



UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICA

Algunas Ideas de Aceleración del Método de Proyecciones Alternantes

Trabajo Especial de Grado presentado ante la ilustre Universidad Central de Venezuela por el **Br. Gledys Sulbarán** para optar al título de Licenciada en Matemática.

Tutores: Prof. Luis Manuel Hernandez y Prof René Escalante

Caracas, Venezuela

Abril 2008

Nosotros, los abajo firmantes, designados por la Universidad Central de Venezuela como integrantes del Jurado Examinador del Trabajo Especial de Grado titulado “**Algunas ideas de Aceleración del Método de Proyecciones Alternantes**”, presentado por la **Br. Gledys Sulbarán**, titular de la Cédula de Identidad **17.313.160**, certificamos que este trabajo cumple con los requisitos exigidos por nuestra Magna Casa de Estudios para optar al título de **Licenciada en Matemática**.

Prof. Luis Manuel Hernández.

Tutor

Prof. René Escalante.

Tutor

Prof. Francisco Tovar.

Jurado

Prof. José Luís Sánchez.

Jurado

Dedicatorias

A mi madre, ella me dá las fuerzas para seguir estudiando.

A mi hermana Glendy y a mi novio Anthony

.

Agradecimientos

A Dios por darme salud y fuerzas para levantarme cuando estoy decaída.

A mi tutores, el profesor Luis Manuel Hernández y el profesor René Escalante por toda la atención, la paciencia y el conocimiento que siempre me brindaron. Muchas gracias por creer en mí.

A mi madre Gladys Goyo, por estar a mi lado en todo momento, por levantarse temprano todos los días para tener todo listo, por tanta dedicación, por ser más que una madre, es mi mejor amiga.

A mi hermanita Glendy por apoyarme y ayudarme cuando la necesito.

A mi tío Orlando Goyo, por ser el mejor tío del mundo y estar siempre pendiente de mí.

A mi novio Anthony (Mi pollito) por su eterna paciencia y comprensión, por su amor incondicional que cada día me das, llenándome de alegrías.

A mis amigos Gisell Marcano y Freddy Hernandez, por todas las noches que compartimos estudiando, a Gisell por explicarme cuando no entendía nada y a Freddy por su alegría y ocurrencias que siempre me hacen reír.

A todos y a cada uno de mis compañeros: Alejandra Ruiz y Andrés Contreras por ayudarme cuando estaba en el hospital, no olvidaré ese detalle que tuvieron. A Nelson y Julio los primeros amigos que conocí en la universidad, a Mariolys y Adolfo por estar siempre dispuestos ayudarme vía internet. a Lili y Adelis por contagiarme sus risas.

Índice general

Introducción	1
1. Preliminares	4
1.1. Espacio de Hilbert	4
1.2. Operadores en Espacio de Hilbert	5
1.3. Ángulo entre Subespacios	6
2. El Método de Proyecciones Alternantes	11
2.1. Introducción	11
2.2. Proyecciones Alternantes	11
2.2.1. Algoritmo MAP	14
2.3. Velocidad de Convergencia	16
2.3.1. Velocidad de Convergencia para Proyecciones Alternantes	16
2.4. Aceleración del Método MAP	18
2.5. El Método de Kaczmarz	19
2.6. Áreas de Aplicación	21
3. Métodos tipo Gradiente	22
3.1. Introducción	22
3.1.1. Direcciones de Descenso	22
3.2. Método de Cauchy	23
3.3. Barzilai-Borwein	25

3.4. Método del Gradiente Espectral	26
3.4.1. Caso Cuadrático	27
3.5. Método Gradiente Proyectado	28
3.6. Método del Gradiente Espectral Proyectado	29
4. MAP Acelerado con Optimización	31
4.1. Introducción	31
4.2. Preliminares	31
4.3. El método de Von-Neumann desde el punto de vista de optimización . . .	33
4.4. Equivalencia Koshy-Gearhart con Gradiente Proyectado	33
4.5. Propuestas del tamaño de paso α_k	35
5. Experimentación Numérica	37
5.1. Experimentos	42
5.1.1. Experimento 1	43
5.1.2. Experimento 2	45
5.1.3. Experimento 3	47
5.1.4. Experimento 4	49
5.1.5. Experimento 5	51
5.1.6. Experimento 6	52
5.1.7. Experimento 7	53
6. Conclusiones y Recomendaciones	54
Bibliografía	56

Introducción

El método de proyecciones alternantes (MAP) fue propuesto en 1933 por John Von Neumann [26], quien trató el problema de encontrar la proyección de un punto dado x_0 en un espacio de Hilbert H sobre la intersección de dos subespacios cerrados. El método MAP tiene variadas aplicaciones en campos como el tratamiento de imágenes, medicina, probabilidad y estadística, problema de Dirichlet, solución de ecuaciones lineales. etc [13, 16, 8, 22]. En el artículo [11] se puede encontrar un resumen de todo lo relacionado con el MAP y sus aplicaciones y asimismo en el libro en preparación de Escalante y Raydan [12].

La velocidad de convergencia de este método depende del ángulo entre los sub-espacios involucrados. Cuando el ángulo es pequeño, el esquema iterativo propuesto por Von Neumann presenta convergencia lenta. Esto significa un alto costo computacional ya que es necesario realizar muchas iteraciones del método para conseguir una buena aproximación. Diversos esquemas han sido propuestos para acelerar el MAP, entre estos podemos destacar el esquema de aceleración propuesto por Koshy-Gearhart [15].

En este trabajo, se reinterpreta el método de proyecciones alternantes como un método tipo gradiente proyectado para cierto problema de optimización. Nos damos cuenta que el método clásico MAP es equivalente al método gradiente proyectado a paso constante y que el esquema de aceleración propuesto por Koshy-Gearhart es equivalente al

algoritmo clásico de gradiente proyectado. Basados en esto, proponemos otras escogencias de la longitud de paso, especialmente proponemos una versión espectral del MAP como esquema de aceleración, con muy buenos resultados experimentales. Esta versión está basada en el método de Gradiente Espectral Proyectado propuesto por Raydan et al. [3].

Finalmente, mostramos algunos resultados experimentales de cálculos de proyecciones sobre la intersección de ciertos subespacios vectoriales de matrices.

Capítulo 1

Preliminares

El objetivo de éste capítulo es presentar algunas definiciones, teoremas y proposiciones importantes que serán de gran utilidad para los capítulos siguientes.[12]

1.1. Espacio de Hilbert

Definición 1.1. *Sea A un espacio métrico. Decimos que A es completo cuando toda sucesión de Cauchy converge a un punto de A .*

Definición 1.2. *Un Espacio de Hilbert H , es un espacio con producto interno que es completo con respecto a la norma dada por el producto interno.*

Teorema 1.3. *Sea H un espacio de Hilbert. Todo subconjunto cerrado y no vacío de H contiene un único elemento de norma mínima.*

Definición 1.4. *Dado un espacio de Hilbert H y una variedad lineal M de H , el ortogonal de M es:*

$$M^\perp = \{z \in H : \langle z, x \rangle = 0\}.$$

Para todo $x \in M$.

Proposición 1.5. *Sean H un espacio de Hilbert y M una variedad lineal de H entonces $M^\perp$ es un subespacio cerrado de H .*

Teorema 1.6. *Sea H un espacio de Hilbert y M un subespacio cerrado de H entonces para cada $x \in H$ existe un único $w \in M$ y un único $z \in M^\perp$ tal que $x = w + z$.*

Corolario 1.7. *Sean H un espacio de Hilbert y M un subespacio cerrado de H entonces*

$$H = M + M^\perp$$

donde $M^\perp$ es un subespacio cerrado de H , que es ortogonal a M .

Proposición 1.8. *En un espacio de Hilbert el único vector que es ortogonal a todos los vectores es el vector cero*

1.2. Operadores en Espacio de Hilbert

A continuación se dan algunos conceptos básicos, este resumen está basado en [10], [12]. Sea H un espacio de Hilbert y sea M un subespacio cerrado de H . Denotemos por P_M el operador de Proyección ortogonal sobre M . Este es un operador lineal, autoadjunto ($\langle Px, y \rangle = \langle x, Py \rangle$) e idempotente ($P^2 \equiv P$). P_M es tal que:

$$\|x - P_M(x)\| = d(x, M)$$

donde $d(x, M) = \inf\{\|x - y\|, y \in M\}$, es decir $P_M(x)$ es el punto más cercano a x en M . Su propiedad característica es:

$$\langle x - P_M(x), y \rangle = 0, \forall y \in M.$$

El Complemento ortogonal de M , denotado por $M^\perp$ es:

$$M^\perp := \{y \in H : \langle x, y \rangle = 0, \forall x \in M\}$$

Definición 1.9. *Un operador en H es una función T que está definido por algún subespacio S de H y que tiene uno o más valores T_x en H correspondiente a cada elemento x de S .*

Denotaremos S el dominio $D(T)$ de T y el conjunto de todos los valores T_x , $x \in S$, el Rango $R(T)$ de T .

Definición 1.10. Sea $T : X \longrightarrow Y$ un operador, donde X, Y son subespacios normados. T es acotado si existe un escalar M tal que $\|Tx\| \leq M\|x\| \quad \forall x \in X$.

Teorema 1.11. Un operador T es una proyección P_S si y solo si

- (i) T es inyectiva, lineal, acotado y $D(T)=H$.
- (ii) $\langle T_x, y \rangle = \langle x, T_y \rangle \quad \forall x, y \in H$.
- (iii) $T^2 = T$.

Teorema 1.12.

- (i) Si U y V son proyecciones en H . Entonces una condición suficiente y necesaria para que UV sea una proyección es que $UV = VU$
- (ii) Si $U = P_M$ y $V = P_N$ (M y N son subespacios cerrados), Entonces $UV = P_{M \cap N}$

1.3. Ángulo entre Subespacios

Recordemos que si $x, y \in H$, el ángulo θ entre x e y está definido como el ángulo cuyo coseno está dado por:

$$\cos \theta = \frac{\langle x, y \rangle}{\|x\| \|y\|}$$

La siguiente definición introducido originalmente por Friedrichs en 1937, es la más aceptada en la literatura del MAP para trabajar con el ángulo entre subespacios.

Definición 1.13. (Friedrichs, 1939) [12] El ángulo $\theta(M, N)$ entre dos subespacios cerrados M y N de H es el ángulo en $[0, \pi/2]$ cuyo coseno $C(M, N) = \cos \theta(M, N)$ esta dado por:

$$C(M, N) := \sup\{|\langle x, y \rangle| : x \in M \cap (M \cap N)^\perp, \|x\| \leq 1, y \in N \cap (N \cap M)^\perp, \|y\| \leq 1\}$$

Otros autores definen el ángulo $\theta(M, N)$ sin considerar el factor $(M \cap N)^\perp$ en la expresión anterior.

Definición 1.14. (Dixmier, 1947)[12] *El mínimo ángulo $\theta_0(M, N)$ entre dos subespacios cerrados M y N de H es el ángulo en $[0, \pi/2]$ cuyo coseno $C_0(M, N) = \cos \theta_0(M, N)$ esta dado por:*

$$C_0(M, N) := \sup\{|\langle x, y \rangle| : x \in M, \|x\| \leq 1, y \in N, \|y\| \leq 1\}$$

Observación 1.15. *Sean M, N subespacios cerrados en H . Entonces*

1. *Es claro que si $M \cap N = \emptyset$ ambas definiciones coinciden. .*
2. *Las siguientes son algunas consecuencias inmediatas de las definiciones.*
 - a) $0 \leq C(M, N) \leq C_0(N, M) \leq 1$.
 - b) $C(M, N) = C(N, M)$ y $C_0(M, N) = C_0(N, M)$.
 - c) $C(M, N) = C_0(M \cap (M \cap N)^\perp, N \cap (M \cap N)^\perp)$.
 - d) $|\langle x, y \rangle| \leq C_0(M, N) \|x\| \|y\| \forall x \in M, y \in N$.

Los siguientes lemas sobre proyecciones ortogonales son en su mayoría conocidos y se pueden encontrar en [10].

Lema 1.16. *Sea P un operador lineal acotado en H . Entonces P es una proyección ortogonal si y solo si $P^2 = P$ y $P^* = P$.*

Lema 1.17. *Las siguientes afirmaciones son equivalentes*

1. P_M y P_N conmutan: $P_M P_N = P_N P_M$.
2. $P_M P_N = P_{M \cap N}$.
3. $P_M P_N$ es una proyección ortogonal.

Lema 1.18. *Las siguientes afirmaciones son equivalentes*

1. $P_M P_N = P_M$.

2. $P_N P_M = P_M$.

3. $M \subset N$.

En particular, ambos P_M y P_N conmutan con $P_{M \cap N}$ y $P_M P_{M \cap N} = P_{M \cap N} = P_N P_{M \cap N}$.

Lema 1.19. *Las siguientes afirmaciones son equivalentes*

1. P_M conmuta con P_N .

2. $P_{M^\perp}$ conmuta con P_N .

3. P_M conmuta con $P_{N^\perp}$.

4. $P_{M^\perp}$ conmuta con $P_{N^\perp}$.

5. $M = M \cap N + M \cap N^\perp$.

Lema 1.20. *Sea M, N subespacios cerrados en H . Entonces*

1. $C(M, N) = C_0(M, N \cap (M \cap N)^\perp) = C_0(M \cap (M \cap N)^\perp, N)$.

2. $C_0(M, N) = \|P_M P_N\| = \|P_M P_N P_M\|^{1/2}$.

3. $C(M, N) = \|P_M P_N - P_{M \cap N}\| = \|P_M P_N P_{(M \cap N)^\perp}\|$.

Demostración. de 1:

$$\begin{aligned}
C(M, N) &= C_0(M \cap (M \cap N)^\perp, N \cap (M \cap N)^\perp). \\
&= \sup\{|\langle x, y \rangle| : x \in M \cap (M \cap N)^\perp, \|x\| \leq 1, y \in M \cap (M \cap N)^\perp, \|y\| \leq 1\} \\
&= \sup\{|\langle P_{M \cap (M \cap N)^\perp} x, P_{N \cap (M \cap N)^\perp} y \rangle| : \|x\| \leq 1, \|y\| \leq 1\} \\
&= \sup\{|\langle P_{(M \cap N)^\perp} P_M x, P_{(M \cap N)^\perp} P_N y \rangle| : \|x\| \leq 1, \|y\| \leq 1\} \\
&\quad (\text{por los Lemas 1.18 y 1.19}) \\
&= \sup\{|\langle P_M x, P_{(M \cap N)^\perp} P_N y \rangle| : \|x\| \leq 1, \|y\| \leq 1\} \\
&\quad (\text{Por Lema 1.16}) \\
&= \sup\{|\langle P_M x, P_{N \cap (M \cap N)^\perp} y \rangle| : \|x\| \leq 1, \|y\| \leq 1\} \\
&\quad (\text{Por los Lemas 1.18 y 1.19}) \\
&= \sup\{|\langle x, y \rangle| : x \in M, \|x\| \leq 1, y \in N \cap (M \cap N)^\perp, \|y\| \leq 1\} \\
&= C_0(M, N \cap (M \cap N)^\perp).
\end{aligned}$$

Por simetría se obtiene $C(M, N) = C_0(M \cap (M \cap N)^\perp, N)$. Demostración de 2

$$\begin{aligned}
C_0(M, N) &= \sup\{|\langle x, y \rangle| : x \in M, \|x\| \leq 1, y \in N, \|y\| \leq 1, \} \\
&= \sup\{|\langle P_M x, P_N y \rangle| : \|x\| \leq 1, \|y\| \leq 1\} \\
&= \sup\{|\langle x, P_M P_N y \rangle| : \|x\| \leq 1, \|y\| \leq 1\} \\
&= \|P_M P_N\|.
\end{aligned}$$

Lo que demuestra la primera igualdad de (2). La segunda igualdad se desprende de la primera y el hecho de que $\|P^* P\| = \|P\|^2$. Para cualquier operador lineal acotado P en H . Tomando $P = P_N P_M$ vemos que

$$P^* P = P_M P_N P_N P_M = P_M P_N P_M$$

Y por lo tanto

$$C_0(M, N) = \|P\| = \|P^* P\|^{1/2} = \|P_M P_N P_M\|^{1/2}.$$

Demostración de 3

$$\begin{aligned}
C(M, N) &= C_0(M \cap (M \cap N)^\perp, N \cap (M \cap N)^\perp) \text{ por (2) .} \\
&= \|P_{M \cap (M \cap N)^\perp} P_{N \cap (M \cap N)^\perp}\|. \\
&= \|P_M P_{(M \cap N)^\perp} P_N P_{(M \cap N)^\perp}\|. \\
&= \|P_M P_N P_{(M \cap N)^\perp}\| = \|P_M P_N (I - P_{M \cap N})\| \\
&\quad \text{como } M \cap N \subseteq M \text{ y } M \cap N \subseteq N. \\
&= \|P_M P_N - P_{M \cap N}\|.
\end{aligned}$$

□

Capítulo 2

El Método de Proyecciones Alternantes

2.1. Introducción

El Método de Proyecciones Alternantes en su versión original fue desarrollado por Von Neumann en 1933, quien trató el problema de encontrar la proyección de un punto dado en un espacio de Hilbert sobre la intersección de dos subespacios cerrados. Posteriormente Cheney y Goldstein extendieron el análisis de Von Neumann, alternando el esquema de proyección para el caso de dos espacios cerrados y convexos. Estudiaremos este caso primero para motivar el resultado más general. Este resumen se hace basado en [12], [9] y [21]

2.2. Proyecciones Alternantes

Sean A y B subespacios cerrados de H . Entonces $P_A P_B = P_B P_A$ si y sólo si $P_A P_B = P_{A \cap B}$. Es decir, P_A y P_B conmutan si y solo si su composición es también una proyección ortogonal. Von Neumann estuvo interesado en el caso que P_A y P_B no conmutan, probando el siguiente resultado:

Teorema 2.1. (Von Neumann [1933]) [26] Sean A y B subespacios cerrados en un espacio de Hilbert H . Entonces, para cada $x \in H$

$$\lim_{n \rightarrow \infty} (P_B P_A)^n(x) = P_{A \cap B}(x)$$

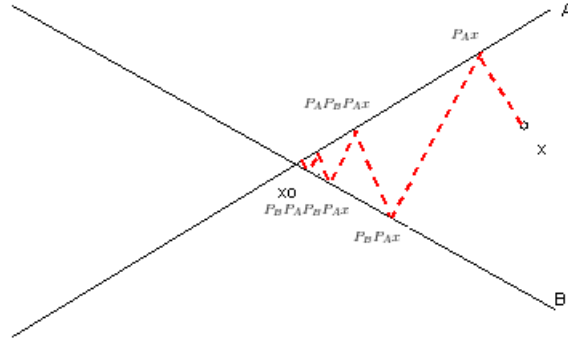


Figura 2.1: Proyecciones Alternantes

Demostración. La prueba que daremos a continuación es la prueba original de Von Neumann [26]. Una prueba diferente puede encontrarse en F. Deutsch [9].

Sea $U = P_M$ y $V = P_N$ y dos sucesiones

$$\Sigma_1 : U, VU, UVU, VUVU, ..$$

$$\Sigma_2 : V, UV, VUV, UVUV, ..$$

Sea T_n el n -ésimo operador de la sucesión Σ_1 . Entonces $\langle T_m x, T_n x \rangle = \langle T_{m+n-\delta} x, y \rangle$

$$\delta = \begin{cases} 1 & \text{si } m \text{ y } n \text{ tienen la misma paridad} \\ 0 & \text{si } m \text{ y } n \text{ tienen paridades opuesta} \end{cases}$$

Veamos que la sucesión $\{T_n\}$ tiene límite

$$\begin{aligned}
 \|T_mx - T_nx\|^2 &= \langle T_mx - T_nx, T_mx - T_nx \rangle \\
 &= \langle T_mx, T_mx \rangle - \langle T_mx, T_nx \rangle - \langle T_nx, T_mx \rangle + \langle T_nx, T_nx \rangle \\
 &= \langle T_{2m-1}x, x \rangle + \langle T_{2n-1}x, x \rangle - 2\langle T_{m+n-\delta}x, x \rangle \\
 &= \langle T_{2m-1}x, x \rangle + \langle T_{2n-1}x, x \rangle - 2\langle T_{2k-1}x, x \rangle
 \end{aligned} \tag{2.1}$$

Puesto $m+n-\delta$ es siempre impar, el último término de esta expresión fue reescrito, con k enteros. Por otro lado, como

$$\langle T_{2k-1}x, x \rangle = \langle T_ix, T_ix \rangle = \|T_ix\|^2$$

se deduce que

$$\|T_{i+1}x\|^2 = \langle T_{2i+1}x, x \rangle.$$

$T_{i+1}x$ es UT_ix o VT_ix , luego

$$\begin{aligned}
 \langle T_{i+1}T_ix, T_ix \rangle &= \langle T_{2i}x, x \rangle \\
 &= \langle T_{i+1}x, T_{i+1}x \rangle \\
 &= \|T_{i+1}x\|^2 \\
 &\leq \|T_ix\|^2.
 \end{aligned}$$

Entonces

$$\|T_{i+1}x\| \leq \|T_ix\|.$$

Por lo tanto

$$\langle T_{2i-1}x, x \rangle \geq \langle T_{2i+1}x, x \rangle,$$

y

$$\langle T_1x, x \rangle \geq \langle T_3x, x \rangle \geq \langle T_5x, x \rangle \geq \dots \geq 0.$$

De modo que $\lim_{i \rightarrow \infty} \langle T_{2i-1}x, x \rangle$ existe y por (2.1) Se deduce que

$$\lim_{m,n \rightarrow \infty} \|T_mx - T_nx\|^2 = 0.$$

En Consecuencia $\{T_n x\}$ es de Cauchy, por lo tanto, $\lim_{n \rightarrow \infty} T_n x$ existe. Llamemoslo x^* .

Definimos $T : H \longrightarrow H$ tal que $Tx = x^*$, entonces $D(T) = H$ y además T es inyectiva.

T es lineal (pues $\lim_{n \rightarrow \infty} T_n(x - \lambda y) = \lim_{n \rightarrow \infty} T_n x - \lambda \lim_{n \rightarrow \infty} T_n y = Tx - Ty$)

Además, como

$$\lim_{m, n \rightarrow \infty} \langle T_m x, T_n y \rangle = \lim_{m, n \rightarrow \infty} \langle T_{m+n-\delta} x, y \rangle$$

Se deduce que

$$\langle Tx, Ty \rangle = \langle Tx, y \rangle.$$

Entonces T es una proyección P_L .

Ahora si $x \in M \cap N$, entonces $Ux = Vx = x$, $T_n x = x$, y $Tx = x \in L$ por lo tanto $M \cap N \subseteq L$.

Por otro lado $UT_{2i} = T_{2i+1}$ y $VT_{2i-1} = T_{2i}$, si $i \rightarrow \infty$ tenemos $UT = T$ y $VT = T$ (ya que U y V son operadores continuos). Ahora, para cualquier $y \in H$ sea $Ty = x \in L$. Entonces, $Ux = UTy = Ty = x \in M$ y $Vx = VTy = Ty = x \in N$. Por lo tanto $x \in M \cap N$ entonces $L \subseteq M \cap N$. En consecuencia, $L = M \cap N$.

Finalmente, intercambiando U y V en la argumento anterior se sigue que Σ_2 tiene límite $T' = P_{M \cap N}$, entonces $T = T'$, por lo que la prueba está completa. $\square$

2.2.1. Algoritmo MAP

El Teorema de Von Neumann induce un algoritmo iterativo (Algoritmo 1) para calcular la proyección en la intersección de dos subespacios cerrados, suponiendo más sencillo proyectar individualmente sobre cada uno de los subespacios involucrados.

La interpretación geométrica de este método consiste en encontrar la mejor aproximación al punto x desde $A \cap B$, primero se proyecta x sobre A , el elemento que resulta entonces se proyecta sobre B , y se continúan proyectando los resultados alternadamente sobre A y B . La sucesión de los elementos generados así converge a $P_{A \cap B}(x)$. La utilidad práctica de este procedimiento proviene del hecho de que, en general, resulta mucho más fácil calcular las proyecciones sobre A o B individualmente que calcular la proyección sobre $A \cap B$.

Algoritmo 1 MAP

- Dado x_0
 - **Para** $n = 1, \dots$, hasta la convergencia **hacer**
 - $x_0 = x$
 - $x_n = P_B P_A x_{n-1} (= (P_B P_A)^n x)$
 - **FinPara**
-

El resultado obtenido por Von Neumann fue extendido posteriormente por Halperin, a una familia finita de subespacio cerrados.

Teorema 2.2. (Halperin 1962). Sean $M_1, M_2, \dots, M_r$ subespacios cerrados en H y $M = \cap_{i=1}^r M_i$. Entonces

$$\lim_{n \rightarrow \infty} (P_{M_r} P_{M_{r-1}} \dots P_{M_2} P_{M_1})^n(x) = P_{\cap_{i=1}^r M_i}(x)$$

para cada $x \in H$.

Demostración. Ver [12]

□

Cabe mencionar que la demostración del teorema de Halperin requiere aplicar un método diferente cuando $k \geq 3$ que el dado por Von Neumann para el caso $k = 2$ donde k es el número de los subespacios involucrados.

En 1983 Frank Deutsch hace la generalización al caso de la intersección de un número finito de variedades lineales.

Teorema 2.3. (Deutsch, 1983 [8]) Si $M_1, M_2, \dots, M_r$ son variedades lineales cerradas en un espacio de Hilbert H . Entonces, para cada $x \in H$.

$$\lim_{n \rightarrow \infty} (P_{M_r} P_{M_{r-1}} \dots P_{M_2} P_{M_1})^n(x) = P_{\cap_{i=1}^r M_i}(x)$$

2.3. Velocidad de Convergencia

La velocidad de convergencia del Método de Proyecciones Alternantes ha sido estudiada por varios autores, entre ellos: Smith, Solmon y Wagner [24], Kayalar y Weinert [17] y F. Deutsch [11]. En general, la velocidad de convergencia de El MAP es r-lineal y depende del ángulo entre los subespacios o variedades lineales involucrados.

2.3.1. Velocidad de Convergencia para Proyecciones Alternantes

Caso dos subespacios

En el próximo teorema se demuestra que la velocidad de convergencia del Método de Proyecciones Alternantes depende del ángulo entre los subespacios involucrados. Del teorema tenemos que si el ángulo entre los subespacios es pequeño el método se hace lento. [9].

Teorema 2.4. (*Aronszajn 1950*): Sean M_1 y M_2 subespacios cerrados en H y $C = C(M_1, M_2)$. Entonces, para cada $x \in H$

$$\| (P_{M_2}P_{M_1})^n(x) - P_{M_1 \cap M_2}(x) \| \leq C^{2n-1} \| x - P_{M_1 \cap M_2}(x) \| \leq C^{2n-1} \| x \|$$

para $n = 1, 2, \dots$ por otra parte, la constante C^{2n-1} es la más pequeña posible.

En la siguiente figura se puede apreciar la dependencia del ángulo en la velocidad de convergencia para el caso de dos subespacios A y B. En el ejemplo de la figura 2.2 el método realiza pocas iteraciones para llegar a la solución a diferencia del ejemplo de la figura 2.3 que converge más lento ya que realiza más iteraciones para llegar a la solución

Caso n subespacios

Smith, Solmon, and Wagner(1977) obtuvieron una cota análoga en el caso de más de dos subespacios.

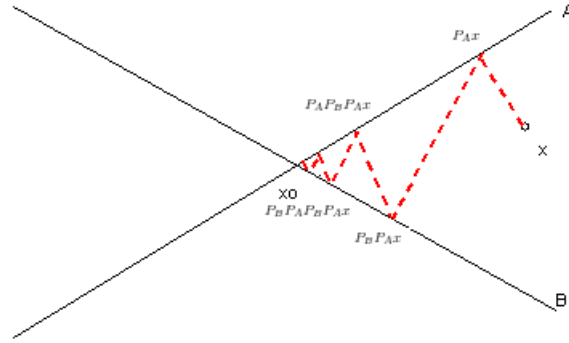


Figura 2.2: Método MAP: Ángulo Grande

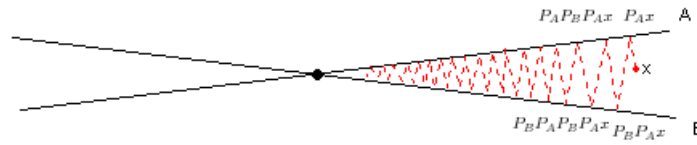


Figura 2.3: Método MAP: Ángulo Pequeño

Teorema 2.5. (Smith, Solmon, Wagner 1977 [8]): Sean $M_1, M_2, \dots, M_k$ subespacios cerrados en H y Sea $M = \cap_{i=1}^k M_i$. Entonces, para cada $x \in H$ y un entero $n \geq 1$

$$\| (P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^n(x) - P_{\cap_{i=1}^k M_i}(x) \| \leq C^n \| x \|$$

Donde

$$C = [1 - \prod_{i=1}^{k-1} \sin^2 \theta_i]^{1/2}$$

Y θ_i es el Ángulo entre el subespacio M_i y $\cap_{j=i+1}^k M_j$

- Kayalar y Weinert (1988)[17] dieron una mejor cota que el teorema 2.4, pero mucho más complicado
- Franchetti y Light (1986) [14] demostró que la convergencia en el teorema 2.1 puede ser arbitrariamente lenta si $A + B$ no es cerrada (es decir si el ángulo entre A y B es

cero). Sin embargo, Gearhart y Koshy (1989) desarrollaron esquemas para acelerar la convergencia del MAP. Dyer (1965)[7] había descrito una técnica para acelerar la convergencia del método de Kaczmarz.

2.4. Aceleración del Método MAP

Anteriormente se vió que el MAP tiene una velocidad de convergencia lenta cuando los ángulos entre los subespacios involucrados son pequeños. Se presentarán una de las técnicas de aceleración más utilizadas con el propósito de encontrar la proyección de un punto dado sobre la intersección de un número finito de subespacios en un espacio de Hilbert [12]. Esta aceleración la propone Gearhart y Koshy [15].

Para describir la técnica de aceleración, denotaremos a x^k como la k -ésima iteración y Qx^k el próximo iterado después de haber aplicado un ciclo de el *Método de Proyecciones Alternantes* es decir $Qx^k = P_{M_k} P_{M_{k-1}} \cdots P_{M_2} P_{M_1}$ donde $M_1, M_2, \dots, M_k$ son subespacios cerrados en H . La idea es, investigar a través de la línea que une los puntos x^k y Qx^k , cual es el punto más cercano a la solución Px . Un punto en dicha línea está representado por:

$$x_t^k = tQx^k + (1 - t)x^k, \text{ para cualquier número real } t$$

y denotaremos por t_k el valor de t para el cual este punto es el más cercano a Px , consecuentemente $(x_{t_k}^k - Px)$ debe ser ortogonal a $(x^k - Qx^k)$, es decir

$$\langle x_{t_k}^k - Px, x^k - Qx^k \rangle = 0. \quad (2.2)$$

Además, puesto que $Px \in M$ y las proyecciones P_i son autoadjuntas, se tiene que:

$$\langle Px, Qx^k \rangle = \langle P_1 \dots P_r Px, x^k \rangle = \langle Px, x^k \rangle.$$

Entonces $\langle Px, x^k - Qx^k \rangle = 0$ y a partir de 2.2 podemos eliminar Px obteniendo

$$\langle x_{t_k}^k, x^k - Qx^k \rangle = 0.$$

Resolviendo para t_k tenemos:

$$t_k = \frac{\langle x^k, x^k - Qx^k \rangle}{\|x^k - Qx^k\|^2}. \quad (2.3)$$

Y el método de la aceleración se puede entonces describir como sigue: Dado $x^0 = x \in H$, se calcula en cada iteración k , Qx^k , aplicando el Método de Proyecciones Alternantes, luego se calcula t_k por 2.3 y se hace $x^{k+1} = x_{t_k}^k$ continuamente.

2.5. El Método de Kaczmarz

Este método iterativo para encontrar soluciones a los sistemas de ecuaciones lineales es llamado el método de Kaczmarz puesto que primero fue estudiado por Kaczmarz (1937).

Unas de las aplicaciones más importante que trataremos en este trabajo es resolver el sistema de ecuaciones lineales que se estudiará a continuación.

Consideremos el problema de resolver el sistema de ecuaciones lineales

$$Ax = b.$$

Donde A es una matriz real $m \times n$, $x \in \mathbf{R}^n$ y $b \in \mathbf{R}^m$. Este problema puede ser generalizado para cualquier espacio de Hilbert H , encontrando un punto (si la solución existe) en la intersección de los m hiperplanos o variedades lineales cerradas, dado por:

$$H_i = \{x \in H : \langle a_i, x \rangle = b_i\}.$$

Donde $a_i = (a_{i1}, a_{i2}, \dots, a_{im}) \in \mathbf{R}^m$, $x \in \mathbf{R}^m$ y $b_i \in \mathbf{R}$ $i = 1, \dots, k$.

Encontrar una solución de $Ax = b$ es equivalente a encontrar un punto en la $\bigcap_{i=1}^k H_i(x)$. Para encontrar tal punto, fije un iterado inicial $x_0 \in \mathbf{R}^m$ arbitrario e inductivamente se define

$$x_n = (P_{H_k} P_{H_{k-1}} \dots P_{H_2} P_{H_1})^n(x) x_{n-1} \quad (n=1,2,\dots)$$

Entonces el Método de Proyecciones Alternantes implica que

$$x_n \longrightarrow P_{\bigcap_{i=1}^k H_i}(x_0) =: y_0.$$

Esto es; “ y_0 ” satisface la ecuación $Ax = b$, donde $y_0 \in \mathbf{R}^m$ es la solución.

Note que a_i es ortogonal a H_i . Más aun si $z \notin H_i$, la proyección sobre H_i viene dada por la siguiente expresión:

$$P_{H_i}(z) = z + \frac{b_i - \langle a_i, z \rangle}{\langle a_i, a_i \rangle} a_i$$

Observación del Método de Kaczmarz

1. Por ser un método “Row-Action” trabaja solamente con una fila por cada iteración.
2. Para matrices “Sparse” el cálculo de P_{H_i} tiene un bajo costo computacional.

Siguiendo la idea del MAP, a partir de un $x_0 \in H$ arbitrario, un paso típico de el Método de Kaczmarz, puede describirse como

$$x_{k+1} = x_k + w_k(P_{H_{i_k}}(x_k) - x_k)$$

Donde $0 < \varepsilon \leq w_k \leq 2 - \varepsilon < 2$, para todo k y para un ε pequeño dado. Además, se sabe que la solución converge al punto más cercano, en el conjunto solución x_0 . Más aún, si el sistema tiene muchas soluciones, los iterados convergen a la solución de la norma mínima, si x_0 está en el espacio del rango de A^T .

Este método fue propuesto originalmente por Kaczmarz, quien estableció la convergencia de matrices cuadradas no singulares con $w_k = 1$ para todo k . Tanabe [25] extendió el análisis para sistemas inconsistentes; Hernan, Lent y Luts [18] introdujeron los parámetros de relajación. Muchos otros autores han dado generalizaciones y extensiones del método de Kaczmarz. Más adelante en este trabajo se mostrará una extensión que nosotros proponemos y se desarrollará para acelerar dicho método.

2.6. Áreas de Aplicación

En esta sección mencionaremos las diferentes áreas de matemáticas donde el MAP ha desempeñado un papel importante. Hemos dado ya algunos detalles de cómo el MAP se puede utilizar para solucionar ecuaciones lineales en la sección anterior (método de Kaczmarz) Enumeramos las diversas áreas del uso del método junto con algunos de los muchos autores que han contribuido a estos estudios.

1. **Solución de Ecuaciones Lineales.** Kaczmarz[8] , Tanabe [25] , Herman, Lent and Luts [18], Eggermont, Herman and Lent .
2. **Probabilidad y Estadística.** Wiener y Masani [27] han utilizado el MAP en la teoría lineal de la predicción. Burkholder y Chow [4] estudiaron cuando la convergencia de la norma en $L_2(\mu)$ para el MAP podría ser sustituido por la convergencia en casi todo punto.
3. **Problema de Dirichlet.** Schwarz [22] Describió lo que él llamó el “Alternierende Verfahren” (= Método alternante) para solucionar el problema de Dirichlet en una región irregular en el plano que es la unión de regiones regulares.
4. **Cálculos del Núcleo de Bergman** En 1983 Skwarczynski [23] demostró cómo utilizar el MAP para calcular el núcleo de Bergman para los espacios de Hilbert $L_2(D)$.
5. **Aproximación de funciones Multivariable.** Deutsch [8] .
6. **Sistema de ecuaciones no lineales.** Martinez [19] .
7. **Restauración de Imágenes** René Escalante utilizó algoritmos numéricos para el problema de restauración de imágenes usando el Método de las Proyecciones Alternantes. [13] .

Capítulo 3

Métodos tipo Gradiente

3.1. Introducción

En este capítulo introduciremos algunos métodos para resolver problemas de minimización. Estos métodos están basados en tomar el negativo del gradiente como dirección de búsqueda, por esa razón se les denomina métodos de tipo gradiente. En particular, nos interesa la forma de estos métodos de minimización en el caso particular donde q es una función cuadrática. Es decir que puede ser escrita como:

$$q(x) = \frac{1}{2}x^tAx - b^tx + c \quad (3.1)$$

donde $A \in \mathbb{R}^{n \times n}$, es una matriz simétrica, los vectores $x, b \in \mathbb{R}^n$ y la constante $c \in \mathbb{R}$. Este capítulo está basado en el libro [20].

3.1.1. Direcciones de Descenso

Dada una función continuamente diferenciable $f : \mathbb{R}^n \rightarrow \mathbb{R}$, se dice que una dirección $d \in \mathbb{R}^n$ es de descenso a partir de un punto $x \in \mathbb{R}^n$ si

$$\nabla f(x)^t d < 0.$$

Si $\nabla f(x) \neq 0$ entonces x no es un minimizador y por ende en cualquier entorno alrededor de x existe $z \in \mathbb{R}^n$ tal que $f(z) < f(x)$.

Lema 3.1. Sean $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$, $x \in \mathbb{R}^n$ tal que $\nabla f(x) \neq 0$, y $d \in \mathbb{R}^n$ tal que $\nabla f(x)^t d < 0$. Entonces existe $\bar{\alpha} > 0$ tal que $f(x + \alpha d) < f(x)$ para todo $\alpha \in (0, \bar{\alpha}]$

Demostración. Definamos $\phi(\alpha) = f(x + \alpha d)$. Claramente $\phi(0) = f(x)$ y $\phi'(0) = \nabla f(x)^t d < 0$. Sin embargo por definición

$$\phi'(0) = \lim_{\alpha \rightarrow 0^+} \frac{\phi(\alpha) - \phi(0)}{\alpha}$$

Entonces para $\bar{\alpha} > 0$ suficientemente pequeño el signo de $\phi'(0)$ y de $(\phi(\alpha) - \phi(0))$ es igual, y por lo tanto $(\phi(\alpha) - \phi(0)) < 0$ para todo $\alpha \in (0, \bar{\alpha}]$. En consecuencia, $f(x + \alpha d) < f(x)$ para todo $\alpha \in (0, \bar{\alpha}]$. $\square$

Por el lema 3.1.1, se puede definir una familia muy amplia de método iterativos para encontrar puntos mínimos, cuya iteración k viene dada por

$$x_{k+1} = x_k + \alpha_k d_k,$$

donde d_k es una dirección de descenso y α_k es una longitud de paso que garantiza cierto tipo de descenso en la función objetivo. Diferentes formas de escoger d_k y diferentes políticas de descenso en f producen diferentes métodos.

El conjunto de direcciones de descenso a partir de un punto x es siempre no vacío, ya que al menos podemos definir la dirección $d = -\nabla f(x)$. Mas aún, esta dirección conocida como la dirección de gradiente negativo, es la que garantiza más rápido descenso local de la función f . En efecto, $d = -\nabla f(x) / \|\nabla f(x)\|$ resuelve el problema de optimización: minimizar $\nabla f(x)^t d$ sujeto a que $d \in \mathbb{R}^n$ y $\|d\|^2 = 1$.

3.2. Método de Cauchy

El método de Cauchy utiliza la dirección del gradiente negativo y longitud de paso dada por:

$$\alpha_k = \min_{\alpha > 0} f(x_k - \alpha g_k).$$

Introducimos la notación $g_k = \nabla f(x) = Ax_k - b$, $\nabla^2 f(x_k) = A$, El método de Cauchy aplicado en el caso en que $f(x)$ es una cuadrática 3.1 se puede escribir como:

$$x_{k+1} = x_k - \alpha_k g_k,$$

explícitamente se puede determinar que la longitud de paso α_k viene por:

$$\alpha_k = \frac{g_k^t g_k}{g_k^t A g_k}.$$

El algoritmo viene dado por:

Algoritmo 2 Cauchy

- **Entrada:** $x^{(0)}, A, b, M$
 - **Salida:** $0, x^{(0)}$
 - **Para** $k = 0, 1, 2, \dots, M - 1$ **hacer**
 1. $v^{(k)} \leftarrow b - Ax^k$
 2. $t_k \leftarrow \frac{\langle v^k, v^k \rangle}{\langle v^k, Av^k \rangle}$
 3. $x^{k+1} \leftarrow x^k + t_k v^k$
 - **Salida:** $k + 1, x^{k+1}$
 - **FinPara**
-

El método de mínimo descenso es raramente usado en el problema 3.1 por su lentitud. A continuación se muestra el recorrido del método del mínimo descenso en una cuadrática.



Figura 3.1: Interpretación geométrica del mínimo descenso

3.3. Barzilai-Borwein

Consideremos ahora la aplicación del método del gradiente espectral al problema general de minimización sin restricciones:

$$\min_{x \in \mathbb{R}^n} f(x) \quad (3.2)$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$. cada iteración del método se define como:

$$x_{k+1} = x_k - \frac{1}{\alpha_k} g_k, \quad (3.3)$$

donde α_k viene dado por

$$\alpha_k = \frac{s_{k-1}^t y_{k-1}}{s_{k-1}^t s_{k-1}}, \quad (3.4)$$

donde $s_{k-1} = x_k - x_{k-1}$ y $y_{k-1} = g_k - g_{k-1}$.

La dirección de búsqueda es siempre la dirección de gradiente negativo, sin embargo la longitud de paso no es la clásica asociada al método de mínimo descenso, también conocido como método de Cauchy. En efecto Barzilai y Borwein observan que esta escogencia de la longitud de paso requiere mucho menos trabajo computacional y acelera fuertemente la convergencia del método del gradiente en el caso de funciones cuadráticas convexas. Más aún, este método es nomonótono, es decir, no garantiza descenso en la función objetivo en cada iteración. En 1988, Barzilai y Borwein usando teoría de números, presentan un análisis de convergencia en el caso de funciones cuadráticas en solo dos variables. Más adelante, Raydan establece convergencia global en el caso de funciones

cuadráticas estrictamente convexas con cualquier número de variables. Este resultado se extiende un poco después al caso de cuadráticas convexas (no necesariamente convexas) por Friedlander, Martinez y Raydan.

3.4. Método del Gradiente Espectral

Con el deseo de resolver problemas de minimización sin restricciones, consideremos el siguiente problema asociado.

$$\text{encontrar } x_* \in \mathbb{R}^n \text{ tal que } \nabla f(x_*) = 0. \quad (3.5)$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es la función diferenciable a la cual se le desea encontrar mínimos locales. La solución numérica del sistema de ecuaciones no lineales 3.5 es obtenida, generalmente, mediante un esquema iterativo, moviéndose en cada iteración de un estimado x_k de x_* a un mejor estimado x_{k+1} . En muchos algoritmos, cada iteración envuelve el cálculo de un paso *Casi-Newton* $S_{NC} = -A_k^{-1} \nabla f(x_k)$, donde $A_k \in \mathbb{R}^{n \times n}$ es una aproximación del Hessiano de f en x_k . Al finalizar cada iteración, A_k se actualiza y se obtiene A_{k+1} , una aproximación del Hessiano de f en x_{k+1} . Esta aproximación, usualmente, se escoge de forma tal que satisfaga la *ecuación de la secante*,

$$A_{k+1} s_k = y_k, \quad (3.6)$$

donde $s_k = x_{k+1} - x_k$ y $y_k = g_{k+1} - g_k$, donde $g_k = \nabla f(x_k)$.

En el caso unidimensional, la ecuación de la secante define completamente el valor de A_{k+1} ; sin embargo, si $n > 1$, muchas matrices satisfacen la ecuación de la secante. Por lo tanto, además de satisfacer 3.6, la matriz A_{k+1} debe restringirse a un conjunto de matrices que posean propiedades deseables (simétricas, positivo definidas, etc.). La propiedad fundamental que caracteriza a los métodos espectrales está basada en la siguiente observación: el escalar $\alpha_{k+1} \in \mathbb{R}$ que resuelve de forma única el sistema lineal sobredeterminado $y_k = \alpha_{k+1} s_k$, en el sentido de los mínimos cuadrados, está dado por

$$\alpha_{k+1} = \frac{s_k^t y_k}{s_k^t s_k} \quad (3.7)$$

si $s_k \neq 0$. Es decir, restringiendo la matriz A_{k+1} a la familia de múltiplos escalares de la identidad y exigiendo que la ecuación de la secante se satisfaga en el sentido de los mínimos cuadrados, se obtiene el siguiente algoritmo:

Algoritmo 3 Gradiente Espectral

- Dados $x_0 \in \mathbb{R}^n$ y $\alpha_0 \in \mathbb{R}$

 - **Para** $k = 1, \dots$, hasta la convergencia **hacer**
 1. $s_k = -\frac{1}{\alpha_k} g_k$
 2. $x_{k+1} = x_k + s_k$
 3. $y_k = g_{k+1} - g_k$
 4. $\alpha_{k+1} = \frac{s_k^t y_k}{s_k^t s_k}$

 - **FinPara**
-

3.4.1. Caso Cuadrático

Si consideramos el problema 3.5 cuando $q(x) = \frac{1}{2}x^t A x - b^t x + c$ es una función cuadrática y A es una matriz simétrica y PD, entonces α_{k+1} en 3.7 se escribe como

$$\alpha_{k+1} = \frac{s_k^t A s_k}{s_k^t s_k} = \frac{g_k^t A g_k}{g_k^t g_k}. \quad (3.8)$$

Es importante notar que, en el método del gradiente espectral, la dirección de búsqueda en cada iteración es siempre el gradiente negativo de q en x_k , tal como sucede en el método de mínimo descenso. Sin embargo, en el caso cuadrático, el algoritmo se transformaría en el método clásico del gradiente (método de Cauchy) si en lugar de escoger α_{k+1} como en 3.8, lo escogiésemos como

$$\alpha_{k+1} = \frac{g_{k+1}^t A g_{k+1}}{g_{k+1}^t g_{k+1}}. \quad (3.9)$$

El método del gradiente espectral es mucho más rápido que el método de mínimo descenso al mismo costo computacional. [20].

Algoritmo 4 Método del Gradiente

- Dados $x_0 \in \mathbb{R}^n$
 - **Para** $k = 1, \dots$, hasta la convergencia **hacer**
 1. Escoger longitud de paso $\frac{1}{\alpha_k}$
 2. $s_k = -\frac{1}{\alpha_k} g_k$
 3. $x_{k+1} = x_k + s_k$
 - **FinPara**
-

Tanto el método del mínimo descenso como el método del Gradiente Espectral son casos particulares del Algoritmo 3. Solo se diferencia en la forma de escoger el α_k

3.5. Método Gradiente Proyectado

El método del gradiente proyectado se utiliza en problemas de minimización con restricciones.

$$\min_{x \in \Omega} f(x), \quad (3.10)$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ y Ω es un convexo cerrado.

El algoritmo de gradiente proyectado supone que es práctico y de bajo costo calcular la proyección sobre Ω , y se escribe como

$$x_0 \in \Omega$$

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k)$$

$$\text{donde } \bar{x}_k = P_\Omega(x_k - \lambda_k \nabla f(x)).$$

En el caso particular donde Ω es un subespacio cerrado, entonces el operador de proyección sobre Ω es un operador lineal, suponiendo $x_k \in \Omega$, $k = 1, \dots$, y el método queda como:

$$x_{k+1} = x_k - \alpha_k P_\Omega(\nabla f(x)).$$

La elección del tamaño de paso α_k puede hacerse mediante diferentes criterios. El método clásico de gradiente proyectado utiliza la selección de Cauchy para α_k , es decir α_k es tal que:

$$f(x_k - \alpha_k \nabla f(x_k)) = \min_{\alpha \geq 0} f(x_k - \alpha \nabla f(x_k)).$$

Como sabemos, en el caso cuadrático, este tamaño de paso corresponde a

$$\alpha_k = \frac{g_k^t g_k}{g_k^t H g_k}$$

donde $g_k = \nabla f(x_k)$, $H = \nabla^2 f(x)$

3.6. Método del Gradiente Espectral Proyectado

El Método del Gradiente Espectral Proyectado [3] surgió de la unión de la idea monótona de Barzila-Borwein (espectral) con las estrategias clásicas del gradiente Proyectado. Este método es aplicable a los problemas convexos limitada en la que la proyección sobre un conjunto factible es fácil de calcular. Desde su aparición, el método ha sido utilizado intensamente en muchas aplicaciones.

Raydan et al. proponen realizar la escogencia del tamaño de paso α_k según el método Gradiente Espectral en vez de la escogencia Cauchy. Este método se conoce como Gradiente Espectral Proyectado.

El método del gradiente espectral proyectado tiene como objetivo minimizar f en un conjunto cerrado y convexo Ω . El método tiene la forma:

$$x_{k+1} = x_k + \alpha_k d_k. \quad (3.11)$$

La dirección de búsqueda d_k ésta definido como:

$$d_k = P(x_k - \frac{1}{\sigma_k} \nabla f(x_k)) - x_k, \quad (3.12)$$

Donde P es la proyección Euclídeana sobre ω . La dirección d_k es una dirección de descenso, lo que significa que, si $d_k \neq 0$, se tiene que $f(x_k + \alpha d_k) \ll f(x_k)$ para α suficientemente pequeño. Esto significa que en principio se podría definir la convergencia del método imponiendo disminución suficiente en cada iteración. Sin embargo, esto lleva a desastrosos resultados prácticos. Por esta razón, los métodos espectrales emplean una búsqueda lineal no monótona que no impone disminución funcional en cada iteración.

Capítulo 4

MAP Acelerado con Optimización

4.1. Introducción

Cuando el ángulo entre los subespacios es “pequeño”, el método de Proyecciones Alternantes puede tener convergencia “lenta”. Esto significa un alto costo computacional ya que es necesario realizar muchas iteraciones del método para conseguir una buena aproximación. Diversos esquemas de aceleración han sido utilizados. Podemos citar los esquemas de aceleración de Pierro y Iusem [6], Gearhart y Koshy [15] y más recientemente los esquemas de H. Bauschke , F. Deutsch [2], de Censor [5] y de Smolarski y Appleby [1]. En este capítulo presentaremos un esquema para acelerar la convergencia del MAP desde el punto de vista de optimización con el fin de encontrar la proyección de un punto dado sobre la intersección de dos subespacios en un espacio de Hilbert.

4.2. Preliminares

Sea H un espacio de Hilbert, y sea P un operador lineal en H . Definimos la función

$$q(x) = \frac{1}{2} \|Px\|^2.$$

Calculamos el gradiente $\nabla q(x)$, definiendo para $d \in H$ dado la función auxiliar,

$$g(t) = q(x + td) = \frac{1}{2} \langle P(x + td), P(x + td) \rangle \quad (4.1)$$

$$= \frac{1}{2} (\|Px\|^2 + 2t \langle Px, Pd \rangle + t^2 \|Pd\|^2). \quad (4.2)$$

Derivamos la función auxiliar,

$$g'(t) = \langle Px, Pd \rangle + t \|Pd\|^2$$

$$g'(0) = \langle Px, Pd \rangle$$

Luego $\langle \nabla q(x), d \rangle = g'(0)$, por lo tanto,

$$\nabla q(x) = P^* P(x),$$

con P^* el operador adjunto de P .

Para calcular la Hessiana, derivamos de nuevo,

$$g''(t) = \|Pd\|^2 = \langle Pd, Pd \rangle.$$

Como $g''(0) = \langle d, \nabla^2 q(x) d \rangle$, en consecuencia tenemos que la Hessiana en este caso es una constante dada por el operador

$$\nabla^2 q(x) = P^* P.$$

En el caso particular donde P sea el operador de proyección en un subespacio, entonces P es idempotente y autoadjunta, por lo tanto,

$$\nabla q(x) = Px,$$

y además

$$\nabla^2 q(x) = P.$$

4.3. El método de Von-Neumann desde el punto de vista de optimización

Sean M, N subespacios de un espacio de Hilbert H . Y sean P_M y P_N los operadores de proyección ortogonal sobre M y N respectivamente. Nos interesamos ahora en el siguiente problema de optimización con restricciones:

$$\min_{x \in M} f(x), \quad (4.3)$$

donde $f(x) = \frac{1}{2} \|x - P_N(x)\|^2$ y P_N es el operador de proyección ortogonal sobre N .

Consideremos el método de Gradiente Proyectado:

$$\begin{aligned} x_{k+1} &= P_M(x_k - \alpha_k \nabla f(x)), \quad \text{para } k = 0, \dots, \\ x_0 &\in M. \end{aligned} \quad (4.4)$$

Como sabemos, $\nabla f(x) = x - P_N(x)$, y como todos los iterados $x_k \in M$ entonces $P_M(x_k) = x_k$, por ende el método de Gradiente Proyectado puede re-escribirse como:

$$\begin{aligned} x_{k+1} &= x_k - \alpha_k P_M \nabla f(x) \\ &= x_k - \alpha_k P_M P_N^\perp x_k. \end{aligned} \quad (4.5)$$

En el caso particular $\alpha \equiv 1$ el método (4.5) corresponde al método de Proyecciones Alternantes propuesto por Von-Neumann [26].

4.4. Equivalencia Koshy-Gearhart con Gradiente Proyectado

En el esquema propuesto por Gearhart y Koshy explicado en el capítulo 2 se tiene que:

$$x_{k+1} = t_k Q x_k + (1 - t_k) x_k.$$

Para M y N subespacios y $Q = P_M P_N$

$$\begin{aligned}
x_{k+1} &= x_k - t_k x_k + t_k P_M P_N x_k \\
&= x_k - t_k (x_k - P_M P_N x_k) \\
&= x_k + t_k (P_M P_N x_k - x_k) \\
&= x_k - t_k (P_M P_N^\perp x_k).
\end{aligned} \tag{4.6}$$

Veamos como es el parámetro de aceleración t

$$\begin{aligned}
t_k &= \frac{\langle x_k, x_k - Qx_k \rangle}{\|x_k - Qx_k\|^2} \\
&= \frac{\langle x_k, x_k - P_M P_N x_k \rangle}{\|x_k - P_M P_N x_k\|^2} \\
&= \frac{\langle x_k, P_M x_k - P_M P_N x_k \rangle}{\|P_M x_k - P_M P_N x_k\|^2} \\
&= \frac{\langle x_k, P_M (x_k - P_N) x_k \rangle}{\|P_M P_N^\perp x_k\|^2} \\
&= \frac{\langle x_k, P_M P_N^\perp x_k \rangle}{\|P_M P_N^\perp x_k\|^2}.
\end{aligned} \tag{4.7}$$

Como la ecuación (4.6) es equivalente a la ecuación (4.5) si, el parámetro $\alpha_k = t_k$ es escogido como el método de gradiente proyectado (escogencia de Cauchy), entonces hemos demostrado que el esquema de aceleración propuesto por Koshy y Gearhart [15] es equivalente en el caso de dos subespacios al Método del Gradiente Proyectado en el problema 4.3.

Por lo tanto, según la escogencia del parámetro α_k en el Método del Gradiente Proyectado pueden obtenerse los siguientes esquemas de Proyecciones Alternantes como se muestra a continuación:

$$\begin{aligned}
& \alpha \text{ fijo} \left\{ \begin{array}{ll} \alpha = 1 & \text{Von Neumann} \\ & x_{k+1} = P_M P_N x_k \\ \alpha = 2 & \text{Reflexión - Proyección} \\ & x_{k+1} = P_M R_N x_k, \quad R_N = 2P_A - I \end{array} \right. \\
& \alpha \text{ variable} \left\{ \begin{array}{ll} \alpha_k = \frac{\langle x_k, P_M P_N^\perp x_k \rangle}{\|P_M P_N^\perp x_k\|^2} & \text{Koshy - Gearhart} \\ & \text{corresponde a la escogencia Cauchy} \end{array} \right.
\end{aligned}$$

4.5. Propuestas del tamaño de paso α_k

Basado en lo anterior en este trabajo proponemos otras escogencia para el tamaño de paso en la aceleración del Método de Proyecciones Alternantes. En [20] Raydan et al. propone la utilización del Método del Gradiente Espectral Proyectado para resolver el problema de minimización de funciones cuadráticas sobre un conjunto convexo, que es el caso del problema de minimización con restricciones de la ecuación 4.3. El Método del Gradiente Espectral como se explicó anteriormente utiliza la misma dirección de búsqueda del Método de Gradiente Proyectado y varía en la escogencia del tamaño de paso α_k . Asi mismo, se proponen igualmente otras selecciones intuitivas del tamaño de paso α_k para su experimentación.

Los tamaños de paso que se proponen son los siguientes:

$$\alpha_k = \frac{\|x_k - x_{k-1}\|^2}{\langle x_k - x_{k-1}, P_N^\perp (x_k - x_{k-1}) \rangle}$$

que corresponde al Método del Gradiente Espectral Proyectado y su equivalente es:

$$\alpha_k = \frac{\langle P_M P_N^\perp x_{k-1}, P_M P_N^\perp x_{k-1} \rangle}{\langle P_M P_N^\perp x_{k-1}, P_N^\perp P_M P_N^\perp x_{k-1} \rangle}.$$

También proponemos probar:

$$\alpha_k \begin{cases} \frac{\langle P_M P_N^\perp x_k, P_M P_N^\perp x_k \rangle}{\langle P_M P_N^\perp x_k, P_N^\perp P_M P_N^\perp x_k \rangle} \\ \frac{\langle P_N^\perp x_k, P_N^\perp x_k \rangle}{\langle P_M P_N^\perp x_k, P_N^\perp x_k \rangle}. \end{cases}$$

El siguiente algoritmo se utilizará en el siguiente capítulo para la experimentación numérica utilizando diversos tamaño de pasos α_k propuestos.

Algoritmo 5 General

- Dados $x_0 \in M$ y $\alpha_k \in \mathbb{R}$
 - **Para** $k = 1, \dots$, hasta la convergencia **hacer**
 - Calcular $s_k = -\alpha_k P_M P_N^\perp x_k$
 - Calcular $x_{k+1} = x_k + s_k$
 - Escoger la longitud de paso α_k propuestas
 - **Si** $\|x - P_N x_k\|_F < tol$ **entonces parar**
 - **FinPara**
-

Capítulo 5

Experimentación Numérica

En este capítulo, haremos algunos experimentos para hallar la proyección de un punto x_0 dado, en la intersección de dos subespacios cerrados M y N . En esta experimentación observaremos el comportamiento de las iteraciones de los diferentes esquemas de aceleración propuestos y los compararemos con los esquemas clásicos de MAP y la aceleración propuesta por Koshy-Gearhart. En los experimentos que realizaremos, los subespacios en cuya intersección vamos a proyectar son subespacios de un espacio vectorial de matrices (los vectores pertenecen a $\mathbb{R}^{n \times n}$).

Los subespacios vectoriales de $\mathbb{R}^{n \times n}$ con los que trabajaremos son:

1. $\{A \in \mathbb{R}^{n \times n} / A = A^t\}$ (Subespacio de las matrices simétricas). Donde el operador de proyección “en la norma de Frobenius” viene dada por:

$$P(A) = \frac{A + A^t}{2}.$$

2. $\{A \in \mathbb{R}^{n \times n} / a(1, 1) = a(n, 1) = a(1, n) = a(n, n) = 0\}$. Donde el operador de

proyección “en la norma de Frobenius” viene dada por:

$$P(A) = \begin{pmatrix} 0 & a_{(1,2)} & \cdots & a_{(1,n-1)} & 0 \\ a_{(2,1)} & a_{(2,2)} & \cdots & a_{(2,n-1)} & a_{(2,n)} \\ \vdots & & \cdots & & \vdots \\ a_{(n-1,1)} & \cdots & & \cdots & a_{(n-1,n)} \\ 0 & a_{(n,2)} & \cdots & a_{(n,n-1)} & 0 \end{pmatrix}.$$

3. $\{A \in \mathbb{R}^{n \times n} / a(i, j) = 0 \text{ para todo } i > j \text{ } (i, j = 1, \dots, n)\}$. (Subespacio de las matrices triangulares superiores). Donde el operador de proyección “en la norma de Frobenius” viene dada por:

$$P(A) = \begin{pmatrix} a_{(1,1)} & a_{(1,2)} & a_{(1,3)} & \cdots & a_{(1,n)} \\ 0 & a_{(2,2)} & a_{(2,3)} & \cdots & a_{(2,n)} \\ 0 & 0 & a_{(3,3)} & \cdots & a_{(3,n)} \\ \vdots & & & \ddots & a_{(n-1,n)} \\ 0 & & 0 & \cdots & a_{(n,n)} \end{pmatrix}.$$

4. $\{A \in \mathbb{R}^{n \times n} / a(j, \frac{n+1}{2}) = \text{promedio para todo } j = 1, \dots, n. \text{ } n; \text{ impar, donde } \text{promedio} = \frac{\sum_{i=1}^n a(i, \frac{n+1}{2})}{n}\}$. Donde el operador de proyección “en la norma de Frobenius” viene dada por:

$$P(A) = \begin{pmatrix} a_{(1,1)} & a_{(1,2)} & \cdots & \text{promedio} & a_{(1, \frac{n+1}{2}+1)} & \cdots & a_{(1,n)} \\ a_{(2,1)} & a_{(2,2)} & \cdots & \text{promedio} & a_{(2, \frac{n+1}{2}+1)} & \cdots & a_{(2,n)} \\ \vdots & & & \vdots & & & \vdots \\ a_{(n,1)} & a_{(n,2)} & \cdots & \text{promedio} & a_{(n, \frac{n+1}{2}+1)} & \cdots & a_{(n,n)} \end{pmatrix}.$$

5. $\{A \in \mathbb{R}^{n \times n} / a(\frac{n+1}{2}, j) = \text{promedio}$ para todo $j = 1, \dots, n$. n ; impar, donde $\text{promedio} = \frac{\sum_{i=1}^n a(\frac{n+1}{2}, i)}{n}\}$. Donde el operador de proyección “en la norma de Frobenius” viene dada por:

$$P(A) = \begin{pmatrix} a_{(1,1)} & a_{(1,2)} & \cdots & \cdots & a_{(1,n)} \\ a_{(2,1)} & a_{(2,2)} & \cdots & \cdots & a_{(2,n)} \\ \vdots & & & & \vdots \\ a_{(\frac{n+1}{2}-1,1)} & & \cdots & \cdots & a_{(\frac{n+1}{2}-1,n)} \\ \text{promedio} & \text{promedio} & \cdots & \cdots & \text{promedio} \\ a_{(\frac{n+1}{2}+1,1)} & & \cdots & \cdots & a_{(\frac{n+1}{2}+1,n)} \\ \vdots & & & & \vdots \\ a_{(n,1)} & a_{(n,2)} & \cdots & \cdots & a_{(n,n)} \end{pmatrix}.$$

6. $\{A \in \mathbb{R}^{n \times n} / a(j, \frac{n+1}{2}) = a(\frac{n+1}{2}, k) = \text{promedio}$ para todo $k, j = 1, \dots, n$. n ; impar donde $\text{promedio} = \frac{\sum_{i=1}^n a(i, \frac{n+1}{2}) + \sum_{i=1}^n a(\frac{n+1}{2}, i)}{2n}\}$. Donde el operador de proyección “en la norma de Frobenius” viene dada por:

$$P(A) = \begin{pmatrix} a_{(1,1)} & a_{(1,2)} & \cdots & \text{promedio} & a_{(1, \frac{n+1}{2}+1)} & \cdots & a_{(1,n)} \\ a_{(2,1)} & a_{(2,2)} & \cdots & \text{promedio} & a_{(2, \frac{n+1}{2}+1)} & \cdots & a_{(2,n)} \\ \vdots & & & \vdots & & & \vdots \\ \text{promedio} & \text{promedio} & \cdots & \vdots & \cdots & & \text{promedio} \\ \vdots & & & \vdots & & & \vdots \\ a_{(n,1)} & a_{(n,2)} & \cdots & \text{promedio} & a_{(n, \frac{n+1}{2}+1)} & \cdots & a_{(n,n)} \end{pmatrix}.$$

7. $\{A \in \mathbb{R}^{n \times n} / a(1, j) = \text{promedio1}, \quad a(\frac{n+1}{2}, k) = \text{promedio2}$ para todo $k, j = 1, \dots, n$. n ; impar donde $\text{promedio1} = \frac{\sum_{i=1}^n a(1, i)}{n}, \text{promedio2} = \frac{\sum_{i=1}^n a(\frac{n+1}{2}, i)}{n}\}$. Donde

el operador de proyección “en la norma de Frobenius” viene dada por:

$$P(A) = \begin{pmatrix} \text{promedio1} & \text{promedio1} & \cdots & \cdots & \text{promedio1} \\ a_{(2,1)} & a_{(2,2)} & \cdots & \cdots & a_{(2,n)} \\ \vdots & & & & \vdots \\ a_{(\frac{n+1}{2}-1,1)} & & \cdots & \cdots & a_{(\frac{n+1}{2}-1,n)} \\ \text{promedio2} & \text{promedio2} & \cdots & \cdots & \text{promedio2} \\ a_{(\frac{n+1}{2}+1,1)} & & \cdots & \cdots & a_{(\frac{n+1}{2}+1,n)} \\ \vdots & & & & \vdots \\ a_{(n,1)} & a_{(n,2)} & \cdots & \cdots & a_{(n,n)} \end{pmatrix}.$$

Los parámetros del tamaño de paso con los cuales haremos la experimentación son los propuestos del capítulo anterior que son:

$$\alpha_k \begin{cases} \frac{\|x_k - x_{k-1}\|^2}{\langle x_k - x_{k-1}, P_N^\perp(x_k - x_{k-1}) \rangle} & \text{Gradiente Espectral Proyectado (GEP)} \\ \frac{\langle P_M P_N^\perp x_k, P_M P_N^\perp x_k \rangle}{\langle P_M P_N^\perp x_k, P_N^\perp P_M P_N^\perp x_k \rangle} & \text{(a)} \\ \frac{\langle P_N^\perp x_k, P_N^\perp x_k \rangle}{\langle P_M P_N^\perp x_k, P_N^\perp x_k \rangle} & \text{(b)}. \end{cases}$$

Y además haremos la comparación con el MAP propuesto por Von Neumann y la aceleración propuesta por Koshy-Gearhart.

El parámetro de Koshy-Gearhart calculado en el capítulo 4 viene dado por:

$$\alpha_k = \frac{\langle x_k, P_M P_N^\perp x_k \rangle}{\|P_M P_N^\perp x_k\|^2}.$$

Ahora bien,

$$\begin{aligned} \frac{\langle x_k, P_M P_N^\perp x_k \rangle}{\|P_M P_N^\perp x_k\|^2} &= \frac{\langle P_M x_k, P_N^\perp x_k \rangle}{\langle P_M P_N^\perp x_k, P_M P_N^\perp x_k \rangle} \text{ por ser } P_M \text{ autoadjunto} \\ &= \frac{\langle x_k, P_N^\perp x_k \rangle}{\langle P_M P_N^\perp x_k, P_N^\perp x_k \rangle} \text{ ya que } P_M \text{ autoadjunto e idempotente} \\ &= \frac{\langle P_N^\perp x_k, P_N^\perp x_k \rangle}{\langle P_M P_N^\perp x_k, P_N^\perp x_k \rangle} \text{ ya que } P_N^\perp \text{ autoadjunto e idempotente.} \end{aligned} \quad (5.1)$$

Analíticamente hemos demostrado en 5.1 que (b) es equivalente al método de Koshy-Gearhart. Sin embargo, como muestra los resultados numéricos el cálculo del coeficiente α_k según la fórmula $\alpha_k = \frac{\langle x_k, P_M P_N^\perp x_k \rangle}{\|P_M P_N^\perp x_k\|^2}$ presenta algunos problemas numéricos en el experimento 3 y 4 y que no sucede con el coeficiente calculado según (b).

En todo método iterativo es necesario especificar mediante ciertas condiciones cuando hay que detener el proceso. El criterio de parada utilizado tiene en cuenta dos factores:

1. La solución encontrada satisface los requerimientos de precisión (tolerancia). Se fijó una tolerancia de 1×10^{-8} .
2. La cantidad de iteraciones empleadas supera el máximo establecido.

Para el primer criterio, como cada iterado $x_k \in M$ entonces el criterio de parada del algoritmo puede tomarse como la distancia entre el iterado y su proyección sobre el subespacio N y utilizando la norma de Frobenius, es decir:

$$\|x - P_N x_k\|_F \quad (5.2)$$

Se muestra graficamente la distancia al cual se minimizará en la figura 5.1.

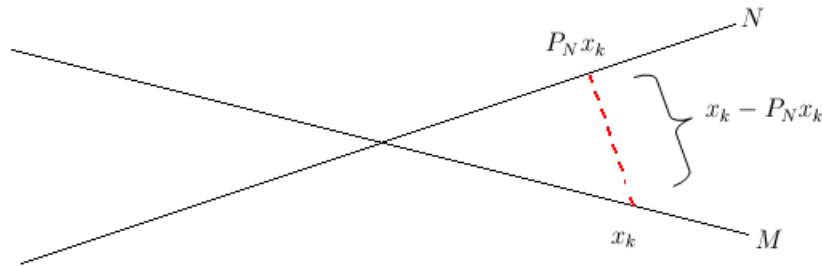


Figura 5.1: Distancia a minimizar

Para el segundo criterio el máximo número de iteraciones que establecimos es de 5000, el cual consideramos que el algoritmo no converge si pasa el dicho límite número de iteraciones.

Las matrices $A \in \mathbb{R}^{n \times n}$ fueron generadas en Matlab aleatoriamente. Los algoritmos requeridos para realizar la pruebas numéricas se desarrollaron en Matlab 7.0 y han sido efectuados con un computador Intel Core Duo 1.86 GHz 2Gb de RAM con Windows XP.

5.1. Experimentos

A continuación muestra el comportamiento de las convergencias en una tabla y gráficamente, donde el eje de ordenadas es el error de la solución y el eje de abcisas son las iteraciones realizadas por los métodos.

5.1.1. Experimento 1

$$x_{k+1} = x_k - \alpha_k P_M P_N^\perp x_k$$

$M := \{A \in \mathbb{R}^{n \times n} / a(i, j) = 0 \text{ para todo } i > j \text{ } (i, j = 1, \dots, n)\}$. (Subespacio de las matrices triangulares superiores).

$$N := \{A \in \mathbb{R}^{n \times n} / a(j, \frac{n+1}{2}) = \frac{\sum_{i=1}^n a(i, \frac{n+1}{2})}{n} \text{ para todo } j = 1, \dots, n; \text{ } n \text{ impar}\}.$$

Iteraciones						
Orden	MAP, $\alpha_k = 1$	$\alpha_k = 2$	Koshy-Gearhart	GEP	(a)	(b)
5×5	45	no converge	16	3	20	16
25×25	36	no converge	22	3	23	22
49×49	36	no converge	23	3	22	23
81×81	36	no converge	22	3	23	22

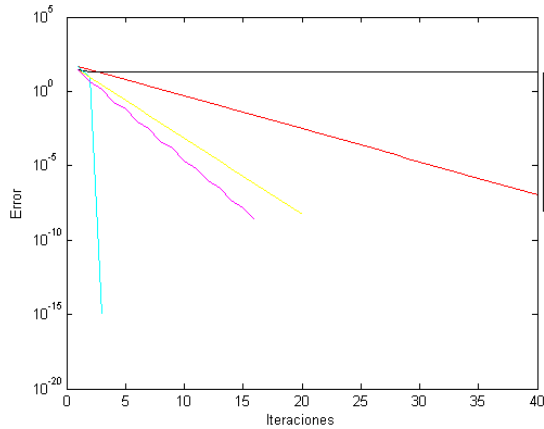


Figura 5.2: matriz 5×5

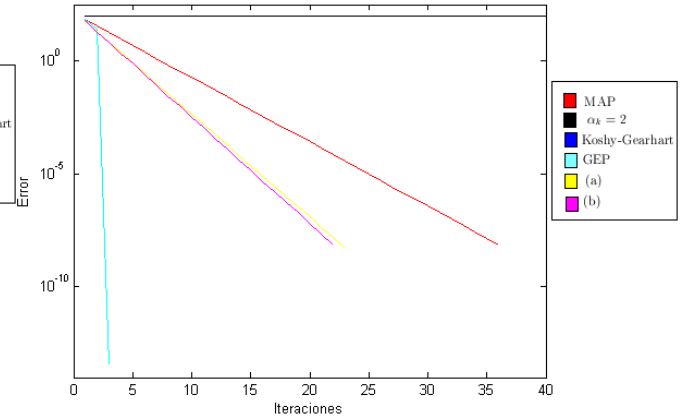
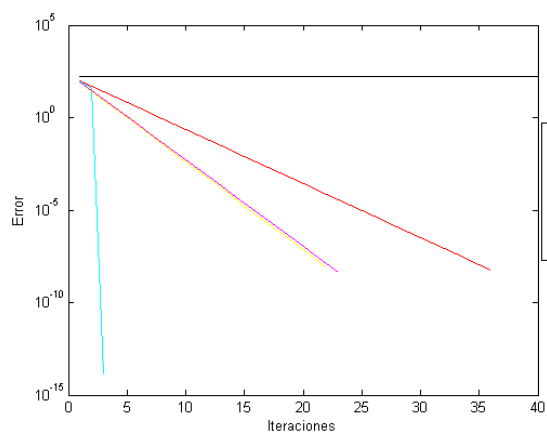
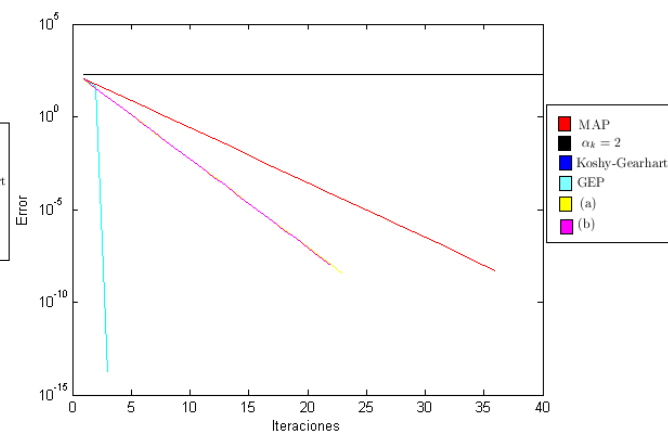


Figura 5.3: matriz 25×25

En las gráficas y en las tablas para el **experimento 1** se observa:

- Koshy-Gearhart se comporta igual en este experimento a (b) lo que corresponde con el análisis teóricos en la parte anterior.
- Se puede evidenciar que el método GEP converge en pocas iteraciones a la *tol* independientemente del orden de la matriz.

Figura 5.4: matriz 49×49 Figura 5.5: matriz 81×81

- El método con $\alpha_k = 2$ (fijo) no converge.
- Los parámetros propuestos GEP, (a) y (b) aceleran el MAP.

5.1.2. Experimento 2

$$x_{k+1} = x_k - \alpha_k P_M P_N^\perp x_k$$

$M := \{A \in \mathbb{R}^{n \times n} / a(i, j) = 0 \text{ para todo } i > j \text{ (} i, j = 1, \dots, n \text{)}\}$. (Subespacio de las matrices triangulares superiores).

$N := \{A \in \mathbb{R}^{n \times n} / a(j, \frac{n+1}{2}) = a(\frac{n+1}{2}, k) = \textit{promedio} \text{ para todo } k, j = 1, \dots, n. \textit{ } n \text{ impar donde } \textit{promedio} = \frac{\sum_{i=1}^n a(i, \frac{n+1}{2}) + \sum_{i=1}^n a(\frac{n+1}{2}, i)}{2n}\}$.

Iteraciones						
Orden	MAP, $\alpha_k = 1$	$\alpha_k = 2$	Koshy-Gearhart	GEP	(a)	(b)
5×5	46	no converge	19	3	16	21
25×25	37	no converge	23	3	23	16
50×50	36	no converge	23	3	22	23
100×100	36	no converge	23	3	23	22

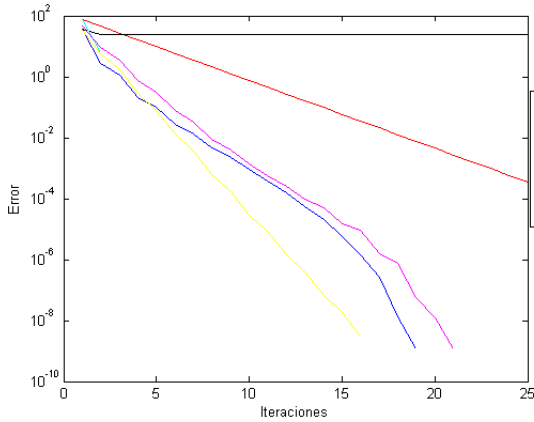


Figura 5.6: matriz 5×5

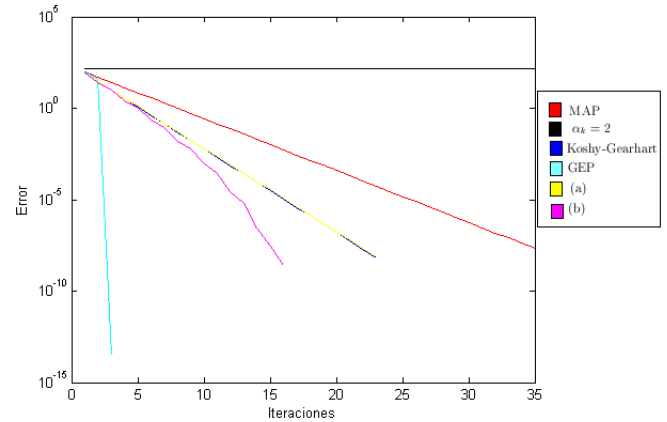
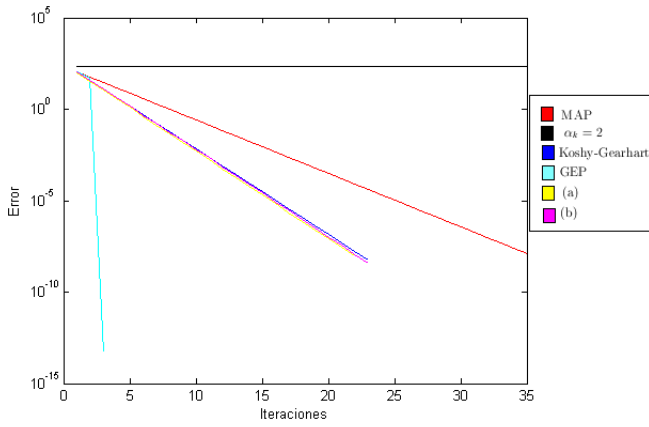
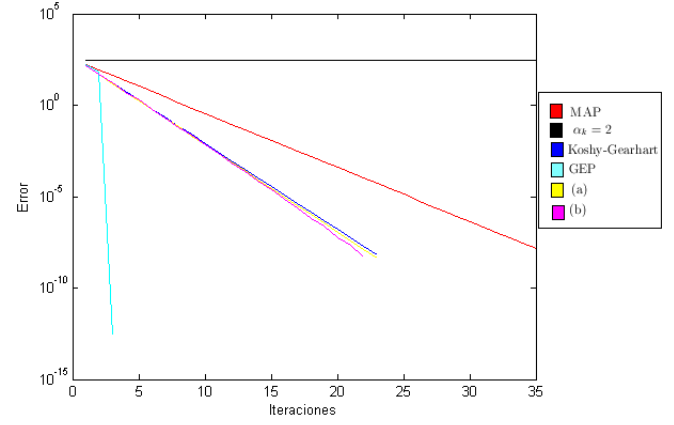


Figura 5.7: matriz 25×25

En las gráficas y en las tablas para el **experimento 2** se observa:

- Se puede evidenciar que el método GEP converge en pocas iteraciones a la *tol* independientemente del orden de la matriz.
- El método con $\alpha_k = 2$ (fijo) no converge.

Figura 5.8: matriz 49×49 Figura 5.9: matriz 81×81

- Se observa que a mayor orden de la matriz los parámetros de Koshy-Gearhart, (a) y (b) se comportan igual
- Los parámetros de GEP, (a) y (b) aceleran la convergencia.

5.1.3. Experimento 3

$$x_{k+1} = x_k - \alpha_k P_M P_N^\perp x_k$$

$M := \{A \in \mathbb{R}^{n \times n} / A = A^t\}$ (Subespacio de las matrices simétricas).

$N := \{A \in \mathbb{R}^{n \times n} / a(j, \frac{n+1}{2}) = \frac{\sum_{i=1}^n a(i, \frac{n+1}{2})}{n} \text{ para todo } j = 1, \dots, n; n \text{ impar}\}.$

Iteraciones						
Orden	MAP, $\alpha_k = 1$	$\alpha_k = 2$	Koshy-Gearhart	GEP	(a)	(b)
5×5	32	98	47	8	18	17
25×25	34	545	158	8	20	18
49×49	34	1001	2096	7	9	8
81×81	34	1782	2852	7	16	13

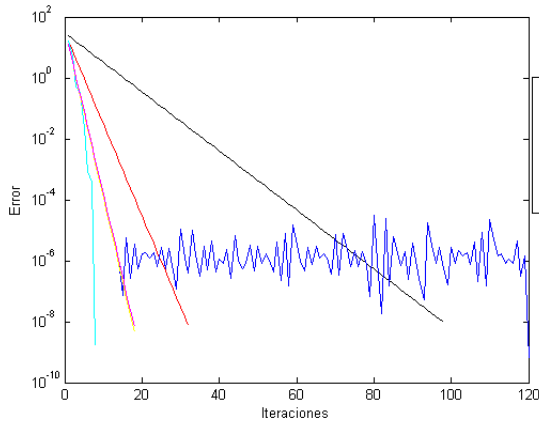


Figura 5.10: matriz 5×5

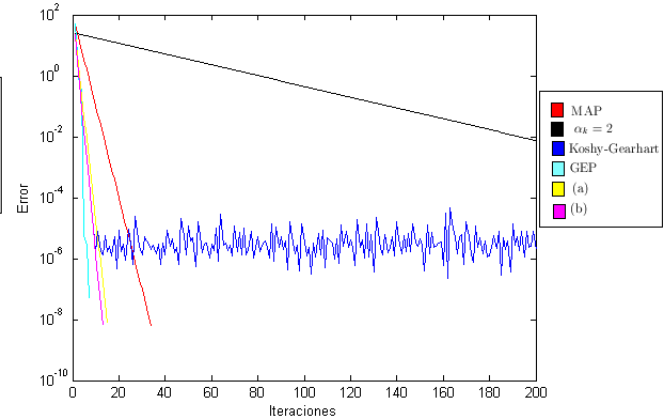
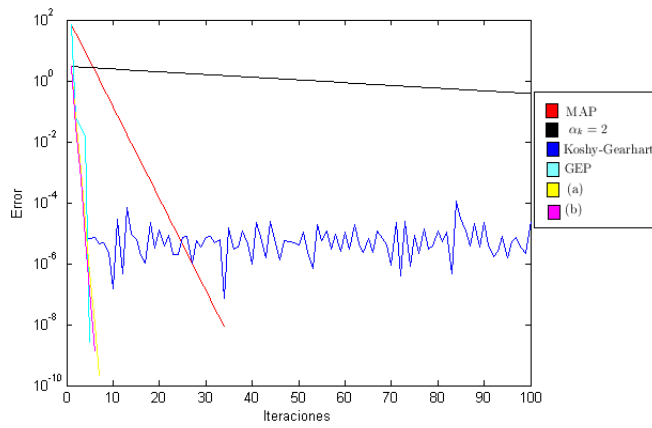
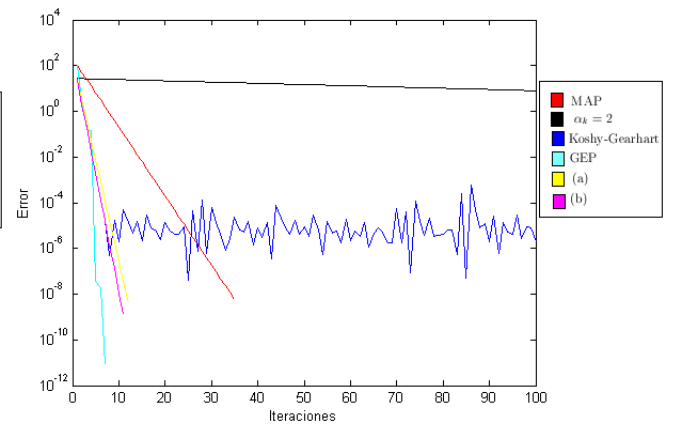


Figura 5.11: matriz 25×25

En las gráficas y en las tablas para el **experimento 3** se observa:

- A pesar que se había probado que el parámetro (b) y Koshy-Gearhart son equivalentes, su gráfica difiere después de una cierta cantidad de iteraciones, se presume que esto se debe a problemas numéricos en el cálculo de dicho parámetro.
- Se puede evidenciar que el método GEP converge en pocas iteraciones a la *tol* independientemente del orden de la matriz.

Figura 5.12: matriz 49×49 Figura 5.13: matriz 81×81

- El método con $\alpha_k = 2$ (fijo) no converge.
- Los parámetros de GEP, (a) y (b) aceleran la convergencia.

5.1.4. Experimento 4

$$x_{k+1} = x_k - \alpha_k P_M P_N^\perp x_k$$

$M : \{A \in \mathbb{R}^{n \times n} / A = A^t\}$ (Subespacio de las matrices simétricas).

$N : \{A \in \mathbb{R}^{n \times n} / a(\frac{n+1}{2}, j) = \frac{\sum_{i=1}^n a(\frac{n+1}{2}, i)}{n} \text{ para todo } j = 1, \dots, n; n \text{ impar}\}.$

Iteraciones						
Orden	MAP, $\alpha_k = 1$	$\alpha_k = 2$	Koshy-Gearhart	GEP	(a)	(b)
5×5	32	98	120	8	18	18
25×25	34	532	212	8	15	13
49×49	34	947	1978	5	7	6
81×81	35	1750	2693	7	12	11

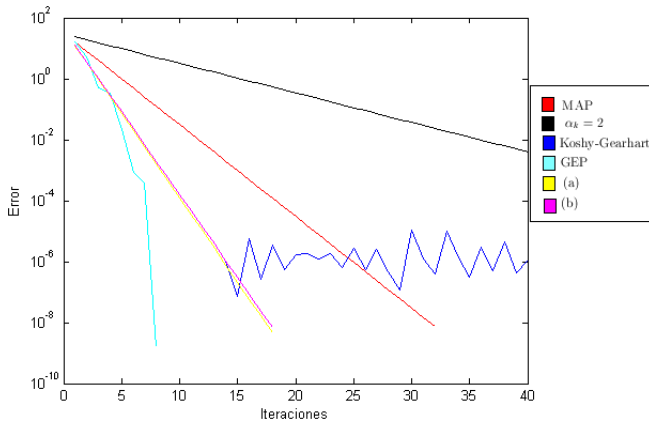


Figura 5.14: matriz 5×5

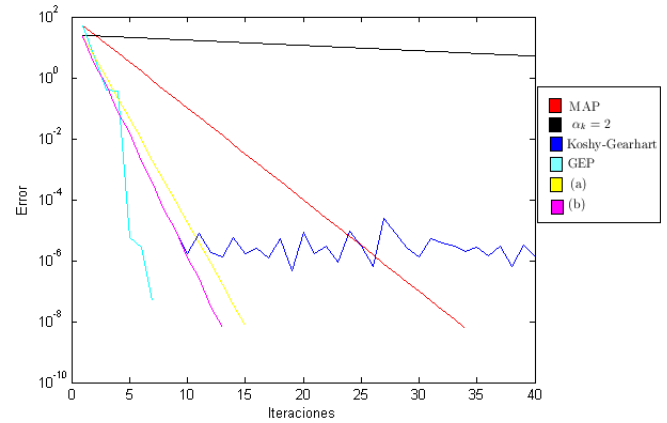
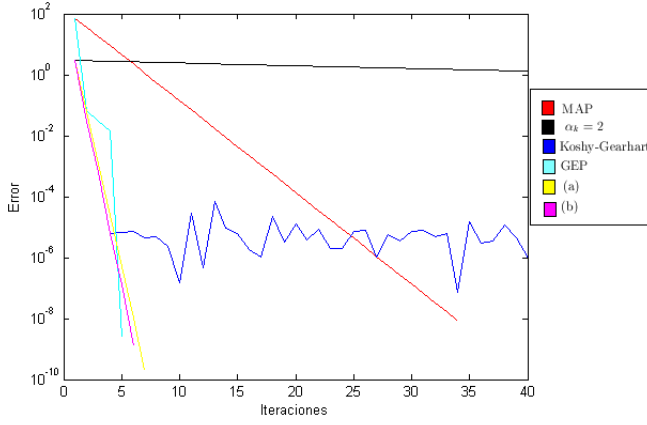
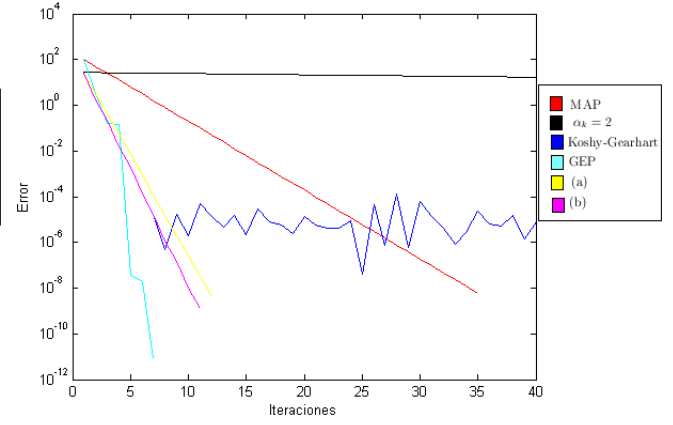


Figura 5.15: matriz 25×25

En las gráficas y en las tablas para el **experimento 4** se observa:

- Al igual que en el experimento anterior sucede que en Koshy-Gearhart y (b) su gráfica difiere después de una cierta cantidad de iteraciones, a pesar que se había demostrado analíticamente que son equivalente, se presume que esto se debe a problemas numéricos en el cálculo de dicho parámetro.

Figura 5.16: matriz 49×49 Figura 5.17: matriz 81×81

- Se puede evidenciar que el método GEP converge en pocas iteraciones a la *tol* independientemente del orden de la matriz.
- El método con $\alpha_k = 2$ (fijo) no converge.
- Los parámetros de GEP, (a) y (b) aceleran la convergencia.

5.1.5. Experimento 5

$$x_{k+1} = x_k - \alpha_k P_M P_N^\perp x_k$$

$M := \{A \in \mathbb{R}^{n \times n} / a(i, j) = 0 \text{ para todo } i > j \text{ } (i, j = 1, \dots, n).\}$ (Subespacio de las matrices triangulares superiores).

$N := \{A \in \mathbb{R}^{n \times n} / a(1, j) = \text{promedio1}, \quad a(\frac{n+1}{2}, k) = \text{promedio2} \text{ para todo } k, j = 1, \dots, n. \text{ } n; \text{ impar} \text{ donde } \text{promedio1} = \frac{\sum_{i=1}^n a(1, i)}{n}, \text{promedio2} = \frac{\sum_{i=1}^n a(\frac{n+1}{2}, i)}{n}\}.$

Iteraciones						
Orden	MAP, $\alpha_k = 1$	$\alpha_k = 2$	Koshy-Gearhart	GEP	(a)	(b)
5×5	46	no converge	37	3	28	26
25×25	36	no converge	387	3	20	23
49×49	35	no converge	623	3	18	21
81×81	36	no converge	461	3	20	22

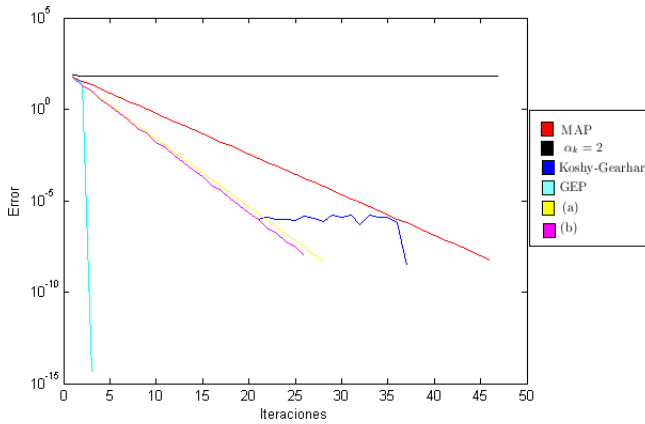


Figura 5.18: matriz 5×5

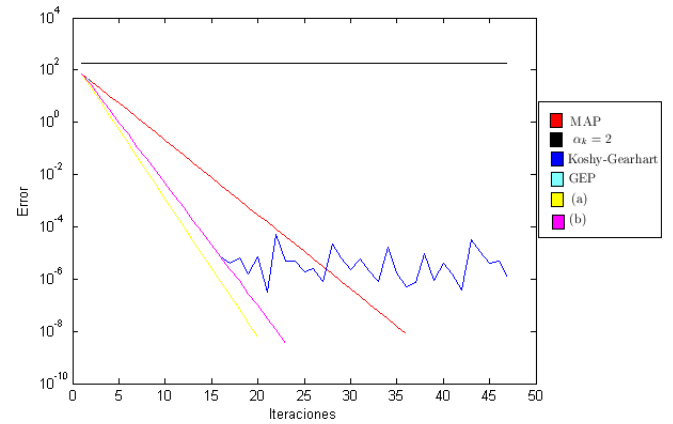
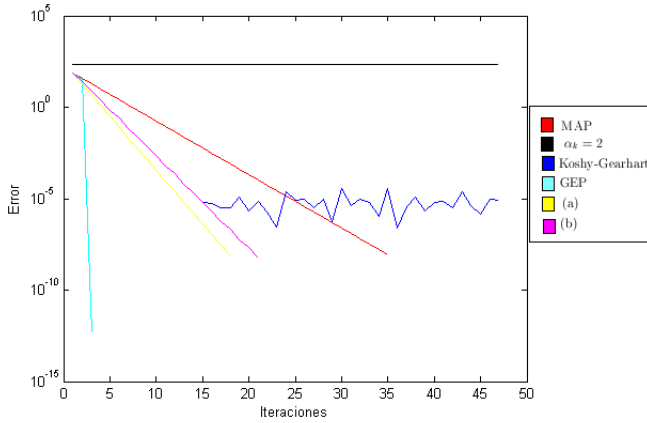
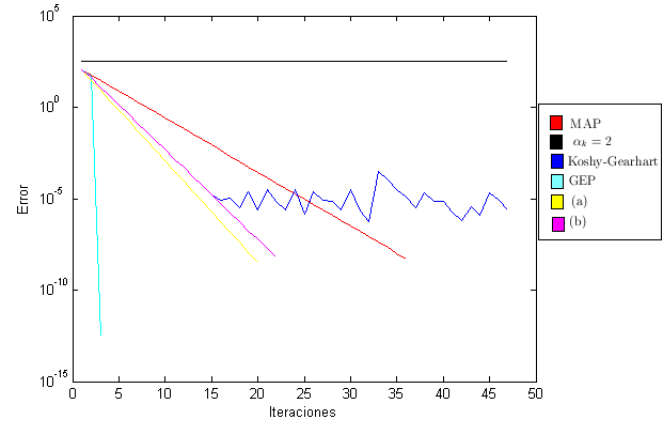


Figura 5.19: matriz 25×25

En las gráficas y en las tablas para el **experimento 5** se observa:

- Al igual que en el experimento anterior sucede que en Koshy-Gearhart y (b) su gráfica difiere después de una cierta cantidad de iteraciones, a pesar que se había demostrado analíticamente que son equivalente, se presume que esto se debe a problemas numéricos en el cálculo de dicho parámetro.

Figura 5.20: matriz 49×49 Figura 5.21: matriz 81×81

- Se puede evidenciar que el método GEP converge en pocas iteraciones a la *tol* independientemente del orden de la matriz.
- El método con $\alpha_k = 2$ (fijo) no converge.
- Los parámetros de GEP, (a) y (b) aceleran la convergencia.

5.1.6. Experimento 6

$$x_{k+1} = x_k - \alpha_k P_M P_N^\perp x_k$$

$$M : \{A \in \mathbb{R}^{n \times n} / a(1,1) = a(n,1) = a(1,n) = a(n,n) = 0\}.$$

$N : \{A \in \mathbb{R}^{n \times n} / a(i,j) = 0 \text{ para todo } i > j \text{ (} i, j = 1, \dots, n \text{)}\}$. (Subespacio de las matrices triangulares superiores)

Iteraciones						
Orden	MAP, $\alpha_k = 1$	$\alpha_k = 2$	Koshly-Gearhart	GEP	(a)	(b)
5×5	1	no converge	1	1	1	1
25×25	1	no converge	1	1	1	1
50×50	1	no converge	1	1	1	1
100×100	1	no converge	1	1	1	1

- En este experimento todos los métodos iterativos convergieron optimamente en 1 iteración salvo $\alpha_k = 2$, como el MAP convergió en 1 iteración significa que nos encontramos con el caso ideal en que los subespacios son ortogonales.
- Los parámetro de aceleración no modifican el hecho de que convergan en 1 iteración.

5.1.7. Experimento 7

$$x_{k+1} = x_k - \alpha_k P_M P_N^\perp x_k$$

$$M := \{A \in \mathbb{R}^{n \times n} / a(1, 1) = a(n, 1) = a(1, n) = a(n, n) = 0\}.$$

$$N := \{A \in \mathbb{R}^{n \times n} / a(j, \frac{n+1}{2}) = \frac{\sum_{i=1}^n a(i, \frac{n+1}{2})}{n} \text{ para todo } j = 1, \dots, n; n \text{ impar}\}.$$

Iteraciones						
Orden	MAP, $\alpha_k = 1$	$\alpha_k = 2$	Koshy-Gearhart	GEP	(a)	(b)
5×5	1	no converge	1	1	1	1
25×25	1	no converge	1	1	1	1
49×49	1	no converge	1	1	1	1
81×81	1	no converge	1	1	1	1

- En este experimento todos los métodos iterativos convergieron optimamente en 1 iteración salvo $\alpha_k = 2$, como el MAP convergió en 1 iteración significa que nos encontramos con el caso ideal en que los subespacios son ortogonales.
- Los parámetro de aceleración no modifican el hecho de que convergan en 1 iteración.

Capítulo 6

Conclusiones y Recomendaciones

En este trabajo pudimos observar el Método de Proyecciones Alternantes como un problema de minimización con restricciones, el cual nos proporcionó grandes herramientas para el estudio de la convergencia. Se demostró que la aceleración propuesta por Koshy-Gearhart es un caso particular de un método clásico de optimización con restricciones que es el Método del Gradiente Proyectado. Igualmente propusimos una aceleración espectral del Método de Proyecciones Alternantes.

Los parámetros propuestos para realizar los experimentos, utilizando subespacios de un espacio vectorial de matrices (los vectores pertenecen a $\mathbb{R}^{n \times n}$) arrojaron los siguientes resultados:

- El esquema de aceleración espectral (GEP) propuesto para el MAP, converge más rápido que los demás esquemas visto en este trabajo.
- El esquema (b) que es equivalente analíticamente al esquema de aceleración propuesto por Koshy-Gearhart es más estable numéricamente.
- Todos los esquemas propuestos en este trabajo aceleran el Método de Proyecciones Alternantes.
- En el caso ideal del MAP donde converge en 1 iteración, los esquemas de aceleración no alteran el resultado, es decir también convergen en 1 iteración.

Además proponemos como trabajo futuro:

- Generalizarlo a más de dos subespacios.
- Proponemos como trabajo futuro utilizar este resultados en las diversas aplicaciones que tiene el Método de Proyecciones Alternantes, como por ejemplo el tratamiento de imágenes.

Bibliografía

- [1] G. Appleby and D. Smolarki *A linear acceleration row action metho for projecting onto subspaces*, Electronic Transactions on Numerical Analysis (ETNA), 20 (2005) pp. 253-275.
- [2] H. Bauschke, F. Deutsch, H. Hundal, and S. H. Park *Acelerating the convergence of the method of alternating projections*, Transactions of the American Methemathical Society, 2003
- [3] E.G Birgin, J.M Martinez, M. Raydan *Spectral Projected Gradient Methods* Lecturas en Ciencias de la Computación, Centro de Cálculo Científico y Tecnológico de la UCV. Caracas, Enero, 2007.
- [4] D.L. Burkholder and Y.S. Chow., *Iterates of conditional expectation operators.*,Proc. Amer. Math. Soc., 12:490–495, 1961.
- [5] Y. Censor *Computational acceleration of projection algorithms for the linear best approximation problem*, Linear Algebra and its Applications, 416 (2006), pp 111-123
- [6] A. De Piero and A. Iusem *Parallel Projection Method of Finding a Common Point of a family of convex sets*, Pesquisa Operacional, 5 (1985).
- [7] J. Dyer *Acceleration of the Convergence of the Kaczmarz Method and Iterated Homogeneous Transformations*,doctoral dissertation. Univ. Calif. at Los Angeles 1965.

- [8] F. Deutsh. *The Method of Alternating Orthogonal Projections*. Approximation Theory, Splines Functions and Applications. 105-121, 1992.
- [9] Frank Deutsch. *Best Approximation in Inner Product Spaces*. Canadian Mathematical Society.
- [10] Frank Deutsch *The Angle Between Subspaces of a Hilbert Space* Department of Mathematics Pennsylvania State University, pp 107-130.
- [11] F. Deutsch *Von Neumann's Alternating Method: The rate of Convergence*, in Approximation Theory IV, C. Chi, L. Schumaker, and J. Ward, eds, Academic Press, New York, 1983, pp 427-434.
- [12] R. Escalante, M. Raydan. *Alternating Projections Methods: Theory and Applications*, en preparación.
- [13] René Escalante, *Algoritmos numéricos para el problema de restauración de imágenes usando el método de las proyecciones alternantes*, Revista de Matemática: Teoría y Aplicaciones, 11(1):41-58, 2004.
- [14] C. Franchetti and W. Light, *The alternating algorithm in uniformly convex spaces*, J. London Math. Soc. 29 no. (2). pp. 545-555.
- [15] W. Gearhart and M. Koshy *Acceleration schemes for the method of alternating projections*, J. Comp. Appl. Math., 26 (1989), pp. 235-249
- [16] L. Hernández Ramos *Alternating oblique projections for coupled linear systems*, Numerical Algorithms, 38 (2005), pp. 285-303.
- [17] S. Kayalar and H. Weinert *Error bounds for the method of alternating projections*, Math. Control Signals Systems, 1 (1988), pp 43-59.
- [18] A. Lent and G.T Herman. *Iterative relaxation methods for image reconstruction*. Communications of the ACM, 1978.

- [19] J.M. Martinez., *The projection method for solving nonlinear systems of equations under the most violated constraint control*, Computers and Mathematics with Appl., 11:987–993, 1985.
- [20] Marcos Raydan, *Minimización Irrestriccta de gran Escala*, noviembre, 2005.
- [21] Pulido Rosalba. *Precondicionadores Inversos Mediante Proyecciones Alternantes*. Trabajo Especial de Grado, 2003.
- [22] H.A. Schwarz, *Ueber einen grenzübergang durch alternirendes Verfahren*. Vierteljahrsschrift der Naturforschenden Gesellschaft in Zurich, 15:272–286.
- [23] M. Skwarczynski, *Alternating projections in complex analysis*. *Complex Analysis and Applications*, 83, Sofia, 1985.
- [24] K. Smith, D. Solmon, and S. Wagner. *Practical and mathematical aspects og the problem of reconstructing objects from radiographs*. Bull. Amer.Math. Soc., 83 (1977), pp. 1227-1270.
- [25] K. Tanabe. *Projection Method for Solving a Singular System of Linear Equations and its Applications*. 1971.
- [26] J. Von Neumann. *The Geometry of Orthogonal Spaces, Vol II*. Princeton University Press, Princenton, New Jersey, 1950.
- [27] N. Wiener and P. Masani, *The prediction theory of multivariate stochastic processes, II: The linear predictor*, Acta Math., 93:95–137, 1957.