

Universidad Central de Venezuela  
Facultad de Ciencias  
Escuela de Computación  
Postgrado en Ciencias de la Computación

**TESIS DOCTORAL:  
MÉTODOS DE PROYECCIONES  
ALTERNANTES VÍA  
OPTIMIZACIÓN NUMÉRICA**



**Tesis doctoral presentada por el**

MSc. Luis Manuel Hernández Ramos

Ante la Ilustre  
Universidad Central de Venezuela  
Para optar por el título de

**Doctor en Ciencias de la Computación**

**Tutores:** Dr. René Escalante  
Dr. Marcos Raydan

Caracas, 30 de Junio de 2011

## Resumen

Los métodos de proyecciones alternantes son estudiados utilizando herramientas de la optimización numérica. Algunos de los métodos de proyecciones son interpretados como esquemas iterativos ya existentes para resolver problemas de la optimización, permitiendo explicar de otro modo algunas de sus características. Igualmente, se han podido crear versiones aceleradas de los métodos de proyecciones alternantes y simultáneas, basadas en técnicas más eficientes y novedosas. Los métodos de proyecciones alternantes se estudian vía optimización numérica en diferentes escenarios que son: el problema de mejor aproximación, tanto en subespacios como en variedades lineales; y el problema de factibilidad convexa.

# Agradecimientos

Quisiera agradecer infinitamente a todas las personas que ayudaron de una u otra manera para que este trabajo se pudiera llevar a cabo.

En especial, quisiera agradecer a mis tutores, los profesores René Escalante y Marcos Raydan. Este agradecimiento no es solo el que se le tiene normalmente a un tutor de Tesis. El mío va más allá, les agradezco por haberme tendido la mano y por haberme defendido en momentos en que no todos querían que mis progresos se volvieran realidad. Quiero agradecerles también por su libro, que es el culpable de que yo me enamorase de los métodos de proyecciones alternantes.

Quiero agradecer igualmente a los miembros del comité del Postgrado en Ciencias de la Computación y de las diversas instancias de la Facultad de Ciencias de la Universidad Central de Venezuela que me apoyaron para que yo pudiera llevar a cabo mis estudios doctorales. Un especial agradecimiento al profesor Alfredo Matteo.

Finalmente, el más grande agradecimiento a mi esposa Meli y a mi hija Lucina, quienes han estado conmigo soportándome, compartiendo y apoyándome en todos estos momentos. A ellas dedico este trabajo.

# Índice general

<b>1. Introducción</b>	<b>4</b>
<b>2. Métodos de proyecciones alternantes</b>	<b>6</b>
2.1. Definición del método . . . . .	6
2.1.1. El método de von Neumann . . . . .	7
2.1.2. Métodos de Kaczmarz y Cimmino . . . . .	8
2.2. Velocidad de convergencia . . . . .	10
2.2.1. Ángulo entre dos subespacios . . . . .	10
2.2.2. Velocidad de convergencia . . . . .	11
2.3. Aceleración del método . . . . .	13
<b>3. Optimización numérica</b>	<b>15</b>
3.1. Método de Cauchy o mínimo descenso . . . . .	16
3.2. Método del gradiente espectral . . . . .	17
3.3. Método de los gradientes conjugados . . . . .	19
3.3.1. Elección de los vectores de dirección . . . . .	20
3.3.2. Convergencia . . . . .	21
3.3.3. Extensión a espacios de Hilbert . . . . .	22
<b>4. Aceleración en espacios de Hilbert</b>	<b>24</b>
4.1. Introducción . . . . .	24
4.2. Aceleración de los métodos de proyecciones alternantes en sub- espacios . . . . .	25
4.2.1. Aceleración de problemas de punto fijo lineales . . . . .	26
4.2.2. El caso auto-adjunto . . . . .	32
4.3. Experimentos numéricos . . . . .	35
4.3.1. Problemas de punto de ensilladura con métodos de pro- yecciones alternantes . . . . .	35

4.3.2. Descripción de los experimentos . . . . .	37
<b>5. Cimmino acelerado</b>	<b>43</b>
5.1. Introducción . . . . .	43
5.2. Aceleración del método de Cimmino para subespacios . . . . .	44
5.3. Relación entre el espectro de la Hessiana y la geometría del problema	48
5.4. Experimentos numéricos . . . . .	50
5.4.1. Resultados . . . . .	51
<b>6. Aceleración para variedades lineales</b>	<b>58</b>
6.1. Introducción . . . . .	58
6.2. Aceleración para variedades lineales . . . . .	58
6.2.1. Métodos de residual para sistemas no lineales . . . . .	62
6.2.2. Experimentos con variedades lineales . . . . .	65
<b>7. Problema de Factibilidad Convexa</b>	<b>69</b>
7.1. Introducción . . . . .	69
7.2. Problema de factibilidad convexa . . . . .	69
7.3. Aceleración para el CFP . . . . .	70
7.4. Experimentos numéricos . . . . .	71
7.4.1. Problema de factibilidad lineal . . . . .	71
7.4.2. Problemas de factibilidad lineal con restricciones de caja .	74
7.4.3. Problemas de factibilidad lineal con restricciones de esferas	80
7.4.4. Intersectando cajas y esferas . . . . .	84
<b>8. Conclusiones</b>	<b>87</b>
<b>Bibliografía</b>	<b>90</b>

# Capítulo 1

## Introducción

Algunos problemas importantes en las matemáticas tratan de encontrar elementos en la intersección de varios conjuntos. En algunas ocasiones, es suficiente con encontrar un solo elemento perteneciente a esta intersección. A este tipo de problemas se les conoce como: *problemas de factibilidad*. En otras ocasiones, no es suficiente solo con encontrar un elemento factible (i.e. un elemento que resuelva el problema de factibilidad), sino que se pide encontrar un elemento factible que verifique alguna propiedad adicional. Un caso es, en espacios que dispongan de alguna métrica, encontrar el elemento factible que se encuentre más cercano a un elemento dado. A este problema se le conoce como: *problema de mejor aproximación* y los elementos que verifiquen esta propiedad son conocidos como *mejores aproximaciones*.

Una clase importante de métodos para resolver estos problemas son los métodos basados en proyecciones, por ejemplo, el método de las proyecciones alternantes de Von Neumann-Halperin [62] y el método de Cimmino [23]. Estos métodos han sido estudiados tanto desde un contexto puramente teórico, en el análisis de la teoría de operadores, como desde un punto de vista práctico, como métodos numéricos para hallar soluciones a los problemas antes mencionados.

Los métodos de Von Neumann-Halperin y Cimmino han sido utilizados para resolver una amplia variedad de problemas prácticos (v.f. [29, 11, 33, 19, 18, 22, 39]) y su rango de aplicaciones sigue aumentando significativamente. Sin embargo, una dificultad en la aplicación práctica de los métodos de proyecciones alternantes es que estos pueden llegar a requerir muchas iteraciones para dar una buena aproximación numérica, y en algunos casos esto puede resultar en un alto costo computacional.

Han habido diversos intentos por acelerar los métodos de proyecciones alternantes (e.g. [44, 46, 20]). Recientemente Bauschke et al. [10] propusieron un esquema de aceleración en donde se generalizan varios de estos métodos. Sin embargo se sabe que existen casos en los que esta clase de métodos acelerados no funcionan bien, quedando la explicación de su mal funcionamiento como un problema abierto.

Por otra parte, la optimización numérica es una disciplina matemática que cuenta con gran vitalidad y fuerza, y que proporciona grandes herramientas, tanto para la resolución de problemas numéricos clásicos, como para el estudio y comprensión de algunos problemas de la matemática aplicada, e incluso de la matemática pura.

En este trabajo, se pretende hacer uso del conocimiento y de las herramientas que nos proporciona la optimización numérica para lograr un mejor entendimiento de los métodos de proyecciones alternantes, explicar en algunos casos su funcionamiento y de esa manera aprovechar ese entendimiento para crear nuevos métodos acelerados de proyecciones alternantes y simultáneas, tanto para problemas de factibilidad como de mejor aproximación.

También en este trabajo, se trata de aumentar el rango de aplicaciones de los métodos de proyecciones alternantes mostrando que un problema muy importante, como es el de la resolución de problemas de punto de ensilladura, se puede interpretar como un problema de mejor aproximación en subespacios y que por lo tanto puede hacerse su resolución numérica mediante algunas de las aceleraciones propuestas para los métodos de proyecciones. Los problemas de punto de ensilladura son usados en este trabajo como una familia de ejemplos para ilustrar nuestras técnicas de aceleración.

# Capítulo 2

## Métodos de proyecciones alternantes

En este capítulo, hablaremos del método de proyecciones alternantes y sus variantes. Para ello, nos basaremos principalmente en los trabajos de F. Deutsch [31, 33] y en el libro de Escalante y Raydan [39].

### 2.1. Definición del método

El método de proyecciones alternantes (MAP), en su formulación más simple fue propuesto originalmente por John von Neumann [62], quien trató el problema de encontrar la proyección sobre la intersección de dos subespacios cerrados de un punto dado en un espacio de Hilbert. Posteriormente, fueron propuestos métodos análogos para el caso particular de resolución de sistemas de ecuaciones lineales. Estos son los métodos de Kaczmarz [55] y Cimmino [23]; métodos que a su vez fueron generalizados luego.

Sea  $H$  un espacio de Hilbert y sea  $M$  un subespacio cerrado de  $H$ . La proyección ortogonal sobre  $M$  está denotada por  $P_M$ . En particular,  $P_M$  es lineal, autoadjunta ( $P_M^* = P_M$ ), idempotente ( $P_M^2 = P_M$ ) y  $P_M(x)$  es la mejor aproximación (o el punto mas cercano) de  $x$  a  $M$ . Es decir:

$$\|x - P_M(x)\| = d(x, M), \quad (2.1)$$

donde  $d(x, M) = \inf\{\|x - y\| \mid y \in M\}$ ,  $\|\cdot\|$  es la norma inducida por el producto interno del espacio, i.e.,  $\|x\|^2 = \langle x, x \rangle$ , para todo  $x \in H$ , donde  $\langle \cdot, \cdot \rangle$  denota el

producto interno en  $H$ .

$P_M$  se denomina la proyección ortogonal sobre  $M$  por la siguiente propiedad que la define:

$$\langle x - P_M(x), y \rangle = 0, \quad \forall x \in M. \quad (2.2)$$

Esto quiere decir que  $x - P_M(x)$  es ortogonal a  $M$ . Si denotamos

$$M^\perp := \{y \in H \mid \langle x, y \rangle = 0, \quad \forall x \in M\}, \quad (2.3)$$

esto se escribe  $x - P_M(x) \in M^\perp$  o también  $x - P_M(x) \perp M$ .

El lema que sigue a continuación, y que puede ser demostrado sin dificultad, es fundamental en la idea principal del método MAP.

**Lema 2.1.1 (von Neumann [62])** Sean  $A$  y  $B$  dos subespacios cerrados de  $H$ . Luego,  $P_A P_B = P_B P_A$  si, y solo si,  $P_A P_B = P_{A \cap B}$ .

Esto quiere decir que,  $P_A$  y  $P_B$  conmutan, si y solo si, su composición es también una proyección ortogonal.

### 2.1.1. El método de von Neumann

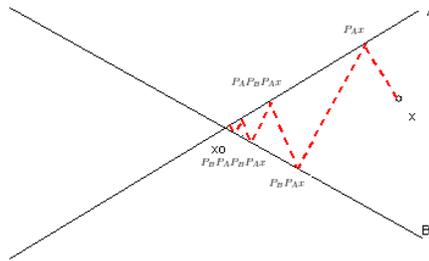


Figura 2.1: Método MAP

Von Neumann [62] se interesó por el caso en que  $P_A$  y  $P_B$  no conmutan, probando el siguiente resultado:

**Teorema 2.1.1 (von Neumann [62])** Sean  $A$  y  $B$  dos subespacios cerrados de  $H$ :

$$\lim_{n \rightarrow \infty} (P_B P_A)^n x = P_{A \cap B} x. \quad (2.4)$$

El Teorema 2.1.1 da origen a un algoritmo conocido como el Método de proyecciones alternantes (MAP) (Algoritmo 1).

---

**Algoritmo 1** Algoritmo MAP
 

---

```

  Seleccionar  $x_0 \in H$ 
  for  $k := 0, 1, \dots$ , do
     $x_n = P_B P_A x_{n-1}$ 
  end for

```

---

La interpretación geométrica del método MAP es la siguiente: para encontrar la mejor aproximación de un  $x \in H$  dado sobre  $A \cap B$ , se proyecta primero  $x$  sobre  $A$ , luego se proyecta el resultado sobre  $B$ , y se continúa proyectando el resultado de manera alternada sobre  $A$  y sobre  $B$ . La sucesión generada converge hacia  $P_{A \cap B}$ .

El método MAP (ver figura 2.1) tiene utilidad práctica cuando resulta menos costoso calcular las proyecciones sobre  $A$  y  $B$  individualmente que calcular directamente la proyección sobre  $A \cap B$ , o cuando la proyección sobre la intersección no se conoce.

### Extensión de Halperin

El método de Von-Neumann puede extenderse a más de dos subespacios. La extensión del Teorema 2.1.1 a más de dos subespacios fue desarrollada por Halperin en 1962 [47], aunque la misma idea había sido propuesta para sistemas de ecuaciones lineales por Kaczmarz [55]. El teorema de Halperin dice lo siguiente:

**Teorema 2.1.2 (Halperin [47])** Sean  $M_1, M_2, \dots, M_m$ ,  $m$  subespacios cerrados del espacio  $H$ . Entonces,

$$\lim_{n \rightarrow \infty} (P_{M_m} P_{M_{m-1}} \dots P_{M_1})^n x = P_{\cap_1^m M_i} x. \quad (2.5)$$

### 2.1.2. Métodos de Kaczmarz y Cimmino

Los métodos de Kaczmarz y de Cimmino fueron propuestos originalmente para resolver sistemas de ecuaciones lineales. Consideremos el problema de resolver sistemas de ecuaciones lineales:

$$Ax = b,$$

donde  $A$  es una matriz real de  $m \times n$ ,  $x \in \mathfrak{R}^n$  y  $y \in \mathfrak{R}^m$ . Este problema puede ser generalizado a cualquier espacio de Hilbert  $H$  como el problema de encontrar una

solución  $x$  (si existe) en la intersección de los  $m$  hiperplanos o variedades lineales cerradas, dadas por:

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

para todo  $i \in I$ , donde  $a_i$  denota la  $i$ -ésima fila de  $A$ , o en general un vector fijo de  $H$  e  $I$  es el conjunto de índices. Note que  $a_i$  es ortogonal a  $H_i$ . Mas aún, si  $z \notin H_i$  entonces la proyección en  $H_i$  está dada por:

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

para comprobar que  $P_{H_i}(z)$  es efectivamente la proyección de  $z$  en  $H_i$  es suficiente con verificar que cumple con la condición (2.2), lo cual es trivial.

### Método de Kaczmarz

Este método fue propuesto originalmente por Kaczmarz [55], quien estableció su convergencia para resolver problemas  $Ax = b$  con matrices cuadradas no singulares. El método de Kaczmarz es el mismo método propuesto posteriormente por Halperin en el contexto particular de resolución de sistemas de ecuaciones lineales. Para un  $x_0 \in \mathfrak{R}^n$  arbitrario, un paso típico del método de Kaczmarz puede ser descrito como sigue:

$$x_{n+1} = P_{H_m} P_{H_{m-1}} \dots P_{H_2} P_{H_1} x_n.$$

Muchos autores han realizado generalizaciones y extensiones del método de Kaczmarz (e.g. [39, 68, 49]).

### Método de Cimmino

Este método fue propuesto inicialmente por Cimmino [23] quien estableció su convergencia para la solución de sistemas lineales  $Ax = b$ , donde  $A$  es una matriz real  $n \times n$  y  $b \in \mathfrak{R}^n$ .

Para un  $x_0 \in \mathfrak{R}^n$  arbitrario, un paso típico del método de Cimmino puede ser descrito mediante,

$$x_{k+1} = \frac{1}{n} \sum_{i=1}^n P_{H_i}(x_k),$$

donde

$$H_i = \{x \in \mathfrak{R}^n : \langle a_i, x \rangle = b_i\}.$$

El método de Cimmino ha sido extendido y generalizado muchas veces; ver [39]. Por ejemplo Kammerer y Nashed en [56] generalizaron este método de la siguiente manera:

Sean  $V_1, V_2, \dots, V_n$  variedades lineales cerradas de un espacio de Hilbert  $H$ , y dado  $x_0 \in H$ . El método de Cimmino,

$$x_{k+1} = \frac{1}{n} \sum_{i=1}^n P_{V_i}(x_k),$$

converge a la proyección del iterado inicial  $x_0$  en la intersección de las variedades lineales. Es decir, el método converge hacia  $P_V(x_0)$ , donde  $V = \bigcap_{i=1}^n V_i$ .

Otras extensiones y generalizaciones del método de Cimmino pueden encontrarse por ejemplo en [9, 21, 20, 1, 2, 54].

## 2.2. Velocidad de convergencia

La velocidad de convergencia del método es determinante para saber si el método MAP es práctico, porque si para obtener una buena aproximación se requieren demasiadas proyecciones, el método deja de ser una opción viable numéricamente.

La velocidad de convergencia del método MAP depende del ángulo entre los subespacios involucrados [67, 39]. Recordemos primero la definición de ángulo entre vectores: si  $x \in H$  y  $y \in H$ , el ángulo  $\theta$  entre los vectores  $x$  y  $y$  está definido como:

$$\cos \theta = \frac{\langle x, y \rangle}{\|x\| \|y\|}. \quad (2.6)$$

Ahora extenderemos esa noción a ángulo entre subespacios.

### 2.2.1. Ángulo entre dos subespacios

La definición de ángulo introducida originalmente por Friedrichs [42] es la noción más utilizada en la literatura al referirse de ángulo entre dos subespacios:

**Definición 2.2.1 (Friedrichs, 1937)** *El ángulo  $\theta(M, N)$  entre dos subespacios cerrados  $M$  y  $N$ , está definido como el ángulo de  $[0, \pi/2]$  cuyo **coseno**  $c(M, N)$  está dado por:*

$$c(M, N) = \sup\{\langle x, y \rangle \mid x \in M \cap (M \cap N)^\perp, \|x\| \leq 1, y \in N \cap (M \cap N)^\perp, \|y\| \leq 1\}. \quad (2.7)$$

**Observación:**

Las siguientes propiedades son consecuencia de la definiciones (v.f. [33]):

1.  $0 \leq c(M, N) \leq 1$ .
2.  $c(M, N) = c(N, M)$ .

El lema a continuación tiene propiedades muy útiles:

**Lema 2.2.1 (Kayalar, Weinert [58] y Deutsch [62])**

$$c(M, N) = \|P_M P_N - P_{M \cap N}\| = \|P_M P_N P_{(M \cap N)^\perp}\|. \quad (2.8)$$

### 2.2.2. Velocidad de convergencia

A continuación, vamos a presentar algunos resultados de convergencia relacionados con el método de Von Neumann - Halperin en subespacios.

De acuerdo con el Teorema 2.1.2,  $(P_m \dots P_2 P_1)^n x$  converge a  $P_M x$  para todo  $x \in H$  (donde  $M = \bigcap_{i=1}^m M_i$  y  $P_i = P_{M_i}$ ).

Podemos notar que, para cada  $i = 1 \dots m$ ,  $P_i P_M = P_M$ , y también

$$\begin{aligned} P_i P_{M^\perp} &= P_i (I - P_M) \\ &= P_i - P_i P_M \\ &= P_i - P_M P_i \\ &= (I - P_M) P_i \\ &= P_{M^\perp} P_i. \end{aligned}$$

Entonces, podemos deducir que para cada  $x \in H$ :

$$\begin{aligned} \|(P_m, \dots, P_2 P_1)^n x - P_M x\| &\leq \|(P_m \dots P_2 P_1)^n - P_M\| \|x\| \\ &= \|(P_m, \dots, P_2 P_1 P_{M^\perp})^n\| \|x\| \end{aligned}$$

$$\leq \|(P_m \dots, P_2 P_1 P_{M^\perp})\|^n \|x\|$$

Luego, la velocidad de convergencia del algoritmo MAP está determinada por la norma del operador  $(P_m \dots P_2 P_1 P_{M^\perp})$ . En particular, para  $m = 2$ , y a partir del Lema 2.2.1 podemos deducir que:

$$\|(P_2 P_1)^n - P_M\| \leq \|P_2 P_1 P_{M^\perp}\| = c(M_1, M_2)^n. \quad (2.9)$$

Sin embargo, en el caso de dos subespacios, esta no es la mejor cota. Aronszajn [3] dio la siguiente cota:

Para todo  $x \in H$ ,  $n \geq 1$ ,  $n \in \mathbb{Z}$ ,

$$\|(P_2 P_1)^n x - P_M x\| \leq c(M_1, M_2)^{2n-1} \|x\|. \quad (2.10)$$

Más tarde, en 1988, Kayalar y Weinert establecieron el siguiente resultado.

**Teorema 2.2.1 (Kayalar y Weinert, 1988 [58])**

$$\|(P_2 P_1)^n - P_M\| = c(M_1, M_2)^{2n-1}, \quad (n = 1, 2, \dots) \quad (2.11)$$

### Velocidad de convergencia en el caso de múltiples subespacios

Para el caso de múltiples subespacios ( $r \geq 2$ ), Smith, Solmon y Wagner [67] probaron una cota para la velocidad de convergencia en función de los ángulos de los subespacios concernientes.

**Teorema 2.2.2 (Smith, Solmon y Wagner [67])**

$$\|(P_{M_m} P_{M_{m-1}} \dots P_{M_1})^n x - P_{\cap_{i=1}^m M_i} x\| \leq c^n \|x - P_{\cap_{i=1}^m M_i} x\| \quad (2.12)$$

donde

$$c = [1 - \prod_{i=1}^{m-1} \sin^2 \theta_i]^{1/2}, \quad (2.13)$$

aquí  $\theta_i$  es el ángulo  $M_i$  y  $N_i = \cap_{j=i+1}^r M_j$ .

Este teorema determina efectivamente que la velocidad de convergencia del método de proyecciones alternantes de Von Neumann - Halperin depende de los ángulos entre los subespacios involucrados.

### 2.3. Aceleración del método

El método de proyecciones alternantes posee una velocidad de convergencia  $r$ -lineal [48]. En el caso cuando el ángulo entre los subespacios es “pequeño”, estos métodos pueden ser lentos. Existen muchas maneras de acelerar la convergencia.

Entre los métodos de aceleración, para el problema de factibilidad convexa, se pueden nombrar las ideas de Gubin et al. [46], De Pierro y Iusem [28], Dos Santos [36], J.M.Martinez [60], Garcia-Palomares y Gonzalez-Castaño [43]. Entre los esquemas de aceleración para el problema de mejor aproximación en subespacios se pueden nombrar los trabajos de Gearhart y Koshy [44] (para el método de Von Neumann - Halperin) y la generalización propuesta por Bauschke et al. [10]. En la resolución de sistemas de ecuaciones lineales  $Ax = b$  podemos citar los esquemas de resolución desarrollados por Bjorck y Elfving et al. en [14] (aceleración de Kaczmarz) y también podemos citar el esquema propuesto por Arioli et al. en [2] (aceleración del método de Cimmino). En estos trabajos se proponen esquemas de aceleración por gradientes conjugados de los métodos de proyecciones para la resolución de sistemas lineales.

Vamos a ejemplificar las ideas de aceleración con el esquema propuesto por Gearhart y Koshy [44] para el problema de mejor aproximación en subespacios, que a su vez es un caso particular del esquema general propuesto por Bauschke et al. [10]:

Denotemos por  $Q$  la composición de operadores de proyección, i.e.  $Q = P_m P_{m-1} \dots P_1$ . Por otra parte denotamos por  $x$  el punto inicial y por  $M$  la intersección de los subespacios  $M_i$ .

Para describir el método de aceleración, sea  $x_k$  la  $k$ -ésima iteración y sea  $Qx_k$  la iteración siguiente MAP, la idea es buscar en la dirección definida por  $x_k$  y  $Qx_k$ , el punto más cercano a la solución  $P_M x$ .

Los puntos se representan en esta dirección:

$$x_t^k = tQx_k + (1-t)x_k, \quad t \in \mathfrak{R}. \quad (2.14)$$

Si denotamos  $t_k$  el valor de  $t$  para aquel punto que se encuentra más cerca de  $P_M x$ ,  $x_{t_k}^k - P_M x$  debe ser ortogonal a  $x_k - Qx_k$ . Lo que quiere decir,

$$\langle x_{t_k}^k - P_M x, x_k - Qx_k \rangle = 0. \quad (2.15)$$

Como las proyecciones  $P_i$  sobre subespacios son autoadjuntas entonces,

$$\langle P_M x, Qx_k \rangle = \langle P_1 P_2 \dots P_m P_M x, x_k \rangle = \langle P_M x, x_k \rangle. \quad (2.16)$$

En consecuencia  $\langle P_M x, x_k - Qx_k \rangle = 0$  y de (2.15) se obtiene  $\langle x_{t_k}^k, x_k - Qx_k \rangle = 0$ . De aquí podemos deducir:

$$t_k = \frac{\langle x_k, x_k - Qx_k \rangle}{\|x_k - Qx_k\|^2}. \quad (2.17)$$

Por ende, la aceleración estará dada por el Algoritmo 2.

---

**Algoritmo 2** Algoritmo MAP acelerado

---

Seleccionar  $x_0 \in H$

**for**  $k := 0, 1, \dots$ , **do**

$$t_k := \frac{\langle x_k, x_k - Qx_k \rangle}{\|x_k - Qx_k\|^2}$$

$$x_{k+1} := t_k Qx_k + (1 - t_k)x_k$$

**end for**

---

# Capítulo 3

## Nociones de optimización numérica

En este capítulo, trataremos sobre algunos métodos numéricos para minimización en gran escala, específicamente nos concentraremos en minimizar, sin restricciones, funciones cuadráticas convexas, que jugarán un papel fundamental en nuestros esquemas de aceleración.

Haremos una breve descripción de los métodos de tipo gradiente, tales como el método de Cauchy [17], el método del gradiente Espectral (GE), llamado también método de Barzilai-Borwein [7, 64], y luego del método de los gradientes conjugados (CG, por sus siglas en inglés) [53].

Consideremos el problema general de minimización sin restricciones:

$$\min_{x \in \mathfrak{R}^n} f(x), \quad (3.1)$$

donde  $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$  es una función cuadrática convexa y diferenciable definida como:

$$f(x) = \frac{1}{2}x^t Ax - b^t x, \quad (3.2)$$

aquí  $A \in \mathfrak{R}^{n \times n}$  es una matriz simétrica y positivo definida (SPD) y los vectores  $x, b \in \mathfrak{R}^n$ . Las dos proposiciones siguientes son muy conocidas:

- Si  $f(x) = \frac{1}{2}x^t Ax - b^t x$  donde  $A$  es SPD, entonces  $\nabla f(x) = Ax - b$ .
- Si  $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$ ,  $f \in C^1(D)$ , es decir,  $f$  pertenece al conjunto de las funciones continuamente diferenciables en  $D$ , con  $D$  abierto y convexo, y  $x_*$  es un mínimo local de  $f$ , entonces  $\nabla f(x_*) = 0$ .

Considerando las proposiciones anteriores, se puede determinar que los puntos que anulan el gradiente de la función objetivo  $f$  son los vectores que resuelven  $Ax = b$ , es decir, los que resuelven el sistema lineal. Por lo tanto, el problema de minimizar una función cuadrática convexa es equivalente al de resolver un sistema de ecuaciones lineales con una matriz SPD.

Los métodos iterativos tipo gradiente, son aquellos cuya iteración es de la forma:

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

donde  $\alpha_k$  es la longitud del paso y  $g_k$  es la dirección llamada de búsqueda, en este caso definida como  $g_k = \nabla f(x_k)$  donde  $f$  es la función objetivo. La dirección de  $-g_k$  es la que garantiza el descenso local más rápido en  $f$  y es conocida como dirección del gradiente negativo [63].

### 3.1. Método de Cauchy o mínimo descenso

El método de mínimo descenso, propuesto por Cauchy [17], es un método iterativo para problemas de optimización multidimensional. Este es un método tipo gradiente (3.3), donde el tamaño del paso  $\alpha_k$  se define como:

$$\alpha_k = \arg \min_{\alpha > 0} f(x_k - \alpha g_k) \quad (3.4)$$

**Proposición 3.1.1** *Si  $x$  y  $u$  son vectores cualesquiera en  $\mathbb{R}^n$  con  $u \neq 0$ , entonces*

$$\arg \min_{\alpha > 0} f(x - \alpha u) = \frac{r(x)^t u}{u^t A u},$$

donde  $r(x) = -\nabla f(x)$ .

Por la proposición anterior, para  $f$  cuadrática, la longitud de paso del método de Cauchy se reduce a:

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

El Algoritmo 3 nos describe el método de Cauchy. Puede comprobarse fácilmente que  $r_{k+1} = b - A(x_k + \alpha_k r_k) = r_k - \alpha_k A r_k$  son formas matemáticamente equivalentes de calcular la actualización del vector residual. Sin embargo, computacionalmente es preferible actualizar  $r_{k+1} = r_k - \alpha_k w_k$  ya que ahorra el cómputo de un producto matriz-vector (si  $w_k = A r_k$  había sido ya previamente calculado).

**Algoritmo 3** Algoritmo del método de Cauchy (cuadráticas)

---

Elegir  $x_0$  y  $\epsilon$ .  
 $r_0 = b - Ax_0$ .  
**for**  $k = 0, 1, 2, 3, \dots$  **do**  
     $w_k = Ar_k$   
     $\alpha_k = \frac{r_k^t r_k}{r_k^t w_k}$ .  
     $x_{k+1} = x_k + \alpha_k r_k$   
     $r_{k+1} = r_k - \alpha_k w_k$   
    Prueba de convergencia: Si  $\|r_{i+1}\| \geq \epsilon$ , continúe.  
**end for**

---

A continuación mostraremos un resultado clásico de convergencia del método de Cauchy para cuadráticas convexas [63]:

**Teorema 3.1.1** *Sea  $A$  simétrica y positivo definida. Entonces la sucesión  $\{x_k\}$  generada por el método de Cauchy converge al mínimo  $x_*$  ( $Ax_* = b$ ), desde cualquier iterado inicial  $x_0 \in \mathfrak{R}^n$ . Además,*

$$\|x_k - x_*\|_A \leq \frac{\text{Cond}(A) - 1}{\text{Cond}(A) + 1} \|x_{k-1} - x_*\|_A,$$

donde:

- $\|x\|_A^2 = \langle x, Ax \rangle$ ,
- $\text{Cond}(A) = \|A\|_2 \|A^{-1}\|_2$  es el número de condición de  $A$  en norma 2.

Nótese que cuando el número de condición de  $A$  es grande, el método de Cauchy puede llegar a tener una velocidad de convergencia muy lenta. En el caso en que  $A$  sea simétrica positivo semidefinida ( $f$  cuadrática convexa, pero no estrictamente), el método de Cauchy converge si  $b \in R(A)$  (si  $b$  pertenece al Rango de  $A$ ) [63].

## 3.2. Método del gradiente espectral

El método del gradiente espectral fue propuesto por Barzilai y Borwein [7] y fue analizado por Raydan [64] para cuadráticas estrictamente convexas. Este método es un método tipo gradiente, por lo que su iteración tiene la forma:

$$x_{x+1} = x_k - \alpha_k g_k, \tag{3.5}$$

sin embargo, difiere del método de Cauchy en la escogencia del paso  $\alpha_k$ .

El método del gradiente espectral es no monótono, es decir, propone una nueva longitud de paso  $\alpha_k = \frac{1}{\delta_k}$  que no garantiza descenso de la función objetivo. Esta longitud de paso es obtenida al resolver el sistema lineal sobredimensionado:

$$y_{k-1} = \delta_k s_{k-1} \quad (3.6)$$

donde  $s_{k-1} = x_k - x_{k-1}$  y  $y_{k-1} = g_k - g_{k-1}$  con  $g_k = \nabla f(x_k)$ .

Para los métodos espectrales, el escalar  $\delta_k$  viene de resolver el sistema (3.6) en el sentido de los mínimos cuadrados, como:

$$\delta_k = \frac{s_{k-1}^t y_{k-1}}{s_{k-1}^t s_{k-1}} \quad (3.7)$$

si  $s_{k-1} \neq 0$ .

Esta forma de escoger  $\delta_k$  está inspirada en el cálculo de un paso Cuasi-Newton  $SCN = -A_k^{-1} \nabla f(x_k)$  donde  $A_k \in \mathfrak{R}^{n \times n}$  es una aproximación del Hessiano de  $f$  en  $x_k$ . Esta aproximación, usualmente se escoge de forma tal que satisfaga la ecuación de la secante,

$$A_k s_{k-1} = y_{k-1} \quad (3.8)$$

restringiendo  $A_{k+1}$  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.

Si consideramos el caso cuando  $f$  es cuadrática, entonces (3.7) puede escribirse como:

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

En este caso,  $\delta_{k+1}$  resulta ser el cociente de Rayleigh de  $A$  evaluado en  $s_k$ . Igualmente, se cumple:

$$\delta_{k+1} = \frac{g_k^t A g_k}{g_k^t g_k}, \quad (3.10)$$

que es el cociente de Rayleigh evaluado en  $g_k$  [63]. El Algoritmo 4 nos describe el método del gradiente espectral. Al igual que en el anterior algoritmo del método de Cauchy, el término  $g_{k+1}$  fue actualizado por una expresión matemáticamente equivalente, pero que computacionalmente permite el ahorro de una operación matriz-vector.

Para funciones cuadráticas estrictamente convexas, se presenta el siguiente resultado de convergencia [64, 63]:

---

**Algoritmo 4** Algoritmo del método del gradiente espectral (cuadráticas)

---

Elegir  $x_0$ ,  $\delta_0$  y  $\epsilon$ .  
 $g_0 = Ax_0 - b$   
**for**  $k = 0, 1, 2, 3, \dots$  **do**  
 $s_k = (1/\delta_k)g_k$   
 $x_{k+1} = x_k + s_k$   
 $w_k = Ag_k$   
 $g_{k+1} = g_k - \frac{1}{\delta_k}w_k$   
Prueba de convergencia: Si  $\|g_{i+1}\| \geq \epsilon$ , continúe  
 $\delta_{k+1} = g_k^t w_k / g_k^t g_k$   
**end for**

---

**Teorema 3.2.1** *Sea  $f(x)$  una función cuadrática estrictamente convexa. Sea  $\{x_k\}$  la sucesión generada por el método del gradiente espectral y  $x_*$  el único minimizador de  $f$ . Entonces, o bien  $x_j = x_*$  para algún  $j$  finito, o la sucesión  $\{x_k\}$  converge a  $x_*$ .*

Este resultado se extiende un poco después al caso de cuadráticas convexas (no necesariamente estrictamente convexas) por Friedlander, Martinez y Raydan [41, 63], al incorporar el método a una técnica de minimización con restricciones tipo “cajas” (restricciones simples). La velocidad de convergencia del método GE se empieza a entender recientemente. Dai y Liao [26] establecieron la convergencia r-lineal para funciones cuadráticas convexas.

### 3.3. Método de los gradientes conjugados

El método de los gradientes conjugados (CG) fue propuesto originalmente por Hestenes y Stiefel [53] para resolver sistemas de ecuaciones lineales  $Ax = b$ , donde la matriz  $A$  es SPD. Este es un método directo en teoría, pero iterativo en la práctica, y está basado en el resultado conocido de optimización ya mencionado anteriormente según el cuál existe una equivalencia entre la resolución de un sistema de ecuaciones lineales  $Ax = b$  para una matriz SPD, y la minimización de una función cuadrática

$$f(x) = \frac{1}{2}x^t Ax - x^t b.$$

◦ El valor mínimo de  $f(x)$  es  $-\frac{1}{2} b^t A^{-1} b$  y es obtenido eligiendo

$$x = A^{-1} b.$$

Existe gran cantidad de métodos iterativos en la literatura de optimización para resolver este problema de minimización. En estos métodos iterativos, las aproximaciones sucesivas  $x_k$  al vector solución  $x$  son calculadas recursivamente:

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

con  $x_0$  dado y donde los vectores  $\{p_k\}$  son llamados vectores de dirección, y los escalares  $\alpha_k$  son elegidos para minimizar la función  $h(\alpha) = f(x_k + \alpha p_k)$ . Es posible mostrar que esto puede lograrse si elegimos,

$$\alpha = \alpha_k = \frac{p_k^t (b - Ax_k)}{p_k^t A p_k} = p_k^t r_k / p_k^t A p_k,$$

donde  $r_k = b - Ax_k$ .

### 3.3.1. Elección de los vectores de dirección

Ahora veremos como se eligen los vectores de dirección  $p_k$ . El método de los gradientes conjugados (GC) es un método que automáticamente genera los vectores de dirección. El vector de dirección de cada paso necesita ser generado en el paso previo. Mas aún, los vectores de dirección  $p_k$  tienen la propiedad notable siguiente:

$$(p_i)^t A p_j = 0, \quad 0 \leq j < i, \quad i = 1, 2, \dots, n-1.$$

Quiere decir esto que los vectores dirección son ortogonales respecto al producto interno  $x^T A y$  definido por  $A$ . Los vectores de dirección  $p_k$  que satisfacen la propiedad anterior se denominan **vectores conjugados** [27].

El Algoritmo 5 describe el algoritmo básico del método de los gradientes conjugados. En él, el paso  $\alpha_k$  ha sido sustituido por una expresión matemáticamente equivalente. Al igual que en los algoritmos de los métodos de Cauchy y del gradiente espectral, la actualización del vector residual ha sido también sustituida por una actualización matemáticamente equivalente  $r_{i+1} = r_i - \alpha_i w_i$  para evitar el cálculo de un producto matriz vector adicional si  $w_i = A p_i$  había sido ya previamente calculado [27].

---

**Algoritmo 5** Algoritmo básico del método de los gradientes conjugados

---

Elegir  $x_0$  y  $\epsilon$ .  
 $p_0 = r_0 = b - Ax_0$ .  
**for**  $i = 0, 1, 2, 3, \dots$  **do**  
     $w_i = Ap_i$   
     $\alpha_i = \|r_i\|^2 / p_i^t w_i$   
     $x_{i+1} = x_i + \alpha_i p_i$   
     $r_{i+1} = r_i - \alpha_i w_i$   
    Prueba de convergencia: Si  $\|r_{i+1}\|^2 \geq \epsilon$ , continúe.  
     $\beta_i = \|r_{i+1}\|^2 / \|r_i\|^2$   
     $p_{i+1} = r_{i+1} + \beta_i p_i$   
**end for**

---

**3.3.2. Convergencia**

Vamos a dar algunos resultados sobre la convergencia del método de los gradientes conjugados para matrices  $A$  simétricas y positivo definidas. En la ausencia de errores de redondeo, el método de los gradientes conjugados debería converger en no más de  $n$  iteraciones, donde  $n$  representa la dimensión de la matriz, por lo que en teoría es un método directo. Sin embargo, en la práctica esto no necesariamente sucede, y en el caso de matrices muy grandes, incluso sería indeseable utilizar este método como un método directo, por lo que se considera normalmente como un método iterativo.

En teoría, puede probarse que el error en cada iteración decrece [27]:

**Lema 3.3.1** *Si  $x$  es la solución exacta del sistema  $Ax = b$  y  $x_{k-1} \neq x$  entonces:*

$$\|x - x_k\|_2 < \|x - x_{k-1}\|_2.$$

El método de los gradientes conjugados es óptimo en un sentido polinomial. En efecto [63]:

**Teorema 3.3.1** *El iterado  $x_k$  generado por el método de los gradientes conjugados cumple con lo siguiente:*

$$\|x_k - x\|_A^2 = \min_{P_{k-1}} (x_0 - x)^t A (I - AP_{k-1}(A))^2 (x_0 - x).$$

sobre todos los posibles polinomios  $P_{k-1}$  de grado  $k - 1$ .

Una manera de reescribir este teorema en función de los autovalores de  $A$  para sacarle mayor provecho es la siguiente [63]:

**Teorema 3.3.2** Sean  $\lambda_1, \lambda_2, \dots, \lambda_n$  y  $v_1, v_2, \dots, v_n$  los autovalores y sus correspondientes autovectores ortonormales de  $A$ . En el método de los gradientes conjugados se cumple:

$$\|x_k - x\|_A^2 \leq \max_{\lambda_i} (1 - \lambda_i P_{k-1}(\lambda_i))^2 \|x_0 - x\|_A^2,$$

para todo polinomio  $P_{k-1}$  de grado  $k-1$ , y donde el máximo se toma sobre todos los autovalores  $\lambda_i$  de  $A$ .

El teorema anterior permite obtener una cota de velocidad de convergencia del método en función del número de condición espectral  $Cond(A)$ . Esto puede verse en el siguiente resultado [27, 63]:

**Teorema 3.3.3**

$$\|x_k - x\|_A \leq 2\alpha^k \|x_0 - x\|_A,$$

donde  $\|y\|_A = \sqrt{y^t A y}$ ,

$$\alpha = (\sqrt{\kappa} - 1) / (\sqrt{\kappa} + 1),$$

y

$$\kappa = Cond(A) = \|A\|_2 \|A^{-1}\|_2.$$

Esta cota es mucho mejor que la que se obtiene para el método de Cauchy. Sin embargo, para matrices  $A$  muy mal condicionadas, el método de los gradientes conjugados puede ser aún de convergencia lenta. El método de los gradientes conjugados converge también en el caso en que  $f$  sea una función cuadrática convexa, pero no estrictamente convexa ( $Ax = b$  con  $A$  positiva semidefinida), si  $b \in R(A)$  [63].

### 3.3.3. Extensión a espacios de Hilbert

El método de los gradientes conjugados puede extenderse también para operadores en espacios de Hilbert, incluso en dimensión infinita. Podemos considerar la generalización a la ecuación de operadores

$$Au = f,$$

donde  $A$  es un operador lineal, acotado, positivo definido, autoadjunto e invertible en un espacio de Hilbert  $H$ . Bajo estas suposiciones, el sistema  $Au = f$  tiene una

única solución  $u^* = A^{-1}f$  [4].

Igualmente, existen resultados para el método de los gradientes conjugados en el caso de que el operador  $A$  sea singular (el algoritmo es el mismo salvo que el producto  $x^t z$  se cambia a  $\langle x, z \rangle$ ). En este caso, una versión del método de los gradientes conjugados fue desarrollada por Kammerer y Nashed en [57]. En el artículo de Kammerer y Nashed se resuelve un problema de mínimos cuadrados:

$$\text{mín } \|f - Ax\|^2.$$

Es importante resaltar, que en el trabajo de Kammerer y Nashed, no solo se desarrolla una versión de gradientes conjugados que resuelve el problema de mínimos cuadrados para un operador singular  $A$  utilizando el método de los gradientes conjugados, sino que se determina a cual de las soluciones de mínimos cuadrados va a converger el algoritmo. En el capítulo siguiente hablaremos sobre este resultado.

# Capítulo 4

## Aceleración de los métodos de proyecciones alternantes en espacios de Hilbert

### 4.1. Introducción

En este capítulo se hace un estudio más general de la aceleración de los métodos de proyecciones alternantes para espacios de Hilbert. De manera más específica se estudian los casos de los métodos de Cimmino y Von Neumann-Halperin en subespacios. Muchos métodos de proyecciones alternantes pueden verse como una iteración de punto fijo  $x_{k+1} = Tx_k$  con  $x_0$  dada, para un cierto operador acotado y no expansivo  $T$ . En [10] se propone un esquema de aceleración general, para este tipo de iteraciones de punto fijo. En este capítulo, se relaciona este esquema de aceleración con la iteración del método de mínimo descenso para un operador singular en espacios de Hilbert, y proponemos el uso de técnicas más efectivas tales como el método de los gradientes conjugados. De esta manera se obtiene un nuevo esquema de aceleración, en general, más eficiente. Luego, se estudia el caso conveniente de la aceleración cuando el operador  $T$  es autoadjunto y semipositivo definido, caso que ocurre cuando el operador  $T$  está asociado con el método de Cimmino. Finalmente, se muestran experimentos numéricos en donde se comparan los métodos clásicos y sus diferentes aceleraciones, en la resolución de problemas de punto de ensilladura.

## 4.2. Aceleración de los métodos de proyecciones alternantes en subespacios

Sea  $H$  un espacio de Hilbert con el producto interno  $\langle \cdot, \cdot \rangle$  y su norma asociada  $\|\cdot\|$ , y sean  $M_1, M_2, \dots, M_m$  subespacios cerrados (lineales) de  $H$  con  $M = \bigcap_{i=1}^m M_i$ . Para cualquier subespacio  $N$  de  $H$ , denotemos  $P_N$  la proyección ortogonal sobre  $N$ . Los métodos de proyecciones alternantes de von Neumann-Halperin y de Cimmino son miembros de la familia de métodos para determinar la mejor aproximación  $P_M x$  de  $x$  en  $M$ . A continuación, presentaremos algunos teoremas asociados que describirán estos esquemas conocidos.

**Teorema 4.2.1** ([62] ( $m = 2$ ), Halperin [47] ( $m \geq 2$ )) Sean  $M_1, M_2, \dots, M_m$  subespacios cerrados (lineales) de un espacio de Hilbert  $H$  con  $M = \bigcap_{i=1}^m M_i$ . Entonces,

$$\lim_{k \rightarrow \infty} \|(P_{M_m} P_{M_{m-1}} \dots P_{M_1})^k x - P_M x\| = 0.$$

Y basado en una idea de Cimmino [23].

**Teorema 4.2.2** (Kammerer y Nashed [56]) Sean  $M_1, M_2, \dots, M_m$  subespacios cerrados (lineales) de un espacio de Hilbert  $H$  con  $M = \bigcap_{i=1}^m M_i$ . Entonces,

$$\lim_{k \rightarrow \infty} \left\| \left( \frac{1}{m} \sum_{i=1}^m P_{M_i} \right)^k x - P_M x \right\| = 0.$$

Las iteraciones de dichos métodos pueden ser vistas como un esquema de punto fijo  $x_k = T x_{k-1}$  para un operador lineal acotado no-expansivo  $T$  (i.e.,  $\|T\| \leq 1$ , (donde  $\|\cdot\|$  es la norma del operador) [10]. Para el método de von Neumann-Halperin este operador viene dado  $T = P_{M_m} P_{M_{m-1}} \dots P_{M_1}$ , y para el método de Cimmino el operador sería  $T = \frac{1}{m} \sum_{i=1}^m P_{M_i}$ . En ambos casos, esos esquemas convergen al conjunto de puntos fijos de  $T$ ,  $\text{Fix } T = M = \bigcap_{i=1}^m M_i$ , i.e.,

$$\lim_{k \rightarrow \infty} \|T^k x - P_{\text{Fix } T} x\| = 0.$$

Cuando el ángulo entre los subespacios es “pequeño”, ambos métodos pueden presentar una velocidad lenta de convergencia [58, 67, 30, 34, 33, 32, 54], de manera que se requerirán muchas iteraciones para obtener una buena aproximación, lo cual implica un alto costo computacional. Diversos esquemas de aceleración han sido propuestos para ambos métodos, e.g. [46, 44, 20, 1, 54].

En [10], Bauschke et al. propusieron un esquema que generaliza las aceleraciones propuestas por Gubin et al. [46] y por Gearhart y Koshy [44], proponiendo un esquema de aceleración para operadores lineales  $T$  acotados y no-expansivos. En dicho esquema, si un operador de punto fijo  $T$ , acotado y lineal, es tal que

$$\lim_{k \rightarrow \infty} \|T^k x - P_{\text{Fix } T} x\| = 0,$$

se propone un esquema de punto fijo acelerado para  $x_0 \in H$ , descrito como sigue:

$$x_{k+1} = A_T(x_k),$$

donde

$$A_T(x) = t_x T x + (1 - t_x)x.$$

y

$$t_x = \begin{cases} \frac{\langle x, x - Tx \rangle}{\|x - Tx\|^2} & \text{si } Tx \neq x \\ 1 & \text{si } Tx = x. \end{cases}$$

Con esta elección de  $t_x$ ,  $A_T(x)$  es el punto, en la línea que pasa por los puntos  $x$  y  $Tx$ , más cercano a  $P_M x$  [10].

Posteriormente, se mostrará que este esquema acelerado está relacionado con el método de Cauchy (o mínimo descenso) para la resolución de sistemas lineales singulares en espacios de Hilbert [61]. Tomando en cuenta que el método de mínimo descenso es conocido por su lentitud en la presencia de problemas mal condicionados, proponemos un nuevo esquema de aceleración basado en el método de los gradientes conjugados para resolver sistemas lineales singulares en espacios de Hilbert [57]. Esta aceleración puede ser utilizada para acelerar cualquier esquema de punto fijo definido por un operador  $T$  de rango cerrado y no expansivo, incluyendo los operadores relacionados con los métodos de von Neumann-Halperin y de Cimmino.

#### 4.2.1. Aceleración de problemas de punto fijo lineales

Sea  $T$  un operador acotado en un espacio de Hilbert  $H$ . El conjunto de los puntos fijos de  $T$  se denota como:

$$\text{Fix } T = \{x \in H : Tx = x\}.$$

El problema de encontrar los puntos fijos de  $T$  es equivalente a resolver el sistema lineal:

$$(I - T)x = 0.$$

Si existe un elemento  $x \neq 0$  en  $\text{Fix } T$ , el operador  $(I - T)$  es claramente singular. Para este caso, una alternativa es resolver el problema de mejor aproximación:

$$\text{mín } f(x) = \frac{1}{2} \|x - Tx\|^2. \quad (4.1)$$

Si  $\text{Fix } T$  es no vacío, claramente el conjunto de minimizadores globales de  $f(x)$  es igual a  $\text{Fix } T$ . Por lo tanto, el problema de encontrar los puntos fijos de  $T$  puede ser visto como un caso particular del problema de mínimos cuadrados (en espacios de Hilbert):

$$\text{mín } f(x) = \frac{1}{2} \|b - Ax\|^2, \quad (4.2)$$

donde el operador  $A = I - T$  y  $b = 0$ .

El problema (4.2), ha sido tratado previamente por Nashed [61] para operadores singulares en espacios de Hilbert utilizando el método de mínimo descenso, y posteriormente por Kammerer y Nashed [57] usando el método de los gradientes conjugados. En los dos teoremas siguientes resumiremos dichos resultados fundamentales.

**Teorema 4.2.3 (Nashed [61])** *Sea  $H$  un espacio de Hilbert,  $A : H \rightarrow H$  un operador lineal acotado de rango cerrado. La sucesión generada por el método de mínimo descenso aplicada al problema (4.2), converge al elemento  $u \in S = \{z : \inf \|Ax - b\| = \|Az - b\|, x \in H\}$  para cualquier iterado inicial  $x_0 \in H$ . La sucesión converge a  $A^\dagger b$  si y solo si  $x_0 \in R(A^*)$ .*

**Teorema 4.2.4 (Kammerer y Nashed [57])** *Sean  $X$  e  $Y$  dos espacios de Hilbert sobre el campo real y sea  $A$  una transformación lineal acotada de  $X$  a  $Y$ . Si el rango de  $A$  es cerrado, el método de los gradientes conjugados, aplicado al problema (4.2) con iterado inicial  $x_0 \in H$ , converge monótonamente a la solución de mínimos cuadrados  $u = A^\dagger b + (I - P)x_0$  de  $Ax = b$ , donde  $P$  denota la proyección ortogonal de  $X$  a  $\overline{R(A^*)}$ .*

Aquí,  $A^*$  denota el operador adjunto de  $A$ ,  $A^\dagger$  es el operador pseudoinversa de  $A$  y  $\overline{R(A^*)}$  denota la clausura del conjunto  $R(A^*)$  (rango del adjunto de  $A$ ).

Una condición para la aplicación de esos resultados para resolver el problema (4.1) es que el rango del operador  $I - T$  sea cerrado. En espacios de dimensión

finita, esta condición es automáticamente satisfecha. Sin embargo, para espacios infinito dimensionales, la condición de rango cerrado para  $I-T$  debe ser verificada.

Los iterados definidos por el esquema acelerado  $A_T$  propuesto en [10] a partir de  $x_0 \in H$ , son:

$$\begin{aligned} A_T(x_k) &= t_{x_k} T x_k + (1 - t_{x_k}) x_k \\ &= x_k + t_{x_k} (T x_k - x_k) \\ &= x_k + t_k d_k, \end{aligned}$$

donde  $d_k = T x_k - x_k$ , y  $t_x$  está dado por

$$t_{x_k} = \frac{\langle -x_k, d_k \rangle}{\langle d_k, d_k \rangle} \quad \text{para } x_k \notin \text{Fix } T.$$

En consecuencia, este esquema acelerado se corresponde con la sucesión de mínimo descenso aplicada a la resolución del problema  $(I - T)x = 0$ . Dicha secuencia es conocida por ser un método lento. Esto explica la observación hecha en [10] de que el esquema acelerado  $A_T$  no es siempre más rápido que el esquema de punto fijo original cuando el operador  $T$  no es un operador auto-adjunto (simétrico). Este es el caso cuando  $T$  es el operador asociado al método de proyecciones alternantes de von Neumann-Halperin [10]. En cambio, para versiones simétricas, tales como aquellas asociadas con el operador autoadjunto  $T^*T$ , la convergencia de los esquemas acelerados es siempre más rápida que sus versiones originales [10]. Sin embargo, para acelerar esos esquemas de punto fijo auto-adjuntos, vamos a estar más interesados en la sucesión del método de los gradientes conjugados analizada en [57], para sistemas lineales singulares en espacios de Hilbert.

Si el rango  $R(I - T)$  es cerrado, el teorema 4.2.4 implica que el método de los gradientes conjugados aplicado a

$$\text{mín } f(x) = \frac{1}{2} \|(I - T)x\|^2, \quad (4.3)$$

converge a la solución de mínimos cuadrados

$$u = (I - T)^\dagger 0 + (I - P)x_0 = (I - P)x_0,$$

donde  $P$  denota la proyección ortogonal de  $x$  en  $\overline{R(I - T)^*}$ . Más aún,

$$\overline{R(I - T)^*} = \eta(I - T)^\perp = (\text{Fix } T)^\perp.$$

Como  $P_{\text{Fix } T} = I - P_{(\text{Fix } T)^\perp}$ , el método de los gradientes conjugados convergerá a la proyección de  $x_0$  en  $\text{Fix } T$ . Resumiendo obtenemos el siguiente teorema.

**Teorema 4.2.5 (Hernández-Ramos et al. [52])** *Sea  $T$  un operador acotado y no expansivo tal que  $R(I - T)$  es cerrado. Entonces, el método de los gradientes conjugados para minimizar (4.1) con iterado inicial  $x_0 \in H$  converge monótonamente a la proyección de  $x_0$  en  $\text{Fix } T$ .*

Sin embargo, para aplicar este resultado es necesario probar que  $R(I - T)$  es un subespacio cerrado. A continuación, caracterizaremos dicha condición de rango cerrado. Para ello, utilizaremos el siguiente resultado.

**Teorema 4.2.6 (Kulkarni y Nair [59])** *Sea  $A : X \rightarrow Y$  un operador lineal acotado no nulo entre dos espacios de Hilbert  $X$  e  $Y$ . El subespacio  $R(A)$  es cerrado en  $Y$  si y solo si existe  $\gamma > 0$  tal que*

$$\sigma(A^*A |_{\eta(A)^\perp}) \subseteq [\gamma, \|A\|^2],$$

donde  $\sigma(A)$  denota el espectro del operador  $A$  y  $\eta(A)$  es el núcleo de  $A$ .

En otras palabras, para que  $R(A)$  sea cerrado, es suficiente que el operador  $A$  verifique

$$0 < \gamma \leq \inf_{x \in \eta(A)^\perp} \frac{\langle x, A^*Ax \rangle}{\langle x, x \rangle} = \inf_{x \in \eta(A)^\perp} \frac{\|Ax\|^2}{\|x\|^2},$$

y

$$\sup_{x \in \eta(A)^\perp} \frac{\langle x, A^*Ax \rangle}{\langle x, x \rangle} = \sup_{x \in \eta(A)^\perp} \frac{\|Ax\|^2}{\|x\|^2} \leq \|A\|^2.$$

Por la definición de norma, la condición del supremo es verificada automáticamente.

Dado que  $\eta(I - T)^\perp = (\text{Fix } T)^\perp$ , para probar que  $A = I - T$  tiene rango cerrado, es suficiente probar que existe  $\gamma > 0$  tal que el cociente de Rayleigh satisface:

$$\inf_{x \in (\text{Fix } T)^\perp} \frac{\|x - Tx\|^2}{\|x\|^2} \geq \gamma > 0.$$

La definición siguiente es una generalización de la definición del coseno del ángulo entre los subespacios  $M_1, M_2, \dots, M_m$ , dada por Bauschke, Borwein y Lewis [8] ( $\cos(M_1, M_2, \dots, M_m)$ ).

**Definición 4.2.1** *Sea  $T$  un operador acotado no expansivo en un espacio de Hilbert. Definiremos el coseno de  $T$  como*

$$\cos(T) = \|TP_{(Fix\ T)^\perp}\|.$$

Dado que  $T$  es un operador no expansivo ( $\|T\| \leq 1$ ) entonces,

$$0 \leq \cos(T) = \|TP_{(Fix\ T)^\perp}\| \leq \|T\| \|P_{(Fix\ T)^\perp}\| \leq 1.$$

Es sencillo probar que  $\cos(T) = \|T - P_{Fix\ T}\|$ , i.e., es la distancia entre el operador  $T$  y la proyección sobre  $Fix\ T$ . De hecho,

$$\cos(T) = \|TP_{(Fix\ T)^\perp}\| = \|T(I - P_{Fix\ T})\| = \|T - TP_{Fix\ T}\|.$$

Como  $P_{Fix\ T}$  es un punto fijo de  $T$ , tenemos que  $TP_{Fix\ T} = P_{Fix\ T}$ , y obtenemos

$$\cos(T) = \|T - P_{Fix\ T}\|.$$

Cuando el operador  $T$  es el asociado con el método de proyecciones alternantes ( $T = P_{M_m}P_{M_{m-1}} \dots P_{M_1}$ ), entonces  $\cos(T) = \cos(M_1, M_2, \dots, M_m)$  en la definición dada por [8].

**Lema 4.2.1** *Si  $c = \cos(T) < 1$  entonces  $R(I - T)$  es cerrado.*

**Prueba.** Sea  $x \in (Fix\ T)^\perp$ . Por la desigualdad de Cauchy-Schwarz,

$$\|Tx\| = \|TP_{(Fix\ T)^\perp}x\| \leq \|TP_{(Fix\ T)^\perp}\| \|x\| = c\|x\|.$$

Por la desigualdad triangular,

$$\|x - Tx\| \geq \|x\| - \|Tx\|.$$

Por ende, para todo  $x \in (Fix\ T)^\perp$  tenemos,

$$\|x - Tx\| \geq (1 - c)\|x\|$$

y

$$\inf_{x \in (\text{Fix } T)^\perp} \frac{\|x - Tx\|^2}{\|x\|^2} \geq (1 - c)^2 = \gamma > 0. \quad \blacksquare \quad (4.4)$$

En consecuencia, si  $c = \cos(T) < 1$ , podemos aplicar la aceleración de los gradientes conjugados. Resumiremos este resultado en el siguiente teorema:

**Teorema 4.2.7 (Hernández-Ramos et al. [52])** *Sea  $T$  un operador acotado no expansivo tal que  $\cos(T) < 1$ . Entonces, el método de los gradientes conjugados para minimizar (4.1) con el iterado inicial  $x_0 \in H$  converge monótonamente hacia la proyección de  $x_0$  en  $\text{Fix } T$ .*

Resumiendo, si  $\cos(T) < 1$  y si aplicamos el método de los gradientes conjugados para minimizar  $f(x) = \frac{1}{2}\|x - Tx\|^2$ , las iteraciones convergerán al punto fijo de  $T$  más cercano al iterado inicial  $x_0$ , i.e., la sucesión de iterados convergerá a la proyección del iterado inicial  $x_0$ , en los puntos fijos  $\text{Fix } T$ .

Para presentar el algoritmo de aceleración por gradientes conjugados, necesitamos el operador gradiente de  $f(x) = \frac{1}{2}\|x - Tx\|^2$

$$\nabla f(x) = (I - T)^*(I - T)x, \quad (4.5)$$

y su operador Hessiano,

$$\nabla^2 f(x) = (I - T)^*(I - T). \quad (4.6)$$

Note que el operador Hessiano es constante y positivo semi-definido, i.e.,  $f$  es una función cuadrática convexa. De hecho, para cualquier  $x \in H$  tenemos,

$$\langle x, \nabla^2 f x \rangle = \|(I - T)x\|^2 \geq 0.$$

En el Algoritmo 6 mostraremos la aceleración por gradientes conjugados que calcula  $P_{\text{Fix } T}x_0$ , i.e., el punto en  $\text{Fix } T$  más cercano a la aproximación inicial  $x_0$ .

La velocidad de convergencia de este algoritmo está relacionada con las cotas espectrales de los operadores restringidos  $U = \{(I - T)^*(I - T) \mid R(I - T)^*\}$  [57]. Sea  $\mu(x) = \langle Ux, x \rangle / \langle x, x \rangle$ ,  $x \neq 0$ , el cociente de Rayleigh de  $U$ . Como  $U$  es un operador lineal simétrico y positivo semidefinido en espacios de Hilbert  $R(I - T)^* = (\text{Fix } T^*)^\perp = (\text{Fix } T)^\perp$ , las cotas espectrales:

$$s = \inf\{\mu(x) : x \in (\text{Fix } T)^\perp\}, \quad (4.7)$$

y

$$S = \sup\{\mu(x) : x \in (\text{Fix } T)^\perp\} \quad (4.8)$$

son positivas y finitas [57].

Una cota para  $s$  está inmediatamente dada por (4.4),

$$s \geq (1 - c)^2, \quad (4.9)$$

donde  $c = \cos(T)$ .

Igualmente, tenemos que  $\|x - Tx\| \leq \|x\| + \|Tx\|$ . Como para  $x \in (\text{Fix } T)^\perp$  tenemos  $\|Tx\| \leq c$ , luego

$$S \leq (1 + c)^2. \quad (4.10)$$

Por lo tanto, una cota del número de condición espectral del operador restringido  $U$  está dado por

$$\text{cond}((I - T)^*(I - T) | (\text{Fix } T)^\perp) \leq \left[ \frac{1 + c}{1 - c} \right]^2.$$

### 4.2.2. El caso auto-adjunto

Un caso importante es cuando el operador  $T$  es auto-adjunto. Para este caso, mostraremos que la condición  $\cos(T) < 1$  es satisfecha automáticamente bajo suposiciones suaves. Primero, consideraremos algunas propiedades importantes de los operadores autoadjuntos, las cuales serán discutidas de manera más general para operadores normales. Para esto, denotaremos  $M := \text{Fix } T$ .

**Lema 4.2.2** *Si  $T$  conmuta con  $P_M$  entonces  $T$  conmuta con  $P_{M^\perp}$ .*

**Demostración.**  $TP_{M^\perp} = T(I - P_M) = T - TP_M = T - P_M T = (I - P_M)T = P_{M^\perp}T$ . ■

**Lema 4.2.3** *Si  $T$  conmuta con  $P_M$  entonces  $T^k - P_M = (TP_{M^\perp})^k$  para todo  $k$ .*

**Demostración.** Por el lema 4.2.2,  $T$  conmuta con  $P_{M^\perp}$ . Entonces,

$$(TP_{M^\perp})^k = T^k P_{M^\perp} = T^k(I - P_M) = T^k - T^k P_M.$$

$TP_M = P_M$  porque  $P_M$  es un punto fijo de  $T$ , y por lo tanto,

$$(TP_{M^\perp})^k = T^k - T^k P_M = T^k - P_M. \quad \blacksquare$$

---

**Algoritmo 6** Aceleración por gradientes conjugados para problemas de punto fijo lineales

---

INPUT:  $x_0 \in H$   
 OUTPUT:  $P_{\text{Fix } T}(x_0)$

$x = x_0$   
 $r = -(I - T^*)(I - T)x$   
 $rr = \langle r, r \rangle$   
 $u = r$

**for**  $k = 1, 2, \dots$  **do**  
   **if** *convergencia* **then**  
     **return**  $P_{\text{Fix } T}(x_0) = x$   
   **else**  
      $v = (I - T^*)(I - T)u$   
      $\alpha = \frac{rr}{\langle u, v \rangle}$   
      $x = x + \alpha u$   
      $r_0 = rr$   
      $r = r - \alpha v$   
      $rr = \langle r, r \rangle$   
      $\beta = \frac{rx}{r_0}$   
      $u = r + \beta u$   
   **end if**  
**end for**

---

**Teorema 4.2.8** (Hernández-Ramos et al. [52]) *Sea  $T$  un operador normal tal que  $T$  conmuta con  $P_M$ . Entonces, si  $\lim_{k \rightarrow \infty} \|T^k x - P_M x\| = 0$  para todo  $x \in H$ , entonces  $\cos(T) < 1$ .*

**Demostración.** Si  $T$  conmuta con  $P_M$  tenemos que para todo  $x \in H$ ,

$$\lim_{k \rightarrow \infty} \|T^k x - P_M x\| = \lim_{k \rightarrow \infty} \|(TP_{M^\perp})^k x\| = 0.$$

Más aún,  $\lim_{k \rightarrow \infty} \|(TP_{M^\perp})^k x\| = 0$  para todo  $x \in H$  si y solo si  $\lim_{k \rightarrow \infty} \|(TP_{M^\perp})^k\| = 0$  (ver, e.g., [6]), lo cuál implica que  $\lim_{k \rightarrow \infty} \|(TP_{M^\perp})^k\| = 0$ . Como  $TP_{M^\perp}$  es también un operador normal entonces ([6])

$$\|(TP_{M^\perp})^k\| = \|(TP_{M^\perp})\|^k,$$

y por ende

$$\lim_{k \rightarrow \infty} \|(TP_{M^\perp})\|^k = 0.$$

Por lo tanto  $\|(TP_{M^\perp})\| < 1$ , y luego, de la Definición 4.2.1, sigue que

$$\cos(T) = \|(TP_{M^\perp})\| < 1. \quad \blacksquare$$

Resumiendo, si  $T$  es normal y  $T$  conmuta con  $P_{\text{Fix } T}$ , y el método de punto fijo converge, entonces el rango de  $I - T$  es cerrado y podremos aplicar la aceleración de gradientes conjugados. Por ejemplo, si  $T = \frac{1}{m} \sum_{i=1}^m P_{M_i}$  (método de Cimmino), como los operadores de proyección son autoadjuntos entonces  $T$  es también un operador autoadjunto. -Adicionalmente,  $P_M P_{M_i} = P_{M_i} P_M = P_M$ , por lo que  $T$  conmuta con  $P_M$ , y el método de punto fijo converge ( $\lim_{k \rightarrow \infty} \|T^k x - P_M x\| = 0$ ), entonces  $\cos(T) < 1$ . En conclusión, podemos aplicar la aceleración de gradientes conjugados directamente a  $I - T$  para acelerar el método de Cimmino.

### El caso finito dimensional

La aplicación del método de los gradientes conjugados al problema de mínimos cuadrados (4.1) tiene la clara desventaja de incrementar el número de condición en los operadores lineales envueltos. Cuando resolvemos un problema, esto incrementa el número de iteraciones requeridas para la convergencia, como discutiremos en la sección de resultados numéricos.

Si  $T$  es no expansivo, el operador  $I - T$  es claramente positivo semi-definido. Si el operador  $(I - T)$  es simétrico y positivo semidefinido, entonces proponemos la estrategia de aplicar directamente los métodos de optimización descritos anteriormente en la resolución del sistema  $(I - T)x = 0$ , para converger a los puntos fijos de  $T$  [51]. Es importante notar, que en el método de Cimmino el operador  $I - T$  es simétrico y positivo semidefinido. Basados en esto, en el resto de este trabajo nos interesaremos en este método en vez de otros métodos donde el operador  $T$  no es simétrico. Es decir, vamos a preferir acelerar métodos del tipo Cimmino, donde el operador es simétrico, de manera que podamos aplicarle el método de los gradientes conjugados para acelerarlo.

### 4.3. Experimentos numéricos

#### 4.3.1. Problemas de punto de ensilladura con métodos de proyecciones alternantes

En años recientes, los problemas de punto de ensilladura (*saddle point problems*) han estado ganando popularidad en muchas áreas de la computación científica. Ellos aparecen en muchas aplicaciones tales como: sistemas KKT en optimización, formulaciones mixtas en dinámica de fluidos y en discretizaciones por descomposición de dominios para la resolución paralela de ecuaciones en derivadas parciales (PDE) [12, 35, 65]. En esta subsección, reformulamos el problema de punto de ensilladura como el problema, equivalente, de calcular la proyección ortogonal en la intersección de muchos subespacios.

Estamos interesados en resolver el sistema de ecuaciones lineales:

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}, \quad (4.11)$$

donde:

(H1)  $B \in \mathfrak{R}^{r \times n}$  es una matriz con  $\text{rango}(B) = r, r \leq n$ ,

(H2)  $A \in \mathfrak{R}^{n \times n}$  es una matriz simétrica positivo definida.

Los siguientes dos lemas son bien conocidos (e.g. [12, 50]). Sin embargo, vamos a presentarlos por razones de completitud.

**Proposición 4.3.1** *Bajo las hipótesis (H1,H2) el sistema (4.11) tiene una única solución  $(x, \lambda)^T$  la cuál satisface:*

$$\begin{cases} f - Ax \perp \ker B, \\ x \in \ker B, \end{cases} \quad (4.12)$$

donde  $\ker B = \{x : Bx = 0\}$ , es el llamado núcleo o kernel de la matriz  $B$ .

Como  $A$  es simétrica positivo definida. Definiendo el producto escalar  $\langle x, y \rangle_A = \langle x, Ay \rangle$ , tenemos:

**Proposición 4.3.2** *Bajo las hipótesis (H1,H2), si  $(x, \lambda)^T$  resuelve (4.11), entonces el vector  $x$  corresponde a la proyección ortogonal de  $x_u = A^{-1}f$  en  $\ker B$ , en el producto escalar  $\langle \cdot, \cdot \rangle_A$ .*

En este caso escribimos la variedad lineal como  $V = \{x_u\} + (\ker B)^{\perp_A}$  [50].

Demos una partición de la matriz  $B \in \mathfrak{R}^{r \times n}$  en  $m$  bloques de filas:

$$B^T = [B_1^T, B_2^T, \dots, B_m^T].$$

Entonces,

$$\ker B = \bigcap_{i=1}^m \ker B_i.$$

De la proposición (4.3.2), se tiene que la solución  $x$  para el problema de punto de ensilladura (4.11) es la proyección A-ortogonal de  $x_u = A^{-1}f$  en el subespacio  $\ker B = \bigcap_{i=1}^m \ker B_i$ . Por lo tanto, este problema puede considerarse como un problema de mejor aproximación o BAP por sus siglas en inglés (*Best Approximation Problem*), en el producto escalar  $\langle \cdot, \cdot \rangle_A$ , y puede ser resuelto por métodos de proyecciones alternantes (o sus variantes aceleradas), tomando como iterado inicial  $x_0 = x_u = A^{-1}f$ , como ejemplo, por el método de Cimmino:

$$x_{k+1} = \frac{1}{m} \sum_{i=1}^m P_{M_i} x_k, \quad (4.13)$$

o por el método de Von Neumann-Halperin:

$$x_{k+1} = P_{M_m} \dots P_{M_1} x_k. \quad (4.14)$$

El operador  $P_{M_i}$  denota la proyección A-ortogonal en  $M_i$ . Como cada  $\ker B_i$  es también un subespacio, podremos obtener la A-proyección de cualquier vector  $y \in \mathfrak{R}^n$  en  $\ker B_i$ ,  $P_{M_i}y$ , calculando  $x$  de un problema de punto de ensilladura más pequeño:

$$\begin{pmatrix} A & B_i^T \\ B_i & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} Ay \\ 0 \end{pmatrix}, \quad (4.15)$$

Esos problemas de punto de ensilladura mucho más pequeños pueden ser resueltos por cualquier método, lo cuál va a depender del tamaño de cada bloque. Las posibles elecciones incluyen métodos de subespacios de Krylov, tales como MINRES [66], o la resolución del sistema simétrico y positivo definido asociado con el complemento de Schur de cada bloque  $B_i A^{-1} B_i^T$  que proviene de (4.15), cuya dimensión es  $m_i \times m_i$ , donde  $m_i$  es el número de filas de cada bloque  $B_i$

[12].

### 4.3.2. Descripción de los experimentos

Para nuestros experimentos, compararemos los métodos de Cimmino y de Von Neumann-Halperin (MAP) con las diferentes versiones aceleradas:

1. Métodos basados en Von-Neumann-Halperin:
  - a) Método clásico de Von-Neumann-Halperin (MAP).
  - b) Esquema acelerado de Bauschke et al. (KG: Aceleración de Koshy y Gearhart en espacios de dimensión finita).
  - c) Gradientes conjugados para mín  $f(x) = \frac{1}{2}\|x - Tx\|^2$  donde  $T$  es el operador asociado con el método de Von Neumann-Halperin (MAP-CG-MC), propuesto en este trabajo.
2. Métodos basados en el método de Cimmino:
  - a) Método de Cimmino clásico (Cimmino).
  - b) Método de Gradientes Conjugados para mín  $f(x) = \frac{1}{2}\|x - Tx\|^2$  donde  $T$  es el operador asociado con el método de Cimmino (CIM-CG-MC).
  - c) Método de Gradientes Conjugados para resolver  $(I - T_S)x = 0$  donde  $T_S$  es el operador simétrico positivo semidefinido asociado con el método de Cimmino (Cimmino-CG) (Capítulo anterior).

Compararemos estos métodos en la resolución de problemas de punto de ensilladura provenientes de:

- Una selección de problemas de punto de ensilladura de la colección CUTER [45].
- Un conjunto de problemas de punto de ensilladura de Stokes generados usando el software IFISS asociado con el libro de Elman et al. [38] (stokes\_testproblems):
  - STOKES1: *Channel domain with natural outflow boundary.*
  - STOKES2: *Flow over a backward facing step.*
  - STOKES3: *Lid driven cavity.*

- STOKES4: *Colliding flow*.

- Varios problemas de punto de ensilladura de descomposición de dominios discutidos en [50] (DD1, DD2, DD3 and DD4).

Todos estos experimentos fueron realizados en una Intel Core 2 Duo usando MATLAB 7.6. En todas nuestras experiencias y en todos los métodos considerados, el criterio de parada del proceso fue cuando la A-norma del residual (la norma definida por  $\|x\|_A^2 = \langle x, Ax \rangle$ ) es menor o igual a  $10^{-8}$ . Para todos los conjuntos de problemas, la matriz  $A \in \mathfrak{R}^{n \times n}$  es simétrica y positivo definida, y  $B \in \mathfrak{R}^{m \times n}$  es una matriz de rango completo por filas. Para los problemas de la biblioteca CUTer, y también para los problemas de Stokes, tomamos el bloque (2,2) de la matriz de punto de ensilladura como la matriz cero, para obtener un sistema como (4.11). Es importante notar que una técnica de Lagrangeano aumentado fue utilizada para mejorar el comportamiento de todos los métodos considerados, y también para garantizar que la matriz  $A$  sea positivo definida. Para ser precisos, la matriz  $A$  en (4.11) fue reemplazada por la matriz  $A + \gamma B^T B$ , donde  $\gamma = \|A\|_F / \|B\|_F^2$  [12]. Más aún, para los problemas de la biblioteca CUTer, añadimos al bloque superior izquierdo de (4.11) el escalar 0,01 veces la matriz identidad, para mejorar el número de condición del bloque. Todos los experimentos fueron obtenidos con una partición de  $B$  de un bloque por fila (número de subespacios  $m = r$ , donde  $r$  es el número de filas de la matriz  $B$ ).

## Resultados

Comparación de resultados.

- Aceleraciones basadas en MAP:

La tabla 4.1 muestra el comportamiento de la convergencia de las diferentes aceleraciones de los métodos de von Neumann-Halperin. Entre los métodos basados en el método de von-Neumann-Halperin, no hay uno del que se pueda decir que sea el mejor en todos los problemas. La aceleración vía gradientes conjugados del MAP tiene las siguientes desventajas: la pérdida de condicionamiento producida por la resolución de un problema de mínimos cuadrados y el alto costo por iteración de hacer cada proyección dos veces por iteración. Sin embargo, fue bastante consistente en todos los problemas. La aceleración propuesta por Koshy y Gearhart [44] fue inconsistente en

varios de esos problemas y requirió un gran número de iteraciones para su convergencia.

- Aceleraciones basadas en el método de Cimmino:

La tabla 4.2 muestra el comportamiento de la convergencia de las diferentes aceleraciones del método de Cimmino. La aceleración por gradientes conjugados aplicada al sistema  $(I - T)x = 0$  es claramente superior que las otras aceleraciones, incluyendo aquellas basadas en el método de Von Neumann. Tanto en el número de iteraciones, como en el tiempo de CPU, esta aceleración fue en general mejor que las otras. Como  $I - T$  es un operador simétrico, se evitó el mal condicionamiento que proviene de resolver un problema de mínimos cuadrados. El costo por iteración no aumenta porque en cada iteración se proyecta sobre cada subespacio solamente una vez.

Problema			MAP		KG		MAP-CG-MC	
Name	$m$	$n$	iter	tiempo	iter	tiempo	iter	tiempo
AUG2DCQP	1600	3280	*	*	*	*	445	746
CVXQP1S	50	100	652	1.42	312	1.70	53	0.39
CVXQP2S	25	100	13	0.03	11	0.03	39	0.09
CVXQP3S	75	100	*	*	*	*	206	1.20
DUALC1	215	223	2	0.06	*	*	2	0.13
DUALC5	278	285	2	0.08	*	*	2	0.16
DUALC8	503	510	1	0.30	*	*	2	0.47
GOULDQP2S	349	659	*	*	*	*	1	0.38
KSIP	1001	1021	*	*	*	*	21	20.06
MOSARQP1	700	3200	2	6.28	*	*	4	6.97
PRIMAL1	85	410	20	0.39	21	0.39	14	0.56
PRIMAL2	96	745	19	0.66	19	0.74	14	0.72
PRIMAL3	111	856	33	1.75	31	1.75	19	1.81
PRIMAL4	75	1564	17	1.19	18	1.23	14	1.94
PRIMALC1	9	239	7	0.03	*	*	4	0.03
PRIMALC2	7	238	6	0.03	*	*	3	0.03
QGROW15	300	645	6	0.53	*	*	6	0.59
QGROW22	440	946	6	0.84	*	*	9	1.41
QSCFXM3	990	1800	*	*	*	*	293	86.98
STOKES1	1024	2178	*	*	*	*	256	173
STOKES2	2816	5890	*	*	*	*	247	94.52
STOKES3	256	578	306	7.31	199	4.84	109	5.45
STOKES4	1024	2178	264	7.25	187	5.14	96	5.36
DD1	80	1600	199	1.66	156	1.32	102	1.80
DD2	85	1925	546	8.81	*	*	134	4.36
DD3	120	3600	215	6.75	151	4.31	135	9.56
DD4	125	3975	*	*	*	*	209	13.73

Cuadro 4.1: Comparación entre las diferentes aceleraciones del método de Von Neumann-Halperin

Problema			Cimmino		CIM-CG-MC		Cimmino-CG	
Name	$m$	$n$	iter	tiempo	iter	tiempo	iter	tiempo
AUG2DCQP	1600	3280	*	*	811	1380	114	99.17
CVXQP1S	50	100	*	*	104	1.70	54	0.33
CVXQP2S	25	100	*	*	25	0.09	17	0.06
CVXQP3S	75	100	*	*	605	3.80	133	0.53
DUALC1	215	223	*	*	2	0.19	2	0.09
DUALC5	278	285	*	*	2	0.16	2	0.16
DUALC8	503	510	*	*	2	0.48	2	0.38
GOULDQP2S	349	659	*	*	2	0.38	2	0.38
KSIP	1001	1021	*	*	11	11.37	6	3.92
MOSARQP1	700	3200	*	*	4	7.05	2	6.38
PRIMAL1	85	410	*	*	21	0.72	10	0.28
PRIMAL2	96	745	*	*	21	1.00	10	0.45
PRIMAL3	111	856	*	*	32	2.98	13	0.88
PRIMAL4	75	1564	*	*	21	1.94	10	1.08
PRIMALC1	9	239	133	0.33	4	0.03	3	0.03
PRIMALC2	7	238	101	0.23	4	0.03	3	0.03
QGROW15	300	645	*	*	7	0.70	5	0.47
QGROW22	440	946	*	*	10	1.61	7	0.91
QSCFXM3	990	1800	*	*	*	*	149	32.16
STOKES1	1024	2178	*	*	111	82.94	72	27.38
STOKES2	2816	5890	*	*	259	99.55	101	22.89
STOKES3	256	578	*	*	139	7.17	44	1.30
STOKES4	1024	2178	*	*	97	5.64	35	1.28
DD1	80	1600	*	*	179	3.34	62	0.63
DD2	85	1925	*	*	182	6.23	79	1.36
DD3	120	3600	*	*	240	16.28	72	2.72
DD4	125	3975	*	*	209	13.73	102	3.80

Cuadro 4.2: Comparación entre los diferentes métodos acelerados de Cimmino

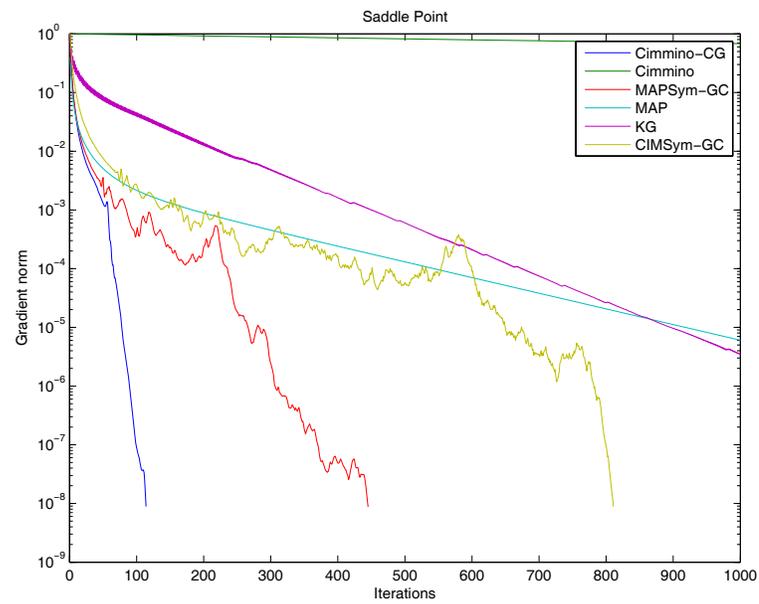


Figura 4.1: Aceleraciones para el problema CUTer (AUG2DCQP)

# Capítulo 5

## El método de Cimmino acelerado para subespacios

### 5.1. Introducción

En esta sección, nos concentraremos en la aceleración del método de Cimmino para subespacios aplicado directamente al sistema  $(I - T)x = 0$  (caso autoadjunto), que fue el que obtuvo mejores resultados numéricos en el capítulo anterior. Este esquema se estudiará desde el punto de vista de optimización, como un método para minimizar una cierta función cuadrática convexa  $f$ . Desde este punto de vista, mostraremos que el método clásico de Cimmino corresponde a la aplicación de un método que va en la dirección del gradiente negativo a paso constante (método de mínimo descenso a paso constante o estacionario). Asimismo, se propone el uso de otras técnicas para minimizar  $f$ , tales como el método del gradiente espectral (GE), y el método de los gradientes conjugados (CG). Se observa que las aceleraciones propuestas no pierden las propiedades de paralelismo del método clásico de Cimmino [11].

Después de considerar este nuevo enfoque desde el punto de vista de optimización, relacionaremos teóricamente el espectro de la Hessiana de la función  $f$  con los ángulos de los subespacios involucrados.

Finalmente, se hacen experimentos numéricos, primero para comparar las diferentes técnicas de aceleración, y luego para comparar el esquema propuesto con otros esquemas de solución de problemas de punto de ensilladura o *saddle point problems* (SPP). La intención es mostrar que las aceleraciones propuestas, pudieran ser competitivas con otros métodos a la hora de resolver problemas de

punto de ensilladura en ambientes paralelos.

## 5.2. Aceleración del método de Cimmino para subespacios

Sean  $M_i$ ,  $i = 1, \dots, m$ , subespacios cerrados de un espacio de Hilbert  $H$ . Sea  $P_{M_i}$  la proyección ortogonal sobre  $M_i$ . Consideremos la siguiente función a minimizar:

$$f(x) = \frac{1}{2} \sum_{i=1}^m \|x - P_{M_i}x\|^2 = \frac{1}{2} \sum_{i=1}^m \|(I - P_{M_i})x\|^2. \quad (5.1)$$

La función  $f$  es una función cuadrática no negativa llamada *función de proximidad* [16]. Si un vector  $x \in \cap_{i=1}^m M_i$ , entonces  $f(x) = 0$  y  $x$  minimiza  $f$ . Para utilizar técnicas de minimización, necesitamos el gradiente de  $f$ ,

$$\nabla f(x) = \sum_{i=1}^m (x - P_{M_i}x), \quad (5.2)$$

y el Hessiano,

$$\nabla^2 f(x) = \sum_{i=1}^m (I - P_{M_i}). \quad (5.3)$$

Es fácil constatar que en este caso el operador Hessiano es una matriz constante positiva semidefinida. De hecho,

$$\langle x, \nabla^2 f x \rangle = \sum_{i=1}^m \|x - P_{M_i}x\|^2 \geq 0.$$

Adicionalmente,  $\langle x, \nabla^2 f x \rangle = 0$  si y solo si  $x \in \cap_{i=1}^m M_i$ . En consecuencia, observamos que el método clásico de Cimmino [23] es el método del gradiente con tamaño de paso constante:

$$x_{k+1} = x_k - \frac{1}{m} \nabla f(x). \quad (5.4)$$

En general, el método del gradiente a paso constante es más lento que el método clásico de mínimo descenso (o método de Cauchy), el cuál es ya famoso por ser lento. Esto explica la bien conocida lentitud del método de Cimmino. La

aceleración del método de Cimmino puede ser llevada a cabo usando otras técnicas más eficientes de optimización para minimizar la función cuadrática convexa  $f(x)$ .

En este capítulo, proponemos acelerar y comparar los métodos clásicos de Cimmino, con aceleraciones basadas en métodos más convenientes de minimización de funciones cuadráticas convexas, como son: el método del gradiente espectral (GE), e inclusive el método de gradientes conjugados (CG) que tiene propiedades óptimas para minimizar funciones cuadráticas convexas y es una elección óptima para nuestra aceleración. Ver, e.g., [5].

Si denominamos  $M = \cap_{i=1}^m M_i$ . Como  $M \subseteq M_i$ , se tiene  $P_M P_{M_i} = P_{M_i} P_M = P_M$ . De allí se deduce que

$$\begin{aligned} P_M \nabla f(x) &= P_M \sum_{i=1}^m (x - P_{M_i} x) \\ &= \sum_{i=1}^m P_M (x - P_{M_i} x) \\ &= \sum_{i=1}^m (P_M x - P_M P_{M_i} x) \\ &= \sum_{i=1}^m (P_M x - P_M x) \\ &= 0. \end{aligned}$$

Por lo tanto, como en el método del gradiente espectral,

$$x_{k+1} = x_k - \frac{1}{\alpha_k} \nabla f(x_k), \quad (5.5)$$

se tiene que

$$P_M x_{k+1} = P_M x_k - \frac{1}{\alpha_k} P_M \nabla f(x_k) = P_M x_k, \quad (5.6)$$

por lo que todos los iterados tienen la misma proyección sobre  $M$ . Como la iteración converge a un punto  $s \in M$ , se tiene:

$$s = P_M(s) = P_M(x_0),$$

y en consecuencia, la iteración GE convergerá hacia  $P_M(x_0)$ . Esto lo podemos resumir en el siguiente teorema:

**Teorema 5.2.1 (Hernández-Ramos [51])** *El método del gradiente espectral aplicado a la minimización de  $f(x) = \frac{1}{2} \sum_{i=1}^m \|x - P_{M_i}x\|^2$ , con iterado inicial  $x_0$ , convergerá al punto  $P_M(x_0)$ .*

De manera similar, podemos desarrollar una versión acelerada del método de Cimmino minimizando la función cuadrática convexa  $f$  usando el conocido método de los gradientes conjugados (CG). El método CG tiene propiedades óptimas para minimizar funciones cuadráticas convexas, y por eso es una opción atractiva para nuestro trabajo. Ver, e.g., [5].

Como vimos en el capítulo anterior, estas aproximaciones, tanto por gradientes conjugados, como por el método del gradiente espectral, son equivalentes a aplicar dichos métodos en la resolución de los sistemas lineales simétrico positivo semidefinidos  $\nabla f(x) = 0$ , i.e.,

$$(I - T_S)x = 0.$$

donde  $T_S = \sum_{i=1}^m P_{M_i}x$  es el operador lineal asociado al método de Cimmino.

En el Algoritmo 7 presentaremos el esquema de aceleración de gradientes conjugados para el método de Cimmino.

De la misma manera a como se hizo con el método del gradiente espectral, llamando  $M = \cap_{i=1}^m M_i$ . Como  $M \subseteq M_i$ , se tiene  $P_M P_{M_i} = P_{M_i} P_M = P_M$ . Si  $T_S = \sum_{i=1}^m P_{M_i}x$ , se puede probar que:

$$P_M(I - T_S)x = 0$$

para todo  $x$ . Por lo que se puede probar inductivamente, que en el Algoritmo 7 siempre se cumple:

$$P_M(r) = P_M(v) = P_M(u) = 0,$$

en cualquier iteración, y que por lo tanto:

$$P_M(x_{k+1}) = P_M(x_k)$$

si llamamos  $x_k$  al valor de la variable  $x$  en la iteración  $k$ .

Esto prueba que la iteración CG converge a un punto  $s = P_M(s) = P_M(x_0)$  y que por lo tanto la aceleración de gradientes conjugados convergerá a la proyección del iterado inicial en el conjunto  $M = \cap_{i=1}^m M_i$ . ■

Lo anterior se puede resumir en el siguiente teorema:

---

**Algoritmo 7** Aceleración por gradientes conjugados del método de Cimmino
 

---

INPUT:  $x_0 \in H$ OUTPUT:  $P_{Fix T}(x_0)$  $x = x_0$  $r = -(I - T_S)x$  $rr = \langle r, r \rangle$  $u = r$ **for**  $k = 1, 2, \dots$  **do**  **if** *convergencia* **then**    **return**  $P_{Fix T}(x_0) = x$   **else**     $v = (I - T_S)u$      $\alpha = \frac{rr}{\langle u, v \rangle}$      $x = x + \alpha u$      $r0 = rr$      $r = r - \alpha v$      $rr = \langle r, r \rangle$      $\beta = \frac{rr}{r0}$      $u = r + \beta u$   **end if****end for**


---

**Teorema 5.2.2 (Hernández-Ramos [51])** *El método de los gradientes conjugados aplicado al sistema  $(I - T_S)x = 0$ , con iterado inicial  $x_0$  convergerá al punto  $P_M(x_0)$ .*

Hay que mencionar que, al igual que en el método original de Cimmino, el principal costo computacional de ambas aceleraciones radica principalmente en el cálculo del gradiente en cada iteración. Para calcular el gradiente, es necesario obtener la proyección del iterado actual  $x_k$  en cada uno de los subespacios. Como ya dijimos, en nuestro caso el cálculo del gradiente puede ser efectuado en paralelo.

La velocidad de convergencia de estos métodos está relacionada con la localización de los autovalores de la Hessiana [66, 64, 26, 40]. Por esta razón, en nuestra próxima sección estableceremos una conexión clave entre el espectro de la Hessiana de  $f$  y los ángulos formados por los subespacios involucrados en el proceso de proyección. Esta conexión añade entendimiento a la velocidad de convergencia del método de Cimmino acelerado.

### 5.3. Relación entre el espectro de la Hessiana y la geometría del problema

En esta sección, mostraremos un interesante resultado teórico que relaciona los cocientes de Rayleigh de la matriz Hessiana y la geometría del problema.

La convergencia de los métodos diseñados para minimizar la función cuadrática

$$f(x) = \frac{1}{2} \sum_{i=1}^m \|x - P_{M_i}x\|^2 = \frac{1}{2} \sum_{i=1}^m \|(I - P_{M_i})x\|^2 \quad (5.7)$$

están usualmente relacionados con el espectro de la matriz Hessiana

$$\nabla^2 f = \sum_{i=1}^m (I - P_{M_i}). \quad (5.8)$$

Para cualquier vector  $x \in H$ , el cociente de Rayleigh  $R(x)$  está dado por:

$$R(x) = \frac{\langle x, \nabla^2 f x \rangle}{\langle x, x \rangle}.$$

Es conocido que:

$$\lambda_{min} \leq R(x) \leq \lambda_{max},$$

donde  $\lambda_{min}$  y  $\lambda_{max}$  son el más pequeño y el más grande autovalor de la Hessiana  $\nabla^2 f$ , respectivamente. Denotamos por  $P_{M_i^\perp} = I - P_{M_i}$  la proyección ortogonal en  $M_i^\perp$ , que es un operador idempotente y autoadjunto en el producto  $\langle \cdot, \cdot \rangle$ . Entonces,

$$\begin{aligned} \langle x, \nabla^2 f x \rangle &= \langle x, \sum_{i=1}^m (x - P_{M_i}x) \rangle \\ &= \sum_{i=1}^m \langle x, x - P_{M_i}x \rangle \\ &= \sum_{i=1}^m \langle x, P_{M_i^\perp}x \rangle. \end{aligned}$$

y

$$\sum_{i=1}^m \langle x, P_{M_i^\perp}x \rangle = \sum_{i=1}^m \langle P_{M_i^\perp}x, P_{M_i^\perp}x \rangle$$

$$= \sum_{i=1}^m \|P_{M_i^\perp} x\|^2.$$

Luego, para todo  $x \in H - \{0\}$ , el cociente de Rayleigh es,

$$R(x) = \sum_{i=1}^m \frac{\|P_{M_i^\perp} x\|^2}{\|x\|^2}.$$

Cuando  $M$  es un subespacio cerrado de un espacio de Hilbert, tenemos para todo  $x \in H - \{0\}$ ,

$$\cos(\text{span}\{x\}, M) = \begin{cases} \|P_M(x)\|/\|x\| & \text{si } x \notin M \\ 0 & \text{if } x \in M \end{cases}$$

[33].

Como consecuencia, para todo  $x \in H - \{0\}$  tenemos,

$$R(x) = \sum_{i=1}^m \sin^2(\text{span}\{x\}, M_i), \quad (5.9)$$

donde  $\sin(\text{span}\{x\}, M_i)$  es el seno del ángulo entre  $\text{span}\{x\}$  and  $M_i$  (en el producto  $\langle \cdot, \cdot \rangle$ ) y  $\text{span}\{x\}$  es el subespacio generado por el vector  $x$ .

De allí tenemos,

$$l \leq R(x) \leq m,$$

donde

$$l = \min_{i \neq j} \sin^2(M_j, M_i), \quad i, j = 1, \dots, m. \quad (5.10)$$

Aquí,  $\sin(M_j, M_i)$  es el seno del ángulo entre los subespacios  $M_j$  y  $M_i$  [33].

El resultado obtenido en esta sección, indica que las propiedades espectrales de la matriz Hessiana de  $f$ , para todas las variantes del método de Cimmino, dependen de los senos de los ángulos formados por los subespacios  $M_i$ . Si los ángulos son *pequeños* entonces todas las variantes del método de Cimmino convergerán *lentamente*. Por otra parte, si los subespacios son *casi* ortogonales entre ellos, entonces todas las variantes del método de Cimmino convergerán *rápidamente*.

## 5.4. Experimentos numéricos

Aplicamos el método de Cimmino y sus versiones aceleradas (Cimmino-gradiente-espectral y Cimmino-gradientes-conjugados) para resolver los siguientes problemas de punto de ensilladura (*saddle point problems*):

- Una selección de problemas de punto de ensilladura de la colección CUTER [45].
- Un conjunto de problemas de punto de ensilladura generados con el software IFISS de flujo incompresible asociado con el libro de Elman et al. [38] (*stokes\_testproblems*):
  - STOKES1: *Channel domain with natural outflow boundary.*
  - STOKES2: *Flow over a backward facing step.*
  - STOKES3: *Lid driven cavity.*
  - STOKES4: *Colliding flow.*
- Varios problemas de punto de ensilladura (*saddle point*) de descomposición de dominios discutidos en [50] (DD1, DD2, DD3 and DD4).

Todos nuestros experimentos fueron realizados en una Intel Core 2 Duo usando MATLAB 7.6. En todos nuestros experimentos y para todos los métodos considerados, el criterio de parada de los procesos fue cuando la norma Euclidea del residual era menor o igual a  $10^{-7}$ . Comparamos las versiones aceleradas de Cimmino contra CG-Uzawa, CG-AOP y el método MINRES. CG-Uzawa es el método de los gradientes conjugados sobre el sistema condensado (también conocido como sistema del complemento de Schur)  $BA^{-1}B^T\lambda = BA^{-1}F$  [12]. El método CG-AOP es una versión mejorada de CG-Uzawa, desarrollada recientemente en [50]. Incluimos el método CG-AOP en nuestra comparación porque es competitivo con algunos de los conocidos esquemas de resolución de problemas de punto de ensilladura que aparecen en descomposición de dominios [50]. También incluimos MINRES porque este método de Krylov ha probado ser competitivo para resolver problemas de punto de ensilladura [12]. Desafortunadamente, para los métodos de Cimmino acelerados propuestos en este trabajo, no tomamos en cuenta las estrategias de preconditionamiento disponibles desarrolladas para los métodos tipo Krylov (e.g., MINRES). Esto porque, si multiplicamos un sistema

de punto de ensilladura por cualquier preconditionador estándar disponible, el sistema resultante no será necesariamente un sistema de punto de ensilladura, y nuestra estrategia de optimización no podrá ser aplicada. Aquí, compararemos MINRES sin preconditionamiento contra las versiones aceleradas del método de Cimmino. Sin embargo, para problemas de Stokes, compararemos nuestros esquemas con una versión preconditionada de MINRES. Este último experimento será descrito a continuación. Para todos los conjuntos de problemas, la matriz  $A \in \mathfrak{R}^{n \times n}$  es simétrica y positivo definida y  $B \in \mathfrak{R}^{r \times n}$  es una matriz de rango completo. Para los problemas del CUTer, y también para los problemas de Stokes, tomamos el bloque  $(2, 2)$  de la matriz de punto de ensilladura como la matriz cero, para obtener un sistema tal como (4.11). Es importante notar, que una técnica de Lagrangeano aumentado ha sido usada para mejorar el rendimiento de todos los métodos considerados y también para garantizar que la matriz  $A$  sea positiva definida. Para ser precisos, la matriz  $A$  en (4.11) ha sido reemplazada por  $A + \gamma B^T B$ , donde  $\gamma = \|A\|/\|B\|^2$  [12]. Más aún, para los problemas: PRIMAL, QGROW, y QSCFXM3, de la biblioteca CUTer, también agregamos en el bloque superior izquierdo de (4.11) el escalar 0,01 veces la matriz identidad, para mejorar el número de condición de ese bloque. Todos los experimentos fueron obtenidos con una partición de una fila-bloque de  $B$  ( $m = r$ ). Este caso representa una ventaja importante porque la matriz de complemento de Schur del sistema (4.15), (necesaria para calcular  $P_{M_i}$ ,  $i = 1, \dots, m$ ), tiene dimensión  $1 \times 1$ . Entonces, la matriz de complemento de Schur es calculada resolviendo el sistema lineal  $Aw_i = B_i^T$ ,  $i = 1, \dots, m$  (que puede ser también resuelto en paralelo). Más aún, la matriz  $A$ , la cual es usualmente grande y rala, puede ser factorizada solamente una vez al inicio del proceso. No es necesario resolver ningún otro sistema lineal con la matriz  $A$  para calcular  $P_{M_i}$ . En nuestros experimentos, la matriz  $A$  es factorizada (usando Cholesky) sin tomar ventaja de una posible estructura adicional de la matriz  $A$ .

### 5.4.1. Resultados

La Tabla 5.1 muestra el comportamiento de la convergencia de las diferentes aceleraciones del método de Cimmino. Aquí podemos ver que las versiones aceleradas son claramente mejores que el método de Cimmino clásico tanto en el número de iteraciones como en el tiempo de CPU. Para todos los problemas, la versión de gradientes conjugados del método de Cimmino presentó el mejor

rendimiento de todas las aceleraciones consideradas. Las Figuras 5.1, 5.2, y 5.3 muestran la evolución de la norma del gradiente por iteración en las versiones aceleradas para un problema de la biblioteca CUTer, un problema de Stokes y un problema de descomposición de dominios. Podemos observar el comportamiento típico de estos rápidos métodos cuando minimizan una función cuadrática convexa.

La Tabla 5.2 compara la versión gradientes conjugados del método de Cimmino versus gradientes conjugados sobre el sistema del complemento de Schur (CG-Uzawa y CG-AOP) y MINRES. Podemos observar que en algunos problemas la versión gradientes conjugados del método de Cimmino (CG-Cimmino) puede ser mejor que los métodos de gradientes conjugados sobre el complemento de Schur o MINRES en una máquina de un solo procesador. El cociente *tiempo/m* en la Tabla 5.2 muestra la potencial aceleración de CG-Cimmino en máquinas paralelas. Las Figuras 5.4, 5.5, y 5.6 muestran la evolución de la norma del gradiente por iteración para un problema de la biblioteca CUTer, un problema de Stokes y uno de descomposición de dominios.

La Tabla 5.3 compara la versión gradientes conjugados del método de Cimmino versus gradientes conjugados sobre el complemento de Schur (CG-Uzawa y CG-AOP) con el MINRES preconditionado y el MINRES para problemas de Stokes. En este caso, el preconditionador seleccionado fue el diagonal por bloques (“*option 2*”) de la rutina *'it-solve'* en el software IFISS [38], que es considerada una opción a elegir para problemas de Stokes. El cociente *tiempo/m* en la Tabla 5.3, de nuevo, muestra la aceleración potencial para CG-Cimmino's en máquinas paralelas.

Problema			Cimmino GE		Cimmino CG		Cimmino	
Nombre	$m$	$n$	iter	tiempo	iter	tiempo	iter	tiempo
AUG2DCQP	1600	3280	157	98.14	85	71.81	471	143.20
CVXQP1S	50	100	79	0.34	33	0.16	237	0.58
CVXQP2S	25	100	13	0.03	11	0.03	39	0.09
CVXQP3S	75	100	242	0.77	85	0.30	726	2.30
DUALC1	215	223	3	0.09	2	0.03	9	0.22
DUALC5	278	285	3	0.13	2	0.08	9	0.25
DUALC8	503	510	3	0.63	2	0.52	9	0.97
GOULDQP2S	349	659	1	0.31	1	0.31	1	0.30
KSIP	1001	1021	1	0.70	1	0.70	1	0.69
MOSARQP1	700	3200	1	0.34	1	0.34	1	0.34
MOSARQP2	600	1500	1	0.16	1	0.16	1	0.16
PRIMAL1	85	410	32	0.56	16	0.38	96	1.28
PRIMAL2	96	745	31	0.78	15	0.53	93	1.77
PRIMAL3	111	856	36	1.89	19	1.08	108	4.69
PRIMAL4	75	1564	31	1.77	15	1.19	93	3.14
PRIMALC1	9	239	6	0.03	4	0.03	18	0.06
PRIMALC2	7	238	6	0.03	3	0.01	18	0.06
QGROW15	300	645	5	0.47	4	0.38	15	0.65
QGROW22	440	946	6	0.83	5	0.66	18	1.44
QSCFXM3	990	1800	552	107.89	154	31.89	1656	317.89
STOKES1	1024	2178	88	28.41	47	16.95	264	51.47
STOKES2	2816	5890	138	381.97	122	372.56	414	1104.34
STOKES3	256	578	164	49.53	79	21.31	*	*
STOKES4	1024	2178	283	86.84	70	26.56	849	259.76
DD1	80	1600	26	0.38	15	0.27	78	1.02
DD2	110	3650	129	3.89	23	0.88	387	11.95
DD3	140	4900	26	2.51	15	1.88	78	6.39
DD4	160	6400	33	4.27	19	3.03	99	11.89

Cuadro 5.1: Comparación entre diferentes aceleraciones del método de Cimmino

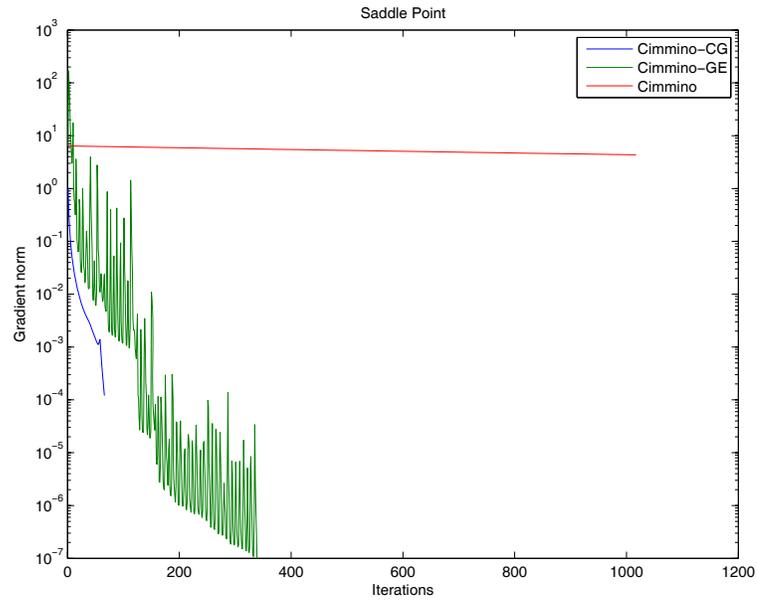


Figura 5.1: Aceleración por Cimmino para un problema CUTer

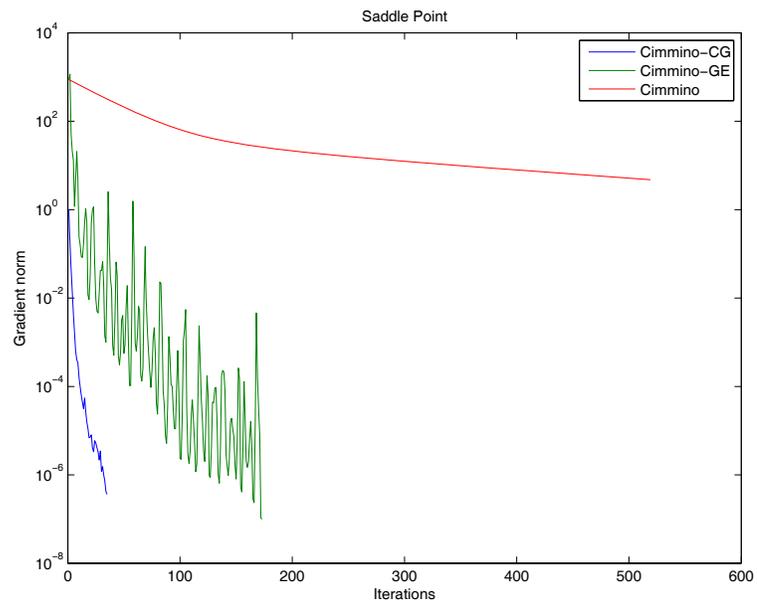


Figura 5.2: Aceleración por Cimmino para un problema de Stokes *Lid driven cavity problem*

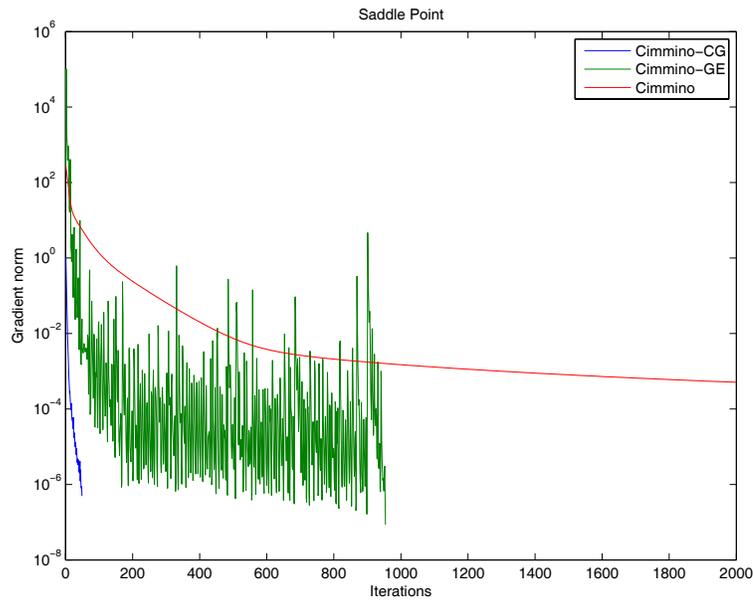


Figura 5.3: Aceleración por Cimmino para un problema de punto de ensilladura de descomposición de dominios

Problema			Cimmino CG			CG-AOP		CG-UZAWA		MINRES	
Nombre	$m$	$n$	iter	tiempo	tiempo/ $m$	iter	tiempo	iter	tiempo	iter	tiempo
AUG2DCQP	1600	3280	85	71.81	0.0449	2	0.53	87	0.64	202	0.17
CVXQP1S	50	100	33	0.16	0.0032	21	0.09	*	*	742	0.14
CVXQP2S	25	100	11	0.03	0.0012	12	0.03	*	*	279	0.06
CVXQP3S	75	100	85	0.30	0.0040	23	0.09	73	0.05	1798	0.17
DUALC1	215	223	2	0.03	0.0001	17	0.41	2	0.01	903	0.13
DUALC5	278	285	2	0.08	0.0003	15	0.48	1	0.01	782	0.11
DUALC8	503	510	2	0.52	0.0010	11	1.03	*	*	1664	0.33
GOULDQP2S	349	659	1	0.31	0.0009	2	0.03	*	*	4879	2.16
KSIP	1001	1021	1	0.80	0.0008	2	1.17	1	0.03	91	0.06
MOSARQP1	700	3200	1	0.48	0.0007	2	0.31	1	0.03	130	0.09
MOSARQP2	600	1500	1	0.29	0.0004	2	0.20	1	0.03	127	0.03
PRIMAL1	85	410	16	0.38	0.0045	39	4.61	16	0.25	1012	0.58
PRIMAL2	96	745	15	0.53	0.0055	30	4.38	15	0.98	672	0.67
PRIMAL3	111	856	19	1.08	0.0097	20	7.17	20	1.80	494	1.98
PRIMAL4	75	1564	15	1.19	0.0159	14	7.13	16	7.67	236	1.89
PRIMALC1	9	239	4	0.03	0.0033	9	0.09	3	0.06	42	0.06
PRIMALC2	7	238	3	0.01	0.0014	8	0.09	3	0.03	26	0.03
QGROW15	300	645	4	0.38	0.0013	14	1.67	5	0.27	139	0.09
QGROW22	440	946	5	0.66	0.0015	13	4.47	5	0.72	243	0.16
QSCFXM3	990	1800	154	31.89	0.0322	118	98.53	*	*	*	*
STOKES1	1024	2178	47	16.95	0.0166	16	5.17	48	1.11	1132	0.84
STOKES2	2816	5890	122	372.56	0.1323	31	21.45	223	15.77	2072	5.67
STOKES3	256	578	79	21.31	0.0832	20	6.23	78	1.92	1339	0.98
STOKES4	1024	2178	70	26.56	0.0259	20	6.01	69	1.56	1141	0.77
DD1	80	1600	15	0.27	0.0034	6	0.09	15	0.06	371	0.09
DD2	110	3650	23	0.88	0.0080	7	0.25	24	0.20	542	0.27
DD3	140	4900	15	1.88	0.0134	6	0.59	14	0.50	534	0.44
DD4	160	6400	19	3.03	0.0189	7	0.53	20	0.91	694	0.77

Cuadro 5.2: Comparación entre MINRES, Cimmino gradientes conjugados y gradientes conjugados aplicados al sistema de complemento de Schur

Problema			Cimmino CG			CG-AOP		CG-UZAWA		Pre-MINRES	
Nombre	$m$	$n$	iter	tiempo	tiempo/ $m$	iter	tiempo	iter	tiempo	iter	tiempo
STOKES1	1024	2178	47	16.95	0.0166	16	5.17	48	1.11	42	0.26
STOKES2	2816	5890	122	372.56	0.1323	31	21.45	223	15.77	61	1.21
STOKES3	256	578	79	21.31	0.0832	20	6.23	78	1.92	36	0.06
STOKES4	1024	2178	70	26.56	0.0259	20	6.01	69	1.56	36	0.23

Cuadro 5.3: Comparación entre MINRES preconditionado, gradientes conjugados Cimmino, y el método de los gradientes conjugados aplicado al sistema del complemento de Schur

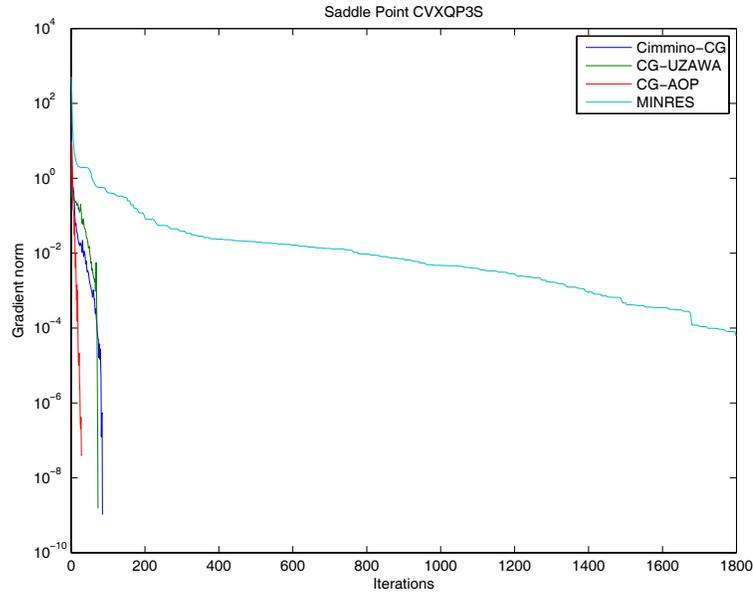


Figura 5.4: MINRES vs. CG-Cimmino vs. gradientes conjugados sobre el sistema condensado para un problema CUTer

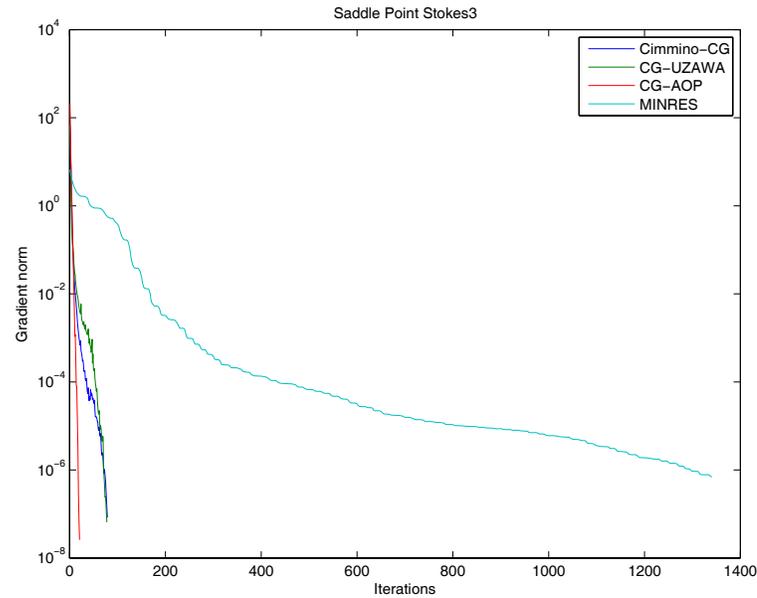


Figura 5.5: MINRES vs. CG-Cimmino vs. gradientes conjugados sobre el sistema condensado para un Stokes *Lid driven cavity problem*

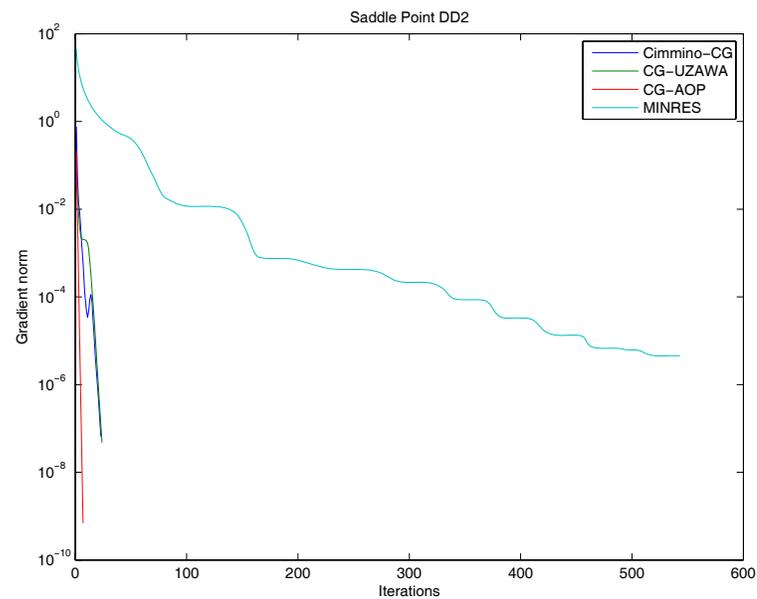


Figura 5.6: MINRES vs. CG-Cimmino vs. gradientes conjugados para un problema de punto de ensilladura de descomposición de dominios,  $m = 110$  y  $n = 3650$

# Capítulo 6

## Extensión de la aceleración del método de Cimmino para variedades lineales

### 6.1. Introducción

En este capítulo, extenderemos la aceleración de los métodos de Cimmino al caso de variedades lineales cerradas. En dicho caso, la proyección sobre variedades lineales es, en general, un operador no lineal. Por ello, para la aceleración del método de Cimmino utilizaremos las técnicas avanzadas para encontrar raíces de funciones no lineales propuestas en [25, 24]. Estas técnicas tienen la ventaja de no necesitar el cálculo de derivadas y en nuestra propuesta, convergen automáticamente al punto factible más cercano al vector inicial dado.

### 6.2. Aceleración para variedades lineales

Una variedad lineal es la traslación de un subespacio, i.e. cualquier variedad lineal  $V$  puede ser escrita como  $V = v_0 + S$  donde  $S$  es un subespacio. El subespacio  $S$  es único en su representación, pero cualquier vector en  $V$  puede jugar el rol de  $v_0$ .

Sea  $H$  un espacio de Hilbert y sean  $V_1, V_2, \dots, V_m$  variedades lineales cerradas de  $H$ ,  $V = \bigcap_{i=1}^m V_i$ ,  $V \neq \emptyset$  y  $x \in H$ . El problema a considerar es: Encontrar la proyección de un vector  $x$  en la variedad lineal  $V$ ,  $P_V(x)$ , i.e. el punto más cercano a  $x$  en  $V$ . Una técnica para resolver este problema es el método de Cimmino, que

consiste en la iteración:

$$x_{k+1} = \frac{1}{m} \sum_{i=1}^m P_{V_i}(x_k),$$

con  $x_0 = x$ . Sin embargo, el método de Cimmino es conocido por tener una velocidad de convergencia lenta.

Como mostramos previamente, para subespacios, una técnica para acelerar el método de Cimmino es aplicar directamente el método de gradientes conjugados al sistema lineal positivo semidefinido,

$$(I - T_S)x = 0,$$

donde  $T_S = \frac{1}{m} \sum_{i=1}^m P_{S_i}$ ,  $I - T_S = \frac{1}{m} \sum_{i=1}^m I - P_{S_i}$ , y  $S_i$  es un subespacio cerrado para  $i = 1, \dots, m$ . Esta versión acelerada converge hacia  $P_{\cap_{i=1}^m S_i} x_0$  donde  $x_0$  es el iterado inicial [51].

En cambio, en el caso de variedades lineales, dada la translación implicada, el operador  $I - T_V = \frac{1}{m} \sum_{i=1}^m x - P_{V_i}(x)$  no es necesariamente un operador lineal, ya que el operador de proyección sobre variedades lineales no es necesariamente un operador lineal. Por ende, proponemos para ello, resolver la ecuación no lineal  $F(x) = 0$ , donde  $F(x) = (I - T_V)x$ , por una técnica de elección para encontrar raíces de una función no lineal. La técnica a elegir será descrita luego. Primero, recordemos que  $V = \cap_{i=1}^m V_i$  es también una variedad lineal de  $H$ .

A continuación, probaremos algunas identidades que nos serán útiles para resultados posteriores.

**Lema 6.2.1** *Sea  $V = S + v_0$  una variedad lineal cerrada con  $S$  un subespacio cerrado. Sea  $P_V$  la proyección en  $V$ . Entonces*

1.  $P_V(x) = P_S(x) + P_{S^\perp}(v_0)$ , para cualquier  $x \in H$ .

2.  $x - P_V(x) = P_{S^\perp}(x - v_0)$ , para cualquier  $x \in H$ .

Donde  $P_{S^\perp} = I - P_S$  es la proyección en el complemento ortogonal de  $S$ .

**Demostración.** Recordemos que para cualquier conjunto  $C$ , y un vector  $y \in H$ , tenemos  $P_{C+y}(x) = P_C(x - y) + y$  (v.f. [33, 39]). Por lo tanto,

1.  $P_V(x) = P_S(x - v_0) + v_0 = P_S(x) - P_S(v_0) + v_0 = P_S(x) + (v_0 - P_S(v_0)) = P_S(x) + P_{S^\perp}(v_0)$ .

2.  $x - P_V(x) = x - v_0 - P_S(x - v_0) = P_{S^\perp}(x - v_0)$ . ■

Sea  $V_i$  una variedad lineal para  $i = 1, \dots, m$ . Ahora definimos  $F : H \mapsto H$  como sigue

$$F(x) = \frac{1}{m} \sum_{i=1}^m x - P_{V_i}(x). \quad (6.1)$$

Nuestro próximo resultado establece que las raíces de  $F(x)$  se corresponden con los puntos en  $V = \cap_{i=1}^m V_i$ .

**Lema 6.2.2**  $F(x) = 0$  si y solo si  $x \in \cap_{i=1}^m V_i$ .

**Demostración.**

Por el Lema 6.2.1,

$$F(x) = \frac{1}{m} \sum_{i=1}^m x - P_{V_i}(x) = \frac{1}{m} \sum_{i=1}^m P_{S_i^\perp}(x - v_0) = (x - v_0) - \frac{1}{m} \sum_{i=1}^m P_{S_i}(x - v_0),$$

con  $v_0 \in \cap_{i=1}^m V_i$ . Por lo tanto,

$$F(x) = 0 \text{ si y solo si } \frac{1}{m} \sum_{i=1}^m P_{S_i}(x - v_0) = x - v_0.$$

Hemos reducido el problema de variedades lineales a subespacios. Podemos observar que  $F(x) = 0$  si y solo si el método de Cimmino (en subespacios) tiene una solución en  $x - v_0$  i.e.

$$x - v_0 \in \cap_{i=1}^m S_i = S$$

pero, como  $v_0 \in V$  entonces  $x \in S + v_0 = V = \cap_{i=1}^m V_i$ . ■

Como consecuencia, las soluciones de  $F(x) = 0$  son vectores en  $V = \cap_{i=1}^m V_i$ . Nosotros no estamos interesados en cualquier vector en  $\cap_{i=1}^m V_i$  arbitrario, sino en el vector más cercano a  $x$  en  $\cap_{i=1}^m V_i$ . Esta tarea especial puede ser llevada a cabo si utilizamos el método apropiado para resolver  $F(x) = 0$ . En otras palabras, en general, los métodos estándar (métodos tipo Newton, por ejemplo) para resolver  $F(x) = 0$  no tienen necesariamente las propiedades de optimalidad requeridas. En contraste, cualquier método iterativo que use la dirección del residual,  $\pm F(x_k)$  como dirección de búsqueda, tiene automáticamente la propiedad de optimalidad deseada. Afortunadamente, los llamados métodos de residual de esta clase, con elementos prácticos adicionales, han sido desarrollados recientemente y serán descritos y analizados en la próxima subsección.

Antes de describir cual método vamos a elegir para resolver el sistema no lineal  $F(x) = 0$  dado por (6.1), necesitamos identificar la derivada de Fréchet,  $F'(x)$  de  $F$ , y algunas de sus propiedades. Para ello, vamos a tomar ventaja de la equivalencia de aplicar Cimmino en las variedades lineales,  $V_i$ 's, y el método de Cimmino en los subespacios  $S_i$ , suponiendo el conocimiento de un hipotético  $v_0 \in V = \cap_{i=1}^m V_i$ .

**Lema 6.2.3** *La ecuación  $F(x) = 0$  es equivalente a  $(I - T_S)x = b$  con  $b = (I - T_S)v_0$  y  $T_S = \frac{1}{m} \sum_{i=1}^m P_{S_i}$ , para cualquier  $v_0 \in V$ . Más aún, la derivada de Fréchet de  $F(x)$  está dada por  $F'(x) = (I - T_S)$ .*

**Demostración.**

$$\begin{aligned} F(x) &= \frac{1}{m} \sum_{i=1}^m x - P_{V_i}(x) = (x - v_0) - \frac{1}{m} \sum_{i=1}^m P_{S_i}(x - v_0) \\ &= \left( \frac{1}{m} \sum_{i=1}^m P_{S_i}(v_0) - v_0 \right) - \left( \frac{1}{m} \sum_{i=1}^m P_{S_i}(x) - x \right), \end{aligned}$$

entonces

$$F(x) = 0 \iff \left( x - \frac{1}{m} \sum_{i=1}^m P_{S_i}(x) \right) = \left( v_0 - \frac{1}{m} \sum_{i=1}^m P_{S_i}(v_0) \right),$$

y por lo tanto

$$F(x) = 0 \iff (I - T_S)x = b = (I - T_S)v_0.$$

Claramente, de esta equivalencia, la derivada de Fréchet de  $F(x)$  está dada por  $F'(x) = (I - T_S)$ . ■

**Corolario 6.2.1** *La derivada de Fréchet  $F'(x) = (I - T_S)$  es auto adjunta y positivo semidefinida.*

**Demostración.** El operador  $(I - T_S)$  es claramente auto adjunto ya que el operador proyección en subespacios es autoadjunto. Para establecer que es positivo semidefinido, consideremos los siguientes cálculos, para cualquier  $x \neq 0$ :

$$\left\langle \frac{1}{m} \sum_{i=1}^m (I - P_{S_i})x, x \right\rangle = \frac{1}{m} \sum_{i=1}^m \langle (I - P_{S_i})x, x \rangle,$$

pero  $(I - P_{S_i}) = P_{S_i^\perp}$  es también una proyección. Por lo tanto, por el resultado

anterior,

$$\langle (I - T_S)x, x \rangle = \frac{1}{m} \sum_{i=1}^m \langle P_{S_i^\perp} x, x \rangle = \frac{1}{m} \sum_{i=1}^m \|P_{S_i^\perp} x\|^2 \geq 0. \quad \blacksquare$$

Si conocemos un vector  $v_0 \in \cap_{i=1}^m V_i$  entonces el problema puede ser visto de nuevo como un problema en subespacios. En este caso,  $F(x) = 0$  se vuelve un sistema lineal positivo semidefinido. De hecho, el problema a ser resuelto sería  $(I - T_S)x = b$ , donde  $T_S = \frac{1}{m} \sum_{i=1}^m P_{S_i}(x)$  y  $b = (I - T_S)v_0$ , y pudiéramos usar la aceleración de gradientes conjugados como describimos antes.

Quisiéramos acotar que la derivada de Fréchet,  $F'(x) = (I - T_S)$ , de la función  $F(x)$  definido en (6.1) pudiera no estar disponible, dado que depende del operador  $T_S$  el cual pudiera no conocerse o ser costoso numéricamente dado que estamos trabajando con variedades lineales. Por lo tanto, el método de los gradientes conjugados pudiera no poder ser usado de manera práctica en presencia de variedades lineales. Sin embargo, las propiedades de  $F'(x)$  que analizaremos en nuestra próxima subsección garantizan la convergencia de un esquema especializado de residual, a ser descrito, para la resolución de sistemas no lineales de ecuaciones (i.e., en el caso de dimensión finita).

### 6.2.1. Métodos de residual para sistemas no lineales

Los métodos de residual que sistemáticamente usan la dirección del residual  $F(x_k)$ , o el negativo de la dirección del residual  $-F(x_k)$ , como dirección de búsqueda para resolver sistemas de ecuaciones grandes no lineales  $F(x) = 0$ , han sido desarrollados recientemente y analizados por La Cruz et al. [25, 24]. Como bien discutiremos más tarde en esta subsección, estos esquemas presentan ventajas en nuestro problema particular, ya que ellos convergen automáticamente al punto más cercano del punto inicial dado.

Las iteraciones de residual (DFSANE o SANE) propuestas y analizadas en [25, 24], para problemas no lineales finito-dimensionales, se encuentran definidas por

$$x_{k+1} = x_k \pm \alpha_k F(x_k), \quad (6.2)$$

donde  $\alpha_k > 0$  es el tamaño del paso y la dirección de búsqueda es  $F(x_k)$  o  $-F(x_k)$ ,

dependiendo de cuál es la dirección de descenso para la función de mérito,

$$f(x) = \|F(x)\|^2 = \langle F(x), F(x) \rangle. \quad (6.3)$$

Estas ideas son bastante efectivas y son esquemas competitivos para resolver sistemas no lineales en gran escala, cuando los tamaños de paso son elegidos de una manera conveniente. La convergencia de (6.2) se alcanza, para sistemas no lineales generales, cuando se asocia con las búsquedas lineales (*line search*) no monótonas y libre de derivadas, descritas ampliamente en [25]. Afortunadamente, como el Jacobiano de  $F(x)$  en (6.1), es simétrico y positivo semidefinido (Lema 6.2.3 y Corolario 6.2.1), habrá convergencia para el método puro, i.e., sin estrategias de globalización, y moviéndose a lo largo de la dirección  $-F(x_k)$ , la cual, en nuestro caso, siempre es una dirección de descenso de  $f(x)$ .

Para la elección del tamaño de paso  $\alpha_k > 0$ , hay muchas opciones que garantizan la convergencia. Una bien conocida es la elección no monótona espectral, que tiene propiedades interesantes, y que está definida como,

$$\alpha_k = \frac{\langle s_{k-1}, s_{k-1} \rangle}{\langle s_{k-1}, y_{k-1} \rangle}, \quad (6.4)$$

donde  $s_{k-1} = x_k - x_{k-1}$ , y  $y_{k-1} = F(x_k) - F(x_{k-1})$ . Mediante la aplicación del tamaño de paso (6.4), se requiere de una cantidad reducida de trabajo computacional, se acelera la convergencia del proceso, y se utilizan las dos últimas iteraciones, de tal manera que, incorpora información de primer orden en la dirección de búsqueda [7, 13, 64, 40].

---

**Algoritmo 8** DFSANE para  $F(x)$ , en su forma pura

---

- 1: Dado  $X_0 \in H, \alpha_0 \in \mathfrak{R}, \alpha_0 \neq 0$
  - 2: **for**  $k = 0, 1, \dots$  **do**
  - 3:      $x_{k+1} = x_k - \alpha_k F(x_k)$
  - 4:      $s_k = x_{k+1} - x_k$
  - 5:      $y_k = F(x_{k+1}) - F(x_k)$
  - 6:      $\alpha_{k+1} = \langle s_k, s_k \rangle / \langle s_k, y_k \rangle$
  - 7: **end for**
- 

**Observación 2.** El Algoritmo 8 produce una sucesión  $\{x_k\}$  que converge globalmente, i.e., para cualquier iterado inicial  $x_0$ , a un vector  $\bar{x}$  tal que  $F(\bar{x}) =$

0. Este resultado teórico sigue directamente de [41, Teorema 2.1], dado que el Jacobiano de  $F$  es simétrico y positivo semidefinido.

Para cerrar esta sección, estableceremos que el punto límite de la sucesión  $\{x_k\}$  generada por el Algoritmo 8 tiene la propiedad optimal de converger hacia el vector en  $\cap_{i=1}^m V_i$ , más próximo al vector inicial  $x_0$ . Primero, mostraremos dos resultados requeridos.

**Lema 6.2.4** *Sea  $V = \cap_{i=1}^m V_i$ ,  $v_0 \in V$ ,  $V_i = S_i + v_0$ ,  $V = S + v_0$ ,  $S = \cap_{i=1}^m S_i$ , entonces*

$$P_S F(x) = 0.$$

**Demostración.**

$$\begin{aligned} P_S F(x) &= P_S \frac{1}{m} \sum_{i=1}^m x - P_{V_i}(x) \\ &= \frac{1}{m} \sum_{i=1}^m P_S(x - v_0 - P_{S_i}(x - v_0)) \\ &= \frac{1}{m} \sum_{i=1}^m P_S(x - v_0) - P_S P_{S_i}(x - v_0), \end{aligned}$$

pero  $P_S P_{S_i} = P_S$  porque  $S \subseteq S_i, i = 1, \dots, m$ .

Por lo tanto,

$$P_S F(x) = \frac{1}{m} \sum_{i=1}^m P_S(x - v_0) - P_S(x - v_0) = 0.$$

■

**Lema 6.2.5** *La sucesión de iterados  $x_{k+1} = x_k - \alpha_k F(x_k)$ , con  $x_0 \in H$ , verifica,*

$$P_V(x_{k+1}) = P_V(x_k), \quad k = 0, \dots,$$

**Demostración.**

Por el resultado anterior y por el Lema 6.2.1,

$$\begin{aligned} P_V(x_{k+1}) &= P_S(x_{k+1}) + P_{S^\perp}(v_0) \\ &= P_S(x_k - \alpha_k F(x_k)) + P_{S^\perp}(v_0) \\ &= P_S(x_k) - \alpha_k P_S F(x_k) + P_{S^\perp}(v_0) \\ &= P_S(x_k) + P_{S^\perp}(v_0) \end{aligned}$$

$$= P_V(x_k).$$

■

Por inducción, todos los iterados de  $\{x_k\}$  tienen la misma proyección en la variedad lineal  $V$ . Como la iteración converge a un vector  $v \in V$ , tenemos,

$$v = P_V(v) = P_V(x_0).$$

Resumiremos esto en el teorema siguiente.

**Teorema 6.2.1 (Hernández-Ramos et al. [52])** *Las iteraciones  $x_{k+1} = x_k - \alpha_k F(x_k)$  generadas por el Algoritmo 8, con el iterado inicial  $x_0 \in H$ , convergen a  $P_V(x_0)$ .*

### 6.2.2. Experimentos con variedades lineales

Diversos experimentos fueron ejecutados en una Intel Core 2 Duo usando MATLAB 7.6. Dado el iterado inicial  $x_0 = (1, 1, \dots, 1)^T$  dichos experimentos numéricos consistieron en encontrar la proyección de  $x_0$  en la variedad lineal  $V = \cap_{i=1}^m H_i$ , donde los  $H_i$  son los hiperplanos definidos por:

$$H_i = \{x : \langle x, a_i \rangle = i\}, \quad i = 1, \dots, m.$$

Escogemos los vectores  $a_i$  de las primeras  $m$  filas de las matrices elegidas. Escogimos algunas matrices de Hilbert (rutina de MATLAB `hilb`) y algunas matrices de la colección Harwell-Boeing [37]. Comparamos el rendimiento de los siguientes métodos: método de Cimmino, método de Kaczmarz (método de proyecciones alternantes) y la aceleración propuesta usando DF-SANE.

### Resultados

En la Tabla 6.1 podemos observar que para los problemas de prueba, la aceleración DF-SANE del método de Cimmino lo acelera notablemente. Más aún, en todos los problemas de prueba muestra un mejor rendimiento que el método de Kaczmarz. De la misma manera que el método de Cimmino, la aceleración DF-SANE mantiene las ventajas sobre máquinas paralelas porque todas las proyecciones sobre las variedades lineales pueden ser hechas independientemente en paralelo (contrariamente al método de Kaczmarz). También observamos que a

medida que el número de variables lineales se incrementa, la mejora comparativa en el rendimiento es mayor.

Matriz			DF-Sane		Kaczmarz		Cimmino	
Name	Dimension	$m$	iter	tiempo	iter	tiempo	iter	tiempo
Hilbert	5	2	8	0.01	407	0.28	1576	1.03
Hilbert	5	3	32	0.04	*	*	*	*
Hilbert	5	4	264	0.25	*	*	*	*
Hilbert	5	5	*	*	*	*	*	*
Hilbert	10	2	8	0.01	304	0.17	1176	0.59
Hilbert	10	3	29	0.03	*	*	*	*
Hilbert	20	2	6	0.01	256	0.22	991	0.41
Hilbert	20	3	23	0.03	*	*	*	*
bcpwr02	49	20	37	0.26	37	0.31	1661	11.89
bcpwr02	49	30	64	0.68	128	1.47	7906	85.97
bcpwr02	49	40	337	4.59	1413	19.70	*	*
bcpwr02	49	49	3668	63.33	*	*	*	*
bcpwr03	118	10	49	0.21	63	0.22	1086	3.93
bcpwr03	118	20	66	0.51	101	0.70	4240	30.05
bcpwr03	118	30	103	1.13	185	2.06	*	*
bcpwr03	118	50	165	2.97	264	4.84	*	*
bcpwr03	118	100	585	20.83	4009	146.42	*	*
bcpwr03	118	118	3834	166.28	*	*	*	*
bcpwr04	274	10	16	0.10	12	0.19	253	1.11
bcpwr04	274	20	34	0.29	38	0.31	1652	12.12
bcpwr04	274	30	49	0.67	59	0.69	3103	34.93
bcpwr04	274	50	545	10.54	*	*	*	*
bcpwr04	274	274	*	*	*	*	*	*
mcca	180	10	29	0.16	60	0.30	1235	4.31
mcca	180	20	37	0.28	42	0.33	1674	11.53
mcca	180	30	35	0.39	40	0.45	2413	26.44
mcca	180	50	40	0.93	40	0.61	4794	87.45
mcca	180	100	298	11.98	1323	51.23	*	*
mcca	180	180	307	22.36	1101	81.18	*	*
mcfe	765	10	10	0.04	5	0.05	157	0.73
mcfe	765	20	137	1.13	565	4.63	*	*
mcfe	765	50	164	3.46	502	10.42	*	*
mcfe	765	100	152	6.64	412	17.71	*	*
mcfe	765	200	125	11.75	246	22.74	*	*
mcfe	765	500	383	114.23	*	*	*	*
mcfe	765	765	404	210.65	*	*	*	*
Sherman2	1080	10	634	2.83	3100	15.38	*	*
Sherman2	1080	20	410	3.53	2955	28.33	*	*
Sherman2	1080	30	3595	46.75	*	*	*	*
bcsstk01	48	10	39	0.20	41	0.12	966	3.67
bcsstk01	48	20	71	0.76	123	1.05	5686	41.59
bcsstk01	48	30	405	4.43	1137	12.82	*	*
bcsstk02	66	10	62	0.43	91	0.59	1641	7.71
bcsstk02	66	20	119	1.07	337	3.12	*	*
bcsstk02	66	30	231	3.24	1254	18.22	*	*
bcsstk02	66	50	520	13.62	4306	121.01	*	*
bcsstk03	112	10	1217	4.57	*	*	*	*
bcsstk03	112	20	*	*	*	*	*	*
bcsstk04	132	10	15	0.12	10	0.11	450	3.67
bcsstk04	132	20	13	0.08	8	0.11	206	1.11
bcsstk04	132	30	54	0.66	131	1.61	6470	81.78
bcsstk04	132	50	157	4.04	2018	44.20	*	*
bcsstk04	132	100	1088	48.82	*	*	*	*

Cuadro 6.1: Comparación entre los diferentes métodos para variedades lineales

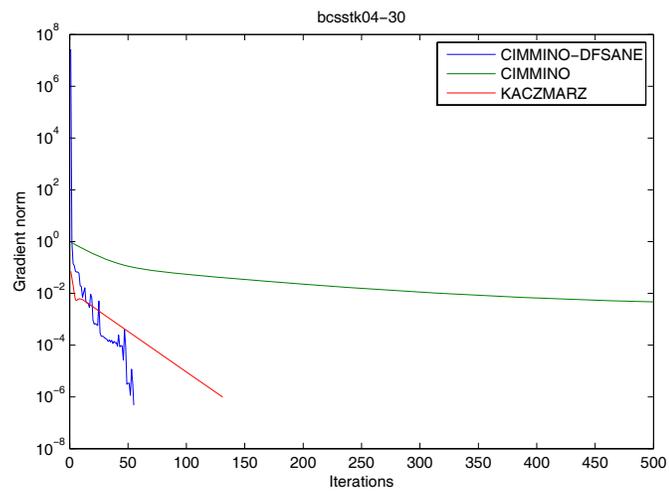


Figura 6.1: Aceleración para 30 filas de la matriz bcsstk04 (variedades lineales)

# Capítulo 7

## Aceleración para el problema de factibilidad convexa

### 7.1. Introducción

En este capítulo proponemos una aceleración del método de Cimmino para el problema de factibilidad convexa basada en métodos de residual tales como el DF-SANE [25, 24]. Realizaremos algunas experiencias numéricas para mostrar como esta propuesta funciona en algunos problemas académicos.

### 7.2. Problema de factibilidad convexa

Dados  $C_1, C_2, \dots, C_m$  conjuntos convexos cerrados en un espacio de Hilbert  $H$ . Llamamos  $C = \bigcap_{i=1}^m C_i$  y suponemos  $C \neq \emptyset$ . El problema de factibilidad convexa (CFP, por sus siglas en inglés) trata de encontrar un punto  $x \in C$ , llamado un punto factible.

Uno de los métodos utilizados para resolver el CFP es el método de Cimmino: Sea  $P_{C_i}$  el operador de proyección en el convexo  $C_i$ ,  $i = 1, \dots, m$ . El método de Cimmino viene dado por la iteración:

$$x_{k+1} = \frac{1}{m} \sum_{i=1}^m P_{C_i} x_k,$$

con  $x_0 \in H$ , iterado inicial dado. El método de Cimmino converge a un punto factible, es decir, a un punto  $x \in C$ . Como ya vimos en el capítulo anterior,

además, cuando los convexos  $C_i$  son variedades lineales o subespacios, para  $i = 1, \dots, m$ , el método de Cimmino converge además al punto  $P_C(x_0)$ , que es el punto  $c \in C$  más cercano a  $x_0$ , pero en el caso de conjuntos convexos en general esta última propiedad no es necesariamente cierta.

Definimos el operador  $T : H \rightarrow H$  asociado con el método de Cimmino,

$$T = \frac{1}{m} \sum_{i=1}^m P_{C_i}$$

y denotamos  $Fix T$  al conjunto de los puntos fijos del operador  $T$  (i.e.  $Fix T = \{x \mid Tx = x\}$ ). El método de Cimmino viene entonces dado por la iteración:

$$x_{k+1} = Tx_k$$

con  $x_0 \in H$  dado. La sucesión  $\{x_k\}$  converge a un punto  $x \in Fix T$ .

La proyección en un conjunto convexo  $K$  cualquiera  $P_K$  es un operador no expansivo:

$$\|P_K x - P_K y\| \leq \|x - y\|,$$

para todo  $x, y \in H$ . Por ende, el operador de proyección sobre un convexo es un operador Lipschitz, con constante igual a 1. Como resultado, el operador  $T = \frac{1}{m} \sum_{i=1}^m P_{C_i}$  es un operador no expansivo al ser una combinación convexa de operadores no expansivos. Es decir:

$$\|Tx - Ty\| \leq \|x - y\|,$$

para todo  $x, y \in H$ , y también el operador  $F = I - T$  es un operador Lipschitz, donde  $I$  es el operador identidad.

### 7.3. Aceleración para el CFP

Los puntos fijos de  $T$  son las soluciones de la ecuación no lineal:

$$F(x) = 0, \tag{7.1}$$

donde  $F = (I - T)$  es un operador  $F : H \rightarrow H$ , Lipschitz no necesariamente diferenciable.

La estrategia que proponemos para resolver el CFP es similar a la del capítulo anterior, es de utilizar un método numérico para resolver (7.1).

Tomemos el caso  $H = \mathbb{R}^n$ . Nuestra propuesta es la de utilizar estrategias para hallar ceros de funciones que no utilicen la derivada de  $F$ . En este caso, proponemos estrategias que utilicen como dirección de búsqueda el residual  $\pm F(x)$ , es decir, utilizar métodos tales como el DF-SANE [25, 24]. En particular, vamos a realizar experimentos numéricos con el DF-SANE para resolver algunos problemas de encontrar un punto factible en la intersección de convexos.

## 7.4. Experimentos numéricos

### 7.4.1. Problema de factibilidad lineal

Un caso importante es cuando los conjuntos  $C_i$  son semiespacios. Es decir, los conjuntos  $C_i$  se pueden escribir (sin pérdida de generalidad) como:

$$C_i = \{x \in \mathfrak{R}^n \mid \langle a_i, x \rangle \leq b_i\}. \quad (7.2)$$

con  $a_i \in \mathfrak{R}^n$  y  $b_i \in \mathfrak{R}$  dados para  $i = 1, \dots, m$ . A este problema se le suele llamar Problema de factibilidad lineal y se denota como:

$$Ax \leq b$$

donde  $A$  es una matriz de  $m \times n$  que tiene por filas los vectores  $a_i$ , y  $b$  es un vector de  $\mathfrak{R}^m$  cuyas componentes son los valores  $b_i$ .

En esta sección, mostramos los resultados numéricos de algunos experimentos que comparan la aceleración del método de Cimmino por DF-SANE versus la aceleración del método de Cimmino propuesta por Iusem y De Pierro [54] (ACCIMIDP). Son pocas las ideas tipo Cimmino implementables de las que tengo conocimiento en la literatura, y en especial esta idea de Iusem y De Pierro se entiende, es programable y posee virtudes en la forma de converger. Por estas razones es el método escogido para nuestras comparaciones.

Diversos experimentos fueron ejecutados en una Intel Core 2 Duo usando MATLAB 7.6. Dado el iterado inicial  $x_0 = (1, 1, \dots, 1)^T$ , dichos experimentos numéricos consistieron en encontrar una solución al problema  $Ax \leq b$ . Como vectores  $a_i$  se tomaron las primeras  $m$  filas de matrices  $n \times n$  dadas de la colección

Harwell-Boeing, y el valor  $b_i$  se tomó como  $b_i = i$ . El criterio de parada utilizado fue  $\|F(x)\| \leq 10^{-8}$ , y el máximo número de iteraciones para parar el algoritmo fue de 500 iteraciones (representado por un \* en la tabla en caso de que el método las sobrepase).

En la tabla 7.1 se muestran las comparaciones. La aceleración del método de Cimmino por DF-SANE funciona siempre, y en la mayoría de los casos necesitó menos iteraciones y menos tiempo para la convergencia que la aceleración propuesta por Iusem y De Pierro en [54]. También es importante hacer notar que en estos experimentos observamos que el método DF-SANE siempre funcionó en su forma pura, permitiendo conjeturar que debe existir teoría de convergencia con el método puro quedando esta como trabajo a futuro.

Matriz			DF-SANE		ACCCIMIDP	
Matriz	$n$	$m$	iter	tiempo	iter	tiempo
bcpwr02	49	30	4	0.03	5	0.02
bcpwr02	49	49	4	0.08	5	0.05
bcpwr03	118	30	5	0.03	5	0.03
bcpwr03	118	118	5	0.05	5	0.06
bcpwr04	274	30	4	0.02	51	0.16
bcpwr04	274	274	4	0.14	51	1.22
bcpwr05	443	30	4	0.02	23	0.06
bcpwr05	443	443	4	0.20	23	0.97
mcca	180	10	18	0.03	263	0.30
mcca	180	20	14	0.05	497	1.16
mcca	180	30	17	0.06	*	*
mcca	180	180	14	0.31	*	*
mcfe	765	10	3	0.01	3	0.03
mcfe	765	20	23	0.08	*	*
mcfe	765	30	26	0.11	*	*
Sherman2	1080	10	7	0.01	118	0.19
Sherman2	1080	20	7	0.02	176	0.48
Sherman2	1080	30	7	0.03	237	0.97
bcsstk01	48	10	7	0.03	87	0.09
bcsstk01	48	20	12	0.05	*	*
bcsstk01	48	30	38	0.11	*	*
bcsstk02	66	10	12	0.02	257	0.36
bcsstk02	66	20	17	0.05	*	*
bcsstk02	66	30	76	0.36	*	*
bcsstk03	112	10	178	0.19	*	*
bcsstk04	132	10	6	0.01	134	0.20
bcsstk04	132	20	7	0.05	171	0.42
bcsstk04	132	30	18	0.08	*	*
bcsstk04	132	50	157	4.04	2018	44.20

Cuadro 7.1: Comparación entre los diferentes métodos para el problema de factibilidad lineal

### 7.4.2. Problemas de factibilidad lineal con restricciones de caja

En esta sección realizaremos los mismos experimentos de resolución de problemas,

$$Ax \leq b,$$

pero ahora sujetos a la restricción de que la solución debe pertenecer también a una caja cerrada dada, proyectando directamente sobre la caja.

Una caja cerrada en  $\mathfrak{R}^n$  es un conjunto convexo que se define de la siguiente manera:

**Definición 7.4.1 (Caja cerrada)** Sean dos vectores  $l, u \in \mathfrak{R}^n$  con la propiedad de que  $l_i \leq u_i$  para  $i = 1, \dots, n$ . La caja cerrada de  $\mathfrak{R}^n$  definida por los vectores  $l, u$  es el conjunto:

$$B_{l,u} = \{x \in \mathfrak{R}^n \mid l_i \leq x_i \leq u_i, i = 1, \dots, n\}.$$

En este sentido, experimentaremos igualmente con los métodos de Cimmino acelerado por el DF-SANE que proponemos aquí, y con la aceleración propuesta por Iusem y De Pierro en [54], pero ahora, al incorporar una restricción más (la restricción de caja), el problema se transforma a hallar un punto en la intersección de  $m + 1$  convexos, en donde  $m$  es el número de filas de la matriz  $A$ . Nuestro problema se transforma entonces en:

Encontrar  $x \in \mathfrak{R}^n$  tal que:

$$Ax \leq b \quad \text{sujeto a: } x \in B_{l,u}, \quad (7.3)$$

donde  $A \in \mathfrak{R}^{m \times n}$ ,  $b \in \mathfrak{R}^m$ ,  $l, u \in \mathfrak{R}^n$ .

Las aceleraciones del método de Cimmino requieren proyectar los iterados en cajas. A continuación veremos como se realizan dichas proyecciones.

#### Proyección de un punto en una caja:

Para proyectar un punto cualquiera  $x \in \mathfrak{R}^n$  en la caja cerrada  $B_{l,u}$  se procede como sigue:

Para todo  $i = 1, \dots, n$ :

- Si  $x_i \leq l_i$  entonces  $p_i = l_i$ ;
- Si  $x_i \geq u_i$  entonces  $p_i = u_i$ ;

- Si  $l_i \leq x_i$  y  $x_i \leq u_i$  entonces  $p_i = x_i$ .

El vector resultante  $p$  viene siendo la proyección de  $x$  en la caja.

Nuestros experimentos son similares a los de la sección anterior, adicionando la restricción de que la solución  $x$  debe pertenecer a la caja  $B_{l,u}$ .

En la Tablas 7.2, 7.3, 7.4 se presentan los resultados de problemas de factibilidad lineal sujetos a restricciones de caja, con cajas de diferentes tamaños. En todos los experimentos se observó que la propuesta del método de Cimmino acelerado por el DF-SANE realiza menos iteraciones y requiere menor tiempo que la aceleración de Iusem y De Pierro [54]. También se observa que ambos métodos aumentan su trabajo computacional requerido a medida que la caja que define la restricción adicional es más pequeña.

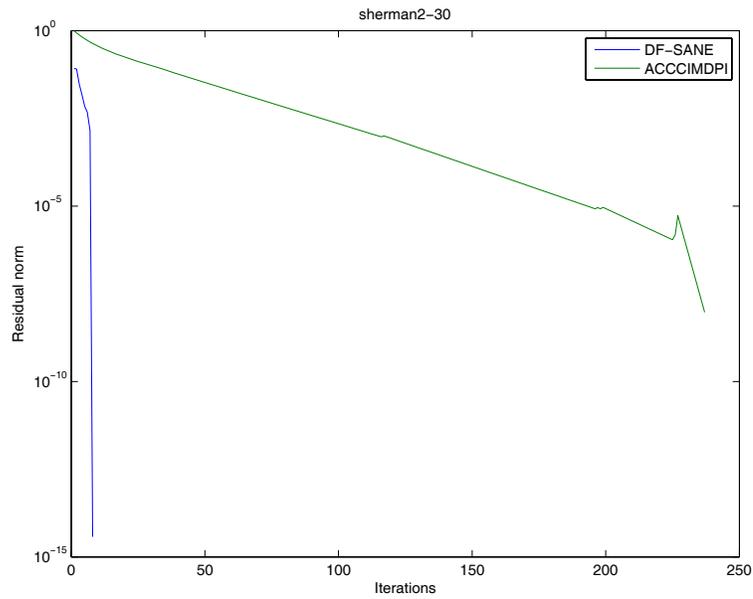


Figura 7.1: Aceleración para las 30 primeras filas de la matriz Sherman2 (Problema de Factibilidad Lineal)

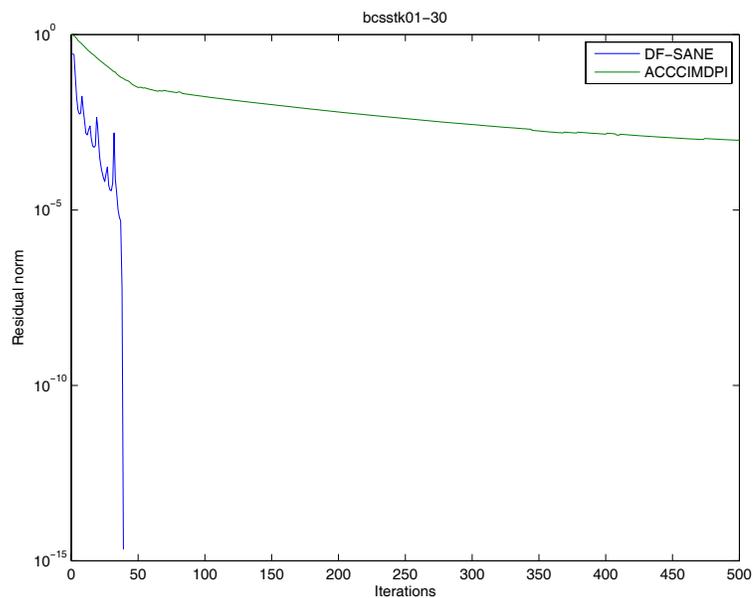


Figura 7.2: Aceleración para las 30 primeras filas de la matriz bcstkt01 (Problema de Factibilidad Lineal) con restricciones de caja  $l = -10 * (1, \dots, 1)^T$  y  $u = 10 * (1, \dots, 1)^T$

Matriz			DF-SANE		ACCCIMIDP	
Matriz	$n$	$m$	iter	tiempo	iter	tiempo
bcpwr02	49	30	4	0.02	5	0.01
bcpwr02	49	49	4	0.02	5	0.02
bcpwr03	118	30	5	0.02	5	0.05
bcpwr03	118	118	5	0.06	5	0.08
bcpwr04	274	30	4	0.02	51	0.19
bcpwr04	274	274	4	0.13	51	1.38
bcpwr05	443	30	4	0.05	23	0.08
bcpwr05	443	443	4	0.20	23	1.06
mcca	180	10	*	*	*	*
mcca	180	20	*	*	*	*
mcca	180	30	*	*	*	*
mcca	180	180	*	*	*	*
mcfe	765	10	4	0.03	*	*
mcfe	765	20	*	0.08	*	*
mcfe	765	30	*	0.11	*	*
Sherman2	1080	10	10	0.05	169	0.53
Sherman2	1080	20	10	0.06	236	1.05
Sherman2	1080	30	10	0.08	324	1.95
bcsstk01	48	10	179	0.25	*	*
bcsstk01	48	20	*	*	*	*
bcsstk01	48	30	*	*	*	*
bcsstk02	66	10	20	0.05	455	0.77
bcsstk02	66	20	36	0.13	*	*
bcsstk02	66	30	111	0.65	*	*
bcsstk03	112	10	464	0.59	*	*
bcsstk04	132	10	*	*	*	*
bcsstk04	132	20	*	*	*	*
bcsstk04	132	30	*	*	*	*
bcsstk04	132	50	*	*	*	*

Cuadro 7.2: Comparación entre los diferentes métodos para el problema de factibilidad lineal sujeto a restricciones de caja  $l = (0, \dots, 0)^T$  y  $u = (1, \dots, 1)^T$

Matriz			DF-SANE		ACCCIMIDP	
Matriz	$n$	$m$	iter	tiempo	iter	tiempo
bcpwr02	49	30	4	0.01	5	0.03
bcpwr02	49	49	4	0.03	5	0.05
bcpwr03	118	30	5	0.03	5	0.01
bcpwr03	118	118	5	0.06	5	0.08
bcpwr04	274	30	2	0.01	51	0.09
bcpwr04	274	274	4	0.14	51	1.20
bcpwr05	443	30	4	0.01	23	0.06
bcpwr05	443	443	4	0.20	23	0.95
mcca	180	10	18	0.02	263	0.34
mcca	180	20	19	0.06	497	1.22
mcca	180	30	18	0.08	*	*
mcca	180	180	*	*	*	*
mcfe	765	10	4	0.03	*	*
mcfe	765	20	*	0.08	*	*
mcfe	765	30	*	0.11	*	*
Sherman2	1080	10	7	0.05	118	0.31
Sherman2	1080	20	7	0.05	176	0.70
Sherman2	1080	30	7	0.05	237	1.22
bcsstk01	48	10	7	0.03	87	0.11
bcsstk01	48	20	12	0.05	*	*
bcsstk01	48	30	37	0.13	*	*
bcsstk02	66	10	12	0.05	257	0.38
bcsstk02	66	20	17	0.06	*	*
bcsstk02	66	30	107	0.50	*	*
bcsstk03	112	10	210	0.25	*	*
bcsstk04	132	10	*	*	*	*
bcsstk04	132	20	*	*	*	*
bcsstk04	132	30	*	*	*	*
bcsstk04	132	50	*	*	*	*

Cuadro 7.3: Comparación entre los diferentes métodos para el problema de factibilidad lineal sujeto a restricciones de caja  $l = -10 * (1, \dots, 1)^T$  y  $u = 10 * (1, \dots, 1)^T$

Matriz			DF-SANE		ACCCIMIDP	
Matriz	$n$	$m$	iter	tiempo	iter	tiempo
bcpwr02	49	30	4	0.01	5	0.01
bcpwr02	49	49	4	0.02	5	0.02
bcpwr03	118	30	5	0.03	5	0.03
bcpwr03	118	118	5	0.06	5	0.08
bcpwr04	274	30	4	0.01	51	0.16
bcpwr04	274	274	4	0.13	51	1.23
bcpwr05	443	30	4	0.03	23	0.06
bcpwr05	443	443	4	0.20	23	0.95
mcca	180	10	18	0.04	263	0.34
mcca	180	20	19	0.06	497	1.25
mcca	180	30	18	0.08	*	*
mcfe	765	10	4	0.03	*	*
mcfe	765	20	*	0.08	*	*
mcfe	765	30	*	0.11	*	*
Sherman2	1080	10	7	0.05	118	0.31
Sherman2	1080	20	7	0.06	176	0.69
Sherman2	1080	30	7	0.06	237	1.27
bcsstk01	48	10	7	0.03	87	0.14
bcsstk01	48	20	12	0.03	*	*
bcsstk01	48	30	37	0.13	*	*
bcsstk02	66	10	12	0.05	257	0.41
bcsstk02	66	20	17	0.06	*	*
bcsstk02	66	30	79	0.39	*	*
bcsstk03	112	10	200	0.30	*	*
bcsstk04	132	10	6	0.03	134	0.23
bcsstk04	132	20	7	0.02	171	0.52
bcsstk04	132	30	19	0.09	*	*
bcsstk04	132	50	*	*	*	*

Cuadro 7.4: Comparación entre los diferentes métodos para el problema de factibilidad lineal sujeto a restricciones de caja  $l = -100 * (1, \dots, 1)^T$  y  $u = 100 * (1, \dots, 1)^T$

### 7.4.3. Problemas de factibilidad lineal con restricciones de esferas

En esta sección realizaremos los mismos experimentos de resolución de problemas:

$$Ax \leq b,$$

pero ahora sujetos a la restricción de que la solución debe pertenecer también a una esfera dada, proyectando directamente sobre la esfera.

Una esfera en  $\mathfrak{R}^n$  es un conjunto convexo que se define de la siguiente manera:

**Definición 7.4.2 (Esfera)** *Dado un vector  $c \in \mathfrak{R}^n$  y dado un valor  $r \in \mathfrak{R}$ . Se define la esfera con centro en  $c$  y de radio  $r$  como el conjunto:*

$$E(c, r) = \{x \in \mathfrak{R}^n \mid \|x - c\| \leq r\}.$$

#### Proyección de un punto en una esfera:

Para proyectar un punto cualquiera  $x \in \mathfrak{R}^n$  en la esfera  $E(c, r)$ , primero se calcula la distancia de  $x$  al centro  $c$  de la esfera. Es decir,  $d = \|x - c\|$ .

- Si la distancia  $d$  es menor al radio de la esfera, es decir, si  $d \leq r$ , significa que el punto  $x$  está dentro de la esfera y por lo tanto la proyección es el mismo punto  $x$ , i.e.  $p = x$ .
- En caso contrario, si el punto  $x$  está fuera de la esfera se calcula la proyección  $p$  de la siguiente manera:

$$p = c + \frac{r}{d}(x - c).$$

Nuestros experimentos son similares a los de la sección anterior, adicionando la restricción de que la solución  $x$  debe pertenecer a la esfera  $E_{c,r}$ .

Matriz			DF-SANE		ACCCIMIDP	
Matriz	n	m	iter	tiempo	iter	tiempo
bcpwr02	49	30	4	0.01	3	0.02
bcpwr02	49	49	4	0.02	3	0.03
bcpwr03	118	30	4	0.03	3	0.03
bcpwr03	118	118	4	0.05	3	0.03
bcpwr04	274	30	4	0.03	3	0.03
bcpwr04	274	274	4	0.13	3	0.11
bcpwr05	443	30	4	0.05	3	0.05
bcpwr05	443	443	4	0.19	3	0.16
mcca	180	10	6	0.03	67	0.08
mcca	180	20	4	0.03	68	0.16
mcca	180	30	6	0.03	68	0.20
mcfe	765	10	3	0.03	3	0.03
mcfe	765	20	8	0.05	6	0.05
mcfe	765	30	8	0.06	6	0.05
Sherman2	1080	10	20	0.03	4	0.03
Sherman2	1080	20	10	0.05	4	0.03
Sherman2	1080	30	26	0.13	4	0.05
bcsstk01	48	10	5	0.02	44	0.06
bcsstk01	48	20	29	0.06	104	0.22
bcsstk01	48	30	62	0.19	122	0.36
bcsstk02	66	10	6	0.03	3	0.03
bcsstk02	66	20	7	0.05	4	0.02
bcsstk02	66	30	50	0.23	4	0.05
bcsstk03	112	10	88	0.11	*	*
bcsstk04	132	10	5	0.02	3	0.03
bcsstk04	132	20	5	0.02	17	0.06
bcsstk04	132	30	5	0.03	38	0.14
bcsstk04	132	132	104	1.67	146	2.31

Cuadro 7.5: Comparación entre los diferentes métodos para el problema de factibilidad lineal sujeto a restricciones de esfera con centro en  $c = (0, \dots, 0)^T$  y radio  $r = 1$

Matriz			DF-SANE		ACCCIMIDP	
Matriz	n	m	iter	tiempo	iter	tiempo
bcpwr02	49	30	4	0.03	2	0.01
bcpwr02	49	49	4	0.03	2	0.03
bcpwr03	118	30	5	0.02	2	0.01
bcpwr03	118	118	5	0.06	2	0.05
bcpwr04	274	30	4	0.03	2	0.03
bcpwr04	274	274	4	0.13	2	0.09
bcpwr05	443	30	4	0.04	4	0.03
bcpwr05	443	443	4	0.19	2	0.13
mcca	180	10	18	0.04	79	0.08
mcca	180	20	19	0.05	79	0.19
mcca	180	30	18	0.06	80	0.23
mcfe	765	10	3	0.03	2	0.03
mcfe	765	20	24	0.08	369	0.89
mcfe	765	30	23	0.11	376	1.28
Sherman2	1080	10	7	0.03	33	0.08
Sherman2	1080	20	7	0.03	33	0.11
Sherman2	1080	30	7	0.06	34	0.19
bcsstk01	48	10	7	0.01	2	0.03
bcsstk01	48	20	12	0.04	116	0.23
bcsstk01	48	30	37	0.13	137	0.41
bcsstk02	66	10	12	0.03	5	0.01
bcsstk02	66	20	17	0.08	7	0.05
bcsstk02	66	30	107	0.50	295	1.38
bcsstk03	112	10	210	0.23	*	*
bcsstk04	132	10	6	0.03	21	0.05
bcsstk04	132	20	7	0.04	20	0.08
bcsstk04	132	30	19	0.06	47	0.17
bcsstk04	132	132	*	*	*	*

Cuadro 7.6: Comparación entre los diferentes métodos para el problema de factibilidad lineal sujeto a restricciones de esfera con centro en  $c = (0, \dots, 0)^T$  y radio  $r = 10^2$

Matriz			DF-SANE		ACCCIMIDP	
Matriz	n	m	iter	tiempo	iter	tiempo
bcpwr02	49	30	4	0.03	2	0.01
bcpwr02	49	49	4	0.03	2	0.03
bcpwr03	118	30	5	0.02	2	0.01
bcpwr03	118	118	5	0.06	2	0.05
bcpwr04	274	30	4	0.03	2	0.03
bcpwr04	274	274	4	0.13	2	0.09
bcpwr05	443	30	4	0.04	4	0.03
bcpwr05	443	443	4	0.19	2	0.13
mcca	180	10	18	0.04	79	0.08
mcca	180	20	19	0.05	79	0.19
mcca	180	30	18	0.06	80	0.23
mcfe	765	10	3	0.03	2	0.03
mcfe	765	20	24	0.08	369	0.89
mcfe	765	30	23	0.11	376	1.28
Sherman2	1080	10	7	0.03	33	0.08
Sherman2	1080	20	7	0.03	33	0.11
Sherman2	1080	30	7	0.06	34	0.19
bcsstk01	48	10	7	0.01	2	0.03
bcsstk01	48	20	12	0.04	116	0.23
bcsstk01	48	30	37	0.13	137	0.41
bcsstk02	66	10	12	0.03	5	0.01
bcsstk02	66	20	17	0.08	7	0.05
bcsstk02	66	30	107	0.50	295	1.38
bcsstk03	112	10	210	0.23	*	*
bcsstk04	132	10	6	0.03	21	0.05
bcsstk04	132	20	7	0.04	20	0.08
bcsstk04	132	30	19	0.06	47	0.17
bcsstk04	132	132	*	*	*	*

Cuadro 7.7: Comparación entre los diferentes métodos para el problema de factibilidad lineal sujeto a restricciones de esfera con centro en  $c = (0, \dots, 0)^T$  y radio  $r = 10^8$

#### 7.4.4. Intersectando cajas y esferas

La siguiente sección, tiene como objetivo mostrar algunos experimentos en donde se compare la aceleración propuesta (DF-SANE) con la aceleración propuesta por Iusem y De Pierro [54]. En este caso la comparación se realizará buscando un punto factible en la intersección de una caja y dos esferas en  $\mathfrak{R}^n$ . En esta serie de experimentos jugaremos con las dimensiones de la caja y con los radios de las esferas para observar el comportamiento de los distintos métodos.

En cada experimento habrá una caja en  $\mathfrak{R}^n$ ,  $B_{l,u}$  donde:  $l = a_1 * (1, 1, \dots, 1, 1)^T$  y  $u = a_2 * (1, 1, \dots, 1, 1)^T$ , con  $a_1, a_2 \in \mathfrak{R}$  a definir. Igualmente, en cada experimento habrán 2 esferas en  $\mathfrak{R}^n$ :  $E(c_1, r_1)$  y  $E(c_2, r_2)$ , donde  $c_1 = (1, 1, \dots, 1, 1)^T$ ,  $c_2 = (0, 0, \dots, 0, 0)^T$  y  $r_1, r_2 \in \mathfrak{R}$  a definir. Los experimentos consistirán en hallar un punto en la intersección  $B_{l,u} \cap E(c_1, r_1) \cap E(c_2, r_2)$  para diferentes valores de los parámetros  $a_1, a_2, r_1, r_2 \in \mathfrak{R}$  y para  $n \in \mathbb{N}$ . El iterado inicial que se tomó siempre fue  $x_0 = (1, 2, 3, \dots, n)^T$ .

#### Resultados

La Tabla 7.8 muestra diferentes resultados de los experimentos antes planteados. En esta tabla *iter* indica el número de iteraciones que realizó el método. También se indica la suma de las normas de las distancias del iterado final hacia cada subespacio, esto lo indicamos por *sdist*, de tal manera que si el valor *sdist* = 0 indica que el método consiguió una solución factible. En los experimentos realizados, cuando existían soluciones factibles, la aceleración propuesta en este trabajo fue capaz de encontrar alguna. Y además, en los casos en los que no existían soluciones factibles, la aceleración por DF-SANE consiguió una solución con el mismo valor *sdist* que la del iterado final de la aceleración propuesta por Iusem y De Pierro, por lo que podemos conjeturar que nuestro método tiene la propiedad de conseguir siempre soluciones en donde la suma de las distancias a cada convexo sea mínima.

Las iteraciones del método de Iusem y De Pierro son ligeramente más costosas que las de la aceleración propuesta en este trabajo. En los experimentos realizados el tiempo total de CPU fue muy pequeño, numéricamente cero en todos los casos, y por esa razón no los reportamos en la tabla. Vale la pena en un futuro estudiar el comportamiento del tiempo de CPU en aplicaciones en donde el costo de proyectar sea mucho mayor. En cuanto al número de iteraciones se observó que en problemas en donde había una región factible muy grande, el comportamiento de ambos

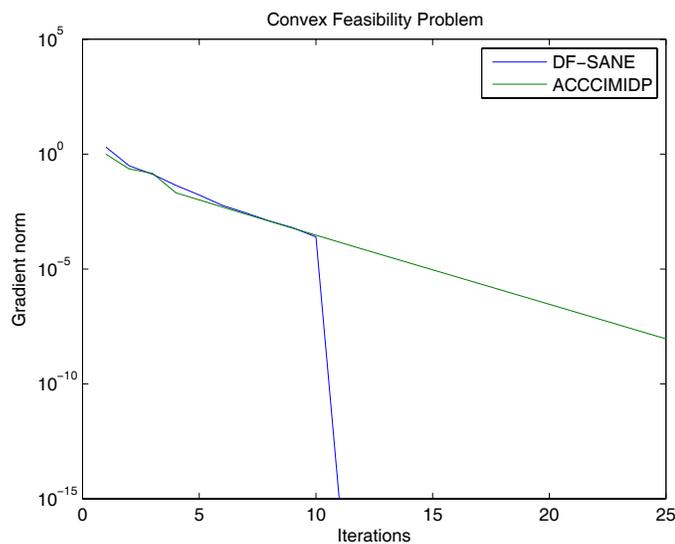


Figura 7.3: Aceleración para Intersección de una caja y dos esferas

n	Caja		Esferas		DF-SANE		ACCCIMIDP	
$n$	$a_1$	$a_2$	$r_1$	$r_2$	iter	$sdist$	iter	$sdist$
2	0	100	1	0.5	2	0	2	0
2	0	0.5	1	0.5	4	0	3	0
2	0	0.1	1	0.5	9	0.27	20	0.27
3	0	100	1	0.5	13	0.23	*	*
3	0	10	1	1	3	0	3	0
3	0.5	1.0	1	1	10	0	25	0
3	0.5	0.8	1	1	8	0	24	0
3	0.7	1.0	1	1	10	0.21	24	0.21
3	0.6	1.0	1	1	14	0.04	31	0.04
3	0.5	1.0	0.5	1	14	0.23	90	0.23
3	0.5	1.0	0.5	0.7	13	0.53	43	0.53
3	0.5	1.0	0.7	0.7	22	0.33	60	0.33
3	0.5	1.0	0.75	0.75	26	0.23	73	0.23
3	0.5	1.0	0.8	0.8	27	0.13	96	0.13
3	0.5	1.0	0.88	0.85	15	0.02	58	0.02
3	0.5	1.0	0.9	0.85	17	0.02	*	*
3	0.5	1.0	0.9	0.9	14	0	30	0
3	0.5	1.0	1.0	1.0	10	0	25	0
3	0.5	1.0	0.95	0.95	10	0	28	0
3	0	10	1	1	3	0	3	0

Cuadro 7.8: Comparación entre los diferentes métodos en la intersección de una caja y dos esferas

métodos fue muy similar. En cambio, en problemas en donde la región factible era más pequeña, resultaba favorecido en cuanto a número de iteraciones el método DF-SANE. También en los casos en donde la región factible era vacía resultaba favorecido el método DF-SANE, pudiendo remarcar algunos casos en donde el comportamiento del método propuesto por Iusem y De Pierro alcanzó el número máximo de iteraciones.

# Capítulo 8

## Conclusiones

En este trabajo, hemos visto como la optimización numérica ha provisto un nuevo marco teórico para el análisis de los métodos de proyecciones alternantes y simultáneas. Algunos de los métodos de proyecciones se pudieron interpretar como esquemas iterativos para resolver problemas de optimización ya existentes, y esto ha permitido explicar de otro modo algunas de sus características. Igualmente, se han podido crear versiones aceleradas de los métodos de proyecciones alternantes y simultáneas, basadas en técnicas más eficientes y novedosas de la optimización numérica [52, 51]. Esto trae nuevos resultados y nuevas perspectivas para afrontar algunos de los problemas que continúan abiertos en el área.

Los métodos de proyecciones alternantes fueron estudiados vía optimización numérica en diferentes escenarios que son: el problema de mejor aproximación, tanto en subespacios como en variedades lineales; y el problema de factibilidad convexa.

Primeramente, para el caso del problema de mejor aproximación en subespacios, se estudiaron los métodos de Von Neumann-Halperin y de Cimmino como esquemas clásicos de iteración de punto fijo (iteraciones de Picard), para operadores lineales en espacios de Hilbert. Luego, este problema fue transformado en un problema de ecuaciones en operadores lineales y singulares. Esta nueva visión permitió responder a interrogantes sobre la poca efectividad de algunas aceleraciones de los métodos de proyecciones propuestas anteriormente (e.g., [10]), dando luces del por qué las aceleraciones de métodos con operadores simétricos o simetrizados (e.g. Cimmino) resultan ser más efectivas. Luego, se demostró bajo que condiciones este problema podía ser resuelto por el método de los gradientes conjugados, tanto en espacios de dimensión finita como infinita. Posteriormente,

se realizaron experimentos numéricos comparativos que validaron los resultados obtenidos teóricamente.

También, en el mismo caso del problema de mejor aproximación en subespacios, se estudió el método de Cimmino y algunas variantes aceleradas propuestas, vistos como la aplicación de técnicas numéricas de minimización existentes para una cierta función cuadrática dada. Este estudio permitió dar un interesante resultado teórico que relaciona la velocidad de convergencia de las diferentes variantes del método de Cimmino y los ángulos de los subespacios involucrados. Se hicieron experimentos numéricos para comparar las diferentes técnicas de aceleración del método de Cimmino en la resolución de problemas de punto de ensilladura (*saddle point problem*), y luego se hicieron comparaciones con otros métodos recientes de resolución de este tipo de problemas, para dar idea del potencial que tiene la aceleración por gradientes conjugados del método de Cimmino en ambientes paralelos.

Posteriormente, se estudió el caso del problema de mejor aproximación en variedades lineales. Para este problema se logró probar primero que, las soluciones factibles eran las raíces de una cierta función  $F$  no lineal, de Jacobiano simétrico y positivo semidefinido. Además, se probó que la utilización de métodos de tipo residual, tales como el DF-SANE, logran no solo encontrar soluciones factibles, sino que resuelven el problema de mejor aproximación, si se toma como iterado inicial al punto a proyectar en la intersección de las variedades lineales. Igualmente, al ser la función  $F$  de Jacobiano positivo semidefinido y simétrico, el método DF-SANE u otros métodos de tipo residual, tienen características óptimas en la resolución de este problema, ya que no requiere de estrategias de globalización. Se hicieron pruebas numéricas hallando el punto más cercano en la intersección de hiperplanos, y comparando contra el método clásico de Cimmino y contra el método de Kaczmarz obteniéndose resultados numéricos alentadores.

Finalmente, se trató el problema de factibilidad convexa en general, proponiendo una aceleración del método de Cimmino por el DF-SANE, que se conjetura que pueda ser utilizada para resolver este problema. Se hicieron experimentos numéricos para el problema de factibilidad lineal, para el problema de factibilidad lineal sujeto a restricciones de caja y de esferas, y de hallar puntos factibles en la intersección de cajas y esferas. En todos estos casos, la aceleración propuesta logró conseguir puntos factibles, si los había, y numéricamente tuvo un mejor desempeño que otros métodos existentes en la literatura.

Como temas futuros a investigar, luego de este trabajo, queda ver si es posible probar la conjetura hecha para acelerar el método de Cimmino en el problema de factibilidad convexa. También, como tema de investigación a futuro está ver si por medio de la optimización numérica pueden acelerarse los métodos existentes para el problema de mejor aproximación en conjuntos convexos, en particular, acelerar el algoritmo de Dykstra [15, 33]. También queda pendiente un estudio de la aceleración por gradientes conjugados del método de Cimmino, para confirmar el potencial que pareciera tener en máquinas paralelas en la resolución de problemas de punto de ensilladura. Finalmente, como trabajo a futuro, queda también el desarrollo de preconditionadores para los métodos de proyecciones y sus nuevas variantes aceleradas.

# Bibliografía

- [1] G. APPLEBY and D. SMOLARSKI. A linear acceleration row action method for projecting onto subspaces. *Electronic Transactions on Numerical Analysis (ETNA)*, 20:253–275, 2005.
- [2] M. ARIOLI, I. DUFF, J. NOAILLES, and D. RUIZ. A block projection method for sparse matrices. *SIAM Journal on Scientific and Statistical Computing*, 13:47, 1992.
- [3] N. ARONSZAJN. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.
- [4] K. ATKINSON and W. HAN. *Theoretical numerical analysis: A functional analysis framework*, volume 39. Springer Verlag, 2009.
- [5] O. AXELSSON. *Iterative solution methods*. Cambridge University Press, Cambridge, 1994.
- [6] G. BACHMAN and L. NARICI. *Functional Analysis*. Dover, 2000.
- [7] J. BARZILAI and J.M. BORWEIN. Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8(1):141, 1988.
- [8] H. H. BAUSCHKE, J. M. BORWEIN, and A. S. LEWIS. The method of cyclic projections for closed convex sets in Hilbert space. In Y. Censor and S. Reich, editors, *Contemporary Mathematics, Volume 204: Recent Developments in Optimization Theory and Nonlinear Analysis*, pages 1–38. American Mathematical Society, 1997.
- [9] H.H. BAUSCHKE and J.M. BORWEIN. On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38:367–426, 1996.

- [10] H.H. BAUSCHKE, F. DEUTSCH, H. HUNDAL, and SUNG-HO PARK. Accelerating the convergence of the method of alternating projections. *Transactions of the American Mathematical Society*, 355(9):3433–3461, 2003.
- [11] M. BENZI. Gianfranco Cimmino’s contributions to numerical mathematics. *Atti del Seminario di Analisi Matematica, Dipartimento di Matematica dell’Università di Bologna. Special Volume: Ciclo di Conferenze in Memoria di Gianfranco Cimmino, March-April*, pages 87–109, 2004.
- [12] M. BENZI, G. GOLUB, and J. LIESEN. Numerical solution of saddle point problems. *Acta Numerica*, 14:1–137, 2005.
- [13] E. G. BIRGIN, J. M. MARTINEZ, and M. RAYDAN. Spectral Projected Gradient Methods. In C. A. Floudas and P. M. Pardalos, editors, *Encyclopedia of Optimization*, chapter 19, pages 3652–3659. Springer, 2nd. edition, 2009.
- [14] A. BJÖRCK and T. ELFVING. Accelerated projection methods for computing pseudoinverse solutions of systems of linear equations. *BIT Numerical Mathematics*, 19(2):145–163, 1979.
- [15] J.P. BOYLE and R.L. DYKSTRA. A method for finding projections onto the intersections of convex sets in Hilbert spaces. *Lecture Notes in Statistics*, 37:28–47, 1986.
- [16] C. BYRNE. A unified treatment of some iterative algorithms in signal processing and image reconstruction. *Inverse Problems*, 20(1):103–120, 2004.
- [17] A. CAUCHY. Méthode générale pour la résolution des systemes d’équations simultanées. *Comp. Rend. Sci. Paris*, 25:536–538, 1847.
- [18] Y. CENSOR. Row-action methods for huge and sparse systems and their applications. *SIAM Review*, 23:444–466, 1981.
- [19] Y. CENSOR. Parallel application of block-iterative methods in medical imaging and radiation therapy. *Math. Prog.*, 42:307–325, 1988.
- [20] Y. CENSOR. Computational acceleration of projection algorithms for the linear best approximation problem. *Linear Algebra and its Applications*, 416(1):111–123, July 2006.

- [21] Y. CENSOR, D. GORDON, and R. GORDON. Component averaging: An efficient iterative parallel algorithm for large and sparse unstructured problems. *Parallel Computing*, 27(6):777–808, 2001.
- [22] Y. CENSOR and S.A. ZENIOS. *Parallel Optimization: Theory, Algorithms, and Applications*. Oxford University Press, USA, 1997.
- [23] G. CIMMINO. Calcolo approssimate per le soluzioni dei sistemi di equazioni lineari. *La Ricerca Scientifica. ed. Il progresso tecnico nell' economia nazionale. Consiglio Nazionale delle Ricerche. Ministero dell' Educatione Nazionale*, 9(2):326–333, 1938.
- [24] W. LA CRUZ, J.M. MARTINEZ, and M. RAYDAN. Spectral residual method without gradient information for solving large-scale nonlinear systems of equations. *Mathematics of Computation*, 75(255):1429–1448, 2006.
- [25] W. LA CRUZ and M. RAYDAN. Nonmonotone spectral methods for large-scale nonlinear systems. *Optimization Methods and Software*, 18(5):583–599, 2003.
- [26] Y.H. DAI and L.Z. LIAO. R-linear convergence of the Barzilai and Borwein gradient method. *IMA Journal of Numerical Analysis*, 22(1):1, 2002.
- [27] B.N. DATTA. *Numerical linear algebra and applications*. Brooks/Cole, 1995.
- [28] A. DE PIERRO and A. IUSEM. A parallel projection method for finding a common point of a family of convex sets. *Pesquisa Operacional*, 5:1–20, 1985.
- [29] F. DEUTSCH. Applications of von Neumann's alternating projections algorithm. In P. Kenderov, editor, *Mathematical Methods in Operations Research*, pages 44–51. Sofia, Bulgaria, 1983.
- [30] F. DEUTSCH. Rate of convergence of the method of alternating projections. *ISNM*, 72, Birkhäuser Verlag, Basel:96–107, 1984.
- [31] F. DEUTSCH. The method of alternating orthogonal projections. In S.P. Singh, editor, *Approximation Theory, Spline Functions and Applications*, pages 105–121. Kluwer Academic Publishers, Netherlands, 1992.

- [32] F. DEUTSCH. The angle between subspaces of a Hilbert space. In *Approximation theory, wavelets and applications (Maratea, 1994)*, pages 107–130. Kluwer Acad. Publ., Dordrecht, 1995.
- [33] F. DEUTSCH. *Best Approximation in Inner Product Spaces*. Springer Verlag, New York, 2001.
- [34] F. DEUTSCH and H. HUNDAL. The rate of convergence for the cyclic projections algorithm I: Angles between convex sets. *Journal of Approximation Theory*, 142(1):36–55, 2006.
- [35] H.S. DOLLAR and A.J. WATHEN. Approximate factorization constraint preconditioners for saddle-point matrices. *SIAM Journal on Scientific Computing*, 27(5):1555–1572, 2006.
- [36] L.T. DOS SANTOS. A parallel subgradient projections method for the convex feasibility problem. *J. Comput. Appl. Math.*, 18:307–320, 1987.
- [37] I.S. DUFF, R.G. GRIMES, and J.G. LEWIS. User’s guide for the Harwell-Boeing sparse matrix collection. *Research and Technology Division, Boeing Computer Services, Seattle, Washington, USA*, 1992.
- [38] H.C. ELMAN, D.J. SILVESTER, and A.J. WATHEN. *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*. Oxford University Press, USA, 2005.
- [39] R. ESCALANTE and M. RAYDAN. *Alternating Projection Methods*. SIAM, 2011 (ISBN 978-1-611971-93-4, por aparecer).
- [40] R. FLETCHER. On the Barzilai-Borwein method. *Optimization and Control with Applications (L. Qi, K. L. Teo, X. Q. Yang, eds.)*, pages 235–256, 2005.
- [41] A. FRIEDLANDER, J. M. MARTINEZ, and M. RAYDAN. A new method for large-scale box constrained convex quadratic minimization problems. *Opt. Methods and Software*, 5:57–74, 1995.
- [42] K. FRIEDRICHS. On certain inequalities and characteristic value problems for analytic functions and for functions of two variables. *Trans. Amer. Math. Soc.*, 41:321–364, 1937.

- [43] U.M. GARCIA-PALOMARES and F.J. GONZALEZ-CASTANO. Incomplete projection algorithms for solving the convex feasibility problem. *Numerical Algorithms*, 18:177–193, 1998.
- [44] W.B. GEARHART and M. KOSHY. Acceleration schemes for the method of alternating projections. *J. Comp. Appl. Math.*, 26:235–249, 1989.
- [45] N.I.M. GOULD, D. ORBAN, and P.L. TOINT. CUTeR and SifDec: A constrained and unconstrained testing environment, revisited. *ACM Transactions on Mathematical Software (TOMS)*, 29(4):373–394, 2003.
- [46] L.G. GUBIN, B.T. POLYAK, and E.V. RAIK. The method of projections for finding the common point of convex sets. *USSR Comp. Math. Phys.*, 7:1–24, 1967.
- [47] I. HALPERIN. The product of projection operators. *Acta Sci. Math. (Szeged)*, 23:96–99, 1962.
- [48] M.T. HEATH. *Scientific computing*. McGraw-Hill, 1997.
- [49] G.T. HERMAN, A. LENT, and P.H. LUTZ. Iterative relaxation methods for image reconstruction. *Communications of the ACM*, 21:152–158, 1978.
- [50] L.M. HERNANDEZ-RAMOS. Alternating oblique projections for coupled linear systems. *Numerical Algorithms*, 38(4):285–303, 2005.
- [51] L.M. HERNANDEZ-RAMOS. Accelerated Cimmino’s method for saddle point problems. Technical Report RT-2008-06, Escuela de Computación. Universidad Central de Venezuela, 2008.
- [52] L.M. HERNANDEZ-RAMOS, R. ESCALANTE, and M. RAYDAN. Unconstrained optimization techniques for the acceleration of alternating projection methods. *Numerical Functional Analysis and Optimization, por aparecer*.
- [53] M.R. HESTENES and E. STIEFEL. Method of conjugate gradient for solving linear equations. *J. Res. Nat. Bur. Stand*, 49:409–436, 1952.
- [54] A.N. IUSEM and A.R. DE PIERRO. Convergence results for an accelerated nonlinear Cimmino algorithm. *Numerische Mathematik*, 49(4):367–378, 1986.

- [55] S. KACZMARZ. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bull. Internat. Acad. Pol. Sci. Lett.*, A 35:355–357, 1937.
- [56] W.J. KAMMERER and M.Z. NASHED. A generalization of a matrix iterative method of G. Cimmino to best approximate solution of linear integral equations of the first kind. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.*(8), 51:20–25, 1971.
- [57] W.J. KAMMERER and M.Z. NASHED. On the convergence of the conjugate gradient method for singular linear operator equations. *SIAM J. Numerical Analysis*, 9(1):165–181, 1972.
- [58] S. KAYALAR and H.L. WEINERT. Error bounds for the method of alternating projections. *Math. Control Signals Systems*, 1:43–59, 1988.
- [59] S.H. KULKARNI and M.T. NAIR. A characterization of closed range operators. *Indian Journal of Pure and Applied Mathematics*, 31(4):353–362, 2000.
- [60] J.M. MARTINEZ. An accelerated successive orthogonal projection method for solving large-scale linear feasibility problems. *Computers and Mathematics with Appl.*, 15:367–373, 1988.
- [61] M.Z. NASHED. Steepest Descent for Singular Linear Operator Equations. *SIAM Journal on Numerical Analysis*, 7:358, 1970.
- [62] J. VON NEUMANN. Functional operators vol. II. The geometry of orthogonal spaces. *Annals of Math. Studies*, 22, 1950. Princeton University Press. This is a reprint of mimeographed lecture notes first distributed in 1933.
- [63] M. RAYDAN. Minimización irrestricta de gran escala. Libro en Preparación.
- [64] M. RAYDAN. On the Barzilai and Borwein choice of steplength for the gradient method. *IMA Journal of Numerical Analysis*, 13(3):321–326, 1993.
- [65] M. ROZLOZNIK. Saddle point problems, iterative solution and preconditioning: A short Overview. *Proceedings of the XVth Summer School Software and Algorithms of Numerical Mathematics, I. Marek (Ed.)*, University of West Bohemia, Pilsen, pages 97–108, 2003.

- [66] Y. SAAD. *Iterative Methods for Sparse Linear Systems*. PSW Publishing Company, 1996.
- [67] K.T. SMITH, D.C. SOLMON, and S.L. WAGNER. Practical and mathematical aspects of the problem of reconstructing objects from radiographs. *Bull. Amer. Math. Soc.*, 83:1227–1270, 1977.
- [68] K. TANABE. Projection method for solving a singular system of linear equations and its applications. *Numer. Math.*, 17:203–214, 1971.