

UNIVERSIDAD CENTROCCIDENTAL
LISANDRO ALVARADO
Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas



COMPARACIÓN ENTRE GAUSS-SEIDEL Y SOBRE-RELAJACIÓN SUSCESIVA
(SOR) EN LA RESOLUCIÓN DEL PROBLEMA DE COMPLETACIÓN DE
MATRICES

TRABAJO ESPECIAL DE GRADO PRESENTADO POR
Br: Harry Oviedo

COMO REQUISITO FINAL
PARA OBTENER EL TITULO DE LICENCIADO
EN CIENCIAS MATEMÁTICAS

ÁREA DE CONOCIMIENTO: OPTIMIZACIÓN

TUTOR: Hugo Lara

Barquisimeto - Venezuela
Octubre 2013



Universidad Centrooccidental Lisandro
Alvarado Decanato de Ciencias y
Tecnología Licenciatura en Ciencias
Matemáticas



ACTA
TRABAJO ESPECIAL DE GRADO

Los suscritos miembros del Jurado designado por el Jefe del Departamento De Matemática del Decanato De Ciencias y Tecnología de la Universidad Central Lisandro Alvarado para examinar y dictar el veredicto sobre el Trabajo especial de grado titulado :

“Comparación entre Gauss-Seidel y sobre-relajación sucesiva (SOR) en la resolución del problema de completación de matrices”

presentado por el ciudadano Br. (HARRY OVIEDO) titular de la Cédula de Identidad No. (19.264.453), con el propósito de cumplir con el requisito académico final para el otorgamiento del título de Licenciado en Ciencias Matemáticas.

Luego de realizada la defensa y en los términos que impone los lineamientos para el Trabajo Especial de Grado de la Licenciatura en Ciencias Matemáticas, se procedió a discutirlo con el interesado habiéndose emitido el veredicto que a continuación se expresa:

con una calificación de _____ puntos.

En fe de lo expuesto firmamos la presente Acta en la Ciudad de Barquisimeto a los _____ días del mes de _____ de _____

TUTOR

FIRMA

PRINCIPAL

FIRMA

PRINCIPAL

FIRMA

Observaciones:

Aprobado ó Reprobado

AGRADECIMIENTOS

En el presente trabajo de tesis primeramente me gustaría agradecer a ti Dios por bendecirme para llegar hasta donde he llegado.

Le doy gracias a mis padres Dina León y Omar Oviedo por apoyarme en todo momento, por los valores que me han inculcado, y por haberme dado la oportunidad de tener una excelente educación a lo largo de mi vida.

A mi hermano Henry Oviedo por ser parte importante de mi vida y por su apoyo a lo largo de esta carrera.

Debo agradecer de manera especial y sincera al profesor Hugo Lara por aceptarme para realizar esta tesis de pregrado bajo su dirección. También por sus conocimientos, su experiencia, su motivación, su apoyo, su confianza en mi trabajo y su capacidad para guiar mis ideas, ha sido un aporte invaluable.

A mis abuelas Dina Ortiz y Carolina Oropeza por su amor incondicional hacia mi como hacia todos sus hijos y nietos.

A todos mis tíos y tías, por tener un amor como de padre o madre hacia mi persona, siempre se que cuento con el apoyo de todos ustedes.

A todos mis primos, por todos aquellos buenos momentos que hemos vivido a lo largo de nuestras vidas, por siempre estar atentos para saber como me iba es mis estudios, a todos los considero como hermanos.

A mis panas y compañeros de clase Eduardo Fuentes, Oscar Sapienza, Jaicer Lopez, Shaday Guerrero, Rafael Chavez, Diana Piñango, Lilibeth Sulvaran, por todos los bueno momentos que pasamos a lo largo de la carrera. En especial a Shaday y Rafael por haberme permitido formar parte de su grupo de estudio, donde resolvimos gran cantidad de problemas, y logramos aprobar las materias más difíciles.

Al profesor Javier Hernandez por su apoyo prestado, para lograr presentar este trabajo de grado y así poder graduarme el presente año.

A todos ellos, muchas gracias.

RESUMEN

En el presente trabajo consideramos el problema de completación de matrices (MC) usando minimización del rango. Como lo hace Wen y colaboradores [1], estudiamos métodos alternantes que buscan la completación a través de una factorización, con el objeto de evitar la descomposición de valor singular. Esta última operación es característica de los métodos de minimización de la norma nuclear, pero es computacionalmente cara. El problema de minimizar la norma nuclear es una relajación convexa eficiente del problema de minimización del rango.

Proponemos un algoritmo alternante híbrido SOR-Lagrangeanos aumentados que utiliza información de dualidad, en contraste al SOR de Wen que genera una sucesión primal pura. Mostramos comparaciones numéricas que muestran las bondades de nuestra propuesta.

Índice general

AGRADECIMIENTOS	v
RESUMEN	vii
Introducción	1
1. Preliminares	3
1.1. Álgebra matricial	3
1.1.1. Proyección ortogonal y la Pseudoinversa	3
1.1.2. Algunas factorizaciones de matrices	4
1.1.3. Normas matriciales	6
1.2. Métodos iterativos para la resolución de sistemas de ecuaciones lineales	8
1.2.1. Método de Gauss-Seidel	9
1.2.2. Método de Sobrerrelajación sucesiva(SOR)	9
1.3. Método del Lagrangeano Aumentado	10
2. Un modelo de factorización de bajo rango para el problema MC	12
2.1. Esquema de minimización alternante	12
2.1.1. El método de Gauss-Seidel no lineal	13
2.1.2. Esquemas Alternativos	14
2.1.3. Un esquema SOR no lineal	16
2.1.4. Extensión para problemas con restricciones lineales generales	19
2.2. Análisis de Convergencia	19
3. Métodos de Lagrangeano aumentado en la factorización de bajo rango	29
3.0.1. El método de Gauss-Seidel no lineal	29
3.0.2. Esquemas Alternativos	30
3.1. Un algoritmo híbrido ALM-SOR para el problema MC	32
4. Experimentación Numérica	35
4.1. Experimentos sobre matrices generadas aleatoriamente	35
4.1.1. Experimentos con datos simulados	36
4.2. Experimentos de data real	37
4.2.1. Jester jokes and Movie Lens.	37
4.2.2. Completación de Matrices aplicado a imágenes	38
Anexos	43

Conclusiones

49

BIBLIOGRAFIA

49

Introducción

El problema de minimizar el rango de una matriz se presenta en muchas aplicaciones, por ejemplo, en teoría de control y sistemas, reducción de dimensionalidad de modelos [10], recuperación de formas y movimientos en imágenes [11], [12], minería de datos y reconocimiento de patrones [13], predicción en colaboración y construcciones de baja dimensionalidad, entre otros. En este trabajo consideramos el problema de completación de matrices (MC) para recuperar una matriz de rango bajo desde un subconjunto de sus componentes. Este problema es NP-duro en virtud de la naturaleza combinatoria de la función rango, pero ha sido demostrado en [9], [23], que bajo condiciones razonables una solución del problema de rango mínimo puede ser encontrada a través de una relajación convexa que involucra la minimización de la norma nuclear.

Se han propuesto varios tipos de algoritmos para recuperar una solución de rango mínimo. Un método es el algoritmo del umbral de valores singulares [8]. Otro enfoque es el algoritmo de contracción de punto fijo [14]. Un algoritmo de gradiente proximal acelerado se desarrolla en [15]. Se ha aplicado también el algoritmo clásico de Lagrangeanos aumentados (AL) en problemas de descomposición de matrices esparsas de rango bajo. Existen otros enfoques que pueden verse en [16, 17, 18, 19, 20, 21]. Todos estos algoritmos encaran el costo computacional del uso de la descomposición de valor singular (SVD). Esta operación se hace mas costosa con el aumento de tamaño y rango de las matrices subyacentes. Es por tanto deseable explorar enfoques alternativos que se adapten mejor a problemas de gran escala.

En [1], Wen, Yin y Zang proponen un modelo alternativo que evita el uso de la SVD. Su enfoque fué el de encontrar la matriz de rango bajo, de forma descompuesta como producto de dos matrices. Esta descomposición se ejecua de forma alternante, usando un problema de optimización fácil de ejecutar en cada etapa. El enfoque es algunas veces llamado Gauss Seidel por bloques no lineal. Para obtener mejores resultados numéricos, el algoritmo fué acelerado usando la llamada Sobrerrelajación Sucesiva (SOR). En el proceso se utilizaron herramientas de álgebra lineal computacional, como la descomposición QR rápida, que permitieron un desempeño práctico competitivo.

En nuestro trabajo modificamos el algoritmo SOR mencionado antes, agregando información de dualidad en cada iteración a través del uso de Lagrangeanos Aumentados. El resultado es un híbrido SOR-AL que se muestra competitivo en las implementaciones.

El resto del trabajo se describe a continuación: En el próximo capítulo mostramos contenidos preliminares relacionados con elementos de álgebra lineal matricial, y métodos iterativos para resolución de sistemas lineales. También estudiamos el método de Lagrangeanos Aumentados

clásico. En el capítulo 2 describimos los resultados de Wen y Colaboradores [1], presentando el algoritmo y describiendo algunas de sus propiedades de convergencia. El capítulo 3 se centra en nuestra propuesta de algoritmo híbrido SOR-AL. Describimos el algoritmo y mostramos propiedades. El capítulo 4 presenta resultados numéricos que comparan nuestro algoritmo con el SOR de Wen. También presentamos análisis de los resultados de la comparación. Culminamos con las conclusiones generales del trabajo.

Capítulo 1

Preliminares

Este capítulo se inicia con un repaso de algunos conceptos de algebra matricial que serán muy importantes para el desarrollo del presente trabajo, entre algunos de ellos se encuentran, la definición de norma matricial, la pseudoinversa de una matriz, proyecciones ortogonales y algunas factorizaciones de matrices tales como la descomposición en valores singulares(SVD) de una matriz y la factorización QR; estos tópicos fueron extraídos del libro *Matrix Computations* cuyos autores son G. H. GOLUB y C. F. VAN LOAN [2]. En la sección 1.2 recordaremos la esencia de los métodos iterativos para la resolución de sistemas de ecuaciones lineales como lo son el método de Gauss-Seidel(GS) y el método de sobrerelajación sucesiva(SOR) este contenido fué extraído de Notas del ITESCAM [4] y también del libro *Numerical Mathematics and Computing* de los autores Ward Cheney, David Kincaid [3]. Por último, en la sección 1.3 estudiaremos el método del Lagrangeano Aumentado para la resolución de problemas de minimización con restricciones de igualdad este contenido fué extraído del libro *Constrained Optimization and Lagrange Multiplier Methods* del autor Dimitri P. Bertsekas [5].

1.1. Algebra matricial

1.1.1. Proyección ortogonal y la Pseudoinversa

Definición 1 Sea $S \subset \mathbb{R}^n$ un subespacio. Se define la proyección ortogonal $P \in \mathbb{R}^{n \times n}$ sobre el subespacio S , a la matriz que satisface:

- $\text{ran}(P) = S$
- $P^2 = P$
- $P = P^\top$

Nota 1 En la definición 1, $\text{ran}(P)$ denota el subespacio imagen de la matriz P , es decir, $\text{ran}(P) = \text{span}\{p_1, p_2, \dots, p_n\}$ donde p_i denota a la columna i -ésima de la matriz P , en ocasiones también denotaremos a este espacio por $\mathcal{R}(P)$.

Observación 1 De la definición es fácil demostrar que si $x \in \mathbb{R}^n$ entonces $Px \in S$ y $(I - P)x \in S^\perp$.

Observación 2 Si las columnas de $V = [v_1 v_2 \dots v_k]$ forman una base ortonormal para el subespacio S , entonces es fácil probar que la única proyección ortogonal sobre S es $P = VV^\top$.

Nota 2 Dada una matriz $A \in \mathbb{R}^{m \times n}$, denotaremos por \mathcal{P}_A a la proyección ortogonal sobre el subespacio imagen $\mathcal{R}(A)$ de A .

Nota 3 Dados $m, n \in \mathbb{N}$, y dado un subconjunto de índices $\Omega \subset \{(i, j) : 1 \leq i \leq m, 1 \leq j \leq n\}$, denotaremos por \mathcal{P}_Ω a la proyección sobre el subespacio de las matrices dispersas con entradas distintas de cero en las posiciones indicadas por Ω . Nótese que de lo mencionado anteriormente, para cualquier matriz $A \in \mathbb{R}^{m \times n}$ se tiene que:

$$\mathcal{P}_\Omega(A) = \begin{cases} a_{ij} & \text{si } (i, j) \in \Omega \\ 0 & \text{en otro caso.} \end{cases}$$

Definición 2 Sea $X \in \mathbb{R}^{n \times m}$, se define la Pseudo-Inversa de X denotada por X^\dagger , como la única matriz que satisface las cuatro condiciones de Moore-Penrose:

- $X^\dagger X X^\dagger = X^\dagger$
- $X X^\dagger X = X$
- $(X^\dagger X)^\top = X^\dagger X$
- $(X X^\dagger)^\top = X X^\dagger$

Observación 3 Se puede probar que la Pseudo-inversa de una matriz $A \in \mathbb{R}^{m \times n}$ es la única solución del problema:

$$\min_{X \in \mathbb{R}^{n \times m}} \|AX - I_m\|_F$$

Además, si $\text{rank}(A) = n$, entonces $A^\dagger = (A^\top A)^{-1} A^\top$, mientras que si $m = n = \text{rank}(A)$, entonces $A^\dagger = A^{-1}$.

1.1.2. Algunas factorizaciones de matrices

Teorema 1 (Descomposición de valores singulares (SVD)) Si $A \in \mathbb{R}^{m \times n}$, entonces existen matrices ortogonales

$$U = [u_1, u_2, \dots, u_m] \in \mathbb{R}^{m \times m} \text{ y } V = [v_1, v_2, \dots, v_n] \in \mathbb{R}^{n \times n}$$

tales que,

$$U^\top A V = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p) \in \mathbb{R}^{m \times n} \text{ con } p = \min(m, n)$$

y $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$.

A los σ_i se les llama valores singulares de A , y los vectores u_i y v_i se les conoce como el i -ésimo vector singular por la izquierda y el i -ésimo vector singular por la derecha respectivamente.

La SVD revela una buena idea acerca de la estructura de una matriz. En el teorema anterior, si se define a r , como el entero positivo tal que:

$$\sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_p = 0$$

entonces:

$$\begin{aligned} \text{rank}(A) &= r \\ \text{null}(A) &= \text{span}\{v_{r+1}, \dots, v_n\} \\ \text{ran}(A) &= \text{span}\{u_1, \dots, u_r\} \end{aligned}$$

y además podemos escribir a la matriz A como:

$$\sum_{i=1}^r \sigma_i u_i v_i^\top.$$

Teorema 2 (Factorización QR) Si $A \in \mathbb{R}^{m \times n}$, entonces existen matrices $Q \in \mathbb{R}^{m \times m}$ y $R \in \mathbb{R}^{m \times n}$ tales que:

$$A = QR$$

donde Q es una matriz ortogonal y R es una matriz triangular superior.

Teorema 3 Si $A = QR$ es la factorización QR de una matriz de rango columna completo $A \in \mathbb{R}^{m \times n}$, donde $m \geq n$ y $A = [a_1, \dots, a_n]$ y $Q = [q_1, \dots, q_k]$ son particiones de columnas, entonces

$$\text{span}\{a_1, \dots, a_k\} = \text{span}\{q_1, \dots, q_k\} \quad k = 1, 2, \dots, n.$$

En particular, si $Q_1 = Q(1 : m, 1 : n)$ y $Q_2 = Q(1 : m, n+1 : m)$ entonces

$$\text{ran}(A) = \text{ran}(Q_1)$$

$$\text{ran}(A)^\perp = \text{ran}(Q_2)$$

y $A = Q_1 R_1$ con $R_1 = R(1 : n, 1 : n)$.

Observación 4 Nótese que el teorema anterior nos garantiza que si A es de rango columna completo entonces las primeras n columnas de la matriz Q forman una base ortonormal para el espacio $\text{ran}(A)$. En efecto, el teorema concluye que:

$$\begin{aligned} \text{ran}(A) &= \text{ran}(Q_1) \\ &= \text{span}\{q_1, \dots, q_n\} \end{aligned}$$

así, se tiene que el conjunto de vectores $\{q_1, \dots, q_n\}$ genera al espacio $\text{ran}(A)$ y como Q es una matriz ortogonal, se cumple que los vectores $\{q_1, \dots, q_n, \dots, q_m\}$ son ortonormales, en particular se tiene que los vectores $\{q_1, \dots, q_n\}$ son ortonormales y por lo tanto linealmente independientes, así obtenemos que el conjunto de vectores $\{q_1, \dots, q_n\}$ forma una base ortonormal para el subespacio $\text{ran}(A)$. De esta manera, se puede observar que el cálculo de la factorización QR es un camino para calcular bases ortonormales para un conjunto de vectores.

Observación 5 Como notamos en la observación anterior, sabemos que se puede calcular una base ortonormal Q_1 para el espacio imagen de una matriz A a través de su factorización QR , y gracias a esto, también se puede calcular la proyección ortogonal sobre el subespacio imagen de la matriz A mediante el producto $Q_1 Q_1^\top$. Este resultado será de gran ayuda para el presente trabajo.

1.1.3. Normas matriciales

Antes de definir lo que es la norma de una matriz, recordaremos algunas normas vectoriales e introduciremos algunas definiciones que serán de mucha utilidad para el desarrollo del presente trabajo.

Definición 3 (Pre-normas) Una función $f: \mathbb{R}^n \rightarrow \mathbb{R}$ que satisface las siguientes propiedades:

- $f(x) \geq 0$, $\forall x \in \mathbb{R}^n$ y $f(x) = 0$ si y solo si $x = 0$.
- $f(\alpha x) = |\alpha|f(x)$, $\forall x \in \mathbb{R}^n$ y $\alpha \in \mathbb{R}$.
- $f(x(z))$ es continua sobre \mathbb{R}^n , donde $z = (z_1, z_2, \dots, z_n) \in \mathbb{R}^n$ y $z(x) = z_1x_1 + z_2x_2 + \dots + z_nx_n$, y donde el conjunto $\{x_1, x_2, \dots, x_n\}$ forma una base de \mathbb{R}^n ;

es llamada una pre-norma.

Nota 4 Nótese que toda norma es una pre-norma, pues por la definición de norma se obtienen las primeras dos propiedades de una pre-norma, y la tercera propiedad se desprende del hecho de que toda norma es una función continua sobre \mathbb{R}^n .

Definición 4 (Norma dual) Sea $f(\cdot)$ una pre-norma. La función $f^D: \mathbb{R}^n \rightarrow \mathbb{R}$ dada por:

$$f^D(u) = \max_{f(x) \leq 1} \langle u, x \rangle.$$

es llamada la norma dual de f .

Ahora, recordemos algunas normas en \mathbb{R}^n , solo recordaremos las tres normas mas importantes, las cuales son: la norma euclideana, definida por:

$$\|x\| := \|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2} \text{ para cada } x \in \mathbb{R}^n.$$

La norma l_1 , definida por:

$$\|x\|_1 = \sum_{i=1}^n |x_i| \text{ para cada } x \in \mathbb{R}^n.$$

La norma del máximo, definida por:

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i| \text{ para cada } x \in \mathbb{R}^n.$$

Nota 5 Nótese que la norma l_1 es la norma dual de la norma del máximo. Esto es fácil de verificar, fíjese que el valor óptimo del siguiente problema de programación lineal:

$$\max_{\|u\|_\infty \leq 1} \langle u, x \rangle \tag{1.1}$$

ocurre cuando: $u^* = (\text{sgn}(x_1), \text{sgn}(x_2), \dots, \text{sgn}(x_n))^\top$, donde $\text{sgn}(x)$ es igual a 1 si $x \geq 0$ y -1 si $x < 0$. Así, como cualquier escalar se puede escribir de la forma $x = \text{sgn}(x)|x|$ y si denotamos por u_i^* a la componente i -ésima del argumento que maximiza a (1.1) obtenemos las siguientes igualdades:

$$\max_{\|u\|_\infty \leq 1} \langle u, x \rangle = \max_{\|u\|_\infty \leq 1} \sum_{i=1}^n u_i x_i = \sum_{i=1}^n u_i^* x_i = \sum_{i=1}^n \text{sgn}(x_i) x_i = \sum_{i=1}^n \text{sgn}(x_i) \text{sgn}(x_i) |x_i| = \sum_{i=1}^n |x_i| = \|x\|_1$$

por lo tanto, se tiene que la norma l_1 es la norma dual de la norma del máximo. Se desprende del teorema de dualidad de programación lineal que la norma del máximo es la norma dual de la norma l_1 .

También es fácil probar que la norma euclídeana es dual de si misma. En efecto, consideremos el problema de programación lineal:

$$\max_{\|u\| \leq 1} \langle u, x \rangle \quad (1.2)$$

Por teorema, sabemos que $\langle u, x \rangle = \cos(\theta) \|u\| \|x\|$, donde θ es el ángulo entre los vectores u y x . Así, el problema (1.2) es equivalente a resolver el problema:

$$\max_{\|u\| \leq 1} \cos(\theta) \|u\| \|x\|.$$

Nótese que el valor óptimo del problema anterior ocurre cuando $\|u^*\| = 1$ y $\theta^* = 0$, esto implica que:

$$\max_{\|u\| \leq 1} \cos(\theta) \|u\| \|x\| = \cos(\theta^*) \|u^*\| \|x\| = \|x\|$$

por lo tanto, la norma euclídeana es dual de si misma.

Definición 5 (Norma Matricial) A toda función $f: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ que satisfaga las siguientes propiedades:

- $f(A) \geq 0$ $A \in \mathbb{R}^{m \times n}$, $(f(A) = 0 \text{ si } A = 0)$
- $f(A+B) \leq f(A) + f(B)$ $A, B \in \mathbb{R}^{m \times n}$
- $f(\alpha A) = |\alpha| f(A)$ $\alpha \in \mathbb{R}, A \in \mathbb{R}^{m \times n}$.

la llamaremos norma matricial.

Observación 6 Al igual que las normas vectoriales, se utilizará la notación de doble barra para denotar a las normas matriciales es decir, $\|A\| = f(A)$.

Entre las normas matriciales mayormente usadas en algebra lineal numérica encontramos: la norma Frobenius dada por,

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$$

y las p-normas dadas por,

$$\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}.$$

Nótese que las p-normas están definidas a partir de las p-normas vectoriales, y que la norma Frobenius es una generalización de la norma euclídiana para vectores, además existe una conexión entre la SVD y la norma Frobenius, dicha conexión es la siguiente: si $A \in \mathbb{R}^{m \times n}$ entonces,

$$\|A\|_F^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_p^2 \text{ donde } p = \min(m, n).$$

Además, de estas normas, existen muchas otras, entre las cuales destacan para la comprensión de este trabajo la norma nuclear que se define por:

$$\|A\|_* = \sum_{k=1}^p \sigma_k$$

y la norma espectral, definida por:

$$\|A\| := \|A\|_{l_\infty} = \max_{1 \leq k \leq p} \sigma_k$$

donde $A \in \mathbb{R}^{m \times n}$, $p = \min(m, n)$ y σ_k es el k -ésimo valor singular de A .

Nota 6 *Nótese que hay una cierta similitud entre las normas vectoriales (la norma del máximo y la norma l_1) y las normas matriciales (norma espectral y la norma nuclear) en sus definiciones. Lo cual hace tender a preguntarse si ¿será cierto que las normas espectral y la nuclear son duales entre sí?, la respuesta a esta pregunta es afirmativa, es decir que se puede probar que la norma nuclear de una matriz $X \in \mathbb{R}^{m \times n}$ es igual al valor óptimo del siguiente problema:*

$$\max_{\|W\| \leq 1} \langle M, W \rangle$$

donde la variable de decisión $W \in \mathbb{R}^{m \times n}$ y $\langle M, W \rangle = \sum_{i,j} m_{i,j} w_{i,j}$. También, nótese que hay una cierta similitud entre la norma Frobenius y la norma euclideana, esto nos indica que la norma Frobenius es dual de sí misma.

1.2. Métodos iterativos para la resolución de sistemas de ecuaciones lineales

Los métodos iterativos o también conocidos como indirectos, son procedimientos utilizados para la resolución de sistemas de ecuaciones lineales de la forma:

$$Ax = b \tag{1.3}$$

en ellos, se comienza con $x^0 = (x_1^0, x_2^0, \dots, x_n^0)$, una aproximación inicial de la solución. A partir de x^0 se construye una nueva aproximación de la solución, $x^1 = (x_1^1, x_2^1, \dots, x_n^1)$. A partir de x^1 se construye x^2 (aquí el superíndice indica la iteración). Así sucesivamente se construye una sucesión de vectores $\{x^k\}_{k=1}^\infty$, con el objetivo, no siempre garantizado, de que

$$\lim_{k \rightarrow \infty} x^k = x^*$$

Un algoritmo iterativo general para resolver (1.3) procede de la siguiente forma: selecciona una matriz no-singular Q y un vector arbitrario x^0 , y luego genera vectores x^1, x^2, \dots respectivamente a partir de la ecuación:

$$Qx^k = (Q - A)x^{k-1} + b \quad (k = 1, 2, \dots) \tag{1.4}$$

Nótese que si la sucesión $\{x^k\}_{k=1}^\infty$ converge a un vector x^* , entonces tomando límite cuando $k \rightarrow \infty$ en (1.4) tenemos que:

$$Qx^* = (Q - A)x^* + b$$

lo cual implica que $Ax^* = b$. Por lo tanto si la sucesión converge, su límite es una solución del sistema original (1.3).

1.2.1. Método de Gauss-Seidel

El método Gauss-Seidel es un método iterativo para la resolución de sistemas de ecuaciones lineales, en cada iteración del Gauss-Seidel hay n subiteraciones, donde n es el número de columnas de la matriz A . En la primera subiteración se modifica únicamente x_1 , las demás coordenadas del vector x_2, x_3, \dots, x_n no son modificadas. El cálculo de x_1 se hace de tal manera que se satisfaga la primera ecuación, es decir:

$$x_1^1 = \frac{b_1 - (a_{12}x_2^0 + a_{13}x_3^0 + \dots + a_{1n}x_n^0)}{a_{11}}$$

$$x_i^1 = x_i^0, \quad (i = 2, 3, \dots, n)$$

En la segunda subiteración se modifica únicamente x_2 , las demás coordenadas x_1, x_3, \dots, x_n no se modifican. El cálculo de x_2 se hace de tal manera que se satisfaga la segunda ecuación, es decir:

$$x_2^2 = \frac{b_2 - (a_{21}x_1^1 + a_{23}x_3^1 + \dots + a_{2n}x_n^1)}{a_{22}}$$

$$x_i^2 = x_i^1, \quad (i = 1, 3, \dots, n)$$

Así sucesivamente, en la n -ésima subiteración se modifica únicamente x_n , y las demás coordenadas x_1, x_2, \dots, x_{n-1} no se modifican. El cálculo de x_n se hace de tal forma que se cumpla la n -ésima ecuación, es decir:

$$x_n^n = \frac{b_n - (a_{n1}x_1^{n-1} + a_{n2}x_2^{n-1} + \dots + a_{(n-1)(n-1)}x_{n-1}^{n-1})}{a_{nn}}$$

$$x_i^n = x_i^{n-1}, \quad (i = 1, 3, \dots, n-1)$$

Una vez que se ha hecho una iteración completa (n subiteraciones), se utiliza el último vector x obtenido como una aproximación inicial y se vuelve a aplicar el procedimiento.

Teóricamente el método de Gauss-Seidel puede ser un proceso infinito. En la práctica el proceso se acaba cuando la distancia entre los vectores x^k y x^{k-1} sea suficientemente pequeña, esto quiere decir que el x actual es casi una solución.

1.2.2. Método de Sobrerrelajación sucesiva(SOR)

Comenzaremos proporcionando la esencia de los métodos de relajación, a través de la siguiente definición:

Definición 6 (Método de Relajación) Sea f una función a valores reales, definida y continua en una vecindad de algún número real ξ . Los métodos de relajación utilizan sucesiones $\{x_k\}$ de la forma,

$$x_{k+1} = x_k - \omega f(x_k), \quad k = 0, 1, 2, \dots,$$

donde $\omega \neq 0$ es un número real fijo, y x_0 es un valor dado al comienzo cercano a ξ . A ω se le llama factor de relajación.

Si la sucesión $\{x_k\}$ converge a ξ , entonces ξ es una solución de la ecuación $f(x) = 0$.

Nota 7 En la selección de ω con $0 < \omega < 1$, los métodos reciben el nombre de subrelajación, y cuando $1 < \omega$ se les llama sobrerrelajación.

Una aceleración del método de Gauss-Seidel se logra, mediante la introducción de un factor de relajación ω , el cual da origen al **método de sobrerrelajación sucesiva (SOR)** que en la k -ésima iteración estima el valor de $x_i^{(k)}$ mediante la siguiente fórmula:

$$x_i^{(k)} = \omega \left\{ \left[- \sum_{1 \leq j < i}^n (a_{ij}/a_{ii}) x_j^{(k)} - \sum_{j > i \geq 1}^n (a_{ij}/a_{ii}) x_j^{(k-1)} + (b_i/a_{ii}) \right] \right\} + (1 - \omega) x_i^{(k-1)}.$$

Nótese que si se escoge $\omega = 1$ el método SOR se reduce al método de Gauss-Seidel.

1.3. Método del Lagrangeano Aumentado

El método del Lagrangeano Aumentado forma parte de los llamados métodos de penalidad, la idea básica en estos métodos no es otra que eliminar algunas o todas las restricciones y adherirlas a la función objetivo en términos de una penalidad que prescribe un alto costo para los puntos infeasibles. Asociado con estos métodos está un parámetro c , el cual determina la severidad de la penalidad y como consecuencia la extensión para la cual el problema resultante sin restricciones se aproxima al problema original restringido.

Consideremos el problema:

$$\min_{x \in X, h(x)=0} f(x) \quad (1.5)$$

donde $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $h: \mathbb{R}^n \rightarrow \mathbb{R}^m$ son dos funciones y X es un subconjunto dado de \mathbb{R}^n . Para cualquier parámetro c , se define la función *Lagrangeana Aumentada* $L_c: \mathbb{R}^n \times \mathbb{R}^m$ por:

$$L_c(x, \lambda) = f(x) + \lambda' h(x) + \frac{1}{2} c \|h(x)\|^2 \quad (1.6)$$

Nos referiremos a c como el parámetro de penalidad y a λ como el vector de multiplicadores.

El método del Lagrangeano Aumentado, consiste en resolver una sucesión de problemas de la forma:

$$\min_{x \in X} L_c(x, \lambda_k)$$

donde $\{\lambda_k\}$ es una sucesión en \mathbb{R}^m y $\{c_k\}$ es una sucesión de parámetros de penalidad que satisface:

$$0 < c_k < c_{k+1} \quad \forall k, \quad c_k \rightarrow \infty.$$

En las implementaciones prácticas del método del *Lagrangeano Aumentado* se utiliza el siguiente algoritmo:

Método del Lagrangeano Aumentado ALM

ENTRADA: $\rho > 1$, $k = 1$

1. Mientras no converge, hacer
2. resolver $x_{k+1} := \operatorname{argmin}_x L_{c_k}(x, \lambda_k)$;
3. $\lambda_{k+1} := \lambda_k + c_k h(x_{k+1})$;

4. $c_{k+1} := \rho c_k; k := k + 1;$
5. fin (Mientras)

SALIDA: x_k, λ_k

Proposición 1 *Asumamos que $X = \mathbb{R}^n$ y que $f, h \in \mathbf{C}^1$. Además, supongamos que para cada $k = 0, 1, 2, \dots$, el vector x_k satisface*

$$|\nabla_x L_{c_k}(x_k, \lambda_k)| \leq \varepsilon_k,$$

donde $\{\lambda_k\}$ es acotada, $0 < c_k < c_{k+1}$ para todo k , $c_k \rightarrow \infty$, y $0 \leq \varepsilon_k$ para todo k , $\varepsilon_k \rightarrow 0$. También asumamos que existe una subsucesión $\{x_k\}_K$ que converge al vector x^ tal que $\nabla h(x^*)$ tiene rango m . Entonces para algún vector λ^* , se cumple que:*

$$\{\lambda_k + c_k h(x_k)\}_K \rightarrow \lambda^*, \quad \nabla f(x^*) + \nabla h(x^*) \lambda^* = 0 \quad \text{y} \quad h(x^*) = 0$$

Para la siguiente proposición, asumiremos que x^* es un mínimo local que satisface las siguientes condiciones de suficiencia de segundo orden:

El vector x^* es un mínimo local estricto y un punto regular de (1.3), y $f, h \in \mathbf{C}^2$ sobre alguna bola abierta centrada en x^* . Además, x^* junto con su vector de multiplicadores de Lagrange asociado λ^* satisfacen:

$$z^\top \nabla_{xx}^2 L_0(x^*, \lambda^*) z > 0,$$

para todo $z \neq 0$ con $\nabla h(x^*)^\top z = 0$.

Proposición 2 *Sea \bar{c} un escalar positivo tal que: $\nabla_{xx}^2 L_{\bar{c}}(x^*, \lambda^*) > 0$. Entonces existen escalares positivos ε , δ y M tales que:*

- *para todo $(\lambda, c) \in D \subset \mathbb{R}^{m+1}$ donde $D = \{(\lambda, c) / |\lambda - \lambda^*| < \delta c, \bar{c} \leq c\}$, el problema:*

$$\min_{x \in B(x^*, \varepsilon)} L_c(x, \lambda)$$

tiene una única solución denotada por $x(\lambda, c)$. La función $x(\cdot, \cdot)$ es continuamente diferenciable en el interior de D , y para todo $(\lambda, c) \in D$, tenemos que:

$$|x(\lambda, c) - x^*| \leq M |\lambda - \lambda^*| / c.$$

- *para todo $(\lambda, c) \in D$, se tiene que:*

$$|\tilde{\lambda}(\lambda, c) - \lambda^*| \leq M |\lambda - \lambda^*| / c,$$

donde

$$\tilde{\lambda}(\lambda, c) = \lambda + cH[x(\lambda, c)].$$

- *para todo $(\lambda, c) \in D$, la matriz $\nabla_{xx}^2 L_c[x(\lambda, c), \lambda]$ es definida positiva y la matriz $\nabla h[x(\lambda, c)]$ es de rango m .*

Capítulo 2

Un modelo de factorización de bajo rango para el problema MC

En este capítulo describiremos de manera detallada y precisa lo contenido en el artículo “*Solving a Low-Rank Factorization Model For Matrix Completion by a Nonlinear Successive Over-Relaxation Algorithm*” [1], de los autores Zaiwen Wen, Wotao Yin y Yin Zhang. Más explícitamente este artículo propone un modelo de factorización de bajo rango y construye un algoritmo no lineal de sobre-relajación sucesiva (SOR), para resolver el problema de completación matricial (MC), sin utilizar la descomposición de valores singulares de la matriz.

2.1. Esquema de minimización alternante

El problema de completación matricial es el de recuperar una matriz de rango bajo de un subconjunto de las entradas de la matriz original, es decir:

$$\min_{W \in \mathbb{R}^{m \times n}} \text{rank}(W), \quad s.a. \quad W_{ij} = M_{ij}, \forall (i, j) \in \Omega \quad (2.1)$$

donde $\text{rank}(W)$ denota el rango de la W , y M_{ij} son dadas por $(i, j) \in \Omega \subset \{(i, j) : 1 \leq i \leq m, 1 \leq j \leq n\}$.

En [7,8,9] se ha probado que bajo ciertas condiciones la solución del problema (2.1) puede ser encontrada resolviendo el siguiente problema de optimización convexa:

$$\min_{W \in \mathbb{R}^{m \times n}} \|W\|_*, \quad s.a. \quad W_{ij} = M_{ij}, \forall (i, j) \in \Omega \quad (2.2)$$

En [1] los autores plantean resolver un modelo más explícito que (2.2) y de esta manera evitar el cálculo de la descomposición de valores singulares (SVD). Dicho modelo es el siguiente:

$$\min_{X, Y, Z} \frac{1}{2} \|XY - Z\|_F^2, \quad s.a. \quad Z_{ij} = M_{ij}, \forall (i, j) \in \Omega \quad (2.3)$$

donde $X \in \mathbb{R}^{m \times K}$, $Y \in \mathbb{R}^{K \times n}$, $Z \in \mathbb{R}^{m \times n}$.

Al igual que en muchos otros casos similares, la estructura de (2.3) sugiere un esquema de minimización alternante. En este caso, uno puede actualizar cada una de las variables X , Y o Z

eficientemente mientras se fijan las otras dos. Los subproblemas con respecto a la variable X o bien a la variable Y son problemas lineales de mínimos cuadrados que solo involucran $K \times K$ coeficientes matriciales en sus ecuaciones normales, y para obtener la solución del subproblema para la variable Z solo es necesario el producto XY . Este procedimiento de minimización alternante es también conocido como un esquema Gauss-Seidel(GS) por bloque.

2.1.1. El método de Gauss-Seidel no lineal

Consideremos la función Lagrangeana para el problema (2.3), la cual esta definida por:

$$L(X, Y, Z, \Lambda) = \frac{1}{2} \|XY - Z\|_F^2 - \langle \Lambda, \mathcal{P}_\Omega(Z - M) \rangle \quad (2.4)$$

donde la matriz de multiplicadores de Lagrange $\Lambda \in \mathbb{R}^{m \times n}$ cumple que $\Lambda = \mathcal{P}_\Omega(\Lambda)$.

Como es bien conocido en programación matemática, las condiciones de optimalidad de primer orden para el problema (2.3) derivan en un sistema de ecuaciones no lineales, tal sistema de ecuaciones se obtiene diferenciando la función (2.4) e igualando a cero, de esta manera obtenemos:

$$\begin{aligned} \frac{\partial L}{\partial X} = 0 &\Rightarrow (XY - Z)Y^\top = 0 \\ \frac{\partial L}{\partial Y} = 0 &\Rightarrow X^\top(XY - Z) = 0 \\ \frac{\partial L}{\partial Z} = 0 &\Rightarrow \mathcal{P}_\Omega(Z - M) = 0 \\ \frac{\partial L}{\partial \Lambda} = 0 &\Rightarrow \mathcal{P}_\Omega^c(Z - XY) = 0 \text{ y } \mathcal{P}_\Omega(Z - XY) = \Lambda \end{aligned} \quad (2.5)$$

Ahora, aplicaremos el método de minimización alternante de Gauss-Seidel(GS) al sistema de ecuaciones no lineal anterior. Para la instancia \bar{Y} y \bar{Z} fijos, obtenemos X_+ resolviendo el problema $X_+ = \operatorname{argmin}_{X \in \mathbb{R}^{m \times K}} \frac{1}{2} \|X\bar{Y} - \bar{Z}\|_F^2$. Para obtener el valor de X_+ basta minimizar la función lagrangeana para el problema de $\min_{X \in \mathbb{R}^{m \times K}} \frac{1}{2} \|X\bar{Y} - \bar{Z}\|_F^2$ con respecto de X , y como vimos anteriormente, diferenciando tal función respecto a X e igualando a cero tenemos que:

$$(X\bar{Y} - \bar{Z})\bar{Y}^\top = 0,$$

luego si distribuimos se obtiene,

$$X\bar{Y}\bar{Y}^\top - \bar{Z}\bar{Y}^\top = 0,$$

lo cual implica,

$$X\bar{Y}\bar{Y}^\top = \bar{Z}\bar{Y}^\top,$$

multiplicando por $(\bar{Y}\bar{Y}^\top)^\dagger$ por la derecha en ambos lados obtenemos,

$$X_+ = \bar{Z}\bar{Y}^\top (\bar{Y}\bar{Y}^\top)^\dagger$$

Similarmente, para el X_+ calculado anteriormente y el \bar{Z} anterior, obtenemos el valor de Y_+ resolviendo el problema $Y_+ = \operatorname{argmin}_{Y \in \mathbb{R}^{K \times n}} \frac{1}{2} \|X_+Y - \bar{Z}\|_F^2$. Para obtener tal valor, basta minimizar la función lagrangeana para el problema $\min_{Y \in \mathbb{R}^{K \times n}} \frac{1}{2} \|X_+Y - \bar{Z}\|_F^2$ con respecto de Y , y como vimos en el sistema (2.5) se debe cumplir que:

$$X_+^\top(X_+Y - \bar{Z}) = 0,$$

luego si distribuimos se obtiene,

$$X_+^\top X_+Y - X_+^\top \bar{Z} = 0,$$

lo que implica,

$$X_+^\top X_+ Y = X_+^\top \bar{Z},$$

multiplicando por $(X_+^\top X_+)^{\dagger}$ por la izquierda en ambos lados obtenemos,

$$Y_+ = (X_+^\top X_+)^{\dagger} X_+^\top \bar{Z}.$$

De manera similar se procede para Z . Mediante este procedimiento se obtiene el siguiente esquema iterativo:

Esquema Básico 1

$$X_+ := ZY^{\dagger} \equiv ZY^{\top}(YY^{\top})^{\dagger}, \quad (2.6)$$

$$Y_+ := (X_+)^{\dagger} Z \equiv (X_+^\top X_+)^{\dagger} (X_+^\top Z), \quad (2.7)$$

$$Z_+ := X_+ Y_+ + \mathcal{P}_{\Omega}(M - X_+ Y_+) \quad (2.8)$$

2.1.2. Esquemas Alternativos

En esta sección presentaremos dos esquemas equivalentes al esquema básico 1, como alternativas de dicho esquema que pueden ser más eficientes a la hora de implementarlos en la práctica. Para establecer cada esquema primero introduciremos el siguiente lema:

Lema 1 Sean (X_+, Y_+) actualizaciones generadas por el esquema básico 1. Entonces se cumple que:

$$X_+ Y_+ = \mathcal{P}_{ZY^{\top}} Z = ZY^{\top}(YZ^{\top}ZY^{\top})^{\dagger}(YZ^{\top})Z. \quad (2.9)$$

donde $\mathcal{P}_A = A(A^{\top}A)^{\dagger}A^{\top} = QQ^{\top}$ es la proyección ortogonal de la matriz A sobre el subespacio imagen $\mathcal{R}(A)$ y $Q = \text{orth}(A)$ es una base ortonormal para $\mathcal{R}(A)$.

DEMOSTRACIÓN:

Afirmación: $\mathcal{R}(ZY^{\top}) = \mathcal{R}(ZV) = \mathcal{R}(X_+)$.

Procedamos a demostrar la afirmación anterior. Consideremos la descomposición en valores singulares de la matriz Y ; es decir:

$$Y = U\Sigma V^{\top}$$

donde $U \in \mathbb{R}^{k \times k}$, $V \in \mathbb{R}^{n \times n}$ son matrices ortogonales y $\Sigma \in \mathbb{R}^{k \times n}$ es una matriz diagonal. Por definición sabemos que:

$$\mathcal{R}(X_+) = \{t \in \mathbb{R}^m : t = X_+ s, \text{ para algun } s \in \mathbb{R}^k\}$$

$$\mathcal{R}(ZY^{\top}) = \{p \in \mathbb{R}^m : p = ZY^{\top} q, \text{ para algun } q \in \mathbb{R}^k\}$$

$$\mathcal{R}(ZV) = \{r \in \mathbb{R}^m : r = ZVi, \text{ para algun } i \in \mathbb{R}^n\}.$$

Primero veamos que: $\mathcal{R}(ZY^{\top}) = \mathcal{R}(ZV)$.

En efecto, tomemos $p \in \mathcal{R}(ZY^\top)$ arbitrario, así, existe $q \in \mathbb{R}^k$ tal que,

$$p = ZY^\top q = ZV\Sigma^\top U^\top q \quad (\text{ya que } Y^\top = V\Sigma^\top U^\top)$$

o equivalentemente, $p = ZVi$, donde $i = \Sigma^\top U^\top q \in \mathbb{R}^n$.

Lo cual implica que $p \in \mathcal{R}(ZV)$, debido a que p es arbitrario obtenemos:

$$\mathcal{R}(ZY^\top) \subset \mathcal{R}(ZV) \quad (1)$$

Por otro lado, tomemos $r \in \mathcal{R}(ZV)$ arbitrario, así, existe $i \in \mathbb{R}^n$ tal que,

$$r = ZVi = Z[Y^\top U(\Sigma^\top)^\dagger]i \quad (\text{ya que } V = Y^\top U(\Sigma^\top)^\dagger)$$

o equivalentemente $r = ZY^\top q$, donde $q = U(\Sigma^\top)^\dagger i \in \mathbb{R}^k$.

Lo cual implica que $r \in \mathcal{R}(ZY^\top)$, y debido a la arbitrariedad de r obtenemos:

$$\mathcal{R}(ZV) \subset \mathcal{R}(ZY^\top) \quad (2)$$

Luego, de (1) y (2) obtenemos la siguiente igualdad $\mathcal{R}(ZV) = \mathcal{R}(ZY^\top)$ (3)

Ahora, veamos que $\mathcal{R}(ZV) = \mathcal{R}(X_+)$.

En efecto, nótese que $X_+ = ZY^\dagger$ así, $X_+ = Z(U\Sigma V^\top)^\dagger$, lo cual implica que $X_+ = ZV\Sigma^\dagger U^\top$ ya que tanto U como V son matrices ortogonales. Sea $t \in \mathcal{R}(X_+)$ arbitrario, así que existe $s \in \mathbb{R}^k$ tal que

$$t = X_+s = ZV\Sigma^\dagger U^\top s = ZVi \quad \text{donde } i = \Sigma^\dagger U^\top s \in \mathbb{R}^n.$$

Lo cual implica que $t \in \mathcal{R}(ZV)$, y debido a que t fue escogido arbitrariamente obtenemos:

$$\mathcal{R}(X_+) \subset \mathcal{R}(ZV) \quad (4)$$

Por otro lado, sea $r \in \mathcal{R}(ZV)$, así que existe $i \in \mathbb{R}^n$ tal que

$$r = ZVi = X_+U\Sigma i \quad (\text{Despejando } ZV \text{ de la igualdad } X_+ = ZV\Sigma^\dagger U^\top)$$

o equivalentemente, $r = X_+s$ donde $s = U\Sigma i \in \mathbb{R}^k$. Lo cual implica que $r \in \mathcal{R}(X_+)$, y debido a que r fue escogido de manera arbitraria obtenemos:

$$\mathcal{R}(ZV) \subset \mathcal{R}(X_+) \quad (5)$$

De (4) y (5) obtenemos la siguiente igualdad $\mathcal{R}(ZV) = \mathcal{R}(X_+)$ (6)

De (3) y (6) se concluye que $\mathcal{R}(ZY^\top) = \mathcal{R}(ZV) = \mathcal{R}(X_+)$, con lo cual queda probada la afirmación.

Ahora bien, como $\mathcal{R}(ZY^\top) = \mathcal{R}(X_+)$ entonces se tiene que $\mathcal{P}_{ZY^\top} = \mathcal{P}_{X_+}$. Además, nótese que:

$$\begin{aligned}
X_+Y_+ &= (X_+(X_+^\top X_+)^{\dagger}X_+^\top)Z \\
&= \mathcal{P}_{X_+}Z \\
&= \mathcal{P}_{ZY^\top}Z \\
&= ZY^\top(YZ^\top ZY^\top)^{\dagger}(YZ^\top)Z
\end{aligned}$$

Con lo cual queda demostrado el lema 1. \square

Como vimos en el capítulo 1 se sabe que a través de la factorización QR y la descomposición en valores singulares SVD de una matriz A , se puede recuperar la pseudoinversa de A , la base ortonormal de $\text{ran}(A)$ y la proyección ortogonal sobre $\text{ran}(A)$. Nótese que la función objetivo (2.3) esta determinada por el producto X_+Y_+ , claramente diferentes valores de X_+ y Y_+ son esencialmente equivalentes siempre y cuando se obtenga el mismo valor del producto X_+Y_+ . Así, el Lema anterior prueba que la pseudoinversa $(YY^\top)^{\dagger}$ puede ser recuperada al calcular la proyección \mathcal{P}_{ZY^\top} . La primera variante del esquema básico 1 es la siguiente:

Esquema alternativo 1

$$X_+ := ZY^\top, \quad (2.10)$$

$$Y_+ := (X_+)^{\dagger}Z \equiv (X_+^\top X_+)^{\dagger}(X_+^\top Z), \quad (2.11)$$

$$Z_+ := X_+Y_+ + \mathcal{P}_\Omega(M - X_+Y_+). \quad (2.12)$$

La característica principal de este nuevo esquema, es que solo en cada iteración esta involucrado un único problema de mínimos cuadrados.

La segunda variante del esquema básico 1 calcula la proyección ortogonal $\mathcal{P}_{ZY^\top} = VV^\top$, donde $V := \text{orth}(ZY^\top)$ es una base ortonormal del subespacio $\mathcal{R}(ZY^\top)$. Por tanto, gracias al lema 1 podemos escribir $X_+Y_+ = VV^\top Z$, así obtenemos es siguiente esquema:

Esquema alternativo 2

$$X_+ := V, \quad (2.13)$$

$$Y_+ := V^\top Z, \quad (2.14)$$

$$Z_+ := X_+Y_+ + \mathcal{P}_\Omega(M - X_+Y_+). \quad (2.15)$$

Este último esquema, frecuentemente es preferido debido a que el cálculo del paso (2.14) mediante la factorización QR es generalmente más estable que resolver las ecuaciones normales. Nótese que los tres esquemas pueden ser usados de forma intercambiable.

2.1.3. Un esquema SOR no lineal

En algebra lineal numérica, el método de sobre-relajación sucesiva (SOR) para resolver un sistema de ecuaciones es divisado mediante una aplicación de una extrapolación del método GS, esto es, el nuevo punto de partida es un promedio ponderado entre el previo iterado y el iterado sucesivo calculado por el GS para cada componente. Un valor apropiado del peso ω frecuentemente resulta en una convergencia mas rápida. Aplicando esta idea a el esquema básico 1 obtenemos:

Esquema SOR no lineal

$$X_+ := ZY^\top(YY^\top)^\dagger, \quad (2.16)$$

$$X_+(\omega) := \omega X_+ + (1 - \omega)X, \quad (2.17)$$

$$Y_+ := (X_+(\omega)^\top X_+(\omega))^\dagger (X_+(\omega)^\top Z), \quad (2.18)$$

$$Y_+(\omega) := \omega Y_+ + (1 - \omega)Y, \quad (2.19)$$

$$Z_+(\omega) := X_+(\omega)Y_+(\omega) + \mathcal{P}_\Omega(M - X_+(\omega)Y_+(\omega)). \quad (2.20)$$

donde el peso $\omega \geq 1$. Es claro que, para $\omega = 1$ obtenemos el esquema básico GS.

Suponiendo que la matriz Y tiene rango filas completo, los dos problemas de mínimos cuadrados en el esquema SOR no lineal pueden ser reducidos en un segundo esquema básico análogo al esquema alternativo 1.

Denotemos el residual por,

$$S = \mathcal{P}_\Omega(M - XY), \quad (2.21)$$

el cual será utilizado para medir optimalidad. Luego de cada iteración, la variable Z , la cual es factible, puede ser expresada como $Z = XY + S$. Definamos Z_ω como una suma ponderada de las matrices XY y S , esto es,

$$\begin{aligned} Z_\omega &:= XY + \omega S \\ &= \omega Z + (1 - \omega)XY \end{aligned} \quad (2.22)$$

Ahora, nótese que:

$$Z_\omega Y^\top (YY^\top)^\dagger = (\omega Z + (1 - \omega)XY)Y^\top (YY^\top)^\dagger,$$

distribuyendo se tiene que,

$$Z_\omega Y^\top (YY^\top)^\dagger = \omega ZY^\top (YY^\top)^\dagger + (1 - \omega)XY Y^\top (YY^\top)^\dagger,$$

de (2.16) y como $YY^\top (YY^\top)^\dagger = I$ obtenemos,

$$Z_\omega Y^\top (YY^\top)^\dagger = \omega X_+ + (1 - \omega)X,$$

el cual es exactamente el paso (2.17). Similarmente, nótese que

$$(X_+(\omega)^\top X_+(\omega))^\dagger (X_+(\omega)^\top Z_\omega) = (X_+(\omega)^\top X_+(\omega))^\dagger [X_+(\omega)^\top (\omega Z + (1 - \omega)X_+(\omega)Y)],$$

distribuyendo se tiene que,

$$(X_+(\omega)^\top X_+(\omega))^\dagger (X_+(\omega)^\top Z_\omega) = (X_+(\omega)^\top X_+(\omega))^\dagger [\omega X_+(\omega)^\top Z + (1 - \omega)X_+(\omega)^\top X_+(\omega)Y],$$

nuevamente distribuyendo y usando el hecho de que $(X_+(\omega)^\top X_+(\omega))^\dagger (X_+(\omega)^\top X_+(\omega)) = I$ obtenemos,

$$(X_+(\omega)^\top X_+(\omega))^\dagger (X_+(\omega)^\top Z_\omega) = \omega (X_+(\omega)^\top X_+(\omega))^\dagger X_+(\omega)^\top Z + (1 - \omega)Y,$$

luego por (2.18) tenemos que,

$$(X_+(\omega)^\top X_+(\omega))^\dagger (X_+(\omega)^\top Z_\omega) = \omega Y_+ + (1 - \omega)Y,$$

el cual es exactamente el paso (2.19). De esta manera concluimos que el esquema SOR no lineal es equivalente al siguiente esquema:

Esquema Iterativo Nsor

$$X_+(\boldsymbol{\omega}) := Z_\omega Y^\dagger \equiv Z_\omega Y^\top (Y Y^\top)^\dagger, \quad (2.23)$$

$$Y_+(\boldsymbol{\omega}) := (X_+(\boldsymbol{\omega})^\top X_+(\boldsymbol{\omega}))^\dagger (X_+(\boldsymbol{\omega})^\top Z_\omega) \quad (2.24)$$

$$\mathcal{P}_{\Omega^c}(Z_+(\boldsymbol{\omega})) := \mathcal{P}_{\Omega^c}(X_+(\boldsymbol{\omega})Y_+(\boldsymbol{\omega})), \quad (2.25)$$

$$\mathcal{P}_\Omega(Z_+(\boldsymbol{\omega})) := \mathcal{P}_\Omega(M). \quad (2.26)$$

Nuevamente, se pueden utilizar implementaciones análogas a los esquemas alternativos 1 y 2 para este esquema.

Ahora, presentaremos una estrategia para escoger el valor más conveniente del peso $\boldsymbol{\omega}$ de tal manera que el esquema iterativo Nsor pueda converger. Por supuesto, un peso fijo no es suficientemente bueno para funciones no-lineales y no convexas. La estrategia será ajustar el peso $\boldsymbol{\omega}$ dinámicamente de acuerdo a los cambios de los valores de la función objetivo. Luego que la terna $(X_+(\boldsymbol{\omega}), Y_+(\boldsymbol{\omega}), Z_+(\boldsymbol{\omega}))$ sea estimada, calcularemos el radio residual:

$$\gamma(\boldsymbol{\omega}) = \frac{\|S_+(\boldsymbol{\omega})\|_F}{\|S\|_F}, \quad (2.27)$$

donde,

$$S_+(\boldsymbol{\omega}) := \mathcal{P}_\Omega(M - X_+(\boldsymbol{\omega})Y_+(\boldsymbol{\omega})). \quad (2.28)$$

Si $\gamma(\boldsymbol{\omega}) < 1$, ese nuevo par de puntos es aceptado y el paso es llamado *exitoso*, en otro caso, el paso es *no exitoso*; en caso de que esto ocurra se actualiza a $\boldsymbol{\omega} = 1$ y se repite el procedimiento. Nótese que un valor pequeño de $\gamma(\boldsymbol{\omega})$ indica que el valor del peso actual $\boldsymbol{\omega}$ esta trabajando bien. Por lo tanto, $\boldsymbol{\omega}$ será incrementado solo si el punto calculado es aceptable pero el radio residual $\gamma(\boldsymbol{\omega})$ es considerado "muy grande", el cual será cuando, $\gamma(\boldsymbol{\omega}) \in [\gamma_1, 1)$ para algun $\gamma_1 \in (0, 1)$. Tomando en cuenta estas consideraciones y el esquema iterativo Nsor, se establece el siguiente algoritmo:

Un algoritmo apropiado para matrices de bajo rango (LMaFit)

INPUT Ω , $\mathcal{P}_\Omega(M)$ y $K \geq r = \text{rank}(M)$.

1. $Y^0 \in \mathbb{R}^{K \times n}$, $Z^0 = \mathcal{P}_\Omega(M)$, $\boldsymbol{\omega} = 1$, $\tilde{\boldsymbol{\omega}} > 1$, $\gamma_1 \in (0, 1)$ y $k = 0$.
2. **while** not convergent **do**
3. calcular $(X_+(\boldsymbol{\omega}), Y_+(\boldsymbol{\omega}), Z_+(\boldsymbol{\omega}))$ de acuerdo al (esquema iterativo Nsor) con $(X, Y, Z) = (X^k, Y^k, Z^k)$.
4. calcular el radio residual $\gamma(\boldsymbol{\omega})$ de acuerdo a (2.27).
5. **if** $\gamma(\boldsymbol{\omega}) \geq 1$ **then** hacer $\boldsymbol{\omega} = 1$ y vaya al paso 4.
6. Actualizar $(X^{k+1}, Y^{k+1}, Z^{k+1}) = (X_+(\boldsymbol{\omega}), Y_+(\boldsymbol{\omega}), Z_+(\boldsymbol{\omega}))$ e incrementar k.

7. **if** $\gamma(\omega) \geq \gamma_1$ **then** hacer $\delta = \max(\delta, 0, 25(\omega - 1))$ y $\omega = \min(\omega + \delta, \tilde{\omega})$.
 8. **end**(while)

OUTPUT XY

2.1.4. Extensión para problemas con restricciones lineales generales

Tanto la aproximación GS como la aproximación SOR pueden ser aplicadas para resolver problemas de minimización de rango con restricciones lineales generales, es decir:

$$\min_{W \in \mathbb{R}^{m \times n}} \text{rank}(W), \text{ s.a. } \mathcal{A}(W) = b, \quad (2.29)$$

donde el operador lineal $\mathcal{A} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^q$ y el vector $b \in \mathbb{R}^q$ son dados. Similarmente a (2.3), el problema (2.29) se convierte mediante una relajación no-convexa en:

$$\min_{X, Y, Z} \frac{1}{2} \|XY - Z\|_F^2 \text{ s.a. } \mathcal{A}(W) = b, \quad (2.30)$$

donde $X \in \mathbb{R}^{m \times K}$, $Y \in \mathbb{R}^{K \times n}$, $Z \in \mathbb{R}^{m \times n}$. Para valores fijos cualesquiera de X y Y , el minimizador Z_+ , de (2.30) con respecto a Z es:

$$Z_+ = XY + \mathcal{A}^\top (\mathcal{A} \mathcal{A}^\top)^\dagger (b - \mathcal{A}(XY)).$$

Sea $S := \mathcal{A}^\top (\mathcal{A} \mathcal{A}^\top)^\dagger (b - \mathcal{A}(XY))$ y $Z_\omega = XY + \omega S$. Similarmente a el esquema iterativo Nsor, obtenemos:

$$X_+(\omega) := Z_\omega Y^\dagger \equiv Z_\omega Y^\top (Y Y^\top)^\dagger, \quad (2.31)$$

$$Y_+(\omega) := (X_+(\omega)^\top X_+(\omega))^\dagger (X_+(\omega)^\top Z_\omega), \quad (2.32)$$

$$Z_+(\omega) := X_+(\omega) Y_+(\omega) + \mathcal{A}^\top (\mathcal{A} \mathcal{A}^\top)^\dagger (b - \mathcal{A}(X_+(\omega) Y_+(\omega))). \quad (2.33)$$

2.2. Análisis de Convergencia

Ahora analizaremos el algoritmo *LMaFit* para revelar la relación entre los residuales $\|S\|_F$ y $\|S_+(\omega)\|_F$. Denotemos por $V(\omega) := \text{orth}(X_+(\omega))$ y por $U := \text{orth}(Y^\top)$ a las bases ortonormales de los subespacios imagen de $\mathcal{R}(X_+(\omega))$ y $\mathcal{R}(Y^\top)$, respectivamente. Consecuentemente, las proyecciones ortogonales sobre $\mathcal{R}(X_+(\omega))$ y $\mathcal{R}(Y^\top)$ pueden se expresadas como:

$$Q(\omega) := V(\omega) V(\omega)^\top = X_+(\omega) (X_+(\omega)^\top X_+(\omega))^\dagger X_+(\omega)^\top,$$

$$P := U U^\top = Y^\top (Y Y^\top)^\dagger Y.$$

De la definición de pseudoinversa se desprenden varias identidades importantes, estas son:

$$A^\dagger = A^\dagger (A^\dagger)^\top A^\top = A^\top (A^\dagger)^\top A^\dagger = (A^\top A)^\dagger A^\top = A^\top (A A^\top)^\dagger, \quad (2.34)$$

$$A = (A^\dagger)^\top A^\top A = A A^\top (A^\dagger)^\top. \quad (2.35)$$

El siguiente lema y su demostración nos proveerá una ecuación clave.

Lema 2 Sea $(X_+(\omega), Y_+(\omega))$ generado por el esquema iterativo Nsor. Entonces se cumple:

$$\langle \omega S, (X_+(\omega)Y_+(\omega) - XY) \rangle = \|X_+(\omega)Y_+(\omega) - XY\|_F^2. \quad (2.36)$$

DEMOSTRACIÓN: De (2.35) se tiene que $Y = YY^\top(Y^\dagger)^\top$, aplicando la traspuesta en ambos lados de la igualdad se obtiene, $Y^\top = Y^\dagger YY^\top$. De (2.23) sabemos que,

$$X_+(\omega) = Z_\omega Y^\dagger$$

multiplicando por YY^\top en ambos lados de la igualdad obtenemos,

$$X_+(\omega)[YY^\top] = Z_\omega Y^\dagger[YY^\top]$$

usando el hecho de que $Y^\top = Y^\dagger YY^\top$ se tiene que:

$$X_+(\omega)YY^\top = Z_\omega Y^\top$$

pero por definición sabemos que $Z_\omega = XY + \omega S$, luego,

$$X_+(\omega)YY^\top = XYY^\top + \omega SY^\top.$$

Multiplicando a esta última ecuación por $(YY^\top)^\dagger Y$ en ambos lados obtenemos,

$$X_+(\omega)YY^\top[(YY^\top)^\dagger Y] = [XYY^\top + \omega SY^\top](YY^\top)^\dagger Y,$$

distribuyendo y usando el hecho de que $(YY^\top)(YY^\top)^\dagger = I$ tenemos que:

$$X_+(\omega)Y = XY + \omega SY^\top(YY^\top)^\dagger Y$$

reagrupando y usando la definición de P se obtiene,

$$(X_+(\omega) - X)Y = \omega SP. \quad (2.37)$$

Por otro lado, de (2.35) sabemos que $X_+(\omega) = (X_+(\omega)^\dagger)^\top X_+(\omega)^\top X_+(\omega)$, aplicando la traspuesta en ambos lados de la igualdad se obtiene, $X_+(\omega)^\top = X_+(\omega)^\top X_+(\omega)X_+(\omega)^\dagger$. De (2.24) sabemos que:

$$Y_+(\omega) = (X_+(\omega)^\top X_+(\omega))^\dagger (X_+(\omega)^\top Z_\omega),$$

multiplicando en ambos lados de la igualdad por $[X_+(\omega)^\top X_+(\omega)]$ tenemos que,

$$[X_+(\omega)^\top X_+(\omega)]Y_+(\omega) = X_+(\omega)^\top X_+(\omega)^\dagger (X_+(\omega)^\top Z_\omega),$$

sustituyendo $X_+(\omega)^\top = X_+(\omega)^\top X_+(\omega)X_+(\omega)^\dagger$ en la ecuación anterior se obtiene:

$$[X_+(\omega)^\top X_+(\omega)]Y_+(\omega) = X_+(\omega)^\top X_+(\omega)^\dagger [(X_+(\omega)^\top X_+(\omega)X_+(\omega)^\dagger)Z_\omega],$$

usando el hecho de que $(X_+(\omega)^\top X_+(\omega))^\dagger (X_+(\omega)^\top X_+(\omega)) = I$, tenemos,

$$[X_+(\omega)^\top X_+(\omega)]Y_+(\omega) = [X_+(\omega)^\top X_+(\omega)]X_+(\omega)^\dagger Z_\omega,$$

nuevamente sustituyendo la ecuación $X_+(\omega)^\top = X_+(\omega)^\top X_+(\omega)X_+(\omega)^\dagger$ en la igualdad anterior, se cumple que,

$$X_+(\omega)^\top X_+(\omega)Y_+(\omega) = X_+(\omega)^\top Z_\omega,$$

luego, de la definición de Z_ω se tiene que:

$$X_+(\omega)^\top X_+(\omega)Y_+(\omega) = X_+(\omega)^\top (XY + \omega S),$$

así, si sumamos y restamos el término $X_+(\omega)Y$ se obtiene,

$$X_+(\omega)^\top X_+(\omega)Y_+(\omega) = X_+(\omega)^\top (XY + \omega S + X_+(\omega)Y - X_+(\omega)Y),$$

asociando y extrayendo el factor comun,

$$X_+(\omega)^\top X_+(\omega)Y_+(\omega) = X_+(\omega)^\top (X_+(\omega)Y - (X_+(\omega) - X)Y + \omega S),$$

utilizando (2.37) y sustituyendola en la ecuación anterior tenemos,

$$X_+(\omega)^\top X_+(\omega)Y_+(\omega) = X_+(\omega)^\top (X_+(\omega)Y - \omega SP + \omega S),$$

o equivalentemente,

$$X_+(\omega)^\top X_+(\omega)Y_+(\omega) = X_+(\omega)^\top (X_+(\omega)Y + \omega S(I - P)),$$

luego, distribuyendo se tiene que,

$$X_+(\omega)^\top X_+(\omega)Y_+(\omega) = X_+(\omega)^\top X_+(\omega)Y + \omega X_+(\omega)^\top S(I - P).$$

Multiplicando esta última ecuación en ambos lados por $[X_+(\omega)(X_+(\omega)^\top X_+(\omega))^\dagger]$ se tiene,

$$\begin{aligned} & [X_+(\omega)(X_+(\omega)^\top X_+(\omega))^\dagger]X_+(\omega)^\top X_+(\omega)Y_+(\omega) = \\ & = [X_+(\omega)(X_+(\omega)^\top X_+(\omega))^\dagger][X_+(\omega)^\top X_+(\omega)Y + \omega X_+(\omega)^\top S(I - P)], \end{aligned}$$

distribuyendo obtenemos,

$$\begin{aligned} & [X_+(\omega)(X_+(\omega)^\top X_+(\omega))^\dagger]X_+(\omega)^\top X_+(\omega)Y_+(\omega) = \\ & = X_+(\omega)(X_+(\omega)^\top X_+(\omega))^\dagger X_+(\omega)^\top X_+(\omega)Y + \omega X_+(\omega)(X_+(\omega)^\top X_+(\omega))^\dagger X_+(\omega)^\top S(I - P), \end{aligned}$$

como $(X_+(\omega)^\top X_+(\omega))^\dagger (X_+(\omega)^\top X_+(\omega)) = I$, entonces

$$X_+(\omega)Y_+(\omega) = X_+(\omega)Y + \omega X_+(\omega)(X_+(\omega)^\top X_+(\omega))^\dagger X_+(\omega)^\top S(I - P),$$

reagrupando y usando la definición de $Q(\omega)$ llegamos a,

$$X_+(\omega)(Y_+(\omega) - Y) = \omega Q(\omega)S(I - P). \quad (2.38)$$

Ahora, nótese que,

$$X_+(\omega)Y_+(\omega) - XY = (X_+(\omega) - X)Y + X_+(\omega)(Y_+(\omega) - Y),$$

sustituyendo (2.37) y (2.38) en la ecuación anterior obtenemos,

$$X_+(\omega)Y_+(\omega) - XY = \omega SP + \omega Q(\omega)S(I - P), \quad (2.39)$$

distribuyendo tenemos que,

$$X_+(\omega)Y_+(\omega) - XY = \omega SP + \omega Q(\omega)S - \omega Q(\omega)SP,$$

luego,

$$X_+(\omega)Y_+(\omega) - XY = \omega(I - Q(\omega))SP + \omega Q(\omega)S. \quad (2.40)$$

Por lo tanto,

$$\|X_+(\omega)Y_+(\omega) - XY\|_F^2 = \omega^2\|(I - Q(\omega))SP\|_F^2 + \omega^2\|Q(\omega)S\|_F^2. \quad (2.41)$$

ya que $Q(\omega)$ proyecta sobre $\mathcal{R}(X_+(\omega))$, y $I - Q(\omega)$ proyecta sobre $\mathcal{R}(X_+(\omega))^\perp$.

Finalmente, de (2.40) y las propiedades de proyecciones ortogonales, tenemos:

$$\langle \omega S, (X_+(\omega)Y_+(\omega) - XY) \rangle = \omega^2 \langle S, (I - Q(\omega))SP \rangle + \omega^2 \langle S, Q(\omega)S \rangle,$$

luego,

$$\langle \omega S, (X_+(\omega)Y_+(\omega) - XY) \rangle = \omega^2 \langle SP, (I - Q(\omega))SP \rangle + \omega^2 \langle S, Q(\omega)S \rangle,$$

así,

$$\langle \omega S, (X_+(\omega)Y_+(\omega) - XY) \rangle = \omega^2\|(I - Q(\omega))SP\|_F^2 + \omega^2\|Q(\omega)S\|_F^2,$$

sustituyendo (2.41) en la ecuación anterior obtenemos,

$$\langle \omega S, (X_+(\omega)Y_+(\omega) - XY) \rangle = \|X_+(\omega)Y_+(\omega) - XY\|_F^2.$$

lo que prueba el lema. \square

Nótese que,

$$\begin{aligned} \frac{1}{\omega} \|XY - Z_\omega\|_F &= \frac{1}{\omega} \|XY - XY - \omega S\|_F \\ &= \frac{1}{\omega} \|\omega S\|_F \end{aligned}$$

así,

$$\frac{1}{\omega} \|XY - Z_\omega\|_F = \|S\|_F. \quad (2.42)$$

Además, luego de los dos primeros pasos del esquema iterativo Nsor se tiene que,

$$\frac{1}{\omega} \|X_+(\omega)Y_+(\omega) - Z_\omega\|_F \leq \|S\|_F,$$

y la desigualdad se satisface de forma estricta a menos que las dos primeras ecuaciones de las condiciones de optimalidad de (2.5) se cumplan. O equivalentemente,

$$\frac{1}{\omega^2} (\|\mathcal{P}_{\Omega^c}(X_+(\omega)Y_+(\omega) - Z_\omega)\|_F^2 + \|\mathcal{P}_\Omega(X_+(\omega)Y_+(\omega) - Z_\omega)\|_F^2) \leq \|S\|_F^2. \quad (2.43)$$

Lema 3 *Sea $(X_+(\omega), Y_+(\omega))$ generado por el esquema iterativo Nsor para cualquier $\omega \geq 1$, entonces*

$$\frac{1}{\omega^2} \|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \|(I - Q(\omega))S(I - P)\|_F^2 = \|S\|_F^2 - \rho_{12}(\omega), \quad (2.44)$$

donde,

$$\rho_{12}(\omega) := \|SP\|_F^2 + \|Q(\omega)S(I - P)\|_F^2 \quad (2.45)$$

es la cantidad de reducción del residual $\|S\|_F^2$ después de los dos primeros pasos del esquema iterativo Nsor.

DEMOSTRACIÓN: De la definición de Z_ω se cumple que,

$$X_+(\omega)Y_+(\omega) - Z_\omega = X_+(\omega)Y_+(\omega) - XY - \omega S,$$

sustituyendo (2.39) en la ecuación anterior se tiene que,

$$X_+(\omega)Y_+(\omega) - Z_\omega = [\omega SP + \omega Q(\omega)S(I-P)] - \omega S,$$

asociando y sacando factor común se obtiene,

$$X_+(\omega)Y_+(\omega) - Z_\omega = \omega Q(\omega)S(I-P) - \omega S(I-P),$$

luego,

$$X_+(\omega)Y_+(\omega) - Z_\omega = -\omega(I-Q(\omega))S(I-P),$$

aplicando norma Frobenius al cuadrado en ambos lados obtenemos,

$$\|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \omega^2 \|(I-Q(\omega))S(I-P)\|_F^2,$$

por lo tanto,

$$\frac{1}{\omega^2} \|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \|(I-Q(\omega))S(I-P)\|_F^2.$$

lo cual prueba la primera igualdad de (2.44). Por otro lado, por definición de Z_ω tenemos que,

$$\|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \|X_+(\omega)Y_+(\omega) - XY - \omega S\|_F^2,$$

por definición sabemos que,

$$\|X_+(\omega)Y_+(\omega) - XY - \omega S\|_F^2 = \langle (X_+(\omega)Y_+(\omega) - XY - \omega S), (X_+(\omega)Y_+(\omega) - XY - \omega S) \rangle,$$

así, de las dos ecuaciones anteriores obtenemos,

$$\|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \langle (X_+(\omega)Y_+(\omega) - XY - \omega S), (X_+(\omega)Y_+(\omega) - XY - \omega S) \rangle,$$

luego por propiedades de producto interno obtenemos,

$$\|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \|X_+(\omega)Y_+(\omega) - XY\|_F^2 + \omega^2 \|S\|_F^2 - 2\omega \langle S, (X_+(\omega)Y_+(\omega) - XY) \rangle,$$

utilizando el Lema 2 se obtiene,

$$\|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \|X_+(\omega)Y_+(\omega) - XY\|_F^2 + \omega^2 \|S\|_F^2 - 2\|X_+(\omega)Y_+(\omega) - XY\|_F^2,$$

así,

$$\|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \omega^2 \|S\|_F^2 - \|X_+(\omega)Y_+(\omega) - XY\|_F^2,$$

sustituyendo (2.39) en la ecuación de arriba llegamos a,

$$\|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \omega^2 \|S\|_F^2 - \|\omega SP + \omega Q(\omega)S(I-P)\|_F^2,$$

o equivalentemente,

$$\|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \omega^2 \|S\|_F^2 - \langle \omega SP + \omega Q(\omega)S(I-P), \omega SP + \omega Q(\omega)S(I-P) \rangle$$

luego,

$$\|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \omega^2 \|S\|_F^2 - \omega^2 \|SP\|_F^2 - 2\omega^2 \langle SP, Q(\omega)S(I-P) \rangle - \omega^2 \|Q(\omega)S(I-P)\|_F^2,$$

como $\langle SP, Q(\omega)S(I-P) \rangle = 0$, obtenemos

$$\|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \omega^2\|S\|_F^2 - \omega^2\|SP\|_F^2 - \omega^2\|Q(\omega)S(I-P)\|_F^2,$$

sacando factor comun de $-\omega^2$ tenemos que,

$$\|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \omega^2\|S\|_F^2 - \omega^2[\|SP\|_F^2 + \|Q(\omega)S(I-P)\|_F^2]$$

o equivalentemente,

$$\|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \omega^2\|S\|_F^2 - \omega^2\rho_{12}(\omega)$$

por lo tanto,

$$\frac{1}{\omega^2}\|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \|S\|_F^2 - \rho_{12}(\omega)$$

lo cual prueba la segunda igualdad de (2.44). \square

Ahora, luego de los tres primeros pasos en el esquema iterativo Nsor, tenemos que,

$$\|\mathcal{P}_{\Omega^c}(X_+(\omega)Y_+(\omega) - Z_+(\omega))\|_F = 0,$$

ya que $\mathcal{P}_{\Omega^c}(Z_\omega) \equiv \mathcal{P}_{\Omega^c}(XY)$ independientemente del valor de ω . Consecuentemente, la reducción del residual después de los tres pasos es,

$$\rho_3(\omega) := \frac{1}{\omega^2}\|\mathcal{P}_{\Omega^c}(X_+(\omega)Y_+(\omega) - XY)\|_F^2. \quad (2.46)$$

Otra formula para ρ_3 es la siguiente:

$$\rho_3(\omega) := \|\mathcal{P}_{\Omega^c}(SP + Q(\omega)S(I-P))\|_F^2, \quad (2.47)$$

la cual se desprende de la igualdad (2.39).

El cambio del valor del residual luego del cuarto paso esta dado por:

$$\rho_4(\omega) := \frac{1}{\omega^2}\|\mathcal{P}_{\Omega}(X_+(\omega)Y_+(\omega) - Z_\omega)\|_F^2 - \|S_+(\omega)\|_F^2;$$

o equivalentemente,

$$\rho_4(\omega) := \frac{1}{\omega^2}\|S_+(\omega) + (\omega - 1)S\|_F^2 - \|S_+(\omega)\|_F^2 \geq 0. \quad (2.48)$$

Claramente, $\rho_4(1) = 0$.

Proposición 3 *Para valores de $\omega > 1$ se satisface la siguiente igualdad:*

$$\frac{\omega^2\rho_4(\omega)}{\omega - 1} = (\omega - 1)(\|S\|_F^2 - \|S_+(\omega)\|_F^2) - 2S_+(\omega) \bullet (S_+(\omega) - S). \quad (2.49)$$

DEMOSTRACIÓN:

En efecto, de la igualdad (2.48) se tiene que:

$$\frac{\omega^2 \rho_4(\omega)}{\omega - 1} = \frac{1}{\omega - 1} \|S_+(\omega) + (\omega - 1)S\|_F^2 - \frac{\omega^2}{\omega - 1} \|S_+(\omega)\|_F^2$$

luego por la definición de la norma Frobenius tenemos que:

$$\frac{\omega^2 \rho_4(\omega)}{\omega - 1} = \frac{1}{\omega - 1} \langle S_+(\omega) + (\omega - 1)S, S_+(\omega) + (\omega - 1)S \rangle - \frac{\omega^2}{\omega - 1} \langle S_+(\omega), S_+(\omega) \rangle$$

por las propiedades del producto interno obtenemos,

$$\begin{aligned} \frac{\omega^2 \rho_4(\omega)}{\omega - 1} &= \frac{1}{\omega - 1} \langle S_+(\omega), S_+(\omega) \rangle + \frac{1}{\omega - 1} \langle S_+(\omega), (\omega - 1)S \rangle + \frac{1}{\omega - 1} \langle (\omega - 1)S, S_+(\omega) \rangle + \\ &\quad + \frac{1}{\omega - 1} \langle (\omega - 1)S, (\omega - 1)S \rangle - \frac{\omega^2}{\omega - 1} \langle S_+(\omega), S_+(\omega) \rangle \end{aligned}$$

asociando, usando propiedades del producto interno y reacomodando la expresión anterior obtenemos,

$$\frac{\omega^2 \rho_4(\omega)}{\omega - 1} = \frac{1 - \omega^2}{\omega - 1} \langle S_+(\omega), S_+(\omega) \rangle + \langle S_+(\omega), S \rangle + \langle S, S_+(\omega) \rangle + (\omega - 1) \langle S, S \rangle$$

o equivalentemente,

$$\frac{\omega^2 \rho_4(\omega)}{\omega - 1} = -(1 + \omega) \langle S_+(\omega), S_+(\omega) \rangle + 2 \langle S_+(\omega), S \rangle + (\omega - 1) \langle S, S \rangle$$

así,

$$\frac{\omega^2 \rho_4(\omega)}{\omega - 1} = -(\omega - 1 + 2) \langle S_+(\omega), S_+(\omega) \rangle + 2 \langle S_+(\omega), S \rangle + (\omega - 1) \langle S, S \rangle$$

distribuyendo en la ecuación anterior tenemos que,

$$\frac{\omega^2 \rho_4(\omega)}{\omega - 1} = -(\omega - 1) \langle S_+(\omega), S_+(\omega) \rangle - 2 \langle S_+(\omega), S_+(\omega) \rangle + 2 \langle S_+(\omega), S \rangle + (\omega - 1) \langle S, S \rangle$$

reordenando obtenemos que,

$$\frac{\omega^2 \rho_4(\omega)}{\omega - 1} = (\omega - 1) \langle S, S \rangle - (\omega - 1) \langle S_+(\omega), S_+(\omega) \rangle - \langle 2S_+(\omega), S_+(\omega) \rangle + \langle 2S_+(\omega), S \rangle$$

o equivalentemente,

$$\frac{\omega^2 \rho_4(\omega)}{\omega - 1} = (\omega - 1) (\|S\|_F^2 - \|S_+(\omega)\|_F^2) - 2S_+(\omega) \bullet (S_+(\omega) - S)$$

con lo cual queda probada la proposición. \square

Proximamente, nos disponemos a probar que la tasa de cambio de $\rho_4(\omega)$ en $\omega = 1^+$ es no-negativa.

Lema 4

$$\lim_{\omega \rightarrow 1^+} \frac{\rho_4(\omega)}{\omega - 1} = 2 \| \mathcal{P}_{\Omega^c}(X_+(1)Y_+(1) - XY) \|_F^2 \geq 0. \quad (2.50)$$

DEMOSTRACIÓN:

Sea $\omega \rightarrow 1$ y denotemos por S_+ a $S_+(1)$. Nótese que,

$$\lim_{\omega \rightarrow 1^+} \frac{\rho_4(\omega)}{\omega - 1} = \lim_{\omega \rightarrow 1^+} \frac{\omega^2 \rho_4(\omega)}{\omega - 1}$$

ahora, de la proposición 3 se deduce que,

$$\lim_{\omega \rightarrow 1^+} \frac{\omega^2 \rho_4(\omega)}{\omega - 1} = -2S_+ \bullet (S_+ - S)$$

por lo tanto,

$$\lim_{\omega \rightarrow 1^+} \frac{\rho_4(\omega)}{\omega - 1} = -2S_+ \bullet (S_+ - S)$$

de las propiedades del producto interno tenemos que,

$$\lim_{\omega \rightarrow 1^+} \frac{\rho_4(\omega)}{\omega - 1} = -2\langle S_+, S_+ \rangle + 2\langle S_+, S \rangle$$

sumando y restando tanto $2\langle S, S_+ \rangle$ como $2\langle S, S \rangle$ en la ecuación anterior obtenemos,

$$\lim_{\omega \rightarrow 1^+} \frac{\rho_4(\omega)}{\omega - 1} = -2\langle S_+, S_+ \rangle + 2\langle S_+, S \rangle + 2\langle S, S_+ \rangle - 2\langle S, S_+ \rangle + 2\langle S, S \rangle - 2\langle S, S \rangle$$

asociando se tiene que,

$$\lim_{\omega \rightarrow 1^+} \frac{\rho_4(\omega)}{\omega - 1} = [-2\langle S_+, S_+ \rangle + 2\langle S_+, S \rangle + 2\langle S, S_+ \rangle - 2\langle S, S \rangle] - \langle 2S, S_+ \rangle + 2\langle S, S \rangle$$

o equivalentemente,

$$\lim_{\omega \rightarrow 1^+} \frac{\rho_4(\omega)}{\omega - 1} = -2\langle S_+ - S, S_+ - S \rangle - \langle 2S, S_+ \rangle + 2\langle S, S \rangle$$

así,

$$\lim_{\omega \rightarrow 1^+} \frac{\rho_4(\omega)}{\omega - 1} = -2\|S_+ - S\|_F^2 - 2S \bullet (S_+ - S)$$

de las definiciones de S y de S_+ obtenemos,

$$\lim_{\omega \rightarrow 1^+} \frac{\rho_4(\omega)}{\omega - 1} = -2\|\mathcal{P}_\Omega(X_+Y_+ - XY)\|_F^2 + 2S \bullet (X_+Y_+ - XY)$$

o equivalentemente,

$$\lim_{\omega \rightarrow 1^+} \frac{\rho_4(\omega)}{\omega - 1} = 2[S \bullet (X_+Y_+ - XY) - \|\mathcal{P}_\Omega(X_+Y_+ - XY)\|_F^2]$$

luego, por el lema 2 y usando el hecho de que $\|X_+Y_+ - XY\|_F^2 = \|\mathcal{P}_\Omega(X_+Y_+ - XY)\|_F^2 + \|\mathcal{P}_{\Omega^c}(X_+Y_+ - XY)\|_F^2$ se concluye que,

$$\lim_{\omega \rightarrow 1^+} \frac{\rho_4(\omega)}{\omega - 1} = 2\|\mathcal{P}_{\Omega^c}(X_+(1)Y_+(1) - XY)\|_F^2 \geq 0$$

lo cual completa la demostración del lema 4. \square

A continuación, probaremos el siguiente resultado acerca de la propiedad de reducción del residual del algoritmo no lineal SOR.

Teorema 4 *Asumase que $\text{rank}(Z_\omega) = \text{rank}(Z), \forall \omega \in [1, \omega_1]$ para algun $\omega_1 \geq 1$. Sea $(X_+(\omega), Y_+(\omega), Z_+(\omega))$ denerado por el esquema iterativo *Nsor* comensando en un punto no estacionario (X, Y, Z) , y ρ_{12} , ρ_3 y ρ_4 definidos como en (2.45), (2.46) y (2.48) respectivamente. Entonces existe algun $\omega_2 \geq 1$ tal que*

$$\|S\|_F^2 - \|S_+(\omega)\|_F^2 = \rho_{12} + \rho_3 + \rho_4 > 0, \forall \omega \in [1, \omega_2] \quad (2.51)$$

donde $\rho_{12}(\omega), \rho_3(\omega) \geq 0$ por definici3n. Adem3s, siempre que $\rho_3(1) > 0$ (o equivalentemente $\mathcal{P}_{\Omega^c}(X_+(1)Y_+(1) - XY) \neq 0$) y $\omega_1 > 1$, entonces $\omega_2 > 1$ puede ser escogido tal que $\rho_4(\omega) > 0, \forall \omega \in (1, \omega_2]$.

Pr3ximamente, presentamos un resultado de convergencia del algoritmo *LMaFit*. Como el modelo (2.3) es no-convexo, solo seremos capaces de establecer la convergencia para un punto estacionario bajo una leve suposici3n. Notemos que la funci3n objetivo es acotada inferiormente por cero y se reduce en al menos en una cantidad de ρ_3 en cada iteraci3n. Tiene que ocurrir (vease (2.46)) que:

$$\mathcal{P}_{\Omega}^C(X^{k+1}Y^{k+1} - X^kY^k) \rightarrow 0.$$

Debido a lo mencionado arriba, es completamente razonable asumir que $\{\mathcal{P}_{\Omega}^C(X^kY^k)\}$ sigue siendo acotada, salvo la alternativa poco probable que $\|\mathcal{P}_{\Omega}^C(X^kY^k)\| \rightarrow \infty$.

Teorema 5 *Sea $\{(X^k, Y^k, Z^k)\}$ una sucesi3n generada por el algoritmo *LMaFit* y $\{\mathcal{P}_{\Omega^c}(X^kY^k)\}$ acotada. Entonces existe por lo menos una subsucesi3n de $\{(X^k, Y^k, Z^k)\}$ que satisface las condiciones de optimalidad de primer orden (2.5) en el limite.*

DEMOSTRACI3N:

Como consecuencia del algoritmo *LMaFit* y del hecho de que la sucesi3n $\{\mathcal{P}_{\Omega^c}(X^kY^k)\}$ es acotada se tiene que tanto la sucesi3n $\{Z^k\}$ como la sucesi3n $\{X^kY^k\}$ son acotadas. Es suficiente probar las dos primeras ecuaciones de las condiciones de optimalidad (2.5) ya que las otras dos ecuaciones son satisfechas por la propia construcci3n del algoritmo *LMaFit*. Sin perdida de generalidad, asumamos que la sucesi3n $\{X^k\}$ es generada por el esquema an3logo al *esquema alternativo 2*: Identifiquemos por $(X, Y) = (X^k, Y^k)$ y sea $\omega \in [1, \tilde{\omega}]$

$$Z_\omega = \omega Z + (1 - \omega)XY, \quad X_+ = \text{orth}(Z_\omega Y^\top), \quad Y_+ = X_+^\top Z_\omega,$$

Obviamente, la sucesi3n $\{X^k\}$ es acotada. Adem3s, la sucesi3n $\{Y^k\}$ tambi3n es acotada debido a que tanto $\{Z^k\}$ como $\{X^kY^k\}$ son sucesiones acotadas.

Sea $\mathcal{I} = \{k : \rho_4^k(\omega^k) \geq 0\}$, y \mathcal{I}^c el complemento del conjunto \mathcal{I} . Como consecuencia de (2.51) tenemos que,

$$\|S^0\|_F^2 \geq \sum_{i \in \mathcal{I}} \rho_{12}^i(\omega) = \sum_{i \in \mathcal{I}} (\|S^i P^i\|_F^2 + \|Q^i S^i (I - P^i)\|_F^2). \quad (2.52)$$

Ahora, consideraremos los siguientes tres casos.

Caso1 : supongamos que $|\mathcal{J}^C| < \infty$. Así, de (2.52) obtenemos que,

$$\lim_{i \rightarrow \infty} \|S^i P^i\|_F^2 = 0 \text{ y } \lim_{i \rightarrow \infty} \|Q^i S^i\|_F^2 = 0. \quad (2.53)$$

A partir de la construcción del esquema iterativo Nsor obtenemos las siguientes igualdades:

$$\mathcal{P}_\Omega(M) = \mathcal{P}_\Omega(Z^i), \quad \mathcal{P}_{\Omega^c}(Z^i) = \mathcal{P}_{\Omega^c}(X^i Y^i), \quad P^i = U^i (U^i)^\top,$$

donde $U^i = \text{orth}((Y^i)^\top)$. Por lo tanto,

$$S^i P^i = \mathcal{P}_\Omega(M - X^i Y^i) P^i = \mathcal{P}_\Omega(Z^i - X^i Y^i) P^i = (Z^i - X^i Y^i) P^i = (Z^i - X^i Y^i) U^i (U^i)^\top,$$

lo cual produce que $\lim_{i \rightarrow \infty} (Z^i - X^i Y^i) U^i = 0$ debido a la primera igualdad de (2.53). Ahora, como U^i es una base ortonormal para $\mathcal{R}((Y^i)^\top)$ y la sucesión $\{Y^i\}$ es acotada, conseguimos lo siguiente,

$$\lim_{i \rightarrow \infty} (Z^i - X^i Y^i) (Y^i)^\top = 0. \quad (2.54)$$

Utilizando $Q^i = V^i (V^i)^\top$, donde V^i es una base ortonormal para (X^{i+1}) , obtenemos

$$Q^i S^i = Q^i S^{i+1} + Q^i (S^i - S^{i+1}) = V^i (V^i)^\top (Z^{i+1} - X^{i+1} Y^{i+1}) + Q^i (S^i - S^{i+1}). \quad (2.55)$$

Usando (2.41) y (2.53) tenemos que,

$$\|S^i - S^{i+1}\|_F^2 \leq \|X^{i+1} Y^{i+1} - X^i Y^i\|_F^2 \leq (\tilde{\omega})^2 (\|S^i P^i\|_F^2 + \|Q^i S^i (I - P^i)\|_F^2) \rightarrow 0,$$

por lo tanto, $\lim_{i \rightarrow \infty} \|S^i - S^{i+1}\|_F = 0$. Este resultado, junto con (2.53) y (2.55), prueba que

$$(V^i)^\top (Z^{i+1} - X^{i+1} Y^{i+1}) \rightarrow 0.$$

En vista de que la sucesión $\{X^i\}$, es acotada, se concluye que

$$\lim_{i \rightarrow \infty} (X^i)^\top (X^i Y^i - Z^i) = 0. \quad (2.56)$$

Caso2 : Supongamos que $|\mathcal{J}^C| = \infty$ y que $|\{k \in \mathcal{J}^C : \gamma(\omega^k) \geq \omega_1\}| < \infty$. Así, para $k \in \mathcal{J}^C$ suficientemente grande tenemos que,

$$\|S^{k+1}\|_F < \gamma_1 \|S^k\|_F.$$

Consecuentemente, $\lim_{k \rightarrow \infty, k \in \mathcal{J}^C} \|S^k\|_F = 0$. Como $\|S^k\|$ es no-creciente, la sucesión completa converge a el minimizador objetivo de (2.3).

Caso3 : Supongamos que $|\mathcal{J}^C| = \infty$ y que $|\{i \in \mathcal{J}^C : \gamma(\omega^i) \geq \omega_1\}| = \infty$. Entonces el algoritmo *LMaFit* resetea $\omega^i = 1$ para un número infinito de iteraciones. De (2.52) se obtiene que,

$$\|S^0\|_F^2 \geq \sum_{i \in \mathcal{J}_1} \rho_{12}^i(\omega) = \sum_{i \in \mathcal{J}_1} (\|S^i P^i\|_F^2 + \|Q^i S^i (I - P^i)\|_F^2). \quad (2.57)$$

Por lo tanto, la sucesión en \mathcal{J}_1 satisface (2.54) y (2.56) mediante la repetición de los pasos análogos de la prueba del caso 1. \square

Capítulo 3

Métodos de Lagrangeano aumentado en la factorización de bajo rango

En este capítulo desarrollaremos nuestra propuesta, la cual esta fundamentada en la aplicación del método del lagrangeano aumentado (ALM) al problema (2.3), con ella buscamos evitar el método de sobrerrelajación sucesiva (SOR) para la resolución del sistema de ecuaciones que se genera a partir de las condiciones de optimalidad. Esperamos un mejor desempeño con esta nueva propuesta debido a que en dicha propuesta se emplea la dualidad.

3.0.1. El método de Gauss-Seidel no lineal

En la presente sección, estableceremos el método de Gauss-Seidel no lineal similarmente al esquema básico 1, para resolver el sistema de ecuaciones que se obtiene de las condiciones de optimalidad que resultan de la minimización de la función lagrangeana aumentada asociada al problema (2.3) en cada iteración. Consideremos la función lagrangeana aumentada para el problema (2.3), la cual viene dada por:

$$L(X, Y, Z, \Lambda, \mu) = \frac{1}{2} \|XY - Z\|_F^2 + \langle \Lambda, \mathcal{P}_\Omega(Z - M) \rangle + \frac{\mu}{2} \|\mathcal{P}_\Omega(Z - M)\|_F^2 \quad (3.1)$$

Se puede establecer las condiciones de optimalidad de primer orden para la minimización de la función Lagrangeana Aumentada por medio del siguiente sistema:

$$(XY - Z)Y^\top = 0 \quad (3.2)$$

$$X^\top(XY - Z) = 0 \quad (3.3)$$

$$\mathcal{P}_\Omega(Z - XY) + \mathcal{P}_\Omega(\Lambda) + \mu \mathcal{P}_\Omega(Z - M) = 0 \quad (3.4)$$

$$\mathcal{P}_{\Omega^c}(Z - XY) = 0 \quad (3.5)$$

$$\mathcal{P}_\Omega(Z - M) = 0 \quad (3.6)$$

Primero estableceremos el ALM para el problema (2.3).

ALM para (2.3)

ENTRADA: M

1. $\Lambda_0^* = 0; \mu_0 = 1; \rho > 1; k = 0.$
2. Mientras no converge, hacer
3. resolver $(X_{k+1}^*, Y_{k+1}^*, Z_{k+1}^*) := \operatorname{argmin}_{X,Y,Z} L_{\mu_k}(X, Y, Z, \Lambda_k^*);$
4. $\Lambda_{k+1}^* := \Lambda_k^* + \mu_k \mathcal{P}_\Omega(Z_{k+1}^* - M);$
5. $\mu_{k+1} := \rho \mu_k; k := k + 1;$
6. fin (Mientras)

SALIDA: x_k, λ_k

Notemos que en cada iteración el algoritmo del Lagrangeano Aumentado para problema (2.3) fija las variables Λ y μ y luego calcula el argumento que minimiza $L_\mu(X, Y, Z, \Lambda)$ para las variables X, Y y Z , después actualiza las variables Λ y μ y sigue haciendo el mismo procedimiento hasta que converja. Por ejemplo, para valores fijos $\bar{\Lambda}$ y $\bar{\mu}$ es necesario obtener la solución para $\min_{X,Y,Z} L(X, Y, Z, \bar{\Lambda}, \bar{\mu})$. Esta minimización se puede acompañar con un esquema alternante sobre las condiciones de optimalidad. Por ejemplo, para la instancia \bar{Y} y \bar{Z} fijos, obtenemos X_+ resolviendo el problema $\min_X L(X, \bar{Y}, \bar{Z}, \bar{\Lambda}, \bar{\mu})$. Luego fijando X_+ y \bar{Z} obtenemos $Y_+ = \operatorname{argmin}_Y L(X_+, Y, \bar{Z}, \bar{\Lambda}, \bar{\mu})$, y finalmente fijando X_+ y Y_+ obtenemos $Z_+ = \operatorname{argmin}_Z L(X_+, Y_+, Z, \bar{\Lambda}, \bar{\mu})$.

Usando las condiciones de optimalidad adecuadamente para cada problema, obtenemos el siguiente esquema iterativo (Método Gauss-Seidel no-lineal para nuestro problema):

Esquema ALM-GS

$$X_+ := ZY^\dagger \equiv ZY^\top (YY^\top)^\dagger, \quad (3.7)$$

$$Y_+ := (X_+)^\dagger Z \equiv (X_+^\top X_+)^\dagger (X_+^\top Z), \quad (3.8)$$

$$Z_+ := X_+ Y_+ + \frac{\mu}{1 + \mu} \mathcal{P}_\Omega(M - X_+ Y_+ - \frac{1}{\mu} \Lambda). \quad (3.9)$$

3.0.2. Esquemas Alternativos

En esta sección presentaremos dos esquemas equivalentes al esquema ALM-GS, como alternativas de dicho esquema, que se obtienen de manera similar a como hicimos en el capítulo 2, que pueden ser más eficientes a la hora de implementarlos en la práctica. Primero notemos que el **lema 1** se sigue satisfaciendo para el esquema ALM-GS, ya que en tal esquema, los iterados X_+ y Y_+ son iguales a los presentados en el capítulo 2, así que de manera similar como se hizo en el capítulo 2, podemos establecer los dos siguientes esquemas alternativos:

Esquema alternativo 1

$$X_+ := ZY^\top, \quad (3.10)$$

$$Y_+ := (X_+)^\dagger Z \equiv (X_+^\top X_+)^\dagger (X_+^\top Z), \quad (3.11)$$

$$Z_+ := X_+ Y_+ + \frac{\mu}{1+\mu} \mathcal{P}_\Omega(M - X_+ Y_+ - \frac{1}{\mu} \Lambda). \quad (3.12)$$

La segunda variante del esquema ALM-GS, calcula la proyección ortogonal $\mathcal{P}_{ZY^\top} = VV^\top$, donde $V := \text{orth}(ZY^\top)$ es una base ortonormal del subespacio $\mathcal{R}(ZY^\top)$. Por tanto podemos escribir $X_+ Y_+ = VV^\top Z$, así obtenemos el siguiente esquema:

Esquema alternativo 2

$$X_+ := V, \quad (3.13)$$

$$Y_+ := V^\top Z, \quad (3.14)$$

$$Z_+ := X_+ Y_+ + \frac{\mu}{1+\mu} \mathcal{P}_\Omega(M - X_+ Y_+ - \frac{1}{\mu} \Lambda). \quad (3.15)$$

Este último esquema, con frecuencia es preferido debido a que el cálculo del paso (3.14) mediante la factorización QR es generalmente más estable que resolver las ecuaciones normales. Usando el esquema ALM-GS construimos el siguiente algoritmo:

Alternating ALM for LRMC (Lagrangean12)

ENTRADA: M

1. $\Lambda_0 = 0; Y_0 = I; Z_0 = \mathcal{P}_\Omega(M); \mu_0 = 1; \rho > 1; k = 0.$
2. Mientras no converge, hacer
3. **líneas 4-9 resolver $(X_{k+1}, Y_{k+1}, Z_{k+1}) := \text{argmin}_{X,Y,Z} L_{\mu_k}(X, Y, Z, \Lambda_k);$
4. Mientras no converge, hacer
5. $X_k^{l+1} := Z_k^l (Y_k^l)^\top (Y_k^l (Y_k^l)^\top)^\dagger;$
6. $Y_k^{l+1} := ((X_k^{l+1})^\top X_k^{l+1})^\dagger (X_k^{l+1})^\top Z_k^l;$
7. $Z_k^{l+1} := X_k^{l+1} Y_k^{l+1} + \frac{\mu_k}{1+\mu_k} \mathcal{P}_\Omega(M - X_k^{l+1} Y_k^{l+1} - \frac{1}{\mu_k} \Lambda_k);$
8. $l := l + 1;$
9. Fin (Mientras).
10. $(X_{k+1}, Y_{k+1}, Z_{k+1}) := (X_{k+1}^l, Y_{k+1}^l, Z_{k+1}^l);$
11. $\Lambda_{k+1} := \Lambda_k + \mu_k \mathcal{P}_\Omega(Z_{k+1} - M);$
12. $\mu_{k+1} := \rho \mu_k$

13. $k := k + 1$;

14. Fin (Mientras).

SALIDA: $(X_k, Y_k, Z_k), \Lambda_k$.

Fijémonos que en el algoritmo anterior, la minimización de la función lagrangeana aumentada se resuelve entre las líneas 4 y 9 por medio del método del Gauss-Seidel no lineal, este método puede ser cambiado por cualquiera de los dos esquemas alternativos que mencionamos anteriormente o incluso, se pueden utilizar de forma intercambiada al igual que en el algoritmo de LMaFit que presentamos en el Capítulo 2.

3.1. Un algoritmo híbrido ALM-SOR para el problema MC

De acuerdo a los experimentos numéricos que realizamos a lo largo de esta investigación (algunos son presentados en el capítulo 3), nos dimos cuenta que el algoritmo **Lagrangean12** se desempeña bastante bien en lo que respecta a la calidad de la solución, inclusive en algunos casos es más rápido que el algoritmo LMaFit, pero no es suficientemente rápido en cuanto a las iteraciones y al tiempo que realiza para conseguir tal solución en el caso de problemas de gran tamaño, en vista de este hecho, la siguiente idea (la cual no era la idea inicial) es implementar un nuevo algoritmo el cual será una especie de híbrido que utilizará el método del lagrangeano aumentado (ALM), y en lugar del esquema alternante de Gauss-Seidel, utilizará de manera similar al paper [1] el método de sobrerrelajación sucesiva (SOR).

Concorde a lo anteriormente dicho, lo que buscamos es resolver en cada iteración la minimización de la función lagrangeana aumentada, resolviendo el sistema de ecuaciones que obtenemos a partir de las condiciones de optimalidad, por medio del método de sobrerrelajación sucesiva, es decir:

ALM-SOR teórico

$$X_+ := ZY^\dagger \equiv ZY^\top (YY^\top)^\dagger, \quad (3.16)$$

$$X_+(\omega) := \omega X_+ + (1 - \omega)X, \quad (3.17)$$

$$Y_+ := (X_+(\omega))^\dagger Z \equiv (X_+(\omega)^\top X_+(\omega))^\dagger (X_+(\omega)^\top Z), \quad (3.18)$$

$$Y_+(\omega) := \omega Y_+ + (1 - \omega)Y, \quad (3.19)$$

$$Z_+ := X_+(\omega)Y_+(\omega) + \frac{\mu}{1 + \mu} \mathcal{P}_\Omega(M - X_+(\omega)Y_+(\omega) - \frac{1}{\mu}\Lambda). \quad (3.20)$$

donde $\omega \geq 1$. Obviamente si $\omega = 1$ recuperamos el Método Gauss-Seidel no-lineal para nuestro problema. De manera similar a lo expuesto en el capítulo 2, si denotamos al residual por:

$$S := P_\Omega(M - XY) \quad (3.21)$$

el cual puede ser usado para medir la optimalidad. Notemos que luego de cada iteración la variable Z puede ser expresada como:

$$Z = XY + \frac{1}{1 + \mu} (\mu S - \mathcal{P}_\Omega(\Lambda)) \quad (3.22)$$

Definamos a nuestro Z_ω por:

$$Z_\omega := XY + \omega \left(\frac{\mu}{1+\mu} \right) \mathcal{P}_\Omega(M - XY - \frac{1}{\mu} \Lambda)$$

ahora bien, por la definición de S obtenemos que:

$$\begin{aligned} Z_\omega &= XY + \omega \frac{\mu}{1+\mu} S - \frac{\omega}{1+\mu} \mathcal{P}_\Omega(\Lambda) \\ &= XY + \omega Z - \omega XY + \frac{\omega}{1+\mu} \mathcal{P}_\Omega(\Lambda) - \frac{\omega}{1+\mu} \mathcal{P}_\Omega(\Lambda) \\ &= \omega Z + (1 - \omega)XY \end{aligned}$$

fijémonos que en la última igualdad obtuvimos,

$$Z_\omega = \omega Z + (1 + \omega)XY$$

al igual que en el capítulo 2, así que los cálculos que se realizaron en tal capítulo también se satisfacen para nuestro caso, lo cual nos permite recuperar el esquema **ALM-SOR teórico** mediante el siguiente esquema:

ALM-SOR modificado

$$X_+(\omega) := Z_\omega Y^\dagger \equiv Z_\omega Y^\top (Y Y^\top)^\dagger, \quad (3.23)$$

$$Y_+(\omega) := (X_+(\omega))^\dagger Z_\omega \equiv (X_+(\omega))^\top X_+(\omega)^\dagger (X_+(\omega))^\top Z_\omega, \quad (3.24)$$

$$Z_+(\omega) := X_+(\omega) Y_+(\omega) + \frac{\mu}{1+\mu} \mathcal{P}_\Omega(M - X_+(\omega) Y_+(\omega) - \frac{1}{\mu} \Lambda). \quad (3.25)$$

Usando este último esquema de aproximación y estimando el parámetro ω de forma análoga a como fué mostrado en el capítulo 2, construimos el siguiente algoritmo (**Lagrangeansor**)

Lagrangeansor

ENTRADA: M

1. $\Lambda_0 = 0; Y_0 = I; Z_0 = \mathcal{P}_\Omega(M); \mu_0 = 1; \rho > 1; k = 0; \omega = 1; \tilde{\omega} > 1; \delta > 0; \gamma_1 \in (0, 1)$.
2. Mientras no converge, hacer
3. **líneas 4-12 resolver $(X_{k+1}, Y_{k+1}, Z_{k+1}) := \operatorname{argmin}_{X, Y, Z} L_{\mu_k}(X, Y, Z, \Lambda_k)$;
4. Mientras no converge, hacer
5. $X_k^{l+1}(\omega) := Z_\omega^l (Y_k^l)^\top (Y_k^l (Y_k^l)^\top)^\dagger$;
6. $Y_k^{l+1}(\omega) := ((X_k^{l+1}(\omega))^\top X_k^{l+1}(\omega))^\dagger (X_k^{l+1}(\omega))^\top Z_\omega^l$;
7. $Z_k^{l+1}(\omega) := X_k^{l+1}(\omega) Y_k^{l+1}(\omega) + \frac{\mu_k}{1+\mu_k} \mathcal{P}_\Omega(M - X_k^{l+1}(\omega) Y_k^{l+1}(\omega) - \frac{1}{\mu_k} \Lambda_k)$;
8. calcular $\gamma(\omega)$ de acuerdo a (2.27).

9. si $\gamma(\omega) \geq 1$ entonces hacemos $\omega = 1$ y volvemos al paso 5
 10. $l := l + 1$;
 11. si $\gamma(\omega) \geq \gamma_1$ entonces hacemos $\delta = \max(\delta, 0,25(\omega - 1))$ y $\omega = \min(\omega + \delta, \tilde{\omega})$
 12. Fin (Mientras).
 13. $(X_{k+1}, Y_{k+1}, Z_{k+1}) := (X_{k+1}^l, Y_{k+1}^l, Z_{k+1}^l)$;
 14. $\Lambda_{k+1} := \Lambda_k + \mu_k \mathcal{P}_\Omega(Z_{k+1} - M)$;
 15. $\mu_{k+1} := \rho \mu_k$
 16. $k := k + 1$;
 17. Fin (Mientras).
- SALIDA:* $(X_k, Y_k, Z_k), \Lambda_k$.

Notemos que en el algoritmo anterior, la minimización de la función lagrangeana aumentada se resuelve entre las líneas 4 y 12 por medio del método SOR no lineal, este método puede ser sustituido por cualquiera de los dos esquemas alternativos análogos a los que mencionamos anteriormente o incluso, se pueden utilizar de forma intercambiada al igual que en el algoritmo de LMafit que se mostró en el Capítulo 2.

Capítulo 4

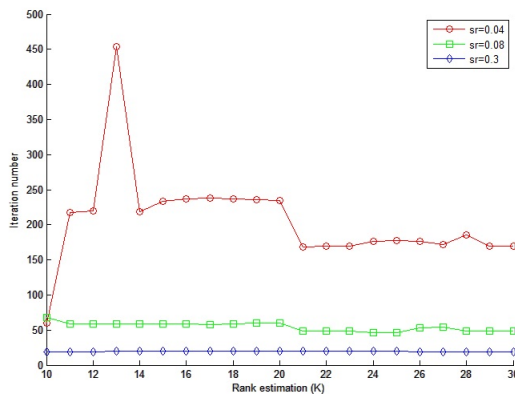
Experimentación Numérica

4.1. Experimentos sobre matrices generadas aleatoriamente

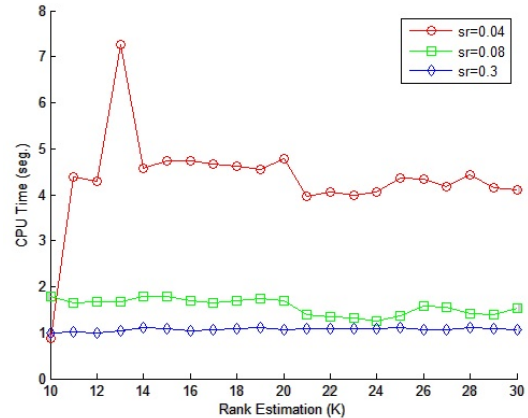
En esta sección crearemos matrices $M \in \mathbb{R}^{m \times n}$ de rango r aleatoriamente, mediante el siguiente procedimiento, primero, generaremos dos matrices aleatorias $M_L \in \mathbb{R}^{m \times r}$ y $M_R \in \mathbb{R}^{n \times r}$ con entradas Gaussianas estandar y entonces se ensambla la matriz $M = M_L M_R^T$, luego construimos el subconjunto Ω de p entradas escogidas aleatoriamente con distribución uniforme de la matriz M . La tasa $\frac{p}{(mn)}$ entre el número de medidas y el número total de entradas de la matriz es denotado por "SR" (Sampling ratio). La tasa $\frac{r(m+n-r)}{p}$ entre el grado de libertad en una matriz de rango r y el número de muestras o medidas es denotado por "FR".

Primero, examinaremos la sensibilidad de nuestro algoritmo *Lagrangeansor* para el rango inicial K , utilizando la estrategia del rango decreciente para la estimación del rango. En esta prueba, utilizamos matrices con $m = n = 1000$ y de rango $r = 10$. Generamos tres pruebas para tres valores diferentes de SR , es decir, para los valores $SR = 0,04$, $SR = 0,08$ y para $SR = 0,3$ respectivamente. En cada caso, corrimos nuestro *Lagrangeansor* para cada $K = 10, 11, 12, \dots, 30$; en las siguientes gráficas, mostraremos el número de iteraciones y el CPU time para cada uno de dichos valores de K . Cabe destacar que tomamos los siguientes parámetros $\rho = 2$, $\mu = 10$ y $reschgxy < 1 * tol$.

En la gráfica (a) podemos observar que el número de iteraciones se mantiene casi al mismo nivel y el CPU time también parece mantenerse aproximadamente estable a medida que se incrementa el valor de K .



(a) Número de iteraciones



(b) Número de iteraciones

4.1.1. Experimentos con datos simulados

Ahora consideraremos aplicar nuestros algoritmos y el **LMaFit** a varios tipos de problemas, generados aleatoriamente en MATLAB. La información de los distintos experimentos fué recolectada en varias tablas (vease anexos).

En el primer experimento, nuestro proposito fué el de comparar los algoritmos *Lagrangean12* y *Lagrangeansor*, es decir comparar el desempeño del método del Lagrangeano aumentado con el esquema GS en contraste con el método del lagrangeano aumentado junto con el esquema SOR. En dicho experimento generamos aleatoreamente matrices cuadradas de diversos tamaños (tamaño m), además, para ambos solvers resolvimos los problemas para un rango inicial igual a $k = [1,5r]$. Los valores del sampling ratio (sr) se pueden visualizar en la *TABLA 1*, además, para ambos algoritmos utilizamos los siguientes parámetros: $miu = 10$, $rho = 2,5$ y $reschgxy < 10tol$, también, aclararemos que los últimos 4 problemas son esparcos y el resto son problemas completos ($Zfull = 1$).

En la *TABLA 1* recogimos los siguientes datos: Rgest (representa el rango de la matriz que genera cada software), Tiempo (representa el tiempo de corrida), Err.rel (representa el error relativo) y Iter (representa el numero de iteraciones). La estadística contenida en dicha tabla verifica claramente que el algoritmo *Lagrangeansor* es mucho mas rápido y hace menos iteraciones que el algoritmo *Lagrangean12* sobre todo cuando la matriz es de mayor orden. También notamos que el algoritmo *Lagrangeansor* es un poco más preciso en la calidad de la solución que el algoritmo *Lagrangean12* en la mayoría de los casos.

Próximamente, compararemos el desempeño del algoritmo *Lagrangeansor* frente al algoritmo *LMaFit* en diversos tipos de problemas: en problemas muy pequeños donde el orden de la matriz es $m = n = 10$, la información de estos experimentos esta contenida en la *TABLA 2*, también se estudia la comparación de dichos algoritmos para problemas de data pequeña ($m = n = 100$) y mediana ($m = n = 500$), cuya información esta contenida en la *TABLA 3* y la *TABLA 4* respectivamente.

En todos estos experimentos, generamos aleatoreamente los diversos problemas y escogimos el sampling ratio (sr) en el siguiente conjunto $\{0,04, 0,8, 0,15, 0,20, 0,25, 0,30, 0,35, 0,40, 0,50, 0,55\}$, además, tanto para el **LMaFit** como para el **Lagrangeansor**, resolvimos los problemas para valores del rango inicial de $k = [1,25r]$ y $k = [1,5r]$. Notemos también que para la *TABLA 2* el rango de la matriz generada fué de $r = 2$; en los problemas de la *TABLA 3* el rango fué de $r = 2$, $r = 10$ y $r = 20$, y para la *TABLA 4* el rango tomó los valores de $r = 10, 50, 100$.

En todas estas tablas recogimos los siguientes datos: Rgest (representa el rango de la matriz que genera cada software), Tiempo (representa el tiempo de corrida), Err.rel (representa el error relativo) y Iter (representa el numero de iteraciones).

Por último, comparamos el desempeño de los tres algoritmos en problemas generados aleatoriamente y de diversos tamaños. La información de este último experimento se puede observar en la *TABLA 5*.

4.2. Experimentos de data real

4.2.1. Jester jokes and Movie Lens.

En esta subsección, consideraremos problemas de aproximación de matrices de rango bajo basados en dos conjuntos de "datos reales": el conjunto de datos *Jester joke* [23] y el conjunto de datos *MovieLens* [22]. En ambos conjuntos de datos, solo cierta cantidad de datos esta disponible de las matrices subyacentes desconocidas las cuales son poco probables que sean exactamente de rango bajo. Sin embargo, los "solvers" de completación de matrices se han aplicado a tales problemas para probar su habilidad de producir aproximaciones de rango bajo.

Ambos problemas tanto el *Jester Joke* como el *MovieLens* son ejemplos de problemas del área de sistemas de recomendación, el primero (*Jester Jokes*) trata acerca de una cierta lista de "chistes" una cierta cantidad de personas o usuarios que deben asignarle un valor o calificación a cada "chiste", pero como la cantidad de chistes es muy grande los usuarios solo califican una cierta cantidad de ellos, de esta forma se puede construir una matriz de datos donde las filas representan a los usuarios y las columnas representan a los chistes, y donde la entrada (i,j) representa la calificación que el usuario i le asigna a el chiste j , pero dicha matriz esta incompleta, así que la idea es obtener una aproximación de bajo rango para la matriz "original", y de esta forma "predecir" la calificación que le asignaría cada usuario a cada chiste. El segundo problema (*MovieLens*) trata acerca de un conjunto de películas de diversos temas (ciencia ficción, fantasía, acción, terror, entre otros) y de una cierta cantidad de usuarios que deben calificar las películas, pero como nuevamente la cantidad de películas es muy grande, los usuarios solo califican algunas de ellas, obtenemos así un problema similar al *Jester Jokes*.

El conjunto de datos *Jester Joke* consiste de cuatro problemas "jester-1", "jester-2", "jester-3z" "jester-all" donde el último es obtenido mediante la combinación de los tres primeros conjuntos de datos; el conjunto de datos *MovieLens* esta conformado por tres problemas: "movie-100k", "movie-1Mz" "movie-10M". Estos estan disponibles en www.ieor.berkeley.edu/~Egoldberg/jester-data y en www.grouplens.org respectivamente. Para **LMaFit**, **Lagrangean12** y para **Lagrangeansor** nuestro conjunto de parámetros fue $tol = 10^{-3}$, $est_rank = 2$, $K = 1$ y $rk_ink = 2$. Además, escogimos como máxima estimación permitida para el rango de 80 para los problemas del *Jester jokes* y 100 para los problemas del *MovieLens* para todos los "solvers". Esto se debe a que como los problemas *Jester joke* solo tienen 100 columnas, no es significativo ajustar una matriz de rango 100 al conjunto de datos *Jester*. Como las entradas de la matriz subyacente M son disponibles o conocidas solo en el conjunto Ω , para medir precisión calculamos el error absoluto medio normalizado (NMAE) mediante la fórmula:

$$NMAE = \frac{1}{(r_{max} - r_{min})|\Omega|} \sum_{(i,j) \in \Omega} |W_{ij} - M_{ij}|$$

donde r_{min} y r_{max} son las cotas inferior y superior para las calificaciones. Específicamente, $r_{min} = -10$, $r_{max} = 10$ para los problemas del *Jester Joke* y $r_{min} = 1$, $r_{max} = 5$ para los problemas *MovieLens*. El reporte de los resultados numéricos de este experimento esta contenido en la *TABLA 6* ubicada en los anexos.

4.2.2. Completación de Matrices aplicado a imágenes

En esta subsección aplicaremos tanto el algoritmo **LMaFit** como también nuestros algoritmos **Lagrangean12** y **Lagrangeansor** para mejorar la calidad de imágenes tanto a color como en escala de grises, mediante la eliminación de ruidos impulsivos. Aquí la tarea es rellenar los valores que deben ir en los pixeles perdidos de una imagen para lograr una buena aproximación a dicha imagen. Este proceso es conocido como “inpainting”, específicamente cuando las posiciones de los pixeles perdidos no están distribuidos aleatoriamente. En su forma original, estos problemas no son verdaderos problemas de completación de matrices, pero los solvers de completación de matrices pueden ser aplicados para obtener aproximaciones de bajo rango.

En nuestro primer experimento, la imagen (800×533) (c) representa la imagen original de la bandera de Venezuela, la imagen (d) fue construida a partir de la imagen original (c) extrayéndole un 50% de los pixeles de forma aleatoria, la imagen (e) muestra la aproximación obtenida del algoritmo **LMaFit** a partir de la imagen (d), la imagen (f) representa la aproximación obtenida del algoritmo **Lagrangean12** a partir de la imagen (d) y por último la imagen (g) representa la aproximación obtenida del algoritmo **Lagrangeansor** a partir de la imagen (d). Para la realización de tal experimento escogimos los siguientes parámetros: $tol = 10^{-3}$, $k = 20$, $est_rank = 2$ y $rank_max = 90$ para los tres algoritmos. Además, para nuestros algoritmos se tomaron valores de $miu = 10$ y $rho = 3$. Un resumen de los resultados computacionales se muestra la siguiente tabla:

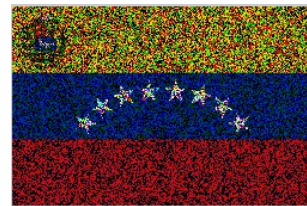
Ouput	LMaFit			Lagrangean12			Lagrangeansor		
	C1	C2	C3	C1	C2	C3	C1	C2	C3
Time	9.956	6.516	8.621	32.281	31.974	32.417	7.588	14.586	7.588
Rel Err.	0.0036	0.0084	0.0071	0.0034	0.0041	0.0028	0.0101	0.0047	0.0054
Iter	291	200	273	500	500	500	131	231	132

A partir de los datos mostrados en la tabla anterior, podemos decir que los algoritmos **LMaFit** y **Lagrangeansor** necesitan básicamente la misma cantidad de tiempo para resolver cada problema, mientras que el algoritmo **Lagrangean12** necesita mucho más tiempo que los anteriores, además, el algoritmo **Lagrangeansor** realiza menos iteraciones en casi todos los problemas, mientras que el algoritmo **Lagrangean12** necesita el máximo de las iteraciones en cada problema. En cuanto, al error relativo, podemos apreciar que los tres algoritmos se comportan bastante parecido.

A continuación mostramos las imágenes obtenidas en dicho experimento.



(c) Imagen original



(d) Imagen corrupta al 50%



(e) Aprox. de LMaFit



(f) Aprox. de Lagrangean12



(g) Aprox. de Lagrangeansor

En la tabla anterior y en la tabla siguiente, las columnas C1, C2 y C3 corresponden a tres problemas que se generan al intentar recuperar una imagen a color, esto se debe a que MATLAB representa las imágenes a color por medio de un arreglo tridimensional, es decir, un arreglo formado por tres matrices que unidas crean la imagen a color. Así, que en estos experimentos (el primero y el segundo) C_i con $i = 1, 2, 3$ es un problema construido a partir de una de esas tres matrices del arreglo tridimensional extrayéndole el 50% de sus entradas de forma aleatoria. Luego de obtener las aproximaciones para cada problema generado (C1,C2,C3), se puede construir la imagen a color aproximada a la original por cada algoritmo, mediante la “unión”(formando un arreglo tridimensional) de las tres soluciones aproximadas de las matrices que se generan en cada uno de estos problemas.

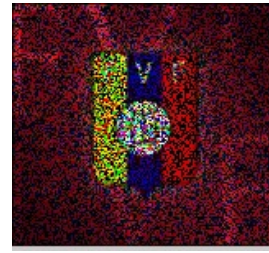
En el segundo experimento, la imagen (h) representa la imagen original de la FVF de escala 160×148 , la imagen (i) fué construida a partir de la imagen original (h) extrayéndole un 50% de los pixeles de forma aleatoria, la imagen (j) muestra la aproximación obtenida del algoritmo **LMaFit** a partir de la imangen (i), la imagen (k) representa la aproximación obtenida del algoritmo **Lagrangean12** a partir de la imangen (i) y por último, la imagen (l) representa la aproximación del algoritmo **Lagrangeansor** a partir de la imagen (i). Para la realización de tal experimento escogimos los siguientes parámetros: $tol = 10^{-3}$, $k = 20$, $est_rank = 2$ y $rank_max = 90$ para los tres algoritmos. Además, para nuestros algoritmo, tomamos valores de $miu = 10$ y $rho = 3$. Un resumen de los resultados computacionales se muestra la siguiente tabla:

Ouput	LMaFit			Lagrangean12			Lagrangeansor		
	C1	C2	C3	C1	C2	C3	C1	C2	C3
Time	0.297	0.247	0.253	0.544	0.482	0.461	0.410	0.342	0.433
Rel Err.	8.75e-4	9.30e-4	9.18e-4	9.62e-4	9.75e-4	9.66e-4	8.98e-4	9.98e-4	9.17e-4
Iter	152	122	128	175	161	160	130	118	141

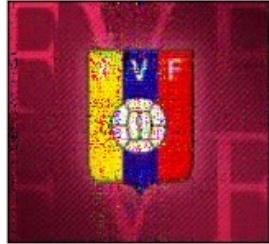
De acuerdo con los datos mostrados en la tabla de arriba, podemos notar que el algoritmo **LMaFit** es más rápido que los otros dos algoritmos, además, se puede observar que el algoritmo **Lagrangeansor** hace menos iteraciones que los otros dos algoritmos, y en lo que se refiere a la calidad de la solución podemos decir que los tres algoritmos se desempeñan bastante bien, prácticamente, los tres algoritmos obtienen la misma calidad de solución. Además, de estos datos también presentamos las siguientes imágenes.



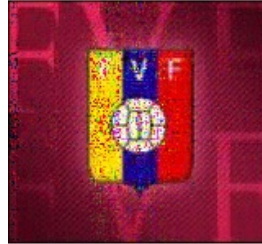
(h) Imagen original



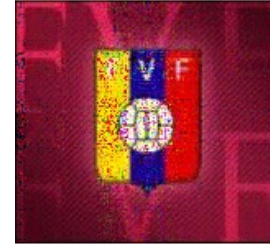
(i) Imagen corrupta al 50%



(j) Aprox. de LMaFit



(k) Aprox. de Lagrangean12



(l) Aprox. de Lagrangeansor

Claramente, las imágenes obtenidas por los tres algoritmos, no representan una buena calidad de imagen, esto se debe a que la imagen original no es de rango bajo, sin embargo, representan una aproximación modesta, considerando que la imagen (i) es de muy baja calidad.

En el tercer experimento, la imagen (m) representa la imagen original de la ajedrez de escala 266×192 , la imagen (n) fué construida a partir de la imagen original (m) truncandola hasta su rango 40, la imagen (\tilde{n}) fué construida a partir de la imagen original (m) extrayéndole un 50% de los pixeles de forma aleatoria, la imagen (o) muestra la aproximación obtenida del algoritmo **LMaFit** a partir de la imagen (\tilde{n}), la imagen (p) representa la aproximación obtenida del algoritmo **Lagrangean12** a partir de la imagen (\tilde{n}), la imagen (q) representa la aproximación obtenida por el algoritmo **Lagrangeansor** a partir de la imagen (\tilde{n}), la imagen (r) fué construida a partir de la imagen (n) extrayendole un 50% de los pixeles de forma aleatoria, la imagen (s) representa la aproximación obtenida del algoritmo **LMaFit** a partir de la imagen (r), la imagen (t) representa la aproximación obtenida del algoritmo **Lagrangean12** a partir de la imagen (r), y por último, la imagen (u) representa la aproximación obtenida del algoritmo **Lagrangeansor** a partir de la imagen (r).

Para la realización de tal experimento escogimos los siguientes parámetros: $tol = 10^{-3}$, $k = 20$, $est_rank = 2$ y $rank_max = 80$ para los tres algoritmos. Además, para nuestros algoritmos tomamos valores de $miu = 10$ y $rho = 3$. En la siguiente tabla mostramos los resultados obtenidos por los algoritmos, para recuperar la matriz que representa la imagen (m) a partir de la matriz que representa a la imagen (\tilde{n}).

Ouput	LMaFit	Lagrangean12	Lagrangeansor
Time	0.866	1.341	0.950
Rel Err.	9.913e-4	9.858e-4	9.792e-4
Iter	183	197	150

En la siguiente tabla mostramos los resultados obtenidos por los algoritmos, para recuperar la matriz que representa la imagen (\mathbf{n}) a partir de la matriz que representa a la imagen (\mathbf{r}).

Ouput	LMaFit	Lagrangean12	Lagrangeansor
Time	0.509	1.043	0.639
Rel Err.	9.99e-4	9.84e-4	9.92e-4
Iter	128	184	109

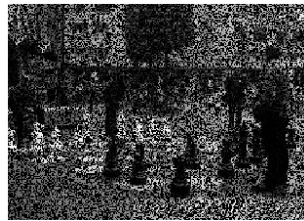
En la estadística contenida en estas dos tablas podemos observar que a los algoritmos **LMaFit** y **Lagrangeansor** les toma un poco menos de tiempo resolver los problemas, y que entre los tres algoritmo, es el algoritmo **LMaFit**, el que los resuelve más rápidamente, aunque solo por una ligera diferencia. En cuanto al número de iteraciones podemos notar que el algoritmo **Lagrangeansor** requiere de menos iteraciones para alcanzar una buena aproximación de la solución en ambos problema, mientras que el algoritmo **Lagrangean12** es al que le toma más iteraciones. También observamos que los tres algoritmos son bastante buenos en lo que respecta a la calidad de la solución. A continuación, mostramos las imágenes obtenidas en dicho experimento.



(m) Imagen original



(n) Rango 40 de la imagen original



(ñ) Imagen corrupta al 50% de la imagen original



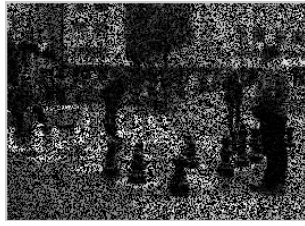
(o) Aprox. de LMaFit



(p) Aprox. de Lagrangean12



(q) Aprox. de Lagrangeansor



(r) Imagen corrupta al 50% de la imagen a rango 40



(s) Aprox. de LMaFit



(t) Aprox. de Lagrangean12



(u) Aprox. de Lagrangean-sor

En las imágenes anteriores, podemos apreciar que las aproximaciones generadas por cada algoritmos a partir de la imagen (\tilde{n}), es decir, las imágenes (o), (p) y (q) no son muy buenas aproximaciones de la imagen original (m) en lo que respecta a calidad de imagen, esto es debido a que la matriz original no es una matriz de rango bajo; pero podemos notar que las imágenes (s), (t) y (u) son muy buenas aproximaciones de la imagen (n), este resultado era de esperarse ya que la imagen (n) tiene un rango bajo (su rango es igual a 40).

Anexos

Problema		Lagrangean12				Lagrangeansor				
m	r	sr	Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter
10	2	0.15	2	0.0241	0.1438	57	2	0.0285	0.1438	79
10	2	0.25	2	0.0705	6.0471e-4	500	2	0.0918	1.1849e-4	323
10	2	0.35	2	0.0748	0.0031	500	3	0.0414	1.1687e-4	41
10	2	0.55	1	0.0253	6.2132e-5	110	1	0.0140	0.3553	111
100	10	0.15	13	0.2839	1.2494e-4	363	4	0.0887	0.3739	104
100	10	0.25	1	0.0629	0.8540	87	1	0.0572	0.8533	63
100	10	0.35	10	0.2711	1.2312e-4	307	10	0.1503	1.2201e-4	153
100	10	0.55	10	0.1456	1.2326e-4	78	10	0.0316	9.9925e-5	28
1000	30	0.15	30	12.243	1.2212e-4	205	30	3.0601	1.1327e-4	52
1000	30	0.25	30	4.4785	1.2392e-4	65	30	1.5223	9.4369e-5	23
1000	30	0.35	30	2.8779	1.0765e-4	36	30	1.4004	6.2161e-5	18
1000	30	0.55	30	1.7515	8.5597e-5	17	30	1.3276	6.8091e-5	12
1000	15	0.04	12	11.629	0.2414	500	15	13.750	4.5372e-4	500
1000	15	0.1	15	8.3452	1.2467e-4	237	15	1.4792	1.0250e-4	42
2000	15	0.1	15	34.805	1.2193e-4	188	15	6.9849	9.8830e-5	38
2000	15	0.15	15	23.293	1.2193e-4	102	15	4.0600	1.2460e-4	21

TABLA 1

Problema		LMaFit						Lagrangeansor										
sr	miu	$K = [1,25r]$			$K = [1,5r]$			$K = [1,25r]$			$K = [1,5r]$							
		Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter	
0.04	5	1.01	2	0.002	0.9932	6	3	0.015	0.8946	6	2	0.0137	0.9932	6	3	0.0016	0.8946	6
0.08	100	2	1	0.003	0.8315	12	2	0.008	0.2779	21	1	0.0133	0.8315	12	2	0.0135	0.2779	16
0.15	100	2	1	0.004	0.7066	23	3	0.007	0.5923	24	1	0.0130	0.7070	15	3	0.0148	0.5923	20
0.30	5	1.5	1	0.004	0.2291	47	2	0.045	0.0466	352	1	0.0202	0.2953	45	3	0.0192	1.1538e-4	32
0.35	10	1.2	1	0.008	0.1642	52	1	0.005	0.1653	29	1	0.0177	0.1645	27	1	0.0171	0.1641	31
0.40	5	1.5	1	0.006	0.2801	53	1	0.025	0.2800	116	1	0.0052	0.2796	47	1	0.0417	0.2803	86
0.45	5	1.1	1	0.017	0.3077	130	2	0.057	0.0026	500	1	0.0216	0.3071	48	2	0.0493	1.2488e-4	193
0.50	20	1.1	1	0.005	0.4056	26	2	0.056	0.0054	500	1	0.0147	0.4093	19	3	0.0194	0.0123	38
0.55	10	1.1	1	0.009	0.4198	50	2	0.051	0.0012	500	1	0.0262	0.4186	40	2	0.0396	1.2303e-4	134

TABLA 2

Problema		LMaFit						Lagrangeansor											
miu	rho	$K = [1,25r]$			$K = [1,5r]$			$K = [1,25r]$			$K = [1,5r]$								
r	sr	Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter		
12	2	0.04	1	0.025	0.3839	106	1	0.064	0.3961	153	1	0.0434	0.3796	96	1	0.0917	0.3946	130	
12	2	0.08	1	0.037	0.5839	102	2	0.171	1.487e-4	500	1	0.0385	0.5857	52	2	0.1667	1.2459e-4	339	
100	4	2	0.15	1	0.0141	0.6217	32	0.050	1.219e-4	67	1	0.0314	0.6221	24	2	0.0590	1.1729e-4	67	
100	10	2	0.30	1	0.026	0.5892	13	0.017	1.189e-4	26	1	0.0297	0.5892	13	2	0.0338	9.8973e-5	25	
100	10	10	0.20	7	0.159	0.1942	253	1	0.083	0.8483	65	7	0.1146	0.2027	140	1	0.0610	0.8483	65
1000	1.5	10	0.25	10	0.274	2.719e-4	500	9	0.140	0.1199	197	10	0.2853	1.2268e-4	410	10	0.2533	1.2246e-4	282
1000	1.5	10	0.30	10	0.227	1.249e-4	324	10	0.247	1.244e-4	395	10	1.388	1.2241e-4	149	10	0.1795	1.1834e-4	190
1000	1.5	10	0.40	10	0.100	1.155e-4	98	10	0.102	1.245e-4	113	10	0.0888	1.1193e-4	76	10	0.0995	1.235e-4	78
100	10	20	0.35	2	0.074	0.8363	64	3	0.075	0.7632	66	2	0.0958	0.8363	65	3	0.0932	0.7632	66
100	1.5	20	0.40	6	0.083	0.5763	72	29	0.241	1.242e-4	255	21	0.6006	0.0185	500	29	0.2463	1.2247e-4	139
100	1.5	20	0.50	20	0.160	1.187e-4	158	19	0.124	0.0594	105	20	0.1281	1.1886e-4	89	20	0.1573	1.0840e-4	99
500	1.5	20	0.55	20	0.123	1.193e-4	105	20	0.123	1.139e-4	117	20	0.1124	1.1420e-4	74	20	0.1373	1.1126e-4	84

TABLA 3

Problema			LMaFit						Lagrangeansor									
miu	rho	r sr	$K = [1,25r]$			$K = [1,5r]$			$K = [1,25r]$			$K = [1,5r]$						
			Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter
12	1.01	10 0.04	4	0.619	0.4954	117	5	0.733	0.3891	163	4	0.4511	0.5086	77	11	3.2904	0.0513	500
5	1.2	10 0.08	10	1.176	1.225e-4	200	10	1.081	1.209e-4	208	10	0.8381	1.245e-4	97	10	0.9786	1.1635e-4	125
5	1.2	10 0.15	10	0.286	1.059e-4	38	10	0.325	1.136e-4	42	10	0.3054	1.165e-4	33	10	0.5351	8.6751e-5	48
5	2.7	10 0.3	10	0.177	1.231e-4	20	10	0.212	1.005e-4	21	10	0.2253	6.858e-5	18	10	0.2251	5.095e-5	18
10	1.1	50 0.20	61	6.959	0.0014	500	7	1.070	0.7964	74	61	3.4787	1.1623e-4	165	7	1.3305	0.7968	74
15	1.2	50 0.25	50	4.448	1.223e-4	340	48	2.027	0.0691	151	50	2.4482	1.1714e-4	118	48	2.1932	0.0679	93
10	1.4	50 0.30	50	1.723	1.242e-4	119	50	2.292	1.188e-4	158	50	1.2962	1.1743e-4	59	50	2.4598	1.0150e-4	85
10	2	50 0.40	50	0.738	1.184e-4	48	50	0.960	1.104e-4	57	50	0.7228	1.1566e-4	28	50	1.1937	1.0773e-4	40
100	1.1	100 0.35	13	1.639	0.7841	62	15	1.780	0.7546	65	117	7.4423	1.2497e-4	165	146	22.366	0.0317	500
300	3	100 0.40	67	3.242	0.2261	142	59	2.959	0.2905	135	99	8.8866	0.0219	257	59	6.8179	0.2905	239
300	1.5	100 0.50	100	3.445	1.198e-4	135	100	3.910	1.233e-4	147	100	3.6655	1.1294e-4	97	100	4.9047	1.1209e-4	105
300	1.1	100 0.55	100	1.886	1.189e-4	66	100	3.340	1.209e-4	103	100	2.2733	1.0122e-4	55	100	4.5079	1.1851e-4	96

TABLA 4

Problema		LMaFit					Lagrangean12					Lagrangeansor																
m	r	sr	minu	rho	$K = [1, 25]$					$K = [1, 5]$					$K = [1, 25]$					$K = [1, 5]$								
					Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter	Rgest	Tiempo	Err.Rel	Iter
1000	25	0.15	100	1.5	25	1.47	1.18e-4	43	25	1.46	1.18e-4	42	25	4.16	1.15e-4	79	25	4.25	1.12e-4	80	25	1.90	1.19e-4	35	25	1.93	1.01e-4	36
1000	25	0.25	100	1.5	25	0.97	1.11e-4	23	25	0.98	1.07e-4	23	25	3.22	9.36e-5	57	25	2.94	1.09e-4	48	25	1.24	2.50e-5	22	25	1.35	9.99e-5	22
1000	25	0.35	100	1.5	25	0.99	9.22e-5	21	25	1.02	1.03e-4	21	25	2.32	4.39e-5	35	25	2.41	4.45e-5	35	25	1.29	4.15e-5	19	25	1.35	4.59e-5	19
1000	25	0.55	100	1.5	25	0.71	1.21e-4	12	25	0.85	9.29e-5	14	25	1.63	2.54e-5	18	25	1.61	2.75e-5	18	25	1.18	3.18e-5	12	25	1.14	4.94e-5	12
2000	30	0.15	100	1.5	30	3.17	1.07e-4	27	30	3.88	9.53e-5	27	30	34.26	1.22e-4	137	30	19.27	1.09e-4	109	30	5.73	1.05e-4	25	30	4.87	5.99e-5	27
2000	30	0.25	100	1.5	30	3.26	1.13e-4	21	30	3.59	9.03e-5	22	30	10.46	1.01e-4	45	30	10.54	1.01e-4	45	30	4.45	4.13e-5	12	30	4.09	4.13e-5	18
2000	30	0.35	100	1.5	30	2.97	1.24e-4	16	30	3.30	8.31e-5	17	30	6.81	8.70e-5	26	30	7.03	8.75e-5	26	30	4.00	6.20e-5	15	30	4.27	7.31e-5	15
2000	30	0.55	1000	10	30	2.34	8.34e-5	10	30	3.06	6.09e-5	13	30	6.06	9.60e-5	16	30	6.28	1.05e-4	16	30	3.69	6.68e-5	10	30	5.35	7.11e-5	14
5000	50	0.15	1000	10	50	27.40	1.06e-4	25	50	28.36	9.29e-5	25	50	217.08	1.20e-4	104	50	241.34	1.19e-4	104	50	35.96	9.51e-5	18	50	41.88	9.24e-5	19
5000	50	0.25	1000	1.2	50	24.90	1.15e-4	22	50	27.68	8.54e-5	23	50	63.76	1.11e-4	41	50	73.36	7.58e-5	44	50	29.00	6.67e-5	19	50	43.15	1.12e-4	25
5000	50	0.35	1000	1.2	50	18.26	1.09e-4	14	50	20.18	6.92e-5	15	50	54.32	7.79e-5	29	50	55.26	7.89e-5	29	50	23.83	5.62e-5	13	50	24.62	6.62e-5	13
5000	50	0.55	1000	1.2	50	14.80	8.55e-5	9	50	18.86	7.63e-5	11	50	42.21	6.36e-5	17	50	43.10	6.83e-5	17	50	21.22	4.75e-5	9	50	33.36	2.91e-5	13

TABLA 5

Problema		LMaFit					Lagrangean12					Lagrangeansor											
name	m	n	Tiempo	Iter	Rgest	Err.Rel	NMAE	Tiempo	Iter	Rgest	Err.Rel	miu	rho	reschagxy	NMAE	Tiempo	Iter	Rgest	Err.Rel	miu	rho	reschagxy	NMAE
Jester-1	24983	100	22.43	117	80	0.186	0.024	43.95	111	80	0.174	100	3	10xtol	0.023	40.19	109	80	0.263	100	3	10xtol	0.043
Jester-2	23500	100	20.85	118	80	0.186	0.025	37.75	114	80	0.176	1000	10	1xtol	0.023	34.48	111	80	0.188	1000	10	1xtol	0.026
Jester-3	24938	100	24.59	235	55	9.312e-4	4.054e-5	77.71	392	39	9.927e-4	100	10	0.1xtol	3.410e-5	43.39	236	55	9.273e-4	100	10	0.1xtol	4.137e-5
Jester-all	73421	100	55.38	114	80	0.164	0.020	95.67	117	80	0.165	100	10	0.1xtol	0.020	77.51	101	80	0.179	20	2	0.6xtol	0.023
Movie-100k	943	1682	20.43	507	100	2.1e-3	9.952e-4	50.147	507	100	2.6e-3	20	5	0.03xtol	1.4e-3	21.279	279	100	6.1e-3	20	5	0.03xtol	3.5e-3
Movie-1M	6040	3706	46.715	170	100	0.1022	7.10e-2	189.193	206	100	0.1040	20	3	0.7xtol	7.45e.2	101.649	141	100	0.1114	20	3	0.7xtol	7.96e-2

TABLE 6

Conclusiones

A continuación describimos los logros del presente trabajo:

Hemos descrito elementos teóricos asociados a cálculo sobre matrices; enfatizando en pseudoinversas, proyecciones, normas matriciales y descomposiciones.

También revisamos métodos iterativos para resolver sistemas de ecuaciones, enfocándonos en el método de Gauss Seidal y sus variantes.

El método de Lagrangeanos Aumentados para resolver problemas con restricciones de igualdad también fué estudiado.

Revisamos el desarrollo del artículo de Wang y colaboradores, donde se describe el problema de completación de matrices, se propone el algoritmo SOR (LMaFit), y se analiza su convergencia. El uso de herramientas de Algebra lineal matricial computacional es crucial en la descripción del paper.

Propusimos un algoritmo híbrido SOR-ALM para acelerar el desempeño del SOR. mostramos algunas propiedades del algoritmo. Su análisis de convergencia es una tarea por ser concluída.

Elaboramos experimentos computacionales con datos simulados, donde son comparados los algoritmos SOR y SOR-ALM. También comparamos los algoritmos con datos experimentales provenientes de otros trabajos, como son Jester-Jokes [23] y MovieLens [22].

Comparamos tres algoritmos: Lmafit, una implementación del SOR; lagrangean12, donde programamos el Lagrangeano Aumentado clásico; y el Lagrangeansor, donde nuestro algoritmo híbrido esta implementado.

Al revisar las tablas comparativas observamos que no hay un dominio total de alguno de los algoritmos, pero se evidencia un orden relativo de desempeño donde Lagrangeansor domina, seguido de LmaFit, y tercero la implementación del Lagrangeano.

Como trabajo futuro, pretendemos hacer el análisis de convergencia del algoritmo híbrido, y aplicarlo a otros problemas de cálculo de matrices de rango bajo. Otro asunto pendiente es el tratamiento que se da al cálculo del rango efectivo, que usa en esta versión una heurística que en nuestra opinión no está bien explicado.

Bibliografía

- [1] ZAIWEN WEN, WOTAO YIN AND YIN ZHANG, *Solving a Low-Rank Factorization Model for Matrix Completion by a Nonlinear Successive Over-Relaxation Algorithm* Pre-print, en publicación, dirección web: <http://www.caam.rice.edu/zhang/reports/tr1007.pdf>
- [2] GENE H. GOLUB AND CHARLES F. VAN LOAN, *Matrix Computations*, third edition, The Johns Hopkins University Press, 1996.
- [3] WARD CHENEY AND DAVID KINCAID, *Numerical Mathematics and Computing*.sixth edition,Thomson Brooks/Cole, MA, 2004.
- [4] NOTAS DEL ITESCAM. *Método de Gauss-Seidel*, dirección web: <http://www.itescam.edu.mx/principal/sylabus/fpdb/recursos/r71981.PDF>
- [5] DIMITRI P. BERTSEKAS, *Constrained Optimization and Lagrange Multiplier Methods*,Massachusetts Institute of Technology,1996.
- [6] ROGER A. HORN AND CHARLES R. JOHNSON *Matrix Analysis*,Cambridge University Press,1985.
- [7] B. RECHT, M. FAZEL, AND P. A. PARRILO, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*,tech. rep., 2007.
- [8] E. J. CANDÉS AND B. RECHT, *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics, (2009).
- [9] E. J. CANDÉS AND T. TAO, *The power of convex relaxation: Near-optimal matrix completion*, tech. rep. 2009 Applied and Computational Mathematics, Caltech, 2009.
- [10] Z. LIU AND L. VANDENBERGHE, *Interior-point method for nuclear norm approximation with application to system identification*, SIAM Journal on Matrix Analysis and Applications, 31 (2009), pp. 1235-1256.
- [11] T. MORITA AND T. KANADE, *A sequential factorization method for recovering shape and motion from image streams*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 (1997), pp. 858-867.
- [12] C. TOMASI AND T. KANADE, *Shape and motion from image streams under orthography: a factorization method*, *Int. J. Comput. Vision*, 9 (1992), pp. 137 154.
- [13] L. ELDÉN. *Matrix Methods in Data Mining and Pattern Recognition (Fundamentals of Algorithms)*, *Society for Industrial and Applied Mathematics*, Philadelphia, PA, USA, 2007.

- [14] D. GROSS, *Recovering low-rank matrices from few coefficients in any basis*, tech. rep., Leibniz University, 2009.
- [15] K.C. TOH AND S. YUN, *An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems*, tech. rep., Department of Mathematics, National University of Singapore, 2009.
- [16] R. H. KESHAVAN AND S. OH, *A gradient descent algorithm on the grassman manifold for matrix completion*, tech. rep., Dept. of Electrical Engineering, Stanford University, 2009.
- [17] Y. LIU, D. SUN, AND K. C. TOH, *An implementable proximal point algorithmic framework for nuclear norm minimization*, tech. rep., National University of Singapore, 2009.
- [18] R. MEKA, P. JAIN, AND I. S. DHILLON, *Guaranteed rank minimization via singular value projection*, tech. rep., University of Texas at Austin, 2009.
- [19] R. MAZUMDER, T. HASTIE, AND R. TIBSHIRANI, *Regularization methods for learning incomplete matrices*, tech. rep., Stanford University, 2009.
- [20] W. DAI AND O. MILENKOVIC, *Set: an algorithm for consistent matrix completion*, tech. rep., Department of Electrical and Computer Engineering, University of Illinois, 2009.
- [21] K. LEE AND Y. BRESLER, *Admira: Atomic decomposition for minimum rank approximation*, tech. rep., University of Illinois, Urbana-Champaign, 2009.
- [22] J. L. HERLOCKER, J. A. KONSTAN, A. BORCHERS, AND J. RIEDL, *An algorithmic framework for performing collaborative filtering*, in SIGIR 99: procedente de la Vigésima segunda conferencia anual internacional ACM SIGIR sobre investigación y desarrollo en la recuperación de información, New York, NY, USA, 1999, ACM, pp. 230-237.
- [23] K. GOLDBERG, T. ROEDER, D. GUPTA, AND C. PERKINS, *Eigentaste: A constant time collaborative filtering algorithm*, *Inf. Retr.*, 4 (2001), pp. 133-151.