UNIVERSIDAD NACIONAL DE LA PLATA

Facultad de Ciencias Exactas

Departamento de Matemática

Métodos de Lagrangiano Aumentado basados en funciones de penalidad no cuadráticas

Trabajo de Tesis Doctoral

MARIA DANIELA SANCHEZ

Directora: María Laura Schuverdt

Año 2017

Índice general

Agradecimientos

Resumen

El proceso del método de Lagrangiano Aumentado genera una sucesión de iteraciones $\{x_k\}$ donde x_k es la solución aproximada de un subproblema que involucra una función Lagrangiana Aumentada. El estudio de la convergencia global de la sucesión $\{x_k\}$ depende fuertemente de la información utilizada para resolver el subproblema.

Cunado se usa información de las primeras derivadas para resolver el subproblema, bajo condiciones de calidad apropiadas, se puede demostrar la convergencia a puntos que satisfacen las condiciones de Karush-Kuhn-Tucker (puntos KKT). Cuando se utiliza información tanto de las primeras como de las segundas derivadas, se demuestra la convergencia, bajo condiciones de calidad adecuadas, a puntos KKT que verifican además una condición de optimalidad de segundo orden. El estudio de condiciones de calidad y condiciones de calidad secuenciales han crecido de manera notoria en los últimos tiempos. La condición más débil que encontamos en la literatura, considerando el método de Lagrangiano Aumentado, involucra la función de penalidad cuadrática. El propósito de este trabajo es estudiar la convergencia global del algoritmo de Lagrangiano Aumentado que considera funciones de penalidad no cuadráticas. Analizamos la convergencia del algoritmo propuesto a puntos que satisfacen las condiciones KKT y, también, la condición de optimalidad necesaria de segundo orden débil. El esquema de generación de las funciones de penalidad Lagrangianas incluye, por ejemplo, la función de penalidad exponencial y la barrera logarítmica sin utilizar hipótesis de convexidad. Para la función de penalidad exponencial, la acotación del parámetro de penalidad es probada utilizando condiciones clásicas.

Introducción

Los problemas generales de optimización no lineal son de la forma

$$\begin{array}{lc} Minimizar & f(x) \\ sujeto\ a & h_i(x) = 0,\ i = 1, \dots, m, \\ & g_j(x) \leq 0,\ j = 1, \dots, p \end{array} \tag{0.0.1}$$

donde las funciones $f : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p$ son dos veces continuamente diferenciables en $\mathbb{R}^n$.

La función f se denomina función objetivo y el conjunto $\Omega = \{x \in \mathbb{R}^n : h_i(x) = 0, g_j(x) \leq 0, i = 1, \dots, m; j = 1, \dots, p\}$ se llama conjunto factible. Denotaremos $A(x^*) = \{j = 1, \dots, p : g_j(x^*) = 0\}$ al conjunto de restricciones de desigualdad activas en x^*.

Una solución global del problema (0.0.1) es un punto $x^* \in \Omega$ tal que $f(x^*) \leqslant f(x)$ para todo $x \in \Omega$. Pero, encontrar soluciones globales es, en muchos casos, muy difícil. La mayoría de los algoritmos populares se conforman con encontrar candidatos a minimizadores, puntos que satisfacen las condiciones de Karush-Kuhn-Tucker. Un $x^* \in \Omega$ es un punto Karush-Kuhn-Tucker (KKT) para el problema (0.0.1) si existen $\lambda^* \in \mathbb{R}^m, \mu^* \in \mathbb{R}^p$, $\mu_j^* \geqslant 0$ tales que

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla g_j(x^*) = 0$$

y

$$\mu_j^* = 0 \quad \text{si} \quad g_j(x^*) \leqslant 0.$$

Los escalares λ_i, μ_j son conocidos como los multiplicadores de Lagrange, [14, 17, 41, 44].

En este trabajo, resolveremos el problema original (0.0.1) usando una técnica popular en optimización con restricciones: el método Lagrangiano Aumentado. El enfoque de Lagrangiano Aumentado ha sido ampliamente estudiado y utilizado en las últimas décadas, (ver el libro recientemente publicado [48] y las referencias que hay en él).

Para resolver el problema (0.0.1), la idea básica de un método de Lagrangiano Aumentado consiste en una sucesión de iteraciones externas. En cada iteración externa de estos métodos se considera un problema para el cual existe en la literatura una variedad importante de algoritmos eficientes que pueden ser utilizados. En cada iteración, fijado un parámetro de penalidad $\rho > 0$ y estimados de los multiplicadores de Lagrange $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p, \mu_j \geqslant 0$, una función Lagrangiana Aumentada es minimizada de manera aproximada:

$$\textit{Minimizar } L(x, \lambda, \mu, \rho).$$

Se pueden penalizar todas las restricciones o bien solo algunas de ellas dejando las restantes en un conjunto factible, como se hizo en [2].

Una vez que la solución aproximada es encontrada, se definen nuevos parámetros y una nueva iteración comienza.

En la implementación práctica, la función Lagrangiana Aumentada más clásica y ampliamente utilizada para resolver el subproblema es la de Powell-Hestenes-Rockafellar (PHR) [33, 53, 55], basada en la función de penalidad cuadrática:

$$L_{PHR}(x, \lambda, \mu, \rho) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{i=1}^{m} \rho(h_i(x))^2 + \sum_{j=1}^{p} \frac{1}{2\rho}(\text{máx}\{0, \mu_j + \rho g_j(x)\}^2 - \mu_j^2).$$

En la literatura se pueden encontrar muchos algorítmos eficientes basados en la función PHR, ver por ejemplo los algorítmos definidos en [1, 2, 16, 17, 18, 19, 23, 24, 25] para resolver problemas regulares no lineales. Por ejemplo, [26] para resolver problemas libres de derivadas, [35, 37, 42, 43] para problemas matemáticos con restricciones de complementariedad y [20, 21] para problemas de optimización global.

Aunque, por su forma simple, la función de penalidad cuadrática mantiene su atractivo cuando se presentan restricciones de desigualdad, en el presente trabajo queremos enfatizar el uso de funciones de penalidad no cuadráticas para penalizar las restricciones de desigualdad.

De esta manera, dados $x \in \mathbb{R}^n, \rho \in \mathbb{R}, \rho > 0, \mu \in \mathbb{R}^p, \mu_j \geqslant 0$ en este trabajo consideraremos la función Lagrangiana Aumentada general de la forma

$$L(x, \lambda, \mu, \rho) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{i=1}^{m} \rho(h_i(x))^2 + \sum_{i=1}^{p} P(g_j(x), \rho, \mu_j), \qquad (0.0.2)$$

donde P es una función no cuadrática que penaliza las restricciones de desigualdad.

Creemos que en ciertas situaciones, dependiendo del problema que tiene que resolverse, las funciones de penalidad no cuadráticas pueden ser más adecuadas para fines de minimización. Este tipo de funciones se han considerado previamente, en el contexto de la penalidad externa y los métodos Lagrangiano Aumentado, en [12, 15, 16, 18, 27, 28, 60]. En el presente trabajo, para penalizar las restricciones de igualdad consideramos la función de penalidad cuadrática, mientras que para penalizar las restricciones de desigualdad consideramos funciones de penalidad Lagrangianas no cuadráticas definidas en el Capítulo 1.

En [18] los autores implementan 65 clases de métodos de Lagrangiano Aumentado para problemas de optimización no lineal con restricciones de desigualdad en el mismo marco con respecto a los criterios de parada, la precisión, la resolución del subproblema y otros parámetros algorítmicos. Analizando los resultados obtenidos, los autores concluyen la superioridad del uso de la función de penalidad cuadrática sobre las no cuadráticas. Sin embargo, consideramos que otros factores deben ser tenidos en cuenta. Por ejemplo, aunque la superioridad del uso de la función de penalidad cuadrática parece sugerir que la discontinuidad de las segundas derivadas de la función Lagrangiana Aumentada no es un inconveniente grave, esta conclusión depende en gran medida del algoritmo interno utilizado para resolver el subproblema. En nuestra opinión, esta es una razón suficientemente importante para estudiar el enfoque Lagrangiano Aumentado considerando otra función de penalidad. Funciones de penalidad no cuadráticas han sido ampliamente estudiadas durante las últimas décadas, véase por ejemplo el libro [16], y los trabajos [22, 27, 36, 60, 61]. Otra motivación importante para considerar funciones de penalidad no cuadráticas es el desarrollo un método Lagrangiano Aumentado con convergencia a puntos que verifican las condiciones de optimalidad de segundo orden. En este caso, es deseable tener una función Lagrangiana Aumentada dos veces diferenciable. Por lo tanto, tal como se menciona en [60], los métodos de tipo Newton se pueden utilizar con más eficacia para minimizar el subproblema de manera más eficiente.

En primera instancia, fuertemente basados en los comentarios anteriores consideramos la aplicación del método de Lagrangiano Aumentado utilizando la función de penalización exponencial. Esta función se introdujo en el contexto de la programación convexa, [39, 40].

El método Lagrangiano Aumentado definido en esta tesis sigue la idea introducida en [1, 2] con la diferencia en la función de penalidad utilizada. Esto significa que los subproblemas se resuelven en una forma aproximada, las estimaciones de los multiplicadores se proyectan en una caja fija con el fin de generar una sucesión acotada y la medida usada para aumentar los parámetros de penalidad se actualiza a partir de la utilizada con la función cuadrática adaptada a este caso.

En la práctica, hallar soluciones globales para un problema de optimización no lineal de la forma (0.0.1) es, en muchas oportunidades, muy difícil y en algunos casos imposible. El método de Lagrangiano Aumentado propone encontrar candidatos a minimizadores: puntos estacionarios (puntos que satisfacen las condiciones de Karush-Kuhn-Tucker). Pero no todos los minimizadores satisfacen las condiciones

KKT. Para esto es necesario que el punto satisfaga alguna condición de calidad de las restricciones. Desde el punto de vista práctico testear la optimalidad es muy difícil, en la implementación del algoritmo debemos decidir en cada iteración si se termina con la ejecución del algoritmo o no. Usualmente, esto está basado en la verificación de alguna condición necesaria de optimalidad. Teniendo en cuenta que el método de Lagrangiano Auemntado es iterativo estudiamos condiciones secuenciales de optimalidad bajo las cuales podemos asegurar la convergencia del algoritmo. Además, establecimos un nuevo criterio de parada del algoritmo bajo el cual se puede obtener la convergencia de segundo orden.

En este contexto, en el Capítulo 2 de [28], hemos sido capaces de obtener la convergencia global a los puntos estacionarios de primer orden utilizando la función de penalidad exponencial bajo una condición de calidad débil y sin hipótesis de convexidad. Hemos demostrado que un punto límite que satisface la condición de calidad del Generador Positivo Constante (CPG) [8] cumple con las condiciones de Karush-Kuhn-Tucker. Posteriormente, demostramos este mismo resultado utilizando la Propiedad del Cono Contínuo (CCP), definida recientemente en [9].

En [8] los autores muestran que CPG puede reemplazar otras condiciones de calidad de las restricciones más estrictas que aseguran la convergencia global de muchos algoritmos de optimización. Muestran resultados específicos para algoritmos de Lagrangianos Aumentados usando la función de penalidad cuadrática definida en [1, 2], programación cuadrática secuencial (SQP), métodos de puntos interiores y restauración inexacta. Por lo tanto, la convergencia global bajo la condición de calidad CPG de un algoritmo Lagrangiano Aumentado basado en la función exponencial es un resultado novedoso en el campo de optimización. Este resultado muestra que, teóricamente hablando, las funciones de penalidad cuadrática y exponencial tienen propiedades de convergencia similares.

Cuando se utiliza información de las primeras derivadas para resolver el subproblema, bajo condiciones de calidad apropiadas, se puede demostrar la convergencia a puntos estacionarios de primer orden. La condición de calidad que consideraremos en este trabajo para establecer la convergencia global de primer orden es la propiedad del cono continuo (CCP) [9]. En [9] los autores prueban que CCP es más débil que la condición Generador Positivo Constante (CPG) [8] y que está naturalmente asociada con la condición de Karush-Kuhn-Tucker Aproximada (AKKT) definida en el Capítulo 2. Esto hace a la condición de calidad CCP adecuada para el análisis de convergencia de algoritmos de optimizaión con resticciones prácticos que generan sucesiones AKKT; por ejemplo, el algoritmo de Lagrangiano Aumentado que utiliza la función de penalidad cuadrática [2, 48].

Está bien establecido en la literatura que cuando el parámetro de penalidad es muy grande, los subproblemas pueden ser muy difíciles de resolver de manera eficiente. Por esta razón, es importante establecer condiciones adecuadas para garantizar que la sucesión de parámetros de penalidad no necesita aumentar indefinidamente para encontrar una solución. Este análisis de convergencia local depende, en gran parte, de la medida definida para controlar el parámetro de penalidad.

En este trabajo la medida utilizada para aumentar el parámetro de penalidad proviene de la medida utilizada en el caso cuadrático [1, 2] adaptado en primera instancia al caso exponencial y luego al caso no cuadrático general. De esta manera, basándonos en las ideas propuestas en [2, 27], podremos obtener la acotación de la sucesión de parámetros de penalidad en el caso del algoritmo Lagrangiano Aumentado Exponencial utilizando la independencia lineal de los gradientes de las restricciones activas, el Hessiano de la función Lagrangiana definido positivo en el subespacio ortogonal a los gradientes de las restricciones de desigualdad activas y la condición de complementariedad estricta.

Cuando la sucesión generada por el algoritmo converge a un punto KKT, que además satisface una condición de optimalidad de Segundo Orden, el subproblema puede ser resuelto con información tanto de primer como de segundo orden. Recientemente, una condición de optimalidad de Segundo Orden para problemas de optimización fue definida en [6]. Observamos que la definición de AKKT2 está directamente relacionada con el uso del método de penalidad externa que considera la función de penalidad cuadrática para penalizar la restricciones de desigualdad. Con el fin de ser más generales y permitir el uso de las funciones de la penalidad no cuadráticas hemos reescrito la definición de AKKT2 de una manera diferente. La condición de calidad apropiada para probar la convergencia de segundo orden es la condición del Cono Contínuo de Segundo Orden (CCP2) definida en [6]. En [6] los autores prueban que CCP2 es más débil que Mangasarian-Formovitz + Rango Constante Débil definido en [4] y la condición de Calidad Rango Constante Relajado (RCRCQ) definida en [50].

Este trabajo esta organizado de la siguiente manera. En el Capítulo 1 describiremos las funciones Lagrangianas Aumentadas no cuadráticas consideradas en este trabajo e introduciremos un algoritmo de Lagrangiano Aumentado. En el Capítulo 2 presentamos condiciones de optimalidad secuenciales de primer y segundo orden para problemas generales y las condiciones de calidad adoptadas para demostrar convergencia global. Basándonos en [6], presentaremos una reformulación de la condición de optimalidad secuencial de segundo orden AKKT2 [6], lo cual es más apropiado cuando se consideran funciones de penalidad no cuadráticas. En el Capítulo 3 presentaremos los resultados de la convergencia global para el caso particular de la función de penalidad exponencial, mostraremos que, para esta función de penalidad, la sucesión de parámetros de penalidad no crece indefinidamente y daremos tres ejemplos ilustrativos. Finalmente, en el Capítulo 4 estableceremos la convergencia global a puntos estacionarios de primer orden y probaremos la convergencia global a puntos estacionarios de segundo orden del método de Lagrangiano Aumentado utilizando funciones de penalidad no cuadráticas. . Por último, en el Capítulo 5 daremos las conclusiones.

Notación.

Denotamos

$\mathbb{R}_+ = \{t \in \mathbb{R} \,| t \geqslant 0\}$,

$\mathbb{R}_{++} = \{t \in \mathbb{R} \,| t > 0\}$,

$\mathbb{N} = \{0, 1, 2, \ldots\}$,

$\| \cdot \|$ la norma arbitraria de un vector.

v_i es la $i-$ésima componente del vector v.

Si $F : \mathbb{R}^n \to \mathbb{R}^m$, $F = (f_1, \ldots, f_m)$, denotamos $\nabla F(x) = (\nabla f_1(x), \ldots, \nabla f_m(x)) \in \mathbb{R}^{n\times m}$.

Si $K = \{k_0, k_1, k_2, \ldots\} \subset \mathbb{N}$ $(k_{j+1} > k_j \forall j)$, denotamos $\lim\limits_{k\in K} x^k = \lim\limits_{j\to\infty} x^{k_j}$.

Si J es un subconjunto de $\{1, \ldots, n\}$, B_J la matriz formada tomando las columnas de B para los índices de J. Análogamente, y_J es el vector formado tomando las coordenadas de y para los índices de J.

Para todo $y \in \mathbb{R}^n$, $y_+ = (\max\{0, y_1\}, \ldots, \max\{0, y_n\})$.

Si $K = \{k_0, k_1, k_2, \ldots\} \subset \mathbb{N}$ $(k_{j+1} > k_j \forall j)$, denotamos $\lim\limits_{k\in K} x^k = \lim\limits_{j\to\infty} x^{k_j}$.

$A \succeq B$ si $d^T A d \geqslant d^T B d$.

$Sym(n)$ al espacio de las matrices reales y simétricas de dimensión n.

Capítulo 1

Los métodos de Penalidad Externa y de Lagrangiano Aumentado

La idea básica de los métodos de penalidad es eliminar algunas o todas las restricciones y agregar a la función objetivo un término de penalidad que tiene un costo alto para los puntos que no son factibles. Asociados con estos métodos hay un parámetro de penalidad que determina la severidad de la penalidad y como consecuencia, el alcance para el cual el problema sin restricciones aproxima al problema original. Como el parámetro de penalidad toma valores cada vez mayores, la aproximación resulta cada vez más precisa.

Consideramos el problema general de programación no lineal con restricciones de igualdad y desigualdad de la forma

$$\begin{array}{cc} Minimizar & f(x) \\ \textit{sujeto a} & h_i(x) = 0,\ i = 1, \dots, m, \\ & g_j(x) \leq 0,\ j = 1, \dots, p \end{array} \tag{1.0.1}$$

donde las funciones $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$ y $g : \mathbb{R}^n \to \mathbb{R}^p$ son funciones dos veces continuamente diferenciables. La clásica función de Lagrange, definida de la forma

$$l(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{p} \mu_j g_j(x),$$

juega un papel importante al momento de describir las condiciones de optimalidad y definir algoritmos que busquen soluciones para el problema (1.0.1).

Para problemas de programación no convexa, $l(x, \lambda^k, \mu^k)$ no suele ser convexa, ni siquiera para λ^k cercano a λ^*, μ^k cercano a μ^* y para x en un entorno de x^*, donde (x^*, λ^*, μ^*) es un punto Karush- Kuhn-Tucker para el problema (1.0.1). Para resolver este problema, en 1969 Hestenes [33] y Powell [53] introdujeron método de

Lagrangiano Aumentado para problemas con restricciones de igualdad. Mas tarde, entre 1973 y 1974, Rockafellar [55, 56, 58] desarrolló el método para problemas con restricciones de desigualdad.

Para problemas de programación convexa, Polyak y Teboulle, en un artículo publicado en 1997, [52], propusieron una clase de funciones de Lagrange para problemas considerando solo restricciones de desigualdad de la forma:

$$F_k(x,\mu) = f(x) + \frac{1}{k}\sum_{j=1}^{p} \mu_j \varphi(kg_j(x)) \tag{1.0.2}$$

donde $k > 0$ es el parámetro de penalidad y φ es una función dos veces continuamente diferenciable.

Auslender, Cominetti y Haddou en 1997, [13], y Ben-Tal y Zibulevsky en 1997, [15], estudiaron otras funciones Lagrangianas no lineales y obtuvieron resultados interesantes de convergencia para problemas convexos. Para problemas no convexos, una clase de funciones Lagrangianas no lineales para problemas con restricciones de desigualdad fue introducido por Mangasarian en 1975 [46]; en 1982, Bertsekas en su libro [16] propone la función de penalidad exponencial de la forma:

$$\varphi(kg_j(x)) = e^{kg_j(x)} - 1. \tag{1.0.3}$$

Polyak (1992) dió dos funciones de barrera modificadas, la función de Frisch modificada:

$$\varphi(kg_j(x)) = \begin{cases} ln(-kg_j(x)+1) & \text{si} \quad x \in Int\Omega_k \\ +\infty & \text{si} \quad x \notin Int\Omega_k, \end{cases}$$

y la función de Carroll modificada

$$\varphi(kg_j(x)) = \begin{cases} 1-(-kg_j(x)+1^{-1}) & \text{si} \quad x \in Int\Omega_k \\ +\infty & \text{si} \quad x \notin Int\Omega_k, \end{cases}$$

donde $k > 0$ es el parámetro de penalidad y $\Omega_k = \{x | 1 + kg_j(x) \geqslant 0, j = 1, \ldots, p\}$. Polyak [51] (2001) definió la función Log-Sigmoid

$$\varphi(x,\mu) = -\left[2ln(2) - 2ln(1+e^{kg_j(x)})\right].$$

Además, bajo condiciones necesarias de optimalidad de segundo orden, Jean-Pierre Dussault [27] (2004) analizó el comportamiento asintótico de Lagrangianos no lineales de forma (1.0.2) y obtuvo resultados interesantes de convergencia.

1.1. El método de penalidad externa

Sea $\mathcal{C}$ un conjunto cerrado y $f : \mathcal{C} \to \mathbb{R}^n$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^p$ funciones continuas. Consideramos el siguiente problema de programación no lineal:

$$\begin{array}{rl} Minimizar & f(x) \\ sujeto \;\; a & h(x) = 0, \;\; g(x) \leq 0, \;\; x \in \mathcal{C}. \end{array} \tag{1.1.1}$$

Sea $\{\rho_k\}_{k\in\mathbb{N}}$, $\rho_k > 0$, una sucesión de parámetros de penalidad tales que

$$\lim_{k\to\infty} \rho_k = +\infty.$$

Para cada $k \in \mathbb{N}$ definimos

$$F_k(x) = f(x) + \frac{\rho_k}{2}\|h(x)\|^2 + \sum_{j=1}^{p} \frac{1}{\rho_k}\varphi(\rho_k g_j(x)) \tag{1.1.2}$$

siendo φ una función continua que penaliza las restricciones de desigualdad. De esta manera, la función que penaliza las restricciones de igualdad es la cuadrática y otras funciones pueden ser consideradas para penalizar las restricciones de desigualdad.

El método de penalidad externa asociado con la sucesión $\{\rho_k\}$ se define de la siguiente manera:

Para cada $k \in K$, se calcula x^k como la solución del subproblema

$$\begin{array}{rl} Minimizar & F_k(x) \\ sujeto \;\; a & x \in \mathcal{C}. \end{array} \tag{1.1.3}$$

Observar que en la descripción del método no exite un criterio de parada. Esto significa que, teóricamente, la sucesión $\{x^k\}_{k\in\mathbb{N}}$ tiene infinitos términos, donde no se excluye la posibilidad de que muchos de ellos sean iguales. Por otro lado, las implementaciones prácticas deben estar asociadas a criterios de parada como veremos más adelante.

En las siguientes secciones incluimos los resultados que se pueden demostrar sobre la convergencia del método de penalidad externa utilizando la función de penalidad cuadrática y exponencial.

1.1.1. Convergencia utilizando la función de penalidad cuadrática

En esta subsección analizamos la convergencia del método de penalidad externa utilizando la función

$$\varphi_1(\rho_k g_j(x)) = \frac{1}{2}(\text{máx}\{0, \rho_k g_j(x)\})^2.$$

En este caso se tiene

$$F_k(x) = f(x) + \frac{\rho_k}{2}\|h(x)\|^2 + \sum_{j=1}^{p} \frac{1}{2\rho_k}(\text{máx}\{0, \rho_k g_j(x)\})^2.$$

Teorema 1.1.4. *[16] Supongamos que el problema (1.1.1) tiene por lo menos un punto factible. Supongamos también que la sucesión $\{x^k\}$ generada por el método de penalidad externa utilizando la función de penalidad cuadrática está bien definida para todo $k \in \mathbb{N}$ y que admite un punto límite x^*. Entonces, x^* es un minimizador global del problema (1.1.1).*

Demostración. Sea $K \subset \mathbb{N}$ un subconjunto infinito tal que

$$\lim_{k \in K} x^k = x^*.$$

Mostraremos primero que el punto límite x^* es factible, es decir, que $x^* \in \mathcal{C}$, $h(x^*) = 0$ y $g(x^*) \leq 0$.

Como $x^k \to x^*$, $x^k \in \mathcal{C}$ para todo k y $\mathcal{C}$ es cerrado tenemos que $x^* \in \mathcal{C}$.

Recordemos que denotamos $g(x)_+ = \text{máx}\{0, g(x)\}$.

Supongamos que

$$\|h(x^*)\|^2 + \|g(x^*)_+\|^2 > 0.$$

Luego,

$$\lim_{k \in K} F_k(x^k) = \lim_{k \in K} f(x^k) + \frac{\rho_k}{2}\left[\|h(x^k)\|^2 + \|g(x^k)_+\|^2\right] = +\infty$$

Sea z un punto factible de (1.1.1). Luego,

$$f(z) + \frac{\rho_k}{2}\left[\|h(z)\|^2 + \|g(z)_+\|^2\right] = f(z) \quad \forall k \in \mathbb{N},$$

y, para $k \in K$ suficientemente grande, tenemos que

$$f(x^k) + \frac{\rho_k}{2}\left[\|h(x^k)\|^2 + \|g(x^k)_+\|^2\right] > f(z) + \frac{\rho_k}{2}\left[\|h(z)\|^2 + \|g(z)_+\|^2\right].$$

Esto es una contradicción, ya que el punto x^k fue definido como un minimizador global del subproblema (1.1.3).

Por lo tanto, x^* es un punto factible.

Probaremos ahora que x^* es un minimizador global de (1.1.1).

Sea y un punto factible arbitrario de (1.1.1). Por la definición del método,

$$f(x^k) + \frac{\rho_k}{2}\left[\|h(x^k)\|^2 + \|g(x^k)_+\|^2\right] \leq f(y) + \frac{\rho_k}{2}\left[\|h(y)\|^2 + \|g(y)_+\|^2\right] = f(y)$$

para todo $k \in K$.

Luego,

$$f(x^k) \leq f(y) \quad \forall k \in K.$$

Tomando límite en esta última desigualdad, obtenemos que $f(x^*) \leq f(y)$. Como y es un punto factible arbitrario, tenemos que x^* es un minimizador global de (1.1.1) como queríamos probar. □

1.1.2. Convergencia utilizando la función de penalidad exponencial

Una desventaja del método de penalidad cuadrática es que, cuando se consideran restricciones de desigualdad, la función a minimizar en cada iteración puede no ser dos veces continuamente diferenciable, a pesar de que tanto la función objetivo como las restricciones lo sean. Esto puede generar problemas al momento de resolver el subproblema. Esto nos motiva a estudiar otras funciones de penalidad, como por ejemplo la función de penalidad exponencial que consideramos a continuación:

$$\varphi_2(\rho_k g_i(x)) = e^{\rho_k g_i(x)} - 1.$$

Por la definición del método,

$$F_k(x) = f(x) + \frac{\rho_k}{2}\|h(x)\|^2 + \sum_{j=1}^{p} \frac{1}{\rho_k}(e^{\rho_k g_j(x)} - 1)$$

Teorema 1.1.5. *Supongamos que el problema (1.1.1) tiene por lo menos un punto factible. Supongamos también que la sucesión $\{x^k\}$ generada por el método de penalidad externa para la función de penalidad exponencial está bien definida para todo $k \in \mathbb{N}$ y que admite un punto límite x^*. Entonces, x^* es un minimizador global del problema (1.1.1).*

Demostración. Sea $K \subset \mathbb{N}$ un subconjunto infinito tal que $\lim_{k\in K} x^k = x^*$.

Mostraremos primero que x^* es factible, es decir, $x^* \in \mathcal{C}$, $h(x^*) = 0$ y $g(x^*) \leq 0$.

Como $x^k \to x^*$, $x^k \in \mathcal{C}$ para todo k y $\mathcal{C}$ es cerrado, tenemos que $x^* \in \mathcal{C}$.

Supongamos que x^* no es un punto factible. Entonces, existe $i \in \{1, ..., m\}$ tal que $h_i(x^*) \neq 0$ o existe $j \in \{1, ..., p\}$ tal que $g_j(x^*) > 0$. Luego, para k suficientemente grande

$$\liminf_{k\to\infty} \frac{\rho_k}{2}\|h(x^k)\|^2 + \sum_{j:g_j(x^*)>0} \frac{e^{\rho_k g_j(x^k)} - 1}{\rho_k} + \sum_{j:g_j(x^*)\leq 0} \frac{e^{\rho_k g_j(x^k)} - 1}{\rho_k} \geq 0 \qquad (1.1.6)$$

pues, si $h_i(x^*) \neq 0$,

$$\liminf_{k\to\infty} \frac{\rho_k}{2}\|h(x^k)\|^2 = +\infty;$$

sino

$$\liminf_{k\to\infty} \sum_{j:g_j(x^*)>0} \frac{e^{\rho_k g_j(x^k)} - 1}{\rho_k} = +\infty$$

y

$$\liminf_{k\to\infty} \sum_{j:g_j(x^*)\leq 0} \frac{e^{\rho_k g_j(x^k)} - 1}{\rho_k} \geq \liminf_{k\to\infty} \sum_{j:g_j(x^*)>0} -\frac{log(2)}{\rho_k} = 0.$$

Esto implica que,

$$\liminf_{k\to\infty} F_k(x^k) = \liminf_{k\to\infty} f(x^k) + \frac{\rho_k}{2}\|h(x^k)\|^2 + \sum_{j=1}^{p} \frac{e^{\rho_k g_j(x^k)} - 1}{\rho_k} = +\infty. \qquad (1.1.7)$$

Sea z un punto factible de (1.1.1). Luego, como $g_j(z) \leq 0$ $\forall j = 1, ..., p$,

$$f(z) + \frac{\rho_k}{2}\|h(z)\|^2 + \sum_{j=1}^{p} \frac{e^{\rho_k g_j(z)} - 1}{\rho_k} \leq f(z) \quad \forall k \in \mathbb{N}$$

y, para k suficientemente grande, por (1.1.7) tenemos que

$$f(x^k) + \frac{\rho_k}{2}\|h(x^k)\|^2 + \sum_{j=1}^{p} \frac{e^{\rho_k g_j(x^k)} - 1}{\rho_k} > f(z) \geq f(z) + \frac{\rho_k}{2}\|h(z)\|^2 + \sum_{j=1}^{p} \frac{e^{\rho_k g_j(z)} - 1}{\rho_k}.$$

Esto no puede suceder, pues x^k fue definido como un minimizador global del subproblema. Por lo tanto, x^* es un punto factible.

Probaremos ahora que x^* es un minimizador global de (1.1.1).

Sea y un punto factible arbitrario de (1.1.1). Por la definición del método,

$$f(x^k) + \frac{\rho_k}{2}\|h(x^k)\|^2 + \sum_{j=1}^{p} \frac{e^{\rho_k g_j(x^k)} - 1}{\rho_k} \leq f(y) + \frac{\rho_k}{2}\|h(y)\|^2 + \sum_{j=1}^{p} \frac{e^{\rho_k g_j(y)} - 1}{\rho_k} \leq f(y)$$

para todo $k \in K$.

Luego,

$$f(x^k) + \frac{\rho_k}{2}\|h(x^k)\| + \sum_{j=1}^{p} \frac{e^{\rho_k g_j(x^k)} - 1}{\rho_k} \leq f(y).$$

Tomando límite en esta última desigualdad, y por (1.1.6) tenemos que $f(x^*) \leq f(y)$.
Como y es un punto factible arbitrario, tenemos que x^* es un minimizador global de (1.1.1) como queriamos probar. □

1.2. El método de Lagrangiano Aumentado

Consideremos el problema:

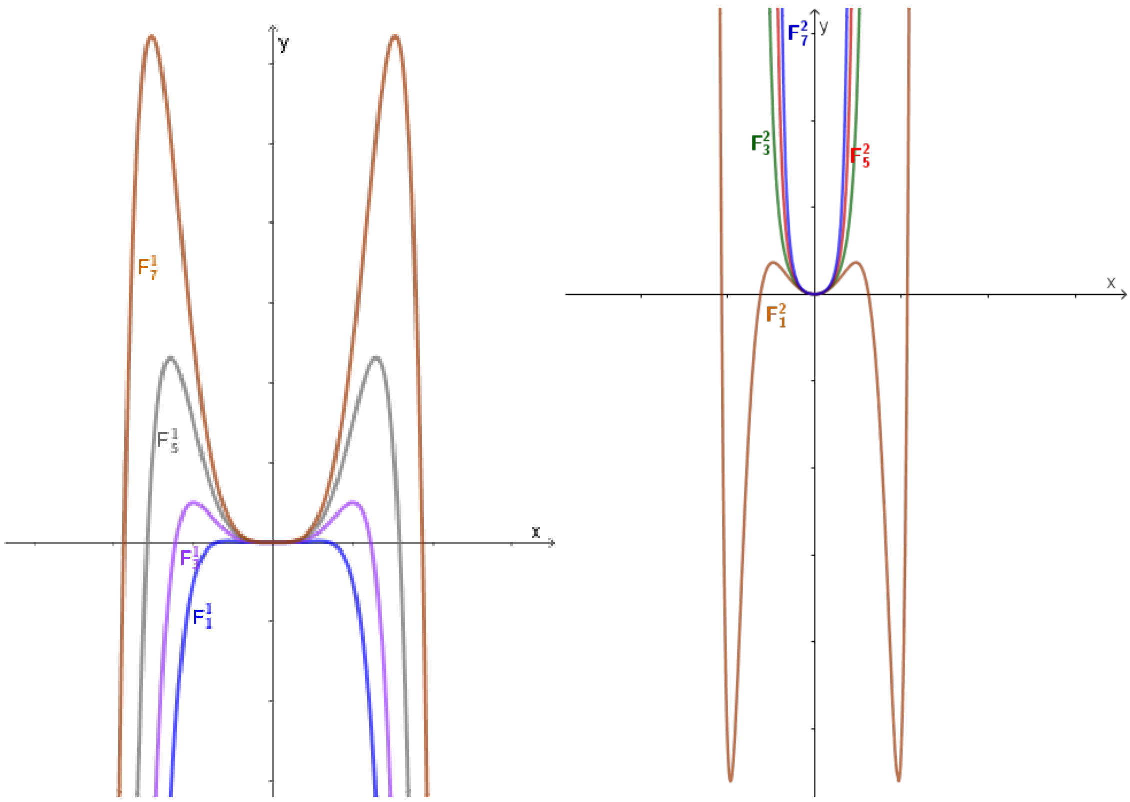
$$\begin{array}{ll} Minimizar & -x^6 \\ sujeto\ \ a & x^2 \leq 0. \end{array}$$

Claramente, $x^* = 0$ es un minimizador global de este problema.

Si consideramos la función de penalidad cuadrática φ_1 tenemos que la función a minimizar en cada subproblema es

$$F_k^1(x) = -x^6 + \frac{1}{2\rho_k}(\text{máx}\{0, \rho_k x^2\})^2.$$

Notemos que, como se observa en el gráfico izquierdo de la Figura 1.1, $F_k^1(x)$ no tiene un mínimo global ya que el $\text{ínf}_x F_k^1(x) = -\infty$. Por lo tanto, los subproblemas



(a) Penalidad Externa Cuadrática (b) Penalidad Externa Exponencial

Figura 1.1: Gráfica de las funciones generadas por el Método de Penalidad Externa.

generados por el método de penalidad externa utilizando la función de penalidad cuadrática no tienen solución.

Si consideramos la función de penalidad exponencial φ_2 tenemos que

$$F_k^2(x) = -x^6 + \frac{1}{\rho_k}\left(e^{\rho_k x^2} - 1\right).$$

Como se puede observar en el gráfico izquierdo de la Figura 1.1,

$$\lim_{x\to\infty} F_k^2(x) = x^6\left(-1 + \frac{e^{\rho_k x^2}}{\rho_k x^6} - \frac{1}{x^6}\right) = \lim_{x\to-\infty} x^6\left(-1 + \frac{e^{\rho_k x^2}}{\rho_k x^6} - \frac{1}{x^6}\right) = +\infty.$$

Esto nos dice que la función está acotada inferiormente y el subproblema tiene solución.

Este ejemplo simple muestra la importancia de considerar el estudio de los métodos de Lagrangiano Aumentado utilizando otras funciones de penalidad no cuadráticas.

La idea básica de un método de Lagrangiano Aumentado consiste en una sucesión de iteraciones externas. En cada iteración externa de estos métodos se resuelve un subproblema de optimización sin restricciones o uno para el cual existe en la literatura una variedad importante de algoritmos eficientes que pueden ser usados. El algoritmo de Lagrangiano Aumentado más famoso está basado en la función de penalidad cuadrática conocida como función de Powell-Hestenes-Rockafellar (PHR) [34, 53, 57]:

$$L_{PHR}(x,\lambda,\mu,\rho) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{i=1}^{m} \rho(h_i(x))^2 + \sum_{j=1}^{p} \frac{1}{2\rho}\left(\text{máx}\{0, \mu_j + \rho g_j(x)\}^2 - \mu_j^2\right)$$

donde los escalares λ_i, μ_j son estimados de los multiplicadores de Lagrange y ρ es el parámetro de penalidad. En la definición de L_{PHR} se pueden penalizar todas las restricciones o bien solo algunas de ellas dejando las restantes en un conjunto factible Ω, como se hizo en [2].

Como mencionamos anteriormente, una de las desventajas del método de Lagrangiano Aumentado utilizando esta función es que, cuando es aplicado a problemas con restricciones de desigualdad, la función Lagrangiana Aumentada P_{PHR} puede no ser dos veces continuamente diferenciable, a pesar de serlo la función objetivo y las restricciones. Como resultado, esto podría traer serias dificultades al momento de resolver el subproblema. Esto nos motiva a estudiar diferentes funciones Lagrangianas Aumentadas dos veces continuamente diferenciables para penalizar las restricciones de desigualdad. Por otro lado, la velocidad de convergencia puede ser más rápida o más lenta dependiendo de la función de penalidad elegida.

Consideraremos una función Lagrangiana aumentada de la forma

$$L(x,\lambda,\mu,\rho) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{i=1}^{m} \rho(h_i(x))^2 + \sum_{j=1}^{p} P(g_j(x), \mu_j, \rho)$$

donde $P : \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{R}_{++} \to \mathbb{R}$ es la función de penalidad para las restricciones de desigualdad de la forma:

- $P_1(y,\mu,\rho) = \frac{\mu}{\rho}\varphi(\rho y)$
- $P_2(y,\mu,\rho) = \frac{1}{\rho}\varphi(\rho\mu y)$
- $P_3(y,\mu,\rho) = \frac{\mu^2}{\rho}\varphi(\frac{\rho y}{\mu})$

donde $\varphi : \mathbb{R} \to \mathbb{R}$ es una función a valores reales que cumple las siguientes condiciones:

C1 φ es dos veces continuamente diferenciable.

C2 $\varphi'(t) \neq 1$ cuando $t < 0$.

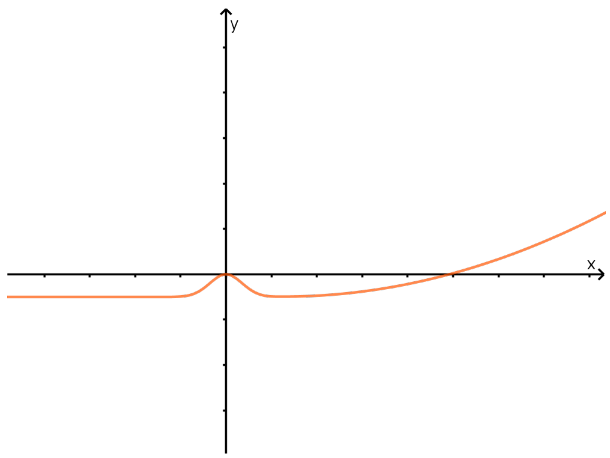
C3 $\lim\limits_{t\to-\infty} \varphi'(t) = 0.$

C4 $\lim\limits_{t\to-\infty} t\varphi''(t) = 0.$

Las funciones de penalidad Lagrangianas P_i, para $i = 1, 2, 3$, fueron previamente consideradas en [18]. Podemos observar que, obviamente, la función de penalidad cuadrática no cumple estas condiciones. Algunas funciones que si las cumplen son la función exponencial (1.0.3), la barrera logarítmica y las funciones $\theta_4, \ldots, \theta_9, \theta_{11}, \theta_{12}, \theta_{13}$, θ_{17}, θ_{18} consideradas en [18].

Algunos de estos axiomas han sido utilizados anteriormente en la literatura, por ejemplo en [12, 15, 16, 38]. En todos los casos, se pide que las funciones de penalidad sean estrictamente convexas, sin embargo, para nuestro objetivo no es necesaria la hipótesis de convexidad. Por ejemplo, consideremos la siguiente función partida, exponencial-cuadrática

$$\varphi(t) = \begin{cases} e^{-t^2} - 1 & \text{si } t \leqslant 2 \\ -1 + \frac{1}{e^4} - \frac{1}{e^4}(t-2) + \frac{1}{e^4}(t-2)^2 & \text{si } t > 2. \end{cases}$$



Esta función no es estrictamente convexa y sin embargo cumple con los axiomas antes mencionados.

Otros ejemplos que pueden ser considerados son:

- La función de penalidad exponencial (1.0.3). Esta función, por su rápido crecimiento, puede causar problemas cuando el algoritmo comienza desde un punto inicial no factible. Por este motivo, Martínez en [47] propone considerar la función partida

$$\varphi(t) = \begin{cases} e^t - 1 & \text{si } t \leq \beta \\ e^\beta - 1 + e^\beta(t-\beta) + e^\beta(t-\beta)^2/2 & \text{si } t > \beta \end{cases}$$

donde β es un número real estrictamente positivo.

- La función Logarítmica considerada en [18]:

$$\varphi(t) = log(1 + e^t).$$

- La función barrera-logarítmica modificada

$$\varphi(t) = \begin{cases} -log(-t) - 1 & \text{si } t \leqslant -\frac{1}{2} \\ e^{\frac{1}{2}}(\frac{1}{2}t^2 + \frac{1}{2}t + \frac{5}{8}) & \text{si } t > -\frac{1}{2}. \end{cases}$$

- La función logarítmica-exponencial

$$\varphi(t) = \begin{cases} -log(1-t) - 1 & \text{si } t \leqslant \frac{1}{2} \\ e^{2t-1} + log(2) - 1 & \text{si } t > \frac{1}{2}. \end{cases}$$

El algoritmo Lagrangiano Aumentado utilizando funciones de penalidad no cuadráticas es el siguiente:

Algoritmo 1. Sea $x^0 \in \mathbb{R}^n$ un punto inicial arbitrario. Los parámetros iniciales para la ejecución del algoritmo son los siguientes:

$$\tau \in [0,1), \gamma > 1,$$

$$-\infty < \bar{\lambda}_i^{\text{mín}} < \bar{\lambda}_i^{\text{máx}} < \infty, \quad \forall i = 1, ..., m,$$

$$0 < \bar{\mu}_j^{\text{mín}} < \bar{\mu}_j^{\text{máx}} < \infty, \quad \forall j = 1, \ldots, p,$$

$$\rho_0 \in \mathbb{R}_{++},$$

$$\bar{\lambda}_i^1 \in [\bar{\lambda}_i^{\text{mín}}, \bar{\lambda}_i^{\text{máx}}] \quad \forall i = 1, ..., m,$$

$$\bar{\mu}_j^0, \bar{\mu}_j^1 \in [\bar{\mu}_j^{\text{mín}}, \bar{\mu}_j^{\text{máx}}], \forall j = 1, \ldots, p,$$

$$\sigma_j^0 = \frac{\bar{\mu}_j^1 - \bar{\mu}_j^0}{\rho_0}.$$

Iniciación $k \leftarrow 1$.

Paso 1. Resolver el subproblema

Calcular (de ser posible) $x^k \in \Omega$ como una solución aproximada del subproblema

$$\textit{Minimizar } L(x, \lambda, \mu, \rho) \text{ sujeto a } x \in \bar{\Omega}. \tag{1.2.1}$$

Si no es posible, parar la ejecución del algoritmo.

Paso 2. Estimar los nuevos multiplicadores y definir una nueva medida de infactibilidad y complementariedad.

Para $i = 1, \dots, m$ calcular

$$\lambda_i^{k+1} = \bar{\lambda}_i^k + \rho_k h_i(x^k) \tag{1.2.2}$$

$$\bar{\lambda}_i^{k+1} = \text{máx}\{\text{mín}\{\lambda_i^{k+1}, \bar{\lambda}_i^{\text{máx}}\}, \bar{\lambda}_i^{\text{mín}}\} \tag{1.2.3}$$

Para $j = 1, \dots, p$ calcular

$$\mu_j^{k+1} = \frac{\partial P}{\partial y}(g_j(x^k), \bar{\mu}_j^k, \rho_k) \tag{1.2.4}$$

$$\bar{\mu}_j^{k+1} = max\{\text{mín}\{\mu_j^{k+1}, \bar{\mu}_j^{max}\}, \bar{\mu}_j^{\text{mín}}\} \tag{1.2.5}$$

$$\sigma_j^k = \frac{\mu_j^{k+1} - \bar{\mu}_j^k}{\rho_k} \tag{1.2.6}$$

Paso 3. Adaptar el parámetro de penalidad.

Si $k = 1$ o

$$\text{máx}\{\|h(x^k)\|_\infty, \|\sigma^k\|_\infty\} \leq \tau \, \text{máx}\{\|h(x^{k-1})\|_\infty, \|\sigma^{k-1}\|_\infty\} \tag{1.2.7}$$

elegir

$$\rho_{k+1} \geq \rho_k. \tag{1.2.8}$$

Sino, definir

$$\rho_{k+1} = \gamma \rho_k. \tag{1.2.9}$$

Paso 4. Comenzar una nueva iteración. Elegir $k \leftarrow k + 1$ e ir al paso 1.

Observación: El conjunto $\bar{\Omega}$ está formado por las restricciones fáciles de manipular, es decir, para las cuales existen algoritmos eficientes para resolver el subproblema.

Observemos que en el Paso 1 del Algoritmo 1 no se establece ninguna condición explícita para la solución aproximada x^k.

En el Capítulo 2 estableceremos condiciones secuenciales de optimalidad de primer y segundo orden apropiadas las cuales x^k debe satisfacer en el Algoritmo. Estas condiciones secuenciales establecerán los criterios de parada para probar la convergencia a puntos estacionarios de primer y segundo orden del problema (1.0.1).

Observaciones:

1. Las fórmulas (1.2.2) y (1.2.4) son las usuales para adaptar los estimados de los multiplicadores. Observar que provienen de que la igualdad

$$\nabla_x L(x^k, \lambda^k, \mu^k, \rho_k) = \nabla_x l(x^k, \lambda^{k+1}, \mu^{k+1}),$$

se debe satisfacer para todo k. Además, el multiplicador inicial satisface $\mu^0 \geqslant 0$ y por lo tanto $\mu^{k+1} \geqslant 0$ para todo k. Además, (1.2.4) está definida de manera multiplicativa, en contraste con la penalidad cuadrática que es aditiva:

$$\mu_j^{k+1} = \text{máx}\{0, \bar{\mu}_j^k + \rho_k g_j(x^k)\}, \tag{1.2.10}$$

ver [1, 2, 17, 19, 25].

2. Los estimados de los multiplicadores definidos en (1.2.3) y (1.2.5) fueron introducidos en [2] y son muy utilizados en el contexto de Lagrangiano Aumentado para garantizar que los estimados de los multiplicadores sean acotados. Observar que estas son las proyecciones de λ^{k+1} sobre $[\bar{\lambda}^{\text{mín}}, \bar{\lambda}^{max}]$ y μ^{k+1} sobre $[\bar{\mu}^{\text{mín}}, \bar{\mu}^{max}]$ respectivamente. Al proyectar los multiplicadores λ^{k+1}, μ^{k+1} se genera una sucesión $\{(\bar{\lambda}^{k+1}, \bar{\mu}^{k+1})\}$ de estimados de los multiplicadores de Lagrange que es acotada.

3. La fórmula (1.2.6) mide el progreso en términos de infactibilidad y complementariedad. La hemos adaptado a partir de la medida usada en el caso del método de Lagrangiano Aumentado cuadrático definido en [1, 2] para el caso no cuadrático [28].

Para el caso cuadrático la fórmula que se tiene es:

$$\sigma_j^k = \text{máx}\left\{g_j(x^k), -\frac{\bar{\mu}_j^k}{\rho_k}\right\}$$

Para obtenerlo, se reemplaza la restricción de desigualdad $g_j(x) \leq 0$ por la restricción de igualdad utilizando variables de holgura $g_j(x) + z_j^2 = 0$ y se calcula z_j optimizando la función Lagrangiano Aumentado cuadrática obteniendo

$$z_j = \sqrt{\text{máx}\left\{0, -\frac{\mu_j}{\rho} - g_j(x)\right\}}$$

Luego, se define $\sigma_j^k = g_j(x_k) + z_j^2 = \text{máx}\left\{g_j(x^k), -\frac{\bar{\mu}_j^k}{\rho_k}\right\}$.

Por otro lado, esta fórmula corresponde a la fórmula (1.2.6) usada en la definición (1.2.10). Por lo tanto, consideramos que (1.2.6) es la medida apropiada para el caso del método de Lagrangiano Aumentado no cuadrático.

En [18], donde los autores comparan 65 métodos de Lagrangiano Aumentado, se consideran otra actualización. El parámetro de penalidad no es aumentado cuando $\forall j = 1, \ldots, p$:

$$\text{máx}\{0, g_j(x^k)\} \leq \tau \,\text{máx}\{0, g_j(x^{k-1})\} \quad \text{y} \quad |g_j(x^k)\mu_j^k| \leq \tau |g_j(x^{k-1})\mu_j^{k-1}|.$$

Sin embargo, con esta medida, [18] no incluye un análisis de convergencia local.

4. Las condiciones (1.2.8) y (1.2.9) fueron consideradas previamente en [48] y [19]. La manera en que se decide si el parámetro de penalidad es adaptado o no, es general ya que considera la posibilidad de aumentar el parámetro de penalidad, incluso cuando el progreso es bueno, es decir, cuando la ecuación (1.2.7) se satisface.

Capítulo 2

Condiciones de optimalidad para problemas generales

Como mencionamos en la Introducción, desde el punto de vista práctico hallar soluciones globales es, en muchas oportunidades, muy difícil, y hasta a veces imposible. La mayoría de los algoritmos populares de optimización se conforman con encontrar candidatos a minimizadores: puntos estacionarios (puntos que satisfacen las condiciones KKT) del problema general de optimización. Si bien los puntos estacionarios son candidatos a minimizadores, no todos los minimizadores satisfacen las condiciones KKT. Por ejemplo, las condiciones KKT no se satisfacen en la solución del problema Minimizar x, sujeto a $x^2 = 0$, pero la condición KKT Aproximada si. Para que un minimizador cumpla las condiciones KKT es necesario que el punto satisfaga alguna condición de calidad de las restricciones. Luego, una condición de calidad (CQ) es una propiedad sobre los puntos factibles del problema, que cuando es verificada por un minimizador local, garantiza que las condiciones KKT se cumplen.

Por lo tanto, dada una condición de calidad de primer orden, decimos que un punto factible verifica una condición de optimalidad puntual de primer orden si este punto verifica la condición KKT o no verifica la condición de calidad.

Consideremos el problema

$$\begin{array}{cc} \textit{Minimizar} & f(x) \\ \textit{sujeto a} & h_i(x) = 0,\ i = 1, \ldots, m, \\ & g_j(x) \leqslant 0,\ j = 1, \ldots, p \end{array} \tag{2.0.1}$$

y denotamos $\Omega = \{x \in \mathbb{R}^n : h_i(x) = 0, g_j(x) \leqslant 0, i = 1, \ldots, m; j = 1, \ldots, p\}$ al conjunto factible.

Decimos que un punto factible x^* es un punto Karush-Kuhn-Tucker (KKT) para el problema (2.0.1) si existen multiplicadores $\lambda^* \in \mathbb{R}^m, \mu^* \in \mathbb{R}^p_+$, tales que

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla g_j(x^*) = 0 \tag{2.0.2}$$

y

$$\mu_j^* = 0 \quad \text{si} \quad g_j(x^*) \leqslant 0. \tag{2.0.3}$$

La condición (2.0.3) es la condición de complementariedad, y decimos que se satisface de manera estricta si

$$\mu_j^* = 0 \quad \text{cuando} \quad g_j(x^*) < 0.$$

Las condiciones de optimalidad secuenciales para optimización con restricciones son necesariamente satisfechas por los minimizadores locales, independientemente de las condiciones de calidad.

Condiciones de optimalidad secuenciales han sido previamente estudiadas en [5, 9, 30, 31, 49] para problemas de optimización no lineal.

La mayoría de los métodos numéricos son iterativos, en particular el método de Lagrangiano Aumentado que se estudia en esta tesis. Por lo tanto, en su implementación debemos decidir, en cada iteración, si se para la ejecución del algoritmo o no. Usualmente, esto está basado en la verificación de alguna condición, estas condiciones son las condiciones necesarias de optimalidad. A diferencia de las condiciones KKT, el objetivo de estas condiciones no es hacer una lista de candidatos a minimizadores locales, sino justificar los criterios de parada, los tipos de convergencia y la robustez de los algoritmos prácticos.

En este Capítulo estableceremos condiciones de optimalidad de primer y segundo orden bajo las cuales podemos probar la convergencia del algoritmo de Lagrangiano Aumentado, definido en el Capítulo 1, a puntos estacionarios de primer y segundo orden. Para ello utilizaremos el método de penalidad externa y algunos resultados asociados a él.

2.1. Condiciones necesarias de optimalidad de primer orden

A continuación definimos una de las condiciones de optimalidad más utilizada: la condición de Karush-Kuhn-Tucker Aproximada (AKKT) definida en [48].

Definición 2.1.1. [9, 48] Supongamos que $h(x^*) = 0$ y $g(x^*) \leqslant 0$. Decimos que x^* satisface la condición de Karush-Kuhn-Tucker Aproximada (AKKT) con respecto al problema (2.0.1) si existen sucesiones $\{x^k\} \subseteq \mathbb{R}^n, \{\lambda^k\} \subseteq \mathbb{R}^m$ y $\{\mu^k\} \subseteq \mathbb{R}^p_+$ tales que $\lim_{k\to\infty} x^k = x^*$,

$$\lim_{k\to\infty} \left\| \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) \right\| = 0, \tag{2.1.2}$$

y

$$\lim_{k\to\infty} \min\{\mu_j^k, -g_j(x^k)\} = 0 \text{ para todo } j = 1, \ldots, p. \tag{2.1.3}$$

Diremos que $\{x^k\}$ es la sucesión AKKT.

Normalmente, los teoremas que respaldan las condiciones de optimalidad son de la forma: "Si un minimizador local satisface una condición de calidad, entonces satisface la condición KKT". Sin embargo, la mayoría de los métodos numéricos populares no verifican que se satisfagan las condiciones de optimalidad, aunque las condiciones KKT (aproximadas) siempre se prueban.

Una propiedad importante de las condiciones de optimalidad secuenciales es que proveen herramientas teóricas adecuadas para justificar el criterio de parada para soluciones de problemas de optimización no lineales.

El siguiente Teorema prueba que la condición AKKT es una condición de optimalidad secuencial genuina en el siguiente sentido: si x^* es un minimizador local del problema (2.0.1) luego x^* satisface la condición AKKT.

Teorema 2.1.4. *[48] Supongamos que x^* es un minimizador local del problema (2.0.1). Entonces, existen sucesiones $\{x^k\} \subseteq \mathbb{R}^n, \{\lambda^k\} \subseteq \mathbb{R}^m$ y $\{\mu^k\} \subseteq \mathbb{R}^p_+$ tales que $\lim_{k\to\infty} x^k = x^*$,*

$$\lim_{k\to\infty} \left\| \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) \right\| = 0, \tag{2.1.5}$$

y

$$\lim_{k\to\infty} \min\{\mu_j^k, -g_j(x^k)\} = 0 \text{ para todo } j = 1, \ldots, p. \tag{2.1.6}$$

Demostración. Por hipótesis, existe $\varepsilon > 0$ tal que x^* es un minimizador global de

$$\textit{Minimizar } f(x) \text{ sujeto a } x \in \Omega \cap B(x^*, \varepsilon). \tag{2.1.7}$$

Entonces x^* es el único minimizador global de

$$\begin{aligned} &\textit{Minimizar } f(x) + \|x - x^*\|_2^2 \\ &\textit{sujeto a } x \in \Omega \cap B(x^*, \varepsilon). \end{aligned} \tag{2.1.8}$$

Para todo $k \in \mathbb{N}$, consideramos el problema

$$\begin{array}{l} \textit{Minimizar } f(x) + k\,[\|h(x)\|_2^2 + \|g_+(x)\|_2^2] + \|x - x^*\|_2^2 \\ \textit{sujeto a } \|x - x^*\| \leqslant \varepsilon. \end{array} \tag{2.1.9}$$

Como $B(x^*, \varepsilon)$ es cerrado y acotado, el Teorema de Bolzano-Weierstras nos asegura que el problema admite una solución $x^k \in B(x^*, \varepsilon)$. Como $\|x^k - x^*\| \leqslant \varepsilon$ para todo $k \in \mathbb{N}$ y $B(x^*, \varepsilon)$ es compacto, existen $K \subset \mathbb{N}$ y $z^* \in B(x^*, \varepsilon)$ tales que $\lim\limits_{k \to \infty} x^k = z^*$. Por la definición de x^k, tenemos que

$$\begin{aligned} & f(x^k) + k\left[\|h(x^k)\|_2^2 + \|g_+(x^k)\|_2^2\right] + \|x^k - x^*\|_2^2 \\ \leqslant\ & f(x^*) + k\left[\|h(x^*)\|_2^2 + \|g_+(x^*)\|_2^2\right] + \|x^* - x^*\|_2^2. \end{aligned}$$

Por lo tanto, como x^* es factible,

$$f(x^k) + k[\|h(x^k)\|_2^2 + \|g(x^k)_+\|_2^2] + \|x^k - x^*\|_2^2 \leqslant f(x^*) \tag{2.1.10}$$

para todo $k \in K$. Dividiendo por k a ambos lados de la desigualdad, usando la continuidad de las funciones y tomando límite para $k \in K$ obtenemos que

$$[\|h(z^*)\|_2^2 + \|g_+(z^*)\|_2^2 = 0.$$

Luego, $z^* \in \Omega$.

Por (2.1.10) tenemos que para todo $k \in K$

$$f(x^k) + \|x^k - x^*\|_2^2 \leqslant f(x^*).$$

Tomando límites en ambos lados de la desigualdad obtenemos que

$$f(z^*) + \|z^* - x^*\|_2^2 \leqslant f(x^*).$$

Pero, como x^* es el único minimizador global de (2.1.8) y $\|z^* - x^*\| < \varepsilon$, tenemos que $z^* = x^*$. Por lo tanto,

$$\lim_{k \to \infty} x^k = x^*. \tag{2.1.11}$$

Claramente, para $k \in K$ suficientemente grande, tenemos que $\|x^k - x^*\| < \varepsilon$. Luego, calculando el gradiente de la función objetivo del problema (2.1.9) se tiene que

$$\nabla f(x^k) + \sum_{i=1}^{m} \left[2k h_i(x^k)\right] \nabla h_i(x^k) + \sum_{g_j(x^k) > 0} \left[2k g_j(x^k)\right] \nabla g_j(x^k) + 2(x^k - x^*) = 0. \tag{2.1.12}$$

Definimos $\lambda_i^k = 2kh_i(x^k)$ y $\mu_j^k = 2kg_j(x^k)_+$. Luego de (2.1.11) y (2.1.12) se tiene que

$$\lim_{k\to\infty} \left\| \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) \right\| = 0,$$

donde $\mu_j^k = 0$ si $g_j(x^k) < 0$. Por lo tanto, $\min\{-g_j(x^k), \mu_j^k\} = 0$ para todo $j = 1, \ldots, p$ y $k \in K$. Esto completa la demostración. □

Este teorema muestra que AKKT es una condición necesaria de optimalidad independientemente de las condiciones de calidad. A pesar de esto, con el fin de verificar que AKKT implica KKT, se deben considerar condiciones de calidad. En [9] los autores definen la condición de calidad más débil para la cual esta implicación es válida:

Dado $x^* \in \mathbb{R}^n$ tal que $h(x^*) = 0, g(x^*) \leqslant 0$, consideramos el cono convexo cerrado definido como:

$$K(x) = \left\{ \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{j\in A(x^*)} \mu_j \nabla g_j(x) : \quad \lambda_i \in \mathbb{R}, \mu_j \in \mathbb{R}_+ \right\}.$$

Dada una multifunción $F : \mathbb{R}^s \rightrightarrows \mathbb{R}^d$, el límite exterior de $F(z)$ cuando $z \to z^*$ es denotado por

$$\limsup_{z\to z^*} F(z) = \{w^* \in \mathbb{R}^d : \exists\ (z^k, w^k) \to (z^*, w^*) \text{ con } w^k \in F(z^k)\}.$$

Luego, la multifunción F se dice que es semicontinua exterior en z^* si

$$\limsup_{z\to z^*} F(z) \subset F(z^*).$$

Definición 2.1.13. [9] Dado $x^* \in \mathbb{R}^n$ tal que $h(x^*) = 0, g(x^*) \leqslant 0$. Decimos que x^* satisface la Propiedad del Cono Contínuo (CCP) si la multifunción $x \rightrightarrows K(x)$ es semicontínua exterior en x^*, esto es

$$\limsup_{x\to x^*} K(x) \subset K(x^*).$$

La condición de calidad más popular es la independencia lineal de los gradientes de las restricciones activas (LICQ) ya que es fácilmente verificable. La condición de calidad de Mangasarian-Fromovitz (MFCQ)[45], que establece que los gradientes de las restricciones activas son positivamente linealmente independientes en el punto factible considerado, es obviamente más débil que LICQ [17]. Qi y Wei [54] introdujeron la condición de Dependencia Lineal Positiva Constante (CPLD) [11], que dice que si algunos gradientes de las restricciones activas son positivamente linealmente dependientes en un punto, entonces los mismos gradientes son linealmente dependientes en un entorno del punto. Andreani, Martínez, y Schuverdt en [11] demostraron que CPLD implica quasinormalidad. CPLD, además es más débil que MFCQ y necesariamente se satisface si las restricciones del problema son lineales (una propiedad que no es compartida por MFCQ). En [8] se definió la condición Generador Positivo Constante (CPG) más débil que CPLD. En [9] lo autores prueban que la condición CPG es estrictamente más fuerte que la propiedad CCP.

El siguiente teorema muestra que CCP es la condición más débil sobre las restricciones bajo la cual se puede probar que AKKT implica KKT.

Teorema 2.1.14. *[9] La condición CCP es la condición más débil bajo la cual se puede probar que AKKT implica KKT, independientemente de la función objetivo.*

Demostración. Probaremos primero que, si se satisface la condición CCP, la condición secuencial AKKT implica la condición KKT independientemente de la función objetivo. Sea f una función objetivo tal que la condición secuencial AKKT se satisface en x^*, luego existen sucesiones $\{x^k\} \subseteq \mathbb{R}^n, \{\lambda^k\} \subseteq \mathbb{R}^m$ y $\{\mu^k\} \subseteq \mathbb{R}^p_+$ tales que $\lim\limits_{k\to\infty} x^k = x^*$,

$$\mu_j^k = 0 \text{ para } j \notin A(x^*)$$

y

$$\zeta^k = \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k) \to 0. \tag{2.1.15}$$

Definimos $\omega^k = \sum\limits_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum\limits_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k)$, luego

$$\omega^k \in K(x^k) \text{ y } \omega^k = \zeta^k - \nabla f(x^k). \tag{2.1.16}$$

Tomando límites en (2.1.15) cuando k tiende a infinito y usando la continuidad del gradiente de f y que $\zeta^k \to 0$ se tiene que

$$-\nabla f(x^*) = \lim \omega^k \in \limsup_{k\to\infty} K(x^k) \subset \limsup_{x\to x^*} K(x) \subset K(x^*), \tag{2.1.17}$$

donde la última inclusión viene de la pripiedad del cono contínuo. Por lo tanto,

$$-\nabla f(x^*) \in K(x^*),$$

lo que es equivalente a decir que x^* satisface la condición KKT.

Ahora probaremos que que si AKKT implica KKT para cualquier función objetivo, entonces se satisface la condición CCP. Sea $\omega^* \in \limsup_{x \to x^*} K(x)$, por la definición de límite exterior, existen sucesiones $\{x^k\}$, $\{\omega^k\}$ tales que $x^k \to x^*$, $\omega^k \to \omega^*$ y $\omega^k \in K(x^k)$. Definimos $f(x) = -\langle \omega^*, x \rangle$ para todo $x \in \mathbb{R}^n$. Luego la condición AKKT se satisface en x^* para esta función con $\{x^k\}$ como sucesión AKKT ya que $\nabla f(x^k) + \omega^k = -\omega^* + \omega^k \to 0$. Entonces, por hipótesis, x^* satisface la condición KKT, esto es, $-\nabla f(x^*) = \omega^* \in K(x^*)$. □

2.2. Condiciones necesarias de optimalidad de segundo orden

Las condiciones necesarias de optimalidad pueden ser de primer orden o de segundo orden. Sabemos que cuando un minimizador local de un problema general no lineal verifica cualquier condición de calidad de primer orden, este punto verifica la condición KKT. Cuando, además, un minimizador local verifica una condición de calidad de segundo orden, es posible probar que el Hessiano de a función de Lagrange es semi-definido positivo (PSD) en el subespacio tangente [10]:

Decimos que x^* verifica la Condición Necesaria de Optimalidad Débil (WSONC) cuando x^* es factible para el problema (2.0.1) y existen multiplicadores de Lagrange $\lambda^* \in \mathbb{R}^m, \mu^* \in \mathbb{R}^p_+$, tales que

$$\mu_j^* = 0 \text{ para } j \notin A(x^*),$$

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla g_j(x^*) = 0$$

y

$$\nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla^2 g_j(x^*) \succeq 0 \text{ en } T(x^*)$$

donde $T(x^*) = \{d \in \mathbb{R}^n : \nabla h_i(x^*)^T d = 0, i = 1, \ldots, m; \nabla g_j(x^*)^T d = 0 \text{ para } j \in A(x^*)\}$

Queremos establecer condiciones bajo las cuales se verifique la convergencia a puntos que satisfagan la condición necesaria de optimalidad se segundo orden débil.

En [6] los autores introducen por primera vez una condición necesaria de optimalidad secuencial llamada AKKT2 (Definición 2.2.1). Además, demostraron que esta condición es una condición de optimalidad necesaria genuina.

Sea $l(x, \lambda, \mu)$ la función de Lagrange asociada al problema de optimización (2.0.1):

$$l(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{p} \mu_j g_j(x).$$

Definición 2.2.1. [6] Supongamos que $h(x^*) = 0$ y $g(x^*) \leqslant 0$. Decimos que el punto x^* satisface la condición estacionaria de segundo orden aproximada (AKKT2) con respecto al problema (2.0.1) si existen sucesiones $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m$, $\{\eta^k\} \subseteq \mathbb{R}^m$, $\{\mu^k\} \subseteq \mathbb{R}^p_+$, $\{\theta^k\} \subseteq \mathbb{R}^p_+$, $\{\delta^k\} \subseteq \mathbb{R}_+$ tales que $\lim\limits_{k\to\infty} x^k = x^*$, $\lim\limits_{k\to\infty} \delta^k = 0$ con

$$\mu_j^k = 0 \text{ para } j \notin A(x^*), \tag{2.2.2}$$

$$\theta_j^k = 0 \text{ para } j \notin A(x^*), \tag{2.2.3}$$

$$\lim_{k\to\infty} \nabla_x l(x^k, \lambda^k, \mu^k) = \lim_{k\to\infty} \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k) = 0 \tag{2.2.4}$$

y, para $k \in \mathbb{N}$ suficientemente grande,

$$\nabla^2_{xx} l(x^k, \lambda^k, \mu^k) + \sum_{i=1}^{m} \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j\in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succeq -\delta_k Id \tag{2.2.5}$$

en $\mathbb{R}^n$.

La expresión (2.2.5) dice que la matriz del lado izquierdo de la desigualdad debe ser semidefinida positiva en el límite.

El siguiente teorema, demostrado en [6], prueba que AKKT2 es una condición necesaria de optimalidad secuencial: si x^* es un minimizador local del problema (2.0.1) entonces x^* satisface la condición AKKT2.

Lema 2.2.6. *[7, 10] Sean $\bar{f} : \mathbb{R}^n \to \mathbb{R}$, $\bar{g}_j : \mathbb{R}^n \to \mathbb{R}$ para $j \in \{1, \ldots, p\}$ funciones con derivadas contínuas de segundo orden en un entorno del punto $\bar{x}$. Definimos*

$$\bar{F}(x) := \bar{f}(x) + \frac{1}{2} \sum_{j=1}^{p} \text{máx}\{0, \bar{g}_j(x)\}^2$$

para todo x en un entorno de $\bar{x}$. Supongamos que $\bar{x}$ es un minimizador local de $\bar{F}$. Entonces la matriz simétrica definida por

$$H(x) := \nabla^2 \bar{f}(x) + \sum_{j=1}^{p} \text{máx}\{0, \bar{g}_j(x)\} \nabla^2 \bar{g}_j(x) + \sum_{j: g_j(x) \geqslant 0} \nabla \bar{g}_j(x) \nabla \bar{g}_j(x)^T$$

es semidefinida positiva en $\bar{x}$.

Demostración. Definimos para todo $x \in \mathbb{R}^n$,

$$\Gamma(x) = \bar{f}(x) + \sum_{\bar{g}_i(x) \geqslant 0} \bar{g}_i(x)^2.$$

Claramente, $\Gamma(x^*) = \bar{F}(x^*)$ y $\nabla\Gamma(x^*) = \nabla\bar{F}(x^*)$. Mas aún, Γ tiene segundas derivadas contínuas y $\nabla\Gamma(x^*) = H(x^*)$.

Supongamos, por contradicción, que $H(x^*)$ no es semidefinida positiva. Luego, x^* no es un minimizador local de $\bar{F}$. Entonces, para todo $\varepsilon > 0$ existe $x \in \mathbb{R}^n$ tal que $\|x - x^*\| \leqslant \varepsilon$ y

$$\Gamma(x) < \Gamma(x^*) = \bar{F}(x^*).$$

Pero, por continuidad de $\bar{g}$, en un entorno de x^* se tiene que $\Gamma(x) \geqslant \bar{F}(x)$. Entonces, x^* no es un minimizador local de $\bar{F}$, contradiciendo la hipótesis del lema. Por lo tanto, $H(x^*)$ es semidefinida positiva como queríamos probar.

□

Teorema 2.2.7. *[6] Si x^* es un minimizador local de (2.0.1), entonces x^* satisface la condición AKKT2.*

Demostración. Como x^* es un minimizador local de (2.0.1) existe $\varepsilon > 0$ tal que

$$f(x^*) \leqslant f(x) \text{ para todo } x \in \Omega \cap B(x^*, \varepsilon).$$

Luego, x^* es la única solución del problema:

$$\textit{Minimizar } f(x) + \frac{1}{4}\|x - x^*\|^4 \textit{ sujeto a } x \in \Omega \cap B(x^*, \varepsilon). \tag{2.2.8}$$

Sea $\{\rho_k\}$ una sucesión de escalares positiva tal que $\rho_k \to +\infty$. Consideremos el método de penalidad externa utilizando la función cuadrática para el problema (2.2.8). Cada subproblema es de la forma

$$\begin{aligned} &\textit{Minimizar } f(x) + \frac{1}{4}\|x - x^*\|^4 + \frac{\rho_k}{2}\left[\sum_{i=1}^{m} h_i(x)^2 + \sum_{j=1}^{p} g_j(x)_+^2\right] \\ &\textit{sujeto a } x \in B(x^*, \varepsilon). \end{aligned} \tag{2.2.9}$$

Sea x^k la solución global del subproblema (2.2.9) que está bien definido ya que $B(x^*, \varepsilon)$ es compacto y las funciones son contínuas. Más aún, por el Teorema 1.1.4, la sucesión $\{x^k\}$ converge a x^* y $x^k \in IntB(x^*, \varepsilon)$ para k suficientemente grande. Luego, el gradiente de la función objetivo de (2.2.9) en x^k debe ser cero para k suficientemente grande:

$$\nabla f(x^k) + \rho_k \sum_{i=1}^{m} h_i(x^k)\nabla h_i(x^k) + \rho_k \sum_{j=1}^{p} g_j(x^k)_+ \nabla g_j(x^k) + \|x^k - x^*\|^2(x^* - x^k) = 0. \tag{2.2.10}$$

Por el Lema 2.2.6 con $\bar{F}(x) = f(x) + \frac{\rho_k}{2}\sum_{i=1}^{m} h_i(x)^2 + \frac{1}{4}\|x - x^*\|^4$, $\bar{g}_j(x) = \sqrt{\rho_k} g_j(x)$ para $j \in \{1, \ldots, p\}$ y $\bar{x} = x^k$, podemos decir que

$$\begin{aligned}
&\nabla^2 f(x^k) + \rho_k \sum_{i=1}^{m} h_i(x^k)\nabla^2 h_i(x^k) + \rho_k \sum_{j=1}^{p} g_j(x^k)_+ \nabla^2 g_j(x^k) \\
&+\rho_k \sum_{i=1}^{m} \nabla h_i(x^k)\nabla h_i(x^k)^T + \rho_k \sum_{j:g_j(x^k)\geqslant 0} g_j(x^k)_+ \nabla g_j(x^k)^T \\
&\qquad +2(x^k - x^*)(x^k - x^*)^T + \|x^k - x^*\|^2 \succeq 0.
\end{aligned} \tag{2.2.11}$$

Definimos

$$\begin{aligned}
\lambda_i^k &:= \rho_k h_i(x^k), \\
\eta_i^k &:= \rho_k \text{ para } i \in \{1, \ldots, m\}, \\
\mu_j^k &:= \rho_k \max\{0, g_j(x^k)\} \text{ para } j \in \{1, \ldots, p\}, \\
\theta_j^k &= \begin{cases} \rho_k \max\{0, g_j(x^k)\} & \text{si} \quad g_j(x^k) \geqslant 0 \\ 0 & \text{sino.} \end{cases}
\end{aligned}$$

Finalmente definimos $\delta_k := 3\|x^k - x^*\|^2$. Claramente $\delta_k \to 0$ cuando k tiende a infinito, $\mu_j^k = 0$ para $j \notin A(x^*)$ y $\theta_j^k = 0$ para $j \notin A(x^*)$. Ahora, eligiéndolos de esta manera, por (2.2.9) y (2.2.11), tenemos que la condición AKKT2 se satisface. □

Es interesante observar que el hecho de que en la Definición 2.2.1 las sucesiones $\{\theta^k\}$ y $\{\delta^k\}$ sean necesariamente no negativas proviene del uso del método de penalidad externa utilizando la función de penalidad cuadrática. Si consideramos la posibilidad de usar una función de penalidad no cuadrática en el método de penalidad externa, esto no es necesariamente así. La Definición 2.2.1 puede reescribirse de la siguiente manera:

Definición 2.2.12. Supongamos que $h(x^*) = 0$ y $g(x^*) \leqslant 0$. Decimos que el punto x^* satisface la condición estacionaria de segundo orden aproximada relajada (rAKKT2) con respecto al problema (2.0.1) si existen sucesiones $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m$, $\{\eta^k\} \subseteq \mathbb{R}^m$, $\{\mu^k\} \subseteq \mathbb{R}^p_+$, $\{\theta^k\} \subseteq \mathbb{R}^p$, $\{\delta^k\} \subseteq \mathbb{R}$ tales que $\lim_{k\to\infty} x^k = x^*$, $\lim_{k\to\infty} \delta^k = 0$ verificando

$$\mu_j^k = 0 \text{ para } j \notin A(x^*), \theta_j^k = 0 \text{ para } j \notin A(x^*), \tag{2.2.13}$$

$$\lim_{k\to\infty} \nabla_x l(x^k, \lambda^k, \mu^k) = \lim_{k\to\infty} \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k) = 0 \tag{2.2.14}$$

y, para $k \in \mathbb{N}$ suficientemente grande,

$$\nabla^2_{xx} l(x^k, \lambda^k, \mu^k) + \sum_{i=1}^{m} \eta_i^k \nabla h_i(x^k)\nabla h_i(x^k)^T + \sum_{j\in A(x^*)} \theta_j^k \nabla g_j(x^k)\nabla g_j(x^k)^T \succeq -\delta_k Id \tag{2.2.15}$$

en $\mathbb{R}^n$.

Teorema 2.2.16. *Si x^* es un minimizador local de (2.0.1), entonces x^* satisface la condición rAKKT2.*

Demostración. Como x^* es un minimizador local de (2.0.1) existe $\varepsilon > 0$ tal que

$$f(x^*) \leqslant f(x) \text{ para todo } x \in \Omega \cap B(x^*, \varepsilon).$$

Luego, x^* es la única solución de

$$\textit{Minimizar } f(x) + \frac{1}{4}\|x - x^*\|^4 \textit{ sujeto a } x \in \Omega \cap B(x^*, \varepsilon). \tag{2.2.17}$$

Sea $\{\rho_k\}$ una sucesión de escalares positiva tal que $\rho_k \to +\infty$.

Consideremos el método de penalidad externa utilizando una función no cuadrática, por ejemplo, $P(g_j(x), \rho) = \frac{1}{\rho}\varphi(\rho g_j(x))$ definida en el Capítulo 1, para el problema (2.2.17).

$$\begin{aligned} &\textit{Minimizar } f(x) + \frac{1}{4}\|x - x^*\|^4 + \frac{\rho_k}{2}\left[\sum_{i=1}^{m} h_i(x)^2 + \sum_{j=1}^{p} \frac{1}{\rho_k}\varphi(\rho_k g_j(x))\right] \\ &\textit{sujeto a } x \in B(x^*, \varepsilon). \end{aligned} \tag{2.2.18}$$

Sea x^k la solución global del subproblema (2.2.18) que está bien definido ya que $B(x^*, \varepsilon)$ es compacto y las funciones son contínuas. Más aún, la sucesión $\{x^k\}$ converge a x^* y $x^k \in IntB(x^*, \varepsilon)$ para k suficientemente grande. Luego, el gradiente de la función objetivo de (2.2.18) en x^k debe ser cero para k suficientemente grande:

$$\nabla f(x^k) + \rho_k \sum_{i=1}^{m} h_i(x^k)\nabla h_i(x^k) + \sum_{j=1}^{p} \varphi'(\rho_k g_j(x^k))\nabla g_j(x^k) + \|x - x^*\|^2(x^* - x^k) = 0. \tag{2.2.19}$$

A partir de que φ tiene segundas derivadas contínuas podemos afirmar que

$$\begin{aligned} &\nabla^2 f(x^k) + \rho_k \sum_{i=1}^{m} h_i(x^k)\nabla^2 h_i(x^k) + \rho_k \sum_{j=1}^{p} \varphi'(\rho_k g_j(x^k))\nabla^2 g_j(x^k) \\ &+\rho_k \sum_{i=1}^{m} \nabla h_i(x^k)\nabla h_i(x^k)^T + \rho_k \sum_{j: g_j(x^k) \geqslant 0} \rho_k \varphi''(\rho_k g_j(x^k))\nabla g_j(x^k)\nabla g_j(x^k)^T \\ &+2(x^k - x^*)(x^k - x^*)^T + \|x^k - x^*\|^2 \succeq 0. \end{aligned} \tag{2.2.20}$$

Definimos

$$\lambda_i^k := \rho_k h_i(x^k),$$

$$\eta_i^k := \rho_k \text{ para } i \in \{1,\ldots,m\},$$
$$\mu_j^k := \varphi'(\rho_k g_j(x^k))\} \text{ para } j \in \{1,\ldots,p\},$$
$$\theta_j^k = \rho_k \varphi''(\rho_k g_j(x^k))$$

Finalmente definimos $\delta_k := 3\|x^k - x^*\|^2$. Claramente $\delta_k \to 0$ cuando k tiende a infinito, $\mu_j^k \to 0$ para $j \notin A(x^*)$ y $\theta_j^k \to 0$ para $j \notin A(x^*)$. Ahora, podemos considerar $\mu_j^k = 0$ y $\theta_j^k = 0$ para $j \notin A(x^*)$ y la desigualdad (2.2.20) se mantiene para k suficientemente grande. Eligiéndolos de esta manera, por (2.2.18) y (2.2.20), tenemos que la condición rAKKT2 se satisface. □

La diferencia entre AKKT2 y rAKKT2 es que las sucesiones $\{\theta^k\} \subseteq \mathbb{R}^p$ y $\{\delta^k\} \subseteq \mathbb{R}$ no son necesariamente no negativas en rAKKT2, pues por las características de φ, $\theta_j^k = \rho_k \varphi''(\rho_k g_j(x^k))$, no es un escalar necesariamente no negativo.

En la siguiente proposición mostraremos que las Definiciones 2.2.1 y 2.2.12 son equivalentes.

Proposición 2.2.21. *Un punto factible x^* satisface la condición estacionaria de segundo orden aproximada (AKKT2) con respecto al problema (2.0.1) si y solo si satisface la condición estacionaria de segundo orden aproximada reformulada (rAKKT2).*

Demostración. Probaremos que rAKKT2 implica AKKT2. La recíproca es inmediata.

Supongamos que x^* satisface la condición rAKKT2 con respecto al problema (2.0.1). Entonces, existen sucesiones $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m$, $\{\eta^k\} \subseteq \mathbb{R}^m$, $\{\mu^k\} \subseteq \mathbb{R}^p_+$, $\{\tilde{\theta}^k\} \subseteq \mathbb{R}^p$, $\{\tilde{\delta}^k\} \subseteq \mathbb{R}$ tales que $\lim\limits_{k\to\infty} x^k = x^*$, $\lim\limits_{k\to\infty} \tilde{\delta}^k = 0$ verificando

$$\mu_j^k = 0 \text{ para } j \notin A(x^*), \theta_j^k = 0 \text{ para } j \notin A(x^*),$$

$$\lim_{k\to\infty} \nabla_x l(x^k,\lambda^k,\mu^k) = \lim_{k\to\infty} \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k) = 0$$

y,

$$\nabla^2_{xx} l(x^k,\lambda^k,\mu^k) + \sum_{i=1}^m \eta_i^k \nabla h_i(x^k)\nabla h_i(x^k)^T + \sum_{j\in A(x^*)} \tilde{\theta}_j^k \nabla g_j(x^k)\nabla g_j(x^k)^T \succeq -\delta_k Id$$

en $\mathbb{R}^n$, para $k \in \mathbb{N}$ suficientemente grande. Supongamos que $\tilde{\theta}^k \notin \mathbb{R}^p_+$ para todo $k \in K \subset \mathbb{N}$, es decir, que existe $j \in A(x^*)$ tal que $\tilde{\theta}_j^k < 0$, para todo $k \in K$.

Consideramos los conjuntos de índices $I_k = \{j \in A(x^*) : \tilde{\theta}_j^k \geq 0, k \in K\}$ y $J_k = \{j \in A(x^*) : \tilde{\theta}_j^k < 0, k \in K\}$. Como los posibles conjuntos I_k, J_k son finitos, existe $K_1 \subset K$ tal que, para todo $k \in K_1$, $I_k = I \subset A(x^*)$, $J_k = J \subset A(x^*)$. Luego, para $k \in K_1$ suficientemente grande, por (2.2.15)

$$\begin{aligned} & d^T \nabla^2_{xx} l(x^k, \lambda^k, \mu^k) d + \sum_{i=1}^m \eta_i^k \|\nabla h_i(x^k)^T d\|^2 + \sum_{j \in I} \tilde{\theta}_j^k \|\nabla g_j(x^k)^T d\|^2 \\ \geq & - \tilde{\delta}_k \|d\|^2 - \sum_{j \in J} \tilde{\theta}_j^k \|\nabla g_j(x^k)^T d\|^2, \quad \forall d \in \mathbb{R}^n. \end{aligned}$$

Como $\tilde{\theta}_j^k < 0$ para todo $j \in J$ tenemos que

$$d^T \nabla^2_{xx} l(x^k, \lambda^k, \mu^k) d + \sum_{i=1}^m \eta_i^k \|\nabla h_i(x^k)^T d\|^2 + \sum_{j \in I} \tilde{\theta}_j^k \|\nabla g_j(x^k)^T d\|^2 \geq -\tilde{\delta}_k \|d\|^2.$$

Luego, podemos definir

$$\theta_j^k = \begin{cases} \tilde{\theta}_j^k & \text{si} \quad j \in I \\ 0 & \text{si} \quad j \in J, \end{cases}$$

y obtenemos que $\{\theta^k\} \subset \mathbb{R}^p_+$.

Si $\tilde{\delta}^k \leqslant 0$ para todo $k \in K_1$ consideramos $\delta^k = 0$ y por lo tanto las sucesiones $\{x^k\}_{k \in K_1} \subseteq \mathbb{R}^n$, $\{\lambda^k\}_{k \in K_1} \subseteq \mathbb{R}^m$, $\{\eta^k\}_{k \in K_1} \subseteq \mathbb{R}^m$, $\{\mu^k\}_{k \in K_1} \subseteq \mathbb{R}^p_+$, $\{\theta^k\}_{k \in K_1} \subseteq \mathbb{R}^p_+$, $\{\delta^k\}_{k \in K_1} \subseteq \mathbb{R}_+$ verifican la condición AKKT2.

Si no, existe $K_2 \subset K_1$ tal que $\tilde{\delta}^k \geq 0$ para todo $k \in K_2$. Luego consideramos las mismas sucesiones en el conjunto de índices K_2 y se verifica la condición AKKT2 como queríamos probar. □

Definiremos, para cada $x, y \in \mathbb{R}^n$ el siguiente cono [6]:

$$C^W(x, y) := \{d \in \mathbb{R}^n : \nabla h_i(x)^T d = 0, \nabla g_j(x)^T d = 0 \text{ para } j \in A(y)\}.$$

Observar que el cono tangente en WSONC es $T(x^*) = C^W(x^*, x^*)$.

En [6] los autores también presentan una nueva condición de calidad de segundo orden inmediatamente asociada a la condición secuencial de segundo orden AKKT2: la propiedad del cono contínuo de segundo orden CCP2.

Dado x^* tal que $h(x^*) = 0$ y $g(x^*) \leqslant 0$, para cada $x \in \mathbb{R}^n$ definimos el siguiente cono

$$K_2^W(x) := \bigcup_{\substack{(\lambda,\mu)\in\mathbb{R}^m\times\mathbb{R}^p_+ \\ \mu_j=0,\, j\notin A(x^*)}} \left\{ \begin{array}{c} \left(\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j\in A(x^*)} \mu_j \nabla g_j(x), H\right) : \\ H \preceq \sum_{i=1}^m \lambda_i \nabla^2 h_i(x) + \sum_{j\in A(x^*)} \mu_j \nabla^2 g_j(x) \text{ en } C^W(x,x^*) \end{array} \right\}.$$

Definición 2.2.22. [6] Supongamos que $h(x^*) = 0$ y $g(x^*) \leqslant 0$. Decimos que x^* satisface la propiedad del cono contínuo de segundo orden (CCP2) si

$$\limsup_{x\to x^*} K_2^W(x) \subset K_2^W(x^*). \tag{2.2.23}$$

Esto es, la multifunción $x \in \mathbb{R}^n \rightrightarrows K_2^W(x) \subseteq \mathbb{R}^n \times Sym(n)$ es semicontínua exterior en x^*.

Como se menciona en [6], $K_2^W(x)$ es un subconjunto del cono convexo de $\mathbb{R}^n \times Sym(n)$ y por lo tanto permite escribir la condición de optimalidad necesaria de segundo orden débil (WSONC) de la siguiente forma compacta

$$(-\nabla f(x^*), -\nabla^2 f(x^*)) \in K_2^W(x^*).$$

Para mostrar que CCP2 es la condición de calidad segundo orden más débil que garantiza que AKKT2 implica WSONC, se ultiliza el siguiente lema:

Lema 2.2.24. *[16] Sea P una matriz simétrica de $n\times n$ y Q una matriz simétrica y semidefinida positiva de $n \times n$. Supongamos que $x^TPx > 0$ $\forall x \neq 0$ satisfaciendo $x^TQx = 0$. Luego, existe un escalar c tal que*

$$P + cQ > 0.$$

Demostración. Por el absurdo. Supongamos que para todo entero k, existe un vector x^k con $\|x^k\| = 1$ tal que

$$(x^k)^TPx^k + k(x^k)^TQx^k \leqslant 0. \tag{2.2.25}$$

La sucesión $\{x^k\}$ tiene una subsucesión $\{x^k\}_{k\in K}$ que converge a un vector x^* con $\|x^*\| = 1$. Tomando límite superior en (2.2.25), tenemos que

$$(x^*)^TPx^* + \limsup_{k\in K}(k(x^k)^TQx^k) \leqslant 0. \tag{2.2.26}$$

Como $(x^k)^TQx^k \geq 0$, (2.2.26) implica que $\{(x^k)^TQx^k\}_k$ converge a cero y entonces $(x^*)^TQx^* = 0$. Por hipótesis teníamos que $(x^*)^TPx^* > 0$ y eso contradice (2.2.26). $\square$

Teorema 2.2.27. *[6] Supongamos que $h(x^*) = 0$ y $g(x^*) \leqslant 0$. Las siguientes condiciones son equivalentes:*

1. *CCP2 se cumple en x^* ;*
2. *Para toda función objetivo $f : \mathbb{R}^n \to \mathbb{R}$ del problema (2.0.1) tal que AKKT2 se cumple en x^*, la condición WSONC se satisface en x^*.*

Demostración. Primero supongamos que se satisface CCP2 en x^* y que existe una función f de manera que se satisface la condición AKKT2. Por la definición de AKKT2, existen sucesiones $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m$, $\{\eta^k\} \subseteq \mathbb{R}^m$, $\{\mu^k\} \subseteq \mathbb{R}^p_+$, $\{\theta^k\} \subseteq \mathbb{R}^p_+$, $\{\delta^k\} \subseteq \mathbb{R}_+$ tales que $\mu^k_j = 0$ para $j \notin A(x^*), \theta^k_j = 0$ para $j \notin A(x^*)$, $\lim\limits_{k\to\infty} x^k = x^*$, $\lim\limits_{k\to\infty} \delta^k = 0$ con

1. $\varepsilon_k := \nabla f(x^k) + \sum_{i=1}^{m} \lambda^k_i \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu^k_j \nabla g_j(x^k) \to 0$
2. $\nabla^2_{xx} l(x^k, \lambda^k, \mu^k) + \sum_{i=1}^{m} \eta^k_i \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j\in A(x^*)} \theta^k_j \nabla g_j(x^k) \nabla g_j(x^k)^T \succeq -\delta_k Id.$

De 1. y 2. podemos ver que

$$(-\nabla f(x^k) + \varepsilon_k, -\nabla^2 f(x^k) - \delta_k Id) \in K^W_2(x^k).$$

Usando la continuidad de ∇f y $\nabla^2 f(x)$ y la semicontinuidad exterior en x^* obtenemos que

$$(-\nabla f(x^*), -\nabla^2 f(x^*)) \in K^W_2(x^*)$$

y en consecuencia WSONC se cumple.

Probaremos ahora la otra implicación. Sea (w, W) un elemento de $\limsup K^W_2(x)$ cuando $x \to x^*$. Veamos que $(w, W) \in K^W_2(x^*)$. Por definición de límite exterior existen sucesiones $\{x^k\}$, $\{\lambda^k_i\}$, $\{\mu^k_j\}$ con $\mu^k_j = 0$ para $j \notin A(x^*)$ y $\{H^k\} \subset Sym(n)$ tales que $x^k \to x^*$,

$$\left(\sum_{i=1}^{m} \lambda^k_i \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu^k_j \nabla^2 g_j(x^k), H^k \right) \to (w, W)$$

y

$$H^k \preceq \sum_{i=1}^{m} \lambda^k_i \nabla^2 h_i(x^k) + \sum_{j\in A(x^*)} \mu^k_j \nabla^2 g_j(x^k) \text{ en } C^W(x^k, x^*).$$

Definimos $f(x) := -\langle w, x - x^* \rangle - \frac{1}{2} W(x - x^*, x - x^*)$. Mostraremos que AKKT2 se cumple en x^* con $f(x)$ como función objetivo. Claramente, tenemos que $\nabla f(x) = -w - W(x - x^*)$ y $\nabla^2 f(x) = -W$.

Para probar (2.2.14) es suficiente con ver que $\lim\limits_{k\to\infty} \nabla_x l(x^k, \lambda^k, \mu^k) = 0$, pero esto es trivial ya que

$$\lim_{k\to\infty} \left(\sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k) - w \right) - \lim_{k\to\infty} W(x^k - x^*) = 0.$$

Para probar (2.2.15) usaremos el Lema 2.2.24 con

$$P^k := \sum_{i=1}^{m} \lambda_i^k \nabla^2 h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla^2 g_j(x^k) - H^k + \frac{1}{k} Id \tag{2.2.28}$$

y Q la matriz formada por las columnas de la matriz

$$\left[\nabla h_i(x^k), i \in \{1, \dots, m\}; \nabla g_j(x^k), j \in A(x^*)\right].$$

Por el Lema 2.2.24 existen sucesiones positivas $\{\theta^k\}$ y $\{\eta^k\}$ tales que

$$S^k := P^k + \sum_{i=1}^{m} \eta^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j\in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succ 0. \tag{2.2.29}$$

Elegimos $\theta_j^k = 0$ para $j \notin A(x^k)$. Usando (2.2.28), (2.2.29) y que $\nabla^2 f(x) = -W$ tenemos que

$$\begin{gathered} \nabla^2 l(x^k, \lambda^k, \mu^k) + \sum_{i=1}^{m} \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j\in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T = \\ -W + H^k + S^k - \tfrac{1}{k} Id. \end{gathered} \tag{2.2.30}$$

Buscaremos ahora una cota superior para esta matriz. Por el principio de Rayleigh tenemos que

$$-W + H^k \succeq -|\lambda_1(W - H^k)| Id,$$

donde $\lambda_1(W - H^k)$ denota el menos autovalor de $(W - H^k)$. Por (2.2.29), $S^k \succ 0$ y por lo tanto tenemos que

$$\begin{gathered} -W + H^k + S^k - \tfrac{1}{k} Id \succeq -|\lambda_1(W - H^k)| Id + S^k - \tfrac{1}{k} Id \succeq \\ -|\lambda_1(W - H^k)| Id - \tfrac{1}{k} Id = -\delta_k Id, \end{gathered} \tag{2.2.31}$$

donde $\delta_k = -|\lambda_1(W - H^k)| + \frac{1}{k}$. Como $H^k \to W$ cuando k tiende a infinito, δ_k tiende a cero. Por (2.2.31) y (2.2.30) vemos que la condición (2.2.15) se satisface, y por lo tanto x^* es un punto AKKT2. Luego, por hipótesis, WSONC se cumple y por (2.2.23) se tiene que

$$(w, W) = (-\nabla f(x^*), -\nabla^2 f(x^*)) \in K_2^W(x^*)$$

como queríamos probar. □

En [6] los autores prueban que CCP2 es una condición de calidad más débil que la condición MFCQ+WCR [4] y que RCRCQ [50].

En relación con las definiciones de AKKT2 y rAKKT2 podemos observar que

- La desiguladad (2.2.15) se debe cumplir en todo $\mathbb{R}^n$,
- La desigualdad (2.2.15) incluye las matrices de rango uno: $\nabla h_i(x^k)\nabla h_i(x^k)^T$ y $\nabla g_j(x^k)\nabla g_j(x^k)^T$.

Como la condición de optimalidad necesaria de segundo orden (WSONC) establece que el Hessiano de la función de Lagrange debe ser semidefinido positivo en el subespacio tangente, creemos que es interesante estudiar condiciones secuenciales de segundo orden considerando la desigualdad (2.2.15) en el subespacio tangente o en una aproximación del subespacio tangente. En ese contexto, en la siguiente definición sugerimos una nueva condición de optimalidad de segundo orden.

Definición 2.2.32. Supongamos que $h(x^*) = 0$ y $g(x^*) \leqslant 0$. Decimos que un punto x^* satisface la nueva condición aproximada estacionaria de segundo orden en un tangente aproximado (AKKT2AT) con respecto al problema (2.0.1) si existen sucesiones $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m$, $\{\mu^k\} \subseteq \mathbb{R}^p_+$, $\{\delta^k\} \subseteq \mathbb{R}_+$, $\{\varepsilon_k\} \subseteq \mathbb{R}_+$ tales que $\lim_{k\to\infty} x^k = x^*$, $\lim_{k\to\infty} \delta^k = 0$, $\lim_{k\to\infty} \varepsilon_k = 0$ verificando

$$\mu_j^k = 0 \text{ para } j \notin A(x^*),$$

$$\lim_{k\to\infty} \nabla_x l(x^k, \lambda^k, \mu^k) = \lim_{k\to\infty} \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k) = 0$$

y, para $k \in \mathbb{N}$ suficientemente grande,

$$\nabla^2_{xx} l(x^k, \lambda^k, \mu^k) \succeq -\delta_k Id \qquad \text{en } C_k \tag{2.2.33}$$

donde C_k es el cono definido por

$$C_k = \{d \in \mathbb{R}^n : \|d\| = 1, |\nabla h_i(x^k)^T d| \leqslant \varepsilon_k, i = 1, \ldots, m; \\ |\nabla g_j(x^k)^T d| \leqslant \varepsilon_k, j : g_j(x^*) = 0\}.$$

En la siguiente proposición probaremos que la nueva condición secuencial de optimalidad es AKKT2AT es más débil que rAKKT2. Como consecuencia, podemos concluir que es una condición genuina.

Proposición 2.2.34. *Si un punto factible x^* satisface la condición estacionaria aproximada de segundo orden (rAKKT2) con respecto al problema (2.0.1) entonces x^* satisface la condición estacionaria aproximada de segundo orden en un tangente aproximado (AKKT2AT).*

Demostración. Supongamos que x^* satisface la condición rAKKT2 con respecto al problema (2.0.1). Entonces, existen sucesiones $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m$, $\{\eta^k\} \subseteq \mathbb{R}^m$, $\{\mu^k\} \subseteq \mathbb{R}^p_+$, $\{\tilde{\theta}^k\} \subseteq \mathbb{R}^p$, $\{\tilde{\delta}^k\} \subseteq \mathbb{R}$ tales que $\lim\limits_{k\to\infty} x^k = x^*$, $\lim\limits_{k\to\infty} \tilde{\delta}^k = 0$ verificando

$$\mu_j^k = 0 \text{ para } j \notin A(x^*), \tilde{\theta}_j^k = 0 \text{ para } j \notin A(x^*),$$

$$\lim_{k\to\infty} \nabla_x l(x^k, \lambda^k, \mu^k) = \lim_{k\to\infty} \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k) = 0$$

y, para $k \in \mathbb{N}$ suficientemente grande,

$$\nabla^2_{xx} l(x^k, \lambda^k, \mu^k) + \sum_{i=1}^m \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j\in A(x^*)} \tilde{\theta}_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succeq -\tilde{\delta}_k Id.$$

en $\mathbb{R}^n$ para $k \in \mathbb{N}$ suficientemente grande. Luego, para todo $d \in C_k$

$$-\tilde{\delta}_k \leqslant d^T \nabla^2_{xx} l(x^k, \lambda^k, \mu^k) d + \sum_{i=1}^m \eta_i^k |\nabla h_i(x^k)^T d|^2 + \sum_{j\in A(x^*)} \tilde{\theta}_j^k |\nabla g_j(x^k)^T d|^2$$

$$\leqslant d^T \nabla^2_{xx} l(x^k, \lambda^k, \mu^k) d + \|\eta^k\| \sum_{i=1}^m |\nabla h_i(x^k)^T d|^2 + \|\tilde{\theta}^k\| \sum_{j\in A(x^*)} |\nabla g_j(x^k)^T d|^2$$

$$\leqslant d^T \nabla^2_{xx} l(x^k, \lambda^k, \mu^k) d + \text{máx}\{\|\eta^k\|, \|\tilde{\theta}^k\|\} \left(\sum_{i=1}^m |\nabla h_i(x^k)^T d|^2 + \sum_{j\in A(x^*)} |\nabla g_j(x^k)^T d|^2 \right).$$

Si $\text{máx}\{\|\eta^k\|, \|\tilde{\theta}^k\|\}$ es acotada, consideramos ε_k tal que $\lim\limits_{k\to\infty} \varepsilon_k = 0$ y definimos

$$\delta_k = \tilde{\delta}_k + (m+p)\,\text{máx}\{\|\eta^k\|, \|\tilde{\theta}^k\|\} \varepsilon_k^2$$

.

Luego, para todo $d \in C_k$,

$$-\delta_k \leqslant d^T \nabla^2_{xx} l(x^k, \lambda^k, \mu^k) d.$$

Sino, $\lim\limits_{k\to\infty} \max\{\|\eta^k\|, \|\tilde{\theta}^k\|\} = +\infty$. En este caso, consideramos

$$\varepsilon_k = \frac{1}{\max\{\|\eta^k\|, \|\tilde{\theta}^k\|\}} \quad \text{y} \quad \delta_k = \tilde{\delta}_k + \varepsilon_k(m+p).$$

En consecuencia, las sucesiones $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m$, $\{\mu^k\} \subseteq \mathbb{R}^p_+$, $\{\varepsilon_k\} \subseteq \mathbb{R}_+$, $\{\delta^k\} \subseteq \mathbb{R}_+$ verifican la condición AKKT2AT como queríamos probar. $\square$

Capítulo 3

Método de Lagrangiano Aumentado utilizando la función de penalidad exponencial

En este Capítulo consideremos el problema de programación no lineal con restricciones de desigualdad

$$\textit{Minimizar } f(x) \textit{ sujeto a } g(x) \leqslant 0. \tag{3.0.1}$$

Comenzaremos estudiando la función de penalidad exponencial, que cumple los axiomas **C1-C4** definidos en la sección 1.2 del Capítulo 1. De esta manera la función Lagrangiana Aumentada será de la siguiente manera:

$$L(x,\mu,\rho) = f(x) + \sum_{j=1}^{p} \frac{\mu_j}{\rho}(e^{\rho g_j(x)} - 1), \tag{3.0.2}$$

para $x \in \mathbb{R}^n$, $\rho > 0$ y $\mu_j \geqslant 0$.

3.1. Convergencia global

En el teorema en el que demostraremos la convergencia del Algoritmo 1, definido en la sección 1.2 del Capítulo 1, cuando se utiliza la función de penalidad exponencial, también demostraremos que si la condición de calidad de Mangasarian-Fromovitz se satisface, el conjunto de multiplicadores de Lagrange es acotado. Para eso necesitamos las siguientes definiciones:

Definición 3.1.1. Sean $A = \{a_1, \ldots, a_m\}$, $B = \{b_1, \ldots, b_p\}$ dos conjuntos finitos de vectores en $\mathbb{R}^n$. Decimos que $A \cup B$ es positivo linealmente dependiente (PLD) si existen escalares $\alpha \in \mathbb{R}^m$, $\beta \in \mathbb{R}^p$ tal que $\beta \geqslant 0$, $(\alpha, \beta) \neq 0$ y

$$\sum_{j=1}^{m} \alpha_j a_j + \sum_{j=1}^{p} \beta_j b_j = 0.$$

En caso contrario, se dice que $A \cup B$ es positivo linealmente independiente (PLI).

Definición 3.1.2. [45] Decimos que un punto factible x^* satisface la condición de calidad de Mangasarian-Fromovitz (MFCQ) si los gradientes

$$\{\nabla h_j(x^*)\}_{j \in \{1,\ldots,m\}} \cup \{\nabla g_j(x^*)\}_{j \in A(x^*)}$$

son PLI.

Cuando se considera la función de penalidad cuadrática, se puede demostrar que

$$\mu_j^* = 0 \text{ cuando } j \notin A(x^*).$$

Si se consideran funciones de penalidad no cuadráticas, esto no es necesariamente así, sin embargo, demostraremos que la sucesión de multiplicadores tiende a cero, cuando k tiende a infinito.

Observar que para la función de penalidad exponencial se tiene que el estimado del multiplicador de Lagrange, (1.2.4), asociado a la restricción $g_j(x) \leqslant 0$ es

$$\mu_j^{k+1} = \bar{\mu}_j^k e^{\rho_k g_j(x^k)}$$

Lema 3.1.3. *Sea $\{x^k\}$ una sucesión generada por el Algoritmo 1, definido en la sección 1.2 del Capítulo 1. Supongamos que x^* es un punto límite factible de $\{x^k\}$ y que $K \subset \mathbb{N}$ es tal que $\lim_{k \in K} x^k = x^*$. Luego,*

$$\lim_{k \in K} \mu_j^{k+1} = 0 \qquad \forall j \notin A(x^*).$$

Demostración. Para cada $j \notin A(x^*)$, consideramos dos casos:

- Si $\{\rho_k\}$ es acotada, por el Paso 3 del Algoritmo 1,

$$\lim_{k \in K} \sigma_j^k = \lim_{k \in K} \bar{\mu}_j^k \left(\frac{e^{\rho_k g_j(x^k)} - 1}{\rho_k}\right) = 0.$$

Como $g_j(x^*) < 0$, $\lim_{k \in K} \dfrac{e^{\rho_k g_j(x^k)} - 1}{\rho_k} \neq 0$ y esto implica que $\lim_{k \in K} \bar{\mu}_j^k = 0$ y como, para este caso, $\mu_j^{k+1} = \bar{\mu}_j^k e^{\rho_k g_j(x^k)}$ se cumple que $\lim_{k \in K} \mu_j^{k+1} = 0$.

- Si $\{\rho_k\}$ no es acotada, como $g_j(x^*) < 0$, tenemos que $\lim\limits_{k\in K} e^{\rho_k g_j(x^k)} = 0$, y como $\{\bar{\mu}_j^k\}_{k\in K}$ es acotada, obtenemos que

$$\lim_{k\in K} \mu_j^{k+1} = \lim_{k\in K} \bar{\mu}^k e^{\rho_k g_j(x^k)} = 0.$$

Por lo tanto, el lema esta probado. □

Veamos ahora el teorema de la convergencia global. Para demostrar este resultado en el trabajo [28] utilizamos la condición de calidad de Generador Positivo Constante [8], pero en [9] se demostró que la condición CCP es más débil y por tal motivo el teorema de convergencia global lo probaremos con esta última condición.

En el Paso 1 del Algoritmo 1, consideramos como criterio de parada la condición inexacta de primer orden definida en el Capítulo 2, es decir,

$$\left\|\nabla_x L(x^k, \bar{\mu}^k, \rho_k)\right\| \leq \varepsilon_k, \tag{3.1.4}$$

$$\mu_j^k = 0 \text{ cuando } g_j(x^k) < -\varepsilon_k, \; j = 1, \ldots, p, \tag{3.1.5}$$

para todo $k \in \mathbb{N}$.

Teorema 3.1.6. *Sea $\{x^k\}$ una sucesión generada por el Algoritmo 1 definido en la sección 1.2 del Capítulo 1, donde cada x^k verifica las condiciones (3.1.4)-(3.1.5). Sea x^* un punto límite de la sucesión $\{x^k\}$ que satisface la Propiedad del Cono Continuo (CCP) para el problema (3.0.1). Luego, x^* es un punto Karush-Kuhn-Tucker del problema (3.0.1).*

Más aún, si x^* satisface la condición de calidad de Mangasarian-Fromovitz y $\{x^k\}_{k\in K}$ es una subsucesión que converge a x^*, el conjunto

$$\{\|\mu^{k+1}\|\}_{k\in K} \quad \textit{es acotado.} \tag{3.1.7}$$

Demostración. Para todo $k \in \mathbb{N}$, por (1.2.1) y (1.2.4) existe $\delta_k \in \mathbb{R}^n$ tal que $\|\delta_k\| \leqslant \varepsilon_k$ y

$$\nabla f(x^k) + \sum_{j=1}^{p} \mu_j^{k+1} \nabla g_j(x^k) = \delta_k.$$

Sea $K_1 \subset \mathbb{N}$ tal que $\lim\limits_{k\in K_1} x^k = x^*$. Luego,

$$\nabla f(x^k) + \sum_{j\in A(x^*)} \mu_j^{k+1} \nabla g_j(x^k) = \delta_k - \sum_{j\notin A(x^*)} \mu_j^{k+1} \nabla g_j(x^k). \tag{3.1.8}$$

Además, por el Lema 3.1.3, tenemos que

$$\zeta^k = \nabla f(x^k) + \sum_{j \in A(x^*)} \mu_j^{k+1} \nabla g_j(x^k) \to 0.$$

Definiendo $\omega^k = \sum_{j \in A(x^*)} \mu_j^{k+1} \nabla g_j(x^k)$ vemos que

$$\omega^k \in K(x^k) \qquad \text{y} \qquad \omega^k = \zeta^k - \nabla f(x^k) - \sum_{j \notin A(x^*)} \mu_j^{k+1} \nabla g_j(x^k). \tag{3.1.9}$$

Tomando límites en esta ecuación cuando k tiende a infinito, usando la continuidad de f y que $\zeta^k \to 0$, tenemos que

$$-\nabla f(x^*) = \lim \omega^k \in \limsup_{k \to \infty} K(x^k) \subset \limsup_{x \to x^*} K(x) \subset K(x^*).$$

La última inclusión viene de aplicar la propiedad del Cono Continuo.

Por lo tanto,

$$-\nabla f(x^k) \in K(x^*),$$

lo que es equivalente a decir que x^* es un punto KKT.

Para probar la segunda parte, supongamos que la sucesión $\{\mu^{k+1}\}_{k \in K_1}$ no es acotada. Definimos, para todo $k \in K$,

$$M_k = \max\{\mu_j^{k+1}, j \in A(x^*)\}.$$

Por lo tanto, $\lim_{k \in K} M_k = +\infty$. Luego podemos dividir (3.1.8) por M_k y se tiene

$$\frac{\nabla f(x^k)}{M_k} + \sum_{j \in A(x^*)} \frac{\mu_j^{k+1}}{M_k} \nabla g_j(x^k) = \frac{\delta_k}{M_k} - \sum_{j \notin A(x^*)} \frac{\mu_j^{k+1}}{M_k} \nabla g_j(x^k).$$

Definimos $\tilde{\mu}^k = \mu^{k+1}/M_k$. La sucesión $\{\tilde{\mu}^k\}_{k \in K}$ es acotada. Resulta que tiene alguna subsucesión convergente y, por como elegimos M_k, no puede ser nula. Tomando límites para esta subsucesión, se tiene que

$$\sum_{j \in A(x^*)} \tilde{\mu}_j^* \nabla g_j(x^*) = 0$$

y por lo tanto que x^* no satisface la condición de calidad de Mangasarian-Fromovitz.

□

3.2. Acotación del parámetro de penalidad utilizando la función de penalidad exponencial

Es conocido que los métodos Lagrangiano Aumentado, a diferencia de los métodos de penalidad, intentan localizar el valor óptimo del problema manteniendo la sucesión $\{\rho_k\}$ acotada para evitar el mal condicionamiento en el límite. Debido a esto, en esta sección estudiaremos condiciones bajo las cuales la sucesión de parámetros de penalidad en el Algoritmo 1 de Lagrangiano Aumentado es acotada.

Cuando se consideran restricciones de desigualdad, las hipótesis usuales para demostrar la convergencia son: la condición suficiente de optimalidad de segundo orden (el Hessiano de la función de Lagrange definido positivo en el subespacio ortogonal de los gradientes de las restriciones activas) y la independencia lineal de las restricciones activas junto con la condición de complementariedad estricta. Estas son las condiciones usadas en [2] para probar que el parámetro de penalidad es acotado en el método de Lagrangiano Aumentado cuadrático ALGENCAN y en [27] para analizar el comportamiento asintótico cuando se consideran funciones de penalidad no cuadráticas.

Similares, pero ligeramente más restrictivas son las condiciones consideradas en [25] para obtener la acotación del parámetro de penalidad para LANCELOT. Recientemente, en [19], los autores demostraron que el parárametro de penalidad permanece acotado para un método de Lagrangiano Aumentado cuadrático bajo las siguientes hipótesis: en lugar de la independencia lineal de las restricciones activas asumen que se satisface la condición de calidad de Mangasarian-Fromovitz y que el vector de multiplicadores de Lagrange es único y, además, utiliza una condición suficiente de optimalidad de segundo orden que no involucra a la condición de complementariedad estricta. Es importante mencionar que en [19] la medida usada para actualizar el parámetro de penalización es diferente de la usada en [1, 2, 18, 25].

En [19] los autores aumentan el parámetro de penalidad usando la medida

$$\sigma(x,\mu) = \left\| \begin{pmatrix} \nabla l(x,\mu,\rho) \\ \min\{-g(x),\mu\} \end{pmatrix} \right\| \tag{3.2.1}$$

y la cota es obtenida usando la teoría de la cota del error local [29, 32].

Para demostrar la acotación del parámetro de penalidad utilizaremos las siguientes hipótesis:

Hipótesis 1. La sucesión $\{x^k\}$ es generada por el Algoritmo 1 y $\lim_{k\to\infty} x^k = x^*$. Además, las funciones f y g admiten segundas derivadas continuas en un entorno de x^*.

Supongamos que el conjunto de índices de restricciones activas en x^* es $A(x^*) =$

$\{1, \ldots, q\}$ y $J = \{q+1, \ldots, p\}$.

Hipótesis 2. Los gradientes $\{\nabla g_1(x^*), \ldots, \nabla g_q(x^*)\}$ son linealmente independientes.

Hipótesis 3. La condición suficiente de segundo orden para minimizadores locales se satisface en (x^*, μ^*). Para todo $z \in T(x^*)$, $z \neq 0$,

$$z^T \nabla^2 l(x^*, \mu^*) z > 0, \tag{3.2.2}$$

donde $T(x^*)$ es el subespacio tangente:

$$T(x^*) = \{z \in \mathbb{R}^n : \nabla g_j(x^*)^T z = 0, j \text{ tal que } g_j(x^*) = 0\}. \tag{3.2.3}$$

También asumimos que la condición de complementariedad estricta se cumple:

$$\mu_j^* > 0 \text{ para todo } j = 1, \ldots, q.$$

Hipótesis 4. Para todo $j = 1, \ldots, p$, $\mu_j^* \in [0, \bar{\mu}_j^{\text{máx}})$.

El siguiente lema asegura que, si x^* es un punto que verifica las condiciones de Karush-Kuhn-Tucker y además que verifica la Hipótesis 3, entonces el punto x^* es un minimizador local de la función Lagrangiana Aumentada exponencial $L(x, \mu^*, \rho)$ para todo $\rho \geqslant \bar{\rho}$. Este resultado fue probado para el caso cuadrático. Ver, por ejemplo, [2, 17].

Lema 3.2.4. *Supongamos que x^* es un punto que satisface las condiciones de Karush -Kuhn-Tucker para el problema (3.0.1) y que la Hipótesis 3 se cumple. Luego, existe $\bar{\rho} > 0$ tal que, para $\rho \geqslant \bar{\rho}$,*

$$z^T \nabla^2_{xx} L(x^*, \mu^*, \rho) z > 0$$

para todo $z \in T(x^*)$, $z \neq 0$.

Demostración. Por la definición de $L(x, \mu, \rho)$ y las hipótesis tenemos que

$$\nabla^2_{xx} L(x^*, \mu^*, \rho) = \nabla^2_{xx} l(x^*, \mu^*) + \rho \sum_{j \in A(x^*)} \mu_j^* \nabla g_j(x^*) \nabla g_j(x^*)^T.$$

Definimos $P = \nabla^2_{xx} l(x^*, \mu^*)$ y $Q = \sum_{j \in A(x^*)} \mu_j^* \nabla g_j(x^*) \nabla g_j(x^*)^T$. Luego, para todo $z \neq 0, z \in T(x^*)$, $z^T Q z = \sum_{j \in A(x^*)} \mu_j^* \|\nabla g_j(x^*)^T z\|^2 = 0$ y, utilizando la Hipótesis 3, se tiene que $z^T P z > 0$. Por lo tanto, usando el Lema 2.2.24 obtenemos el resultado deseado. □

En lo que sigue daremos algunos resultados necesarios para proporcionar estimaciones de una solución aproximada de $L(x, \mu, \rho)$ a x^* y la estimación del multiplicador de Lagrange correspondiente a μ^*. El análisis sigue las ideas de [2, 27].

Lema 3.2.5. *Supongamos que se satisfacen las Hipótesis 2 y 3. Consideremos $\bar{\rho} > 0$ como en el Lema 3.2.5. Luego, para todo $r \in [0, \frac{1}{\bar{\rho}}]$, la matriz*

$$H = \begin{pmatrix} \nabla^2_{xx} l(x^*, \mu^*) & [\nabla g(x^*)]_{A(x^*)} & [\nabla g(x^*)]_J \\ [\nabla g(x^*)]^T_{A(x^*)} & [-\frac{r}{\mu^*}]_{A(x^*)} & 0 \\ 0 & 0 & I \end{pmatrix}$$

es no singular.

Demostración. Consideremos primero el caso en que $r = 0$. Si H es singular existe un vector $\bar{x} = (y, z) \in \mathbb{R}^{n+p}$, $(y, z) \neq 0$ tal que $H\bar{x} = 0$. Entonces se tiene el sistema de ecuaciones

$$\nabla^2_{xx} l(x^*, \mu^*)y + \sum_{j=1}^{p} \nabla g_j(x^*)z_j = 0, \tag{3.2.6}$$

$$\nabla g_j(x^*)^T y = 0 \qquad j = 1, \ldots, q, \tag{3.2.7}$$

$$z_j = 0 \qquad j = q + 1, \ldots, p, \tag{3.2.8}$$

premultiplicando por y^T en la (3.2.6) y usando (3.2.7) y (3.2.8) obtenemos que

$$y^T \nabla^2_{xx} l(x^*, \mu^*)y = 0.$$

Por lo tanto, por la Hipótesis 3 $y \in T(x^*)$, luego $y = 0$. Finalmente, por (3.2.6) y la Hipótesis 2, $z = 0$ y obtenemos una contradicción. Por lo tanto, H es no singular cuando $r = 0$.

Ahora analicemos el caso en que $r \in (0, \frac{1}{\bar{\rho}}]$. Supongamos que existe un vector $\bar{x} = (y, z) \in \mathbb{R}^{n+p}$, $(y, z) \neq 0$ tal que $H\bar{x} = 0$. Luego,

$$\nabla^2_{xx} l(x^*, \mu^*)y + \sum_{j=1}^{p} \nabla g_j(x^*)z_j = 0 \tag{3.2.9}$$

$$\nabla g_j(x^*)^T y - \frac{r}{\mu_j^*} z_j = 0 \qquad j = 1, \ldots, q \tag{3.2.10}$$

$$z_j = 0 \qquad j = q + 1, \ldots, p. \tag{3.2.11}$$

Para $j = 1, \ldots, q$, usando (3.2.10) tenemos que

$$z_j = \frac{\mu_j^* \nabla g_j(x^*)^T y}{r}$$

y reemplazando en (3.2.9):

$$\nabla^2_{xx} l(x^*, \mu^*)y + \sum_{j=1}^{q} \frac{\mu_j^* \nabla g_j(x^*) \nabla g_j(x^*)^T y}{r} = 0.$$

Luego,

$$y^T \nabla^2_{xx} l(x^*, \mu^*)y + \sum_{j=1}^{q} \frac{\mu_j^*}{r} ||\nabla g_j(x^*)^T y||^2 = 0.$$

En consecuencia, $y^T \nabla^2_{xx} L(x^*, \mu^*, \frac{1}{r})y = 0$ para todo $r = \frac{1}{\rho}$ tal que $\rho \geqslant \bar{\rho}$, y usando el lema previo, $y = 0$. Por lo tanto, por la definición de z, se tiene que $z = 0$ obteniendo una contradicción. □

Teorema de la Función Implícita. [17] *Sea $F : \mathbb{R}^{m+n} \to \mathbb{R}^m$ una función de $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^m$ tal que:*

1. *$F(\bar{x}, \bar{y}) = 0$.*
2. *F es contínua, y existe $\nabla_y F(x, y)$ contínua en un conjunto abierto que contiene a $(\bar{x}, \bar{y})$.*

Luego, exiten conjuntos abiertos $S_{\bar{x}} \subset \mathbb{R}^n$ y $S_{\bar{y}} \subset \mathbb{R}^m$ conteniendo a $\bar{x}$ y $\bar{y}$, respectivamente, y una función contínua $\phi : S_{\bar{x}} \to S_{\bar{y}} \subset \mathbb{R}^m$ tal que $\bar{y} = \phi(\bar{x})$ y $F(x, \phi(x)) = 0$ para todo $x \in S_{\bar{x}}$. La función ϕ es única en el sentido que si $x \in S_{\bar{x}}$, $y \in S_{\bar{y}}$ y $f(x, y) = 0$, luego, $y = \phi(x)$. Más aún, si para algún $p > 0$, F es p veces coninuamente diferenciable, lo mismo es válido para ϕ y tenemos que

$$\nabla \phi(x) = -\nabla_x F(x, \phi(x))(\nabla_y F(x, \phi(x)))^{-1}, \quad \forall x \in S_{\bar{x}}.$$

Teorema de Valor Medio. *Dada $F : U \to \mathbb{R}$ diferenciable en un abierto $U \subset \mathbb{R}^n$, si el segmento de recta $[a, a+v]$ está contenido en U, entonces existe $\theta \in (0, 1)$ tal que*

$$\begin{aligned} F(a+v) - F(a) &= \frac{\partial F}{\partial v}(a + \theta v) = \langle \nabla^F(a + \theta v), v \rangle \\ &= \sum_{i=1}^{n} \frac{\partial F}{\partial x_i}(a + \theta v)\alpha_i \end{aligned}$$

donde $v = (\alpha_1, \ldots, \alpha_n)$.

Lema 3.2.12. *Supongamos que se satisfacen las Hipótesis 1-3. Luego, existen funciones diferenciables $x(\mu, r, \alpha)$ y $y(\mu, r, \alpha)$ tales que para $r > 0$:*

1. Las funciones son soluciones del siguiente sistema

$$\begin{cases} \nabla f(x) + \sum_{j=1}^{p} y_j \nabla g_j(x) = \alpha \\ g_j(x) - rln(\dfrac{y_j}{\mu_j}) = 0, i = 1, \ldots, q \\ y_j - \mu_j e^{\frac{g_j(x)}{r}} = 0, i = q+1, \ldots, p. \end{cases}$$

2. Para r, $\|\alpha\|$ y $\|\mu - \mu^*\|$ son suficientemente chicos, se tienen las siguientes cotas:

a) $\|x(\mu, r, \alpha) - x^*\| \leqslant M \text{máx}\{r\|\mu - \mu^*\|, \|\alpha\|\}$;

b) $\|y(\mu, r, \alpha) - \mu^*\| \leqslant M \text{máx}\{r\|\mu - \mu^*\|, \|\alpha\|\}$.

Demostración. Para $\rho > 0$ consideremos el siguiente sistema de ecuaciones en las variables $(x, y, \mu, \rho, \alpha) \in \mathbb{R}^{n+p+p+1+n}$:

$$\begin{cases} \nabla f(x) + \sum_{j=1}^{p} y_j \nabla g_j(x) - \alpha = 0 & \\ g_j(x) - \frac{1}{\rho} ln(\frac{y_j}{\mu_j}) = 0 & j = 1, \ldots, q \\ y_j - \mu_j e^{\rho g_j(x)} = 0 & j = q+1, \ldots, p. \end{cases}$$

Definiendo la variable $r = \frac{1}{\rho}$ obtenemos el sistema

$$\Theta(x, y, \mu, r, \alpha) = \begin{pmatrix} \nabla f(x) + \sum_{j=1}^{p} y_j \nabla g_j(x) - \alpha \\ g_1(x) - rln(\frac{y_1}{\mu_1}) \\ \vdots \\ g_q(x) - rln(\frac{y_q}{\mu_q}) \\ y_{q+1} - \mu_{q+1} e^{\frac{g_{q+1}(x)}{r}} \\ \vdots \\ y_p - \mu_p e^{\frac{g_p(x)}{r}} \end{pmatrix} = 0.$$

Por la Hipótesis 2, para $r \in (0, \frac{1}{\bar{\rho}}]$ tenemos que $\Theta(x^*, \mu^*, \mu^*, r, 0) = 0$. Más aún, la matriz Jacobiana de Θ con respecto a (x, y) evaluada en $(x^*, \mu^*, \mu^*, r, 0)$ es la matriz

$$H = \begin{pmatrix} \nabla^2_{xx} l(x^*, \mu^*) & [\nabla g(x^*)]_{A(x^*)} & [\nabla g(x^*)]_J \\ [\nabla g(x^*)]^T_{A(x^*)} & [-\frac{r}{\mu^*}]_{A(x^*)} & 0 \\ 0 & 0 & I \end{pmatrix}.$$

Por el Lema 3.2.5, esta matriz es no singular para $r \in [0, \frac{1}{\bar{\rho}}]$. Utilizando la continuidad, se tiene que la norma de la matriz inversa es acotada uniformemente en un entorno de $(x^*, \mu^*, \mu^*, r, 0)$. Usando el Teorema de la Función Implícita, podemos asegurar que existen un entorno V_r de $(\mu^*, r, 0)$ y funciones diferenciables $x(\mu, r, \alpha), y(\mu, r, \alpha)$ tales que para todo $(\mu, r, \alpha) \in V_r, r \in (0, \frac{1}{\bar{\rho}}]$:

$$\begin{cases} \nabla f(x(\mu, r, \alpha)) + \sum_{j=1}^{p} \nabla g(x(\mu, r, \alpha)) y(\mu, r, \alpha) - \alpha = 0 & \\ g_j(x(\mu, r, \alpha)) - rln\left(\dfrac{y_j(\mu, r, \alpha)}{\mu_j}\right) = 0 & j \in A(x^*) \\ y_j(\mu, r, \alpha) - \mu_j e^{\frac{g_j(x(\mu, r, \alpha))}{r}} = 0 & j \in J. \end{cases} \tag{3.2.13}$$

Por el Teorema de Valor Medio existe $c \in \mathbb{R}^{m+2}$ tal que

$$y(\mu, r, \alpha) - \mu^* = \nabla y(c)(\mu - \mu^*, r, \alpha) = \nabla_\mu y(c)(\mu - \mu^*) + \nabla_r y(c) r + \nabla_\alpha y(c)\alpha.$$

y obtenemos que

$$\|y(\mu, r, \alpha) - \mu^*\| \leqslant \|\nabla_\mu y(c)\| \|\mu - \mu^*\| + \|\nabla_r y(c)\| r + \|\nabla_\alpha y(c)\| \|\alpha\|. \tag{3.2.14}$$

Calculando las primeras derivadas de $\Theta(x(\mu, r, \alpha), y(\mu, r, \alpha), \mu, r, \alpha) = 0$ con respecto a (μ, r, α) tenemos que existe una constante C tal que

- $\|\nabla_\alpha y(c)\| \leqslant C \|\nabla_\alpha \Theta\|$,
- $\|\nabla_r y(c)\| \leqslant C \|\nabla_r \Theta\|$,
- $\|\nabla_\mu y(c)\| \leqslant C \|\nabla_\mu \Theta\|$.

Buscaremos cotas para $\|\nabla_\mu \Theta\|$, $\|\nabla_r \Theta\|$ y $\|\nabla_\alpha \Theta\|$.

(i) $\nabla_\alpha \Theta(x, y, \mu, r, \alpha) = (-I \;\; 0 \;\; 0)$, luego, $\|\nabla_\alpha \Theta\| = 1$.

(ii) $\nabla_\mu \Theta(x, y, \mu, r, \alpha) = \begin{pmatrix} 0 & \left[\frac{r}{\mu}\right]_{A(x^*)} & \left[-e^{\frac{g(x)}{r}}\right]_J \end{pmatrix}$. Entonces,

- si $j \in J$, para $r > 0$ suficientemente chico,

$$\left| -e^{\frac{g_j(x)}{r}} \right| \leqslant m_1 r;$$

- si $j \in A(x^*)$, tenemos que $\mu_j^* > 0$ y para todo μ_j en un entorno de μ_j^*,

$$\left| \frac{r}{\mu_j} \right| \leqslant m_2 r.$$

Por lo tanto, $\|\nabla_\mu \Theta\| \leqslant mr$.

(iii) $\nabla_r \Theta(x, y, \mu, r, \alpha) = \begin{pmatrix} 0 & \left[-ln(\frac{y}{\mu}) \right]_{A(x^*)} & \left[\frac{\mu e^{\frac{g(x)}{r}}}{r^2} \right]_J \end{pmatrix},$

- si $j \in J$, para $r > 0$ suficientemente chico,

$$\left| \frac{\mu_j e^{\frac{g_j(x)}{r}} g_j(x)}{r^2} \right| \leqslant m_3 \mu_j.$$

- si $j \in A(x^*)$,

$$\left| -ln(\frac{y_j}{\mu_j}) \right| \leqslant m_4 |y_j - \mu_j|.$$

Por lo tanto,

$$\begin{aligned} \|\nabla_r \Theta\| &\leqslant \tilde{m} \max \left\{ \max_{j \in A(x^*)} \{|y_j(\mu, r, \alpha) - \mu_j|\}, \max_{j \in J} \{\mu_j\} \right\} \\ &\leqslant \tilde{m} \left(\|y - \mu\| + \|\mu - \mu^*\| \right). \end{aligned}$$

Reemplazando las cotas obtenidas en (i)-(iii) en (3.2.14) podemos asegurar que existe $\tilde{M} > 0$ tal que

$$\begin{aligned} \|y(\mu, r, \alpha) - \mu^*\| &\leqslant \tilde{M} \left(r\|\mu - \mu^*\| + r\|y(\mu, r, \alpha) - \mu\| + \|\alpha\| \right) \\ &\leqslant \tilde{M}(2r\|\mu - \mu^*\| + r\|y(\mu, r, \alpha) - \mu^*\| + \|\alpha\|) \end{aligned}$$

y, eligiendo r suficientemente chico tal que $1 - r\tilde{M} > 0$, obtenemos que

$$\|y(\mu, r, \alpha) - \mu^*\| \leqslant \frac{\tilde{M}}{1 - r\tilde{M}} (\|\alpha\| + 2r\|\mu - \mu^*\|).$$

Luego, existe un entorno N de $(\mu^*, r, 0)$ tal que, para todo $(\mu, r, \alpha) \in N$,

$$\|y(\mu, r, \alpha) - \mu^*\| \leqslant M \max \{r\|\mu - \mu^*\|, \|\alpha\|\}$$

como queríamos probar.

Usando las mismas ideas es posible probar $2(b)$. □

Proposición 3.2.15. *Supongamos que las Hipótesis 1 y 2 se cumplen.*

1. *Luego,* $\lim_{k\to\infty} \mu^{k+1} = \mu^*$.

2. *Si, además, suponemos que la Hipótesis 4 también se satisface, entonces* $\mu^k = \bar{\mu}^k$ *para* k *suficientemente grande.*

Demostración. 1. Como x^* es un punto Karush-Kuhn-Tucker, tenemos que

$$\nabla f(x^*) + \nabla g_A(x^*)\mu_A^* = 0,$$

donde $\nabla g_A(x^*) = [\nabla g(x^*)]_{A(x^*)}$.

Por la Hipótesis 2, la matriz $\nabla g_A(x^*)$ tiene rango completo. Luego, obtenemos que

$$\mu_A^* = (\nabla g_A(x^*)^T \nabla g_A(x^*))^{-1}(-\nabla g_A(x^*)^T \nabla f(x^*)). \tag{3.2.16}$$

Usando el criterio de parada del Algoritmo 1 tenemos que

$$\nabla f(x^k) + \nabla g_A(x^k)\mu_A^{k+1} = \delta_k - \sum_{j\notin A(x^*)} \mu_j^{k+1} \nabla g_j(x^k).$$

Usando la Hipótesis 2, consideramos k suficientemente grande de manera que el conjunto de gradientes $\{\nabla g_j(x^k)\}_{j\in A(x^*)}$ sea linealmente independiente. Luego, como hicimos antes, tenemos que

$$\mu_A^{k+1} = (\nabla g_A(x^k)^T \nabla g_A(x^k))^{-1} \left[\nabla g_A(x^k)^T\right] \delta_k - \textstyle\sum_{j\notin A(x^*)} \mu_j^{k+1} \nabla g_j(x^k) \left[-\nabla g_A(x^k)^T \nabla f(x^k)\right].$$

Tomando límites en la última igualdad y usando que $\delta_k \to 0$ y la Proposición 3.1.3 obtenemos que

$$\lim_{k\to\infty} \mu_A^{k+1} = (\nabla g_A(x^*)^T \nabla g_A(x^*))^{-1}(-\nabla g_A(x^*)^T \nabla f(x^*)). \tag{3.2.17}$$

Comparando con (3.2.16) y (3.2.17) tenemos que $\lim_{k\to\infty} \mu_A^{k+1} = \mu_A^*$.

Si $j \notin A(x^*)$, por la Proposición 3.1.3, podemos concluir que

$$\lim_{k\to\infty} \mu_j^{k+1} = 0 = \mu_j^*$$

como queríamos probar.

2. Sea $\varepsilon = \frac{1}{2}\,\text{mín}\{\mu^*, \bar{\mu}^{\text{máx}} - \mu^*\}$, que por la Hipótesis 4, $\varepsilon > 0$. Por el item 1, tenemos que $\mu^{k+1} \to \mu^*$, y por lo tanto existe k_1 tal que $\forall k \geqslant k_1$

$$\|\mu^{k+1} - \mu^*\|_\infty \leqslant \varepsilon$$

Luego, como $|\mu_j^{k+1} - \mu_j^*| \leqslant \varepsilon$, $\forall j = 1, \ldots, p$, por la definición de ε tenemos que

$$\mu_j^{k+1} \in [0, \bar{\mu}_j^{\text{máx}}].$$

Como $\bar{\mu}^{k+1}$ es la proyección de μ^{k+1} en el intervalo $[0, \bar{\mu}^{\text{máx}}]$, implica que $\bar{\mu}^{k+1} = \mu^{k+1}$ para todo k suficientemente grande. □

Corolario 3.2.18. *Supongamos que se satisfacen las Hipótesis 1-3 y además que* $\lim_{k\to\infty} \rho_k = +\infty$. *Definimos* $\alpha^k \in \mathbb{R}^n$ *tal que* $\|\alpha^k\| \leqslant \varepsilon_k$. *Entonces, existe* $M > 0$ *tal que, para todo* $k \in \mathbb{N}$,

$$\|x^k - x^*\| \leqslant M\,\text{máx}\left\{\frac{\|\bar{\mu}^k - \mu^*\|}{\rho_k}, \varepsilon_k\right\}.$$

Y, si además $\mu^{k+1} = \bar{\mu}^k e^{\rho_k g(x^k)}$,

$$\|\mu^{k+1} - \mu^*\| \leqslant M\,\text{máx}\left\{\frac{\|\bar{\mu}^k - \mu^*\|}{\rho_k}, \varepsilon_k\right\}. \tag{3.2.19}$$

Demostración. Consideremos $k_0 \in \mathbb{N}$ tal que $\rho_k \geqslant \bar{\rho}$ para todo $k \geqslant k_0$ y para $\bar{\rho}$ definido en el Lema 3.2.4.

Como la sucesión $\{\bar{\mu}^k\}$ es acotada y $\|\alpha^k\| \to 0$ existe $k_1 \geqslant k_0$ tal que $\|\alpha_k\|$ y $\frac{1}{\rho_k}\|\bar{\mu}^k - \mu^*\|$ son suficientemente chicos para todo $k \geqslant k_1$.

Por el Lema 3.2.12, definimos $(x^k, y^k) = (x(\bar{\mu}^k, \frac{1}{\rho_k}, \alpha_k), y(\bar{\mu}^k, \frac{1}{\rho_k}, \alpha_k))$. Luego,

$$\nabla f(x^k) + \sum_{j=1}^{p} y_j^k \nabla g_j(x^k) = \alpha^k$$

y usando que $\|\alpha^k\| \leqslant \varepsilon_k$ se tiene que

$$\|x^k - x^*\| \leqslant M\,\text{máx}\left\{\frac{\|\bar{\mu}^k - \mu^*\|}{\rho_k}, \varepsilon_k\right\}.$$

Si $y^k = y(\bar{\mu}^k, \frac{1}{\rho_k}, \alpha_k)$, por el item 1 del Lema 3.2.12 tenemos que

$$y^k = \bar{\mu}^k e^{\rho_k g(x^k)} = \mu^{k+1}$$

y

$$\|\mu^{k+1} - \mu^*\| \leqslant M \max\left\{\frac{\|\bar{\mu}^k - \mu^*\|}{\rho_k}, \varepsilon_k\right\}$$

como queríamos probar. □

Teorema 3.2.20. *Sea $\{x^k\}$ la sucesión generada por el Algoritmo 1 definido en la sección 1.2 del Capítulo 1, donde cada x^k verifica las condiciones (3.1.4)-(3.1.5). Supongamos que se satisfacen las Hipótesis 1-4 y que, además, existe una sucesión $\eta_k \to 0$ tal que para todo $k \in \mathbb{N}$*

$$\varepsilon_k \leqslant \eta_k \|\sigma^k\|. \tag{3.2.21}$$

Entonces, la sucesión de parámetros de penalidad $\{\rho_k\}$ es acotada.

Demostración. Supongamos, por contradicción, que $\lim\limits_{k\to\infty} \rho_k = +\infty$.

Consideremos $k_1 \geqslant k_0$ suficientemente grande tal que, por el item 2 de la Proposición 3.2.15, $\bar{\mu}^k = \mu^k$ for all $k \geqslant k_1$.

Por la definición de σ^k y usando el Corolario 3.2.18, para todo $k \geqslant k_1$

$$\begin{aligned}\|\sigma^k\| = \left\|\frac{\mu^{k+1} - \mu^k}{\rho_k}\right\| &\leqslant \left\|\frac{\mu^{k+1} - \mu^*}{\rho_k}\right\| + \left\|\frac{\mu^* - \mu^k}{\rho_k}\right\| \\ &\leqslant \frac{M}{\rho_k} \max\left\{\frac{\|\mu^k - \mu^*\|}{\rho_k}, \varepsilon_k\right\} + \frac{\|\mu^k - \mu^*\|}{\rho_k}.\end{aligned}$$

Consideramos dos posibilidades:

1. Si $\max\left\{\frac{\|\mu^k-\mu^*\|}{\rho_k}, \varepsilon_k\right\} = \frac{\|\mu^k-\mu^*\|}{\rho_k}$ entonces

$$\|\sigma^k\| \leqslant \frac{M}{\rho_k}\frac{\|\mu^k - \mu^*\|}{\rho_k} + \frac{\|\mu^k - \mu^*\|}{\rho_k} = \left(\frac{M}{\rho_k} + 1\right)\frac{\|\mu^k - \mu^*\|}{\rho_k}.$$

2. Si $\max\left\{\frac{\|\mu^k-\mu^*\|}{\rho_k}, \varepsilon_k\right\} = \varepsilon_k$ entonces

$$\|\sigma^k\| \leqslant \frac{M}{\rho_k}\varepsilon_k + \frac{\|\mu^k - \mu^*\|}{\rho_k} \leqslant \frac{M}{\rho_k}\eta_k\|\sigma^k\| + \frac{\|\mu^k - \mu^*\|}{\rho_k},$$

luego

$$\|\sigma^k\| - \frac{M}{\rho_k}\eta_k\|\sigma^k\| \leqslant \frac{\|\mu^k - \mu^*\|}{\rho_k},$$

y por lo tanto,
$$\frac{\rho_k - M\eta}{\rho_k}\|\sigma^k\| \leqslant \frac{\|\mu^k - \mu^*\|}{\rho_k}.$$

Luego, para k suficientemente grande tal que $\rho_k - M\eta_k > 0$, tenemos que

$$\|\sigma^k\| \leqslant \frac{\rho_k}{\rho_k - M\eta_k}\frac{\|\mu^k - \mu^*\|}{\rho_k}.$$

Entonces, para k suficientemente grande tal que $\rho_k - M\eta_k > 0$, de 1. y 2. obtenemos que

$$\|\sigma^k\| \leqslant \frac{\|\mu^k - \mu^*\|}{\rho_k}\,\text{máx}\left\{\frac{M}{\rho_k} + 1, \frac{\rho_k}{\rho_k - M\eta_k}\right\}. \tag{3.2.22}$$

Por otro lado, usando (3.2.19) y (3.2.21) se tiene que

$$\|\mu^k - \mu^*\| \leqslant M\left(\frac{\|\mu^{k-1} - \mu^*\|}{\rho_{k-1}} + \eta_{k-1}\|\sigma^{k-1}\|\right).$$

Luego,
$$\frac{\|\mu^{k-1} - \mu^*\|}{\rho_{k-1}} \geqslant \frac{\|\mu^k - \mu^*\|}{M} - \eta_{k-1}\|\sigma^{k-1}\|. \tag{3.2.23}$$
Por la definición de σ^{k-1}, y usando (3.2.23) se tiene que

$$\begin{aligned}\|\sigma^{k-1}\| = \frac{\|\mu^k - \mu^{k-1}\|}{\rho_{k-1}} &\geqslant \frac{\|\mu^{k-1} - \mu^*\|}{\rho_{k-1}} - \frac{\|\mu^k - \mu^*\|}{\rho_{k-1}} \\ &\geqslant \frac{\|\mu^k - \mu^*\|}{M} - \eta_{k-1}\|\sigma^{k-1}\| - \frac{\|\mu^k - \mu^*\|}{\rho_{k-1}}.\end{aligned}$$

Por lo tanto,

$$(1 + \eta_{k-1})\|\sigma^{k-1}\| \geqslant \|\mu^k - \mu^*\|\left[\frac{1}{M} - \frac{1}{\rho_{k-1}}\right] \geqslant \frac{1}{2M}\|\mu^k - \mu^*\|$$

y tenemos que
$$\|\mu^k - \mu^*\| \leqslant 2M(1 + \eta_{k-1})\|\sigma^{k-1}\|.$$

Entonces, por (3.2.22)

$$\|\sigma^k\| \leqslant \frac{2M(1 + \eta_{k-1})}{\rho_k}\,\text{máx}\left\{\frac{M}{\rho_k} + 1, \frac{\rho_k}{\rho_k - M\eta_k}\right\}\|\sigma^{k-1}\|.$$

Si asumimos que $\rho_k \to +\infty$ y $\eta_k \to 0$, existe k_2 tal que

$$\frac{2M(1+\eta_{k-1})}{\rho_k} \text{máx} \left\{ \frac{M}{\rho_k} + 1, \frac{\rho_k}{\rho_k - M\eta_k} \right\} < \tau$$

para todo $k \geqslant k_2$, y esto muestra que

$$\|\sigma^k\| \leqslant \tau \|\sigma^{k-1}\|, \forall k \geqslant k_2.$$

En consecuencia, $\rho_{k+1} = \rho_k$ para todo $k \geqslant k_2$ contradiciendo la hipótesis. Por lo tanto, probamos que $\{\rho_k\}$ es acotada. □

3.3. Ejemplos ilustrativos

El objetivo de esta sección es ilustrar el comportamiento del método de Lagrangiano Aumentado Exponencial con tres problemas chicos. La idea principal es mostrar que existen problemas espécificos donde la función de penalidad exponencial puede ser más útil que la función de penalidad cuadrática para encontrar minimizadores locales o globales.

Consideramos el algoritmo ALGENCAN del software TANGO en *www.ime.usp.br/ egbirgin/tango/*. ALGENCAN es un código de Fortran para resolver problemas de programación no lineal, desarrollado en el Departamento de Matemática Aplicada de la Universidad de Campinas y el Departamento de Ciencias de la Computación de la Universidad de Sao Paulo, Brasil. El algoritmo general es del tipo Lagrangiano Aumentado usando, esencialmente, el método definido en [2] y en [48], donde se usa la función de penalidad cuadrática.

Hemos realizado modificaciones en algunas subrutinas de ALGENCAN con el objetivo de considerar la función de penalidad exponencial y sus derivadas en lugar de la función cuadrática.

Otras modificaciones fueron:

1. El estimado de los multiplicadores es calculado considerando la fórmula (1.2.4) adaptado para la función de penalidad exponencial:

$$\mu_j^{k+1} = \bar{\mu}_j^k e^{\rho_k g_j(x^k)}.$$

2. El parámetro que mide el progreso en término de la infactibilidad y la complementariedad definido en (1.2.6):

$$\sigma_j^k = \frac{\mu_j^{k+1} - \bar{\mu}_j^k}{\rho_k}.$$

Como hemos mencionado previamente, esta medida para el caso cuadrático es

$$\sigma_j^k = \text{máx}\left\{g_j(x^k), -\frac{\bar{\mu}_j^k}{\rho_k}\right\};$$

y en los experimentos de [19] la medida considerada fue

$$\sigma_j^k = \text{máx}\{g_j(x^k), -\bar{\mu}_j^k\},$$

diferente a la utilizada para obtener los resultados de convergencia citados en [19].

Como definimos en el Paso 3 del Algoritmo 1, el parámetro de penalidad para las restricciones de desigualdad es adaptado de acuerdo al siguiente criterio: Si $\|\sigma^k\| \leqslant \tau\|\sigma^{k-1}\|$ definir $\rho_{k+1} \geqslant \rho_k$, sino definir $\rho_{k+1} = \gamma\rho_k$, donde $\tau < 1$, $\gamma > 1$.

No cambiamos los parámetros de ALGENCAN con respecto al máximo número de iteraciones tanto internas como externas, el criterio de parada, precisión de subproblemas ni de otros parámetros definidos para resolver el subproblema. También se utilizaron los mismos parámetros iniciales μ_0, ρ_0, τ,γ y $\bar{\mu}^{\text{máx}}$. Llamaremos ALExp a la versión de AlGENCAN modificada para el caso exponencial.

Ejemplo 1. Para $x \in \mathbb{R}^2$ consideramos el siguiente problema cuadrático indefinido:

$$\textit{Minimizar } x_1^2 - x_2^2 \qquad \textit{s. a.} \qquad \|x\|^2 \leqslant 1.$$

Usando el punto inicial $(0{,}5, 0)$, ALGENCAN converge al punto silla interior $(0,0)$ en una iteración mientras que ALExp converge al minimizador global $(0,-1)$ en dos iteraciones.

Ejemplo 2. Para $x \in \mathbb{R}^2$ consideramos el siguiente problema (Problema 19 de [59]):

$$\begin{array}{ll} \textit{Minimizar} & x_1^4 - 14x_1^2 + 24x_1 - x_2^2 \\ \textit{s. a.} & -x_1 + x_2 - 8 \leqslant 0 \\ & x_2 - x_1^2 - 2x_1 + 2 \leqslant 0 \\ & -8 \leqslant x_1 \leqslant 10 \\ & 0 \leqslant x_2 \leqslant 11. \end{array}$$

Usando el punto inicial $(0,1)$, ALGENCAN converge al punto silla $(2,0)$ con el valor funcional $f = 8{,}00$ en dos iteraciones mientras que ALExp converge al minimizador local $(2{,}702; 10{,}702)$ con valor funcional $f = -98{,}597$ en 10 iteraciones.

Observamos que un comportamiento similar se obtiene considerando diferentes puntos iniciales, por ejemplo: $(0,a)$ con $a = 2,3,\ldots,7$.

También consideramos el siguiente ejemplo con el objetivo de mostrar la necesidad de considerar información de segundo orden para obtener minimizadores locales en lugar de puntos silla.

Ejemplo 3. Para $x \in \mathbb{R}^2$ consideramos el siguiente problema (Problema 1 de [59], ejemplo 2 de [3]):

$$\begin{array}{lr} \textit{Minimizar} & -x_1 - x_2 \\ \textit{s. a.} & x_1 x_2 \leqslant 4 \\ & 0 \leqslant x_1 \leqslant 6 \\ & 0 \leqslant x_2 \leqslant 4. \end{array}$$

Para este problema observamos que, usando el punto inicial factible $(2, 2)$, ambos métodos llegan a un punto silla.

Capítulo 4

Convergencia global

Consideraremos ahora el problema general de programación no lineal de la forma

$$\textit{Minimizar } f(x) \textit{ sujeto a } c_I(x) \leqslant 0, x \in \Omega \tag{4.0.1}$$

donde las funciones $f : \mathbb{R}^n \to \mathbb{R}, c_I : \mathbb{R}^n \to \mathbb{R}^l$ son dos veces continuamente diferenciables en $\mathbb{R}^n$ y el conjunto factible Ω es de la forma general

$$\Omega = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leqslant 0\}$$

donde las funciones $h : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p$ son dos veces continuamente diferenciables en $\mathbb{R}^n$. El conjunto factible Ω esta definido de modo que las restricciones pueden ser más fáciles de manipular, por ejemplo, restricciones de cota o lineales.

Dado x^* un punto factible del problema (4.0.1), definimos el conjunto factible como: $A_{c_I}(x^*) = \{j \in \{1, \ldots, l\} : [c_I(x^*)]_j = 0\}$ y $A(x^*) = \{j \in \{1, \ldots, p\} : g_j(x^*) = 0\}$.

En este Capítulo penalizaremos, usando funciones de penalidad no cuadráticas, las restricciones de desigualdad $c_I(x) \leqslant 0$, manteniendo el conjunto factible Ω en el subproblema.

De esta manera, el Algoritmo de Lagrangiano Aumentado utilizando las funciones de penalidad no cuadráticas generales P_1, P_2, P_3 definidas en la sección 1.2 del Capítulo 1 es el siguiente:

Algoritmo 2. Sea $x^0 \in \mathbb{R}^n$ un punto inicial arbitrario. Los parámetros de penalidad para la ejecución del Algoritmo son los siguientes:

$$\tau \in [0, 1), \gamma > 1,$$

$$0 < \bar{\mu}_j^{left(} < \bar{\mu}_j^{\text{máx}} < \infty, \ \forall\, j = 1, \ldots, l,$$

$$\rho_0 \in \mathbb{R}_{++},$$
$$\bar{\mu}_j^0, \bar{\mu}_j^1 \in [\bar{\mu}_j^{\text{mín}}, \bar{\mu}_j^{\text{máx}}],\ \forall\, j = 1, \ldots, l,$$
$$\sigma_j^0 = \frac{\bar{\mu}_j^1 - \bar{\mu}_j^0}{\rho_0}, \quad \forall j = 1, \ldots, l.$$

Inicio $k \leftarrow 1$.

Paso 1. Resolver el subproblema

Calcular (de ser posible) $x^k \in \Omega$ como la solución aproximada del subproblema

$$\textit{Minimizar } L(x, \bar{\mu}^k, \rho_k) \textit{ sujeto a } \ x \in \Omega. \tag{4.0.2}$$

Si no es posible, para la ejecución del Algoritmo.

Paso 2. Estimar nuevos multiplicadores y definir una nueva medida de infactibilidad y complementariedad.

Para $j = 1, \ldots, l$ calcular

$$\mu_j^{k+1} = \frac{\partial P}{\partial y}([c_I(x^k)]_j, \rho_k, \bar{\mu}_j^k) \tag{4.0.3}$$

$$\bar{\mu}_j^{k+1} = \text{máx}\{\text{mín}\{\mu_j^{k+1}, \bar{\mu}_j^{\text{máx}}\}, \bar{\mu}_j^{\text{mín}}\} \tag{4.0.4}$$

$$\sigma_j^k = \frac{\mu_j^{k+1} - \bar{\mu}_j^k}{\rho_k} \tag{4.0.5}$$

Paso 3. Adaptar el parámetro de penalidad.

Si $k = 1$ o

$$\|\sigma^k\|_\infty \leqslant \tau \|\sigma^{k-1}\|_\infty \tag{4.0.6}$$

elegir

$$\rho_{k+1} \geqslant \rho_k. \tag{4.0.7}$$

Sino, definir

$$\rho_{k+1} = \gamma \rho_k. \tag{4.0.8}$$

Paso 4. Comenzar una nueva iteración. Definir $k \leftarrow k + 1$ e ir al Paso 1.

4.1. Convergencia global de primer orden para el problema general

En esta sección asumiremos que la solución x^k del subproblema (4.0.2) generada por el Algoritmo 2 satisface, aproximadamente, la condición de optimalidad de primer orden en el siguiente sentido: Dada una sucesión de parámetros de tolerancia $\{\varepsilon_k\} \subset \mathbb{R}_+$ tal que $\lim\limits_{k\to\infty} \varepsilon_k = 0$, suponemos que $\{x^k\}$ es la sucesión generada por el Algoritmo 2 para el cual existen vectores $v^k \in \mathbb{R}^m, u^k \in \mathbb{R}^p$ satisfaciendo

$$\left\| \nabla_x L(x^k, \bar{\mu}^k, \rho_k) + \sum_{i=1}^{m} v_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} u_j^k \nabla g_j(x^k) \right\| \leq \varepsilon_k, \tag{4.1.1}$$

$$\|h(x^k)\| \leqslant \varepsilon_k, \tag{4.1.2}$$

$$u_j^k \geqslant 0 \text{ y } g_j(x^k) \leqslant \varepsilon_k \text{ para todo } j = 1, \dots, p, \tag{4.1.3}$$

$$u_j^k = 0 \text{ cuando } g_j(x^k) < -\varepsilon_k, \ \text{ para todo } j = 1, \dots, p, \tag{4.1.4}$$

para cada $k \in \mathbb{N}$.

Supongamos también que la función real φ es diferenciable y verifica las condiciones **C2** y **C3** definidas en la Sección 1.2 del Capítulo 1.

Las condiciones (4.1.1)-(4.1.4) fueron previamente definidas en [2] como criterio de parada de primer orden. Estas condiciones miden el grado de factibilidad y optimalidad de la solución aproximada del subproblema (4.0.6).

Si bien en el libro recientemente publicado, [48], se pide que la solución aproximada x^k debe satisfacer (4.1.1) y que además debe existir u^k tal que $u^k \geqslant 0$ y

$$\|h(x^k)\| \leqslant \varepsilon_k, \quad \|\min\{u^k, -g(x^k)\}\| \leqslant \varepsilon_k, \tag{4.1.5}$$

nosotros en el siguiente lema probaremos que las condiciones (4.1.2)-(4.1.4) implican la condición (4.1.5) mostrando así que (4.1.1)-(4.1.4) son condiciones secuenciales aproximadas mas débiles que las condiciones propuestas en [48].

Lema 4.1.6. *Sea $\{x^k\}$ la sucesión generada por el Algoritmo 2 y supongamos que $\{x^k\}$ converge a x^*. Luego, si x^* es un punto factible, para cada $j = 1, \dots, l$ se tiene que*

$$\lim_{k\to\infty} \min\{\mu_j^{k+1}, -[c_I(x^k)]_j\} = 0; \tag{4.1.7}$$

y, para cada $j = 1, \dots, p$ se tiene que

$$\lim_{k\to\infty} \min\{u_j^k, -g_j(x^k)\} = 0. \tag{4.1.8}$$

Demostración. Probaremos primero la condición (4.1.7). Sea $j \in A_{c_I}(x^*)$. Como $\min\{\mu_j^{k+1}, -[c_I(x^k)]_j\} \leqslant -[c_I(x^k)]_j$ y $\lim\limits_{k\to\infty} -[c_I(x^k)]_j = 0$ tenemos que la condición (4.1.7) se satisface.

Consideremos ahora el conjunto de índices de las restricciones inactivas, $j \notin A_{c_I}(x^*)$. Para cada función de penalidad P_1, P_2 y P_3, tenemos que

- $\mu_j^{k+1} = \bar{\mu}_j^k \varphi'\left(\rho_k [c_I(x^k)]_j\right)$ para $P_1(y,\mu,\rho) = \frac{\mu}{\rho}\varphi(\rho y)$,

- $\mu_j^{k+1} = \bar{\mu}_j^k \varphi'\left(\rho_k \bar{\mu}_j^k [c_I(x^k)]_j\right)$ para $P_2(y,\mu,\rho) = \frac{1}{\rho}\varphi(\rho\mu y)$,

- $\mu_j^{k+1} = \bar{\mu}_j^k \varphi'\left(\dfrac{\rho_k [c_I(x^k)]_j}{\bar{\mu}_j^k}\right)$ para $P_3(y,\mu,\rho) = \frac{\mu^2}{\rho}\varphi(\frac{\rho y}{\mu})$.

Para este caso probaremos que $\lim\limits_{k\to\infty} \mu_j^{k+1} = 0$ y para hacer esto contemplaremos dos situaciones:

(a) La sucesión de parámetros de penalidad $\{\rho_k\}$ es acotada;

(b) La sucesión de parámetros de penalidad $\{\rho_k\}$ no es acotada.

Para el caso (a) mostraremos que $\lim\limits_{k\to\infty} \bar{\mu}_j^k = 0$. Por el Paso 3 del Algoritmo 2, tenemos que $\lim\limits_{k\to\infty} \sigma_j^k = 0$. Luego, reemplazando (4.0.3) en (4.0.5) para cada función de penalidad P_i se tiene que

- $\sigma_j^k = \dfrac{\bar{\mu}_j^k}{\rho_k}\left(\varphi'(\rho_k [c_I(x^k)]_j) - 1\right)$,

- $\sigma_j^k = \dfrac{\bar{\mu}_j^k}{\rho_k}\left(\varphi'(\rho_k \bar{\mu}_j^k [c_I(x^k)]_j) - 1\right)$,

- $\sigma_j^k = \dfrac{\bar{\mu}_j^k}{\rho_k}\left(\varphi'\left(\dfrac{\rho_k [c_I(x^k)]_j}{\bar{\mu}_j^k}\right) - 1\right)$

respectivamente.

Como φ verifica la condición **C2**, es decir, $\varphi'(t) \neq 1$ para todo $t < 0$, se tiene que $\lim\limits_{t\to\infty} \bar{\mu}_j^k = 0$ y por la definición de μ_j^{k+1}, se tiene que $\lim\limits_{k\to\infty} \mu_j^{k+1} = 0$.

Para el caso (b), analizaremos cada caso por separado.

P_1: En este caso tenemos que $\mu_j^{k+1} = \bar{\mu}_j^k \varphi'\left(\rho_k[c_I(x^k)]_j\right)$. Como $[c_I(x^*)]_j < 0$, se tiene que $\lim\limits_{k\to\infty} \rho_k[c_I(x^k)]_j = -\infty$. Y como además $\{\bar{\mu}_j^k\}_{k\in\mathbb{N}}$ es una sucesión acotada y se satisface el axioma **C3**, $\lim\limits_{t\to-\infty} \varphi'(t) = 0$, podemos concluir que $\lim\limits_{k\in K} \mu_j^{k+1} = 0$.

P_2: Para esta función tenemos dos casos:

- Si $\lim\limits_{k\to\infty} \rho_k\bar{\mu}_j^k[c_I(x^k)]_j = -\infty$, entonces utilizando el axioma **C3** se tiene que $\lim\limits_{k\to\infty} \mu^{k+1} = 0$.
- Sino, como la sucesión $\{\bar{\mu}^k\}_k$ es acotada, $\lim\limits_{k\to\infty} \bar{\mu}^k = 0$. Y por lo tanto $\lim\limits_{k\to\infty} \mu^{k+1} = 0$.

P_3: Para este caso, $\lim\limits_{k\to\infty} \dfrac{\rho_k[c_I(x^k)]_j}{\bar{\mu}_j^k} = -\infty$. Y como estamos suponiendo que **C3** se satisface, se tiene que $\lim\limits_{k\to\infty} \mu_j^{k+1} = 0$.

Finalmente, probaremos que la condición (4.1.8) se cumple para cada función P_i.

Si $j \in A(x^*)$, para k suficientemente grande $|g_j(x^k)| \leqslant \varepsilon_k$. Luego,

$$-g_j(x^k) \leqslant \varepsilon_k \quad \text{y} \quad \min\{u_j^k, -g_j(x^k)\} \leq -g_j(x^k) \leqslant \varepsilon_k.$$

Si $j \notin A(x^*)$ entonces $g_j(x^k) < -\varepsilon_k$ para k suficientemente grande y usando (4.1.4) tenemos que $u_j^k = 0$. Luego,

$$\min\{u_j^k, -g_j(x^k)\} = 0 \leqslant \varepsilon_k.$$

Y por lo tanto, (4.1.8) se cumple. □

Observar que, como consecuencia de este lema, los estimados de los multiplicadores de Lagrange correspondientes a las restricciones de desigualdad inactivas son nulos en el límite, independientemente de la función de penalidad P_i que se utilice.

En el siguiente teorema probaremos que cuando el Algoritmo 2 admite un punto factible, este punto satisface la condición de optimalidad secuencial de primer orden AKKT de la Definición 2.1.1 asociada al problema original (4.0.1), es decir, que existen sucesiones $\{x^k\} \subseteq \mathbb{R}^n$, $\{\mu^k\} \subseteq \mathbb{R}^l_+$, $\{u^k\} \subseteq \mathbb{R}^p_+$ y $\{v^k\} \subseteq \mathbb{R}^m$ tales que $\lim\limits_{k\to\infty} x^k = x^*$,

$$\lim_{k\to\infty} \left\| \nabla f(x^k) + \sum_{j=1}^{l} \mu_j^{k+1} \nabla [c_I(x^k)]_j + \sum_{i=1}^{m} v_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} u_j^k \nabla g_j(x^k) \right\| = 0, \quad (4.1.9)$$

$$\lim_{k\to\infty} \min\{\mu_j^{k+1}, -[c_I(x^k)]_j\} = 0, \quad (4.1.10)$$

y, para cada $j = 1, \ldots, p$,

$$\lim_{k\to\infty} \min\{u_j^k, -g_j(x^k)\} = 0. \quad (4.1.11)$$

Teorema 4.1.12. *Supongamos que la sucesión $\{x^k\}$ es generada por el Algoritmo 2 satisfaciendo, en el Paso 1, la condición inexacta de primer orden*

$$\left\| \nabla_x L(x^k, \bar{\mu}^k, \rho_k) + \sum_{i=1}^{m} v_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} u_j^k \nabla g_j(x^k) \right\| \leq \varepsilon_k, \quad (4.1.13)$$

$$\|h(x^k)\| \leqslant \varepsilon_k, \quad (4.1.14)$$

$$u_j^k \geqslant 0 \text{ y } g_j(x^k) \leqslant \varepsilon_k \text{ para todo } j = 1, \ldots, p, \quad (4.1.15)$$

$$u_j^k = 0 \text{ cuando } g_j(x^k) < -\varepsilon_k,\ j = 1, \ldots, p, \quad (4.1.16)$$

para todo $k \in \mathbb{N}$. Supongamos que $K \subset \mathbb{N}$ es tal que $\lim\limits_{k\in K} x^k = x^*$ y que x^* es factible. Más aún, si la sucesión $\{\varepsilon_k\}$ es tal que $\lim\limits_{k\in K} \varepsilon_k = 0$. Entonces x^* satisface la condición AKKT de la Definición 2.1.1 con respecto al problema (4.0.1).

Demostración. Para todo $k \in K$, por (4.0.3) y (4.1.13) tenemos que

$$\left\| \nabla f(x^k) + \sum_{j=1}^{l} \mu_j^{k+1} \nabla [c_I(x^k)]_j + \sum_{i=1}^{m} v_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} u_j^k \nabla g_j(x^k) \right\| \leq \varepsilon_k.$$

Como $\{\varepsilon_k\} \subseteq \mathbb{R}_+$ es una sucesión tal que $\lim\limits_{k\in K} \varepsilon_k = 0$ tenemos que existen sucesiones $\{\mu^{k+1}\}_{k\in K} \subseteq \mathbb{R}^l_+, \{v^k\}_{k\in K} \subseteq \mathbb{R}^m$ y $\{u^k\}_{k\in K} \subseteq \mathbb{R}^p_+$ tales que

$$\lim_{k\in K} \left\| \nabla f(x^k) + \sum_{j=1}^{l} \mu_j^{k+1} \nabla [c_I(x^k)]_j + \sum_{i=1}^{m} v_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} u_j^k \nabla g_j(x^k) \right\| = 0$$

y (4.1.9) se cumple.

Por el Lema 4.1.6 tenemos que las ecuaciones (4.1.10) y (4.1.11) se satisfacen para las restricciones de desigualdad $c_I(x) \leqslant 0$ y $g(x) \leqslant 0$.

Por lo tanto x^* satisface la condición KKT Aproximada con respecto al problema (4.0.1) como queriamos probar. $\square$

De los Teoremas 2.1.14 y 4.1.12 podemos concluir que:

Teorema 4.1.17. *Supongamos que la sucesión $\{x^k\}$ es generada por el Algoritmo 2 satisfaciendo, en el Paso 1, la condición de primer orden inexacta*

$$\left\| \nabla_x L(x^k, \bar{\mu}^k, \rho_k) + \sum_{i=1}^{m} v_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} u_j^k \nabla g_j(x^k) \right\| \leq \varepsilon_k,$$

$$\|h(x^k)\| \leqslant \varepsilon_k,$$

$$u_j^k \geqslant 0 \text{ y } g_j(x^k) \leqslant \varepsilon_k \text{ para todo } j = 1, \ldots, p,$$

$$g_j(x^k) < -\varepsilon_k \text{ luego } u_j^k = 0,\ j = 1, \ldots, p,$$

para todo $k \in \mathbb{N}$. Supongamos que $K \subset \mathbb{N}$ es tal que $\lim_{k \in K} x^k = x^*$ y x^* es factible y satisface la condición CCP. Entonces x^* es un punto KKT del problema (4.0.1).

4.2. Convergencia de Segundo Orden

En esta sección estudiaremos la convergencia del Algoritmo 2 a puntos estacionarios de segundo orden. Para ello, estamos interesados en encontrar condiciones para la solución aproximada del subproblema (4.0.2) bajo las cuales el Algoritmo 2 es capaz de encontrar puntos que satisfagan las condiciones de optimalidad necesarias tanto de primer, como de segundo orden del problema (4.0.1). Estos son puntos factibles para las cuales existen multiplicadores $\mu^* \in \mathbb{R}^l_+$, $v^* \in \mathbb{R}^m$ y $u^* \in \mathbb{R}^p_+$ satisfaciendo las condiciones KKT:

$$\nabla f(x^*) + \sum_{j=1}^{l} \mu_j^* \nabla [c_I(x^*)]_j + \sum_{i=1}^{m} v_i^* \nabla h_i(x^*) + \sum_{j=1}^{p} u_j^* \nabla g_j(x^*) = 0, \tag{4.2.1}$$

$$\mu_j^* [c_I(x^*)]_j = 0, j = 1, \ldots, l,$$

$$u_j^* g_j(x^*) = 0, j = 1, \ldots, p$$

y la condición necesaria débil de segundo orden (WSONC):

$$\nabla^2 f(x^*) + \sum_{j=1}^{l} \mu_j^* \nabla^2 [c_I(x^*)]_j + \sum_{i=1}^{m} v_i^* \nabla^2 h_i(x^*) + \sum_{j=1}^{p} u_j^* \nabla^2 g_j(x^*) \succeq 0 \text{ en } T_1(x^*)$$

donde el cono $T_1(x^*)$ es el subespacio tangente de las restricciones de desigualdad activas asociadas al problema general (4.0.1):

$$\begin{aligned} T_1(x^*) = \{d \in \mathbb{R}^n | \quad & \nabla h_i(x^*)^T d = 0 \text{ para } i = 1, \dots, m; \\ & \nabla [c_I(x^*)]_j^T d = 0 \text{ para } j \text{ tal que } j \in A_{c_I}(x^*); \\ & \nabla g_j(x^*)^T d = 0 \text{ para } j \text{ tal que } j \in A(x^*)\}. \end{aligned}$$

Para obtener este tipo de puntos y siguiendo la idea de [6] podemos suponer que la solución x^k del subproblema (4.0.2) satisface, aproximadamente, una condición de optimalidad secuencial de segundo orden. Observar que, para cada $k \in \mathbb{N}$, la función de Lagrange del subproblema (4.0.2) es una combinación de $L(x, \bar{\mu}^k, \rho_k)$ y las funciones en $\bar{\Omega} : h_i(x), i = 1, \dots, m$ y $g_j(x), j = 1, \dots, p$. Dada $\{\varepsilon_k\} \subseteq \mathbb{R}_+$ una sucesión de parámetros de tolerancia tal que $\lim\limits_{k\to\infty} \varepsilon_k = 0$, suponemos que $\{x^k\}$ es la sucesión generada por el Algoritmo 2 para la cual existen vectores $v^k \in \mathbb{R}^m, u^k \in \mathbb{R}^p$ satisfaciendo la condición secuencial de primer orden (4.1.1)-(4.1.4) y, además, existen vectores $\eta^k \in \mathbb{R}^m, \theta^k \in \mathbb{R}^p$ satisfaciendo

$$\nabla^2_{xx} L(x^k, \bar{\mu}^k, \rho_k) + \sum_{i=1}^m v_i^k \nabla^2 h_i(x^k) + \sum_{j=1}^p u_j^k \nabla^2 g_j(x^k) + \tag{4.2.2}$$

$$+ \sum_{i=1}^m \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j=1}^p \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succeq -\varepsilon_k Id \text{ en } \mathbb{R}^n$$

y

$$\text{si } g_j(x^k) < -\varepsilon_k \text{ entonces } \theta_j^k = 0, \ j = 1, \dots, p, \tag{4.2.3}$$

para cada $k \in \mathbb{N}$.

Observar que si x^* es un punto límite factible de la sucesión $\{x^k\}$ que verifica (4.2.2)-(4.2.3) entonces

$$\theta_j^k = 0 \text{ para } j \notin A(x^*)$$

y por lo tanto la condición (4.2.3) se satisface.

En esta sección asumimos que la función real φ verifica las condiciones **C1-C4** definidas en la Sección 1.2 del Capítulo 1.

El siguiente lema es necesario para probar el resultado de convergencia de segundo orden.

Lema 4.2.4. *Sea $\{x^k\}$ una sucesión generada por el Algoritmo 2 y supongamos que $\{x^k\}$ converge a x^*. Entonces, si x^* es factible, para todo $j \notin A_{c_I}(x^*)$*

$$\lim_{k\to\infty} \frac{\partial^2 P}{\partial y^2}([c_I(x^k)]_j, \rho_k, \bar{\mu}_j^k) = 0. \tag{4.2.5}$$

Demostración. Para cada función de penalidad P_i tenemos que

- $\dfrac{\partial^2 P_1}{\partial y^2}([c_I(x^k)]_j, \rho_k, \bar{\mu}_j^k) = \bar{\mu}_j^k \rho_k \varphi''(\rho_k [c_I(x^k)]_j)$, para $P_1(y, \mu, \rho) = \frac{\mu}{\rho}\varphi(\rho y)$

- $\dfrac{\partial^2 P_2}{\partial y^2}([c_I(x^k)]_j, \rho_k, \bar{\mu}_j^k) = (\bar{\mu}_j^k)^2 \rho_k \varphi''(\rho_k \bar{\mu}_j^k [c_I(x^k)]_j)$, para $P_2(y, \mu, \rho) = \frac{1}{\rho}\varphi(\rho\mu y)$

- $\dfrac{\partial^2 P_3}{\partial y^2}([c_I(x^k)]_j, \rho_k, \bar{\mu}_j^k) = \rho_k \varphi''\left(\dfrac{\rho_k [c_I(x^k)]_j}{\bar{\mu}_j^k}\right)$, para $P_3(y, \mu, \rho) = \frac{\mu^2}{\rho}\varphi(\frac{\rho y}{\mu})$.

Contemplaremos dos situaciones:

(a) $\{\rho_k\}$ es una sucesión acotada;

(b) $\{\rho_k\}$ no es acotada.

Para el caso (a), como ya probamos en el Lema 4.1.6, podemos establecer que $\lim_{k\to\infty} \bar{\mu}_j^k = 0$ y esto implica que (4.2.5) se satisface.

En el caso (b), si consideramos cada función P_i:

P_1**:** A partir de que $j \notin A_{c_I}(x^*)$, para k suficientemente grande, se tiene que $c_I(x^k) < 0$, luego

$$\bar{\mu}_j^k \rho_k \varphi''(\rho_k [c_I(x^k)]_j) = \frac{\bar{\mu}_j^k}{[c_I(x^k)]_j} \rho_k [c_I(x^k)]_j \varphi''(\rho_k [c_I(x^k)]_j).$$

Como la sucesión $\{\bar{\mu}^k\}$ es acotada, y $\rho_k [c_I(x^k)]_j \to -\infty$, utilizando el axioma **C4**, se tiene que la condición (4.2.5) se satisface.

P_2**:** Para las funciones de penalidad del tipo P_2, analizaremos dos casos:

Si $\lim_{k\to\infty} \rho_k \bar{\mu}_j^k [c_I(x^k)]_j = -\infty$, entonces utilizando el axioma **C4** se tiene que inmediatamente que $\lim_{k\to\infty} \dfrac{\partial^2 P}{\partial y^2}([c_I(x^k)]_j, \rho_k, \bar{\mu}_j^k) = 0$.

Sino, la sucesión $\lim_{k\to\infty} \bar{\mu}^k = 0$. Por lo tanto,

$$\lim_{k\to\infty} \frac{\partial^2 P}{\partial y^2}([c_I(x^k)]_j, \rho_k, \bar{\mu}_j^k) = (\bar{\mu}_j^k)^2 \rho_k \varphi''(\rho_k \bar{\mu}_j^k [c_I(x^k)]_j) = 0.$$

P_3**:** Como $j \notin A_{c_I}(x^*)$, para k suficientemente grade, se tiene que $[c_I(x^k)]_j < 0$, luego

$$\rho_k \varphi'' \left(\frac{\rho_k [c_I(x^k)]_j}{\bar{\mu}_j^k} \right) = \frac{\bar{\mu}_j^k}{[c_I(x^k)]_j} \frac{\rho_k [c_I(x^k)]_j}{\bar{\mu}_j^k} \varphi'' \left(\frac{\rho_k [c_I(x^k)]_j}{\bar{\mu}_j^k} \right).$$

Como sabemos que la sucesión $\{\bar{\mu}^k\}$ es acotada, se tiene que $\rho_k \bar{\mu}_j^k [c_I(x^k)]_j \to -\infty$. Utilizando el axioma **C4**, se tiene que la condición (4.2.5) se satisface.

De esta manera hemos demostrado que la condición (4.2.5) se satisface para cada función P_i. □

En el siguiente teorema probaremos que cuando el Algoritmo 2 admite un punto factible, este punto satisface la condición de optimalidad secuencial de segundo orden rAKKT2 asociada a todas las restricciones del problema original (4.0.1).

Teorema 4.2.6. *Supongamos que la sucesión $\{x^k\}$ es generada por el Algoritmo 2 satisfaciendo,en el Paso 1, la condición inexacta de primer orden (4.1.1)-(4.1.4) y la condición inexacta de segundo orden (4.2.2)-(4.2.3) para todo $k \in \mathbb{N}$. Supongamos que $K \subset \mathbb{N}$ es tal que $\lim_{k \in K} x^k = x^*$ y x^* es factible. Más aún, supongamos que la sucesión $\{\varepsilon_k\}$ es tal que $\lim_{k \in K} \varepsilon_k = 0$. Entonces, x^* satisface la condición rAKKT2 con respecto al problema (4.0.1).*

Demostración. Para todo $k \in \mathbb{N}$, por (4.0.3), (4.1.1) y (4.2.2) tenemos que

$$\left\| \nabla f(x^k) + \sum_{j=1}^{l} \mu_j^{k+1} \nabla [c_I(x^k)]_j + \sum_{i=1}^{m} v_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} u_j^k \nabla g_j(x^k) \right\| \leq \varepsilon_k,$$

y

$$\nabla_{xx}^2 L(x^k, \bar{\mu}^k, \rho_k) + \sum_{i=1}^{m} v_i^k \nabla^2 h_i(x^k) + \sum_{j=1}^{p} u_j^k \nabla^2 g_j(x^k) + \tag{4.2.7}$$

$$+ \sum_{i=1}^{m} \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j=1}^{p} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succeq -\varepsilon_k Id \text{ en } \mathbb{R}^n$$

Como $\{\varepsilon_k\} \subseteq \mathbb{R}_+$ es tal que $\lim_{k \in K} \varepsilon_k = 0$, existen sucesiones

$\{\mu^{k+1}\}_{k \in K} \subseteq \mathbb{R}_+^l$, $\{v^k\}_{k \in K} \subseteq \mathbb{R}^m$ y $\{u^k\}_{k \in K} \subseteq \mathbb{R}_+^p$ tales que

$$\lim_{k \in K} \left\| \nabla f(x^k) + \sum_{j=1}^{l} \mu_j^{k+1} \nabla [c_I(x^k)]_j + \sum_{i=1}^{m} v_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} u_j^k \nabla g_j(x^k) \right\| = 0,$$

y, por (4.2.7)

$$\nabla^2 f(x^k) + \sum_{j=1}^{l} \mu_j^{k+1} \nabla^2 [c_I(x^k)]_j + \sum_{i=1}^{m} v_i^k \nabla^2 h_i(x^k) + \sum_{j=1}^{p} u_j^k \nabla^2 g_j(x^k) +$$

$$\sum_{j=1}^{l} \frac{\partial^2 P}{\partial y^2}([c_I(x^k)]_j, \rho_k, \bar{\mu}_j^k) \nabla [c_I(x^k)]_j \nabla [c_I(x^k)]_j^T +$$

$$\sum_{i=1}^{m} \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j=1}^{p} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succeq -\varepsilon_k Id \text{ en } \mathbb{R}^n.$$

Consideremos $d \in \mathbb{R}^n$ y

$$\delta^k = \varepsilon_k + \sum_{j \notin A_{c_I}(x^*)} \frac{\partial^2 P}{\partial y^2}([c_I(x^k)]_j, \rho_k, \bar{\mu}_j^k) \|\nabla [c_I(x^k)]_j^T\|^2 + \sum_{j \notin A(x^*)} \theta_j^k \|\nabla g_j(x^k)\|^2.$$

Luego, obtenemos que

$$d^T \left(\nabla^2 f(x^k) + \sum_{j=1}^{l} \mu_j^{k+1} \nabla^2 [c_I(x^k)]_j + \sum_{i=1}^{m} v_i^k \nabla^2 h_i(x^k) + \sum_{j=1}^{p} u_j^k \nabla^2 g_j(x^k) \right) d +$$

$$\sum_{j \in A_{c_I}(x^*)} \frac{\partial^2 P}{\partial y^2}([c_I(x^k)]_j, \rho_k, \bar{\mu}_j^k) \|\nabla [c_I(x^k)]_j^T d\|^2 +$$

$$\sum_{i=1}^{m} \eta_i^k \|\nabla h_i(x^k)^T d\|^2 + \sum_{j \in A(x^*)} \theta_j^k \|\nabla g_j(x^k) d^T\| \geqslant -\delta_k \|d\|^2.$$

Como vale (4.2.3) tenemos que, para k suficientemente grande,

$$\theta_j^k = 0 \text{ para } j \notin A(x^*).$$

Luego, por la hipótesis acerca de ε_k y el Lema 4.2.4 se tiene que

$$\lim_{k \in K} \delta_k = 0.$$

Por lo tanto, considerando la familia de sucesiones $\{v^k\} \subseteq \mathbb{R}^m$, $\{\eta^k\} \subseteq \mathbb{R}^m$ $\{\tilde{\mu}^k\} \subseteq \mathbb{R}_+^{l+p}$, $\{\tilde{\theta}^k\} \subseteq \mathbb{R}^{l+p}$, $\{\delta^k\} \subseteq \mathbb{R}_+$ de la Definición 2.2.12 (con respecto al problema (4.0.1)) definidos de la siguiente manera:

$$\tilde{\mu}_j^k = \begin{cases} \mu_j^{k+1} & \text{para } j \in A_{c_I}(x^*) \\ 0 & \text{para } j \notin A_{c_I}(x^*) \\ u_j^k & \text{para } j \in A(x^*) \\ 0 & \text{para } j \notin A(x^*) \end{cases}$$

$$\tilde{\theta}_j^k = \begin{cases} \frac{\partial^2 P}{\partial y^2}([c_I(x^k)]_j, \rho_k, \bar{\mu}_j^k) & \text{para } j \in A_{c_I}(x^*) \\ 0 & \text{para } j \notin A_{c_I}(x^*) \\ \theta_j^k & \text{para } j \in A(x^*) \\ 0 & \text{para } j \notin A(x^*) \end{cases}$$

obtenemos que x^* satisface la condición rAKKT2 con respecto al problema original como queríamos probar. $\square$

De los Teoremas 2.2.27 y 4.2.6 tenemos la siguiente conclusión.

Teorema 4.2.8. *Supongamos que la sucesión* $\{x^k\}$ *generada por el Algoritmo 2 satisface, en el Paso 1, la condición inexacta de primer orden (4.1.1)-(4.1.4) y la condición inexacta de segundo orden (4.2.2)-(4.2.3) para todo* $k \in \mathbb{N}$. *Supongamos que* $K \subset \mathbb{N}$ *es tal que* $\lim_{k \in K} x^k = x^*$ *y que* x^* *es factible y satisface la condición CCP2. Luego, la condición WSONC se cumple en* x^* *para el problema original (4.0.1).*

Capítulo 5

Conclusiones

En primera instancia, propusimos un algoritmo de Lagrangiano Aumentado basado en la función de penalidad exponencial para problemas de optimización con restricciones de desigualdad. Una característica potencial de este algoritmo es que posee convergencia global a puntos estacionarios de primer orden bajo las condiciones de calidad Generador Positivo Constante (CPG) y la Propiedad del Cono Continuo (CCP). Probamos que el algoritmo propuesto garantiza que el parámetro de penalidad es acotado cuando se satisface la condición de calidad de Independencia Lineal (LICQ), la complementariedad estricta y la condición de optimalidad suficiente de segundo orden.

Observando que la mayor ventaja de la función de penalidad exponencial es que tiene segundas derivadas continuas, a diferencia de la conocida función de penalidad cuadrática, concluimos que tomar información de segundo orden podría ser importante para definir un algoritmo con convergencia a puntos que verifican las condiciones de optimización necesarias de segundo orden. Este nuevo marco podría aplicarse para resolver problemas en los que los algoritmos de primer orden no convergen en minimizadores locales.

Con este argumento, en este trabajo hemos extendido el análisis de convergencia de un método Lagrangiano Aumentado Inexacto al caso en que se consideran las funciones de penalidad no cuadráticas, incluyendo la función de penalidad exponencial. Hemos demostrado que, como ocurre en el caso cuadrático, la convergencia a puntos estacionarios de primer orden se puede obtener usando la condición de calidad débil CCP. También hemos demostrado que la condición de calidad CCP2 puede ser utilizada para demostrar la convergencia global a puntos que satisfacen la condición de optimalidad necesaria de segundo orden débil WSONC. Finalmente, hemos establecido un nuevo criterio de parada del algoritmo bajo el cual se puede obtener la convergencia global. La nueva condición utiliza un subespacio tangente aproximado apropiado para resolver el subproblema en una forma inexacta.

Bibliografía

[1] Andreani, R., Birgin E. G., Martínez J. M. and Schuverdt, M. L. , Augmented Lagrangian methods under the Constant Positive Linear Dependence constraint qualification. *Mathematical Programming* 111, pp. 5-32, 2008.

[2] Andreani, R., Birgin, E. G., Martínez, J. M., and Schuverdt, M. L., On augmented Lagrangian methods with general lower-level constraints. *SIAM Jornal Optimization* 18(4), pp. 1286-1309, 2007.

[3] Andreani, R., Birgin, E.G., Martínez, J.M. and Schuverdt, M.L., Second-order negative-curvature methods for box-constrained and general constrained optimization. *Computational Optimization and Applications* 45, pp. 209-236, 2010.

[4] Andreani, R., Echagüe, C.E. and Schuverdt, M.L., Constant-rank condition and second-order constraint qualification. *Journal of Optimization Theory and Applications* 146, pp. 155-266, 2009.

[5] Andreani, R., Haeser, G. and Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. *Optimization* 60(5), pp. 627-641, 2011.

[6] Andreani, R., Haeser, G., Ramos, A. and Silva, P.J.S., On second-order sequential optimality conditions for nonlinear optimization and applications. Submitted, 2015.

[7] Andreani, R., Haeser, G., Schuverdt, M. L. and Silva, P. J. S., A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programing* 135 , no. 1-2, Ser. A, pp. 255-273, 2012.

[8] Andreani, R., Haeser, G., Schuverdt, M. L., and Silva, P. J. S., Two new weak constraint qualifications and applications. *SIAM Jornal Optimization* 22(3), pp. 1109-1135, 2012.

[9] Andreani, R., Martínez, J. M., Ramos, A. and Silva, P.J.S., A Cone-Continuity Constraint qualification and Algorithmic Consequences. *SIAM Jornal Optimization* 26(1), pp. 96-110, 2016.

[10] Andreani, R., Martínez, J. M. and Schuverdt,M. L., On Second-Order Optimality Conditions for Nonlinear Programming. *Optimization* 56, pp. 529-542, 2007.

[11] Andreani, R., Martínez, J. M. and Schuverdt, M. L., On the relation between the Constant Positive Linear Dependence condition and quasinormality constraint qualification. *Journal of Optimization Theory and Applications* 125, pp. 473-485, 2005.

[12] Audet, C. and Dennis Jr., J., Penalty/barrier multiplier algorithm for semidefinite programming. *Optimization Methods and Software* 13(4) pp. 235-261, 2000.

[13] Auslender, A, Cominetti, R. and Haddou, M., Asymptotic analysis of penalty and barrier methods in convex and linear programming. *Mathematics of Operations Research* 22, pp. 43-62, 1997.

[14] Bazaraa, M.S., Sherali, H.D. and Shetty, C.M., Nonlinear Programming: Theory and Algorithms. *Wiley,* third edition, 2006.

[15] Ben-Tal, A. and Zibulevsky, M., Penalty/barrier multiplier methods for convex programming problems. *SIAM Jornal Optimization* 7, pp. 347-366, 1997.

[16] Bertsekas, D.P., Constrained Optimization and Lagrange Multiplier Methods. *Athena Scientific,* 1996.

[17] Bertsekas, D.P., Nonlinear Programming: 2nd Edition. *Athena Scientific,* 1999.

[18] Birgin, E. G., Castillo, R. A. and Martínez, J. M., Numerical comparison of augmented Lagrangian algorithms for non-convex problems. *Computational Optimization and Applications* 31(1), pp. 31-55, 2005.

[19] Birgin, E. G., Fernández, D. and Martínez, J. M., The boundedness of penalty parameters in an augmented Lagrangian method with constrained subproblems. *Optimization Methods and Software* 27(6), pp.1001-1024, 2012.

[20] Birgin E. G., Floudas, C. A. and Martinez, J. M., Global minimization using an augmented Lagrangian method with variable lower-level constraints. *Mathematical Programming,* 125(1, Ser. A), pp. 139-162, 2010.

[21] Birgin E. G., Martínez, J. M. and Prudente, L. F., Augmented Lagrangians with possible infeasibility and finite termination for global nonlinear programming. *Journal of Global Optimization* 58(2), pp. 207-242 2014.

[22] Cominetti, R. and Dussault, J.P. Stable exponential-penalty algorithm with superlinear convergence. *Journal of Optimization Theory and Applications* 83(2). pp. 285-309, 1994.

[23] Conn, A. R., Gould, N. I. M., Sartenaer, A. and Toint Ph. L.. Convergence properties of an augmented lagrangian algorithm for optimization with a combination of general equality and linear constraints. *SIAM Journal on Optimization* 6, pp. 674-703, 1996.

[24] Conn, A. R., Gould, N. I. M. and Toint, Ph. L. A globally convergent augmented lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Numerical Analysis* 28, pp.545-572, 1991.

[25] Conn, A. R., Gould, N. I. M. and Toint,Ph. L., LANCELOT: A Fortran package for large scale nonlinear optimization. *Springer-Verlag*, Berlin, 1992.

[26] Diniz-Ehrhardt, M.A., Martínez, J.M. and Pedroso, L.G., Derivative-free methods for nonlinear programming with general lower-level constraints. *Computational and Applied Mathematics* 30, pp. 19-52, 2011.

[27] Dussault, J. P., Augmented non-quadratic penalty algorithms. *Mathematical Programming* 99, pp. 467-486, 2004.

[28] Echebest, N., Sánchez, M. D. and Schuverdt, M. L., Convergence results of an augmented lagrangian method using the exponential penalty function. *Jornal Optimization Theory and Applications,* pp. 92-108, 2015.

[29] Fischer, A., Local behavior of an iterative framework for generalized equations with nonisolated solutions. *Mathematical Programming* 94(1, Ser. A), pp. 91-124, 2002.

[30] Haeser G., ~~Condicoes~~ sequenciais de otimalidade,IMECC-UNICAMP", Departamento de Matemática Aplicada, Campinas-SP, Brazil, note = "http://libdigi.unicamp.br/document/?code=000466897, 2009.

[31] Haeser, G. and Schuverdt, M. L., On approximate KKT condition and its extension to continuous variational inequalities. *Jornal Optimization Theory and Applications* 23, pp. 707-716, 2011.

[32] Hager, W. W. and Gowda, M. S.: Stability in the presence of degeneracy and error estimation. *Mathematical Programming* 85(1, Ser. A), pp. 181-192, 1999.

[33] Hestenes, M. R., Multiplier and gradient methods. *Journal of Optimization Theory and Applications* 4, pp. 303-320, 1969.

[34] Hestenes, M.R., Optimization theory: the finite dimensional case. *John Wiley,* 1975.

[35] Huang, X. X., Yang, X. Q. and Teo, K. L., Partial augmented Lagrangian method and mathematical programs with complementarity constraints. *Journal of Global Optimization* 35(2), pp. 235-254, 2006.

[36] Iusem, A. N., Augmented lagrangian methods and proximal point methods for convex optimization. *Investigación Operativa* 8, pp. 11-50, 1999.

[37] Izmailov, A. F., Solodov, M. V. and Uskov, E. I., Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. *SIAM Journal on Optimization* 22(4), pp. 1579-1606, 2012.

[38] Kočvara, M. and Stingl, M., PENNON: A code for convex nonlinear and semidefinite programming, *Optimization Methods Software,* 18(3), pp. 317-333, 2003.

[39] Kort, B. W. and Bertsekas, D. P.. Combined primal-dual and penalty methods for convex programming. *SIAM Journal on Control and Optimization,* 35, pp. 1142-1168, 1976.

[40] Kort, B. W. and Bertsekas, D. P., Multiplier methods for convex programming. *Proceedings of the IEEE Descicion and Control Conference,* San Diego, CA, pp. 260-264, 1973.

[41] Luenberger, D.G. and Ye, Y., Linear and Nonlinear Programming. *Springer,* third edition edition, 2007.

[42] Luo, H. Z., Sun, X. L. and Xu, Y. F., Convergence properties of modified and partially-augmented Lagrangian methods for mathematical programs with complementarity constraints. *Journal of Optimization Theory and Applications,* 145(3), pp. 489-506, 2010.

[43] Luo, H. Z., Sun, X. L., Xu, Y. F. and Wu, H. X., On the convergence properties of modified augmented Lagrangian methods for mathematical programming with complementarity constraints. *Journal of Global Optimization,* 46(2), pp. 217-232, 2010.

[44] Nocedal, J. and Wright, S. J., Numerical Optimization. *Springer Series in Operations Research, Springer,* 1999.

[45] Mangasarian, O. L. and Fromovitz, S., The Fritz-John necessary optimality conditions in presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications* 17, pp. 37-47, 1967

[46] Mangasarian, OL Unconstrained Lagrangians in nonlinear programming. *SIAM Journal on Control* 13, pp. 772-791, 1975.

[47] Martínez, J.M.,BOX-QUACAN and the implementation of Augmented Lagrangian algorithms for minimization with inequality constraints. *Computational and Applied Mathematics* 19, pp. 31-56, 2000.

[48] Martínez, J.M. and Birgin, E.G., Practical Augmented Lagrangian Methods for Constrained Optimization. *SIAM Publications,* pp. xiv+220, 2014.

[49] Martínez, J.M. and Svaiter, B.F. A practical optimality condition without constraint qualifications for nonlinear programming. *Jornal Optimization Theory and Applications* 118, pp. 117-133, 2003.

[50] Minchenko, L. and Stakhovski, S., On relaxed constant rank regularity condition in mathematical programming. *Optimization* 60(4), pp. 429-440, 2011.

[51] Polyak, R.A. Modified barrier function: theory and mehtods. *Mathematical Programming* 54, pp. 177-222, 1992.

[52] Polyak, R.A. and Teboulle, M., Nonlinear rescaling and proximal-like methods in convex optimization. *Mathematical Programming* pp. 76, 265-284, (1997).

[53] 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.

[54] Qi, L. and Wei, Z., On the constant positive linear dependence condition and its application to SQP methods. *SIAM Journal on Optimization* 10, pp. 963-981, 2000.

[55] Rockafellar, R. T., Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control and Optimization* 12, pp. 268-285, 1974. Collection of articles dedicated to the memory of Lucien W. Neustadt.

[56] Rockafellar, R.T. A dual approach to solving nonlinear programming problems by unconstrained optimization. *Mathematical Programming* 5,pp. 354-373, 1973.

[57] Rockafellar, R. T., Lagrange multipliers and optimality. *SIAM Review* 35, pp. 183-238, 1993.

[58] Rockafellar, R.T., The multiplier method of Hestenes and Powell applied to convex programming. *Journal of Optimization Theory and Applications,* 12, 555-562, (1973).

[59] Ryoo, N.V., Sahinidis, N.D., Global Optimization of Nonconvex NLPs and MINLPs with applications in process desig. *Computers & Chemical Engineering* 19, pp. 551-566, 1995.

[60] Tseng, P. and Bertsekas, D.P., On the convergence of the exponential multiplier method for convex programming. *Mathematical Programming* 60, pp. 1-19, 1993.

[61] Wang, C. and Li, D., Unified theory of augmented lagrangian methods for constrained global optimization. *Journal of Global Optimization* 44, pp. 433 -458, 2009.