PESQUIMAT, Revista de la F.C.M. de la
Universidad Nacional Mayor de San Marcos
Vol. XIV N°2, pp. 25-43, Lima - Perú, Agosto 2011

UN MÉTODO MULTIPLICADOR PROXIMAL PARA OPTIMIZACIÓN CONVEXA SEPARABLE

Erik Papa[*] & *Orlando Sarmiento*[**]

Resumen: El objetivo de este trabajo es establecer la convergencia de un método multiplicador proximal utilizando distancias generalizadas para resolver problemas de minimización convexa con estructura separable, motivados en particular en la solución de problemas de optimización de gran tamaño que surgen en redes de telecomunicaciones y en la gestión de producción de energía eléctrica. Los procedimientos utilizados fueron la recopilación de información en revistas científicas y textos especializados, el estudio de los mismos y finalmente el uso de herramientas matemáticas para estudiar la convergencia de la sucesión del método propuesto. Los resultados del estudio nos muestran que, bajo algunas hipótesis adecuadas, las iteraciones generadas por el método están bien definidas y la sucesión converge a una solución óptima del problema. Por la generalidad del estudio resultan casos particulares algunos trabajos de investigación relacionados con métodos proximales, como por ejemplo, el de Chen y Teboulle (1994), Kyono y Fukushima (2000) y Auslender y Teboulle (2001).

Palabras clave: Métodos proximales, problemas convexos separables, ortante no negativo, distancia generalizada, funciones convexas.

MULTIPLIER METHOD FOR PROXIMAL CONVEX OPTIMIZATION SEPARABLE

Abstract: The aim of this work is to prove the convergence of a proximal multiplicator method using generalized distances to solve convex minimization problems with separable structure, motivated in particular by the solution of optimization problems that arising in telecommunication networks and management of electrical energy production. The used procedures were the collection of information in scientific journals and specialized books, the study of the same and finally the use of mathematical tools to study the convergence of the sequence of the proposed method. The results show that, under some appropriate assumptions, the iterations generated by the method are well defined and the sequence converges to an optimal solution of the problem. Due to the generality of the study some papers related to proximal methods such as the works of Chen and Teboulle (1994), Kyono and Fukushima (2000) and Auslender and Teboulle (2001) are particular cases of our approach.

Key words: Proximal point methods, separable convex problems, nonnegative orthant, proximal distances, convex functions.

1. Introducción

En los problemas de ruteamiento (transporte de diversos servicios a diferentes clientes) en las redes de telecomunicaciones y de la gestión de producción de energía eléctrica aparecen problemas de optimización de gran dimensión y complejidad que los softwares actuales tienen dificultad para resolverlos.

Existen varios estudios de como afrontar esta problemática, uno de estos, y que aparentemente es el más atractivo por su naturaleza y simplicidad, es usar una metodología de descomposición de los problemas de optimización en problemas más acsesibles y que puedan ser resueltos paralelamente en

[*] UNMSMS, Facultad de Ciencias Matemáticas, e-mail: erikpapa@gmail.com
[**] UNAC, e-mail: orlando.unac@gmail.com

tiempos razonables por algunos métodos conocidos. Esta idea se remonta al trabajo de Dantzig y Wolfe (1960), [?] quienes fueron los pioneros de los métodos de descomposición para optimización lineal usando el método simplex.

Lamentablemente no todo problema es lineal, como sucede en los casos mencionados anteriormente, y aunque existen técnicas de linealización puede suceder que al aplicarlas se pierdan algunas propiedades fundamentales e intrinsecas que conduscan a soluciones equivocadas del problema original. Esto permitió desarrollar abordages no lineales de métodos de descomposición.

Probablemente, el primer trabajo relacionado a un método de descomposición para problemas de optimización no lineal convexo se remonta a Uzawa (1958), [?]. El autor introdujo un método de descomposición de gradiente aplicado al problema dual pero lamentablemente las condiciones de convergencia exigian que la función objetivo cumpla condiciones muy restrictivas para las aplicaciones (estricta convexidad). Apartir de entonces se han desarrollado varias investigaciones para mejorar las propiedades de convergencia y eficiencia computacional.

Actualmente, es bien conocido que una importante clase de métodos de descomposición para optimización convexa pueden ser considerados como casos particulares del método del punto proximal para encontrar un cero de un operador monótono maximal, ver Eckstein y Bertsekas (1992), [?].

El método del punto proximal es uno de los métodos más estudiados e importantes para resolver problemas de optimización, de singularidades, problemas de desigualdad variacional, de equilibrio, entre otros. Este método fue introducido por Martinet (1970),[?], motivado por el trabajo de Moreau (1962, 1965) [?] [?], y fue extensamente estudiado por Rockafellar [?] en los años setenta del siglo anterior para encontrar ceros de operadores monótonos maximales.

Para el problema de minimización

$$\min\{f(x) : x \in H\}, \tag{1}$$

donde H es un espacio de Hilbert y $f : H \to \mathbb{R} \cup \{+\infty\}$ es una función propia convexa y semicontinua inferior, el método del punto proximal, para resolver el problema (??) genera una sucesión de puntos $\{x^k\}$ definidos por

$$x^0 \in H, \tag{2}$$

$$x^k = \mathbf{arg}\min\{f(x) + (\lambda_k/2)||x - x^{k-1}||^2 : x \in H\}. \tag{3}$$

donde λ_k es algún parametro positivo. Observe que la notación $\mathbf{arg}\min$ de la ecuación anterior es bien utilizada en la comunidad de optimización y significa que x^k es el mínimo (global) de $f(\cdot) + (\lambda_k/2)||\cdot -x^{k-1}||^2$ sobre H.

Los resultados de convergencia del método para el caso convexo son muy bien conocidos y puede ser probado, ver Guler (1991), [?], que si $\{\lambda_k\}$ satisface $\sum_{k=1}^{+\infty}(1/\lambda_k) = +\infty$, entonces $\lim_{k\to\infty} f(x^k) = \inf\{f(x) : x \in M\}$. Además, si la solución existe la sucesión $\{x^k\}$ converge (débilmente) a un punto de mínimo. En la actualidad se están realizando investigaciones para extender las aplicaciones del método para funciones no necesariamente convexas, motivados por las múltiples aplicaciones en Economía.

La importancia del método del punto proximal radica principalmente en dos aspectos: se puede probar que los métodos de multiplicadores, usado generalmente en todos los softwares de optimización coinciden con el método del punto proximal aplicado al dual del problema original y es así que la teoría de convergencia del método se puede utilizar para garantizar la convergencia de los métodos de multiplicadores que directamente e independientemente es complicado de demostrar. El otro aspecto es con respecto a los métodos de descomposición, Eckstein y Bertsekas (1992), [?], han demostrado que una gran clase de métodos de descomposición de problemas convexos son casos particulares del método del punto proximal para encontrar un cero de un operador monótono maximal, especificamente el método de Douglas-Rachford estudiado por Lions y Mercier (1979), [?], que abarca una gran clase de algoritmos de optimización es una versión particular del método del punto proximal.

Entre los métodos de descomposición que usan técnicas proximales podemos resaltar los siguientes algoritmos: método de multiplicadores de direcciones alternadas [?, ?, ?], el método de inverso parcial de spingarn [?, ?, ?, ?, ?] y metodos de multiplicadores proximales predictor-corrector de Chen y Teboulle y sus extensiones [?, ?, ?].

La motivación de este trabajo es estudiar la última clase de problemas, es decir, el método de multiplicadores proximales predictor corrector. Considere el problema de minimización convexa con estructura separable:

$$(PCS) \quad \text{mín}\{f(x) + g(z) \, : \, Ax + Bz = b, x \in \mathbb{R}^n_+, \; z \in \mathbb{R}^p\},$$

donde $f : \mathbb{R}^n \to (-\infty, +\infty]$ y $g : \mathbb{R}^p \to (-\infty, +\infty]$ son funciones propias convexas cerradas y $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times p}$ y $b \in \mathbb{R}^m$.
El problema dual Fenchel para (PCS) es definido por:

$$(D) \quad \text{mín}\{(f + \delta_{\mathbb{R}^n_+})^*(-A^T y) + g^*(-B^T y) : y \in \mathbb{R}^m\}$$

donde $(f + \delta_{\mathbb{R}^n_+})^*$ y g^* son las funciones conjugadas de $f + \delta_{\mathbb{R}^n_+}$ y g, respectivamente, y y es un vector multiplicador de Lagrange asociado con las restricciones en (PCS). El lagrangiano $L : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \to (-\infty, +\infty]$ para el problema (PCS) es definido por:

$$L(x, z, y) = (f + \delta_{\mathbb{R}^n_+})(x) + g(z) + \langle y, Ax + Bz - b \rangle .$$

Si (x^*, z^*) es una solución óptima de (CP) y y^* es el correspondiente vector multiplicador de Lagrange, entonces y^* es una solución óptima de (D) y (x^*, z^*, y^*) es un punto de silla de L (ver [?], Teorema 28.3); es decir, para todo $x \in domf \cap \mathbb{R}^n_+$, $z \in domg$ y $y \in \mathbb{R}^m$, se tiene que

$$L(x^*, z^*, y) \leq L(x^*, z^*, y^*) \leq L(x, z, y^*). \tag{4}$$

Para el caso cuando $\mathbb{R}^n_+$ es sustituido por $\mathbb{R}^n$, $B = -I$ y $b = 0$, Chen y Teboulle [?] presentaron el siguiente método el cual llamaron método multiplicador proximal predictor-corrector: dado (x^0, z^0, y^0) arbitrario el método en su versión exacta genera una sucesión de puntos $\{(x^k, z^k, y^k)\}$ dado por las siguientes iteraciones:

$$p^{k+1} = y^k + \lambda_k(Ax^k - z^k)$$

$$x^{k+1} = \arg\text{mín}\left\{f(x) + \langle p^{k+1}, Ax \rangle + \frac{1}{2\lambda_k}||x - x^k||^2\right\} \tag{5}$$

$$z^{k+1} = \arg\text{mín}\left\{g(z) - \langle p^{k+1}, z \rangle + \frac{1}{2\lambda_k}||z - z^k||^2\right\} \tag{6}$$

$$y^{k+1} = y^k + \lambda_k(Ax^{k+1} - z^{k+1}).$$

Asumiendo la hipótesis de que existe una solución óptima primal-dual (x^*, y^*, z^*) Chen y Teboulle [?] probaron que la sucesión $\{x^k\}$ converge a x^*, $\{z^k\}$ converge a Ax^* y $\{y^k\}$ converges to y^*.
Kyono y Fukushima (2000) [?], consideraron reemplazar los términos cuadráticos de (??) y (??) por distancias más generales dadas por funciones de Bregman, es así que los autores introdujeron las siguientes iteraciones para reemplazar respectivamente (??) y (??):

$$x^{k+1} = \arg\text{mín}\left\{f(x) + \langle p^{k+1}, Ax \rangle + \frac{1}{\lambda_k}D_h(x, x^k)\right\}$$

$$z^{k+1} = \arg\text{mín}\left\{g(z) - \langle p^{k+1}, z \rangle + \frac{1}{\lambda_k}D_h(z, z^k)\right\}$$

Bajo las mismas condiciones de Chen y Teboulle [?] y algunas hipótesis sobre h los autores consiguieron obtener los mismos resultados de convergencia.
Paralelamente al trabajo de Kyono y Fukushima (2000), Auslender y Teboulle (2001), [?] susttituyeron sólo el término cuadrático de (??) por una distancia log-cuadrática motivados por un trabajo anterior de Auslender, Teboulle y Ben-Tiba (1999), [?], esto es, el término cuadrático fué sustituido por:

$$x^{k+1} = \arg\text{mín}\left\{f(x) + \langle p^{k+1}, Ax \rangle + \frac{1}{\lambda_k}d(x, x^k)\right\}$$

donde

$$d(u,v) = \begin{cases} \sum_{i=1}^{n} \frac{\sigma}{2}(u_i - v_i)^2 + \mu(v_i^2 \log(\frac{v_i}{u_i}) + u_i v_i - v_i^2) & \text{si } u \in \mathbb{R}^n_{++} \\ +\infty & \text{caso contrario} \end{cases}$$

Algunas extensiones de estos trabajos se han desarrollado para problemas más generales como complementaridad no lineal, desigualdades variacionales, entre otras.

El objetivo de este trabajo es extender las propiedades de convergencia del método multiplicador proximal predictor corrector de Chen y Teboulle [?] para el problema (PCS) usando una pequeña variación de las distancias proximales que fueron introducidas en Auslender y Teboulle(2006), [?], y que llamamos distancias generalizadas, cuya definición extiende las distancias anteriormente comentadas e incluye también otra clase de distancias llamadas $\phi-$ divergencias y que aún no han sido estudiadas en este contexto.

El artículo está organizado de la siguiente manera: En la sección ?? presentamos la metodología empleada en el estudio de investigación. En la sección ?? mostramos los resultados del trabajo, lo cuales están divididas en subsecciones que detallamos a seguir: en la subsección ?? damos una motivación del modelo (PCS), en la subsección ?? recordamos algunos elementos básicos de análisis convexo, en la subsección ?? introducimos la definición de distancias generalizadas, en la subsección ?? presentamos el método propuesto para resolver el problema (PCS), damos las hipótesis y demostramos la buena definición y la convergencia de las iteraciones a la solución del problema. En la sección ?? presentamos las discusiones del trabajo. En la sección ?? los agradecimientos respectivos y finalmente presentamos las referencias bibliográficas empleadas para esta investigación.

2. Metodología

La metodología de trabajo fue dividida en las siguientes partes:

2.1. El universo

El universo de nuestra investigación es la optimización convexa en espacios euclidianos. La optimización convexa es una línea de la matemática aplicada que estudia el problema de hallar los valores mínimos o máximos de funciones convexas sujetas a restricciones que forman un conjunto convexo.

2.2. Técnicas de recopilación de datos

Para el inicio de la investigación fue necesario una búsqueda de información científica en revistas publicadas, uso de bibliotecas y hemerotecas especializadas como también viajes a centros de investigación en el extranjero como Brasil y finalmente Francia.

Debido que en el país la mayoria de universidades no estan indexadas a las revistas internacionales se torna imposible obtener información apropiada para desarrollar una primera etapa en la investigación, es así como el primer autor visitó el Programa de Engenharia de Sistemas e Computação (PESC) de la COPPE-Universidad Federal de Rio de Janeiro y el Instituto de Matemática Pura y aplicada (IMPA).

2.3. Técnicas descriptivas para la contrastación o demostración de las hipótesis

La metodología usada en esta etapa fue un enfoque de tipo inductivo-deductivo de los resultados de recientes investigaciones, con la finalidad de adaptarlo a nuestros requerimientos y que nos llevó a la obtención de los resultados planteados en los objetivos de la investigación.

También se participó como expositores en algunos eventos nacionales e internacionales como por ejemplo El Encuentro Cientifico Internacional (ECI 2012v) del 3 al 6 de enero del 2012 como también en el II Workshop Internacional de Matemáticas: Fundamentos y aplicaciones, desarrollado del 4 al 6 de enero en la Universidad Nacional del Callao.

Finalmente, el primer autor realizó una visita al Laboratorio LIMOS-ISIMA de la Universidad Blaise Pascal, Clermont-Ferrand, Francia para una contrastación del método propuesto, demostrandonos que este abordage no ha sido desarrollado por otros investigadores y que nos dá un campo de investigación para futuros trabajos.

3. Resultados

3.1. Un Modelo de Redes en Telecomunicaciones

En esta sección, para motivar el problema estudiado, presentamos un modelo de flujo de bienes múltiples y lo expresamos como un problema del tipo (PCS). Para lectores interesados en una revisión de modelos y métodos de flujo de bienes múltiples recomendamos la lectura de [?, ?, ?, ?].

Los problemas de flujo de bienes múltiples pertenecen a una clase de modelos de la optimización en grafos y se refiere a la distribución de varios bienes o productos en una red (finita) común y que deben satisfacer los siguientes requisitos:

a. Los bienes deben ser indivisibles, es decir a lo largo de la red no se debe dividir el bien.

b. Los bienes deben originarse en nodos-fuente específicos, y deben llegar a nodos-destino determinados.

El modelo matemático de tráfico de problemas de flujo de bienes múltiples se puede expresar de la siguiente forma: Consideremos un grafo dirigido $G(V, E)$ donde V es el conjunto de nodos y E es el conjunto de arcos, supongamos que el grafo tiene m nodos, n arcos y capacidades c_u sobre cada arco u. Un arco del nodo i al nodo j será denotado por (i, j) o alternativamente por μ donde $\mu = 1, 2, ..., m$. Por ejemplo, la figura ?? representa un grafo dirigido de 4 nodos y 5 arcos donde $V = \{1, 2, 3, 4\}$,

$$E = \{(1, 2), (2, 3), (3, 4), (4, 1), (4, 2)\} = \{\mu_1, \mu_2, \mu_3, \mu_4, \mu_5\},$$

y las capacidades sobre cada arco son $c_{(1,2)} = c_{\mu_1} = 4$, $c_{(2,3)} = c_{\mu_2} = 3$, $c_{(3,4)} = c_{\mu_3} = 7$, $c_{(4,1)} = c_{\mu_4} = 1$, $c_{(4,2)} = c_{\mu_5} = 5$. Para modelar el tráfico requerido supongamos que K es el número de bienes que se

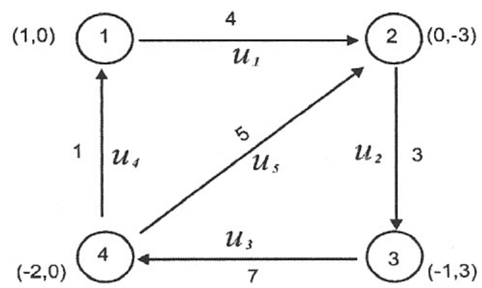

Figura 1: Grafo dirigido de 4 nodos y 5 arcos

desea trasladar a travéz del grafo y $k = 1, 2, ..., K$. Por cada bien k se tiene un nodo fuente s_k y un nodo destino t_k llamado también un $k-$par (s_k, t_k). Estos datos son exogenos al problema, es decir, son datos dados por el usuario. Para el ejemplo de la grafo consideremos que tenemos dos bienes, esto es, $K = 2$; para el bien $k = 1$ consideremos que el nodo fuente es el nodo 1 y el nodo destino es el nodo 3 y para el bien $k = 2$ consideremos que el nodo fuente es 3 y el nodo destino es 2; en este caso tenemos respectivamente que $(s_1, t_1) = (1, 3)$ y $(s_2, t_2) = (3, 2)$.

Para cada k sea:

$$x_{ij}^k = \text{ el flujo del bien } k \text{ sobre el arco } (i, j)$$

$$d^k = \text{ el valor del flujo a ser enviado de } s_k \text{ a } t_k$$

$$x_{ij} = \sum_{k=1}^{K} x_{ij}^k \text{ el total del flujo sobre el arco } (i,j).$$

Para cada $i \in N$ sea

$$A(i) = \{j \in V : (i,j) \in E\} \text{ y } B(i) = \{j \in V : (j,i) \in E\}.$$

En nuestro ejemplo tenemos:

$$\begin{gathered}
x_{12}^1 \text{ el flujo del bien 1 sobre el arco } (1,2) \\
x_{23}^1 \text{ el flujo del bien 1 sobre el arco } (2,3) \\
x_{34}^1 \text{ el flujo del bien 1 sobre el arco } (3,4) \\
x_{41}^1 \text{ el flujo del bien 1 sobre el arco } (4,1) \\
x_{42}^1 \text{ el flujo del bien 1 sobre el arco } (4,2) \\
x_{12}^2 \text{ el flujo del bien 2 sobre el arco } (1,2) \\
x_{23}^2 \text{ el flujo del bien 2 sobre el arco } (2,3) \\
x_{34}^2 \text{ el flujo del bien 2 sobre el arco } (3,4) \\
x_{41}^2 \text{ el flujo del bien 2 sobre el arco } (4,1) \\
x_{42}^2 \text{ el flujo del bien 2 sobre el arco } (4,2) \\
d^1 = 1 \text{ el valor del flujo a ser enviado de } s_1 = 1 \text{ a } t_1 = 3 \\
d^2 = 3 \text{ el valor del flujo a ser enviado de } s_2 = 3 \text{ a } t_2 = 2. \\
x_{12} = x_{12}^1 + x_{12}^2 \\
x_{23} = x_{23}^1 + x_{23}^2 \\
x_{34} = x_{34}^1 + x_{34}^2 \\
x_{41} = x_{41}^1 + x_{41}^2 \\
x_{42} = x_{42}^1 + x_{42}^2 \\
A(1) = \{2\}, A(2) = \{3\}, A(3) = \{4\}, A(4) = \{1,2\} \\
B(1) = \{4\}, B(2) = \{1,4\}, B(3) = \{2\}, B(4) = \{3\}
\end{gathered}$$

Observe que en la gráfica las demandas de cada bien son representados por el vector (a,b) del lado de la gráfica del nodo, en este caso a es la demanda del bien 1 y b es la del bien 2, si a es positivo quiere decir que este nodo es nodo fuente y si es negativo quiere decir que es un nodo destino, de forma análoga el valor de b representa la demanda del bien 2.

Un flujo de bienes múltiples es un conjunto de flujos de arcos $\{x_{ij}^k : (i,j) \in E, k = 1, ..., K\}$ satisfaciendo:

$$\sum_{j \in A(i)} x_{ij}^k - \sum_{j \in B(i)} x_{ij}^k = \begin{cases} d^k, & \text{si } i = s^k \\ -d^k, & \text{si } i = t^k \\ 0, & \text{en otros casos} \end{cases}$$

$$x_{ij}^k \geq 0, \forall (i,j)$$

Eligiendo una enumeración de nodos y arcos de E podemos definir el vector $x^k = (x_1^k, x_2^k, ..., x_m^k)$ donde x_μ^k es el flujo del bien k sobre el arco μ, $\mu = 1, ..., m$. Con esta notación podemos establecer la matriz de incidencia nodo-arco el cual será denotada por M.
En el ejemplo dado tenemos

$$x^1 = (x_1^1, x_2^1, x_3^1, x_4^1, x_5^1) = (x_{\mu_1}^1, x_{\mu_2}^1, x_{\mu_3}^1, x_{\mu_4}^1, x_{\mu_5}^1)$$

y la matriz de incidencia nodo-arco es:

$$M = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 1 \end{bmatrix}$$

Entonces el problema general de bienes múltiples puede formularse como:

$$\begin{cases} \text{mín } f(x^1, \ldots, x^K) \\ sa: \\ \quad x^k \in \Omega_k, \text{ para } k = 1, 2, ..., K \\ \quad g(x^1, \ldots, x^k) \leq 0 \end{cases}$$

donde $f : \mathbb{R}^{Kn} \to \mathbb{R}$ es una función (posiblemente no lineal), $g : \mathbb{R}^{Kn} \to \mathbb{R}$ es la función de restricciones impuestas en los flujos y Ω_k es el conjunto de vectores de flujo para la demanda k satisfaciendo la conservación de flujo:

$$\Omega_k = \{x^k \in \mathbb{R}^n \; : \; Mx^k = D^k, 0 \leq x^k\},$$

M es la matriz de incidencia del grafo,

$$D^k = (0, \ldots, d^k, \ldots, -d^k, \ldots, 0),$$

y d^k es la cantidad requerida de flujo entre el k-ésimo par fuente-destino.
Para el ejemplo dado el problema de flujo de bienes múltiples queda expresado de la siguiente forma:

$$\begin{cases} \text{mín } f(x^1, \ldots, x^K) \\ sa: \\ \begin{pmatrix} 1 & 0 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1^1 \\ x_2^1 \\ x_3^1 \\ x_4^1 \\ x_5^1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} \\ \begin{pmatrix} 1 & 0 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1^2 \\ x_2^2 \\ x_3^2 \\ x_4^2 \\ x_5^2 \end{pmatrix} = \begin{pmatrix} 0 \\ -3 \\ 3 \\ 0 \end{pmatrix} \\ g(x^1, x^2) \leq 0 \end{cases}$$

La mayoria de modelos no lineales de tráfico de flujo de bienes múltiples tienen la forma de

$$\begin{cases} \text{mín } f(x) = \sum_{j=1}^{n} f_j(x_{0j}) + \sum_{k=1}^{K} \sum_{j=1}^{n} f_{kj}(x_j^k) \\ sa: \\ \quad x_{0j} = \sum_{k=1}^{K} x_j^k, \quad \text{para } j = 1, 2, ..., n \\ \quad Mx^k = D^k, \quad \text{para } k = 1, 2, ..., K \\ \quad 0 \leq x_k, \quad \text{para } k = 1, 2, ..., K \\ \quad 0 \leq x_{0j} \leq c_j, \quad \text{para } j = 1, 2, ..., n \end{cases}$$

Entre los modelos relacionados a ruteamiento de datos en redes, el problema de ruteamiento de mensajes juega un papel importante en la optimización de redes. Este problema consiste en la determinación de conjunto de rutas sobre el cual los paquetes van ha transitar de tal manera de optimizar algún costo que mide la calidad global del servicio. Los especialistas en esta área usan en general la función de demora media del mensaje bajo tráfico que fue introducida por Kleinrock (1972), [?], definida por

$$f_j(x_{0j}) = \frac{x_{0j}}{c_j - x_{0j}}, \text{ y } f_{kj}(x_j^k) = 0$$

donde $c_j > x_{0j}$.

Así, el problema de envio de mensajes bajo tráfico usando la función de Kleinrock es establecida por:

$$\begin{cases} \text{mín} f(x) = \sum\limits_{j=1}^{n} \frac{x_{0j}}{c_j - x_{0j}} \\ sa: \\ \quad x_{0j} = \sum\limits_{k=1}^{K} x_j^k, \quad \text{para } j = 1, 2, ..., n \\ \quad Mx^k = D^k, \quad \text{para } k = 1, 2, ..., K \\ \quad 0 \leq x_k, \quad \text{para } k = 1, 2, ..., K \\ \quad 0 \leq x_{0j} \leq c_j, \quad \text{para } j = 1, 2, ..., n \end{cases}$$

Definiendo $z = x_{0j}$, $y = (x^1, x^2, ..., x^K)$ y

$$F(y, z) = \begin{cases} \sum\limits_{j=1}^{n} \frac{z_j}{c_j - z_j} & \text{si } My^k = D^k, k = 1, ..., K; 0 \leq z_j < c_j, j = 1, ..., n \\ +\infty, & \text{caso contrario} \end{cases}$$

el problema se puede expresar como (PCS) donde A es el vector fila compuesta por elementos todos de valor 1 y $B = -I$ (negativo de la matriz identidad) y $b = 0$.

3.2. Elementos Básicos del Análisis Convexo

A lo largo del artículo $\mathbb{R}^n$ es el espacio euclidiano dotado con el producto interno canónico $\langle , \rangle$ y la norma de x dado por $||x|| := \langle x, x \rangle^{1/2}$. Dada una función a valores reales extendidos $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ denotaremos su dominio por $\mathtt{domf} := \{x \in \mathbb{R}^n : f(x) < +\infty\}$ y su epigrafo por $\mathtt{epi}f := \{(x, \beta) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq \beta\}$. f es denominado propia, si $\mathtt{domf} \neq \emptyset$ y $\forall x \in \mathtt{domf}$ se tiene que $f(x) > -\infty$. También denotemos por $r.i.(X)$ al interior relativo del conjunto $X \subset \mathbb{R}^n$.
Finalmente, f es semicontinua inferior si para cada $x \in \mathbb{R}^n$ tenemos que toda sucesión $\{x^l\}$ tal que $\text{lím}_{l \to +\infty} x^l = x$ implica que $f(x) \leq \text{lím inf}_{l \to +\infty} f(x^l)$. Podemos probar fácilmente que la semicontinuidad inferior de f es equivalente a que el conjunto de nivel inferior $L_f(\alpha) = \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$ es cerrado, para cada $\alpha \in \mathbb{R}$.

Definición 3.1. *Sea C un conjunto convexo no vacio, el cono recesión de C es definido por*

$$0^+(C) = \{z \in \mathbb{R}^n : w + tz \in C, \ \forall w \in C, t \geq 0\},$$

Además, definimos la función de recesión $f0^+ : \mathbb{R}^n \to \overline{\mathbb{R}}$ tal que $\boldsymbol{epi}(f0^+) = 0^+(\boldsymbol{epi}f)$

Definición 3.2. *Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ una función propia, la conjugada de f, denotado por f^*, es definido por*

$$f^*(x^*) = \sup_x [\langle x, x^* \rangle - f(x)].$$

Proposición 3.1. *Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ una función propia convexa y cerrada. Sea x^* un vector fijo y sea $g(x) = f(x) - \langle x, x^* \rangle$. Entonces, $x^* \in \boldsymbol{int}(domf^*)$ si y sólo si $(g0^+)(y) > 0$ para cada $y \neq 0$.*

Prueba. Ver [?], Corolario 13.3.4, **c**, pg. 117. ∎

Proposición 3.2. *Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ una función propia convexa y cerrada. El conjunto de minimizadores de f es no vacio y acotado si y sólo si $0 \in \boldsymbol{int}(domf^*)$.*

Prueba. Ver [?], Teorema 27.1, **d**, pg. 265. ∎

Proposición 3.3. *Sean $f_1, \ldots, f_m$ funciones propias convexas en $\mathbb{R}^n$. Si cada f_i es cerrada y $f_1 + \cdots + f_m$ no es identicamente $+\infty$, entonces $f_1 + \cdots + f_m$ es una función propia convexa cerrada y*

$$(f_1 + \cdots + f_m)0^+ = f_1 0^+ + \cdots + f_m 0^+.$$

Prueba. Ver [?], Teorema 9.3, pg. 77. ■

Proposición 3.4. *Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ una función propia convexa. La función de recesión $f0^+$ de f es una función propia, positivamente homogénea y convexa. Para cada vector $y \in$ **dom** f, se tiene que*

$$(f0^+)(y) = \sup\{f(x+y) - f(x) \,:\, x \in dom f\}.$$

Si f es cerrada, $f0^+$ es también cerrada y para cualquier $x \in dom f$, $f0^+$ es dado por la fórmula:

$$(f0^+)(y) = \sup_{\lambda>0} \frac{f(x+\lambda y) - f(x)}{\lambda} = \lim_{\lambda\to+\infty} \frac{f(x+\lambda y) - f(x)}{\lambda}.$$

Prueba. Ver [?], Teorema 8.5, pg. 66. ■

Proposición 3.5. *Sean $f_1, \ldots, f_m$ funciones propias convexas en $\mathbb{R}^n$ y sea $f = f_1 + \cdots + f_m$. Si los conjuntos convexos $ri(dom f_i)$, $i = 1, 2, \ldots, m$ tienen un punto en común, entonces*

$$\partial f(x) = \partial f_1(x) + \cdots + \partial f_m(x), \quad \forall x.$$

Prueba. Ver [?], Teorema 23.8, pg. 223. ■

3.3. Distancia Generalizada

En esta subsección presentamos una variante de la definición de distancia proximal y distancia proximal inducida, introducida por Auslender y Teboulle [?], pero adaptado para el conjunto $\mathbb{R}^n_+$ que es la restricción del problema convexo (PCS).

Definición 3.3. *Una función $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ es denominada distancia proximal en $\mathbb{R}^n_{++}$ si para cada $y \in \mathbb{R}^n_{++}$ se satisface las siguientes propiedades:*

i. *$d(\cdot, y)$ es propia, semicontinua inferior, estrictamente convexa y continuamente diferenciable en $\mathbb{R}^n_{++}$;*

ii. **dom** *$d(\cdot, y) \subset \mathbb{R}^n_+$ y* **dom** *$\partial_1 d(\cdot, y) = \mathbb{R}^n_{++}$, donde $\partial_1 d(\cdot, y)$ denota el clásico subdiferencial de la función $d(\cdot, y)$ con respecto a la primera variable;*

iii. *$d(\cdot, y)$ es coerciva en $\mathbb{R}^n$ (es decir, $\lim_{\|u\|\to\infty} d(u, y) = +\infty$).*

iv. *$d(y, y) = 0$.*

Denotamos por $D(\mathbb{R}^n_{++})$ a la familia de funciones satisfaciendo esta definición.

La propiedad **i.** es necesaria para preservar la convexidad de $d(\cdot, y)$, la propiedad **ii.** forzará que la iteración del método proximal permanezca en $\mathbb{R}^n_{++}$, y la propiedad **iii.** será usada para garantizar las existencias de las iteraciones proximales del método propuesto. Para cada $y \in \mathbb{R}^n_{++}$, $\nabla_1 d(\cdot, y)$ denota el gradiente de la función $d(\cdot, y)$ con respecto a la primera variable. Notemos que por definición $d(\cdot, \cdot) \geq 0$ y de **iv.** el mínimo global de $d(\cdot, y)$ se obtiene en y, y así tenemos que $\nabla_1 d(y, y) = 0$.

Definición 3.4. *Dado $d \in D(\mathbb{R}^n_{++})$, una función $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ es denominada distancia proximal inducida por d si H tiene valores finitos en $\mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$ y para cada $a, b \in \mathbb{R}^n_{++}$ se satisface*

(Ii) *$H(a, a) = 0$.*

(Iii) *$\langle c - b, \nabla_1 d(b, a)\rangle \leq H(c, a) - H(c, b) - H(b, a), \quad \forall c \in \mathbb{R}^n_{++}$.*

Escribiremos $(d, H) \in \mathcal{F}(\mathbb{R}^n_{++})$ a la distancia proximal y distancia proximal inducida que satisface las premisas de la Definición **??**.
Denotemos por $(d, H) \in \mathcal{F}(\mathbb{R}^n_+)$ si existe una función H tal que:

(Iiii) H tiene valores finitos en $\mathbb{R}^n_+ \times \mathbb{R}^n_{++}$ satisfaciendo **(Ii)** y **(Iii)**, para cada $c \in \mathbb{R}^n_+$.

(Iiv) Para cada $c \in \mathbb{R}^n_+$, $H(c,\cdot)$ tiene conjuntos de nivel acotados en $\mathbb{R}^n_{++}$.

Finalmente, escribiremos $(d,H) \in \mathcal{F}_+(\mathbb{R}^n_+)$ si

(Iv) $(d,H) \in \mathcal{F}(\mathbb{R}^n_+)$.

(Ivi) $\forall\, y \in \mathbb{R}^n_+$ y $\forall\, \{y^k\} \subset \mathbb{R}^n_{++}$ acotada con $\lim_{k\to+\infty} H(y,y^k) = 0$, entonces $\lim_{k\to+\infty} y^k = y$.

(Ivii) $\forall\, y \in \mathbb{R}^n_+$, y $\forall\, \{y^k\} \subset \mathbb{R}^n_{++}$ tal que $\lim_{k\to+\infty} y^k = y$, entonces $\lim_{k\to+\infty} H(y,y^k) = 0$.

El resultado principal del método propuesto será cuando $(d,H) \in \mathcal{F}_+(\mathbb{R}^n_+)$. Varios ejemplos de distancias proximales que se ajusten a las definiciones anteriores, por ejemplo distancias Bregman, distancias proximal sobre la base de $\phi-$ divergencias, distancias auto-proximales y las distancias sobre la base de distancias proximales homogéneas de segundo orden, fueron dadas por Auslender y Teboulle (2006).

3.4. Un método multiplicar proximal con distancias generalizadas

El Problema

Estamos interesados en resolver el siguiente problema de minimización convexa con estructura separable:

$$(PCS)\quad \min\{f(x) + g(z) \,:\, Ax + Bz = b, x \in \mathbb{R}^n_+,\, z \in \mathbb{R}^p\},$$

donde $f : \mathbb{R}^n \to (-\infty,+\infty]$ y $g : \mathbb{R}^p \to (-\infty,+\infty]$ son funciones propias convexas cerradas y $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times p}$ y $b \in \mathbb{R}^m$.
El problema dual Fenchel para (PCS) es definido por:

$$(D)\quad \min\{(f + \delta_{\mathbb{R}^n_+})^*(-A^T y) + g^*(-B^T y) : y \in \mathbb{R}^m\}$$

donde $(f + \delta_{\mathbb{R}^n_+})^*$ y g^* son las funciones conjugadas de $f + \delta_{\mathbb{R}^n_+}$ y g, respectivamente, y y es un vector multiplicador de Lagrange asociado con las restricciones en (PCS). El lagrangiano $L : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \to (-\infty,+\infty]$ para el problema (PCS) es definido por:

$$L(x,z,y) = (f + \delta_{\mathbb{R}^n_+})(x) + g(z) + \langle y, Ax + Bz - b\rangle\,.$$

Si (x^*,z^*) es una solución óptima de (PCS) y y^* es el correspondiente vector multiplicador de Lagrange, entonces y^* es una solución óptima de (D) y (x^*,z^*,y^*) es un punto de silla de L (ver [?], Teorema 28.3); es decir, para todo $x \in domf \cap \mathbb{R}^n_+$, $z \in domg$ y $y \in \mathbb{R}^m$, se tiene que

$$L(x^*,z^*,y) \leq L(x^*,z^*,y^*) \leq L(x,z,y^*). \tag{7}$$

El Método (MMPDG)

El método propuesto es una extensión del método multiplicador proximal predictor-corrector introducido por Chen y Teboulle (1994), [?], la extensión está basado en la utilización de la clase de distancias generalizadas $(d_0,H_0) \in \mathcal{F}_+(\mathbb{R}^n_+)$. A este método lo llamaremos Método Multiplicador Proximal con Distancias Generalizadas o simplemente (MMPDG).
Dado $\rho > 0$ definimos las siguientes funciones:

$$d(x,y) = d_0(x,y) + (\rho/2)\left\|x-y\right\|^2, \tag{8}$$

$$H(x,y) = H_0(x,y) + (\rho/2)\left\|x-y\right\|^2. \tag{9}$$

Es fácil verificar que $(d,H) \in \mathcal{F}_+(\mathbb{R}^n_+)$.

Observación 3.1. *Dada una matriz M, definimos la norma de M como $\|M\| := \max_{\|x\|\neq 0} \|Mx\|$.*

El método (MMPDG) es como sigue:

Método (MMPDG).

Paso 0. Elegir un par $(d_0, H_0) \in \mathcal{F}_+(\mathbb{R}^n_+)$, un parámetro $\rho > 0$ y definir (d, H) dado por (??)-(??). Tomar un punto arbitrario $(x^0, z^0, y^0) \in \mathbb{R}^n_{++} \times \mathbb{R}^p \times \mathbb{R}^m$ y una constante $\eta > 0$.

Paso 1. Terminar si el criterio de parada es obtenido. Caso contrario, elegir una sucesión $\{\lambda_k\}$ tal que

$$\lambda_k \geq \eta,\ \lambda_k \|A\| \leq \frac{\sqrt{\rho}}{2},\ \lambda_k \|B\| \leq \frac{\sqrt{2}}{2} \tag{10}$$

y calcular

$$p^{k+1} = y^k + \lambda_k(Ax^k + Bz^k - b), \tag{11}$$

Paso 2. Resolver

$$x^{k+1} = \arg \min_{x \geq 0} \left\{ f(x) + \langle p^{k+1}, Ax \rangle + (1/\lambda_k) d(x, x^k) \right\}, \tag{12}$$

$$z^{k+1} = \arg \min_{z \in \mathbb{R}^p} \left\{ g(z) + \langle p^{k+1}, Bz \rangle + (1/\lambda_k) \|z - z^k\|^2 \right\}. \tag{13}$$

Paso 3. Calcular

$$y^{k+1} = y^k + \lambda_k(Ax^{k+1} + Bz^{k+1} - b), \tag{14}$$
$$k = k+1, \text{ e ir al paso 1.}$$

Análisis de Convergencia

En esta subsección se demuestra que las iteraciones dadas por el método están bien definidas y convergen a una solución óptima del problema primal (PCS) y dual (D).

A lo largo del artículo asumiremos las siguiente hipótesis:

(H1) El conjunto optimal del problema (PCS) es no vacio.
(H2) Existen $x \in \mathbb{R}^n_{++} \cap r.i(dom f)$ y $z \in r.i(dom g)$ tal que $Ax + Bz = b$.

Teorema 3.1. *Supongamos que las hipótesis* $(H1)$ *y* $(H2)$ *se satisfacen, entonces para cualquier* $(x^k, z^k, y^k) \in \mathbb{R}^n_{++} \times \mathbb{R}^p \times \mathbb{R}^m$, $\lambda_k > 0$, *existe* $(x^{k+1}, z^{k+1}) \in \mathbb{R}^n_{++} \times \mathbb{R}^p$ *satisfaciendo* (??) *y* (??).

Demostración. Sea $F_k(x) = f(x) + \langle p^{k+1}, Ax \rangle + (1/\lambda_k) d(x, x^k)$, entonces de la Hipótesis (H2) se tiene que $dom F_k = dom f \cap dom\, d(\cdot, x^k) \neq \emptyset$. Definamos $S_k = \arg \min_{x \geq 0} \{F_k(x)\} = \arg \min_{x \in \mathbb{R}^n} \{F_k(x) + \delta_{\mathbb{R}^n_+}(x)\}$, donde $\delta_{\mathbb{R}^n_+}$ es la función indicatriz, probaremos que S_k es no vacio. En efecto, como f, $\langle A^T p^{k+1}, \cdot \rangle$ y $(1/\lambda_k) d(\cdot, x^k)$ son funciones propias convexas y cerradas, entonces F_k es también propia convexa y cerrada y de las Proposiciones ?? y ??, es suficiente mostrar que $(F_k + \delta_{\mathbb{R}^n_+} 0^+)(y) > 0$, $\forall y \neq 0$.
Sea $y \neq 0$, entonces por la Proposición ?? tenemos que

$$((F_k + \delta_{\mathbb{R}^n_+}) 0^+)(y) = (f0^+)(y) + ((\langle A^T p^{k+1}, \cdot \rangle 0^+)(y) + (1/\lambda_k)(d(\cdot, x^k) 0^+)(y) + (\delta_{\mathbb{R}^n_+} 0^+)(y). \tag{15}$$

Es fácil probar que

$$(f0^+)(y) > -\infty,\ (\delta_{\mathbb{R}^n_+} 0^+)(y) > -\infty,\ ((\langle A^T p^{k+1}, \cdot \rangle 0^+)(y) = \langle A^T p^{k+1}, y \rangle.$$

Así, nos queda probar que $(d(\cdot, x^k) 0^+)(y) = +\infty$.
Como $d(\cdot, z)$ es propia convexa y cerrada, entonces

$$(d(\cdot, x^k) 0^+)(y) = (d_0(\cdot, x^k) 0^+)(y) + (\| \cdot - x^k \|) 0^+)(y)$$

Debido a que d es coerciva, ver Definición ??,**iii**, $(d_0(\cdot, x^k) 0^+)(y) > 0$ y como $(\| \cdot - x^k \|) 0^+)(y) = +\infty$ obtenemos que $(d(\cdot, x^k) 0^+)(y) = +\infty$ y así

$$((F_k + \delta_{\mathbb{R}^n_+}) 0^+)(y) = +\infty > 0,\ \forall y \neq 0.$$

Esto es, existe $\bar{x} := x^{k+1} \in \mathbb{R}^n$ tal que

$$0 \in \partial\Big(f(\cdot) + \langle A^T p^{k+1}, \cdot\rangle + (1/\lambda_k)d(\cdot, x^k) + \delta_{\mathbb{R}^n_+}(\cdot)\Big)(x^{k+1}).$$

Ahora por la hipótesis **(H2)**, existe $x \in \mathbb{R}^n_{++} \cap r.i(domf)$, entonces de la Proposición ??, tenemos que

$$0 \in \partial f(x^{k+1}) + A^T p^{k+1} + (1/\lambda_k)\partial_1 d(x^{k+1}, x^k) + \partial\delta_{\mathbb{R}^n_+}(x^{k+1}),$$

de la Definición ??,**(ii)**, tenemos que $\texttt{dom}\partial_1 d(\cdot, y) = \mathbb{R}^n_{++}$, entonces $x^{k+1} \in \mathbb{R}^n_{++}$. Finalmente de la estrictamente convexidad de la función $d(\cdot, y)$, obtenemos

$$x^{k+1} = \arg\min_{x\geq 0}\{f(x) + \langle p^{k+1}, Ax\rangle + (1/\lambda_k)d(x, x^k)\}.$$

Para la existencia de z^{k+1} el análisis es similar. Por lo tanto, la prueba esta completada. ■

Lema 3.1. *Sea $F : \mathbb{R}^m \to (-\infty, +\infty]$ una función propia convexa y cerrada, $\tau > 0$ y se define:*

$$u^{k+1} = \arg\min_{u\in\mathbb{R}^m}\{F(u) + (1/(2\tau))\|u - u^k\|^2\}.$$

Entonces para cualquier entero $k \geq 0$,

$$2\tau[F(u^{k+1}) - F(u)] \leq \|u^k - u\|^2 - \|u^{k+1} - u\|^2 - \|u^{k+1} - u^k\|^2, \ \forall u \in \mathbb{R}^m.$$

Prueba. Ver [?], Lema 3.1. ■

Usaremos también la siguiente identidad útil:

$$\langle b - a, c - b\rangle = (1/2)(\|c - a\|^2 - \|c - b\|^2 - \|b - a\|^2) \tag{16}$$

y por conveniencia usamos la notación

$$\Delta_k(x, z) = H(x, x^k) - H(x, x^{k+1}) - H(x^{k+1}, x^k) + \|z - z^k\|^2 - \|z - z^{k+1}\|^2 - \|z^{k+1} - z^k\|^2 \tag{17}$$

Lema 3.2. *Supongamos que $\{x^k\}$, $\{z^k\}$, $\{y^k\}$, $\{p^k\}$ son sucesiones generadas por el (MMPDG). Entonces, para todo $x \in \mathtt{dom}f$, $z \in \mathtt{dom}g$, $y \in \mathbb{R}^m$ las siguientes desigualdades se satisfacen:*

(i) $\lambda_k\{L(x^{k+1}, z^{k+1}, p^{k+1}) - L(x, z, p^{k+1})\} \leq \Delta_k(x, z)$

(ii) $2\lambda_k\{L(x^k, z^k, y) - L(x^k, z^k, p^{k+1})\} \leq \|y^k - y\|^2 - \|p^{k+1} - y\|^2 - \|p^{k+1} - y^k\|^2$

(iii) $2\lambda_k\{L(x^{k+1}, z^{k+1}, y) - L(x^{k+1}, z^{k+1}, y^{k+1})\} \leq \|y^k - y\|^2 - \|y^{k+1} - y\|^2 - \|y^{k+1} - y^k\|^2$.

Prueba.
(i) Sea $x \in domf$, $z \in domg$, $y \in \mathbb{R}^m$ y denotemos por:

$$f^k(x) = f(x) + \langle p^{k+1}, Ax\rangle \ \text{ y } \ g^k(z) = g(z) + \langle p^{k+1}, Bz\rangle,$$

entonces, de (??) y (??) obtenemos, respectivamente,

$$0 \in \partial f^k(x^{k+1}) + (1/\lambda_k)\nabla_1 d(x^{k+1}, x^k) \ \text{ y } \ 0 \in \partial g^k(z^{k+1}) + (2/\lambda_k)(z^{k+1} - z^k),$$

entonces existe $g_1^{k+1} \in \partial f^k(x^{k+1})$ y $g_2^{k+1} \in \partial g^k(z^{k+1})$ tal que

$$g_1^{k+1} = -(1/\lambda_k)\nabla_1 d(x^{k+1}, x^k) \tag{18}$$
$$g_2^{k+1} = -(2/\lambda_k)(z^{k+1} - z^k). \tag{19}$$

Por otro lado, de la definición de L, es claro que

$$L(x^{k+1}, z^{k+1}, p^{k+1}) - L(x, z, p^{k+1}) = (f^k(x^{k+1}) - f^k(x)) + (g^k(z^{k+1}) - g^k(z)). \tag{20}$$

Como $g_1^{k+1} \in \partial f^k(x^{k+1})$, entonces

$$f^k(x) \geq f^k(x^{k+1}) + \langle g_1^{k+1}, x - x^{k+1}\rangle,$$

luego de (??) y de la Definición ?? **(Iii)**, obtenemos

$$\lambda_k(f^k(x^{k+1}) - f^k(x)) \leq \langle \nabla_1 d(x^{k+1}, x^k), x - x^{k+1}\rangle \leq H(x, x^k) - H(x, x^{k+1}) - H(x^{k+1}, x^k). \tag{21}$$

También, como $g_2^{k+1} \in \partial g^k(z^{k+1})$, entonces

$$g^k(z) \geq g^k(z^{k+1}) + \langle g_z^{k+1}, z - z^{k+1}\rangle,$$

luego de (??) y de la identidad (??), tenemos que

$$\lambda_k(g^k(z^{k+1}) - g^k(z)) \leq 2\langle z^{k+1} - z^k, z - z^{k+1}\rangle = \|z - z^k\|^2 - \|z - z^{k+1}\|^2 - \|z^{k+1} - z^k\|^2, \tag{22}$$

sumando (??) y (??), y de (??) y la notación en (??), obtenemos

$$\lambda_k\{L(x^{k+1}, z^{k+1}, p^{k+1}) - L(x, z, p^{k+1})\} \leq \Delta_k(x, z).$$

Para la prueba de **(ii)** y **(iii)** notemos que los Pasos 1 y 3 del (MMPDG) pueden ser escritos equivalentemente como:

$$p^{k+1} = \arg\min_{y \in \mathbb{R}^m}\{-L(x^k, z^k, y) + (1/2\lambda_k)\|y - y^k\|^2\}$$

$$y^{k+1} = \arg\min_{y \in \mathbb{R}^m}\{-L(x^{k+1}, z^{k+1}, y) + (1/2\lambda_k)\|y - y^k\|^2\}.$$

Entonces, usando el Lema ??, dos veces con $\tau = \lambda_k$, $F(y) = -L(x^k, z^k, y)$ y $F(y) = -L(x^{k+1}, z^{k+1}, y)$ respectivamente, la primera y segunda ecuación anterior producen las desigualdades deseadas **(ii)** y **(iii)**.
■

Observación 3.2. *En lo que resta del artículo, denotemos $w = (x, z, y)$ y definamos la función $\hat{H} : X \times X \to \mathbb{R}_+ \cup \{+\infty\}$ por*

$$\hat{H}(w_1, w_2) = \hat{H}((x_1, z_1, y_1), (x_2, z_2, y_2)) = H(x_1, x_2) + \|z_1 - z_2\|^2 + (1/2)\|y_1 - y_2\|^2, \tag{23}$$

donde $X = \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ y H es la distancia proximal inducida por d.

Proposición 3.6. *Supongamos que las hipótesis $(H1)$ y $(H2)$ se satisfacen. Sean $\{w^k\}$ y $\{p^k\}$ sucesiones generadas por el método (MMPDG); sea (x^*, z^*) un solución óptima de (PCS), y sea y^* su correspondiente multiplicador de Lagrange, entonces tenemos, para cada k,*

$$\begin{aligned}\hat{H}(w^*, w^{k+1}) \leq\ & \hat{H}(w^*, w^k) - [H_0(x^{k+1}, x^k) + (1/2)(\rho - 2\lambda_k^2\|A\|^2)\|x^{k+1} - x^k\|^2] + \\ & -(1 - 2\lambda_k^2\|B\|^2)\|z^{k+1} - z^k\|^2 - (1/2)(\|p^{k+1} - y^{k+1}\|^2 - \|p^{k+1} - y^k\|^2),\end{aligned} \tag{24}$$

donde $w^* = (x^*, z^*, y^*)$. En particular, se tiene que, para todo k,

$$\hat{H}(w^*, w^{k+1}) \leq \hat{H}(w^*, w^k). \tag{25}$$

Prueba. Del Lema ??, en el punto $(x, z) = (x^*, z^*)$, obtenemos

$$\begin{aligned}\lambda_k(L(x^{k+1}, z^{k+1}, p^{k+1}) - L(x^*, z^*, p^{k+1})) \leq\ & H(x^*, x^k) - H(x^*, x^{k+1}) - H(x^{k+1}, x^k) + \\ & + \|z^* - z^k\|^2 - \|z^* - z^{k+1}\|^2 - \|z^{k+1} - z^k\|^2.\end{aligned} \tag{26}$$

Como (x^*, z^*, y^*) es un punto de silla de L, obtenemos de (??)

$$\lambda_k(L(x^*, z^*, p^{k+1}) - L(x^{k+1}, z^{k+1}, y^*)) \leq 0. \tag{27}$$

Sumando (??) y (??) reordenando algunos términos, conseguimos

$$\begin{array}{rcl} H(x^*,x^{k+1})+\|z^*-z^{k+1}\|^2 & \leq & H(x^*,x^k)+\|z^*-z^k\|^2-\left[H(x^{k+1},x^k)+\|z^{k+1}-z^k\|^2\right] \\ & & -\lambda_k\left\langle p^{k+1}-y^*,Ax^{k+1}+Bz^{k+1}-b\right\rangle. \end{array} \tag{28}$$

Ahora, usando el Lema ??, **(ii)**, con $y:=y^{k+1}$ y $y:=y^*$ por **(iii)**, obtenemos, respectivamente,

$$2\lambda_k(L(x^k,z^k,y^{k+1})-L(x^k,z^k,p^{k+1}))\leq\|y^k-y^{k+1}\|^2-\|p^{k+1}-y^{k+1}\|^2-\|p^{k+1}-y^k\|^2, \tag{29}$$

$$2\lambda_k(L(x^{k+1},z^{k+1},y^*)-L(x^{k+1},z^{k+1},y^{k+1}))\leq\|y^k-y^*\|^2-\|y^{k+1}-y^*\|^2-\|y^{k+1}-y^k\|^2, \tag{30}$$

sumando (??) y (??) y reordenando algunos términos, conseguimos

$$\begin{array}{rcl} (1/2)\left\|y^{k+1}-y^*\right\|^2 & \leq & (1/2)\left\|y^k-y^*\right\|^2-(1/2)\left[\left\|p^{k+1}-y^{k+1}\right\|^2+\left\|p^{k+1}-y^k\right\|^2\right]+ \\ & - & \lambda_k(\langle y^*-y^{k+1},Ax^{k+1}+Bz^{k+1}-b\rangle+\langle y^{k+1}-p^{k+1},Ax^k+Bz^k-b\rangle). \end{array} \tag{31}$$

Luego adicionando (??) y (??) y de (??), obtenemos

$$\begin{array}{rcl} \hat{H}(w^*,w^{k+1}) & \leq & \hat{H}(w^*,w^k)-(H(x^{k+1},x^k)+\|z^{k+1}-z^k\|^2)+ \\ & & -(1/2)\left(\|p^{k+1}-y^{k+1}\|^2+\|p^{k+1}-y^k\|^2\right) \\ & & +\lambda_k\left\langle y^{k+1}-p^{k+1},A(x^{k+1}-x^k)+B(z^{k+1}-z^k)\right\rangle \\ & = & \hat{H}(w^*,w^k)-(H(x^{k+1},x^k)+\|z^{k+1}-z^k\|^2)+ \\ & & -(1/2)\left(\|p^{k+1}-y^{k+1}\|^2+\|p^{k+1}-y^k\|^2\right) \\ & & +\lambda_k^2\left\|A(x^{k+1}-x^k)+B(z^{k+1}-z^k)\right\|^2, \end{array} \tag{32}$$

donde la última igualdad se sigue de (??) y (??). Notando la desigualdad

$$\left\|A(x^{k+1}-x^k)+B(z^{k+1}-z^k)\right\|^2\leq 2\left(\|A\|^2\|x^{k+1}-x^k\|^2+\|B\|^2\|z^{k+1}-z^k\|^2\right),$$

y también,

$$H(x^{k+1},x^k)=H_0(x^{k+1},x^k)+(\rho/2)\|x^{k+1}-x^k\|^2,$$

donde H_0 es la distancia proximal inducida por d_0.
De (??), obtenemos

$$\begin{array}{rcl} \hat{H}(w^*,w^{k+1}) & \leq & \hat{H}(w^*,w^k)-[H_0(x^{k+1},x^k)+(1/2)(\rho-2\lambda_k^2\|A\|^2)\|x^{k+1}-x^k\|^2]+ \\ & & -(1-2\lambda_k^2\|B\|^2)\|z^{k+1}-z^k\|^2-(1/2)\left(\|p^{k+1}-y^{k+1}\|^2+\|p^{k+1}-y^k\|^2\right). \end{array}$$

Entonces, la desigualdad (??) se sigue inmediatamente de (??) y (??). ∎

Proposición 3.7. *Sea $\{w^k\}$ una sucesión generada por el método (MMPDG), entonces $\{w^k\}$ es acotado y cada punto de acumulación de $\{w^k\}$ es un punto de silla del lagrangiano L.*

Prueba. De (??), tenemos

$$\hat{H}(w^*,w^{k+1})\leq\hat{H}(w^*,w^k),\quad \text{for all } k$$

entonces

$$w^k\in L_{\hat{H}}(w^*,\bar{\alpha}):=\{w:\hat{H}(w^*,w)\leq\bar{\alpha}\},\quad \text{for all } k,$$

donde $\bar{\alpha}=\hat{H}(w^*,w^0)$.
Esto implica que $\{w^k\}$ es acotada debido a la condición (Iiv) en la Definición ??.
Ahora, sea (x^*,z^*) una solución óptima de (PCS), sea y^* un correspondiente multiplicador de Lagrange, y sea $w^*=(x^*,z^*,y^*)$.
De (??), vemos que la sucesión $\{\hat{H}(w^*,w^k)\}$ es monótona y acotada, por lo tanto, existe $\beta\geq 0$ tal que

$$\lim_{k\to+\infty}\hat{H}(w^*,w^k)=\beta, \tag{33}$$

así, tomando límite en ambos miembros de (??) y reordenando términos, obtenemos

$$\limsup_{k\to+\infty}(H_0(x^{k+1},x^k)+(1/2)(\rho-2\lambda_k^2\|A\|^2)\|x^{k+1}-x^k\|^2+(1-2\lambda_k^2\|B\|^2)\|z^{k+1}-z^k\|^2+ \\ (1/2)(\|p^{k+1}-y^{k+1}\|^2-\|p^{k+1}-y^k\|^2))\leq 0. \tag{34}$$

Luego se sigue de (??) y (??) que

$$\begin{array}{ll} H_0(x^{k+1},x^k)\to 0, & \|x^{k+1}-x^k\|\to 0 \\ \|z^{k+1}-z^k\|\to 0, & \|p^{k+1}-y^{k+1}\|\to 0 \\ \|p^{k+1}-y^k\|\to 0. & \end{array} \tag{35}$$

Ahora, sea $\bar{w}$ un punto de acumulación arbitrario de $\{w^k\}$, y sea $\{w^k\}_{k\in K}$ una sucesión convergiendo a $\bar{w}$. Entonces, de la condición (Ivii) en la Definición ??, obtenemos

$$\lim_{k\to\infty,\,k\in K}\hat{H}(\bar{w},w^k)=0$$

Tomando límite en la desigualdad del Lema ??,(**i**), obtenemos que para todo $x\in domf$ y $z\in domg$,

$$L(\bar{x},\bar{z},\bar{y})-L(x,z,\bar{y})\leq 0. \tag{36}$$

Similarmente, tomando límite en el Lema ??, (**ii**), tenemos que, para todo $y\in\mathbb{R}^m$,

$$L(\bar{x},\bar{z},y)-L(\bar{x},\bar{z},\bar{y})\leq 0. \tag{37}$$

De(??) y (??), obtenemos que, para todo $x\in domf$, $z\in domg$, $y\in\mathbb{R}^m$,

$$L(\bar{x},\bar{z},y)\leq L(\bar{x},\bar{z},\bar{y})\leq L(x,z,\bar{y}).$$

Por lo tanto, $\bar{w}$ es un punto de silla de L. ■

Teorema 3.2. *Supongamos que las hipótesis (H1) y (H2) se satisfacen y sea $\{(x^k,z^k,y^k)\}$ la sucesión generada por el método MMPDG, entonces la sucesión $\{(x^k,z^k,y^k)\}$ converge globalmente a (x^*,z^*,y^*), con (x^*,z^*) óptimo para (PCS) y y^* su correspondiente multiplicador de Lagrange*

Demostración. De la Proposición ??, $\{w^k\}=\{(x^k,z^k,y^k)\}$ es acotada y cualquier punto de acumulación $\bar{w}$ es un punto de silla de L.
Sea $\{w^k\}_{k\in K}$ la subsucesión la cual converge a $\bar{w}$, entonces $(\bar{x},\bar{z})$ es una solución óptima de (PCS) y $\bar{y}$ su correspondiente multiplicador de Lagrange (ver [?], Teorema 28.3). Entonces, es suficiente mostrar que $\{w^k\}$ converge a $\bar{w}$.
En efecto, como

$$\lim_{k\to+\infty,\,k\in K}x^k=\bar{x},$$

de la Definición ??, (**Ivii**), tenemos que

$$\lim_{k\to+\infty,\,k\in K}H(\bar{x},x^k)=0$$

entonces, tomando $w_1=\bar{w}$ y $w_2=w^k$ en (??) se tiene que

$$\lim_{k\to+\infty,\,k\in K}\hat{H}(\bar{w},w^k)=0. \tag{38}$$

De (??) sustituyendo w^* por $\bar{w}$ obtenemos

$$0\leq\hat{H}(\bar{w},w^{k+1})\leq\hat{H}(\bar{w},w^k),\quad\text{para todo } k,$$

esto es, la sucesión $\{\hat{H}(\bar{w},w^k)\}$ es monótona y acotada inferiormente, entonces converge y como existe una subsucesión convergiendo a cero, ver ((??)), toda la sucesión converge, esto es,

$$\lim_{k\to+\infty}\hat{H}(\bar{w},w^k)=0,$$

Finalmente, de la Definición ??, (**Ivi**) obtenemos que

$$\lim_{k\to+\infty}w^k=\bar{w}.$$

Con lo que la demostración está concluida. ■

4. Discusiones

En este trabajo obtenemos la convergencia del método **(MMPDG)** usando hipótesis adecuadas pero naturales al problema (PCS). El estudio de esta investigación fue motivada por problemas de optimización de gran dimensión que aparecen en redes de telecomunicación y gestión de energía eléctrica.

Obviamente, existen otros abordages para resolver estos modelos y en cierto sentido podemos separarlos en clases. La primera clase corresponde a los métodos directos que tratan de adaptar las técnicas conocidas de la optimización matemática para la estructura de redes (tenemos por ejemplo los métodos de desviación de flujo basado en el método del gradiente [?], método de Newton proyectado [?], métodos de puntos interiores [?]).

La segunda clase agrupa a los métodos duales (cóncavos) que se diferencian por la elección de la relajación lagrangeana (con el objetivo de tener mejores propiedades de diferenciabilidad de tal manera que puedan ser aplicados métodos eficientes de optimización) y por el uso de las direcciones de ascenso (métodos subgradientes [?]).

La tercera clase de métodos son relacionados con métodos proximales como el método multiplicador de direcciones alternadas [?, ?], método de inverso parcial, [?, ?, ?, ?] y los métodos multiplicador proximal predictor-corrector los cuales hemos mencionado en la introducción. Debemos observar que este trabajo se encuentra en la última clase descrita.

Este trabajo permite abrir la posibilidad de futuras investigaciones que detallamos a seguir:

a) De forma análoga al análisis de convergencia presentada en este trabajo una versión inexacta del método (MMPDG) es deseable por cuestiones computacionales. Esto es natural en optimización numérica ya que todo método no está libre de errores computacionales, la metodologia debe ser similar a la desarrollada por Kyono y Fukushima, [?].

b) En problemas más generales es posible que las restricciones usadas en el artículo $\mathbb{R}^n_+$ sean más generales, por ejemplo un cono o un conjunto convexo por un futuro trabajo debe considerar sustituir $\mathbb{R}^n_+$ por uno más general y por lo tanto la definición de distancia debe ser extendida a ese nivel, en este caso la metodologia seguida por Auslender y Teboulle (2006), [?], es la más cercana a este tratamiento.

c) Todo algoritmo debe ser implementado y comparado con metodologias existentes, por eso creemos que es necesario tener experimentos computacionales con problemas de mediano y gran tamaño y su respectiva comparaciones con otros métodos.

d) El estudio de la velocidad de convergencia de un método es importante del punto de vista teórico ya que eso nos dá una idea de la eficiencia y efectividad del método. En nuestro método, el parámetro que está libre es el λ_k, entonces se puede estudiar que intervalos de λ_k son más óptimos para acelerar la convergencia del método. Para esto recomendamos el artículo de Mahey et al. [?]

e) Siguiendo el artículo de Auslender y Teboulle [?], creemos que es posible extender el método para resolver problemas de desigualdad variacional en espacios euclidianos y así ampliar el campo de aplicación del método. Ahora un trabajo que puede ser un desafio es la extensión del método para problemas de desigualdad variacional en Optimización Semidefinida, ver [?].

f) Existen variantes de los métodos proximales para mejorar las propiedades de convergencia usando las métricas variables. Una futura investigación en este contexto seria reemplazar la distancia proximal en $\mathbb{R}^n_{++}$ por

$$d(x,y) = d_0(x,y) + (\rho_k/2)||x-y||_y^2,$$

donde $||a||_y = a^T Y^{-r} a$, con $Y = \texttt{diag}(y_1, y_2, ..., y_n)$, $r > 0$, y $\rho_k > 0$. Está métrica es motivada por la métrica diagonal Riemaniana $(\mathbb{R}^n_{++}, X^{-r})$ que tiene mucha relación con geometria diferencial y métodos de puntos interiores.

5. Agradecimientos

Los autores agradecemos al Programa de Engenharia de Sistemas e Computação (PESC) de la COPPE-Universidad Federal de Rio de Janeiro, Brasil, por el apoyo económico para los viajes y estadía del primer autor a esta institución y al CONCYTEC-Embajada de Francia por el financiamiento para un estagio Post-Doctoral también del primer autor al Laboratorio LIMOS-ISIMA de la Universidad Blaise Pascal, Clermont-Ferrand, Francia.

Bibliografía

[1] Auslender A, Teboulle M, Ben-Tiba S (1999) Interior proximal and multiplier methods based on second order homogeneous functionals. Math Oper. Research 24, 3: 645-668.

[2] Ahuja R.V. Magnanti T.L. and Orlin J.B. (1993). Network flows: theory, algorithms and applications. Prentice Hall, Englewood Cliffs, N.J.

[3] Auslender A., Teboulle M., (2001). Entropic proximal decomposition methods for convex programs and variational inequalities, Math. Program., Ser A, Vol. 91: 33-47.

[4] Auslender A., Teboulle M. (2006) Interior gradient and proximal methods for convex and conic optimization. SIAM J. Optim., Vol. 16, 3: 697-725, 2006.

[5] Assad A.A. (1978) Multicommodity network flows-A survey, Networks, vol 8: 37-91.

[6] Bazaraa M.S. (2005). Programación lineal y flujo en redes, Limusa, N.J.

[7] Bertsekas D.P. and Tsitsiklis J.N. (1989). Parallel and distributed computation, international ed., Prentice Hall, Englewood Cliffs.

[8] Bertsekas D.P. and Gafni E.M. (1983). Projected Newton methods and optimization of multicommodity flows. IEEE Trans. Automat. Contr. AC-28: 1090-1096.

[9] Bonnans F.J, Haddou M., Lisser A. and Reba R. (2000) Interior point methods with decomposition for multicommodity flow problems, Institut National de Recherche en Informatique et en Automatique (INRIA), Rapport de Recherche.

[10] Chen, G., and Teboulle, M. (1994). A proximal-Based decomposition method for convex minimization problems, Mathematical Programming, Serie A, 1, Vol. 64: 81-101.

[11] Chifflet J. Mahey P. and Reiner V. (1994). Proximal decomposition for multicommodity flow problems with convex costs, Telecommun. Syst. Vol. 3:1-10.

[12] Dantzig G. and Wolfe P. (1960) Decomposition principle for linear programs. Operations Research Vol. 8:1–111.

[13] Eckstein J. y Bertsekas D.P. (1992) On the Douglas-Rachford splitting method and the proximal algorithm for maximal monotone operators. Mathematical Programming Vol 55, N°3: 293-318.

[14] Fratta L. Gerla M. and Cleinrock. (1973). The flow deviation method: An approach to store-and-forward communication networks design. Networks 3: 97-133.

[15] Gabay, D. (1983). Applications of the method of multipliers to variational inequalities. In M. Fortin, R. Glowinski, editors, Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems (Amsterdam: North-Holland).

[16] Gabay, D., Mercier B. (1976). A Dual algorithm for the solution of nonlinear variational problems via finite element approximations. Computers and Mathematics with Applications Vol.2: 17-40.

[17] Guler O.(1991). On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. Vol. 29: 403-419.

[18] Kleinrock l.(1972). Communications, nets, stochastic message flow and delay.

[19] Kyono M., and Fukushima M. (2000). Nonlinear proximal decomposition method for convex programming, JOTA, Vol. 106, N° 2:357-372.

[20] Lions, P. L., Mercier, B. (1979), Splitting algorithms for the sum of two nonlinear operators. SIAM Journal on Numerical Analysis Vol. 16 (6):964-979.

[21] Mahey P., Oualibouch, and Pham D.T. (1995). Proximal decomposition of the graph of a maximal monotone operator. SIAM J. Optim. Vol. 5, N° 2: 454-466.

[22] Mahey P. Ouorou A. LeBlanc L. and Chifflet J. (1997) A new proximal decomposition algorithm for routing in telecommunication networks. Networks, Vol. 31: 227-238.

[23] Martinet B. (1970). Régularization, d'inéquations variationelles par approximations successives, Revue Francaise q'Informatique et de Recherche Operationelle: 154-159.

[24] Moreau J.-J. (1962). Fonctiones convexes duales et points proximaux dans un espace hilbertien. Comptes Rendus de l'Acadmie des Sciences de Paris 255: 2897–2899.

[25] Moreau J.-J.. (1965). Proximité et dualité dans un espace hilbertien. Bulletin de la Société Mathématique de France 93: 273–299.

[26] Ouorou A. Mahey P. Vial J.-Ph. (2000). A survey of algorithms for convex multicommodity flow problems. INFORMS Vol.46, N° 1: 126-147.

[27] Rockafellar R.T. (1970). Convex Analysis, Princeton University Press, Princeton New Jersey.

[28] Rockafellar R.T. (1976). Monotone operators and the proximal point algorithm, SIAM J. Control and Optimization, Vol. 14 N° 5: 877-898.

[29] Spingarn J. E. (1983). Partial Inverse of a Monotone Operator. Applied Mathematics and Optimization 10:247-265.

[30] Spingarn J.E. (1985). Applications of the method of partial inverse to convex programming decomposition. Math. Program. Vol. 32: 199-223.

[31] Uzawa, H. (1958). Iterative methods for concave programming. In: Arrow, K.J., Hurwicz, L., Uzawa, H. (eds). Estudies in Linear and Nonlinear Programming. Standford University Press, Standfor, C.A.: 154-165.