

UNIVERSIDAD CENTROCCIDENTAL
“LISANDRO ALVARADO”

Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas



“PROBLEMAS INVERSOS, REGULARIZACIÓN Y
BÚSQUEDAS INEXACTAS EN LA CURVA L”

TRABAJO ESPECIAL DE GRADO PRESENTADO POR

BR. MARIANGELA ANDREINA PINTO OCANTO

COMO REQUISITO FINAL
PARA OBTENER EL TÍTULO DE LICENCIADA
EN CIENCIAS MATEMÁTICAS

ÁREA DE CONOCIMIENTO: OPTIMIZACIÓN.

TUTOR: DR. HUGO LARA

Barquisimeto, Venezuela. Junio de 2011



Universidad Centroccidental
 “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 Matemáticas del Decanato de Ciencias y Tecnología de la Universidad Centroccidental “Lisandro Alvarado”, para examinar y dictar el veredicto sobre el Trabajo Especial de Grado titulado:

“PROBLEMAS INVERSOS, REGULARIZACIÓN Y BÚSQUEDAS
 INEXACTAS EN LA CURVA L”

Presentado por la ciudadana BR. MARIANGELA ANDREINA PINTO OCANTO titular de la Cédula de Identidad No. 17.572.120, con el propósito de cumplir con el requisito académico final para el otorgamiento del título de Licenciada en Ciencias Matemáticas.

Luego de realizada la Defensa y en los términos que imponen 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:

¹ _____

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:

¹ Aprobado ó Reprobado

A mis Padres Marlene y Orlando
A mi futuro Esposo Germán

AGRADECIMIENTOS

Primeramente a Dios, que me permitió estar aquí, vivir este momento, que obro a través de la gente que me rodea, para completar este camino que pronto culmina, y empezar uno nuevo, en el que se, me acompañara y bendecirá, pues nada en mi vida ocurre sin su presencia.

A mis Padres... Mamá y Papá sin ustedes, no lo hubiese logrado nunca, son la presencia viva de Dios cuidándome en la tierra, todo su esfuerzo, todos los sacrificios que han hecho por mi, lo mucho que me han cuidado a lo largo de toda mi vida, su amor hacia mi, sus concejos, sus abrazos, todo su trabajo, absolutamente todo me ha hecho la persona que soy y humildemente les expreso mi agradecimiento en estas líneas, Infinitas Gracias mamá y papá.

A mi querido Profe. Hugo, por darme la oportunidad de trabajar al lado de una persona tan preparada, con tantos conocimientos, que camino de la mano conmigo en mis primeros pasos en la investigación matemática, todo lo que he aprendido con usted se que me servirá eternamente. Muchas Gracias Profe.

A mi tía Luz, mi segunda mamá, que siempre ha estado pendiente de mi, que siempre me ha apoyado y ayudado.

A ti Germán que tanto me has dado, pero sobre todo tu amor y tu apoyo, han sido el motor de este proyecto, tu que me animabas cuando no veía el final del camino, que estuviste en todo el proceso junto a mi, como amigo, como compañero, como mi futuro esposo. TE AMO.

A mis Queridas Amigas, Keyserling, Maura, Yeli, Yesika, Will, por haberme tenido la paciencia que tuvieron cuando la necesite durante toda la carrera, por ayudarme y darme animo.

A mis compañeros y ex-compañeros de la Universidad, en especial a Marcos, Maria José, José Antonio, José Angel, porque con ustedes compartí muchas cosas durante mucho tiempo y eso siempre sera valioso para mi.

A mis Tíos y Hermanos, a mis bellos sobrinos, a todas las personas que de un modo u otro colaboraron para que esta meta se cumpliera, este logro no solo es mio, cada uno de ustedes tiene un pedacito de el, estuvieron ahí, y hoy les doy una pequeñísima muestra de mi agradecimiento. Simplemente gracias a todos.

Universidad Centroccidental “Lisandro Alvarado”
“PROBLEMAS INVERSOS, REGULARIZACIÓN Y
BÚSQUEDAS INEXACTAS EN LA CURVA L”

Br. Mariangela A. Pinto O.

Tutor: Dr. Hugo Lara

RESUMEN

Estudiamos problemas inversos lineales discretos mal puestos, donde se desea encontrar el la causa x que produce el efecto b relacionados como

$$b = Mx + n$$

usando el método de regularización de Tikhonov para tratarlo, según Hansen [5]. Revisamos el criterio de la Curva L para seleccionar el parámetro de regularización y búsquedas inexactas que hacen una rápida selección del parámetro, apoyándonos en [6]. Usamos el toolbox de regularización de Hansen [4] en problemas de prueba, y finalmente construimos búsquedas inexactas para elegir el parámetro de regularización, como en [3], pero en el ambiente del toolbox de Hansen.

Palabras claves: Problemas Mal Puestos, Problemas Inversos, Regularización de Tikhonov, Búsquedas inexactas, Curva L.

ÍNDICE

Agradecimientos	i
Resumen	ii
Introducción	1
1. Preliminares	5
1.1. Conceptos Básicos	5
1.2. Sistema de Ecuaciones Lineales	6
1.2.1. Sistema de Ecuaciones Lineales	6
1.3. Aplicaciones Lineales	7
1.3.1. Núcleo e Imagen de una Aplicación Lineal	7
1.4. Autovalores y Autovectores	7
1.4.1. Diagonalización de una Matriz	9
1.5. Descomposición Valor Singular	10
1.5.1. Interpretación de la descomposición valor singular de una matriz	14
1.5.2. Geometría de una Transformación Lineal	16
1.5.3. La Descomposición Valor Singular en Problemas de Modelos de Ajustes	18
1.5.4. Relación con la inversa Moore-Penrose	20
1.5.5. Los efectos del ruido y valores singulares pequeños	22
1.6. Mínimos Cuadrados	24
1.6.1. Solución del Sistema $Ax = b$ mediante el método de mínimos cuadrados	24
1.7. Búsquedas Inexactas	25
1.7.1. Conceptos Básicos	25

1.7.2. Condiciones de Optimalidad	25
1.7.3. Condiciones de Wolfe	27
2. Problemas Inversos, Problemas Mal Puestos y Regularización	31
2.1. Introducción a Problemas Inversos, Regularización y Curva L	31
2.2. Problemas Inversos	33
2.3. Regularización	37
2.3.1. La Data Perturbada y la Norma de La Solución	38
2.4. Algunos Métodos de Regularización	39
2.4.1. Descomposición Valor Singular Truncada	39
2.4.2. Regularización de Tikhonov	40
2.4.3. La Descomposición Valor Singular y El método de Regulari- zación de Tikhonov	42
2.5. Análisis de la Descomposición valor Singular	45
3. La Curva-L y Búsquedas Inexactas en la Curva-L	48
3.1. La Curva-L	48
3.1.1. Introducción	48
3.1.2. La Curva-L	49
3.2. Análisis de la Curva-L	50
3.2.1. Un Criterio de Mínimo Local	65
3.3. Búsquedas Inexactas en La Curva-L Rotada.	68
3.3.1. Derivadas en La Curva L	68
3.3.2. Curva L Rotada	69
4. Experimentación Numérica y Resultados	73
Conclusiones	90
A. Códigos generados en Matlab	92
B. Ejemplos de Problemas Mal Puestos	95

Referencias

99

ÍNDICE DE TABLAS

2.1. Distintos valores de x_α	33
2.2. Clasificación de Problemas Directos vs. Problemas Inversos	36
4.1. Resumen de la Experimentación	75
4.2. Test Problema Deriva2 caso1	77
4.3. Test Problema Deriva2 caso2	78
4.4. Test Problema Deriva2 caso3	78
4.5. Data Evans	79
4.6. Test Problema Foxgood	79
4.7. Test Problema Gravity caso1	80
4.8. Test Problema Gravity caso2	80
4.9. Test Problema Gravity caso3	81
4.10. Test Problema Heat	81
4.11. Test Problema Ilaplace caso1	82
4.12. Test Problema Ilaplace caso2	82
4.13. Test Problema Ilaplace caso3	83
4.14. Test Problema Ilaplace caso4	83
4.15. Data Lanczos	84
4.16. Test Problema Phillips	84
4.17. Test Problema Shaw	85
4.18. Test Problema Spikes	85
4.19. Test Problema stay-week	86
4.20. Test Problema Tomo	86
4.21. Test Problema Wing	87
4.22. Suma de una Exponencial	87
4.23. Suma de dos Exponenciales	88

4.24. Suma de tres exponenciales	88
4.25. Suma de cuatro exponenciales	89
B.1. Resultados del problema	97

Índice de figuras

1.1.	Acción de la Matriz M sobre el vector x	8
1.2.	Descomposición Valor Singular de la Matriz M	15
1.3.	Descomposición Valor Singular de la Matriz M^T	16
1.4.	Núcleo e Imagen del Operador M	17
1.5.	Tamaño de Paso	27
1.6.	Condición de suficiente decaimiento	28
1.7.	Condición de Curvatura	29
4.1.	Test problema Baart	76
B.1.	Gráficas de la aproximación de la solución $\phi(t) = t$	98

INTRODUCCIÓN

Muchos problemas en ciencias e ingeniería son problemas inversos mal puestos. Dada una matriz M de orden $n \times m$ y un vector de datos $b \in \mathbb{R}^n$, un problema lineal inverso requiere de resolver de manera aproximada el siguiente sistema de ecuaciones lineales:

$$Mx = b. \tag{1}$$

Según Hadamard un problema inverso es bien puesto si satisface los requerimientos de existencia, estabilidad y unicidad de sus soluciones. Si uno de dichos requerimientos no es satisfecho, entonces el problema se dice que es mal puesto.

Una solución de mínimos cuadrados para el sistema (1) es una solución de

$$\min_{x \in \mathbb{R}^N} \|b - Mx\|^2. \tag{2}$$

Cuando un problema inverso es mal puesto, la solución de mínimos cuadrados para el problema (2) no es satisfactoria, dando frecuentemente reconstrucciones pobres. Para sortear estas dificultades, se introducen esquemas de regularización (ver por ejemplo [4]). Métodos de regularización para calcular soluciones estables de problemas inversos hacen un balance entre el tamaño de la solución regularizada (o su diferencia con una solución por defecto conocida) y la calidad del ajuste que esta proporciona a unos datos dados. El bien conocido esquema de regularización de Tikhonov (ver por ejemplo [5]) consiste en reemplazar el problema de mínimos cuadrados (2) por

$$\min_{x \in \mathbb{R}^N} \|b - Mx\|^2 + \lambda^2 \|L(x - x_c)\|^2, \tag{3}$$

donde λ es llamado parámetro de regularización. Aquí, el “tamaño” de la solución regularizada, para una matriz L dada, es medida por $\|L(x - x_c)\|$; donde el ajuste es medido por $\|b - Mx\|$; x_c es un estimador a priori de x el cual representa nuestro conocimiento previo de la solución. Si no hay información a priori, entonces x_c es

cero.

Denotamos por $x(\lambda)$ la única solución optimal para el problema (3). Se necesita la regularización al resolver problemas inversos porque las soluciones “ingenuas” de mínimos cuadrados ($\lambda = 0$), denotadas por x_{LS} , es completamente dominada por contribuciones que provienen de los errores de datos y errores de redondeo.

Durante el desarrollo de este trabajo también se estudiarán conceptos, métodos y procedimientos relacionados con las búsquedas inexactas y regularización.

- **Regularización de Tikhonov y la curva L :** la regularización es introducida para amortiguar esas contribuciones y mantener la norma $\|L(x - x_c)\|$ de un tamaño razonable. Si mucha regularización, o amortiguación, se impone a la solución, entonces no ajustará bien los datos dados y y el residual $\|b - Mx\|$ será muy grande. Por otra parte, si se impone muy poca regularización, entonces el ajuste será bueno, pero la solución será dominada por los errores, y así la norma $\|L(x - x_c)\|$ será muy grande. El parámetro de regularización λ juega un rol importante en el balance de dichas normas. Una herramienta crucial en la elección de un valor propio del parámetro λ es el llamado criterio de la curva L (ver [5]). La curva L es el gráfico $(\|b - Mx(\lambda)\|, \|L(x(\lambda) - x_c)\|)$ para $\lambda > 0$. Es una curva parametrizada por λ que ayuda a controlar el balance entre las dos cantidades. La curva L es de nuestro interés porque muestra como la solución regularizada cambia cuando el parámetro de regularización varía. Se distingue una esquina en forma de L localizada exactamente donde la solución $x(\lambda)$ cambia de ser dominada por los errores de los datos. Esta es la razón de porque la esquina de la curva L corresponde a un buen balance entre la minimización de estas dos cantidades, y el valor del parámetro correspondiente es bueno.
- **Descomposición de valor singular y regularización:** la cual es una herramienta estándar para analizar soluciones regularizadas es la descomposición de valor singular (SVD) de la matriz M , la cual es una descomposición de la

forma

$$M = \sum_{i=1}^N u_i \sigma_i v_i^T. \quad (4)$$

donde los vectores singulares izquierdos y derechos u_i y v_i son ortonormales, y los valores singulares σ_i son no negativas y aparecen en un orden no decreciente. Es fácil mostrar que, si $L = I$ y $x_c = 0$, la solución regularizada es dada por

$$x(\lambda) = \sum_{i=1}^N f_i \frac{u_i^T b}{\sigma_i} v_i.$$

donde f_1, \dots, f_N son los llamados factores de filtro de Tikhonov $f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda}$. La norma de las soluciones y la norma del vector residual que caracteriza el desajuste son dadas en términos de la descomposición de valor singular por

$$\|x(\lambda)\|^2 = \sum_{i=1}^N \left(f_i \frac{u_i^T b}{\sigma_i} \right)^2. \quad (5)$$

y $\|b - Mx(\lambda)\|^2 = \sum_{i=1}^N ((1 - f_i) u_i^T b)^2$. Estas expresiones forman la base para analizar la curva L en [5, 4, 6] y [1]. Sin embargo, en problemas grandes la descomposición de valor singular pudiera ser computacionalmente cara, y sería preferible encontrar soluciones iterativas para el problema (3). Daremos nuestro análisis en el contexto de estas soluciones iterativas.

- Búsquedas inexactas:** Cuando calculamos puntos $x(\lambda)$ sobre la curva L , debemos resolver el problema de optimización (3) para cada valor de λ . Una búsqueda sobre la curva para elegir λ sobre la esquina involucra posiblemente un proceso iterativo donde se define una sucesión λ_k que converge a λ^* en la esquina. Cuando el problema inverso es grande, es deseable escoger λ^* en tan pocas de estas costosas iteraciones como sea posible. Existen en la literatura búsquedas inexactas (ver por ejemplo [8], [10]). Una búsqueda inexacta tiene por objetivo escoger una solución aproximada en un número pequeño de pasos, cuando no se necesita mucha precisión. En lugar de encontrar una solución con pequeño error, se admite un error mayor, pero garantizando un

decrecimiento suficiente de la función objetivo en cada paso. Búsquedas como Armijo o Wolfe usan primeras derivadas. En [3] se definen procedimientos de búsquedas inexactas para encontrar un buen valor del parámetro de regularización de manera rápida, en problemas lineales inversos mal puestos de sumas de exponenciales.

- **Toolbox de regularización de Hansen:** Hansen [4] elaboró una herramienta en Matlab para tratar problemas inversos mal puestos, usando métodos de regularización. Esta herramienta presenta varias funciones Matlab que resumen extensivamente la teoría de regularización clásica. Hay rutinas para resolver problemas lineales inversos mal condicionados, con herramientas estándar, como la descomposición de valor singular, que tratan problemas pequeños y medianos, y también herramientas complementarias que permiten comparar las diferentes estrategias de solución, y las distintas soluciones encontradas. Entre otras, el paquete presenta rutinas para analizar la curva L.

CAPÍTULO 1

PRELIMINARES

Para el desarrollo de este trabajo se consultó diversas fuentes de información, se estudió Algebra Lineal (Autovalores y Autovectores, transformaciones lineales, y sistemas de Ecuaciones Lineales) y para ello se utilizó ([2]) y ([9]). También se revisó fuentes bibliográficas correspondientes a Optimización No lineal como lo son ([10])y ([8]), esto en relación al desarrollo de la Teoría de Búsquedas Inexactas. Además de esto se consulto algunas publicaciones previas a la investigación que origino este Trabajo Especial De Grado ([3]), como lo son ([6], [5], [7]) con lo cual se pudo entender la teoría relacionada con Métodos de Regularización y Problemas Inversos. También se puede mencionar que para la puesta en practica computacional de este trabajo se revisó manuales de Matlab, disponibles en internet, esto para conocer un poco mas de cerca el lenguaje de programación ahí utilizado, así como la familiarización con el ambiente Matlab, y desde luego se utilizó para el desarrollo de los programas ([4]).

1.1. Conceptos Básicos

Definición 1.1.1. Se le llama vector a los elementos de un Espacio Vectorial.

Definición 1.1.2. Se conoce como Norma al operador que determina la longitud o magnitud de un vector.

Para efectos de este trabajo la norma utilizada será la norma que se deriva del producto interno la cual denotaremos como $\| \cdot \|$, y se definirá a continuación. Es importante acotar que existen otras normas.

Definición 1.1.3. El Producto Interno en \mathbb{R}^n : Sean x, y dos vectores en \mathbb{R}^n , el Producto Interno denotado como $\langle x, y \rangle$, es el escalar $x^T y$, es decir:

$$\langle x, y \rangle = x^T y = x_1 y_1 + x_2 y_2 + \dots + x_n y_n.$$

Definición 1.1.4. Norma en \mathbb{R}^n : Sea x un vector en \mathbb{R}^n , la norma de x se denotara como $\|x\|_2$ y se definirá como:

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

1.2. Sistema de Ecuaciones Lineales

Una ecuación lineal sobre un campo Real, es una expresión de la forma:

$$a_1 x_1 + a_2 x_2 + \dots + a_n x_n = b. \quad (1.1)$$

donde $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ y $b \in \mathbb{R}$ y los x_i son indeterminados. Los escalares a_i son los coeficientes de los x_i respectivamente y b es el término constante.

Se dice que $K \in \mathbb{R}^n$ una solución de (1.1) si la proposición que se obtiene al sustituir k_i por x_i

$$a_1 k_1 + a_2 k_2 + \dots + a_n k_n = b.$$

es cierta.

1.2.1. Sistema de Ecuaciones Lineales

Sea el siguiente sistema:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1. \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2. \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots &\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m. \end{aligned}$$

el cual es un sistema de m ecuaciones lineales con n incógnitas x_1, \dots, x_n , donde los a_{ij}, b_i pertenecen a un cuerpo ($K = \mathbb{R}$ ó \mathbb{C}). El Sistema se dice Homogéneo si las

constantes b_1, \dots, b_m son todas 0.

Una n -upla $u = (k_1, \dots, k_n)$ de números reales es una solución, si esta satisface cada una de las ecuaciones.

1.3. Aplicaciones Lineales

Sean \mathbb{V} y \mathbb{W} dos espacios Vectoriales sobre \mathbb{R}^n . Una aplicación $F : \mathbb{V} \rightarrow \mathbb{W}$ es una aplicación lineal si satisface las siguientes condiciones:

- Para todo $v, w \in \mathbb{V}$, $F(v + w) = F(v) + F(w)$.
- Para todo $k \in \mathbb{R}$ y todo $v \in \mathbb{V}$, $F(kv) = kF(v)$.

1.3.1. Núcleo e Imagen de una Aplicación Lineal

Sea $F : \mathbb{V} \rightarrow \mathbb{W}$ una Aplicación Lineal. Denotaremos la imagen de F como **Imag**(F), y se define como el conjunto de las imágenes de los puntos de \mathbb{V} en \mathbb{W}

$$\mathbf{Imag}(F) = \{w \in \mathbb{W} : F(v) = w \text{ para algún } v \in \mathbb{V}\}.$$

El núcleo de F , lo denotaremos **Ker**(F), y se define como el conjunto de elementos de \mathbb{V} que se aplican en $0 \in \mathbb{W}$:

$$\mathbf{Ker}(F) = \{v \in \mathbb{V} : F(v) = 0\}.$$

1.4. Autovalores y Autovectores

Dada una matriz cuadrada M de orden $n \times n$, podemos decir que M es un operador lineal que actúa sobre un espacio vectorial \mathbb{V} de dimensión n . De esta forma el operador M transforma los vectores $x \in \mathbb{V}$ en $Mx \in \mathbb{V}$, se dice que un subespacio \mathbb{W} de \mathbb{V} es M -invariante si los vectores $w \in \mathbb{W}$ siguen estando en \mathbb{W} al transformarse por M , esto es:

$$w \in \mathbb{W} \rightarrow Mw \in \mathbb{W}$$

Así notamos que necesariamente el vector Mw es un múltiplo escalar de w , esto es la existencia de un escalar $\lambda_0 \in \mathbb{K} = \mathbb{R} \text{ ó } \mathbb{C}$ fijo tal que:

$$Mw = \lambda_0 w \tag{1.2}$$

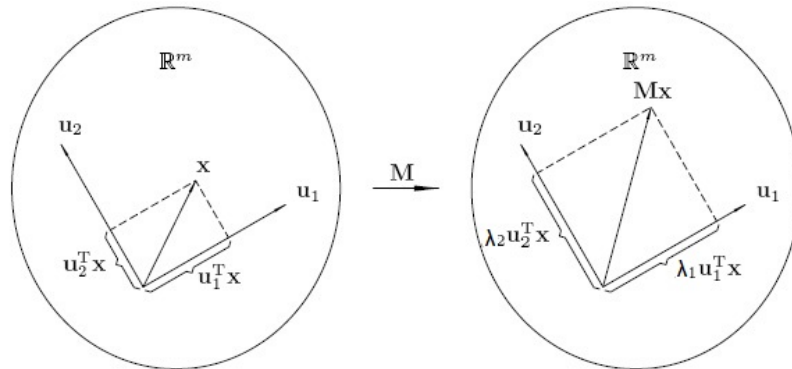


Figura 1.1: Acción de la Matriz M sobre el vector x

Definición 1.4.1. Al escalar λ_0 que cumple con la relación (1.2) se denomina Valor Propio de M .

Definición 1.4.2. Al vector w que cumple con la relación (1.2) se denomina vector propio de M asociado λ_0

Observemos que la ecuación (1.2) es equivalente a

$$(M - \lambda I)w = 0. \tag{1.3}$$

con $\lambda_0 = \lambda$ e I la matriz identidad. La ecuación (1.3) admite soluciones no-nulas si $M - \lambda I$ es una matriz singular esto se traduce en:

$$\det(M - \lambda I) = 0. \tag{1.4}$$

Los valores propios de M son las soluciones de (1.4). Además (1.4) se le llama polinomio característico de M , este polinomio es de orden n y es de la forma

$$P_M(\lambda) = \det(M - \lambda I).$$

Definición 1.4.3. Si λ_0 es un valor propio de M , la solución de la ecuación (1.2) forman un subespacio vectorial de soluciones llamado subespacio propio.

1.4.1. Diagonalización de una Matriz

Sea M una matriz real, simétrica de orden $m \times m$, todos los autovalores asociados a M son reales y sus autovectores forman una base ortonormal de \mathbb{R}^m . Si los valores propios son denotados por μ_i y los vectores propios correspondientes se indican u_i , entonces afirmar que u_i es el autovector asociado al autovalor μ_i , es decir:

$$Mu_i = u_i\mu_i. \quad (1.5)$$

Si ahora escribimos cada uno de los vectores u_1, \dots, u_m en forma de vectores columnas y formamos una matriz cuadrada:

$$U = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1m} \\ u_{21} & u_{22} & \dots & u_{2m} \\ \vdots & \vdots & & \vdots \\ u_{m1} & u_{m2} & \dots & u_{mm} \end{pmatrix}$$

La relación (1.5) para $i = 1, \dots, m$ puede ser escrita como:

$$\begin{aligned} MU &= \begin{pmatrix} \vdots & \vdots & & \vdots \\ Mu_1 & Mu_2 & \cdots & Mu_m \\ \vdots & \vdots & & \vdots \end{pmatrix} \\ &= \begin{pmatrix} \vdots & \vdots & & \vdots \\ \mu_1 u_1 & \mu_2 u_2 & \cdots & \mu_m u_m \\ \vdots & \vdots & & \vdots \end{pmatrix} \\ &= \begin{pmatrix} \vdots & \vdots & & \vdots \\ u_1 & u_2 & \cdots & u_m \\ \vdots & \vdots & & \vdots \end{pmatrix} \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \mu_m \end{pmatrix} \\ &= UD \end{aligned}$$

En donde D es una matriz Diagonal, tal que esa diagonal esta formada por los autovalores de la matriz M , además como U es una matriz ortonormal, ella es invertible y así

$$M = UDU^{-1} = UDU^T.$$

1.5. Descomposición Valor Singular

Sea M una matriz real de orden $m \times n$ con ($m \geq n$) la cual es un operador que envía vectores de \mathbb{R}^m en vectores de \mathbb{R}^n . Tenemos que tanto MM^T como $M^T M$ son matrices cuadradas tales que MM^T es de orden m y $M^T M$ es de orden n , así podemos hablar sobre autovalores y autovectores de ellas.

Definición 1.5.1. Sea M una matriz real de orden $m \times n$ con ($m \geq n$), los valores singulares de M son las raíces cuadradas de los autovalores de $M^T M$ (respectivamente MM^T).

Proposición 1.5.1. Si tanto la matriz $M^T M$ y MM^T son semidefinida positivas entonces todos los autovalores son positivos

Demostración:

Supongamos sin pérdida de generalidad que $M^T M$ es semidefinida positiva, y consideremos v un vector singular, fijo pero arbitrario de $M^T M$, asociado al valor singular λ , entonces se cumple por (1.2):

$$M^T M v = \lambda v \tag{1.6}$$

Multiplicando al lado izquierdo por v^T y agrupando los términos en (1.6)

$$\begin{aligned} v^T M^T M v &= v^T \lambda v \\ \Rightarrow (v^T M^T)(M v) &= \lambda (v^T v) \\ \Rightarrow \|M v\|^2 &= \lambda \|v\|^2 \end{aligned}$$

Luego como $\|M v\|^2$ es no negativa por hipótesis y $\|v\|^2$ es positiva entonces $\lambda > 0$. De igual forma ocurre con MM^T .

□

La matriz MM^T es una matriz simétrica de orden m , así tiene m vectores ortonormales propios y $M^T M$ es una matriz simétrica de orden n , por lo cual tiene n vectores propios ortonormales.

Proposición 1.5.2. *Si todos los autovalores de $M^T M$ (respectivamente MM^T) son distintos de ceros, entonces los autovalores de $M^T M$ (respectivamente MM^T) son los mismos que MM^T (respectivamente $M^T M$).*

Demostración:

Denotemos v_i con $i = 1, \dots, n$ los n autovectores ortonormales de $M^T M$ asociados a λ_i y supongamos que

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$$

y denotemos a u_j con $j = 1, \dots, m$ los m autovectores ortonormales de MM^T , y tal que

$$\mu_1 \geq \mu_2 \geq \dots \geq \mu_m \geq 0$$

sea i fijo pero arbitrario y consideremos el autovector v_i de $M^T M$ asociado a λ_i , con $\lambda_i \neq 0$. Se quiere mostrar que Mv_i es un autovector de MM^T , en efecto

$$(MM^T)(Mv_i) = M(M^T M)v_i = M\lambda_i v_i = \lambda_i Mv_i.$$

Esto último muestra que Mv_i es un Autovector de MM^T el cual pertenece al autovalor λ_i .

Si normalizamos Mv_i

$$\frac{Mv_i}{\|Mv_i\|}.$$

obtenemos un autovector propio de MM^T con norma unitaria, y este corresponde a alguno de los u_i mencionados anteriormente, siempre que los autovalores de MM^T sean no degenerados. Si existiese degeneración entre los autovalores, es posible tomar una combinación lineal apropiada de ellos y así trabajar.

Así obtenemos para cada i un autovector de $M^T M$ que es también un autovector de MM^T . De este modo tenemos que si existen r autovectores de $M^T M$ que son distintos de cero, se tiene que $\lambda_1 = \mu_1, \dots, \lambda_r = \mu_r$ y el resto de los autovalores iguales a cero, $\lambda_i = 0$ con $i = r + 1, \dots, n$ y $\mu_j = 0$ para $j = r + 1, \dots, m$.

□

Hemos obtenemos los autovectores normalizados

$$u_k = \frac{Mv_k}{\|Mv_k\|} \text{ y } v_k = \frac{M^T u_k}{\|A^T u_k\|}. \quad (1.7)$$

Notemos que

$$\|Mv_k\|^2 = (Mv_k)^T(Mv_k) = v_k^T(M^T M)v_k = v_k^T \lambda_k v_k = \lambda_k > 0 \quad (1.8)$$

donde los u_i forman una base ortonormal. Así podemos definir σ_k como la raíz cuadrada del autovalor (1.8), luego para $k = 1, 2, \dots, r$

$$\|Mv_k\| = \|M^T u_k\| = \sigma_k = \sqrt{\lambda_k} = \sqrt{\mu_k}.$$

y de (1.7) se obtiene

$$Mv_k = \sigma_k u_k. \quad (1.9)$$

$$M^T u_k = \sigma_k v_k. \quad (1.10)$$

La transformación Lineal M actúa sobre el vector unitario $v_k \in \mathbb{R}^n$ llevándolo al otro espacio como $\sigma_k u_k \in \mathbb{R}^m$ cuya longitud es σ_k y dirección $u_k \in \mathbb{R}^m$ de igual forma la transformación lineal M^T sobre los vectores unitarios de la base $u_k \in \mathbb{R}^m$ actúa enviándolos al otro espacio como $\sigma_k v_k \in \mathbb{R}^n$, un vector de longitud σ_k en dirección de $v_k \in \mathbb{R}^n$. Ahora bien, hemos supuesto que r -autovalores son no nulos para cada una de las matrices MM^T y $M^T M$, así que cuando consideramos a los $k > r$, el autovalor λ_k de $M^T M$ es nulo, y donde además esta asociado a un vector v_k , de donde se obtiene que

$$M^T Mv_k = 0. \quad (1.11)$$

Multipliquemos ahora la expresión (1.11) por v_k^T a la izquierda y se obtiene:

$$\begin{aligned} v_k^T M^T Mv_k = 0 &\Rightarrow (Mv_k)^T(Mv_k) = 0 \\ &\Rightarrow \|Mv_k\| = 0 \\ &\Rightarrow Mv_k = \vec{0}. \end{aligned}$$

y con el mismo argumento se verifica que $MM^T u_k = 0$. Así obtenemos

$$Mv_k = \vec{0} \text{ para } k = r + 1, \dots, n \quad (1.12)$$

$$M^T u_k = \vec{0} \quad \text{para } k = r + 1, \dots, m \quad (1.13)$$

y podemos ahora apreciar de una forma explícita como actúa la matriz M y M^T sobre las bases $\{v_k\}_{k=1}^n$ y $\{u_k\}_{k=1}^m$.

Ahora bien, si se tiene cualquier operador lineal que haga la misma acción de M sobre los vectores de la base $\{v_k\}_{k=1}^n$ $k = 1$, este operador ha de ser el mismo M . En efecto, solo basta ver que este nuevo operador realice la misma acción de la matriz M para todo vector en \mathbb{R}^n . Con los vectores básicos ya se satisface, y para cualquier otro vector w este es posible escribirlo como combinación lineal de la base $\{v_k\}_{k=1}^n$ y es claro que tanto M como el otro supuesto operador harán exactamente lo mismo, de donde se tiene el resultado.

Obtengamos ahora expresiones explícitas para el operador M , probemos que el operador

$$\sum_{k=1}^r \sigma_k u_k v_k^T \quad (1.14)$$

para $i = 1, \dots, r$ Realiza las mismas acciones de M sobre los elementos de la base $\{v_k\}_{k=1}^n$ y $\{u_k\}_{k=1}^m$ respectivamente M^T

$$\left(\sum_{k=1}^r \sigma_k u_k v_k \right) u_i = \sum_{k=1}^r \sigma_k u_k v_k^T u_i \quad (1.15)$$

$$= \sum_{k=1}^r \sigma_k u_k v_k^T u_i \quad (1.16)$$

$$= \sigma_1 u_1 v_1^T u_i + \sigma_2 u_2 v_2^T u_i + \dots + \sigma_r u_r v_r^T u_i \quad (1.17)$$

$$= \sigma_k u_i \quad (1.18)$$

donde en (1.18) ocurre por

$$v_k^T u_i = \begin{cases} 1 & \text{si } k = i \\ 0 & \text{si } k \neq i \end{cases}$$

y para $i = r + 1, \dots, n$

$$\left(\sum_{k=1}^r \sigma_k u_k v_k^T \right) u_i = \sum_{k=1}^r \sigma_k u_k v_k^T u_i \quad (1.19)$$

$$= 0. \quad (1.20)$$

y en (1.20) ya que $i \neq k$. Así hemos probado que el operador (1.14) es el mismo M así

$$M = \sum_{k=1}^r \sigma_k u_k v_k^T. \quad (1.21)$$

y aplicando traspuesta se obtiene el operador M^t

$$M^T = \sum_{k=1}^r \sigma_k v_k u_k^T. \quad (1.22)$$

Definición 1.5.2. A los vectores $\{v_k\}$ de (1.21) se le denominan Vectores Singulares Derechos asociados al valor singular σ_k .

Definición 1.5.3. A los vectores $\{u_k\}$ de (1.22) se le denominan Vectores Singulares Izquierdos asociados al valor singular σ_k .

Ahora de veremos que existen matrices ortonormales U y V tales que

$$M = UDV^T.$$

donde D es una matriz diagonal de orden $m \times n$ cuyas entradas d_{ii} son los llamados valores singulares, es decir, $d_{ii} = \sigma_i$, y donde U es la matriz cuya k -ésima columna es u_k , V es la matriz cuya k -ésima columna es el vector v_k

1.5.1. Interpretación de la descomposición valor singular de una matriz

Veamos ahora que sentido toma Mx cuando se tiene una expresión explícita de M como la dada en (1.21) y describir como M actúa sobre $x \in \mathbb{R}^n$. De (1.21), se tiene que:

$$Mx = \left(\sum_{k=1}^r \sigma_k u_k v_k^T \right) x \quad (1.23)$$

$$= \sum_{k=1}^r \sigma_k u_k (v_k^T x). \quad (1.24)$$

Dado que $x \in \mathbb{R}^n$, x puede ser escrito como combinación lineal de la base $\{v_k\}_{k=1}^r$. Luego el producto $v_k^T x$ dado en (1.24), representa el valor de la componente de x

a lo largo del vector v_k y al multiplicar dicho producto por el valor singular σ_k y por el vector u_k , nos sugiere en la ecuación (1.24), que la componente del vector x sobre la base de los v_k ha sido enviada a la componente de Mx en la base de los u_k . Observemos la situación sobre 2 dimensiones:

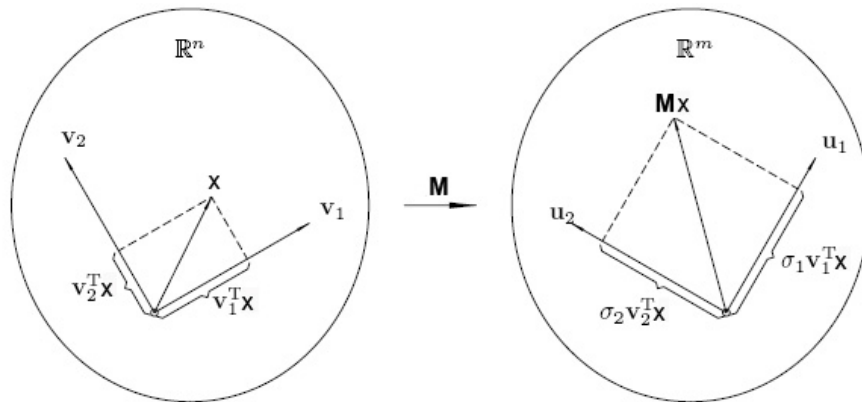


Figura 1.2: Descomposición Valor Singular de la Matriz M

De forma similar se estudia el efecto de M^T sobre $b \in \mathbb{R}^m$ con el resultado obtenido en (1.22), lo cual podemos apreciar en el gráfico

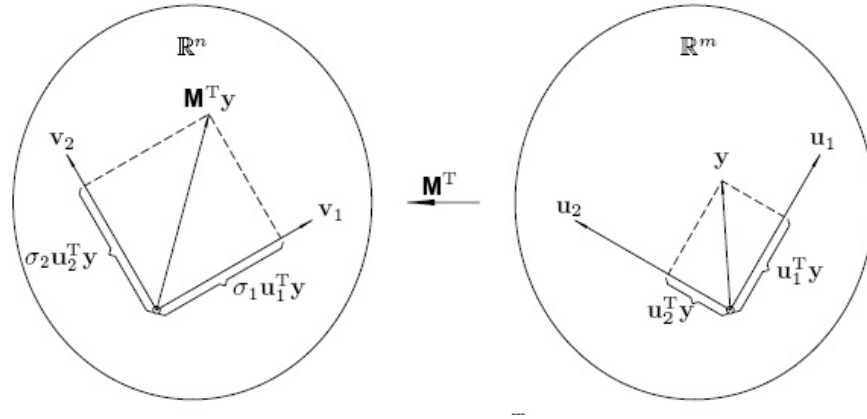


Figura 1.3: Descomposición Valor Singular de la Matriz M^T

1.5.2. Geometría de una Transformación Lineal

Consideremos las ecuaciones (1.9) y la proposición (1.5.2) indican que

$\mathbf{Imag}(M) = \mathbf{span}\{u_1, u_1, \dots, u_r\}$ y $\mathbf{Ker}(M^T) = \mathbf{span}\{u_{r+1}, u_{r+2}, \dots, u_m\}$ respectivamente.

Como $\{u_1, \dots, u_m\}$ es una base del espacio \mathbb{R}^m , dicho espacio podemos escribirlo de la siguiente forma:

$$\mathbb{R}^m = \mathbf{Imag}(M) \oplus \mathbf{Ker}(M^T).$$

y de manera similar

$$\mathbb{R}^n = \mathbf{Imag}(M^T) \oplus \mathbf{Ker}(M).$$

Veamos ahora como actúa el operador M sobre x para que este sea enviado al espacio de los datos. Sea $x \in \mathbb{R}^n$ y consideremos como base de \mathbb{R}^n el conjunto de los vectores singulares derechos $\{v_k\}_{k=1}^n$, así escribimos:

$$x = \sum_{k=1}^n x_k v_k = \sum_{k=1}^n \langle x, v_k^T \rangle v_k \quad x_k = \langle x, v_k^T \rangle \quad (1.25)$$

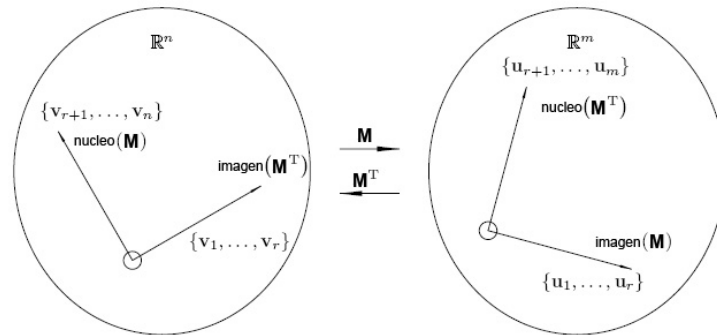


Figura 1.4: Núcleo e Imagen del Operador M

y esto es la suma vectorial de las componentes de x a lo largo de las direcciones básicas y (1.24) puede escribirse como

$$Mx = \sum_{k=1}^r \sigma_k x_k u_k \quad (1.26)$$

Esta última expresión nos dice que la información de la data no se encuentra de forma explícita, y por ende las proyecciones de x a lo largo de los vectores v_k . Notemos que en la ecuación (1.26) el valor singular σ_k ejerce una influencia para indicar que tan presente se encuentra la k -ésima componente de la imagen en la data.

Sabemos que el valor de r representa el rango de la matriz M , y considerando $r < n$ señala que solo las primeras r componentes de x tienen presencia en la data. Esto nos dice que la información aportada por la data no es suficiente, en el sentido que si se tienen dos imágenes distintas x y x' las cuales son iguales en las primeras r componentes, difiriendo en las restantes $n - r$ componentes, entonces se observa que en la misma data hallaremos dos imágenes que se satisfacen con el mismo b y son distintas.

El objetivo es encontrar una x única y cuando solo considerando la data podríamos obtener varias soluciones para la misma data, esto crea la necesidad estudiar técnicas que nos permitan determinar las componentes invisibles a la data.

1.5.3. La Descomposición Valor Singular en Problemas de Modelos de Ajustes

En esta sección se vera un poco como relacionar la Descomposición valor singular, con la solución a problemas mal puestos. Consideremos una Colección de m datos o puntos de la forma $\{(z_k, y_k)\}$ y un recta que se ajusta a ellos como es el caso de la recta $y = x_o + x_i z$, la cual está representada de la siguiente forma:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} 1 & z_1 \\ 1 & z_2 \\ \vdots & \vdots \\ 1 & z_3 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \quad (1.27)$$

es notorio que la dimensión del espacio imagen es pequeña, de dimensión $n = 2$ (es importante destacar que no se trata de la dimensión de la imagen de la matriz), y la dimensión del espacio de los datos es m . Si el rango de la matriz fuese n nos encontraríamos que el problema arroja respuestas bien definidas, y se tiene que la imagen de M es un subespacio de dimensión 2.

La aproximación por mínimos cuadrados tiene como objetivo encontrar el parámetro \hat{x} de tal manera que Mx se aproxime tanto como sea posible a b , es decir:

$$\hat{x} = \arg \min \|b - Mx\|^2. \quad (1.28)$$

ahora bien utilizando la descomposición valor singular de la matriz M :

$$M = \sum_{k=1}^r \sigma_k u_k v_k^T.$$

y considerando que la data es de tamaño m y la ecuación (1.26), podemos escribir la data b como combinación lineal de la base formada por los u_k es decir $\{u_k\}_{k=1}^m$. Así, la data se escribe como la suma vectorial de los componentes a lo largo de cada vector de la base:

$$b = \sum_{k=1}^m u_k (u_k^T b). \quad (1.29)$$

Utilizando este último resultado (1.29), minimizar (1.28), para algún x dado puede ser estudiado como a continuación:

$$\|b - Mx\|^2 = \left\| \sum_{k=1}^m u_k(u_k^T b) - \sum_{k=1}^r \sigma_k u_k(v_k^T x) \right\|^2; \quad (1.30)$$

$$= \left\| \sum_{k=1}^r [u_k(u_k^T b) - \sigma_k u_k(v_k^T x)] + \sum_{k=r+1}^m u_k(u_k^T b) \right\|^2; \quad (1.31)$$

$$= \left\| \sum_{k=1}^r u_k[(u_k^T b) - \sigma_k(v_k^T x)] + \sum_{k=r+1}^m u_k(u_k^T b) \right\|^2; \quad (1.32)$$

$$= \left\| \sum_{k=1}^r u_k[(v_k^T b) - \sigma_k(v_k^T x)] \right\|^2 + \left\| \sum_{k=r+1}^m u_k(u_k^T b) \right\|^2; \quad (1.33)$$

$$= \sum_{k=1}^r |(u_k^T b) - \sigma_k(v_k^T x)|^2 + \sum_{k=r+1}^m |u_k^T b|^2. \quad (1.34)$$

Donde hemos utilizado la ortonormalidad de los vectores u_k , $k = 1, 2, \dots, m$ para la última igualdad.

Supongamos que tenemos el vector \hat{x} , el cual es el vector que minimiza (1.28), tal que es un mínimo cuando \hat{x} es tomado de la siguiente forma:

$$v_k^T \hat{x} = \frac{u_k^T b}{\sigma_k}. \quad \text{para } k=1, 2, \dots, r \quad (1.35)$$

Además en la igualdad (1.34) notamos que uno de los términos que conforman esta igualdad, no depende del vector \hat{x} que tomamos, así que la solución afecta solo el término restante.

La expresión (1.35) determinara al vector solución \hat{x} si $r = n$ ó $r < n$, para el modelo de ajuste que hemos considerado y la condición de que $r = n$ podemos decir que la

solución viene dada de la siguiente manera:

$$\hat{x} = \sum_{k=1}^n (v_k^T \hat{x}) v_k; \quad \text{por (1.25)} \quad (1.36)$$

$$= \sum_{k=1}^n \left(\frac{u_k^T b}{\sigma_k} \right) v_k; \quad \text{utilizando (1.35)} \quad (1.37)$$

$$= \sum_{k=1}^n \left(\frac{u_k^T b}{\sigma_k} \right) v_k; \quad (1.38)$$

$$= \left(\sum_{k=1}^n \frac{1}{\sigma_k} u_k^T v_k \right) b. \quad (1.39)$$

1.5.4. Relación con la inversa Moore-Penrose

Estudiamos ahora la Inversa de Moore-Penrose, que es muy importante para calcular la solución. Recordamos que el objetivo es encontrar una solución \hat{x} que minimiza (1.28). Ahora reescribamos esto de la siguiente forma:

$$C = \|b - Mx\|^2. \quad (1.40)$$

y notemos que $\sum_{l=1}^n m_{kl} x_l$ es la k -ésima componente de Mx . Por otra parte la norma (1.40) no es mas que la sumatoria de las componentes del vector $b - Mx$ al cuadrado, es por ello que también podemos ver (1.40) como:

$$C = \sum_{k=1}^m \left(b_k - \sum_{l=1}^n m_{kl} x_l \right)^2. \quad (1.41)$$

Como es típico en las teorías de optimización, para minimizar la expresión (1.40) se considera la utilización de la derivada para minimizar respecto a cada componente:

$$\frac{dC}{dx_i} = \sum_{k=1}^m 2 \left(b_k - \sum_{l=1}^n m_{kl} x_l \right) (-m_{ki}) = 0. \quad i = 1, 2, \dots, n \quad (1.42)$$

Reacomodando esta ultima expresión tenemos:

$$\implies 2 \sum_{k=1}^m \left(-m_{ki} b_k + \sum_{l=1}^n m_{kl} m_{ki} x_l \right) = 0; \quad i = 1, 2, \dots, n \quad (1.43)$$

$$\implies - \sum_{k=1}^m m_{ki} b_k + \sum_{l=1}^n \left(\sum_{k=1}^m m_{ki} m_{kl} \right) x_l = 0; \quad (1.44)$$

$$\implies \sum_{l=1}^n \left(\sum_{k=1}^m m_{ki} m_{kl} \right) x_l = \sum_{k=1}^m m_{ki} b_k. \quad (1.45)$$

Notemos que la igualdad (1.45) escrita en forma matricial queda como:

$$(M^T M)x = M^T b; \quad (1.46)$$

esta representación matricial es la que se obtiene al solucionar el problema (1.40) vía mínimos cuadrados, y se denominan ecuaciones normales.

Ahora bien para que la solución sea única la matriz $(M^T M)$ debe ser invertible, así la solución estará dada por:

$$\hat{x} = (M^T M)^{-1} M^T b; \quad (1.47)$$

así definamos según [9] la Inversa Moore-Penrose:

Definición 1.5.4. Sea M una matriz de orden $m \times n$. La matriz $(M^T M)^{-1} M^T$ (1.47), se le denomina Inversa Moore-Penrose de M .

Ahora dada la definición queremos buscar una expresión para ella utilizando la descomposición valor singular. Calculemos la descomposición valor singular de la Matriz $M^T M$:

$$M^T M = \left(\sum_{k=1}^r \sigma_k v_k u_k^T \right) \left(\sum_{l=1}^r \sigma_l u_l v_l^T \right); \quad \text{por (1.21) y (1.22)} \quad (1.48)$$

$$= \sum_{k=1}^r \sum_{l=1}^r \sigma_k \sigma_l v_k (u_k^T u_l) v_l^T; \quad (1.49)$$

$$= \sum_{k=1}^r \sigma_k^2 v_k v_k^T. \quad \text{por la ortogonalidad de los } u_k \quad (1.50)$$

$M^T M$ es una matriz invertible (ya que es cuadrada de orden $n \times n$) si $r = n$, en cuyo caso su inversa estaría escrita como:

$$(M^T M)^{-1} = \sum_{k=1}^r \frac{1}{\sigma_k^2} v_k v_k^T. \quad (1.51)$$

Luego utilizando (1.51) y (1.22), la inversa de Moore-Penrose se escribe:

$$(M^T M)^{-1} M^T = \left(\sum_{k=1}^r \frac{1}{\sigma_k^2} v_k v_k^T \right) \left(\sum_{l=1}^r \sigma_l v_l u_l^T \right); \quad (1.52)$$

$$= \sum_{k=1}^r \sum_{l=1}^r \frac{1}{\sigma_k^2} \sigma_l v_k (v_k^T v_l) u_l^T; \quad (1.53)$$

$$= \sum_{k=1}^r \frac{1}{\sigma_k} v_k u_k^T. \quad (1.54)$$

Notemos que el resultado obtenido para \hat{x} es igual tanto trabajando con la inversa de Moore-Penrose, como con solo la descomposición valor singular como fue hecho en la sección anterior.

1.5.5. Los efectos del ruido y valores singulares pequeños

Es importante observar la acción que ejercen los valores singulares pequeños, en las direcciones de sus respectivos valores propios durante el proceso de reconstruir la data.

La data puede ser escrita como:

$$b = Mx + n;$$

donde M representa la data verdadera y n el ruido, de la descomposición valor singular del operador M se obtiene

$$b = \left(\sum_{k=1}^n \sigma_k u_k v_k^T \right) x + n. \quad (1.55)$$

Supongamos que el rango es n , el cual es el tamaño del espacio imagen ($n < m$), la aplicación directa es inyectiva, así el problema tiene solución única en el sentido de

mínimos cuadrados, la cual viene dada por:

$$\hat{x} = \left(\sum_{k=1}^n \frac{1}{\sigma_k} v_k u_k^T \right) b. \quad (1.56)$$

Si en (1.56) sustituimos (1.55) obtenemos

$$\hat{x} = \sum_{k=1}^n \left(\frac{1}{\sigma_k} v_k u_k^T b \right); \quad (1.57)$$

$$= \sum_{k=1}^n \left(\frac{1}{\sigma_k} u_k^T \left[\sum_{k=1}^n (\sigma_k u_k v_k^T) x + n \right] \right) v_k; \quad (1.58)$$

$$= \sum_{k=1}^n \left(\frac{1}{\sigma_k} u_k^T \sigma_k u_k v_k^T x + \frac{u_k^T n}{\sigma_k} \right) v_k; \quad (1.59)$$

$$= \sum_{k=1}^n \left(v_k^T x v_k + \frac{v_k^T n}{\sigma_k} v_k \right); \quad (\{v_k\} \text{ es una base ortonormal}) \quad (1.60)$$

$$= x + \sum_{k=1}^n \frac{1}{\sigma_k} (u_k^T n) v_k; \quad (\{v_k\} \text{ es una base ortonormal}) \quad (1.61)$$

Observamos en la expresión (1.61) que la reconstrucción se basa en la imagen real más un término el cual indica la presencia de ruido. Esto nos indica la presencia que el Ruido en la k -ésima componente de la reconstrucción hecha depende del ruido presente en la k -ésima componente de la data, el cual se encuentra dividido por el valor singular correspondiente. Notemos que las componentes donde los valores singulares son pequeños, el ruido es amplificado y esto trae como consecuencia que la componente de la imagen real, es decir x , sea opacada a lo largo de esa dirección. En pocas palabras la data contiene poca información de la imagen en esas componente. Así si se intenta dar mas importancia al papel de la data en el modelo entonces el ruido se amplifica, esto nos indica que el método de mínimos cuadrados nos aporta malas reconstrucciones para modelos cuyo operador lineal contiene valores singulares muy pequeños.

1.6. Mínimos Cuadrados

1.6.1. Solución del Sistema $Ax = b$ mediante el método de mínimos cuadrados

La condición para la solución exacta del sistema de n ecuaciones lineales con m incógnitas $Ax = b$, donde A es una matriz $n \times m$, se expresa diciendo que b está en el espacio columna de A , dado que el producto Ax es una combinación lineal de columnas de A . Si el vector b no cumple esta condición, el sistema $Ax = b$ no tiene solución exacta. Sin embargo es posible buscar una condición aproximada.

La solución x' de $Ax = b$ en el sentido de mínimos cuadrados es aquella que minimiza el error

$$E = \|Ax - b\|. \quad (1.62)$$

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{pmatrix} X = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

La solución x' que cumple con lo anterior tal que Ax' es la proyección ortogonal de b sobre el espacio columna de A . Esto nos permite calcular x' como sigue: $Ax' - b$ es ortogonal al espacio de las columnas de A . Sea w un vector en este espacio, entonces w es una combinación lineal de las columnas de A , es decir: $w = Ac$, donde $c = [c_1, c_2, \dots, c_m]$. Entonces para cualquier c arbitrario :

$$\langle Ac, Ax' - b \rangle = 0;$$

es decir,

$$\begin{aligned} (Ac)^T(Ax' - b) &= 0, \\ \Rightarrow c^T A^T(Ax' - b) &= 0, \end{aligned}$$

entonces

$$c^T(A^T Ax' - A^T b) = 0,$$

finalmente

$$A^T A x' = A^T b.$$

Estas ecuaciones son llamadas ecuaciones normales. Si A tiene columnas linealmente independientes entonces $A^T A$ tiene inversa y las ecuaciones normales se convierten en:

$$x' = (A^T A)^{-1} A^T b$$

1.7. Búsquedas Inexactas

La optimización sin restricciones según [10], [8] trata el problema de minimizar una función objetivo que depende de variables reales sin restricciones sobre los valores de esas variables. La formulación matemática es:

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

donde f es una función continua y diferenciable.

1.7.1. Conceptos Básicos

Definición 1.7.1. Un punto x^* es un Mínimo Global de la función f si $f(x^*) \leq f(x)$ para todo $x \in \mathbb{R}^n$.

Definición 1.7.2. Un punto x^* es un Mínimo Local si existe un entorno \mathcal{B} de x^* tal que $f(x^*) \leq f(x)$ para todo $x \in \mathcal{B}$.

Si una función f es continua y diferenciable existen modos eficientes para identificar mínimos locales.

1.7.2. Condiciones de Optimalidad

- CONDICIÓN NECESARIA DE PRIMER ORDEN:

Si x^ un mínimo local y f es continuamente diferenciable en un entorno abierto de x^* , entonces $\nabla f(x^*) = 0$. (x^* es un punto estacionario)*

- **CONDICIÓN NECESARIA DE SEGUNDO ORDEN:**
Si x^ es un mínimo local y $\nabla^2 f$ es continua en un entorno abierto de x^* , entonces $\nabla f(x^*) = 0$ y $\nabla^2 f(x^*)$ es semidefinida positiva.*
- **CONDICIÓN SUFICIENTE DE SEGUNDO ORDEN:**
Si $\nabla^2 f$ es continua en un entorno abierto de x^ , $\nabla f(x^*) = 0$ y $\nabla^2 f(x^*)$ es definida positiva, entonces x^* es un mínimo local estricto de f .*

Cuando f es convexa, cualquier mínimo local x^* es un mínimo global de f . Si además f es diferenciable, entonces cualquier punto estacionario x es un mínimo global de f .

En cada iteración, los métodos de búsqueda lineal calculan una dirección de Descenso d_k y deciden que tanto moverse a lo largo de la dirección. Estas direcciones están dadas por:

$$x_{k+1} = x_k + \alpha_k d_k. \quad (1.63)$$

Definición 1.7.3. El vector d_k de la ecuación (1.63) es llamada la dirección de descenso.

Se dice que la dirección d de \mathbb{R}^n es una dirección de descenso en x si

$$f(x + \alpha d) < f(x),$$

para α positivo suficientemente pequeño.

Una condición suficiente para que d sea una dirección de descenso es que se verifique:

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

Entre todas las direcciones en que podemos movernos desde x_k la dirección $d_k = -\nabla f(x_k)$ es aquella a lo largo de la cuál f decrece más de prisa. Sin embargo si el condicionamiento de la matriz hessiana en el óptimo es grande la convergencia es muy lenta.

Definición 1.7.4. El escalar α_k de la ecuación (1.63) es llamado Tamaño de Paso.

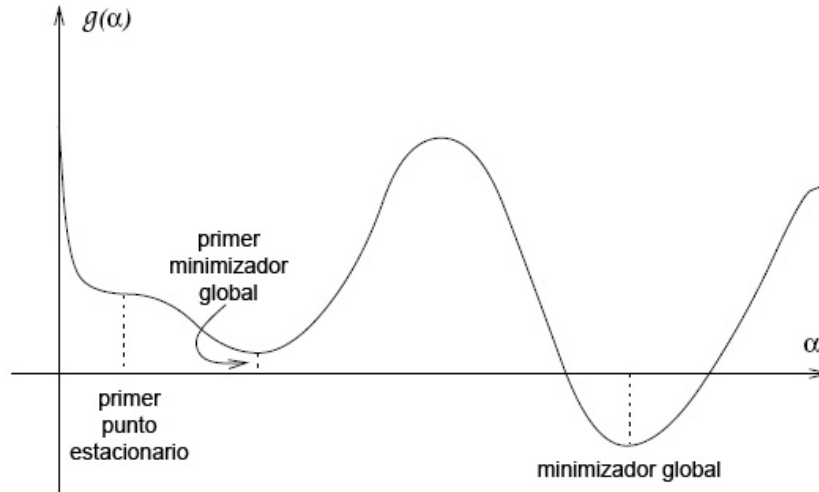


Figura 1.5: Tamaño de Paso

Sea $g : \mathbb{R} \rightarrow \mathbb{R}^n$ definida por

$$g(\alpha) = f(x_k + \alpha p_k) \quad (1.64)$$

Con $\alpha > 0$, y f es la función Objetivo. La idea básica es encontrar un minimizador global de la función f , sin embargo esto puede ser costoso computacionalmente, así recurrimos a estrategias más prácticas denominadas métodos de Búsqueda Lineal Inexactas, para identificar el tamaño de paso que alcance adecuadamente reducciones en f al mínimo costo.

1.7.3. Condiciones de Wolfe

Un método de Búsqueda Lineal inexacta muy usado es la Búsqueda de Wolfe, la cual consiste en que el Tamaño de Paso α_k cumpla las siguientes condiciones

$$g(\alpha_k) \leq g(\alpha_0) + c_1 g'(\alpha_0) \alpha_k \quad (1.65)$$

$$g'(\alpha_k) \geq c_2 g'(\alpha_0) \quad (1.66)$$

con $0 < c_1 < c_2 < 1$.

Definición 1.7.5. La Condición (1.65) es llamada Condición de Armijo

La condición de Armijo busca una reducción de la función en el iterado actual que sea proporcional al Tamaño de Paso y al Gradiente de la Función.

Gráficamente podemos apreciar que los subintervalos que cumplen la condición (1.65) La condición de Armijo es implementada al considerar la función $g(0) +$

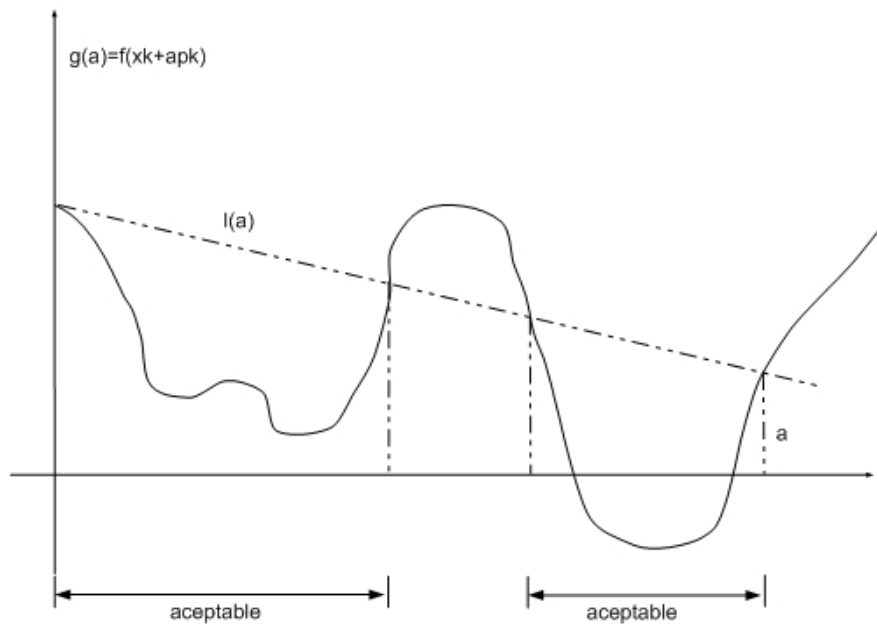


Figura 1.6: Condición de suficiente decaimiento

$cg'(0)\alpha$. Un valor de α se considera que no es muy grande si el valor de la función cae debajo de la línea punteada. Para asegurar que α no sea muy pequeño, un valor de $\eta > 1$ es seleccionado, y α no es considerado muy pequeño si

$$g(\eta\alpha) > g(o) + cg'(0)\eta\alpha.$$

si α crece η veces falla el criterio anterior. Esta sola condición no es suficiente puesto que pueden existir valores de α muy pequeños que cumplan con (1.65), de ahí la necesidad de la segunda condición.

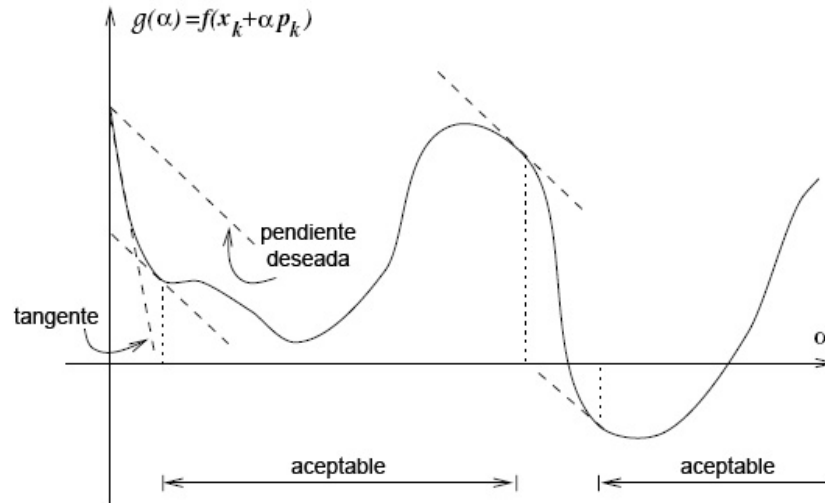


Figura 1.7: Condición de Curvatura

Definición 1.7.6. A la condición (1.66) se le denomina Condición de Curvatura

Gráficamente podemos apreciar, que corresponde a aquellos puntos cuyas derivadas son proporcionalmente mayor a la derivada en el punto inicial.

Los criterios de selección que siguen las condiciones de Wolfe, para la elección del paso adecuado son:

- Si $g(\alpha) \leq g(0) + c_1 \alpha g'(0)$ lo cual nos indica un paso suficientemente pequeño para tener descenso, y $g'(\alpha) \geq c_2 g'(0)$ lo que indica un paso no demasiado pequeño, para que el algoritmo no se relentece, entonces el paso es adecuado y se termina.
- Si $g(\alpha) > g(0) + c_1 \alpha g'(0)$ entonces $\alpha_g = \alpha$ lo que indica que α es demasiado grande.
- Si $g(\alpha) \leq g(0) + c_1 \alpha g'(0)$ y $g'(\alpha) < c_2 g'(0)$ entonces $\alpha_p = \alpha$ nos indica que el paso es demasiado pequeño.

La interpretación es que se pide por una parte que f descienda suficientemente y por la otra parte la derivada debe aumentar lo bastante para que el iterado siguiente es decir x_{k+1} no sea excesivamente próximo a x_k , lo cual podría generar una convergencia muy lenta de tal método.

CAPÍTULO 2

PROBLEMAS INVERSOS, PROBLEMAS MAL PUESTOS Y REGULARIZACIÓN

2.1. Introducción

El concepto de Problemas mal puestos, viene de principios de siglo *XX* cuando el afamado matemático francés Jacques Hadamard (1865-1963) lo definió por primera vez. El creía que esta clase de problemas eran “artificiales” ya que según el, no describían ninguna situación física real. Hoy en día se ha comprobado que en ese punto en particular estaba equivocado, pues existen múltiples aplicaciones para ellos, como veremos más adelante.

Estudemos un pequeño ejemplo tomado de [4].

$$\min_x \|b - Mx\|.$$

con

$$M = \begin{pmatrix} 0,16 & 0,10 \\ 0,17 & 0,11 \\ 2,02 & 1,29 \end{pmatrix} \quad b = \begin{pmatrix} 0,27 \\ 0,25 \\ 3,38 \end{pmatrix};$$

donde b se escribe como la solución exacta dada por $\bar{x} = \begin{pmatrix} 1,00 \\ 1,00 \end{pmatrix}$; más una pequeña perturbación, esto es:

$$b = \begin{pmatrix} 0,16 & 0,10 \\ 0,17 & 0,11 \\ 2,02 & 1,29 \end{pmatrix} \begin{pmatrix} 1,00 \\ 1,00 \end{pmatrix} + \begin{pmatrix} 0,01 \\ -0,03 \\ 0,02 \end{pmatrix}; \quad (2.1)$$

Calculando el número de condición de M , dada por la siguiente ecuación:

$$\mu(M) = \|M^{-1}\| \|M\| = 1,110^3;$$

Ellos explican que el calculo de la solución es potencialmente muy sensitiva a perturbaciones de la data. La solución de mínimos cuadrados, vía ecuaciones normales, esta dada por:

$$x_{lsq} = \begin{pmatrix} 7,01 \\ -8,40 \end{pmatrix};$$

También afirman que un número de condición grande, es indicación que las columnas de M son linealmente dependientes y proponen reemplazar las columnas de la matriz mal condicionada $M = (a_1 \ a_2)$ por $(a_1 \ 0)$ o $(0 \ a_2)$ Las cuales están mejor condicionadas. Las dos soluciones que proporcionan para estas nuevas matrices son:

$$x_B^{(1)} = \begin{pmatrix} 1,65 \\ 0,00 \end{pmatrix} \quad x_B^{(2)} = \begin{pmatrix} 0,00 \\ 2,58 \end{pmatrix};$$

Si bien estas soluciones son mucho menos sensibles a las perturbaciones de los datos, y si bien las normas residuales son pequeñas:

$$\|Mx_B^{(1)} - b\| = 0,031 \quad \|Mx_B^{(2)} - b\| = 0,036.$$

estas soluciones no tienen nada en común con la solución exacta.

La mayor dificultad con la solución de mínimos cuadrados es que su norma es significativamente mayor que la norma de la solución exacta, en [4] proponen resolver el problema de mínimos cuadrados adicionando una restricción a la norma de la solución es decir

$$\min_x \|Mx - b\| \quad \text{sujeto a} \quad \|x\| \leq \alpha$$

Así calculan la solución x_α , que depende de α , vemos la siguiente tabla:

Se aprecia que una adecuada solución de $\alpha = 1,385$ se puede calcular una solución bastante aproximada a la solución exacta \bar{x} .

α	x_α
0,1	$\begin{pmatrix} 0,08 \\ 0,05 \end{pmatrix}$
1	$\begin{pmatrix} 0,84 \\ 0,54 \end{pmatrix}$
1,385	$\begin{pmatrix} 1,17 \\ 0,74 \end{pmatrix}$
10	$\begin{pmatrix} 6,51 \\ -7,60 \end{pmatrix}$

Tabla 2.1: Distintos valores de x_α

2.2. Problemas Inversos

Podemos entender de manera general, a un problema inverso, como un problema en el cual se conocen las consecuencias, pero no las causas. En este tipo de problemas tienen más dificultad para ser solucionado que un problema directo. Ahora Bien sería natural preguntarse: **¿A qué puede ser inverso este problema?**, diremos entonces dos problemas son inversos cuando para la formulación de uno involucramos al otro.

Según [9] podemos definir un problema Inverso como un problema donde se predicen los datos a partir de parámetros y de un modelo.

Según [7] Un problema inverso consiste en determinar las causas de un fenómeno mediante el análisis de sus efectos.

Menke (1989) dice que el problema inverso es simplemente el conjunto de métodos usados para extraer información útil de nuestro entorno a partir de medidas físicas o datos.

Existen diversas clasificaciones de los problemas inversos (mas o menos arbitrarias), pero entre las mas difundidas tenemos:

- Problemas retrospectivos: Determinar el estado presente de un sistema basado en observaciones futuras. Un ejemplo de este tipo de problemas es determinar la distribución de calor en una barra metálica para un instante de tiempo pa-

sado, basado en la distribución de calor actual.

- Problema de identificación de parámetros: Las leyes que gobiernan el sistema son conocidas, pero no así los parámetros involucrados.

- Problema de condiciones de frontera: Hace falta información acerca de las condiciones en la frontera de un dominio dado.

Definición 2.2.1. El Conjunto de todos los Posibles datos se le llama Espacio de la Data.

Los datos en un problema los denotaremos por b . Denotando por \bar{b} a la data libre de ruido, el problema directo se considera como la aplicación que va de la imagen a \bar{b} . Para nosotros el ruido sera la diferencia $\bar{b} - b$.

Definición 2.2.2. La colección de Valores que queremos reconstruir son llamados Imagen.

Definición 2.2.3. El Conjunto de todas las Imágenes es llamado Espacio de Imágenes

Ahora podemos definir un Problema Inverso:

Definición 2.2.4. El Problema de hallar la imagen original de una data dada, conociendo el problema directo es llamado Problema Inverso

Veamos algunas definiciones: Según (Hadamard (1923)) un problema es bien puesto si:

- Existe una solución para cualquier data en el espacio de datos.

- La solución es única en el espacio de imágenes.

- La aplicación inversa $b \rightarrow x$ es continua.

Las primeras dos condiciones son equivalentes a exigir que la inversa de M esté bien definida y que su dominio sea todo el espacio de datos.

La continuidad exigida en la tercera condición es una condición necesaria pero no suficiente para la estabilidad de la solución. Así definimos un problema Bien puesto como:

Definición 2.2.5. Un problema es bien puesto, si cumple con las siguientes condiciones:

- Su solución existe.
- La solución es única.
- La solución depende en forma continua de los datos.

Definición 2.2.6. Un problema es mal puesto (ill-posed) si no es bien puesto.

Es decir, generalizando un poco este último concepto, en ([4]) afirman:

Dado el siguiente sistema de ecuaciones

$$Mx = b, \quad M \in \mathbb{R}^{n \times n} \quad (2.2)$$

El Problema de mínimos cuadrados lineales:

$$\min_x \|Mx - b\|, \quad M \in \mathbb{R}^{m \times n}, \quad m > n \quad (2.3)$$

es Discreto Mal Puesto, si los siguientes criterios son dados:

- (a) Los valores Singulares de M decaen Gradualmente a Cero.
- (b) El radio entre el mayor y el menor valor singular distinto de cero es muy grande.

En general todo problema Inverso es un problema mal puesto. Dependiendo de si la data y la Imagen sean funciones de variables continuas o discretas, podemos clasificar de una manera básica el problema directo, tal como lo indicamos a continuación:

Imagen	Data
Continuo	Continuo
Continuo	Discreto
Discreto	Continuo
Discreto	Discreto

Tabla 2.2: Clasificación de Problemas Directos vs. Problemas Inversos

En la práctica la clasificación (Discreto - Discreto) es la mas usual y utilizada, ya que los datos que se pueden medir siempre serán finitos, y por la parte de la imagen se trabaja con una discretización de esta.

Queremos resolver sistemas de la forma:

$$b - Mx = 0. \tag{2.4}$$

Supongamos que queremos resolver un sistema como el (2.4), pero los valores de los coeficientes de b tienen errores, el número de condicionamiento de M nos da una idea de lo sensible que es el problema a errores en los datos.

Entre más alto sea el número de condicionamiento, más se amplifican los errores en los datos, es decir, el número de condicionamiento esta relacionado con que tan bien puesto es el sistema de ecuaciones.

Un número de condicionamiento bajo, nos lleva a un problema bien puesto, mientras que uno alto, nos lleva a un problema mal puesto.

Definición 2.2.7. Sea M una matriz de orden $n \times m$ de coeficientes reales, entonces su número de condicionamiento μ se define como

$$\mu(M) = \|M^{-1}\| \|M\|. \tag{2.5}$$

Definición 2.2.8. Una matriz con un número de condicionamiento pequeño es una matriz bien condicionada.

Ahora bien, Cuando se presenta un problema bien puesto, el error de propagación relativo de la data a la solución es controlado por el número de condición de M .

Si Δb es la variación de b y Δx la variación de x , entonces

$$\frac{\|\Delta x\|}{\|x\|} \leq \text{cond}(M) \frac{\|\Delta b\|}{\|b\|} \quad (2.6)$$

de allí que se desee un número de condición pequeño.

Hadamaard dijo: *Un problema mal puesto es aquel que deja de satisfacer alguna de las tres condiciones*, estas condiciones están dadas en la página 34, pues entonces, diremos que estamos frente a un problema mal puesto si la inversa de M no existe y esto ocurre porque la data no se encuentra en el rango de M , o la imagen que es solución del problema, no es única, ya que más de una imagen es enviada a la misma data, o porque pequeñas perturbaciones de la data produce grandes perturbaciones de la imagen.

2.3. Regularización

Regularización de manera general es la aproximación de un problema mal puesto, por medio de problemas bien puestos similares. En general la regularización conlleva a realizar ciertas suposiciones acerca de la solución (con respecto a su tamaño o suavidad, por ejemplo).

Debido a que no se cuenta con toda la información sobre la x , debemos considerar información adicional que nos permita tener criterios de selección dentro de un conjunto factible de reconstrucciones.

En [4] Afirman que una de las principales dificultades de los problemas mal puestos, es que son esencialmente indeterminado para valores singulares pequeños de las matriz M , es por ello que es necesario incorporar más información sobre la solución, para la estabilización del problema y poder seleccionar una solución útil y estable, siendo este es el propósito de la Regularización.

El problema puede complicarse aun más debido a que M depende de el modelo

de medición, y dado que este proceso no es preciso, se tiene como resultado valores singulares imprecisos.

2.3.1. La Data Perturbada y la Norma de La Solución

Consideremos el problema lineal

$$b = Mx + n. \quad (2.7)$$

Donde $M \in \mathbb{R}^{m \times n}$. Si la imagen es reconstruida a través de datos simulados, cercanos a los datos observados entonces, da pie a una buena reconstrucción de los mismos, denotado por $M\hat{x}$, de esta forma para ([9]) se puede medir la calidad de \hat{x} con la norma del residual, la cual es:

$$C(x) = \|b - Mx\|^2. \quad (2.8)$$

Podemos apreciar, que la selección de \hat{x} para la minimización (2.8) nos proporciona reconstrucciones pobres, si además de lo anterior se considera, que el rango de la matriz M ($M \in \mathbb{R}^{m \times n}$) es menor que n tendremos infinitas reconstrucciones de la imagen que minimizan (2.8), y ya que la data no nos aporta ninguna información sobre algunos aspectos de la imagen x es necesario incluir información adicional que nos permita seleccionar entre varias reconstrucciones factibles. Esto se hará mediante la inclusión de una nueva función $\Omega(x)$ la cual representa nuestra versión de la solución para una reconstrucción en particular. Por ejemplo podemos elegir reconstrucciones que estén cerca de la solución por defecto como es el caso de:

$$\Omega(x) = \|x - x_c\|^2. \quad (2.9)$$

lo que se desea en (2.9) es que esa norma, sea pequeña, para ello es necesaria la inclusión de un operador lineal que actúe sobre esta diferencia, así:

$$\Omega(x) = \|L(x - x_c)\|^2; \quad (2.10)$$

$$= (L(x - x_c))^T(L(x - x_c)); \quad (2.11)$$

$$= (x - x_c)^T L^T L(x - x_c). \quad (2.12)$$

Tal como se dice en [9] Si el espacio de la Imagen es de dimensión n y el espacio de la data es de dimensión m , la matriz M es de dimensión $m \times n$ y la matriz L es de dimensión $p \times n$, donde $p \leq n$, en general esta matriz es la Identidad, pero también es posible usar una aproximación a la $(n - p)$ ésima derivada.

También es posible usar una combinación lineal de las derivadas, tal como:

$$\Omega(x) = \alpha_0 \|x - x_c\|^2 + \sum_{k=1}^q \alpha_k \|L_k(x - x_c)\|^2. \quad (2.13)$$

donde L_k es una matriz la cual es una aproximación de la k -ésima derivada, y α_k es una constante no negativa.

2.4. Algunos Métodos de Regularización

2.4.1. Descomposición Valor Singular Truncada

Sea la Matriz M que es un operador cuya descomposición valor singular está dada por:

$$M = \sum_{l=1}^r \sigma_l u_l v_l^T. \quad (2.14)$$

La descomposición valor singular Truncada es un método basado en la descomposición valor singular de la matriz M , la cual es truncada en cierto nivel.

Consideremos los mayores valores singulares de la matriz M , sea $k \leq n$ tal que los valores singulares son significativos, hasta el valor k . Las componentes de la reconstrucción hecha a lo largo de los vectores singulares correspondientes están bien determinadas por la data, las componentes a lo largo de los restantes direcciones, es decir $l = k + 1, \dots, n$ serán seleccionadas de forma tal que que la solución reconstruida satisfaga algún criterio de optimización, como por ejemplo la minimización de la seminorma (2.9).

Sea la matriz V de orden n formada por todos los vectores singulares derechos

de la matriz M , y denotemos a V_k la $n \times (n - k)$ matriz cuyas columnas son los vectores $\{v_l\}$ para $l = k + 1, \dots, n$, de esta forma se dice que las columnas de la matriz V_k , generan el espacio nulo de la matriz M , así tenemos una reconstrucción de la solución considerando solo los vectores que no contribuyen con la generación del espacio nulo de la matriz M . Dicha reconstrucción viene dada por:

$$x' = \sum_{l=1}^k \left(\frac{u_l^T d}{\sigma_l} \right) v_l. \quad (2.15)$$

La reconstrucción que se desea de la solución puede ser vista como la suma de x' , con un vector que es la superposición de los vectores columnas que forman V_k , así la solución reconstruida se escribiría como:

$$\hat{x} = x' + v_k c. \quad (2.16)$$

y reescribiendo la solución de la seminorma dada en (2.9) se tiene que:

$$\Omega(\hat{x}) = \|L(\hat{x} - x_c)\|^2 = \|L(x' - V_k c - x_c)\|^2 = \|L(x' - x_c) + (LV_k)c\|^2. \quad (2.17)$$

y donde el vector c esta definido como:

$$c = -(LV_k)^\dagger L(x' - x_c). \quad (2.18)$$

la notación \dagger indica que estamos trabajando con la inversa Moore-Penrose, la cual estudiamos en secciones anteriores. Finalmente una expresión para la solución esta dada por:

$$\hat{x} = f' - V_k(LV_k)^\dagger L(x' - x_c). \quad (2.19)$$

2.4.2. Regularización de Tikhonov

Consiste en la suma de dos funciones mencionadas anteriormente, como lo son (2.8) y (2.9). Recordemos que la función (2.9) corresponde es la que aporta la parte histórica del problema, la función (2.8) corresponde a las contribuciones de la data completa es decir incluyendo el ruido, recordemos entonces que a esta data la denotamos como b .

Ahora consideremos un nuevo parámetro llamado λ , que actuara de la siguiente forma:

$$\hat{x}_\lambda = \arg \min\{\|Mx - b\|^2 + \lambda^2\|(x - x_c)\|^2\}. \quad (2.20)$$

este nuevo parámetro definido, es un factor de moderación que da mas o menos peso al factor histórico del problema.

Definición 2.4.1. Al factor λ^2 de (2.20) se le llamara, parámetro de regularización.

Estudiando la dependencia de (2.20) al parámetro de regularización, vemos que es totalmente dependiente de este, y que varia notoriamente si λ varia.

Si hacemos tender $\lambda \rightarrow \infty$ la información proveniente de la data (2.8) se minimizara casi por completo, y entonces (2.20) tomara el valor de la información por defecto.

Por el contrario si $\lambda \rightarrow 0$, la solución por defecto influye poco siendo mas relevante la información aportada por la data.

Vemos entonces el importante rol de λ , ahora bien el objetivo es encontrar un λ apropiado que minimice ambas funciones, y obtener un resultado mas óptimo. Estudiemos el lado derecho de la expresión (2.20):

$$\lambda^2\|L(x - x_c)\|^2 + \|b - Mx\|^2. \quad (2.21)$$

Se escribe, desarrollando la definición de la norma como:

$$\lambda^2(x - x_c)^T L^T L(x - x_c) + (b - Mx)^T (b - Mx); \quad (2.22)$$

utilizando la derivada para encontrar el argumento mínimo, obtenemos:

$$\frac{d}{dx_k} (\lambda^2(x - x_c)^T L^T L(x - x_c) + (b - Mx)^T (b - Mx)) = 0; \quad (2.23)$$

esto es:

$$2\lambda^2 L^T L(x - x_c) - 2M^T (b - Mx) = 0; \quad (2.24)$$

ahora resolvamos el sistema para encontrar una solución x .

$$\begin{aligned} 2\lambda^2 L^T L(x - x_c) - 2M^T(b - Mx) = 0 &\Rightarrow 2\lambda^2 L^T Lx - 2\lambda^2 Lx_c - 2M^T b + 2M^T Mx = 0; \\ &\Rightarrow (2\lambda^2 L^T L + 2M^T M)x = 2\lambda^2 Lx_c + 2M^T b; \\ &\Rightarrow (\lambda^2 L^T L + M^T M)x = \lambda^2 Lx_c + M^T b. \end{aligned}$$

y así obtendremos una solución única del sistema si $\lambda^2 L^T L + M^T M$ es no singular. Luego la solución de Tikhonov esta dada por:

$$x_\lambda = (M^T M + \lambda^2 L^T L)^{-1} M^T b. \quad (2.25)$$

2.4.3. La Descomposición Valor Singular y El método de Regularización de Tikhonov

Dada la descomposición valor singular de la matriz M como (2.14), la forma de la solución dada en (2.25) y la inversa de Moore-Penrose se tiene que en el problema de mínimos cuadrados determinado por el planteamiento de Tikhonov para la resolución de problemas inversos mal puestos, x_λ esta dada por:

$$x_\lambda = \sum_{i=1}^r f_i \frac{u_i^T b}{\sigma_i} v_i; \quad (2.26)$$

donde f_1, \dots, f_n son los factores de filtro de Tikhonov, el cual depende de σ_i y λ tal que:

$$f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \approx \begin{cases} 1, & \sigma_i \gg \lambda \\ \frac{\sigma_i^2}{\lambda^2}, & \sigma_i \ll \lambda \end{cases} \quad (2.27)$$

[5] afirma que particularmente la solución de mínimos cuadrados x_{LSQ} también viene dada por (2.26), cuando $\lambda = 0$ y siguiendo (2.27) se tienen que los factores de filtro de Tikhonov son iguales a uno, es decir, la solución de mínimos cuadrados satisface la siguiente ecuación normal:

$$Mx = b \implies M^T Mx = M^T b.$$

de acá utilizando (1.22) y la inversa de Moore-Penrose se obtiene:

$$x_{LSQ} = \sum_{k=1}^r \frac{u_k^T b}{\sigma_k} v_k. \quad (2.28)$$

De esta forma, se hace una comparación entre la x_λ de Tikhonov, con la x_{LSQ} del problema de mínimos cuadrados, y concluyen que los factores de filtro de Tikhonov son los encargados filtrar las contribuciones de los valores singulares pequeño a la solución x_λ , y a su vez dejando las componentes contribuyentes de los valores singulares grandes intactas.

Por otra parte el vector residual $(b - Mx_\lambda)$ el cual lleva consigo el ruido, utilizando la descomposición valor singular, (2.26) y se puede ver como:

$$b - Mx_\lambda = \sum_{k=1}^r (1 - f_k) u_k^T b u_k + b_0. \quad (2.29)$$

esto se obtiene de:

$$\begin{aligned} b - Mx_\lambda &= \sum_{j=1}^n u_j u_j^T b - \sum_{k=1}^r \sigma_k u_k v_k^T \left(\sum_{i=1}^r f_i \frac{u_i^T b}{\sigma_i} u_i \right); \\ &= \left(\sum_{j=1}^r u_j u_j^T b + \sum_{j=r+1}^n u_j u_j^T b \right) - \sum_{k=1}^r \sigma_k u_k v_k^T \left(f_k \frac{u_k^T b}{\sigma_k} u_k \right); \\ &= \sum_{j=r+1}^n u_j u_j^T b + \left(\sum_{k=1}^r u_k u_k^T b - \sigma_k u_k v_k^T \left(f_k \frac{u_k^T b}{\sigma_k} u_k \right) \right); \\ &= \sum_{j=r+1}^n u_j u_j^T b + \left(\sum_{k=1}^r u_k u_k^T b - (u_k f_k u_k^T b) \right); \\ &= \sum_{j=r+1}^n u_j u_j^T b + \left(\sum_{k=1}^r (1 - f_k) u_k u_k^T b \right); \\ &= \sum_{k=1}^r (1 - f_k) u_k u_k^T b + b_0. \end{aligned}$$

en donde el vector $b_0 = b - \sum_{k=1}^r u_k u_k^T b$ la cual representa las componentes de b que están fuera del rango de M , y $1 - f_k = \frac{\lambda^2}{(\sigma_k^2 + \lambda^2)}$. Finalmente obtendremos las dos

expresiones básicas para el análisis que se hará de la Curva-L en el próximo capítulo, en términos de la descomposición valor singular:

$$\|x_\lambda\|^2 = (x_\lambda)^T(x_\lambda); \quad (2.30)$$

Para facilitar los cálculos, denotaremos $\beta_k = \|u_k^T b\|$. Además considerando esto se tiene que:

$$\begin{aligned} (x_\lambda)^T &= \left(\sum_{k=1}^r f_k \frac{b_k}{\sigma_k} v_k \right)^T; \\ &= \sum_{k=1}^r f_k \frac{b_k^T}{\sigma_k} v_k^T. \end{aligned}$$

De esta forma:

$$\begin{aligned} \|x_\lambda\|^2 = (x_\lambda)^T(x_\lambda) &= \sum_{k=1}^r f_k \frac{b_k^T}{\sigma_k} v_k^T \left(\sum_{k=1}^r f_k \frac{b_k}{\sigma_k} v_k \right); \\ &= \sum_{k=1}^r \left(f_k \frac{b_k^T}{\sigma_k} v_k^T \right) \left(f_k \frac{b_k}{\sigma_k} v_k \right); \\ &= \sum_{k=1}^r \left(f_k^2 \frac{\beta^2}{\sigma_k^2} \right); \\ &= \sum_{k=1}^r \left(f_k \frac{\beta}{\sigma_k} \right)^2. \end{aligned}$$

luego:

$$\|x_\lambda\|^2 = \sum_{k=1}^r \left(f_k \frac{\|u_k^T b\|}{\sigma_k} \right)^2. \quad (2.31)$$

$$\begin{aligned}
 \|Mx_\lambda - b\|^2 &= (Mx_\lambda - b)^T (Mx_\lambda - b); \\
 &= \left(\sum_{k=1}^r (1 - f_k) b_k u_k \right)^T \left(\sum_{k=1}^r (1 - f_k) b_k u_k \right); \\
 &= \left(\sum_{k=1}^r (1 - f_k) u_k^T b_k^T \right) \left(\sum_{k=1}^r (1 - f_k) b_k u_k \right); \\
 &= \sum_{k=1}^r (1 - f_k)^2 \|u_k^T b\|^2; \\
 &= \sum_{k=1}^r ((1 - f_k) \|u_k^T b\|)^2.
 \end{aligned}$$

finalmente obtenemos:

$$\|Mx_\lambda - b\|^2 = \sum_{k=1}^r ((1 - f_k) \|u_k^T b\|)^2. \quad (2.32)$$

2.5. Análisis de la Descomposición valor Singular

Dado el problema:

$$\text{mín } \|b - Mx\|. \quad (2.33)$$

supongamos que los errores de este problema esta restringidos solo a lado derecho, es decir, podemos escribir los datos en los siguientes términos:

$$b = \bar{b} + n \quad \bar{b} = M\bar{x}. \quad (2.34)$$

donde \bar{b} representa la data no perturbada y $\bar{x} = M^\dagger \bar{b}$ representa la solución exacta y donde los errores en los datos están representados por el vector n , la solución proporcionada por Tikhonov puede ser escrita como:

$$x_\lambda = \bar{x}_\lambda + x_\lambda^n. \quad (2.35)$$

y donde la solución de Tikhonov para la parte no perturbada esta dada por:

$$\bar{x}_\lambda = (M^T + \lambda^2 L)^{-1} M^T \bar{b}. \quad (2.36)$$

y la solución para la parte de los errores que se consigue al aplicar regularización a ellos esta dada por:

$$x_\lambda^n = (M^T M + \lambda^2 L)^{-1} M^T. \quad (2.37)$$

Ahora bien consideremos la Curva-L generada solo por los datos no perturbados \bar{b} . Es así como en este punto se presenta la llamada Condición Discreta de Picard, que dice:

Los Coeficientes exactos de la Descomposición Valor Singular $|u_i^T \bar{b}|$ decaen más rápidamente que los valores singulares σ_i .

Esta condición nos asegura que la solución proporcionada por el esquema de mínimos cuadrados $\bar{x} = M \dagger \bar{b}$ del problema no perturbado, no tiene una norma grande, porque los coeficientes de la solución exacta $\|v_i^T \bar{x}\| = \|u_i^T \bar{b} / \sigma_i\|$ también decaen. A manera mas general esta condición se refiere a que la solución se puede aproximar por una solución regularizada con la condición de una adecuada selección del parámetro de regularización.

Supongamos que elegimos un parámetro de regularización entre σ_1 y σ_n tal que alguno de los factores de filtro son pequeños f_i y algunos están cercanos a uno. Sea k el número de factores de filtro cercanos a uno, entonces podemos ver que la relación entre k y λ esta dado por la siguiente expresión: $\lambda \simeq \sigma_k$ y de la expresión obtenida para la solución por defecto (2.31) obtenemos la siguiente expresión:

$$\|\bar{x}_\lambda\|^2 \simeq \sum_{i=1}^k (v_i^T \bar{x})^2 \simeq \sum_{i=1}^n (v_i^T \bar{x})^2 = \|\bar{x}\|^2. \quad (2.38)$$

Si tenemos que $\lambda \rightarrow \infty$ y $k \rightarrow 0$ se obtiene que $\bar{x}_\lambda \rightarrow 0$ de esta forma obtenemos que $\|\bar{x}_\lambda\| \rightarrow 0$ y en del otro lado se obtiene a partir de $\lambda \rightarrow 0$ entonces $\bar{x}_\lambda \rightarrow \bar{x}$ y finalmente se tiene que $\|\bar{x}_\lambda\| \rightarrow \|\bar{x}\|$.

Utilizando el hecho de que los coeficientes $\|v_i^T \bar{x}\|$ disminuyen tal que en las últimas $n - k$ términos casi no contribuyen a la suma. Notemos que (2.38) será valida

siempre y cuando λ no sea muy grande.

Ahora bien la Norma residual correspondiente a la solución regularizada no perturbada, \bar{x}_λ satisface lo siguiente:

$$\|M\bar{x}_\lambda - \bar{b}\|^2 \simeq \sum_{i=k}^n (u_i^T \bar{b})^2. \quad (2.39)$$

esto por la expresión (2.32) y el hecho de que k es el número de los factores de filtro cercanos a uno. El incremento de esta norma nos muestra que ella crece constantemente desde $\|\bar{b}_0\| = 0$ a $\|\bar{b}\|$ cuando λ incrementa. De esta forma la curva-L correspondiente al problema no perturbado es una curva plana sobre $\|\bar{x}_\lambda\| \simeq \|\bar{x}_\lambda\|$ excepto para valores grandes de la norma residual en donde la curva se aproxima al eje de las abscisas.

Finalmente consideramos la curva-L correspondiente al ruido del lado derecho $b = \bar{b} + n$. Dependiendo de λ , o de los componentes libre de ruido $u_i^T \bar{b}$, o las componentes que dominan el ruido puro $u_i^T n$, y el resultado consiste esencialmente en uno de los extremos de la curva-L.

CAPÍTULO 3

LA CURVA-L Y BÚSQUEDAS INEXACTAS EN LA CURVA-L

3.1. La Curva-L

3.1.1. Introducción

En capítulos previos hemos estudiado los problemas mal puestos y su solución (y el porqué de esta), se plantea mediante esquemas de regularización, los cuales buscan hacer un equilibrio entre el tamaño de la solución regularizada (o la diferencia de esta con una solución conocida), y la calidad del ajuste dado por los datos ([3]).

Se estudió a fondo el esquema de Regularización de Tikhonov que consiste en reemplazar el problema de mínimos cuadrados por:

$$\min_{x \in \mathbb{R}^n} \|b - Mx\|^2 + \lambda^2 \|L(x - x_c)\|^2. \quad (3.1)$$

donde λ se definió como el parámetro de regularización, y x_c es una solución que se conoce a priori, si por algún motivo no se tiene información sobre esta solución, entonces $x_c = 0$.

Un difícil problema es el de encontrar el parámetro de regularización, esto se debe a que casi todas las estrategias requieren de algún conocimiento previo del problema y de la solución [6].

El método para seleccionar el parámetro de regularización que pretendemos estudiar, esta basado en el gráfico de la norma de la solución regularizada versus la

correspondiente norma del residual, cuya forma, revela una L.

Lo que propone el criterio de esquina de la Curva-L, para la selección del parámetro de regularización, es utilizar la forma de “L” del gráfico de la curva, y localizar la esquina de dicha “L”.

La parte negativa de esto radica en que en general, conocer la esquina de la curva es una gran dificultad, puesto que la localización de la esquina en la Curva-L, depende de la escala con que la cual se grafiqué la misma.

Ahora bien se pretende hacer un estudio de la geometría de la Curva-L mediante criterios matemáticos, para así conocer sus propiedades y a partir de ahí utilizar eso y algunas otras herramientas para la búsqueda de la esquina de la Curva-L.

3.1.2. La Curva-L

Comencemos por conocer la parametrización que da origen a la Curva-L. Basados en (3.1) obtenemos las siguientes funciones auxiliares:

$$\tau(\lambda) = \|Mx_\lambda - b\|^2. \quad (3.2)$$

$$\eta(\lambda) = \|x_\lambda\|^2. \quad (3.3)$$

Donde x_λ es la solución de (3.1) la cual puede ser la solución del problema de mínimos cuadrados asociado.

Definición 3.1.1. Llamaremos Curva-L al conjunto de puntos dados por la siguiente parametrización:

$$\{\tau(\lambda), \eta(\lambda)\}_{\lambda \in (0, \infty)}. \quad (3.4)$$

Denotemos por L_ϕ la Curva-L representada mediante una escala genérica, así de la definición (3.1.1) se obtiene que la Curva-L puede definirse como:

$$L_\phi = \{\phi(\tau(\lambda)), \phi(\eta(\lambda))\}_{\lambda \in (0, \infty)}. \quad (3.5)$$

donde $\phi = \sqrt{\cdot}$ ó $\phi = \ln$ En el caso de este trabajo nos centraremos en el estudio de la Curva-L trabajada desde la escala logarítmica, puesto que en esta escala, la curva posee algunas propiedades geométricas que juegan a nuestro favor.

Recordemos las dos expresiones básicas para este análisis obtenidas en el capítulo anterior:

$$\|x_\lambda\|^2 = \sum_{k=1}^n \left(x_k \frac{u_k^T b}{\sigma_k} \right)^2 = \eta \quad (3.6)$$

$$\|Mx_\lambda - b\|^2 = \sum_{k=1}^n ((1 - x_k)u_k^T b)^2 = \tau \quad (3.7)$$

3.2. Análisis de la Curva-L

A continuación presentaremos algunos resultados obtenidos al hacer un estudio de la Curva-L que se encuentran en [6].

Lema 3.2.1. *Si $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}$ es una función positiva, creciente entonces para cualquier punto $(\hat{\tau}, \hat{\eta})$ en la L_ϕ son soluciones de las siguientes dos inecuaciones restrictas del problema de mínimos cuadrados.*

$$\hat{\tau} = \text{mín } \phi(\|Mx - b\|^2) \text{ sujeto a } \phi(\|L(x - x_c)\|^2) \leq \hat{\eta} \quad (3.8)$$

$$\hat{\eta} = \text{mín } \phi(\|L(x - x_c)\|^2) \text{ sujeto a } \phi(\|Mx - b\|^2) \leq \hat{\tau} \quad (3.9)$$

Demostración:

Dado $(\hat{\tau}, \hat{\eta}) \in L_\phi$, $\exists \lambda \in (0, \infty)$, y $x(\lambda)$ tal que

$$\hat{\tau} = \phi(\|Mx(\lambda) - b\|^2),$$

$$\hat{\eta} = \phi(\|x(\lambda)\|^2),$$

debemos probar que $x(\lambda)$ es minimizador (3.6) y (3.7). Notemos que si \hat{x} minimiza (3.8) entonces también minimiza

$$\text{mín } \|Mx - b\|^2 \quad (3.10)$$

$$s.a. \|x\|^2 \leq \|x(\lambda)\|^2 \quad (3.11)$$

ya que ϕ es creciente y positiva y $\hat{\eta} = \phi(\|x(\lambda)\|^2)$, similarmente, si \hat{x} minimiza (3.9) entonces también

$$\text{mín} \quad \|x\|^2 \tag{3.12}$$

$$\text{s.a.} \quad \|Mx - b\|^2 \leq \|Mx(\lambda) - b\|^2 \tag{3.13}$$

Supongamos por reducción al Absurdo que $x(\lambda)$ no minimiza (3.8) (respectivamente (3.9)), entonces $\exists \hat{x}$ minimizador de (3.10) tal que:

$$\|M\hat{x} - b\|^2 < \|Mx(\lambda) - b\|^2$$

y

$$\|\hat{x}\|^2 \leq \|x(\lambda)\|^2$$

luego

$$\|M\hat{x} - b\|^2 + \lambda^2 \|\hat{x}\|^2 < \|Mx(\lambda) - b\|^2 - \lambda^2 \|x(\lambda)\|^2;$$

lo cual es una contradicción ya que $x(\lambda)$ minimiza

$$\|Mx - b\|^2 - \lambda^2 \|x\|;$$

por (2.21) de igual forma para (3.9).

□

Comencemos el estudio buscando $\frac{dx}{d\lambda}$, utilizando (3.6), para ello primero encontremos una igualdad que nos permita trabajar la expresión término a término para su

derivación, por (2.32):

$$\begin{aligned}
\tau &= \sum_{k=1}^r ((1 - f_k) \|u_k^T b\|)^2; \\
&= \sum_{k=1}^r \left(\left(1 - \frac{\sigma_k^2}{(\sigma_k^2 + \lambda^2)} \right) \|u_k^T b\| \right)^2; \\
&= \sum_{k=1}^r \left(\left(\frac{\sigma_k^2 + \lambda^2 - \sigma_k^2}{\sigma_k^2 + \lambda^2} \right) \|u_k^T b\| \right)^2; \\
&= \sum_{k=1}^r \frac{\lambda^4 \|u_k^T b\|^2}{(\sigma_k^2 + \lambda^2)^2}; \\
&= \sum_{k=1}^r \frac{\lambda^4 \beta_k^2}{(\sigma_k^2 + \lambda^2)^2}.
\end{aligned}$$

Así

$$\frac{d\tau}{d\lambda} = \sum_{k=1}^r \frac{4\lambda^3 \beta_k^2 (\sigma_k^2 + \lambda^2)^2 - (2(\sigma_k^2 + \lambda^2) 2\lambda\lambda^4) \beta_k^2}{(\sigma_k^2 + \lambda^2)^4}; \quad (3.14)$$

$$= \sum_{k=1}^r \frac{4\lambda^3 \beta_k^2 (\sigma_k^2 + \lambda^2)^2 - (4\lambda^5 (\sigma_k^2 + \lambda^2) \beta_k^2)}{(\sigma_k^2 + \lambda^2)^4}; \quad (3.15)$$

$$= 4\lambda^3 \sum_{k=1}^r \frac{\beta_k^2 \sigma_k^2}{(\sigma_k^2 + \lambda^2)^3}; \quad (3.16)$$

$$= 4 \sum_{k=1}^r \frac{\lambda^3 \beta_k^2 \sigma_k^2}{(\sigma_k^2 + \lambda^2)^3}; \quad (3.17)$$

$$= 4 \sum_{k=1}^r \frac{1}{\lambda} \left(\frac{\lambda^4}{(\sigma_k^2 + \lambda^2)^2} \right) \left(\frac{\sigma_k^2}{(\sigma_k^2 + \lambda^2)} \right) \beta_k^2; \quad (3.18)$$

$$= 4 \sum_{k=1}^r \frac{1}{\lambda} \left(\frac{\lambda^2}{(\sigma_k^2 + \lambda^2)} \right)^2 \left(\frac{\sigma_k^2}{(\sigma_k^2 + \lambda^2)} \right) \beta_k^2; \quad (3.19)$$

$$= \frac{4}{\lambda} \sum_{k=1}^r (1 - f_k)^2 f_k \beta_k^2. \quad (3.20)$$

donde $\beta_k = \|u_k^T b\|$. Por otra parte notemos que en (3.16),

$$4\lambda^3 \sum_{k=1}^r \frac{\beta_k^2 \sigma_k^2}{(\sigma_k^2 + \lambda^2)^3} > 0. \quad (3.21)$$

esto indica que la función τ es estrictamente creciente, luego ella es invertible y se denota su inversa como $\lambda(\tau)$, y la curva $\{\tau(\lambda), \eta(\lambda)\}_{\lambda \in \mathbb{R}^+}$ tiene una única rama la cual puede ser representada como la función $\eta(\tau(\lambda)) = \eta(\lambda(\tau(\lambda)))$ y esto es lo que se utilizara para probar el siguiente lema.

Lema 3.2.2. *Si τ y η son definidas como en (3.6) y (3.7), entonces η como función de τ es decreciente y estrictamente convexa.*

Demostración:

La función $\tau : (0, \infty) \rightarrow (\|b\|^2, \|b_0\|^2)$, como se vio en (3.21) es una función que es creciente, su función inversa denotada como $\lambda(\tau)$ y definida como:

$$\lambda(\tau) : (\|b\|^2, \|b_0\|^2) \rightarrow (0, \infty).$$

también es creciente. Ahora calculemos $\frac{d\eta}{d\lambda}$, pero antes encontremos una expresión que permita su manipulación término a término:

$$\begin{aligned} \eta &= \sum_{k=1}^r \left(\frac{f_k \beta_k}{\sigma_k} \right)^2; \\ &= \sum_{k=1}^r \left(\frac{\sigma_k^2}{\sigma_k^2 + \lambda^2} \right)^2 \frac{\beta_k^2}{\sigma_k^2}; \\ &= \sum_{k=1}^r \frac{\sigma_k^4}{(\sigma_k^2 + \lambda^2)^2} \frac{\beta_k^2}{\sigma_k^2}; \\ &= \sum_{k=1}^r \frac{\sigma_k^2 \beta_k^2}{(\sigma_k^2 + \lambda^2)^2}. \end{aligned}$$

$$\frac{d\eta}{d\lambda} = \sum_{k=1}^r \sigma_k^2 \beta_k^2 (\sigma_k^2 + \lambda^2)^{-2}; \quad (3.22)$$

$$= \sum_{k=1}^r \sigma_k^2 \beta_k^2 (-2) (\sigma_k^2 + \lambda^2)^{-3} 2\lambda; \quad (3.23)$$

$$= -4 \sum_{k=1}^r \frac{\lambda \sigma_k^2 \beta_k^2}{(\sigma_k^2 + \lambda^2)^3}; \quad (3.24)$$

$$= -4 \sum_{k=1}^r \frac{1}{\lambda} \left(\frac{\lambda^2}{(\sigma_k^2 + \lambda^2)} \right) \left(\frac{\sigma_k^4}{(\sigma_k^2 + \lambda^2)^2} \right) \frac{\beta_k^2}{\sigma_k^2}; \quad (3.25)$$

$$= -4 \sum_{k=1}^r \frac{1}{\lambda} \left(\frac{\lambda^2}{(\sigma_k^2 + \lambda^2)} \right) \left(\frac{\sigma_k^2}{(\sigma_k^2 + \lambda^2)} \right)^2 \frac{\beta_k^2}{\sigma_k^2}; \quad (3.26)$$

$$= -\frac{4}{\lambda} \sum_{k=1}^r (1 - f_k) f_k^2 \frac{\beta_k^2}{\sigma_k^2}. \quad (3.27)$$

donde $\beta_k = \|u_k^T b\|$. Por otra parte observemos que:

$$-\frac{4}{\lambda} \sum_{k=1}^r (1 - f_k) f_k^2 \frac{\beta_k^2}{\sigma_k^2} < 0. \quad (3.28)$$

Por lo que η es una función decreciente. Pretendemos estudiar la razón de cambio

$$\frac{d\eta}{d\tau}$$

$$\frac{d\eta}{d\tau} = \frac{d\eta}{d\lambda} \frac{d\lambda}{d\tau} = -\frac{1}{\lambda^2}. \quad (3.29)$$

Luego η como función de τ es una función decreciente se tiene que:

$$\text{Si } \tau \rightarrow \|b_0\|^2 \text{ entonces } \frac{d\eta}{d\tau} \rightarrow -\infty. \quad (3.30)$$

y esto ocurre dado que

$$\text{Si } \lambda \rightarrow 0 \text{ entonces } \tau \rightarrow \|b_0\|^2$$

también

$$\text{Si } \tau \rightarrow \|b\|^2 \text{ entonces } \frac{d\eta}{d\tau} \rightarrow 0$$

porque

$$\text{Si } \lambda \rightarrow \infty \text{ Entonces } \tau \rightarrow \|b\|^2$$

Así tenemos que:

$$\frac{d^2\eta}{d\tau^2} = \frac{d\eta}{d\tau} \left(\frac{d\eta}{d\tau} \right) = -\frac{d\eta}{d\tau} \frac{1}{\lambda^2} = \frac{2}{d\lambda(\tau)} \tau > 0.$$

ya que $\lambda(\tau)$ es una función creciente. Luego Por la caracterización de funciones convexas y su Hessiano, η como función de τ es una función estrictamente convexa.

□

De esta forma obtenemos que la Curva L, es una curva Decreciente en cualquier escala diferenciable y estrictamente monótona ϕ y siguiendo (3.5) obtenemos que:

$$\frac{d\hat{\eta}}{d\hat{\tau}} = -\frac{1}{\lambda^2} \frac{\phi'(\eta(\lambda))}{\phi'(\tau(\lambda))}. \quad (3.31)$$

donde $\phi = \sqrt{\cdot}$ ó $\phi = \ln$,y además definimos:

$$\hat{\eta} = \ln \eta \quad (3.32)$$

$$\hat{\tau} = \ln \tau \quad (3.33)$$

Lema 3.2.3. *Si $b_o = 0$, La curva L en la escala logarítmica es estrictamente cóncava para $\lambda \in (0, \sigma_r)$ y para $\lambda \in (\sigma_1, \infty)$.*

Demostración:

Dada la expresión (3.5), obtenemos que la razón de cambio entre $\hat{\eta}$ y $\hat{\tau}$ esta dada por:

$$\frac{d\hat{\eta}}{d\hat{\tau}} = -\frac{1}{\lambda^2} \frac{\tau}{\eta}. \quad (3.34)$$

donde $\tau = e^{\hat{\tau}}$ y $\eta = e^{\hat{\eta}}$. Derivando (3.34) obtenemos:

$$\frac{d^2 \hat{\eta}}{d\hat{\tau}^2} = \frac{d\hat{\eta}^2}{d\lambda} \frac{d\lambda}{d\hat{\tau}}; \quad (3.35)$$

$$= \left(\frac{-e^{\hat{\tau}} \lambda^2 e^{\hat{\eta}} \frac{d\hat{\tau}}{d\lambda} + e^{\hat{\tau}} 2\lambda e^{\hat{\eta}} + e^{\hat{\tau}} \lambda^2 e^{\hat{\eta}} \frac{d\hat{\eta}}{d\lambda}}{(\lambda^2 e^{\hat{\eta}})^2} \right) \left(\frac{d\lambda}{d\hat{\tau}} \right); \quad (3.36)$$

$$= \frac{-e^{\hat{\tau}} \lambda^2 e^{\hat{\eta}} + 2\lambda e^{\hat{\tau}} \frac{1}{\frac{d\hat{\tau}}{d\lambda}} + \lambda^2 e^{\hat{\tau}} e^{\hat{\eta}} \frac{d\hat{\eta}}{d\lambda}}{(\lambda^2 e^{\hat{\eta}})^2}; \quad (3.37)$$

$$= \frac{-\lambda^2 e^{\hat{\tau}} e^{\hat{\eta}}}{\lambda^4 e^{2\hat{\eta}}} + \frac{2\lambda e^{\hat{\tau}} e^{\hat{\eta}} \frac{1}{\frac{d\hat{\tau}}{d\lambda}}}{\lambda^4 e^{2\hat{\eta}}} + \frac{\lambda^2 e^{\hat{\tau}} e^{\hat{\eta}}}{(\lambda^4 e^{2\hat{\eta}})} \frac{-1 e^{\hat{\tau}}}{\lambda^2 e^{\hat{\eta}}}; \quad (3.38)$$

$$= \frac{-e^{\hat{\tau}}}{\lambda^2 e^{\hat{\eta}}} + \frac{2e^{\hat{\eta}} \frac{1}{\frac{d\hat{\tau}}{d\lambda}}}{\lambda^3 e^{\hat{\eta}}} - \frac{e^{2\hat{\tau}}}{\lambda^4 e^{2\hat{\eta}}}; \quad (3.39)$$

$$= -\frac{\tau}{\lambda^2 \eta} - \frac{\tau^2}{\lambda^4 \eta^2} + \frac{2\tau \frac{1}{\frac{d\hat{\tau}}{d\lambda}}}{\lambda^3 \eta}; \quad (3.40)$$

$$= -\frac{\tau}{\lambda^4 \eta^2} \left[\lambda^2 \eta + \tau - 2\lambda \eta \frac{1}{\frac{d\hat{\tau}}{d\lambda}} \right]. \quad (3.41)$$

Ahora bien por la expresión obtenida en (3.16) tenemos:

$$\frac{d\hat{\tau}}{d\lambda} = \frac{1}{\tau} \frac{d\tau}{d\lambda}; \quad (3.42)$$

$$= \frac{4\lambda^3}{\tau} \delta. \quad (3.43)$$

donde llamaremos

$$\delta = \sum_{k=1}^r \frac{\beta_k^2 \sigma_k^2}{(\sigma_k^2 + \lambda^2)^3}. \quad (3.44)$$

luego a partir de esta última expresión, sustituyéndola en (3.41) obtenemos lo si-

guiente:

$$\frac{-\tau}{\lambda^4\eta^2} \left[\lambda^2\eta + \tau - \frac{2\lambda\eta}{\frac{d\hat{r}}{d\lambda}} \right] = \frac{2\tau}{\lambda^3\eta} \frac{1}{\frac{d\hat{r}}{d\lambda}} - \frac{\tau}{\lambda^2\eta} - \frac{\tau^2}{\lambda^4\eta}; \quad (3.45)$$

$$= \frac{-\tau}{\lambda^2\eta} - \frac{\tau^2}{\lambda^4\eta^2} + \frac{2\tau}{\lambda^3\eta \frac{4\lambda^3}{\tau} \delta}; \quad (3.46)$$

$$= \frac{-\tau}{\lambda^2\eta} - \frac{\tau^2}{\lambda^4\eta^2} + \frac{\tau^2}{2\lambda^6\eta\delta}; \quad (3.47)$$

$$= -\frac{\tau^2}{\lambda^6\eta} \left[\frac{\lambda^4}{\tau} + \frac{\lambda^2}{\eta} - \frac{1}{2\delta} \right]. \quad (3.48)$$

Ahora bien estudiemos la expresión (3.48), el signo de ella esta determinado por el signo de $\frac{1}{2\delta}$, observemos que para $\lambda \leq \sigma_r$ tenemos que $\sigma_k^2 + \lambda^2 \leq \sigma_k^2 + \sigma_r^2 \leq 2\sigma_k^2$, esto recordando que $\sigma_r \leq \sigma_k$ para $\forall k$ así, el cociente:

$$\frac{2\sigma_k^2}{\sigma_k^2 + \lambda^2} \geq 1 \quad k = 1, \dots, r. \quad (3.49)$$

se tiene que $2\delta \geq \frac{\tau}{\lambda^4}$, utilizando (2.27), en efecto;

$$\begin{aligned} 2\delta &= \sum_{k=1}^r \frac{2\beta_k^2\sigma_k^2}{(\sigma_k^2 + \lambda^2)^3}; \\ &= \sum_{k=1}^r \frac{2\beta_k^2\sigma_k^2}{(\sigma_k^2 + \lambda^2)^2(\sigma_k^2 + \lambda^2)}; \\ &> \sum_{k=1}^r \frac{\beta^2}{(\sigma_k^2 + \lambda^2)^2} \\ &= \frac{\tau}{\lambda^4}. \end{aligned}$$

de igual forma tenemos que si $\lambda \geq \sigma_1$, así por (2.27),

$$\frac{2}{\sigma_k^2 + \lambda^2} \geq \frac{1}{\lambda^2}, \quad i = 1, \dots, r$$

por lo que, $2\delta \geq \frac{\eta}{\lambda^2}$, en efecto;

$$\begin{aligned} 2\delta &= \sum_{k=1}^r \frac{2\beta_k^2 \sigma_k^2}{(\sigma_k^2 + \lambda^2)^3}; \\ &\geq \sum_{k=1}^r \frac{\beta_k^2 \sigma_k^2}{(\sigma_k^2 + \lambda^2)^2} \frac{1}{\lambda^2} \\ &= \frac{\eta}{\lambda^2} \end{aligned}$$

todo esto implica que:

$$\frac{\lambda^4}{\tau} + \frac{\lambda^2}{\eta} - \frac{1}{2\delta} > 0.$$

de donde se tiene que

$$\frac{d^2\eta}{d\tau^2} < 0.$$

de igual forma se verifica para el caso $\lambda \geq \sigma_1$ considerando $2\delta \geq \frac{\eta}{\lambda^2}$, por lo que para ambos casos:

$$\frac{d^2\eta}{d\tau^2} < 0.$$

por lo que la curva en escala logarítmica es cóncava para $\lambda \leq \sigma_r$ y para $\lambda \geq \sigma_1$.

□

El siguiente lema da una condición suficiente (en el caso de dos componentes de b) sobre la cual la curva L en escala logarítmica no posea la propiedad de tener esquina.

Lema 3.2.4. *Sea $i_1 < i_2$ tales que $\beta_i \neq 0$ Para $i \in \{i_1, i_2\}$ y $\beta_i = 0$ para $i \notin \{i_1, i_2\}$. Si*

$$\frac{\beta_{i_2}}{\sigma_{i_2}^2} < \frac{\beta_{i_1}}{\sigma_{i_1}^2}. \tag{3.50}$$

entonces la curva L en escala logarítmica es estrictamente cóncava.

Primero demostraremos la siguiente proposición, que se utilizara en la demostración del Lema:

Proposición 3.2.1. *Si $\sigma_{i_1}, \sigma_{i_2}, \beta_{i_1}, \beta_{i_2} > 0$ y $\sigma_{i_2} \leq \lambda \leq \sigma_{i_1}$ y $\frac{\beta_{i_2}}{\sigma_{i_2}^2} < \frac{\beta_{i_1}}{\sigma_{i_1}^2}$, entonces $\beta_{i_1} > \beta_{i_2}$*

Demostración:

En efecto, dado que:

$$\sigma_{i2} \leq \sigma_{i1} \Rightarrow \frac{1}{\sigma_{i1}} \leq \frac{1}{\sigma_{i2}}$$

Así

$$\begin{aligned} \frac{\beta_{i2}}{\sigma_{i2}^2} &< \frac{\beta_{i1}}{\sigma_{i1}^2} \Rightarrow \frac{\beta_{i2}}{\sigma_{i1}^2} < \frac{\beta_{i1}}{\sigma_{i1}^2} \\ &\Rightarrow \beta_{i2} < \beta_{i1} \end{aligned}$$

por lo tanto $\beta_{i2} < \beta_{i1}$

□

Ahora procederemos con la prueba del Lema.

Demostración:

Por el lema anterior tenemos que la curva L en escala logarítmica es estrictamente cóncava para $\lambda \leq \sigma_{i2}$ y para $\lambda \geq \sigma_{i1}$, así que solo bastara probar que es estrictamente cóncava para $\sigma_{i2} \leq \lambda \leq \sigma_{i1}$.

Sea $\{\hat{\tau}, \hat{\eta}\} \in L_{\ln}$, donde según se definió, se tiene que $\hat{\tau} = \ln \tau(\lambda)$ y $\hat{\eta} = \ln \eta(\lambda)$, calculemos $\frac{d\hat{\tau}}{d\hat{\eta}}$:

$$\begin{aligned} \frac{d\hat{\tau}}{d\hat{\eta}} &= \frac{\frac{d\hat{\tau}}{d\lambda}}{\frac{d\hat{\eta}}{d\lambda}}, \\ &= \frac{\frac{4\lambda^3}{\tau} \left(\frac{\beta_{i1}^2 \sigma_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^3} \right)}{\frac{-4\lambda}{\eta} \left(\frac{\sigma_{i2}^2 \beta_{i2}^2}{(\sigma_{i2}^2 + \lambda^2)^3} \right)}, \\ &= -\lambda^2 \frac{\eta}{\tau}. \end{aligned}$$

Ahora calculemos $\frac{d^2 \hat{\tau}}{d\hat{\eta}^2}$, tomando en cuenta que $\tau = e^{\hat{\tau}}$ y $\eta = e^{\hat{\eta}}$

$$\begin{aligned} \frac{d^2 \hat{\tau}}{d\hat{\eta}^2} &= \left[\frac{(-2\lambda e^{\hat{\eta}} - \lambda^2 e^{\hat{\eta}} \frac{d\hat{\eta}}{d\lambda}) e^{\hat{\tau}} + \lambda^2 e^{\hat{\eta}} e^{\hat{\tau}} \frac{d\hat{\tau}}{d\lambda}}{e^{2\hat{\tau}}} \right]; \\ &= \left[\frac{-2\lambda e^{\hat{\eta}} e^{\hat{\tau}} - \lambda^2 e^{\hat{\eta}} e^{\hat{\tau}} \frac{d\hat{\eta}}{d\lambda} + \lambda^2 e^{\hat{\eta}} e^{\hat{\tau}} \frac{d\hat{\tau}}{d\lambda}}{e^{2\hat{\tau}}} \right]; \end{aligned}$$

así

$$\begin{aligned}
\frac{d^2\hat{\tau}}{d\hat{\eta}^2} &= \left[\frac{-2\lambda e^{\hat{\eta}} e^{\hat{\tau}} - \lambda^2 e^{\hat{\eta}} e^{\hat{\tau}} \frac{d\hat{\eta}}{d\lambda} + \lambda^2 e^{\hat{\eta}} e^{\hat{\tau}} \frac{d\hat{\tau}}{d\lambda}}{e^{2\hat{\tau}}} \right] \frac{d\lambda}{d\hat{\eta}}; \\
&= \frac{-2\lambda e^{\hat{\eta}} e^{\hat{\tau}} \frac{1}{\frac{d\hat{\eta}}{d\lambda}} - \lambda^2 e^{\hat{\eta}} e^{\hat{\tau}} + \lambda^2 e^{\hat{\eta}} e^{\hat{\tau}} \left(-\lambda^2 \frac{e^{\hat{\eta}}}{e^{\hat{\tau}}}\right)}{e^{2\hat{\tau}}}; \\
&= -\frac{2\lambda e^{\hat{\eta}} \frac{1}{\frac{d\hat{\eta}}{d\lambda}}}{e^{\hat{\tau}}} - \frac{\lambda^2 e^{\hat{\eta}}}{e^{\hat{\tau}}} - \frac{\lambda^4 e^{2\hat{\eta}}}{e^{2\hat{\tau}}}; \\
&= -\frac{2\lambda\eta}{\tau} \frac{1}{\frac{d\hat{\eta}}{d\lambda}} - \frac{\lambda^2\eta}{\tau} - \frac{\lambda^4\eta^2}{\tau^2}.
\end{aligned}$$

ahora bien

$$\begin{aligned}
\frac{d\hat{\eta}}{d\lambda} &= \frac{1}{\eta} \frac{d\eta}{d\lambda}; \\
&= -\frac{1}{\eta} \left(\frac{\lambda\sigma_{i1}^2\beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^3} + \frac{\lambda\sigma_{i2}^2\beta_{i2}^2}{(\sigma_{i2}^2 + \lambda^2)^3} \right); \\
&= -\frac{4\lambda z}{\eta}.
\end{aligned}$$

donde $z = \frac{\lambda\sigma_{i1}^2\beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^3} + \frac{\lambda\sigma_{i2}^2\beta_{i2}^2}{(\sigma_{i2}^2 + \lambda^2)^3}$ sustituyendo este resultado obtenemos:

$$\begin{aligned}
\frac{d^2\hat{\tau}}{d\hat{\eta}^2} &= -\frac{2\lambda\eta}{\tau} \frac{1}{-\frac{4\lambda z}{\eta}} - \frac{\lambda^2\eta}{\tau} - \frac{\lambda^4\eta^2}{\tau^2}; \\
&= -\left[\lambda^2 \frac{\eta}{\tau} + \frac{\lambda^4\eta^2}{\tau^2} - \frac{\eta^2}{2\tau z} \right]; \\
&= -\frac{\eta}{\tau} \left[\frac{\lambda^2}{\eta} + \frac{\lambda^4}{\tau} - \frac{1}{2z} \right].
\end{aligned}$$

Ahora debemos verificar que dada la hipótesis, se cumple:

$$\frac{\lambda^2}{\eta} + \frac{\lambda^4}{\tau} - \frac{1}{2z} > 0.$$

Recordemos que para este caso:

$$\begin{aligned}
\eta &= \frac{\sigma_{i1}^2\beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^2} + \frac{\sigma_{i2}^2\beta_{i2}^2}{(\sigma_{i2}^2 + \lambda^2)^2}. \\
\tau &= \frac{\lambda^4\beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^2} + \frac{\lambda^4\beta_{i2}^2}{(\sigma_{i2}^2 + \lambda^2)^2}.
\end{aligned}$$

$$z = \frac{\lambda \sigma_{i1}^2 \beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^3} + \frac{\lambda \sigma_{i2}^2 \beta_{i2}^2}{(\sigma_{i2}^2 + \lambda^2)^3}.$$

Ahora acotemos inferiormente cada uno de los términos de esta suma

$$\begin{aligned} \frac{\lambda^4}{\tau} &= \frac{\lambda^4}{\frac{\lambda^4 \beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^2} + \frac{\lambda^4 \beta_{i2}^2}{(\sigma_{i2}^2 + \lambda^2)^2}}; \\ &= \frac{1}{\frac{\beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^2} + \frac{\beta_{i2}^2}{(\sigma_{i2}^2 + \lambda^2)^2}}; \\ &= \frac{(\sigma_{i1}^2 + \lambda^2)^2 (\sigma_{i2}^2 + \lambda^2)^2}{\beta_{i1}^2 (\sigma_{i2}^2 + \lambda^2)^2 + \beta_{i2}^2 (\sigma_{i1}^2 + \lambda^2)^2}; \\ &> \frac{(\sigma_{i2}^2 + \lambda^2)^2 (\sigma_{i2}^2 + \lambda^2)^2}{\beta_{i1}^2 (\sigma_{i1}^2 + \lambda^2)^2 + \beta_{i1}^2 (\sigma_{i1}^2 + \lambda^2)^2}; \\ &> \frac{(2\sigma_{i2}^2)^4}{2\beta_{i1}^2 (2\sigma_{i1}^2)^2}; \\ &= \frac{2\sigma_{i2}^8}{\beta_{i1}^2 \sigma_{i1}^4}. \end{aligned}$$

$$\begin{aligned} \frac{\lambda^2}{\eta} &= \frac{\lambda^2}{\frac{\sigma_{i1}^2 \beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^2} + \frac{\sigma_{i2}^2 \beta_{i2}^2}{(\sigma_{i2}^2 + \lambda^2)^2}}; \\ &= \frac{\lambda^2 (\sigma_{i1}^2 + \lambda^2)^2 (\sigma_{i2}^2 + \lambda^2)^2}{\sigma_{i1}^2 \beta_{i1}^2 (\sigma_{i2}^2 + \lambda^2)^2 + (\sigma_{i1}^2 + \lambda^2)^2 \sigma_{i2}^2 \beta_{i2}^2}; \\ &> \frac{\sigma_{i2}^2 (\sigma_{i2}^2 + \sigma_{i2}^2)^2 (\sigma_{i2}^2 + \sigma_{i2}^2)^2}{\sigma_{i1}^2 \beta_{i1}^2 (\sigma_{i1}^2 + \sigma_{i1}^2)^2 + (\sigma_{i1}^2 + \sigma_{i1}^2)^2 \sigma_{i1}^2 \beta_{i1}^2}; \\ &> \frac{2\sigma_{i2}^{10}}{\sigma_{i1}^6 \beta_{i1}^2}. \end{aligned}$$

$$\begin{aligned} -z &= -\frac{\sigma_{i1}^2 \beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^3} - \frac{\sigma_{i2}^2 \beta_{i2}^2}{(\sigma_{i2}^2 + \lambda^2)^3}; \\ &= \frac{-\sigma_{i1}^2 \beta_{i1}^2 (\sigma_{i1}^2 + \lambda^2)^3 - \sigma_{i2}^2 \beta_{i2}^2 (\sigma_{i1}^2 + \lambda^2)^3}{(\sigma_{i1}^2 + \lambda^2)^3 (\sigma_{i2}^2 + \lambda^2)^3}; \\ &< \frac{-\sigma_{i2}^2 \beta_{i2}^2 (\sigma_{i2}^2 + \sigma_{i2}^2)^3 - \sigma_{i2}^2 \beta_{i2}^2 (\sigma_{i2}^2 + \sigma_{i2}^2)^3}{(\sigma_{i1}^2 + \sigma_{i1}^2)^3 (\sigma_{i1}^2 + \sigma_{i1}^2)^3}; \\ &< -\frac{1}{4} \frac{\beta_{i2}^2 \sigma_{i2}^8}{\sigma_{i1}^{12}}. \end{aligned}$$

de acá que:

$$\begin{aligned} -2z < -\frac{1}{2} \frac{\beta_{i_2}^2 \sigma_{i_2}^8}{\sigma_{i_1}^{12}} &\Rightarrow \frac{-2\sigma_{i_1}^{12}}{\beta_{i_2}^2 \sigma_{i_2}^8} < \frac{-1}{2z}; \\ &\Rightarrow -\frac{1}{2z} > -\frac{2\sigma_{i_1}^{12}}{\beta_{i_2}^2 \sigma_{i_2}^8}. \end{aligned}$$

finalmente tenemos que:

$$\begin{aligned} \frac{\lambda^2}{\eta} + \frac{\lambda^4}{\tau} - \frac{1}{2z} &> \frac{2\sigma_{i_2}^{10}}{\sigma_{i_1}^6 \beta_{i_1}^2} + \frac{2\sigma_{i_2}^8}{\beta_{i_1}^2 \sigma_{i_1}^4} - \frac{2\sigma_{i_1}^{12}}{\beta_{i_2}^2 \sigma_{i_2}^8}; \\ &= \frac{2\sigma_{i_2}^{10} \beta_{i_2}^2 \sigma_{i_2}^8 + 2\sigma_{i_2}^8 \sigma_{i_1}^2 \beta_{i_2}^2 \sigma_{i_2}^8 - 2\sigma_{i_1}^{12} \sigma_{i_1}^6 \beta_{i_1}^2}{\sigma_{i_1}^6 \beta_{i_1}^2 \beta_{i_2}^2 \sigma_{i_2}^8}; \\ &= \frac{2\sigma_{i_2}^{18} \beta_{i_2}^2 + 2\sigma_{i_2}^{16} \sigma_{i_1}^2 \beta_{i_2}^2 - 2\sigma_{i_1}^{18} \beta_{i_1}^2}{\sigma_{i_1}^6 \beta_{i_1}^2 \beta_{i_2}^2 \sigma_{i_2}^8}. \end{aligned}$$

Nos interesa el signo de este numerador entonces tenemos que

$$\begin{aligned} \sigma_{i_1} > \sigma_{i_2} \quad y \quad \beta_{i_1} > \beta_{i_2} &\Rightarrow \sigma_{i_1} \beta_{i_1} > \sigma_{i_2} \beta_{i_2}; \\ &\Rightarrow -\sigma_{i_1} \beta_{i_1} < -\sigma_{i_2} \beta_{i_2} < \sigma_{i_2} \beta_{i_2}; \\ &\Rightarrow -2\sigma_{i_1}^{18} \beta_{i_1}^2 < 2\sigma_{i_2}^{18} \beta_{i_2}^2. \end{aligned}$$

por lo que:

$$2\sigma_{i_2}^{18} \beta_{i_2}^2 + 2\sigma_{i_2}^{16} \sigma_{i_1}^2 \beta_{i_2}^2 - 2\sigma_{i_1}^{18} \beta_{i_1}^2 > 0.$$

Así

$$\frac{2\sigma_{i_2}^{18} \beta_{i_2}^2 + 2\sigma_{i_2}^{16} \sigma_{i_1}^2 \beta_{i_2}^2 - 2\sigma_{i_1}^{18} \beta_{i_1}^2}{\sigma_{i_1}^6 \beta_{i_1}^2 \beta_{i_2}^2 \sigma_{i_2}^8} > 0.$$

finalmente

$$\frac{\lambda^2}{\eta} + \frac{\lambda^4}{\tau} - \frac{1}{2z} > 0.$$

Por lo tanto $\frac{d^2 \hat{\tau}}{d\hat{\eta}^2} < 0$, por lo que la Curva L es estrictamente cóncava.

□

Lema 3.2.5. *Sea $i_1 < i_2$ tal que $\beta_i \neq 0$ para $i \in \{i_1, i_2\}$ y $\beta_i = 0$ para $i \notin \{i_1, i_2\}$. Entonces la curva parametrizada*

$$L_{\sqrt{\cdot}} = \{\|Mx_\lambda - b\|, \|x_\lambda\|\}_{\lambda \in (0, \infty)}$$

es convexa.

Demostración:

Sea $\{\bar{\tau}, \bar{\eta}\} \in L_{\sqrt{\cdot}}$, donde $\bar{\tau} = \sqrt{\tau(\lambda)}$ y $\bar{\eta} = \sqrt{\eta(\lambda)}$, de (3.31), tenemos que:

$$\begin{aligned} \frac{d\bar{\eta}}{d\bar{\tau}} &= -\frac{1}{\lambda^2} \frac{\bar{\eta}'}{\bar{\tau}'}; \\ &= -\frac{1}{\lambda^2} \sqrt{\frac{\tau(\lambda)}{\eta(\lambda)}}. \end{aligned}$$

debemos estudiar ambos casos $\sigma_1 = \sigma_2$ y $\sigma_1 > \sigma_2$, supongamos que $\sigma_1 = \sigma_2$, entonces, dado que:

$$\begin{aligned} \eta &= \frac{\sigma_{i1}^2 \beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^2} + \frac{\sigma_{i2}^2 \beta_{i2}^2}{(\sigma_{i2}^2 + \lambda^2)^2}. \\ \tau &= \frac{\lambda^4 \beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^2} + \frac{\lambda^4 \beta_{i2}^2}{(\sigma_{i2}^2 + \lambda^2)^2}. \end{aligned}$$

$$\begin{aligned} \frac{d\bar{\eta}}{d\bar{\tau}} &= -\frac{1}{\lambda^2} \sqrt{\frac{\tau(\lambda)}{\eta(\lambda)}}; \\ &= -\frac{1}{\lambda^2} \sqrt{\frac{\frac{\lambda^4 \beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^2} + \frac{\lambda^4 \beta_{i2}^2}{(\sigma_{i1}^2 + \lambda^2)^2}}{\frac{\sigma_{i1}^2 \beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^2} + \frac{\sigma_{i1}^2 \beta_{i2}^2}{(\sigma_{i1}^2 + \lambda^2)^2}}}; \\ &= -\frac{1}{\lambda^2} \sqrt{\frac{\frac{\lambda^4 \beta_{i1}^2 + \lambda^4 \beta_{i2}^2}{(\sigma_{i1}^2 + \lambda^2)^2}}{\frac{\sigma_{i1}^2 (\beta_{i1}^2 + \beta_{i2}^2)}{(\sigma_{i1}^2 + \lambda^2)^2}}}; \\ &= -\frac{1}{\lambda^2} \sqrt{\frac{1}{\sigma_{i1}^2}}; \\ &= -\frac{1}{\sigma_{i1}}. \end{aligned}$$

De acá obtenemos que $\frac{d\bar{\eta}}{d\bar{\tau}}$ es constante, por lo que es inyectiva.

Ahora supongamos que $\sigma_{i1} > \sigma_{i2}$:

$$\begin{aligned}
\frac{d\bar{\eta}}{d\bar{\tau}} &= -\frac{1}{\lambda^2} \sqrt{\frac{\tau(\lambda)}{\eta(\lambda)}}; \\
&= -\frac{1}{\lambda^2} \sqrt{\frac{\frac{\lambda^4 \beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^2} + \frac{\lambda^4 \beta_{i2}^2}{(\sigma_{i1}^2 + \lambda^2)^2}}{\frac{\sigma_{i1}^2 \beta_{i1}^2}{(\sigma_{i1}^2 + \lambda^2)^2} + \frac{\sigma_{i1}^2 \beta_{i2}^2}{(\sigma_{i1}^2 + \lambda^2)^2}}}; \\
&= -\frac{1}{\lambda} \sqrt{\frac{\lambda^4 (\beta_{i1}^2 (\sigma_{i1}^2 + \lambda^2)^2 + \beta_{i2}^2 (\sigma_{i1}^2 + \lambda^2)^2)}{\sigma_{i1}^2 \beta_{i1}^2 (\sigma_{i2}^2 + \lambda^2)^2 + \sigma_{i2}^2 \beta_{i2}^2 (\sigma_{i1}^2 + \lambda^2)^2}}}. \\
&= -\sqrt{\frac{\beta_{i1}^2 (\sigma_{i1}^2 + \lambda^2)^2 + \beta_{i2}^2 (\sigma_{i1}^2 + \lambda^2)^2}{\sigma_{i1}^2 \beta_{i1}^2 (\sigma_{i2}^2 + \lambda^2)^2 + \sigma_{i2}^2 \beta_{i2}^2 (\sigma_{i1}^2 + \lambda^2)^2}}}.
\end{aligned}$$

Luego elevando al cuadrado ambos términos:

$$\left(\frac{d\bar{\eta}}{d\bar{\tau}}\right)^2 = \frac{\beta_{i1}^2 (\sigma_{i1}^2 + \lambda^2)^2 + \beta_{i2}^2 (\sigma_{i1}^2 + \lambda^2)^2}{\sigma_{i1}^2 \beta_{i1}^2 (\sigma_{i2}^2 + \lambda^2)^2 + \sigma_{i2}^2 \beta_{i2}^2 (\sigma_{i1}^2 + \lambda^2)^2}.$$

luego al resolver la ecuación se tiene que:

$$\beta_{i1}^2 (\sigma_{i1}^2 + \lambda^2)^2 \left[1 - \left(\frac{d\bar{\eta}}{d\bar{\tau}}\right)^2 \sigma_{i1}^2\right] = \beta_{i2}^2 (\sigma_{i1}^2 + \lambda^2)^2 \left[\sigma_{i2}^2 \left(\frac{d\bar{\eta}}{d\bar{\tau}}\right)\right].$$

como queremos probar que la derivada es una función inyectiva, supongamos que $\frac{d\bar{\eta}}{d\bar{\tau}}(\lambda_1) = \frac{d\bar{\eta}}{d\bar{\tau}}(\lambda_2)$, así se tiene que:

$$\frac{\beta_{i1}^2 (\sigma_{i2}^2 + \lambda_1^2)^2}{\beta_{i2}^2 (\sigma_{i1}^2 - \lambda_1^2)^2} = \frac{\beta_{i1}^2 (\sigma_{i2}^2 + \lambda_2^2)^2}{\beta_{i2}^2 (\sigma_{i1}^2 - \lambda_2^2)^2};$$

luego

$$\frac{2\sigma_{i1}^2 \lambda_1^2 + \lambda_1^4}{-2\sigma_{i1}^2 \lambda_1^2 + \lambda_1^4} = \frac{2\sigma_{i2}^2 \lambda_2^2 + \lambda_2^4}{-2\sigma_{i1}^2 \lambda_2^2 + \lambda_2^4};$$

esto implica que:

$$(2\sigma_{i1}^2 \lambda_1^2 + \lambda_1^4)(-2\sigma_{i1}^2 \lambda_2^2 + \lambda_2^4) = (2\sigma_{i2}^2 \lambda_2^2 + \lambda_2^4)(-2\sigma_{i1}^2 \lambda_1^2 + \lambda_1^4);$$

de acá que:

$$\lambda_2^4 (\sigma_{i2}^2 \lambda_1^2 + \sigma_{i1}^2 \lambda_1^2) = \lambda_1^4 (\sigma_{i2}^2 \lambda_2^2 + \lambda_2^2 \sigma_{i1}^2);$$

y esto ocurre solamente si $\lambda_1 = \lambda_2$.

Así $\frac{d\bar{\eta}}{d\bar{\tau}}$ es una función inyectiva, que es además continua.

Sea $f : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$ una función auxiliar, definida como:

$$f(t, s) = \frac{t + s}{t\sigma_{i1}^2 + s\sigma_{i2}^2};$$

para estudiar la parte interna de la función $\frac{d\bar{\eta}}{d\bar{\tau}}$, tomemos un punto $t_0 = \beta_{i1}^2 \frac{\sigma_{i2}^4}{\sigma_{i1}^4}$, y $s_0 = \beta_{i2}^2$, $t_k = \beta_{i1}^2$, de donde obtenemos que:

$$\lim_{\lambda \rightarrow 0} \frac{d\bar{\eta}}{d\bar{\tau}} = -\sqrt{f(t_0, s_0)};$$

$$\lim_{\lambda \rightarrow \infty} \frac{d\bar{\eta}}{d\bar{\tau}} = -\sqrt{f(t_k, s_0)};$$

como $\sigma_{i1} > \sigma_{i2}$ tenemos que, $\frac{df}{dt} < 0$ es decir f es una función que decrece en t . Como $t_0 < t_k$ tenemos que $f(t_0, s_0) > f(t_k, s_0)$, Así

$$\lim_{\lambda \rightarrow 0} \frac{d\bar{\eta}}{d\bar{\tau}} < \lim_{\lambda \rightarrow \infty} \frac{d\bar{\eta}}{d\bar{\tau}};$$

luego $\frac{d\bar{\eta}}{d\bar{\tau}}$ es creciente, por lo tanto $\bar{\eta}$ es una función estrictamente convexa.

□

3.2.1. Un Criterio de Mínimo Local

Consideremos el problema discreto mal puesto $\min_{x \in \mathbb{C}} \|b - Mx\|^2$ y la solución x_λ (solución del problema regularizado mediante Tikhonov).

Los Datos b están contaminado por los errores, mientras que \bar{b} satisface la condición discreta de Picard, sin embargo la perturbación $b - \bar{b}$ no necesariamente cumple con esta condición, como consecuencia existe una gran influencia de los errores $b_i - \bar{b}_i$ correspondiente a los valores singulares pequeños sobre la norma de la solución.

Cualquier $\lambda > 0$ reduce la norma de x , el problema es seleccionar un λ , tal que:

(a) $\eta(\lambda)$ no sea muy grande, $\eta(\lambda) \leq \eta(\lambda_0)$, $\lambda \geq \lambda_0$

(b) Que la norma Residual sea pequeña, esto es que, $\tau(\lambda) \leq \tau(\lambda_1)$, con $\lambda \leq \lambda_1$

Como $\tau(\lambda)$, crece cuando $\eta(\lambda)$ decrece en función de λ , consideremos la siguiente función:

$$\Psi_\alpha(\lambda) = \tau(\lambda)\eta^\alpha(\lambda). \quad (3.51)$$

donde $\alpha > 0$, y busquemos el mínimo sobre el Intervalo $[\lambda_0, \lambda_1]$.

Teorema 3.2.1. *La Función Ψ_α tiene mínimo local en $\bar{\lambda}$ si y solo si:*

(a) *La Curva L_{\ln} es tangente en el punto $\{\hat{\tau}(\bar{\lambda}), \hat{\eta}(\bar{\lambda})\}$, sobre la línea recta $\tau + \alpha\eta = \text{const.}$ y*

(b) *La L_{\ln} es convexa en una vecindad de $\{\hat{\tau}(\bar{\lambda}), \hat{\eta}(\bar{\lambda})\}$.*

Demostración:

Supongamos que la función Ψ_α tiene mínimo local en $\bar{\lambda}$, usando la notación $\hat{\tau} = \ln \tau$ y $\hat{\eta} = \ln \eta$, sustituyendo tenemos que:

$$\Psi_\alpha(\lambda) = e^{\hat{\tau}(\lambda) + \alpha\hat{\eta}(\lambda)}. \quad (3.52)$$

derivando respecto de λ tenemos:

$$\begin{aligned} \frac{d}{d\lambda} \Psi_\alpha(\lambda) &= e^{\hat{\tau}(\lambda) + \alpha\hat{\eta}(\lambda)} \frac{d}{d\lambda} (\hat{\tau}(\lambda) + \alpha\hat{\eta}(\lambda)); \\ &= e^{\hat{\tau}(\lambda) + \alpha\hat{\eta}(\lambda)} \frac{\hat{\tau}'(\lambda)}{d\lambda} + \alpha \frac{d\hat{\eta}(\lambda)}{d\lambda}. \end{aligned}$$

Como $\frac{\hat{\tau}'(\lambda)}{d\lambda} > 0$ se tiene que la condición para que $\bar{\lambda}$ sea un extremo local de Ψ_λ

$$\begin{aligned} \frac{d\hat{\tau}}{d\lambda} + \alpha \frac{d\hat{\eta}}{d\lambda} = 0 &\Rightarrow \frac{d\hat{\tau}}{d\lambda} = -\alpha \frac{d\hat{\eta}}{d\lambda}; \\ &\Rightarrow \frac{1}{\alpha} = -\frac{d\hat{\eta}}{d\hat{\tau}}; \\ &\Rightarrow \frac{d\hat{\eta}}{d\hat{\tau}} + \frac{1}{\alpha} = 0; \\ &\Rightarrow \frac{d\hat{\eta}}{d\hat{\tau}} = -\frac{1}{\alpha}. \end{aligned}$$

Luego como

$$\bar{\lambda} \in \min_{\alpha > 0} \{\Psi_\alpha(\lambda)\}.$$

el punto $(\hat{\tau}(\bar{\lambda}), \hat{\eta}(\bar{\lambda})) \in L_{\text{in}}$ y la recta tangente en este punto, esta dada por

$$\begin{aligned} \hat{\eta}(\bar{\lambda}) &= -\frac{1}{\alpha}\hat{\tau}(\bar{\lambda}) + ctte \Rightarrow \alpha\hat{\eta}(\bar{\lambda}) = -\hat{\tau}(\bar{\lambda}) + ctte \\ &\Rightarrow \hat{\tau}(\bar{\lambda}) + \alpha\hat{\eta}(\bar{\lambda}) = ctte \end{aligned}$$

Así la recta dada, tiene la misma pendiente que la recta tangente a la curva, por lo que la curva L_{in} es tangente a la recta $\hat{\tau} + \alpha\hat{\eta} = ctte$

Si $\Psi_\alpha(\bar{\lambda})$ alcanza un mínimo local en $\bar{\lambda}$ entonces por (3.52), $\hat{\tau}(\bar{\lambda}) + \alpha\hat{\eta}(\bar{\lambda})$ alcanza un mínimo local en $\bar{\lambda}$, por la primera implicación tenemos que L_{in} es tangente a la recta $\hat{\tau} + \alpha\hat{\eta}$ en el punto $(\hat{\tau}(\bar{\lambda}), \hat{\eta}(\bar{\lambda}))$. Así existe un vecindad U de $(\hat{\tau}(\bar{\lambda}), \hat{\eta}(\bar{\lambda}))$, tal que todos los puntos de L_{in} , están por encima de $\hat{\tau} + \alpha\hat{\eta} = \hat{\tau}(\bar{\lambda}) + \alpha\hat{\eta}(\bar{\lambda})$ la cual representa la recta tangente en el punto $(\hat{\tau}(\bar{\lambda}), \hat{\eta}(\bar{\lambda}))$. Luego por definición de convexidad, L_{in} es convexa en U .

Supongamos ahora que se cumple (1) y (2). Por (1) tenemos que la curva L_{in} es tangente a la recta $\hat{\tau} + \alpha\hat{\eta}$ en el punto $(\hat{\tau}(\bar{\lambda}), \hat{\eta}(\bar{\lambda}))$, por lo que la recta tangente en este punto tiene pendiente:

$$\frac{d\hat{\eta}}{d\hat{\tau}}(\bar{\lambda}) = -\frac{1}{\alpha}. \quad (3.53)$$

Calculemos la derivada de $\Psi_\alpha(\bar{\lambda})$ con respecto de λ

$$\frac{d}{d\lambda}\Psi_\alpha(\bar{\lambda}) = e^{\hat{\tau}(\bar{\lambda}) + \alpha\hat{\eta}(\bar{\lambda})} \left[\frac{d\hat{\tau}}{d\lambda}(\bar{\lambda}) + \alpha \frac{d\hat{\eta}}{d\lambda}(\bar{\lambda}) \right], \quad (3.54)$$

Como $e^{\hat{\tau}(\bar{\lambda}) + \alpha\hat{\eta}(\bar{\lambda})} > 0$ y $\frac{d\hat{\tau}}{d\lambda}(\bar{\lambda}) > 0$, podemos reescribir (3.54) como,

$$\frac{d}{d\lambda}\Psi_\alpha(\bar{\lambda}) = e^{\hat{\tau}(\bar{\lambda}) + \alpha\hat{\eta}(\bar{\lambda})} \left[\frac{d\hat{\eta}}{d\hat{\tau}}(\bar{\lambda}) + \frac{1}{\alpha} \right];$$

sustituyendo (3.53) tenemos que:

$$\frac{d}{d\lambda}\Psi_\alpha(\bar{\lambda}) = 0.$$

Por lo que $\bar{\lambda}$ es un extremo local de Ψ_α , luego por (2) como L_{\ln} , es convexa en una vecindad U del punto $(\hat{\tau}(\bar{\lambda}), \hat{\eta}(\bar{\lambda}))$, se tiene que $\bar{\lambda}$ es un mínimo local de Ψ_α .

□

3.3. Búsquedas Inexactas en La Curva-L Rotada.

En [3] Lara, Torrealba, Alvarez, estudian la curva L rotada y las búsquedas Inexactas.

Estamos interesados en establecer un procedimiento de búsqueda inexacta de λ en la curva $L_{\ln - \ln}$, para desarrollar el mismo deberemos calcular algunas derivadas.

3.3.1. Derivadas en La Curva L

Recordando la definición dada en (3.32) y la expresión obtenida en (3.29), obtenemos que:

$$d\hat{\eta} = -\frac{\frac{1}{\eta} d\eta}{\frac{1}{\tau} d\tau} = -\frac{\tau}{\lambda\eta}.$$

además de, $d\hat{\tau} = \frac{d\tau}{\tau}$ y $d\hat{\eta} = \frac{d\eta}{\eta}$

$$d\eta = [L^T L(x - x_c)]^T dx.$$

$$d\tau = [M^T (Mx - b)]^T dx. \quad (3.55)$$

Al calcular la segunda derivada obtenemos:

$$\frac{d^2 \hat{\eta}}{d\hat{\tau}^2} = \frac{d}{d\hat{\tau}} \left(\frac{d\hat{\eta}}{d\hat{\tau}} \right) = \frac{d}{d\tau} \left(-\frac{\tau}{\lambda\eta} \right) \frac{d\tau}{d\hat{\tau}}. \quad (3.56)$$

Las derivadas para $x(\lambda)$ pueden ser calculadas en un punto que satisfaga:

$$M^T (Mx - b) + \lambda L^T L(x - x_c) = 0.$$

al diferenciar implícitamente esta expresión se obtiene:

$$(M^T M + \lambda L^T L) dx = -L^T L(x - x_c) d\lambda.$$

de esta forma se obtiene

$$dx = -(M^T M + \lambda L^T L)^{-1} L^T L(x - x_c) d\lambda.$$

y de la expresión obtenida en (3.55) así tenemos:

$$\frac{d\tau}{d\lambda} = \lambda\kappa.$$

donde

$$\kappa = (x(\lambda) - x_c)^T L^T L (M^T M + \lambda L^T L)^{-1} L^T L (x(\lambda) - x_c).$$

luego de (3.56) tenemos que:

$$\frac{d}{d\tau} \left(-\frac{\tau}{\lambda\eta} \right) = \frac{-\frac{d\tau}{d\lambda} \lambda\eta + \tau \left(\eta \frac{d\lambda}{d\tau} + \lambda \frac{d\eta}{d\tau} \right)}{(\lambda\eta)^2}.$$

y sustituyendo todo esto en (3.56)

$$\frac{d^2 \hat{\eta}}{d\hat{\tau}^2} = \left[\frac{-\lambda^2 \kappa \eta + \tau \eta - \lambda \kappa \tau}{\lambda^3 \kappa \eta^2} \right] \tau.$$

Se necesitan procedimientos eficientes para calcular todas estas derivadas, para calcular la solución $x(\lambda)$, además que para resolver κ , requiere de la resolución de un sistema de ecuaciones lineales, además de calcular la inversa del Hessiano. Es por ello que nos planteamos un procedimiento menos infructuoso o complicado si se quiere en la proxima sección.

3.3.2. Curva L Rotada

Al rotar la Curva-L dada por, [3] se obtiene la curva $(\theta, G(\theta))$, donde $\theta = \hat{\tau} - \hat{\eta}$ y $G(\theta) = \hat{\theta} - \hat{\eta}$.

El objetivo es usar la minimización de la función G como criterio para seleccionar el parámetro de regularización λ , evitando seleccionar aquellos λ que estén en regiones de la curva dominadas por la solución regularizada o el ruido. Para el desarrollo de este procedimiento se utilizan primeras derivadas de la función G .

Lema 3.3.1. Dada la Curva-L rotada $(\theta, G(\theta))$. La primera derivada esta dada por:

$$\frac{dG}{d\theta} = \frac{-\tau + \lambda\eta}{\tau + \lambda\eta}.$$

Demostración:

Observemos que $d\theta = d\hat{\tau} - d\hat{\eta}$ y que $dG(\theta) = d\hat{\tau} + d\hat{\eta}$, además usando las derivadas encontradas en la sección anterior $d\hat{\tau} = \frac{d\tau}{\tau}$, $d\hat{\eta} = \frac{d\eta}{\eta}$ y $d\eta = -\frac{d\tau}{\lambda}$, se obtiene:

$$\begin{aligned} \frac{dG(\theta)}{d\theta} &= \frac{d\hat{\tau} + d\hat{\eta}}{d\hat{\tau} - d\hat{\eta}}; \\ &= \frac{\frac{d\tau}{\tau} + \frac{d\eta}{\eta}}{\frac{d\tau}{\tau} - \frac{d\eta}{\eta}}; \\ &= \frac{\frac{d\tau}{\tau} - \frac{d\tau}{\lambda\eta}}{\frac{d\tau}{\tau} + \frac{d\tau}{\lambda\eta}}; \\ &= \frac{\frac{d\tau(\lambda\eta) - \tau d\tau}{\tau\lambda\eta}}{\frac{\lambda\eta d\tau + \tau d\tau}{\tau\lambda\eta}}; \\ &= \frac{-\tau + \lambda\eta}{\tau + \lambda\eta}. \end{aligned}$$

□

Definición 3.3.1. Llamaremos Intervalo de Wolfe, al intervalo formado por el λ_{k-1} , λ_k , en donde se encuentra el mínimo de la función G y se obtienen como resultado del Algoritmo Diseñado.

Definición 3.3.2. Llamaremos Óptimo Aproximado a cualquier λ , perteneciente al Intervalo de Wolfe.

Se apela a las búsquedas inexactas cuando se puede prescindir de la exactitud, es decir podemos buscar el λ en un intervalo, donde se aproxime lo suficiente al λ en la Esquina de la curva. Aquí se implementará la Búsqueda de Wolfe, detallada en capítulos anteriores y en [8], [10].

En [3], diseñaron un algoritmo de búsqueda inexacta, que calcula de forma aproximada, el parámetro λ , que minimiza la curva L, rotada. En el algoritmo se implementa

una versión de la búsqueda de Wolfe, que usa segundas derivadas para evitar valores altos de la función G .

En este trabajo se pone en practica el algoritmo, con una versión de la búsqueda de Wolfe, que usa solo primeras derivadas y presentamos a continuación.

Algoritmo 1 Algoritmo para Buscar el Parámetro de Regularización Utilizando Búsquedas Inexactas

Dado $c_1, c_2 \in (0, 1), \epsilon > 0$ $\lambda_0 = \sigma_r$ satisfaciendo $\frac{dG}{d\theta} < 0$

```

1: para  $i = 1, 2, \dots$  hacer
2:    $\lambda_i := 2^i \lambda_0$ 
3:   si  $\frac{dG}{d\theta}(\lambda_i) > 0$  entonces
4:     Parar
5:   fin si
6: fin para
7:  $\underline{\lambda}_0 := \lambda_{i-1}, \bar{\lambda}_0 := \lambda_i$ 
8: mientras  $[G(\lambda_k) \geq G(\lambda_0) + c_1 \lambda_k \frac{dG}{d\theta}(\lambda_0)$  y  $\frac{dG}{d\theta}(\lambda_k) \leq c_2 \frac{dG}{d\theta}(\lambda_0)]$  o  $[|\underline{\lambda}_0 - \bar{\lambda}_0| > \epsilon]$ 
   hacer
9:    $\lambda_k := \frac{\bar{\lambda}_0 - \underline{\lambda}_0}{2}$ 
10:  si  $\frac{dG}{d\theta}(\lambda_k) > 0$  entonces
11:     $\bar{\lambda}_0 := \lambda_k$ 
12:  si no
13:     $\underline{\lambda}_0 := \lambda_k$ 
14:  fin si
15: fin mientras

```

La búsqueda inexacta y el método de Bisección garantizan que el algoritmo se detendrá en un número finito de pasos. Para apreciarlo, tomemos $r_0 = \lambda_k - \lambda_{k-1}$, entonces:

$$r_1 = \frac{1}{2}r_0; \quad (3.57)$$

$$r_2 = \frac{1}{2}r_1; \quad (3.58)$$

$$\vdots \quad (3.59)$$

$$r_k = \left(\frac{1}{2}\right)^k r_0 \leq \epsilon \quad (3.60)$$

$$= 2^{-k} \leq \frac{\epsilon}{r_0}; \quad (3.61)$$

de (3.61), se tendrá que:

$$-k \log(2) \leq \log\left(\frac{\epsilon}{r_0}\right),$$

así el número de pasos necesarios, para el cual el algoritmo se detendrá será:

$$k \geq \log\left(\frac{\epsilon}{r_0}\right).$$

Teorema 3.3.1. *El algoritmo 1, termina en un número finito de pasos, con un óptimo aproximado.*

Para ver la demostración de este teorema, consultar [3].

CAPÍTULO 4

EXPERIMENTACIÓN NUMÉRICA Y RESULTADOS

Para la realización y ejecución de la experimentación numérica correspondiente a este trabajo, desarrollamos en primer lugar el algoritmo en Matlab, con la utilización del toolbox de regularización para problemas discretos mal puestos desarrollado por Christian Hansen, en el 2008, de donde se extrajo los test-problemas para la obtención de datos correspondientes a distintos problemas discretos mal puestos.

Adicionalmente para la presentación de dichos resultados, desarrollamos una interface gráfica llamada ABICL, con la utilización del entorno GUIDE de Matlab.

La primera parte de la experimentación consistió en escribir en lenguaje de programación Matlab nuestro algoritmo, utilizando módulos independientes para cada operación y así optimizar la ejecución del mismo por Matlab. Luego procedimos al desarrollo de módulos para la obtención de datos generados por los test problemas del toolbox de regularización.

Finalmente incluimos en la interface desarrollada, el calculo de los respectivos errores relativos y absolutos para la comparación de la soluciones dadas por el toolbox de Hansen y el de nuestro algoritmo. Cabe destacar que el número de datos por defecto es de 64, puesto que es un número suficientemente alto para apreciar los resultados y además es múltiplo de 4, para algunos test Problemas del toolbox requieren esa condición, además para algunos test Problemas relacionados con la Suma de exponenciales, estos datos corresponden a problemas inversos reales, por tal motivo los

datos se introducen manualmente.

El toolbox de regularización de Hansen, ofrece una serie de rutinas de utilidad, para el análisis de problemas discretos mal puestos en detalles, la visualización de las propiedades, y para la generación fácil de problemas de prueba.

Nombre del Test Problema	Parametro de Regularización de Hansen		Parametro de Regularización ABICL		Error Absoluto de hansen y la solución real	Error relativo de hansen y la solución real	Error Absoluto de ABICL y la solución real	Error Relativo de ABICL y la solución real
	Valor_Lambda	Tiempo_seg.	Valor_Lambda	Tiempo_seg.				
Baart	0,00901074	2,307	0,00683187	0,1141	0,1451	0,1173	0,1421	0,1141
Deriva2 caso 1	0,012027	0,991807	0,019531	0,249911	0,2676	0,5501	0,2839	0,6179
Deriva2 caso 2	0,004181	0,893171	0,0045573	0,243222	0,6242	0,3714	0,6238	0,3731
Deriva2 Caso3	0,016554	0,934062	0,03125	0,262591	0,0456	0,1562	0,0376	0,1386
Foxgood	0,0022713	1,034177	0,0020726	0,217012	0,0993	0,0215	0,1204	0,0261
Gravity Caso 1	0,0015458	1,436501	0,0015214	0,225466	0,4527	0,0714	0,46	0,0725
Gravity Caso 2	0,00069077	4,505038	0,00076072	0,258323	1,2851	0,1204	1,2267	0,115
Gravity Caso3	0,00093173	1,018418	0,00076072	0,238186	0,7703	0,0685	0,783	0,0696
Heat	0,24915	0,92023	0,24917	0,2663	0,5527	0,2683	0,7543	0,3519
l Laplace Caso 1	0,0039213	0,8975	0,0039168	0,232436	0,2713	0,132	0,2715	0,1321
l Laplace Caso 2	0,0015988	0,909874	0,0019584	0,228415	6,0563	1,3826	6,0522	1,3894
l Laplace Caso 3	0,0017442	0,901161	0,0019584	0,221315	0,2587	0,0475	0,2412	0,0443
l Laplace Caso 4	0,0015988	0,895832	0,0019584	0,241051	6,0431	1,3212	6,0421	1,3428
Phillips	0,00192	1,060452	0,0016635	0,224439	0,57	0,1866	45,8373	0,9979
shaw	0,0010317	0,903045	0,0014947	0,25201	0,328	0,0411	0,3378	0,0423
Spikes	0,00037243	0,960332	0,00050289	0,25751	24,0732	2,7188	24,0701	2,7381
Tomo	0,01503	5,695135	0,022545	0,224923	0,4102	0,0357	0,7604	0,0665
Wing	0,0069232	4,56719	0,015976	0,24731	0,3487	0,7062	0,3467	0,7686
Lanczos	7,385E-07	1,776449	4,1812E-06	0,237002	0,6103	1,5198	0,6119	1,8546
Evans	1,1556E-06	1,778654	2,4475E-06	0,240427	0,2872	0,5123	0,3029	0,5583
Suma de 1 Exponenciales	0,000015255	1,942145	0,00002295	0,36441	0,9732	4,2667	0,9734	4,2961
Suma de 2 Exponenciales	0,000016162	2,904388	2,6438E-05	0,243101	0,7014	3,7165	0,7027	3,802
Suma de 3 Exponenciales	0,000019103	1,709,148	2,6438E-05	0,24939	0,6365	4,256	0,6366	4,2762
Suma de 4 Exponenciales	0,0045	2,627815	0,0069	0,232652	0,5222	3,6503	0,5226	3,7161
steyrn-wyk	0,1384	2,555397	0,15732	0,234031	0,3475	1	0,3467	1

Tabla 4.1: Resumen de la Experimentación

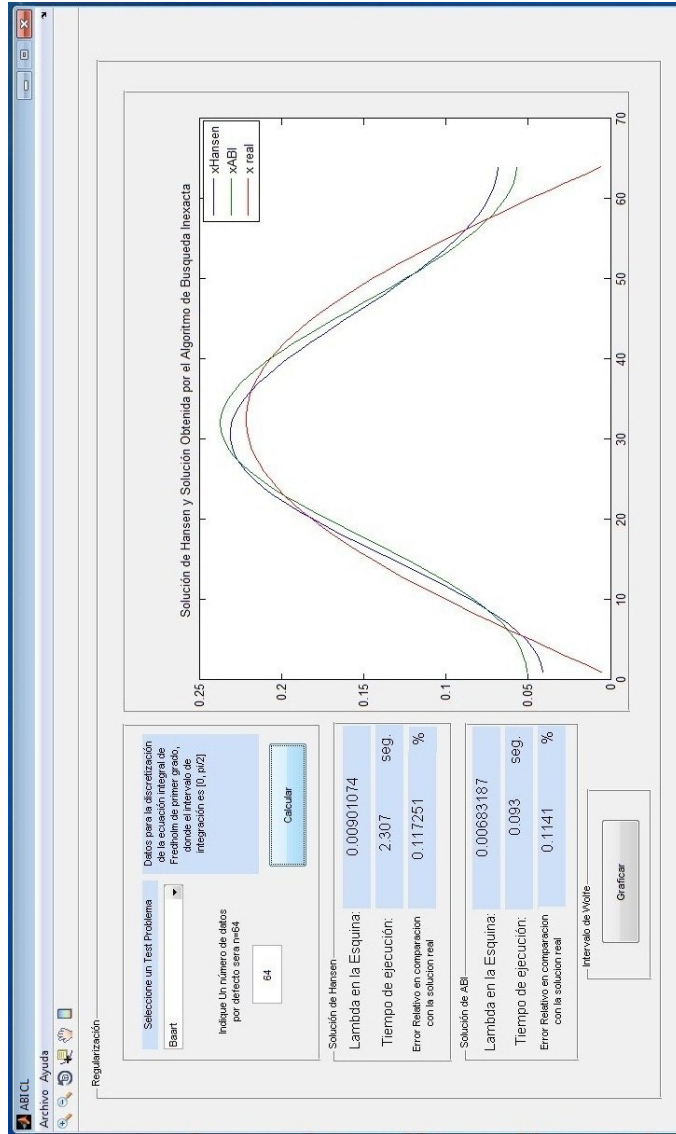


Figura 4.1: Test problema Baart

Para esta Data, La solución dada por el nuestro algoritmo (ABI CL), se aproxima más a la solución real, y además de eso lo hace en menos tiempo que el toolbox de Hansen, por lo que sería mas efectivo para este problema en particular.

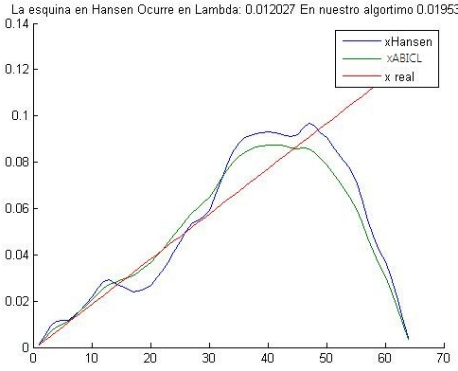
	<p>Descripción: Es una Discretización de la ecuación integral de Fredholm de primer orden, cuyo kernel K, es la función de Green's para la segunda derivada.</p> <p>Para este Test Problema, tanto el toolbox de Hansen, como nuestro algoritmo se acercan a la solución de una misma forma, logrando el toolbox de Hansen conseguir estar mas cerca de esta solución, sin embargo la diferencia en el tiempo que lo hace es bastante significativa, lo que hace a nuestro algoritmo una alternativa viable, ya que logra un solución muy similar, en mucho menos tiempo.</p>
---	---

Tabla 4.2: Test Problema Deriva2 caso1

	<p>Descripción: Es una Discretización de la ecuación integral de Fredholm de primer orden, cuyo kernel K, es la función de Green's para la segunda derivada.</p> <p>En este Test Problema, ambos métodos consiguen el mismo resultado, pero con la salvedad que al toolbox de Hansen le toma mucho mas tiempo que a nuestro algoritmo, lo que hace del ABICL un opción viable para resolver esta Data.</p>
--	--

Tabla 4.3: Test Problema Deriva2 caso2

	<p>Descripción: Es una Discretización de la ecuación integral de Fredholm de primer orden, cuyo kernel K, es la función de Green's para la segunda derivada.</p> <p>Para estos Datos, nuestro algoritmo se aproxima de mejor forma a la solución al conseguir un parámetro de regularización más indicado, y en mucho menos tiempo que el toolbox de Hansen, lo que hace del ABICL un método atractivo para resolver esta data.</p>
--	---

Tabla 4.4: Test Problema Deriva2 caso3

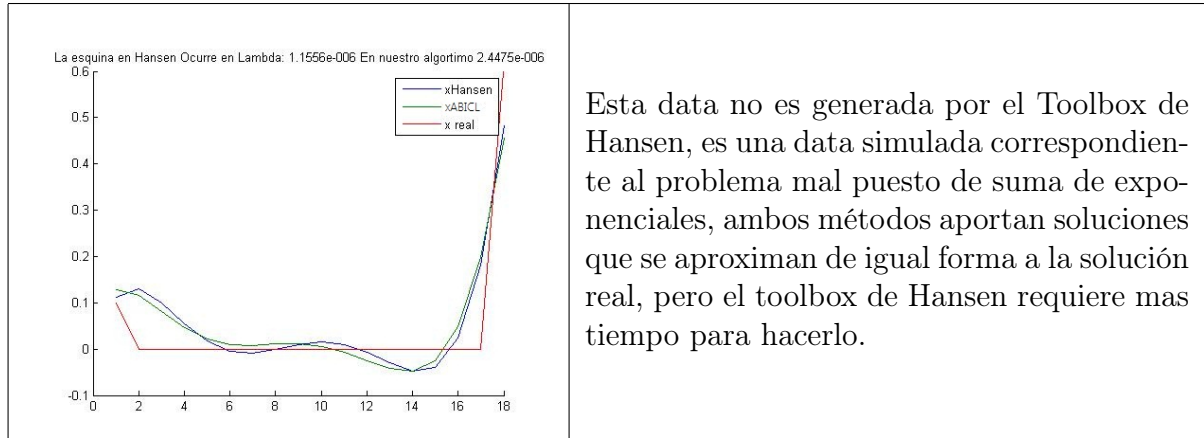


Tabla 4.5: Data Evans

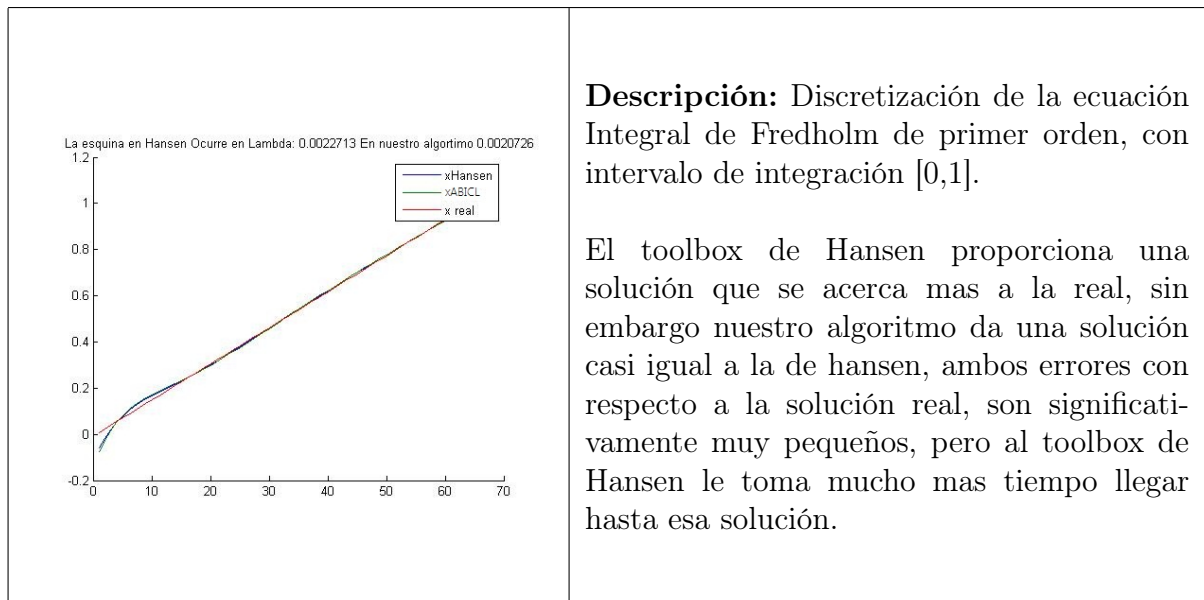


Tabla 4.6: Test Problema Foxgood

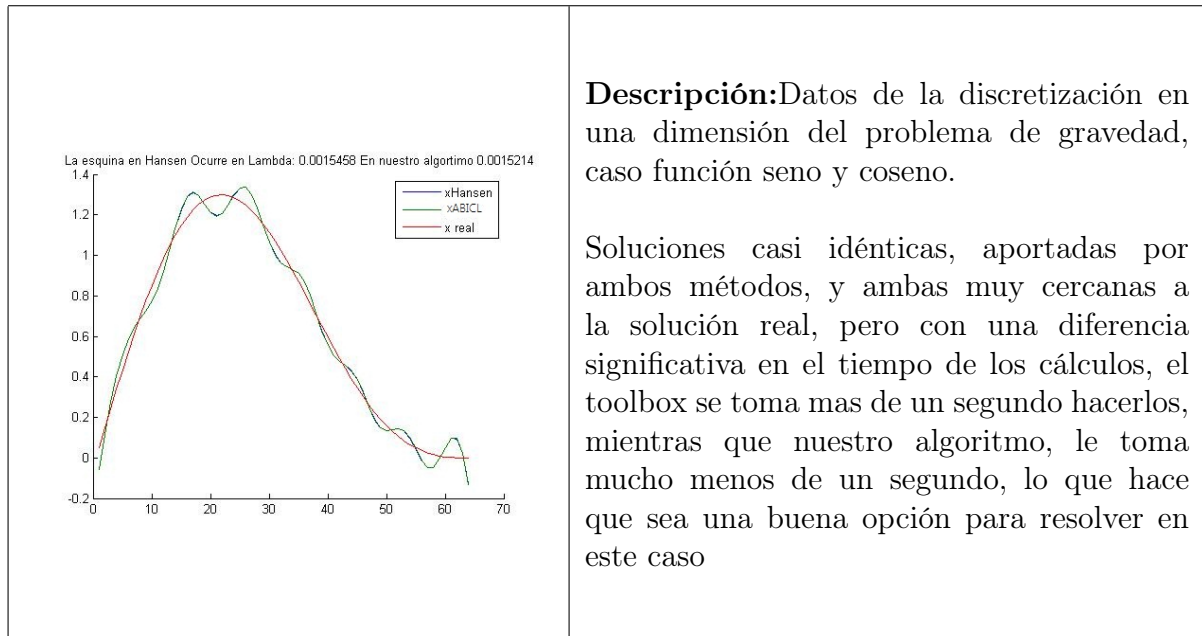


Tabla 4.7: Test Problema Gravity caso1

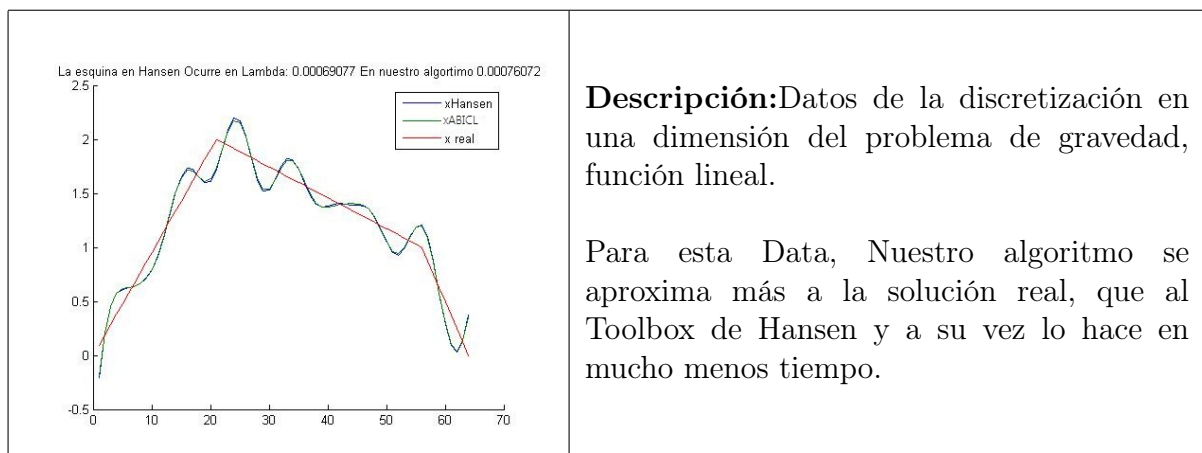


Tabla 4.8: Test Problema Gravity caso2

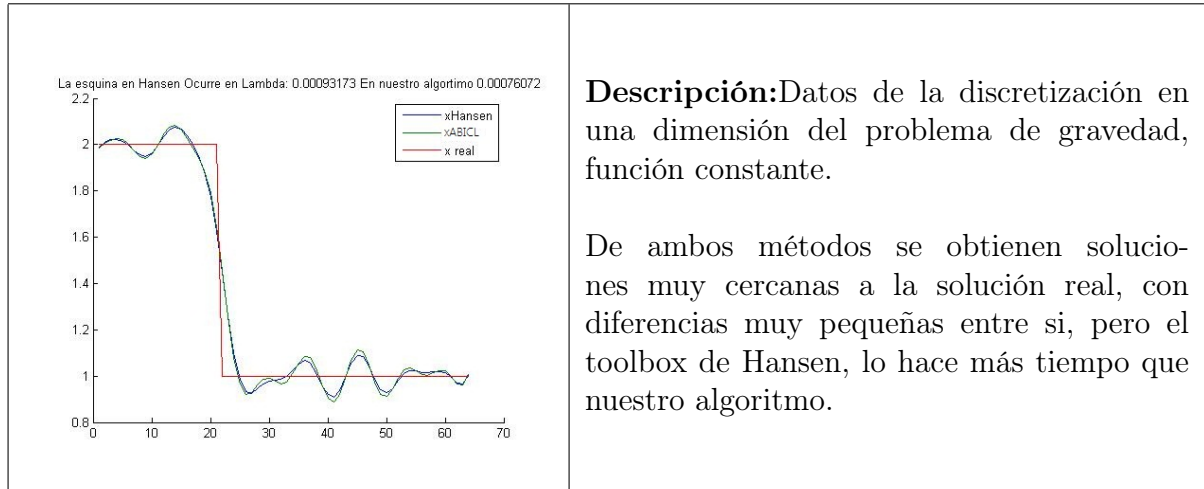


Tabla 4.9: Test Problema Gravity caso3

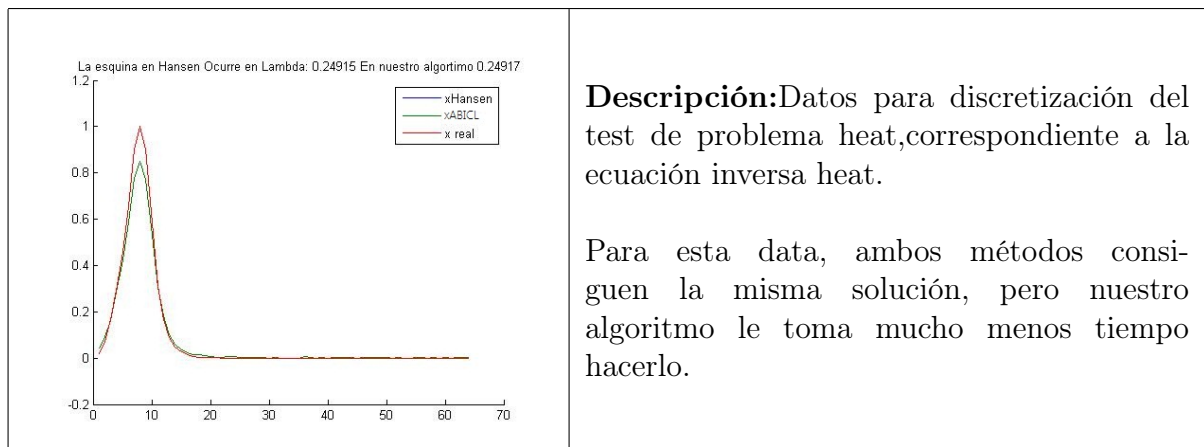


Tabla 4.10: Test Problema Heat

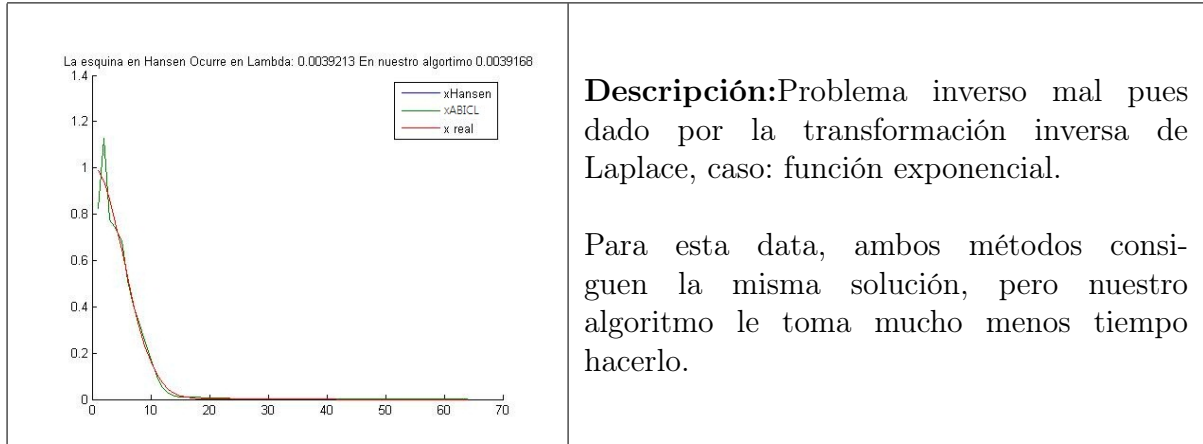


Tabla 4.11: Test Problema Ilaplace caso1

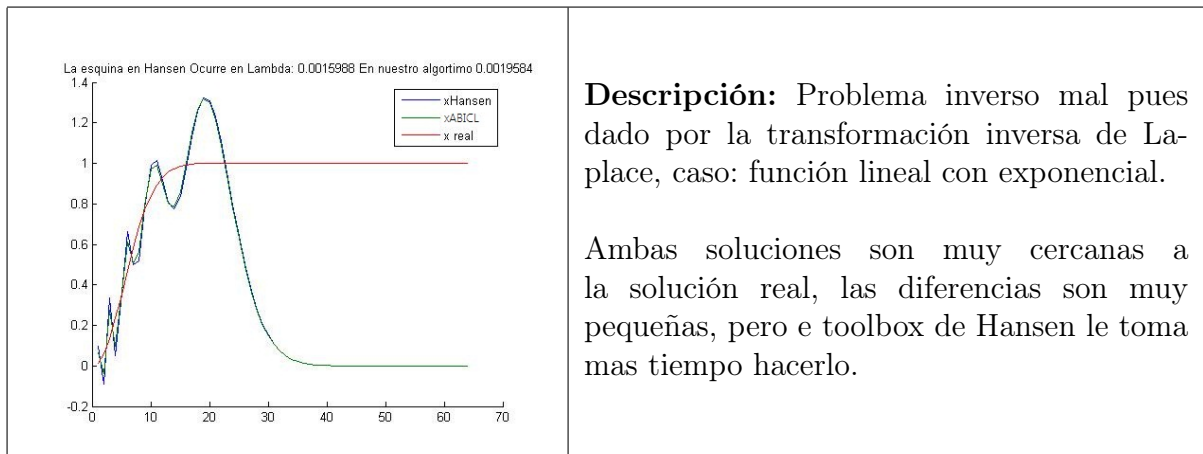


Tabla 4.12: Test Problema Ilaplace caso2

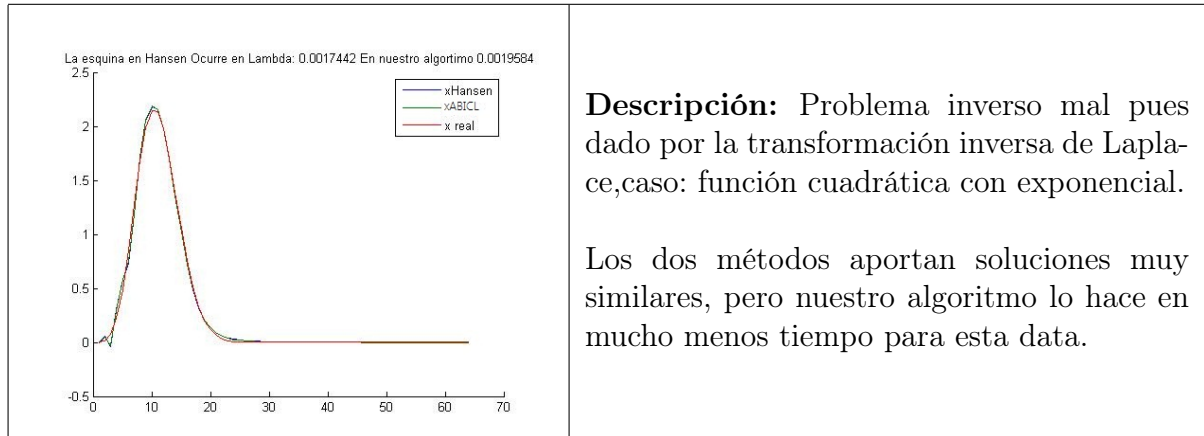


Tabla 4.13: Test Problema Ilaplace caso3

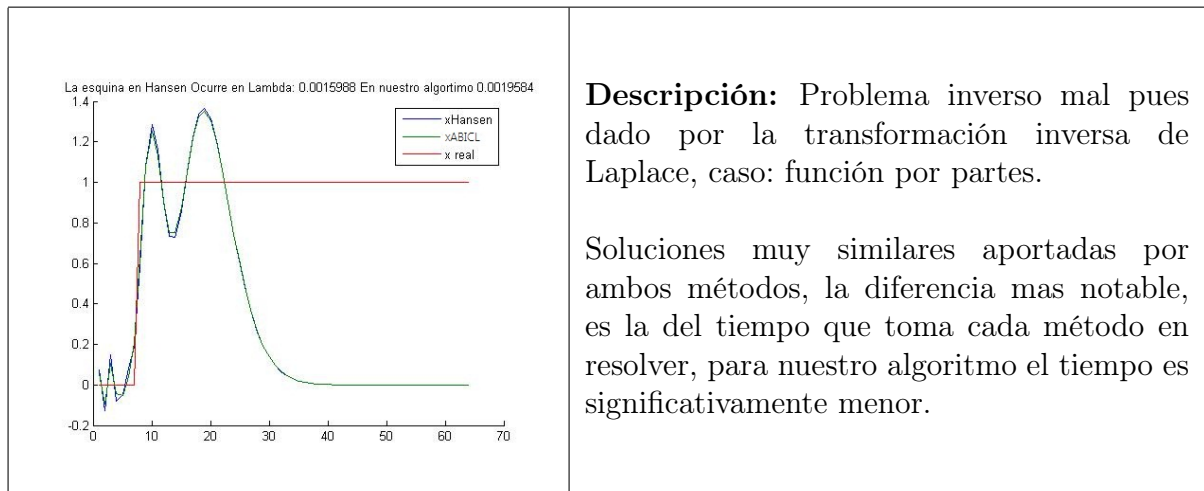


Tabla 4.14: Test Problema Ilaplace caso4

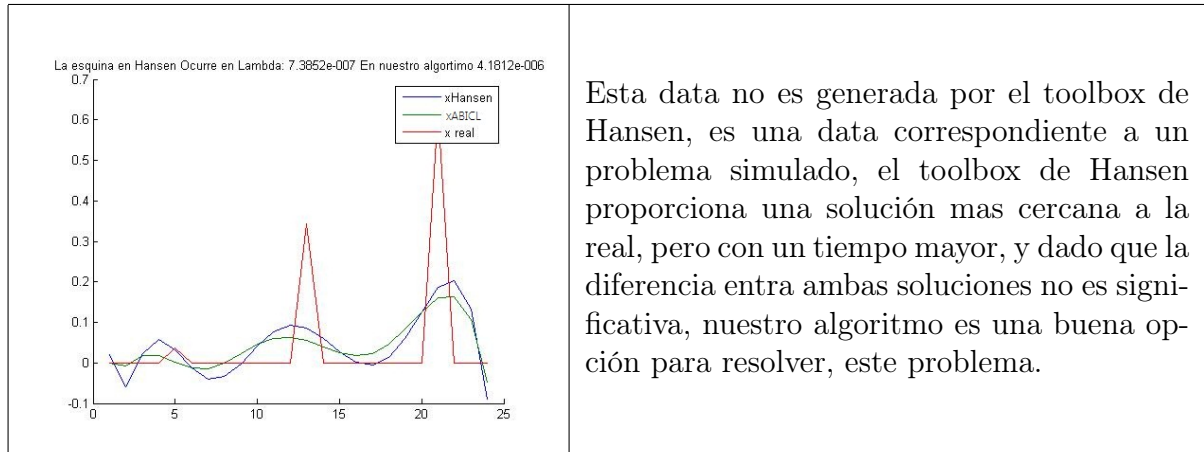


Tabla 4.15: Data Lanczos

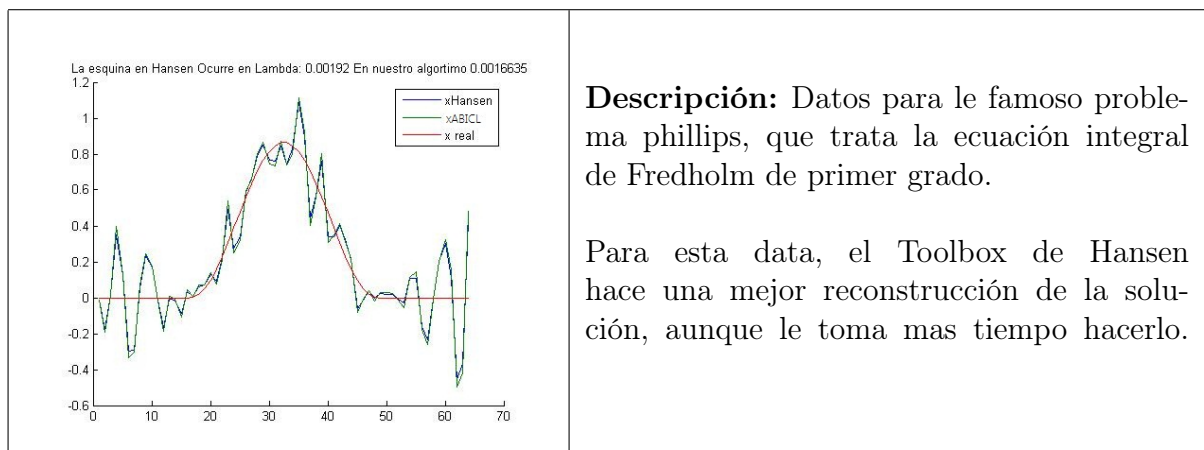


Tabla 4.16: Test Problema Phillips

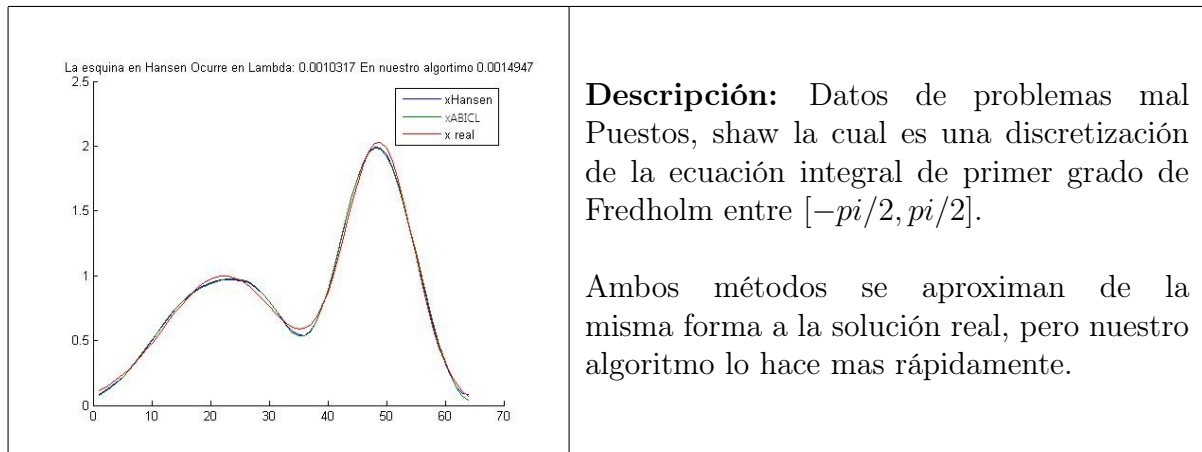


Tabla 4.17: Test Problema Shaw

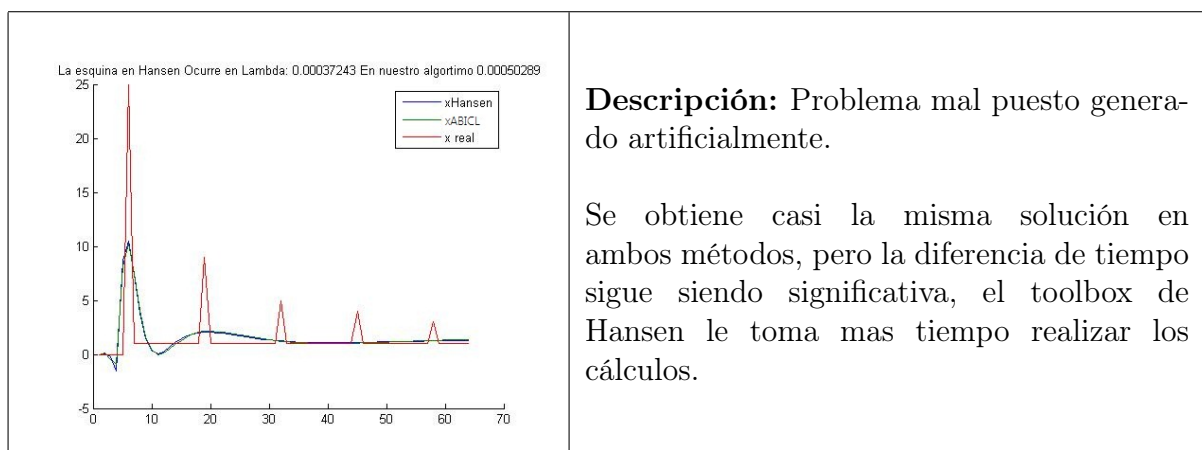


Tabla 4.18: Test Problema Spikes

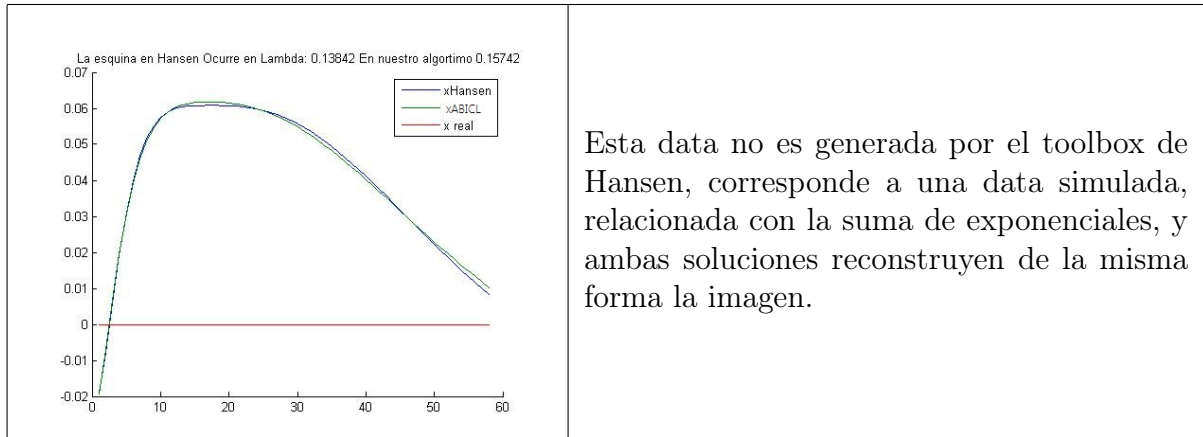


Tabla 4.19: Test Problema stay-week

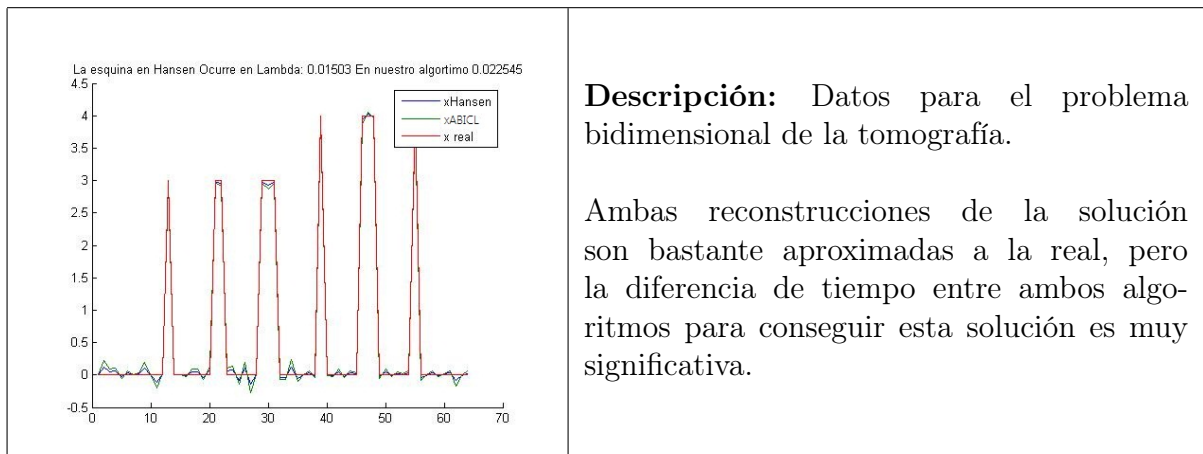


Tabla 4.20: Test Problema Tomo

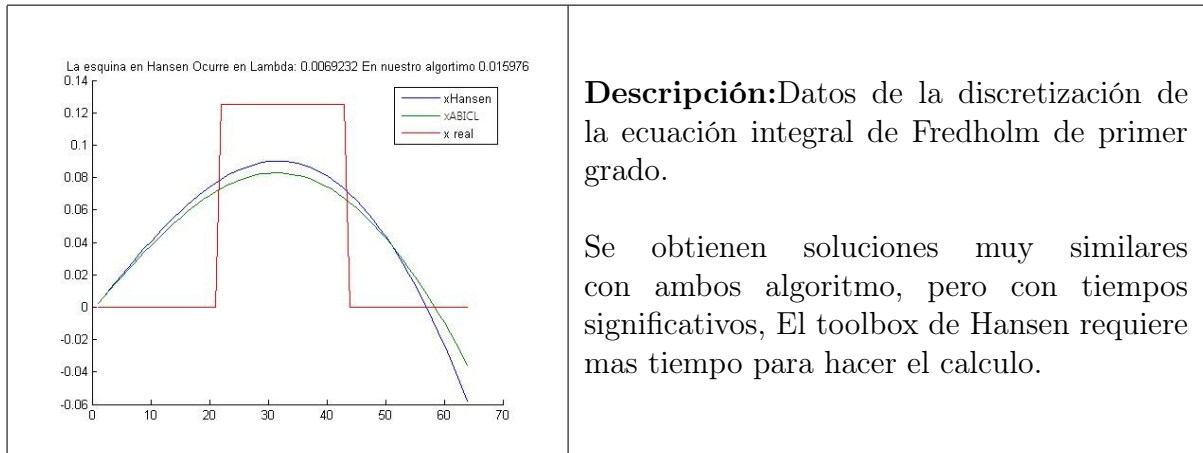


Tabla 4.21: Test Problema Wing

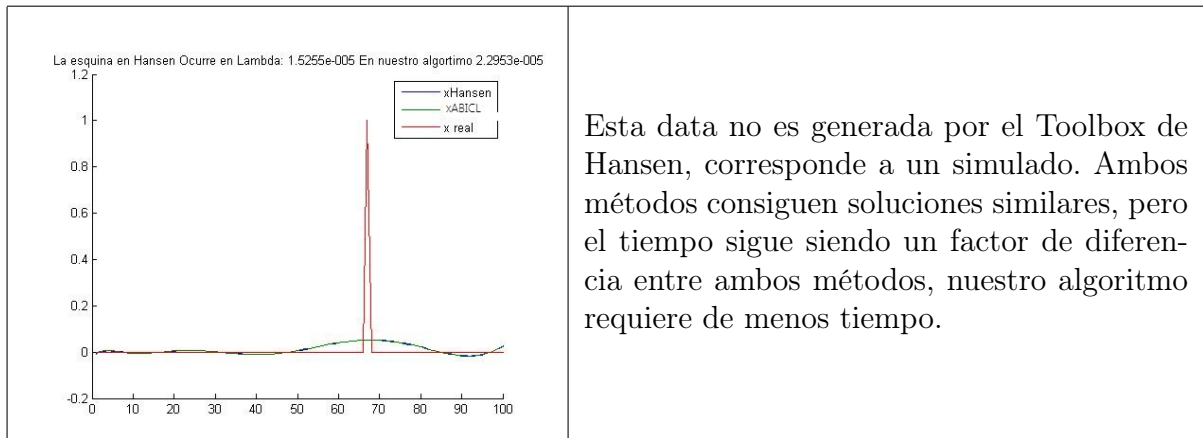


Tabla 4.22: Suma de una Exponencial

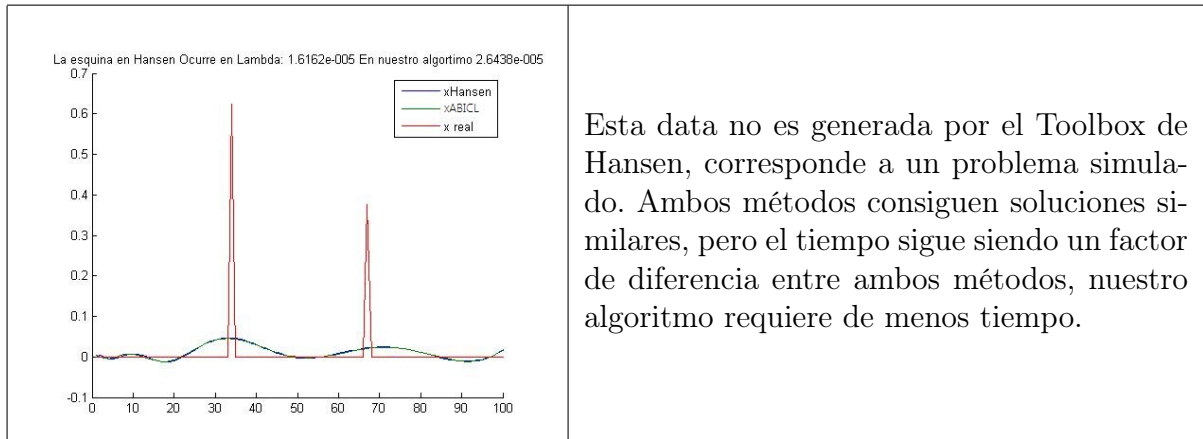


Tabla 4.23: Suma de dos Exponenciales

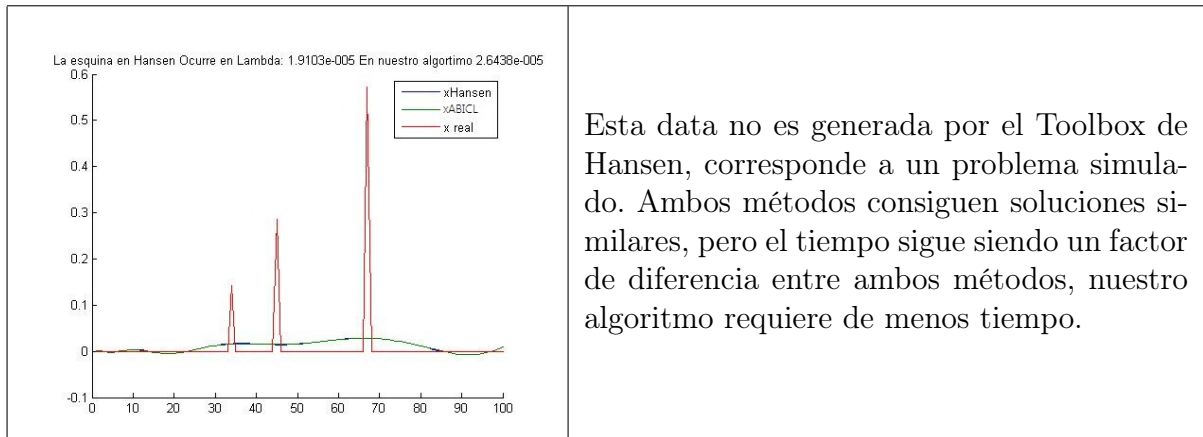
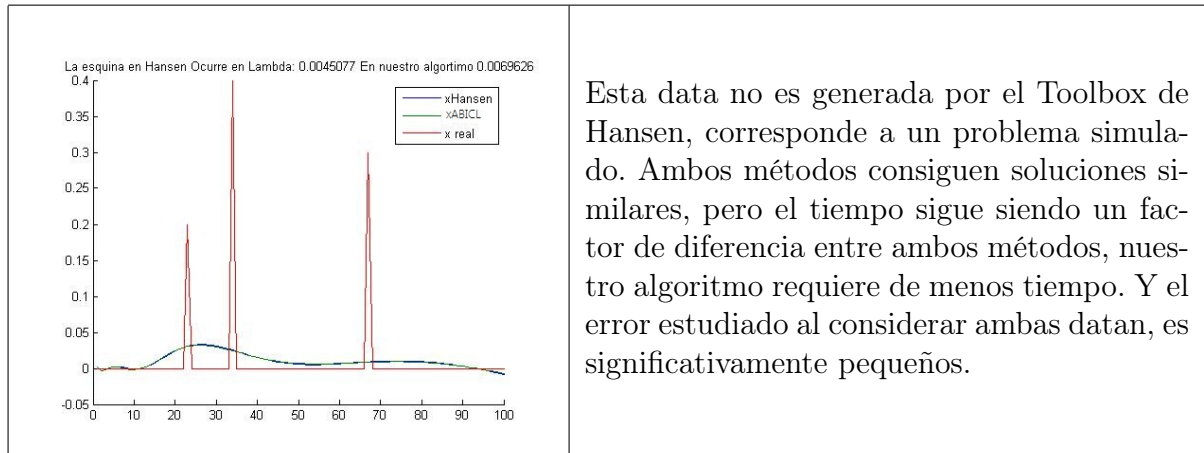


Tabla 4.24: Suma de tres exponenciales



Esta data no es generada por el Toolbox de Hansen, corresponde a un problema simulado. Ambos métodos consiguen soluciones similares, pero el tiempo sigue siendo un factor de diferencia entre ambos métodos, nuestro algoritmo requiere de menos tiempo. Y el error estudiado al considerar ambas datan, es significativamente pequeños.

Tabla 4.25: Suma de cuatro exponenciales

CONCLUSIONES

Este trabajo especial de grado incluye una revisión exhaustiva, sobre la bibliografía correspondiente a los problemas inversos mal puestos y la regularización como método para resolver los sistemas de mínimos cuadrados mal condicionados, presentándose con mayor hincapié la Regularización de Tikhonov, puesto que es el método base para solucionar problemas inversos, planteado en este trabajo.

Los Problemas Inversos aparecen en prácticamente todas las ramas de las matemáticas. Las aplicaciones cubren campos dispares como la medicina, la geofísica o la ciencia de los materiales. La formulación matemática de estos problemas da lugar a problemas difíciles de tratar, tanto analítica como numéricamente.

Los métodos hoy conocidos utilizados para el arduo trabajo que implica, encontrar el parámetro de regularización, son eficaces sin embargo son procedimientos, que pudiesen ser mucho mas rápidos, óptimos y sencillos, como lo es el método que proponemos en este trabajo.

La escogencia del parámetro de regularización es parte fundamental para obtener una buena aproximación de la solución problema inverso $Mx = b$. Los métodos de escogencia de dicho parámetro son una mezcla entre teoría y heurística.

El algoritmo de Búsquedas inexactas en la Curva L, (ABICL), se presenta como una alternativa viable, comprobado con todos lo experimentos realizados, para la

búsqueda del parámetro de regularización.

Es un algoritmo sencillo, y liviano, que hace su selección basada en el método de Wolfe para la búsqueda inexactas de mínimos en funciones, se obtienen, parámetros de regularización muy aproximados a los conseguidos con los métodos aportados por Hansen, como se aprecia en la tabla 5.1, sin embargo se logra una mejora sustancial en el tiempo de calculo que llevan estos algoritmo, a cambio de un costo en exactitud que si bien no es mucho, ya que en algunos casos el ABICL, obtiene una mejor aproximación del parámetro de regularización que los métodos de Hansen, (como lo es el caso del problema de prueba Deriva2 caso3, o del problema de prueba Baart), viene siendo, en algunos casos algo negativo que mencionar.

La experimentación con el el algoritmo ABICL proporciona evidencia clara, de la eficiencia y rapidez de este algoritmo, y lo muestra como algo novedoso, en lo que hasta ahora, había sido la teoría de Regularización.

APÉNDICE A

CÓDIGOS GENERADOS EN MATLAB

Los códigos y Scrip en esta sección, fueron desarrollados en Matlab, con la ayuda de [4], que como se menciona, proporcionan herramientas para la generación de problemas de prueba con los que pudimos experimentar.

```
1  %Programa principal del algoritmo de búsqueda inexacta en la curva L
2  %log-log
3  tic
4  clear all
5  clc
6  load resultados.mat
7  [U, s, V] = csvd(M);
8  lambdainicial=s(n,1);
9  xlamba=tikhonov(U,s,V,b,lambdainicial);
10 [k,r]=size(xlamba);
11 L=eye(size(k));
12 xc=0;
13 [dGdt,taob,etab]=dGdtheta(lambdainicial,b,M,xlamba,xc,L);
14 i=1;
15 %modulo para encontrar el lambda0 donde la función es decreciente
16 %(el inicial)
17 if (dGdt>0)
18     while (dGdt>0)
19         lambdai=(2^i)*lambdainicial;
20         xlamba=tikhonov(U,s,V,b,lambdai);
21         [dGdt,taob,etab]=dGdtheta(lambdai,b,M,xlamba,xc,L);
22         lambdaaux=lambdai;
23         i=i+1;
24     end
25     %Actualizando el punto inicial para wolfe
26     lambdainicial=lambdaaux;
27     dGdt0=dGdt;
28     [glambdainicial]=gdelambda(taob,etab);
29 else
30     dGdt0=dGdt;
31     [glambdainicial]=gdelambda(taob,etab);
```

```

32 end
33 j=1;
34 %modulo para encontrar el punto donde la funcion es creciente
35 while (dGdt<0)
36     lambda_0=(2^(j-1))*lambdainicial;
37     lambdaj=(2^j)*lambdainicial;
38     xlambda=tikhonov(U,s,V,b,lambdaj);
39     [dGdt,taob,etab]=dGdtheta(lambdaj,b,M,xlambda,xc,L);
40     lambda0=lambdaj;
41     j=j+1;
42 end
43 %modulo de la bisección y de búsqueda inexacta
44 alfa=-1;
45 while alfa===-1
46     lambdabk=(lambda0+lambda_0)/2;
47     xlambda=tikhonov(U,s,V,b,lambdabk);
48     [dGdt,taob,etab]=dGdtheta(lambdabk,b,M,xlambda,xc,L);
49     glambda=gdelambda(taob,etab);
50     if (sign(dGdt)== sign(dGdt0))
51         lambda_0=lambdabk;
52
53     else
54         lambda0=lambdabk;
55     end
56     %criterio de parada
57     alfa=wolfe(lambdabk,dGdt,glambda,dGdt0,glambdainicial,lambda_0,lambda0);
58 end
59 toc
60 %figure
61 %hold on
62 %plot([x_tik,xlambda,x]);
63 %title([' La esquina en Hansen Ocurre en Lambda: ',num2str(reg_corner),
64 %' En nuestro algortimo ',num2str(lambdabk)])
65 %legend('xHansen','xLaraP','x real');
66 %hold off
67 %errorabs=norm(x-xlambda);
68 %errorel=errorabs/norm(xlambda);
69 save Equis x_tik xlambda U s V b lambda0 lambda_0 x

1 %script para la partición del intervalo y para encontrar el intervalo de
2 %wolfe
3 clear all
4 clc
5 load Equis
6 for i=1:10
7     lambdas(1,i)=lambda_0+((lambda0-lambda_0)/10)*(i-1);
8 end

```



```
9 xtk=tikhonov(U,s,V,b,lambdas);
10 figure
11 hold on
12 plot(xtk,'k')
13 plot(x_tik,'y')
14 plot(xlambda,'r')
15 plot(x,'c')
16 legend('lambda1','lambda2','lambda3','lambda4','lambda5','lambda6','lambda7',
17 'lambda8','lambda9','lambda10','xtikhonov','xlamnda','xoriginal')
18 hold off

1 %Este módulo calcula las derivadas de la función G
2 function [dGdt,etab, taob]=dGdtheta(lambda,b,M,xlambda,xc,L)
3 taob=0.5*norm(b-M*xlambda);
4 etab=0.5*norm(L*(xlambda-xc));
5 dGdt=(-taob + lambda*etab)/(taob+lambda*etab);
6 end

1 %Este modulo corresponde a la evaluación de la función G
2 function [glambda]=gdelambda(taob,etab)
3 glambda=taob+etab;
4 end
```

APÉNDICE B

EJEMPLOS DE PROBLEMAS MAL PUESTOS

El ejemplo clásico de un problema mal puesto es una ecuación integral de Fredholm de primera especie con un núcleo cuadrado integrable.

$$\int_a^b K(s, t)f(t)dt = g(s), \quad c \leq s \leq d, \quad (\text{B.1})$$

donde el lado derecho de g y el núcleo de K , son dados, y donde f es la solución desconocida. Si la solución f es perturbada por:

$$\Delta f(t) = \epsilon \sin(2\pi pt), \quad p = 1, 2, \dots, \quad \epsilon = \text{constant} \quad (\text{B.2})$$

entonces la correspondiente perturbación del lado derecho de g es dada por:

$$\Delta g(s) = \epsilon \int_a^b K(s, t) \sin(2\pi pt) dt, \quad p = 1, 2, \dots \quad (\text{B.3})$$

y por el lema de Riemann-Lebesgue se deduce que $\Delta g \rightarrow 0$ cuando $p \rightarrow \infty$

Ejemplo Resolvamos la ecuación integral unidimensional.

$$\int_0^1 (s^2 + t^2)^{\frac{1}{2}} \phi(t) dt = \frac{1}{3} [(1 + s^2)^{\frac{3}{2}} - s^3] \quad s \in [0, 1] \quad (\text{B.4})$$

con solución exacta, $\phi(t) = t$.

Si se discretiza mediante la regla de cuadratura del punto medio dividiendo en forma

uniforme el intervalo $[0, 1]$ en n subintervalos de amplitud $\frac{1}{n}$, y tomando los puntos medios en cada uno de ellos se tienen los siguientes valores:

$$\begin{aligned} t_1 &= \frac{0 + \frac{1}{n}}{2} = \frac{1}{2n}; \\ t_2 &= \frac{\frac{1}{n} + \frac{2}{n}}{2} = \frac{3}{2n}; \\ &\vdots \\ t_j &= \frac{\frac{j-1}{n} + \frac{j}{n}}{2} = \frac{2j-1}{2n}. \end{aligned}$$

La fórmula de aproximación de B.4 es

$$\sum_{j=1}^n (s_i^2 + t_j^2)^{\frac{1}{2}} \phi(t_j) \Delta t_j = \frac{1}{3} [(1 + s_i^2)^{\frac{3}{2}} - s_i^3].$$

Como $\Delta t_j = \frac{1}{n}$ y tomando $\phi(t_j) = x_j$ y $b_i = [(1 + s_i^2)^{\frac{3}{2}} - s_i^3]$, se obtiene el sistema lineal de la forma $Ax = b$

$$\frac{1}{n} \sum_{j=1}^n (s_i^2 + t_j^2)^{\frac{1}{2}} x_j = b_i \quad i = 1, \dots, n \quad (\text{B.5})$$

con

$$A = \{a_{ij}\} = \left\{ \frac{1}{n} (s_i^2 + t_j^2)^{\frac{1}{2}} \right\}.$$

Al calcular la matriz A para distintos valores de n y tomando s también en el intervalo $[0, 1]$ con el mismo paso de t para simplificar los cálculos se obtienen los valores de la tabla siguiente. El error del problema inverso corresponde a la diferencia entre la aproximación del vector b , calculada por medio de la solución \tilde{x} para cada n , y el vector b exacto: $\|\tilde{b} - b\| = \|A\tilde{x} - b\|$.

Después de resolver el sistema lineal por eliminación de Gauss, se obtienen los resultados que se ilustran a continuación.

n	$\text{cond}(A)$	Error Problema Inverso
4	$3,8x10^3$	0,008
6	$7,4x10^4$	0,02
10	$1,9x10^{10}$	2,03
15	$5,5x10^{15}$	$1,9x10^4$
20	$2,5x10^{17}$	$1,2x10^5$
50	$1,8x10^{19}$	$1,2x10^5$
100	10^{19}	$7,1x10^4$
200	$3,6x10^{19}$	$1,3x10^6$
1000	$6,7x10^{20}$	$1,4x10^7$

Tabla B.1: Resultados del problema

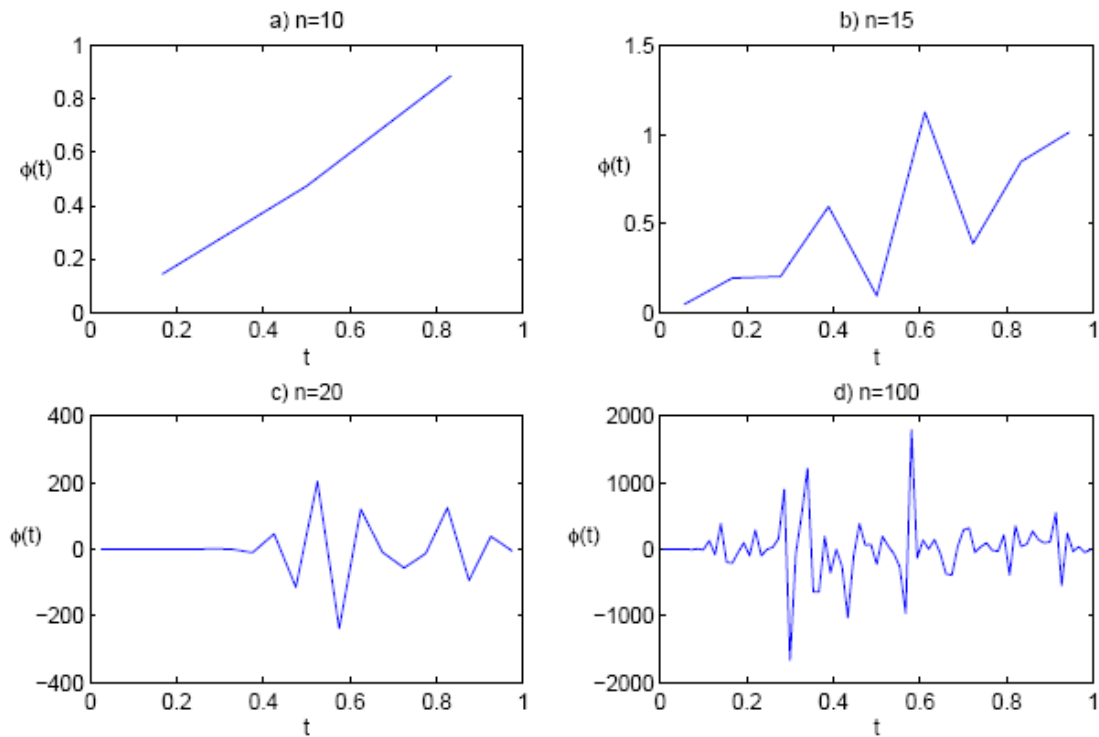


Figura B.1: Gráficas de la aproximación de la solución $\phi(t) = t$.

En ellos se ve claramente como se va perdiendo la solución en el caso del problema inverso a medida que crecen los valores de n . La distorsión en los valores de la función $\phi(t)$ se manifiestan incluso para valores relativamente pequeños de n . En particular podemos ver que cuando $n = 15$ se ven algunos valores lejanos de la identidad. Se esperaría que si aumentamos el tamaño de n a 20, éstos valores se acercaran más a la solución, sin embargo, vemos que la solución es más distorsionada.

REFERENCIAS

- [1] M. Rudnicki D.Krawczyk-Stando, *Regularization parameter selection in discrete ill-posed problems, the use of the u-curve*, Int. J. Appl. Math. Comput. Sci. **17** (2007), no. 2, 157–164.
- [2] S. Grossman, *Algebra lineal*, Grupo Editorial Iberoamericana, 1987.
- [3] F.Torrealba. H. Lara, A. Álvarez, *Búsquedas inexactas para seleccionar el parámetro de regularización de tikhonov*, VI Encuentro Colombia Venezuela de Estadística (2009), –.
- [4] P. C. Hansen, *Regularization tools version 3.0 for matlab 5.2*, Numerical Algorithms **20** (1999), 195–196.
- [5] P.C.Hansen, *The l–curve and its use in the numerical treatment of inverse problems,computational inverse problems in electrocardiology*, Advances in Computational Bioengineering **5** (2001), –.
- [6] T. Reginska, *A regularization parameter in discrete ill-posed problems*, SIAM J. Sci. Comput **17** (1996), no. 3, 740–749.
- [7] K. A. Woodbury., *Inverse problems*, <http://www.me.ua.edu/inverse/whatis.html>. (1995), –.
- [8] S. Nash y A. Sofer, *Linear and nonlinear programming*, McGraw Hill, 1996.
- [9] S.M. Tan y Colin Fox, *Physics 707 inverse problems*, The University of Auckland.
- [10] Jorge. Nocedal y Sheplen J. Wright., *Numerical optimization*, New York: Springer, 2006.