# Un análisis comparativo de unas formulaciones de Optimización Condicionada aplicadas a Desigualdades Variacionales.

M.A. Gómez, mago@udc.es, L. P Pedreira, lucky@udc.es, F. Rey, ferey@udc.es, J.A. Seijas, jasm@udc.es

*Facultad de Ciencias Económicas y Empresariales. Universidade da Coruña*

**Resumen.** Las condiciones KKT pueden ser consideradas para resolver problemas de desigualdades variacionales. En este trabajo analizaremos dos propuestas que resuelven desventajas de propuestas recientes de reformulación de sistemas KKT. Examinamos los sistemas KKT como problemas de minimización con condiciones de no negatividad. Probamos que, bajo ciertas condiciones débiles, todo punto estacionario de estos problemas es una solución de KKT. Basándonos en estas formulaciones, sugerimos algoritmos para resolver el problema de KKT y estudiamos algunas condiciones de convergencia.

**Palabras clave**. Condiciones KKT, Desigualdades variacionales, Problemas de optimización condicionada

## 1. Introducción

Sea $F:\Re^n \rightarrow \Re$ de clase 1 y $h:\Re^n \rightarrow \Re^m$ de clase 2. Definimos la Lagrangiana $\boldsymbol{L}:\Re^{n+p+m} \rightarrow \Re^n$ por

$$L(x,y,z) := F(x) + \nabla h(x)y - \nabla g(x)z \ ,$$

donde h son restricciones de igualdad, g son funciones de restricción de desigualdad y z e *y* son los multiplicadores de Lagrange, y consideramos el siguiente sistema KKT

$$\begin{aligned} &L(x,y,z) = 0, \\ &h(x) = 0, \\ &g(x) \geq 0, z \geq 0, \ < z, g(x) >= 0. \end{aligned} \qquad (1)$$

Un sistema de este tipo se puede utilizar para distintos casos. Por ejemplo, bajo una condición de cualificación estándar, el sistema (1) representa las condiciones necesarias de KKT para que un vector x*∈X sea una solución del problema de desigualdad variacional VIP(X,F),

$$< F(x^*), x - x^* > \geq 0, \quad \forall x \in X,$$

donde

$$X := \left\{ x \in \Re^n : h(x) = 0, g(x) \geq 0 \right\}$$

En particular, si $F = \nabla f$ para una función $f:\Re^n \to \Re$, las condiciones de KKT representan, bajo una condición de cualificación, las condiciones necesarias de primer orden para el problema de la minimización

$$\min f(x) \qquad \text{sujeto a } \; h(x) = 0, g(x) \geq 0,$$

En este trabajo estudiamos el problema de hallar un punto KKT $w^*=(x^*,y^*,z^*)\in\Re^{n+p+m}$, es decir una tripla que satisface el sistema KKT (1). En esta en dirección actualmente se construyen algoritmos para obtener la solución de problemas de desigualdad variacional y problemas de programación no lineal.

Los métodos que describimos en este trabajo están relacionados con las recientes propuestas hechas por Facchinei y Kanzow, donde el sistema (1) se transforma en una problema diferenciable de minimización sin condiciones. Estas reformulaciones están basadas en el uso de la función convexa decreciente $\varphi:\Re^2\to\Re$ definida por

$$\varphi(a,b) := \sqrt{a^2 + b^2} - a - b \tag{2}$$

Es fácil comprobar que

$$\varphi(a,b) = 0 \Leftrightarrow a \geq 0, b \geq 0, ab = 0. \tag{3}$$

Por tanto podemos considerar ahora el sistema (1) como un sistema no lineal de ecuaciones $\Phi(w)=0$, donde la aplicación $\Phi:\Re^{n+p+m}\to\Re^{n+p+m}$ está definida por

$$\Phi(w) := \Phi(x,y,z) := \begin{pmatrix} L(x,y,z) \\ h(x) \\ \varphi\,(\mathrm{g(x),z}) \end{pmatrix}$$

y donde

$$\varphi(g(x),z) := \left(\varphi(g_1(x),z_1),....,\varphi(g_m(x),z_m)\right)^T \in \Re^m.$$

Podemos asociar a este sistema una función merit,

$$\Psi(w) := \frac{1}{2}\Phi(w)^T\Phi(w) = \frac{1}{2}\|\Phi(w)\|^2,$$

lo que hace que resolver el sistema (1) sea equivalente a calcular la solución global del problema

$$\min \Psi(w) \tag{4}$$

Notemos que, para calcular una solución del sistema (1), se tienen que obtener soluciones globales del problema de minimización (4), mientras que los algoritmos usuales de minimización no condicionada pueden únicamente calcular puntos estacionarios de (4). Una de las cuestiones centrales planteadas por Facchinei y

Kanzow es el estudiar las condiciones que implican que un punto estacionario de $\Psi$ es un mínimo global de $\Psi$. Las condiciones dadas en [5], relativamente débiles, incluyen la suposición de que la Jacobiana de la Lagrangiana con respecto a las x-variables,

$$\nabla_x L(x,y,z) = \nabla F(x)^T + \sum_{j=1}^{p} y_j \nabla^2 h_j(x) - \sum_{i=1}^{m} z_i \nabla^2 g_i(x),$$

está definida positiva. Esta condición es satisfecha, por ejemplo, si $F$ es monótona y las condiciones son tales que $\nabla F(x)$ es semidefinida positiva y las Hessianas $\nabla^2 h_j(x)$ y $\nabla^2 g_i(x)$ son vacías. Sin embargo, si uno considera el caso más natural, es decir, que $F$ sea monótona, $h$ lineal, y $g$ no lineal y cóncava, es fácil comprobar que, puesto que las matrices $\nabla^2 g_i(x)$ son semidefinidas negativas, si $z_i$ es negativo y suficientemente grande, el $\nabla_x \boldsymbol{L}(x,y,z)$ no puede ser definido positivo. Notemos también que esta conclusión es independiente de la estructura de $F$ y de $h$.

Este hecho es a veces perturbador, ya que si resolvemos una desigualdad variacional monótona sobre un conjunto convexo definido por desigualdades no lineales, no podemos asegurar la convergencia de la solución única de la desigualdad variacional. Esto nos lleva a considerar la siguiente variante del problema (4)

$$\min \Psi(w) \quad \text{sujeto a } z \geq 0\,, \tag{5}$$

En este trabajo damos condiciones que aseguran que todo punto estacionario del problema (5) es una solución del sistema (1). También proponemos unos algoritmos específicos para la solución del problema (5). Algunos resultados preliminares tratados en este trabajo se pueden hallar en [11]. Un caso análogo fue propuesto por Fischer para el contexto de problemas de complementariedad no lineal.

Este trabajo lo estructuramos de la siguiente forma. En la siguiente sección recordamos algunas propiedades acerca de las funciones $\Psi$ y $\Phi$ y del método de Newton. Luego, en la sección 3, damos condiciones que aseguran que todo punto estacionario del problema (5) es una solución del sistema KKT (1). En la sección 4 introducimos los algoritmos para la solución de (1). Estos algoritmos reducen la función merit $\Psi$ en cada paso a la vez que mantienen la variable z no negativa. Estudiamos resultados de convergencia local y global para estos algoritmos en la sección 5.

NOTACIONES.
Denotamos por $J$ e $I$ a los conjuntos índices $\{1,\ldots p\}$ y $\{1,\ldots m\}$ respectivamente.

Si $w^*=(x^*,y^*,z^*) \in \Re^{n+p+m}$ es un KKT-punto, denotamos por $I_0$ el conjunto de condiciones activas de desigualdad no estricta y por $I_+$ el conjunto de condiciones activas de desigualdad estricta, es decir

$$I_0 = \{i \in I : g_i(x^*) = 0\} \qquad I_+ = \{\in I_0 : z_i^* > 0\}$$

El símbolo $\nabla$ indica la traspuesta de la Jacobiana de una función. Si la función es de variable real la traspuesta de la Jacobiana coincide con el vector gradiente, es decir, utilizando el gradiente como vector columna. Además, usaremos $\nabla^2$ para indicar la matriz Hessiana.

**2.- Preliminares.**

En esta sección veremos propiedades de diferencibilidad de las funciones $\Phi$ y $\Psi$

Debido a las suposiciones que hemos hecho de las funciones $F$, $h$ y $g$, y por la convexidad de $\varphi$, es obvio que la aplicación $\Phi$ es localmente Lipschitziana y por tanto diferenciable en casi todas partes (teorema de Rademacher). Denotamos por $D_\Phi$ el conjunto de puntos $w \in \Re^{n+p+m}$ en el que $\Phi$ es diferenciable. Entonces podemos considerar la B-subdiferencial (o matriz jacobiana pregeneralizada en el sentido de Clarke) de $\Phi$ en $w$,

$$\partial_B \Phi(w) := \left\{ H : H = \lim_{w^k \to w, w^k \in D_\Phi} \nabla \Phi(w^k)^T \right\}$$

que es un conjunto no vacío y compacto cuya envolvente convexa

$$\partial \Phi(w) := conv \partial_B \Phi(w)$$

es la jacobiana generalizada de Clarke de $\Phi$ en w. Véase [1].

**Definición 2.1.** El vector $w^*$ se dice BD-regular para $\Phi$ si todos los elementos $H \in \partial_B \Phi(w^*)$ son no singulares

El siguiente resultado da una estimación de la Jacobiana generalizada de $\Phi$. Esto se sigue básicamente de las reglas conocidas del cálculo de la Jacobiana generalizada [1]. La prueba viene dada en la proposición 3.2 de [5].

**Proposición 2.2.** Sea $w=(x,y,z) \in \Re^{n+p+m}$. Entonces

$$\partial \Phi(w)^T \subseteq \begin{pmatrix} \nabla_x L(w) & \nabla h(x) & \nabla g(x) D_a(w) \\ \nabla h(x)^T & 0 & 0 \\ -\nabla g(x)^T & 0 & D_b(w) \end{pmatrix},$$

donde $D_a(w)=\mathrm{diag}(a_1(w),\ldots,a_m(w))$, $D_b(w)=\mathrm{diag}(b_1(w),\ldots,b_m(w)) \in \Re^{mxn}$ son matrices diagonales cuyos i-esimos elementos vienen dados por

$$a_i(w) = \frac{g_i}{\sqrt{g_i(x)^2 + z_i^2}} - 1, \qquad b_i(w) = \frac{z_i}{\sqrt{g_i(x)^2 + z_i^2}} - 1$$

si $(g_i(x), z_i) \neq 0$,

y $a_i(w)=\xi_i-1$, $b_i(w)=\zeta_i-1$ para cualquier $(\xi_i, \zeta_i)$ con $\|(\xi_i, \zeta_i)\| \leq 1$ si $(g_i(x),z_i)=0$.

Para la siguiente proposición usamos la condición de regularidad estricta de Robinson [12].

**Proposición 2.3.** Una solución $w^*=(x^*,y^*,z^*)\in\Re^{n+p+m}$ del sistema (1) es estrictamente regular si y solo si todas las matrices de la Jacobiana generalizada de Clarke son no singulares. En particular, la regularidad estricta de $w^*$ es condición suficiente para que $w^*$ sea solución BD-regular del sistema $\Phi(w)=0$.

Prueba. Véase el Corolario 4.7 de [6]

A partir de las suposiciones de diferenciabilidad de las funciones $F$, $g$ y $h$ establecidas en la sección 1, y el hecho de que $\varphi$ es estrictamente semismooth, se llega al siguiente resultado [6, proposición 3.1]

**Proposición 2.4.** Se cumplen las siguientes propiedades

a) La aplicación $\Phi$ es semismooth

b) Si $\nabla F$, $\nabla^2 h_j$ $(j\in J)$, y $\nabla^2 g_i$ $(i\in I)$ son localmente Lipschitzianas, entonces $\Phi$ es estrictamente semismooth.

Como una consecuencia, obtenemos la siguiente proposición.

**Proposición 2.5.** Se cumplen las siguientes propiedades

a) Se verifica que

$$\|\Phi(w+h)-\Phi(w)-Hh\| = O(\|h\|)$$

para $h\to 0$ y $H\in\partial\Phi(w+h)$.

b) Si $\Phi$ es estrictamente semismooth, entonces

$$\|\Phi(w+h)-\Phi(w)-Hh\|=O(\|h\|^2)$$

para $h\to 0$ y $H\in\partial\Phi(w+h)$.

La primera parte del siguiente resultado es consecuencia de la semicontinuidad superior de la Jacobiana generalizada [1, Proposición 2.6.2 (c)] y de la suposición de la BD-regularidad; la segunda parte es un resultado estándar que es consecuencia de la suposición de la BD-regularidad.

**Proposición 2.6.** Sea $w^*$ una solución BD-regular de $\Phi(w)=0$. Entonces se verifican las siguientes propiedades:

a) Existen números $c_1>0$ y $\delta_1>0$ tales que las matrices $H\in\partial_B\Phi(w)$ son no singulares, y satisfacen
$\|H^{-1}\|\le c_1$, para todo $w$ con $\|w-w^*\|\le\delta_1$.

b) Existen números $c_2>0$ y $\delta_2>0$ tales que

$$\|\Phi(w)\| \geq c_2 \|w - w^*\|,$$

para todo $w$ con $\|w\text{-}w^*\|\leq\delta_2$.

Concluimos esta sección recordando un resultado no obvio, pero simple [6] que juega un papel crucial en el diseño y análisis de nuestros algoritmos.

**Preposición 2.7.** $\psi$ es continuamente diferenciable, y $\nabla\psi=H^T\Phi(w)$ para todo $H$ en $\partial\Phi(w)$.

## 3. Una reformulación de los KKT-sistemas

En esta sección consideramos puntos estacionarios del problema (5) y su relación con las soluciones del sistema (1). Sea el problema (5):

$$\min \Psi(w) \quad \text{sujeto a } z \geq 0,$$

recordamos que un punto $w^*=(x^*,y^*,z^*)\in\Re^{n+p+m}$ con $z^*\geq 0$, es un punto estacionario del problema sí $\nabla_x\Psi(w^*)=0$, $\nabla_y\Psi((w^*)=0$, y

$$z_i^* > 0 \Rightarrow \frac{\partial\Psi(\mathrm{w}^*)}{\partial \mathrm{z}_\mathrm{i}} = 0$$

$$z_i^* = 0 \Rightarrow \frac{\partial\Psi(\mathrm{w}^*)}{\partial \mathrm{z}_\mathrm{i}} \geq 0$$

En lo que sigue indicaremos por $I_>$ el conjunto de los índices para los que $z_i^* > 0$.

En el siguiente teorema damos condiciones que aseguran que un punto estacionario es una solución global y, por tanto, una solución del sistema (1)

**Teorema 3.1.** Sea $w^*=(x^*,y^*,z^*)\in\Re^{n+m+p}$ un punto estacionario. Supongamos que

I. $\nabla_x L(w^*)$ es semidefinida positiva en $\Re^n$

II. $\nabla_x L(w^*)$ es definida positiva en el cono

$$C(x^*) = \left\{ v\in\Re^n : \nabla h(x^*)^T v = 0, \begin{matrix} \nabla g_i(x^*)^T v = 0, (i\in I_>) \\ \nabla g_i(x^*)^T v \leq 0, (i\notin I_>) \end{matrix} \right\},$$

y que se cumple una de las dos condiciones siguientes

$c_1$) $\nabla h(x^*)$ tiene el rango columna completo

$c_2$) $h$ es una función afín, y el sistema $h(x)=0$ esta bien definido

entonces $w^*$ es una solución del KKT-sistema

La prueba la podemos ver en [2]

**Corolario 3.2.** Sea $w^*=(x^*,y^*,z^*)\in\Re^{n+p+m}$ un punto estacionario de (5). Supongamos que

I. $F$ es monótona, $h$ afín, y $g$ es cóncava (es decir cada función componente de $g$ es cóncava).

II. $\nabla_x L(w^*)$ está definido positivo en el cono X(x*)

III. el sistema $h(x)=0$ esta bien definido

Entonces $w^*$ es una solución del KKT-sistema.

Notemos que la suposición II del teorema 3.1 es tan buena como la suposición del corolario 3.2 usada en [11] para el caso de desigualdades variacionales monótonas. Aquí, al $\nabla F(x^*)$ le exigimos que sea definido positivo. Finalmente, puede ser interesante observar que la suposición II se satisfece en el caso de que $F$ sea estrictamente monótona

## 4. Algoritmos

### 4.1. Algoritmo de Facchinei

En esta sección describimos el algoritmo para la solución del problema (5) propuesto por Facchinei y otros [2].

Antes de establecer el algoritmo formalmente vamos a hacer algunas anotaciones. Primeramente recordemos que queremos resolver el sistema condicionado nonsmooth de ecuaciones

$$\Phi(w) = 0, \qquad z \geq 0. \tag{6}$$

Una forma simple de resolver (6) será el resolver iterativamente el sistema linealizado

$$H_k \Delta w = -\Phi(w^k), \qquad z^k + \Delta z \geq 0, \tag{7}$$

Donde $w^k$ es el vector iterante k-esimo, $H_k\in\partial_B\Phi(w^k)$, y $\Delta w=(\Delta x,\Delta y,\Delta z)\in\Re^{n+p+m}$. Es razonable resolver (7) por mínimos cuadrados, reemplazando el sistema condicionado lineal (7) por el siguiente problema lineal condicionado de mínimos cuadrados.

$$min\frac{1}{2}\left\|\Phi(w^k) - H_k\Delta w\right\|^2 \qquad \text{sujeto a } z^k + \Delta z \geq 0. \tag{8}$$

Ahora, teniendo en cuenta la proposición 2.7, es fácil ver que la función merit de (8) puede ser sustituida por

$$\begin{aligned}\frac{1}{2}\left\|\Phi(w^k) + H_k\Delta w\right\|^2 &= \frac{1}{2}\left(\Phi(w^k) + H_k\Delta w\right)^T\left(\Phi(w^k) + H_k\Delta w\right)\\ &= \Psi(w^k) + \nabla\Psi(w^k)^T\Delta w + \frac{1}{2}\Delta w^T H_k^T H_k \Delta w.\end{aligned}$$

Puesto que $\Psi(w^k)$ es una constante, el problema (8) es equivalente a

$$\min \nabla\Psi(w^k)^T \Delta w + \frac{1}{2}\Delta w^T H_k^T H_k \Delta w \qquad \text{sujeto a } z^k + \Delta z \geq 0. \qquad (9)$$

Si el sistema $\Phi(w)$ =0 fuese diferenciable sería simplemente una versión condicionada del método de Gauss-Newton. En este caso (9) podría no tener solución única. Por lo tanto, en base a resultados análogos en el caso smooth, vemos conveniente considerar alguna clase de modificación en la dirección de búsqueda del subproblema (9). En este trabajo consideramos una modificación tipo Levengerg-Maquardt. Para este fin, sea $\rho:\Re\to\Re_+$ una función fuerte, es decir, una función continua que toma valores no negativos y es 0 si y solo si su argumento es cero. El subproblema actualmente usado en el algoritmo sería el siguiente, versión regularizada de (9):

$$\min \nabla\Psi(w^k)^T \Delta w + \frac{1}{2}\Delta w^T \left(H_k^T H_k + \rho\,(\Psi(\mathrm{w}^k))I\right)\Delta w \quad \text{sujeto a } z^k + \Delta z \geq 0. \quad (10)$$

Nótese que si $w^k$ no es solución de (6) la matriz $H_k^T H_k + \rho\,(\Psi(w^k))I$ es definida positiva. Puesto que, el conjunto factible del programa cuadrático (10) es obviamente no vacío, el problema (10) admite siempre una solución única.

Podemos dar ahora una descripción formal del algoritmo. Básicamente resuelve el KKT-sistema resolviendo una sucesión de problemas (10).

**Algoritmo 4.1.1.** (Algoritmo basado en QP nonsmooth)

(S.0) (Dato inicial)

Elegir $w^0=(x^0,y^0,z^0)\in\Re^{n+p+m}$ con $z^0\geq 0$, $\sigma\in(0,1)$, $\beta\in(0,1)$ y $k:=0$

(S.1) Fin del criterio

Si $w^k$ es un punto estacionario de (5): Stop

(S.2) (Subproblema de programación cuadrática)

Seleccionamos un elemento $H_k\in\partial_B\Phi(w^k)$. Sea $\Delta w^k=(\Delta x^k,\Delta y^k,\Delta z^k)\in\Re^{n+p+m}$ la solución única del problema de programación cuadrática ($QP_k$):

$$\min \Psi(w^k)^T \Delta w + \frac{1}{2}\Delta w^T (H_k^T H_k + \rho(\Psi(w^k))I)\Delta w \qquad \textit{sujeto } \text{a } z^k + \Delta z \geq 0.$$

(S.3) (línea de búsqueda)

Sea $t_k:=\max\left\{\beta^l : l = 0,1,2,..\right\}$ tal que

$$\Psi(w^k + t_k \Delta w^k) \leq (1-\sigma t_k^2)\Psi(w^k) \qquad (11)$$

(S.4) Calcular

$$w^{k+1} = w^k + t_k \Delta w^k, \quad k = k+1. \text{ E ir a (S.1).}$$

Se observa fácilmente que cualquier sucesión $\{w^k\}=\{(x^k,y^k,z^k)\}\subset\Re^{n+p+m}$ generada por el algoritmo 4.1 mantiene factible el problema 5, es decir, $z^k\geq 0$ para todo $k$. También notemos que el $t_k$ en lado izquierdo de la inecuación (11) está elevado al cuadrado, lo que no es usual en las reglas de selección de paso.

El siguiente teorema prueba que el algoritmo está bien definido

**Teorema 4.1.2.** Sea $w^k=(x^k,y^k,z^k)\in\Re^{n+p+m}$ con $z^k\geq 0$ un vector arbitrario y sea $\Delta w^k\in\Re^{n+p+m}$ una solución del ($QP_k$). Tenemos que

$$\nabla\Psi(w^k)^T \Delta w^k \leq 0$$

en el caso de que $w^k$ no sea un punto estacionario del problema (5), se cumple que

$$\nabla\Psi(w^k)^T \Delta w^k < 0$$

Además, el algoritmo 4.1 está bien definido, en el sentido de que puede ser siempre obtenido un paso $t_k$ positivo (S.3).

La prueba del teorema se puede ver en [2].

### 4.2. Algoritmo de Kanzow

En esta sección, damos con detalle una descripción del algoritmo de Kanzow y la prueba de que está bien definido para cualquier problema de desigualdad variacional VIP(X,F).

**Algoritmo 4.2.**

(S.0) Inicialización

Elección de $w^0=(x^0,y^0,z^0)\in\Re^{n+p+m}$ con $z\geq 0$, $\sigma\in(0,1)$, $\beta\in(0,1)$, $\gamma\in(0,1)$, $c>0$, $\delta>0$, $\varepsilon\geq 0$, y $k:=0$.

(S.1) Estrategia del conjunto Activo

Sea

$$\delta_k := \min\left\{\delta, c\sqrt{\|\Phi(w^k)\|}\right\}$$

y definimos $J_k := \left\{j\in\Im \,\middle|\, w_j^k \leq \delta_k\right\}$

donde $\Im = \{n+p+1,\ldots,n+p+m\}$

(S.2) (Fin del criterio)

Sea

$$h^k = \begin{pmatrix} h^k_{J_k} \\ h^k_{\bar{J}_k} \end{pmatrix}$$

definimos por

$$h^k_{J_k} = \min\left\{w^k_{J_k}, g^k_{J_k}\right\} \text{y } h^k_{\bar{J}_k} = g^k_{\bar{J}_k},$$

Donde $\bar{J} =: I \cup P \cup (J - J_k)$. Siendo $I:=\{1,...,n\}$ y $P:=\{n+1,...,n+p\}$.
Si $\|h^k\| \leq \varepsilon$, stop

(S.3) (Solución del subproblema)

Seleccionamos $H_k \in \partial\Phi(w^k)$ y escribimos $H_k=(H^h_{..J_k}, H^k_{.\bar{J}_k})$. Sea $d^k_{\bar{J}_k}$ la única solución del sistema lineal reducido

$$\left(\left(H^k_{.\bar{J}_k}\right)^T H^k_{.\bar{J}_k} + \rho(\Psi(w^k))I\right) d^k_{\bar{J}_k} = -h^k_{\bar{J}_k} \qquad (12)$$

(S.4) Condición de seguridad de la factibilidad
Calculo de las direcciones de búsqueda

$$d^k := \begin{pmatrix} -w^k_{J_k} \\ d^k_{\bar{J}_k} \end{pmatrix} \text{ y } \bar{d}^k := \begin{pmatrix} -h^k_{J_k} \\ d^k_{\bar{J}_k} \end{pmatrix},$$

y definimos

$$\bar{\tau}_k := \sup\left\{\tau \middle| w^k_{J-J_k} + \tau d^k_{J-J_k} \geq 0\right\}$$

$$\tau_k := \min\{1, \bar{\tau}_k\}$$

(S.5) (Cálculo de la nueva iterante)

Sí $\Psi(w^k + \tau_k d^k) \leq \gamma\Psi(w^k)$ (13)
Entonces (*paso rápido)*
$w^{k+1} := w^k + \tau_k d^k$
o (*paso seguro)*
calculamos $t_k := \max\left\{\beta^l \middle| l = 0,1,2,...\right\}$ tal que

$$\Psi(w^k + \tau_k t_k \bar{d}^k) \leq (1 - \sigma\tau_k t_k^2)\Psi(w^k) \qquad (14)$$

y $w^{k+1} := w^k + \tau_k \bar{d}^k$ .

(S.6) Sea $k \leftarrow k+1$ e ir a (S.1)

Antes de analizar las propiedades del Algoritmo 4.2, vamos a hacer algunas aclaraciones: Primeramente, el conjunto $J_k$ definido en el paso (S.1) es utilizado como una aproximación para el conjunto de condiciones activas:

$$J_* := \left\{ j \in \Im \middle| w_j^* = 0 \right\}$$

La definición de $h^k$ en el paso (S.2) juega un papel crucial en la definición de la dirección de búsqueda $\bar{d}^k$ del paso (S.4) y por tanto en el apartado del paso seguro (S.5). En (S.2), esto sólo se usa para terminar la iteración. Finalizar el criterio será motivada por el lema 4.2.1 que nos dice que $h^k=0$ sólo si $w^k$ está cercano al punto estacionario.

La motivación básica de considerar sistema lineal del paso (S.3) es usar la ecuación estándar de Newton

$$H_k d = -\Phi(w^k)$$

para el operador $\Phi$ y usando una regularización tipo Levengerg-Marquardt de este

$$(H_k^T H_k + \rho(\Psi(w^k))I)d = -H_k^T \Phi(w^k) = -g^k, \qquad (15)$$

El sistema (1) puede ser ahora derivado de (15) teniendo en cuenta que $J_k$ es una aproximación de $J_*$.

La principal idea del paso (S.4) es calcular el paso más grande que garantice que la siguiente iterante $w^{k+1}=(x^{k+1},y^{k+1},z^{k+1})$ también satisfagan la condición de factibilidad $z^{k+1} \geq 0$. Nótese que el valor de $\tau_k$ es siempre positivo puesto que $w_j^k > 0$ para todo $j \in J - J_k$, y que esto es una simple expresión para $\tau_k$, a saber

$$\tau_k = \min\left\{1, \min_{\langle j \in J - J_k \left| d_j^k < 0 \right. \rangle} \left\{ w_j^k \middle| d_j^k \right\} \right\}.$$

En el paso (S.5), utilizamos las dos direcciones de búsqueda $d^k$ y $\bar{d}^k$ calculadas en el paso anterior. Nótese, sin embargo, que a veces no hay una gran diferencia entre estas dos direcciones; ambas dependen de la solución $d_{J_k}^k$ del sistema lineal (12) y, en efecto, coinciden completamente si $w_j^k \leq g_j^k$ para todo $j \in J_k$. En general, no obstante, estos dos vectores son diferentes para cada componente $j \in J_K$, y esta diferencia es esencial en nuestro análisis de la convergencia del algoritmo 4.2. La dirección $d^k$ se usa para probar la convergencia local bajo la condición de regularidad fuerte de Robinson. Esto garantizará la convergencia global al punto estacionario.

Iniciamos el análisis del algoritmo 4.2. Siempre supondremos implícitamente que el parámetro $\varepsilon$ en el algoritmo 4.2 es igual a cero, y que el algoritmo no termina después de un número finito de iteraciones. Esto es una suposición razonable puesto que el resultado siguiente prueba que el iterante $w^k=(x^k,y^k,z^k)$ estará alrededor de un punto estacionario de (5).

**Lema 4.2.1.-** Sea $w^k=(x^k,y^k,z^k)\in\Re^{n+p+m}$ con $z^k \geq 0$ un punto arbitrario. Entonces las siguientes propiedades son equivalentes

a) $w^k$ es un punto estacionario de (5)
b) $(g^k)^T h^k = 0$
c) $h^k=0$

**Lema 4.2.2.-** Sea $w^k=(x^k,y^k,z^k) \in\Re^{n+p+m}$ con $z^k \geq 0$ un punto arbitrario. Entonces

$$\nabla\Psi\left(w^k\right)^T \bar{d}^k \leq 0.$$

Además, si $w^k$ no es un punto estacionario de (5), entonces

$$\nabla\Psi\left(w^k\right)^T \bar{d}^k < 0.$$

**Corolario 4.2.1.** Sea $w^k=(x^k,y^k,z^k) \in\Re^{n+p+m}$ con $z^k \geq 0$ un punto arbitrario, y supongamos que $w^k$ no es un punto estacionario de (5). Entonces hay un $l_k$ tal que la longitud de paso $t_k=\beta^{l_k}$ satisfaga el test de la línea de búsqueda (13) en el algoritmo 4.2.

**Lema 4.2.3.-** Sea $w^k=(x^k,y^k,z^k) \in\Re^{n+p+m}$ con $z^k \geq 0$ un punto arbitrario, y supongamos que $w^k$ no es un punto estacionario de (14). Entonces la iterante $w^{k+1}$ puede ser calculada por el algoritmo 4.2., y se cumple que $z^{k+1} \geq 0$.

Las pruebas de estos lemas pueden ser examinadas en [7]

Puesto que la matriz del sistema lineal (12) está definida positiva cuando $w^k$ no es una solución global de (5), podemos aplicar un argumento inductivo y resumir los resultados previos en el siguiente

**Teorema 4.2. -** El algoritmo 4.2. está bien definido y genera una sucesión $\{w^k\} = \left\{\left(x^k, y^k, z^k\right)\right\}$ con $z^k \geq 0$ para todo $k$.

## 5. Convergencia de los algoritmos.

La convergencia del algoritmo de Facchinei viene garantizada por el siguiente teorema [2]

**Teorema 5.1.** Sea $\{w^k\}$ una sucesión finita generada por el Algoritmo 4.1, y sea $w^*$ un punto de acumulación de esta sucesión. Si $w^*$ es una BD-regular solución del sistema $\Phi(w) \geq 0$, entonces las siguientes propiedades se cumplen

a) La sucesión $\{w^k\}$ converge a $w^*$.
b) Existe un índice $k_0$ tal que $t_k=1$ para todo $k \geq k_0$.
c) La razón de convergencia es Q-superlineal

d ) La razón de convergencia es Q-cuadrática, cumpliéndose que $\nabla F, \nabla^2 h_j (j \in J),$ y $\nabla^2 g_i \ (i \in I)$ son localmente Lipschtzianas, y si $\rho(\Psi(w^k)) = O(\Psi(w^k))$.

La convergencia del algoritmo de Kanzow viene dada por los siguientes resultados [7]

**Teorema 5.2.** Sea $\{w^k\}$ una sucesión generada por el Algoritmo 4.2.. Todo punto de acumulación de $\{w^k\}$ es un punto estacionario de (5)

**Teorema 5.3.-** Supongamos que $w^*$ es un KKT punto regular fuerte del VIP($X,F$), y supongamos que $w^*$ es un punto de acumulación de una sucesión $\{w^k\}$ generada por el algoritmo 4.2. Entonces las siguientes propiedades se cumplen

*a)* La sucesión $\{w^k\}$ converge a $w^*$

b) Finalmente, el algoritmo toma sólo pasos rápidos

c) La razón de convergencia es Q-superlineal

d ) La razón de convergencia es Q-cuadrática si se cumple que $\nabla F, \nabla^2 h_j (j \in J),$ y $\nabla^2 g_i \ (i \in I)$ son localmente Lipschtzianas, y si $\rho(\Psi(w^k)) = O(\Psi(w^k))$.

## 6. Conclusiones

En este trabajo hemos llegado a la conclusión que ambos algoritmos llegan a las mismas razones de convergencia tanto superlineal como cuadráticamente. Las soluciones del problema general de desigualdades variacionales usando métodos tipo-Newton tienen un número considerable de propiedades. El hecho de que estos métodos resuelvan sistemas de ecuaciones lineales en cada iteración es también aplicable a problemas de amplia-escala que frecuentemente aparecen para la discretización de problemas continuos.

La relación existente entre problemas de desigualdades variacionales y problemas de complementariedad y complementariedad mixta hace que es posible utilizar estos algoritmos para esta clase de problemas. Sin embargo debido a la estructura especial de estos problemas, sería más conveniente usar cualquier método de multiplicadores de Lagrange.

La sencillez de los métodos del conjunto activo hace que nos inclinemos por el segundo algoritmo puesto que la literatura existente es mucho más amplia y el número de implementaciones existente de estos métodos nos permite una aplicación más efectiva.

## Bibliografía.

[1] Clarke FH (1983) "Optimization and Nonsmooth Análisis". Wiley, New York

[2] Facchinei, F. Fisher, A. Kanzow, C. Peng, JM (1999) "A Simply Constrained Optimization Reformulation of KKT Systems Arising from Variational Inequalities". Applied Mathematics and Optimization 40, 19-37. Spring-Verlag. New York Inc

[3] Facchinei, F. Pang, JS (1998) "Total Stability of Variational Inequalities"

[4] Facchinei, F.Fisher, A. Kanzow, C. (1996) “Inexact newton methods for semismoth equations with applications to variational inequality problems”. Inc G.di Pillo and F. Giannessi (eds.): Nonlinear Optimization and Applications. Plenum Press, New York, pp. 125-149
[5] Facchinei, F. Fisher, A. Kanzow, C. (1997) “A semismooth Newton method for variational inequalities: the case of bos constraints”. In Ferris MC, Pang Js(eds) Complementary an Variational Problems: State of the Art. SIAM, Philadelfia, PA, pp 76-90
[6] Facchinei, F. Fisher, A. Kanzow, C. (1998) “Regularity properties of a semismooth reformulation of variational inequalities”, SIAM J Optim 8:850-869
[7] Kanzow, C. Qi, HD. (1997) “A QP-free constrained Newton-type method for variational inequality problems”. Hamburger Beiträge zur Angewandten Mathematik. Reihe A. Preprint 121.
[8] Kanzow, C. Kleinmichel H (1995) “A class of Newton-type methods for equality and inequality constrained optimization. Optim Methods Sftware 5:173-198
[9] Kanzow, C. Jiang, H. (1994) “A continuation method for (strongly) monotone variational inequalities. Preprint 90, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany (revised January and November 1996).
[10] Peng, JM, Kanzow, C. Fukushima, M. (1998) “Ahibrid Josephy-Newton method for solving box contrained variation inequality problems via the D-Gap function.
[11] Peng JM (1995) “A global method for monotone variational inequality problems with ineauality constraints”. Technical Report, State Key Laboaroty of Scientfic and Engineering Computing. Academic Sinica. Beijing, 1995
[12]Robinson SM (1980) “Strongly regular generalized equations”. Math Oper Res 5. 43-62