Journal homepage http://revistas.unitru.edu.pe/index.php/SSMM

SELECCIONES MATEMÁTICAS
Universidad Nacional de Trujillo
ISSN: 2411-1783 (Online)
2024; Vol.11(2):222-235.

Improved global convergence results for augmented Lagrangian method with exponential penalty function

Resultados de convergencia global mejorados para el Método de Lagrangiano aumentado con función de penalidad exponencial

Frank Navarro Rojas[*] and Luis G. Ticona Quispe[†]

Received, Set. 06, 2024; *Accepted, Nov. 10, 2024;* *Published, Dec. 27, 2024*

How to cite this article:
Navarro R. Frank, Ticona Q. Luis G. *Improved global convergence results for augmented Lagrangian method with exponential penalty function*. Selecciones Matemáticas. 2024;11(2):222–235. http://dx.doi.org/10.17268/sel.mat.2024.02.02

Abstract

The purpose of the present paper is to improve the global convergence results established so far concerning the Augmented Lagrangian Algorithm with exponential penalty function for solving nonlinear programming problems with equality and inequality constraints. We prove global convergence for KKT points under the PAKKT-regular constraint qualifications, which results as a consequence that accumulation points generated by the algorithm are PAKKT points. This convergence result is new for the augmented Lagrangian Method based on the exponential penalty function. An interesting consequence is that the estimates of the Lagrange multipliers computed by the method remain bounded in the presence of the quasi-normality condition. Finally we give optimality and feasibility results for the convex case.
Keywords . Nonlinear programming, Augmented Lagrangian methods, exponential penalty function, global convergence, constraint qualifications, sequential optimality conditions.

Resumen

El propósito del presente artículo es el de mejorar los resultados de convergencia global establecidos hasta ahora, referentes al Algoritmo de Lagrangiano aumentado con función de penalidad exponencial para resolver problemas de programación no lineal con restricciones de igualdad y de desigualdad. Demostramos la convergencia global para puntos KKT bajo la condición de calificación PAKKT-regular, que resulta como una consecuencia de que puntos de acumulación generados por el algoritmo son puntos PAKKT. Este resultado de convergencia es nuevo para el Método Lagrangiano aumentado basado en la función de penalización exponencial. Una consecuencia interesante es que las estimaciones de los multiplicadores de Lagrange calculadas por el método permanecen acotados en presencia de la condición de cuasi-normalidad. Finalmente damos resultados de optimalidad y viabilidad para el caso convexo.
Palabras clave. Programación no lineal, métodos Lagrangiano aumentado, función de penalización exponencial, convergencia global, condiciones de calificación, condiciones de optimalidad secuencial.

[*]Facultad de Ciencias. Universidad Nacional de Ingeniería, Lima, Perú. **Correspondence author** (fnavarro@uni.edu.pe).
[†]Facultad de Ciencias. Universidad Nacional de Ingeniería, Lima, Perú. (luis.ticona.q@uni.pe).

1. Introducción. Los métodos Lagrangianos Aumentados son uno de los métodos mas populares y usados en optimización numérica. Las ideas principales fueron introducidas por Powell [1], Hestenes [2] y Rockafellar [3]. La idea principal del método es que en cada iteración para un parámetro de penalización fijo y una estimación de los multiplicadores de Lagrange, se minimiza aproximadamente la función objetivo más un término que penaliza el incumplimiento de las restricciones con respecto a tolerancias desplazadas adecuadamente; en la práctica estos subproblemas son considerablamente mas fáciles a resolver que el problema original. Una vez encontrada la solución aproximada, se definen nuevas estimativas a los multiplicadores, se actualiza el parámetro de penalidad y se inicia una nueva iteración.

A diferencia de los métodos clásicos de penalización externa [4, 5] que necesita emplear parámetros de penalización que tienden al infinito, la técnica de desplazamiento usada por el método de Lagrangano aumentado tiene como objetivo producir convergencia mediante desplazamientos de las restricciones que generan aproximaciones a una solución con parámetros de penalización moderados [6]. Algunas variantes del método, han sido estudiadas ampliamente, por ejemplo la versión salvaguardada del método [7] descarta las aproximaciones de los multiplicadores de Lagrange cuando se hacen muy grandes. La teoría de convergencia de los métodos lagrangianos aumentados Lagrangianos Salvaguardados fue dada en [7, 6].

En [8] se han dado ejemplos que ilustran la ventajas del método de lagrangiano aumentado salvaguardado en relación al lagrangiano aumentado estándar.
Para problemas de programación no lineal, se ha demostrado que esta variante posee fuertes propiedades de convergencia incluso bajo hipótesis muy débiles [9, 10]. Desde entonces se ha aplicado para resolver otros problemas, incluyendo programas matemáticos con restricciones de complementariedad (MPCC) [11, 12], desigualdades cuasi variacionales [13], problemas de equilibrio de Nash generalizados [14, 15], programación semidefinida [16, 17] problemas de optimización con restricciones de cardinalidad [18], programación cónica no lineal [16], problemas de optimización multiobjetivo [19] y problemas de cuasi equilibrio [20].

La función lagrangiana aumentada más ampliamente utilizada, es debido a Powell–Hestenes-Rockafellar [1, 2, 3] , que se basa en la función de penalización cuadrática (llamada PHR función). La penalidad cuadrática goza de interesantes propiedades teóricas y es la más utilizada en implementaciones prácticas (ver [21] para una comparación numérica entre diferentes funciones de penalidad).

En [21] los autores concluyeron la superioridad del uso de la función de penalización cuadrática sobre otras no cuadráticas; creemos que esta conclusión depende en gran medida del algoritmo interno utilizado para resolver los subproblemas en cada iteración. La discontinuidad de las segundas derivadas de la penalidad cuadrática puede volver lenta la tasa de convergencia y provocar fallos, además limita el uso de métodos rapidos que usan en general información de segundo orden, así, como se menciona en [22], los métodos de tipo Newton se pueden utilizar de forma más eficaz para minimizar los subproblemas correspondientes para penalidades dos veces diferenciables.

Segun nuestra opinión, esto es suficiente para estudiar el enfoque Lagrangiano Aumentado considerando una opción diferente a la función de penalización cuadrática.

Otra motivación importante para considerar funciones de penalización no cuadráticas en el método lagrangiano aumentado, es el de establecer resultados de convergencia a puntos que verifiquen condiciones de optimmalidad de segundo orden. En este caso, es deseable tener una función Lagrangiana Aumentada dos veces diferenciable.

En la presente investigación, consideramos el métododo de lagrangiano aumentado salvaguardado con función de penalización exponencial. Esta función fue introducida en el contexto de la programación convexa [23, 24]. Las funciones de penalización no cuadráticas han sido ampliamente estudiadas durante las últimas décadas; véase, por ejemplo, [22, 25, 26, 27, 28, 29, 30, 31] y sus referencias.

El método lagrangiano aumentado definido aquí sigue la idea introducida en [6], pero la diferencia principal radica en la función de penalización utilizada. Un estudio un poco mas detallado del métododo de lagrangiano aumentado salvaguardado con el uso de la función de penalización exponencial fue dado en [29] donde se demostro que un punto límite viable de la sucesión generada por el método que satisface la condición de calificación: generador positivo constante (CPG) [32], cumple las condiciones de Karush-Kuhn-Tucker (KKT); también se establecieron condiciones suficientes para garantizar la acotación de los parámetros de penalidad que facilitan la resolución de los subproblemas generados por el método en cada iteración.

Posteriormente en [31] se estudia un modelo del lagrangiano aumentado, con una función de penalidad general que tiene como casos particulares a la penalidad cuadrática y la exponencial, en este contexto se estableció la convergencia global para un punto KKT bajo la condición de calificación de cono continuo (CCP) que es más débil que CPG [33], este resultado se obtuvo como consecuencia de demostrar que puntos límites factibles de sucesiones generadas por el método son puntos AKKT [34].

En [10] se define una nueva condición de optimalidad secuencial para optimización diferenciable con restricciones , llamada PAKKT. Se demuestra que todo punto que satisface la condición de optimalidad secuencial PAKKT satisface la condición de optimalidad secuencial AKKT. Esto prueba la eficiencia de métodos que generan sucesiones cuyos puntos límites son puntos que cumplen PAKKT por sobre los que cunplen AKKT, por ejemplo el algoritmo Lagrangiano aumentado con penalidad cuadrática [10]. También en [10] se establece la condición de calificación estricta asociada con la condición de optimalidad secuencial PAKKT, llamada PAKKT-regular, y se demostró que es estrictamente más débil que tanto la cuasi-normalidad [35] como de la propiedad de cono continuo (CCP). PAKKT-regular conecta ambas ramas de estas condiciones de calificación independientes.

Recientemente nuevos resultados de convergencia para el método de lagrangiano aumentado cuadrático fueron presentados en [36, 37, 38, 39], estos resultados establecen consecuencias algorítmicas provenientes de nuevas condiciones de optimalidad halladas, con sus respectivas condiciones de calificación estrictas.
En particular, las siguientes son las principales contribuciones de este artículo:

- Bajo la condición de calificación de cuasi-normalidad demostramos la acotación de las sucesiones duales generadas por el método. Esta propiedad es particularmente relevante para el análisis de complejidad y para aplicaciones donde tambien se busca la solución dual.
- Mejoramos la teoría de convergencia proporcionada en [29] y [31] utilizando condiciones de calificación más débiles, en particular bajo la condición de calificación PAKKT-regular [34], que se obtiene como consecuencia de que puntos límites viables generados por el método son puntos PAKKT.
- Para el caso convexo aseguramos la viabilidad del punto límite, y establecemos optimalidad sin el uso de alguna condición de calificación.

Este artículo está organizado de la siguiente manera: la Sección 2 presenta los resultados preliminares y las definiciones relevantes que usaremos. En la Sección 3 presentamos el algoritmo de Lagrangiano aumentado salvaguardado con penalidad exponencial y establecemos los resultados de convergencia global. En la Sección 4 presentamos resultados de optimalidad y viabilidad para el caso convexo y en la Sección 5 se presentan algunas observaciones finales.

Notación.

- $\mathbb{R}_+ = \{t \in \mathbb{R} \mid t \geq 0\}$, $\|\cdot\|$ denota una norma vectorial arbitraria, $\|\cdot\|_\infty$ la norma del máximo, y $\|\cdot\|_2$ la norma euclidiana.
- v_i es el $i-$ésimo componente del vector v.
- Para todo $y \in \mathbb{R}^n, y_+ = (\text{máx}\{0, y_1\}, \ldots, \text{máx}\{0, y_n\})$.
- Si $K = \{k_0, k_1, k_2, \ldots\} \subset \mathbb{N}\,(k_{j+1} > k_j)$, escribimos $\text{lím}_{k\in K} y^k = \text{lím}_{j\to\infty} y^{k_j}$. En particular, $\text{lím}_k y^k = \text{lím}_{k\in\mathbb{N}} y^k$.
- Si $\{\gamma_k\} \subset \mathbb{R}, \gamma_k > 0$ y $\gamma_k \to 0$, escribimos $\gamma_k \downarrow 0$.
- Para una función vectorial $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m, F := (f_1, \ldots, f_m)$, la matriz jacobiana se denota por $\nabla F(x)$ y se define por $\nabla F(x) := (\nabla f_1(x), \ldots, \nabla f_m(x)) \in \mathbb{R}^{n\times m}$.

2. Notación y prelimares. Consideraremos el problema general de programación no lineal con restricciones:

$$\begin{array}{lll} (\text{P}) & \text{minimizar} & f(x) \\ & \text{sujeto a} & x \in X \end{array}$$

donde $f : \mathbb{R}^n \to \mathbb{R}$ y X es el conjunto viable formado por restricciones de igualdad y desigualdad de la forma

$$X = \{x \in \mathbb{R}^n \mid h(x) = 0, g(x) \leq 0\},$$

con $h : \mathbb{R}^n \to \mathbb{R}^m$ y $g : \mathbb{R}^n \to \mathbb{R}^p$. Donde las funciones f, h y g son continuamente diferenciables en $\mathbb{R}^n$. Dado $x^* \in X$, denotamos el conjunto de restricciones de desigualdad activa en x^* por $I_g(x^*) = \{j \in \{1, \ldots, p\} \mid g_j(x^*) = 0\}$.

El método Lagrangiano aumentado utiliza una sucesión iterativa de subproblemas que en la práctica deben de ser considerablemente más fáciles de resolver que el problema original. En cada subproblema, con un parámetro de penalización fijo $\rho > 0$ y estimaciones de los multiplicadores de Lagrange $\lambda \in \mathbb{R}^m$ y $\mu \in \mathbb{R}^p, \mu \geq 0$, se calcula un punto estacionario aproximado de una función lagrangiana aumentada . Una vez que se encuentra el punto estacionario aproximado, el parámetro de penalización ρ y las estimaciones de los multiplicadores se actualizan y comienza una nueva iteración. Como ya mencionamos, la función Lagrangiana Aumentada más usada se basa en la función de penalización cuadrática, llamada función lagrangiana aumentada de Powell-Hestenes-Rockafellar (PHR) que, para el problema (P), toma la forma

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

donde $x \in \mathbb{R}^n, \rho > 0, \lambda \in \mathbb{R}^m$ y $\mu \in \mathbb{R}^p, \mu \geq 0$.

En este trabajo, asociado a (P) consideramos la función Lagrangiana Aumentada basada en la función de penalización exponencial:

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

donde $x \in \mathbb{R}^n, \rho > 0, \lambda \in \mathbb{R}^m$ y $\mu \in \mathbb{R}^p, \mu \geq 0$.

En los últimos años, se ha dedicado especial atención a las llamadas condiciones secuenciales de optimalidad para optimización no lineal con restricciones (ver, por ejemplo, [36, 34, 40, 41]). Estas condiciones ademas estan relacionadas con los criterios de parada de algoritmos en optimización y tienen como objetivo unificar los resultados teóricos de convergencia. En particular, se han utilizado para estudiar ampliamente la convergencia del método lagrangiano aumentado con penalidad cuadratica (ver [6, 36, 37, 38] y las referencias allí).

Una característica importante de las condiciones de optimalidad secuencial es que son necesarias para optimalidad: un minimizador local de (P) verifica dicha condición independientemente del cumplimiento de cualquier condición de calificación. Una de las primeras condiciones de optimalidad secuencial mas estudiadas y populares es la condición aproximada de Karush-Kuhn-Tucker (AKKT), definida en [34].

Definición 2.1. *Decimos que $x^* \in X$ satisface la condición AKKT si existen sucesiones $\{x^k\} \subset \mathbb{R}^n, \{\lambda^k\} \subset \mathbb{R}^m$, y $\{\mu^k\} \subset \mathbb{R}^p, \mu^k \geq 0$, tal que*

$$\lim_{k\to\infty} x^k = x^*, \tag{2.1}$$

$$\lim_{k\to\infty} \|\nabla_x L(x^k,\lambda^k,\mu^k)\| = 0, \tag{2.2}$$

$$\lim_{k\to\infty} \|\text{mín}\left\{-g(x^k),\mu^k\right\}\| = 0, \tag{2.3}$$

donde $L(\cdot)$ es la función lagrangiana asociada con (P). Este tipo de puntos x^* son llamados puntos AKKT y $\{x^k\}$ una sucesión AKKT.

Cuando decimos que una condición de optimalidad secuencial unifica los resultados de convergencia, nos referimos a que cuando se demuestra que un punto AKKT que cumple una determinada condición de calificación (CQ) es un punto KKT, todos los algoritmos que alcanzan puntos AKKT, como el método Lagrangiano aumentado con penalidad cuadrática [6], tienen su convergencia teórica establecida automáticamente con el mismo CQ.
En los ultimos años, se ha demostrado que puntos AKKT son puntos estacionarios bajo diferentes CQs, como la dependencia lineal positiva constante (CPLD) [42], la dependencia lineal positiva constante relajada (RCPLD) [43] y el generador positivo constante (CPG). [44]. Finalmente, se demostró que la propiedad de continuidad del cono (CCP) (también llamada AKKT-regular [9]) es la condición de calificación mas débil para esta propiedad. Todas estas CQ tratan con hipótesis de rango y/o dependencia lineal positiva. Las CQ de pseudonormalidad y cuasinormalidad [35], por el contrario, tienen una naturaleza diferente. Imponen un control sobre los multiplicadores alrededor del punto de interés.

En [10] se introduce una nueva condición de optimalidad secuencial la: Positiva Aproximada Karush-Kuhn-Tucker(PAKKT).

Definición 2.2. *Decimos que $x^* \in X$ es un punto KKT aproximado positivo (PAKKT) si* *existen sucesiones* $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, *y* $\{\mu^k\} \subset \mathbb{R}^p_+$ *tal que*

$$\lim_k x^k = x^*, \tag{2.4}$$

$$\lim_k \|\nabla_x L(x^k, \lambda^k, \mu^k)\| = 0, \tag{2.5}$$

$$\lim_k \|\min\{-g(x^k), \mu^k\}\| = 0, \tag{2.6}$$

$$\lambda_i^k h_i(x^k) > 0 \; si \quad \lim_k \frac{|\lambda_i^k|}{\delta_k} > 0, \tag{2.7}$$

$$\mu_j^k g_j(x^k) > 0 si \quad \lim_k \frac{\mu_j^k}{\delta_k} > 0, \tag{2.8}$$

donde $\delta_k = \left\|(1, \lambda^k, \mu^k)\right\|_\infty$. *En este caso,* $\{x^k\}$ *se llama sucesión PAKKT.*

Las expresiones (2.4)-(2.6) están relacionadas con las condiciones KKT y se utilizan en la condición de optimización aproximada de KKT (AKKT). Las expresiones (2.7) y (2.8) pretenden controlar el signo de los multiplicadores de Lagrange, justificando el nombre de la nueva condición. Como siempre $|\lambda_i^k|/\delta_k, \mu_j^k/\delta_k \in [0,1]$, y podemos suponer, tomando una subsecuencia si es necesario, que estos límites existen.

Teorema 2.1. *PAKKT es una condición necesaria de optimalidad.*

Demostración: Ver [10] □

Teorema 2.2. *Cada sucesión PAKKT es también una sucesión AKKT. En particular, cada punto PAKKT es un punto AKKT.*

Demostración: Ver [10] □

Decimos que una condición de calificación es estricta (SCQ) para la condición de optimalidad secuencial A si A + SCQ implica KKT (ver [17]). Dado que toda condición de optimalidad secuencial se satisface en cualquier minimizador local independientemente del cumplimiento de una CQ, un SCQ es de hecho una condición de calificación.

Por otro lado, la condición de calificación SCQ proporciona una medida de la calidad de la condición de optimimalidad secuencial A. Específicamente, A es mejor, cuando el SCQ asociado sea más débil. En [9], los autores presentaron la condición de calificación estricta más débil asociada con AKKT, llamada propiedad de continuidad del cono (CCP), denominada en [41] como .ᴬKKT-regular", además en [45], [44] y [46] se establecieron los SCQ más débiles relacionados con otras condiciones secuenciales de optimalidad.

En [10] se estableció que la SCQ más débil asociada a la condición PAKKT, es la llamada PAKKT-regular. Para definir esta CQ primero definimos para cada $x^* \in X$ y $x \in \mathbb{R}^n$, $\alpha, \beta \geq 0$, el conjunto

$$K_+(x,\alpha,\beta) = \left\{ \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in I_g(x^*)} \mu_j \nabla g_j(x) \;\middle|\; \begin{array}{l} \lambda_i h_i(x) \geq \alpha \text{ si } |\lambda_i| > \beta \|(1,\lambda,\mu)\|_\infty, \\ \mu_j g_j(x) \geq \alpha \text{ si } \mu_j > \beta \|(1,\lambda,\mu)\|_\infty, \\ \lambda \in \mathbb{R}^m, \mu_j \in \mathbb{R}_+, j \in I_g(x^*) \end{array} \right\}.$$

Tenga en cuenta que las condiciones KKT para (P) se pueden escribir como $-\nabla f(x^*) \in K_+(x^*, 0, 0)$.

Dada una multifunción $\Gamma : \mathbb{R}^s \rightrightarrows \mathbb{R}^q$, el límite exterior/superior secuencial de Painlevé-Kuratowski de $\Gamma(z)$ cuando $z \to z^*$ se denota por

$$\limsup_{z \to z^*} \Gamma(z) = \left\{ y^* \in \mathbb{R}^q \mid \exists (z^k, y^k) \to (z^*, y^*) \text{ con } y^k \in \Gamma(z^k) \right\}. \tag{2.9}$$

(ver [47]). Decimos que Γ es semicontinuo exterior en z^* si $\limsup_{z \to z^*} \Gamma(z) \subseteq \Gamma(z^*)$.

Definimos la condición regular PAKKT, que impone una condición externa similar a la semicontinuidad de la multifunción $(x,\alpha,\beta) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}_+ \rightrightarrows K_+(x,\alpha,\beta)$. De manera análoga a (2.9), consideramos el siguiente conjunto:

$$\begin{aligned} &\limsup_{x \to x^*, \alpha \downarrow 0, \beta \downarrow 0} K_+(x,\alpha,\beta) \\ &= \left\{ y^* \in \mathbb{R}^n \mid \exists (x^k, y^k) \to (x^*, y^*), \alpha_k \downarrow 0, \beta_k \downarrow 0 \text{ con } y^k \in K_+(x^k, \alpha_k, \beta_k) \right\}. \end{aligned} \tag{2.10}$$

Definición 2.3. *Decimos que* $x^* \in X$ *satisface la condición regular PAKKT si*

$$\limsup_{x \to x^*, \alpha \downarrow 0, \beta \downarrow 0} K_+(x,\alpha,\beta) \subseteq K_+(x^*, 0, 0).$$

el resultado que garantiza que PAKKT-regular es la SCQ más débil para la condición de optimalidad secuencial PAKKT fue demostrado en [10]; y que todo punto KKT es un punto PAKKT.

Ademas en [10] se demuestra que PAKKT-regular es estrictamente más débil que la cuasinormalidad y que CCP.

Como consecuencia algoritmica de estos resultados, fueron generalizados todos los resultados teóricos de convergencia anteriores a [10] para el método lagrangiano aumentado PHR. De hecho, demostraron que este método alcanza puntos PAKKT. Estos puntos son más fuertes que los puntos KKT aproximados (AKKT) , que se habían utilizado para analizar la convergencia de esta popular técnica.

También en [10], se demuestra que la sucesión dual generada por el método de lagrangiano aumentado cuadrático, permanece acotada bajo la condición de cuasi-normalidad, esta CQ esta asociada con los puntos mejorados de Fritz Jhon (EFJ).

Definición 2.4. *Un punto viable $\bar{x}$ para (P) satisface la condición de cuasinormalidad (QN) si no existe $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$ tal que:*

1. $\nabla h(\bar{x})\lambda + \nabla g(\bar{x})\mu = 0$;
2. $(\lambda, \mu) \neq 0$;
3. *Existe una sucesión $\left\{x^k\right\}$ que converge a $\bar{x}$ tal que, para todo k,*
   (a) $\lambda_i h_i(x^k) > 0, \forall i \in I_{\neq}, \quad$ *y* $g_j(x^k) > 0, \forall j \in J_+$;
   donde $I_{\neq} = \{i \mid \lambda_i \neq 0\}$, $J_+ = \{j \mid \mu_j > 0\}$.

Tambien en [10] se demuestra que AKKT-regular es mas débil que cuasi-normalidad.

Teorema 2.3. *La condición de calificación de cuasi-normalidad implica PAKKT-regular.*

Finalmente de [10], obtenemos que suceciones PAKKT que convergen a un punto satisfaciendo la condicion de cuasi-normalidad tiene sucesiones duales acotadas.

Teorema 2.4.

Sea $\left\{x^k\right\}$ una sucesión PAKKT asociada al problema (P), con $\left\{(\lambda^k, \mu^k)\right\}$ como sucesión dual asociada. Si $x^* = \lim_k x^k$ satisface la CQ de cuasi-normalidad entonces la sucesión dual es acotada.

En el siguiente capitulo demostramos que estos resultados algoritmicos hallados en [10] para método lagrangiano aumentado PHR, tambien se dan para el método lagrangiano aumentado con penalidad exponencial.

3. El método lagrangiano aumentado exponencial y su convergencia global.

Con respeto al problema (P), la función lagrangiana $\mathscr{L}$ es definida por:

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

para todo $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$, y $\mu \in \mathbb{R}^p_+$, mientras que la función Lagrangiana Aumentada exponencial, que usaremos en lo sucesivo es dada por:

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

para todo $\rho > 0, x \in \mathbb{R}^n, \lambda \in \mathbb{R}^m$, y $\mu \in \mathbb{R}^p_+$.

En la literatura sobre métodos numéricos para optimización global, los algoritmos están diseñados generalmente para converger a puntos estacionarios y no necesariamente a minimizadores globales; estos algoritmos son conocidos como algoritmos locales. La convergencia de este tipo de algoritmos generalmente es obtenida con un punto incial arbitrario y suelen ser globalmente convergentes en el sentido de que puntos de acumulación de la sucesión generada son puntos estacionarios.

En esta sección, presentamos un algoritmo local basado en el Lagrangiano Aumentado exponencial. El algoritmo halla un punto estacionario aproximado de la función lagrangiana aumentada exponencial, actualiza los multiplicadores de Lagrange y los parámetros de penalización en cada iteración.

Algoritmo 1 (Método Lagrangiano Aumentado Exponencial).

Sea $x^1 \in \mathbb{R}^n$ un punto inicial arbitrario. Los parámetros dados para la ejecución del algoritmo son los siguientes: $\tau \in [0, 1), \gamma > 1, \lambda_{\text{mín}} < \lambda_{\text{máx}}, \mu_{\text{máx}} \geq \mu_{\text{mín}} > 0$ *y* $\rho_1 > 0$. *Además,* *sean* $\bar{\lambda}^1 \in [\lambda_{\text{mín}}, \lambda_{\text{máx}}]^m$ *y* $\bar{\mu}^1 \in [\mu_{\text{mín}}, \mu_{\text{máx}}]^p$ *las estimaciones iniciales de los multiplicadores de Lagrange. Finalmente,* $\{\varepsilon_k\} \subset \mathbb{R}_+$*es una sucesión de parámetros de tolerancia tal que* $\text{lím}_k\, \varepsilon_k = 0$.

Inicialice $k \leftarrow 1$.

Paso 1. (Resolver el subproblema).
Calcular (si es posible) $x^k \in \mathbb{R}^n$ *tal que*

$$\|\nabla_x L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)\| \leq \varepsilon_k.$$

Si no es posible, detener la ejecución del algoritmo, declarando fallo y declarar la finalización del algoritmo.

Paso 2. (Estime los nuevos multiplicadores).
Calcule $\lambda^k \in \mathbb{R}^m$ *y* $\mu^k \in \mathbb{R}^p$ *tal que :*

$$\lambda^k = \bar{\lambda}^k + \rho_k h(x^k) \quad \text{y} \quad \mu_j^k = \bar{\mu}_j^k e^{\rho_k g_j(x^k)}, j = 1, ..., p.$$

Paso 3. (Actualización del parámetro de penalización).
Definir

$$V^k = \frac{\mu^k - \bar{\mu}^k}{\rho_k}.$$

Si $k = 1$ *o*

$$\text{máx}\left\{\|h\left(x^k\right)\|, \|V^k\|\right\} \leq \tau\, \text{máx}\left\{\|h\left(x^{k-1}\right)\|, \|V^{k-1}\|\right\},$$

elija $\rho_{k+1} = \rho_k$. *De lo contrario, defina* $\rho_{k+1} = \gamma\rho_k$.

Paso 4. (Actualice las estimaciones de los multiplicadores).
Calcule $\bar{\lambda}^{k+1} \in [\lambda_{\text{mín}}, \lambda_{\text{máx}}]^m$ *y* $\bar{\mu}^{k+1} \in [\mu_{\text{mín}}, \mu_{\text{máx}}]^p$.

Paso 5. (Comience una nueva iteración).
Hacer $k \leftarrow k + 1$ *e ir al paso 1.*

Tenga en cuenta que, a pesar de la acotación de $(\bar{u}^k)$ y $(\bar{\lambda}^k)$, las sucesiones correspondientes (μ^k) y (λ^k) generadas por el algoritmo podrían ser no acotadas. Además, enfatizamos que el algoritmo permite cierta libertad en la elección de las sucesiones $(\bar{u}^k)$ y $(\bar{\lambda}^k)$. En el caso extremo, podemos tomar $\bar{u}^k := 0$ y $\bar{\lambda}^k = 0$ para todos los k, en cuyo caso el algoritmo se reduce escencialmente al método de penalización clásico. Una forma más natural y numéricamente más usada en implementaciones es tomar $\bar{u}^k$ y $\bar{\lambda}^k$ como la proyección de u^k y λ^k sobre las cajas $\left[\mu^{\text{mín}}, \mu^{\text{máx}}\right]^p$ y $\left[\lambda^{\text{mín}}, \lambda^{\text{máx}}\right]^m$ respectivamente.

Otra diferencia a comentar entre el algoritmo lagrangiano con penalidad cuadrática, como tambien el del lagrangiano exponencial dado en [29], con Algoritmo 1 es que en ellos se consideran $\mu^{\text{mín}} = 0$ en nuestro caso y tambien en [31] tomamos $\mu^{\text{mín}} > 0$; en implementaciones computacionales elegimos $\mu^{\text{mín}}$ suficientemente pequeño.

Un primer resultado referente al algoritmo es que los multiplicadores asociados a las restricciones de desigualdad que son inactivas en el punto límite son asintóticamente nulos, independientemente de la viabilidad del punto límite.

Este resultado es una adaptación directa del Teorema 4.1 de [6] y que también fue demostrado en [29]. Nótese que en el enunciado del Teorema siguiente, se supone la existencia de un punto límite x^*.

Teorema 3.1. *Sea* $\left\{x^k\right\}$ *una sucesión generada por el Algoritmo 1. Supongamos que* x^* *es un punto límite de* $\left\{x^k\right\}$ *y que* $K \subseteq \mathbb{N}$ *es tal que* $\text{lím}_{k\in K}\, x^k = x^*$, *entonces* $\forall j : g_j(x^*) < 0$ *tenemos que* $\text{lím}_{k\in K}\, \mu_j^k = 0$,

Demostración: Consideremos los dos casos siguientes:

a) Si $\{\rho_k\}$ está acotado, por el Paso 3 del Algoritmo tenemos que $\text{lím}_{k\in K}\, V_j^k = 0$ para todo $j = 1, ..., p$. Si $g_j(x^*) < 0$, obtenemos que $\text{lím}_{k\in K}\, \frac{e^{\rho_k g_j(x^k)} - 1}{\rho_k} \neq 0$ y esto implica que $\text{lím}_{k\in K}\, \bar{\mu}_j^k = 0$. Por lo tanto, según la definición de μ_j^k se tiene que que $\text{lím}_{k\in K}\, \mu_j^k = 0$.

b) Si $\{\rho_k\}$ no es acotado, ya que $g_j(x^*) < 0$, tenemos que $\lim_{k\in K} e^{\rho_k g_j(x^k)} = 0$. Como $\{\bar{\mu}_j^k\}_{k\in K}$ está acotado, obtenemos que $\lim_{k\in K} \mu_j^k = 0$. Que es lo que queriamos demostrar.

□

Es conocido que cuando el algoritmo no se detiene por falla (Paso 1), genera una sucesión AKKT para el problema (P) siempre que tenga un punto límite viable (ver [31]). En particular, cada punto de acumulación viable de este algoritmo es un punto AKKT. Este resultado tambien fue demostrado en un contexto mas general en [48] donde se considera una familia de funciones de penalidad, del cual la penalidad cuadrática y exponencial son casos particulares.

A continuación demostramos que el Algoritmo 1 alcanza los puntos mas fuertes PAKKT. En lo sucesivo, supondremos que el método genera una sucesión primal infinita.

Teorema 3.2. *Todo punto de acumulación viable x^* generado por el Algoritmo 1 es un punto PAKKT.*

Demostración: Sean $\{x^k\}, \{\bar{\lambda}^k\}, \{\bar{\mu}^k\}$, y $\{\rho_k\}$ sucesiones generadas por el Algoritmo 1 y sea x^* un punto de acumulación viable de $\{x^k\}$. Podemos suponer, tomando una subsucesión si es necesario, que $\lim_k x^k = x^*$. Tenemos

$$\nabla_x L_{\rho_k}(x^k, \lambda^k, \mu^k) = \nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k \to 0 \tag{3.1}$$

dónde

$$\lambda^k = \bar{\lambda}^k + \rho_k h(x^k) \quad \text{ y } \quad \mu_j^k = \bar{\mu}^k e^{\rho_k g_j(x^k)} \quad j = 1, ..., m$$

Puesto que $-|g_j(x^k)| \leq \min\{-g_j(x^k), \mu_j^k\} \leq |g_j(x^k)|$ y usando el Teorema 3.1, vemos que (2.6) se cumple.

Defina $\delta_k = \left\|(1, \lambda^k, \mu^k)\right\|_\infty$ como en la definición de PAKKT. Si $\{\delta_k\}$ no es acotado, entonces podemos suponer, tomando una subsucesión si es necesario, que, además de $\{\delta_k\}$ ser no acotado, $\lim_k \lambda_i^k/\delta_k$ existe para todo i y $\lim_k \mu_j^k/\delta_k$ existe para todo j.

Entonces, de la acotación de $\{\bar{\lambda}^k\}$ tenemos que, para cada i tal que $\lim_k \lambda_i^k/\delta_k \neq 0$,

$$0 \neq \lim_k \frac{\lambda_i^k}{\delta_k} = \lim_k \left[\frac{\bar{\lambda}_i^k}{\delta_k} + \frac{\rho_k h_i(x^k)}{\delta_k}\right] = \lim_k \frac{\rho_k h_i(x^k)}{\delta_k} \Rightarrow \lambda_i^k h_i(x^k) > 0 \;\forall k \geq k_i,$$

por algunos $k_i \geq 1$. Entonces (2.7) se satisface con una subsucesión de $\left\{(x^k, \lambda^k, \mu^k)\right\}$ que se inicializa desde el índice $\max_i k_i$.

Sea j tales que $g_j\,(x^*) = 0$, si $\lim_k \mu_j^k/\delta_k > 0$ entonces $0 < \lim_k \mu_j^k/\delta_k = \lim_k \frac{\bar{\mu}_j^k e^{\rho_k g_j(x^k)}}{\delta_k}$, como $\lim_k \frac{\bar{\mu}^k}{\delta_k} = 0$ entonces (tomando una subsucesión si es necesario) $(e^{\rho_k g_j(x^k)})$ no es acotada, luego $g_j(x^k) > 0$ por algunos $k_j \geq 1$, por otro lado tambien tenemos que, $\mu_j^k > 0$, luego $\mu_j^k g_j(x^k) > 0 \;\forall k \geq k_j$. Ahora si $g_j(x^*) < 0$, entonces $\lim_k \mu_j^k/\delta_k \leq \lim_k \mu_j^k = 0$, y estos índices j no violan (2.8). Por lo tanto, hemos demostrado que $\left\{x^k\right\}$ es una secuencia PAKKT y, en consecuencia, que x^* es un punto PAKKT cuando $\{\delta_k\}$ es acotada.

Si $\{\delta_k\}$ está acotada, entonces, tomando el límite en (3.1) en una sucesión apropiada, concluimos que x^* es un punto KKT, y entonces un punto PAKKT según el Lema 2.6 en [10] (no necesariamente con la misma sucesión dual-primal generada por el método). □

Es importante señalar que el Algoritmo 1 genera puntos PAKKT, pero no necesariamente sucesiones PAKKT. Es decir, afirmamos que para cada punto de acumulación viable x^* hay una sucesión PAKKT correspondiente, pero la sucesión generada $\{x^k\}$ no necesariamente tiene una subsucesión con esta propiedad. Específicamente, cuando la sucesión dual $\{\delta_k\}$ no es acotada, $\{x^k\}$ tiene una subsucesión PAKKT asociada con x^*. Pero en la prueba del teorema anterior no tenemos ninguna garantía de que la sucesión generada por el método tenga una subsucesión PAKKT si $\{\delta_k\}$ está acotada. Por supuesto, esto no es un incoveniente porque en el último caso el punto límite ya es un punto KKT.

El siguiente resultado es una consecuencia directa del teorema anterior y del teorema 2.4 en [10]

Corolario 3.1. *Sea x^* un punto de acumulación viable de (P) generado por el Algoritmo 1. Si x^* satisface la condición de calificación PAKKT-regular, entonces x^* es un punto KKT.*

Una consecuencia importante de la discusión anterior y del teorema 2.4 permite obtener el siguiente resultado de forma directa .

Corolario 3.2.

Sea $\{x^k\}$ la sucesión generada por el Algoritmo 1 aplicado a (P). Supóngase que $x^* \in X$ es un punto de acumulación viable de $\{x^k\}$ con un conjunto de índices infinito asociado K, es decir, $\lim_{k\in K} x^k = x^*$. Además, supóngase que x^* satisface la condición de cuasi-normalidad. Entonces, las sucesiones duales asociadas $\{\lambda^k\}_{k\in K}$ y $\{\mu^k\}_{k\in K}$ calculadas en el paso 2 están acotadas.

Ejemplo 3.1.
Proporcionamos un ejemplo que tambien fue usado en [8] y que adaptamos para el caso de la penalidad exponencial. Veremos que el Algoritmo 1 genera una sucesión cuyos terminos son puntos estacionarios (de hecho, minimizadores locales) de los subproblemas, y que la sucesión tiene dos puntos estacionarios, uno de los cuales es inviable y viola básicamente cualquier condición de calificación y, por lo tanto, no se puede esperar que sea un punto KKT del problema de optimización , mientras que el otro punto de acumulación es viable satisface esencialmente todas las condiciones de calificación y, por lo tanto, es necesariamente un punto KKT del problema de optimización por el Corolario 3.1. Sea $n = m = 1$ y consideremos el problema de optimización dado por:

$$\begin{array}{ll} \textit{minimizar} & x \\ \textit{sujeto a} & 1 - x^3 \leq 0 \end{array}$$

tenemos que $f(x) = x$ y $g(x) = 1 - x^3$. Es fácil ver que $\bar{x} := 1$ es la única solución de este problema de optimización; además, un cálculo sencillo muestra que $(\bar{x}, \bar{\lambda}) := (1, 1/3)$ es el único punto KKT.

Consideremos el algoritmo 1 aplicado a este problema. El análisis siguiente es bastante general y supongamos por conveniencia que $\rho_0 > ln(27)$, $\bar{\mu}^{\text{mín}} = \dfrac{1}{81}$, donde $\bar{\mu}^{\text{máx}} = 1$ y $\bar{\mu}^0 \leq 1/3$.

Además, definimos $\bar{u}^{k+1}$ como la proyección de μ^k sobre el intervalo $\left[\bar{\mu}^{\text{mín}}, \bar{\mu}^{\text{máx}}\right]$. Se ve fácilmente que, para todo $\mu > 0$ y $\rho > 0$, la función $L_\rho(\cdot, \mu) = x + \dfrac{\mu}{\rho}(e^{\rho(1-x^3)} - 1)$ es coerciva sobre $\mathbb{R}$, $L'_\rho(x, \mu) = 1 - 3x^2\mu e^{\rho(1-x^3)}$, $\lim_{x\to-\infty} L'_\rho(x, \mu) = -\infty$ y $\lim_{x\to\infty} L'_\rho(x, \mu) = 1$ se obtiene ademas que $L'_\rho(0, \mu) = 1$ y $L'_\rho(1, \mu) = 1 - 3\mu$. De ello se deduce usando el criterio de la segunda derivada que L_ρ siempre alcanza un mínimo local en $(-\infty, 0)$ y si $\mu > 1/3$, alcanza otro mínimo local en $(1, +\infty)$. Sea $\left(x^k\right)_{k\geq 1}$ la sucesión de minimizadores locales tales que:

- para k impar, x^k es el minimizador local más grande en $(-\infty, 0)$,
- para k par, x^k es el minimizador local más pequeño en $(1, +\infty)$.

Si k es impar, tenemos $x^k < 0$ y $g(x^k) > 1$. De ello se deduce que:

$$\mu^k = \bar{\mu}^k e^{\rho_k g(x^k)} > \bar{\mu}^k e^{\rho_k} > \frac{1}{3},$$

luego $\bar{u}^{k+1} > \frac{1}{3}$, concluimos que x^{k+1} está bien definido. Otra propiedad de la sucesión (x^k) es la acotación.

Lema 3.1. $x^k \in [-1, 2]$ *para todo* $k \geq 1$.

Demostración: Si k es impar, entonces $x^k < 0$. Además usando las condiciones iniciales tenemos que, $L'_{\rho_k}(-1, \bar{\mu}^k) = 1 - 3\bar{\mu}^k e^{2\rho_k} < 0$, lo que implica que $x^k > -1$, ya que se supone que x^k es el mínimo local más grande en $(-\infty, 0)$.

Ahora demostraremos que $x^k \leq 2$ para k par. Observe que

$$L'_{\rho_k}(2, \bar{\mu}^k) = 1 - 12\bar{\mu}^k e^{-7\rho_k},$$

sea $k > 1$ par, entonces

$$L'_{\rho_k}(2, \bar{\mu}^k) = 1 - 12\bar{\mu}^k e^{-7\rho_k} = 1 - 12e^{-7\rho_k} > 0,$$

□

La acotación de $\left(x^k\right)$ implica que la sucesión tiene al menos un punto límite en $[-1,0]$ y uno en $[1,2]$. En particular, tenemos $\rho_k \rightarrow \infty$, ya que de lo contrario cada punto límite de $\left(x^k\right)$ tendría que ser viable, por el paso 3 del algoritmo. Utilizando el teorema 3.2 (de convergencia), sabemos que cada punto límite viable de $\left(x^k\right)$ donde g satisface PAKKT-regular es un punto KKT. Claramente, el intervalo $[-1,0]$ consta solo de puntos inviables. Por otra parte, el intervalo $[1,2]$ consiste enteramente de puntos viables, y la condición PAKKT-regular se cumple en cada uno de estos puntos. Por lo tanto, la subsucesión de $\left(x^k\right)$ que consiste en k pares converge a $x=1$ que es la solución (y único punto estacionario) del problema de optimización .

Notese tambien que la condición de cuasi-normalidad es tambien verificada en $x=1$, luego por el corolario 3.2, la sucesión $\{\mu^k\}_{k\in K}$ es acotada, con K siendo el conjunto de números pares; este hecho puede ser tambien verificado directamente; en efecto si $k\in K$, de la definición de x^k como el mínimo local más pequeño en $(1,+\infty)$ implica que $0=L'_{\rho_k}(x^k,\bar{\mu}^k)=1-3\left(x^k\right)^2\bar{\mu}^k e^{\rho_k g(x^k)}=1-3\left(x^k\right)^2\mu^k$, lo que implica $\mu^k=1/\left(3\left(x^k\right)^2\right)\leq 1/3$.

4. El caso convexo. En el ejemplo dado en la sección anterior, observamos que la sucesión generada por el algoritmo puede llevarnos a puntos límite que no son viables para (P).

En el caso cuadratico se demuestra en [6] que todo punto límite x^* de la sucesión generada por el método de Lagrangiano aumentado es un punto KKT de la medida de inviabilidad

$$\|h(x)\|_2^2+\|g(x)_+\|_2^2\,.$$

Según nuestro conocimiento, no existe un resultado similar para el caso de la penalidad exponencial. En la demostración de este resultado puede verse que esta medida de inviabilidad se adapta mejor a la forma que tiene la penalidad cuadrática en comparación con otras funciones de penalidad. Creemos que ese resultado es una ventaja de la penalidad cuadrática sobre otro tipo de penalidades.

En [8], considerando (P) con $X=\{x\in\mathbb{R}^n: g(x)\leq 0\}$ y usando el resutado anterior demuestra que cuando x^* satisface la extendida condición de Mangazarian Fromovitz (EMFCQ) entonces x^* es viable. Por otro lado no es difícil ver que cuando g es convexa y $X\neq\emptyset$, el resultado anterior asegura también que x^* es viable, ya que $\|g(x)_+\|_2^2$ es convexa.

En esta sección veremos que cuando g tiene componentes convexas y $X=\{x\in\mathbb{R}^n: g(x)\leq 0\}$ todo punto límite de la sucesión generada por el Algoritmo 1 es viable.

Consideraremos el problema de optimización definido por:

$$\begin{array}{ll}(PC)\ \text{Minimizar} & f(x)\\ \quad\text{sujeto a} & g(x)\leq 0\end{array}.$$

donde $g:\mathbb{R}^n\rightarrow\mathbb{R}^p$, $f:\mathbb{R}^n\rightarrow\mathbb{R}$ son continuamente diferenciable en $\mathbb{R}^n$ y g es convexa.

Cuando el Algoritmo 1 es aplicado al problema (PC), podemos probar un resultado de viabilidad del punto límite de la sucesión generada por el método.

Teorema 4.1. *Sea $\left\{x^k\right\}$ una sucesión generada por el Algoritmo 1, que tiene un punto límite x^*. Entonces, x^* es un punto viable del problema (PC), es decir, $g_j\left(x^*\right)\leq 0$ para todo $j\in\{1,\cdots,p\}$.*

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

1. la sucesión $\{\rho_k\}$ es acotada.
2. la sucesión $\{\rho_k\}$ no está acotada, es decir, $\lim_{k\rightarrow\infty}\rho_k=\infty$.

Caso 1. Como $\{\rho_k\}$ está acotada, por el paso 3 del algoritmo, debe existir $\bar{k}$ tal que, para todo $k\geq\bar{k}$, tenemos $\|V^k\|\leq\tau\|V^{k-1}\|$. Esto significa que

$$\lim_{k\longrightarrow\infty}\|V^k\|=0,$$

entonces para todo $j=1,...,p$ tenemos que:

$$\lim_{k\rightarrow\infty}V_j^k=\lim_{k\rightarrow\infty}\frac{\mu_j^k-\bar{\mu}_j^k}{\rho_k}=0.$$

como $\bar{\mu}_j^k \geq u^{min} > 0$ para todo k y para todo $j \in \{1, \cdots, p\}$ tenemos:

$$\lim_{k\to\infty} \frac{e^{\rho_k g_j(x^k)} - 1}{\rho_k} = 0.$$

de donde obtenemos $\lim_{k\to\infty} e^{\rho_k g_j(x)} = 1$ que implica que

$$g_j(x^*) = \lim_{k\to\infty} g_j(x^k) = 0.$$

entonces x^* es viable.

Caso 2. Por el paso 1 del Algoritmo y tomando un punto $\widehat{x} \in X$ tenemos

$$\|\widehat{x} - x^k\| \left\| \nabla f(x^k) + \sum_{j=1}^{p} \bar{\mu}_j^k e^{\rho_k g_j(x^k)} \nabla g_j(x^k) \right\| \leq \|\widehat{x} - x^k\| \varepsilon_k.$$

y usando la desigualdad de Cauchy-shwarz tenemos que:

$$\nabla f(x^k)^T (\widehat{x} - x^k) \geq -\varepsilon_k \|\widehat{x} - x^k\| - \left(\sum_{j=1}^{p} \bar{\mu}_j^k e^{\rho_k g_j(x^k)} \nabla g_j(x^k) \right)^T (\widehat{x} - x^k). \tag{4.1}$$

Usando la convexidad de los g_j, tenemos que $g_j(\widehat{x}) \geq g_j(x^k) + \nabla g_j(x^k)^T (\widehat{x} - x^k)$, y como $-g(\widehat{x})$ es no negativo, podemos acotar el ultimo termino de (4.1):

$$\begin{aligned} -\left(\sum_{j=1}^{p} \bar{\mu}_j^k e^{\rho_k g_j(x^k)} \nabla g_j(x^k) \right)^T (\widehat{x} - x^k) &\geq \sum_{j=1}^{p} \bar{\mu}_j^k e^{\rho_k g_j(x^k)} \left(g_j(x^k) - g_j(\widehat{x}) \right) \\ &\geq \sum_{j=1}^{p} \bar{\mu}_j^k e^{\rho_k g_j(x^k)} g_j(x^k). \end{aligned}$$

Ahora, supongamos por contradicción que $\{j \in \{1, \cdots, q\} : g_j(x^*) > 0\} \neq \emptyset$.

Sea $a = \max\{g_j(x^*) : j = 1, ..., p\} > 0$. Usando (4.1) y dividiendo entre $e^{\frac{a}{2}\rho_k}$, obtenemos:

$$\frac{\nabla f(x^k)^T \left(\widehat{x} - x^k \right)}{e^{\frac{a}{2}\rho_k}} \geq -\frac{\varepsilon_k}{e^{\frac{a}{2}\rho_k}} \|\widehat{x} - x^k\| + \sum_{j=1}^{p} \bar{\mu}_j^k e^{\rho_k (g_j(x^k) - \frac{a}{2})} g_j(x^k), \tag{4.2}$$

dado que, $\rho_k \to \infty, \bar{\mu}^k \geq u^{min}$, y las g_j son continuas, tenemos:

$$\lim_{k\to\infty} \sum_{j: g_j(x^*)<0} \bar{\mu}_j^k e^{\rho_k (g_j(x^k) - \frac{a}{2})} g_j(x^k) = 0.$$

Además

$$\lim_{k\to\infty} \sum_{j: g_j(x^*)=0} \bar{\mu}_j^k e^{\rho_k (g_j(x^k) - \frac{a}{2})} g_j(x^k) = 0,$$

tenemos que $0 = \lim_{k\to\infty} \left(\nabla f(x^k)^T (\widehat{x} - x^k) / \ e^{\frac{a}{2}\rho_k} \right)$. Luego, usando la desigualdad (4.2), se tiene:

$$0 \geq \lim_{k\to\infty} \sum_{j: g_j(x^*)>0} \bar{\mu}_j^k e^{\rho_k (g_j(x^k) - \frac{a}{2})} g_j(x^k),$$

lo cual es absurdo. Luego, el conjunto$\{j \in \{1, \cdots, q\} : g_j\,(x^*) > 0\} = \emptyset$, i.e., x^* es viable. □

Finalmente, demostramos que un punto límite de la sucesión generada por el algoritmo es un punto estacionario en el sentido que:

$$\nabla f\,(x^*)^T\,(z - x^*) \geq 0 \quad \forall z \in X;$$

con una suposición adicional sobre los parámetros de penalización y precisión. Nótese que este contexto no pedimos el cumplimiento de aguna condición de calificación.

Proposición 4.1. *Sea $\{x^k\}$ una sucesión generada por el Algoritmo 1. Supongamos que la sucesión $\{\varepsilon^k \rho_k\}$ es acotada. Entonces, cada punto de acumulación x^* de $\{x^k\}$ es un punto estacionario del problema.*

Demostración: Sea $K \subseteq \{0, 1, \cdots\}$ un subconjunto infinito tal que:

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

Según el teorema anterior, tenemos $x^* \in X$. Supongamos por contradicción que x^* no es un punto estacionario entonces, existe $z \in X$ tal que:

$$\nabla f\left(x^*\right)^T \left(z - x^*\right) \geq 0,$$

de forma análoga como en la demostración del teorema anterior tenemos

$$\nabla f(x^k)^T (z - x^k) \geq -\varepsilon_k \|z - x^k\| + \sum_{j=1}^{m} \bar{\mu}_j^k e^{\rho_k g_j(x^k)} g_j(x^k). \tag{4.3}$$

Considere los términos $\bar{\mu}_j^k e^{\rho_k g_j(x^k)} g_j(x^k)$. Si $g_j(x^k) \geq 0$, entonces

$$\bar{\mu}_j^k e^{\rho_k g_j(x^k)} g_j(x^k) \geq 0.$$

De lo contrario, si $g_j(x^k) < 0$, tenemos

$$\bar{\mu}_j^k e^{\rho_k g_j(x^k)} g_j(x^k) \geq \bar{\mu}_j^k g_j(x^k),$$

luego tenemos que

$$\bar{\mu}_j^k e^{\rho_k g_j(x^k)} g_j(x^k) \geq \text{mín}\{0, \bar{\mu}_j^k g_j(x^k)\}.$$

Por lo tanto, de (4.3) tenemos:

$$\nabla f(x^k)^T (z - x^k) \geq -\varepsilon_k \|z - x^k\| + \sum_{j: g_j(x^*) < 0} \bar{\mu}_j^k e^{\rho_k g_j(x^k)} g_j(x^k) + \sum_{j: g_j(x^*) = 0} \text{mín}\{0, \bar{\mu}_j^k g_j(x^k)\}, \tag{4.4}$$

tomando limite en (4.4) tenemos :

$$\nabla f(x^*)^T (z - x^*) \geq 0,$$

lo cual es una contradicción. □

El siguiente resultado es consecuencia del teorema anterior y de resultados conocidos de optimización convexa.

Corolario 4.1. *En las condiciones del teorema anterior si ademas f es convexa, entonces x^* es solución global del problema (PC).*

5. Conclusiones.

Estudiamos un algoritmo lagrangiano aumentado basado en la función de penalidad exponencial para optimización con restricciones. Mejoramos la convergencia global del algoritmo a puntos estacionarios de primer orden bajo la condición de calificación : positiva aproximada KKT regular (PAKKT-regular). Probamos que bajo la condición de cuasinormalidad en el punto limite viable de la sucesión generada por el Algoritmo 1 la sucesión dual es acotada. Finalmente damos condiciónes suficientes para la viabilidad y optimalidad del punto límite, para el caso en que las restricciones son convexas.

Contribución de los Autores. F.N. Planteamiento del problema, metodologia de desarrollo y coordinación del trabajo. L. T. Recolección de datos, análisis y revisión de resultados.

Financiamiento. La presente investigación no ha recibido ningún tipo de financiamiento para su realización.

Conflicto de intereses. Los autores declaran no tener conflicto de intereses.

Agradecimientos. Queremos agradecer profundamente a los árbitros anónimos por expresar comentarios y sugerencias que permitieron mejorar este manuscrito.

ORCID and License

Frank Navarro Rojas https://orcid.org/0000-0002-9847-0822
Luis G. Ticona Quispe https://orcid.org/0000-0002-6209-9427

Referencias

[1] Powell M JD. A method for nonlinear constraints in minimization problems. Optimization. 1969; 283-298.

[2] Hestenes M R. Multiplier and gradient methods. Journal of optimization theory and applications. 1969; 4(5):303-320.

[3] Rockafellar R T. Augmented Lagrange multiplier functions and duality in nonconvex programming. SIAM Journal on Control. 1974; 12(2): 268-285.

[4] Fiacco A V, Garth P. McCormick. Nonlinear programming: sequential unconstrained minimization techniques. Society for Industrial and Applied Mathematics, 1990.

[5] Fletcher R. Practical methods of optimization. John Wiley and Sons, 2000.

[6] Birgin EG, Martínez JM. Practical augmented Lagrangian methods for constrained optimization. Society for Industrial and Applied Mathematics, 2014.

[7] Andreani R, et al. On augmented Lagrangian methods with general lower-level constraints. SIAM Journal on Optimization. 2008; 18(4): 1286-1309.

[8] Kanzow Ch, Steck D. An example comparing the standard and safeguarded augmented Lagrangian methods. Operations Research Letters. 2017; 45(6):598-603.

[9] Andreani R, et al. A cone-continuity constraint qualification and algorithmic consequences. SIAM Journal on Optimization. 2016; 26(1):96-110.

[10] Andreani R. et al. A sequential optimality condition related to the quasi-normality constraint qualification and its algorithmic consequences. SIAM Journal on Optimization. 2019; 29(1):743-766.

[11] Andreani R, et al. New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences. SIAM Journal on Optimization. 2019; 29(4):3201-3230.

[12] Izmailov AF, Solodov MV, Uskov EI. Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. SIAM Journal on Optimization. 2012; 22(4):1579-1606.

[13] Kansow Chr. On the multiplier-penalty-approach for quasi-variational inequalities. Mathematical Programming. 2016; 160: 33-63. Doi: 10.1007/s10107-015-0973-3

[14] Kanzow Chr, Steck D. Augmented Lagrangian methods for the solution of generalized Nash equilibrium problems. SIAM Journal on Optimization. 20216; 26(4):2034-2058.

[15] Bueno LF, Haeser G, Lara F, Rojas FN. An augmented Lagrangian method for quasi-equilibrium problems. Computational Optimization and Applications. 2020; 76: 737-766. Doi: 10.1007/s10589-020-00180-4

[16] Andreani R, Haeser G, Viana DS. Optimality conditions and global convergence for nonlinear semidefinite programming. Mathematical Programming. 2020; 180:203-235.

[17] Birgin EG, et al. An Augmented Lagrangian algorithm for nonlinear semidefinite programming applied to the covering problem. Comp. and Appl. Math. 2020; 39:1-21.

[18] Kanzow, Chr., Raharja AB, Schwartz A. An augmented Lagrangian method for cardinality-constrained optimization problems. J. of Optimization Theory and Appl. 2021; 189(3):793-813.

[19] Fazzio NS. Teoría y métodos para problemas de optimización multiobjetivo[Tesia de Doctorado]. Universidad Nacional de La Plata, 2018.

[20] Bueno LF, Haeser G, Lara F, Rojas FN. An augmented Lagrangian method for quasi-equilibrium problems. Computational Optimization and Appl. 2020; 76:737-766. Doi: 10.1007/s10589-020-00180-4

[21] Birgin EG, Castillo RA, Martínez JM. Numerical comparison of augmented Lagrangian algorithms for nonconvex problems. Comp. Optimization and Appl. 2005; 31(1):31-55.

[22] Tseng P, Bertsekas PD. On the convergence of the exponential multiplier method for convex programming. Mathematical programming. 1993; 60(1):1-19.

[23] Kort BW, Bertsekas DP. Multiplier methods for convex programming. In 1973 IEEE Conference on Decision and Control including the 12th Symposium on Adaptive Processes (pp. 428-432). IEEE. 1973.

[24] Kort BW, Bertsekas DP. Combined primal–dual and penalty methods for convex programming. SIAM Journal on Control and Optimization. 1976; 14(2):268-294.

[25] Bertsekas DP. Constrained optimization and Lagrange multiplier methods. Academic press, 2014.

[26] Birgin EG, Castillo RA, Martínez JM. ”Numerical comparison of augmented Lagrangian algorithms for nonconvex problems.Çomputational Optimization and Applications 31.1 (2005): 31-55.

[27] Iusem AN. Augmented Lagrangian methods and proximal point methods for convex optimization. Investigación Operativa; 1999; 8(7):11-49.

[28] Wang Ch-Yu, Duan L. Unified theory of augmented Lagrangian methods for constrained global optimization. J. of Global Optimization. 2009; 44: 433-458.

[29] Echebest N, Sánchez MD, and Schuverdt ML. Convergence results of an augmented Lagrangian method using the exponential penalty function. J. of Optimization Theory and Applications. 2016; 168:92-108.

[30] Ramirez LM. The hyperbolic augmented lagrangian algorithm (the carioca algorithm)[Thesis doutorado]. Universidade Federal do Rio de Janeiro, 2022.

[31] Sanchez MD, Schuverdt ML. A second-order convergence augmented Lagrangian method using nonquadratic penalty functions. Opsearch. 2019; 56: 390-408.

[32] Andreani R, et al. Two new weak constraint qualifications and applications."SIAM Journal on Optimization. 2012; 22(3):1109-1135.

[33] Andreani R, et al. A cone-continuity constraint qualification and algorithmic consequences. SIAM J. on Optimization. 2016; 26(1):96-110.

[34] Andreani R, Haeser G, Martínez JM. On sequential optimality conditions for smooth constrained optimization. Optimization; 2011; 60(5):627-641.

[35] Bertsekas DP. Nonlinear programming. J. of the Operational Research Society. 1997; 48(3): 334-334.

[36] Andreani R, et al. On the best achievable quality of limit points of augmented Lagrangian schemes. Numerical Algorithms. 2022; 1-27.

[37] Andreani R, et al. On scaled stopping criteria for a safeguarded augmented Lagrangian method with theoretical guarantees. Math. Programming Comp. 2022; 14(1):121-146.

[38] Andreani R, et al. On the convergence of augmented Lagrangian strategies for nonlinear programming. IMA J. of Numerical Analysis. 2022; 42(2):1735-1765.

[39] Prado RW, Santos SA, Simões LEA. A novel sequential optimality condition for smooth constrained optimization and algorithmic consequences. Optimization. 2024; 73(5): 1447-1476.

[40] Andreani R, Martínez JM, Svaiter BF. A new sequential optimality condition for constrained optimization and algorithmic consequences. SIAM J. on Optimization. 2010; 20(6):3533-3554.

[41] Andreani R, et al. Strict constraint qualifications and sequential optimality conditions for constrained optimization. Mathematics of Operations Research, 2018; 43(3):693-717.

[42] Qi L, Wei Z. On the constant positive linear dependence condition and its application to SQP methods. SIAM J. on Optimization. 2000; 10(4):963-981.

[43] Andreani R, et al. A relaxed constant positive linear dependence constraint qualification and applications. Mathematical Programming. 2012; 135(1):255-273.

[44] Andreani R, Martínez JM, Svaiter BF. A new sequential optimality condition for constrained optimization and algorithmic consequences. SIAM J. on Optimization. 2010; 20(6):3533-3554.

[45] Haeser G, Schuverdt ML. On approximate KKT condition and its extension to continuous variational inequalities. J. of Optimization Theory and Appl. 2011; 149(3):528-539.

[46] Martínez JM, Svaiter BF. A practical optimality condition without constraint qualifications for nonlinear programming. J. of Optimization Theory and Appl. 2003; 118:117-133.

[47] Rockafellar RT, Wets RJ-B. Variational analysis. Vol. 317. Springer Science and Business Media, 2009.

[48] Sánchez MD. Métodos de Lagrangiano Aumentado basados en funciones de penalidad no cuadráticas[Tésis Doctorado]. Universidad Nacional de La Plata, 2017.