Centro de Investigación en Matemáticas, A. C.

CIMAT

Departamento de Ciencias de la Computación

Algoritmos de Optimización sobre Variedades con Restricciones de Ortogonalidad

por

Harry F. Oviedo Leon

Sometida a revisión al Departamento de Ciencias de la Computación en el cumplimiento parcial de los requisitos para obtener el grado de

Maestro en Ciencias de la Computación y Matemáticas Industriales

en el

Centro de Investigación en Matemáticas, A. C.

Firma del autor. . . . . .

Departamento de Ciencias de la Computación

Certificado por. . . . . .

Dr. Oscar S. Dalmau Cedeño
Director de Tesis

Certificado por. . . . . .

Dr. Francisco Javier Solís Lozano
Presidente del comité de evaluación

Guanajuato, Guanajuato, México Junio de 2016

Comité de evaluación

Dr. Francisco Javier Solís Lozano (Presidente)
Centro de Investigación en Matemáticas, CIMAT A. C.

Dr. Rafael Herrera Guzmán (Secretario)
Centro de Investigación en Matemáticas, CIMAT A. C.

Dr. Oscar S. Dalmau Cedeño (Vocal)
Centro de Investigación en Matemáticas, CIMAT A. C.

Resumen

Los problemas de optimización sobre variedades han sido estudiados por más de 40 años para diferentes conjuntos de restricciones, sin embargo, es hasta los años 90's que surgen trabajos que encaran el problema de minimizar una función objetivo sujeto a restricciones de ortogonalidad, a este conjunto de restricciones se le conoce como variedad de Stiefel. Investigar este tipo de problemas ha generado gran interés en años recientes debido al gran número de aplicaciones que han aparecido en diferentes áreas. En esta tesis se abordan problemas de optimización con restricciones de ortogonalidad con la finalidad de proponer eficientes algoritmos para resolver esta clase de problemas.

Utilizando el enfoque de algoritmos de optimización sobre variedades matriciales, se estudian e implementan un total de cuatro algoritmos de búsqueda lineal que resuelven problemas generales de optimización sobre la variedad de Stiefel, tres de los cuales son métodos basados en proyecciones, inspirados en los métodos *Adams-Bashforth* y *Adams-Moulton* que se utilizan para resolver ecuaciones diferenciales numéricamente, y el cuarto método es una generalización de un algoritmo propuesto en años recientes. Además, se estudia un problema particular sobre la variedad de Stiefel llamado *Weighted Orthogonal Procrustes Problem*, para el cual se obtiene una reformulación de dicho problema y se implementa el conocido algoritmo *Iteración de Bregman* para resolver dicha reformulación. Asimismo, para acelerar dichos métodos empleamos una técnica no monótona para seleccionar el tamaño de paso combinada con el paso de Barzilai-Borwein. Algunos resultados teóricos y de convergencia son estudiados para los algoritmos presentados en esta tesis. Igualmente, en este trabajo se realiza un estudio comparativo de los algoritmos propuestos versus algunos de los algoritmos publicados recientemente por diversos autores, con la finalidad de analizar el desempeño y la eficiencia de los métodos como algoritmos generales que resuelven problemas de optimización sobre la variedad de Stiefel.

Agradecimientos

Quisiera hacer constar mi más profundo agradecimiento a todas las personas que me ayudaron o que estuvieron involucradas para que mi persona pueda realizar este viaje, en especial a mi abuela Dina Ortiz, mis padrinos Isabel León y Jose Luis Hernández, a mis tíos Victor León y Ana García, así como también a mi prima María Carolina Peña por su apoyo económico y confianza en mi. También quiero agradecer a los profesores: Msc. Jhonny Escalona y Dr. Joaquín Ortega Sánchez por prestar sus servicios para realizar el examen de admisión en mi país y por hacer posible la conexión entre mi anterior universidad y el CIMAT.

Le doy gracias a toda mi familia, a mis padres y a mi hermano, que fueron mi fuente de inspiración y de esperanza, así como también por apoyarme en todo momento y por los valores que me han inculcado.

Mi especial agradecimiento a mi asesor de tesis Dr. Oscar S. Dalmau Cedeño por su labor de dirección.

Asimismo, quisiera agradecer al todo el colectivo de profesores, investigadores y trabajadores del CIMAT. En particular agradezco a aquellos profesores que con trabajo y dedicación lograron trasmitirme sus conocimientos a través de sus excelentes clases: Dr. Johan Van Horebeek, Dr. Mariano José Juan Rivera Meraz, Dr. Jean-Bernard Hayet, Dr. Rafael Herrera Guzmán y al Dr. Salvador Botello Rionda.

Igualmente quiero agradecer a Karla por su amor incondicional y por ser la mujer que me hace feliz.

A mis amigos Shaday Guerrero y Diana Piñango los cuales al igual que mi persona emprendimos este viaje a este hermoso país para seguir formándonos académicamente y con los cuales he vivido grandes experiencias. Siempre los tengo presentes.

Además, deseo agradecer a mis compañeros de maestría con los que compartí agradables momentos de esparcimiento así como conocimientos e ideas que ayudaron a mi formación, Jorge López, Salvador Botello, Miguel Angel Ochoa, David Dobarro, Fernando Cervantes, José Angel Neria, Mario Ocampo, Ulises Rodriguez, Dora Alvarado y Emmanuel Ovalle.

Finalmente, quiero reconocer al Consejo Nacional de Ciencia y Tecnología (CONACYT) y a la Organización de los Estados Americanos (OEA), ya que nada de esto hubiera sido posible sin el apoyo económico brindado.

Índice general

Índice de tablas

Índice de figuras

Capítulo 1

Introducción

En este trabajo consideramos el problema de optimización con restricciones de ortogonalidad el cual se formula de la siguiente manera:

$$\min_{X \in \mathbb{R}^{n \times p}} \mathcal{F}(X) \quad \text{s.a. } X^\top X = I, \tag{1.1}$$

donde $\mathcal{F} : \mathbb{R}^{n \times p} \to \mathbb{R}$ es una función continuamente diferenciable y donde $I \in \mathbb{R}^{p \times p}$ denota a la matriz identidad de orden p. Al conjunto factible $St(n, p) = \{X \in \mathbb{R}^{n \times p} : X^\top X = I\}$ se le conoce como la variedad de Stiefel (ver [19]), el cual se convierte en la "*esfera unitaria*" cuando $p = 1$ y en el caso cuando $p = n$ se le conoce con el nombre de "*grupo ortogonal*". Es conocido que el conjunto factible del problema (1.1) es un conjunto compacto (ver [19]), esto garantiza la existencia de un óptimo global, sin embargo la variedad de Stiefel no es convexa lo cual convierte a (1.1) en un problema difícil de resolver, de hecho existen algunos aplicaciones del problema (1.1) (con funciones objetivos particulares) como por ejemplo el "*maxcut problem*" y el "*leakage interference minimization*" que son NP-hard (ver [18]). Por otra parte, diseñar algoritmos eficientes de optimización sobre variedades para resolver el problema (1.1), es bastante complicado debido a que preservar la factibilidad de cada iterado es muy costoso computacionalmente. Otra característica que convierte a (1.1) en un problema difícil de resolver es que en general el conjunto de factible hace que existan muchos mínimos locales y por lo tanto no hay garantía de encontrar un minimizador global exceptuando unos pocos casos.

1.1. Motivación

Existe un vasto número de aplicaciones que engloba el problema (1.1) tales como *nearest low-rank correlation matrix problem* [1, 2, 3], el *problema de autovalores lineal* [4, 5], *Kohn-Sham total energy minimization* [6], *orthogonal procrustes problem* [7, 8], *weighted orthogonal procrustes problem* [9], *sparse principal component analysis* [10, 11, 12], *joint diagonalization* (*blind-source separation*) [14, 15], entre otras. Además, muchos problemas bien conocidos tales como *PCA*, *LDA*, *escalamiento multidimensional*, *orthogonal neighborhood preserving projection* pueden ser formulados como el problema (1.1) (ver [16]) y la mayoría de estos problemas se reducen a resolver un problema de valores propios, estas aplicaciones son muy utilizadas en análisis de datos y minería de datos. La ventaja de utilizar algoritmos de optimización para resolver diversas aplicaciones que pueden formularse como el problema (1.1) y que surgen en minería de datos e ingeniería, recae en el hecho de que la teoría de optimización provee una base sólida para el análisis de convergencia, además, la velocidad de convergencia es una propiedad intrínseca de los algoritmos de optimización, por ejemplo, es bien conocido que los algoritmos de optimización

tipo gradiente convergen linealmente y algoritmos de optimización tipo Newton gozan de convergencia cuadrática, si se inicia cerca del óptimo.

Desde los años 70's diversos autores han propuestos algoritmos para resolver problemas de optimización sobre diferentes variedades, sin embargo, es hasta los años 90's que aparecieron los primeros algoritmos para resolver el problema (1.1), por lo tanto el diseño de algoritmos de optimización sobre variedades para resolver problemas sobre la variedad de Stiefel es un problema recientemente estudiado y que sigue siendo de gran interés en la actualidad. Desde entonces, se han propuestos métodos que construyen geodésicas (ver [13, 23, 24, 26]) sobre la variedad de Stiefel y que generan una secuencia de puntos descendiendo por dichas geodésicas hasta obtener un minimizador local, este tipo de métodos encaran el problema (1.1) desde un punto de vista geométrico (ver [13]), a pesar de que dichos métodos resuelven el problema (1.1) su principal desventaja es que la construcción de geodésicas sobre la variedad de Stiefel es muy costoso computacionalmente debido a que, en general, dicha construcción implica calcular exponenciales matriciales. Otros enfoques existentes en la literatura para resolver el problema (1.1), extienden los métodos clásicos de optimización no-lineal vectorial para el caso de variedades tales como métodos tipo Newton [27], métodos Quasi-Newton [19], un modificado descenso del gradiente sobre la variedad de Stiefel [28], entre otros. Por otra parte, en [9] Francisco y Martini proponen un método del gradiente proyectado espectral, el cual es un algoritmo de descenso no-monótono [40, 44, 45] que garantiza el descenso de la función objetivo con respecto a al menos un punto de los calculados en un número fijo de pasos previos; además este algoritmo utiliza un operador de proyección basado en la descomposición SVD para asegurar la factibilidad de cada iterado. Diferente a los putos de vista anteriores, existen métodos basados en retracciones [19] los cuales construyen una secuencia factible de puntos que converge a una solución local, dichos métodos pueden ser vistos como métodos de proyecciones o como una relajación de los métodos basados en geodésicas pero con la particularidad de que no es necesaria la construcción explícita de la geodésica. Por ejemplo, en [21] Wen-Yin utilizan una retracción basada en la *transformada de Cayley*, este método se muestra muy eficiente computacionalmente para resolver el problema (1.1).

A pesar de que en la literatura existen varios algoritmos propuestos para abordar el problema (1.1), muchos de dichos algoritmos son poco eficientes para resolver problemas grandes y algunos inclusive son ineficientes para problemas de talla mediana, esto hace interesante investigar y desarrollar nuevos algoritmos más eficiente para resolver el problema (1.1).

1.2. Objetivos

En este trabajo de investigación estamos interesados en cubrir los siguientes objetivos:

1. Estudiar e implementar métodos de optimización sobre variedades para resolver problemas de optimización sobre la variedad de Stiefel.

2. Proponer algoritmos eficientes para resolver problemas de optimización con restricciones de ortogonalidad.

1.3. Visión general y organización

En este trabajo proponemos y estudiamos cuatro algoritmos para resolver el problema (1.1), donde dos de los cuales emplean esquemas de actualización inspirados en los métodos de *Adams-*

Bashforth y *Adams-Moulton* que son métodos utilizados para resolver ecuaciones diferenciales numéricamente. Nuestro tercer algoritmo propuesto utiliza una fórmula de actualización que se construye a partir de una combinación lineal relacionada con los dos iterados previos. Estas tres primeras propuestas se valen de un operador de proyección basado en la SVD, con la finalidad de asegurar la factibilidad de cada iterado. El cuarto algoritmo que proponemos para resolver el problema (1.1), hace uso de una modificación de la fórmula de actualización del método propuesto en [21] para calcular el nuevo iterado, dicha modificación emplea la factorización de Cholesky en cada paso y garantiza factibilidad de cada punto generado sin necesidad de recurrir a un operador de proyección. Por otra parte, presentamos un quinto algoritmo para resolver el *Weighted Orthogonal Procrustes Problem* (ver [9]) el cual es un problema de mínimos cuadrados sobre la variedad de Stiefel, para resolverlo proponemos utilizar el conocido *algoritmo de Bregman* o también llamado *Iteración de Bregman* para resolver un problema equivalente al *Weighted Orthogonal procrustes Problem*. Además hacemos el análisis de convergencia de nuestros algoritmos y realizamos comparaciones numéricas con algunos algoritmos del estado del arte para estudiar la eficiencia y el desempeño de nuestros algoritmos.

El resto del trabajo se describe a continuación: en el capítulo 2 mostramos contenidos preliminares relacionados con álgebra lineal matricial, elementos de análisis matemático, algunos métodos para resolver ecuaciones diferenciales numéricamente y métodos de optimización sobre variedades, los cuales son necesarios para comprender el contenido de este trabajo. En el capítulo 3 describimos algunos algoritmos propuestos por diferentes investigadores para resolver el problema (1.1) o algunos casos particulares de dicho problema. El capítulo 4 se centra en nuestras propuestas. Describimos cada uno de nuestros algoritmos y mostramos algunas propiedades de convergencia, así como también algunos resultados teóricos de cada método. El capítulo 5 presenta resultados numéricos que comparan nuestros algoritmos con varios algoritmos del estado del arte. También presentamos un análisis comparativo de los resultados obtenidos en los experimentos numéricos entre los diferentes algoritmos. En el capítulo 6 presentamos las conclusiones generales del trabajo y el trabajo futuro. Y por último culminamos con los anexos.

Capítulo 2

Preliminares

Dado que el problema (1.1) corresponde a un problema de optimización sobre una variedad matricial, se enunciarán algunas definiciones y teoremas fundamentales sobre análisis matricial, análisis matemáticos, ecuaciones diferenciales y optimización sobre variedades, con la finalidad de facilitar la comprensión y el análisis de los temas que serán desarrollados en los próximos capítulos. El propósito es establecer una base teórica suficiente para describir tanto los métodos de optimización sobre variedades que han sido investigados por diversos autores, como también los que nosotros proponemos para resolver el problema (1.1). En particular se estudiarán algunos conceptos básicos de análisis matricial, elementos básicos de análisis matemático, diferenciabilidad, algunos métodos para resolver ecuaciones diferenciales numéricamente y métodos de búsqueda lineal para optimización sobre variedades. Cada uno de estos tópicos pueden encontrarse en [4, 19, 25, 31, 33].

En lo que resta del trabajo denotaremos por: $\mathbb{R}^{m\times n}$ al conjunto de las matrices de tamaño $m \times n$ con entradas reales, $I_{n,p}$ a la matriz identidad de tamaño $n \times p$, $\mathbf{0}$ a la matriz nula de apropiadas dimensiones, $A^\top$ denotará la traspuesta de la matriz A, $\mathcal{L}(\mathbb{R}^{m\times n}, \mathbb{R})$ al conjunto formado por todas las funciones lineales continuas que van de $\mathbb{R}^{m\times n}$ en $\mathbb{R}$.

2.1. Análisis matricial

En esta sección mostramos varias definiciones básicas y algunos resultados de análisis matricial, los cuales serán de gran utilidad para la comprensión del contenido presentado en los capítulos posteriores.

Definición 1 (Matriz simétrica y matriz anti-simétrica) *Una matriz cuadrada $A \in \mathbb{R}^{n\times n}$ se dice que es simétrica si satisface que $A^\top = A$, y se dice que es anti-simétrica si $A^\top = -A$.*

Definición 2 (Traza) *Sea $A \in \mathbb{R}^{n\times n}$, la traza de A se define como la suma de los elementos de la diagonal principal, es decir:*

$$Tr[A] := \sum_{i=1}^{n} a_{ii}$$

Teorema 1 (Propiedades de la traza) *Sean $A, B \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{m\times n}$, $D \in \mathbb{R}^{n\times m}$, y $k \in \mathbb{R}$ entonces:*

- $Tr[kA] = kTr[A]$, *en particular:* $Tr[-A] = -Tr[A]$

- $Tr[A+B] = Tr[A] + Tr[B]$
- $Tr[A] = Tr[A^\top]$
- $Tr[CD] = Tr[DC]$
- $Tr[C^\top C] = Tr[CC^\top] = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}^2$
- $Tr[A] = \sum_{i=1}^{n} \lambda_i$, *donde* λ_i *es el i-ésimo valor propio de* A.

Definición 3 (Producto interno Euclidiano matricial) *Dadas dos matrices* $A, B \in \mathbb{R}^{m\times n}$, *se define el producto interno Euclidiano entre* A *y* B *como:*

$$\langle A, B\rangle := \sum_{i,j} a_{ij}b_{ij} = Tr[A^\top B] \tag{2.1}$$

Definición 4 (Norma Frobenius) *Dada una matriz* $A \in \mathbb{R}^{m\times n}$, *se define la norma Frobenius de* A *como:*

$$||A||_F := \sqrt{\sum_{i,j} a_{ij}^2} = \sqrt{Tr[A^\top A]} = \sqrt{\langle A, A\rangle} \tag{2.2}$$

Proposición 1 *Sean* $L, R \in \mathbb{R}^{n\times p}$. *Si definimos por* $W := LR^\top - RL^\top$ *entones:*

$$||W||_F^2 = 2\,Tr[\,L^\top W R\,] \tag{2.3}$$

Demostración.

$$\begin{aligned}
||W||_F^2 &= Tr[\,W^\top W\,] \\
&= Tr[\,(LR^\top - RL^\top)^\top (LR^\top - RL^\top)\,] \\
&= Tr[\,(RL^\top - LR^\top)(LR^\top - RL^\top)\,] \\
&= Tr[\,RL^\top LR^\top\,] - Tr[\,RL^\top RL^\top\,] - Tr[\,LR^\top LR^\top\,] + Tr[\,LR^\top RL^\top\,] \\
&= Tr[\,RL^\top LR^\top\,] + Tr[\,LR^\top RL^\top\,] - 2\,Tr[\,RL^\top RL^\top\,] \quad (\text{ya que } Tr[A] = Tr[A^\top]) \\
&= 2\,Tr[\,LR^\top RL^\top\,] - 2\,Tr[\,RL^\top RL^\top\,] \quad (\text{ya que } Tr[AB] = Tr[BA]) \\
&= 2\,(\,Tr[\,L^\top LR^\top R\,] - Tr[\,L^\top RL^\top R\,]\,) \quad (\text{ya que } Tr[AB] = Tr[BA]) \\
&= 2\,(\,Tr[\,L^\top LR^\top R - L^\top RL^\top R\,]\,) \qquad (\text{ya que } Tr[A+B] = Tr[A] + Tr[B]) \\
&= 2\,(\,Tr[\,L^\top\,(\,LR^\top - RL^\top\,)\,R\,]\,) \\
&= 2\,Tr[\,L^\top W R\,].\,\square
\end{aligned} \tag{2.4}$$

Lema 1 (Desigualdad de Cauchy-Schwarz) *Sean* $A, B \in \mathbb{R}^{m\times n}$ *y consideremos el producto interno definido en* (2.1) *entonces,*

$$|\langle A, B\rangle| \leq ||A||_F ||B||_F$$

Definición 5 (Complemento ortogonal) *El complemento ortogonal de un subespacio vectorial $S \subseteq \mathbb{R}^m$ se define como:*

$$S^{\perp} := \{y \in \mathbb{R}^m : \langle y, x \rangle = 0, \forall x \in S\} \tag{2.5}$$

donde $\langle \cdot, \cdot \rangle$ denota el producto interno estándar de $\mathbb{R}^m$.

Definición 6 (Matriz ortogonal) *Una matriz $Q \in \mathbb{R}^{n \times n}$ es ortogonal si sus columnas forman una base ortonormal de $\mathbb{R}^n$.*

Teorema 2 *Sea $Q \in \mathbb{R}^{n \times n}$ arbitraria, entonces las siguientes afirmaciones son equivalentes:*

- *Q es ortogonal*
- *$Q^{\top}Q = QQ^{\top} = I$*
- *Las filas de Q forman una base ortonormal de $\mathbb{R}^n$*
- *Q preserva los productos internos, es decir: $\langle Qx, Qy \rangle = \langle x, y \rangle$, $\forall x, y \in \mathbb{R}^n$*
- *Q preserva norma, es decir: $||Qx|| = ||x||$, $\forall x \in \mathbb{R}^n$*

La siguiente proposición establece que la norma Frobenius es invariante bajo transformaciones ortogonales.

Proposición 2 *Sea $A \in \mathbb{R}^{m \times n}$. Entonces se cumple que:*

$$||A||_F = ||QAZ||_F \tag{2.6}$$

para cualquier par de matrices ortogonales $Q \in \mathbb{R}^{m \times m}$ y $Z \in \mathbb{R}^{n \times n}$.

Teorema 3 (Fórmula de Sherman-Morrison-Woodbury) *Sean $A \in \mathbb{R}^{n \times n}$ y $U, V \in \mathbb{R}^{n \times k}$, entonces:*

$$(A + UV^{\top})^{-1} = A^{-1} - A^{-1}U(I + V^{\top}A^{-1}U)^{-1}V^{\top}A^{-1}.$$

Definición 7 (Transformada de Cayley) *Sea E un operador sobre un espacio complejo de Hilbert tal que $E + iI$ tiene un kernel trivial, entonces se define la transformada de Cayley $C(E)$ como:*

$$C(E) = (E - iI)(E + iI)^{-1}$$

donde I es el operador identidad.

Nota 1 *La definición de espacio de Hilbert se encuentra en la próxima sección (ver definición 18).*

Nota 2 *Para el caso de operadores lineales reales, la transformada de Cayley se reduce a:*

$$Q = (I - A)(I + A)^{-1}$$

donde $A \in \mathbb{R}^{n \times n}$ es la matriz de representación del operador lineal. Además, en el caso cuando A es una matriz anti-simétrica, se tiene que la transformada de Cayley Q es una matriz ortogonal.

2.2. Elementos de análisis matemático

En esta sección presentamos algunas definiciones y resultados que serán usados en los capítulos posteriores. Dichos resultados se presentan en el espacio de número reales y sobre el espacio $\mathbb{R}^m$, sin embargo también son válidos para el caso del espacio de matrices $\mathbb{R}^{n\times p}$.

Definición 8 (Sucesión convergente) *Una sucesión $\{x_n\}$ de números reales se dice que* converge *a un número real x si para todo $\epsilon > 0$ existe un número $N \in \mathbb{N}$ tal que:*

$$n > N \quad \Rightarrow \quad |x_n - x| < \epsilon. \tag{2.7}$$

Si $\{x_n\}$ converge a x, escribiremos $\lim_{n\to\infty} x_n = x$ o $x_n \to x$.

Definición 9 (Sucesión monótona) *Sea $\{x_n\}$ una sucesión de número reales. Diremos que la sucesión es creciente si se cumple:*

$$x_n < x_{n+1}, \quad \forall n \in \mathbb{N}. \tag{2.8}$$

Diremos que la sucesión es decreciente si satisface que:

$$x_n > x_{n+1}, \quad \forall n \in \mathbb{N}. \tag{2.9}$$

Diremos que la sucesión $\{x_n\}$ es monótona si es creciente o decreciente.

Definición 10 (Sucesión acotada) *Sea $\{x_n\}$ una sucesión de número reales. Diremos que la sucesión es acotada si existe un número real $M > 0$ tal que: $|x_n| \leq M,\ \forall n \in \mathbb{N}$.*

Teorema 4 *Toda sucesión monótona y acotada es convergente.*

Definición 11 (Sucesión de Cauchy) *Consideremos (X, d) un espacio métrico con métrica $d : X \times X \to \mathbb{R}_+$. Una sucesión $\{x_n\}$ de puntos en X es llamada* sucesión de Cauchy *en (X, d) si dicha sucesión tiene la propiedad de que dado $\epsilon > 0$, existe un entero N tal que:*

$$d(x_n, x_m) < \epsilon \quad \text{siempre que } m, n \geq N.$$

Definición 12 (Conjunto acotado) *Diremos que el conjunto $X \subset \mathbb{R}^m$ es acotado si existe un número real $M > 0$ tal que: $||x||_2 \leq M,\ \forall x \in X$.*

Definición 13 (Clausura) *Sea $X \subseteq \mathbb{R}^m$. Se define la clausura de X como el siguiente conjunto:*

$$\overline{X} := \{x \in \mathbb{R}^m : N_\epsilon(x) \cap X \neq \emptyset,\ \forall \epsilon > 0\}. \tag{2.10}$$

donde $N_\epsilon(x) = \{s \in \mathbb{R}^m : ||x - s||_2 < \epsilon\}$.

Definición 14 (Punto límite) *Sean $X \subseteq \mathbb{R}^m$ y $x \in X$. Diremos que x es un "punto límite" de X si dicho punto pertenece a la clausura del conjunto $X - \{x\}$. Al conjunto de todos los puntos límites de X lo denotaremos por X'.*

Definición 15 (Conjunto cerrado) *Diremos que $X \subset R^m$ es cerrado si y solo si $X = \overline{X}$.*

Definición 16 (Compacidad en $\mathbb{R}^m$) *Sea $X \subset \mathbb{R}^m$. Diremos que X es compacto si y solo si X es cerrado y acotado.*

Definición 17 (Espacio completo) *El espacio métrico (X, d) es llamado* completo *si toda sucesión de Cauchy en X es convergente.*

Nota 3 *Dado un número natural n, se puede demostrar que el conjunto $X = \mathbb{R}^n$, con la métrica Euclidiana es un espacio métrico completo, y también es un hecho conocido que el espacio de matrices $X = \mathbb{R}^{m\times p}$ con la extensión de la métrica Euclidiana (es decir con la norma Frobenius) es un espacio métrico completo.*

Definición 18 (Espacio de Hilbert) *Un espació vectorial H con producto interno $\langle\cdot,\cdot\rangle$ que puede ser real o complejo da lugar a una norma:*

$$||x|| = \sqrt{\langle x, x\rangle}$$

la norma inducida por el producto interno. Diremos que H es un espacio de Hilbert, si H es completo con respecto a esta norma.

Teorema 5 *Sea $f : X \to \mathbb{R}$ una función continua sobre un conjunto compacto $X \subset \mathbb{R}$. Entonces existen números reales $m, M \in \mathbb{R}$ tales que:*

$$m \leq f(x) \leq M, \quad \forall x \in X. \tag{2.11}$$

Además, existen $x_1, x_2 \in X$ tales que $f(x_1) = m$ y $f(x_2) = M$, es decir, f alcanza máximo y mínimo en X.

Definición 19 (Conjunto convexo) *Un conjunto X es llamado* convexo *si dados $x, y \in X$ y $\theta \in [0, 1]$ arbitrarios se satisface que:*

$$\theta x + (1 - \theta)y \in X.$$

Definición 20 (Función convexa) *Una función $f : \mathbb{R}^n \to \mathbb{R}$ es* convexa *si su dominio $dom(f)$ es un conjunto convexo y si además se satisface que para todo $x, y \in dom(f)$, y todo $\theta \in [0, 1]$ se verifica que:*

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y).$$

Teorema 6 (Teorema del valor medio) *Sea $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ una función continua en $[a, b]$ y diferenciable en (a, b). Entonces existe $c \in (a, b)$ tal que,*

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \tag{2.12}$$

2.3. Derivadas

En este apartado presentamos algunas definiciones y resultados acerca de diferenciabilidad para el caso de funciones que van del espacio de matrices en el conjunto de los números reales. Comenzaremos introduciendo los conceptos de la diferencial, derivadas parciales y derivada direccional, para culminar con un resultado que relaciona la derivada direccional con el producto interno entre la diferencial y la matriz que representa la dirección, dicho resultado será bastante utilizado en el capítulo 4.

Definición 21 (La Diferencial) *Sean E, F dos espacios vectoriales normados, y sea $f : A \subset E \to F$ una función definida sobre un abierto $A \subset E$ y $a \in A'$. Diremos que la función f es diferenciable en "a" si existe una aplicación lineal $J : E \to F$ tal que:*

$$\lim_{h \to \boldsymbol{0}} \frac{f(a+h) - f(a) - J(h)}{||h||} = \boldsymbol{0} \tag{2.13}$$

donde $\boldsymbol{0}$ denota al vector nulo del espacio vectorial F.
A la función Lineal y continua J se le llama "diferencial de f en a" y la denotaremos por $\mathcal{D}f(a)$.

Nota 4 *La diferencial $\mathcal{D}f$ es única.*

Definición 22 (Derivada direccional) *Dada una función f definida en un abierto A de un espacio vectorial normado E con valores en un espacio vectorial normado F, es decir $f : A \subset E \to F$, se llamará "Derivada direccional de f en $a \in A$ con respecto al vector $\overrightarrow{v} \in E$" al elemento de F definido por:*

$$\mathcal{D}f(a, \overrightarrow{v}) := \lim_{t \to 0} \frac{f(a + t\overrightarrow{v}) - f(a)}{t} \tag{2.14}$$

siempre y cuando el límite exista.

Nota 5 *Sea $\varepsilon > 0$, si definimos la función $\phi : (-\varepsilon, \varepsilon) \to F$ por $\phi(t) = f(a + t\overrightarrow{v})$, es fácil ver que se tiene la fórmula $\mathcal{D}f(a, \overrightarrow{v}) = \phi'(0)$.*

Definición 23 *La función f de la Definición de Derivada Direccional se dirá parcialmente derivable en "a", si $\mathcal{D}f(a, v)$ existe para todo $v \in E$.*

Teorema 7 *Si f es una función definida en un abierto A de un espacio vectorial normado E, con valores en un espacio vectorial normado F, diferenciable en $a \in A$, entonces ella es continua en "a", parcialmente derivable en "a" y además se satisface que:*

$$\mathcal{D}f(a)(v) = \mathcal{D}f(a, v), \forall v \in E \tag{2.15}$$

Demostración.
Ver anexos.□

Teorema 8 (Teorema de representación de Riesz) *Sean H un espacio de Hilbert, y H^* su espacio dual (que consiste en todas las funciones lineales continuas de H en el cuerpo base $\mathbb{R}$ o $\mathbb{C}$). Para cualquier $\phi \in H^*$, existe un único $x_0 \in H$ tal que $\phi(x) = \langle x, x_0 \rangle$ para todo $x \in H$, en este caso decimos que x_0 representa a ϕ.*

Para ver la demostración del teorema de representación de Riesz ver [34].

Consideremos ahora el espacio de matrices con entradas reales de tamaño $m \times n$ denotado por $\mathbb{R}^{m\times n}$, sobre este espació definimos un producto interno dado por (2.1) este producto interno induce una norma llamada "Norma Frobenius" sobre $\mathbb{R}^{m\times n}$, dada por (2.2). Se puede probar que el espacio vectorial $\mathbb{R}^{m\times n}$ es un espacio de Hilbert con respecto a este producto interno. Además, para funciones de valores reales cuyo dominio sea el espacio de matrices $\mathbb{R}^{m\times n}$ o un subconjunto de dicho espacio vectorial, es decir funciones del tipo $\mathcal{F} : A \subseteq \mathbb{R}^{m\times n} \to \mathbb{R}$ se puede redefinir "La Diferencial" gracias al Teorema de Representación de Riesz como sigue:

Definición 24 *Sea $\mathcal{F} : A \subset \mathbb{R}^{m\times n} \to \mathbb{R}$ una función definida sobre un abierto $A \subset \mathbb{R}^{m\times n}$ y $X \in A'$. Diremos que la función $\mathcal{F}$ es diferenciable en X si existe una matriz $G_X \in \mathbb{R}^{m\times n}$ tal que:*

$$\lim_{H\to \boldsymbol{0}} \frac{\mathcal{F}(X+H) - \mathcal{F}(X) - \langle G_X, H\rangle}{||H||} = 0. \tag{2.16}$$

Próximamente definimos la derivada parcial para funciones del tipo $F : A \subset \mathbb{R}^{m\times n} \to \mathbb{R}$, para esto, recordaremos que el conjunto $\mathcal{B} = \{X \in \mathbb{R}^{m\times n} : X = E_{ij} \text{ para algun } (i,j) \in \{1,2,\ldots,m\} \times \{1,2,\ldots,n\}\}$, donde E_{ij} denota a la matriz cuyas entradas son todas iguales a cero excepto la entrada e_{ij} que es igual a uno, forma una base para el espacio vectorial de matrices $\mathbb{R}^{m\times n}$.

Definición 25 (Derivada parcial) *Sean $\mathcal{F} : A \subset \mathbb{R}^{m\times n} \to \mathbb{R}$ una función definida sobre un abierto $A \subset \mathbb{R}^{m\times n}$ y $X \in A'$. Se define la "Derivada Parcial de $\mathcal{F}$ en X respecto a x_{ij}" denotada por $\frac{\partial \mathcal{F}(X)}{\partial x_{ij}}$ como:*

$$\frac{\partial \mathcal{F}(X)}{\partial x_{ij}} := \lim_{t\to 0} \frac{\mathcal{F}(X + tE_{ij}) - \mathcal{F}(X)}{t}. \tag{2.17}$$

Teorema 9 *Sea $\mathcal{F} : \mathbb{R}^{m\times n} \to \mathbb{R}$. Si $\mathcal{F}$ es diferenciable en $X \in \mathbb{R}^{m\times n}$ entonces:*

$$\mathcal{D}\mathcal{F}(X) := G_X = \Big[\frac{\partial \mathcal{F}(X)}{\partial x_{ij}}\Big]_{1\le i\le m, 1\le j\le n}.$$

Demostración.
Ver anexos.□

Teorema 10 *Sea $\mathcal{F} : \mathbb{R}^{m\times n} \to \mathbb{R}$ una función diferenciable en $X \in \mathbb{R}^{m\times n}$, y $Z \in \mathbb{R}^{m\times n}$. Entonces,*

$$\mathcal{D}\mathcal{F}(X,Z) := \lim_{t\to 0} \frac{\mathcal{F}(X+tZ) - \mathcal{F}(X)}{t} = \langle G_X, Z\rangle$$

donde G_X denota a la matriz de derivadas parciales de $\mathcal{F}$ en X.

Demostración.
Ver anexos.□

Nota 6 *En el capítulo 4 denotaremos a la derivada direccional de $\mathcal{F}$ en X en la dirección Z por: $\mathcal{D}\mathcal{F}(X)[Z]$ en lugar de $\mathcal{D}\mathcal{F}(X,Z)$.*

2.3.1. Algunas derivadas de funciones matriciales

Proposición 3 *La derivada de la función $F:\mathbb{R}^{m\times n}\to R$ dada por*

$$F(X)=\frac{1}{2}||AXC-B||_F^2$$

donde $A\in\mathbb{R}^{p\times m}$, $B\in\mathbb{R}^{p\times q}$ y $C\in\mathbb{R}^{n\times q}$ es:

$$G_X=A^\top(AXC-B)C^\top$$

Demostración.
Ver anexos.□

Proposición 4 *La derivada de la función $F:\mathbb{R}^{m\times n}\to R$ dada por*

$$F(X)=Tr[X^\top AX]$$

donde $A\in\mathbb{R}^{m\times m}$ es una matriz simétrica, es:

$$G_X=2AX$$

Demostración.
Ver anexos.□

2.4. Ecuaciones diferenciales

En esta sección presentamos algunos métodos numéricos que se utilizan para resolver ecuaciones diferenciales tanto ordinarias como parciales, los métodos mostrados a continuación son el método de *Crank-Nicolson*, *Adams-Bashforth* y el método de *Adams-Moulton*, una descripción detallada de dichos métodos se encuentra en [25].

2.4.1. Método de Crank-Nicolson

El método de Crank-Nicolson es un método que utiliza *diferencias finitas* para la resolución numérica de ecuaciones en derivadas parciales. Se trata de un método de segundo orden en tiempo, implícito y numéricamente estable. Este método resuelve numéricamente ecuaciones diferenciales del tipo:

$$\frac{\partial u}{\partial t}=F(u,x,t,\frac{\partial u}{\partial x},\frac{\partial^2 u}{\partial x^2}).$$

Si denotamos por $u(i\Delta x,n\Delta t)=u_i^n$, el método de *Crank-Nicolson* realiza la actualización como:

$$u_i^{n+1}=u_i^n+\frac{\Delta t}{2}[F_i^{n+1}(u,x,t,\frac{\partial u}{\partial x},\frac{\partial^2 u}{\partial x^2})+F_i^n(u,x,t,\frac{\partial u}{\partial x},\frac{\partial^2 u}{\partial x^2})].$$

2.4.2. Métodos de Adams-Bashforth y de Adams-Moulton

El método de *Adams-Bashforth* es un método multipaso utilizado para resolver numéricamente ecuaciones diferenciales ordinarias del tipo:

$$\frac{\partial y}{\partial x} = f(x, y) \tag{2.18}$$

este método al igual que muchos otros métodos para resolver ecuaciones diferenciales ordinarias de forma numérica (como por ejemplo el muy conocido método de *Euler*, ver [25]), procede particionando el intervalo [0,1] en un gran número N de subintervalos, todos de longitud $h = \frac{1}{N}$ y entonces desarrolla una fórmula recursiva que relaciona a y_n con $\{y_{n-1}, y_{n-2}, \ldots\}$, donde y_n es la aproximación a $y(x_n = nh)$; dicha recursividad primite integrar numéricamente a (2.18) sobre el intervalo [0,1], puesto que (2.18) es equivalente a:

$$y = y_0 + \int_0^1 f(x, y)\partial x \quad \text{(Ecuación Integral)} \tag{2.19}$$

para el intervalo, $[x_n, x_{n+1}]$, se tiene:

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} f(x, y)dx.$$

A este procedimiento se le conoce como *diferencias finitas*. El método de *Adams-Bashforth* de dos pasos, utiliza *diferencias finitas* y aproxima a y_{n+1} como:

$$y_{n+1} = y_n + h(\frac{3}{2}f_n - \frac{1}{2}f_{n-1}) \tag{2.20}$$

con un error de $O(h^3)$.

De manera similar, el método de *Adams-Moulton* de dos paso, es un método multipaso e implícito usado para resolver numéricamente una ecuación diferencial ordinaria, la fórmula de actualización de dicho método viene dada por:

$$y_{n+1} = y_n + \frac{h}{12}(5f_{n+1} + 8f_n - f_{n-1}) \tag{2.21}$$

con un error de $O(h^4)$.

2.5. Optimización sobre variedades

Optimización sobre variedades hace referencia a algoritmos que mantienen factibilidad durante las iteraciones donde el conjunto de restricciones del problema de optimización es una variedad diferencial, muy parecido al caso de los algoritmos clásicos de "*Punto Interior*" (ver [30]), sin embargo la principal diferencia entre estos métodos, es que en los métodos de *Punto Interior* se actualizan los iterados por medio de una búsqueda lineal, mientras que en el caso de optimización en variedades puede que la actualización de búsqueda lineal no genere un punto que viva en la variedad. La generalización de búsqueda lineal sobre variedades viene siendo el movimiento a lo largo de geodésicas que son curvas de longitud mínima que unen a dos puntos que están sobre la variedad, esta es claramente la generalización puesto que en el espacio Euclídeo la recta es la curva de menor distancia que une a dos puntos.

Al igual que los métodos clásicos de búsqueda lineal, se recurre a métodos iterativos que de cierta forma relajan la idea de construir el camino de descenso mediante la solución de un sistema de ecuaciones diferenciales (como el sistema gradiente, ver [32]), esta relajación hace más eficiente los algoritmos. Generalmente en el caso de optimización sobre variedades, construir una geodésica sobre la variedad es computacionalmente costoso, porque esto implica la solución de ecuaciones diferenciales que a su vez involucra el computo de exponenciales y en gran cantidad de aplicaciones conllevan a calcular exponenciales matriciales que es bastante caro, para hacer más eficiente a los algoritmos de optimización sobre variedades, se puede relajar la construcción de la geodésica por medio del concepto de *retracción*, el cual es una función que mapea vectores tangentes en la variedad, es decir que nos movemos en dirección de un vector tangente mientras al mismo tiempo nos mantenemos en la variedad.

La importancia de optimización en variedades surge en aquellas aplicaciones donde la función objetivo este definida solo en el conjunto factible, lo cual hace imposible utilizar métodos infactibles tales como algoritmos de penalización, o el bien estudiado "Método del Lagrangiano Aumentado (ALM)" (ver [30]). Además, los métodos de optimización sobre variedades buscan no romper la estructura del problema a resolver, es decir, en gran cantidad de aplicaciones existen problemas donde la variedad es una variedad matricial (son los casos de estudios en esta investigación), si bien es cierto que estos problemas los podemos resolver con los métodos clásicos de optimización como por ejemplo el *Descenso del Gradiente*, *Método de Newton*, *Métodos Quasi-Newton*, *Gradiente Conjugado* entre otros, mediante la vectorización del problema y resolverlo como si estuviésemos en $\mathbb{R}^n$, lo que buscan los métodos de optimización sobre variedades es no emplear la vectorización sino utilizar las variables como matrices y emplear las propiedades matriciales, y factorizaciones que tenemos disponibles de análisis numérico matricial.

2.5.1. La variedad de Stiefel

El conjunto $\mathrm{St}(n,p) := \{X \in \mathbb{R}^{n\times p} : X^\top X = I_p\}$ es conocido como *Variedad de Stiefel*, una prueba de que dicho conjunto en realidad es una variedad se encuentra en [19]. Claramente St(n,p) esta contenida en $\mathbb{R}^{n\times p}$, además, es conocido que la variedad de Stiefel es una subvariedad embebida de la variedad $\mathbb{R}^{n\times p}$ [19, ver pág. 26-27] también es sabido que la dimensión de dicha variedad es $np - \frac{1}{2}p(p+1)$, una demostración de este hecho se encuentra en [19, ver pág. 27]. Debido a que la variedad de Stiefel es una subvariedad embebida de $\mathbb{R}^{n\times p}$ su topología es la topología del subconjunto inducida por $\mathbb{R}^{n\times p}$. También es conocido que St(n,p) es un conjunto cerrado y acotado, por lo tanto es un conjunto compacto [19, ver pág. 27]. Nótese que si $p = 1$ entonces la variedad de Stiefel se reduce a la *esfera unitaria* ($\mathcal{S}^{n-1}$) y cuando $p = n$ entonces coincide con el *grupo ortogonal* denotado por $\mathcal{O}_n$ y en este caso su dimensión es $\frac{1}{2}n(n+1)$.

Se puede demostrar fácilmente [19, ver pág. 41-42] que dada una matriz $X \in \mathbb{R}^{n\times p}$, el espacio tangente de la variedad de Stiefel en X denotado por $T_X\mathrm{St}(n,p)$ viene dado por:

$$T_X\mathrm{St}(n,p) \;=\; \{Z \in \mathbb{R}^{n\times p} : X^\top Z + Z^\top X = \mathbf{0}\}. \tag{2.22}$$

Una caracterización alternativa para (2.22) es:

$$T_X\mathrm{St}(n,p) \;=\; \{X\Omega + X_\perp K : \Omega^\top = -\Omega,\ K \in \mathbb{R}^{(n-p)\times p}\}. \tag{2.23}$$

donde $X_\perp$ denota a cualquier matriz de orden $n \times (n-p)$ tal que el $span(X_\perp)$ es el complemento ortogonal del $span(X)$. Para el caso particular cuando $n = p$ es decir, para el *grupo ortogonal*, tenemos que el espacio tangente de $\mathcal{O}_n$ en X esta dado por:

$$T_X\mathcal{O}_n \;=\; \{Z = X\Omega : \Omega^\top = -\Omega\} = X\mathcal{S}_{skew}(n). \tag{2.24}$$

donde $\mathcal{S}_{skew}(n) := \{A \in \mathbb{R}^{n\times n} : A^\top = -A\}$. Asimismo, es conocido [19, ver pág. 48] que dada $X \in \mathbb{R}^{n\times p}$, el espacio normal de la variedad de Stiefel en X es:

$$(\, T_X\mathrm{St}(n,p)\,)^\perp \;=\; \{XS : S \in \mathcal{S}_{sym}(p)\}. \tag{2.25}$$

donde $\mathcal{S}_{sym}(p) := \{A \in \mathbb{R}^{p\times p} : A^\top = A\}$. Igualmente es sabido [19, ver pág. 48] que dado $\varepsilon \in T_X\mathrm{St}(n,p)$, los operadores de proyección sobre el espacio tangente y sobre el espacio normal de la variedad de Stiefel en X son:

$$P_X\varepsilon \;=\; (I - XX^\top)\varepsilon + X\, skew(X^\top\varepsilon), \tag{2.26}$$

$$P_X^\perp\varepsilon \;=\; X\, sym(X^\top\varepsilon), \tag{2.27}$$

respectivamente, donde $sym(M) := \frac{1}{2}(M + M^\top)$ y $skew(M) := \frac{1}{2}(M - M^\top)$.

A continuación definimos el operador de proyección sobre la variedad de Stiefel que será de gran importancia en algunos de nuestros métodos que presentaremos en el capítulo 4.

Definición 26 *Sea $X \in \mathbb{R}^{n\times p}$ una matriz de rango p. El operador de proyección $\pi : \mathbb{R}^{n\times p} \to St(n,p)$ se definido como:*

$$\pi(X) := \arg\min_{Q\in St(n,p)} ||X - Q||_F^2. \tag{2.28}$$

La siguiente proposición nos provee una forma explícita para calcular $\pi(X)$.

Proposición 5 *Sea $X \in \mathbb{R}^{n\times p}$ una matriz de rango p. Entonces $\pi(X)$ esta bien definida. Además, si $X = U\Sigma V^\top$ es la SVD de la matriz X, entonces $\pi(X) = UI_{n,p}V^\top$.*

Demostración.
Ver anexos.□

2.5.2. Algoritmos de optimización en variedades basados en retracciones

En esta subsección introducimos el concepto de *retracción*, el cual es la base de los métodos de *búsqueda lineal* sobre variedades, posterior a esto, presentamos algunas retracciones conocidas sobre la *esfera unitaria*, el *grupo ortogonal* y sobre la variedad de Stiefel puesto que son las variedades que aparecen en las restricciones del problema de interés de este trabajo; por último finalizamos mostrando el algoritmo general de *búsqueda linial* sobre variedades Rimenianas. En nuestro caso, la variedad de Stiefel junto con el producto interno estándar definido a través de la traza forma una variedad Rimeniana (ver [19]).

Retracciones

Uno de los algoritmos más populares y muy bien estudiados teóricamente es el llamado *Descenso del Gradiente*, este algoritmo es utilizado para resolver problemas de optimización sin restricciones donde la función objetivo f es continuamente diferenciable, dicho método traslada un punto de prueba $x(t)$ en la dirección de máximo descenso es decir en dirección de $-\nabla f(x)$ y el método se detiene cuando el gradiente de la función objetivo evaluado en el iterado actual se anule. Una implementación numérica y continua de este método construye una curva que satisface la siguiente ecuación diferencial:

$$\dot{\gamma(t)} \;=\; -\nabla f(\gamma(t)) \tag{2.29}$$

el cual es llamado sistema gradiente que ha sido estudiado ampliamente (ver [32]). Una forma de construir la curva γ es vía los métodos numéricos para resolver ecuaciones diferenciales, pero hacer esto, es muy costoso computacionalmente lo cual lo hace poco práctico. En la práctica, se recurre a los llamados *métodos de búsqueda lineal* los cuales son métodos iterativos que construyen una sucesión $\{x_k\}_{\mathcal{K}}$ de puntos en $\mathbb{R}^n$ recursivamente de la siguiente manera:

$$x_{k+1} \quad = \quad x_k + \alpha\eta \tag{2.30}$$

donde $\alpha > 0$ es llamado *tamaño de paso* y η es una dirección de descenso en x_k, es decir cualquier vector que satisfaga $\nabla f(x_k)^\top \eta < 0$. En el caso del descenso del gradiente, se escoge $\eta = -\nabla f(x_k)$.

En general, para el caso de optimización continua sobre variedades la idea es análoga al caso de optimización clásica, lo que se desea es construir una geodésica es decir, una curva que se mueva a lo largo de un vector tangente sobre la variedad, nuevamente la construcción de dicha geodésica implica resolver una ecuación diferencial (ver [13, 23, 24]) que en general conlleva a calcular exponenciales matriciales, lo cual es muy costoso computacionalmente. Para tratar con este problema, se recurre a un método iterativo donde se relaja la metodología de construir una geodésica, por un mapeo suave que traslada el punto que vive en la variedad (el iterado que se obtuvo del paso anterior) a lo largo de una dirección tangente a la variedad en dicho punto (un vector tangente al punto anterior) y a su vez dicho mapeo se mantiene sobre la variedad, es decir, dicha función mapea vectores tangente a la variedad; estas funciones son llamadas *retracciones*.

Definición 27 (Retracción) *Una "Retracción" sobre una variedad $\mathcal{M}$, es un mapeo suave (de clase $\mathcal{C}^\infty$ sobre su dominio) R que va del haz $T\mathcal{M}$ sobre $\mathcal{M}$ con las siguientes propiedades. Sea R_x la restricción de R a $T_x\mathcal{M}$.*

1. *$R_x(0_x) = x$, donde 0_x denota el elemento nulo de $T_x\mathcal{M}$.*
2. *Con la identificación canónica $T_{0_x}T_x\mathcal{M} \simeq T_x\mathcal{M}$, R_x satisface*

$$\mathcal{D}R(0_x) \quad = \quad id_{T_x\mathcal{M}}, \tag{2.31}$$

donde $id_{T_x\mathcal{M}}$ denota el mapeo identidad sobre $T_x\mathcal{M}$.

Para el caso de variedades $\mathcal{M}$ que son subvariedades embebidas en un espacio vectorial $\mathcal{E}$ como por ejemplo la variedad de Stiefel, se puede calcular una retracción R_x de forma más fácil. Dado $x \in \mathcal{M}$ y $\varepsilon \in T_x\mathcal{M}$, calcular $R_x(\varepsilon)$ por

1. moverse a lo largo de ε para obtener el punto $x + \varepsilon$ que vive en $\mathcal{E}$;
2. "proyectar" el nuevo punto $x + \varepsilon$ sobre la variedad $\mathcal{M}$.

Nótese que para este tipo especial de variedades la retracción es básicamente moverse a lo largo de un vector tangente y luego proyectar sobre la variedad, esto es muy parecido al método clásico del *Gradiente proyectado* (ver [30]). Un estudio detallado de este tipo especial de retracciones se encuentra en [20]. Para ver más detalles de la definición 27 y para el caso de la construcción de retracciones cuando la variedad $\mathcal{M}$ es una subvariedad embebida en un espacio vectorial [19, ver pág. 55-57]. Próximamente mostramos algunas retracciones sobre la esfera unitaria, la variedad de Stiefel y sobre el grupo ortogonal.

Ejemplo 1 (Retracción sobre la esfera unitaria $\mathcal{S}^{n-1}$) *Sea $x \in \mathcal{S}^{n-1}$ y $\varepsilon \in T_x\mathcal{S}^{n-1}$ entonces se puede probar (ver pág.57 en [19]) que la siguiente función,*

$$R_x(\varepsilon) := \frac{x+\varepsilon}{||x+\varepsilon||_2}, \tag{2.32}$$

es una retracción sobre la esfera unitaria. Nótese que para ε fijo,

$$R_x(\varepsilon) := \arg \min_{||R||_2=1} ||(\, x+\varepsilon\,) - R||_2^2 \tag{2.33}$$

es decir, $R_x(\varepsilon)$ es el punto sobre la esfera unitaria que minimiza la distancia a $x+\varepsilon$.

Ejemplo 2 (Retracciones sobre el grupo ortogonal) *Sea $X \in \mathcal{O}_n$ y $X\Omega \in T_x\mathcal{O}_n$ entonces las siguientes funciones:*

$$R_X(X\Omega) := X\,\mathrm{qf}(I+\Omega), \tag{2.34}$$
$$R_X(X\Omega) := X(I+\Omega)(I-\Omega^2)^{-1/2}, \tag{2.35}$$
$$R_X(X\Omega) := X\mathrm{Giv}(\Omega), \tag{2.36}$$
$$R_X(X\Omega) := X(I-\frac{1}{2}\Omega)(I+\frac{1}{2}\Omega)^{-1}, \tag{2.37}$$
$$Exp_X(X\Omega) := X\exp(\Omega), \tag{2.38}$$

son todas retracciones sobre el grupo ortogonal, donde $\mathrm{qf}(A) := Q$ representa la matriz ortogonal Q que se obtiene de la factorización QR *de la matriz A, y con $\mathrm{Giv}(\Omega) = \Pi_{1\le i<j\le n} G(i,j,\Omega_{ij})$ donde $G(i,j,\theta)$ es la* rotación de Givens *de ángulo θ en el plano (i,j), y donde $\exp(A)$ representa la exponencial matricial de A. Una descripción más detallada de como verificar que estas funciones son realmente retracciones la encontramos en [19, ver pág. 58-59].*

Ejemplo 3 (Retracciones sobre la variedad de Stiefel) *Sea $X \in St(n,p)$ y consideremos $\varepsilon \in T_X St(n,p)$ entonces una retracción basada en la "descomposición polar" es,*

$$R_X(\varepsilon) := (X+\varepsilon)(I_p + \varepsilon^\top\varepsilon)^{-1/2}, \tag{2.39}$$

otra retracción sobre la variedad de Stiefel basada en la factorización QR es,

$$R_X(\varepsilon) := \mathrm{qr}(X+\varepsilon). \tag{2.40}$$

Como podemos apreciar, se pueden definir muchas retracciones sobre la variedad de Stiefel y sobre el grupo ortogonal, aquí solo mostramos algunos ejemplos. Al momento de implementar algoritmos de optimización basados en retracción, la tarea fundamental es la elección de la retracción apropiada que tenga el mejor desempeño, es decir, aquella que requiera de poco cómputo por iteración para tener un algoritmo más eficiente. Por ejemplo las retracciones basadas en la descomposición polar (2.35) y (2.39) en general son poco eficientes puesto que dichas retracción necesitan calcular la inversa de una matriz y además, deben calcular la raíz cuadrada matricial que en general implica encontrar la descomposición espectral de una matriz lo cual es poco eficiente. Sin embargo cuando p es pequeño (o n es pequeño en el caso de $\mathcal{O}_n$) hacer dicho cómputo no es tan difícil y puede ser una buena opción para el algoritmo.

Otras retracciones como (2.38) y (2.36) también son poco eficientes en problemas de gran tamaño puesto que el cálculo de rotaciones de Givens y de exponenciales matriciales es muy costoso computacionalmente. Entre las retracciones más eficientes que mostramos en esta sección están aquellas basadas en la factorización QR, es decir, las retracciones (2.34) y (2.40) dichas retracciones emplean la factorización QR que al menos es más eficientes que las mencionadas anteriormente. La retracción que al parecer es la más eficiente sobre el grupo ortogonal de las que presentamos aquí, es aquella que utiliza la *transformada de Cayley* (2.37), esto se debe a que solo involucra multiplicaciones matriciales y el cálculo de una inversa matricial, de hecho uno de los trabajos del estado del arte (ver [21]), utiliza precisamente la retracción basada en la *trasformada de Cayley* para resolver eficientemente problemas de optimización con restricciones de ortogonalidad y la extiende para el caso de la variedad de Stiefel .

Algoritmos de búsqueda lineal sobre variedades

Antes de explicar los métodos de búsqueda lineal introduciremos la siguiente definición.

Definición 28 (Sucesión gradiente-relacionada) *Dada una función objetivo f de una variedad diferencial $\mathcal{M}$ en $\mathbb{R}$, una sucesión $\{\varepsilon_k\}$ tal que $\varepsilon_k \in T_{x_k}\mathcal{M}$ se dice que es* gradiente-relacionada *si para cualquier subsucesión $\{x_k\}_{k\in\mathcal{K}}$ de $\{x_k\}$ que converja a un punto no crítico de f, la correspondiente subsucesión $\{x_k\}_{k\in\mathcal{K}}$ es acotada y satisface que:*

$$\lim_{k\to\infty} \sup_{k\in\mathcal{K}} \langle \nabla f(x_k), \varepsilon_k \rangle < 0. \tag{2.41}$$

Los métodos de búsqueda lineal sobre variedades, son métodos iterativos que construyen una sucesión de punto $\{x_k\}_{k\in\mathcal{K}}$ que viven en la variedad, de manera recursiva como sigue,

$$x_{k+1} \quad = \quad R_{x_k}(t_k\varepsilon_k), \tag{2.42}$$

donde ε_k es un vector tangente a la variedad en x_k, $t_k > 0$ es el *tamaño de paso*, y $R_x(\varepsilon)$ es una retracción sobre la variedad. Además, el vector tangente ε_k se escoge de tal forma que descienda monótonamente la función objetivo, es decir que la sucesión de vectores tangente $\{\varepsilon_k\}_{k\in\mathcal{K}}$ debe ser gradiente-relacionada, y el *tamaño de paso* t_k se selecciona generalmente usando la *regla de Armijo*, aunque existen muchos otros criterios para seleccionar el *tamaño de paso* adecuado.

El algoritmo general de búsqueda lineal para resolver problemas de minimización de una función objetivo sobre una variedad se presenta a continuación:

Algoritmo 1 Método de búsqueda lineal general LSM

Entrada: Variedad Rimeniana $\mathcal{M}$; una función continuamente diferenciable $\mathcal{F} : \mathcal{M} \to \mathbb{R}$ (función objetivo); una retracción $R : T\mathcal{M} \to \mathcal{M}$; $\rho, \sigma \in (0, 1)$; $\tau > 0$; $k = 0$. Iterado inicial $X_0 \in \mathcal{M}$.

mientras No halla convergencia **hacer**

Seleccionar $\eta_k \in \mathtt{T}_{X_k}\mathcal{M}$ tal que la sucesión $\{\eta_i\}_{i=0,1,\ldots}$ sea una sucesión gradiente-relacionada.

mientras $\mathcal{F}(R_{X_k}(\tau\eta_k)) \geq \mathcal{F}(X_k) + \sigma\tau\mathcal{D}\mathcal{F}(X_k)[\dot{R}_{X_k}(0)]$ **hacer**

$\tau = \rho\tau$;

fin mientras

Actualización del iterado: $X_{k+1} = R_{X_k}(\tau\eta_k)$;

$k = k + 1$;

fin mientras

$X^* = X_k$.

Salida: X^* el minimizador local.

Capítulo 3

Estado del arte

Si bien es cierto que desde los años 70's existen diversos trabajos que han estudiado el problema de minimizar una función objetivo sobre una variedad (ver [35, 36]), es hasta los años 90's donde se proponen algoritmos generales para resolver problemas de optimización sobre la variedad de Stiefel (ver [13, 17, 37]). A partir de dicha década se han propuesto algoritmos que se mueven a lo largo de curvas que localmente minimizan la longitud de arco entre dos puntos de la variedad, a dichas curvas se le conocen como *Geodésicas*, también se han propuesto métodos factibles que utilizan diferentes operadores de proyección para garantizar la construcción de una sucesión factible de puntos, asimismo se han estudiado también métodos basados en retracciones, igualmente algunos investigadores han propuestos otros métodos tales como el *método penalidad cuadrática* y la *iteración de Bregman* que no siguen la filosofía de los algoritmos de optimización en variedades.

3.1. Esquemas de actualización basados en geodésicas

En 1998, Edelman y otros [13] desarrollan el algoritmo del gradiente conjugado y el algoritmo de Newton con el propósito de resolver problemas sobre la variedad de Stiefel, los mencionados autores utilizan un esquema de actualización que construye una geodésica en la cual los iterados están sobre la curva definida por:

$$Y(\tau, X) := [X, Q] \ \texttt{exp}\left(\tau \begin{bmatrix} -X^\top H & -R^\top \\ R & 0 \end{bmatrix}\right) \begin{bmatrix} I_p \\ 0 \end{bmatrix}, \tag{3.1}$$

donde $QR = -(I_n - XX^\top)H$ es la factorización QR de $-(I_n - XX^\top)H$, y donde la matriz $H \in \mathbb{R}^{n\times p}$ es construida mediante la fórmula de la dirección de Newton y la del gradiente conjugado pero sobre variedades, es decir, usando el gradiente y Hessiano Rimeniano (si desea entrar más en detalle ver la definición de gradiente y Hessiano Rimeniano en [19]). En [13], se estudia desde un punto de vista geométrico dichas propuestas que funcionan bien en problemas pequeños, sin embargo en problemas grandes, la fórmula de actualización (3.1) es poco eficiente debido a que esta involucra el cálculo de una exponencial matricial de una matriz de tamaño $2p \times 2p$, y además, necesita calcular la factorización QR de una matriz de tamaño $n \times p$, lo cual hace que los algoritmos propuestos en [13] sean muy lentos para el caso cuando $p >= n/2$.

Otra propuesta que utiliza geodésicas es presentado por Abrudan y colabradores en [24], dichos autores estudian el caso de minimización de una función objetivo matricial con restricciones unitarias, es decir la variedad es el conjunto de matrices unitarias cuadradas, esto quiere decir que estudian el caso cuando $p = n$ y las entradas de la matriz X pueden ser reales o complejas.

Específicamente, Abrudan y colaboradores proponen el método de gradiente conjugado utilizando el siguiente esquema de actualización:

$$Y(\tau, X) = \quad \texttt{exp}(-\tau M_1)X, \tag{3.2}$$

donde $M_1 \in \mathbb{C}^{n\times n}$ es una matriz anti-hermitiana (anti-simétrica en el caso real) dada por la actualización de la dirección del gradiente conjugado Rimeniano usando el parámetro propuesto por Polak-Ribièrre, para más detalles sobre como dichos autores construyen la matriz M_1 ver [24]. La fórmula (3.2) resuelve más eficientemente el problema (1.1) para el caso cuando $p >= n/2$ que el método propuesto por Edelman, sin embargo, todavía es poco eficiente puesto que necesita calcular la exponencial matricial de un matriz de tamaño $n \times n$ lo cual es muy costoso computacionalmente si n es grande. También un algoritmo de descenso del gradiente ha sido propuesto por Abrudan y otros en [23] que también emplea la construcción de una geodésica. Otro algoritmo que utiliza una geodésica ha sido propuestos en [26].

3.2. Esquemas de actualización basados en proyecciones y en retracciones

Existen varios métodos propuestos que emplean operadores de proyección para garantizar que cada iterado se mantenga en la variedad de Stiefel, Manton en 2002 (ver [28]) propone un *Modificado descenso del gradiente* dado por el siguiente esquema de actualización:

$$Y(\tau, X) = \pi(\, X - \tau Z^{(sd)}\,), \tag{3.3}$$

donde $\pi(\cdot)$ es el operador de proyección definido en (2.28), y $Z^{(sd)} := D_X - XD_X^\top X$ con $D_X \in \mathbb{R}^{n\times p}$ es cualquier matriz que satisface:

$$\mathcal{F}(X - Z) = \mathcal{F}(X) - Tr[D_X^\top Z] + O(||Z||_F^2), \quad \forall Z \in T_X St(n, p).$$

Nótese que una selección válida para D_X puede ser $D_X = \mathcal{DF}(X_k)$, además observe que (3.3) es una *retracción* debido a que la matriz $Z^{(sd)} \in T_X St(n, p)$. Manton en [28] también propone un *Modificado método de Newton* el cual se mueve a lo largo de la dirección de Newton Z^{New} y luego calcula el nuevo iterado como la proyección de $X + \tau Z^{New}$ usando el operador de proyección $\pi(\cdot)$ al igual que en (3.3).

Absil y otros en [19], también proponen un método que para la actualización de los iterados, emplea un operador de proyección basado en la factorización QR, dicho método utiliza la siguiente fórmula de actualización:

$$Y(\tau, X) = \texttt{qf}(X + D), \tag{3.4}$$

donde $D \in T_X St(n, p)$ y $\texttt{qf}(A)$ denota a la matriz ortogonal obtenida mediante la factorización QR de la matriz A, nótese que la función $\texttt{qf}(\cdot)$ es también un operador de proyección sobre la variedad de Stiefel. Además, en [19] Absil y otros proponen otro método que emplea una *retracción* basada en la descomposición polar, dada por:

$$Y(\tau, X) = (X - \tau D)(I_p + \tau^2 D^\top D)^{-1/2}, \tag{3.5}$$

donde $D \in T_X St(n, p)$.

Los esquemas de actualización propuestos por Absil, Manton y en general los métodos basados en proyecciones, casi siempre muestran mejor desempeño que los métodos que construyen

geodésicas, debido a que por lo general estos últimos requieren calcular exponenciales matriciales, lo cual es más costoso computacionalmente que proyectar sobre la variedad de Stiefel. Sin embargo para problemas grandes el solo cálculo de la proyección es también muy costoso, y esto convierte lentos a los algoritmos que utilizan operadores de proyección.

En 2012, Wen y Yin en [21] proponen un esquema de actualización parecido a la fórmula que utiliza el método de Crank-Nicolson (ver capítulo 2), el cual viene dado por:

$$Y(\tau, X) = X + \frac{\tau}{2} A_X (X + Y(\tau, X)), \tag{3.6}$$

donde $A_X = \mathcal{DF}(X)X^\top - X\mathcal{DF}(X)^\top$, observe que (3.6) es un esquema implícito, sin embargo Wen y Yin demuestran en [21] que la fórmula (3.6) es equivalente al siguiente esquema explícito que emplea la *transformada de Cayley*:

$$Y(\tau, X) = (I + \frac{\tau}{2} A_X)^{-1}(I - \frac{\tau}{2} A_X)X, \tag{3.7}$$

el cual es una *retracción* sobre la variedad de Stiefel. Nótese que este último esquema de actualización, no requiere la utilización de un operador de proyección, lo cual lo hace bastante eficiente computacionalmente, sin embargo, la fórmula (3.7) todavía necesita invertir una matriz de tamaño $n \times n$ que es bastante difícil si n es grande, para solventar este problema los autores de [21] proponen utilizar la fórmula de *Sherman-Morrison-Woodbury* y a partir de esta, obtienen una fórmula equivalente a (3.7) dada por:

$$Y(\tau, X) = X - \tau U(I + \frac{\tau}{2} V^\top U)^{-1} V^\top X, \tag{3.8}$$

donde $A_X = UV^\top$, con $U = [\mathcal{DF}(X), X]$ y $V = [X, -\mathcal{DF}(X)]$. Nótese que (3.8) es una fórmula de actualización más eficiente que (3.7) para el caso cuando $p < \frac{n}{2}$, debido a que esta requiere invertir una matriz de tamaño $2p \times 2p$. De los algoritmos existentes en la literatura para resolver el problema (1.1), el método propuesto por Wen-Yin es uno de los que presenta un mejor desempeño y eficiencia en la actualidad.

En 2014 Francisco y Martini [9] estudiaron el comportamiento de un algoritmo de región de confianza para resolver el problema *Weighted Orthogonal Procrustes Problem(WOPP)*, es cual es un problema particular del esquema general de optimización (1.1). Dicho algoritmo, puede ser visto como una variación del método *Levenberg-Marquardt* analizado por Birgin y otros en [38] (ver también [42]). Más específicamente, Francisco y Martini proponen ajustar el siguiente modelo cuadrático local:

$$Q_\rho^k(X) = Tr[\mathcal{DF}(X_k)^\top (X - X_k)] + \frac{\sigma_\rho^k + \rho}{2} ||X - X_k||_F^2, \tag{3.9}$$

donde σ_ρ^k es un parámetro actualizado en cada iteración y ρ es un parámetro de regularización. Dichos autores seleccionan el siguiente iterado como aquel que resuelve el modelo cuadrático (3.9) sujeto a la variedad de Stiefel, es decir, como aquel que resuelve el siguiente problema de optimización:

$$\min_{X \in \mathbb{R}^{n \times p}} Q_\rho^k(X) \quad \text{s.a.} \quad X^\top X = I, \tag{3.10}$$

dicho problema (3.10) tiene solución cerrada dada por:

$$W_{k+1} = X_k - \frac{1}{\rho + \sigma_\rho^k} \mathcal{DF}(X_k), \tag{3.11}$$

$$X_{k+1} = \pi(W_{k+1}), \tag{3.12}$$

donde $\pi(\cdot)$ es el operador de proyección definido en (2.28). Al esquema anterior, lo combinan con una búsqueda lineal no-monótona (ver [43]) junto al paso de *Barzilai-Borwein* (ver [41]) para seleccionar el *tamaño de paso* adecuado, dicho *tamaño de paso* t_k debe satisfacer la siguiente condición:

$$\mathcal{F}(X_{k+1}) \leq \mathcal{F}(X_{l(k)}) + \gamma t_k ||\mathcal{D}\mathcal{F}(X_k)||_F^2, \tag{3.13}$$

donde $\gamma \in (0,1)$ y $\mathcal{F}(X_{l(k)}) = \max\{\mathcal{F}(X_{k-j}) : 0 \leq j \leq m(k)\}$ con $m(k)$ definido recursivamente como sigue: $m(0) = 0$ y para $k \geq 1$ se define como $0 \leq m(k) \leq \min\{m(k-1)+1, M\}$, con $M \in \mathbb{Z}^+$. Existen muchas estrategias no-monótonas para la selección del *tamaño de paso*, el algoritmo propuesto por Francisco y Martini emplea la técnica descrita arriba, nosotros a diferencia de dichos autores, utilizaremos en nuestro algoritmo 3 (ver Capítulo 4), otra técnica no-monótona que es propuesta en [45] para la selección del *tamaño de paso*, debido a que esta funcionó mejor en la práctica. La técnica de selección del *tamaño de paso* empleada por los autores Francisco y Martini en su algoritmo, garantiza convergencia global de la sucesión de los iterados a un punto estacionario bajo ciertas hipótesis, una demostración de este resultado se encuentra en [43]. Nótese que en general, el algoritmo propuesto por Francisco y Martini es un algoritmo de optimización basado en proyección, puesto que para resolver el modelo cuadrático (3.9) necesita utilizar un operador de proyección para garantizar factibilidad de la sucesión de iterados, por otra parte, dicho método puede llegar a ser bastante lento en problemas grandes, puesto que requiere el cálculo de la SVD en cada iteración para resolver cada modelo cuadrático.

3.3. Otras propuestas

Existen otras algoritmos propuestos para resolver problemas de optimización sobre la variedad de Stiefel que han sido investigados por diversos autores, pero que no son algoritmos de optimización sobre variedades, es decir, algunos son métodos infactibles como lo es el bien conocido método de penalidad cuadrática, también se ha propuesto la novedosa *iteración de Bregman*, entre otros. En esta subsección mencionamos alguno de estos estudios realizados actualmente.

En 2014 R. Lai y S. Osher [29] estudian el siguiente problema general de optimización muy relacionado al problema (1.1):

$$\min_{X \in \mathbb{R}^{n\times p}} \mathcal{F}(X) \quad \text{s.a.} \quad X^\top Q X = I, \tag{3.14}$$

donde $\mathcal{F} : \mathbb{R}^{n\times p} \to \mathbb{R}$ es considerada una función convexa y $Q \in \mathbb{R}^{n\times n}$ es una matriz definida positiva conocida. Directamente la *iteración de Bregman* no puede utilizarse para resolver el problema (3.14), puesto que este no tiene restricciones lineales, para aplicar dicho algoritmo, Lai y Osher introducen una variable auxiliar $P \in \mathbb{R}^{n\times p}$ dada por $P = LX$, donde $L^\top L = Q$ es la factorización de Cholesky de la matriz Q. Así, los autores de [29] construyen una formulación equivalente al problema (3.14) dada por:

$$\min_{X,P \in \mathbb{R}^{n\times p}} \mathcal{F}(X) \quad \text{s.a.} \quad LX = P \text{ y } P^\top P = I, \tag{3.15}$$

El algoritmo *iteración de Bregman* propuesto por R. Lai y S. Osher en [29] para resolver el problema (3.15) procede de la siguiente forma:

$$1. - X^{k+1} = \arg\min_X \mathcal{F}(X) + \frac{r}{2}||LX - P^k + B^k||_F^2, \tag{3.16}$$

$$2. - P^{k+1} = \arg\min_P \frac{1}{2}||P - LX^{k+1} - B^k||_F^2 \quad P^\top P = I,, \tag{3.17}$$

$$3. - B^{k+1} = B^k + LX^{k+1} - P^{k+1}, \tag{3.18}$$

donde $r > 0$ es un parámetro dado. La ventaja del algoritmo anterior recae en el hecho de que el problema del paso 1, es un problema de optimización convexa sin restricciones, el cual en general es fácil de resolver y en muchas aplicaciones tiene solución cerrada, además el problema de optimización sobre la variedad de Stiefel del paso 2 tiene solución analítica, estas dos característica hacen que dicho algoritmo sea muy eficiente y veloz.

Wen y otros en [22] estudian el problema de valores propios lineales (ver [19]) y proponen resolverlo mediante el conocido y bien estudiado *método de penalidad cuadrática* (ver [30]), es decir que los autores de [22], plantean resolver el problema *Linear Eigenvalues Problem* (ver [19]), mediante la resolución del siguiente problema de optimización irrestricto en cada iteración:

$$\min_{X \in \mathbb{R}^{n \times p}} \mathcal{F}_\mu(X) := -\frac{1}{2} Tr[X^\top A X] + \frac{\mu}{4} ||X^\top X - I||_F^2, \tag{3.19}$$

donde $\mu > 0$ es el parámetro de penalidad. Es conocido que el modelo clásico de penalidad cuadrática se aproxima al problema con restricciones original solo cuando el parámetro de penalidad tiende a infinito, sin embargo los autores de [22] muestran que el problema (3.19) es equivalente al problema de valores propios lineales cuando el parámetro de penalidad es escogido apropiadamente.

Para resolver cada sub-problema de optimización sin restricciones Wen y colaboradores utilizan el bien estudiado *descenso del gradiente* (ver [30]), el cual actualiza los iterados mediante la siguiente fórmula recursiva:

$$X_{k+1} = X_k - \alpha_k \nabla \mathcal{F}_\mu(X_k), \tag{3.20}$$

donde $\nabla \mathcal{F}_\mu(\cdot)$ es la derivada de la función de penalidad y $\alpha_k > 0$ es el *tamaño de paso* que es seleccionado usando *Backtracking* combinado con el paso de *Barzilai-Borwein*.

Como podemos notar se han propuesto algunos métodos infactibles para resolver problemas de optimización sobre la variedad de Stiefel, este último es decir el método de penalidad cuadrática también se puede proponer para resolver el problema general (1.1), aunque es conocido que dicho método genera mal condicionamiento de la matriz Hessiana de la función de penalidad cuando el parámetro de penalidad se acerca a infinito, lo cual impide utilizar métodos tipo *Newton* para acelerar la convergencia al resolver los sub-problemas de optimización sin restricciones de cada iteración. Otra método infactible que puede aplicarse como alternativa de método de solución para resolver el problema (1.1) es el conocido *Método del Lagrangiano Aumentado* (ALM) (ver [30]), este método resuelve el problema del mal condicionamiento de la matriz Hessiana, sin embargo en este trabajo de investigación estamos interesados en estudiar y proponer algoritmos de optimización sobre variedades para resolver el problema (1.1), a los cuales les concierne generar una sucesión factible de iterados que converja a un óptimo local del problema (1.1).

Capítulo 4

Métodos propuestos para resolver problemas de optimización con restricciones de ortogonalidad

En este capítulo presentamos nuestros cinco métodos propuestos para resolver problemas de optimización con restricciones de ortogonalidad, todos ellos construyen una sucesión de puntos factibles que converge a un mínimo local. Tres de nuestros métodos utilizan la proyección sobre la variedad de Stiefel definida en (2.28), para asegurar que la sucesión creada sea factible, mientras que nuestro cuarto método propuesto utiliza una fórmula de actualización que emplea la factorización de Cholesky de tal forma que garantiza factibilidad. Nuestras cuatro primeras propuestas resuelven el problema general (1.1), mientras que nuestra quinta propuesta utiliza un *Splitting Bregman Algorithm* para resolver el *Weighted Orthogonal Procrustes Problem* (WOPP). En las siguientes secciones, explicamos de forma detallada cada una de las fórmulas de actualización de todas nuestras cinco propuestas, también revelamos las estrategias que utilizaremos para seleccionar el tamaño de paso de nuestros algoritmos. Además, mostramos algunos resultados de convergencia de nuestros métodos y presentamos dichos algoritmos.

4.1. Condiciones de optimalidad

La función Lagrangiana asociada al problema de optimización (1.1) viene dada por:

$$\mathcal{L}(X, \Lambda) = \mathcal{F}(X) - \frac{1}{2} Tr[\Lambda(X^\top X - I)] \tag{4.1}$$

donde Λ es la matriz de multiplicadores de Lagrange, dicha matriz es simétrica debido a que la matriz $X^\top X$ también lo es. De la función Lagrangiana se desprenden las condiciones de optimalidad de primer orden para el problema (1.1):

$$G - X\Lambda = \mathbf{0}, \tag{4.2a}$$

$$X^\top X - I_p = \mathbf{0}, \tag{4.2b}$$

donde $G := \mathcal{D}\mathcal{F}(X)$.

Lema 2 *(cf. Wen-Yin [21], 2012, page 7) Supongamos que X es un minimizador local del problema (1.1). Entonces X satisface las condiciones de optimalidad de primer orden* (4.2a) *y* (4.2b)

con multiplicador de Lagrange asociado $\Lambda = G^\top X$. *Definamos,*

$$\nabla \mathcal{F}(X) := G - XG^\top X \quad \textit{and} \quad A := GX^\top - XG^\top,$$

entonces $\nabla \mathcal{F}(X) = AX$. *Además,* $\nabla \mathcal{F} = \boldsymbol{0}$ *si y solo si* $A = \boldsymbol{0}$.

Demostración.
Ver anexos.□

Nótese que el lema 2 nos provee una equivalencia de la condición (4.2a), es decir, si suponemos que X es factible y que además satisface que $\nabla \mathcal{F}(X) = \mathbf{0}$ entonces la condición (4.2a) también se satisface, por otro lado, si suponemos que las condiciones (4.2a) y (4.2b) se cumplen para algun X entonces $\nabla \mathcal{F}(X) = \mathbf{0}$. Este hecho nos permite utilizar como criterio de parada para nuestros algoritmos la condición $\nabla \mathcal{F}(X) = \mathbf{0}$, es decir nos detenemos cuando encontremos un $X \in \mathrm{St}(n,p)$ tal que $\nabla \mathcal{F}(X) = \mathbf{0}$, de esta forma se garantiza que se satisfagan las condiciones (4.2a)-(4.2b). Observe que la condición (4.2b) también se va a satisfacer debido a que nuestros algoritmos construyen una secuencia factible y $\mathrm{St}(n,p)$ es un conjunto compacto.

Próximamente, estableceremos las condiciones de optimalidad de segundo orden para el problema (1.1):

Lema 3 *1) (Condición necesaria de segundo orden, [30]) Supongamos que* $X \in St(n,p)$ *es un minimizador local del problema* (1.1). *Entonces* X *satisface:*

$$Tr[\, Z^\top \mathcal{D}(\mathcal{D}\mathcal{F}(X))[Z]\,] - Tr[\, \Lambda Z^\top Z\,] \geq 0, \quad \forall Z \in T_X St(n,p), \quad \textit{donde } \Lambda = G^\top X. \tag{4.3}$$

2) (Condición suficiente de segundo orden, [30]) Supongamos que para $X \in St(n,p)$, *existe un multiplicador de Lagrange* Λ *tal que se satisfacen las condiciones de optimalidad de primer orden. Además supongamos que*

$$Tr[\, Z^\top \mathcal{D}(\mathcal{D}\mathcal{F}(X))[Z]\,] - Tr[\, \Lambda Z^\top Z\,] > 0, \tag{4.4}$$

para cualquier matriz $Z \in T_X St(n,p)$. *Entonces* X *es un minimizador local estricto para el problema* (1.1).

4.2. Esquemas de actualización

En las siguientes subsecciones presentamos las fórmulas de actualización para cuatro de nuestros métodos. La primera se construye a partir de una combinación lineal del iterado actual con otro término, las siguientes dos fórmulas de actualización están inspiradas en las fórmulas de los métodos para resolver ecuaciones diferenciales numéricamente *Adams-Bashforth* y *Adams-Moulton* que fueron presentadas en el capítulo 2, la idea de estudiar estas últimas dos propuesta surgió de hacer una analogía a la investigación presentada por Wen-Yin en [21] donde el esquema de actualización del método propuesto por dichos autores (ver (3.6)) esta inspirada en la fórmula del método de *Crank-Nicolson* (ver capítulo 2).

Por último en esta sección presentamos la fórmula de actualización de nuestro cuarto método propuesto que emplea una factorización tipo Cholesky de forma inteligente para garantizar factibilidad sin necesidad de utilizar un operador de proyección sobre la variedad de Stiefel, dicha

fórmula de actualización generaliza el método propuesto por Wen-Yin en [21], que actualiza los iterados usando la transformada de Cayley, ver (3.7).

Los cuatro métodos que describiremos en esta sección emplearan el siguiente esquema general de actualización:

$$X_{k+1} = Z_k(\tau), \tag{4.5}$$

donde $\tau > 0$ representa el *tamaño de paso*, y $Z_k : \mathbb{R}^{n\times p} \times \mathbb{R} \to \mathtt{St}(n,p)$ es una función cuyos que depende del iterado anterior X_k y del *tamaño de paso*. Cada uno de nuestros cuatro métodos utiliza una función $Z_k(\cdot)$ diferente.

4.2.1. Esquema de actualización basado en una combinación lineal

Nuestra primera propuesta utiliza la siguiente fórmula de actualización:

$$Y_k^{CL}(\tau) := X_k - \tau\,(\,\lambda B_k L + \mu C_k R\,), \tag{4.6}$$

donde $B_k = G_k L^\top - L G_k^\top$, $C_k = G_k R^\top - R G_k^\top$, $L, R \in \mathbb{R}^{n\times p}$, τ es el *tamaño de paso* y donde λ y μ son escalares que satisfacen:

$$\lambda ||B_k||_F^2 + \mu ||C_k||_F^2 \;>\; 0.$$

El siguiente lema muestra que la curva $Y_k^{CL}(\tau)$ definida en (4.6) es una curva de descenso en $\tau = 0$.

Lema 4 *$Y_k^{CL}(\tau)$ definido en (4.6) es una curva de descenso en $\tau = 0$, esto es:*

$$\mathcal{F}'_\tau(\,Y_k^{CL}(0)\,) := \frac{\partial \mathcal{F}(\,Y_k^{CL}(\tau)\,)}{\partial \tau}\Big|_{\tau=0} = -\frac{\lambda}{2}||B_k||_F^2 - \frac{\mu}{2}||C_k||_F^2 < 0. \tag{4.7}$$

Demostración.
Del teorema 10 sabemos que la diferencial la podemos escribir como:

$$\mathcal{F}'_\tau(\,Y_k^{CL}(0)\,) = \langle \mathcal{D}\mathcal{F}(X_k), \dot{Y}_k^{CL}(0)\rangle,$$

por notacion sabemos que $G_k = \mathcal{D}\mathcal{F}(X_k)$ y además $\dot{Y}_k^{CL}(0) = -(\,\lambda B_k L + \mu C_k R\,)$, luego,

$$\mathcal{F}'_\tau(\,Y_k^{CL}(0)\,) = -Tr[\,G_k^\top(\,\lambda B_k L + \mu C_k R\,)\,],$$

o equivalentemente,

$$\mathcal{F}'_\tau(\,Y_k^{CL}(0)\,) = -\lambda Tr[\,G_k^\top B_k L\,] - \mu Tr[\,G_k^\top C_k R],$$

luego por (2.3) obtenemos,

$$\mathcal{F}'_\tau(\,Y_k^{CL}(0)\,) = -\frac{\lambda}{2}||B_k||_F^2 - \frac{\mu}{2}||C_k||_F^2 \;<\; 0,$$

por lo tanto, $Y_k^{CL}(\tau)$ es una curva de descenso en $\tau = 0$. □

Nota 7 *Observe que en la fórmula de actualización* (4.6)*, se pueden utilizar inclusive las matrices L y R construidas de forma aleatoria junto con los parámetros μ y λ seleccionados de forma apropiada (por ejemplo ambos positivos); y de esta forma se garantiza que el método haga descender la función objetivo y eventualmente puede converger a un mínimo local. Sin embargo, lo más adecuado es escoger las matrices L y R de tal forma que estén relacionadas con el problema, es decir que contengan cierta información del problema, en general se deberían comportar mejor aquellas escogencia de estas matrices que estén relacionadas con direcciones que lleven la información del gradiente o del hessiano de la función objetivo en el iterado actual. En vista de esto, para la implementación de este método, tomaremos $L = X_k$ y $R = X_{k-1}$ con $\lambda = 2/3$ y $\mu = 1/3$, es decir la dirección de descenso es una combinación lineal del gradiente en la iteración actual $\nabla\mathcal{F}(X_k) = B_k L = B_k X_k$ y un término que se aproxima al gradiente de la iteración anterior si τ es pequeño. Escogeremos los parámetros λ y μ de tal forma que se le de mayor peso al gradiente en la iteración actual en lugar de la otra dirección. Es decir que para efectos de la implementación proponemos utilizar la siguiente fórmula de actualización:*

$$Y_k^{CL}(\tau) := X_k - \tau\,(\,\lambda A_k X_k + \mu B_k X_{k-1}\,), \tag{4.8}$$

donde $A_k = G_k X_k^\top - X_k G_k^\top$ *y* $B_k = G_k X_{k-1}^\top - X_{k-1} G_k^\top$.

4.2.2. Los esquemas Adams-Bashforth y Adams-Moulton

Nuestra segunda propuesta esta inspirada en el método de *Adams-Bashforth* el cual fue presentado en el capítulo 2, esta propuesta actualiza el nuevo iterado utilizando los dos iterados previos, más específicamente:

$$Y_k^{AB}(\tau) := X_k - \frac{\tau}{2} A_k\,(\,3X_k - X_{k-1}), \tag{4.9}$$

donde A_k esta definido igual que en nuestro esquema basado en una combinación lineal.

Por otro lado, nuestra tercera propuesta esta inspirada en el método de *Adams-Moulton* el cual también fue presentado en el capítulo 2, esta propuesta actualiza el siguiente iterado como sigue:

$$Y_k^{AM}(\tau) := X_k - \frac{\tau}{12} A_k\,(\,5Y_k^{AM}(\tau) + 8X_k - X_{k-1}\,). \tag{4.10}$$

En nuestras tres primeras propuestas, los esquemas tipo *Adams-Bashforth*, tipo *Adams-Moulton* y el esquema (4.6) construyen puntos $Y_k^{AB}(\tau)$, $Y_k^{AM}(\tau)$, $Y_k^{CL}(\tau)$ que no necesariamente viven sobre la variedad de Stiefel, así que para garantizar que estos métodos construyan una secuencia factible, proyectaremos estos puntos sobre $\mathtt{St}(n,p)$, es decir que para dichos métodos, emplearemos el esquema general de actualización del siguiente iterado (4.5) con $Z_k(\cdot)$ definida como:

$$Z_k(\tau) := \pi(Y_k(\tau)), \tag{4.11}$$

donde $Y_k(\tau)$ es cualquiera de las tres actualizaciones (4.6), (4.9) o (4.10) y donde $\pi(\cdot)$ es el operador de proyección sobre la variedad de Stiefel que presentamos en la definición 26 (ver capítulo 2).

Note que en la ecuación (4.10), $Y_k^{AM}(\tau)$ esta definida en forma implícita, para implementar esta fórmula necesitamos un esquema explícito, el siguiente lema resuelve este problema y además nos provee una expresión cerrada para calcular la derivada de la curva $Y_k^{AM}(\tau)$.

Lema 5 *1) Sea $W \in \mathbb{R}^{n\times n}$ una matriz anti-simétrica, entonces la matriz $Q = (I+W)^{-1}$ esta bien definida.*
2) $Y_k(\tau)$ definida como en (4.10) puede escribirse como:

$$Y_k^{AM}(\tau) = (I + \frac{5\tau}{12}A_k)^{-1}(X_k - \frac{\tau}{12}A_k\,(\,8X_k - X_{k-1}\,)\,), \tag{4.12}$$

3) y su derivada con respecto a τ es:

$$\dot{Y}_k^{AM}(\tau) = -\frac{1}{12}(\,I + \frac{5\tau}{12}A_k\,)^{-1}A_k\,(\,5Y_k^{AM}(\tau) + 8X_k - X_{k-1}\,), \tag{4.13}$$

en particular, $\dot{Y}_k^{AM}(0) = -\frac{1}{12}A_k\,(\,13X_k - X_{k-1}\,)$.

Demostración.
1) Supongamos que W es anti-simétrica, veamos que la matriz $Q = (I+W)^{-1}$ es invertible. Sea $v \in \mathbb{R}^n$, como $v^\top Wv \in \mathbb{R}^n$ entonces

$$\begin{aligned} v^\top Wv &= (\,v^\top Wv\,)^\top \\ &= v^\top W^\top v \\ &= -v^\top Wv \quad (\text{ya que } W^\top = -W\,), \end{aligned}$$

así, $v^\top Wv = 0,\ \forall v \in \mathbb{R}^n$. Ahora dado $v \in \mathbb{R}^n$,

$$v^\top(\,I+W\,)v = v^\top v + v^\top Wv = v^\top v = ||v||_2^2,$$

lo cual implica que $(\,I+W\,)$ es una matriz definida positiva y por lo tanto $Q = (\,I+W\,)$ es invertible. □.

2) Del esquema de actualización (4.10) tenemos que:

$$Y_k^{AM}(\tau) = X_k - \frac{\tau}{12}A_k\,(\,5Y_k^{AM}(\tau) + 8X_k - X_{k-1}\,),$$

o equivalentemente,

$$(\,I + \frac{5\tau}{12}A_k\,)\,Y_k^{AM}(\tau) = X_k - \frac{\tau}{12}A_k\,(\,8X_k - X_{k-1}\,),$$

por lo tanto,

$$Y_k^{AM}(\tau) = (\,I + \frac{5\tau}{12}A_k\,)^{-1}(X_k - \frac{\tau}{12}A_k\,(\,8X_k - X_{k-1}\,)\,).\ \square$$

3) Derivando la curva $Y_k^{AM}(\tau)$ definida en (4.10) con respecto a τ tenemos que:

$$\dot{Y}_k^{AM}(\tau) = -\frac{5}{12}A_kY_k^{AM}(\tau) - \frac{5\tau}{12}A_k\dot{Y}_k^{AM}(\tau) - \frac{1}{12}A_k(\,8X_k - X_{k-1}\,),$$

o equivalentemente,

$$(\,I + \frac{5\tau}{12}A_k\,)\,\dot{Y}_k^{AM}(\tau) = -\frac{1}{12}A_k(\,5Y_k^{AM}(\tau) + 8X_k - X_{k-1}\,),$$

luego,

$$\dot{Y}_k^{AM}(\tau) \;=\; -\frac{1}{12}(\, I + \frac{5\tau}{12}A_k\,)^{-1}\, A_k\,(\,5Y_k^{AM}(\tau) + 8X_k - X_{k-1}\,). \tag{4.14}$$

En particular, cuando $\tau = 0$ tenemos en (4.10) que $Y_k^{AM}(0) = X_k$ luego, sustituyendo $\tau = 0$ en (4.14) tenemos finalmente,

$$\dot{Y}_k^{AM}(0) = -\frac{1}{12}A_k\,(\,13\,X_k - X_{k-1}\,).$$

Con lo cual hemos probado el lema 5.□

Lema 6 *Sean $U = [G_k, X_k]$ y $V = [X_k, -G_k]$. Si la matriz $I + \frac{5\tau}{12}V^\top U$ es invertible entonces (4.12) es equivalente a:*

$$Y_k^{AM}(\tau) = X_k - W(\tau)\,(\,13X_k - X_{k-1}\,) \tag{4.15}$$

donde $W(\tau) = \frac{\tau}{12}U(\,I + \frac{5\tau}{12}V^\top U\,)^{-1}V^\top$.

Demostración.
Nótese que de la definición de U, V y A_k tenemos que $I + \frac{5\tau}{12}A_k = I + \frac{5\tau}{12}UV^\top$, si aplicamos la fórmula SMW:

$$(B + \alpha UV^\top)^{-1} = B^{-1} - \alpha B^{-1}U(I + \alpha V^\top B^{-1}U)^{-1}V^\top B^{-1}, \tag{4.16}$$

con $B = I$ y $\alpha = \frac{5\tau}{12}$ obtenemos:

$$(\,I + \frac{5\tau}{12}A_k\,)^{-1} = I - \frac{5\tau}{12}U(\,I + \frac{5\tau}{12}V^\top U\,)^{-1}V^\top,$$

además, nótese que $(I - \frac{8\tau}{12}A_k)X_k = (I - \frac{8\tau}{12}UV^\top)X_k$ así, si denotamos por $T_1 = (I + \frac{5\tau}{12}A_k)^{-1}(I - \frac{8\tau}{12}A_k)X_k$ tenemos que:

$$\begin{aligned}
T_1 &= [\,I - \frac{5\tau}{12}U(\,I + \frac{5\tau}{12}V^\top U\,)^{-1}V^\top\,]\,(\,I - \frac{8\tau}{12}UV^\top\,)X_k \\
&= X_k - \frac{5\tau}{12}U\Big(\,\frac{8}{5}I + (\,I + \frac{5\tau}{12}V^\top U\,)^{-1}(\,I - \frac{8\tau}{12}V^\top U\,)\Big)V^\top X_k \\
&= X_k - \frac{5\tau}{12}U(\,I + \frac{5\tau}{12}V^\top U\,)^{-1}\Big[\frac{8}{5}(\,I + \frac{5\tau}{12}V^\top U\,) + (\,I - \frac{8\tau}{12}V^\top U\,)\Big]V^\top X_k \\
&= X_k - \frac{5\tau}{12}U(\,I + \frac{5\tau}{12}V^\top U\,)^{-1}(\,\frac{13}{5}I\,)V^\top X_k \\
&= X_k - \frac{13\tau}{12}U(\,I + \frac{5\tau}{12}V^\top U\,)^{-1}V^\top X_k \\
&= X_k - 13\,W(\tau)X_k,
\end{aligned} \tag{4.17}$$

$T_2 = \frac{\tau}{12}(\, I + \frac{5\tau}{12}A\,)^{-1}(\, AX_{k-1}\,)$ obtenemos:

$$\begin{aligned}
T_2 &= \frac{\tau}{12}[\, I - \frac{5\tau}{12}U(\, I + \frac{5\tau}{12}V^\top U\,)^{-1}V^\top\,](\, UV^\top X_{k-1}\,) \\
&= \frac{\tau}{12}UV^\top X_{k-1} - \frac{5\tau^2}{12^2}U(\, I + \frac{5\tau}{12}V^\top U\,)^{-1}V^\top UV^\top X_{k-1} \\
&= \frac{\tau}{12}U\Big(\, I - \frac{5\tau}{12}(\, I + \frac{5\tau}{12}V^\top U\,)^{-1}V^\top U\,\Big)V^\top X_{k-1} \\
&= \frac{\tau}{12}U(\, I + \frac{5\tau}{12}V^\top U\,)^{-1}\Big[\,(\, I + \frac{5\tau}{12}V^\top U\,) - \frac{5\tau}{12}V^\top U\,\Big]V^\top X_{k-1} \\
&= \frac{\tau}{12}U(\, I + \frac{5\tau}{12}V^\top U\,)^{-1}V^\top X_{k-1} \\
&= W(\tau)X_{k-1}, \qquad (4.18)
\end{aligned}$$

Ahora, del lema 5 sabemos que,

$$\begin{aligned}
Y_k^{AM}(\tau) &= (\, I + \frac{5\tau}{12}A_k\,)^{-1}\Big(\, X_k - \frac{\tau}{12}A_k(\, 8X_k - X_{k-1}\,)\,\Big) \\
&= (\, I + \frac{5\tau}{12}A_k\,)^{-1}(\, I - \frac{8\tau}{12}A_k\,)X_k + \frac{\tau}{12}(\, I + \frac{5\tau}{12}A_k\,)^{-1}A_kX_{k-1} \\
&= T_1 + T_2 \\
&= (\, X_k - 13W(\tau)X_k\,) + (\, W(\tau)X_{k-1}\,) \quad (\text{por}(4.17)\text{ y }(4.18)) \\
&= X_k - W(\tau)(13X_k - X_{k-1}),
\end{aligned}$$

lo cual prueba el lema. □

Observación 1 *El lema anterior nos provee otra forma de calcular $Y_k^{AM}(\tau)$, para el caso cuando $p \ll n$ invertir la matriz $(I + \frac{5\tau}{12}V^\top U) \in \mathbb{R}^{2p\times 2p}$ es mucho menos costoso computacionalmente que invertir la matriz $(I + \frac{5\tau}{12}A_k) \in \mathbb{R}^{n\times n}$, así que en este caso conviene usar la fórmula de actualización* (4.15)*. Más específicamente, para efectos de la implementación de nuestros Algoritmos 2 y 3 (dichos algoritmos son presentados más adelante en este mismo capítulo) cuando utilicemos nuestro método basado en* Adams-Moulton *usaremos la fórmula de actualización* (4.10) *para el caso cuando $p \geq \frac{n}{2}$ y en caso contrario emplearemos la fórmula de actualización* (4.15)*, con la finalidad de reducir el cómputo de invertir la matriz $I + \frac{5\tau}{12}A_k$.*

Observación 2 *En las fórmulas de actualización* (4.9) *y* (4.10) *podemos introducir parámetros calculados de forma adecuada para garantizar que las direcciones utilizadas por estos métodos sean de descenso en X_k, es decir podemos modificar dichas fórmulas de actualización por:*

$$Y_k^{AB}(\tau) := X_k - \frac{\tau}{2}A_k\,(\, 3\alpha_k X_k - X_{k-1}) \qquad (4.19)$$

$$Y_k^{AM}(\tau) := X_k - \frac{\tau}{12}A_k\,(\, 5\beta_k Y_k^{AM}(\tau) + 8X_k - X_{k-1}\,) \qquad (4.20)$$

donde los parámetros α_k y β_k los calculamos como:

$$\alpha_k = \begin{cases} 1 + (\, 2Tr[G_k^\top A_k X_{k-1}]\,)\,/\,3||A_k||_F^2 & si\ \mathcal{DF}(X_k)[\dot{Y}_k^{AB}(0)] \geq 0 \\ 1 & en\ otro\ caso. \end{cases}$$

y

$$\beta_k = \begin{cases} 2Tr[G_k^\top A_k X_{k-1}]\,/\,5||A_k||_F^2 & si\ \mathcal{DF}(X_k)[\dot{Y}_k^{AM}(0)] \geq 0 \\ 1 & en\ otro\ caso. \end{cases}$$

y de esta forma se puede probar que las curvas (4.19) *y* (4.20) *son curvas de descenso en* $\tau = 0$. *En efecto, consideremos la fórmula de actualización* (4.19), *si* $\mathcal{DF}(X_k)[\dot{Y}_k^{AB}(0)] < 0$ *no hay nada que probar y en este caso se toma* $\alpha_k = 1$ *y así la fórmula de actualización* (4.19) *coincide con* (4.9), *en caso contrario se selecciona* $\alpha_k = 1 + (\, 2Tr[G_k^\top A_k X_{k-1}]\,) \,/\, 3||A_k||_F^2$ *y en este caso tenemos que:*

$$\begin{aligned}
\mathcal{DF}(X_k)[\dot{Y}_k^{AB}(0)] &= Tr[G_k^\top \dot{Y}_k^{AB}(0)] \\
&= Tr[G_k^\top (\, \frac{-3\alpha_k}{2} A_k X_k + \frac{1}{2} A_k X_{k-1}\,)] \\
&= -\frac{3\alpha_k}{2} Tr[G_k^\top A_k X_k] + \frac{1}{2} Tr[G_k^\top A_k X_{k-1}] \\
&= -\frac{3\alpha_k}{4} ||A_k||_F^2 + \frac{1}{2} Tr[G_k^\top A_k X_{k-1}] \quad (\ por\ (2.3)\ ) \\
&= -\frac{3}{4}(1 + \frac{2}{3}\frac{Tr[G_k^\top A_k X_{k-1}]}{||A_k||_F^2}) ||A_k||_F^2 + \frac{1}{2} Tr[G_k^\top A_k X_{k-1}] \\
&= -\frac{3}{4} ||A_k||_F^2 - \frac{1}{2} Tr[G_k^\top A_k X_{k-1}] + \frac{1}{2} Tr[G_k^\top A_k X_{k-1}] \\
&= -\frac{3}{4} ||A_k||_F^2 \\
&< 0
\end{aligned}$$

y por lo tanto (4.19) *es una curva de descenso en* $\tau = 0$.

Por otro lado consideremos la fórmula de actualización (4.20), *en el caso cuando* $\mathcal{DF}(X_k)[\dot{Y}_k^{AM}(0)] < 0$ *no hay nada que demostrar y en este caso se toma* $\beta_k = 1$ *y así la fórmula de actualización* (4.19) *es exactamente igual a* (4.10), *en caso contrario, el método escoge* $\beta_k = 2Tr[G_k^\top A_k X_{k-1}] \,/\, 5||A_k||_F^2$ *y en este caso tenemos que:*

$$\begin{aligned}
\mathcal{DF}(X_k)[\dot{Y}_k^{AB}(0)] &= Tr[G_k^\top \dot{Y}_k^{AM}(0)] \\
&= Tr[G_k^\top (\, -\frac{5\beta_k + 8}{12} A_k X_k + \frac{1}{12} A_k X_{k-1}\,)] \\
&= -\frac{5\beta_k + 8}{12} Tr[G_k^\top A_k X_k] + \frac{1}{12} Tr[G_k^\top A_k X_{k-1}] \\
&= -\frac{5\beta_k + 8}{24} ||A_k||_F^2 + \frac{1}{12} Tr[G_k^\top A_k X_{k-1}] \quad (\ por\ (2.3)\ ) \\
&= -\frac{1}{3} ||A_k||_F^2 - \frac{5}{24}(\frac{2}{5}\frac{Tr[G_k^\top A_k X_{k-1}]}{||A_k||_F^2}) ||A_k||_F^2 + \frac{1}{12} Tr[G_k^\top A_k X_{k-1}] \\
&= -\frac{1}{3} ||A_k||_F^2 - \frac{1}{12} Tr[G_k^\top A_k X_{k-1}] + \frac{1}{12} Tr[G_k^\top A_k X_{k-1}] \\
&= -\frac{1}{3} ||A_k||_F^2 \\
&< 0
\end{aligned}$$

y por lo tanto (4.20) *es una curva de descenso en* $\tau = 0$.

En nuestra observación del comportamiento numérico de ambos métodos hemos visto que dichos métodos siempre seleccionan $\alpha_k = \beta_k = 1$ *en todas las iteraciones, lo cual puede ser evidencia de que en realidad las direcciones utilizadas por las fórmulas de actualización* (4.9) *y* (4.10) *son de descenso en* X_k, *sin embargo no hemos podido demostrar esto y se dejará como trabajo futuro.*

4.2.3. Método basado en la factorización de Cholesky

Antes de presentar la fórmula de actualización de nuestro cuarto método introducimos la siguiente definición:

Definición 29 *Denotemos por $\mathcal{C}^\infty(\mathcal{S}_{Skew(n)})$ al conjunto de funciones infinitamente continuamente diferenciables que van del conjunto de las matrices anti-simétricas de orden n al conjunto $\mathbb{R}^{n\times n}$. Ahora definimos el siguiente subconjunto de $\mathcal{C}^\infty(\mathcal{S}_{Skew(n)})$ por:*

$$\Phi := \{L \in \mathcal{C}^\infty(\mathcal{S}_{Skew(n)}) : L(\mathbf{0}) = I_n, \quad L(\Omega)^\top L(\Omega) = I_n - \Omega^2, \, \forall \Omega \in \mathcal{S}_{Skew(n)}\}. \tag{4.21}$$

Ejemplo 4 *. A continuación mostraremos algunos ejemplos de funciones que pertenecen a Φ, definido en (4.21). Se puede verificar fácilmente que las siguientes funciones $L_i(\cdot)$ pertenecen al conjunto Φ.*

a)

$$L_1(\Omega) = I + \Omega, \ \forall \Omega \in \mathcal{S}_{Skew(n)}, \tag{4.22}$$

b) *Además, L_1 no es única existen diferentes elecciones de funciones $L \in \Phi$:*

$$L_2(\Omega) = R(\Omega), \ \forall \Omega \in \mathcal{S}_{Skew(n)}. \tag{4.23}$$

donde $R(\Omega)$ es la matriz obtenida de la factorización QR de la matriz $(I + \Omega)$, es decir $Q(\Omega)R(\Omega) = I + \Omega$.

c)

$$L_3(\Omega) = (I - \Omega^2)^{\frac{1}{2}}, \ \forall \Omega \in \mathcal{S}_{Skew(n)}. \tag{4.24}$$

dada por la descomposición polar.

d) *También la matriz triangular R obtenida mediante la factorización tipo Cholesky de la matriz $(I - \Omega^2)$ dada por:*

$$L_4(\Omega) \ = \ R(\Omega), \tag{4.25}$$

con $R(\Omega)$ siendo la matriz triangular obtenida de la factorización QR de la matriz $\Sigma(\Omega)^{1/2}U(\Omega)^\top$, donde las matrices $\Sigma(\Omega)$ y $U(\Omega)$ son obtenidas a través de la SVD de la matriz $I_n - \Omega^2$. Veamos que esta elección (4.25) pertenece al conjunto Φ, en efecto, supongamos que $I_n - \Omega^2 = U(\Omega)\Sigma(\Omega)V(\Omega)^\top$ es una descomposición de valores singulares de la matriz $I_n - \Omega^2$, entonces como $I_n - \Omega^2$ es simétrica y definida positiva tenemos que $U(\Omega) = V(\Omega), \ \forall \Omega \in \mathcal{S}_{Skew(n)}$, luego consideremos la factorización QR de la matriz $\Sigma(\Omega)^{1/2}U(\Omega)^\top = Q(\Omega)R(\Omega)$, así $R(\Omega) = Q(\Omega)^\top\Sigma(\Omega)^{1/2}U(\Omega)^\top$, luego:

$$\begin{aligned} R^\top(\Omega)R(\Omega) &= (\, U(\Omega)(\Sigma(\Omega)^{1/2})^\top Q(\Omega)\,)(\, Q(\Omega)^\top\Sigma(\Omega)^{1/2}U(\Omega)^\top\,), \\ &= U(\Omega)\Sigma(\Omega)U(\Omega)^\top, \\ &= I_n - \Omega^2, \ \forall \Omega \in \mathcal{S}_{Skew(n)}, \end{aligned}$$

por lo tanto $R(\Omega)$ es una factorización de Cholesky de la matriz $I_n - \Omega^2$, además $R(\cdot) \in C^\infty(\mathcal{S}_{Skew(n)})$ puesto que la que la factorización QR puede obtenerse mediante un proceso C^∞ vía Gram-Schmidt *y la descomposición SVD también es un proceso C^∞, observe también que $R(\boldsymbol{0}) = \boldsymbol{0}$ y por lo tanto concluimos que $R \in \Phi$.*

Observación 3 *Consideremos la factorización clásica de Cholesky $L^\top(\Omega)L(\Omega) = I_n - \Omega^2$, observe que esta función $L(\cdot)$ también satisface que $L(\mathbf{0}) = I_n$ sin embargo no puede garantizarse que dicha función pertenezca a $\mathcal{C}^\infty(\mathcal{S}_{Skew(n)})$. Pese a este inconveniente existen diferentes factorizaciones tipo Cholesky que pueden obtenerse mediante procesos diferenciables tales como $L_2(\cdot)$, $L_4(\cdot)$ entre otras.*

Fórmula de actualización

Nuestro cuarto método propuesto para resolver el problema (1.1) utiliza la siguiente fórmula de actualización

$$Z_k(\tau) := \Gamma(\tau)\,(I + \tau A_k)^{-1}\,X_k, \tag{4.26}$$

donde la curva $\Gamma(\cdot)$ esta dada por $\Gamma(\tau) := L(\tau A_k)$ con $L \in \Phi$. Nótese que a diferencia de las fórmulas de actualización de nuestros tres primeros métodos, la fórmula (4.26) no necesita el uso de un operador de proyección.

En nuestros experimentos (ver capítulo 5), usamos como función $L(\cdot)$ la que corresponde a factorización clásica de Cholesky de la matriz $I - \tau^2 A_k^2$, en lugar de la función $L_4(\cdot)$ mostrada en el ejemplo anterior, no obstante para efectos teóricos de nuestro método supondremos que se utiliza cualquier función $L(\cdot) \in \Phi$.

Aspectos teóricos

El siguiente lema establece que siempre es posible calcular la factorización de Cholesky de la matriz $I - \tau^2 A_k^2$, también establece que si X_k es factible entonces $Z_k(\tau)$ calculado mediante el esquema (4.26) también es factible, además prueba que la derivada direccional de $\mathcal{F}$ en X_k en la dirección de $\dot{Z}_k(0)$ es negativa.

Lema 7 *1) Dada cualquier matriz anti-simétrica $W \in \mathbb{R}^{n\times n}$, entonces, la matriz $P = (I - W^2)$ es definida positiva.*
2) Dada $X_k \in \mathtt{St}(n,p)$ entonces $Z_k(\tau)^\top Z_k(\tau) = I$.
3) Si $\Gamma(\tau)$ es diferenciable entonces la derivada respecto a τ de $Z_k(\tau)$ satisface:

$$\Gamma(\tau)^\top \dot{Z}_k(\tau) \;\; = \;\; -A_k X_k - \dot{\Gamma}(\tau)^\top Z_k(\tau), \tag{4.27}$$

y en particular en $\tau = 0$ tenemos,

$$\dot{Z}_k(0) = -\nabla\mathcal{F}(X_k). \tag{4.28}$$

4) La curva $Z_k(\tau)$ es una curva de descenso en $\tau = 0$, esto es:

$$\mathcal{D}\mathcal{F}(X_k)[\dot{Z}_k(0)] = -\frac{1}{2}||A_k||^2 < 0. \tag{4.29}$$

Demostración.
1) Supongamos que W es anti-simétrica, veamos que la matriz $P = (I - W^2)$ es definida positiva. En efecto, sea $v \in \mathbb{R}^n$ no nulo, entonces:

$$v^\top P v = v^\top (I - W^2)v = ||v||_2^2 - v^\top W^2 v = ||v||_2^2 + ||Wv||_2^2 > 0,$$

por lo tanto $P = (I - W^2)$ es definida positiva. $\square$

2) Tomemos $X_k \in \mathtt{St}(n,p)$ arbitraria, entonces:

$$\begin{aligned} Z_k(\tau)^\top Z_k(\tau) &= X_k^\top (I-\tau A_k)^{-1}\Gamma(\tau)^\top\Gamma(\tau)(I+\tau A_k)^{-1}X_k \\ &= X_k^\top (I-\tau A_k)^{-1}(I-\tau^2 A_k^2)(I+\tau A_k)^{-1}X_k \\ &= X_k^\top (I-\tau A_k)^{-1}[(I-\tau A_k)(I+\tau A_k)](I+\tau A_k)^{-1}X_k \\ &= X_k^\top X_k \\ &= I. \square \end{aligned} \tag{4.30}$$

3) Nótese que $\Gamma(0) = I$, además como $\Gamma(\tau)^\top\Gamma(\tau) = I - \tau^2 A_k^2$ entonces derivando esta ecuación con respecto a τ obtenemos que $\dot\Gamma(\tau)^\top\Gamma(\tau) + \Gamma(\tau)^\top\dot\Gamma(\tau) = -2\tau A_k^2$ y en particular para $\tau = 0$ tenemos que $\dot\Gamma(0)^\top + \dot\Gamma(0) = \mathbf{0}$, como $\dot\Gamma(\tau)$ es triangular superior (suponiendo que se utiliza una $L \in \Phi$ obtenida mediante una factorización tipo Cholesky) entonces se concluye que $\dot\Gamma(0) = \mathbf{0}$. Por otra lado, de la definición de $Z_k(\tau)$ tenemos:

$$\Gamma(\tau)^\top Z_k(\tau) = (I - \tau A_k)X_k,$$

derivando con respecto de τ esta última ecuación obtenemos,

$$\dot\Gamma(\tau)^\top Z_k(\tau) + \Gamma(\tau)^\top \dot Z_k(\tau) = -A_k X_k$$

o equivalentemente,

$$\Gamma(\tau)^\top \dot Z_k(\tau) = -A_k X_k - \dot\Gamma(\tau)^\top Z_k(\tau) \tag{4.31}$$

en particular para $\tau = 0$ de la ecuación (4.31) y el hecho de que $\dot\Gamma(0) = \mathbf{0}$ tenemos que:

$$\dot Z_k(0) = -A_k X_k = -\nabla\mathcal{F}(X_k). \square \tag{4.32}$$

4) Consideremos la diferencial de $\mathcal{F}$ en X_k en dirección de $\dot Z_k(0)$:

$$\begin{aligned} \mathcal{D}\mathcal{F}(X_k)[\dot Z_k(0)] &= Tr[G_k^\top \dot Z_k(0)], \quad \text{(por teorema 10)} \\ &= -Tr[G_k^\top A_k X_k] \quad \text{(por (4.32))} \\ &= -\frac{1}{2}||A_k||_F^2 \quad \text{(por (2.3))} \\ &< 0. \square \end{aligned}$$

Con lo cual queda demostrado el lema.$\square$

Proposición 6 *Sea $X \in \mathcal{O}(n)$, definamos la aplicación $R_X : T_X(\mathcal{O}(n)) \to \mathcal{O}(n)$ por:*

$$R_X(\xi = X\Omega) = XL(\Omega)(I-\Omega)^{-1} \tag{4.33}$$

donde $L \in \Phi$. Además supongamos que las curvas definidas por $\Gamma(t) := L(t\Omega), \forall\Omega \in \mathcal{S}_{Skew(n)}$ satisfacen que $\dot\Gamma(0) = \mathbf{0}$, entonces $R_X(\xi)$ es una retracción en X sobre el grupo ortogonal.

Demostración.
Solo debemos probar que las dos condiciones de la definición 27 se cumplen para (4.33) debido

a que claramente $R_X(\xi)$ es suave por ser un producto de funciones C^∞.

Nótese que $\mathbf{0} = 0_X \in T_X(\,\mathcal{O}(n)\,)$ es el cero del espacio tangente en X del grupo ortogonal, como $L(\mathbf{0}) = I$ entonces:

$$R_X(0_X) = XL(\mathbf{0})(I - \mathbf{0})^{-1} = X \tag{4.34}$$

por otro lado, sea $\xi = X\Omega \in T_X(\,\mathcal{O}(n)\,)$ arbitrario.

$$\begin{aligned}
\mathcal{D}R_X(0_X)[\xi] &= \lim_{t\to 0} \frac{R_X(0_X + t\xi) - R_X(0_X)}{t} \\
&= \lim_{t\to 0} \frac{XL(t\Omega)(I - t\Omega)^{-1} - X}{t} \\
&= X(\lim_{t\to 0} \frac{L(t\Omega)(I - t\Omega)^{-1} - I}{t}\,)
\end{aligned}$$

$$\begin{aligned}
&= X(\lim_{t\to 0} \frac{L(t\Omega)(I + t\Omega)(L(t\Omega)^\top L(t\Omega))^{-1} - I}{t}\,) \\
&= X(\lim_{t\to 0} \frac{[L(t\Omega)(I + t\Omega) - L(t\Omega)^\top L(t\Omega)](L(t\Omega)^\top L(t\Omega))^{-1}}{t}\,) \\
&= X(\lim_{t\to 0} \frac{[L(t\Omega) + tL(t\Omega)\Omega - L(t\Omega)^\top L(t\Omega)](L(t\Omega)^\top L(t\Omega))^{-1}}{t}\,) \\
&= X(\lim_{t\to 0} \frac{[(\,L(t\Omega) - L(t\Omega)^\top L(t\Omega)\,) + tL(t\Omega)\Omega](L(t\Omega)^\top L(t\Omega))^{-1}}{t}\,) \\
&= X(\lim_{t\to 0} \Big\{ \frac{[(I - L(t\Omega)^\top)L(t\Omega)](L(t\Omega)^\top L(t\Omega))^{-1}}{t} + L(t\Omega)\Omega(L(t\Omega)^\top L(t\Omega))^{-1} \Big\}\,)
\end{aligned} \tag{4.35}$$

como $\lim_{t\to 0} L(t\Omega) = \lim_{t\to 0} L(t\Omega)^\top = I$ entonces tenemos que:

$$\lim_{t\to 0} L(t\Omega)\Omega(L(t\Omega)^\top L(t\Omega))^{-1} = \Omega \tag{4.36}$$

$$\lim_{t\to 0} L(t\Omega)(L(t\Omega)^\top L(t\Omega))^{-1} = I \tag{4.37}$$

además, si definimos $f(t) = I - h_\Omega(t)^\top$ y $g(t) = t$ tenemos que:

$$\begin{aligned}
\lim_{t\to 0} \frac{I - L(t\Omega)^\top}{t} &= \lim_{t\to 0} \frac{f(t)}{g(t)} \\
&= \lim_{t\to 0} \frac{f'(t)}{g'(t)} \qquad (4.38) \\
&= \lim_{t\to 0} f'(t) \\
&= \mathbf{0} \quad (\text{ ya que } \dot\Gamma(0) = \mathbf{0}\,) \qquad (4.39)
\end{aligned}$$

donde la segunda igualdad (4.38) se obtiene aplicando la Regla de L'Hopital componente a componente. Luego, de (4.35), (4.36), (4.37) y (4.39) obtenemos:

$$\mathcal{D}R_X(0_X)[\xi] = X\Omega = \xi = \mathrm{id}_{T_X(\mathcal{O}(n))}[\xi]$$

como ξ es arbitrario entonces tenemos que:

$$\mathcal{D}R_X(0_X) = \mathrm{id}_{T_X(\mathcal{O}(n))} \tag{4.40}$$

por lo tanto de (4.34) y (4.40) se concluye que $R_X(X\Omega)$ es una retracción de X sobre el grupo ortogonal.□

Nota 8 *Nótese que al construir las curvas* $\Gamma(t) := L(t\Omega), \forall \Omega \in \mathcal{S}_{Skew(n)}$ *donde la función* $L(\cdot) \in \Phi$ *es obtenida mediante la factorización tipo Cholesky de la matriz* $I_n - \Omega^2, \forall \Omega \in \mathcal{S}_{Skew(n)}$, *entonces se verifica que* $\dot{\Gamma}_\Omega(0) = \boldsymbol{0}$ *esto se obtuvo en la demostración del lema anterior. Por lo tanto nuestra fórmula de actualización es una retracción sobre el grupo ortogonal. Esto implica que para nuestro algoritmo en el caso cuando* $n = p$ *en el problema* (1.1) *se tienen los resultados de convergencia presentados en [19].*

4.3. Estrategias para seleccionar el tamaño de paso

Es conocido que los métodos de búsqueda lineal pueden no converger cuando se utiliza un *tamaño de paso* τ fijo para todas las iteraciones, en vista de esto, conviene seleccionar el *tamaño de paso* de forma que se garantice convergencia. Existen muchos criterios para escoger el *tamaño de paso*, es por esta razón, que en esta sección presentamos tanto un criterio clásico que garantiza descenso de la función objetivo en cada iteración, como también, otro criterio que no obliga el descenso de la función objetivo en cada iteración, los métodos que calculan el *tamaño de paso* de esta manera se conocen como algoritmos no-monótonos.

4.3.1. Condición de Armijo

Los algoritmos monótonos construyen una sucesión $\{X_k\}$ tal que la sucesión $\{\mathcal{F}(X_k)\}$ es monótona decreciente, generalmente estos métodos calculan el *tamaño de paso* τ como la solución del siguiente problema de optimización:

$$\min_{\tau>0} \phi(\tau) := \mathcal{F}(\, Z_k(\tau)\,). \tag{4.41}$$

En la mayoría de los casos el problema de optimización (4.41) no tiene una solución cerrada, lo cual hace que sea necesario estimar el *tamaño de paso* satisfaciendo condiciones que relajen este problema de optimización, pero que a su vez garantice descenso de la función objetivo. Una de las condiciones de descenso más populares en la llamada *condición de suficiente descenso* o también *condición de Armijo*, usando esta condición se escoge el *tamaño de paso* de la iteración k-ésima τ_k como el mayor número real positivo que satisfaga:

$$\mathcal{F}(\, Z_k(\tau_k)\,) \leq \mathcal{F}(X_k) + \rho_1 \tau_k Tr[\mathcal{D}\mathcal{F}(X_k)^\top \dot{Z}_k(0)], \tag{4.42}$$

donde $0 < \rho_1 < 1$.

Esta *condición de Armijo* para seleccionar el *tamaño de paso* de la iteración k-ésima, la emplearemos únicamente para nuestro método que utiliza la fórmula de actualización basada en la factorización de Cholesky (4.26) y la implementaremos por medio de la heurística llamada *backtracking*, para nuestras otras propuestas basadas en proyecciones es decir, las que utilizan las fórmulas de actualización (4.9), (4.10) y (4.6) emplearemos otra condición que será explicada en la siguiente subsección.

Si bien es cierto que existen otras condiciones para escoger el *tamaño de paso*, que también relajan el problema (4.41) tales como *condiciones débiles de Wolfe*, *condiciones fuertes de Wolfe*, *condición de Goldstein*, (ver [30]) entre otras, usaremos la *condición de Armijo* puesto que requiere menos cómputo y además basta con esta condición para hacer el análisis de convergencia del método, usando la fórmula de actualización (4.26).

4.3.2. Una condición de descenso

En vista de la dificultad de calcular la derivada de la curva $Z_k(\tau)$ en $\tau = 0$ (debido al operador de proyección) para nuestros métodos tipo *Adams-Bashforth*, *Adams-Moulton* y el basado en la combinación lineal (4.6), introducimos una nueva condición de descenso como sigue, escogeremos el *tamaño de paso* de la iteración k-ésima τ_k como el mayor número real positivo que satisfaga simultáneamente las siguientes condiciones:

$$\mathcal{F}(Z_k(\tau_k)) \leq \mathcal{F}(X_k) + \sigma\tau_k Tr[\mathcal{D}\mathcal{F}(X_k)^\top \dot{Y}_k(0)], \tag{4.43a}$$

$$Tr[\mathcal{D}\mathcal{F}(Z_k(\tau_k))^\top \dot{Y}_k(0)] \geq c_2\, Tr[\mathcal{D}\mathcal{F}(X_k)^\top \dot{Y}_k(0)], \tag{4.43b}$$

con $0 < \sigma < c_2 < 1$ y donde $\dot{Y}_k(0)$ es la derivada de cualquiera de las curvas $Y_k^{AB}(\tau)$, $Y_k^{AM}(\tau)$ o $Y_k^{CC}(\tau)$ en $\tau = 0$.

Observación 4 *Note que las ecuaciones* (4.43a), (4.43b) *no son exactamente las bien conocidas* condiciones débiles de Wolfe *(ver [30]), puesto que en lugar de* $\dot{Z}_k(0)$ *utilizamos* $\dot{Y}_k(0)$ *y esto no concuerda con las* condiciones débiles de Wolfe*. Por otro lado, si utilizamos únicamente la condición* (4.43a) *con el propósito de calcular el* tamaño de paso*, entonces esta garantiza el descenso de la función objetivo siempre y cuando aseguremos que la derivada direccional* $Tr[\mathcal{D}\mathcal{F}(X_k)^\top \dot{Y}_k(0)]$ *sea negativa y así siempre podremos encontrar un* tamaño de paso *que satisfaga esta condición. Observe que al emplear la expresión* $Tr[G_k \dot{Y}_k(0)]$ *en la nueva condición de descenso, estamos utilizando información de la función objetivo, además empleamos dicha expresión puesto a que esta es más fácil de calcular que* $Tr[G_k^\top \dot{Z}_k(0)]$.

Usando la heurística de *Backtracking* con las condiciones anteriores (4.43a)-(4.43b), no podemos garantizar que exista un *tamaño de paso* que satisfaga ambas condiciones simultáneamente. En este caso, podemos garantizar descenso usando únicamente la condición (4.43a). Así, para la implementación de nuestro algoritmo 2, escogeremos el *tamaño de paso* satisfaciendo (4.43a)-(4.43b) siempre que sea posible y de no ser posible entonces escogeremos dicho *tamaño de paso* satisfaciendo solo (4.43a).

4.3.3. Búsqueda no monótona con tamaño de paso de Barzilai-Borwein

Además del problema (4.41) para calcular τ, existen otros criterios de optimización para escoger el *tamaño de paso*. Entre estos se encuentra un criterio que se fundamenta en la *ecuación de la secante*. Más específicamente, se escoge el $\tau > 0$ que resuelve alguno de los siguientes problemas de optimización:

$$\min_{\tau} ||B(\tau)S_k - R_k||_F, \tag{4.44}$$

o,

$$\min_{\tau} ||S_k - B(\tau)^{-1}R_k||_F, \tag{4.45}$$

donde $S_k = X_{k+1} - X_k$, $R_k = \nabla\mathcal{F}(\,X_{k+1}\,) - \nabla\mathcal{F}(\,X_k\,)$ y donde la matriz $B(\tau) = (\tau I)^{-1}$ es considerada una aproximación del Hessiano de la función objetivo, así, el *tamaño de paso* es obtenido forzando una propiedad Quasi-Newton. Se puede probar fácilmente que las soluciones de los problemas de optimización (4.44)-(4.45) son:

$$\tau_k^{BB1} = \frac{||S_k||_F^2}{Tr[S_k^\top R_k]} \qquad y \qquad \tau_k^{BB2} = \frac{Tr[S_k^\top R_k]}{||R_k||_F^2} \tag{4.46}$$

respectivamente. Para asegurar que τ sea estrictamente positivo, en general, se toma el valor absoluto en alguna o ambas opciones (4.46). A estas selecciones de *tamaños de paso* (4.46) se le conoce como *paso de Barzilai-Borwein* o simplemente como *paso BB*. Cuando el método del descenso del gradiente se combina con el *paso BB* al método se le conoce como *gradiente espectral*, además cuando el método del gradiente proyectado emplea este *tamaño de paso* se le conoce como *gradiente proyectado espectral* estos métodos son estudiados ampliamente en [38, 39, 40].

Es conocido que esta escogencia del *tamaño de paso* BB puede acelerar, en ocasiones, a los métodos de búsqueda lineal (ver [41]). Sin embargo, estos *tamaños de paso* no garantizan el descenso de la función objetivo en cada iteración, lo cual puede implicar que el algoritmo no converja. Para solventar este problema, se suele combinar esta escogencia del *tamaño de paso* con estrategias de globalización no monótonas, las cuales garantizan convergencia al óptimo global bajo ciertas hipótesis (ver [40, 44]). Para uno de nuestros algoritmos (ver *Algoritmo 3*) recurriremos a un método de búsqueda lineal no-monótono que es estudiado en [45].

Más concretamente, usamos *Backtracking* para estimar el *tamaño de paso*, iniciando en τ_k^{BB1} o τ_k^{BB1}. El criterio de paro es

$$\mathcal{F}(Z_k(\tau_k)) \leq C_k + \rho_1 \tau_k Tr[\mathcal{D}\mathcal{F}(X_k)^\top \dot{Z}_k(0)], \tag{4.47}$$

y las τ_k, se actualizan en cada iteración h según $\tau_k = \tau_k^{BB1}\delta^h$ o $\tau_k = \tau_k^{BB2}\delta^h$, donde $0 < \delta < 1$. La $Z_k(\tau)$ es calculada mediante alguna de las ecuaciones (4.11), (4.26); C_{k+1} es calculado mediante una combinación convexa de C_k y $\mathcal{F}(X_{k+1})$ de la siguiente forma: $C_{k+1} = (\eta Q_k C_k + \mathcal{F}(X_{k+1}))/Q_{k+1}$, con $Q_{k+1} = \eta Q_k + 1$ y $Q_0 = 1$, $C_0 = \mathcal{F}(X_0)$.

4.4. Algoritmos de busqueda lineal propuestos para resolver problemas de optimización sobre la variedad de Stiefel

El algoritmo 2, es un algoritmo monótono general que construye una sucesión factible que hace descender la función objetivo en cada iteración, el cual engloba nuestras cuatro primeras propuestas para resolver el problem (1.1), es decir nuestros métodos basados en proyecciones (*Adams-Bashforth*, *Adams-Moulton* y el de la combinación lineal) y nuestro método que emplea la factorización de Cholesky en cada iteración. El *tamaño de paso* para este último método (el basado en la factorización de Cholesky) es calculado usando la técnica de *Backtracking* seleccionando dicho *tamaño de paso* satisfaciendo la *condición de Armijo* ver (4.42), mientras que para nuestros tres métodos basados en proyecciones se escoge el *tamaño de paso* también usando la técnica de *Backtracking* pero seleccionado el *tamaño de paso* más grande que satisfaga las condiciones (4.43a) y (4.43b) simultáneamente o bien que satisfaga únicamente la condición (4.43a).

Algoritmo 2 Algoritmo monótono

Entrada: $X_0 \in \mathtt{St}(n,p)$, $0 < \epsilon < 1$, $\tau_0 > 0$, $X_{-1} = X_0$, k=0.
Salida: X^* el optimizador.

1: **mientras** $||\nabla \mathcal{F}(X_k)||_F > \epsilon$ **hacer**
2: Calcular $G_k = \mathcal{D}\mathcal{F}(X_k)$
3: Calcular $A_k = G_k X_k^\top - X_k G_k^\top$.
4: Calcular $B_k = G_k X_{k-1}^\top - X_{k-1} G_k^\top$ si se desea utilizar la fórmula de actualización (4.8).
5: Calcular el tamaño de paso τ_k satisfaciendo (4.43a) y (4.43b) o satisfaciendo la regla de Armijo (4.42).
6: Calcular $Y_k(\tau_k)$ como en (4.8), (4.19), o (4.20), en caso de que no se desee utilizar el método basado en la factorización de Cholesky.
7: $X_{k+1} = Z_k(\tau_k)$ con $Z_k(\cdot)$ definida como en (4.11) o bien en (4.26).
8: $k = k + 1$
9: **fin mientras**

Nuestro algoritmo 3, es un algoritmo no-monótono el cual construye una secuencia factible utilizando cualquiera de las fórmulas de actualización de nuestras primeras cuatro propuestas, es decir (4.8), (4.19), (4.20) o (4.26) que resuelve el problema que se estudia en el presente trabajo. El *tamaño de paso* se obtiene utilizando la técnica de *Backtracking* combinada con el paso de *Barzilai-Borwein* y con la estrategia de búsqueda no-monótona presentada en la sección previa. Dicho algoritmo es presentado a continuación:

Algoritmo 3 Algoritmo no-monótono con paso BB

Entrada: $X_0 \in \mathtt{St}(n,p)$, $\tau > 0$, $0 < \tau_m < \tau_M$, $\rho_1, \epsilon, \eta, \delta \in (0,1)$, $X_{-1} = X_0$, $C_0 = \mathcal{F}(X_0)$, $Q_0 = 1$, k=0.
Salida: X^* el optimizador.

1: **mientras** $||\nabla \mathcal{F}(X_k)||_F > \epsilon$ **hacer**
2: **mientras** $\mathcal{F}(Z_k(\tau)) \geq C_k + \rho_1 \tau \mathcal{F}_\tau'(Z_k(0))$ **hacer**
3: $\tau = \delta \tau$
4: **fin mientras**
5: $X_{k+1} = Z_k(\tau)$, $Q_{k+1} = \eta Q_k + 1$ y $C_{k+1} = (\eta Q_k C_k + \mathcal{F}(X_{k+1}))/Q_{k+1}$
6: Seleccionar $\tau = |\tau_k^{BB1}|$ o bien $\tau = |\tau_k^{BB2}|$, donde τ_k^{BB1} y τ_k^{BB2} se calculan como en (4.46).
7: Asignar, $\tau = \text{máx}(\text{mín}(\tau, \tau_M), \tau_m)$
8: $k = k + 1$
9: **fin mientras**

En dicho algoritmo se calcula $Z_k(\tau)$ como la fórmula de actualización presentada en (4.26) o bien como la proyección de cualquiera de nuestros tres primeros métodos, es decir como en (4.11), donde $Y_k(\tau)$ es cualquiera de las fórmulas (4.8), (4.19) o (4.20).

4.5. Análisis de convergencia

En esta sección mostramos algunos resultados teóricos acerca de la convergencia de nuestros métodos propuestos tanto los basados en proyecciones ((4.6), (4.19) y (4.20)) como el que emplea la factorización de Cholesky.

Lema 8 *Sea* $\{X_k\}$ *una sucesión infinita generada por el* Algoritmo 2 *usando la condición* (4.43a) *para seleccionar el tamaño de paso. Entonces la sucesión* $\{F(X_k)\}$ *converge, además cualquier punto de acumulación* X_* *de* $\{X_k\}$ *es factible, es decir* $X_*^\top X_* = I$.

Demostración.
Representemos por $Y_k(\tau)$ a cualquiera de los esquemas de actualización (4.6), (4.19) o (4.20). Por construcción de los algoritmos tenemos, para todo $k \in \mathbb{N}$,

$$F(X_{k+1}) \quad \leq \quad F(X_k) + \sigma\tau_k Tr[G_k^\top \dot{Y}_k(0)], \tag{4.48}$$

o equivalentemente,

$$\begin{aligned} F(X_k) - F(X_{k+1}) &\geq -\sigma\tau_k Tr[G_k^\top \dot{Y}_k(0)] && (4.49)\\ &> 0 \quad (\text{ ya que } Y_k(\tau) \text{ es curva de descenso en } \tau = 0\,), && (4.50) \end{aligned}$$

así, la sucesión $\{F(X_k)\}$ es monótona decreciente. Ahora como la Variedad de Stiefel es un conjunto compacto y $F(\cdot)$ es continua entonces $F(\cdot)$ alcanza máximo y mínimo sobre la $\mathtt{St}$(n,p) luego como X_k es factible para todo k tenemos que la sucesión $\{F(X_k)\}$ es acotada, así $\{F(X_k)\}$ es monótona y acotada, por lo tanto la sucesión $\{F(X_k)\}$ converge.

Por otro lado, sea $\{X_k\}_{k\in\mathcal{K}}$ una subsucesión convergente de $\{X_k\}$ que converge a X_* es decir $\{X_k\}_{k\in\mathcal{K}} \to X_*$, como X_k es factible para todo $k \in \mathcal{K}$ y $\mathtt{St}$(n,p) es compacto entonces tenemos que $X_* \in \mathtt{St}(n,p)$ es decir:

$$X_*^\top X_* = I$$

por lo tanto todo punto de acumulación es factible.□

El lema anterior y próximo teorema son válidos para el *Algoritmo 2* usando como fórmula de actualización $Z_k(\tau) = \pi(Y_k(\tau))$ donde $Y_k(\tau)$ es cualquiera de los esquemas de actualización (4.6), (4.19), o (4.20).

Teorema 11 *Sea* $\{X_k\}$ *una sucesión de puntos generada por el* Algoritmo 2 *escogiendo el tamaño de paso satisfaciendo las condiciones* (4.43a) *y* (4.43b) *simultáneamente. Supongamos que* $\lim_{k\to\infty} X_k = X^*$ *y que la derivada de la función objetivo* $\mathcal{F}$*, es Lipschitz continua, es decir, que existe una constante* $M > 0$ *tal que:*

$$||\mathcal{DF}(X) - \mathcal{DF}(Y)||_F \quad \leq \quad M\,||X - Y||_F, \quad \forall X, Y \tag{4.51}$$

Entonces X^* *satisface las condiciones de optimalidad de primer orden.*

Demostración.

De el lema 8 sabemos que X_* es factible, falta probar que $\nabla\mathcal{F}(X_*) = \mathbf{0}$. Por la desigualdad de Cauchy-Schwarz y usando la hipótesis de que $\mathcal{DF}(\cdot)$ es Lipschitz continua tenemos que:

$$\begin{aligned} Tr[(\,\mathcal{DF}(X_{k+1}) - \mathcal{DF}(X_k)\,)^\top \dot{Y}_k(0)] &\leq ||\mathcal{DF}(X_{k+1}) - \mathcal{DF}(X_k)||_F ||\dot{Y}_k(0)||_F \\ &\leq M||X_{k+1} - X_k||_F ||\dot{Y}_k(0)||_F \end{aligned}$$

así,

$$Tr[(\,\mathcal{DF}(X_{k+1}) - \mathcal{DF}(X_k)\,)^\top \dot{Y}_k(0)] \quad \leq \quad M||X_{k+1} - X_k||_F ||\dot{Y}_k(0)||_F \tag{4.52}$$

por otro lado, de la condición (4.43b) se tiene que:

$$Tr[(\,\mathcal{D}\mathcal{F}(X_{k+1}) - \mathcal{D}\mathcal{F}(X_k)\,)^\top \dot{Y}_k(0)] \quad \geq \quad (c_2 - 1)Tr[\mathcal{D}\mathcal{F}(X_k)^\top \dot{Y}_k(0)] \tag{4.53}$$

de (4.52) y (4.53) obtenemos:

$$(c_2 - 1)Tr[\mathcal{D}\mathcal{F}(X_k)^\top \dot{Y}_k(0)] \quad \leq \quad M||X_{k+1} - X_k||_F||\dot{Y}_k(0)||_F \tag{4.54}$$

ahora, como $Tr[\cdot]$, $\mathcal{D}\mathcal{F}(\cdot)$ y $\dot{Y}_k(\cdot)$ son continuas, al aplicar límite cuando $k \to \infty$ en (4.54) obtenemos,

$$(c_2 - 1)Tr[\mathcal{D}\mathcal{F}(X_*)^\top(-A_*X_*)] \quad \leq \quad 0 \tag{4.55}$$

o equivalentemente,

$$\frac{1 - c_2}{2}||A_*||_F^2 \quad \leq \quad 0 \tag{4.56}$$

y como $c_2 < 1$ obtenemos,

$$||A_*|| = 0$$

lo cual implica que $\mathbf{0} = A_*X_* = \nabla\mathcal{F}(X_*)$.□

El siguiente teorema establece la convergencia de nuestro *Algoritmo 2* cuando se utiliza la fórmula de actualización (4.26) y se selecciona el *tamaño de paso* satisfaciendo la *condición de Armijo* en cada iteración.

Teorema 12 *Sea* $\{X_k\}$ *una sucesión infinita de iterados generados por el* Algoritmo 2*, empleando la fórmula de actualización* (4.26) *en el paso 7 de dicho algoritmo y seleccionando el tamaño de paso como el mayor número real positivo que satisface la condición de Armijo. Entonces todo punto de acumulación de* $\{X_k\}$ *satisface las condiciones de optimalidad de primer orden.*

Demostración.
Supongamos por absurdo que existe una subsucesión $\{X_k\}_{k\in\mathcal{K}}$ tal que $X_k \to X_*$ y que $\nabla\mathcal{F}(X_*) \neq \mathbf{0}$. Como $L \in \Phi$, es decir, $L(\cdot)$ es infinitamente continuamente diferenciable sobre el conjunto de las matrices anti-simétricas, tenemos que la función $\Gamma(\tau) \in \mathcal{C}^\infty(\mathbb{R})$ y por lo tanto, la función $Z_k(\cdot)$ es también de clase $\mathcal{C}^\infty(\mathbb{R})$ para todo $k \in \mathcal{K}$, por ser un producto de funciones continuamente diferenciables.

Ahora, como $\{F(X_k)\}$ es estrictamente decreciente y acotada sobre la variedad de Stiefel, entonces $\{F(X_k)\}$ converge, y como $F(\cdot)$ es continua entonces $F(X_k) \to F(X_*)$ así, $F(X_k) - F(X_{k+1}) \to 0$. Por construcción del algoritmo tenemos:

$$F(X_k) - F(X_{k+1}) \geq -\sigma\tau_k Tr[\mathcal{D}\mathcal{F}(X_k)^\top \dot{Z}_k(0)], \tag{4.57}$$

luego, como $\{\dot{Z}_k(0)\}$ es gradiente-relacionada y $F(X_k) - F(X_{k+1}) \to 0$ entonces tenemos que $\{\tau_k\}_{k\in\mathcal{K}} \to 0$.

Además, como los $\tau_k's$ son determinados por la *condición de Armijo* (usando Backtracking) esto implica que para todo k mayor que algún $\overline{k}$, $\tau_k = \beta^{m_k}\overline{\tau}$ donde $m_k \in \mathbb{N}$ es el número natural más pequeño de tal forma que se satisface la Regla de Armijo para el k-ésimo iterado, y donde

$0 < \overline{\tau}$ y $0 < \beta < 1$ son el tamaño de paso inicial con el que empieza a iterar el Backtracking y el parámetro de contracción utilizado por el Backtracking respectivamente. Esto significa que para el tamaño de paso $\frac{\tau_k}{\beta}$ no se satisface (4.57), es decir:

$$F(X_k) - F(Z_k(\frac{\tau_k}{\beta})) \quad < \quad -\sigma\frac{\tau_k}{\beta}Tr[\mathcal{DF}(X_k)^\top \dot{Z}_k(0)] \quad \forall k \in \mathcal{K}, k \geq \overline{k}, \tag{4.58}$$

Definamos por $\widetilde{\tau}_k := \frac{\tau_k}{\beta} > 0$ y $W_k(\tau) = F \circ Z_k(\tau)$, entonces (4.58) puede escribirse como:

$$-\frac{W_k(\widetilde{\tau}_k) - W_k(0)}{\widetilde{\tau}_k - 0} \quad < \quad -\sigma Tr[\mathcal{DF}(X_k)^\top \dot{Z}_k(0)] \quad \forall k \in \mathcal{K}, k \geq \overline{k}, \tag{4.59}$$

como $F(\cdot)$ es $C^1(\mathcal{M}(n,p))$ y $Z_k(\cdot)$ es $C^\infty(\mathbb{R})$, $\forall k \geq \overline{k}$ entonces la función $W_k(\cdot) : \mathbb{R} \to \mathbb{R}$ es continua sobre $[0, \widetilde{\tau}_k]$ y diferenciable en $(0, \widetilde{\tau}_k)$, luego por el Teorema del Valor Medio existe $t \in (0, \widetilde{\tau}_k)$ tal que:

$$-\dot{W}_k(t) < -\sigma Tr[\mathcal{DF}(X_k)^\top \dot{Z}_k(0)], \quad \forall k \in \mathcal{K}, k \geq \overline{k},$$

o equivalentemente,

$$-\mathcal{DF}(Z_k(t))[\dot{Z}_k(t)] < -\sigma Tr[\mathcal{DF}(X_k)^\top \dot{Z}_k(0)] \quad \forall k \in \mathcal{K}, k \geq \overline{k},$$

es decir,

$$-Tr[\mathcal{DF}(Z_k(t))^\top \dot{Z}_k(t)] < -\sigma Tr[\mathcal{DF}(X_k)^\top \dot{Z}_k(0)] \quad \forall k \in \mathcal{K}, k \geq \overline{k}. \tag{4.60}$$

Como $\mathcal{DF}(\cdot)$, $Z_k(\cdot)$, $\dot{Z}_k(\cdot)$, $Tr[\cdot]$ son continuas y $\{\widetilde{\tau}_k\} \to 0$ entonces al aplicar límite en la desigualdad (4.60) obtenemos:

$$-Tr[\mathcal{DF}(X_*)^\top(-\nabla F(X_*))] \leq -\sigma Tr[\mathcal{DF}(X_*)^\top(-\nabla F(X_*))], \tag{4.61}$$

o equivalentemente,

$$Tr[\mathcal{DF}(X_*)^\top \nabla F(X_*)] \leq \sigma Tr[\mathcal{DF}(X_*)^\top \nabla F(X_*)], \tag{4.62}$$

lo cual implica que $Tr[\mathcal{DF}(X_*)^\top \nabla F(X_*)] \leq 0$ ya que $\sigma < 1$, pero como $\{-\nabla F(X_k)\}_k$ es gradiente-relacionada entonces tenemos $Tr[\mathcal{DF}(X_*)^\top \nabla F(X_*)] > 0$, lo cual es una contradicción. Por lo tanto todo punto de acumulación X_* de la sucesión $\{X_k\}$ satisface:

$$\mathbf{0} = \nabla F(X_*) = A_* X_*.$$

Por otro lado, por construcción del algoritmo se tiene que $X_k^\top X_k = I_p$, $\forall\, k \in \mathcal{K}$ y como $\mathtt{St}(n,p)$ es un conjunto compacto entonces se tiene que $X_*^\top X_* = I_p$, así, X^* es factible. Por lo tanto X_* satisface:

$$\mathbf{0} = \nabla F(X_*) = A_* X_* = \mathcal{DF}(X_*) - X_* \mathcal{DF}(X_*)^\top X_*,$$

y

$$X_*^\top X_* = I_p,$$

es decir, todo punto de acumulación de $\{X_k\}$ satisface las condiciones de optimalidad de primer orden.□

4.6. Un Splitting Bregman Algorithm para resolver WOPP

El *algoritmo de Bregman* o también conocido como *iteración de Bregman* es un método novedoso que apareció en ciencias de la información el cual fue propuesto por *S. Osher* y colaboradores en [46] para resolver problemas relacionados con *variación total* que surgen en procesamiento digital de imágenes. Este algoritmo es utilizado para resolver el siguiente problema de optimización con restricciones:

$$\min_x \mathcal{J}(x) \quad s.a. \quad \mathcal{A}x = b, \tag{4.63}$$

donde $\mathcal{J}(\cdot)$ es una función convexa y $\mathcal{A}$ es un operador lineal. El optimizador del problema (4.63) puede ser aproximado eficientemente usando el *método de la iteración de Bregman* el cual procede como sigue:

$$\begin{aligned} x_{k+1} &= arg\min_x \mathcal{B}_{\mathcal{J}}^{p_k}(x, x_k) + \frac{\tau}{2}||\mathcal{A}x - b||_2^2 \\ p_{k+1} &= p_k - \tau\mathcal{A}^\top(\mathcal{A}x_{k+1} - b), \end{aligned} \tag{4.64}$$

donde $\mathcal{B}_{\mathcal{J}}^{p_k}(x, x_k) = \mathcal{J}(x) - \mathcal{J}(x_k) - \langle p_k, x - x_k\rangle$ es la distancia de Bregman (ver [47]). Los autores en [29] demostraron que el esquema iterativo (4.64) es equivalente al siguiente método el cual es mucho más simple:

$$\begin{aligned} x_{k+1} &= arg\min_x \mathcal{J}(x) + \frac{\tau}{2}||\mathcal{A}x - b + d_k||_2^2 \\ d_{k+1} &= d_k - \mathcal{A}x_{k+1} - b, \end{aligned} \tag{4.65}$$

es conocido que el esquema iterativo (4.65) es equivalente al bien estudiado *método del Lagrangiano aumentado* (ver [48, 50]).

En esta sección presentamos otra de nuestras propuestas que consiste en utilizar la *iteración de Bregman* para resolver el *Weighted Orthogonal Procrustes Problem* (WOPP), el cual se formula como sigue: dados $X \in \mathbb{R}^{m\times n}$, $A \in \mathbb{R}^{p\times m}$, $B \in \mathbb{R}^{p\times q}$ y $C \in \mathbb{R}^{n\times q}$ el WOPP consiste en resolver el siguiente problema de optimización,

$$\begin{aligned} \min_X \quad & \tfrac{1}{2}||AXC - B||_F^2 \\ \text{sujeto a} \quad & X^\top X = I_n. \end{aligned} \tag{4.66}$$

Nótese que no podemos aplicar el *algoritmo de Bregman* directamente sobre(WOPP) puesto que las restricciones no son lineales. Así, nuestra propuesta consiste en reformular el WOPP, en un problema de optimización para el cual si podamos resolverlo eficientemente utilizando la *iteración de Bregman*. Para ello, consideremos la descomposición en valores singulares de la matriz A como $A = U\Sigma V^\top$, debido a que la norma *Frobenius* es invariante bajo transformaciones ortogonales tenemos que el WOPP es equivalente a:

$$\begin{aligned} \min_Z \quad & \tfrac{1}{2}||\Sigma ZC - H||_F^2 \\ \text{sujeto a} \quad & Z^\top Z = I_n, \end{aligned} \tag{4.67}$$

donde $H = U^\top B \in \mathbb{R}^{p\times q}$ y $Z \in \mathbb{R}^{m\times n}$, el cual es otro WOPP del mismo tamaño pero con una matriz diagonal en lugar de la matriz A que puede ser densa. Observe que si Z^* es el optimizador de (4.67), entonces $X^* = VZ^*$ es optimizador para el problema (4.66). Introduciendo una variable

auxiliar Y (usando la técnica "Splitting") obtenemos un problema equivalente al problema (4.67) como sigue:

$$\begin{aligned} \min_{Y,Z} \quad & J(Y) = \tfrac{1}{2}||\Sigma Y - H||_F^2 \\ \text{sujeto a} \quad & Y = ZC \text{ y } Z^\top Z = I_n. \end{aligned} \tag{4.68}$$

La siguiente proposición muestra que la función $J(Y) = \frac{1}{2}||\Sigma Y - H||_F^2$ es convexa, el cual es un requisito para aplicar el *algoritmo de Bregman.*

Proposición 7 *Consideremos la función $J : \mathbb{R}^{m\times q} \to \mathbb{R}$ dada por $J(Y) = \frac{1}{2}||\Sigma Y - H||_F^2$, entonces $J(Y)$ es una función convexa.*

Demostración.

Primero probemos que la función dada por $F(Y) = ||\Sigma Y - H||_F$ es convexa. En efecto, sean $Y_1, Y_2 \in \mathbb{R}^{m\times q}$ y $\alpha \in (0,1)$, entonces:

$$\begin{aligned} F(\alpha Y_1 + (1-\alpha)Y_2) &= ||\alpha\Sigma Y_1 + (1-\alpha)\Sigma Y_2 - H||_F && (4.69)\\ &= ||\alpha(\Sigma Y_1 - H) + (1-\alpha)(\Sigma Y_2 - H)||_F && (4.70)\\ &\leq \alpha||(\Sigma Y_1 - H)||_F + (1-\alpha)||(\Sigma Y_2 - H)||_F && (4.71)\\ &= \alpha F(Y_1) + (1-\alpha)F(Y_2), && (4.72) \end{aligned}$$

por lo tanto $F(\cdot)$ es convexa. Ahora, como la función $g(x) = x^2$ es convexa y creciente sobre $\mathbb{R}^+$ tenemos que:

$$\begin{aligned} g(\, F(\alpha Y_1 + (1-\alpha)Y_2)\,) &\leq g(\, \alpha F(Y_1) + (1-\alpha)F(Y_2)\,) && (4.73)\\ &\leq \alpha g(\, F(Y_1)\,) + (1-\alpha) g(\, F(Y_2)\,), && (4.74) \end{aligned}$$

o equivalentemente,

$$J(\alpha Y_1 + (1-\alpha)Y_2) \leq \alpha J(Y_1) + (1-\alpha)J(Y_2), \tag{4.75}$$

por lo tanto la función $J(\cdot)$ es convexa. □

Observe que el problema (4.68) el cual es equivalente al problema (4.66) ahora si tiene restricciones lineales con función objetivo convexa, por lo tanto podemos aplicar la *Iteración de Bregman.* Para resolver el problema de optimización (4.68) haremos uso de un *Alternating Algorithm* , es decir, lo haremos de forma desacoplada, primero resolvemos (4.68) considerando el grupo de variables Z fijas, y luego resolvemos para la variable Z considerando las Y fijas las cuales fueron obtenidas en el paso anterior. Nótese que cuando Z esta fija, el problema (4.68) puede resolverse utilizando la *iteración de Bregman*, puesto que las restricciones son lineales. A continuación mostramos el *algoritmo de Bregman* para resolver el problema específico (4.68):

Algoritmo 4 Iteración de Bergman para WOPP genérico

Entrada: $\tau > 0$, k=0, $Z_0 \in \mathtt{St}(m,n)$, $D_0 = \mathbf{0}$, $A \in \mathbb{R}^{p\times m}$, $B \in \mathbb{R}^{p\times q}$ y $C \in \mathbb{R}^{n\times q}$.
Salida: X^* el optimizador.

1: Calcular la SVD de la matriz A como: $A = U\Sigma V^\top$,
2: Fijar $H = U^\top B$,
3: **para** $k = 0, 1, 2, \ldots$ **hacer**
4: $\quad Y_{k+1} = arg\,\mín_Y J(Y) + \frac{\tau}{2}||Y - Z_kC + D_k||_F^2$,
5: $\quad Z_{k+1} = arg\,\mín_Z \frac{1}{2}||ZC - Y_{k+1} - D_k||_F^2$ st. $Z^\top Z = I_n$,
6: $\quad D_{k+1} = D_k + Y_{k+1} - Z_{k+1}C$,
7: $\quad k = k + 1$,
8: **fin para**
9: $Z^* = Z_k$,
10: $X^* = VZ^*$.

Gracias a la forma particular de la función $J(Y)$, podemos calcular una expresión cerrada que resuelve el problema de optimización sin restricciones que aparece en el paso 4 del algoritmo anterior, la siguiente proposición muestra este hecho.

Proposición 8 *El optimizador global del problema de optimización sin restricciones* $\mín_Y J(Y) + \frac{\tau}{2}||Y - Z_kC + D_k||_F^2$ *viene dado por:*

$$Y^* = (\Sigma^\top\Sigma + \tau I)^{-1}(\Sigma^\top H + \tau(Z_kC - D_k)), \tag{4.76}$$

Demostración.
Consideremos el problema de optimización:

$$\mín_{Y\in\mathbb{R}^{m\times q}} \bar{I}(Y) = J(Y) + \frac{\tau}{2}||Y - Z_kC + D_k||_F^2 \tag{4.77}$$

Sea Y^* un minimizador local del problema (4.77), entonces Y^* debe satisfacer que:

$$\begin{aligned}
\nabla\bar{I}(Y^*) = \mathbf{0} &\Leftrightarrow \Sigma^\top(\,\Sigma Y^* - H\,) + \tau(\,Y^* - Z_kC + D_k\,) = \mathbf{0} \\
&\Leftrightarrow (\Sigma^\top\Sigma + \tau I)Y^* = \Sigma^\top H + \tau(Z_kC - D_k) \\
&\Leftrightarrow Y^* = (\Sigma^\top\Sigma + \tau I)^{-1}(\,\Sigma^\top H + \tau(Z_kC - D_k)\,)
\end{aligned}$$

Por otro lado, de la proposición 7 sabemos que la función $J(\cdot)$ es convexa lo cual implica que la función objetivo $\bar{I}(\cdot)$ es convexa por ser una suma de funciones convexas, así cualquier óptimo local del problema (4.77) es también un óptimo global. Por lo tanto concluimos que $Y^* = (\Sigma^\top\Sigma + \tau I)^{-1}(\,\Sigma^\top H + \tau(Z_kC - D_k)\,)$ es un minimizador global del problema (4.77).□

Por otra parte, el problema de optimización sobre la variedad de Stiefel que aparece en el paso 5 del algoritmo anterior, es un problema bien conocido llamado "*Orthogonal Procrustes Problem*"(OPP) el cual tiene también una expresión cerrada (ver [51]). Este hecho se establece en la siguiente proposición adaptada a la notación del problema que aparece en el paso 5 del Algoritmo 4.

Proposición 9 *Si* $rank(C) = n$ *entonces la solución del problema de optimización* $\mín_Z \frac{1}{2}||ZC - Y_{k+1} - D_k||_F^2$ *st.* $Z^\top Z = I_n$ *es:*

$$Z^* = VI_{n,m}U^\top,$$

donde $C(Y_{k+1} + D_k)^\top = USV^\top$ *es la descomposición en valores singulares de la matriz* $C(Y_{k+1} + D_k)^\top$.

Demostración.
Ver anexos.□

De las proposiciones 8 y 9 podemos reescribir el algoritmo 4 como sigue:

Algoritmo 5 Iteración de Bergman para WOPP

Entrada: $\tau > 0$, k=0, $Z_0 \in \mathtt{St}(m,n)$, $D_0 = \mathbf{0}$, $A \in \mathbb{R}^{p\times m}$, $B \in \mathbb{R}^{p\times q}$ y $C \in \mathbb{R}^{n\times q}$.
Salida: X^* el optimizador.
1: Calcular la SVD de la matriz A como: $A = U\Sigma V^\top$,
2: Fijar $H = U^\top B$,
3: **para** $k = 0, 1, 2, \ldots$ **hacer**
4: $Y_{k+1} = (\Sigma^\top\Sigma + \tau I)^{-1}(\Sigma^\top H + \tau(Z_k C - D_k))$,
5: $Z_{k+1} = V I_{n,m} U^\top$ donde $C(Y_{k+1} + D_k)^\top = USV^\top$.
6: $D_{k+1} = D_k + Y_{k+1} - Z_{k+1}C$,
7: $k = k + 1$
8: **fin para**
9: $Z^* = Z_k$,
10: $X^* = VZ^*$,

La eficiencia de esta propuesta recae en el hecho de que no se resuelven problemas de optimización en cada iteración gracias a que ambos problemas de optimización que aparecen en el algoritmo *Iteración de Bergman para WOPP genérico* tienen soluciones cerradas. Próximamente mostramos un resultado teórico que obtuvimos probando nuestro algoritmo *Iteración de Bergman para WOPP*.

Proposición 10 *Sean $A \in \mathbb{R}^{p\times m}$, $B \in \mathbb{R}^{p\times q}$ y $C \in \mathbb{R}^{n\times q}$, además, supongamos que $A = U\Sigma V^\top$ es una SVD de la matriz A. Y sea X^* la solución de* (4.66)*, es decir:*

$$X^* := \arg\min_{X\in \mathcal{St}(m,n)} \frac{1}{2}||AXC - B||_F^2. \tag{4.78}$$

Consideremos la primera iteración del algoritmo 4, usando como parámetro inicial $\tau = 0$, es decir:

$$Y^* := \arg\min_{Y\in\mathbb{R}^{m\times q}} \frac{1}{2}||\Sigma Y - H||_F^2, \tag{4.79}$$

y,

$$Z^* := \arg\min_{Z\in \mathcal{St}(m,n)} \frac{1}{2}||ZC - Y^*||_F^2. \tag{4.80}$$

Entonces se satisface que: Si $||AX^*C - B||_F = 0$ entonces $Y^* = Z^*C$.

Demostración.
Consideremos una descomposición en valores singulares de la matriz A dada por $A = U\Sigma V^\top$, y hagamos $Z = V^\top X^*$, como $||AX^*C - B||_F^2 = 0$ entonces debemos tener que $||\Sigma ZC - H||_F^2 = 0$. Supongamos por absurdo que $Y^* \neq Z^*C$, así $||Y^* - Z^*C||_F^2 > 0$, luego por (4.80) tenemos que:

$$0 < ||Y^* - Z^*C||_F^2 \leq ||Y^* - ZC||_F^2,$$

ahora,

$$0 < ||Y^* - ZC||_F^2 \leq ||\Sigma^\dagger||_F^2 ||\Sigma ZC - \Sigma Y^*||_F^2, \tag{4.81}$$

donde $\Sigma^{\dagger}$ denota a la pseudo-inversa de la matriz Σ. Por otro lado, de (4.79) obtenemos:

$$||\Sigma Y^* - H||_F^2 \leq ||\Sigma ZC - H||_F^2 = 0, \tag{4.82}$$

luego, de (4.81) y (4.82) tenemos que:

$$0 < ||\Sigma^{\dagger}||_F^2 ||\Sigma ZC - H||_F^2,$$

o equivalentemente,

$$0 < ||\Sigma ZC - H||_F^2,$$

lo cual es una contradicción puesto que $||\Sigma ZC - H||_F^2 = 0$. □

Observación 5 *Nótese que la proposición 10 establece que si el valor de la función objetivo evaluada en el optimizador X^* del problema* (4.66) *es igual a cero, entonces el algoritmo 4 converge en una iteración, y la solución del problema* (4.66) *esta dada por $X^* = VZ^*$ donde Z^* es calculado como en el enunciado de la proposición anterior.*

4.7. Resumen del capítulo

En este capítulo se presentaron de forma detallada cuatro métodos de busqueda lineal propuestos para resolver porblemas generales de optimización sobre la variedad de Stiefel, donde tres de los cuales son métodos basados en proyecciones y el otro es una generalización del esquema de actualización (3.7) porpuesto en [21], dicha generalización, emplea una factorización tipo Cholesky de tal forma que garantiza factibilidad del nuevo iterado. Asimismo, se obtuvo una reformulación equivalente del problema *Weighted Orthogonal Procrustes Problem* (WOPP) para la cual se propuso el conocido algoritmo *Iteración de Bergman* para resolver el WOPP eficientemente.

Por otra parte, parece haber suficiente evidencia experimental de que las fórmula de actualización (4.9) y (4.10) son curvas de descenso en $\tau = 0$, sin embargo, esto quedará como trabajo futuro. Además, se estudiaron diferentes estrategias para la selección del tamaño de paso para nuestros algoritmos 2 y 3, entre ellas se introdujo una nueva condición de descenso tipo Armijo-Wolfe (ver (4.43a)-(4.43b)) que funcionó bien en la práctica, no obstante, no se realizó un estudio teórico para determinar si siempre existen tamaños de paso que satifagan dichas condiciones simultaneamente, esto último también quedará como trabajo futuro.

Capítulo 5

Experimentos numérico

En este capítulo analizamos el desempeño de todos nuestros métodos mediante la solución de varios experimentos simulados de diversos problemas con el formato del problema (1.1), para diferentes funciones objetivos y diferentes tamaños de problemas. Además realizamos comparaciones entre algunos métodos del estado del arte contra nuestros métodos, con el propósito de medir el comportamiento y la eficiencia de dichos algoritmos.

5.1. Detalles de Implementación

Todos nuestros experimentos fueron ejecutados usando Matlab R2013a en un procesador intel CORE i3-380M, CPU 2.53 GHz con 500Gb de HD y 8Gb de Ram. Para las distintas constantes de nuestro Algoritmo 3, utilizamos los siguiente valores: *tamaño de paso* inicial $\tau =$ 1e-2, $\rho_1 =$ 1e-4, $\eta = 0.85$, $\delta = 0.1$ (estos son los valores por defectos empleados por el algoritmo propuesto en [21]). Por otra parte, como la convergencia de los métodos de primer orden (métodos que utilizan la primera derivada de la función objetivo) puede ser muy lenta a medida que dichos métodos se acercan al óptimo local, es importante detectar esta cercanía al óptimo para detener el algoritmo apropiadamente. Para lidiar con esto, es necesario emplear varios criterios de parada. En nuestra implementación de todos nuestros algoritmos, además de considerar la norma del gradiente $||\nabla \mathcal{F}(X_k)||_F$ revisamos el error absoluto entre dos iterados consecutivos, así como también, el error relativo entre los valores objetivos de dos iterados consecutivos. Más específicamente, nuestros algoritmos ejecutaran un máximo de M iteraciones y se detendrán en la iteración $k < M$ si se satisface alguno de los siguientes tres criterios de parada:

1. $\text{tol}_k^x := ||\nabla \mathcal{F}(X_k)||_F < \epsilon$
2. $\text{tol}_k^f := \frac{||X_{k+1} - X_k||_F}{\sqrt{n}} < \text{xtol} \quad y \quad \frac{\mathcal{F}(X_k) - \mathcal{F}(X_{k+1})}{|\mathcal{F}(X_k)|+1} < \text{ftol}$
3. $\text{mean}([\text{tol}_{k-\min(k,T)+1}^x, \ldots, \text{tol}_k^x]) < 10\text{xtol} \quad y \quad \text{mean}([\text{tol}_{k-\min(k,T)+1}^f, \ldots, \text{tol}_k^f]) < 10\text{ftol}.$

Los valores por defecto son xtol = 1e-6, ftol = 1e-12, $T = 5$ y $K = 1000$.

En todos los experimentos que presentaremos en las siguientes subsecciones denotaremos por:

- *Nfe* : Al número de evaluaciones de la función objetivo.
- *Nitr* : El número de iteraciones realizadas por el algoritmo hasta convergencia.
- *Time* : El tiempo (en segundos) utilizado por el algoritmo hasta converger.

- *NrmG* : La norma del gradiente de la función Lagrangiana respecto a las variables primales evaluado en el "óptimo" estimado por el algoritmo.
- *Fval* : La evaluación de la función objetivo en el "óptimo" estimado por el algoritmo.
- *Feasi*: La distancia entre la matriz identidad y la matriz producto resultante de multiplicar la traspuesta del "óptimo" estimado por el algoritmo multiplicado con sigo misma, es decir, $Feasi = ||X_*^\top X_* - I_p||_F$, donde X_* denota al "óptimo" estimado por el algoritmo.

Además, denotaremos por: *OptiStiefel* al método propuesto en [21], *Sgmin* al método propuesto en [13], *PGST* al algoritmo presentado en [9], *MSteepest* al "Modified Steepest Descent Method" propuesto en [28], utilizando el paso de Barzilai Borwein como *tamaño de paso* inicial con el cual empieza a iterar el *Bactracking*. Mientras que: *Ad-Bash*, *Ad-Moul*, *Linear-Co*, *Chol-Retrac* denotaran a nuestro algoritmo 3, cuando este utilice como $Y_k(\tau)$ las fórmulas de actualización (4.19), (4.20), (4.8) (con los parámetros (λ, μ) fijados como $\lambda = 2/3$ y $\mu = 1/3$) y (4.26) respectivamente. Y por último llamaremos *BregmanWOPP* a nuestro algoritmo 5.

5.2. Weighted Orthogonal Procrustes Problem (WOPP)

Dados $X \in \mathbb{R}^{m\times n}$, $A \in \mathbb{R}^{p\times m}$, $B \in \mathbb{R}^{p\times q}$ y $C \in \mathbb{R}^{n\times q}$, el WOPP consiste en resolver el siguiente problema de optimización con restricciones:

$$\begin{aligned} \text{mín} \quad & \tfrac{1}{2}||AXC - B||_F^2 \\ \text{sujeto a} \quad & X^\top X = I. \end{aligned}$$

Para los experimento numéricos, consideramos $n = q$, $p = m$, $A = PSR^\top$ y $C = Q\Lambda Q^\top$, donde P y Q son matrices ortogonales generadas de manera aleatoria con $Q \in \mathbb{R}^{n\times n}$ y $R, P \in \mathbb{R}^{m\times m}$, $\Lambda \in \mathbb{R}^{n\times n}$ es una matriz diagonal con entradas generadas siguiendo una distribución uniforme en el intervalo $[\frac{1}{2}, 2]$ y S es una matriz diagonal definida para cada tipo de problema (mas abajo se explica como se generará la matriz S para cada problema). Como punto inicial $X_0 \in \mathbb{R}^{m\times n}$ para que comiencen a iterar los algoritmos, generamos una matriz aleatoria perteneciente a la variedad de Stiefel, donde cada entrada de X_0 sigue una distribución uniforme en el intervalo [0,1]. Cuando no se especifique como se generan los valores aleatorios, serán generados bajo una distribución *Normal estándar*.

Para tener control sobre el óptimo del problema generado, creamos la solución $Q_* \in \mathbb{R}^{m\times n}$ de manera aleatoria en la variedad de Stiefel, y construimos la matriz B de problema como $B = AQ_*C$, de esta forma, podemos medir el error de convergencia al óptimo global. La construcción de la matriz S se hace de tres formas diferentes como se explica a continuación:

Problema 1: Cada elemento de la diagonal de la matriz S es generado siguiendo una distribución *Normal truncada* en el intervalo [10,12].

Problema 2: La diagonal de S está dada por $S_{ii} = i + 2r_i$, donde cada r_i es una número aleatorio uniformemente distribuido en el intervalo $[0, 1]$.

Problema 3: Cada elemento de la diagonal de S se genera de la siguiente manera: $S_{ii} = 1 + \frac{99(i-1)}{m+1} + 2r_i$, donde cada r_i es una número aleatorio uniformemente distribuido en el intervalo $[0, 1]$.

5.2.1. Comparación entre la búsqueda no monótona y la búsqueda monótona

Primero que todo, realizamos un estudio comparativo de nuestros métodos utilizando tanto la técnica monótona como la técnica no monótona para la selección del *tamaño de paso* en problemas sobre la variedad de Stiefel, así como también en problemas sobre el *Grupo Ortogonal*, dicho estudio es presentado en la Tabla 5.1. Para este estudio se construyeron 100 *Orthogonal Procrustes Problems* (OPP), el cual se formula de forma similar que el WOPP pero cuando la matriz C es igual a la identidad, se tomo como número máximo de iteraciones de 3000 y como tolerancia de $\epsilon =$ 1e-4 (tolerancia para la norma del gradiente) para todos los algoritmos. Para este experimento se generaron las matrices $A \in \mathbb{R}^{m\times m}$ y $B \in \mathbb{R}^{m\times n}$ aleatoriamente donde cada entrada de dichas matrices siguen una distribución normal estándar, y como punto inicial se generó una matriz $X_0 = UI_{m,n}V^\top$, donde $R = U\Sigma V^\top$ es la descomposición en valores singulares de la matriz aleatoria $R \in \mathbb{R}^{m\times n}$, con entradas escogidas mediante una distribución uniforme en el intervalo [0,1].

En dicha tabla la palabra *monótono* se refiere a que emplearemos el algoritmo 2 y la palabra *no monótono* se refiere al algoritmo 3. En la Tabla 5.1 se observa claramente que la estrategia no monótona es más eficiente que la estrategia monótona para cada uno de nuestros métodos, sin embargo se muestra un desempeño muy similar cuando se utiliza como fórmula de actualización a (4.26), es decir, en el caso del método basado en la factorización de Cholesky.

Nota 9 *En los próximos experimentos mostrados en las siguientes secciones y subsecciones emplearemos nuestros métodos usando la estrategia no monótona para la selección del tamaño de paso, debido a que en este estudio se observa que nuestros métodos logran un mejor desempeño cuando utilizan dicha estrategia.*

5.2.2. Estudio comparativo de los métodos resolviendo el problema WOPP

Para cada una de las siguientes 12 tablas (Tablas: 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10, 5.11, 5.12 y 5.13) se construyeron un total de 300 problemas WOPP's generados como se explicó al inicio de esta sección, creando la matriz S de acuerdo a los problemas: **Problema 1**, **Problema 2** y **Problema 3** respectivamente; y se utilizó un número máximo de iteraciones de 8000 con una tolerancia de $\epsilon =$ 1e-4 (tolerancia para la norma del gradiente) para todos los métodos. En dichas tablas se denotará por *Error* a la norma del error en la solución, es decir a: $||X_k - Q_*||$, donde X_k es el óptimo estimado por el algoritmo, además, *min*, *mean*, *max* denotan el valor mínimo, máximo y el promedio obtenido por cada algoritmo en los 300 problemas, mientras que *var* denotará la varianza de cada valor a comparar en dichas 300 corridas.

En dichas tablas estudiamos el desempeño de nuestros algoritmos en problemas tanto bien condicionados (**Problema 1**) como en problemas mal condicionados (**Problema 2** y **Problema 3**) de diversos tamaños, además, comparamos nuestros 5 algoritmos con 2 de los algoritmos del estado del arte.

Nota 10 *En todos los experimentos mostrados en las 12 tablas (Tablas: 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10, 5.11, 5.12 y 5.13), se obtuvo una error en la factibilidad promedio del orden 1e-14 para todos los métodos.*

En todos los experimentos mostrados en esta subsección, observamos que el algoritmo más eficiente es nuestro *BregmanWOPP*, el cual solo requiere a lo sumo 5 iteraciones para llegar a

Tabla 5.1: Estudio comparativo entre el algoritmo monótono (Algoritmo 2) y el algoritmo no-monótono (Algoritmo 3) en el problema WOPP

	Ad-Bash no monótono						**Ad-Bash monótono**					
Problema	**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Error**	**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Error**
m = 50, n = 20	751.22	917.61	1.8206	1.20e-03	2.10e-03	0.1567	1423.9	3042.1	6.7112	1.05e-02	3.60e-03	0.1854
m = 75, n = 75	818.75	829.48	3.8688	9.07e-04	3.21e-06	0.0168	1532.4	3594.6	17.9253	6.96e-02	1.20e-03	0.1065
	Ad-Moul no monótono						**Ad-Moul monótono**					
m = 50, n = 20	700.92	729.5	0.9215	8.97e-04	2.15e-04	0.0578	1355.6	2923.6	6.1105	1.17e-02	4.40e-03	0.1188
m = 75, n = 75	771.58	782.82	3.9633	8.81e-04	3.11e-06	0.0166	1481.1	3793.3	21.7005	2.76e-01	4.40e-03	0.1565
	Linear-Co no monótono						**Linear-Co monótono**					
m = 50, n = 20	728.97	752.7	0.8252	9.37e-04	4.54e-04	0.1062	1417.9	3070.7	6.6712	1.49e-02	3.90e-03	0.1766
m = 75, n = 75	832.91	843.38	4.0208	1.70e-03	4.67e-06	0.019	1472.7	3295.5	16.4805	5.39e-02	5.61e-04	0.0985
	Chol-Retrac no monótono						**Chol-Retrac monótono**					
m = 50, n = 20	817.77	976.75	0.6687	1.00e-03	3.20e-03	0.2888	845.02	1050.3	0.6869	1.20e-03	3.20e-03	0.29
m = 75, n = 75	1313.6	1313.6	3.0609	9.99e-04	6.91e-06	0.0211	1306.8	1550.4	3.4537	1.87e-02	5.79e-04	0.0532

convergencia, además, dicho método parece comportarse igual de bien tanto en problemas bien condicionados como en los mal condicionados. El resto de los métodos en general resuelven rápidamente los problemas bien condicionados, sin embargo les tomas mas iteraciones, evaluaciones de la función objetivo y más tiempo en converger.

En los problemas bien condicionados (Tablas: 5.2, 5.3, 5.4, 5.5), notamos que todos los métodos muestran similar eficiencia en cuanto a iteraciones y evaluaciones de la función objetivo, aunque el algoritmo *PGST* es el que realiza en promedio menor cantidad de iteraciones (con excepción de *BregmanWOPP*), sin embargo todos los algoritmos llegan al óptimo global con bastante precisión en promedio.

En problemas mal condicionados (Tablas: 5.6, 5.7, 5.8, 5.9, 5.10, 5.11, 5.12 y 5.13), observamos que en ocasiones nuestros métodos propuestos (con excepción de *BregmanWOPP*) pueden conseguir mucho error al óptimo global en promedio (ver Tabla 5.6 y Tabla 5.10) que los métodos del estado del arte, sin embargo en el restos de los experimentos (Ex.6, Ex.7, Ex.8, Ex.10, Ex.11 y Ex.12), obtienen un *Error* promedio similar al error logrado por los métodos del estado del arte. Por otra parte, notamos que nuestro método *Chol-Retrac* le toma menos tiempo en converger que al método *PGST*, aunque en ocasiones puede realizar más iteraciones nuestro método basado en Cholesky. También nuestros métodos *Ad-Bash* y *Ad-Moul* convergen más rápido en tiempo y en a veces también en iteraciones que el método *PGST*.

En adición a esto, observamos que el método *OptiStiefel* es el segundo método que mejores resultados obtiene en cada uno de los problemas mostrados en dichas 12 tablas, sin embargo algunos de nuestros métodos (Por ejemplo *Ad-Bash*, *Ad-Moul*, en las Tablas 5.6 y 5.8) en ocasiones pueden realizar menos iteraciones que los métodos del estado del arte *OptiStiefel* y *PGST*.

En la Figura 5.1 presentamos las gráficas de la función objetivo promedio y de la norma del gradiente promedio versus las primeras 25 iteraciones, correspondientes a los experimentos presentados en las tablas: 5.2, 5.4, 5.6, 5.9, 5.11 y 5.13. En dichas gráficas se observa que el método que converge más rápido es nuestro método *BregmanWOPP*, también se nota claramente que todos nuestros métodos muestran un comportamiento muy similar. Además, vemos que el algoritmo *PGST* en los problemas bien condicionado (Figura 5.1(a) y Figura 5.1(b)) desciende la norma del gradiente un poco más rápido que el resto de los métodos(exceptuando a *BregmanWOPP*), sin embargo para problemas mal condicionados (Figura 5.1(c), Figura 5.1(d), Figura 5.1(e) y Figura 5.1(f)), varios de nuestros métodos descienden la norma del gradiente un poco más rápido que el algoritmo *PGST*.

Nota 11 *Para el resto de los otros 6 experimentos (Ex.2, Ex.4, Ex.6, Ex.7, Ex.9 y Ex.11) mostrados en las tablas: 5.3, 5.5, 5.7, 5.8, 5.10 y 5.11 se observó un comportamiento similar entre los métodos y por tanto no mostramos sus gráficas.*

5.2.3. Comparación entre los métodos OptiStiefel y BregmanWOPP

En esta subsección comparamos solo los métodos *OptiStiefel* y *BregmanWOPP*, puesto que fueron los más eficientes en las comparaciones previas. En las tablas: 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10, 5.11, 5.12 y 5.13, vimos que el algoritmo *BregmanWOPP* fue mucho más eficiente (tanto en problemas bien condicionados como en los mal condicionados) que el método

Tabla 5.2: Desempeño de los métodos sobre problemas WOPP's bien condicionados (**Problema 1**) con $m = 100$ y $n = 50$

		Ex.1: **Problema 1** con m=100 y n = 50					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Error**
OptiStiefel	min	46	47	0.0690	2.93e-06	1.10-13	7.78e-08
	mean	56.51	57.51	0.0946	6.74e-05	4.06e-11	1.24e-06
	max	64	65	0.1542	9.95e-05	1.62e-10	3.27e-06
	var	17.89	17.88	0.0005	5.83e-10	1.82e-21	7.43e-13
Ad-Bash	min	45	46	0.1653	5.52e-06	8.08e-14	5.30e-08
	mean	57.35	58.35	0.2219	6.63e-05	4.00e-11	1.25e-06
	max	69	70	0.3642	9.99e-05	1.61e-10	3.25e-06
	var	24.61	24.61	0.0024	5.64e-10	1.60e-21	7.15e-13
Ad-Moul	min	47	48.00	0.1712	2.04e-05	5.53e-13	9.00e-08
	mean	56.13	57.13	0.2523	6.58e-05	4.02e-11	1.26e-06
	max	69	70	0.4386	9.90e-05	1.57e-10	3.17e-06
	var	19.35	19.35	0.0033	4.93e-10	1.67e-21	7.06e-13
Linear-Co	min	48	49	0.1621	4.60e-06	1.59e-13	9.56e-08
	mean	58.23	59.23	0.2273	6.94e-05	4.37e-11	1.33e-06
	max	72	73	0.3724	9.97e-05	1.49e-10	3.10e-06
	var	21.15	21.15	0.0021	5.80e-10	1.81e-21	7.13e-13
PGST	min	31	34	0.1461	1.51e-05	3.36e-13	1.09e-07
	mean	38.83	39.49	0.1964	1.00e-04	4.37e-11	1.23e-06
	max	46	42	0.3597	4.24e-04	1.44e-10	2.97e-06
	var	9.11	2.96	2.1e-03	4.87e-09	1.32e-21	5.72e-13
Chol-Retrac	min	51	52	0.0966	8.85e-06	6.12e-13	9.00e-08
	mean	61.41	62.41	0.1404	7.24e-05	3.41e-11	1.07e-06
	max	75	76	0.2374	5.15e-04	2.10e-10	2.97e-06
	var	19.92	19.92	1.00e-03	3.09e-09	1.56e-21	6.59e-13
BregmanWOPP	min	3	4	0.0091	1.53e-09	3.64e-21	6.43e-12
	mean	3	4	0.0118	2.15e-09	5.87e-21	7.81e-12
	max	3	4	0.0212	2.61e-09	8.04e-21	9.09e-12
	var	0	0	7.49e-06	4.55e-20	8.64e-43	3.16e-25

Tabla 5.3: Desempeño de los métodos sobre problemas WOPP's bien condicionados (**Problema 1**) con m = 500 y n = 70

		Ex.2: **Problema 1** con m=500 y n = 70					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Error**
OptiStiefel	min	49	50	0.8021	1.23e-05	5.98e-13	1.45e-07
	mean	57.46	58.46	1.0532	6.79e-05	5.01e-11	1.45e-06
	max	70	71	1.4797	9.86e-05	1.43e-10	3.08e-06
	var	24.78	24.7841	0.0368	5.31e-10	2.01e-21	8.57e-13
Ad-Bash	min	51	52	2.4564	1.35e-05	7.81e-13	1.55e-07
	mean	57.44	58.44	3.1662	8.44e-05	5.23e-11	1.36e-06
	max	66	67	4.9007	7.61e-04	4.77e-10	4.15e-06
	var	15.03	15.02	0.3259	1.01e-08	5.66e-21	1.00e-12
Ad-Moul	min	49	50	2.0818	2.79e-05	6.20e-13	1.19e-07
	mean	56.66	57.66	2.7779	6.91e-05	4.94e-11	1.45e-06
	max	65	66	3.8186	9.93e-05	1.57e-10	3.19e-06
	var	16.27	16.26	0.2226	5.43e-10	2.09e-21	8.27e-13
Linear-Co	min	48	49	2.6009	1.33e-05	7.13e-13	1.44e-07
	mean	59.72	60.72	3.7167	6.10e-05	3.63e-11	1.27e-06
	max	71	72	5.0644	9.95e-05	1.34e-10	2.92e-06
	var	24	24.0016	0.3924	6.22e-10	1.28e-21	5.56e-13
PGST	min	36	38	1.8673	9.63e-06	7.96e-13	1.75e-07
	mean	41.62	44.04	2.3654	7.85e-05	3.68e-11	1.12e-06
	max	49	53	3.2158	2.23e-04	1.35e-10	2.98e-06
	var	11.02	1.8678	0.1415	1.83e-09	1.48e-21	7.15e-13
Chol-Retrac	min	52	53	4.8960	1.89e-05	1.95e-13	3.93e-08
	mean	60.40	61.4	6.1676	8.60e-05	4.83e-11	1.42e-06
	max	69	70	8.9264	4.30e-04	1.31e-10	2.90e-06
	var	15.06	15.0612	0.5813	4.29e-09	1.50e-21	6.71e-13
BregmanWOPP	min	3	4	0.1272	2.32e-09	9.07e-21	9.33e-12
	mean	3	4	0.1676	2.57e-09	1.06e-20	1.02e-11
	max	3	4	0.2659	2.95e-09	1.29e-20	1.09e-11
	var	0	0	0.0011	1.99e-20	7.85e-43	1.09e-25

Tabla 5.4: Desempeño de los métodos sobre problemas WOPP's bien condicionados (**Problema 1**) con m = 150 y n = 150

		Ex.3: **Problema 1** con m = 150 y n = 150					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Error**
OptiStiefel	min	42	43	0.3025	6.55e-06	5.83e-14	4.66e-08
	mean	50.46	51.66	0.4711	6.04e-05	8.53e-12	5.59e-07
	max	62	66	0.8487	9.88e-05	3.59e-11	1.50e-06
	var	11.20	12.81	0.0080	5.39e-10	7.93e-23	1.47e-13
Ad-Bash	min	45	46	0.8317	4.30e-06	4.17e-14	4.14e-08
	mean	51.90	52.95	1.1888	7.79e-05	1.82e-11	6.52e-07
	max	66	67	1.7733	1.60e-03	7.91e-10	2.56e-06
	var	15.06	15.40	0.0440	2.58e-08	6.19e-21	2.17e-13
Ad-Moul	min	42	43	0.8411	1.76e-06	4.40e-15	1.33e-08
	mean	50.24	51.28	1.2423	6.55e-05	1.08e-11	6.41e-07
	max	61	63	1.7344	9.99e-05	3.88e-11	1.57e-06
	var	10.55	11.03	0.0486	5.53e-10	1.08e-22	1.95e-13
Linear-Co	min	44	45	0.8931	1.09e-05	1.85e-13	4.60e-08
	mean	53.40	54.45	1.2279	6.99e-05	1.11e-11	6.37e-07
	max	66	68	1.9491	4.23e-04	5.18e-11	1.61e-06
	var	15.54	16.23	0.0402	1.96e-09	1.24e-22	1.79e-13
PGST	min	33	35	0.8992	1.68e-04	1.89e-11	4.16e-07
	mean	37.64	41.02	1.2803	6.62e-04	1.07e-09	6.45e-06
	max	43	44	1.7722	9.97e-04	3.63e-09	1.49e-05
	var	6.17	2.91	0.0437	4.89e-08	9.55e-19	1.78e-11
Chol-Retrac	min	47	48	0.4415	1.47e-05	2.08e-13	6.33e-08
	mean	56.07	57.57	0.6303	5.97e-05	8.36e-12	5.41e-07
	max	66	68	1.0052	2.43e-04	3.55e-11	1.50e-06
	var	19.90	21.00	0.0148	9.69e-10	8.71e-23	1.43e-13
BregmanWOPP	min	3	4	0.0531	3.46e-09	6.76e-21	8.54e-12
	mean	3	4	0.0750	4.16e-09	9.80e-21	1.02e-11
	max	3	4	0.1202	5.10e-09	1.45e-20	1.23e-11
	var	0	0	0.0002	8.90e-20	1.58e-42	3.95e-25

Tabla 5.5: Desempeño de los métodos sobre problemas WOPP's bien condicionados (**Problema 1**) con m = 200 y n = 200

		Ex.4: **Problema 1** con m=200 y n = 200					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Error**
OptiStiefel	min	45	46	0.6945	5.25e-06	8.09e-14	2.87e-08
	mean	51.66	52.74	1.0213	5.97e-05	9.53e-12	6.01e-07
	max	63	65	1.5491	9.97e-05	3.59e-11	1.48e-06
	var	14.71	16.01	0.0311	6.37e-10	9.39e-23	1.70e-13
Ad-Bash	min	45	46	1.6943	9.72e-06	2.27e-13	5.39e-08
	mean	51.53	52.54	2.4482	6.58e-05	1.20e-11	6.84e-07
	max	66	67	4.0893	9.95e-05	3.70e-11	1.53e-06
	var	13.69	13.69	0.1839	5.83e-10	1.23e-22	2.03e-13
Ad-Moul	min	44	45	1.7656	1.08e-05	2.98e-13	6.90e-08
	mean	51.07	52.13	2.6669	7.38e-05	1.25e-11	6.41e-07
	max	61	63	3.9080	9.16e-04	2.25e-10	1.51e-06
	var	12.33	13.02	0.2343	7.77e-09	5.47e-22	1.59e-13
Linear-Co	min	46	47	1.7694	1.35e-05	9.35e-14	4.51e-08
	mean	53.06	54.10	2.6375	6.16e-05	1.08e-11	6.40e-07
	max	63	65	3.8103	9.97e-05	3.94e-11	1.58e-06
	var	13.27	13.79	0.1975	6.02e-10	1.12e-22	1.89e-13
PGST	min	33	36	1.8616	1.64e-04	2.25e-11	5.48e-07
	mean	38.23	42.01	2.7528	6.56e-04	9.62e-10	5.95e-06
	max	43	45	3.7282	9.99e-04	3.83e-09	1.55e-05
	var	6.22	2.86	0.2112	5.84e-08	8.13e-19	1.54e-11
Chol-Retrac	min	49	50	0.9660	6.12e-06	1.29e-13	2.96e-08
	mean	57.68	59.72	1.4378	6.25e-05	1.02e-11	6.23e-07
	max	77	79	2.5175	9.96e-05	3.49e-11	1.49e-06
	var	25.65	26.73	0.0729	5.66e-10	1.08e-22	1.83e-13
BregmanWOPP	min	3	4	0.1045	4.20e-09	1.05e-20	1.08e-11
	mean	3	4	0.1585	4.81e-09	1.31e-20	1.19e-11
	max	3	4	0.2367	5.29e-09	1.61e-20	1.35e-11
	var	0	0	0.0012	6.78e-20	1.58e-42	3.19e-25

Tabla 5.6: Desempeño de los métodos sobre problemas WOPP's mal condicionados (**Problema 2**) con m = 100 y n = 50

		Ex.5: **Problema 2** con m=100 y n = 50					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Error**
OptiStiefel	min	496	516	0.6950	7.18e-05	4.18e-12	6.74e-07
	mean	947.64	983.67	1.4822	9.40e-04	3.68e-09	2.30e-05
	max	4202	4320	6.3298	1.67e-02	2.74e-08	7.69e-05
	var	2.04e+05	2.14e+05	0.4881	4.83e-06	1.93e-17	2.57e-10
Ad-Bash	min	615	643	2.0454	1.14e-04	2.81e-11	1.79e-06
	mean	1449.40	1496.00	5.8582	2.00e-03	3.78e+00	0.661
	max	8000	8176	44.0102	4.90e-02	17.7	1.61
	var	1.23e+06	1.28e+06	33.9137	4.12e-05	23.8	4.54e-01
Ad-Moul	min	586	610	2.2492	1.20e-04	4.26e-11	1.56e-06
	mean	998.17	1034.80	4.3623	9.34e-04	7.99e-01	0.237
	max	3468	3551	18.5796	7.80e-03	11.3	1.53e+00
	var	1.97e+05	2.04e+05	5.6050	2.42e-06	4.74e+00	0.278
Linear-Co	min	596	628	2.1075	6.25e-05	3.29e-12	8.24e-07
	mean	1470.50	1520.60	37.7982	1.70e-03	3.62	0.688
	max	8000	8223	2232.9	2.70e-02	17.7	1.61
	var	1.50e+06	1.59e+06	7.09e+04	1.37e-05	20.6	4.64e-01
PGST	min	757	831	3.2030	1.72e-05	1.40e-12	2.10e-07
	mean	1341.50	1403.30	5.9058	1.57e-04	3.78e-10	7.17e-06
	max	2471	2526	11.7753	2.28e-04	1.21e-09	2.04e-05
	var	1.53e+05	1.73e+05	3.5353	2.62e-09	1.19e-19	2.50e-11
Chol-Retrac	min	644	668	1.3122	6.57e-05	5.65e-12	8.21e-07
	mean	1368.20	1412.90	2.9728	9.47e-04	2.66e-05	5.98e-03
	max	3797	3905	7.9177	1.04e-02	1.49e-03	1.59e-02
	var	5.41e+05	5.67e+05	2.8141	1.92e-06	1.37e-02	4.63e-03
BregmanWOPP	min	3	4	9.00e-03	1.58e-05	1.56e-14	4.67e-09
	mean	3.70	4.70	0.0135	2.07e-04	5.93e-12	5.22e-08
	max	4	5	0.0369	1.20e-03	7.28e-11	2.97e-07
	var	0.21	0.21	1.47e-05	6.96e-08	1.73e-22	4.43e-15

Tabla 5.7: Desempeño de los métodos sobre problemas WOPP's mal condicionados (**Problema 2**) con m = 300 y n = 20

		Ex.6: **Problema 2** con m=300 y n = 20					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Error**
OptiStiefel	min	2002	2065	2.6180	4.51e-04	1.03e-08	4.15e-05
	mean	4719.30	4848.40	6.0856	1.27e-02	5.33e-02	5.32e-02
	max	8000	8261	11.9580	0.425	0.886	0.5
	var	2.52e+06	2.64e+06	4.3966	2.50e-03	2.92e-02	1.99e-02
Ad-Bash	min	1986	2053	14.0136	4.34e-04	6.93e-10	6.80e-06
	mean	4804.10	4.94e+03	37.1042	4.73e-02	9.08e-02	7.80e-02
	max	8000	8229	67.9388	3.25	0.991	0.489
	var	2.58e+06	2.71e+06	196.2429	0.132	5.07e-02	2.74e-02
Ad-Moul	min	2240	2310	12.8783	3.81e-04	2.80e-09	2.60e-05
	mean	4772.30	4.90e+03	30.3373	4.00e-03	5.05e-02	4.82e-02
	max	8000	8259	55.2121	4.03e-02	0.886	0.5
	var	2.62e+06	2.75e+06	141.5461	5.10e-05	2.89e-02	1.87e-02
Linear-Co	min	2078	2133	16.3560	5.20e-04	4.17e-09	2.38e-05
	mean	4732.20	4.86e+03	40.2486	1.01e-02	9.57e-02	8.04e-02
	max	8000	8229	72.3710	0.340	0.991	0.489
	var	2.62e+06	2.75e+06	234.4277	1.50e-03	5.33e-02	2.71e-02
PGST	min	3118	3280	18.1472	6.52e-05	1.59e-13	1.46e-07
	mean	6373.10	6.74e+03	37.3798	0.467	8.66e-02	8.14e-02
	max	8000	8678	53.7511	26.20	1.22	0.496
	var	1.95e+06	2.11e+06	73.6521	8.76	5.15e-02	2.89e-02
Chol-Retrac	min	2370	2453	21.8324	7.58e-04	5.56e-09	3.46e-05
	mean	5002.90	5.14e+03	45.9841	4.16e-02	0.114	8.54e-02
	max	8000	8226	75.2888	1.49	1.51	0.575
	var	3.05e+06	3.20e+06	256.4498	3.99e-02	7.16e-02	2.89e-02
BregmanWOPP	min	3	4	0.0162	2.03e-06	4.06e-17	2.30e-10
	mean	3.025	4.025	0.0179	2.29e-05	1.40e-14	2.21e-09
	max	4	5	0.0308	1.50e-04	2.10e-13	1.39e-08
	var	0.0247	0.0249	7.32e-06	7.84e-10	1.29e-27	8.15e-18

Tabla 5.8: Desempeño de los métodos sobre problemas WOPP's mal condicionados (**Problema 2**) con m = 100 y n = 100

		Ex.7: **Problema 2** con m=100 y n = 100					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Error**
OptiStiefel	min	470	491	1.2599	7.31e-05	5.21e-12	6.25e-07
	mean	698.21	732.67	2.2601	7.32e-04	3.74e-09	2.23e-05
	max	1150	1202	4.5162	8.40e-03	1.29e-07	1.86e-04
	var	2.36E+04	2.55e+04	0.4272	1.17e-06	1.65e-16	4.41e-10
Ad-Bash	min	428	453	3.2538	4.90e-05	3.95e-12	7.30e-07
	mean	684.99	716.53	5.9783	7.90e-04	2.56e-09	2.02e-05
	max	1290	1335	12.5499	5.90e-03	2.12e-08	6.56e-05
	var	2.92e+04	3.07e+04	3.4566	9.89e-07	7.73e-18	1.83e-10
Ad-Moul	min	454	469	3.8384	1.21e-04	3.09e-11	2.30e-06
	mean	696.31	728.49	6.4396	9.82e-04	2.65e-09	2.00e-05
	max	1491	1538	17.4251	1.20e-02	1.88e-08	7.22e-05
	var	2.99e+04	3.17e+04	4.1027	2.95e-06	1.15e-17	2.14e-10
Linear-Co	min	423	443	3.2542	4.62e-05	1.20e-11	1.13e-06
	mean	702.17	734.43	6.0978	9.42e-04	3.01e-09	2.15e-05
	max	1513	1561	16.1990	1.57e-02	4.82e-08	1.02e-04
	var	3.04e+04	3.22e+04	3.7451	3.52e-06	2.70e-17	2.57e-10
PGST	min	639	675	7.0615	1.74e-04	6.07e-12	6.87e-08
	mean	1155.10	1174.10	14.1745	8.16e-04	4.57e-09	2.46e-05
	max	2782	3177	46.8151	1.00e-03	2.01e-08	8.91e-05
	var	1.47e+05	1.92e+05	33.6195	2.60e-08	1.96e-17	3.81e-10
Chol-Retrac	min	526	551	1.7912	5.39e-05	8.28e-12	8.21e-07
	mean	926.40	965.53	3.5897	7.13e-04	2.54e-09	2.03e-05
	max	1764	1839	6.8566	4.60e-03	3.27e-08	1.07e-04
	var	5.69e+04	6.08e+04	1.1152	6.74e-07	1.30e-17	2.15e-10
BregmanWOPP	min	3	4	0.0209	3.70e-05	4.52e-14	6.72e-09
	mean	3.54	4.54	0.0330	5.16e-04	1.66e-11	7.47e-08
	max	4	5	0.0694	2.90e-03	2.01e-10	3.97e-07
	var	0.25	0.26	7.28e-05	4.11e-07	1.37e-21	8.70e-15

Tabla 5.9: Desempeño de los métodos sobre problemas WOPP's mal condicionados (**Problema 2**) con m = 150 y n = 150

		Ex.8: **Problema 2** con m=150 y n = 150					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Error**
OptiStiefel	min	686	716	4.7328	1.40e-04	8.25e-12	7.19e-07
	mean	1140.70	1188.00	8.4452	1.40e-03	1.49e-08	4.02e-05
	max	2245	2316	17.1876	2.18e-02	6.17e-07	3.65e-04
	var	7.76e+04	8.24e+04	4.9983	8.17e-06	4.10e-15	1.90e-09
Ad-Bash	min	662	696	11.5057	1.01e-04	1.39e-11	1.53e-06
	mean	1146.50	1190.60	20.1947	1.30e-03	7.69e-09	3.43e-05
	max	2186	2255	38.0823	1.27e-02	7.96e-08	1.60e-04
	var	9.40e+04	9.85e+04	30.6897	3.14e-06	1.06e-16	5.67e-10
Ad-Moul	min	676	711	12.4192	2.06e-04	2.87e-11	1.90e-06
	mean	1159	1205.10	21.9395	1.20e-03	7.97e-09	3.54e-05
	max	1872	1942	35.6667	1.51e-02	1.17e-07	1.36e-04
	var	8.59e+04	9.13e+04	31.2668	3.18e-06	1.54e-16	4.34e-10
Linear-Co	min	576	775	13.4751	1.20e-04	6.74e-11	2.66e-06
	mean	1164.10	1210.60	20.8039	1.30e-03	1.06e-08	3.71e-05
	max	1881	1945	33.5554	1.29e-02	2.56e-07	2.53e-04
	var	6.85e+04	7.26e+04	21.5239	3.60e-06	9.38e-16	1.25e-09
PGST	min	1125	1162	27.1616	1.67e-04	3.66e-12	6.59e-08
	mean	2039.60	2101.10	50.6247	8.52e-04	5.50e-09	2.85e-05
	max	3521	3558	116.3565	1.00e-03	1.98e-08	8.50e-05
	var	3.30e+05	3.65e+05	247.4626	2.51e-08	2.29e-17	3.71e-10
Chol-Retrac	min	744	770	6.5042	8.78e-05	1.38e-11	1.28e-06
	mean	1568.30	1624.80	13.9113	1.50e-03	9.28e-09	3.77e-05
	max	3424	3550	32.6171	1.12e-02	9.84e-08	1.25e-04
	var	2.17e+05	2.30e+05	18.6499	4.26e-06	2.00e-16	4.53e-10
BregmanWOPP	min	3	4	0.0525	9.54e-05	1.02e-13	6.49e-09
	mean	3.70	4.70	0.0749	1.30e-03	5.04e-11	8.60e-08
	max	4	5	0.1074	7.20e-03	5.27e-10	4.39e-07
	var	0.2100	0.2121	1.49e-04	2.95e-06	1.31e-20	1.18e-14

Tabla 5.10: Desempeño de los métodos sobre problemas WOPP's bien condicionados (**Problema 3**) con m = 100 y n = 50

		Ex.9: **Problema 3** con m=100 y n = 50					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Error**
OptiStiefel	min	565	593	0.8862	7.60e-05	5.69e-11	3.42e-06
	mean	922.90	957.44	1.5990	6.42e-04	3.50e-09	2.46e-05
	max	1561	1610	3.0633	2.80e-03	1.13e-08	6.06e-05
	var	6.42e+04	6.82e+04	0.3025	4.26e-07	6.70e-18	1.73e-10
Ad-Bash	min	690	722	2.4038	7.13e-05	3.93e-11	2.64e-06
	mean	1354.20	1399.60	6.0294	1.20e-03	2.52	0.590
	max	6249	6410	35.7016	9.20e-03	12.6	1.54
	var	7.94e+05	8.34e+05	25.6153	2.45e-06	11.3	0.458
Ad-Moul	min	590	610	2.8954	9.29e-05	5.58e-11	2.56e-06
	mean	1235.60	1277.60	6.2601	1.30e-03	1.35	0.287
	max	6149	6319	39.94	1.98e-02	9.31	1.47
	var	1.01e+06	1.06e+06	41.9638	1.08e-05	7.87	0.303
Linear-Co	min	590	616	2.3169	9.75e-05	7.52e-11	3.20e-06
	mean	1221.20	1263.20	5.4358	7.40e-04	1.80	0.441
	max	3682	3792	19.3750	3.50e-03	10.3	1.54
	var	3.91e+05	4.13e+05	10.1415	4.26e-07	8.95	0.423
PGST	min	730	751	3.8403	6.84e-05	1.54e-12	6.66e-08
	mean	1497.70	1546.80	7.5341	1.61e-04	4.17e-10	7.59e-06
	max	5676	5918	27.0817	2.41e-04	1.50e-09	2.12e-05
	var	5.59e+05	6.62e+05	13.7644	2.16e-09	1.51e-19	3.04e-11
Chol-Retrac	min	666	686	1.5634	1.81e-04	3.72e-10	5.73e-06
	mean	1726.20	1778.90	4.2702	1.50e-03	3.96e-03	7.80e-03
	max	7360	7545	19.5379	1.31e-02	1.48e-02	0.154
	var	1.66e+06	1.75e+06	11.7044	6.14e-06	0.177	0.458
BregmanWOPP	min	3	4	9.00e-03	2.05e-05	2.22e-14	5.34e-09
	mean	3.72	4.72	0.0149	2.49e-04	9.55e-12	6.17e-08
	max	4	5	0.0231	1.60e-03	1.08e-10	3.14e-07
	var	0.21	0.21	1.49e-05	1.28e-07	4.91e-22	7.17e-15

Tabla 5.11: Desempeño de los métodos sobre problemas WOPP's bien condicionados (**Problema 3**) con m = 500 y n = 20

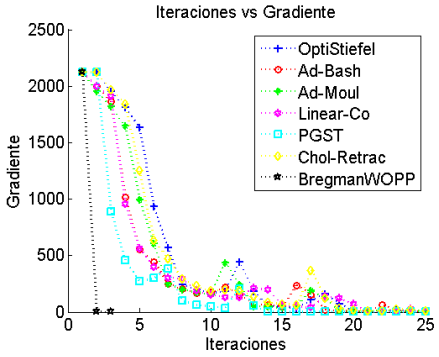
		Ex.10: **Problema 3** con m=500 y n = 20					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Error**
OptiStiefel	min	983	1020	4.1092	1.16e-04	4.81e-10	1.66e-05
	mean	1889.50	1948.60	7.5236	8.79e-04	3.87e-08	1.49e-04
	max	3330.00	3423	13.6319	1.05e-02	7.37e-07	6.28e-04
	var	1.95e+05	2.04e+05	3.3653	2.57e-06	8.13e-15	8.09e-09
Ad-Bash	min	1082	1125	27.0043	1.28e-04	3.49e-09	5.27e-05
	mean	1848.50	1.91E+03	46.0724	9.64e-04	3.84e-08	1.49e-04
	max	3288	3314	81.2410	1.19e-02	9.22e-07	9.67e-04
	var	1.61e+05	1.68e+05	103.5515	3.89e-06	1.19e-14	1.35e-08
Ad-Moul	min	1146	1192	19.7332	1.07e-04	3.23e-10	1.43e-05
	mean	1808.30	1.87e+03	31.0771	1.00e-03	4.26e-08	1.63e-04
	max	3132	3223	53.9565	1.25e-02	5.10e-07	7.26e-04
	var	1.54e+05	1.62e+05	47.0689	3.78e-06	5.72e-15	1.35e-08
Linear-Co	min	1193	1247	35.3059	1.23e-04	2.74e-10	1.47e-05
	mean	1856.80	1.92e+03	55.1726	7.26e-04	3.59e-08	1.48e-04
	max	3514	3608	105.2558	7.20e-03	6.33e-07	6.30e-04
	var	1.85e+05	1.94e+05	167.9982	1.43e-06	5.82e-15	8.48e-09
PGST	min	1494	1615	28.1761	4.21e-05	4.54e-12	1.74e-06
	mean	2562.70	2.63e+03	47.0365	8.88e-05	1.12e-09	2.76e-05
	max	4256	4456	77.1138	1.07e-04	2.66e-09	5.53e-05
	var	3.57e+05	6.50e+04	109.7266	2.64e-10	6.22e-19	2.15e-10
Chol-Retrac	min	1269	1323	93.0690	6.08e-05	6.33e-10	1.80e-05
	mean	1886.10	1.95e+03	138.5358	1.10e-03	3.58e-08	1.54e-04
	max	3208	3304	227.5429	1.19e-02	2.87e-07	4.94e-04
	var	1.70e+05	1.79e+05	900.33	3.07e-06	2.03e-15	6.74e-09
BregmanWOPP	min	3	4	0.0510	1.10e-06	9.72e-17	6.76e-10
	mean	3	4	0.0570	4.51e-06	3.48e-15	3.40e-09
	max	3	4	0.0738	2.32e-05	4.61e-14	1.44e-08
	var	0	0	1.90e-05	1.72e-11	6.53e-29	9.17e-18

Tabla 5.12: Desempeño de los métodos sobre problemas WOPP's bien condicionados (**Problema 3**) con m = 150 y n = 150

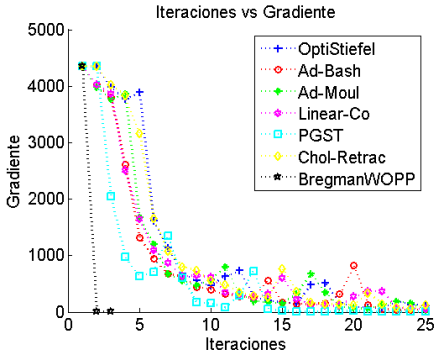
		Ex.11: **Problema 3** con m=150 y n = 150					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Error**
OptiStiefel	min	483	513	2.6585	9.07e-05	7.13e-11	2.12e-06
	mean	772.3	808.79	4.6259	7.79e-04	3.24e-09	2.44e-05
	max	1192	1245	13.8069	1.05e-02	4.72e-08	1.01e-04
	var	2.67e+04	2.85e+04	2.7324	1.35e-06	2.55e-17	2.15e-10
Ad-Bash	min	474	503	6.8173	8.78e-05	5.27e-11	2.36e-06
	mean	775.61	809.81	11.9164	9.68e-04	3.15e-09	2.47e-05
	max	1324	1371	26.482	1.33e-02	1.62e-08	6.48e-05
	var	3.57e+04	3.83e+04	11.7727	3.24e-06	7.72e-18	1.93e-10
Ad-Moul	min	458	480	8.0387	7.67e-05	2.38e-12	4.38e-07
	mean	767.94	803.25	12.6462	8.08e-04	4.01e-09	2.43e-05
	max	1162	1223	27.0107	1.43e-02	1.10e-07	1.46e-04
	var	2.78e+04	2.95e+04	8.9717	2.60e-06	1.30e-16	3.98e-10
Linear-Co	min	502	527	7.955	1.28e-04	2.96e-11	1.95e-06
	mean	781.4	817.03	12.1716	9.23e-04	3.94e-09	2.64e-05
	max	1378	1434	29.2381	7.60e-03	6.03e-08	1.21e-04
	var	3.02e+04	3.02e+04	11.7465	1.48e-06	4.53e-17	3.40e-10
PGST	min	666	681	12.7661	2.91e-04	2.56e-11	2.20e-07
	mean	1333.6	1376.5	26.6159	8.05e-04	5.89e-09	3.13e-05
	max	3450	3374	65.4974	9.98e-04	2.58e-08	1.14e-04
	var	2.18e+05	2.28e+05	92.5951	3.91e-08	3.48e-17	6.01e-10
Chol-Retrac	min	650	682	4.29	4.99e-05	6.81e-12	1.03e-06
	mean	1049.8	1091.1	7.1067	1.20e-03	1.85e-08	2.99e-05
	max	2239	2318	20.6407	4.19e-02	1.53e-06	4.74e-04
	var	7.93e+04	8.42e+04	6.1716	1.76e-05	2.33e-14	2.22e-09
BregmanWOPP	min	3	4	0.0423	6.06e-05	1.02e-13	8.79e-09
	mean	3.57	4.57	0.0547	5.50e-04	1.96e-11	7.86e-08
	max	4	5	0.0944	3.20e-03	2.25e-10	4.39e-07
	var	0.25	0.25	1.00e-4	5.08e-07	2.14e-21	9.91e-15

Tabla 5.13: Desempeño de los métodos sobre problemas WOPP's bien condicionados (**Problema 3**) con m = 200 y n = 200

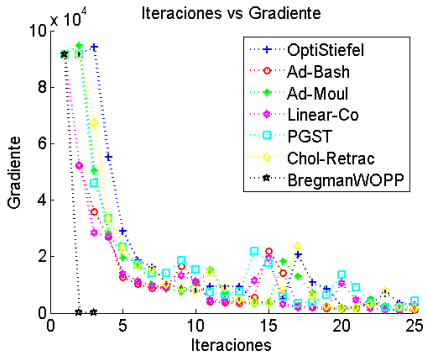
		Ex.12: **Problema 3** con m=200 y n = 200					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Error**
OptiStiefel	min	506	533	5.6492	1.02e-04	2.09e-12	6.22e-07
	mean	830.89	869.09	9.2928	7.84e-04	4.82e-09	2.99e-05
	max	1308	1365	14.611	6.10e-03	6.53e-08	1.49e-04
	var	3.42e+04	3.64e+04	4.3007	9.26e-07	7.96e-17	6.41e-10
Ad-Bash	min	478	507	13.7358	6.61e-05	2.45e-12	7.30e-07
	mean	825.51	862.65	23.3411	1.00e-03	4.56e-09	2.95e-05
	max	1381	1427	39.0457	8.20e-03	8.42e-08	1.44e-04
	var	3.29e+04	3.49e+04	25.7105	1.63e-06	9.25e-17	4.80e-10
Ad-Moul	min	509	529	15.3845	1.24e-04	3.38e-11	2.35e-06
	mean	829.83	866.26	25.154	1.10e-03	5.57e-09	3.13e-05
	max	1307	1343	39.5962	1.39e-02	8.88e-08	2.05e-04
	var	2.91e+04	3.06e+04	25.8994	4.27e-06	1.42e-16	8.08e-10
Linear-Co	min	530	565	15.4902	1.29e-04	2.32e-11	2.61e-06
	mean	843.9	880.68	24.1485	9.31e-04	4.60e-09	3.06e-05
	max	1784	1534	42.8174	5.80e-03	5.38e-08	1.51e-04
	var	3.65e+04	3.85e+04	29.0578	1.32e-06	5.61e-17	5.77e-10
PGST	min	735	784	27.1321	8.20e-05	5.84e-13	3.43e-07
	mean	1407	1465.6	52.646	8.26e-04	6.65e-09	3.44e-05
	max	4473	4817	174.0447	9.99e-04	2.39e-08	9.95e-05
	var	2.44e+05	3.14e+05	374.7758	2.98e-08	4.07e-17	6.45e-10
Chol-Retrac	min	653	678	8.3250	8.02e-05	7.07e-12	9.00e-07
	mean	1124.3	1168.8	14.2194	1.20e-03	5.43e-09	3.04e-05
	max	1968	2014	24.5911	2.10e-02	1.52e-07	2.69e-04
	var	7.67e+04	8.10e+04	11.9617	7.29e-06	2.90e-16	9.75e-10
BregmanWOPP	min	3	4	0.0846	6.03e-05	9.38e-14	8.94e-09
	mean	3.65	4.65	0.1035	5.67e-04	1.81e-11	7.96e-08
	max	4	5	0.1293	3.20e-03	2.01e-10	3.75e-07
	var	0.23	0.23	1.79e-04	4.55e-07	1.71e-21	8.59e-15



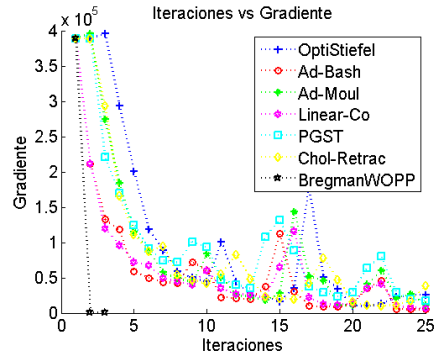
(a) Norma del Gradiente promedio, Ex.1, Tabla 5.2



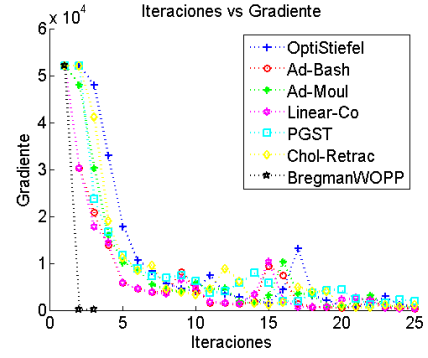
(b) Norma del Gradiente promedio, Ex.3, Tabla 5.4



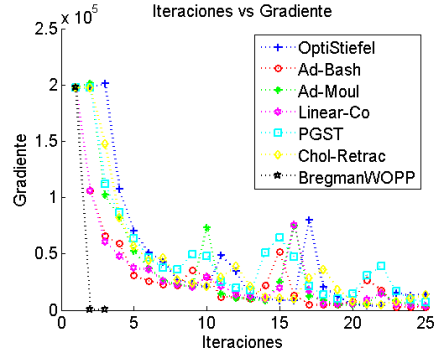
(c) Norma del Gradiente promedio, Ex.5, Tabla 5.6



(d) Norma del Gradiente promedio, Ex.8, Tabla 5.9



(e) Norma del Gradiente promedio, Ex.10, Tabla 5.11



(f) Norma del Gradiente promedio, Ex.12, Tabla 5.13

Figura 5.1: Gráficas comparativas de la norma del gradiente promedio de los métodos sobre problemas WOPP's

OptiStiefel sobre problemas para los cuales el óptimo evaluado en la función objetivo es exactamente igual a cero. Para realizar un estudio comparativo más completo entre ambos métodos, estudiamos el desempeño de los métodos mencionados, en problemas para los cuales el valor optimal no sea igual a cero, dicho estudio se muestra en las tablas: 5.14, 5.15, 5.16, 5.17, 5.18 y 5.16.

Para los problemas presentados en dichas tablas (tablas: 5.14, 5.15, 5.16, 5.17, 5.18 y 5.16), se construyeron un total de 100 WOPP's, siguiendo la construcción de las matrices A, X_0 y C descrita en los problemas: **problema 1** y **problema 2** explicados al inicio de esta sección, y la matriz $B \in \mathbb{R}^{m\times n}$ se generó de manera aleatoria, donde cada entrada de la matriz B sigue una distribución uniforme en el intervalo [0,1]; además tomamos como número máximo de iteraciones de 5000 y tolerancia de $\epsilon =$ 1e-3. En las tablas 5.14, 5.15 y 5.16, se comparan ambos métodos en problemas bien condicionados (generando la matriz A como se explica en **problema 1**), mientras que en las tablas 5.17, 5.18 y 5.16 se muestra la comparación en problemas mal condicionados (generando la matriz A como se explica en **problema 2**).

En dichas tablas se denota por τ, al parámetro que recibe de entrada nuestro algoritmo 5 (*BregmanWOPP*). Por otra parte, en las tablas: 5.17, 5.18 y 5.16, denotaremos por *OptiStiefel1*, al método propuesto en [21] cuando este comienza a iterar con un punto inicial X_0^{rand} elegido aleatoriamente, y denotamos por *OptiStiefel2* al mismo método cuando el punto inicial es un punto X_{aprox} que esta próximo a un óptimo local del problema; de igual manera para nuestro algoritmo, es decir denotaremos por *BregmanWOPP1* cuando este inicie en un punto aleatorio (el mismo punto X_0^{rand}) y por *BregmanWOPP2* cuando inicie en X_{aprox}. En vista de que no tenemos conocimiento del óptimo verdadero, construiremos al punto X_{aprox} como sigue: primero resolvemos el problema usando el método *OptiStiefel1* con el punto inicial X_0^{rand}, obteniendo X^* como la solución estimada por dicho método, luego construimos una matriz aleatoria R con entradas todas menores o iguales a 1e-2, después calculamos la SVD de la matriz $X^* + R$, de tal forma que $X^* + R = U\Sigma V^\top$ y por último tomamos $X_{aprox} = UI_{m,n}V^\top$.

En los experimentos presentados en las tablas: 5.14, 5.15 y 5.16 (problemas bien condicionados), se observa que nuestro algoritmo *BregmanWOPP* necesita en promedio un menor número de iteraciones y de evaluaciones de la función objetivo para converger, en contraste con el método *OptiStiefel*. Además, ambos algoritmos convergen en un tiempo similar. En contraparte, en problemas mal condicionados (ver tablas: 5.17, 5.18 y 5.19), nuestro algoritmo *BregmanWOPP* obtiene un pobre desempeño y realiza el número máximo de iteraciones sin llegar a converger cuando el punto inicial es elegido al azar (ver los resultados del algoritmo *BregmanWOPP1* en las tablas: 5.17, 5.18 y 5.19), a diferencia del método *OptiStiefel* que si logra llegar a un óptimo local y en un tiempo razonable. Sin embargo, nuestro algoritmo *BregmanWOPP* si logra converger a un óptimo local cuando el punto inicial esta próximo a un óptimo local (ver los resultados del algoritmo *BregmanWOPP2* en las tablas: 5.17, 5.18 y 5.19), pero sigue realizando en promedio muchas más iteraciones y evaluaciones de la función objetivo que el método *OptiStiefel*, y en consecuencia, también le toma más tiempo en converger a nuestro algoritmo 5.

En la Figura 5.2, mostramos las gráficas promedios de las evaluaciones de la función objetivo y de la norma del gradiente versus las iteraciones (a lo sumo 500 iteraciones), correspondientes a los experimentos Ex.5 y Ex.6 cuyos resultados están contenidos en las tablas 5.16 y 5.17 respectivamente. En dichas gráficas notamos que para el problema bien condicionado (Ex.5), a nuestro algoritmo le toma menos iteraciones en converger en promedio que al algoritmo *OptiStiefel*, sin embargo en las gráficas asociadas al problema mal condicionado (Ex.6) vemos que ocurre lo contrario, no obstante, nuestro algoritmo *BregmanWOPP* va descendiendo la función objetivo pero

lo hace muy lentamente.

Nota 12 *En la Figura 5.2 solo mostramos las gráficas de los experimentos Ex.5 y Ex.6, sin embargo, para las gráficas del resto de los experimentos se obtuvo un comportamiento similar y por tanto solo mostramos las gráficas de los experimentos Ex.5 y Ex.6.*

De los experimentos realizados en esta sección concluimos que nuestro método *Bregman-WOPP* es más eficiente que el resto de los métodos comparados en problemas para los cuales el valor optimal de la función objetivo es exactamente igual a cero y también en casos donde el problema de optimización (WOPP) esta bien condicionado, en el resto de los casos nuestro *BregmanWOPP* es poco eficiente y por tanto, conviene utilizar el método *OptiStiefel*.

5.3. Problemas de valores propios lineales

Dada una matriz simétrica $A \in \mathbb{R}^{n\times n}$ y una matriz $X \in \mathtt{St}(n,p)$ arbitraria, la traza de $X^\top AX$ es maximizada cuando X es una base ortonormal de espacio generado por los vectores propios asociados a los p mayores valores propios de la matriz A. Sean $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ los valores propios de A. El problema de valores propios lineales, puede ser formulado como:

$$\sum_{i=1}^{p} \lambda_i =: \max_{X\in\mathtt{St}(n,p)} Tr[X^\top AX], \tag{5.1}$$

el cual, es un problema que podemos resolver minimizando la función $\mathcal{F}(X) = -Tr[X^\top AX]$ sobre la variedad de Stiefel. Observe que la función objetivo del problema (5.1) es una generalización del *cociente de Raleigh*.

En esta subsección comparamos el desempeño de cuatro de nuestros algoritmos junto con dos algoritmos del estado del arte (*OptiStiefel*, *Sgmin*). Para dicha comparación se construyeron tres grupos de problemas, con $n = 100, 500, 1000$ y variando el valor de p, los resultados para cada uno de estos grupos de problemas (problemas con $n = 100, 500, 1000$) son presentados en las tablas 5.20, 5.21 y 5.22 respectivamente. Para todos los experimentos mostrados en dichas tablas se crearon un total de 100 problemas de valores propios lineales, generando la matriz A de la siguiente forma: $A = \bar{A}^\top \bar{A}$, donde la matriz $\bar{A} \in \mathbb{R}^{n\times n}$ fue generada aleatoriamente con cada una de sus entradas siguiendo una distribución *Normal estándar*, y generando al punto inicial X_0 (con el cual empieza a iterar los algoritmos) aleatoriamente en la variedad de Stiefel. Tomamos como número máximo de iteraciones 2000 y una tolerancia para la norma del gradiente de $\epsilon =$ 1e-4.

En las tablas 5.20, 5.21 y 5.22, se presentan el promedio para cada valor a comparar (*Nitr*,*Nfe*,...) obtenido por cada método, además como conocemos teóricamente el valor óptimo del problema (la suma de los p mayores valores propios de A) también comparamos el error relativo entre el valor objetivo dado por la función *eig* de Matlab y el valor objetivo del óptimo estimado por cada algoritmo, dicho error relativo será denotado en la tablas por *Error*, es decir:

$$Error = |\frac{\sum_{i=1}^{p} \lambda_i^{eig} - Tr[X_*^\top AX_*]}{Tr[X_*^\top AX_*]}|, \tag{5.2}$$

donde λ_i^{eig} denota el i-ésimo valor propio obtenido con la función *eig* de Matlab y X_* es la aproximación del óptimo estimado por cada algoritmo.

Tabla 5.14: Comparación entre los métodos *OptiStiefel* y *BregmanWOPP* en problemas WOPP's bien condicionados (**Problema 1**)

		Ex.1: **Problema 1** con m=50 y n = 30					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Feasi**
OptiStiefel	min	59	60	0.0325	6.49e-05	1.70e+03	2.50e-14
	mean	76.31	77.38	0.0528	1.10e-03	2.41e+03	3.04e-14
	max	123	125	0.1798	8.30e-03	3.17e+03	4.18e-14
	var	126.18	128.60	3.72e-04	1.83e-06	1.10e+05	9.78e-30
BregmanWOPP	min	22	23	0.0186	1.30e-04	1.70e+03	1.17e-14
	mean	27.09	28.09	0.0296	2.62e-04	2.41e+03	1.33e-14
	max	56	57	0.0545	6.96e-04	3.17e+03	1.56e-14
	var	22.61	22.608	7.51e-05	1.29e-08	1.10e+05	5.89e-31
		Ex.2: **Problema 1** con m=50 y n = 50					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Feasi**
OptiStiefel	min	97	98	0.0798	3.01e-04	3.48e+03	2.70e-15
	mean	181.43	190.25	0.1549	3.30e-03	4.19e+03	6.91e-14
	max	536	566	0.3551	1.62e-02	5.14e+03	9.70e-14
	var	4124.9	4690.1	0.0033	1.11e-05	1.24e+05	2.27e-28
BregmanWOPP	min	32	33	0.0588	3.93e-04	3.48e+03	1.99e-14
	mean	240.5	241.5	0.5036	1.50e-03	4.19e+03	2.16e-14
	max	2756	2757	5.1397	3.60e-03	5.14e+03	2.43e-14
	var	1.32e+05	1.32e+05	0.5715	5.71e-07	1.24e+05	8.52e-31

Los valores obtenidos de τ en el Ex.1 fueron: (min,mean,max,var) = (117,118,81,120,0,36)
Los valores obtenidos de τ en el Ex.2 fueron: (min,mean,max,var) = (121,126,74,150,37,81)

Tabla 5.15: Comparación entre los métodos *OptiStiefel* y *BregmanWOPP* en problemas WOPP's bien condicionados (**Problema 1**)

		Ex.3: **Problema 1** con m=100 y n = 50					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Feasi**
OptiStiefel	min	60	61	0.0815	4.26e-05	2.94e+03	4.25e-14
	mean	71.17	72.18	0.1277	1.90e-03	3.79e+03	5.36e-14
	max	87	88	0.4152	1.62e-02	4.80e+03	6.89e-14
	var	33.6	33.75	3.00e-03	6.82e-06	1.34e+05	2.78e-29
BregmanWOPP	min	19	20	0.0494	1.51e-04	2.94e+03	1.94e-14
	mean	21.03	22.0300	0.0710	4.27e-04	3.79e+03	2.15e-14
	max	26	27	0.2013	7.84e-04	4.80e+03	2.44e-14
	var	2.45	2.4536	7.28e-04	1.72e-08	1.34e+05	9.38e-31
		Ex.4: **Problema 1** con m=100 y n = 100					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Feasi**
OptiStiefel	min	153	290	0.3216	8.88e-04	6.80e+03	4.94e-15
	mean	270.79	162	0.5836	5.50e-03	8.21e+03	5.54e-15
	max	858	927	1.8731	2.82e-02	9.90e+03	6.49e-15
	var	13826.00	16204.00	0.0657	2.96e-05	3.37e+05	5.36e-32
BregmanWOPP	min	38	39	0.2260	9.20e-04	6.80e+03	3.61e-14
	mean	487.54	488.54	2.9613	3.10e-03	8.21e+03	3.87e-14
	max	3446	3447	20.8523	6.50e-03	9.90e+03	4.11e-14
	var	2.54e+05	2.54e+05	9.9785	1.37e-06	3.37e+05	7.46e-31

Los valores obtenidos de τ en el Ex.3 fueron: (min,mean,max,var) = (115,116,69,118,0,32)
Los valores obtenidos de τ en el Ex.4 fueron: (min,mean,max,var) = (121,133,74,170,103,12)

Tabla 5.16: Comparación entre los métodos *OptiStiefel* y *BregmanWOPP* en problemas WOPP's bien condicionados (**Problema 1**)

		Ex.5: **Problema 1** con m=300 y n = 70					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Feasi**
OptiStiefel	min	48	50	0.2424	1.02e-04	4.61e+03	4.55e-15
	mean	57.29	59.37	0.2970	2.30e-03	5.29e+03	5.52e-15
	max	69	71	0.3564	2.06e-02	6.76e+03	6.84e-15
	var	19.14	19.29	5.38e-04	1.13e-05	1.47e+05	1.68e-31
BregmanWOPP	min	18	19	0.1499	1.10e-03	4.61e+03	2.53e-14
	mean	18.12	19.12	0.1559	1.90e-03	5.29e+03	2.71e-14
	max	19	20	0.3061	2.30e-03	6.76e+03	2.91e-14
	var	0.11	0.1067	2.42e-04	9.00e-08	1.47e+05	7.51e-31

Los valores obtenidos de τ en el Ex.5 fueron: (min,mean,max,var) = (107,108,110,0,45)

Tabla 5.17: Comparación entre los métodos *OptiStiefel* y *BregmanWOPP* en problemas WOPP's mal condicionados (**Problema 2**)

		Ex.6:**Problema 2** con m=50 y n = 30					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Feasi**
OptiStiefel1	min	524	560	0.4328	4.8e-03	2.70e+03	1.56e-15
	mean	1.04e+03	1.12e+03	0.8554	0.0165	4.24e+03	1.95e-14
	max	2902	3076	2.0927	0.0839	6.14e+03	9.93e-14
	var	2.07e+05	2.31e+05	0.1187	1.15e-04	6.14e+03	1.27e-27
BregmanWOPP1	min	5000	5001	5.1104	0.3269	2.78e+03	1.14e-14
	mean	5000	5001	6.5169	5.4345	4.54e+03	1.35e-14
	max	5000	5001	9.2219	182.7906	8.16e+03	1.58e-14
	var	0	0	0.6269	406.8315	5.14e+05	8.93e-31
OptiStiefel2	min	60	63	0.0376	6.7e-03	2.70e+03	1.75e-15
	mean	82.9625	88.5	0.0703	0.0138	4.24e+03	4.67e-14
	max	154	163	0.1414	0.046	4.24e+03	7.65e-14
	var	236.2897	280.4557	4.49e-04	2.86e-05	5.13e+05	1.69e-28
BregmanWOPP2	min	273	274	0.2868	9.1e-03	2.70e+03	1.16e-14
	mean	430.2	4.31e+02	0.5569	0.0242	4.24e+03	1.35e-14
	max	2849	2850	3.8029	0.0312	6.14e+03	1.35e-14
	var	9.80e+03	8.98e+04	0.1626	1.48e-05	5.13e+05	8.39e-31
		Ex.7: **Problema 2** con m=50 y n = 50					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Feasi**
OptiStiefel1	min	453	484	0.3439	0.0108	1.31e+04	2.50e-15
	mean	9.38e+02	9.85e+02	0.8959	0.0351	1.94e+04	2.95e-15
	max	3478	3615	3.3669	0.1859	2.83e+04	3.25e-15
	var	2.17e+05	2.33e+05	0.1905	7.02e-04	7.73e+06	1.88e-32
BregmanWOPP1	min	5000	5001	9.2945	4.5625	1.35e+04	1.94e-14
	mean	5000	5001	11.664	9.4457	1.99e+04	2.14e-14
	max	5000	5001	19.8812	18.215	2.93e+04	2.34e-14
	var	0	0	2.5311	7.3711	7.73e+06	7.91e-31
OptiStiefel2	min	44	46	0.0302	0.0158	1.31e+04	3.69e-14
	mean	64.6375	67.875	0.0605	0.038	1.94e+04	4.93e-14
	max	108	112	0.1402	0.182	2.83e+04	8.17e-14
	var	224.8416	252.3892	4.15e-04	9.14e-04	7.73e+06	8.92e-29
BregmanWOPP2	min	264	265	0.5967	0.0535	1.31e+04	1.90e-14
	mean	355.3625	3.56e+02	0.8213	0.0771	1.94e+04	2.17e-14
	max	544	545	1.3819	0.0956	2.83e+04	2.42e-14
	var	2.50e+03	2.50e+03	0.0218	7.38e-05	7.73e+06	1.22e-30

Los valores obtenidos de τ en el Ex.6 fueron: (min,mean,max,var) = (1165,1291,3,1300,603,18)
Los valores obtenidos de τ en el Ex.7 fueron: (min,mean,max,var) = (2691,2703,1,2710,20,72)

Tabla 5.18: Comparación entre los métodos *OptiStiefel* y *BregmanWOPP* en problemas WOPP's mal condicionados (**Problema 2**)

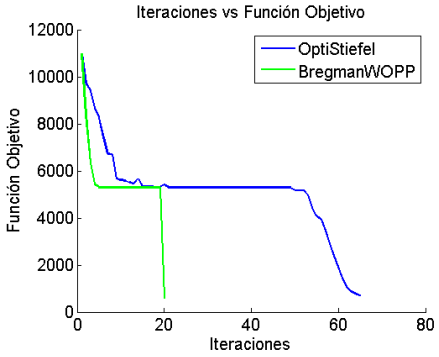
		Ex.8:**Problema 2** con m=75 y n = 40					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Feasi**
OptiStiefel1	min	644	699	0.5815	0.0166	6.38e+03	2.14e-15
	mean	1.72e+03	1.83e+03	1.6645	0.039	9.98e+03	2.48e-15
	max	5000	5246	7.4802	0.2144	1.46e+04	2.86e-15
	var	6.25e+05	6.90e+05	1.0409	5.89e-04	2.61e+06	2.80e-32
BregmanWOPP1	min	5000	5001	8.2706	2.9797	7.38e+03	1.62e-14
	mean	5000	5001	8.8013	6.9968	1.05e+04	1.80e-14
	max	5000	5001	20.8063	13.1069	1.65e+04	2.04e-14
	var	0	0	3.3725	5.987	2.60e+06	9.46e-31
OptiStiefel2	min	71	76	0.0627	0.0122	6.38e+03	2.14e-15
	mean	116.95	126.0625	0.1104	0.0421	9.98e+03	2.58e-14
	max	900	951	0.8111	0.3551	1.46e+04	9.87e-14
	var	9.47e+03	1.06e+04	7.90e-03	1.80e-03	2.61e+06	1.41e-27
BregmanWOPP2	min	341	342	0.5594	0.0326	6.38e+03	1.59e-14
	mean	605.4125	6.06e+02	1.0548	0.0541	9.98e+03	1.78e-14
	max	4349	4350	8.3407	0.1326	1.46e+04	2.03e-14
	var	3.13e+05	1.13e+05	0.8988	4.61e-04	2.61e+06	8.08e-31
		Ex.9: **Problema 2** con m=75 y n = 75					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Feasi**
OptiStiefel1	min	7.10e+02	746	0.9698	0.0439	4.45e+04	3.84e-15
	mean	1.55e+03	1.62e+03	2.1147	0.102	6.28e+04	4.20e-15
	max	3302	3438	4.5274	0.6227	7.87e+04	4.65e-15
	var	2.94e+05	3.15e+05	0.5656	5.90e-03	5.37e+07	2.84e-32
BregmanWOPP1	min	5000	5001	17.5784	17.2538	4.84e+04	2.77e-14
	mean	5000	5001	17.9971	36.352	6.60e+04	2.96e-14
	max	5000	5001	18.7796	180.1288	8.37e+04	3.19e-14
	var	0	0	0.0348	956.1632	5.37e+07	7.57e-31
OptiStiefel2	min	46	49	0.0601	0.0616	4.45e+04	3.86e-15
	mean	77.9875	82.7375	0.1054	0.1071	6.28e+04	5.76e-14
	max	198	206	0.267	1.097	7.87e+04	9.99e-14
	var	569.2277	631.0821	1.00e-03	1.47e-02	5.37e+07	1.31e-27
BregmanWOPP2	min	340	341	1.2055	0.0758	4.45e+04	2.70e-14
	mean	723.875	7.25e+02	2.5905	0.638	6.28e+04	2.95e-14
	max	4989	4990	18.0812	1.2853	7.87e+04	3.21e-14
	var	1.26e+06	1.26e+06	16.0548	5.26e-02	5.37e+07	8.80e-31

Los valores obtenidos de τ en el Ex.8 fueron: (min,mean,max,var) = (2210,2368,3,2390,2273,3)
Los valores obtenidos de τ en el Ex.9 fueron: (min,mean,max,var) = (5610,5794,6,5810,522,44)

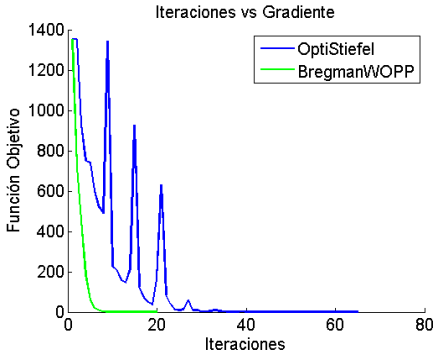
Tabla 5.19: Comparación entre los métodos *OptiStiefel* y *BregmanWOPP* en problemas WOPP's mal condicionados (**Problema 2**)

		Ex.10:**Problema 2** con m=120 y n = 70					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Feasi**
OptiStiefel1	min	1090	1152	2.2387	0.0852	3.85e+04	3.74e-15
	mean	2.68e+03	2.84e+03	5.5401	0.2497	5.06e+04	4.25e-15
	max	5000	5314	10.376	4.2202	6.28e+04	4.92e-15
	var	1.02e+06	1.12e+06	4.3163	0.274	2.76e+07	5.77e-32
BregmanWOPP1	min	5000	5001	19.5731	16.1145	3.86e+04	2.57e-14
	mean	5000	5001	19.8349	57.8981	5.07e+04	2.76e-14
	max	5000	5001	20.6273	1.29e+03	6.28e+04	2.94e-14
	var	0	0	0.0247	3.83e+04	2.75e+07	7.00e-31
OptiStiefel2	min	74	79	0.1482	0.0323	3.85e+04	3.58e-15
	mean	190.0875	204.6625	0.3891	0.1306	5.06e+04	4.23e-15
	max	2026	2139	4.1554	0.329	6.28e+04	4.85e-15
	var	1.07e+05	1.19e+05	4.49e-01	8.97e-04	2.76e+07	7.46e-32
BremanWOPP2	min	446	447	1.7492	0.0977	3.85e+04	2.56e-14
	mean	944.525	9.46e+02	3.7661	0.1587	5.06e+04	2.76e-14
	max	4957	4958	20.8576	0.780	6.28e+04	3.02e-14
	var	1.19e+06	1.19E+06	19.4639	1.97e+04	2.76e+07	9.56e-31

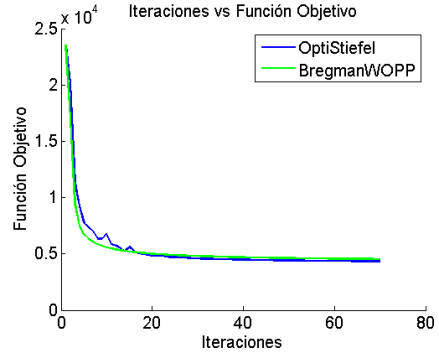
Los valores obtenidos de τ en el Ex.10 fueron: (min,mean,max,var) = (6400,6540,6,6570,2049,1)



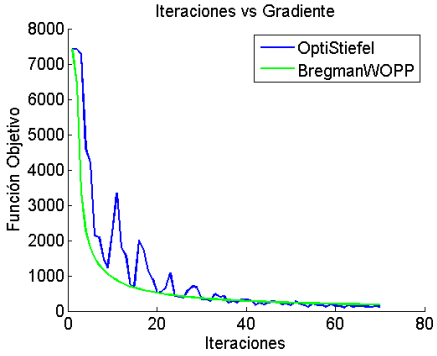
(a) Función objetivo promedio, Ex.5, Tabla 5.16



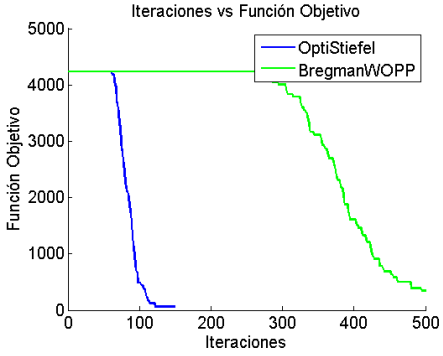
(b) Norma del Gradiente promedio, Ex.5, Tabla 5.16



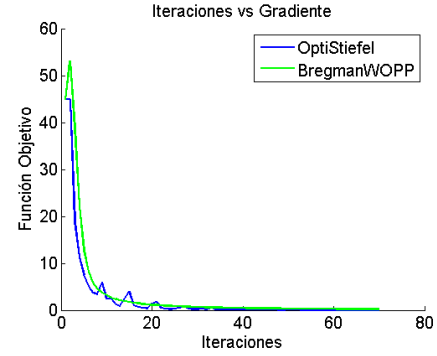
(c) Función objetivo promedio, Ex.6, Tabla 5.17



(d) Norma del Gradiente promedio, Ex.6, Tabla 5.17



(e) Función objetivo promedio, Ex.6, Tabla 5.17



(f) Norma del Gradiente promedio, Ex.6, Tabla 5.17

Figura 5.2: Gráficas comparativas de la función objetivo y de la norma del gradiente promedio entre los métodos *OptiStiefel* y *BregmanWOPP* sobre problemas WOPP's

Nota 13 *Observe que tres de nuestros cuatro métodos (*Ad-Bash*,* Ad-Moul *y* Linear-Co*) utilizan la SVD en cada iteración, como el cálculo de la factorización SVD es más costoso que calcular directamente los valores propios de la matriz A del problema* (5.1)*, no tiene sentido utilizar estos métodos para resolver dicho problema en aplicaciones reales. Sin embargo, en esta sección realizamos experimentos sobre este problema, con la finalidad de estudiar el potencial de nuestros algoritmos, como algoritmos que resuelven problemas generales de optimización sobre variedades.*

*Así, lo que queremos demostrar en esta sección de experimentos es que nuestros métodos propuestos obtienen buenos resultados en cuanto al error (*Error*) de la solución. Note además que nuestro algoritmo* Chol-Retrac*, no emplea uso de un operador de proyección (y por tanto no usa SVD), así que para dicho método si deseamos realizar un estudio comparativo en cuanto a velocidad de convergencia e iteraciones promedio con los algoritmos del estado del arte.*

En los resultados de los experimentos presentados en la tablas 5.20, 5.21 y 5.22 se observa que nuestros cuatro métodos obtienen al menos un error relativo (*Error*) del orden de 1e-7 en promedio y en la mayoría de los experimentos, dicho error es del orden de 1e-10. Además, en los experimentos mostrados en dichas tablas, todos los métodos obtienen un error de factibilidad (*Feasi*) promedio al menos del orden de 1e-14, y en la mayoría de los experimentos, la norma del gradiente (*NrmG*) estuvo por el orden de 1e-2 para todos los métodos, note que no se obtuvo mucha precisión para *NrmG* puesto que se uso una tolerancia de 1e-4 y por tanto todos los métodos en algunos experimentos se detuvieron con los otros criterios de parada en lugar de la norma del gradiente, a pesar de esto, todos los métodos lograron conseguir un punto factible con un error relativo de la función objetivo muy pequeño.

En cuanto a velocidad, observamos claramente que el método que realiza menos tiempo en todos los experimentos es *OptiStiefel*, mientras que el segundo mejor fue *Chol-Retrac* (en los experimentos mostrados en las tablas 5.20 y 5.21), sin embargo para los experimentos de la tabla 5.22 el segundo mejor método en cuanto a tiempo fue *Ad-Moul*. En contra parte, los métodos más lentos fueron *Linear-Co* y *Sgmin* para los problemas con $n = 500$ y $n = 1000$. El algoritmo *Sgmin* fue el más rápido en lo que respecta al número de iteraciones y al número de evaluaciones de la función objetivo promedio, no obstante, dicho algoritmo es muy lento en cuanto al tiempo cuando la dimensión del problema es cada vez más grande.

En general, todos los algoritmos comparados realizan más iteraciones y evaluaciones de la función objetivo a medida que la dimensión del problema crece, en los experimentos presentados en las tablas 5.21 y 5.22, se nota claramente que nuestros métodos *Linear-Co* y *Chol-Retrac* se vuelven muy ineficiente cuando el valor de p aumenta. A pesar de que el algoritmo de *Chol-Retrac* realiza un número cercano de iteraciones que el realizado por el método *OptiStiefel* sobre problemas grandes $n = 1000$, a nuestro algoritmo (*Chol-Retrac*) le toma mucho más tiempo en converger, por tanto para poder obtener beneficio de nuestro algoritmo (*Chol-Retrac*), es recomendable utilizarlo solo en aplicaciones de dimensiones pequeñas.

5.4. Joint diagonalization problem on the Stiefel manifold (JDP)

El "Joint diagonalization problem" (JDP) para N matrices A_1, A_2,..., A_N reales simétricas de tamaño $n \times n$, por lo general es un problema de optimización donde el conjunto de restricciones es el *grupo ortogonal* $\mathcal{O}(n)$. El problema consiste en encontrar una matriz ortogonal de tamaño $n \times n$, que minimize la suma de los cuadrados de las entradas fuera de la diagonal principal de

Tabla 5.20: Resumen computacional de los métodos resolviendo el problema de valores propios lineales sobre matrices densas generadas aleatoriamente fijando $n = 100$ y variando p

p	1	2	3	10	50	95
Sgmin						
Nitr	11.21	12.81	13.43	18.66	29.98	17.02
Nfe	45.84	52.24	54.72	75.64	120.92	69.08
Time	0.14	0.3045	0.3329	0.4521	3.2073	6.3478
NrmG	0.0036	0.0057	0.0072	0.0135	0.0246	0.0238
Fval	383.96	743.4321	1.09e+03	3.16e+03	8.92e+03	9.98e+03
Feasi	5.77e-17	1.89e-16	2.48e-16	6.22e-16	2.60e-15	4.73e-15
Error	1.65e-10	2.65e-10	3.44e-10	9.78e-10	2.28e-09	5.95e-09
OptiStiefel						
Nitr	41.9	51.66	54.56	73.87	101.24	294.21
Nfe	44.86	55.33	58.1	77.94	103.8	313.79
Time	0.0076	0.013	0.0148	0.0242	0.1152	0.5647
NrmG	3.97e-04	4.72e-04	8.82e-04	0.0021	0.0034	0.0044
Fval	383.9603	743.4321	1.09e+03	3.16e+03	8.92e+03	9.98e+03
Feasi	1.80e-14	8.27e-15	1.38e-15	1.16e-15	6.03e-14	5.25e-15
Error	3.96e-12	3.39e-12	7.33e-12	5.56e-11	6.46e-11	1.86e-09
Ad-Bash						
Nitr	138.96	143.9	143.27	146.57	111.53	289.48
Nfe	169.73	177.09	175.74	177.4	120.44	307.29
Time	0.0642	0.0997	0.1083	0.1609	0.3453	1.4694
NrmG	4.60e-04	5.65e-04	8.78e-04	0.0012	0.0037	0.0049
Fval	383.9603	743.4321	1.09e+03	3.16e+03	8.92e+03	9.98e+03
Feasi	2.89e-16	6.90e-16	1.07e-15	3.21e-15	1.21e-14	1.54e-14
Error	2.64e-12	3.04e-12	4.17e-12	2.39e-11	6.50e-11	1.92e-09
Ad-Moul						
Nitr	41.89	50.21	54.34	70.14	100.97	291.99
Nfe	44.35	52.9	56.95	73.02	103.69	310.64
Time	0.0159	0.0282	0.0336	0.0681	0.3537	1.681
NrmG	3.74e-04	6.67e-04	0.0011	0.0015	0.0032	0.0043
Fval	383.96	743.43	1.09e+03	3.16e+03	8.92e+03	9.98e+03
Feasi	3.26e-16	6.49e-16	1.07e-15	3.21e-15	1.14e-14	1.54e-14
Error	3.84e-12	7.20e-12	7.47e-12	3.58e-11	5.26e-11	1.98e-09
Linear-Co						
Nitr	84.04	149.48	157.03	337.01	536.72	4487.70
Nfe	85.18	157.78	165.75	342.4	651.08	5065.10
Time	0.04	0.1115	0.1245	0.4883	1.7922	29.2412
NrmG	0.0014	0.0026	0.0031	0.0073	0.0144	0.0234
Fval	383.9603	743.4321	1.09e+03	3.16e+03	8.92e+03	9.98e+03
Feasi	2.84e-16	6.72e-16	1.30e-15	2.67e-15	1.00e-14	1.52e-14
Error	1.22e-11	2.39e-11	2.86e-11	2.18e-10	4.26e-10	1.88e-07
Chol-Retrac						
Nitr	45.58	53.8	56.7	80.38	122.3	398.67
Nfe	48.5	57.13	60.03	84.4	126.12	416.51
Time	0.0285	0.036	0.0401	0.0701	0.1899	0.9115
NrmG	4.44e-04	5.53e-04	8.15e-04	0.0018	0.0036	0.0051
Fval	383.9603	743.4321	1.09e+03	3.16e+03	8.92e+03	9.98e+03
Feasi	1.87e-14	2.62e-15	2.38e-15	1.09e-15	6.78e-14	5.10e-15
Error	3.65e-12	8.53e-12	7.56e-12	1.03e-10	5.92e-11	3.78e-08

Tabla 5.21: Resumen computacional de los métodos resolviendo el problema de valores propios lineales sobre matrices densas generadas aleatoriamente fijando $n = 500$ y variando p

p	**1**	**2**	**3**	**6**	**10**	**100**
			Sgmin			
Nitr	17.78	19.25	23.16	23.48	23.57	46.8
Nfe	72.12	78	93.64	94.92	95.28	188.2
Time	1.1965	1.3799	2.1639	3.5806	3.8909	34.5653
NrmG	0.0231	0.0343	0.0433	0.0615	0.0787	0.2219
Fval	1.98e+03	3.91e+03	5.80e+03	1.14e+04	1.85e+04	1.34e+05
Feasi	8.88e-17	1.82e-16	2.84e-16	3.63e-16	4.54e-16	4.42e-15
Error	6.65e-10	6.60e-10	1.41e-09	1.36e-09	2.05e-09	5.78e-09
			OptiStiefel			
Nitr	67.61	75.88	83.58	98.36	100.13	145.58
Nfe	70.62	79.26	86.97	102.22	103.79	152
Time	0.037	0.0465	0.0645	0.1664	0.1958	1.8035
NrmG	0.0035	0.005	0.0051	0.0073	0.0095	3.37e-02
Fval	1.98e+03	3.91e+03	5.80e+03	1.14e+04	1.85e+04	1.34e+05
Feasi	4.54e-15	7.59e-16	1.00e-15	1.51e-15	2.02e-15	8.79e-15
Error	2.76e-11	4.52e-11	3.62e-11	4.53e-11	8.07e-11	3.04e-10
			Ad-Bash			
Nitr	146.03	147.9	146.03	147.61	151.49	163.09
Nfe	186.71	187.47	185.03	189.18	192.3	198.44
Time	2.0154	2.0125	2.6025	3.0504	74.7037	7.3169
NrmG	0.0021	0.0036	0.0043	0.0059	0.0071	0.0255
Fval	1.98e+03	3.91e+03	5.80e+03	1.14e+04	1.85e+04	1.34e+05
Feasi	5.33e-16	9.11e-16	1.39e-15	2.57e-15	4.05e-15	2.30e-14
Error	1.13e-11	1.26e-11	9.35e-12	1.02e-11	2.74e-11	1.42e-10
			Ad-Moul			
Nitr	68.21	74.16	83.3	95.56	99.95	146.18
Nfe	70.36	76.38	85.57	98.5	102.49	151.63
Time	0.3638	0.4188	0.6376	0.9622	1.1275	5.5688
NrmG	0.0025	0.0043	0.0059	0.0072	0.0107	0.0392
Fval	1.98e+03	3.91e+03	5.80e+03	1.14e+04	1.85e+04	1.34e+05
Feasi	5.37e-16	9.93e-16	1.36e-15	2.51e-15	3.82e-15	2.27e-14
Error	2.53e-11	3.62e-11	3.47e-11	5.18e-11	7.57e-11	2.70Ee-10
			Linear-Co			
Nitr	200.63	251.75	264.71	352.99	384.44	1035.60
Nfe	201.75	339.03	417.6	483.23	515.27	1134.80
Time	3.5778	4.7737	7.124	9.5033	10.0098	52.166
NrmG	0.0149	0.0148	0.0165	0.0327	0.0415	0.1413
Fval	1.98e+03	3.91e+03	5.80e+03	1.14e+04	1.85e+04	1.34e+05
Feasi	5.57e-16	8.97e-16	1.62e-15	2.76e-15	3.64e-15	1.78e-14
Error	8.87e-11	6.52e-11	7.61e-11	1.32e-10	2.08e-10	1.20e-09
			Chol-Retrac			
Nitr	74.41	87.21	95.05	111.1	112.29	180.69
Nfe	77.36	90.36	98.23	114.97	115.91	188.65
Time	4.9314	5.2923	2.6346	8.1163	7.077	7.7161
NrmG	0.0032	0.0046	0.0055	0.0089	0.011	0.0315
Fval	1.98e+03	3.91e+03	5.80e+03	1.14e+04	1.85e+04	1.34e+05
Feasi	2.86e-15	7.38e-16	9.95e-16	1.54e-15	1.95e-15	8.53e-15
Error	2.20e-11	2.51e-11	3.24e-11	5.52e-11	1.17e-10	3.45e-10

Tabla 5.22: Resumen computacional de los métodos resolviendo el problema de valores propios lineales sobre matrices densas generadas aleatoriamente fijando $n = 1000$ y variando p

p	**1**	**2**	**3**	**6**	**10**	**100**
			Sgmin			
Nitr	21.78	22.24	26.7	30.84	30.88	54.44
Nfe	88.12	89.96	107.8	124.36	124.52	218.76
Time	4.9506	6.2456	9.0335	14.0035	15.487	85.1332
NrmG	0.05	0.0734	0.0894	0.1281	0.1656	0.4794
Fval	3.97e+03	7.89e+03	1.18e+04	2.32e+04	3.81e+04	3.14e+05
Feasi	8.10e-17	2.45e-16	2.82e-16	5.55e-16	7.98e-16	4.33e-15
Error	1.11e-09	1.22e-09	1.84e-09	1.21e-06	2.59e-09	6.89e-09
			OptiStiefel			
Nitr	90.7	89.9	104.96	119.99	123.88	161.69
Nfe	94.46	93.1	109.4	125.4	129.38	170.18
Time	0.1602	0.1824	0.2993	0.5403	0.6618	4.2221
NrmG	0.005	0.0101	0.0176	0.0177	0.0229	0.0768
Fval	3.97e+03	7.89e+03	1.18e+04	2.32e+04	3.81e+04	3.14e+05
Feasi	1.23e-15	1.09e-15	1.50e-15	2.40e-15	3.15e-15	1.20e-14
Error	4.04e-11	3.59e-11	7.77e-11	1.07e-10	1.32e-10	1.09e-07
			Ad-Bash			
Nitr	156.99	149.39	158.4	157.29	159.33	178.81
Nfe	199.45	190.03	202.57	202.89	205.01	228.46
Time	8.0621	9.3993	11.0763	14.0338	16.7469	25.2045
NrmG	0.0054	0.0075	0.0075	0.0147	0.0134	0.0659
Fval	3.97e+03	7.89e+03	1.18e+04	2.32e+04	3.81e+04	3.14e+05
Feasi	6.89e-16	1.28e-15	1.80e-15	3.19e-15	4.61e-15	2.38e-14
Error	3.28e-11	8.20e-12	1.11e-11	4.17e-11	7.84e-11	1.69e-10
			Ad-Moul			
Nitr	92.08	89.76	102.67	120.17	126.47	160.4
Nfe	94.81	92.11	105.79	124.4	131.21	167.45
Time	1.8272	2.201	3.0029	5.175	7.0678	16.3578
NrmG	0.0092	0.0095	0.014	0.021	0.0228	0.0734
Fval	3.97e+03	7.89e+03	1.18e+04	2.32e+04	3.81e+04	3.14e+05
Feasi	7.13e-16	1.22e-15	1.80e-15	3.01e-15	4.57e-15	2.32e-14
Error	7.61e-11	3.79e-11	1.04e-10	1.03e-10	1.32e-10	1.09e-07
			Linear-Co			
Nitr	430.13	346.89	505.19	621.39	686.04	1.18e+03
Nfe	431.14	348.29	507.8	625.81	693.67	1.22e+03
Time	29.6307	29.7896	46.8031	67.5162	82.9327	186.7298
NrmG	0.0292	0.0413	0.0506	0.071	0.0926	0.3123
Fval	3.97e+03	7.89e+03	1.18e+04	2.32e+04	3.81e+04	3.14e+05
Feasi	7.43e-16	1.28e-15	1.81e-15	3.07e-15	3.95e-15	1.86e-14
Error	1.34e-10	1.05e-10	1.76e-10	2.44e-10	2.92e-10	1.13e-09
			Chol-Retrac			
Nitr	86.15	86.6	103.7	132.22	142.1	187.26
Nfe	87.27	89.2	106.8	136.6	147.2	196.2
Time	31.4545	30.9952	38.6503	53.665	61.1489	99.0271
NrmG	0.0017	0.0133	0.0134	0.0508	0.0165	0.0592
Fval	3.98e+03	7.90e+03	1.18e+04	2.33e+04	3.81e+04	3.13e+05
Feasi	4.44e-15	1.01e-15	1.58e-15	2.72e-15	2.87e-15	9.76e-15
Error	8.27e-13	2.15e-11	5.06e-11	6.21e-11	4.23e-10	2.43e-10

las matrices $X^\top A_l X,\ l = 1, 2, \ldots, N$, o equivalentemente, maximize la suma de los cuadrados de las entradas de las diagonales principales de las matrices $X^\top A_l X,\ l = 1, 2, \ldots, N$, ver [53]. Conseguir una solución para este problema (JDP) es de gran valor para aplicaciones como *Independent Component Analysis* (ICA) y también para *The Blind Source Separation Problem* ver [49].

En [15, 27] se han estudiado distintos métodos sobre variedades para resolver el problema JDP sobre la variedad de Stiefel en lugar del grupo ortogonal, más específicamente, consideran el siguiente problema:

$$\text{minimizar } \mathcal{F}(X) = -\sum_{l=1}^{N} ||\text{diag}(X^\top A_l X)||_F^2, \quad \text{s.a.} \quad X \in \text{St}(n, p), \tag{5.3}$$

donde A_1, A_2,..., A_N son matrices reales simétricas de tamaño $n \times n$ y donde $diag(Y)$ es la matriz cuya diagonal principal coincide con la diagonal principal de la matriz Y y el resto de entradas son nulas.

En esta sección probamos la eficiencia de nuestros algoritmos al momento de resolver el problema (5.3), además comparamos nuestros métodos con dos de los métodos del estado del arte (*OptiStiefel* y *MSteepest*). Para probar los métodos, fijamos $N = 10$ y construimos 100 problemas (100 JDPs) generando cada una de las N matrices $A_1, A_2, \ldots, A_N$ de la siguiente forma: primero generamos aleatoriamente una matriz $\bar{A} \in \mathbb{R}^{n \times n}$ con todas sus entradas siguiendo una distribución *Normal estándar*, luego construimos A_l por $A_l = \bar{A}^\top \bar{A}$ y así aseguramos que cada A_l sea simétrica. Por otra parte, generamos al punto inicial X_0 (con el cual empezaran a iterar los algoritmos) aleatoriamente en la variedad de Stiefel. En las tablas: 5.23, 5.24, 5.25 y 5.26 se muestran los resultados obtenidos por los métodos para diferentes tamaños de problemas, en dichas tablas mostramos el mínimo, media, máximo y varianza (denotados por *min*, *mean*, *max*, *var*) de los valores a comparar. Además, el número máximo de iteraciones lo fijamos en 10000, utilizamos una tolerancia para la norma del gradiente de $\epsilon =$ 1e-5 y tomamos los siguientes valores: $xtol =$ 1e-12 y $ftol =$1e-15 como tolerancias para los otros criterios de parada. Asimismo, para nuestro método *Linear-Co* usamos la fórmula de actualización (4.8) pero ajustando los parámetros (λ y μ) por $\lambda = 3/4$ y $\mu = 1/4$ (en lugar de $\lambda = 2/3$ y $\mu = 1/3$ que fue como se utilizaron en todos los experimentos anteriores).

En las tablas 5.23, 5.24, 5.25 y 5.26 los métodos que obtienen el peor desempeño son nuestros métodos *Linear-Co* y *Ad-bash* puesto que son a los que les tomas más tiempo e iteraciones en promedio en converger, de hecho en los experimentos Ex.1 y Ex.2 mostrados en las tablas 5.23 y 5.24 en ocasiones dichos métodos realizan el número máximo de iteraciones. Por otra parte, en los cuatro experimentos mostrados en dichas cuatro tablas, vemos que nuestro método *Ad-Moul* muestra un comportamiento muy similar al método del estado del arte *MSteepest*, inclusive le toma ligeramente menos tiempo en promedio llegar a converger que al método *MSteepest*. En general, los dos métodos más eficientes en los experimentos mostrados en las tablas 5.23, 5.24 fueron *OptiStiefel* y *Chol-Retrac* debido a que fueron los métodos que convergieron más rápido en promedio, sin embargo el método que realizó menos iteraciones en los cuatro experimentos fue el método *MSteepest*. En las tablas 5.23, 5.24 observamos que los métodos *OptiStiefel* y *Chol-Retrac*, muestran resultados muy parecidos y que inclusive en la tabla 5.24, nuestro método *Chol-Retrac* le toma menos iteraciones y menos evaluaciones de la función objetivo en promedio para converger que al método *OptiStiefel*; mientras que el tiempo promedio fue casi el mismo para ambos métodos. Por otro lado, para experimentos de tallas más grandes (ver tablas 5.25 y 5.26),

observamos que nuestro método *Chol-Retrac* obtiene buenos resultados en cuanto a predicción, no obstante dura mucho tiempo en converger. Note que en los experimentos Ex.3 y Ex.4 (tablas 5.25 y 5.26) se observa además que los tres métodos más eficientes fueron los dos del estado del arte junto con nuestro método *Ad-Moul*.

De los resultados mostrados en las tablas 5.23, 5.24, 5.25 y 5.26 podemos concluir que para problemas de talla pequeña, los mejores métodos fueron *OptiStiefel* y *Chol-retrac*, mientras que para problemas de talla más grande (con $n \gg p$), los dos mejores métodos fueron *OptiStiefel* y nuestro *Ad-Moul*. Sin embargo es de hacer notar que todos los métodos funcionan bien puesto que obtienen una matriz que satisface las condiciones de optimalidad de primer orden con cierta tolerancia si se utiliza un número máximo de iteraciones y tolerancia apropiados.

En la Figura 5.3 presentamos tanto la gráfica de las evaluaciones de la función objetivo promedio versus las primeras 25 iteraciones como también la gráfica de la norma del gradiente promedio versus las primeras 60 iteraciones, para cada uno de los métodos comparados en los experimentos Ex.2 y Ex.3 cuyos resultados están contenidos en las tablas 5.24 y 5.25 respectivamente. En dichas gráficas se observa claramente que todos los métodos a excepción del método *Linear-Co* muestran un comportamiento muy similar tanto en las evaluaciones de la función objetivo como en la norma del gradiente promedio. Además, observamos que nuestro método *Linear-Co* desciende la función objetivo más lento que el resto de los métodos en ambos experimentos y de igual manera ocurre con la norma del gradiente, sin embargo notamos que dicho método muestra una tendencia decreciente para las dos gráficas en ambos experimentos.

Nota 14 *En la Figura 5.3 solo mostramos las gráficas de los experimentos Ex.2 y Ex.3. Sin embargo, las gráficas para los experimentos Ex.1 y Ex.4 mostraron un comportamiento muy similar para todos los métodos y por tanto las omitimos.*

Tabla 5.23: Desempeño de los métodos en problemas JDP's sobre la variedad de Stiefel generados aleatoriamente con n = 100 y p = 75

		Ex.1: n=100 y p = 75					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Feasi**
OptiStiefel	min	560	596	2.81	4.42e-06	-1.2664e+07	3.72e-15
	mean	1060.00	1128.90	5.38	2.97e-05	-1.2373e+07	4.30e-15
	max	6295	6455	31.91	5.20e-04	-1.2087e+07	4.81e-15
	var	3.79e+05	4.01e+05	9.72	3.77e-09	1.6781e+10	5.02e-32
Ad-Bash	min	572	831	5.89	1.55e-06	-1.2669e+07	1.10e-14
	mean	3281.40	4302.00	36.03	2.30e-03	-1.2374e+07	1.18e-14
	max	10000	12761	120.10	3.30e-02	-1.2095e+07	1.26e-14
	var	1.01e+07	1.55e+07	1369.00	3.03e-05	1.6618e+10	1.31e-31
Ad-Moul	min	602	670	5.27	4.60e-06	-1.2668e+07	1.11e-14
	mean	1074.30	1177.80	9.56	2.08e-04	-1.2373e+07	1.18e-14
	max	2392	2545	20.94	4.10e-03	-1.2083e+07	1.27e-14
	var	1.12e+05	1.31e+05	9.02	3.29e-07	1.6859e+10	1.08e-31
Linear-Co	min	2541	3164	24.06	2.14e-05	-1.2663e+07	1.12e-14
	mean	6397.10	7865.40	70.89	2.66e-01	-1.2373e+07	1.23e-14
	max	10000	12819	131.80	9.27	-1.2090e+07	2.28e-14
	var	6.15e+06	9.66e+06	1178.00	1.65	1.6689e+10	3.94e-30
Msteepest	min	456	609	5.04	2.30e-03	-1.2667e+07	1.08e-14
	mean	930.80	1169.90	9.66	1.48e-02	-1.2373e+07	1.18e-14
	max	1592	1917	15.98	2.67e-02	-1.2095e+07	1.24e-14
	var	7.54e+04	1.11e+05	7.75	3.77e-05	1.6688e+10	1.22e-31
Chol-Retrac	min	528	559	2.87	3.37e-06	-1.2661e+07	3.73e-15
	mean	1116.10	1185.10	6.09	8.12e-06	-1.2373e+07	4.30e-15
	max	2241	2336	12.52	9.99e-06	-1.2096e+07	4.96e-15
	var	1.30e+05	1.43e+05	3.79	3.23e-12	1.6488e+10	4.87e-32

Tabla 5.24: Desempeño de los métodos en problemas JDP's sobre el grupo ortogonal generados aleatoriamente con n = 100 y p = 100

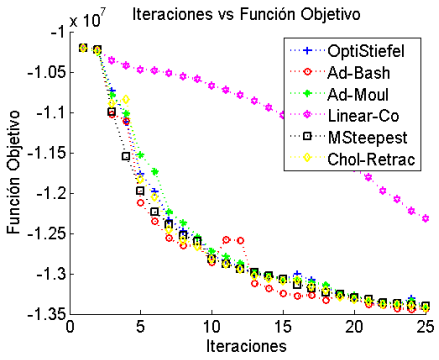
		Ex.2: n=100 y p = 100					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Feasi**
OptiStiefel	min	780	837	7.26	6.10e-06	-1.3974e+07	4.93e-15
	mean	1578.20	1663.60	16.42	1.17e-04	-1.3706e+07	5.23e-15
	max	4432	4578	44.90	2.60e-03	-1.3290e+07	5.91e-15
	var	4.53e+05	4.78e+05	46.03	1.75e-07	1.9696e+10	2.81e-32
Ad-Bash	min	939	1329	21.14	4.88e-06	-1.3969e+07	1.40E-14
	mean	2597.40	3482.30	54.94	3.40e-03	-1.3707e+07	1.49e-14
	max	10000	12807	224.79	1.22e-01	-1.3300e+07	1.57e-14
	var	3.59e+06	5.59e+06	1723.40	3.02e-04	1.9721e+10	1.43e-31
Ad-Moul	min	713	776	11.80	6.71e-06	-1.3972e+07	1.40e-14
	mean	1314.30	1426.20	23.36	4.71e-04	-1.3706e+07	1.48e-14
	max	2564	2773	46.09	5.20e-03	-1.3296e+07	1.55e-14
	var	1.66e+05	1.90e+05	54.84	8.92e-07	1.9502e+10	1.13e-31
Linear-Co	min	5159	6273	105.09	2.39e-05	-1.3973e+07	1.42e-14
	mean	8926.90	10897.00	194.71	3.88e+01	-1.3706e+07	1.80e-14
	max	10000	12505	236.72	9.35e+02	-1.3292e+07	3.30e-14
	var	2.09e+06	3.23e+06	1543.70	2.40e+04	1.9770e+10	3.37e-29
Msteepest	min	661	810	12.30	5.00e-03	-1.3972e+07	1.35e-14
	mean	1263.40	1565.70	25.05	1.63e-02	-1.3706e+07	1.49e-14
	max	2534	3058	48.64	2.93e-02	-1.3296e+07	1.59e-14
	var	1.87e+05	2.66e+05	67.97	3.30e-05	1.9892e+10	1.61e-31
Chol-Retrac	min	854	914	9.84	3.84e-06	-1.3964e+07	4.66e-15
	mean	1543.60	1622.90	16.98	1.06e-04	-1.3708e+07	5.08e-15
	max	3264	3424	36.13	1.80e-03	-1.3299e+07	5.44e-15
	var	3.18e+05	3.46e+05	37.48	1.10e-07	1.9508e+10	3.22e-32

Tabla 5.25: Desempeño de los métodos en problemas JDP's sobre la variedad de Stiefel generados aleatoriamente con n = 300 y p = 20

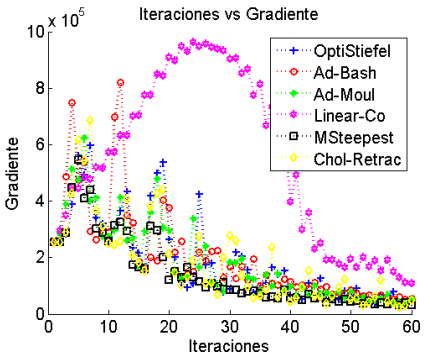
		Ex.3: n=300 y p = 20					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Feasi**
OptiStiefel	min	320	343	3.11	4.80e-03	-4.9755e+07	1.65e-15
	mean	619.20	674.74	6.65	9.44e-02	-4.9146e+07	2.42e-15
	max	1196	1296	13.39	7.13e-01	-4.8452e+07	3.19e-15
	var	3.51e+04	4.09e+04	4.33	2.45e-02	8.6397e+10	1.26e-31
Ad-Bash	min	330	451	7.74	3.66e-05	-4.9755e+07	3.74e-15
	mean	696.96	1003.70	16.35	1.83e-02	-4.9152e+07	5.08e-15
	max	1633	2256	38.28	3.73e-01	-4.8452e+07	6.47e-15
	var	7.52e+04	1.55e+05	42.54	2.80e-03	8.7358e+10	2.64e-31
Ad-Moul	min	284	308	4.70	6.30e-04	-4.9755e+07	4.31e-15
	mean	554.64	622.08	9.85	7.84e-02	-4.9146e+07	5.09e-15
	max	1538	1679	26.63	6.33e-01	-4.8452e+07	6.92e-15
	var	4.95e+04	6.02e+04	15.54	1.45e-02	8.7175e+10	2.42e-31
Linear-Co	min	969	1131	19.84	2.48e-04	-4.9737e+07	3.88e-15
	mean	2155.40	2745.90	49.74	1.80e-02	-4.9156e+07	5.05e-15
	max	8094	9050	174.72	8.55e-02	-4.8443e+07	6.09e-15
	var	2.16e+06	3.23e+06	1227.90	4.54e-04	8.6413e+10	2.22e-31
Msteepest	min	311	406	6.88	5.41e-02	-4.9755e+07	3.98e-15
	mean	513.58	654.06	11.78	2.12e-01	-4.9148e+07	5.07e-15
	max	777	981	17.36	4.21e-01	-4.8463e+07	6.60e-15
	var	1.44e+04	2.05e+04	6.71	8.70e-03	8.8309e+10	3.00e-31
Chol-Retrac	min	356	376	11.07	3.34e-05	-4.9755e+07	2.12e-15
	mean	586.02	631.54	18.38	4.80e-03	-4.9152e+07	4.56e-14
	max	1088	1145	32.68	4.94e-02	-4.8440e+07	9.90e-14
	var	3.15e+04	3.65e+04	30.25	9.01e-05	8.8413e+10	4.87e-28

Tabla 5.26: Desempeño de los métodos en problemas JDP's sobre la variedad de Stiefel generados aleatoriamente con n = 500 y p = 15

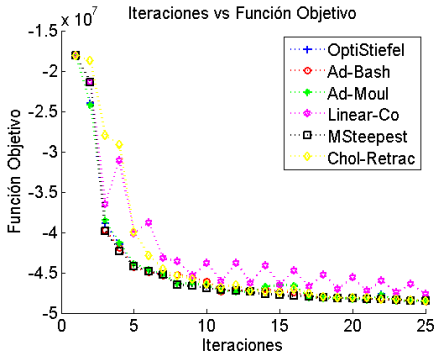
		Ex.4: n=500 y p = 15					
Método		**Nitr**	**Nfe**	**Time**	**NrmG**	**Fval**	**Feasi**
OptiStiefel	min	304	328	10.22	1.45e-02	-1.0749e+08	1.73e-15
	mean	581.68	631.28	20.74	1.48e-01	-1.0639e+08	2.54e-15
	max	1488	1629	56.64	4.98	-1.0550e+08	3.50e-15
	var	4.11e+04	4.86e+04	56.68	5.44e+01	1.7494e+11	1.63e-31
Ad-Bash	min	289	391	19.59	1.10e-03	-1.0749e+08	2.79e-15
	mean	654.88	957.30	47.27	6.80e-02	-1.0640e+08	3.77e-15
	max	1321	1912	93.92	2.27e-01	-1.0547e+08	5.03e-15
	var	4.90e+04	1.11e+05	273.33	5.20e-03	1.8007e+11	2.45e-31
Ad-Moul	min	294	324	14.49	5.59e-04	-1.0749e+08	2.77e-15
	mean	567.38	624.44	28.65	1.22e-01	-1.0639e+08	3.84e-15
	max	2354	2479	117.62	1.14	-1.0538e+08	5.66e-15
	var	1.11e+05	1.21e+05	276.59	4.15e-02	1.8057e+11	3.95e-31
Linear-Co	min	651	1041	54.14	1.90e-03	-1.0749e+08	2.48e-15
	mean	1694.30	2551.80	136.83	9.48e-02	-1.0640e+08	3.85e-15
	max	7608	13068	733.25	3.46e-01	-1.0542e+08	5.23e-15
	var	1.40e+06	3.43e+06	10823.00	5.00e-03	1.8517e+11	3.44e-31
Msteepest	min	244	326	17.02	8.39e-02	-1.0749e+08	2.82e-15
	mean	441.88	559.44	29.60	6.62e-01	-1.0640e+08	3.77e-15
	max	1074	1299	69.18	1.48	-1.0545e+08	5.31e-15
	var	2.31e+04	3.37e+04	91.11	1.01e-01	1.8446e+11	3.00e-31
Chol-Retrac	min	339	365	40.08	2.79e-04	-1.0749e+08	1.33e-15
	mean	545.40	593.52	67.48	2.27e-02	-1.0640e+08	5.22e-14
	max	904	984	117.08	2.35e-01	-1.0548e+08	9.30e-14
	var	1.86e+04	2.25e+04	305.99	1.90e-03	1.8100e+11	3.13e-28



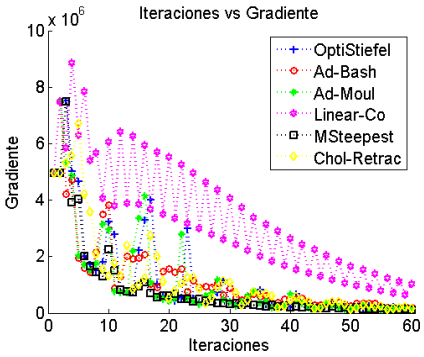
(a) Función objetivo promedio, Ex.2, Tabla 5.24



(b) Norma del Gradiente promedio, Ex.2, Tabla 5.24



(c) Función objetivo promedio, Ex.3, Tabla 5.25



(d) Norma del Gradiente promedio, Ex.3, Tabla 5.25

Figura 5.3: Gráficas comparativas de la función objetivo y de la norma del gradiente promedio de los métodos sobre problemas JDP's

Capítulo 6

Conclusiones generales y trabajo futuro

En esta tesis se abordó el problema de optimización con restricciones de ortogonalidad, con el interés de implementar y proponer algoritmos eficientes para resolver dicho problema. El trabajo realizado se basó en estudiar y proponer algoritmos de búsqueda lineal sobre variedades, inspirado en los métodos para resolver ecuaciones diferenciales numéricamente de *Adams-Bashforth* y *Adams-Moulton*, combinado con un esquema de proyección, así como también, se propuso un método basado en *retracción* para resolver problemas generales de optimización sobre la variedad de Stiefel. Por otra parte, se investigó un problema particular *Weighted Orthogonal Procrustes Problem* (WOPP), para el cual se obtuvo una reformulación equivalente a dicho problema y se propuso utilizar el conocido algoritmo *Iteración de Bregman* para resolver dicha reformulación. Asimismo, en este trabajo especial de grado, se realizó el análisis de convergencia para el Algoritmo 2. Además se demostraron varios resultados teóricos para cada uno de los métodos propuestos.

6.1. Principales contribuciones

1. Se propusieron métodos tipo *Adams-Bashforth*, *Adams-Moulton* y uno basado es una combinación lineal para resolver problemas sobre la variedad de Stiefel. De los experimentos numéricos realizados, se observó que entre estas tres propuestas se recomienda utilizar el método tipo *Adams-Moulton* puesto que este obtuvo un mejor desempeño.

2. Se Generalizó el método basado en la *transformada de Cayley* presentado en [21] a través de la factorización de Cholesky de una matriz determinada, además, se demostró que dicha generalización es una *retracción* sobre el *grupo ortogonal*, mientras que para el caso de optimización sobre la variedad de Stiefel, se obtuvo un resultado de convergencia.

3. Se propuso el conocido algoritmo *Iteración de Bregman* para resolver una reformulación equivalente del problema *Weighted Orthogonal Procrustes Problem* (WOPP). De los experimentos numéricos realizados, se concluye que este método propuesto es más eficiente que los métodos del estado del arte considerados en este trabajo sobre problemas bien condicionados y en problemas donde el valor optimal es igual a cero.

4. Se realizó un estudio teórico detallado de cada uno de nuestros métodos propuestos.

5. Se desarrolló un estudio comparativo de los métodos propuestos en esta tesis con algunos de los algoritmos del estado del arte, donde se obtuvieron buenos resultados en cuanto

a precisión de la solución y en ocasiones en tiempo de convergencia de los métodos propuestos versus algunos métodos del estado del arte, resolviendo problemas simulados de optimización sobre la variedad de Stiefel y sobre el grupo ortogonal.

6.2. Trabajo futuro

En esta sección se señalan algunos problemas abiertos, y algunos temas que no fueron considerados en esta tesis y que pudieran investigarse para desarrollar mejores métodos para resolver problemas con restricciones de ortogonalidad:

1. Estudiar si en realidad las curvas construidas por los métodos tipo Adams-Bashforth y Adams-Moulton (ver ecuaciones (4.9) y (4.10)) son curvas de descenso en $\tau = 0$.

2. Responder la siguiente pregunta: ¿Siempre existen valores del tamaño de paso τ que satisfacen las condiciones de descenso tipo Armijo-Wolfe presentadas en las ecuaciones (4.43a)-(4.43b) simultáneamente?.

3. Estudiar posibles mejoras para los métodos tipo Adams-Bashforth y el de la combinación lineal, puesto que fueron las propuestas que tuvieron un peor desempeño.

4. Estudiar e implementar algoritmos tipo Newton, Quasi-Newton, Gradiente Conjugado, Región de Confianza, entre otros, sobre variedades para abordar problemas de optimización sobre la variedad de Stiefel.

5. Investigar el desempeño de algoritmos infactibles como por ejemplo el algoritmo del Lagrangiano aumentado (ALM) y los algoritmos de penalización para resolver el problema investigado en esta tesis.

6. Estudiar el desempeño de los algoritmos propuestos en aplicaciones tales como PCA, LDA, separación de imágenes entre otras.

Anexos

En este apartado, presentamos las demostraciones de algunos teoremas, lemas y proposiciones enunciadas en los capítulos 2 y 4 con la finalidad de complementar el presente trabajo de investigación. Todos estas demostraciones se encuentran contenidas en [28, 33, 51, 52].

Teorema 13 *Si f es una función definida en un abierto A de un espacio vectorial normado E, con valores en un espacio vectorial normado F, diferenciable en $a \in A$, entonces ella es continua en a, parcialmente derivable en a y, se tiene la igualdad*

$$\mathcal{D}f(a)(v) = \mathcal{D}f(a, v), \forall v \in E \tag{6.1}$$

Demostración.
Como f es diferenciable en a de la definición 21 haciendo $x = a + h$ obtenemos que para todo $\epsilon > 0$ existe $\delta > 0$ tales que:

$$\begin{aligned} ||x - a|| < \delta &\Rightarrow ||f(x) - f(a) - \mathcal{D}f(a)(x - a)|| < \epsilon||x - a|| \\ &\Rightarrow ||f(x) - f(a)|| - ||\mathcal{D}f(a)(x - a)|| < \epsilon||x - a|| \end{aligned} \tag{6.2}$$

y como $\mathcal{D}f(a)$ es una función lineal continua (Lipschitz con constante de Lipschitz $||\mathcal{D}f(a)||$), obtenemos

$$\begin{aligned} ||x - a|| < \delta &\Rightarrow ||f(x) - f(a)|| < ||\mathcal{D}f(a)(x - a)|| + \epsilon||x - a|| \\ &\Rightarrow ||f(x) - f(a)|| < ||\mathcal{D}f(a)||||x - a|| + \epsilon||x - a|| \\ &\Rightarrow ||f(x) - f(a)|| < (||\mathcal{D}f(a)|| + \epsilon)||x - a|| \end{aligned} \tag{6.3}$$

lo que implica que f es continua en a.
Ahora, dado $v \in E$ no nulo, para $h = tv$ en la definición 21, tenemos que:

$$\lim_{t \to 0} \frac{f(a + tv) - f(a) - \mathcal{D}f(a)(tv)}{|t|||v||} = 0$$

multiplicando por $||v||$ y usando la linealidad de $\mathcal{D}f(a)$ obtenemos,

$$\lim_{t \to 0} \frac{f(a + tv) - f(a)}{t} = \mathcal{D}f(a)(v)$$

y por lo tanto, $\mathcal{D}f(a)(v) = \mathcal{D}f(a, v), \forall v \in E$. El caso para $v = 0$ es trivial.□

Teorema 14 *Sea $\mathcal{F}:\mathbb{R}^{m\times n}\to\mathbb{R}$. Si $\mathcal{F}$ es diferenciable en $X\in\mathbb{R}^{m\times n}$ entonces:*

$$G_X=\Big[\frac{\partial\mathcal{F}(X)}{\partial x_{ij}}\Big]_{1\leq i\leq m,1\leq j\leq n}$$

Demostración.
Como $\mathcal{F}$ es diferenciable en X por la definición 24 tenemos que existe una matriz $G_X\in\mathbb{R}^{m\times n}$ tal que:

$$\lim_{H\to\mathbf{0}}\frac{\mathcal{F}(X+H)-\mathcal{F}(X)-\langle G_X,H\rangle}{||H||}=0$$

Como $G_X\in\mathbb{R}^{m\times n}$, entonces:

$$G_X=\begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1n}\\ g_{21} & g_{22} & \cdots & g_{2n}\\ \vdots & \vdots & \ddots & \\ g_{m1} & g_{m2} & \cdots & g_{mn}\end{pmatrix}$$

luego, de la definición de derivada direccional y de derivada parcial tenemos que para la matriz canónica E_{ij} con $1\leq i\leq m, 1\leq j\leq n$ arbitrarios se cumple:

$$\frac{\partial\mathcal{F}(X)}{\partial x_{ij}}=\mathcal{D}\mathcal{F}(X,E_{ij}) \tag{6.4}$$

Por otro lado, por el teorema 7 y como $\mathcal{D}\mathcal{F}(X)(E_{ij})=\langle G_X,E_{ij}\rangle$ obtenemos:

$$\mathcal{D}\mathcal{F}(X,E_{i,j})=\langle G_X,E_{ij}\rangle$$

o equivalentemente,

$$\mathcal{D}\mathcal{F}(X,E_{i,j})=g_{ij} \tag{6.5}$$

de (6.4) y (6.5) obtenemos:

$$\frac{\partial\mathcal{F}(X)}{\partial x_{ij}}=g_{ij}$$

como i,j son arbitrarios concluimos que G_X es la matriz de derivadas parciales de la función $\mathcal{F}$ en X.□

Teorema 15 *Sea $\mathcal{F}:\mathbb{R}^{m\times n}\to\mathbb{R}$ una función diferenciable en $X\in\mathbb{R}^{m\times n}$, y $Z\in\mathbb{R}^{m\times n}$. Entonces,*

$$\mathcal{D}\mathcal{F}(X,Z):=\lim_{t\to0}\frac{\mathcal{F}(X+tZ)-\mathcal{F}(X)}{t}=\langle G_X,Z\rangle$$

donde G_X denota a la matriz de derivadas parciales de $\mathcal{F}$ en X.

Demostración.
Como $\mathbb{R}^{m\times n}$ es un espacio vectorial normado y $\mathcal{F}$ es diferenciable en X, entonces por el teorema 7 tenemos que:

$$\mathcal{D}\mathcal{F}(X,Z)=\mathcal{D}\mathcal{F}(X)(Z),\quad\forall Z\in\mathbb{R}^{m\times n}$$

además, por el Teorema de Representación de Riesz y el teorema 9 tenemos que $\mathcal{DF}(X) \in \mathcal{L}(\mathbb{R}^{m\times n}, \mathbb{R})$ se puede escribir como:

$$\mathcal{DF}(X)(Z) = \langle G_X, Z\rangle, \quad \forall Z \in \mathbb{R}^{m\times n}$$

donde G_X es la matriz de derivadas parciales de F en X. Luego,

$$\mathcal{DF}(X, Z) = \langle G_X, Z\rangle, \quad \forall Z \in \mathbb{R}^{m\times n}.$$

□

Proposición 11 *La derivada de la función $F : \mathbb{R}^{m\times n} \to R$ dada por*

$$F(X) = \frac{1}{2}||AXC - B||_F^2$$

donde $A \in \mathbb{R}^{p\times m}$, $B \in \mathbb{R}^{p\times q}$ y $C \in \mathbb{R}^{n\times q}$ es:

$$G_X = A^\top(AXC - B)C^\top$$

Demostración.
Sean $X, H \in \mathbb{R}^{m\times n}$,

$$\begin{aligned} F(X+H) - F(X) &= \frac{1}{2}(\,||A(X+H)C - B||_F^2 - ||AXC - B||_F^2\,) \\ &= \frac{1}{2}(\,\langle A(X+H)C - B, A(X+H)C - B\rangle - ||AXC - B||_F^2\,) \\ &= \frac{1}{2}(\,||AXC - B||_F^2 + 2\langle AXC - B, AHC\rangle + ||AHC||_F^2 - ||AXC - B||_F^2\,) \\ &= \langle AXC - B, AHC\rangle + \frac{1}{2}||AHC||_F^2 \end{aligned} \tag{6.6}$$

luego,

$$\begin{aligned} F(X+H) - F(X) - \langle G_X, H\rangle &= \langle AXC - B, AHC\rangle + \frac{1}{2}||AHC||_F^2 - \langle G_X, H\rangle \\ &= Tr[(C^\top X^\top A^\top - B^\top)AHC] + ||AHC||_F^2 \\ &\quad -Tr[C(C^\top X^\top A^\top - B^\top)AH] \\ &= \frac{1}{2}||AHC||_F^2 \qquad (\text{ya que } Tr[XY] = Tr[YX]) \end{aligned}$$

Así,

$$\begin{aligned} |F(X+H) - F(X) - \langle G_X, H\rangle| &= \frac{1}{2}||AHC||_F^2 \\ &\leq \frac{q}{2}||A||_F^2||H||_F^2||C||_F^2 \end{aligned} \tag{6.7}$$

luego, dado $\epsilon > 0$ existe $\delta = \frac{2\epsilon}{q||A||_F^2||C||_F^2}$ tal que:

$$
\begin{aligned}
||H||_F < \delta \Rightarrow \frac{|F(X+H) - F(X) - \langle G_X, H\rangle|}{||H||_F} &\leq \frac{q||A||_F^2||H||_F^2||C||_F^2}{2||H||_F} \\
&= \frac{q}{2}||A||_F^2||H||_F||C||_F^2 \\
&< \delta\frac{q}{2}||A||_F^2||C||_F^2 \\
&= \frac{2\epsilon}{q||A||_F^2||C||_F^2}\frac{q}{2}||A||_F^2||C||_F^2 \\
&= \epsilon
\end{aligned}
$$

Por lo tanto,

$$\lim_{H\to\mathbf{0}} \frac{F(X+H) - F(X) - \langle G_X, H\rangle}{||H||} = 0.$$

□

Proposición 12 *La derivada de la función $F : \mathbb{R}^{m\times n} \to R$ dada por*

$$F(X) = Tr[X^\top AX]$$

donde $A \in \mathbb{R}^{m\times m}$ es una matriz simétrica, es:

$$G_X = 2AX$$

Demostración.
Sean $X, H \in \mathbb{R}^{m\times n}$,

$$
\begin{aligned}
F(X+H) - F(X) &= Tr[(X+H)^\top A(X+H)] - Tr[X^\top AX] \\
&= Tr[(X^\top + H^\top)A(X+H)] - Tr[X^\top AX] \\
&= (\,Tr[X^\top AX] + Tr[X^\top AH] + Tr[H^\top AX] + Tr[H^\top AH]\,) - Tr[X^\top AX] \\
&= Tr[X^\top AH] + Tr[H^\top AX] + Tr[H^\top AH] \\
&= 2Tr[X^\top AH] + Tr[H^\top AH], \qquad \text{(ya que } Tr[Z] = Tr[Z^\top] \text{ y } A^\top = A)
\end{aligned}
$$

luego,

$$
\begin{aligned}
|\,F(X+H) - F(X) - \langle G_X, H\rangle| &= |2Tr[X^\top AH] + Tr[H^\top AH] - Tr[(2AX)^\top H]\,| \\
&= |\,2Tr[X^\top AH] + Tr[H^\top AH] - 2Tr[X^\top AH]\,| \\
&= |\,Tr[H^\top AH]\,| \\
&= |\langle H, AH\rangle| \\
&\leq ||H||_F||AH||_F \qquad \text{(por el lema 1)} \\
&\leq ||A||_F||H||_F^2, \qquad (6.8)
\end{aligned}
$$

así, dado $\epsilon > 0$ existe $\delta = \frac{\epsilon}{||A||_F} > 0$ tal que:

$$
\begin{aligned}
||H||_F < \delta \Rightarrow \frac{|F(X+H) - F(X) - \langle G_X, H\rangle|}{||H||_F} &\leq \frac{||A||_F ||H||_F^2}{||H||_F} \\
&= ||A||_F ||H||_F \\
&< \delta ||A||_F \\
&= \frac{\epsilon}{||A||_F} ||A||_F \\
&= \epsilon
\end{aligned}
\tag{6.9}
$$

Por lo tanto,

$$
\lim_{H \to \mathbf{0}} \frac{F(X+H) - F(X) - \langle G_X, H\rangle}{||H||} = 0.
$$

□

La siguiente proposición junto con su demostración se encuentra contenida en [28].

Proposición 13 *Sea $X \in \mathbb{R}^{n\times p}$ una matriz de rango p. Entonces $\pi(X)$ esta bien definida. Además, si $X = U\Sigma V^\top$ es la SVD de la matriz X, entonces $\pi(X) = UI_{n,p}V^\top$.*

Demostración.

De la definición 26 y de la proposición 2 se desprende inmediatamente que dado $X \in \mathbb{R}^{n\times p}$,

$$
\pi(X) = U\pi(\Sigma)V^\top \tag{6.10}
$$

donde $X = U\Sigma V^\top$ es la SVD de la matriz X. Así, basta demostrar que $Q_* = I_{n,p}$ es la solución de :

$$
\min_{Q \in \mathrm{St}(n,p)} ||\Sigma - Q||_F^2. \tag{6.11}
$$

En efecto, como

$$
\begin{aligned}
||\Sigma - Q||_F^2 &= Tr[(\Sigma - Q)^\top (\Sigma - Q)], \\
&= p + Tr[\Sigma^\top \Sigma] - 2Tr[\Sigma^\top Q],
\end{aligned}
\tag{6.12}
$$

entonces el problema 6.11 es equivalente al siguiente problema:

$$
\max_{Q \in \mathrm{St}(n,p)} Tr[\Sigma^\top Q]. \tag{6.13}
$$

Así que para probar que $Q_* = I_{n,p}$ es la solución del problema 6.11, es suficiente demostrar que $Tr[\Sigma^\top Q] \leq Tr[\Sigma^\top I_{n,p}],\ \forall Q \in \mathrm{St}(n,p)$ y que la igualdad se satisface únicamente cuando $Q = I_{n,p}$.

Si denotamos por σ_{ij} y q_{ij} denotan las entradas (i,j) de las matrices Σ y Q respectivamente, entonces:

$$
\begin{aligned}
Tr[\Sigma^\top Q] &= \sum_{i=1}^{p} \sigma_{ii} q_{ii} && (6.14)\\
&\leq \sum_{i=1}^{p} \sigma_{ii} |q_{ii}| && (6.15)\\
&\leq \sum_{i=1}^{p} \sigma_{ii} \quad (\text{ya que } Q \in \mathrm{St}(n,p)) && (6.16)\\
&= Tr[\Sigma^\top I_{n,p}]. && (6.17)
\end{aligned}
$$

Además, como $q_{ii} = 1$, para $i = 1, 2, \ldots, p$ si y solo si $Q = I_{n,p}$ entonces la igualdad se satisface en (6.17) solo cuando $Q = I_{n,p}$. Por lo tanto $Q_* = I_{n,p}$ es el óptimo global del problema (6.11), lo cual prueba la proposición. □

El próximo lema y su demostración es proveída por Wen-Yin en [21].

Lema 9 *Supongamos que X es un minimizador local del problema* (1.1)*. Entonces X satisface las condiciones de optimalidad de 1er orden: $\mathcal{D}_X\mathcal{L}(X,\Lambda) = G - XG^\top X = \boldsymbol{0}$ y $X^\top X = I$ con la asociada matriz de multiplicadores de Lagrange $\Lambda = G^\top X$. Además, si definimos por:*

$$\nabla\mathcal{F}(X) := G - XG^\top X \ \text{ y } A := GX^\top - XG^\top$$

entonces $\nabla\mathcal{F}(X) = AX$ y se satisface que: $\nabla\mathcal{F} = \boldsymbol{0}$ si y solo si $A = \boldsymbol{0}$.

Demostración.
Derivando la función Lagrangiana (4.1) respecto a X tenemos que:

$$\mathcal{D}_X\mathcal{L}(X,\Lambda) = G - \frac{1}{2}(2X)\Lambda = G - X\Lambda$$

y derivando la función Lagrangiana respecto a Λ tenemos que:

$$\mathcal{D}_\Lambda\mathcal{L}(X,\Lambda) = X^\top X - I$$

luego las condiciones de optimalidad de 1er orden son:

$$
\begin{aligned}
G - X\Lambda &= \mathbf{0} && (6.18)\\
X^\top X &= I && (6.19)
\end{aligned}
$$

multiplicando la ecuación (6.18) por la izquierda por la matriz $X^\top$ se tiene,

$$X^\top G - X^\top X\Lambda = \mathbf{0}$$

de la ecuación de arriba y la ecuación (6.19) se obtiene,

$$\Lambda = X^\top G \tag{6.20}$$

y por lo tanto, como la matriz de multiplicadores de Lagrange es simétrica (debido a la forma particular del conjunto de restricciones de problema (1.1)), las condiciones de optimalidad de 1er orden se pueden escribir de la siguiente forma:

$$\mathcal{D}_X\mathcal{L}(X,\Lambda) = G - XG^\top X = \mathbf{0} \text{ y } X^\top X = I$$

con la asociada matriz de multiplicadores de Lagrange $\Lambda = G^\top X$.

Por otro lado, de la notación del enunciado del lema tenemos que $\nabla\mathcal{F}(X) = G - XG^\top X$ así, $\nabla\mathcal{F}(X) = GX^\top X - XG^\top X$ ya que $X^\top X = I$, luego $\nabla\mathcal{F}(X) = (GX^\top - XG^\top)X$ y por lo tanto se obtiene que:

$$\nabla\mathcal{F}(X) = AX.$$

de este último resultado se desprende claramente que $\nabla\mathcal{F} = \mathbf{0}$ si y solo si $A = \mathbf{0}$.□

La siguiente proposición junto con su demostración es presentada en [51].

Proposición 14 *Sea $X \in \mathbb{R}^{n\times k}$ con $rank(X) = n$ y $B \in \mathbb{R}^{m\times k}$. Entonces la solución del Orthogonal Procrustes Problem:*

$$\min_{Q\in \mathtt{St}(m,n)} ||QX - B||_F^2, \tag{6.21}$$

viene dada por $Q_* = VI_{m,n}U^\top$, donde $U\Sigma V^\top = XB^\top$ es la SVD de la matriz $XB^\top$.

Demostración.
De la definición de la norma Frobenius tenemos que:

$$\begin{aligned} ||QX - B||_F^2 &= Tr[(QX - B)^\top (QX - B)], && (6.22)\\ &= Tr[X^\top Q^\top QX] - 2Tr[QXB^\top] + Tr[B^\top B], && (6.23)\\ &= ||X||_F^2 - 2Tr[QXB^\top] + ||B||_F^2 \quad (\text{ya que } Q^\top Q = I_n), && (6.24)\end{aligned}$$

así, el problema de optimización (6.21) es equivalente a:

$$\max_{Q\in \mathtt{St}(m,n)} Tr[QXB^\top], \tag{6.25}$$

para resolver (6.25), consideremos la SVD de la matriz $XB^\top$, dada por $XB^\top = U\Sigma V^\top$, entonces

$$Tr[QXB^\top] = Tr[QU\Sigma V^\top] = Tr[V^\top QU\Sigma] = Tr[\Sigma Z] = \sum_{i=1}^{n} \sigma_{ii} z_{ii}, \tag{6.26}$$

con $Z = V^\top QU$, y donde σ_{ij} y z_{ij} denotan las entradas (i, j) de las matrices Σ y Z respectivamente.

Como $Z \in \mathbb{R}^{m\times n}$ es ortonormal, entonces se satisface que $z_{ij} \leq |z_{ij}| \leq 1$ para todo $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$. Luego, la sumatoria en (6.26) es maximizada cuando $Z = I_{m,n}$, y por lo tanto, el valor optimal del problema 6.21 se alcanza cuando $Q_* = VI_{m,n}U^\top$.□

Bibliografía

[1] Grubišić, I.,Pietersz, R., *Efficient rank reduction of correlation matrices.* Linear Algebra and its Applications 422(2), 629-653 (2007).

[2] Pietersz, R. and Groenen, P. J., *Rank reduction of correlation matrices by majorization.* Quantitative Finance 4(6), 649-662 (2004).

[3] Rebonato, R and Jäckel, P., *The most general methodology to creating a valid correlation matrix for risk manage-ment and option pricing purposes.* Journal of Risk 2, 17-27 (1999).

[4] Golub, G. H. and Van Loan, C. F., (3 edn) *Matrix computations.* Johns Hopkins Univerisity Press (1996).

[5] Saad, Y., *Numerical methods for large eigenvalue problems.* Manchester University Press (1992).

[6] Yang, C.,Meza, J. C.,Lee, B. and Wang, L.W. *KSSOLV - a Matlab toolbox for solving the Kohn-Sham equations.* ACM Transactions on Mathematical Software 36, 1-35 (2009).

[7] Eldén, L. and Park, H., *A Procrustes problem on the Stiefel manifold.* Numerische Mathematik 82(4), 599-619 (1999).

[8] Schönemann, P.H., *A generalized solution of the orthogonal Procrustes problem.* Psychometrika 31(1), 1-10 (1966).

[9] Francisco, J. B. and Martini, T., *Spectral Projected Gradient Method for the Procrustes Problem.* TEMA (São Carlos) vol.15 no.1 São Carlos (2014).

[10] d'Aspremont, A., Ei Ghaoui, L., Jordan, M.I. and Lanckriet, G.R., *A direct formulation for sparse PCA using semidefinite programming.* SIAM Review 49(3), 434-448 (2007).

[11] Journée, M., Nesterov, Y., Richtárik, P. and Sepulchre, R., *Generalized power method for sparse principal component analysis.* The Journal of Machine Learning Research 11, 517-553 (2010).

[12] Zou, H., Hastie, T., and Tibshirani, R., *Sparse principal component analysis.* Journal of Computational and Graphical Statistics 15(2), 265-286 (2006).

[13] Edelman, A., Arias, T.A. y Smith, S.T., *The geometry of algorithms with orthogonality constraints.* SIAM Journal on Matrix Analysis and Applications 20(2), 303-353 (1998).

[14] Joho, M. and Mathis, H., *Joint diagonalization of correlation matrices by using gradient methods with application to blind signal separation.* In: Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002, pp. 273-277. IEEE (2002).

[15] Theis, F., Cason, T. and Absil, P.A., *Soft dimension reduction for ICA by joint diagonalization on the stiefel manifold.* In: T. Adali, C. Jutten, J.M.T. Romano, A. Barros (eds.) Independent Component Analysis and Signal Separation, Lecture Notes in Computer Science, vol. 5441, pp. 354-361. Springer Berlin Heidelberg (2009).

[16] Kokiopoulou, E., Chen, J. and Saad, Y., *Trace optimization and eigenproblems in dimension reduction methods.* Numerical Linear Algebra with Applications, Volume 18, Issue 3, pages 565-602 (2011).

[17] Mahony, R. E., *Optimization algorithms on homogeneous spaces: with applications in linear systems theory.* Ph.D. dissertation, Australian Nat. Univ., Canberra, (1994).

[18] Liu, Y. F., Dai, Y. H. and Luo, Z.Q., *On the complexity of leakage interference minimization for interference alignment.* In: 2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications, pp. 471-475 (2011)

[19] P.-A. Absil, R. Mahony and R. Sepulchre, *Optimization algorithms on matrix manifolds.* Princeton University Press, Princeton, NJ, 2008.

[20] Absil, P. A., Malick, J. C., *Projection-like retractions on matrix manifolds.* SIAM Journal on Optimization 22(1), 135-158 (2012).

[21] Wen, Z. W. and Yin, W. T., *A feasible method for optimization with orthogonality constraints.* Mathematical Programming pp. 1-38 (2012). DOI 10.1007/s10107-012-0584-1.

[22] Wen, Z., Yang, C., Liu, X. y Zhang, Y., *Trace-penalty minimization for large-scale eigenspace computation.* Journal of Scientific Computing, 1-29, (2015).

[23] Abrudan, T., Eriksson, J. y Koivunen, V., *Steepest descent algorithms for optimization under unitary matrix constraint.* IEEE Transactions on Signal Processing 56(3), 1134-1147 (2008).

[24] Abrudan, T., Eriksson, J. y Koivunen, V., *Conjugate gradient algorithm for optimization under unitary matrix constraint.* Signal Processing 89(9), 1704-1714 (2009).

[25] Koonin, S. E. y Meredith, D. C., *Computational Physics.* Westview Press, a member of the Perseus Books Group, Fortran version, 1990.

[26] Nishimori, Y. y Akaho, S., *Learning algorithms utilizing quasi-geodesic ows on the Stiefel manifold.* Neurocomputing 67, 106-135 (2005).

[27] Hiroyuki S., *Riemannian Newton's method for joint diagonalization on the Stiefel manifold with application to ICA.* ArXiv:1403.8064v2 [math.OC] 27 May 2014.

[28] Manton, J. H., *Optimization algorithms exploiting unitary constraints.* Signal Processing, IEEE Transactions on, 50(3), 635-650, (2002).

[29] Lai, R. y Osher, S., *A splitting method for orthogonality constrained problems.* Journal of Scientific Computing, 58(2), 431-449, (2014).

[30] Nocedal, J. y Wright, S. J., *Numerical Optimization.* Springer Series in Operations Research and Financial Engineering, Springer, New York, second ed., 2006.

[31] Ross, K. A., *Elementary Analysis* (pp. 129-130). New York: Springer-Verlag., (1980).

[32] Perko, L., *Differential Equations and Dynamical Systems.* Springer Science & Business Media, (Vol. 7), 2013.

[33] James R. Munkres, *Analysis on Manifolds.* Addison-Wesley, Redwood City, CA, first ed., 1990.

[34] Corach, G. y Andruchov, E., *Notas de Análisis Funcional.* Universidad de Buenos Aires, Departamento de Matemáticas, Argentina, (1997). http://glaroton.ungs.edu.ar/bsv.pdf.

[35] Luenberger, D. G., *The Gradient Projection Method Along Geodesics.* Management Science, Vol.18, 620-631, (1972).

[36] Gabay, D., *Minimizing a differentiable function over a differential manifold.* Journal of Optimization Theory and Applications 37(2), 177-219, (1982).

[37] Smith, S. T., *Geometric optimization methods for adaptive filtering.* Ph.D. dissertation, Harvard Univ., Cambridge, MA, (1993).

[38] Birgin, E. G., MartÃnez, J. M. y Raydan, M., *Monmonotone spectral projected gradient methods on convex sets.* SIAM Journal on Optimization, 10(2000), 1196-1211.

[39] Birgin, E. G., Martinez, J. M. y Raydan, M. *Spectral projected gradient methods: review and perspectives.* J. Stat. Softw, 60(3), (2014).

[40] Raydan M., *The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem.* SIAM Journal on Optimization, 7(1), 26-33, (1997).

[41] Barzilai, J. y Borwein, J. M., *Two-point step size gradient methods.* IMA J. Numer. Anal., 8(1988), pp. 141-148.

[42] Francisco, J. B. y Bazán, F. S. V., *Nonmonotone algorithm for minimization on closed sets with application to minimization on Stiefel manifolds.* Journal of Computational and Applied Mathematics, 236(10), 2717-2727, (2012).

[43] Grippo, L., Lampariello, F. y S. Lucidi, *A nonmonotone line search technique for Newton's method.* SIAM Journal on Numerical Analysis, 23(4), 707-716, (1986).

[44] Dai, Y.H. y Fletcher R., *Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming.* Numerische Mathematik, 100(1), 21-47, (2005).

[45] Zhang, H. y Hager, W. W., *A nonmonotone line search technique and its application to unconstrained optimization. SIAM Journal on Optimization.* 14(4), 1043-1056, (2004).

[46] Osher, S., Burger, M., Goldfarb, D., Xu, J. y Yin, W., *An iterative regularization method for total variation-based image restoration.* Multiscale Modeling & Simulation, 4(2), 460-489, (2005).

[47] Bregman, L., *The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex optimization.* USSR Computational Mathematics and Mathematical Physics, 7:200-217, (1967).

[48] Esser, E., *Applications of Lagrangian-based alternating direction methods and connections to split Bregman.* CAM report, 9,31, (2009).

[49] Afsari, B. y Krishnaprasad, P. S., *Some gradient based joint diagonalization methods for ICA, in Independent Component Analysis and Blind Signal Separation.* Springer, pp. 437-444, (2004).

[50] Tai, X. C. y Wu, C., *Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model.* In Scale space and variational methods in computer vision (pp. 502-513). Springer Berlin Heidelberg, (2009).

[51] Viklands, T., *Algorithms for the weighted orthogonal Procrustes problem and other least squares problems.* Ph.D. dissertation, Ume University, Sweden, (2006).

[52] Gajardo, A. P., *Diferenciabilidad en espacios vectoriales normados.* Universidad Técnica Federico Santa María, Valparaíso, Chile, http://pgajardo.mat.utfsm.cl/optimizacion/Diferenciabilidad.pdf.

[53] Wax, M. y Sheinvald, J., *A least-squares approach to joint diagonalization.* Signal Processing Letters, IEEE, 4(1997) pp. 52-53.