. Conceptos básicos de Programación bi-objetivo	78
4.3. Problema bi-objetivo p-mediana-centro. El método ε -constraint . .	80
4.4. Resultados computacionales. Análisis bi-objetivo	85
4.5. Conclusiones	88
Conclusiones	89
Bibliografía.	91

AGRADECIMIENTOS

En primer lugar, quiero expresar mi más sincero agradecimiento a mi tutor, Jesús Sáez Aguado, por aceptar dirigirme este trabajo y por su paciencia y entrega a lo largo de todos estos meses. Estoy en deuda personal con él por transmitirme su visión personal y pasión por la Investigación Operativa, y mostrarme la belleza de la misma. Gracias por las amenas conversaciones, tanto personales como acerca de la Investigación Operativa. Gracias Jesús.

No quiero olvidarme de Juan García Laguna, que fue quien me inició en la Investigación Operativa y en Teoría de Inventarios, el capítulo 2 está dedicado a él. También agradecer al Departamento de Estadística e Investigación Operativa por cederme unas instalaciones para llevar a cabo los experimentos numéricos y también agradecer a FICO por proveer XPRESS.

Finalmente, dar las gracias a mis seres queridos por su apoyo constante e incondicional.

RESUMEN (ABSTRACT)

Resumen

En la actualidad, nos encontramos con numerosas situaciones en las que se requiere la mejor ubicación espacial de un cierto objeto que cumpla con una serie de necesidades, donde este proceso implica una toma de decisión. Ejemplos de ello aparecen en la cadena de suministro, donde el objetivo es optimizar tanto las decisiones de tipo estratégico (Localización) como las de tipo táctico (Inventario) o en problemas de localización de servicios tanto públicos como privados. En este trabajo, se formulan con detalle algunos ejemplos de los problemas anteriores, tales como un modelo integrado de Localización-Inventario, el problema de la p -mediana con restricciones de distancia máxima y un problema bi-objetivo. Finalmente, se resuelven los mismos a través de métodos como relajación Lagrangiana, métodos heurísticos o metaheurísticos tales como GRASP y se han obtenido resultados novedosos, proporcionando diferentes procedimientos de resolución que mejoran los existentes en la literatura.

Abstract

Nowadays, we find many situations in which the best spatial location of a certain object that meets a series of needs is required, where this process implies a decision making. Examples of this occur in the supply chain, where the objective is to optimize both strategic (Location) and tactical type (Inventory) decisions or in localization problems of both public and private services. In this project, some examples of previous problems are formulated in detail, such as an integrated Location-Inventory model, the p -median problem with maximum distance constraints and a bi-objective problem. Finally, they are solved through methods such as Lagrangian relaxation, heuristic or metaheuristic methods such as GRASP and new results have been obtained, providing different resolution procedures that improve those existing in the literature.

INTRODUCCIÓN

Hoy en día y más en época de crisis, nos encontramos con multitud de situaciones en las que un cierto objeto debe localizarse dentro de un sentido espacial. Evidentemente, siempre se requiere la mejor ubicación que cumpla con ciertas necesidades. Este proceso de elección implica algún tipo de toma de decisión.

Un ejemplo muy notable de ello, ocurre en la cadena de suministro. Por tanto, es necesario optimizar tanto la parte táctica como estratégica de la misma. La parte táctica de la cadena de suministro está asociada los inventarios y la parte estratégica a la localización. A pesar de que la Teoría de Inventarios es una de las ramas más antiguas de la Investigación Operativa, su estudio ha sido independiente de la Teoría de Localización. Esto implica que al optimizar por separado las decisiones de inventario de las de localización, se obtendrán frecuentemente soluciones sub-óptimas. No sólo en cadena de suministro es necesaria una optimización de la misma, sino también en servicios, sean tanto públicos (localización de servicios de emergencia, colegios, hospitales etc.) como privados (empresas de manufactura, establecimientos etc.). Según Daskin (2013), "*the success or failure of both private and public sector facilities depends in part on the locations chosen for those facilities*". Actualmente, la Teoría de Localización es un área de investigación bastante activa y son necesarias diversas técnicas de Investigación Operativa para resolver los problemas que aparecen en la misma.

El objetivo del trabajo es mostrar la versatilidad y potencia de la relajación Lagrangiana y los métodos heurísticos y metaheurísticos en algunos modelos de Optimización, como pueden ser tanto modelos de Localización como de Inventarios. Este trabajo consta de cuatro capítulos con tres partes claramente diferenciadas:

- La primera parte, está dedicada a presentar diferentes conceptos y aspectos de Investigación Operativa necesarios para los siguientes capítulos. Este capítulo de preliminares consta de lo siguiente. En primer lugar, se presentan diversos aspectos de la teoría de dualidad así como de la relajación Lagrangiana, un tema ya clásico en Programación Matemática. Después se presenta la optimización subgradiente que es un método bastante conocido de Optimización no diferenciable. A continuación, se señalan diversas nociones de Teoría de Inventarios y finalmente unas breves notas acerca de métodos heurísticos y metaheurísticos.
- La segunda parte del trabajo, ya está centrada en un modelo de Optimización, y más concretamente, en el capítulo 2 se estudia un modelo integrado de Localización-Inventario. Se presentan diversos aspectos del mismo como formulaciones y métodos de resolución, además de resultados computacionales.
- La tercera parte del trabajo está formada por dos capítulos. Esta parte, a diferencia de la segunda, se centra más en modelos de Localización. Más concretamente, en el capítulo 3 se estudia en primer lugar el problema de la p -mediana con restricciones de distancia máxima, que presenta una aplicabilidad real en gestión de servicios. Además, se resuelve el mismo con diferentes procedimientos de resolución, tales como relajación Lagrangiana o la metaheurística GRASP. Finalmente, el capítulo 4 presenta diferentes nociones de Programación Multiobjetivo, dirigidas a la resolución de un problema de Localización bi-objetivo, en el que se tiene en cuenta la eficiencia y la equidad. Para su resolución se utilizará el método ε -constraint apoyándose en el GRASP y en los resultados del capítulo 3.

Al final, se incluyen las conclusiones del trabajo y una relación de las referencias bibliográficas utilizadas en el mismo.

Capítulo 1

Preliminares

Este capítulo tiene el objetivo de reunir una serie de resultados o conceptos que son necesarios para el estudio central de este trabajo, o bien que ayudan a tener una mejor visión del mismo. La primera sección está dedicada al estudio de la *relajación Lagrangiana y dualidad*, un tema ya clásico en Optimización. La segunda sección está dedicada a la *optimización subgradiente*, un procedimiento de optimización no diferenciable bastante usado en la práctica. A continuación, se incluyen varios aspectos de *Teoría de Inventarios* relacionados con modelos de inventario de revisión continua. Finalmente, incluimos una sección dedicada a los fundamentos básicos de los *métodos heurísticos y metaheurísticos*, centrándonos especialmente en los usados en este trabajo.

1.1. Relajación Lagrangiana y dualidad

1.1.1. Conceptos básicos

Consideremos el problema (P)

$$(P) \quad \begin{array}{ll} \text{Minimizar} & f(x) \\ \text{s.a.} & g_i(x) \leq 0 \quad i = 1, \dots, m \\ & h_j(x) = 0 \quad j = 1, \dots, p \\ & x \in X \subseteq \mathbb{R}^n \end{array}$$

donde X es un conjunto no vacío y f es una función real. Además, diremos que (P) es el *problema primal*.

Sean las funciones $g(x) := [g_1(x), \dots, g_m(x)]^T$, $h(x) := [h_1(x), \dots, h_p(x)]^T$ y consideremos el conjunto $F := \{x \in X \mid g(x) \leq 0, h(x) = 0\}$. F es la *región factible* del problema primal (P) y cualquier punto de la misma se denomina *solución factible*.

Las restricciones del problema primal (P) pueden ser de varios tipos:

- El grupo de restricciones $g_i(x) \leq 0$, $i = 1, \dots, m$ se denominan *restricciones explícitas de desigualdad*, mientras que el grupo de restricciones $h_j(x) = 0$, $j = 1, \dots, p$ se denominan *restricciones explícitas de igualdad*.
- $x \in X$ se denominan *restricciones implícitas*.

En la formulación del problema primal, se ha separado el conjunto de restricciones en explícitas e implícitas, para enfatizar que las restricciones *difíciles* son las explícitas y que el problema primal sería más fácil de resolver sin las mismas. Esta partición del conjunto de restricciones es lo que justifica la siguiente

Definición 1.1.

Dado el problema primal (P),

- Se define el Lagrangiano como la función

$$L(x, \mu, \lambda) := f(x) + \sum_{i=1}^m \mu_i g_i(x) + \sum_{j=1}^p \lambda_j h_j(x)$$

donde $\mu_i \geq 0$, $i = 1, \dots, m$ y $\lambda_1, \dots, \lambda_p \in \mathbb{R}$.

Además, $\mu := [\mu_1, \dots, \mu_m]$, $\lambda := [\lambda_1, \dots, \lambda_p]$ se denominan variables duales o multiplicadores de Lagrange.

- Se define la función Lagrangiana como $\phi(\mu, \lambda) := \inf_{x \in X} L(x, \mu, \lambda)$.

Definición 1.2 (Condiciones globales de optimalidad).

Se dice que $(\bar{x}, \bar{\mu}, \bar{\lambda})$, con $\bar{x} \in X$, $\bar{\mu} \in \mathbb{R}_+^m$ y $\bar{\lambda} \in \mathbb{R}^p$, satisfacen las condiciones globales de optimalidad para el problema primal (P) si se cumplen las siguientes condiciones:

1. $f(\bar{x}) + \bar{\mu}g(\bar{x}) + \bar{\lambda}h(\bar{x}) = \inf_{x \in X} \{f(x) + \bar{\mu}g(x) + \bar{\lambda}h(x)\}$
2. $\bar{\mu}g(\bar{x}) = 0$
3. $g(\bar{x}) \leq 0$ y $h(\bar{x}) = 0$.

A continuación, enunciamos unas condiciones suficientes de optimalidad para el problema primal (P). Nótese que no son condiciones necesarias¹.

Teorema 1.3.

Si $(\bar{x}, \bar{\mu}, \bar{\lambda})$ satisfacen las condiciones globales de optimalidad, entonces \bar{x} es óptimo para el problema primal.

1.1.2. Propiedades del problema dual Lagrangiano

Teorema 1.4 (Dualidad débil).

Para cada $(\mu, \lambda) \in \mathbb{R}_+^m \times \mathbb{R}^p$ se tiene $\phi(\mu, \lambda) \leq p^*$, donde $p^* = \inf_{x \in F} f(x)$.

El problema dual del problema primal (P), es el problema (D) siguiente

$$(D) \quad \begin{array}{ll} \text{Maximizar} & \phi(\mu, \lambda) \\ & \mu, \lambda \\ \text{s.a.} & \mu \geq 0 \end{array}$$

Sea $d^* := \sup_{\mu \geq 0, \lambda} \phi(\mu, \lambda)$.

¹Se pueden establecer condiciones necesarias, asumiendo convexidad y algunas condiciones de regularidad. Si $X = \mathbb{R}^n$ y se asumen hipótesis de cierta diferenciabilidad se obtienen las condiciones de Karush-Kuhn-Tucker (KKT).

Llamaremos *solución factible dual* a todo (μ, λ) tal que $\mu \geq \mathbf{0}$ y $\phi(\mu, \lambda) > -\infty$.

Además, fijado x , dado que la función $L(x, \mu, \lambda)$ es lineal con respecto a (μ, λ) (luego convexa y cóncava), se deduce que $\phi(\mu, \lambda)$ es *cóncava*. El problema dual (D) es un problema de maximizar una función objetivo cóncava sobre una región factible convexa. Además, en virtud del teorema de dualidad débil, se tiene que $d^* \leq p^*$. Si $d^* < p^*$, se dice que hay un *salto de dualidad* y la magnitud del mismo es igual a $p^* - d^*$.

Nótese que en este trabajo estudiamos problemas de Programación Entera, donde se tiene que X es finito y donde existe habitualmente (salvo casos muy especiales) salto de dualidad mayor que cero. Además, el hecho de que X sea finito implica que $\phi(\mu, \lambda)$ es una función *cóncava, lineal a trozos y no diferenciable*. A continuación señalaremos brevemente unas observaciones acerca de la infactibilidad y no acotación del problema primal y su relación con el problema dual:

- Si el problema primal es no acotado, es decir, $p^* = -\infty$, en virtud del teorema de dualidad débil se tiene $\phi(\mu, \lambda) = -\infty$, para todo $\mu \in \mathbb{R}_+^m$ y todo $\lambda \in \mathbb{R}^p$, luego el problema dual es no factible.
- Supongamos que el problema primal es no factible, es decir, el conjunto $F = \emptyset$. Entonces, tomaremos como convenio que $p^* = \infty$.
- La función Lagrangiana $\phi(\mu, \lambda)$, satisface que $\phi(\mu, \lambda) < \infty$, $\forall \lambda$, pero es posible que $d^* = \infty$, en cuyo caso el problema dual es no acotado, y en consecuencia, se deduce que es no factible. También puede suceder que $d^* < \infty$ o incluso que $d^* = -\infty$.

Dado un problema primal, se pueden obtener varios problemas duales, dependiendo qué restricciones se seleccionan como restricciones explícitas y cuáles como implícitas. En la práctica, dependiendo de la naturaleza del problema, se debe realizar una adecuada selección del conjunto X . A continuación presentamos dos teoremas que proporcionan unas condiciones bajo las cuales el salto de dualidad es cero. Más concretamente:

Teorema 1.5.

Si $(\bar{x}, \bar{\mu}, \bar{\lambda})$ satisfacen las condiciones globales de optimalidad, entonces $(\bar{\mu}, \bar{\lambda})$ es óptimo para el problema dual. Además, $d^* = p^*$.

Teorema 1.6 (del punto de silla).

$(\bar{x}, \bar{\mu}, \bar{\lambda})$ es óptimo si, y sólo si $\bar{x} \in X$, $\bar{\mu} \in \mathbb{R}_+^m$ y

$$L(\bar{x}, \mu, \lambda) \leq L(\bar{x}, \bar{\mu}, \bar{\lambda}) \leq L(x, \bar{\mu}, \bar{\lambda}), \forall x \in X, \mu \in \mathbb{R}_+^m, \lambda \in \mathbb{R}^p.$$

Es posible dar un *teorema de dualidad fuerte*. En la literatura, es frecuente encontrarlo bajo diferentes versiones, dependiendo de las hipótesis, aunque todos suelen tener en común hipótesis de convexidad. En la práctica, y salvo muy raras ocasiones y en modelos muy concretos, no tendremos hipótesis de convexidad y mucho menos salto de dualidad cero. Recordemos, que en este trabajo, nos centraremos en problemas de Programación Entera, donde en general, existirá salto de dualidad mayor que cero.

Si el punto de silla del Lagrangiano no existe, un procedimiento para atacar el problema es la convexificación del mismo. Puede probarse, (ver Shapiro (1979)) que *resolver el problema dual y resolver el problema convexificado son equivalentes*. La dificultad del problema convexificado está en conocer una expresión explícita de la envolvente convexa, si bien, hay algunos casos donde existen métodos que permiten resolver el problema convexificado, como los llamados *cortes Fenchel* en Programación Entera. Nosotros, en principio, nos centraremos en resolver el problema dual.

1.1.3. Subgradientes y el problema dual

En Análisis Convexo, es conocido el hecho de que una función es convexa si, y sólo si su epigrafo es un conjunto convexo. Por tanto, su estudio se reduce al de los conjuntos convexos. Además, los teoremas de separación (vía hiperplanos separadores) aplicados al epigrafo proporcionan un estudio geométrico de las funciones convexas. De este modo, se llega a la noción de hiperplano soporte. Veremos que estos conceptos nos servirán para definir los subgradientes.

Sea $x_o \in \mathbb{R}^n$. Se puede demostrar que una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ que presente un hiperplano soporte en el punto x_o equivale a decir que existe $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ funcional lineal y continuo tal que $f(x) \geq f(x_o) + \psi(x - x_o)$, $\forall x \in \mathbb{R}^n$. El conjunto de funcionales lineales y continuos que verifican lo anterior, se denomina subdiferencial de f en el punto x_o , y se representa por $\partial f(x_o)$. Al ser \mathbb{R}^n espacio de Hilbert, por el teorema de representación de Riesz, ψ se puede representar

unívocamente como $\psi(x) = \langle x, \xi \rangle$ para algún $\xi \in \mathbb{R}^n$, luego

$$\begin{aligned} \partial f(x_o) &= \left\{ \xi \in \mathbb{R}^n \mid f(x) \geq f(x_o) + \langle \xi, x - x_o \rangle, \forall x \in \mathbb{R}^n \right\} \\ &= \left\{ \xi \in \mathbb{R}^n \mid f(x) \geq f(x_o) + (x - x_o)\xi, \forall x \in \mathbb{R}^n \right\} \end{aligned} \quad (1.1)$$

En este caso, diremos que ξ es un *subgradiente de f en x_o* .

A continuación, señalaremos unas propiedades:

1. $\partial f(x_o)$ es cerrado y convexo (puede ser vacío si no se añaden hipótesis). Además, si $X \subseteq \mathbb{R}^n$ es acotado, entonces el conjunto $\bigcup_{x \in X} \partial f(x)$ es acotado. Se pueden dar condiciones suficientes de existencia de subgradientes, vía la *Máx-fórmula* (ver Borwein y Lewis (2000, p.36)).

2. Si f_1, \dots, f_m son funciones convexas, $\lambda_1, \dots, \lambda_m \geq 0$, entonces

$$\partial (\lambda_1 f_1(x) + \dots + \lambda_m f_m(x)) = \lambda_1 \partial (f_1(x)) + \dots + \lambda_m \partial (f_m(x))$$

3. Si f es una función cóncava, entonces diremos que ξ es un subgradiente de f en x_o si $-\xi$ es un subgradiente de la función $-f$ en x_o . Más concretamente, si f es una función cóncava definida en \mathbb{R}^n , entonces ξ es un subgradiente de f en x_o si $f(x) \leq f(x_o) + (x - x_o)\xi, \forall x \in \mathbb{R}^n$.
4. Una función f es diferenciable en el punto x_o con gradiente $\nabla f(x_o)$, si y sólo si $\nabla f(x_o)$ es el único subgradiente de f en el punto x_o .

Definición 1.7.

- Sean $X \subseteq \mathbb{R}^n$ y $x_o \in X$. Se dice que un vector $v \in \mathbb{R}^n$ es una *dirección factible de X en x_o* si existe $\bar{\alpha} > 0$, con $x_o + \alpha v \in X$, para todo $\alpha \in [0, \bar{\alpha}]$. Además, si X es convexo, las direcciones factibles de X en x_o son vectores de la forma $\alpha(v - x_o)$, con $\alpha > 0$ y $v \in X$.
- El conjunto de todas las direcciones factibles de X en x_o es un cono que se denota por $F_X(x_o)$.

Definición 1.8.

- Sean $X \subseteq \mathbb{R}^n$ y $x_o \in X$. Se dice que un vector $y \in \mathbb{R}^n$ es una tangente de X en x_o si o bien $y = 0$, o bien existe una sucesión $(x_n)_n \subset X$, tal que $x_n \neq x_o, \forall n$ y $x_n \rightarrow x_o$, $\frac{x_n - x_o}{\|x_n - x_o\|} \rightarrow \frac{y}{\|y\|}$.
- El conjunto de todas las tangentes de X en x_o es un cono llamado cono tangente de X en x_o que se denota por $T_X(x_o)$.

Proposición 1.9.

Sean $X \subseteq \mathbb{R}^n$ no vacío y $x_o \in X$. Entonces:

- a) $T_X(x_o)$ es un cono cerrado.
- b) $\text{cl}(F_X(x_o)) \subset T_X(x_o)$.
- c) Si X es convexo, entonces $F_X(x_o)$ y $T_X(x_o)$ son ambos convexos y se tiene $\text{cl}(F_X(x_o)) = T_X(x_o)$.

Veamos a continuación el concepto de cono normal.

Definición 1.10.

Sean $X \subseteq \mathbb{R}^n$ y $x_o \in X$. Se define el cono normal de X en x_o como $N_X(x_o) = \text{cl}((T_X(x_o))^*)$, donde $(T_X(x_o))^*$ denota el cono polar de $T_X(x_o)$.

Señalemos a continuación unas propiedades:

- Si $x_o \notin X$, tomaremos como convenio que $N_X(x_o) = \emptyset$.
- Si $N_X(x_o) = (T_X(x_o))^*$ diremos que X es regular en x_o .
- Si X es regular en x_o , entonces $T_X(x_o)$ es un cono convexo.
- Si x es un punto interior a X , entonces $T_X(x_o) = \mathbb{R}^n$ y $N_X(x_o) = \{0\}$.

Proposición 1.11.

Sea $X \subseteq \mathbb{R}^n$ convexo. Entonces, para cada $x_o \in X$ se tiene que $\xi \in (T_X(x_o))^*$ si, y sólo si $(x - x_o)\xi \leq 0, \forall x \in X$. Además, X es regular en todo $x \in X$.

A continuación, se enuncia un teorema importante de cara a la Optimización y al fundamento teórico de la Optimización Subgradiente.

Teorema 1.12.

Sea X^* el conjunto de soluciones óptimas. Entonces, el punto $x \in X^*$ si, y sólo si $0 \in \partial f(x) + N_X(x)$ o equivalentemente, si $-\partial f(x) \cap N_X(x) \neq \emptyset$.

Concluamos esta sección con una serie de propiedades y comentarios relativos al problema dual y su relación con los subgradientes.

Lema 1.13.

Sea \bar{x} una solución óptima para la función Lagrangiana en $(\bar{\mu}, \bar{\lambda})$, es decir, $\phi(\bar{\mu}, \bar{\lambda}) = f(\bar{x}) + \bar{\mu}g(\bar{x}) + \bar{\lambda}h(\bar{x}) = \inf_{x \in X} \{f(x) + \bar{\mu}g(x) + \bar{\lambda}h(x)\}$. Entonces, $\begin{pmatrix} g(\bar{x}) \\ h(\bar{x}) \end{pmatrix}$ es un subgradiente de $\phi(\mu, \lambda)$ en el punto $(\bar{\mu}, \bar{\lambda})$.

Para $(\bar{\mu}, \bar{\lambda})$, con $\bar{\mu} \geq 0$, sea

$$\mathcal{X}(\bar{\mu}, \bar{\lambda}) = \{x \in X \mid \phi(\bar{\mu}, \bar{\lambda}) = f(x) + \bar{\mu}g(x) + \bar{\lambda}h(x)\}$$

el conjunto de soluciones óptimas para la función Lagrangiana en $(\bar{\mu}, \bar{\lambda})$. Además, sea $\Psi(x) = (g(x), h(x))$ y consideremos

$$\Psi[\mathcal{X}(\bar{\mu}, \bar{\lambda})] = \{\gamma \in \mathbb{R}^{m+p} \mid \gamma = \begin{pmatrix} g(\bar{x}) \\ h(\bar{x}) \end{pmatrix}, \forall x \in \mathcal{X}(\bar{\mu}, \bar{\lambda})\}$$

Por el Lema 1.13, se deduce que $\Psi[\mathcal{X}(\bar{\mu}, \bar{\lambda})] \subseteq \partial\phi(\bar{\mu}, \bar{\lambda})$. Además, puede probarse (ver Grinold (1970)) que

$$\text{conv}[\Psi[\mathcal{X}(\bar{\mu}, \bar{\lambda})]] = \partial\phi(\bar{\mu}, \bar{\lambda})$$

1.2. Optimización Subgradiente

Consideremos el problema de optimización no diferenciable (*NDP*)

$$\begin{array}{ll} \text{Maximizar} & f(x) \\ \text{(NDP)} & \\ \text{s.a.} & x \in X \end{array}$$

donde $X \subseteq \mathbb{R}^n$ es un conjunto convexo no vacío y f es una función real cóncava que puede ser no diferenciable. Los métodos gradiente aquí no sirven, y es por ello que se necesitan métodos de optimización no diferenciable. Uno de ellos son los métodos de *Optimización Subgradiente*.

Bajo el nombre de *Optimización Subgradiente* se engloba a una familia de algoritmos de ascenso aproximados, y en esencia, son una simplificación heurística del algoritmo primal-dual generalizado.

Su uso es bastante extendido pues son utilizados como método de optimización no diferenciable para resolver el problema dual. El método de Optimización Subgradiente es como sigue. *Dado un iterante x^k , con $k \geq 1$, se elige $\xi^k \in \partial f(x^k)$ y se calcula el iterante x^{k+1} mediante*

$$x^{k+1} = P_X \left(x^k + T_k \frac{\xi^k}{\|\xi^k\|} \right)$$

donde

- T_k es la longitud o tamaño de paso en la iteración k -ésima (> 0).
- P_X es la proyección sobre X .

En la mayoría de los casos prácticos de interés, X suele ser todo el espacio \mathbb{R}^n o bien un subconjunto del mismo cuya estructura permite fácilmente el cálculo de la proyección, como puede ser el ortante no negativo \mathbb{R}_+^n .

En la literatura, es frecuente encontrarse con diversas propuestas para la elección de la sucesión de los tamaños de paso $\{T_k\}$. Cada elección da lugar a un algoritmo de optimización subgradiente. A continuación vamos a exponer los más comunes junto con sus propiedades de convergencia:

a) $T_k \rightarrow 0$ y $\sum_{k=1}^{\infty} T_k = \infty$.

- Se tiene que $\limsup_k f(x^k) = f^*$ y, si X^* , el conjunto de soluciones óptimas, es compacto, entonces se tiene

$$d(x^k, x^*) = \min_{x^* \in X^*} \|x - x^*\| \rightarrow 0.$$

- Si $\sum_{k=1}^{\infty} T_k^2 < \infty$, entonces $x^k \rightarrow x^* \in X^*$.

b) $T_k = C\rho^k$, con $0 < \rho < 1$ y $C > 0$.

Se tiene que $x^k \rightarrow x^* \in X^*$ si se toman ρ y C *suficientemente grandes*.

c) $T_k = \theta_{sub} \frac{f^* - f(x^k)}{\|\xi^k\|}$, con $0 < \theta_{sub} < 2$ y siendo f^* el valor óptimo del problema (NDP) (Regla de POLYAK).

Se cumple que $x^k \rightarrow x^* \in X^*$ converge sin imponer hipótesis adicionales. En la práctica, no se suele conocer f^* , y un tratamiento heurístico muy usado es el siguiente (conocido en la literatura como *heurística de HELD, WOLFE y CROWDER* (Held *et al.* (1974))). Se sustituye f^* por una cota superior del valor óptimo de la función objetivo, UB . Además, se suelen fijar un número de iteraciones y unas tolerancias, de modo que si en dichas iteraciones el valor de la función objetivo no mejora más de una tolerancia prefijada, se divide θ_{sub} entre 2. Éste es el método que seguiremos.

Si bien en la literatura existen otros procedimientos como el método *bundle* (Hiriart-Urruty y Lemaréchal (1993)), en este trabajo hemos optado por usar optimización subgradiente, ya que en Programación Entera es quizá el método más usado y extendido para resolver el problema dual (Frangioni *et al.* (2017)).

1.3. Modelos de inventario con revisión continua

Esta sección está dedicada a presentar los fundamentos básicos de algunos de los modelos de inventario más importantes de *revisión continua*: el *modelo EOQ* y el *modelo (r, Q)*. Los conceptos básicos de Teoría de Inventarios pueden consultarse en Esteban-Pérez (2016).

1.3.1. El modelo EOQ

El modelo más sencillo de inventario y uno de los más antiguos es el modelo EOQ (*Economic Order Quantity*), el cual fue presentado por Ford W. HARRIS en 1913, aunque también es conocido como *modelo de Wilson*, ya que fue popularizado por R.H. WILSON. A pesar de que impone hipótesis muy restrictivas (por ejemplo, que la demanda por unidad de tiempo y el coste de almacenamiento por artículo y por unidad de tiempo son constantes), es muy popular y es la base de los modelos de inventario de revisión continua.

Este modelo EOQ es determinista, es decir, supondremos que los parámetros del mismo se conocen con certeza y además, tomaremos como unidad de tiempo el año (aunque es válida cualquier otra unidad de tiempo: día, semana, mes, etc.). Contruiremos el modelo partiendo de las siguientes hipótesis iniciales. Se supone que el inventario es de revisión continua y el horizonte de planificación es infinito y continuo, considerando un sólo artículo o producto para el cual no se permite escasez. La reposición se produce cuando el stock disponible se agota, y además dicha reposición es instantánea. Además, supondremos que no hay descuentos en el precio por volumen de compra, la demanda es continua y constante a lo largo del tiempo (representaremos la misma por D unidades al año) y la cantidad de pedido es constante para cada orden (la representaremos por Q). Los siguientes parámetros del modelo son constantes a lo largo del horizonte de planificación. Son los siguientes:

- El coste fijo por lanzar un pedido de cualquier tamaño lo representaremos por K .
- El coste unitario de compra lo representaremos por p .
- El coste unitario de mantenimiento por año lo representaremos por h .

El objetivo es determinar la cantidad óptima de pedido Q^* (es decir, *cuánto pedir*) y el instante de tiempo en el cual debe hacerse el pedido (es decir, *cuándo pedir*) de modo que se *minimice el coste total de la gestión del inventario*. Dado que la reposición es instantánea, se sigue que el pedido debe hacerse cuando el nivel del inventario se reduce a 0. En consecuencia, el objetivo se reduce a *determinar la cantidad económica de pedido Q^** . Ésta es la razón por la cual al modelo EOQ se le conoce también como *modelo de la cantidad económica de pedido*. Además, se denomina **ciclo de inventario** al *intervalo de tiempo, que comienza con la llegada de un pedido y finaliza en el instante previo de recepción del siguiente pedido*. Representaremos el mismo por T y es igual a Q/D años.

Dado que el objetivo es minimizar la función que expresa el coste total anual, debemos calcular los costes siguientes

1. El coste anual de hacer pedidos es igual a al coste fijo de pedido por el número de pedidos anuales:

$$K \cdot \frac{D}{Q}$$

2. El coste anual de compra es igual al precio unitario de compra por el número de unidades compradas al año:

$$p \cdot D$$

3. El coste anual de mantenimiento es el coste de mantenimiento de un ciclo multiplicado por el número anual de ciclos:

- Dado que el nivel del inventario es variable en el tiempo, consideraremos el nivel de existencias promedio, el cual es el mismo en todos los ciclos: $Q/2$. Por tanto, el coste de mantenimiento en un ciclo es $h \frac{Q}{2} \frac{Q}{D}$.
- El número anual de ciclos es D/Q .

Por tanto, el coste anual de mantenimiento es $h \frac{Q}{2} \frac{Q}{D} \frac{D}{Q} = h \frac{Q}{2}$

La función objetivo es por tanto

$$C(Q) = K \frac{D}{Q} + h \frac{Q}{2} + pD$$

Puede comprobarse fácilmente que la función $C(Q)$ es convexa para todo $Q > 0$ y, por tanto, el mínimo global se puede obtener resolviendo $C'(Q) = 0$. Se obtiene así, la llamada *fórmula de Wilson*:

$$Q^* = \sqrt{\frac{2KD}{h}}$$

A continuación señalaremos unas observaciones:

1. En el punto Q^* el coste anual de pedido y el de mantenimiento coincide.
2. Puede probarse (ver Fernández Suárez *et al.* (1998)) que la gráfica de la función $C(Q)$ tiene pendiente más suave para valores de Q mayores que Q^* que para Q menores de Q^* .

Además, se tiene:

- El número óptimo de pedidos es $N^* = D/Q^* = \sqrt{\frac{Dh}{2K}}$.
- La longitud del ciclo óptimo de inventario es $T^* = 1/N^* = \sqrt{\frac{2K}{Dh}}$.
- El coste total óptimo es $C(Q^*) = \sqrt{2KDh} + pD = hQ^* + pD$.
- Si D' , K' y h' son los valores estimados de los valores reales de D , K y h respectivamente, y las variaciones en el cambio se expresan por medio de las relaciones $D' = \alpha D$, $K' = \beta K$ y $h' = \gamma h$ se tiene que la razón entre el EOQ estimado $(Q' = \sqrt{\frac{2K'D'}{h'}})$ y el EOQ óptimo $(Q^* = \sqrt{\frac{2KD}{h}})$ es

$$\frac{Q'}{Q^*} = \sqrt{\frac{\alpha\beta}{\gamma}}$$

Esto permite expresar el efecto sobre Q^* de los errores en la estimación de los parámetros.

1.3.2. El modelo (r, Q)

Una característica de los modelos de revisión continua es que en cualquier instante de tiempo puede realizarse un pedido. Uno de los modelos básicos de revisión continua es el denominado modelo (r, Q) o modelo *reorder-point/order-quantity*: cuando el nivel del inventario alcanza r unidades, se piden Q unidades. En este caso, r se denomina punto de reabastecimiento y Q cantidad de pedido. Una manera muy simple de realizar lo anterior es la siguiente. Se disponen en el almacén de dos apartados o zonas; la zona 1 debe disponer de una capacidad de r unidades. La demanda se sirve de forma prioritaria de la zona 2 y, al agotarse las existencias de la misma se coloca un pedido de Q unidades y hasta su recepción se sirve la demanda de la zona 1 y el resto del pedido se almacena en la zona 2. Así, cada vez que la zona 2 se vacía sabemos que se ha alcanzado el punto de reabastecimiento. Si las existencias se almacenan en una sola pila, una forma fácil de realizar el método anterior es disponer de una especie de marca o señal en la unidad que hace el número r .

Admitiremos las siguientes hipótesis:

1. El inventario es de un único producto o artículo.

2. La demanda (anual) puede expresarse por medio de una v.a. continua \mathcal{D} .
3. Se tiene revisión continua y se usa una política (r, Q) .
4. El plazo de reposición o tiempo de entrega puede expresarse por medio de una v.a. \mathcal{L} .
5. Existe demanda acumulable, y la demanda no servida debe satisfacerse con la recepción del pedido en curso.
6. No existen descuentos en el precio por volumen de compra.
7. El horizonte de planificación es continuo.
8. La demanda durante el plazo de reposición puede expresarse por medio de una v.a. continua \mathcal{X} . Entonces, se tiene que

$$E(\mathcal{X}) = E(\mathcal{L})E(\mathcal{D})$$

$$\sigma_{\mathcal{X}} = \sqrt{E(\mathcal{L})\text{Var}(\mathcal{D}) + \text{Var}(\mathcal{L})(E(\mathcal{D}))^2}$$

En algunos modelos de inventario, con el fin de evitar roturas del stock (provocadas por el carácter estocástico de la demanda y los plazos de reposición) se suele considerar una cantidad extra de inventario añadida al punto de reabastecimiento original, el denominado **stock de seguridad** (*safety stock*) cuya misión es amortiguar los efectos de las fluctuaciones de la demanda durante el plazo de reposición. En algunos casos, lo anterior equivale a descomponer el inventario en dos clases: las *existencias de ciclo* (que corresponderían a la mercancía en movimiento si el almacén se comportase acorde al modelo EOQ) y las *existencias de seguridad* (corresponderían a la diferencia entre el punto de reabastecimiento y la demanda media durante el plazo de reposición, y serían mercancía sin movimiento si el almacén se comportase acorde al modelo EOQ).

Es evidente que si se eleva el punto de reabastecimiento aumenta el stock de seguridad y, en consecuencia, su holding cost, aunque se gana en un mejor servicio al cliente. En la práctica el problema está en hallar un punto de equilibrio entre el servicio al cliente y los costes.

Dar servicio al cliente con la certeza absoluta de que nunca habrá escasez, es una actitud muy poco realista, ya que esto puede conducir a puntos de reabastecimientos muy altos y, por tanto, lleva a un incremento en el holding cost. El nivel de servicio, puede verse como una medida del éxito de una política de inventario para satisfacer la demanda.

Al lanzar un pedido cuando quedan r unidades en el inventario, se tiene que $P(\mathcal{X} \leq r)$ es la probabilidad de satisfacer a tiempo toda la demanda originada durante el plazo de reposición de un ciclo de inventario cualquiera. Por consiguiente, si consideramos un intervalo de tiempo suficientemente grande, $\alpha = P(\mathcal{X} \leq r)$ puede verse como la fracción (esperada) que obtendríamos si dividiéramos el número de ciclos servidos sin roturas entre el número total de ciclos servidos en ese intervalo de tiempo.

Dado el punto de reabastecimiento r , denominaremos **nivel de servicio** α (asociado a r) (o nivel de servicio de tipo I) al valor $P(\mathcal{X} \leq r)$. Evidentemente, se tiene que el nivel de servicio α pertenece al intervalo cerrado $[0, 1]$ y, además, fijada la v.a. \mathcal{X} , α es una función no decreciente de r y es independiente de Q .

En la práctica interesa determinar el valor de r para que el nivel de servicio asociado al mismo sea una cantidad α prefijada. Teóricamente, basta despejar r en la ecuación $P(\mathcal{X} \leq r) = \alpha$. Sea \mathcal{F} la función de distribución asociada a la v.a. estandarizada $\frac{(\mathcal{X} - E(\mathcal{X}))}{\sigma_{\mathcal{X}}}$. Entonces:

- Si F es estrictamente creciente, entonces se tiene que $r = E(\mathcal{X}) + \sigma_{\mathcal{X}}\mathcal{F}^{-1}(\alpha)$
- \mathcal{F} es siempre no decreciente, pero si no es estrictamente creciente, puede suceder que haya más de una forma de despejar r en la ecuación $P(\mathcal{X} \leq r) = \alpha$. Entonces, consideraremos como solución $r = \inf\{r \mid \mathcal{F}(r) \geq \alpha\}$ si \mathcal{X} es una v.a. continua, y $r = \min\{r \mid \mathcal{F}(r) \geq \alpha\}$ si \mathcal{X} es una v.a. discreta.

Además, las unidades de producto adicionales que en el caso estocástico son utilizadas para alcanzar un nivel de inventario en el instante de pedir que nos asegure un nivel de servicio α , es decir, al valor

$$r - E(\mathcal{X}) = \sigma_{\mathcal{X}}\mathcal{F}^{-1}(\alpha) \quad (1.2)$$

es precisamente el *stock de seguridad*. Además,

- Dado un nivel de servicio α , al valor $1 - \alpha$ se le denomina *riesgo de rotura*.
- En la ecuación $r = E(\mathcal{X}) + \sigma_{\mathcal{X}}\mathcal{F}^{-1}(\alpha)$, al valor $\mathcal{F}^{-1}(\alpha)$ (más comúnmente denotado por z_{α}) se le denomina *factor de seguridad*.

Debido a la gran complejidad del modelo (r, Q) es frecuente encontrarse en la literatura con aproximaciones o tratamientos heurísticos del mismo. Aquí señalaremos brevemente algunas de ellas. En este trabajo usaremos la siguiente ya que es bastante usada en la práctica (conocida a veces como *aproximación stock de seguridad*+EOQ). Prefijando un nivel de servicio α , calcularemos r como lo descrito anteriormente. La cantidad de pedido Q la tomaremos como el valor EOQ. Otra aproximación es la llamada EOQB (ver Snyder y Shen (2011,p.73)), si bien presenta mayor complejidad de computación, tiene la ventaja de que se dispone de una cota del error del 11.8 %. Nótese que nuestra aproximación es una solución exacta en cierto sentido, pues es la solución exacta de una aproximación al modelo (r, Q) . En el siguiente capítulo estudiaremos un modelo integrado de Localización-Inventario donde no aparecen explícitamente las variables de decisión correspondientes a los parámetros de política (r, Q) , pues podremos modelar bastante bien los costes y los fenómenos asociados al inventario y a la política (r, Q) . Es por ello, que no desarrollamos aquí un problema de programación no lineal para hallar los valores de la política (r, Q) y el coste total esperado (véase Zipkin (2000) para más cuestiones acerca del modelo (r, Q)).

Finalmente, señalaremos una política próxima a la (r, Q) bastante usada en la práctica, la *política KANBAN*²: Con una política KANBAN³ existen N contenedores, cada uno con Q unidades del producto y con una tarjeta⁴ en la parte inferior. Cuando un contenedor se vacía, la tarjeta se utiliza como orden para las Q unidades. $N - 1$ de las tarjetas siempre están asociadas con contenedores llenos, que están en stock u ordenados pero aún no se han entregado. Brevemente, sin entrar en demasiados detalles, una política KANBAN es muy similar a una política (r, Q) , con $r = (N - 1)Q$.

1.4. Métodos heurísticos y metaheurísticos

Esta sección, está dedicada a presentar los fundamentos básicos de algunos métodos heurísticos y metaheurísticos necesarios para el cuerpo central del trabajo. A pesar de que hoy en día, existen multitud de metaheurísticas, nos centraremos en la metaheurística conocida como GRASP por ser la que se ha utilizado en una gran parte del trabajo.

²Política de inventario originaria de Toyota que a su vez es un subsistema del Just in Time (JIT)

³*Kanban* significa tarjeta en japonés

⁴Hoy en día, suelen usarse señales o algo identificativo del contenedor

1.4.1. Métodos heurísticos

Si bien en Programación Lineal, se dispone del método símplex (y de otros métodos como generación de columnas por ejemplo cuando el problema es de un tamaño grande), en Programación Entera se produce un gran salto cualitativo por la dificultad que encierra la misma. Si bien los métodos heredados de la llamada teoría poliédrica, tales como métodos de planos de corte, o los métodos enumerativos (algoritmos Branch & Bound) forman parte de los fundamentos de los solvers comerciales para resolver problemas de Programación Entera, se han ido desarrollando procedimientos heurísticos para cada tipo de problema, es decir algoritmos que proporcionan soluciones aproximadas de calidad media. La razón principal del uso de procedimientos heurísticos es que se necesitan cuando o bien no se dispone de solver comercial específico para un tipo de problema o el solver comercial necesita elevados tiempos de computación, o bien cuando no se tiene un buen método de resolución. Una posible clasificación de los métodos heurísticos es la siguiente: *heurísticas de construcción* y *heurísticas de mejora*.

- Una **heurística de construcción** *construye, paso a paso, una solución del problema*. Con notable diferencia, los métodos constructivos que más se usan son los algoritmos *greedy* (también conocidos como *voraces* o *miopes*). Los métodos *greedy* se pueden utilizar para resolver, de forma aproximada o heurística, cualquier problema de optimización, incorporando cada vez un elemento a la solución en función de la información disponible en ese momento; una vez elegido el elemento que va a pertenecer a la solución, el método no se replantea sacarlo, ni se hace ningún tipo de re-optimización. Algo fundamental, es que son métodos rápidos y fáciles de implementar pero no garantizan alcanzar la solución óptima. El funcionamiento es muy sencillo, y está basado en la definición de un criterio *greedy*, que normalmente está definido a partir de la función objetivo del problema. El algoritmo tratará de encontrar el elemento candidato que minimiza (o maximiza en su caso) el criterio *greedy*, de forma que, cumpliéndose las restricciones del problema, se pueda incorporar a la solución final. En cada etapa se tomará la decisión que se considera mejor en ese momento, sin considerar las consecuencias futuras, es decir, se escogerá entre todos los elementos candidatos el que minimiza (o maximiza en su caso) el criterio *greedy*. Para el siguiente paso es necesario redefinir el criterio y repetir la etapa hasta obtener una solución completa. En el siguiente pseudocódigo de un método *greedy* genérico, se está denotando por C a la lista de candidatos que pueden incluirse en la solución, y por $S \subseteq C$ una solución parcial.

Algoritmo 1.1: Pseudocódigo de un algoritmo greedy genérico

```
1  $S = \emptyset$ ;  
2 while  $S$  no sea una solución y  $C \neq \emptyset$  do  
3    $x = \text{selección}(C)$ ;  
4    $C = C \setminus \{x\}$ ;  
5   if  $S \cup \{x\}$  es factible then  
6      $S = S \cup \{x\}$   
7 if  $S$  es una solución then  
8   return  $S$ 
```

La selección del elemento x a incluir en la solución depende de la función criterio elegida. Por lo tanto, la elección del criterio es clave para el rendimiento de un método greedy, y es muy dependiente de cada tipo de problema. En la práctica, la selección se efectúa encontrando un índice que minimiza (o maximiza) el criterio. En el caso de que haya varios índices optimizando el criterio, el método no discrimina ninguno en particular. Esto implica que mediante el método greedy es posible obtener varias soluciones diferentes. Existen varias maneras de introducir aleatoriedad a este algoritmo, y así obtener un método greedy aleatorizado. Uno de los procedimientos más conocidos utiliza una lista restringida de candidatos (RCL: *Restricted Candidate List*). Dicha lista está constituida por los elementos candidatos que ofrecen los mejores valores de la función o criterio greedy. El siguiente candidato añadido a la solución se elige aleatoriamente de la lista restringida de candidatos. Tal lista puede tener un número fijo de elementos (restricción por cardinalidad) o consistir de los elementos con los valores de la función greedy comprendidos en un rango dado (restricción por valor).

- Una **heurística de mejora** *parte de una solución factible ya dada, y trata de mejorarla*. Los métodos de mejora más usados están basados en *búsqueda locales* o *búsqueda por entornos*. Un procedimiento de búsqueda local en esencia, es como sigue: Se parte de una solución factible (obtenida por ejemplo con un método greedy), y se trata de mejorarla. Para ello, se explora el entorno o vecindad de la solución, mediante una operación básica llamada movimiento que, aplicada a la solución, proporciona las soluciones de su entorno. Un procedimiento de búsqueda local queda determinado al especificar cómo es el entorno y el criterio de selección de una solución dentro del entorno. La definición de entorno/movimiento, depende en gran medida de la estructura del problema a resolver, así como de la función objetivo. Por ejemplo, es muy habitual definir entornos mediante intercambios (*swap*). El óptimo local alcanzado no puede mejorarse mediante el tipo de movimiento que define

el entorno. Sin embargo, el método empleado no permite garantizar, de ningún modo, que sea el óptimo global del problema. Más concretamente, dado $x \in X$ solución, siendo X aquí el conjunto de todas las soluciones factibles, definamos $N(x)$ como el conjunto de soluciones $N(x) \subseteq X$ asociadas a x . Denominaremos a $N(x)$ entorno de x . En el siguiente pseudocódigo se está denotando por f a la función objetivo a minimizar, y por x^o a una solución inicial.

Algoritmo 1.2: Pseudocódigo de búsqueda local (Busqueda_Local)

Data: $x^o, N(\cdot), f(\cdot)$

```

1  $x := x^o$ ;
2 while  $x$  no sea localmente óptimo con respecto a  $N(x)$  do
3   |   Sea  $y \in N(x)$  tal que  $f(y) < f(x)$  ;
4   |    $x := y$ 
5 return  $x$ .
```

Un frecuente método heurístico en dos etapas consiste de un método greedy que proporciona una solución inicial, y un método de búsqueda local para tratar de mejorar esa solución. Otra clasificación de los métodos heurísticos podría ser entre *métodos de un único paso* (como un método greedy), y *métodos multi-arranque (multistart) o de varias pasadas*. En estos últimos métodos se usa, generalmente, algún tipo de aleatorización para obtener una solución diferente en cada pasada del procedimiento.

Nótese, que aunque la búsqueda local mejore las soluciones que proporciona un método greedy, generalmente se detiene al llegar a un óptimo local, en el que *queda atrapado*. Para poder escaparse de los óptimos locales se han diseñado una serie de métodos heurísticos más complejos, que se conocen como *metaheurísticas*.

1.4.2. Métodos metaheurísticos

Hoy en día, las metaheurísticas forman parte de uno de los procedimientos de resolución más importantes en Investigación Operativa. Unas de las más conocidas son los métodos Simulated Annealing, Tabu Search y los Algoritmos Genéticos. Estos métodos están basados en dos ideas principalmente, la *diversificación* y la *intensificación*; la primera se lleva a cabo mediante aleatorización y la segunda mediante la búsqueda local. Pueden citarse otras metaheurísticas muy populares, entre otras muchas existentes, como colonias de hormigas (Ant colony), búsqueda de entorno variable, búsqueda local guiada, GRASP o scatter search.

1.4.2.1. GRASP

GRASP (*Greedy Randomized Adaptive Search Procedure*) es una metaheurística muy popular de gran éxito en bastantes problemas de Optimización Combinatoria. Como referencia general actual para el GRASP, puede señalarse Resende y Ribeiro (2016). GRASP incorpora las dos características principales que se exigen a una metaheurística: la *diversificación* (obtenida mediante la aleatorización y repetición) y la *intensificación* (obtenida mediante la búsqueda local).

En cada iteración de un algoritmo GRASP se obtiene, como resultado de la búsqueda local, un mínimo local y, si la aleatorización ha hecho su papel correctamente, se habrían examinado múltiples regiones del espacio de soluciones, de forma que la mejor solución estaría cerca del óptimo global.

Esencialmente, una iteración básica del GRASP consta de dos fases:

1. **FASE CONSTRUCTIVA:** Se encuentran unas soluciones factibles iniciales de forma aleatoria (basándose en ciertos criterios), siendo las mismas no necesariamente óptimos locales. Además, los componentes de esta fase son: un criterio o función greedy, un procedimiento de elección aleatorio y un proceso de actualización adaptativo. Generalmente, suele ser un greedy aleatorizado.
2. **BÚSQUEDA LOCAL:** A partir de la solución hallada en la fase constructiva, se examinan los entornos de la misma con el objetivo de mejorarla.

GRASP es un método multistart o multiarranque, de forma que la iteración básica se repite un número predeterminado de veces (`iter_max`), guardando en cada paso la mejor solución encontrada, hasta que se alcanza el criterio de parada.

En el siguiente pseudocódigo de un GRASP genérico, se está denotando por x a la solución parcial en construcción en una iteración dada, por C al conjunto de elementos candidatos (que pueden ser añadidos a x), y por g a la función o criterio greedy. Además, por $RCL(C)$ denotamos a una lista restringida de candidatos de C .

Algoritmo 1.3: Pseudocódigo de GRASP

Data: iter_max

```
1  $f^* := \infty$ ;  
2 for  $i \leq \text{iter\_max}$  do  
3    $x := \emptyset$ ;  
4   Calcular  $C$ ;  
5   while  $x$  no sea solución y  $C \neq \emptyset$  do  
6      $\forall e \in C$ , calcular  $g(e)$ .  
7      $RCL(C) := \{ e \in C \mid g(e) \text{ tiene un } \textit{buen valor} \}$  ;  
8     Seleccionar al azar un elemento  $e^* \in RCL(C)$ .  
9      $x^* := x \cup \{e^*\}$ ;  
10    Actualizar el conjunto de candidatos  $C$ ;  
11    Calcular criterio greedy  $g(e)$ ,  $\forall e \in C$ .  
12     $x := \text{Busqueda\_Local}(x)$  ;  
13    if  $f(x) < f^*$  then  
14       $f^* := f(x)$ ;  
15       $x^* := x$ ;  
16 return  $x^*$ .
```

En este trabajo, optaremos por implementar la línea 7 del pseudocódigo del GRASP del modo siguiente (restricción por cardinalidad):

$$RCL(C) := \{ K \text{ elementos } e \in C \text{ con el menor valor de } g(e) \}.$$

Nótese que existen otras opciones tales como perturbación de costes o uso de una función sesgo por ejemplo (ver Resende y Ribeiro (2016)). De este modo, el GRASP queda definido por dos parámetros: el número en lista restringida de candidatos o cardinal de la lista restringida de candidatos, $KRCL = K$, y el número de iteraciones máximas, iter_max.

Capítulo 2

Un modelo de Localización-Inventario

2.1. Introducción

La estructura tradicional del FCL (*Facility Location Problem*) no considera la relación entre las decisiones de localización y las de inventario ni su impacto en la estructura de la red de distribución. Esto se debe a que el diseño de la cadena de suministro se resuelve secuencialmente, resolviendo primero el problema de localización y luego el problema de inventario. Más concretamente, la razón es la separación natural entre la toma de decisiones estratégicas (decisiones de localización de instalaciones) y tácticas u operacionales (decisiones de inventario). Sin embargo, cuando las mismas se abordan por separado, se suelen obtener soluciones subóptimas. Si bien la mayoría de artículos presentes en la literatura ignoran los costes de inventario o aproximan los costes no lineales por medio de funciones lineales, en la última década ha habido una fuerte tendencia a estudiar modelos integrados de Inventarios y Localización. Estos modelos determinan simultáneamente la ubicación de los centros de distribución (en adelante, CD) o almacenes que se abrirán, la asignación de clientes a los mismos y los parámetros óptimos de la política de inventario para minimizar el coste total del sistema.

Los modelos que aquí presentamos se centran en modelos de Localización-Inventario que integran en su formulación el nivel de servicio, en términos de disponibilidad de producto. En este sentido, Daskin *et al.* (2002) estudian un modelo de Localización-Inventario sin capacidades. La principal dificultad de este modelo es que los costes de inventario en cada CD son no lineales. El modelo está formulado

como un problema no lineal entero y resuelto mediante un procedimiento de relajación Lagrangiana para un caso especial en el que los ratios entre la varianza y la demanda (esperada) son constantes para todos los clientes. Shen *et al.* (2003) analizan el mismo problema que Daskin *et al.* (2002). Su trabajo reformula el modelo como un modelo de programación entera (de hecho de cubrimiento o *set covering*) y utiliza métodos de generación de columnas.

You y Grossmann (2008) relajan la hipótesis de que los ratios entre la varianza y la demanda esperada son constantes para todos los clientes, reformulando el modelo y resolviéndolo con diferentes enfoques de solución, incluyendo un método heurístico y un algoritmo de relajación Lagrangiana. Finalmente, Atamtürk *et al.* (2012) reformulan el modelo como un problema cónico-cuadrático entero mixto, relajando también la hipótesis mencionada anteriormente, aplicando cortes, funciones submodulares y desigualdades válidas.

El problema que aquí presentamos fue motivado por un estudio realizado por Shen *et al.* (2003) y Daskin *et al.* (2002) en un banco de sangre con sede en Chicago. El estudio se centraba en la producción y distribución de plaquetas, el más caro y el más perecedero de todos los productos sanguíneos. Además, debía tenerse en cuenta varios aspectos tales como el hecho de que si una unidad de plaquetas no es utilizada en los cinco días siguientes a la producción de sangre (entera), debe ser destruida y, que la demanda de plaquetas es algo muy variable, ya que se necesitan en variados contextos médicos.

Otra aplicación del modelo (en Ingeniería Química) ha sido a la cadena de suministro del oxígeno líquido (You y Grossmann (2008)).

2.2. Formulación del modelo. Formulación como problema no lineal entero (INLP)

El modelo que presentamos es el siguiente. Consideremos un sistema formado por proveedores, centros de distribución o almacenes y clientes (minoristas o *retailers*). El problema es *determinar el número óptimo de CD, sus localizaciones, los clientes asignados a cada CD y la política óptima de cada CD*. En este modelo, tanto la capacidad de almacenamiento de los CD como la capacidad de suministro de los proveedores, es ilimitada.

Supondremos que el horizonte de planificación es continuo y de duración por ejemplo, un año, y que los CD realizan un método de revisión continua y que la política de inventario es del tipo (r, Q) . No se incluyen explícitamente las variables de decisión r_j y Q_j , pues son innecesarias, ya que se incorpora el coste correspondiente en la función objetivo que resulta cuando se establecen los parámetros r_j y Q_j . Además, haremos una aproximación heurística (stock de seguridad+EOQ) para obtener los parámetros de la política de inventario.

Varios artículos (Daskin *et al.* (2002), Shen *et al.* (2003) por ejemplo) argumentan que la aproximación stock de seguridad+ EOQ provee una cota del error del 11.8%. Esto es *falso*, pues dicha cota sólo es válida para la aproximación EOQB. Salvo mayor conocimiento, no hay una cota del error para la aproximación que usamos aquí. Además, en la literatura existente, es muy usual encontrarse con la hipótesis de normalidad de las demandas de los clientes. Aquí, permitiremos mayor generalidad, pues supondremos que la demanda de los clientes puede expresarse por medio de una v.a. continua (con esperanza finita) no necesariamente normal.

Asumiremos que los costes de transporte dependen sólo de los CD y de la distancia entre los CD y los proveedores, y no de los proveedores en sí. Por lo tanto, la elección del mejor proveedor para un cierto CD puede ser fácilmente hecha *a priori*. Con estas hipótesis, el modelo es matemáticamente equivalente a suponer que existe un único proveedor. Para un ejemplo de la estructura de la red de la cadena de suministro, véase la figura 2.1.

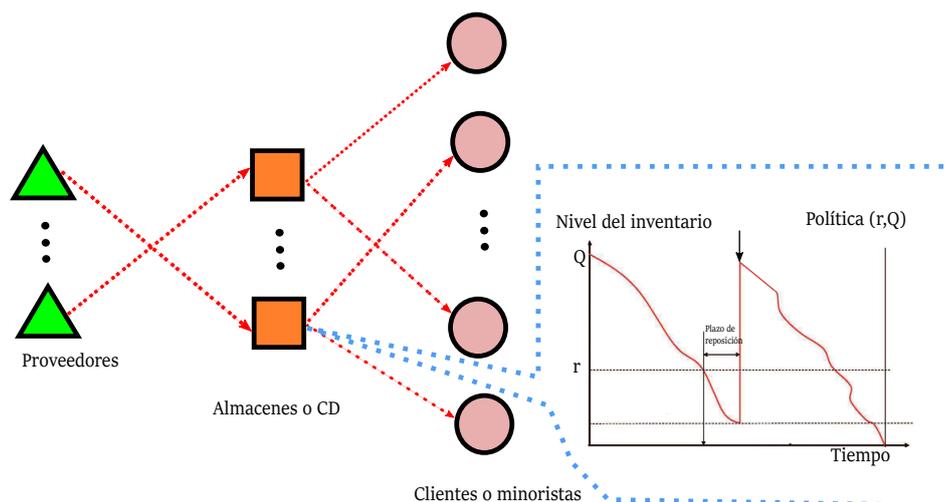


Figura 2.1: Estructura de la red de la cadena de suministro.

NOTACIÓN	
I	conjunto de clientes.
J	conjunto de puntos de posibles localizaciones de los CD.
μ_i	demanda media diaria del cliente i .
σ_i^2	varianza de la demanda diaria del cliente i .
f_j	coste fijo (anual) por apertura de un CD en el punto j .
α	nivel de servicio.
β	peso asociado a los costes de transporte.
θ	peso asociado a los costes de inventario.
z_α	factor de seguridad asociado al nivel de servicio α .
h_j	holding cost unitario (anual) del CD j .
F_j	coste fijo por realizar un pedido (elevar el nivel del inventario) en el CD j .
L_j	plazo de entrega o de reposición (en días) asociado al CD j .
g_j	coste fijo por transporte desde el proveedor al CD j .
a_j	coste unitario por transporte desde el proveedor al CD j .
d_{ij}	coste unitario de transporte desde el CD j al cliente i .
χ	número de días trabajados por año.
VARIABLES DE DECISIÓN	
x_j	= 1, si el punto j se selecciona como CD, y 0 en otro caso.
y_{ij}	= 1 si el cliente i es asignado al CD j , y 0 en otro caso.

- Los pesos β y γ son usados para incrementar o disminuir la importancia relativa de los costes de transporte e inventario presentes en la función objetivo. Esto puede ser interesante de cara a realizar un análisis de la sensibilidad.
- Asumimos que el horizonte de planificación es de un año, aunque uno puede elegir un horizonte de planificación distinto, simplemente ajustando los valores de f_j , h_j y χ adecuadamente.

El modelo de Localización-Inventario que presentamos es el siguiente:

$$\begin{aligned}
\text{Minimizar } & \sum_{j \in J} f_j x_j + \left(\beta \sum_{j \in J} \sum_{i \in I} \chi d_{ij} \mu_i y_{ij} \right) \\
& + \left(\sum_{j \in J} \sqrt{2\theta h_j (F_j + \beta g_j)} \sum_{i \in I} \chi \mu_i y_{ij} + \beta \sum_{j \in J} a_j \sum_{i \in I} \chi \mu_i y_{ij} \right) \quad (2.1) \\
& + \theta z_\alpha \sum_{j \in J} h_j \sqrt{\sum_{i \in I} L_j \sigma_i^2 y_{ij}}
\end{aligned}$$

$$\text{sujeta a } \sum_{j \in J} y_{ij} = 1, \forall i \in I \quad (2.2)$$

$$y_{ij} \leq x_j, \forall i \in I, \forall j \in J \quad (2.3)$$

$$x_j \in \{0, 1\}, \forall j \in J \quad (2.4)$$

$$y_{ij} \in \{0, 1\}, \forall i \in I, \forall j \in J \quad (2.5)$$

El primer término de la función objetivo es el coste fijo total por la apertura de los CD. El tercer término, $\sqrt{2\theta h_j \chi (F_j + \beta g_j) \sum_{i \in I} \mu_i y_{ij}}$, es similar a la clásica expresión para el coste óptimo del modelo EOQ, donde aquí el coste fijo es $\chi (F_j + \beta g_j)$, el holding cost está dado por θh_j y la demanda (esperada) asignada usando las variables de decisión está dada por la expresión $\sum_{i \in I} \mu_i y_{ij}$. El cuarto término, $\theta z_\alpha \sum_{j \in J} h_j \sqrt{\sum_{i \in I} L_j \sigma_i^2 y_{ij}}$, representa el coste de almacenamiento del stock de seguridad de los CD $j \in J$ con un nivel de servicio α (nótese que aquí $\sqrt{L_j \sum_{i \in I} \sigma_i^2 y_{ij}}$ correspondería a σ_χ de la expresión (1.2), la cual es multiplicada por z_α y por el factor h_j , evaluando el coste total sumando en $j \in J$).

El primer grupo de restricciones, (2.2), son las *restricciones de fuente única*: cada cliente esté asignado exactamente a un CD. El segundo grupo de restricciones, (2.3), aseguran que si un CD no está abierto entonces no puede suministrar a ningún cliente. Estas restricciones conforman una familia de facetas del politopo asociado al problema. Es por ello, que a esta formulación se la denomina fuerte. Finalmente, los dos últimos grupos de restricciones, (2.4) y (2.5), se refieren al carácter binario de las variables.

Nos encontramos ante un problema de Programación no lineal entera donde la función objetivo es cóncava. Además, podemos reescribir la función objetivo como sigue:

$$\begin{aligned}
(\text{P0}) \text{ Minimizar } & \sum_{j \in J} f_j x_j + \left(\beta \sum_{j \in J} \sum_{i \in I} \chi d_{ij} \mu_i \gamma_{ij} \right) \\
& + \left(\sum_{j \in J} \sqrt{2\theta h_j (F_j + \beta g_j)} \sum_{i \in I} \chi \mu_i \gamma_{ij} + \beta \sum_{j \in J} a_j \sum_{i \in I} \chi \mu_i \gamma_{ij} \right) \\
& + \theta z_\alpha \sum_{j \in J} h_j \sqrt{\sum_{i \in I} L_j \sigma_i^2 \gamma_{ij}} = \sum_{j \in J} \left(f_j x_j + \sum_{i \in I} \widehat{d}_{ij} \gamma_{ij} + K_j \sqrt{\sum_{i \in I} \mu_i \gamma_{ij}} + q_j \sqrt{\sum_{i \in I} \widehat{\sigma}_i^2 \gamma_{ij}} \right)
\end{aligned} \tag{2.6}$$

donde

$$\begin{aligned}
\widehat{d}_{ij} &= \beta \chi \mu_i (d_{ij} + a_j) \\
K_j &= \sqrt{2\theta h_j \chi (F_j + \beta g_j)} \\
q_j &= \theta h_j z_\alpha \\
\widehat{\sigma}_i^2 &= L_j \sigma_i^2
\end{aligned}$$

Tras esta reescritura, podemos dar una sencilla interpretación a estos nuevos parámetros. Más concretamente:

- \widehat{d}_{ij} está asociado a los costes unitarios de transporte, tanto los relativos a los pedidos al proveedor, como a los de envío a los clientes.
- K_j se asocia a los costes de inventario debidos a los costes fijos relativos a los lanzamientos de pedidos en los CD, y a los de transporte del producto desde el proveedor.
- q_j está relacionado con los costes del stock de seguridad.
- $\widehat{\sigma}_i^2$ engloba el plazo de reposición junto con la varianza.

2.3. Análisis poliédrico

Una primer enfoque en la resolución del problema de Programación no lineal entera que aquí consideramos, puede ser un análisis poliédrico del mismo. Más concretamente, dado que minimizar una función cóncava sobre un poliedro esta alcanza

el óptimo en uno de los puntos extremos del mismo (ver Rockafellar (1972)), estamos interesados en saber si la relajación lineal del problema proporciona soluciones enteras. Por tanto, se resolvería un problema con función objetivo no lineal cóncava con variables continuas y restricciones lineales.

En primer lugar, podemos estudiar el grafo subyacente al problema, sea $G = (V, A)$ tal grafo, donde V está particionado en I y J , siendo I el conjunto de clientes y J el conjunto de puntos de posibles localizaciones de los CD y $A \subseteq I \times J$.

Además, consideremos las restricciones asociadas al problema cuando se toma la relajación lineal del mismo. Más concretamente, se tiene el siguiente politopo:

$$\sum_{j \in J} y_{ij} = 1, \forall i \in I \quad (2.7)$$

$$y_{ij} \leq x_j, \forall i \in I, \forall j \in J \quad (2.8)$$

$$0 \leq x_j \leq 1 \quad (2.9)$$

$$y_{ij} \geq 0 \quad (2.10)$$

Queremos caracterizar cuando (2.7)–(2.10) define un politopo entero. Denotemos por P_G dicho politopo.

Sea $W \subset V$, y definamos $\delta^+(W)$ como el conjunto de arcos $(i, j) \in A$, con $i \in W$ y $j \in V \setminus W$. Además, definamos $\delta^-(W)$ como el conjunto de arcos $(i, j) \in A$, con $j \in W$ y $i \in V \setminus W$. Sean \bar{I} el conjunto de clientes $i \in I$ tales que $|\delta^+(\{i\})| = 1$ y \bar{J} el conjunto de posibles puntos de localización de los CD que son adyacentes a un cliente del tipo \bar{I} . Es claro que $x_j = 1 \forall j \in \bar{J}$. Sea \bar{G} el subgrafo inducido por $V \setminus \bar{J}$. Puede probarse el siguiente teorema:

Teorema 2.1 (Baïou y Barahona (2009)).

P_G es un politopo entero, si y sólo si, \bar{G} no tiene ciclos impares.

A través del teorema anterior, vemos que trasladamos una propiedad de politopos a grafos. A su vez, podemos trasladar el problema de grafos a un problema de conos. Más concretamente, estudiaremos brevemente el cono convexo asociado al grafo \bar{G} . Para ello seguiremos García-Laguna (1984).

Sean $v_{\overline{G}}$ y $E_{\overline{G}}$ el número de vértices y el conjunto de arcos de \overline{G} respectivamente.

A cada arco $(i, j) \in E_{\overline{G}}$ le hacemos corresponder el semiespacio cerrado de $\mathbb{R}^{v_{\overline{G}}}$ $x_i \leq x_j$ si $i \neq j$ o todo el espacio $\mathbb{R}^{v_{\overline{G}}}$ si $i = j$. La intersección de todos estos semiespacios cerrados forma el cono asociado al grafo \overline{G} dado por

$$P_{\overline{G}} = \{x \in \mathbb{R}^{v_{\overline{G}}} / A_{\overline{G}}x \leq 0\}$$

donde $A_{\overline{G}}$ es la matriz traspuesta de la matriz de incidencia asociada al grafo \overline{G} .

Puede probarse lo siguiente (ver García-Laguna (1984)):

- Dos vértices i, j pertenecen al mismo ciclo si, y sólo si en el cono $P_{\overline{G}}$ se tiene $x_i = x_j$.
- Si el grafo \overline{G} no tiene ciclos entonces $\dim(P_{\overline{G}}) = v_{\overline{G}}$.
- Si el grafo \overline{G} es conexo, entonces \overline{G} es un árbol si y sólo si el rango de la matriz $A_{\overline{G}}$ es $v_{\overline{G}}$ y $A_{\overline{G}}$ tiene $v_{\overline{G}}$ filas.
- El grafo \overline{G} carece de ciclos si, y sólo si existe una numeración ϕ (con valores enteros no negativos) de los vértices de \overline{G} , tal que para todo arco $(i, j) \in A_{\overline{G}}$ se tiene $\phi(i) < \phi(j)$.

Siguiendo esta línea, se haría un estudio poliédrico más profundo para diseñar métodos específicos de resolución. Esto no es lo que haremos aquí, pues en general, salvo muy raras ocasiones, el politopo no será entero.

2.4. Propiedades básicas del problema

El problema (P0) es un problema no lineal entero donde todas las variables de decisión son binarias. Además de su naturaleza combinatoria, los términos no lineales son no convexos, lo cual hace que el problema sea muy difícil de resolver. Además, este problema tiene la siguiente propiedad tal como señala Snyder y Shen (2011) o Daskin (2013): a diferencia del UFLP (*Uncapacited Facility Problem*) (nótese que el UFLP es un caso particular de nuestro problema (basta considerar el holding cost igual a 0), y dado que aquel es un problema NP-duro, se sigue que nuestro problema también lo es), los clientes no siempre son asignados a los CD abiertos

más cercanos, ya que influyen los costes de inventario y transporte (es el resultado de la presencia de los términos no lineales presentes en la función objetivo) y además, no sólo es posible asignar un nodo de demanda a un CD no próximo en una solución óptima, sino que también es posible que las demandas en un punto cliente que alberga un CD se asignen a otro CD distinto a el de ese nodo. Con el fin de abordar el problema, en la literatura es frecuente encontrarse con la hipótesis de que los ratios $\frac{\sigma_i^2}{\mu_i}$ son constantes e iguales para cada cliente i (véase por ejemplo Daskin *et al.* (2002)) (puede probarse que si se asume esta hipótesis lo anterior no ocurre (ver Daskin *et al.* (2002) o Daskin (2013) para mayor detalle acerca de esta hipótesis y sus implicaciones). Sin embargo, a pesar de que se gana en simplicidad del modelo pues se puede reducir el problema a uno con un sólo término no lineal en la función objetivo, en la práctica esta hipótesis no es muy realista, pues los ratios $\frac{\sigma_i^2}{\mu_i}$ pueden variar de unos nodos a otros. Por tanto, es necesario un algoritmo eficiente para resolver el problema sin la hipótesis restrictiva mencionada anteriormente. En consecuencia, recurrimos a otro enfoque para atacar el problema: reformulaciones equivalentes y relajación Lagrangiana. Estos son los objetivos de las secciones siguientes.

2.5. Formulación como problema no lineal entero mixto (MINLP)

El problema original (P0) es muy difícil de resolver debido al elevado número (potencial) de variables binarias. La siguiente proposición nos permitirá reformular el problema (P0) como un problema no lineal entero mixto (MINLP) con menos variables binarias, lo cual reduce considerablemente el esfuerzo computacional.

Proposición 2.2 (You y Grossmann (2008)).

Para valores fijos de las variables x_j , las variables continuas y_{ij} toman los valores $\{0, 1\}$ al optimizar globalmente o localmente el problema (P0).

La proposición anterior equivale a que en el siguiente problema

$$(P1) \text{ Minimizar } \sum_{j \in J} \left(f_j x_j + \sum_{i \in I} \hat{d}_{ij} y_{ij} + K_j \sqrt{\sum_{i \in I} \mu_i y_{ij}} + q_j \sqrt{\sum_{i \in I} \hat{\sigma}_i^2 y_{ij}} \right) \quad (2.11)$$

$$\text{sujeta a } \sum_{j \in J} y_{ij} = 1, \forall i \in I \quad (2.12)$$

$$y_{ij} \leq x_j, \forall i \in I, \forall j \in J \quad (2.13)$$

$$x_j \in \{0, 1\}, \forall j \in J \quad (2.14)$$

$$y_{ij} \geq 0, \forall i \in I, \forall j \in J \quad (2.15)$$

las variables binarias de asignación y_{ij} tomen los valores $\{0, 1\}$ al optimizar globalmente o localmente para valores fijos binarios de las variables x_j .

2.6. Reformulación como problema no lineal entero mixto (MINLP)

Con el propósito de incrementar la eficiencia computacional del método de resolución del problema no lineal entero mixto (P1) reformulamos el mismo. Las raíces cuadradas presentes en la función objetivo (2.11) pueden dar lugar a dificultades en el procedimiento de optimización. Cuando el punto j no se selecciona como CD, ambos términos asociados al mismo de la raíz cuadrada tomarían el valor 0, lo que llevaría a gradientes no acotados en la optimización no lineal y por tanto, dificultades en la misma. En consecuencia, reformularemos el modelo con el fin de eliminar las raíces cuadradas. Para ello, primero introduzcamos dos conjuntos de variables (continuas y no negativas) u_j y v_j , para representar los términos de las raíces cuadradas presentes en la función objetivo:

$$u_j^2 = \sum_{i \in I} \mu_i y_{ij}, \forall j \in J \quad (2.16)$$

$$v_j^2 = \sum_{i \in I} \widehat{\sigma}_i^2 y_{ij}, \forall j \in J \quad (2.17)$$

$$u_j \geq 0, v_j \geq 0, \forall j \in J \quad (2.18)$$

Además, dado que las variables no negativas u_j y v_j en la función objetivo presentan coeficientes positivos y estamos ante un problema de minimización, podemos relajar las restricciones (2.16) y (2.17):

$$-u_j^2 + \sum_{i \in I} \mu_i y_{ij} \leq 0, \forall j \in J \quad (2.19)$$

$$-v_j^2 + \sum_{i \in I} \widehat{\sigma}_i^2 y_{ij} \leq 0, \forall j \in J \quad (2.20)$$

En consecuencia, podemos reformular el problema (P1) como el problema no lineal entero mixto (P2):

$$(P2) \text{ Minimizar } \sum_{j \in J} \left(f_j x_j + \sum_{i \in I} \widehat{d}_{ij} y_{ij} + K_j u_j + q_j v_j \right) \quad (2.21)$$

$$\text{sujeta a } \sum_{j \in J} y_{ij} = 1, \forall i \in I \quad (2.22)$$

$$y_{ij} \leq x_j, \forall i \in I, \forall j \in J \quad (2.23)$$

$$-u_j^2 + \sum_{i \in I} \mu_i y_{ij} \leq 0, \forall j \in J \quad (2.24)$$

$$-v_j^2 + \sum_{i \in I} \widehat{\sigma}_i^2 y_{ij} \leq 0, \forall j \in J \quad (2.25)$$

$$x_j \in \{0, 1\}, \forall j \in J \quad (2.26)$$

$$y_{ij} \geq 0, \forall i \in I, \forall j \in J \quad (2.27)$$

$$u_j \geq 0, v_j \geq 0, \forall j \in J \quad (2.28)$$

(P2) es un problema no lineal entero mixto con función objetivo lineal y restricciones no lineales. A continuación, enunciamos una proposición relativa al problema (P2). Nótese la similitud con la proposición 2.2.

Proposición 2.3 (You y Grossmann (2008)).

Para valores fijos de las variables x_j , las variables continuas y_{ij} toman los valores $\{0, 1\}$ al optimizar globalmente o localmente el problema (P2).

2.7. Relajación Lagrangiana

En este capítulo y para el problema que tratamos aquí, hemos optado por usar relajación Lagrangiana puesto que métodos heurísticos de tipo *greedy* o procedimientos de *búsqueda local* (y por ende, algunas metaheurísticas basadas en los mismos, tales como GRASP) han resultado aquí ser bastante lentos en la práctica, debido fundamentalmente a la no linealidad del problema en las variables de asignación.

2.7.1. Descomposición espacial del problema

$$(\mathbf{P2}\lambda) \text{ Minimizar } \sum_{j \in J} \left(f_j x_j + \sum_{i \in I} (\widehat{d}_{ij} - \lambda_i) y_{ij} + K_j u_j + q_j v_j \right) + \sum_{i \in I} \lambda_i \quad (2.29)$$

$$\text{sujeta a } y_{ij} \leq x_j, \forall i \in I, \forall j \in J \quad (2.30)$$

$$-u_j^2 + \sum_{i \in I} \mu_i y_{ij} \leq 0, \forall j \in J \quad (2.31)$$

$$-v_j^2 + \sum_{i \in I} \widehat{\sigma}_i^2 y_{ij} \leq 0, \forall j \in J \quad (2.32)$$

$$x_j \in \{0, 1\}, \forall j \in J \quad (2.33)$$

$$y_{ij} \geq 0, \forall i \in I, \forall j \in J \quad (2.34)$$

$$u_j \geq 0, v_j \geq 0, \forall j \in J \quad (2.35)$$

Denotaremos por V al valor óptimo de la correspondiente función objetivo.

El problema $(\mathbf{P2}\lambda)$ puede descomponerse en $|J|$ subproblemas, uno para cada punto posible de localización de los CD $j \in J$, los cuales denotaremos por $(\mathbf{P2}_{j^*}\lambda)$. Más concretamente, sea $j^* \in J$ fijo. Se tiene

$$(\mathbf{P2}_{j^*}\lambda) \text{ Minimizar } \left(f_{j^*} x_{j^*} + \sum_{i \in I} (\widehat{d}_{ij^*} - \lambda_i) y_{ij^*} + K_{j^*} u_{j^*} + q_{j^*} v_{j^*} \right) + \sum_{i \in I} \lambda_i \quad (2.36)$$

$$\text{sujeta a } y_{ij^*} \leq x_{j^*}, \forall i \in I, \quad (2.37)$$

$$-u_{j^*}^2 + \sum_{i \in I} \mu_i y_{ij^*} \leq 0, \quad (2.38)$$

$$-v_{j^*}^2 + \sum_{i \in I} \widehat{\sigma}_i^2 y_{ij^*} \leq 0, \quad (2.39)$$

$$x_{j^*} \in \{0, 1\}, \quad (2.40)$$

$$y_{ij^*} \geq 0, \forall i \in I, \quad (2.41)$$

$$u_{j^*} \geq 0, v_{j^*} \geq 0, \quad (2.42)$$

Denotaremos por V_{j^*} al valor óptimo de la correspondiente función objetivo.

Para cada $j \in J$, el subproblema $(\mathbf{P2}_j\lambda)$ tiene sólo una variable binaria, x_j . Además, tiene $|I| + 2$ variables continuas, u_j, v_j e y_{ij} , y $2|I| + 2$ restricciones. Esta descomposición se conoce como *descomposición espacial*, ya que se basa en aprovechar la estructura espacial de la red de la cadena de suministro.

Como resultado de esta descomposición, el valor óptimo (global) de la función objetivo de $(\mathbf{P2}\lambda)$, el cual corresponde a una cota inferior del problema $(\mathbf{P2})$, puede calcularse como

$$V = \sum_{j \in J} V_j + \sum_{i \in I} \lambda_i \quad (2.43)$$

Para cada valor fijo de los multiplicadores λ_i , resolveremos el problema $(\mathbf{P2}_j\lambda)$ minimizando para cada $j \in J$. Entonces, en virtud de (2.43), el valor óptimo de la función objetivo puede calcularse para cada valor λ_i . Además, usaremos un método de Optimización Subgradiente para actualizar los multiplicadores λ_i .

2.7.2. Subproblemas de la relajación Lagrangiana

En cada iteración (con valores fijos de los multiplicadores λ_i), las variables x_j se optimizan separadamente en cada subproblema $(\mathbf{P2}_j\lambda)$ de acuerdo con el procedimiento de descomposición mencionado anteriormente. Para cada subproblema $(\mathbf{P2}_j\lambda)$, observemos que el valor de la correspondiente función objetivo es 0 si, y sólo si $x_j = 0$ (es decir, si no seleccionamos el punto j como CD). En otras palabras, hay una solución factible que conduce al valor 0. En consecuencia, el valor óptimo del subproblema $(\mathbf{P2}_j\lambda)$ debe ser menor o igual que 0. Debido a esta observación, es posible que bajo algún λ_i los valores óptimos de la función objetivo para todos los subproblemas $(\mathbf{P2}_j\lambda)$ sean 0 (es decir, $x_j = 0, \forall j \in J$, no seleccionando ningún punto $j \in J$ como CD). En cualquier caso la restricción (2.22) implica

$$\sum_{j \in J} x_j \geq 1 \quad (2.44)$$

Es decir, (2.44) quiere decir que al menos un punto $j \in J$ debe ser seleccionado como CD para satisfacer la demanda. Sin embargo, una vez que se ha relajado la restricción (2.22), la restricción (2.44) no se tiene en el problema, luego esta debe

ser tenida en cuenta en el procedimiento de resolución del problema. Más concretamente, para que se satisfaga la restricción (2.44) en la relajación Lagrangiana, realizaremos las siguientes modificaciones a la etapa antes mencionada para resolver para cada punto $j \in J$ el problema $(\mathbf{P2}_j\lambda)$:

En primer lugar, consideremos para cada $j \in J$, el problema $(\mathbf{P2R}_j\lambda)$, que es en realidad un caso particular de $(\mathbf{P2}_j\lambda)$ cuando $x_j = 1$. La formulación para un $j^* \in J$ fijo es

$$(\mathbf{P2R}_{j^*}\lambda) \text{ Minimizar } \left(f_{j^*} + \sum_{i \in I} (\widehat{d}_{ij^*} - \lambda_i) y_{ij^*} + K_{j^*} u_{j^*} + q_{j^*} v_{j^*} \right) + \sum_{i \in I} \lambda_i \quad (2.45)$$

$$\text{sujeta a } y_{ij^*} \leq 1, \forall i \in I, \quad (2.46)$$

$$-u_{j^*}^2 + \sum_{i \in I} \mu_i y_{ij^*} \leq 0, \quad (2.47)$$

$$-v_{j^*}^2 + \sum_{i \in I} \widehat{\sigma}_i^2 y_{ij^*} \leq 0, \quad (2.48)$$

$$y_{ij^*} \geq 0, \forall i \in I, \quad (2.49)$$

$$u_{j^*} \geq 0, v_{j^*} \geq 0, \quad (2.50)$$

Denotaremos por \widehat{V}_{j^*} al valor óptimo de la correspondiente función objetivo. Como la variable x_j no aparece en $(\mathbf{P2R}_{j^*}\lambda)$, el valor óptimo de la función objetivo del mismo es igual al valor óptimo de la función objetivo correspondiente al problema $(\mathbf{P2}_{j^*}\lambda)$ cuando $x_j = 1$. Nótese que en general, puede no ser igual al valor óptimo (global) de la función objetivo correspondiente al problema $(\mathbf{P2}_{j^*}\lambda)$, pues este puede ser alcanzado para $x_j = 0$.

Para cada valor λ_i fijo,

- Si $\widehat{V}_j < 0$, entonces si $x_j = 1$ se tiene $\widehat{V}_j = V_j < 0$ y, si $x_j = 0$, se llega a que $\widehat{V}_j \geq 0$ lo que es absurdo. En consecuencia, es óptimo que $x_j = 1$.
- Si $\widehat{V}_j \geq 0$ se tiene que si $x_j = 0$ entonces $0 = V_j \leq \widehat{V}_j$. Si $x_j = 1$ se comprueba trivialmente que se llega a un absurdo.

Un posible caso extremo es cuando $\widehat{V}_j > 0, \forall j \in J$ (por ejemplo cuando $\lambda_i = 0, i \in I$). Esto significa que $V_j = 0, \forall j \in J$, es decir que $x_j = 0, \forall j \in J$ (i.e. no se selecciona ningún punto como CD). Para que se satisfaga la restricción (2.44) (es decir,

que al menos se seleccione un punto como CD) instalaremos el CD en el punto $j \in J$ que tenga el menor valor de la función objetivo \widehat{V}_j .

En consecuencia, el algoritmo para resolver los subproblemas de la relajación Lagrangiana es como sigue. Para cada valor fijo de los multiplicadores λ_i , resolveremos para cada $j \in J$, el problema $(\mathbf{P2R}_j^*\lambda)$. Entonces, seleccionaremos aquellos puntos $j \in J$ como posibles puntos de localización de los CD (es decir $x_j = 1$) para los que $\widehat{V}_j \leq 0$. Para el resto de puntos $j \in J$ para los que $\widehat{V}_j > 0$, no serán seleccionados como CD y entonces $x_j = 0$. En otro caso, si $\widehat{V}_j > 0, \forall j \in J$, seleccionaremos sólo un CD con el mínimo $\widehat{V}_j > 0, j \in J$ es decir, $x_{j^*} = 1$ para $j^* \in J$ tal que $\widehat{V}_{j^*} = \min_{j \in J} \{\widehat{V}_j\}$.

Haciendo esto en cada iteración de la relajación Lagrangiana (para cada valor del multiplicador λ_i), aseguramos que la solución óptima siempre satisfaga (2.44). De este modo, podemos calcular V como

$$V = \sum_{j \in J, x_j=1} \widehat{V}_j + \sum_{i \in I} \lambda_i \quad (2.51)$$

2.7.3. Obtención de soluciones factibles: Heurística Lagrangiana

Podemos obtener soluciones factibles iniciales (y por tanto, cotas superiores para la solución óptima) por medio de una heurística Lagrangiana. Un método muy simple para obtener una solución factible es el siguiente. Se selecciona un CD y a continuación asignaremos al mismo todos los clientes y el correspondiente valor de la función objetivo. Es posible obtener cotas superiores más finas que las que se obtendrían con el método anterior. Más concretamente, para obtener soluciones factibles durante las iteraciones, fijados los valores óptimos de las variables x_j correspondientes a los subproblemas de la relajación Lagrangiana, se resuelve el problema $(\mathbf{P2})$. Nótese que en virtud de la proposición 2.3, se tiene que las variables y_{ij} toman valores binarios. Con este método de obtención de soluciones factibles que aquí proponemos, mejoramos notablemente la obtención de soluciones factibles que se obtendrían aplicando el paso 3 del algoritmo 2 de You y Grossmann (2008, p.7809), pues allí se propone resolver el problema $(\mathbf{P2}\lambda)$.

2.7.4. Algoritmo de resolución

A continuación, se resume el algoritmo propuesto para resolver el problema de este capítulo. Es un algoritmo mejorado basado en el que aparece en You y Grossmann (2008).

- **Paso 1** (Inicialización): Inicializar $UB := +\infty$ (cota superior), $LB := -\infty$ (cota inferior), $k := 0$ y tomar unos valores iniciales para los multiplicadores. Nosotros hemos optado por tomar como multiplicadores iniciales las variables duales obtenidas al optimizar el problema (P2) localmente. *Esto proporciona mejores multiplicadores que los que se obtendrían por el método propuesto por You y Grossmann (2008): inicializar los multiplicadores a valores arbitrarios o bien a las variables duales asociadas a la solución de la relajación del problema (P2).*
- **Paso 2** (Resolución de los subproblemas) : En la iteración k -ésima, fijado λ^k , resolveremos para cada $j \in J$, el problema (P2R $_j$ λ^k), cuya solución la denotaremos por $\widehat{y}_{ij}(\lambda^k)$ y por $\widehat{V}_j(\lambda^k)$ al valor objetivo. Entonces, seleccionaremos aquellos puntos $j \in J$ como posibles puntos de localización de los CD (es decir $x_j(\lambda^k) = 1$) para los que $\widehat{V}_j(\lambda^k) \leq 0$. Para el resto de puntos $j \in J$ para los que $\widehat{V}_j(\lambda^k) > 0$, no serán seleccionados como CD y entonces $x_j(\lambda^k) = 0$. En otro caso, si $\widehat{V}_j(\lambda^k) > 0, \forall j \in J$, seleccionaremos sólo un CD con el mínimo $\widehat{V}_j(\lambda^k) > 0, j \in J$ es decir, $x_{j^*}(\lambda^k) = 1$ para $j^* \in J$ tal que $\widehat{V}_{j^*}(\lambda^k) = \min_{j \in J} \{ \widehat{V}_j(\lambda^k) \}$.

Calcular

$$V(\lambda^k) = \sum_{j \in J, x_j(\lambda^k)=1} \widehat{V}_j(\lambda^k) + \sum_{i \in I} \lambda_i^k \quad (2.52)$$

Si $V(\lambda^k) > LB$, entonces $LB := V(\lambda^k)$.

- **Paso 3** (Obtención de soluciones factibles): Fijados los valores de las variables $x_j(\lambda^k)$, resolver el problema (P2), cuya solución la denotaremos por $y_{ij}(\lambda^k)$ y el valor objetivo correspondiente por $\bar{V}(\lambda^k)$. Aquí obtenemos una mejor solución factible que la que se obtendría aplicando el método propuesto en el paso 3, algoritmo 2 de You y Grossmann (2008, p.7809).

Si $\bar{V}(\lambda^k) < UB$, entonces $UB := \bar{V}(\lambda^k)$.

- **Paso 4** (Iteración subgradiente): Calcular el subgradiente ξ_i usando

$$1 - \sum_{j \in J} \hat{y}_{ij}(\lambda^k), \quad i \in I.$$

Calcular la longitud de paso

$$T_k := \frac{\theta_{sub}(UB - LB)}{\sum_{i \in I} \left(1 - \sum_{j \in J} \hat{y}_{ij}(\lambda^k)\right)^2}$$

Actualizar los multiplicadores:

$$\lambda^{k+1} := \max \{0, \lambda^k + T_k \xi^k\}$$

Además, si en más de un número prefijado de iteraciones (pasadas) la cota inferior LB no mejora, entonces $\theta_{sub} := \theta_{sub}/2$. En nuestros experimentos numéricos hemos optado por establecer dos pasadas.

- **Paso 5** (Criterio de parada): Si el salto de dualidad ($GAP = \frac{UB-LB}{UB}$) es menor que una tolerancia prefijada, o si el número de iteraciones máximas se ha alcanzado o si la norma del subgradiente o θ_{sub} son menores que unas tolerancias prefijadas, PARAR, y establecer UB como valor objetivo, y $x_j(\lambda^k)$ y $y_{ij}(\lambda^k)$ como la solución final obtenida. Si no, $k := k + 1$ e ir al paso 2.

2.8. Otras formulaciones equivalentes

2.8.1. Formulación cónica

Es posible formular el problema (P0) como un problema cónico-cuadrático entero mixto. Para ello, seguiremos Atamtürk *et al.* (2012). Más concretamente, dado que $y_{ij} = y_{ij}^2$, consideraremos $t_{1j}, t_{2j} \geq 0$ variables auxiliares para representar los términos no lineales en la función objetivo (2.6):

$$\text{(CQMIP) Minimizar } \sum_{j \in J} \left(f_j x_j + \sum_{i \in I} \widehat{d}_{ij} y_{ij} + K_j t_{1j} + q_j t_{2j} \right) \quad (2.53)$$

$$\text{sujeta a } \sum_{j \in J} y_{ij} = 1, \forall i \in I \quad (2.54)$$

$$y_{ij} \leq x_j, \forall i \in I, \forall j \in J \quad (2.55)$$

$$\sum_{i \in I} \mu_i y_{ij}^2 \leq t_{1j}^2, \forall j \in J \quad (2.56)$$

$$\sum_{i \in I} \widehat{\sigma}_i^2 y_{ij}^2 \leq t_{2j}^2, \forall j \in J \quad (2.57)$$

$$x_j \in \{0, 1\}, \forall j \in J \quad (2.58)$$

$$y_{ij} \in \{0, 1\}, \forall i \in I, \forall j \in J \quad (2.59)$$

$$t_{1j} \geq 0, t_{2j} \geq 0, \forall j \in J \quad (2.60)$$

Nótese que la función objetivo de (CQMIP) es *lineal*. Además las restricciones son cónico-cuadráticas. De hecho, es el problema (P2), donde aquí las variables de asignación y_{ij} son binarias. Atamtürk *et al.* (2012) siguen un método de resolución diferente al que aquí presentamos, aplicando un procedimiento relacionado con desigualdades válidas y funciones submodulares.

2.8.2. Formulación set covering

Es posible formular el problema (P0) como un problema clásico de cubrimiento (*set covering*, en inglés). Este es el enfoque de Shen *et al.* (2003) donde aplican un método conocido en la literatura como *generación de columnas*. Describiremos brevemente la formulación set covering:

Aquí, supondremos eventualmente y sin pérdida de generalidad que $I = J$, ya que por una parte, si hay un cliente o minorista que no es un CD o almacén potencial basta establecer su coste fijo de instalación como ∞ , y por otra parte si hay un CD o almacén potencial que no es cliente o minorista basta establecer su demanda igual a 0. En primer lugar, nótese que cada solución factible de nuestro problema de decisión consiste en una partición del conjunto I de clientes en subconjuntos no vacíos, R_1, R_2, \dots, R_n , junto con un CD designado para cada uno de estos n conjuntos. Sea \mathcal{R} la colección de todos los subconjuntos no vacíos del conjunto I . Para todo $R \in \mathcal{R}$ y todo $j \in R$, sea $C_{R,j}$ el coste de servicio del CD j a todos los clientes del conjunto R :

$$C_{R,j} = f_j + \sum_{i \in R} \widehat{d}_{ij} + K_j \sqrt{\sum_{i \in R} \mu_i} + q \sqrt{\sum_{i \in R} \widehat{\sigma}_i^2}$$

Además, sea $C_R = \min_{j \in R} \{C_{R,j}\}$. Por tanto, el problema de set-covering es

$$\text{Minimizar } \sum_{R \in \mathcal{R}} C_R Z_R \quad (2.61)$$

$$\text{sujeta a } \sum_{R \in \mathcal{R}: i \in R} Z_R \geq 1, \forall i \in I \quad (2.62)$$

$$Z_R \in \{0, 1\}, \forall R \in \mathcal{R} \quad (2.63)$$

donde Z_R es igual a 1 si el conjunto R está en la solución, y 0 en otro caso.

Tal como señala Snyder y Shen (2011, p.223), esta formulación junto con el procedimiento de resolución por generación de columnas, proporciona un método bastante rápido, aunque no tan rápido como la relajación Lagrangiana.

2.9. Resultados computacionales

En esta sección presentamos los resultados computacionales de los experimentos numéricos realizados. Los conjuntos de datos (88 nodos y 150 nodos) provienen (al igual que los artículos de Shen *et al.*(2003) y Atamtürk *et al.* (2012) por ejemplo) del censo de Estados Unidos del año 1990, disponibles en Daskin (1995, 2013).

Además, para estudiar el impacto de los costes de inventario y transporte en el modelo, variaremos los valores de β y θ , los cuales representan los pesos asociados a los costes de transporte e inventario respectivamente. Obsérvese que cuando θ es mayor que β la complejidad del problema aumenta, pues los términos no lineales de la función objetivo adquieren mayor importancia.

Para los parámetros usados en todos los experimentos numéricos de este capítulo, tomaremos d_{ij} la distancia *great circle* (o distancia ortodrómica) calculada a partir de las coordenadas de los puntos que aparecen en Daskin (1995, 2013), $F_j, g_j = 10$, $a_j = 5$, $\forall j \in J$; $h, \chi, L_j = 1, \forall j \in J$, $\alpha = 0.975$, y $z_\alpha = 1.96$. Para el conjunto de datos de 88 nodos tomaremos concretamente como f_j los valores de Daskin (1995) divididos por 100 y como μ_i los valores de la demanda 1 disponibles en Daskin (1995, 2013) divididos por 1000. Para el conjunto de datos de 150 nodos, tomaremos para cada $j \in J$, $f_j = 100$ y como valores de μ_i los valores disponibles en Daskin (1995, 2013) divididos por 1000. En ambos casos, tomaremos $\sigma_i^2 = \mu_i$ con el fin de comparar los resultados con los existentes en la literatura.

Todos los experimentos numéricos se han implementado en un PC Intel (R) Core (TM) i7-4790 CPU @ 3.60 GHz con 16,0 GB de RAM y sistema operativo Windows 7 Professional. En este trabajo nos hemos decantado por usar el software Xpress y el lenguaje de modelización y programación Mosel, para utilizar los mismos a través del entorno disponible en Windows Xpress-IVE. Se ha empleado la versión del programa 8.0 de 64 bits. De los diferentes módulos de optimización disponibles, nos hemos decantado por el módulo de optimización no lineal, FICO Xpress-NonLinear y XSLP (basado en aproximación lineal sucesiva), lanzado al mercado en 2014. A pesar de disponer de este solver de última generación, al aumentar el tamaño y complejidad del problema, el solver necesita cada vez más memoria, es ineficaz en términos de tiempo, y es incapaz de resolver el problema. Esta es una de las razones de necesitar métodos alternativos a la resolución exacta, como la relajación Lagrangiana.

La tabla que aparece a continuación, resume los diferentes experimentos numéricos realizados variando los parámetros β y θ para los casos de 88 nodos y 150 nodos. En la cuarta columna se muestra el valor objetivo obtenido en la literatura anteriormente, el cual sirve para comparar con el valor objetivo que nosotros obtenemos por medio de nuestro algoritmo de relajación Lagrangiana. Además, se muestran los tiempos de computación obtenidos por otros métodos y el promedio de los mismos, para poder compararlos con los obtenidos en este trabajo con nuestro algoritmo. Como ejemplo, representamos en la figura 2.2 la estructura de la red (espacial) de la cadena de suministro correspondiente a 88 nodos, $\beta = 0.001$ y $\theta = 0.1$.

Shen <i>et al.</i> (2003) set covering			Atamtürk <i>et al.</i> (2012) form. cónica			Relajación Lagrangiana		
β	θ	Valor Objetivo	Tiempo (s.)	Tiempo (s.)	Tiempo (s.)	Nº de CD	Valor Objetivo	desv. (%)
88	0.001	0.1	13229.55	107	1	9	13229.55	0
88	0.002	0.1	19975.37	24	1	11	19975.4	0
88	0.003	0.1	25306.68	7	2	14	25307.5	0.003
88	0.004	0.1	28752.64	3	2	19	28752.6	0
88	0.005	0.1	31390.69	1	2	18	31394.8	0.01
88	0.001	0.1	13229.55	107	1	9	13229.55	0
88	0.002	0.2	20491.17	38	1	10	20491.2	0
88	0.005	0.5	33794.94	2	3	17	33865.7	0.2
88	0.005	0.1	31390.69	1	2	18	31394.8	0
88	0.005	0.5	33794.94	2	3	17	33865.7	0.2
88	0.005	1	35876.10	3	2	16	36099.7	0.6
88	0.005	5	47348.38	7	5	12	48342.1	2.1
88	0.005	10	57959.54	14	10	11	61017.7	5.3
88	0.005	20	74760.97	51	30	7	80565.2	7.7
150	0.0004	0.01	3977.27	226	12	15	3981.19	0.099
150	0.0006	0.01	4867.76	89	8	21	4867.79	0.099
150	0.0008	0.01	5580.71	49	11	25	5587.96	0.13
150	0.001	0.01	6162.99	26	10	25	6165.79	0.045
150	0.0005	0.01	4459.32	132	10	19	4470.96	0.26
150	0.001	0.02	6410.09	26	11	23	6417.46	0.11
150	0.002	0.04	8988.62	7	23	25	9000.73	0.13
150	0.001	0.01	6162.99	26	10	25	6165.79	0.045
150	0.001	0.1	7508.49	40	24	20	7569.35	0.81
150	0.001	0.5	10175.71	78	56	12	10500.5	3.19
150	0.001	1	12380.63	260	185	11	12797.6	3.37
Promedio			53.04	17			7.40	

A partir de la tabla, observamos que el número de CD crece cuando aumenta el coste de transporte y decrece cuando el holding cost aumenta. Esto no es un hecho sorprendente, ya que es lo que cabe esperar y es un modo sencillo para validar el modelo. Además, se observa que cuando β y θ aumentan, se tiene que el número de CD aumenta también.

En Escalona *et al.* (2015) aparece "*Atamtürk et al. (2012) show, through a computational study, that the conic formulation outperforms the column generation and Lagrangian based methods considered up to now*". En este trabajo, comprobamos que la relajación Lagrangiana es notablemente superior a la formulación cónica de Atamtürk *et al.* (2012) y su procedimiento de resolución. De un modo general, observamos a partir de los datos de la tabla, que nuestro algoritmo de relajación Lagrangiana encuentra en un tiempo razonable la solución. De hecho, *nuestro algoritmo es un 130 % más rápido (en media) que el de Atamtürk et al. (2012) y además, para el problema más difícil (el problema con 150 nodos, $\beta = 0.001, \theta = 1$), nuestro algoritmo de resolución en menos de 10 segundos proporciona una solución con un error relativo de alrededor del 3 %, un tiempo razonablemente menor que el obtenido por Atamtürk et al. (2012) que son 185 segundos.*

2.10. Conclusiones

En este capítulo se ha estudiado en profundidad un modelo integrado de Localización-Inventario, estudiando las diferentes formulaciones y algoritmos de resolución propuestos en la literatura. Además, se ha formulado rigurosamente el modelo y se ha considerado una extensión del mismo, y en concreto, se ha corregido un error acerca de la política (r, Q) y su aproximación, que aparecía frecuentemente en la literatura; además de admitir cualquier distribución de probabilidad continua (con esperanza finita) no necesariamente normal para expresar la demanda. Hemos propuesto un algoritmo de relajación Lagrangiana basado en el existente en You y Grossmann (2008) mejorando notablemente algunos aspectos del mismo, con el objetivo de obtener mejores soluciones. También, a pesar de que Atamtürk *et al.* (2012) y Escalona *et al.* (2015) señalan que la formulación cónica de Atamtürk *et al.* (2012) es superior a la relajación Lagrangiana, hemos comprobado que esto no es cierto, ya que nuestro algoritmo de relajación Lagrangiana es notablemente superior en tiempo de computación en problemas de tamaño y complejidad grandes.

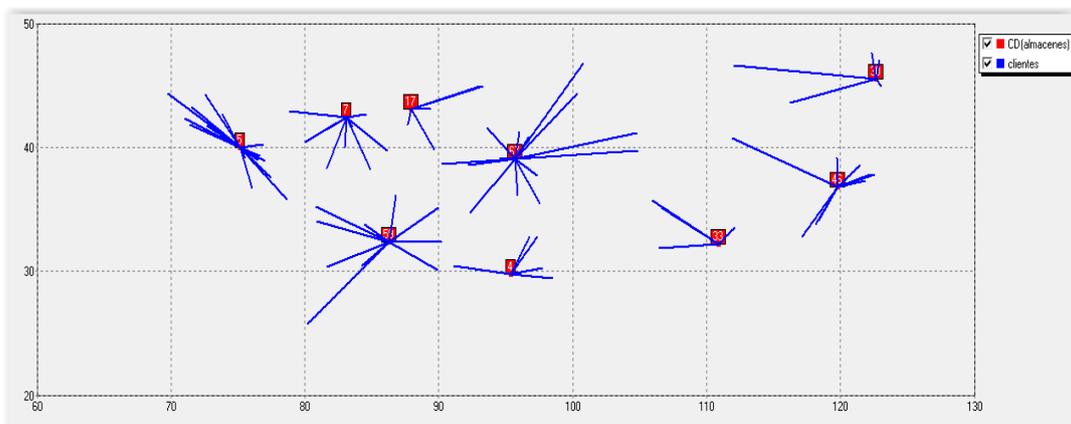


Figura 2.2: Estructura de la red (espacial) de la cadena de suministro correspondiente a 88 nodos, $\beta = 0.001$ y $\theta = 0.1$.

Capítulo 3

El problema de la p -mediana con restricciones de distancia máxima

3.1. Introducción

El problema de la p -mediana, descrito inicialmente por Hakimi (1965) y formulado matemáticamente después por ReVelle y Swain (1970), es un problema NP-duro, clásico en Teoría de Localización. Aunque la literatura acerca del mismo es bastante amplia, el problema de la p -mediana con restricciones de distancia máxima está poco tratado en la misma. Es posible considerar dos enfoques al formular el problema de la p -mediana con restricciones de distancia máxima: el *enfoque indirecto*, introduciendo un valor M suficientemente grande en la matriz de distancias si la distancia (o tiempo) entre un punto de demanda y un punto-candidato de instalación excede una distancia (o tiempo) máxima, y el *enfoque directo*, introduciendo una serie de restricciones de distancia (o tiempo) máxima (en adelante, salvo que se especifique lo contrario consideraremos distancia y tiempo indistintamente). Tal como señala Choi y Chaudry (1993), "*the performance of the indirect approach can greatly be dependent on the choice of the large number for big- M* ". Una aplicación de este problema a servicios públicos (localización de paradas de bus) puede verse en Murray y Wu (2003).

El problema de la p -mediana con restricciones de distancia máxima apareció por primera vez en Toregas *et al.* (1971). En Khumawala (1973,1975) aparece una heurística específica para este problema. En cambio, Hillsman y Rushton (1975) aplican la heurística de Teitz y Bart (1968). Rahman y Smith (1991) señalan "*With maximum distance constraints the performance of the Teitz and Bart method is even better*", refiriéndose a la heurística de Ardalan (1988) (éste último refleja la superioridad frente a la heurística de Khumawala (1973)). Choi y Chaudry (1993) aplican relajación Lagrangiana obteniendo soluciones óptimas en un tiempo razonable. Finalmente, Chaudry *et al.* (1994) corrigen las soluciones heurísticas (aplicando la heurística de Teitz y Bart (1968)) obtenidas por Rahman y Smith (1991), comparando con las soluciones óptimas obtenidas por relajación Lagrangiana en Choi y Chaudry (1993). Desde los años 90 el uso de las metaheurísticas en Investigación Operativa ha ido incrementándose con los años y, actualmente existen multitud de ellas para resolver problemas de gran complejidad en Optimización. Es por ello, que dado que en la literatura existente para este problema no se han aplicado metaheurísticas, consideramos adecuado atacar el problema aplicando las mismas.

3.2. Formulación

En esta sección, mostraremos dos formulaciones equivalentes del problema que tratamos en este capítulo: la *formulación densa* y la *formulación dispersa*.

Sea $M = \{1, 2, \dots, m\}$ un conjunto de puntos de demanda de cierto servicio, con demandas h_i , $i = 1, \dots, m$, y un conjunto $N = \{1, 2, \dots, n\}$ de puntos donde es posible situar o abrir una instalación (puntos de servicio). Supondremos que conocemos la matriz de distancias de cada punto de demanda i a cada punto de instalación j , (d_{ij}). Además, para cada $i \in M$, sea s_i la distancia máxima entre el punto de demanda i y cualquier instalación.

Dado un número total de instalaciones p que van a ser abiertas, el problema de la p -mediana con restricciones de distancia máxima consiste en decidir en qué puntos deben ser abiertas las instalaciones y asignar a cada punto de demanda a una instalación abierta que cumpla unas restricciones de distancia máxima, de forma que la distancia total (promediada con la población) que se atraviesa para servir a toda la demanda sea mínima.

3.2.1. Formulación densa

La formulación densa, añade explícitamente unas restricciones de distancia máxima a la formulación del problema de la p -mediana. Más concretamente, el modelo que presentamos es el siguiente

$$\text{(PMPDC1) Minimizar } \sum_{i=1}^m \sum_{j=1}^n h_i d_{ij} y_{ij} \quad (3.1)$$

$$\text{sujeta a } \sum_{j=1}^n y_{ij} = 1, \quad i = 1, \dots, m \quad (3.2)$$

$$\sum_{j=1}^n d_{ij} y_{ij} \leq s_i, \quad i = 1, \dots, m \quad (3.3)$$

$$y_{ij} \leq x_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n \quad (3.4)$$

$$\sum_{j=1}^n x_j = p, \quad (3.5)$$

$$x_j \in \{0, 1\}, \quad j = 1, \dots, n \quad (3.6)$$

$$y_{ij} \in \{0, 1\}, \quad i = 1, \dots, m, \quad j = 1, \dots, n \quad (3.7)$$

Aquí, m es el número de puntos de demanda, n es el número de puntos-candidato de posibles localizaciones de las instalaciones, p es el número de instalaciones que se abrirán, h_i es la demanda del punto i , d_{ij} es la distancia entre el punto de demanda i al punto-candidato de instalación j , s_i es la distancia máxima entre el punto de demanda i y cualquier instalación, $x_j \in \{0, 1\}$ es una variable de localización, definida para cada $j = 1, \dots, n$, con $x_j = 1$ si el punto j se selecciona como instalación, y 0 en otro caso. Finalmente, $y_{ij} \in \{0, 1\}$ es una variable de asignación, definida para todos $i = 1, \dots, m$, $j = 1, \dots, n$ siendo $y_{ij} = 1$ si el punto de demanda i es asignado a la instalación j , y 0 en otro caso.

En la función objetivo (3.1) se minimiza la distancia total que se va a atravesar por toda la población. Las restricciones (3.2) aseguran que todos los puntos de demanda van a quedar asignados a una instalación, con las restricciones (3.3) se imponen las condiciones de distancia máxima, y con las restricciones (3.4) nos aseguramos que si una instalación no está abierta, no puede servir a ningún punto de demanda. Con la restricción (3.5) nos aseguramos que se abrirán exactamente p instalaciones. Finalmente, las restricciones (3.6) y (3.7) se refieren al carácter binario de las variables de decisión.

La formulación densa presenta un mayor número de restricciones que el enfoque indirecto, el cual está basado en modificar la matriz de distancias introduciendo en la misma un valor M suficientemente grande si la distancia entre un punto de demanda y un punto-candidato de instalación excede una distancia máxima y suprimir la restricción (3.3). Más concretamente:

$$d'_{ij} = \begin{cases} d_{ij}, & \text{si } d_{ij} \leq s_i \\ M, & \text{si } d_{ij} > s_i \end{cases}$$

El problema de la p -mediana con restricciones de distancia máxima usando este enfoque indirecto puede parecer a priori un problema fácil de resolver, pues se reduce a un problema de p -mediana clásico con la matriz de distancias modificada (esto es quizá una de las razones por las que no esté muy estudiado el mismo en la literatura existente), aunque sorprendentemente las heurísticas proporcionan problemas y frecuentemente soluciones no factibles, pues la calidad de las soluciones dependen en gran medida del valor M (Choi y Chaudry (1993)).

3.2.2. Formulación dispersa

A continuación, presentaremos una segunda formulación, que es la que resulta más adecuada de cara a los procedimientos de resolución. Además, a la notación expuesta anteriormente añadimos la siguiente. Definamos N_i como el conjunto de instalaciones a distancia menor que s_i unidades del punto i .

NOTACIÓN	
M	conjunto de puntos de demanda.
N	conjunto de puntos-candidato de posibles localizaciones de las instalaciones.
h_i	demanda del punto i .
p	número de instalaciones que se abrirán.
d_{ij}	distancia entre el punto de demanda i al punto-candidato de instalación j .
s_i	distancia máxima entre el punto de demanda i y cualquier instalación.
N_i	conjunto de instalaciones a distancia menor que s_i unidades del punto i , con $N_i \subseteq N$.
VARIABLES DE DECISIÓN	
x_j	= 1, si el punto j se selecciona como instalación, y 0 en otro caso.
y_{ij}	= 1 si el punto de demanda i es asignado a la instalación j , y 0 en otro caso.

El modelo que presentamos es el siguiente

$$\text{(PMPDC2) Minimizar } \sum_{i=1}^m \sum_{j \in N_i} h_i d_{ij} y_{ij} \quad (3.8)$$

$$\text{sujeta a } \sum_{j \in N_i} y_{ij} = 1, \quad i = 1, \dots, m \quad (3.9)$$

$$y_{ij} \leq x_j, \quad i = 1, \dots, m, \quad j \in N_i \quad (3.10)$$

$$\sum_{j=1}^n x_j = p, \quad (3.11)$$

$$x_j \in \{0, 1\}, \quad j = 1, \dots, n \quad (3.12)$$

$$y_{ij} \in \{0, 1\}, \quad i = 1, \dots, m, \quad j = 1, \dots, n \quad (3.13)$$

En la función objetivo (3.8) se minimiza la distancia total que se va a atravesar por toda la población. Las restricciones (3.9) aseguran que todos los puntos de demanda va a quedar asignado a una instalación dentro de unos valores de distancia máxima s_i asociados a los mismos, y con las restricciones (3.10) nos aseguramos que si una instalación no está abierta, no puede servir a ningún punto de demanda (teniendo en cuenta las restricciones de distancia máxima). Con la restricción (3.11) nos aseguramos que se abrirán exactamente p instalaciones. Finalmente, las restricciones (3.12) y (3.13) se refieren al carácter binario de las variables de decisión.

Es importante notar que para algún par (p, \mathbf{s}) , con $\mathbf{s} = (s_1, \dots, s_m)$, puede no existir solución factible. Además, dado p fijo, si se aumentan los valores de las distancias máximas, existe un valor \bar{s} límite del cual la solución del problema (PMPDC2) no diferirá de la solución del problema de la p -mediana clásico (PMP). En la práctica es usual considerar que $s_i = s$ para todo $i = 1, \dots, m$. Es por ello, que más adelante asumiremos esta hipótesis.

3.3. Relajación Lagrangiana

La relajación Lagrangiana es una metodología bien conocida para resolver problemas de optimización combinatoria de gran tamaño y complejidad. Se basa en explotar la estructura inherente de cada problema con el objetivo de obtener cotas inferiores y cotas superiores sobre el valor óptimo del problema (Guignard (2003)).

Además, relajaremos (3.9) con el fin de obtener un problema sencillo de resolver. Nótese que en el procedimiento de relajación Lagrangiana necesitamos: ser capaces de resolver fácilmente los subproblemas, un método para obtener soluciones factibles (una heurística Lagrangiana, ver Avella *et al.* (2007) o Daskin (2013) para heurísticas Lagrangianas para el problema (PMP)) y un procedimiento para actualizar los multiplicadores (procedimiento de optimización subgradiente, ver Frangioni *et al.* (2017)).

El procedimiento de relajación Lagrangiana que aplicamos aquí es novedoso, y difiere del existente en Choi y Chaudry (1993), ya que ellos recaen en un subproblema de tipo mochila y nosotros sólo necesitamos resolver un problema trivial. La razón de esto, es que hemos sido capaces de adaptar el procedimiento de relajación Lagrangiana para el problema de la p -mediana clásico (PMP) (Daskin (2013)) al problema (PMPDC2).

$$\begin{aligned}
 \text{(PMPDC2}\lambda) \text{ Minimizar } & \sum_{i=1}^m \sum_{j \in N_i} h_i d_{ij} y_{ij} + \sum_{i=1}^m \lambda_i \left[1 - \sum_{j \in N_i} y_{ij} \right] \\
 & = \sum_{i=1}^m \sum_{j \in N_i} (h_i d_{ij} - \lambda_i) y_{ij} + \sum_{i=1}^m \lambda_i
 \end{aligned} \tag{3.14}$$

$$\text{sujeta a } y_{ij} \leq x_j, \quad i = 1, \dots, m, \quad j \in N_i \tag{3.15}$$

$$\sum_{j=1}^n x_j = p, \tag{3.16}$$

$$x_j \in \{0, 1\}, \quad j = 1, \dots, n \tag{3.17}$$

$$y_{ij} \in \{0, 1\}, \quad i = 1, \dots, m, \quad j = 1, \dots, n \tag{3.18}$$

Para valores fijos de los multiplicadores, λ_i , queremos minimizar la función objetivo (3.14). Nótese que el término $\sum_{i=1}^m \lambda_i$ es constante. A continuación, se resume el algoritmo propuesto de relajación Lagrangiana para resolver el problema de la p -mediana con restricciones de distancia máxima.

- **Paso 1 (Inicialización):** Inicializar $UB := +\infty$ (cota superior), $LB := -\infty$ (cota inferior), $k := 0$ y tomar unos valores iniciales para los multiplicadores. Nosotros hemos optado por establecer como LB inicial el valor objetivo de la solución óptima de la relajación lineal del problema y como multiplicadores iniciales las variables duales asociadas a la solución óptima de la relajación lineal.

- **Paso 2** (Resolución de los subproblemas) : En la iteración k -ésima, fijado λ^k , calcularemos

$$V_j = \sum_{i \in M / j \in N_i} \min \{0, h_i d_{ij} - \lambda_i\}, \quad j \in N$$

Encontrar los p menores valores de V_j y poner $x_j(\lambda^k) = 1$ en los correspondientes puntos candidato de servicio y $x_j(\lambda^k) = 0$ en otro caso. Establecer $y_{ij} = 1$ si $x_j(\lambda^k) = 1$ y $h_i d_{ij} - \lambda_i < 0$ y $y_{ij} = 0$ en otro caso. Calcular

$$V(\lambda^k) = \sum_{i=1}^m \sum_{j \in N_i} (h_i d_{ij} - \lambda_i) y_{ij} + \sum_{i \in I} \lambda_i^k \quad (3.19)$$

Si $V(\lambda^k) > LB$, entonces $LB := V(\lambda^k)$.

- **Paso 3** (Obtención de soluciones factibles: Heurística Lagrangiana): Fijados los p valores de las variables $x_j(\lambda^k)$, para cada $i = 1, \dots, m$ encontrar

$$\hat{j}_i = \arg \min \{d_{ij} \mid x_j = 1, j \in N\}$$

Establecer $y_{ij}(\lambda^k) = 1$ si $j = \hat{j}_i$ y $y_{ij}(\lambda^k) = 0$ en otro caso.

Calcular

$$\bar{V}(\lambda^k) = \sum_{i=1}^m h_i d_{i\hat{j}_i}$$

Si $\bar{V}(\lambda^k) < UB$, entonces $UB := \bar{V}(\lambda^k)$.

- **Paso 4** (Iteración subgradiente): Calcular el subgradiente ξ_i usando

$$1 - \sum_{j \in N_i} y_{ij}, \quad i \in M.$$

Calcular la longitud de paso

$$T_k := \frac{\theta_{sub}(UB - LB)}{\sum_{i=1}^m \left(1 - \sum_{j \in N_i} y_{ij}\right)^2}$$

Actualizar los multiplicadores:

$$\lambda^{k+1} := \max \{0, \lambda^k + T_k \xi^k\}$$

Además, si en más de un número prefijado de iteraciones (pasadas) la cota inferior LB no mejora, entonces $\theta_{sub} := \theta_{sub}/2$. En nuestros experimentos numéricos, nosotros hemos optado por establecer 18 pasadas.

- **Paso 5** (Criterio de parada): Si el salto de dualidad ($GAP = \frac{UB-LB}{UB}$) es menor que una tolerancia prefijada, o si el número de iteraciones máximas se ha alcanzado o si la norma del subgradiente o θ_{sub} son menores que unas tolerancias prefijadas, PARAR, y establecer UB como valor objetivo, y $x_j(\lambda^k)$ y $y_{ij}(\lambda^k)$ como la solución final obtenida. Si no, $k := k + 1$ e ir al paso 2.

3.4. GRASP

GRASP (*Greedy Randomized Adaptive Search Procedure*) es una metaheurística que funciona bastante bien para el problema de la p -mediana clásico. La adaptación de la misma a nuestro problema de p -mediana con restricciones de distancia máxima no es del todo inmediata, ya que influyen las restricciones de distancia máxima en la fase constructiva del GRASP y pueden obtenerse soluciones no factibles. Es por ello, que es necesario implementar dos fases en la fase constructiva del GRASP. Nótese que el GRASP que aquí se ha implementado requiere dos parámetros: KRCL (número en lista restringida de candidatos o cardinal de la lista restringida de candidatos) e `iter_max` (número de iteraciones máximas del procedimiento). Además, supondremos conocidos el resto de parámetros propios del problema de la p -mediana con restricciones de distancia máxima y asumiremos que la distancia máxima es $s_i = s$ para todo $i = 1, \dots, m$. A continuación, se explican las dos fases básicas en una iteración del GRASP:

- **FASE CONSTRUCTIVA:** Es un greedy aleatorizado, descompuesto el mismo en dos fases claramente diferenciadas. Más concretamente, se aleatoriza (usando una lista restringida de candidatos de cardinal KRCL) el procedimiento greedy siguiente:

Fase 1 : Se aplica la *heurística greedy de set covering* (ver Murty (1995) o Chvatal (1979)), donde se minimiza el número de instalaciones, y tomando como distancia de cubrimiento $dc = s$. Al finalizar la misma se obtiene un número k de instalaciones. Pueden darse los siguientes casos:

- Si $k > p$, previsiblemente el problema es no factible. En nuestros experimentos numéricos siempre que se ha dado este caso el problema era realmente no factible.
- Si $k = p$, se obtiene ya de hecho una solución factible.

- Si $k < p$, ir a la Fase 2.

Fase 2 : Dadas $k < p$ instalaciones, se completa hasta p instalaciones, con el objetivo original (3.1). Más concretamente, en una iteración cualquiera de esta fase 2 para poder completar hasta p instalaciones, sea X_k el conjunto que contiene las localizaciones asignadas a las k instalaciones. Sea $d(i, X_k)$ la distancia mínima entre el punto de demanda i y el punto más cercano perteneciente al conjunto X_k . De modo similar, sea $d(i, j \cup X_k)$ la distancia mínima entre el punto de demanda i y el punto más cercano, que pertenece al conjunto X_k aumentado por la nueva localización j . El mejor sitio para localizar una única instalación, dado que las primeras k instalaciones se localizan en sitios del conjunto X_k , es la localización j que minimiza $Z_j = \sum_i h_i d(i, j \cup X_k)$. Este procedimiento se repite hasta completar hasta p instalaciones, en cuyo caso PARAR.

- **BÚSQUEDA LOCAL:** Para el caso del problema de la p -mediana clásico, el conocido algoritmo de búsqueda local fue introducido por Teitz y Bart (1968). Se conoce también como *heurística de intercambios* o *heurística swap*, pues en cada iteración intercambia un elemento que está en la solución (punto de servicio abierto), por uno que no esté en la solución (punto de servicio cerrado). La heurística de Teitz y Bart (1968) fue mejorada posteriormente por Whitaker (1983) y conocida en la literatura como *heurística rápida de Whitaker*. Finalmente, Resende y Werneck (2003) proponen un procedimiento de implementación que mejora significativamente la heurística rápida de Whitaker (1983). Aquí se ha implementado la heurística rápida de Whitaker con la *implementación de Resende y Werneck* (ver Resende y Werneck (2003)) adaptando la misma al problema de la p -mediana con restricciones de distancia máxima.

3.5. Resultados computacionales

En esta sección presentamos los resultados computacionales de los experimentos numéricos realizados. Con el fin de comparar los diferentes métodos de resolución, consideraremos dos conjuntos de datos y diferentes problemas test:

- El primer conjunto de datos, está formado por problemas test cuyos parámetros provienen de datos reales sobre cierto servicio sanitario: aint11, aint12 y aint13 con $m = 134$, $n = 121$; $m = 266$, $n = 21$ y $m = 430$, $n = 22$ respectivamente, siendo m el número de puntos de demanda y n el número de puntos-candidato de servicio. Además, tomaremos $s_i = s$ para todo $i = 1, \dots, m$.
- El segundo conjunto de datos ha sido generado aleatoriamente, para obtener problemas test de mayor tamaño que el anterior. Más concretamente, son datos con $n = m$ de tamaños 500, 800 y 1000. Denotaremos a estos problemas test como s500_1, s800_1 y s1000_1. Además, las demandas han sido generadas aleatoriamente siguiendo una distribución uniforme entre 10 y 100 y las distancias han sido obtenidas redondeando las distancias reales euclídeas calculadas a partir de las coordenadas generadas uniformemente entre 0 y 100. Además, tomaremos $s_i = s$ para todo $i = 1, \dots, m$.

3.5.1. Intervalos de factibilidad de distancia máxima

Para analizar la factibilidad de los problemas y su relación con los parámetros p y s nos interesa calcular, el mínimo valor de s que hace factible el problema el cual llamaremos *umbral de factibilidad de distancia máxima* y lo representaremos por \underline{s} . Además, al valor máximo que hace que la solución de (PMPDC2) coincida con la solución de (PMP) lo llamaremos *umbral de invariancia de distancia máxima* y lo representaremos por \bar{s} . Al intervalo cuyos extremos son el umbral de factibilidad de distancia máxima y el umbral de invariancia de distancia máxima lo denominaremos *intervalo de factibilidad de distancia máxima* y lo representaremos por $[\underline{s}, \bar{s}]$.

Además, fijado un valor $s \in [\underline{s}, \bar{s}]$, al subintervalo (puede degenerarse al punto s) donde la solución es constante e igual a la solución asociada a s , lo denominaremos *subintervalo de invariancia (asociado a s)* y lo representaremos por $\text{inv}(s)$. Nótese que el extremo inferior de cualquier subintervalo de invariancia coincide con la distancia máxima de cualquier punto de demanda a su instalación asignada en el óptimo (tal valor lo representaremos por d_{max}^* , el cual depende de s).

Nótese que la máxima distancia mínima (depende de los datos de las distancias y no del valor de p) puede no coincidir con el umbral de factibilidad de distancia máxima (que depende de p).

Fijado p , nos interesa saber cuánto y cómo varía el valor objetivo (de la solución óptima el cual representaremos por z (es la distancia total)) variando el valor de la distancia máxima en el intervalo de factibilidad de distancia máxima. En las figuras 3.1, 3.2 se muestra un gráfico con algunos de estos conceptos, correspondiente al conjunto de datos aint13 con $p = 7$ y $p = 10$. Otro ejemplo puede verse en la figura 3.3, correspondiente al conjunto de datos aint12 con $p = 7$. Puede verse, que aquí hay más subintervalos de invariancia con respecto a los ejemplos de las figuras 3.1 y 3.2.

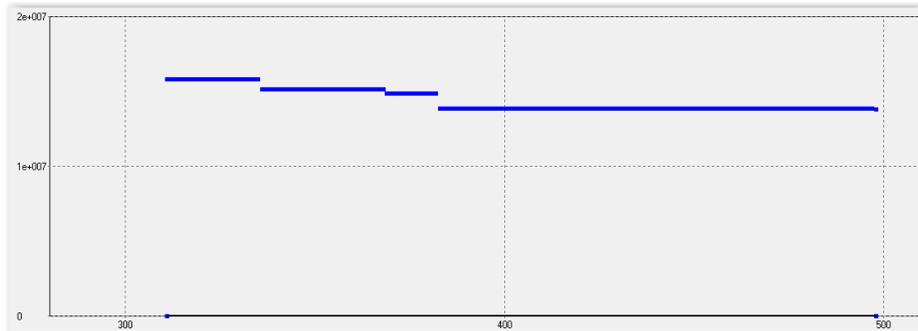


Figura 3.1: z (valor objetivo, distancia total) frente a distancia máxima, correspondiente al conjunto de datos aint13 con $p = 7$. El intervalo de factibilidad de distancia máxima es $[311, 498]$. Por ejemplo, observamos que abriendo $p = 7$ instalaciones, reduciendo en $187 = 498 - 311$ unidades la distancia máxima la distancia total aumenta un 14.48 %.

- Nótese que conforme aumenta p , tanto el umbral de factibilidad de distancia máxima como el umbral de invariancia de distancia máxima se mantienen constantes o disminuyen su valor.
- El número de subintervalos de invariancia disminuye conforme aumenta p .
- Si $p = n$ (es decir, si se abren todas las instalaciones disponibles) se tiene que $s = \text{inv}(s) = d_{max}^*$. A este valor lo denominaremos *umbral límite de factibilidad-invariancia*.
- A partir de un cierto p hasta un cierto $p^* \leq n$ los intervalos de factibilidad coinciden. Además, conforme aumenta p , existe un punto $p^* \leq n$ tal que los intervalos de factibilidad de distancia máxima degeneran en el umbral límite de factibilidad-invariancia.
- Si conforme aumenta p los valores de \underline{s} coinciden, se tiene que el umbral límite de factibilidad-invariancia es \underline{s} .

Para ver la variación de la función objetivo cuando varía p , en la figura 3.4 se muestra para el conjunto de datos aint11, se representa el valor objetivo frente a la variación de p fijado $s = 504$.

En la figura 3.5 se ha dibujado para cada valor de p y s que hicieran factible el problema (correspondiente al conjunto de datos aint13), la distancia promedio (solución óptima dividida entre demanda total). En la figura 3.6 se muestra el plano p-distancia promedio asociado con la figura 3.5.

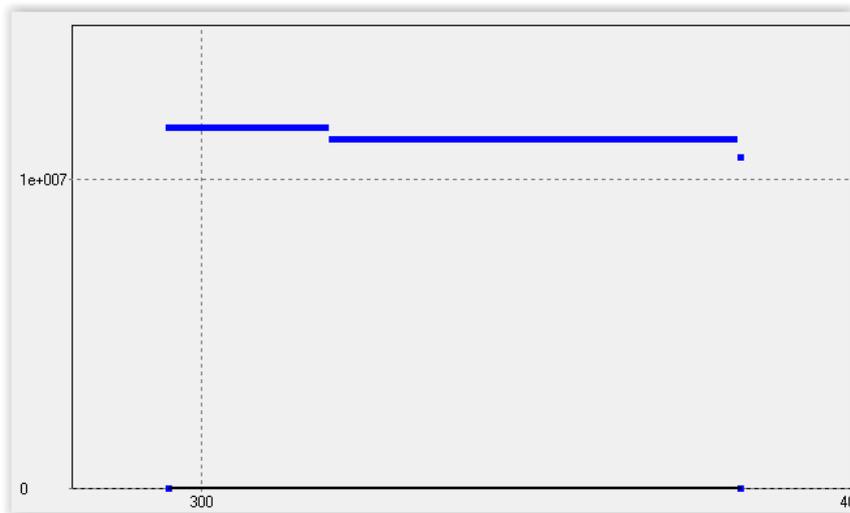


Figura 3.2: z (valor objetivo, distancia total) frente a distancia máxima, correspondiente al conjunto de datos aint13 con $p = 10$. El intervalo de factibilidad de distancia máxima es $[295, 383]$

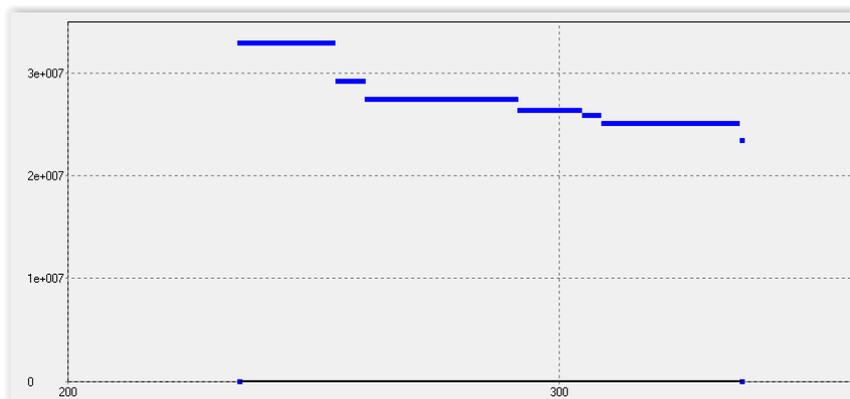


Figura 3.3: z (valor objetivo, distancia total) frente a distancia máxima, correspondiente al conjunto de datos aint12 con $p = 7$. El intervalo de factibilidad de distancia máxima es $[235, 337]$

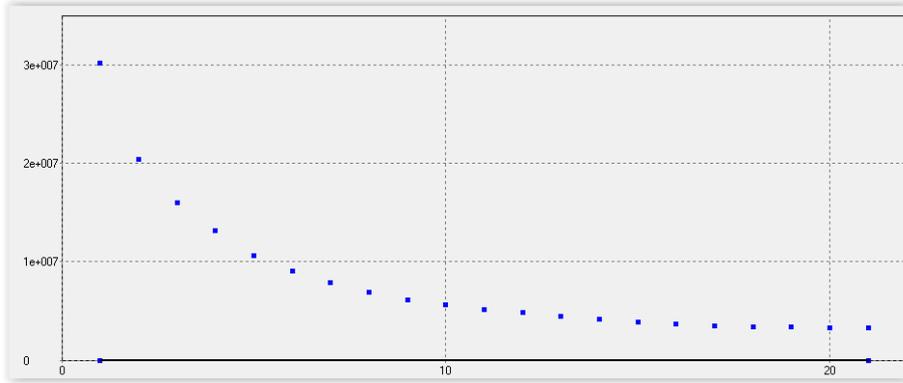


Figura 3.4: z (valor objetivo, distancia total) frente a p fijado $s = 504$, correspondiente al conjunto de datos aint11.

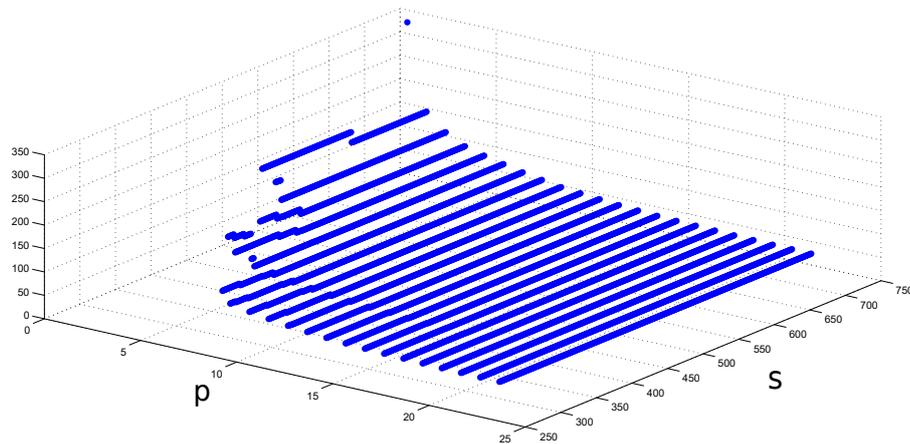


Figura 3.5

El estudio de los intervalos de factibilidad de distancia máxima ayuda al decisor a elegir un valor de p adecuado sujeto a unas restricciones de distancia (o tiempo de servicio) máxima (o viceversa) ya que en la práctica, hay situaciones en las que se prefija unos valores de p , otras en las que se prefija unos tiempos de servicio o distancia máxima o incluso ambas. Es necesario un equilibrio entre los valores de p y los valores de distancia máxima, ya que si se necesitan tiempos de servicio bajos cabe esperar que será necesario un número p más alto de instalaciones (y al revés, si se necesitan valores de p pequeños, cabe esperar que para que haya factibilidad de soluciones se necesitan valores altos de distancia máxima). Además, podemos calcular fijado un valor de p , cuanto aumentaría el coste si disminuyéramos la distancia máxima hasta el umbral de factibilidad.

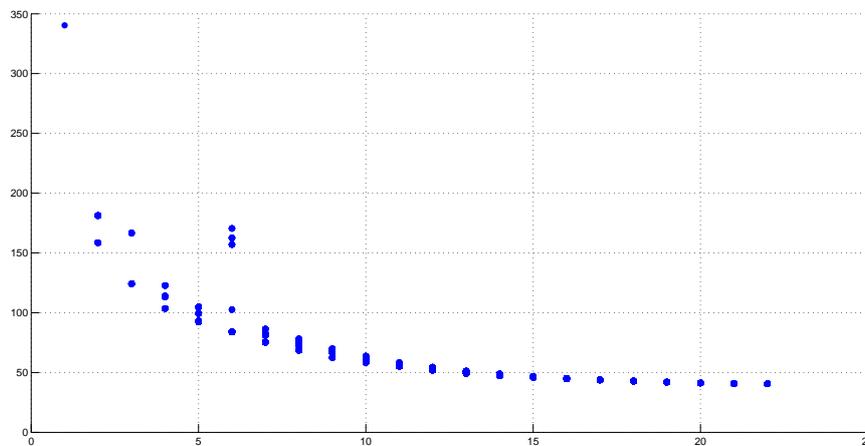


Figura 3.6: En este plano, podemos ver para cada p , el número de subintervalos de invariancia

3.5.2. Análisis de los resultados computacionales

La tabla que se adjunta, muestra los resultados computacionales de los diferentes experimentos numéricos realizados en un PC Intel (R) Core (TM) i7-4790 CPU @ 3.60 GHz con 16,0 GB de RAM y sistema operativo Windows 7 Professional. El procedimiento de relajación Lagrangiana ha sido implementado en Xpress-Mosel (versión del programa 8.0 de 64 bits), al igual que la resolución óptima del modelo que ha sido resulta por XPRESS-Optimizer (presenta una de las más eficientes implementaciones del conocido algoritmo Branch & Cut, e incluye entre otras cosas, numerosos métodos de planos de corte, heurísticas y métodos de preprocesado). El GRASP se ha implementado en C con el fin de explotar la potencia de esta metaheurística. Además, para la misma, hemos establecido 50 iteraciones y 8 como el número en lista restringida de candidatos.

En la primera columna aparecen los parámetros (p, s) , encabezando la columna el conjunto de datos asociado a los experimentos numéricos. Las columnas etiquetadas con z, z^{LR} y z^G muestran la distancia promedio obtenida por XPRESS, relajación Lagrangiana y GRASP respectivamente. Las columnas etiquetadas con t, t^{LR} y t^G muestran el tiempo de computación en segundos por XPRESS, relajación Lagrangiana y GRASP respectivamente. Con XPRESS, hemos calculado la distancia máxima en el óptimo, d_{max}^* y, para la relajación Lagrangiana y GRASP hemos calculado la desviaciones porcentuales de la distancia promedio obtenidas por relajación Lagrangiana y GRASP respectivamente, siendo d^{LR} y d^G las mismas.

Además, a modo de ejemplo, en la figura 3.7 aparece un gráfico representando las coordenadas de los puntos-candidato abiertos y los puntos de demanda asignados a los mismos, correspondientes a la solución del problema test s1000_1, con $p = 15$ y $s = 21$.

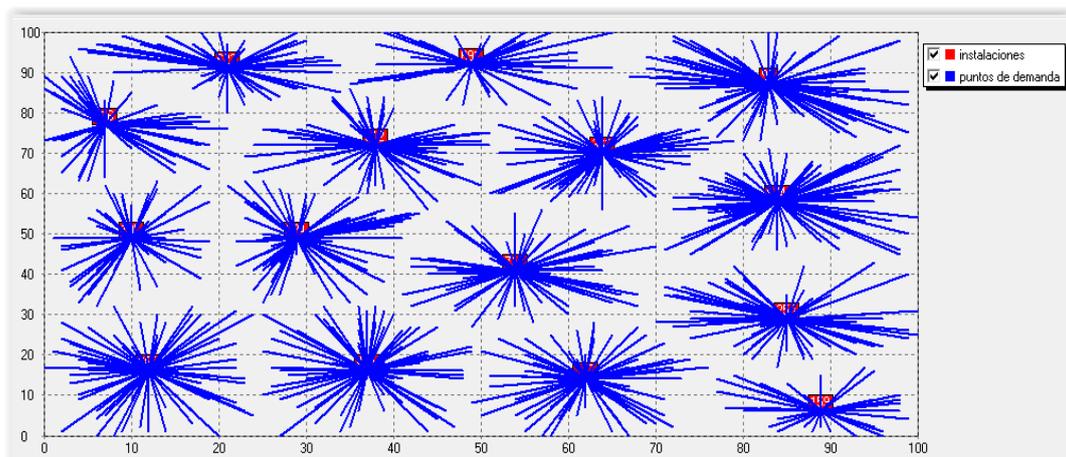


Figura 3.7

	XPRESS			Relajación Lagrangiana			GRASP		
	z	t	d_{max}^*	z^{LR}	t^{LR}	d^{LR}	z^G	t^G	d^G
aint11									
(p, s)									
(4, 300)	57.62	0.08	292	57.62	0.125	0	57.62	0.0066	0
(4, 350)	50.32	0.09	305	50.32	0.125	0	50.32	0.0066	0
(5, 300)	46.68	0.08	292	46.68	0.078	0	46.68	0.0067	0
(5, 350)	40.84	0.08	305	40.84	0.125	0	40.84	0.0070	0
(6, 300)	37.20	0.08	235	37.20	0.125	0	46.68	0.0073	0
(6, 350)	34.54	0.09	305	34.54	0.124	0	34.54	0.0072	0
(7, 220)	40.32	0.08	209	40.32	0.125	0	40.32	0.0077	0
(7, 230)	36.32	0.06	222	36.32	0.125	0	36.32	0.0059	0
aint12									
(p, s)									
(10, 250)	41.12	0.27	235	41.12	0.359	0	41.12	0.01443	0
(10, 300)	37.91	0.30	255	37.91	0.374	0	37.91	0.01384	0
(11, 250)	38.30	0.27	235	38.30	0.92	0	38.3	0.01477	0
(11, 300)	35.35	0.30	255	35.35	0.375	0	41.12	0.01452	0
(12, 250)	35.75	0.27	235	35.75	2.512	0	35.75	0.01598	0
(12, 300)	33.06	0.38	255	33.06	2.652	0	33.06	0.01599	0
aint13									
(p, s)									
(7, 380)	80.94	0.33	369	80.94	0.265	0	80.94	0.01244	0
(7, 400)	75.49	0.37	383	75.49	0.578	0	75.49	0.01253	0
(10, 300)	63.68	0.31	295	63.68	0.53	0	63.68	0.01656	0
(10, 350)	61.77	0.30	320	61.77	0.546	0	61.77	0.01408	0
(12, 300)	54.45	0.28	295	54.45	0.39	0	54.45	0.01467	0
(12, 400)	51.95	0.28	383	51.95	0.39	0	51.95	0.0154	0
s500_1									
(p, s)									
(15, 21)	9.20	15.44	20	9.20	28.61	0	9.20	5.975	0
(15, 50)	9.20	24.06	22	9.20	21.70	0	9.20	4.99	0
(20, 21)	7.73	11.25	17	8.05	26.29	4.13	7.73	4.314	0
(20, 50)	7.73	20.93	17	7.99	32.84	3.25	7.73	5.124	0
s800_1									
(p, s)									
(15, 21)	9.44	140.17	21	9.46	105.27	0.21	9.44	11.71	0
(15, 50)	9.44	293.76	21	9.47	122.98	0.32	9.44	15.83	0
(20, 21)	8.11	147.19	18	8.17	105.12	0.73	8.11	16.08	0
(20, 50)	8.11	185.43	18	8.18	113.90	0.86	8.11	17.13	0
s1000_1									
(p, s)									
(15, 21)	9.55	389.85	20	10.11	213.97	5.53	9.55	23.34	0
(15, 50)	9.55	544.34	22	9.67	234.98	1.24	9.55	27.08	0
(20, 21)	8.12	841.61	19	8.53	194.53	4.81	8.12	31.65	0
(20, 50)	8.12	1540.75	19	8.56	219.50	5.14	8.12	35.74	0

Podemos observar, que cuando crece el tamaño del problema, aumenta el tiempo de computación de XPRESS. Además, observamos que la relajación Lagrangiana con los problemas test aint11, aint12 y aint13 proporciona la solución óptima (además de salto de dualidad 0), aunque para los problemas test s500_1, s800_1 y s1000_1 proporciona soluciones que difieren porcentualmente de la solución óptima a lo sumo aproximadamente un 5 %.

Claramente, se observa que GRASP es, con diferencia, el mejor método de los tres, pues aquí siempre proporciona la solución óptima en muy breve tiempo.

3.6. Conclusiones

En este capítulo, se ha estudiado una extensión del problema clásico de la p -mediana, el problema de la p -mediana con restricciones de distancia máxima. La literatura acerca del mismo es escasa, y hasta la fecha, se había comprobado que para este problema (salvo mayor conocimiento), que el procedimiento heurístico de Teitz y Bart proporcionaba el mejor método de resolución. Es por ello, que dado que la metaheurística GRASP es quizá uno de los mejores procedimientos para resolver el problema de la p -mediana clásico hemos intentado aplicarla a este problema. Además, se ha realizado un estudio detallado acerca de la factibilidad del problema y su relación con los parámetros del mismo y se han propuesto dos nuevos métodos de resolución: relajación Lagrangiana, que difiere del existente en la literatura para este problema y un procedimiento basado en la metodología de la metaheurística GRASP. Finalmente, hemos comprobado que este último ha resultado de gran éxito en los experimentos numéricos realizados, superando a la relajación Lagrangiana y a XPRESS en tiempos de computación.

Capítulo 4

Programación Entera bi-objetivo. Aplicaciones a modelos de Localización

4.1. Introducción

Según Halpern (1978), "*In many real world problems the objective function is a mixture of the two different, possibly adverse objectives mentioned above*". Esta frase de Halpern, está asociada a lo siguiente: *en muchos problemas reales de Localización, existen dos objetivos contrapuestos: equidad y eficiencia*. Estos objetivos, están relacionados con dos problemas clásicos: el primero, con el problema del p -centro (minimiza la distancia máxima o el tiempo máximo de servicio), y se suele asociar clásicamente a la localización de servicios de emergencia; y el segundo, con el problema de la p -mediana (minimiza la distancia total que se atraviesa para servir toda la demanda), y se suele asociar a la localización de puntos de servicio o distribución de personas o mercancías.

Sin embargo, puesto que el problema de la p -mediana está basado en calcular un promedio, puede discriminar áreas remotas y de densidad de población o demanda baja, en contra de áreas o zonas situadas centralmente y caracterizadas por poseer una alta densidad de población o una gran cantidad de demanda, lo que implica ninguna equidad (aunque bastante eficiencia) (ver Ogryczak (1997)). Por otro lado, la localización de un cierto servicio a partir del problema del p -centro puede ocasionar una elevada distancia total, lo que significa ninguna eficiencia en el servicio.

Sin embargo, estos objetivos se suelen combinar. Un ejemplo de ello, es la localización de una serie de establecimientos comerciales (dígase p). Debe considerarse por tanto el problema del p -centro, pues es deseable que los establecimientos se encuentren cerca de los clientes, y también el problema de la p -mediana, pues se necesita un reparto rápido de las existencias.

La combinación convexa de las funciones objetivo asociadas a los problemas del p -centro y la p -mediana, se suele conocer como *función cent-dian*, debido a Halpern (1976) o también conocida como *función medi-center*, debido a Handler (1976).

En la literatura existente, el enfoque predominante de estudio del problema bi-objetivo del p -centro y la p -mediana es por medio de combinaciones convexas o métodos que ponderan ambos objetivos mediante unos pesos. La resolución de estos problemas, o bien no lleva a calcular todos los puntos eficientes, o bien pueden calcularse pero el éxito de encontrar todos los puntos eficientes depende en gran medida de los pesos y del tiempo excesivo de computación. Este trabajo parte de los estudios clásicos del problema del medi-center (o cent-dian), Daskin (2010, 2013) y de Daskin y Maass (2015).

4.2. Conceptos básicos de Programación bi-objetivo

Esta sección tiene el objetivo de presentar una serie de conceptos y resultados básicos de Programación Multiobjetivo necesarios para el cuerpo central del capítulo. Para ello, seguiremos Sáez-Aguado y Trandafir (2016). Para una referencia general dedicada a Programación Multiobjetivo puede consultarse Ehrgott (2005).

Consideremos el problema de Programación Entera bi-objetivo (BOIP)

$$\begin{array}{ll}
 \text{Minimizar} & f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x})) \\
 \text{(BOIP)} & \\
 & \text{s.a.} \quad \mathbf{x} \in X
 \end{array}$$

donde toda solución factible $\mathbf{x} = (x_1, \dots, x_n) \in X$ es tal que $x_j \geq 0$ es entero, para todo $j = 1, \dots, n$. Además, supondremos que las funciones objetivo son lineales de

la forma

$$f_1(\mathbf{x}) = \sum_{j=1}^n c_{1j}x_j$$

$$f_2(\mathbf{x}) = \sum_{j=1}^n c_{2j}x_j$$

donde $c_{ij} \in \mathbb{Z}$, para todo $i = 1, 2$ y todo $j = 1, 2$.

Sea $\mathbf{x}^* \in X$ una solución factible de **(BOIP)**. Entonces, diremos que \mathbf{x}^* es *eficiente* u *optimal Pareto* si no existe $\mathbf{x} \in X$ tal que $f_i(\mathbf{x}) \leq f_i(\mathbf{x}^*)$, $i = 1, 2$, donde la desigualdad estricta se da para algún $i \in \{1, 2\}$. Si no existe $\mathbf{x} \in X$ tal que $f_i(\mathbf{x}) < f_i(\mathbf{x}^*)$, $i = 1, 2$, entonces diremos que \mathbf{x}^* es *débilmente eficiente*.

Dado $\mathbf{y} = f(\mathbf{x}) \in Y = f(X) \subset \mathbb{R}^2$, diremos que \mathbf{y} es *no dominado* si \mathbf{x} es eficiente. Además, sea $X_E \subset X$ el conjunto de soluciones eficientes del problema bi-objetivo y sea $Y_E = f(X_E) \subset Y$ el conjunto de todos los puntos no dominados en el espacio objetivo, también conocido como *frontera eficiente* o *frontera de Pareto*. Supongamos ahora que existen N puntos no dominados y que $Y_E = \{\mathbf{y}^1, \dots, \mathbf{y}^N\}$ está indexado de forma que

$$y_1^1 < y_1^2 < \dots < y_1^N$$

y por lo tanto

$$y_2^1 > y_2^2 > \dots > y_2^N$$

En este caso, a los puntos \mathbf{y}^1 e \mathbf{y}^N se les denomina *puntos extremales no dominados*. De estos puntos, se pueden obtener los puntos *ideales* y *nadir*, ip y np , cuyas componentes proporcionan cotas inferiores y superiores respectivamente para las funciones objetivo. Más concretamente, $ip = (y_1^1, y_2^N)$, $np = (y_1^N, y_2^1)$. Estos cuatro puntos, determinan un rectángulo que contiene al resto de puntos no dominados. El objetivo es hallar el conjunto Y_E , y para cada $\mathbf{y} \in Y_E$ tener $\mathbf{x} \in X_E$ tal que $\mathbf{y} = f(\mathbf{x})$. En la literatura existente, es frecuente encontrarse con diferentes procedimientos para hallar los puntos no dominados. La elección del procedimiento depende de las características del problema multiobjetivo en concreto que estemos tratando.

4.3. Problema bi-objetivo p -mediana-centro. El método ε -constraint

Consideremos el problema entero bi-objetivo formado por el problema de la p -mediana (clásico) y el problema del p -centro (**BO-MCP**) siguiente:

$$\begin{aligned} \text{(BO-MCP)} \quad & \text{Minimizar} \quad \sum_{i=1}^m \sum_{j=1}^n h_i d_{ij} y_{ij} \\ & \text{Minimizar} \quad w \end{aligned} \quad (4.1)$$

$$\text{s.a.} \quad \sum_{j=1}^n y_{ij} = 1, \quad i = 1, \dots, m \quad (4.2)$$

$$\sum_{j=1}^n d_{ij} y_{ij} \leq w, \quad i = 1, \dots, m \quad (4.3)$$

$$y_{ij} \leq x_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n \quad (4.4)$$

$$\sum_{j=1}^n x_j = p, \quad (4.5)$$

$$x_j \in \{0, 1\}, \quad j = 1, \dots, n \quad (4.6)$$

$$y_{ij} \in \{0, 1\}, \quad i = 1, \dots, m, \quad j = 1, \dots, n \quad (4.7)$$

Este problema bi-objetivo puede resolverse por el método ε -constraint (ver Erhgoth (2005) para cuestiones teóricas generales de este método). La idea básica del mismo es minimizar una función objetivo y añadir restricciones a la otra para obtener los puntos eficientes. En consecuencia, se obtiene el siguiente subproblema (**BO-MCP**(ε)):

$$\begin{aligned} \text{(BO-MCP}(\varepsilon)) \quad & \text{Minimizar} \quad \sum_{i=1}^m \sum_{j=1}^n h_i d_{ij} y_{ij} \\ & \text{s.a.} \quad \sum_{j=1}^n d_{ij} y_{ij} \leq \varepsilon, \quad i = 1, \dots, m \\ & \mathbf{x} \in X \end{aligned} \quad (4.8)$$

siendo ε una cota superior de la distancia máxima y X el conjunto de restricciones (4.2),(4.4), (4.5),(4.6) y (4.7).

Afortunadamente, con el método ε -constraint hemos recaído en un problema de p -mediana con restricciones de distancia máxima, el cual hemos tratado en el capítulo anterior y hemos resuelto por varios métodos, resultando el GRASP el método adecuado. Más adelante, observaremos la trascendencia del hecho de que los subproblemas son un problema de p -mediana con restricciones de distancia máxima ya que existirá una relación entre los puntos eficientes y los subintervalos de invariancia correspondientes a los subproblemas.

En la literatura (véase Ehrgott (2005)), existen otros métodos para resolver problemas bi-objetivo, tales como métodos basados en normas de Chebychev, o también métodos basados en ponderaciones o métodos lexicográficos si bien para nuestro problema no resultan adecuados pues o son difíciles de resolver o no permiten obtener todos los puntos eficientes. Más concretamente, un método muy común encontrado en la literatura para resolver el problema bi-objetivo que tratamos aquí es el método de las ponderaciones (*weighted sum method*, en inglés) que se basa en

$$\begin{array}{ll} \text{Minimizar} & f(\mathbf{x}) = (\lambda_1 f_1(\mathbf{x}) + \lambda_2 f_2(\mathbf{x})) \\ \text{(WM-BOIP)} & \\ & s.a. \quad \mathbf{x} \in X \end{array}$$

donde $\lambda_1, \lambda_2 \geq 0$. Si $\lambda > 0$, $i = 1, 2$ entonces toda solución óptima de (WM-BOIP) es eficiente. Las soluciones eficientes que pueden ser encontradas por el método de las ponderaciones con coeficientes positivos se conocen como *soluciones soportadas* y los correspondientes puntos se llaman *puntos no dominados soportados*. Gráficamente, los puntos soportados se encuentran entre los puntos extremos de la envolvente inferior del conjunto Y_E , como por ejemplo los puntos $\mathbf{y}^1, \mathbf{y}^3$ e \mathbf{y}^5 en la figura 4.1. El mayor inconveniente del método de las ponderaciones es que por ejemplo para problemas no convexos tales como (BOIP), puntos no soportados, tales como \mathbf{y}^2 e \mathbf{y}^4 en la figura 4.1, no se pueden obtener por dicho método.

En este problema que tratamos aquí, salvamos las dificultades que encuentran otros autores con las soluciones soportadas que aparecen al utilizar métodos ponderados. Daskin (2010, 2013) señala que con el método de las ponderaciones no se pueden obtener todos los puntos eficientes, y es por ello, que recurre al método ε -constraint y resuelve el problema con el software que acompaña a los textos de Daskin, conocido como SITATION. Además, SITATION está caracterizado por resolver un limitado tipo de problemas y ejemplos de los existentes en Daskin (2010, 2013) presentando además el inconveniente de que es sólo capaz de resolver hasta un limitado número de nodos. A continuación, señalemos cómo aborda Daskin (2010, 2013) el problema que estamos tratando:

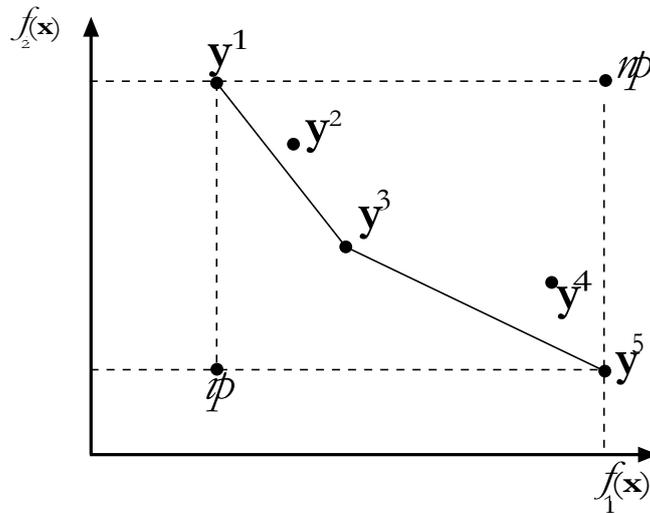


Figura 4.1: (De Sáez-Aguado y Trandafir (2016)).

- En Daskin (2010), aparece el siguiente enfoque basado en introducir

$$\beta_{ij} = \begin{cases} 1, & \text{si } d_{ij} > s \\ 0, & \text{si } d_{ij} \leq s \end{cases}$$

donde se asume que la distancia máxima es igual al valor s para todo $i = 1, \dots, m$ y se resuelve el problema de

$$\text{Minimizar } \sum_{i=1}^m \sum_{j=1}^n h_i (d_{ij} + W \beta_{ij}) y_{ij} \quad (4.9)$$

s.a. las restricciones del problema de la p -mediana clásico, donde W es un parámetro de penalización. Además, la calidad de las soluciones obtenidas dependen en gran medida de la elección del parámetro W . Este enfoque conlleva gran carga computacional.

- En Daskin (2013), se resuelve el problema que tratamos aquí con el método ε -constraint señalando también el enfoque indirecto (basado en modificar la matriz de distancias) conllevando gran carga computacional al resolver un número alto de problemas de p -mediana con restricciones de distancia máxima.

Daskin (2010, 2013), al no poseer ningún método eficiente para resolver el problema de la p -mediana con restricciones de distancia máxima, recurre al uso del software SITTATION, siendo por tanto este enfoque no novedoso de cara a métodos heurísticos de resolución. En Daskin y Maass (2015), aparece "*For small instances it is often possible to solve bi-objective problems using extensions of the Lagrangian algorithm outlined above. For larger instances, using a genetic algorithm is often advisable since the population of solutions in a genetic algorithm automatically gives an initial approximation of the non-dominated set of solutions*". Por tanto, Daskin y Maass (2015) resuelven el problema bi-objetivo por medio del software SITTATION usando para problemas de tamaño pequeño un algoritmo de relajación Lagrangiana+Branch & Bound y para problemas más grandes, un algoritmo genético el cual según lo citado, proporciona una *aproximación inicial* al conjunto de puntos no-dominados. Nótese que Daskin y Maass (2015, p.41) usan el enfoque indirecto, basado en modificar la matriz de distancias, sin atender a los peligros que ello conlleva en los métodos heurísticos. De hecho, para el problema con 250 nodos expuesto en Daskin y Maass (2015, p.42) aparece: "*Obtaining the 22 solutions shown in the figure took nearly 16 h of computing time*". Más adelante, veremos que nuestros procedimientos de resolución mejoran muy notablemente los existentes en la literatura, calculando todos los puntos no dominados en escasos segundos, sin obtener aproximaciones de los mismos, al contrario que Daskin y Maass (2015). Nótese que en Daskin (2010, 2013), Daskin y Maass (2015) no se proporciona ningún método heurístico efectivo de resolución (además de valerse del software SITTATION), y tampoco se señalan los problemas que implica el enfoque indirecto en los métodos heurísticos. A continuación señalamos brevemente algunas propiedades relacionadas con el método ε -constraint:

- Toda solución eficiente puede obtenerse resolviendo (**BO-MCP**(ε)) para algún valor de ε .
- Toda solución \mathbf{x}^* de (**BO-MCP**(ε)) es débilmente eficiente y además, existe alguna solución óptima de (**BO-MCP**(ε)) que es eficiente. En particular, si \mathbf{x}^* es la única solución óptima de (**BO-MCP**(ε)), entonces \mathbf{x}^* es eficiente.
- La gran ventaja del método ε -constraint es que se obtienen todas las soluciones eficientes, aunque presenta un inconveniente, ya que se pueden obtener puntos dominados. En la figura 4.2, los puntos A, B y C minimizan simultáneamente el primer objetivo, pero sólo C es no dominado. Es por ello, que a la hora de implementar el mismo, debe existir un procedimiento o rutina para eliminar tales puntos dominados. Para una revisión detallada de las diferentes implementaciones del método ε -constraint puede consultarse Sáez-Aguado y Trandafir (2016).

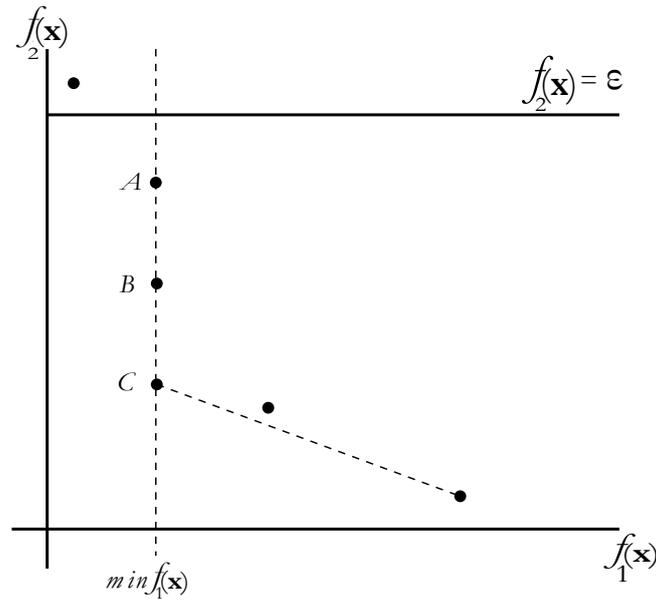


Figura 4.2: (De Sáez-Aguado y Trandafir (2016)).

A continuación, describimos los pasos del algoritmo ε -constraint con eliminación de puntos dominados aplicado a nuestro problema:

ALGORITMO ε -constraint con eliminación de puntos dominados

- **Paso 1** (Inicialización): Establecer $Y_E = \emptyset, X^* = \emptyset, \varepsilon = \infty$ y $k = 0$.
- **Paso 2** (Resolución de los subproblemas) : Resolver (**BO-MCP**(ε)). Si el problema es no factible, poner $N = k$, e ir al paso 4. En otro caso, establecer \bar{x} como la solución óptima obtenida e ir al paso 3.
- **Paso 3**: Establecer $k := k + 1, \mathbf{x}^k = \bar{x}, \mathbf{y}^k = (f_1(\bar{x}), f_2(\bar{x}))$, $Y_E = Y_E \cup \{\mathbf{y}^k\}$, $X^* = X^* \cup \{\mathbf{x}^k\}$, $\varepsilon = f_2(\bar{x}) - 1$ e ir al paso 2.
- **Paso 4**: Eliminar todos los puntos dominados de Y_E , y actualizar N, Y_E y X^* y PARAR.

4.4. Resultados computacionales. Análisis bi-objetivo

Vamos a realizar un análisis bi-objetivo correspondiente a los problemas test aint12 con $p = 7$ y aint13, con $p = 10$ para ilustrar gran parte de todo lo anterior.

- Problema aint12, con parámetros $m = 266, n = 39, p = 7$, y con máxima distancia mínima igual a 185 y tiempo de computación (en segundos) igual a 0.202 seg.

ε	OBJ. 1 (Distancia promedio)	OBJ. 2 (Distancia máxima)
9999999	45.72	337
336	249.088811	309
308	250.551789	305
304	251.428772	292
291	253.618036	261
260	257.046515	255
254	264.324902	235

A continuación, se muestran en un gráfico, la distancia máxima frente a la distancia promedio para el problema de la p -mediana con restricciones de distancia máxima asociado a los mismos datos. Los subintervalos de invariancia son: $[235,254]$, $[255,260]$, $[261,291]$, $[292, 304]$, $[305,308]$, $[309,336]$ y 337.

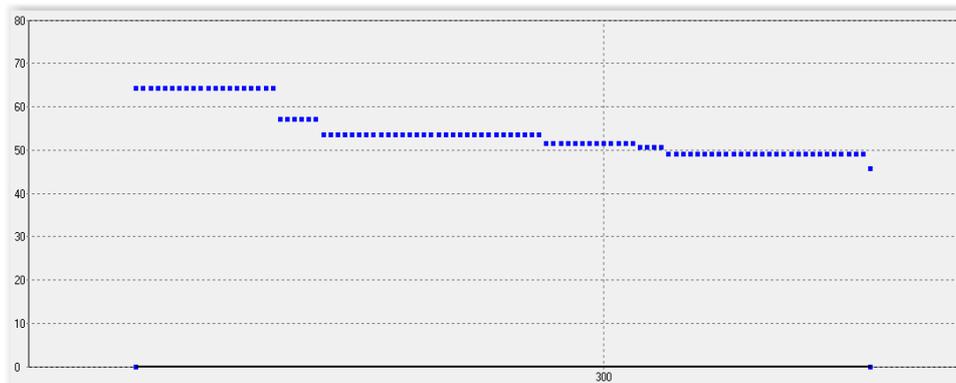


Figura 4.3

La relación de los subintervalos de invariancia con el problema bi-objetivo que tratamos es la siguiente:

- El número de puntos de la frontera eficiente coincide con el número de subintervalos de invariancia.
- Los extremos inferiores de los subintervalos de invariancia coinciden (convenimos que si un subintervalo de invariancia degenera en un punto su extremo inferior es el mismo punto) con los valores hallados del objetivo 2.
- Los valores de ε correspondientes a un punto de la frontera eficiente coinciden salvo el primer valor del mismo (recuérdese que ε se toma como un valor suficientemente *grande* del objetivo 1, nosotros hemos tomado 9999999) con los extremos superiores de los subintervalos de invariancia.
- El primer punto obtenido corresponde con la solución del problema de la p -mediana clásico, y el último punto corresponde a la solución del problema del p -centro. El resto de puntos intermedios (si los hubiera) corresponden a soluciones intermedias de posible interés para el decisor.
- Nótese que a priori (a excepción de cuando se toma ε suficientemente grande), no hay relación entre la distancia promedio asociada a un subintervalo de invariancia y la distancia promedio hallada en el objetivo 1. Esta es la diferencia que aporta el análisis bi-objetivo al análisis de los intervalos de factibilidad. Con el análisis bi-objetivo se obtiene más información para el decisor, y esto se debe fundamentalmente a la influencia del objetivo 2. Al final, el decisor tiene que seleccionar una solución eficiente. A la misma la llamaremos *solución de mejor compromiso*.

A continuación, vamos a representar en un gráfico los puntos eficientes hallados para este problema test:

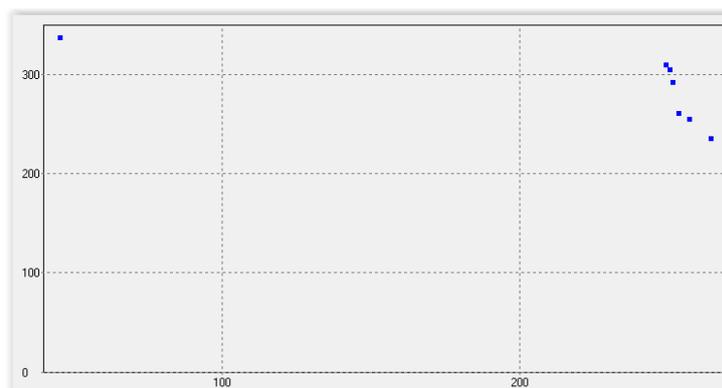


Figura 4.4

Observamos que el segundo punto obtenido hace aumentar mucho el primer objetivo (distancia promedio) manteniendo casi invariante el segundo objetivo (distancia máxima). Esto parece indicar que el segundo punto obtenido no es una opción aconsejable de solución, ya que teniendo el primer punto, con un poco más de distancia máxima se tiene bastante menos distancia promedio y esto indica que es un buen candidato a solución de mejor compromiso.

- Problema aint13, con parámetros $m = 430, n = 22, p = 10$, y con máxima distancia mínima igual a 295 y tiempo de computación (en segundos) igual a 0.083 seg.

ε	OBJ. 1 (Distancia promedio)	OBJ. 2 (Distancia máxima)
9999999	58.44	383
382	261.77	320
319	263.68	295

A continuación, se muestran en un gráfico, la distancia máxima frente a la distancia promedio para el problema de la p-mediana con restricciones de distancia máxima asociado a los mismos datos. Los subintervalos de invariancia son: $[295,319]$, $[320,382]$ y 383.

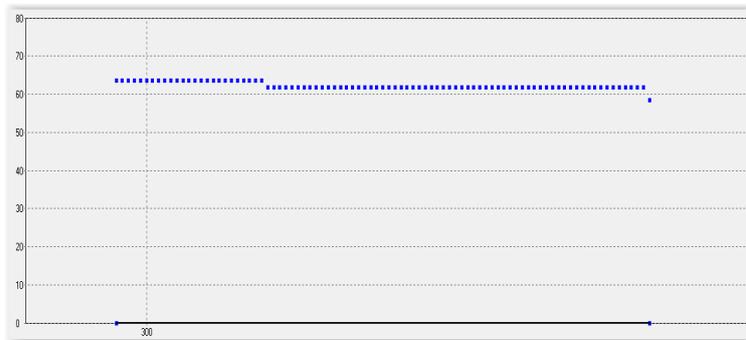


Figura 4.5

A continuación, vamos a representar en un gráfico los puntos eficientes hallados para este problema test:

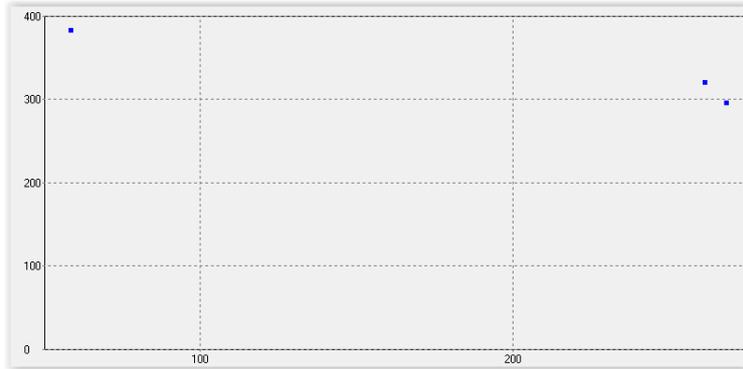


Figura 4.6

Observamos que el segundo punto obtenido hace aumentar mucho el primer objetivo (distancia promedio) manteniendo casi invariante el segundo objetivo (distancia máxima). Esto parece indicar que el segundo punto obtenido no es una opción aconsejable de solución, ya que teniendo el primer punto, con un poco más de distancia máxima se tiene bastante menos distancia promedio y esto indica que es un buen candidato a solución de mejor compromiso. Se observa que hay menos puntos eficientes que en el problema test anterior, y que básicamente el decisor tiene que seleccionar una solución a su problema teniendo en cuenta o bien la eficiencia o bien la equidad ya que la solución intermedia hallada está bastante próxima a la solución del p -centro.

4.5. Conclusiones

En este capítulo, se ha estudiado un problema bi-objetivo correspondiente a los problemas de la p -mediana y del p -centro. Aplicando el método ε -constraint, se recae en unos subproblemas de p -mediana con restricciones de distancia máxima, los cuales se han resuelto muy rápidamente con la metaheurística GRASP para obtener los puntos eficientes. Además de analizar los mismos y su relación con los subintervalos de invariancia, se han obtenido unas soluciones intermedias de gran interés para el decisor aparte de las extremas, que se obtendrían al resolver separadamente los dos objetivos.

CONCLUSIONES

A lo largo de este trabajo, a partir de problemas de Programación Entera y a través de diversas técnicas avanzadas de Programación Matemática tales como la relajación Lagrangiana, o métodos heurísticos y metaheurísticos, se han estudiado algunos modelos que involucran aspectos relacionados con la Teoría de Localización o la Teoría de Inventarios.

Se han obtenido **resultados novedosos**, ya que se han proporcionado diferentes procedimientos de resolución que mejoran los existentes en la literatura. Se ha podido comprobar que la relajación Lagrangiana sigue siendo hoy en día un buen procedimiento para obtener soluciones óptimas en modelos integrados de Localización-Inventario. En otros modelos de Localización, se ha comprobado que la metaheurística conocida como GRASP es de gran éxito en la práctica por la eficiencia computacional. Finalmente, aplicamos el método ε -constraint para resolver un problema bi-objetivo de gran interés en el mundo real.

Bibliografía

- [1] ARDALAN, A. (1988). "A comparison of heuristic methods for service facility locations". *International Journal of Operations & Production Management*, Vol. 8, No. 2, 52-8.
- [2] ATAMTÜRK, A., BERENQUER, G., SHEN, Z.-J.M. (2012). "A conic integer programming approach to stochastic joint location-inventory problems". *Operations Research*. Vol. 60 (2), 366-381.
- [3] AVELLA, P., SASSANO, A., VASILEV, I. (2007). "Computational study of large-scale p-median problems". *Mathematical Programming*, Vol. 109, (1), 89-114.
- [4] BAÏOU, M., BARAHONA, F. (2009). "On the integrality of some facility location polytopes". *SIAM J. Discrete Math.* Vol. 23 (2), 665-679.
- [5] BERTSEKAS, D. P. (2016). *Nonlinear Programming*. 3rd edition, Belmont, Mass.: Athena Scientific.
- [6] BORWEIN, J., LEWIS, A.S. (2006). *Convex Analysis and Nonlinear Optimization. Theory and Examples*. Ed. Springer-Verlag. New York.
- [7] CHAUDRY, S.S., CHOI, I.-C., SMITH, D.K. (1994). "Facility location with and without maximum distance constraints through the p-median problem", *International Journal of Operations & Production Management*, Vol. 15, (10), 75-81.
- [8] CHOI, I.-C., CHAUDHRY, S.S. (1993). "The p-median problem with maximum distance constraints: a direct approach", *Location Science*, Vol. 1 No. 3, 235-43.
- [9] CHVATAL V. (1979). "A greedy heuristic for the set-covering problem", *Mathematics of Operations Research*, 4, 233-5.

- [10] COLEBROOK SANTAMARÍA, M. (2014). *Localización simple de servicios deseados y no deseados en redes con múltiples criterios*. Tesis Doctoral. Universidad de La Laguna.
- [11] DASKIN, M. S. (1995). *Network and Discrete Location: Models, Algorithms and Applications*. Wiley. New York.
- [12] DASKIN, M. S., (2010). *Service Science*, John Wiley and Sons, Inc., New York.
- [13] DASKIN, M. S. (2013). *Network and Discrete Location: Models, Algorithms and Applications*. Second Ed. Wiley. New York.
- [14] DASKIN, M. S., COULLARD, C. R., SHEN. Z.-J. M. (2002). "An Inventory-Location model: Formulation, solution algorithm and computational results". *Annals of Operations Research*, 110(1) 83-106.
- [15] DASKIN, M.S., MAASS, K.L.: "The p-median problem". En: LAPORTE, G., NICKEL, S., SALDANHA DA GAMA, F. (eds.) *Location Science*, 21-45. Springer International Publishing, Berlin (2015).
- [16] EHRGOTT, M. (2005) *Multicriteria Optimization*. Springer.
- [17] ESCALONA, P., ORDÓÑEZ, F., MARIANOV, V. (2015). "Joint location-inventory problem with differentiated service levels using critical level policy". *Transportation Research Part E* , 83, 141-157.
- [18] ESTEBAN PÉREZ, A. (2016). *Estudio de propiedades retrohereditarias en algunos problemas secuenciales de Optimización Estocástica*. Trabajo Fin de Grado. Universidad de Valladolid.
- [19] FERNÁNDEZ SUÁREZ, N., GARCÍA LAGUNA, J., MARTÍNEZ FERRERAS, J., SAN JOSÉ NIETO, L.A. (1999). *Gestión de stocks : modelos de optimización y software*. Secretariado de Publicaciones e Intercambio Editorial, Universidad de Valladolid.
- [20] FRANGIONI, A., GENDRON, B., GORGONE, E. (2017). "On the computational efficiency of subgradient methods: a case study with Lagrangian bounds". *Mathematical Programming Computation*, 1-32.
- [21] GARCÍA LAGUNA, J. (1984). "Sobre un cono convexo asociado a un grafo". *Trabajos de Estadística y de Investigación Operativa* . Consejo Superior de Investigaciones Científicas. Sociedad de Estadística e Investigación Operativa. Vol. 35, (3), 277-292.

- [22] GRINOLD, R.C. (1970). "Lagrangean subgradients". *Management Science*, 17, 185-188.
- [23] GUIGNARD, M. (2003). "Lagrangean relaxation". *TOP*, 11, 151-228.
- [24] HAKIMI, S.L. (1964). "Optimum locations of switching centers and the absolute centers and medians of a graph", *Operations Research*, 12 (3), 450-459.
- [25] HALPERN, J. (1976). "The location of a center-median convex combination on an undirected tree", *Journal of Regional Science*, 16, 237-245.
- [26] HALPERN, J. (1978). "Finding minimal center-median convex combination (cent-dian) of a graph", *Management Science*, 24(5), 535-544.
- [27] HANDLER, G.Y. (1976). "Medi-centers of a tree", Working Paper 278, The Recanati Graduate School of Business Administration, Tel Aviv University, Israel.
- [28] HELD M., WOLFE P., CROWDER H.P. (1974) "Validation of subgradient optimization", *Mathematical Programming*, 6,62-88.
- [29] HILLSMAN E.L., RUSHTON G. (1975). "The p-median problem with maximum distance constraints: A comment". *Geographical Analysis*, 7, 85-89.
- [30] HIRIART-URRUTY, J.-B., LEMARÉCHAL, C. (1993). *Convex Analysis and Minimization Algorithms*, Vols. I y II. Ed. Springer.
- [31] KHUMAWALA B.M. (1973) "An efficient algorithm for the p-median problem with maximum distance constraints". *Geographical Analysis*, 5, 309-321.
- [32] KHUMAWALA B.M. (1975) "Algorithm for the p-median problem with maximum distance constraints: Extension and reply". *Geographical Analysis*, 7, 91-95.
- [33] LI, D., SUN, X. (2006). *Nonlinear Integer Programming*. (International Series in Operations Research & Management Science). Ed. Springer. New York.
- [34] MUCKSTADT, J.A., SAPRA, A. (2010). *Principles of Inventory Management: When You Are Down to Four, Order More* . Ed. Springer.
- [35] MURRAY, A.T., WU, X. (2003). "Accessibility tradeoffs in public transit planning". *Journal of Geographical Systems*, Vol.5, 1, 93-107.
- [36] MURTY, K.G. (1995). *Operations Research : Deterministic Optimization Models*. Prentice-Hall. Upper Saddle River (New Jersey).

- [37] OGRYCZAK, W. (1997), "On cent-dians of general networks", *Location Science*, 5 (1), 15-28.
- [38] PANDO FERNÁNDEZ, V. (2014). *Análisis y optimización de nuevos sistemas determinísticos y estocásticos en gestión de stocks*. Tesis Doctoral. Universidad de Valladolid.
- [39] PORTEUS, E.L. (2002). *Foundations of Stochastic Inventory Theory*. Stanford University Press. Stanford. California.
- [40] RAHMAN, S. AND SMITH, D.K. (1991). "A comparison of two heuristic methods for the p-median problem with and without maximum distance constraints", *International Journal of Operations & Production Management*, Vol. 11 No. 6, 76-84.
- [41] RESENDE M., RIBEIRO C. (2016). *Optimization by GRASP: Greedy Randomized Adaptive Search Procedures*. Springer Verlag.
- [42] RESENDE M., WERNECK, R.F. (2003). "On the Implementation of a Swap-Based Local Search Procedure for the p-Median Problem". En R.E. Ladner (ed.), *Proceedings of the Fifth Workshop on Algorithm Engineering and Experiments (ALENEX'03)*, SIAM, 119-127.
- [43] REVELLE C.S, SWAIN R.W. (1970). "Central facilities location". *Geographical Analysis*, 2, 30-42.
- [44] ROCKAFELLAR, R.T. (1972). *Convex Analysis*. Princeton University Press. Princeton (New Jersey).
- [45] RUSZCZYNSKI, A. (2006). *Nonlinear Optimization*. Princeton University Press.
- [46] SÁEZ-AGUADO, J., TRANDAFIR, P.C. (2016). *Implementations of the ε -Constraint method for bi-objective integer linear programming problems. Application to p-median-cover problems*. Preprint. Departamento de Estadística e Investigación Operativa. Universidad de Valladolid.
- [47] SAN JOSÉ NIETO, L.A. (2006). *Obtención de Algoritmos de Optimización en Modelos EOQ*. Tesis Doctoral. Universidad de Valladolid.
- [48] SHAPIRO, J.F (1979). *Mathematical Programming : Structures and Algorithms*. John Wiley & Sons. New York.
- [49] SHEN, Z.-J.M., COULLARD, C., DASKIN, M.S. (2003). "A joint location-inventory model". *Transportation Science*. 37 (1), 40-55.

- [50] SIMCHI-LEVI, D., CHEN, X., BRAMEL, J. (2014). *The Logic of Logistics. Theory, Algorithms, and Applications for Logistics Management*. Ed. Springer.
- [51] SNYDER, L.V. , SHEN, Z.-J. M. (2011). *Fundamentals of Supply Chain Theory*. Wiley.
- [52] TEITZ, M. B., BART P. (1968) . "Heuristic Methods for Estimating Generalized Vertex Median of a Weighted Graph" . *Operations Research*, 16, 955-961.
- [53] TOREGAS, C., SWAIN, R.W., REVELLE, C.S. AND BERGMAN, L. (1971). "The location of emergency service facilities". *Operations Research*. Vol. 19 (6), 1971, 363-73.
- [54] WHITAKER, R. (1983). "A fast algorithm for the greedy interchange of large-scale clustering and median location problems". *INFOR*. 21, 95-108.
- [55] YOU, F., GROSSMANN, I.E. (2008). "Mixed-integer nonlinear programming models and algorithms for large-scale supply chain design with stochastic inventory management". *Industrial & Engineering Chemistry Research* . 47 (20), 7802-7817.
- [56] ZIPKIN P.H. (2000). *Foundations of Inventory Management*. McGraw-Hill.