



Resolución de los problemas de optimización con restricciones mediante los métodos de penalización y del Lagrangiano aumentado

| Solving constraint optimization problems using the penalty and augmented Lagrangian methods |

 Manuel Vázquez Mourazos

manuelvazquezmourazos@gmail.com

Universidad de Santiago de Compostela
España (Melide, A Coruña)

Recibido: 24 abril 2020

Aceptado: 15 Septiembre 2020

Resumen: En la resolución de los problemas de optimización con restricciones (también conocida con el nombre de programación matemática) se pueden establecer diversidad de algoritmos. Como la investigación en este campo es muy amplia, continuamente se están desarrollando nuevos métodos, cada vez más sofisticados, que permiten resolver este tipo de problemas.

En este artículo se presentarán los métodos de penalización. Estos son los más intuitivos y permiten mostrar una introducción dentro de la resolución de este tipo de problemas. Se estudiarán sus propiedades y, luego, se procederá a su deducción, interpretación y demostración de su convergencia. Finalmente, se presentará el método del Lagrangiano aumentado. Este es un método que mejora a los anteriores y con una mayor velocidad de convergencia. Así mismo, este artículo supone una introducción a los algoritmos de resolución de problemas de optimización con restricciones, mostrando una iniciación a los métodos numéricos empleados en la programación matemática.

Palabras Clave: optimización con restricciones, penalización, penalización exterior, penalización interior, Lagrangiano aumentado, condiciones KKT

Abstract: In solving constrained optimization problems (also by the name of mathematical programming), a variety of algorithms can be established. As the research in this field is very extensive, new and increasingly sophisticated methods are continually being developed to solve this type of problem.

In this article the methods of penalization will be presented. These are the most intuitive and allow to show an introduction within the resolution of this type of problem. Their properties will be studied, and then their deduction, interpretation and demonstration of their convergence will proceed. Finally, the augmented Lagrangian method will be presented. This is a method that improves on the previous ones and allows for greater and better convergence. Likewise, this article supposes an introduction to the optimization algorithms of constraint optimization, showing an introduction to the numerical methods used in mathematical programming.

Keywords: constraint optimization, penalty methods, simple penalty, interior point penalty, Augmented Lagrangian, KKT conditions

1. Introducción

La programación matemática es una rama que pretende estudiar la resolución de problemas de optimización. En este artículo se dará una serie de algoritmos para su resolución, siendo el objetivo de estudio el algoritmo del lagrangiano aumentado. A lo largo de las diferentes secciones se introducirán resultados teóricos sobre los que se apoyan estos métodos, haciendo una revisión básica por los más importantes. Los teoremas y resultados mostrados no son de producción propia, sino que son una recopilación sobre la que se basa el tema de estudio en este artículo. En este sentido, el aporte del artículo se centra en los ejemplos, la ejecuciones y las reflexiones que se vayan mostrando según se introduzcan los contenidos.

Notación 1

Antes de comenzar con la redacción, se hacen las siguientes aclaraciones correspondientes a la notación que se empleará en lo sucesivo:

1. A lo largo del artículo se emplearán escalares, vectores y matrices de forma frecuente. Para diferenciarlos, se escribirá a los vectores en letra negrita minúscula (por ejemplo \mathbf{v}), a las matrices por letra negrita mayúscula (por ejemplo \mathbf{M}) y en letra normal minúscula a los escalares (por ejemplo k).
2. El empleo de los vectores será muy reiterado. Dado un vector \mathbf{v} , si se desea referirse a su componente i -ésima, esta se denotará en letra normal minúscula con el subíndice, es decir, v_i .
3. En una sucesión de vectores, se denotará al vector k -ésimo de la sucesión con la siguiente escritura: $\mathbf{v}^{(k)}$.
4. A los conjuntos se les denotará en letra normal mayúscula, por ejemplo S . El término $\text{Int}(S)$ se refiere al interior del conjunto.

Las diversas situaciones en las que es necesario el estudio de estos problemas, conlleva a que su formulación pueda ser muy variada y compleja. Normalmente, un problema que se asemeja bastante a una formulación genérica de un problema de este tipo es el siguiente:

$$\begin{aligned} & \underset{\mathbf{v} \in \mathbb{R}^n}{\text{mín}} && J(\mathbf{v}) \\ & \text{Sujeto a:} && \\ & \left\{ \begin{array}{ll} \varphi_i(\mathbf{v}) \leq 0 & i = 1, \dots, m \\ \phi_j(\mathbf{v}) = 0 & j = 1, \dots, p \end{array} \right. && \end{aligned} \quad (1)$$

El objetivo del problema (1) es la obtención del mínimo del funcional objetivo, $J(\mathbf{v})$, dentro del conjunto que determinan las restricciones. Es decir, no se busca optimizar el funcional objetivo en todo su dominio, sino que se exige que la solución pertenezca al siguiente conjunto,

$$S := \left\{ \mathbf{v} \in \mathbb{R}^n : \begin{array}{ll} \varphi_i(\mathbf{v}) \leq 0 & i = 1, \dots, m \\ \phi_j(\mathbf{v}) = 0 & j = 1, \dots, p \end{array} \right\},$$

que se denominará como *conjunto admisible* o *conjunto factible*.

La resolución del problema (1) es bastante compleja y, para ello, hay diversas tácticas que permiten obtener una solución numérica del problema. Algunos de los algoritmos abordan el problema directamente, pero, por el contrario, otros métodos optan por el cálculo de puntos que cumplen una serie de condiciones cuyas propiedades hacen que sean candidatos muy probables a ser solución del problema.

Dentro de los algoritmos que opten por encontrar un posible candidato a solución, los más empleados consisten en calcular un punto que verifique las condiciones de Karush-Kuhn-Tucker, también conocidas como condiciones KKT. Estas condiciones permiten concluir, bajo ciertas hipótesis, que los puntos que las verifican son mínimos del problema genérico (1). Su definición se establece a continuación:

Definición 1

Se dice que un punto $\mathbf{u} \in \mathbb{R}^n$ es un punto de Karush-Kuhn-Tucker para el problema (1), si y sólo si, existen multiplicadores de Lagrange y de Karush-Kuhn-Tucker (a los que se denotará por las letras griegas $\lambda = (\lambda_1, \dots, \lambda_p)^\top$ y $\mu = (\mu_1, \dots, \mu_m)^\top$, respectivamente) que verifican las condiciones de KKT, que son las siguientes:

- Condición estacionaria:

$$\nabla J(\mathbf{u}) + \sum_{i=1}^m \mu_i \nabla \varphi_i(\mathbf{u}) + \sum_{j=1}^p \lambda_j \nabla \phi_j(\mathbf{u}) = \mathbf{0}, \quad (2)$$

- Condición de factibilidad

$$\begin{cases} \varphi_i(\mathbf{u}) \leq 0 & i = 1, \dots, m \\ \phi_j(\mathbf{u}) = 0 & j = 1, \dots, p \end{cases} \quad (3)$$

- Condición de holgura:

$$\begin{cases} \mu_i \varphi_i(\mathbf{u}) = 0 & i = 1, \dots, m \\ \mu_i \geq 0 & i = 1, \dots, m \end{cases} \quad (4)$$

Si se observa la Definición 1, esta establece, mediante la condición (3), la factibilidad del punto. Las demás condiciones establecen una caracterización del punto que servirá para demostrar que un punto que verifica las condiciones KKT es un buen candidato a ser solución del problema (1). Dado que este artículo no versa sobre las condiciones KKT, se introducirán los siguientes resultados sin demostración. Además, se incluirá una referencia donde se pueden consultar.

Definición 2

Dado el punto $\mathbf{v} \in S$, se define el conjunto de restricciones activas de \mathbf{v} como sigue:

$$I(\mathbf{v}) := \{i \in \{1, \dots, m\} : \varphi_i(\mathbf{v}) = 0\}.$$

Teorema 1 (Teorema de Karush-Kuhn-Tucker)

Sea \mathbf{u} un mínimo local del problema (1). Si los vectores $\nabla \phi_j(\mathbf{u})$ con $j = 1, \dots, p$ y $\nabla \varphi_i(\mathbf{u})$, con $i \in I(\mathbf{u})$, son linealmente independientes, entonces existen multiplicadores de Lagrange y de KKT que verifican las condiciones KKT.

Demostración. Se puede consultar en [13, Teorema 8.2.7]. □

Este resultado permite establecer que las condiciones KKT son una condición necesaria bajo una hipótesis. La hipótesis establecida en el Teorema 1 se denomina *condición de cualificación* y esta se puede cambiar por otras que permiten demostrar el mismo resultado. Para más información de este tema, consultar [13, Cap. 8] ó [5, Sec. 6.3].

Sin embargo, el resultado visto es necesario, pero no es suficiente. Para ello, se pueden robustecer las hipótesis del problema de optimización haciendo que este sea convexo. Esto quiere decir que tanto el funcional objetivo como el conjunto admisible sean convexos (se puede consultar para ello la Definición 3). El resultado se puede ver en el Teorema 2.

Definición 3

Un conjunto S se dice convexo si verifica que para cualesquiera $a, b \in S$ se cumple que

$$(1 - t)a + tb \in S, \quad \forall t \in [0, 1].$$

Además, dada una función f , definida sobre un conjunto convexo, se dice que es convexa si para cualesquiera dos puntos a y b de su dominio se verifica que

$$f((1 - t)a + tb) \leq (1 - t)f(a) + tf(b), \quad \forall t \in [0, 1].$$

Teorema 2

Se considera el problema de optimización (1). Si este es convexo, se tiene que dado un punto admisible $\mathbf{u} \in S$, para el cual existen multiplicadores de Lagrange y KKT (λ y μ , respectivamente), entonces \mathbf{u} es un mínimo del problema.

Demostración. Se puede consultar en [13, Teorema 8.2.18]. □

De esta forma, se comprueba que los puntos que verifican las condiciones KKT, bajo ciertas condiciones, son puntos que pueden ser mínimos del problema original (1) o incluso ser los únicos mínimos. Así, algunos métodos numéricos optan por calcular numéricamente este tipo de puntos para hacer un estudio posterior y comprobar si efectivamente son soluciones al problema.

En este artículo se presentarán los métodos de penalización. Estos algoritmos se basan en una serie de ideas muy intuitivas que permiten resolver el problema genérico (1) de forma directa. Sin embargo, en la práctica no resultan del todo favorables y, por ello, se mostrará finalmente el método del Lagrangiano aumentado que fue desarrollado primeramente entre 1973 y 1974 por Rockafellar ([10],[11] y [12]) a partir de la idea inicial de Hestenes [7] y Powell [9] en 1969. Aunque este método se basa en parte en la penalización, este opta por la obtención y cálculo de un punto que verifican las condiciones KKT.

2. Métodos de penalización

En esta sección se estudiarán los métodos de penalización, existen diversos tipos, pero una característica común que los define. Esta consiste en reducir el problema con restricciones (1), a un problema sin restricciones. Es decir, se reduce el problema genérico principal a uno del siguiente tipo:

$$\min_{\mathbf{v} \in \mathbb{R}^n} J(\mathbf{v}), \tag{5}$$

donde se optimiza el funcional objetivo en todo su dominio en lugar de en un conjunto admisible. La optimización sin restricciones resulta más sencilla y se ha estudiado con mayor profundidad a lo largo de la historia. Para su resolución existen diversos algoritmos, algunos de ellos muy sofisticados y con muy buenos resultados en la práctica. Por este motivo, se opta por esta reducción de un problema más complejo a uno más sencillo. Si se desean conocer los detalles acerca de los algoritmos que permiten

resolver los problemas de optimización sin restricciones, se puede consultar [2, Cap. 5] para tener una visión general o, también, se puede consultar [8, Cap. 6] donde se hace hincapié en los métodos Quasi-Newton, que son los más empleados.

La idea principal de los métodos de penalización consiste en construir un nuevo funcional objetivo de tal manera que si se toma un elemento fuera del conjunto admisible, se penalice esa elección aumentando el valor que toma el funcional. De ahí proviene el nombre de *penalización*, pues se castiga la toma de un punto no factible y, dependiendo del tipo de penalización, se pueden diferenciar los distintos tipos de métodos que se presentan a continuación.

2.1. Penalización exterior

Este primer método reside en el diseño de una sucesión de subproblemas de optimización sin restricciones. La sucesión formada por las soluciones de los subproblemas converge a la solución del problema original (1). Con este motivo se crea una función $P : \mathbb{R}^n \rightarrow \mathbb{R}$, denominada *función de penalización*, y que viene caracterizada del siguiente modo:

$$P(\mathbf{v}) := \begin{cases} P(\mathbf{v}) > 0 & \text{si } \mathbf{v} \notin S \\ P(\mathbf{v}) = 0 & \text{si } \mathbf{v} \in S \end{cases}.$$

Con base en esta función se define una sucesión de problemas que pretenden minimizar al funcional $J_\varepsilon(\mathbf{v})$ sin restricciones, de tal manera que:

$$\min_{\mathbf{v} \in \mathbb{R}^n} J_\varepsilon(\mathbf{v}) = J(\mathbf{v}) + \frac{1}{\varepsilon} P(\mathbf{v}). \quad (6)$$

Entonces, dada una sucesión de elementos $\{\varepsilon_k\}_{k \in \mathbb{N}}$ positivos convergentes a cero, se obtiene una sucesión $\{\mathbf{u}^{(\varepsilon_n)}\}_{n \in \mathbb{N}}$ de soluciones de (6), la cual convergerá a un elemento \mathbf{u} , que es solución de (1). Para comenzar, se verán algunas propiedades de las funciones de penalización y de las soluciones $\mathbf{u}^{(\varepsilon_k)}$.

Lema 1 Dado $0 < 1/\varepsilon_1 < 1/\varepsilon_2$, se tienen las siguientes propiedades:

1. $J_{\varepsilon_1}(\mathbf{u}^{(\varepsilon_1)}) \leq J_{\varepsilon_2}(\mathbf{u}^{(\varepsilon_2)})$.
2. $J(\mathbf{u}^{(\varepsilon_1)}) \leq J(\mathbf{u}^{(\varepsilon_2)})$.
3. $P(\mathbf{u}^{(\varepsilon_1)}) \geq P(\mathbf{u}^{(\varepsilon_2)})$.

Demostración. Aplicando la definición de solución del problema (6), se tiene que:

$$J_{\varepsilon_1}(\mathbf{u}^{(\varepsilon_1)}) \leq J_{\varepsilon_1}(\mathbf{u}^{(\varepsilon_2)}) \leq J_{\varepsilon_2}(\mathbf{u}^{(\varepsilon_2)}) \leq J_{\varepsilon_2}(\mathbf{u}^{(\varepsilon_1)}). \quad (7)$$

Por tanto, ya se ha obtenido (1). Ahora, usando la propiedad (1) demostrada y (7), se llega a,

$$0 \leq J_{\varepsilon_1}(\mathbf{u}^{(\varepsilon_2)}) - J_{\varepsilon_2}(\mathbf{u}^{(\varepsilon_2)}) - [J_{\varepsilon_1}(\mathbf{u}^{(\varepsilon_1)}) - J_{\varepsilon_2}(\mathbf{u}^{(\varepsilon_1)})] = \frac{1}{\varepsilon_1} P(\mathbf{u}^{(\varepsilon_2)}) - \frac{1}{\varepsilon_2} P(\mathbf{u}^{(\varepsilon_2)}) - \left[\frac{1}{\varepsilon_1} P(\mathbf{u}^{(\varepsilon_1)}) - \frac{1}{\varepsilon_2} P(\mathbf{u}^{(\varepsilon_1)}) \right] = \left(\frac{1}{\varepsilon_1} - \frac{1}{\varepsilon_2} \right) (P(\mathbf{u}^{(\varepsilon_2)}) - P(\mathbf{u}^{(\varepsilon_1)})),$$

de lo que se deduce (3). Y, usando ahora (3) y (7), se tiene que,

$$J(\mathbf{u}^{(\varepsilon_1)}) \leq J(\mathbf{u}^{(\varepsilon_2)}) + \frac{1}{\varepsilon_1} (P(\mathbf{u}^{(\varepsilon_2)}) - P(\mathbf{u}^{(\varepsilon_1)})) < J(\mathbf{u}^{(\varepsilon_2)}),$$

lo cual demuestra (2) y finaliza la prueba. \square

Lema 2 Sea \mathbf{u} la solución del problema (1) considerado. Entonces, para cada $k \in \mathbb{N}$ se tiene que:

$$J(\mathbf{u}) \geq J_{\varepsilon_k}(\mathbf{u}^{(\varepsilon_k)}) \geq J(\mathbf{u}^{(\varepsilon_k)}).$$

Demostración. El resultado es inmediato considerando el siguiente desarrollo:

$$J(\mathbf{u}) = J(\mathbf{u}) + \frac{1}{\varepsilon_k} P(\mathbf{u}) \geq J(\mathbf{u}^{(\varepsilon_k)}) + \frac{1}{\varepsilon_k} P(\mathbf{u}^{(\varepsilon_k)}) \geq J(\mathbf{u}^{(\varepsilon_k)}).$$

□

Lema 3 Se considera el problema (1) y $\mathbf{u}^{(\varepsilon)}$ la solución al problema (6). Entonces, tomando $\delta = P(\mathbf{u}^{(\varepsilon)})$, se tiene que $\mathbf{u}^{(\varepsilon)}$ también es solución del problema:

$$\begin{aligned} & \min_{\mathbf{v} \in \mathbb{R}^n} J(\mathbf{v}) \\ & \text{Sujeto a:} \\ & P(\mathbf{v}) \leq \delta. \end{aligned} \tag{8}$$

Demostración. Para cualquier \mathbf{v} que satisfaga la condición $P(\mathbf{v}) \leq \delta$ se tiene que:

$$\begin{aligned} 0 & \leq \frac{1}{\varepsilon} (P(\mathbf{u}^{(\varepsilon)}) - P(\mathbf{v})) = J_{\varepsilon}(\mathbf{u}^{(\varepsilon)}) - J(\mathbf{u}^{(\varepsilon)}) - J_{\varepsilon}(\mathbf{v}) + J(\mathbf{v}) \\ & = [J_{\varepsilon}(\mathbf{u}^{(\varepsilon)}) - J_{\varepsilon}(\mathbf{v})] + J(\mathbf{v}) - J(\mathbf{u}^{(\varepsilon)}) \leq J(\mathbf{v}) - J(\mathbf{u}^{(\varepsilon)}) \Rightarrow J(\mathbf{v}) \geq J(\mathbf{u}^{(\varepsilon)}), \end{aligned}$$

obteniendo así la definición de mínimo y concluyendo así la demostración. □

Llegados a este punto, se está en condiciones de deducir el algoritmo del método. Entonces, al considerar el problema (1), teniendo en cuenta cómo se define la función de penalización, este se puede reescribir como:

$$\begin{aligned} & \min_{\mathbf{v} \in \mathbb{R}^n} J(\mathbf{v}) \\ & \text{Sujeto a:} \\ & P(\mathbf{v}) = 0 \end{aligned} \tag{9}$$

Ahora bien, tomando δ lo suficientemente pequeño, el problema (8) será una aproximación del problema original (1). Por tanto, aplicando el Lema 3, $\mathbf{u}^{(\varepsilon)}$ es una aproximación de \mathbf{u} (solución del problema inicial).

La idea básica de este método se basa en que, según se reduce el parámetro de penalización ε , se reduce el valor de $P(\mathbf{u}^{(\varepsilon)})$, como se vió en el Lema 1. Por este motivo, si se fija un valor de tolerancia δ para aproximar el problema (1) mediante (8), se puede establecer el Algoritmo 1.

Algoritmo 1 Penalización exterior

Se sigue el siguiente procedimiento:

PASO 0: Se toma $\varepsilon_0 > 0$, $\delta > 0$ y $k = 0$.

PASO 1: Se calcula $\mathbf{u}^{(\varepsilon_{k+1})}$, solución de (6).

PASO 2: Si la penalización $P(\mathbf{u}^{(\varepsilon_{k+1})}) < \delta$, entonces se finaliza. En caso contrario, se toma $\varepsilon_{k+1} = \varepsilon/10$, $k = k + 1$ y se vuelve al PASO 1.

Observación 1

La actualización del valor ε no tiene porqué ser la utilizada en el algoritmo. Esta puede variar en función de nuestros intereses, haciendo que la sucesión $\{\varepsilon_k\}_{k \in \mathbb{N}}$ decrezca con mayor o menor rapidez. Lo ideal es que la sucesión decrezca paulatinamente, para que la sucesión de soluciones no sufra cambios bruscos. En caso de construir una sucesión $\{\varepsilon_k\}_{k \in \mathbb{N}}$ que decrezca muy rápidamente, es posible que alguno de los subproblemas asociados no tenga solución o sea muy complicado calcularla, por lo que es preferible observar un comportamiento progresivo de la sucesión. Además, en la resolución de los subproblemas (6), se tendrá que usar algún método de resolución de problemas de optimización sin restricciones. En dichos métodos se debe aportar un iterante inicial. En este caso, un buen iterante inicial sería la solución $\mathbf{u}^{(\varepsilon_k)}$ calculada en la iteración anterior.

Por otro lado, nótese que la función de penalización debe ser elegida de tal forma que el funcional $J_\varepsilon(\mathbf{v})$ tenga mínimo finito, ya que si no es así, el algoritmo diverge. Esta situación se puede dar en casos donde el funcional objetivo decrezca según $\|\mathbf{v}\| \rightarrow +\infty$ y la función de penalización no crezca con la misma rapidez que decrece $J(\mathbf{v})$.

Ahora bien, una cuestión importante es el estudio de la convergencia del método. Para ello se introduce el Teorema 3.

Teorema 3

Dado el problema (1) con un funcional continuo. Entonces, cualquier punto límite de la sucesión $\{\mathbf{u}^{(\varepsilon_k)}\}$ generada por el Algoritmo 1 es solución del problema original (1).

Demostración. Suponga que $\{\mathbf{u}^{(\varepsilon_{k'})}\}$ es una subsucesión convergente con límite \mathbf{u} . Entonces, por la continuidad del funcional J , se tiene que:

$$\lim_{k' \rightarrow +\infty} J(\mathbf{u}^{(\varepsilon_{k'})}) = J(\mathbf{u}). \quad (10)$$

Denotando por J^* el valor del funcional en la solución del problema (1), y teniendo en cuenta los Lemas 1 y 2, la sucesión $\{J_{\varepsilon_{k'}}(\mathbf{u}^{(\varepsilon_{k'})})\}$ es creciente y está limitada superiormente por J^* . Entonces:

$$\lim_{k' \rightarrow +\infty} J_{\varepsilon_{k'}}(\mathbf{u}^{(\varepsilon_{k'})}) = q \leq J^*. \quad (11)$$

De manera que, extrayendo (10) de la ecuación (11) se tiene:

$$\lim_{k' \rightarrow +\infty} \frac{1}{\varepsilon_{k'}} P(\mathbf{u}^{(\varepsilon_{k'})}) = q - J(\mathbf{u}). \quad (12)$$

Como $P(\mathbf{u}^{(\varepsilon_{k'})}) \geq 0$ y $1/\varepsilon_{k'} \rightarrow +\infty$, por (12), implica que:

$$\lim_{k' \rightarrow +\infty} P(\mathbf{u}^{(\varepsilon_{k'})}) = 0.$$

Usando ahora la continuidad de $P(\mathbf{v})$, se obtiene que $P(\mathbf{u}) = 0$, por lo que \mathbf{u} es un punto admisible para el problema. Además, por el Lema 2, $J(\mathbf{u}^{(\varepsilon_{k'})}) \leq J^*$ y por lo tanto:

$$J(\mathbf{u}) = \lim_{k' \rightarrow +\infty} J(\mathbf{u}^{(\varepsilon_{k'})}) \leq J^*.$$

Por lo que \mathbf{u} es una solución óptima factible del problema original (1). □

Observación 2

Del Teorema 3 se deduce que si la sucesión generada por el Algoritmo 1 converge a un punto, este será solución del problema general (1). Sin embargo, esta sucesión no tiene porqué ser convergente siempre. Por ejemplo, puede darse que el mínimo de algún subproblema no sea finito y, por tanto, la sucesión comience a divergir. Por este motivo sería importante poder demostrar la existencia y unicidad de solución en los subproblemas sin restricciones del Algoritmo 1. Este hecho es bastante complicado en la práctica y se podría consultar [3] para conocer más acerca de esta temática.

Una cuestión interesante es cómo crear una función de penalización $P(\mathbf{v})$. En realidad, cualquier función que cumpla las propiedades dadas en el Lema 1, se puede usar. Sin embargo, algunas tiene mejores propiedades que otras. Una posible opción, es usar la función (13) denominada *penalización cuadrática*:

$$P(\mathbf{v}) = \sum_{i=1}^m (\max\{0, \varphi_i(\mathbf{v})\})^2 + \sum_{j=1}^p |\phi_j(\mathbf{v})|^2. \quad (13)$$

Tal y como está definida esta penalización, el funcional $J_\varepsilon(\mathbf{v})$ será derivable y continuo. Pero al no ser continua la segunda derivada, puede conllevar problemas si se utiliza un método de resolución de optimización sin restricciones que involucre a la segunda derivada del funcional objetivo.

Ejemplo 1

Para ejemplificar visualmente lo que produce la penalización, tomando la función de penalización (13), se va a aplicar ahora al siguiente problema unidimensional:

$$\begin{aligned} \min_{v \in \mathbb{R}} \quad & J(v) = e^v \\ \text{Sujeto a:} \quad & \varphi_1(v) = -x \\ & \varphi_2(v) = x - 1. \end{aligned} \quad (14)$$

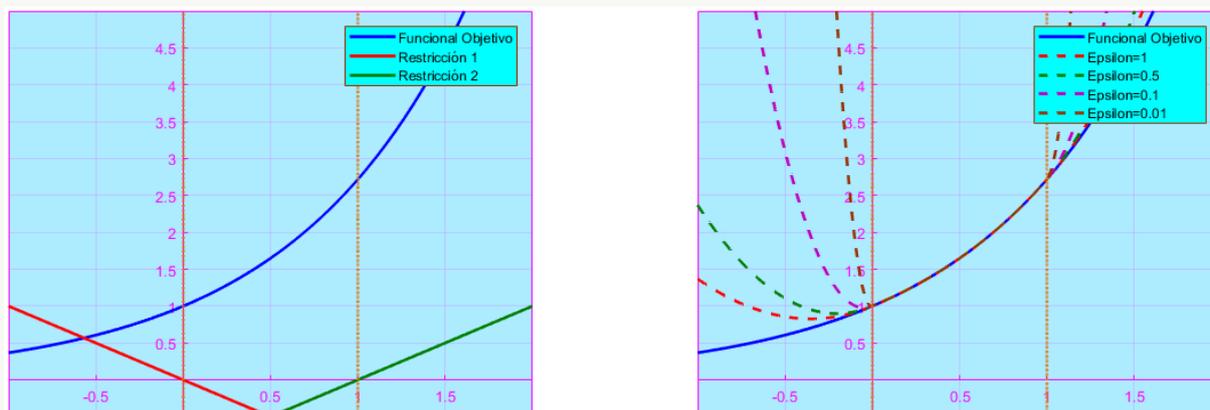


Figura 1: Representación gráfica del problema (14).

En la gráfica de la izquierda se muestra el funcional, las restricciones y el conjunto factible. Mientras que en la segunda gráfica se muestra el funcional objetivo $J(v)$ y los diferentes $J_\varepsilon(v)$, a medida que se afina el valor de ε , según se indica en las gráficas. Se observa que los funcionales $J_\varepsilon(v)$ aumentan su valor fuera de la región factible, mientras que se mantiene el funcional original dentro del conjunto admisible.

Una vez vista la intuición del método de forma gráfica, se puede introducir una visión práctica. En este sentido, se presenta el siguiente problema del que se mostrarán los resultados obtenidos

en algunas iteraciones del método de penalización exterior:

$$\begin{aligned} \min_{\mathbf{v} \in \mathbb{R}^2} \quad & J(\mathbf{v}) = -v_1^2 - v_2^2 \\ \text{Sujeto a:} \quad & \varphi_1(\mathbf{v}) = v_1^2 + v_2^2 - 1. \end{aligned} \quad (15)$$

Iteración k	ε_k	$\mathbf{u}^{(\varepsilon_k)}$	$P(\mathbf{u}^{(\varepsilon_k)})$	$J_{\varepsilon_k}(\mathbf{u}^{(\varepsilon_k)})$	$J(\mathbf{u}^{(\varepsilon_k)})$
0	10	(1.732, 1.732)	2.500489e+01	-3.500000e+00	-6.000489e+00
1	3.33	(1.154, 1.154)	2.777778e+00	-1.833333e+00	-2.666667e+00
2	1.11	(0.881, 0.881)	3.086419e-01	-1.277778e+00	-1.555556e+00
3	0.37	(0.769, 0.769)	3.429353e-02	-1.092593e+00	-1.185185e+00
4	0.123	(0.728, 0.728)	3.810393e-03	-1.030864e+00	-1.061728e+00
5	0.041	(0.714, 0.714)	4.233768e-04	-1.010288e+00	-1.020576e+00
6	0.013	(0.709, 0.709)	4.704181e-05	-1.003429e+00	-1.006859e+00

Tabla 1: Resultados obtenidos con el método de penalización exterior en el problema (15).

En la Tabla 1 se puede ver la convergencia hacia un punto de la circunferencia unidad (pues el mínimo es cualquier punto de la circunferencia unidad). Pero, además, se pueden apreciar las propiedades vistas en el Lema 1.

Observación 3

Los resultados mostrados en la Tabla 1 no son únicos. Dependiendo del algoritmo que se utilice en la resolución de los subproblemas de optimización sin restricciones se obtendrán unos valores u otros, pero en cualquier caso los comportamientos que se observan deben ser similares a los mostrados en la Tabla 1.

2.2. Penalización interior

Otro método similar al anterior es el método de la penalización interior o de la barrera. En este nuevo caso, se vuelve a usar una función de penalización a partir de la cual se definirá un nuevo funcional objetivo y, a partir del funcional, se construye una sucesión de subproblemas cuya sucesión de soluciones converge a la solución del problema inicial.

Sin embargo, se presentan diferencias respecto al caso anterior. En primer lugar, este tipo de método solo es aplicable a problemas de optimización, como el visto en (1), pero considerando únicamente las restricciones de desigualdad. Es decir, únicamente se aplicará el método al siguiente problema:

$$\begin{aligned} \min_{\mathbf{v} \in \mathbb{R}^n} \quad & J(\mathbf{v}) \\ \text{Sujeto a:} \quad & \varphi_i(\mathbf{v}) \leq 0 \quad i = 1, \dots, m. \end{aligned} \quad (16)$$

Entonces, siguiendo una filosofía similar al caso de la penalización exterior, se busca construir una sucesión de soluciones factibles, por lo que cada $\mathbf{u}^{(\varepsilon)} \in S$. Para introducirse en el método se comenzará por la definición de la función de penalización, que se define, para este método, como $P : \mathbb{R}^n \rightarrow \mathbb{R}^+$ tal que:

$$\lim_{d(\mathbf{v}, \partial S) \rightarrow 0} P(\mathbf{v}) = +\infty,$$

donde $d(\mathbf{v}, \partial S)$ denota la distancia del punto \mathbf{v} a la frontera de S , por lo tanto, la penalización aumenta según se aproxima a la frontera del conjunto admisible. Con base en esta penalización, se define el

siguiente problema:

$$\min_{\mathbf{v} \in \mathbb{R}^n} J_\varepsilon(\mathbf{v}) = J(\mathbf{v}) + \varepsilon P(\mathbf{v}), \quad (17)$$

que es un problema sin restricciones. A continuación se presentan varias propiedades de esta función de penalización que son análogas a las del caso anterior.

Lema 4 Dado $0 < \varepsilon_2 < \varepsilon_1$, se tiene:

1. $J_{\varepsilon_2}(\mathbf{u}^{(\varepsilon_2)}) \leq J_{\varepsilon_1}(\mathbf{u}^{(\varepsilon_1)})$.
2. $J(\mathbf{u}^{(\varepsilon_2)}) \leq J(\mathbf{u}^{(\varepsilon_1)})$.
3. $P(\mathbf{u}^{(\varepsilon_2)}) \geq P(\mathbf{u}^{(\varepsilon_1)})$.

Demostración. Teniendo en cuenta la definición de $J_\varepsilon(\mathbf{v})$ y de mínimo, se tiene:

$$J_{\varepsilon_2}(\mathbf{u}^{(\varepsilon_2)}) \leq J_{\varepsilon_2}(\mathbf{u}^{(\varepsilon_1)}) \leq J_{\varepsilon_1}(\mathbf{u}^{(\varepsilon_1)}) \leq J_{\varepsilon_1}(\mathbf{u}^{(\varepsilon_2)}), \quad (18)$$

de lo que se obtiene (1). Además:

$$\begin{aligned} 0 &\leq J_{\varepsilon_2}(\mathbf{u}^{(\varepsilon_1)}) - J_{\varepsilon_1}(\mathbf{u}^{(\varepsilon_1)}) - [J_{\varepsilon_2}(\mathbf{u}^{(\varepsilon_2)}) - J_{\varepsilon_1}(\mathbf{u}^{(\varepsilon_2)})] \\ &= \varepsilon_2 P(\mathbf{u}^{(\varepsilon_1)}) - \varepsilon_1 P(\mathbf{u}^{(\varepsilon_1)}) - [\varepsilon_2 P(\mathbf{u}^{(\varepsilon_2)}) - \varepsilon_1 P(\mathbf{u}^{(\varepsilon_2)})] = (\varepsilon_1 - \varepsilon_2) (P(\mathbf{u}^{(\varepsilon_2)}) - P(\mathbf{u}^{(\varepsilon_1)})) . \end{aligned}$$

Como $\varepsilon_1 - \varepsilon_2 > 0$, entonces se deduce (3). Y, finalmente, usando (18) y (3) se tiene,

$$J(\mathbf{u}^{(\varepsilon_2)}) \leq J(\mathbf{u}^{(\varepsilon_1)}) + \varepsilon_2 (P(\mathbf{u}^{(\varepsilon_1)}) - P(\mathbf{u}^{(\varepsilon_2)})) ,$$

de lo que se deduce (2). □

Lema 5 Siendo $\mathbf{u}^{(\varepsilon)}$ solución del problema (17) y tomando $\delta = P(\mathbf{u}^{(\varepsilon)})$, entonces $\mathbf{u}^{(\varepsilon)}$ es solución del problema

$$\begin{aligned} \min_{\mathbf{v} \in \mathbb{R}^n} J(\mathbf{v}) \\ \text{Sujeto a: } P(\mathbf{v}) \leq \delta . \end{aligned} \quad (19)$$

Demostración. Para todo $\mathbf{v} \in \mathbb{R}^n$ tal que $P(\mathbf{v}) \leq \delta$, se tiene:

$$\begin{aligned} 0 \leq \varepsilon (P(\mathbf{u}^{(\varepsilon)}) - P(\mathbf{v})) &= J_\varepsilon(\mathbf{u}^{(\varepsilon)}) - J(\mathbf{u}^{(\varepsilon)}) - J_\varepsilon(\mathbf{v}) + J(\mathbf{v}) = \\ [J_\varepsilon(\mathbf{u}^{(\varepsilon)}) - J_\varepsilon(\mathbf{v})] + J(\mathbf{v}) - J(\mathbf{u}^{(\varepsilon)}) &\leq J(\mathbf{v}) - J(\mathbf{u}^{(\varepsilon)}) \Rightarrow J(\mathbf{u}^{(\varepsilon)}) \leq J(\mathbf{v}) , \end{aligned}$$

de lo que se obtiene la definición de mínimo y, por tanto, lo que se quería probar. □

Cuando el valor de δ del Lema 5 es lo suficientemente grande, este sirve para aproximar el siguiente problema:

$$\begin{aligned} \min_{\mathbf{v} \in \mathbb{R}^n} J(\mathbf{v}) \\ \text{Sujeto a: } P(\mathbf{v}) < +\infty . \end{aligned} \quad (20)$$

Ahora bien, del modo en que se ha definido la función de penalización, la restricción del problema (20) es equivalente a que su solución sea factible para el problema inicial (1). Sin embargo, dicha

solución no se encontrará sobre la frontera de S , ya que en dichos puntos la penalización tiende a infinito.

Además, si $\varepsilon > 0$ es suficientemente pequeño y el δ del Lema 5 es lo bastante grande, entonces $\mathbf{u}^{(\varepsilon)}$ se encuentra dentro de la región factible S y aproxima el valor de la solución del problema original. Entonces, se puede deducir el Algoritmo 2.

Algoritmo 2 Penalización interior

Se aporta el siguiente pseudocódigo:

PASO 0: Se toma $\varepsilon_1 > 0$, $\delta > 0$ y $k = 1$.

PASO 1: Se calcula $\mathbf{u}^{(\varepsilon_{k+1})}$, solución de (17).

PASO 2: Si $\varepsilon_k P(\mathbf{u}^{(\varepsilon_{k+1})}) < \delta$, se finaliza. En caso contrario, se toma $\varepsilon_{k+1} = \varepsilon_k/10$, $k = k + 1$ y se vuelve al PASO 1.

Observación 4

Al igual que sucedía en el método de penalización exterior, el decrecimiento de ε en el PASO 2, no tiene que ser la indicada, sino que puede ser cualquiera. Lo ideal es que el decrecimiento, al igual que en la Observación 1, sea paulatina debido a que esto ayuda en la resolución de los subproblemas. Si el decrecimiento de ε es demasiado pronunciado, se corre el riesgo de que los algoritmos de resolución de los subproblemas no encuentren una solución finita, lo que conllevaría a no obtener información acerca de la sucesión de soluciones aportada por el Algoritmo 2.

Una vez visto el Algoritmo 2, se introduce a continuación el Teorema 4 que aporta resultados sobre la convergencia de dicho método.

Teorema 4

Sea $J(\mathbf{v})$ un funcional limitado inferiormente en la región factible S . Entonces, el algoritmo de penalización interior termina de forma finita para $\delta > 0$ y, cuando no lo hace, entonces se verifica que:

1. $\lim_{k \rightarrow \infty} \varepsilon_k P(\mathbf{u}^{(k+1)}) = 0$,
2. $\lim_{k \rightarrow \infty} J(\mathbf{u}^{(k+1)}) = \inf_{\mathbf{v} \in \text{Int}(S)} J(\mathbf{v})$.

Y, además, cualquier punto de acumulación de la sucesión $\{\mathbf{u}_k\}_{k \in \mathbb{N}}$ es solución del problema (16).

Demostración. Solo será necesario probar (1) y (2) cuando el algoritmo no termina de forma finita. Para ello, se toma $\eta > 0$ arbitrario, por lo que existirá un elemento (que depende de η), al que se denotará por $\mathbf{u}_\eta \in \text{Int}(S)$, de tal forma que:

$$J(\mathbf{u}_\eta) < \inf_{\mathbf{v} \in \text{Int}(S)} J(\mathbf{v}) + \frac{\eta}{2}, \quad (21)$$

Entonces, como $\{\varepsilon_k\} \rightarrow 0$ y $\mathbf{u}_\eta \in \text{Int}(S)$, debe existir un $\bar{k} \in \mathbb{N}$ tal que:

$$\varepsilon_k P(\mathbf{u}_\eta) < \frac{\eta}{2}, \quad \forall k \geq \bar{k}. \quad (22)$$

Usando ahora que $\mathbf{u}^{(k+1)}$ es el mínimo del funcional $J_{\varepsilon_k}(\mathbf{v})$:

$$J_{\varepsilon_k}(\mathbf{u}^{(k+1)}) \leq J_{\varepsilon_k}(\mathbf{u}_\eta) \Rightarrow \varepsilon_k P(\mathbf{u}^{(k+1)}) \leq J(\mathbf{u}_\eta) + \varepsilon_k P(\mathbf{u}_\eta) - J(\mathbf{u}^{(k+1)}) . \quad (23)$$

Por lo que usando (21), (22) y (23), se obtiene que:

$$\begin{aligned} \varepsilon_k P(\mathbf{u}^{(k+1)}) &\leq J(\mathbf{u}_\eta) + \varepsilon_k P(\mathbf{u}_\eta) - J(\mathbf{u}^{(k+1)}) \\ &\leq \inf_{\mathbf{v} \in \text{Int}(S)} J(\mathbf{v}) + \frac{\eta}{2} + \frac{\eta}{2} - J(\mathbf{u}^{(k+1)}) \leq \eta, \quad \forall k \geq \bar{k}. \end{aligned}$$

Como $\eta > 0$ es arbitrario, se concluye (1). Ahora, para probar (2), utilizando la desigualdad anterior, se tiene:

$$J(\mathbf{u}^{(k+1)}) \leq J(\mathbf{u}_\eta) + \varepsilon_k P(\mathbf{u}_\eta) \leq \inf_{\mathbf{v} \in \text{Int}(S)} J(\mathbf{v}) + \eta, \quad \forall k \geq \bar{k}.$$

Como $\eta > 0$ es arbitrario, se concluye (2). □

Una vez vistos estos resultados, es natural plantear la cuestión sobre cómo tomar la función de penalización $P(\mathbf{v})$. Con este motivo, se plantean las opciones reflejadas en (24) que son las más generales:

$$P(\mathbf{v}) = -\sum_{i=1}^m \frac{1}{\varphi_i(\mathbf{v})} \quad \text{o} \quad P(\mathbf{v}) = -\sum_{i=1}^m \frac{1}{\log(-\varphi_i(\mathbf{v}))}. \quad (24)$$

Para mostrar el efecto de la penalización interior de un modo visual, se considera el siguiente problema:

$$\begin{aligned} \min_{\mathbf{v} \in \mathbb{R}} \quad & J(\mathbf{v}) = v^4 \\ \text{Sujeto a:} \quad & \begin{cases} \varphi_1(\mathbf{v}) = v - 1 \\ \varphi_2(\mathbf{v}) = -v - 1 \end{cases} \end{aligned} \quad (25)$$

del que se obtiene los siguientes resultados.

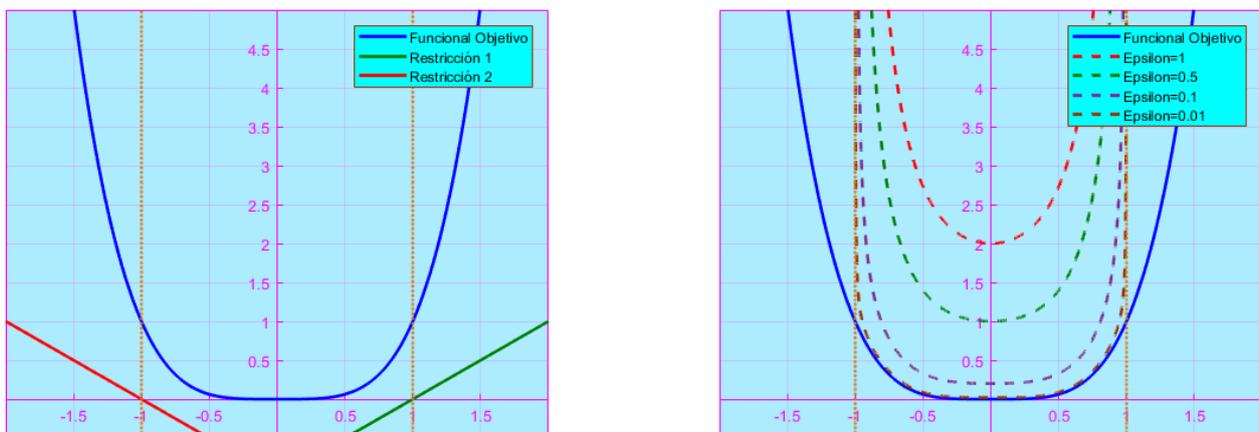


Figura 2: Representación gráfica del problema (25).

En la primera gráfica de la Figura 2 se muestra el funcional objetivo con las restricciones y el conjunto factible. Luego, en la segunda gráfica se muestra de nuevo el funcional objetivo y, además, los funcionales $J_\varepsilon(\mathbf{v})$ para distintos valores de ε . De este modo, se observa la convergencia de los funcionales

$J_\varepsilon(\mathbf{v})$ al funcional del problema original dentro de la región factible. Sin embargo, existe una peculiaridad, dado que en la frontera del conjunto los valores de $J_\varepsilon(\mathbf{v})$ tienden al infinito, lo que muestra que el método sólo convergerá a un punto del interior del conjunto admisible.

De este modo, se muestra visualmente cómo es necesario inicializar el método desde un punto factible, además de que, para restricciones de igualdad, el método no serviría dado que los funcionales $J_\varepsilon(\mathbf{v})$ no estarían definidos en esos puntos.

3. Método del lagrangiano aumentado

Una vez vistos los métodos basados en penalización, se puede observar la necesidad de que el parámetro ε_k tienda a cero cuando k tiende a infinito. Este hecho es una gran desventaja desde el punto de vista numérico, debido a que según se toma ε_k más pequeño, la resolución numérica de los subproblemas generados puede resultar más compleja de lo previsto.

Para solventar esta dificultad se introduce el concepto de Lagrangiano aumentado. Este concepto se deriva de la función Lagrangiano y permite solventar ese problema que presentan los métodos de penalización. Para ello, se planteará, en primer lugar, el concepto de Lagrangiano aumentado para el problema con restricciones de igualdad. Esto es, se considerará en primera estancia el siguiente problema de optimización:

$$\begin{aligned} \min_{\mathbf{v} \in \mathbb{R}^n} \quad & J(\mathbf{v}) \\ \text{Sujeto a:} \quad & \phi_j(\mathbf{v}) = 0 \quad j = 1, \dots, p. \end{aligned} \quad (26)$$

Definición 4

Se define la función Lagrangiano aumentado para el problema (26) como sigue:

$$\mathcal{L}_\varepsilon(\mathbf{v}, \boldsymbol{\xi}) = J(\mathbf{v}) + \sum_{j=1}^p \xi_j \phi_j(\mathbf{v}) + \frac{1}{2\varepsilon} \sum_{j=1}^p (\phi_j(\mathbf{v}))^2 = J(\mathbf{v}) + \boldsymbol{\xi}^\top \Phi(\mathbf{v}) + \frac{1}{2\varepsilon} \|\Phi(\mathbf{v})\|^2.$$

Observación 5

Es fácil observar cómo la definición anterior es precisamente el Lagrangiano del problema de minimización:

$$\begin{aligned} \min_{\mathbf{v} \in \mathbb{R}^n} \quad & J(\mathbf{v}) + \frac{1}{2\varepsilon} \|\Phi(\mathbf{v})\|^2 \\ \text{Sujeto a:} \quad & \Phi(\mathbf{v}) = \mathbf{0}, \end{aligned}$$

que tiene el mismo mínimo local que el problema (26).

El Lema 6 y el Teorema 5 serán la base para justificar el método del Lagrangiano aumentado.

Lema 6 Sean \mathbf{P} y \mathbf{Q} dos matrices simétricas. Se asume que \mathbf{Q} es semidefinida positiva y que \mathbf{P} es definida positiva en el espacio nulo de \mathbf{Q} , es decir, $\mathbf{d}^T \mathbf{P} \mathbf{d} > 0$ para todo $\mathbf{d} \neq \mathbf{0}$, con $\mathbf{d}^T \mathbf{Q} \mathbf{d} = 0$. Entonces, existe un escalar $\bar{\rho}$ tal que:

$$\mathbf{P} + \rho \mathbf{Q} \text{ es definida positiva } \forall \rho > \bar{\rho}.$$

Demostración. Se puede ver en [1, Lema 3.2.1]. □

Teorema 5 Condición suficiente de Lagrange

Se considera el problema (26), bajo las hipótesis de derivabilidad del funcional objetivo y de las restricciones. Entonces, suponiendo que $\mathbf{u} \in S$ es un punto que verifica las condiciones KKT con multiplicadores de Lagrange λ , y que cumple que,

$$\mathbf{d}^T \mathbf{H} \mathcal{L}_{\mathbf{v}}(\mathbf{u}, \lambda) \mathbf{d} > 0 \text{ para todo } \mathbf{d} \neq \mathbf{0}, \text{ con } \Phi(\mathbf{u})' \mathbf{d} = \mathbf{0},$$

entonces \mathbf{u} es un mínimo local estricto del problema (26).

Demostración. Se puede consultar en [1, Proposición 3.2.1]. □

Llegado a este punto, se está en condiciones de demostrar el siguiente resultado que justifica lo que se ha comentado hasta el momento.

Teorema 6

Dado el problema (26), si $\mathbf{u} \in S$ cumple las condiciones suficientes de segundo orden del Teorema 5 con su multiplicador de Lagrange asociado λ . Entonces, existe un valor $\bar{\varepsilon}$ tal que \mathbf{u} es un mínimo local de la función $\mathcal{L}_{\varepsilon}(\mathbf{v}, \lambda)$, para todo $\varepsilon < \bar{\varepsilon}$.

Demostración. Calculando el gradiente y la matriz hessiana de la función $\mathcal{L}_{\varepsilon}(\mathbf{v}, \xi)$ respecto de la variable \mathbf{v} , se tiene:

$$\nabla_{\mathbf{v}} \mathcal{L}_{\varepsilon}(\mathbf{v}, \xi) = \nabla J(\mathbf{v}) + \Phi'(\mathbf{v})^T \left(\xi + \frac{1}{\varepsilon} \Phi(\mathbf{v}) \right),$$

y

$$\mathbf{H}_{\mathbf{v}} \mathcal{L}_{\varepsilon}(\mathbf{v}, \xi) = \mathbf{H} J(\mathbf{v}) + \sum_{j=1}^p \left(\xi_j + \frac{1}{\varepsilon} \phi_j(\mathbf{v}) \right) \mathbf{H} \phi_j(\mathbf{v}) + \frac{1}{\varepsilon} \Phi'(\mathbf{v})^T \Phi'(\mathbf{v}).$$

En particular, si \mathbf{u} y λ satisfacen las condiciones del Teorema 5, se tiene:

$$\nabla_{\mathbf{v}} \mathcal{L}_{\varepsilon}(\mathbf{u}, \lambda) = \nabla J(\mathbf{u}) + \Phi'(\mathbf{u})^T \left(\lambda + \frac{1}{\varepsilon} \Phi(\mathbf{u}) \right) = \nabla_{\mathbf{v}} \mathcal{L}(\mathbf{u}, \lambda) = \mathbf{0}, \quad (27)$$

$$\begin{aligned} \mathbf{H}_{\mathbf{v}} \mathcal{L}_{\varepsilon}(\mathbf{u}, \lambda) &= \mathbf{H} J(\mathbf{u}) + \sum_{j=1}^p \left(\lambda_j + \frac{1}{\varepsilon} \phi_j(\mathbf{u}) \right) \mathbf{H} \phi_j(\mathbf{u}) + \frac{1}{\varepsilon} \Phi(\mathbf{u}) \Phi(\mathbf{u})^T \\ &= \mathbf{H}_{\mathbf{v}} \mathcal{L}(\mathbf{u}, \lambda) + \frac{1}{\varepsilon} \Phi'(\mathbf{u})^T \Phi'(\mathbf{u}). \end{aligned}$$

Usando la condición suficiente del teorema, se sabe que:

$$\mathbf{d}^T \mathbf{H} \mathcal{L}(\mathbf{u}, \lambda) \mathbf{d} > 0 \text{ para todo } \mathbf{d} \neq \mathbf{0} \text{ tal que } \mathbf{d}^T \Phi'(\mathbf{u})^T \Phi'(\mathbf{u}) \mathbf{d} = 0.$$

Entonces, haciendo uso del Lema 6, implica que existe un $\bar{\varepsilon}$ tal que:

$$H_v \mathcal{L}_\varepsilon(\mathbf{u}, \lambda) \text{ es definida positiva para todo } \varepsilon < \bar{\varepsilon}. \quad (28)$$

Teniendo en cuenta (27) y (28), se está en hipótesis del Teorema 5, por lo que se verifica que para todo $\varepsilon < \bar{\varepsilon}$, \mathbf{u} es un mínimo local de la función $\mathcal{L}_\varepsilon(\mathbf{v}, \lambda)$. \square

Por tanto, si se conocen de antemano los valores de los multiplicadores de Lagrange λ bastaría con un número finito de subproblemas sin restricciones que minimicen al funcional $\mathcal{L}_\varepsilon(\mathbf{v}, \lambda)$ para llegar a obtener la solución del problema (26). Sin embargo, esta es una situación idílica, debido a que en la práctica dichos multiplicadores no se conocerán. Por este motivo, se recurre a una estimación de los multiplicadores. Para ver su deducción, se considera el siguiente problema de minimización que, para un valor fijo de λ , es equivalente al problema (26):

$$\min_{\mathbf{v} \in \mathbb{R}^n} J(\mathbf{v}) + \sum_{j=1}^p \lambda_j \phi_j(\mathbf{v})$$

$$\text{Sujeto a:} \\ \Phi(\mathbf{v}) = 0,$$

que al aplicarle el método de penalización exterior, tomando como función penalizadora $P(\mathbf{v}) = 1/2 \sum_{j=1}^p (\phi_j(\mathbf{v}))^2$, se obtiene el siguiente subproblema en cada iteración:

$$\min_{\mathbf{v} \in \mathbb{R}^n} J(\mathbf{v}) + \sum_{j=1}^p \lambda_j \phi_j(\mathbf{v}) + \frac{1}{2\varepsilon} \sum_{j=1}^p (\phi_j(\mathbf{v}))^2.$$

Al suponer que el valor de los λ_j varía, entonces los subproblemas también cambian. Pero, suponiendo que se busca un punto estacionario, entonces:

$$\nabla J(\mathbf{v}) + \sum_{j=1}^p \lambda_j \nabla \phi_j(\mathbf{v}) + \frac{1}{\varepsilon} \sum_{j=1}^p \phi_j(\mathbf{v}) \nabla \phi_j(\mathbf{v}) = 0 \Leftrightarrow \nabla J(\mathbf{v}) + \sum_{j=1}^p \left(\lambda_j + \frac{1}{\varepsilon} \phi_j(\mathbf{v}) \right) \nabla \phi_j(\mathbf{v}) = 0.$$

De este modo, teniendo en cuenta las condiciones KKT del problema original (26), para un valor fijo de λ se observa que un buen estimador de los multiplicadores de Lagrange sería $\lambda + \frac{1}{\varepsilon} \Phi(\mathbf{v})$. Por tanto, se puede plantear el siguiente algoritmo basado en el Lagrangiano aumentado para la resolución de los problemas de tipo (26).

Algoritmo 3 Método de Lagrangiano aumentado

Se procede con el siguiente método iterativo:

PASO 0: Se toma un punto inicial, $\varepsilon_0 > 0$, $\delta > 0$ y $\lambda^{(0)} \in \mathbb{R}^p$.

PASO 1: Se toma como $\mathbf{u}^{(k)}$ la solución del problema de minimización sin restricciones definido por el funcional $\mathcal{L}_{\varepsilon_k}(\mathbf{v}, \lambda^{(k)})$.

PASO 2: Si $\|\nabla_v \mathcal{L}_{\varepsilon_k}(\mathbf{u}^{(k)}, \lambda^{(k)})\|_\infty < \delta$, entonces se para. En caso contrario, se toma $\varepsilon_{k+1} = 1/10\varepsilon_k$, $\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)}$ y $\lambda^{(k+1)} = \lambda^{(k)} + \frac{1}{\varepsilon_k} \Phi(\mathbf{u}^{(k)})$ para volver al PASO 1.

Observación 6

Como se ha comentado en los anteriores métodos de penalización, δ es el parámetro de tolerancia, un valor muy pequeño que permite comprobar que la solución obtenida es factible o está próxima al conjunto admisible. Luego, si este criterio de parada no se cumple, se pasa a reducir el valor de ε , que en este caso se reduce multiplicando por 1/10. Lo ideal es que el decrecimiento del valor de ε , en cada iteración, sea paulatino y no muy pronunciado.

Una vez visto el Algoritmo 3, es necesario probar que efectivamente converge y que la deducción del método es correcta. Para ello, se introducen los siguientes resultados.

Teorema 7

Dado el problema (26), se supone que el funcional objetivo y las restricciones son continuas. Si $\mathbf{u}^{(k)}$ es la solución global de minimizar sin restricciones $\mathcal{L}_{\varepsilon_k}(\mathbf{v}, \boldsymbol{\lambda}^{(k)})$, donde $\{\boldsymbol{\lambda}^{(k)}\}_{k \in \mathbb{N}}$ es una sucesión acotada y $\{\varepsilon_k\} \rightarrow 0$ con $0 < \varepsilon_{k+1} < \varepsilon_k$, entonces cualquier punto límite de la sucesión $\{\mathbf{u}^{(k)}\}$ es un mínimo global del problema original (26).

Demostración. Sea \mathbf{u} el límite de la sucesión $\{\mathbf{u}^{(k)}\}_{k \in \mathbb{N}}$. Como $\mathbf{u}^{(k)}$ se define como el mínimo global de $\mathcal{L}_{\varepsilon_k}(\mathbf{v}, \boldsymbol{\lambda}^{(k)})$, entonces:

$$\mathcal{L}_{\varepsilon_k}(\mathbf{u}^{(k)}, \boldsymbol{\lambda}^{(k)}) \leq \mathcal{L}_{\varepsilon_k}(\mathbf{v}, \boldsymbol{\lambda}^{(k)}), \quad \forall \mathbf{v} \in \mathbb{R}^n. \quad (29)$$

Tomando como χ el valor del funcional objetivo en el óptimo para el problema (26), entonces:

$$\chi = \min_{\mathbf{v} \in S} J(\mathbf{v}) = \min_{\mathbf{v} \in S} \left\{ J(\mathbf{v}) + \left(\boldsymbol{\lambda}^{(k)} \right)^T \Phi(\mathbf{v}) + \frac{1}{2\varepsilon_k} \|\Phi(\mathbf{v})\|^2 \right\} = \min_{\mathbf{v} \in S} \mathcal{L}_{\varepsilon_k}(\mathbf{v}, \boldsymbol{\lambda}^{(k)}).$$

Por lo que, tomando el mínimo en el extremo derecho de la ecuación (29) en $\mathbf{v} \in S$, se obtiene:

$$\mathcal{L}_{\varepsilon_k}(\mathbf{u}^{(k)}, \boldsymbol{\lambda}^{(k)}) = J(\mathbf{u}^{(k)}) + \left(\boldsymbol{\lambda}^{(k)} \right)^T \Phi(\mathbf{u}^{(k)}) + \frac{1}{2\varepsilon_k} \|\Phi(\mathbf{u}^{(k)})\|^2 \leq \chi. \quad (30)$$

Ahora bien, la sucesión $\{\boldsymbol{\lambda}^{(k)}\}_{k \in \mathbb{N}}$ es acotada y tiene como límite $\boldsymbol{\lambda}$. Tomando el límite superior en la ecuación (30) y usando la continuidad de las funciones $J(\mathbf{v})$ y $\Phi(\mathbf{v})$, se obtiene que:

$$J(\mathbf{u}) + \boldsymbol{\lambda}^T \Phi(\mathbf{u}) + \limsup_{k \rightarrow \infty} \frac{1}{2\varepsilon_k} \|\Phi(\mathbf{u}^{(k)})\|^2 \leq \chi.$$

Como $\|\Phi(\mathbf{u}^{(k)})\|^2 \rightarrow 0$ y $\varepsilon_k \rightarrow \infty$, se deduce que $\Phi(\mathbf{u}^{(k)}) \rightarrow 0$ y, por tanto, $\Phi(\mathbf{u}) = 0$. Por consiguiente, \mathbf{u} es factible y además $J(\mathbf{u}) \leq \chi$, por lo que \mathbf{u} es el óptimo del problema (26). \square

Sin embargo, este resultado implica que el mínimo de la función Lagrangiano aumentado es calculado de forma exacta. Pero esto no sucede numéricamente, por lo que se necesita un resultado más fuerte. En este sentido, se introduce el Teorema 8.

Definición 5

Dado el problema (26), un punto \mathbf{u} se dirá que es regular si el conjunto formado por los gradientes de las restricciones en dicho punto es linealmente independiente. Esto es que la matriz $\Phi'(\mathbf{u})$ tenga rango p .

Teorema 8

Sea $\mathbf{u}^{(k)}$ tal que $\|\nabla_{\mathbf{v}} \mathcal{L}_{\varepsilon_k}(\mathbf{v}, \boldsymbol{\lambda}^{(k)})\| \leq \delta_k$, siendo la sucesión $\{\boldsymbol{\lambda}^{(k)}\}_{k \in \mathbb{N}}$ acotada, $\{\varepsilon_k\}_{k \in \mathbb{N}} \rightarrow 0$ decreciente y $\{\delta_k\}_{k \in \mathbb{N}} \rightarrow 0$. Entonces, si la sucesión de $\{\mathbf{u}^{(k)}\}_{k \in \mathbb{N}} \rightarrow \mathbf{u}$ con \mathbf{u} un punto regular, entonces \mathbf{u} es un punto KKT para el problema (26) y, además, $\left\{ \boldsymbol{\lambda}^{(k)} + \frac{1}{\varepsilon_k} \Phi(\mathbf{u}^{(k)}) \right\}_{k \in \mathbb{N}} \rightarrow \boldsymbol{\lambda}$, siendo $\boldsymbol{\lambda}$ los multiplicadores de Lagrange asociados a \mathbf{u} .

Demostración. Sin pérdida de generalidad se supone que $\{\mathbf{u}^{(k)}\}_{k \in \mathbb{N}} \rightarrow \mathbf{u}$ y se define, para todo $k \in \mathbb{N}$, $\tilde{\boldsymbol{\lambda}}^{(k)} = \boldsymbol{\lambda}^{(k)} + 1/\varepsilon_k \Phi(\mathbf{u}^{(k)})$. Entonces se obtiene que:

$$\begin{aligned} \nabla_{\mathbf{v}} \mathcal{L}_{\varepsilon_k}(\mathbf{u}^{(k)}, \boldsymbol{\lambda}^{(k)}) &= \nabla J(\mathbf{u}^{(k)}) + \Phi'(\mathbf{u}^{(k)})^T \left(\boldsymbol{\lambda}^{(k)} + \frac{1}{\varepsilon_k} \Phi(\mathbf{u}^{(k)}) \right) \\ &= \nabla J(\mathbf{u}^{(k)}) + \Phi'(\mathbf{u}^{(k)})^T \tilde{\boldsymbol{\lambda}}^{(k)}. \end{aligned} \quad (31)$$

Como \mathbf{u} es regular, $\Phi'(\mathbf{u})$ tiene rango p y $\Phi'(\mathbf{u}^{(k)})$ tiene rango p para todo k suficientemente grande. Sin pérdida de generalidad se asume que $\Phi'(\mathbf{u}^{(k)})$ tiene rango p para todo k . Entonces, multiplicando la ecuación (31) por $\left(\Phi'(\mathbf{u}^{(k)}) (\Phi'(\mathbf{u}^{(k)}))^T \right)^{-1} \Phi'(\mathbf{u}^{(k)})$ se obtiene que:

$$\tilde{\boldsymbol{\lambda}}^{(k)} = \left(\Phi'(\mathbf{u}^{(k)}) (\Phi'(\mathbf{u}^{(k)}))^T \right)^{-1} \Phi'(\mathbf{u}^{(k)}) \nabla J(\mathbf{u}^{(k)}) \left(\nabla_{\mathbf{v}} \mathcal{L}_{\varepsilon_k}(\mathbf{u}^{(k)}, \boldsymbol{\lambda}^{(k)}) - \nabla J(\mathbf{u}^{(k)}) \right).$$

Luego, por hipótesis, $\nabla_{\mathbf{v}} \mathcal{L}_{\varepsilon_k}(\mathbf{u}^{(k)}, \boldsymbol{\lambda}^{(k)}) \rightarrow 0$, entonces $\tilde{\boldsymbol{\lambda}}^{(k)} \rightarrow \boldsymbol{\lambda}$, donde:

$$\boldsymbol{\lambda} = - \left(\Phi'(\mathbf{u}^{(k)}) (\Phi'(\mathbf{u}^{(k)}))^T \right)^{-1} \Phi'(\mathbf{u}^{(k)}) \nabla J(\mathbf{u}^{(k)}).$$

Nuevamente, como $\nabla_{\mathbf{v}} \mathcal{L}_{\varepsilon_k}(\mathbf{u}^{(k)}, \boldsymbol{\lambda}^{(k)}) \rightarrow 0$, entonces de la ecuación (31) se obtiene que

$$\nabla J(\mathbf{u}^{(k)}) + \Phi'(\mathbf{u}^{(k)})^T \tilde{\boldsymbol{\lambda}}^{(k)} = \mathbf{0}.$$

Ahora bien, como $\{\boldsymbol{\lambda}^{(k)}\}_{k \in \mathbb{N}}$ es acotada y $\boldsymbol{\lambda}^{(k)} + 1/\varepsilon_k \Phi(\mathbf{u}^{(k)}) \rightarrow \boldsymbol{\lambda}$, esto implica que $\{1/\varepsilon_k \Phi(\mathbf{u}^{(k)})\}$ es acotada. Pero como $1/\varepsilon_k \rightarrow +\infty$, se deduce que $\Phi(\mathbf{u}) = \mathbf{0}$. \square

Por tanto, este resultado muestra que si el punto \mathbf{u} , al que converge la sucesión creada por el método del Lagrangiano aumentado, es regular y los multiplicadores de Lagrange convergen a un valor finito, entonces, \mathbf{u} es un punto KKT. Por consiguiente, dicho punto es un buen candidato a ser mínimo del problema (26).

Este método ha sido desarrollado para un problema del tipo (26), pero sería interesante desarrollarlo para resolver un problema general del tipo (1). Con este objetivo, se pretende reescribir el problema (1) de una forma equivalente en la que sólo haya restricciones de igualdad. Así pues, se puede introducir el problema,

$$\begin{aligned} &\underset{\mathbf{v} \in \mathbb{R}^n, \mathbf{z} \in \mathbb{R}^m}{\text{mín}} && J(\mathbf{v}) \\ &\text{Sujeto a:} && \\ &\left\{ \begin{array}{ll} \varphi_i(\mathbf{v}) + z_i^2 = 0 & i = 1, \dots, m \\ \phi_j(\mathbf{v}) = 0 & j = 1, \dots, p, \end{array} \right. && \end{aligned} \quad (32)$$

donde z_i , $i = 1, \dots, m$, son variables adicionales. Esta transformación del problema original (1) en uno de este estilo de tipo (26), se puede consultar en [1, págs. 286-287].

De este modo, se puede aplicar el algoritmo del Lagrangiano aumentado al problema (32) para así obtener un punto KKT de dicho problema y, en consecuencia, del problema original (1). En esta línea, se plantea minimizar el Lagrangiano aumentado asociado al problema (32), que sería el siguiente:

$$\mathcal{L}_\varepsilon(\mathbf{v}, \mathbf{z}, \boldsymbol{\xi}, \boldsymbol{\nu}) = J(\mathbf{v}) + \boldsymbol{\xi}^\top \Phi(\mathbf{v}) + \frac{1}{2\varepsilon} \|\Phi(\mathbf{v})\|^2 + \sum_{i=1}^m \left\{ \nu_i (\varphi_i(\mathbf{v}) + z_i^2) + \frac{1}{2\varepsilon} |\varphi_i(\mathbf{v}) + z_i^2|^2 \right\}, \quad (33)$$

Para resolver el problema (1) se aplicará el Algoritmo 3, tomando como función Lagrangiano aumentado la (33). Además, a esta función se le pueden aplicar los resultados anteriores y, por tanto, si se conocen los valores exactos de los multiplicadores de Lagrange ($\boldsymbol{\lambda}$) y de KKT ($\boldsymbol{\mu}$), se puede aplicar el Teorema 6, que nos permite concluir que bastarían un número finito de subproblemas sin restricciones que minimicen el funcional $\mathcal{L}_\varepsilon(\mathbf{v}, \mathbf{z}, \boldsymbol{\lambda}, \boldsymbol{\mu})$, para hallar la solución del problema (1).

Sin embargo, bastaría con resolver el problema de minimización respecto de la variable $\mathbf{v} \in \mathbb{R}^n$, debido a que la minimización referente a la variable auxiliar $\mathbf{z} \in \mathbb{R}^m$, se puede calcular analíticamente. Pues bien, minimizando en primer lugar $\mathcal{L}_\varepsilon(\mathbf{v}, \mathbf{z}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ respecto de \mathbf{z} , es equivalente a minimizar:

$$\min_{w_i \geq 0} \left\{ \mu_i (\varphi_i(\mathbf{v}) + w_i) + \frac{1}{2\varepsilon} (\varphi_i(\mathbf{v}) + w_i)^2 \right\}, \quad \forall i = 1, \dots, m,$$

que es una función cuadrática en w_i . Su mínimo, teniendo en cuenta las restricciones, es $w_i^* = \max\{0, \tilde{w}_i\}$, donde \tilde{w}_i es el mínimo del problema sin restricciones y que, por tanto, cumple que anula la derivada, es decir,

$$\mu_i + 1/\varepsilon (\varphi_i(\mathbf{v}) + \tilde{w}_i) = 0,$$

luego:

$$w_i^* = \max\{0, -(\varepsilon \mu_i + \varphi_i(\mathbf{v}))\}. \quad (34)$$

Una vez hallado el mínimo respecto de la variable $\mathbf{z} \in \mathbb{R}^m$, se puede substituir la expresión (34) en la ecuación (33), de tal modo que, para los valores fijos de los multiplicadores de Lagrange y KKT, se tiene que minimizar respecto de $\mathbf{v} \in \mathbb{R}^n$ la siguiente función,

$$\mathcal{L}_\varepsilon(\mathbf{v}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = J(\mathbf{v}) + \boldsymbol{\lambda}^\top \Phi(\mathbf{v}) + \frac{1}{2\varepsilon} \|\Phi(\mathbf{v})\|^2 + \frac{\varepsilon}{2} \sum_{i=1}^m \left\{ \left(\max\left\{0, \mu_i + \frac{1}{\varepsilon} \varphi_i(\mathbf{v})\right\} \right)^2 - \mu_i^2 \right\}, \quad (35)$$

que se considerará como la función *Lagrangiano aumentado* para el problema (1). Entonces, llegado este punto, con seguir el mismo procedimiento del Algoritmo 3, con el funcional (35), se obtendría numéricamente un punto KKT para el problema (1).

Claro que, en este caso, se saben los valores fijos de los multiplicadores, que se desconocen en la realidad. Para ello, se deben aproximar mediante un procedimiento iterativo. En el caso de los multiplicadores de Lagrange ya se dedujo anteriormente, pero teniendo en cuenta la conversión hecha para el caso de restricciones de desigualdad, haciendo el mismo procedimiento, se llega a que sus actualizaciones se corresponden respectivamente con los valores:

$$\left\{ \boldsymbol{\lambda}^{(k)} + \frac{1}{\varepsilon_k} \Phi(\mathbf{u}^{(k)}) \right\} \quad \text{y} \quad \max\left\{0, \boldsymbol{\mu}^{(k)} + \frac{1}{\varepsilon_k} \Psi(\mathbf{u}^{(k)})\right\}$$

Hechas estas conclusiones, se puede reescribir el algoritmo del Lagrangiano aumentando para el caso del problema genérico (1). Entonces, se puede concluir el siguiente método:

Algoritmo 4 Método del Lagrangiano aumentado

Se procede con el siguiente método iterativo:

PASO 0: Se toma un punto inicial, $\varepsilon_0 > 0$, $\delta > 0$ y $\lambda^{(0)} \in \mathbb{R}^p$.

PASO 1: Se toma como $\mathbf{u}^{(k)}$ la solución del problema de minimización sin restricciones definido por el funcional $\mathcal{L}_{\varepsilon_k}(\mathbf{v}, \lambda^{(k)})$.

PASO 2: Si $\|\nabla_{\mathbf{v}} \mathcal{L}_{\varepsilon_k}(\mathbf{u}^{(k)}, \lambda^{(k)})\|_{\infty} < \delta$, entonces se para. En caso contrario, se toma $\varepsilon_{k+1} = 1/10\varepsilon_k$, $\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)}$, $\lambda^{(k+1)} = \lambda^{(k)} + \frac{1}{\varepsilon_k} \Phi(\mathbf{u}^{(k)})$ y $\mu^{(k+1)} = \max\left\{0, \mu^{(k)} + \frac{1}{\varepsilon_k} \Psi(\mathbf{u}^{(k)})\right\}$ para volver al PASO 1.

Observación 7

Debido a la construcción del método para el caso genérico, se pueden aplicar los mismos resultados que se mostraron con anterioridad. Esto es debido a que este algoritmo es una simple consecuencia de aplicar el Algoritmo 3 al problema (32).

Ejemplo 2

Para ejemplificar el método propuesto en el Algoritmo 4, se aplicará el código propuesto al problema

$$\begin{aligned} \min_{\mathbf{v} \in \mathbb{R}^2} J(\mathbf{v}) &= (v_1 - 2)^2 + (v_2 - 1)^2 \\ \text{Sujeto a:} & \\ &0.25v_1^2 + v_2^2 \leq 1 \\ &v_1 - 2v_2 = -1, \end{aligned} \tag{36}$$

del que se obtiene los resultado mostrados en la Tabla 2.

Iteración k	ε_k	$\mathbf{u}^{(\varepsilon_k)}$	$J(\mathbf{u}^{(\varepsilon_k)})$
Datos iniciales	1	(2 2)	1
1	1	(1.087946 9.664382e - 01)	0.83297
2	3.333333e - 01	(9.476319e - 01 9.368910e - 01)	1.1115
3	1.111111e - 01	(8.541536e - 01 9.181050e - 01)	1.3197
4	3.703704e - 02	(8.262058e - 01 9.121708e - 01)	1.3855
5	1.234568e - 02	(8.230051e - 01 9.114667e - 01)	1.3932
6	4.115226e - 03	(8.228774e - 01 9.114382e - 01)	1.3935
7	1.371742e - 03	(8.228757e - 01 9.114378e - 01)	1.3935
8	4.572474e - 04	(8.228757e - 01 9.114378e - 01)	1.3935
9	1.524158e - 04	(8.228757e - 01 9.114378e - 01)	1.3935

Tabla 2: Resultados obtenidos con el método del Lagrangiano aumentado para el problema (36).

Como se puede apreciar, según decrece el valor de ε_k en cada iteración, las soluciones obtenidas se aproximan a la solución de forma pronunciada. Aunque el valor del funcional en las soluciones no decrezca en cada iteración, esto es debido a que los puntos obtenidos no son siempre factibles y se van aproximando al conjunto factible según se suceden las iteraciones.

4. Ejecuciones y pruebas prácticas

Con el objetivo de probar el funcionamiento de los anteriores métodos en la práctica, se mostrará un par de ejecuciones en problemas reales. Para ello, se programaron los Algoritmos 1, 2 y 3 en MatLab, donde se utilizó un algoritmo Quasi-Newton en la resolución de los subproblemas de optimización sin restricciones (estos son los más empleados en este ámbito).

4.1. El problema de la geodésica

Un problema usual en las compañías aéreas es el cálculo de la menor trayectoria en el recorrido de los aviones. Intuitivamente dicha trayectoria es la línea recta. Sin embargo, al aproximarse la Tierra a una esfera, la trayectoria más corta es el arco de geodésica que une ambos puntos.

El objetivo de este primer problema es una versión de la situación planteada. Se considerará la esfera unidad y se querrá calcular la curva de longitud mínima que una dos puntos dados. La solución exacta es conocida y se puede consultar sus detalles en [6, Cap. V]. Para plantear el problema de forma numérica se desea hallar una discretización de dicha curva para aproximarla por la curva que une los puntos de la discretización solución. Así, se plantea el siguiente problema:

$$\begin{aligned} & \min_{\mathbf{A}(2:n-1,:) \in \mathbb{R}^{3(n-2)}} \sum_{i=1}^{n-1} \sqrt{(a_{i+1,1} - a_{i1})^2 + (a_{i+1,2} - a_{i2})^2 + (a_{i+1,3} - a_{i3})^2} \\ & \text{Sujeto a:} \\ & \quad a_{i1}^2 + a_{i2}^2 + a_{i3}^2 = 1; \quad \forall i = 2, \dots, n-1 \\ & \quad a_{ij} = \sigma_i \quad \forall j = 2, \dots, n-1 \text{ y algún } i \in \{1, 2, 3\} \end{aligned} \quad (37)$$

donde $\mathbf{A} \in \mathcal{M}_{n \times 3}$ es la matriz que en cada fila contiene las coordenadas de cada punto. Entonces, en la primera fila se tiene el punto de partida y, en la última, el punto de destino. Así, las variables son las coordenadas de los puntos incluidos en las filas interiores de la matriz. De esta forma, se tiene un problema $3(n-2)$ -dimensional con $n-2$ puntos que definen la discretización de la curva buscada.

Respecto a las restricciones, se tienen $n-2$ primeras restricciones de igualdad que aseguran que los puntos calculados pertenezcan a la esfera unidad. Finalmente, para asegurar que la discretización sea útil para visualizar la trayectoria, se fija una de las coordenadas de los puntos a un valor a elegir (σ_i), dejando los demás libres.

En este caso se tomará como punto de partida $(\sqrt{2}/2, 0, \sqrt{2}/2)$ y como punto de destino $(0, \sqrt{2}/2, \sqrt{2}/2)$ con 100 puntos de discretización. Con estas bases se han obtenido los siguientes resultados:

	Número de iteraciones	Distancia	Tiempo de ejecución
Penalización exterior	2	1.047107	34.347507 sec
Lagrangiano aumentado	10	1.047192	13.0308 min

Tabla 3: Ejecuciones del problema de la geodésica

La solución se representa en la siguiente Figura 3, donde se muestra en rojo la geodésica obtenida numéricamente y, en azul, el recorrido en línea recta (el intuitivo en un primer lugar).

La distancia del recorrido azul es de 1.1107 u.d, mientras que la distancia de la geodésica exacta es $\arccos(0.5) = 1.0472$ u.d. Por tanto, dentro de las soluciones obtenidas, sería la del Lagrangiano aumentado la más aproximada. Por otra parte, el método de penalización exterior converge en dos iteraciones, lo que impide la convergencia de $\{\varepsilon_k\}_{k \in \mathbb{N}} \rightarrow 0$ y, por consiguiente, de que sea fiable el método.

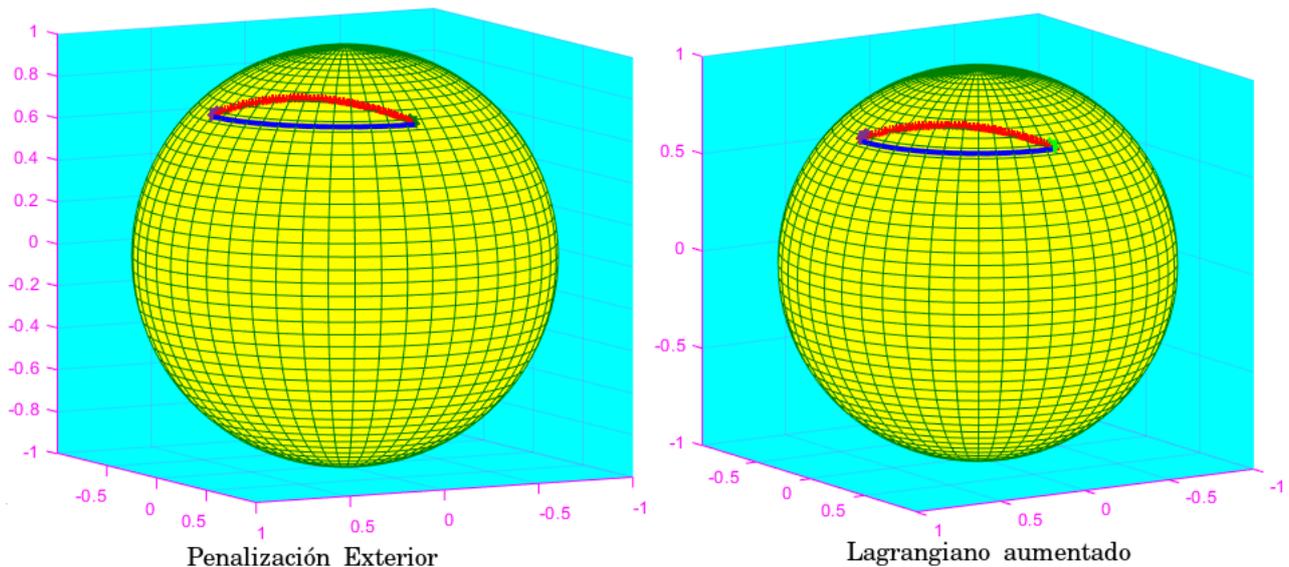


Figura 3: Soluciones al problema de la geodésica

Por ello, sería conveniente aumentar el criterio de tolerancia para ver cómo se suceden las soluciones según siguiese decreciendo ε_k en cada iteración. Aunque si se lleva a cabo este último consejo, entonces el tiempo de ejecución del método de penalización exterior se dispara en más de media hora, por lo que se descarta su utilidad en este caso. Finalmente, el método de penalización interior no se muestra en los recuadros de soluciones debido a que este método no se puede emplear en contextos donde aparezcan restricciones de igualdad.

4.2. El problema de la bancarrota

Otra de las aplicaciones de la optimización es en la Teoría de Juegos. Pues bien, en esta rama de las matemáticas es necesario resolver problemas de optimización con restricciones para obtener las soluciones requeridas. Un ejemplo de ello sería en la obtención de la solución de Nash en los juegos de negociación.

En [4] se pueden ver los detalles relativos a esta disciplina y uno de los problemas que se plantea es el problema de la bancarrota. En él se presenta una empresa en quiebra, cuyo capital es insuficiente para pagar las deudas contraídas. Además, la legislación del país obliga a que el capital de la empresa se reparta entre los acreedores, sin que ninguno de ellos reciba más de lo que se le adeuda y de un modo que sea aceptado por todos.

Entonces, se denota por n el número de acreedores, por $E > 0$ al capital de la empresa y por α a un vector de n componentes que contiene en la componente i -ésima los adeudamientos del i -ésimo acreedor. Así, en [4, págs. 105-112], se plantea como posible respuesta la solución de Nash, la cuál es la solución al problema:

$$\begin{aligned}
 & \underset{v \in \mathbb{R}^n}{\text{máx}} \quad \prod_{i=1}^n v_i \\
 & \text{Sujeto a:} \\
 & \quad \sum_{i=1}^n v_i \leq E \\
 & \quad 0 \leq v_i \leq \alpha_i, \quad \forall i = 1, \dots, n.
 \end{aligned} \tag{38}$$

Este se ha resuelto, en su versión equivalente a un problema de minimización, con los distintos algoritmos estudiados para el caso particular en el que $n = 10$, $E = 5$ millones de euros, y el vector α de

los adeudamientos es aquel que viene dado por

$$\alpha = (1 \quad 0.8 \quad 0.5 \quad 1.1 \quad 0.7 \quad 0.2 \quad 0.9 \quad 1.5 \quad 0.1 \quad 1.2)$$

en millones de euros. Así pues, con una tolerancia común de 10^{-5} , se han obtenido los resultados mostrados en la Tabla 4.

	Número iteraciones	J(\mathbf{u})	Tiempo de ejecución
Penalización exterior	Diverge	Diverge	Diverge
Lagrangiano aumentado	14	2.7991.e-04	0.7227 seg

Tabla 4: Ejecuciones del problema de la bancarrota

Entonces, a pesar de que se trata de un problema aparentemente sencillo (con pocas dimensiones y restricciones lineales), el método de penalización exterior diverge debido a que la penalización 13 no es capaz de compensar el decrecimiento del funcional objetivo (motivo por el cual el mínimo de los funcionales $J_\epsilon(\mathbf{v})$ no es finito). El método de penalización interior no aporta una solución. Esto es debido a que, en la práctica, al resolver en MatLab los subproblemas sin restricciones, estos aportan soluciones no finitas o que se encuentran fuera del conjunto admisible. Así, el método de penalización interior conlleva serios problemas en la implementación del algoritmo en la práctica. Por consiguiente, es el Lagrangiano aumentado el único método estudiado que realmente aporta una solución al problema en un tiempo razonable.

5. Conclusiones

En base al trabajo desarrollado en el presente artículo, se obtienen las siguientes conclusiones:

- Dentro de los métodos de penalización, el método de penalización exterior es el que mejor se adecúa. El método de penalización interior no está definido en todo el dominio y si el mínimo del problema se encuentra en la frontera del conjunto admisible, este nunca podrá llegar a alcanzarse. Por el contrario, el método de penalización exterior, además de permitir las restricciones de igualdad, consta de mejores características para minimizar los subproblemas sin restricciones, siempre y cuando estos tengan solución finita.
- En caso de convergencia del método de penalización exterior, este calcula numéricamente la solución del problema que resuelve, pero esta solución puede que no sea factible. Sin embargo, en el método de penalización interior siempre se van a obtener soluciones factibles, en caso de obtener alguna solución.
- Por lo visto en la sección 4, el método del Lagrangiano aumentado obtiene mucho mejores resultados en la práctica que los métodos de penalización exterior o interior. Este hecho es debido a que cuenta con mejores propiedades que los anteriores métodos debido, en parte, al Teorema 6.
- De obtener una solución con los métodos de penalización y del Lagrangiano aumentado, hay que tener un rango de escepticismo con los resultados obtenidos. Pues bien, los métodos de penalización, en caso de convergencia, lo hacen a una solución del problema estudiado. Sin embargo, el método del Lagrangiano aumentado converge hacia un punto que verifica las condiciones KKT. Por consiguiente, en este último caso sería necesario hacer un estudio local del punto obtenido para comprobar que efectivamente se trata de un punto óptimo.

Referencias

- [1] Bertsekas D.P. *Nonlinear Programming*. 2nd. Athena Scientific. 1999.
- [2] Boglárka G.-Tóth; Eligius M.T.Hendrix. *Introduction to Nonlinear and Global Optimization*. Springer. 2010.
- [3] Bruguera M; Viaño, J.M. *Lecciones de Métodos Numéricos. Volumen 4: Optimización*. Andavira. 2013.
- [4] Casas B., Fiestras Janeiro G., García Jurado I., González Díaz J., *Introducción a la Teoría de Juegos*, USC editora, (2012)
- [5] Clark, F.H. *Optimization and Non-Smooth Analysis*. Society for Industrial and Applied Mathematics. 1990.
- [6] Hernández Cifre M.A., Pastor González J.A., *Un curso de geometría diferencial. Teoría, problemas, soluciones y prácticas con ordenador*, Ediciones Doce Calles S.L (2010)
- [7] Hestenes, M.R, Multiplier and gradient methods. *Journal of Optimization Theory and Applications* 4, pp. 303-320, 1969.
- [8] Nocedal, J; Stephen J.Wright. *Numerical Optimization*. Springer. 2nd. 2006.
- [9] Powell, M.J.D., A method for nonlinear constraints in minimization problems. *Optimization*, R. Fletcher (ed.) Academic Press, New York, NY, pp. 283-298, 1969.
- [10] Rockafellar, R.T., Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control and Optimization* 12, pp. 268-285, 1974. Colletion of articles dedicated to the memory of Lucien W. Neustadt.
- [11] Rockafellar, R.T., A dual approach to solving nonlinear programming problems y unconstrained optimization. *Mathematical Programming* 5, pp. 354-373, 1973.
- [12] Rockafellar, R.T., The multiplier method of Hestenes and Powell applied to convex programming. *Journal of Optimization Theory and Applications*, 12, pp. 555-562, 1973.
- [13] W. Sun; Y-X Yuan. *Optimization Theory and Methods. Nonlinear Programming*. Springer. 2006.