



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

MAESTRÍA EN CIENCIAS
(MATEMÁTICAS APLICADAS E INDUSTRIALES)

Regularización de Problemas Mal Planteados

Tesis para la obtención del grado de:

MAESTRA EN CIENCIAS
(MATEMÁTICAS APLICADAS E INDUSTRIALES)

Alumna: Diana Assaely León Velasco

Asesor:

Dr. L. Héctor Juárez Valencia,
UAM-I.

Coasesor:

Dr. J. Jacobo Oliveros Oliveros,
BUAP.

25 de Octubre de 2010

*“En las matemáticas es donde el espíritu encuentra los elementos que más ansía: la
continuidad y la perseverancia”.*
Jacques Anatole (1844-1924). Escritor francés.

*“En el fondo, los científicos somos gente con suerte: podemos jugar a lo que queramos
durante toda la vida”.*
Lee Smolin Físico teórico y cosmólogo.

*“Aquel que le gusta la práctica sin la teoría, es como el marino que navega barco sin
timón ni brújula y nunca sabe dónde anclar”.*
Leonardo Da Vinci (1452-1519). Artista florentino.

Agradecimientos:

A mi madre, mis hermanos, mi abuela y toda mi familia, que siempre han creído en mi, y que han estado ahí cuando más los he necesitado.

A Mariano que nunca ha dejado de creer en mi, que siempre me ha dado todo su apoyo, confianza y que me alienta a seguir.

Al Dr. L. Héctor Juárez Valencia, por sus enseñanzas, consejos y sobre todo por la dirección de este trabajo, ya que el es parte primordial de este.

Al Dr. J. Jacobo Oliveros Oliveros, por la coasesoría de este trabajo y por sus comentarios para mejorarlo.

A mis sinodales: Dr. L. Héctor Juárez Valencia, Dr. Joaquín Delgado Fernández, Dr. J. Héctor Morales Bárcenas, por tomarse el tiempo de revisar mi trabajo y orientarme con sus comentarios para mejorarlo.

A mis profesores del posgrado por orientarme, con sus buenos consejos y sus enseñanzas para mejorar como persona y como estudiante.

A mis amigos, por apoyarme cuando los he necesitado y distraerme cuando era necesario.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el apoyo otorgado, el cual fue esencial para realizar mis estudios.

1. Introducción.	11
1.1. Objetivo	11
1.2. Motivación	11
1.3. Algunos Conceptos Básicos	13
1.3.1. Problemas Inversos	13
1.3.2. Problemas Bien y Mal Planteados	14
1.3.3. Operadores de primer y segundo tipo	16
2. Conceptos básicos para la Teoría de Regularización de Problemas Inversos.	19
2.1. Ejemplo Ilustrativo de Problemas Inversos.	19
2.2. Regularización de Problemas Inversos	23
2.2.1. Análisis del Error	30
2.2.2. Grado de Convergencia	35
2.2.3. Método de Regularización Variacional	39
2.2.4. Método de Regularización Iterativa	41
3. La Inversa Generalizada, Operadores Compactos, Regularización y Métodos de Optimización.	47

3.1. La Inversa Generalizada de Moore-Penrose	48
3.1.1. Soluciones de Mínimos Cuadrados y la Inversa de Moore-Penrose	49
3.2. Operadores Lineales Compactos, Sistemas Singulares y SVD	56
3.2.1. Descomposición Singular de Operadores Compactos (SVD)	57
3.3. Operadores de Regularización	61
3.4. Herramientas de la Optimización Numérica.	66
3.4.1. Método de Máximo Descenso	68
3.4.2. Método de Gradiente Conjugado	71
4. Aplicación a problemas inversos para ecuaciones diferenciales parciales.	77
4.1. Ecuación de Calor en 1-D sobre un segmento infinito	77
4.1.1. Problema Inverso	77
4.2. Ecuación de Calor en 1-D sobre un segmento finito	78
4.2.1. Problema Inverso	79
4.3. Ecuación de Calor en 2-D sobre un segmento finito	89
4.3.1. Problema Inverso	90
5. Conclusiones	101
A. Resultados Importantes	103
B. Descomposición en Valores Singulares (SVD)	105
C. Método Implícito	109
D. Métodos Implícitos de direcciones alternantes (ADI)	113
E. Códigos Computacionales	117
Bibliografía	124

Resumen.

Los *problemas inversos* están relacionadas con la determinación de causas de un efecto observado ó deseado, sin embargo existen distintas dificultades en el estudio de estos, una de ellas es que generalmente no satisfacen los postulados de un *buen planteamiento de Hadamard*, es decir, puede suceder que no tenga solución ó que las soluciones no sean únicas. Los *problemas mal planteados* ofrecen dificultades numéricas, debidas la mayoría de las veces a la dependencia discontinua de las soluciones respecto de los datos.

Para abordar el mal planteamiento de los problemas inversos se utilizan los *métodos de regularización*, los cuales buscan restaurar la estabilidad de las soluciones con relación a los datos. El objetivo de este trabajo es revisar la teoría de *Regularización de Problemas Mal Planteados*, así como, ejemplificar las dificultades que se presentan en la solución de problemas inversos y la forma en que son resueltas. Así mismo, aplicaremos la teoría vista en este trabajo abordando el problema de calor con retroceso en el tiempo para 1 y 2 dimensiones.

1.1. Objetivo

El presente trabajo de tesis tiene como propósito estudiar la teoría general existente sobre regularización de problemas mal planteados. Asimismo se pretende hacer una revisión de los principales métodos de regularización. Por último, se realizarán los códigos computacionales de dichos métodos y se aplicarán al problema de calor con retroceso en el tiempo para 1 y 2 dimensiones.

1.2. Motivación

Normalmente en matemáticas pensaríamos que los problemas se presentan de forma directa, es decir, tenemos una ecuación y deseamos encontrar la solución, o bien dada una causa, y un modelo asociado, deseamos conocer el efecto. Sin embargo, los problemas no siempre se presentan de esta manera, ya que en ocasiones tenemos la solución o efecto de un modelo, pero no sabemos que es lo que esta causando que el modelo nos de tal efecto. Otro problema que se puede presentar es que dada la causa y el efecto, queremos identificar el modelo. A estos dos problemas se les conocen como problemas inversos.



Un ejemplo sencillo para ilustrar estos conceptos, es el sistema lineal

$$Ax = b, \quad \text{con } A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n.$$

El problema directo es encontrar la solución b (efecto) a partir de x (la causa) y de A (el modelo); mientras que un problema inverso consiste en encontrar x dados A y b .

Los problemas inversos siempre han estado presentes en la ciencia, por ejemplo Aristóteles decía que la Tierra es redonda basándose en un razonamiento inverso. Él observó que las formas que la luna presentaba cada mes son del tipo recto, concavo y convexo, y además notó que durante los eclipses lunares, cuando la sombra de la Tierra se proyecta sobre la Luna, la línea del cono de sombra es siempre curva, lo cual es debido a la forma de la Tierra. En [8], se puede ver con mayor detalle este ejemplo.



Regresemos al ejemplo del sistema lineal. Notemos que la solución es única para el problema directo. Sin embargo, para el problema inverso puede suceder que la solución no exista, tenga una infinidad de soluciones, y en el mejor de los casos que tenga solución única. No obstante, que tenga solución única no garantiza que la solución del problema se pueda obtener computacionalmente en forma estable. Cuando se presenta alguno de estos casos decimos que el problema es mal planteado.

Como se mencionó anteriormente, existen distintas dificultades en el estudio de los problemas inversos, una de ellas es que generalmente no satisfacen los postulados de un buen planteamiento de Hadamard [6]. Los problemas mal planteados ofrecen dificultades numéricas, debidas la mayoría de las veces a la dependencia discontinua de las soluciones respecto de los datos.

Definición 1.2.1. (Postulados de Hadamard).

- i) Para todo dato, la solución existe.*
- ii) Para todo dato, la solución es única.*
- iii) La solución depende continuamente de los datos.*

Para abordar el mal planteamiento de los problemas inversos se utilizan los métodos de regularización, los cuales buscan restaurar la estabilidad de las soluciones con relación a los datos. En términos generales la regularización es una aproximación de un problema mal planteado por medio de una familia cercana de problemas bien planteados.

Algunos avances en la ciencia y la tecnología han sido posibles gracias a la solución de problemas inversos, y por tanto, el campo de estos problemas es uno de los que han

tenido un mayor crecimiento en matemáticas aplicadas e industriales. Algunos problemas involucran la determinación de las leyes físicas a través de observaciones indirectas, sensores remotos, astronomía, la reconstrucción de hechos pasados de la observación de la situación actual, entre otros. No obstante, el crecimiento en la investigación de los problemas inversos también se debe al desarrollo de computadoras más poderosas y a métodos numéricos más efectivos para la solución de los problemas asociados.

En este primer capítulo realizaremos una introducción detallada sobre la teoría elemental de Regularización de problemas mal planteados basándonos en la bibliografía estándar. Así mismo, describiremos que es un problema inverso mostrando ejemplos que ilustren el mal planteamiento en este tipo de problemas.

En el Capítulo 2, describiremos la idea del mal planteamiento y regularización formalizando algunos conceptos básicos y enfocándonos en un ejemplo donde mostraremos gráfica y analíticamente que dificultades se presentan en la solución de problemas inversos. Usando la descomposición en valores singulares (SVD) mostraremos la inestabilidad de los problemas inversos, y además nos enfocamos en dos métodos básicos de regularización: TSVD (SVD truncado) y de Tijonov para resolver estas dificultades. Así mismo, veremos de manera general la relación entre mínimos cuadrados y la regularización de problemas inversos. Cabe mencionar que existen diversos métodos de regularización distintos a los que vamos a estudiar en este trabajo uno de ellos son los métodos de regularización estadísticos los cuales se pueden revisar en [23].

En el Capítulo 3, realizaremos una revisión de la teoría de Operadores Compactos, la inversa generalizada y métodos de optimización, para así tener herramientas para resolver los problemas inversos que abordaremos en el Capítulo 4, los cuales consisten en solucionar el problema de calor con retroceso en el tiempo para 1 y 2 dimensiones.

1.3. Algunos Conceptos Básicos

1.3.1. Problemas Inversos

Los modelos matemáticos son ampliamente usados para la descripción e investigación de diferentes procesos y fenómenos. Una clase importante de problemas son los modelos matemáticos con características desconocidas, las cuales deben ser obtenidas en función de los datos observados; a este proceso se le llama *problema inverso*. Dichos problemas surgen en los casos en que el proceso en sí es inaccesible para la observación o bien es demasiado caro acceder a tal observación. En otras palabras, los problemas inversos deben determinar las causas si se conocen los resultados de las observaciones. Veamos algunos ejemplos.

Ejemplo 1.3.1. *Supongase que tenemos un cierto cuerpo en un espacio inaccesible para nosotros. Supongase también que podemos iluminar este cuerpo desde diferentes puntos y registrar la sombra de este cuerpo. Con estas suposiciones tenemos la siguiente configuración matemática del problema inverso: determinar la forma del cuerpo si las proyecciones ortogonales en diferentes planos son conocidas.*

Ejemplo 1.3.2. *Gran parte del trabajo en la medición de campos físicos puede describirse como se indica a continuación. En la entrada tenemos la señal $x(t)$, en la salida registramos la función $y(t)$. El modelo lineal simple es definido por la fórmula*

$$\int_0^t K(t, \tau)x(\tau) = y(t), \quad (1.1)$$

donde $K(t, \tau)$ es una función conocida. Así, el problema inverso consiste en resolver la ecuación integral de Volterra de primer orden (1.1), donde la función $K(t, \tau)$ y $y(t)$ son dadas y $x(t)$ es desconocida.

El aspecto importante de los problemas inversos es que la información inicial es conocida solo en forma aproximada, dado que surge como resultado de un experimento. Esto es debido a que, los aparatos con los cuales se mide la información tienen un cierto nivel de incertidumbre. De modo que, los métodos para resolver los problemas inversos deben de tener una cierta estabilidad con respecto a perturbaciones en los datos iniciales.

1.3.2. Problemas Bien y Mal Planteados

Los problemas inversos generalmente no satisfacen los postulados del buen planteamiento en el sentido de Hadamard. Estos postulados los introdujo Hadamard en el contexto de estudio de los problemas de la física matemática. A continuación reescribimos los postulados de Hadamard de manera formal en el contexto de espacios métricos [6]:

Sea X, Y espacios métricos, $T : X \rightarrow Y$ un operador lineal y $x \in X$ el dato inicial, tal que

$$Tx = y, \quad y \in Y. \quad (1.2)$$

El problema (1.2) es llamado *bien-planteado* en los espacios Y, X , si satisface las siguientes condiciones:

- (i) Para cada $x \in X$ la solución del problema existe.
- (ii) Para cada $x \in X$ la solución del problema es única.
- (iii) La solución y del problema depende continuamente de los datos, es decir, hay estabilidad respecto de los datos.

Un problema se denomina *mal-planteado* si al menos no cumple una de las condiciones anteriores. No obstante, el buen y mal planteamiento de los problemas no dependen únicamente del operador, también puede depender del espacio X donde el dato inicial es dado. Veamos algunos ejemplos.

Ejemplo 1.3.3. *Considere la ecuación integral de Fredholm de primer tipo*

$$\int_a^b K(t, s)x(s)ds = y(x), \quad c \leq t \leq d, \quad (1.3)$$

donde $K(t, s)$ y $y(t)$ son dados, el problema consiste en encontrar $x(s)$. Supongamos que $K(t, s)$, $K_t(t, s)$, $K_s(t, s)$ son continuas en el rectángulo $c \leq t \leq d$, $a \leq s \leq b$, $y(t) \in \mathcal{C}[c, d]$ y $x(s) \in \mathcal{C}[a, b]$. El problema es mal planteado, ya que falla la condición (iii), como se demuestra a continuación:

La prueba de que la condición (iii) falla para la ecuación (1.3) es la siguiente, consideremos la sucesión

$$x_n(s) = x_0(s) + n \sin(n^2 s), \quad n = 0, 1, \dots$$

tal que

$$y_n(t) = \int_a^b K(t, s)x_n(s)ds \quad n = 0, 1, \dots$$

Ahora vamos a estimar $\|y_n - y_0\|_{\mathcal{C}[c, d]}$.

$$\begin{aligned} |y_n(t) - y_0(t)| &= \left| \int_a^b K(t, s)x_n(s)ds - \int_a^b K(t, s)x_0(s)ds \right| \\ &\leq \left| \int_a^b K(t, s)n \sin(n^2 s)ds \right| \\ &= \left| \left[-\frac{1}{n}K(t, s) \cos(n^2 s) \right]_a^b + \frac{1}{n} \int_a^b K_s(t, s) \cos(n^2 s)ds \right| \\ &\leq \frac{1}{n}k_1 \end{aligned}$$

donde $K_s(t, s)$ denota la derivada parcial de $K(t, s)$ respecto de s , y la constante k_1 no depende de n , entonces

$$\begin{aligned} \|y_n - y_0\|_{\mathcal{C}[c, d]} &\leq \frac{1}{n}k_1 \quad n = 1, 2, \dots \\ \lim_{n \rightarrow \infty} \|y_n - y_0\|_{\mathcal{C}[c, d]} &= 0. \end{aligned}$$

Luego, para n suficientemente grande y_0 y y_n se encuentran arbitrariamente cerca una de otra. Por otra parte, observamos que para las funciones $x_n(s)$

$$\lim_{n \rightarrow \infty} \|x_n - x_0\|_{\mathcal{C}[a, b]} = \lim_{n \rightarrow \infty} \|n \sin(n^2 s)\|_{\mathcal{C}[a, b]} = \infty.$$

Por lo tanto, se concluye que no hay estabilidad respecto de los datos.

Ejemplo 1.3.4. Consideremos el siguiente sistema de ecuaciones lineales

$$Ax = y, \quad (1.4)$$

donde A es una matriz de $n \times n$, $y, x \in \mathbb{R}^n$.

Si $\det A \neq 0$, entonces para cada vector $y \in \mathbb{R}^n$ existe una única solución de (1.4): $x \in \mathbb{R}^n$ tal que $x = A^{-1}y$, donde A^{-1} es la matriz inversa de A . De esta solución se sigue que la solución depende continuamente de los datos. En el caso donde el $\det A = 0$, el sistema (1.4) no tiene solución para cada $y \in \mathbb{R}^n$, y si, para algún y_0 la solución existe, entonces esta no es única. Por lo tanto, si el $\det A = 0$, el problema es mal planteado.

1.3.3. Operadores de primer y segundo tipo

La mayoría de los problemas inversos son formulados como en (1.2), es decir

$$Tx = y.$$

donde T es un operador que puede ser lineal o no lineal. Este problema es bien planteado si cumple con las condiciones del buen planteamiento mencionadas anteriormente. Estas condiciones implican que el problema (1.2) es bien planteado si el operador inverso T^{-1} existe y es continuo en Y . Los siguientes dos teoremas son útiles para poder describir los operadores de primer y segundo tipo en el caso lineal.

Teorema 1.3.1. Sea X y Y espacios de Banach y $T : X \rightarrow Y$ un operador lineal acotado e inyectivo, entonces existe el operador inverso $T^{-1} : Y \rightarrow X$.

De este teorema, se sigue que para un operador lineal continuo e inyectivo T , que mapea sobre espacios de Banach, el problema de resolver (1.2) es bien planteado, [2]. Este Teorema muestra, además, una cierta dependencia entre las condiciones de buen-planteamiento y el problema de resolver $Tx = y$ sobre espacios de Banach.

Por otro lado, si T es un operador lineal completamente continuo, entonces la ecuación (1.2) es llamada ecuación del operador lineal de primer tipo. Un ejemplo importante es la ecuación de la integral de Fredholm de primer tipo

$$Tx = \int_a^b K(t, s)x(s)ds = y(t), \quad c \leq t \leq d. \quad (1.5)$$

donde $K(t, s)$ y $y(t)$ son conocidas, y deseamos determinar a $x(s)$. Este problema es mal planteado (ver ejemplo 1.3.3). Cabe mencionar que muchos de los problemas inversos se pueden reducir a operadores de primer tipo, o en un caso particular de la ecuación integral de Fredholm de primer tipo.

Ahora, la ecuación

$$x - Tx = y \tag{1.6}$$

se define como la ecuación del operador lineal de segundo tipo. Donde $T : X \rightarrow X$ es un operador lineal completamente continuo, y X un espacio de Banach. Formularemos el teorema de la solubilidad de la ecuación (1.6) y de la existencia del operador inverso $I - T$, donde I es el operador identidad.

Teorema 1.3.2. *La ecuación (1.6) es soluble para cada $y \in X$ si, y solo si, la ecuación homogénea $x - Tx = 0$ tiene solución única y esta es la trivial $x = 0$. En este caso, el operador inverso $(I - T)^{-1}$ existe.*

De este teorema se sigue que la solución de la ecuación (1.6) es bien-planteada si la ecuación $x - Tx = 0$ tiene solo la solución trivial, [2]. Sin embargo, si la ecuación $x - Tx = 0$ tiene soluciones distintas a la trivial, entonces el problema (1.6) es mal-planteado.

Un ejemplo importante de (1.6) es la ecuación integral de Fredholm de segundo tipo

$$x(t) - \int_a^b K(t, s)x(s)ds = y(t), \quad a \leq t \leq b. \tag{1.7}$$

Aquí, el kernel $K(t, s)$ satisface la condición

$$\int_a^b \int_a^b K^2(t, s)dsdt < \infty,$$

y $x, y \in L_2[a, b]$.

Conceptos básicos para la Teoría de Regularización de Problemas Inversos.

En este capítulo vamos a introducir la idea general del término de “mal planteamiento” y “regularización”. Por ejemplo, en el modelado de flujo de aguas subterráneas, se estiman parámetros del material de un acuífero a partir de mediciones de la presión de un fluido que sumerge al acuífero. Desafortunadamente, un poco de ruido en los datos puede hacer que las estimaciones tengan errores enormes, lo que provoca el mal planteamiento del problema inverso. Afortunadamente, se han desarrollado técnicas matemáticas conocida como los métodos de regularización, lo cual hacen más estables las estimaciones.

También veremos un enfoque muy general de la relación de mínimos cuadrados y regularización de problemas inversos. En el siguiente capítulo vamos a formalizar algunos de los conceptos que veremos en este.

2.1. Ejemplo Ilustrativo de Problemas Inversos.

Consideremos la ecuación integral de convolución de primer tipo de Fredholm:

$$g(x) = \int_0^1 k(x-s)f(s)ds \equiv (Kf)(x), \quad 0 < x < 1 \quad (2.1)$$

Esta es la versión unidimensional de un modelo de imagen óptica en dos dimensiones. En este modelo, f representa la intensidad de la fuente de luz en función de la posición espacial, y g representa la intensidad de la imagen. El kernel k caracteriza los efectos de difuminación (efectos borrosos) que se producen durante la formación de la imagen. El Gaussiano

$$k(x) = C \exp \left\{ -\frac{x^2}{2\gamma^2} \right\} \quad (2.2)$$

donde C y γ son parámetros positivos, es el kernel que modela los efectos promedio en tiempos largos de la turbulencia atmosférica sobre la propagación de luz, [27].

El problema directo (forward problem), asociado con el modelo de la ecuación (2.1) es el siguiente: Dada la fuente f y el kernel k , determine la imagen borrosa g .

El problema inverso es el de nuestro interés y, es el siguiente: Dado el kernel k y la imagen borrosa g , determinar la fuente f . A primera vista, la solución a este problema parece directo; podemos simplemente discretizar la ecuación (2.1) para obtener un sistema lineal discreto $Kf = d$. Por ejemplo, aplicando la cuadratura del punto medio al operador K , sea $h = 1/n$ y $x = x_j = jh$, donde $1 \leq j \leq n$. Entonces, obtenemos

$$\begin{aligned} (Kf)(x_i) &= h \int_0^1 k(x_i - s)f(s)ds \\ &\approx h \sum_{j=1}^n k(x_i - x_j)f(x_j) \\ &= h \sum_{j=1}^n k(ih - jh)f(jh) \\ &= h \sum_{j=1}^n C \exp\left(-\frac{((i-j)h)^2}{2\gamma^2}\right)f(x_j) \\ &= Kf \end{aligned}$$

por tanto K tiene entradas

$$[K]_{ij} = hC \exp \left\{ -\frac{((i-j)h)^2}{2\gamma^2} \right\}, \quad 1 \leq i, j \leq n. \quad (2.3)$$

Si la matriz K es no singular, entonces podemos hacer la aproximación discreta de $K^{-1}d$ para calcular f . Para obtener una buena aproximación de la cuadratura, n debe ser relativamente grande. A pesar de que K es invertible, su número de condición crece conforme n crece, así que los posibles errores introducidos en d se pueden amplificar mucho cuando se resuelva el sistema. Ciertos errores, como aquellos debidos a la cuadratura, pueden ser controlados. Otros, como el ruido en el dispositivo de la grabación de la imagen, no se pueden controlar en un ajuste práctico. Consecuentemente, este enfoque directo de la solución es probable que falle. El siguiente ejemplo muestra que así es.

Asumimos que la solución exacta del sistema discreto

$$Kf = d \quad (2.4)$$

es

$$f_{true} = \begin{cases} 0.75, & 0.1 < x < 0.25, \\ 0.25, & 0.3 < x < 0.32, \\ \sin^4(2\pi x), & 0.5 < x < 1, \\ 0, & \text{en otro caso.} \end{cases} \quad (2.5)$$

Ahora, generamos los n datos de la imagen borrosa d_{true} , resolviendo el problema directo (2.4), considerando $x_j = jh$ en el intervalo $[0, 1]$ y agregando ruido a los datos, es decir, $d = d_{true} + \eta$, donde $|\eta_i| < 0.01$, con parámetros $\gamma = 0.05$ y $C = 1/\gamma\sqrt{2\pi}$.

En la Figura 2.1 se ilustra la función fuente $f_{true}(x)$, y la imagen borrosa d_{true} con y sin ruido, asumiendo $n = 56$.

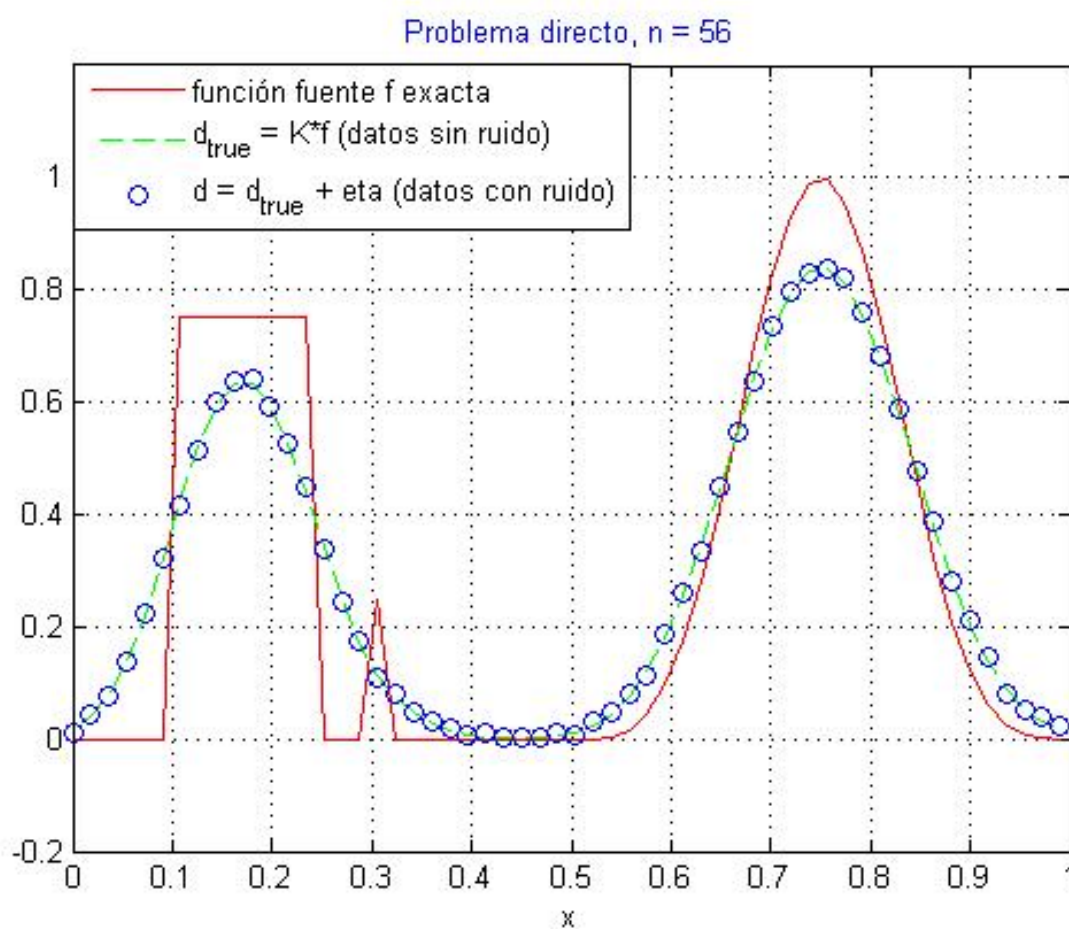


Figura 2.1: Problema Directo

Ejemplo 2.1.1. Supongase $n = 56$ datos, resolvemos el problema inverso del sistema discreto (2.4) cuando se parte de los datos d con ruido. En la Figura 2.2 se muestra que existe un problema de inestabilidad en la aproximación de f . Esto se debe al error que se introduce al calcular la matriz K , y a que su número de condición es muy grande.

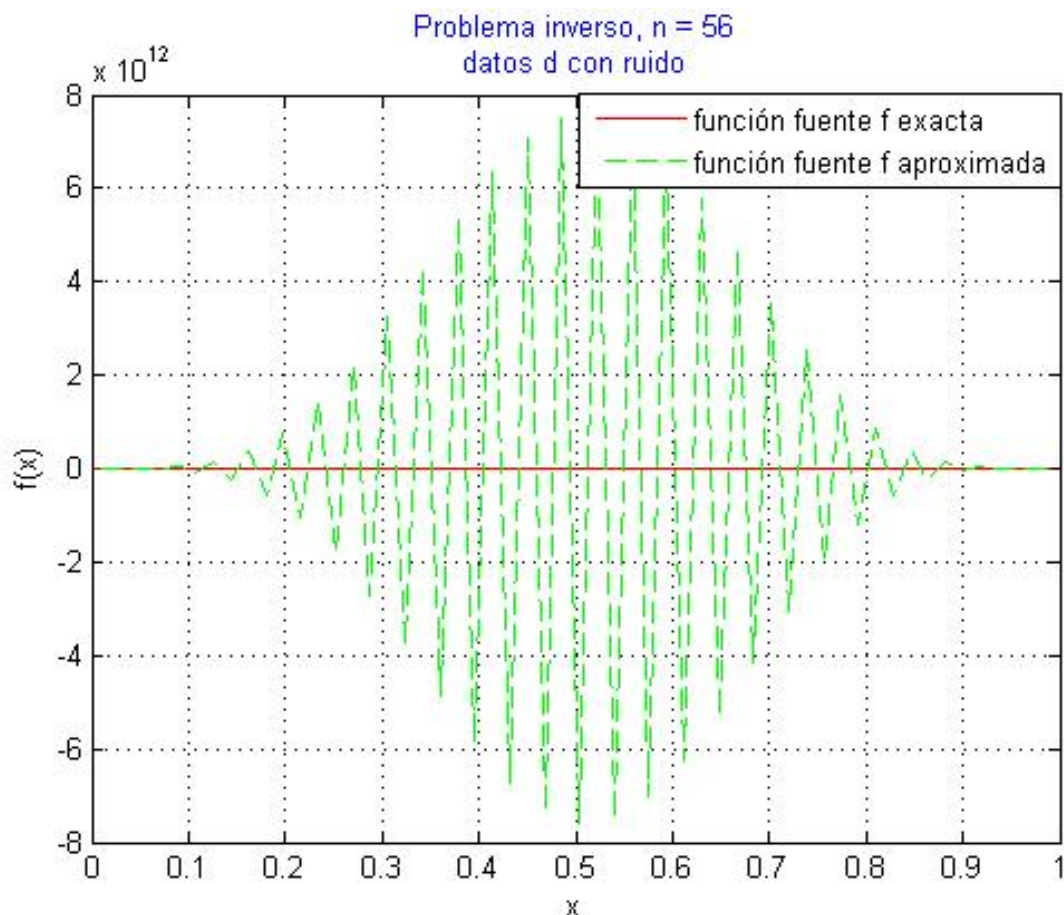


Figura 2.2: Problema Inverso sin Regularización

Este es un claro ejemplo de la inestabilidad que se puede presentar cuando resolvemos un problema inverso. De hecho, al resolver el problema inverso con datos con muy poco "ruido", el resultado aproximado es enormemente lejano al verdadero. La inestabilidad se manifiesta mediante una amplificación tremenda del "ruido" en el proceso de solución.

Nota: En el ejemplo anterior, el sistema $f = K^{-1}d$ se resuelve utilizando la descomposición en valores singulares (SVD) de la matriz K , es decir $K = U\Sigma V^T$ con U y V matrices ortogonales y Σ la matriz diagonal de valores singulares de K . Por lo tanto:

$$f_{true} = V\Sigma^{-1}U^T \quad (2.6)$$

En el Apéndice B se incluye una descripción de la descomposición en valores singulares de una matriz.

2.2. Regularización de Problemas Inversos

Como vimos en el ejemplo anterior un pequeño error en los datos produce una gran desviación de la solución exacta del problema inverso. Esta inestabilidad surge debido a la división por los pequeños valores singulares (σ_i) de K , los cuales amplifican el error en los datos durante el proceso de solución, como se muestra a continuación:

Los datos d son de la forma

$$d = K f_{true} + \eta, \quad (2.7)$$

donde η es el ruido introducido por errores de medición. Como K es invertible y $d_{true} = K f_{true}$, entonces

$$\begin{aligned} f &= K^{-1}d \\ &= f_{true} + K^{-1}\eta \\ &= f_{true} + V\Sigma^{-1}U\eta \\ &= f_{true} + \sum_{i=1}^n \sigma_i^{-1} (u_i^T \eta) v_i. \end{aligned} \quad (2.8)$$

Como puede observarse el ruido η , por muy pequeño que sea, es amplificado en la proporción de $1/\sigma_i$, el cual puede ser muy grande si σ_i es muy pequeño, es decir, si K es mal condicionada. Por lo tanto, la amplificación del error

$$e = f - f_{true} \quad (2.9)$$

debido a σ_i^{-1} puede llegar a ser desastroso, y es lo que provoca el crecimiento desmedido del error (explosión) en la solución del problema inverso anterior. Para controlar el crecimiento de este error es necesario controlar el crecimiento de los términos σ_i^{-1} cuando σ_i es pequeño. Esto se logra mediante una técnica denominada regularización, que consiste en multiplicar σ_i^{-1} en (2.8) por una función filtro de regularización $w_\alpha(\sigma_i^2)$, para la cual el producto $w_\alpha(\sigma_i^2)\sigma_i^{-1} \rightarrow 0$ cuando $\alpha \rightarrow 0$. Esta función filtra (desecha) las componentes de $K^{-1}d$ en (2.8) correspondientes a valores singulares pequeños y produce una aproximación f_α a f_{true} de la forma:

$$\begin{aligned} f_\alpha &= V \text{diag}(w_\alpha(\sigma_i^2)\sigma_i^{-1}) U^T d \\ &= \sum_{i=1}^n w_\alpha(\sigma_i^2)\sigma_i^{-1} (u_i^T d) v_i, \end{aligned} \quad (2.10)$$

donde α es llamado el parámetro de regularización o valor umbral. Para retener algún nivel de precisión, se deben retener las componentes singulares correspondientes a valores singulares grandes y desechar o amortiguar las componentes con valores singulares pequeños en (2.10). Esto se logra tomando $w_\alpha(\sigma_i^2) \approx 1$ para valores σ_i grandes. Un ejemplo de filtro es el denominado filtro TSVD (SVD truncado), dado por:

$$f_\alpha = \sum_{\sigma_i^2 > \alpha} \sigma_i^{-1} (u_i^T d) v_i.$$

El filtro más común es el denominado filtro de Tijonov:

$$f_\alpha = \sum_{i=1}^n \frac{\sigma_i^2}{\sigma_i^2 + \alpha} \sigma_i^{-1} (u_i^T d) v_i$$

Ejemplo 2.2.1. Con el objeto de ilustrar el efecto de la regularización para resolver el problema $Kf = d$ (2.4), consideramos el caso cuando $n = 80$ en (2.3).

La matriz K construida es muy mal condicionada y cualquier intento por resolver el sistema (2.4) sin regularización produce los resultados mostrados en la Figura 2.2. De hecho, en la Figura 2.3 se muestran los valores singulares de la matriz construida. Se observa que los valores singulares de menor valor son extremadamente pequeños, del orden de 10^{-17} , y amplificaran tremendamente el ruido en los datos, por muy pequeño que este sea. Por otro lado, en la Figura 2.4 se muestran algunos de los vectores singulares v_i de la matriz K , concretamente, se muestran las gráficas de v_1 , v_5 , v_{15} y v_{80} . Se observa que los modos de valor pequeño son suaves y de baja frecuencia (pocas oscilaciones), mientras que los vectores singulares de mayor valor, como v_{80} , son de alta frecuencia (altamente oscilatorios). Los vectores singulares de frecuencia alta, corresponden a los valores singulares más pequeños, y afectan la aproximación degradandola en mayor grado que los vectores singulares de menor frecuencia.

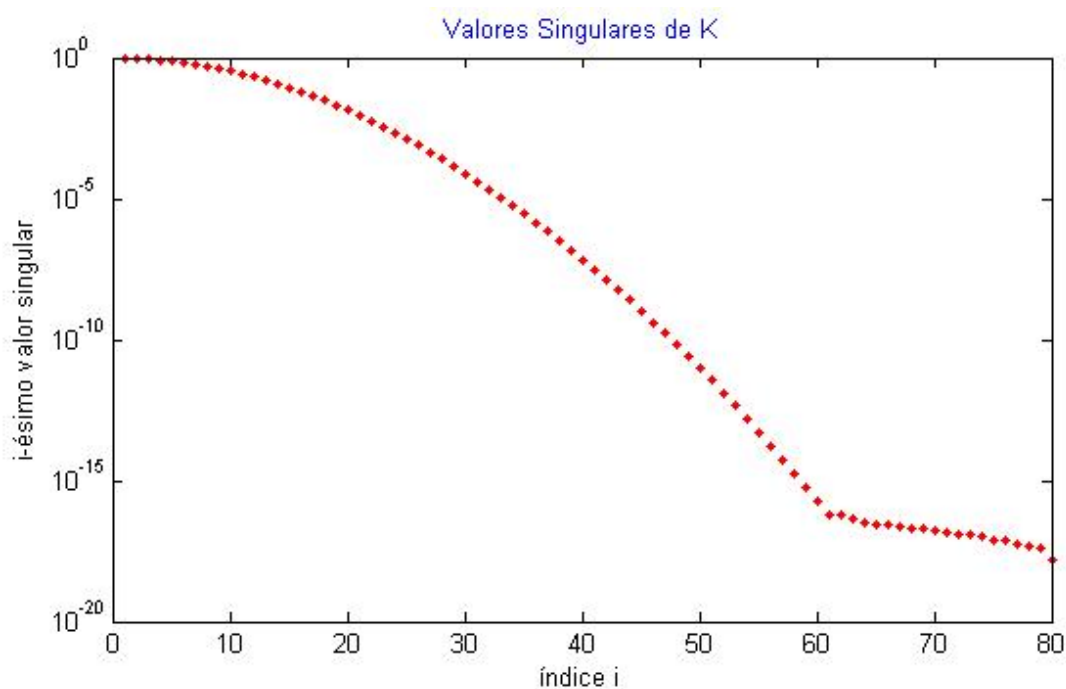


Figura 2.3: Valores Singulares de K para $n = 80$

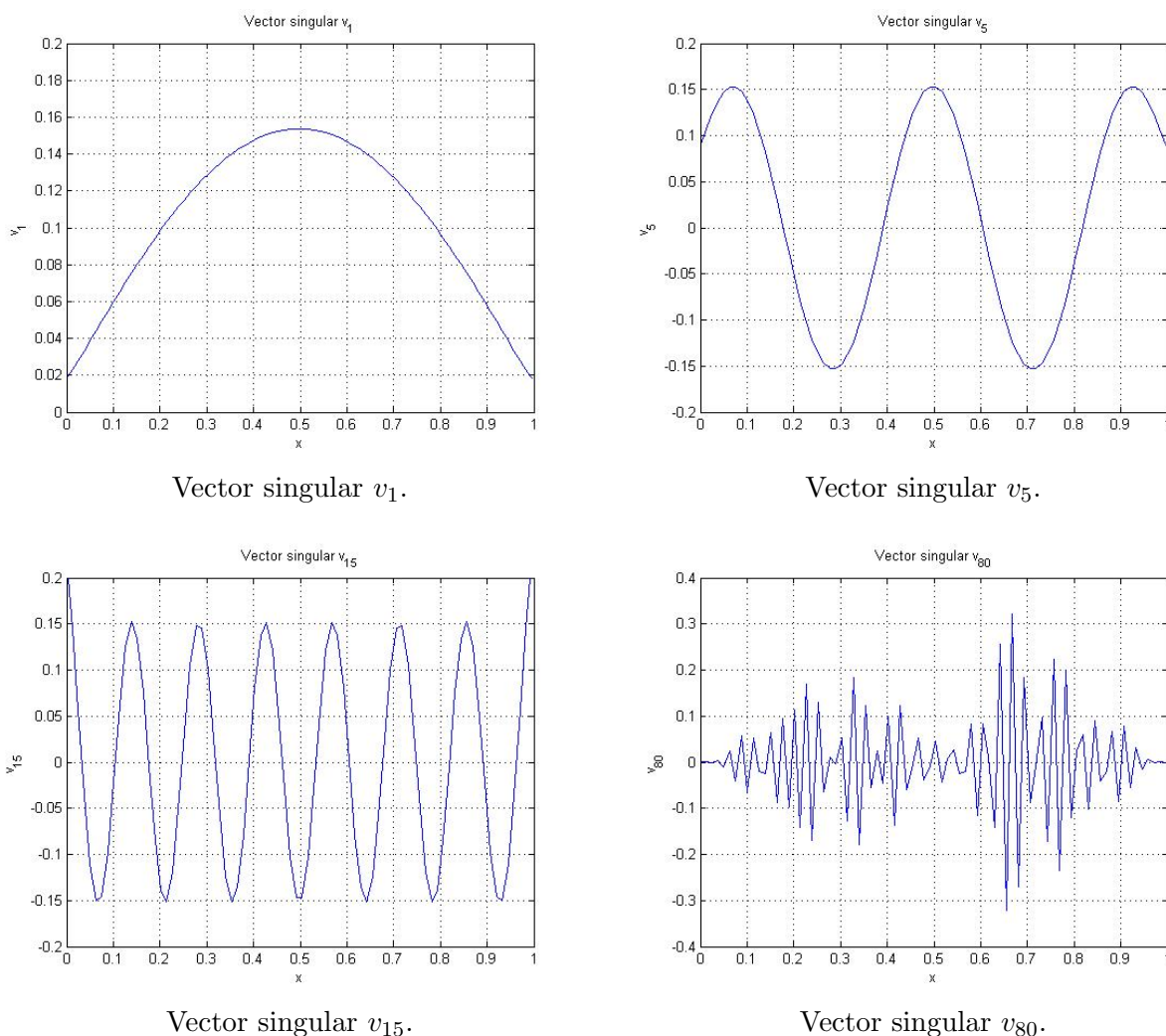


Figura 2.4: Algunos vectores singulares: v_1 , v_5 , v_{15} , v_{80} .

Resultados utilizando la regularización TSVD

Al utilizar el filtro SVD truncado (TSVD), la aproximación de la ecuación (2.10) toma la forma

$$f_\alpha = \sum_{\sigma_i^2 > \alpha} \sigma_i^{-1} (u_i^T d) v_i \quad (2.11)$$

la cual depende del valor $\alpha > 0$, pues este valor indica el valor singular a partir del cual truncamos la serie, permitiendo mayor ó menor regularización del problema. Entre más pequeño se escoja α , menor regularización se obtiene y la solución tendrá oscilaciones fuertes, mientras que a mayor valor de α , mayor regularización se obtiene y la solución aproximada será más suave. En la Figura 2.5 se muestra la gráfica de la solución aproximada f_α considerando $n = 56$ para diferentes valores de α : 0.8, 0.01, 0.001, 10^{-4} , 10^{-5} y 10^{-6} , donde se comprueba que, efectivamente, a mayor α , mayor regularización y mayor suavidad en la solución. El menor error $\|f_\alpha - f_{true}\|$ se obtiene con $\alpha \approx 10^{-2}$. En la Figura 2.6 se muestra la dependencia del error $\|f_\alpha - f_{true}\|$ con respecto al parámetro de regularización α .

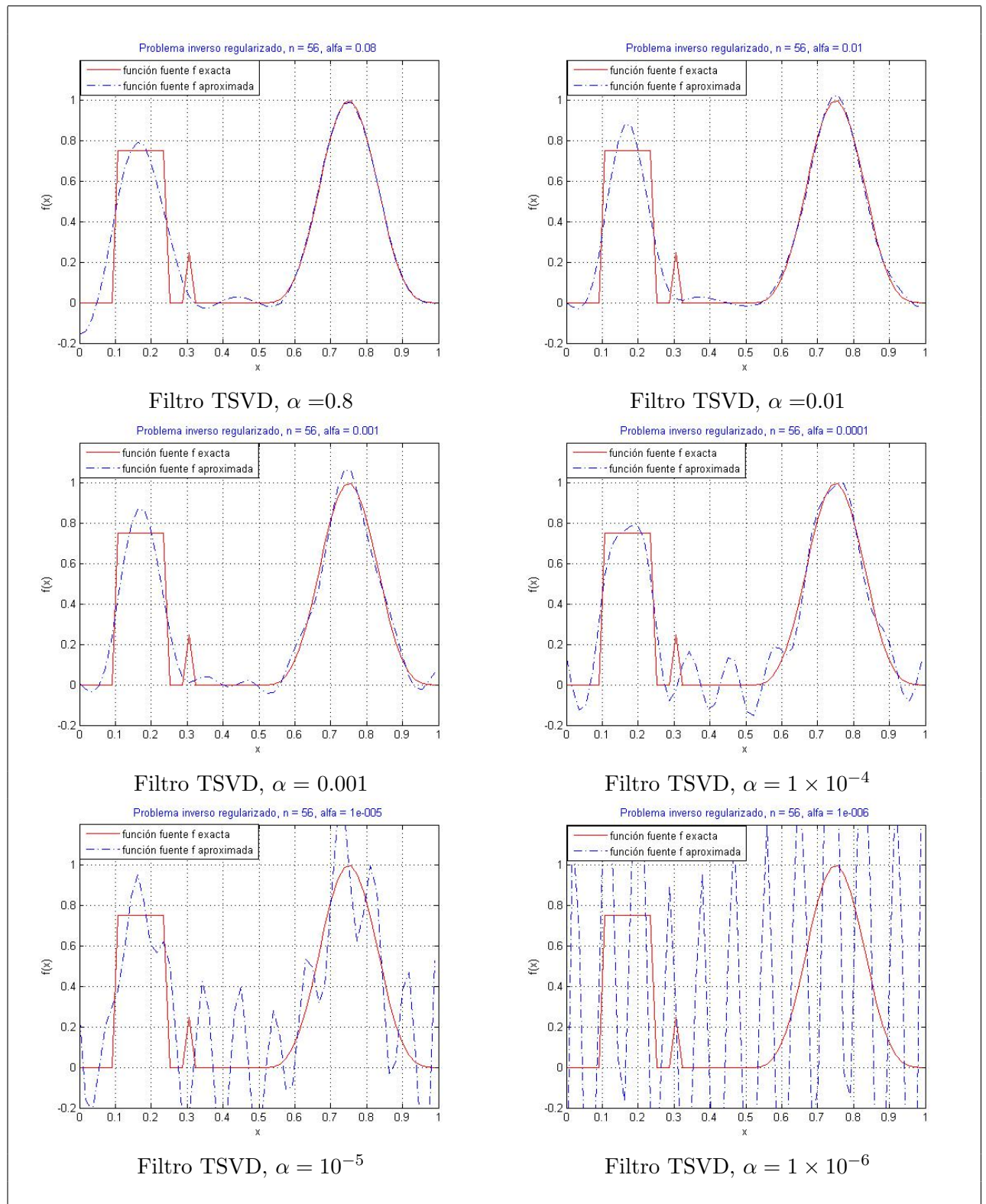


Figura 2.5: Regularización por Filtros de TSVD

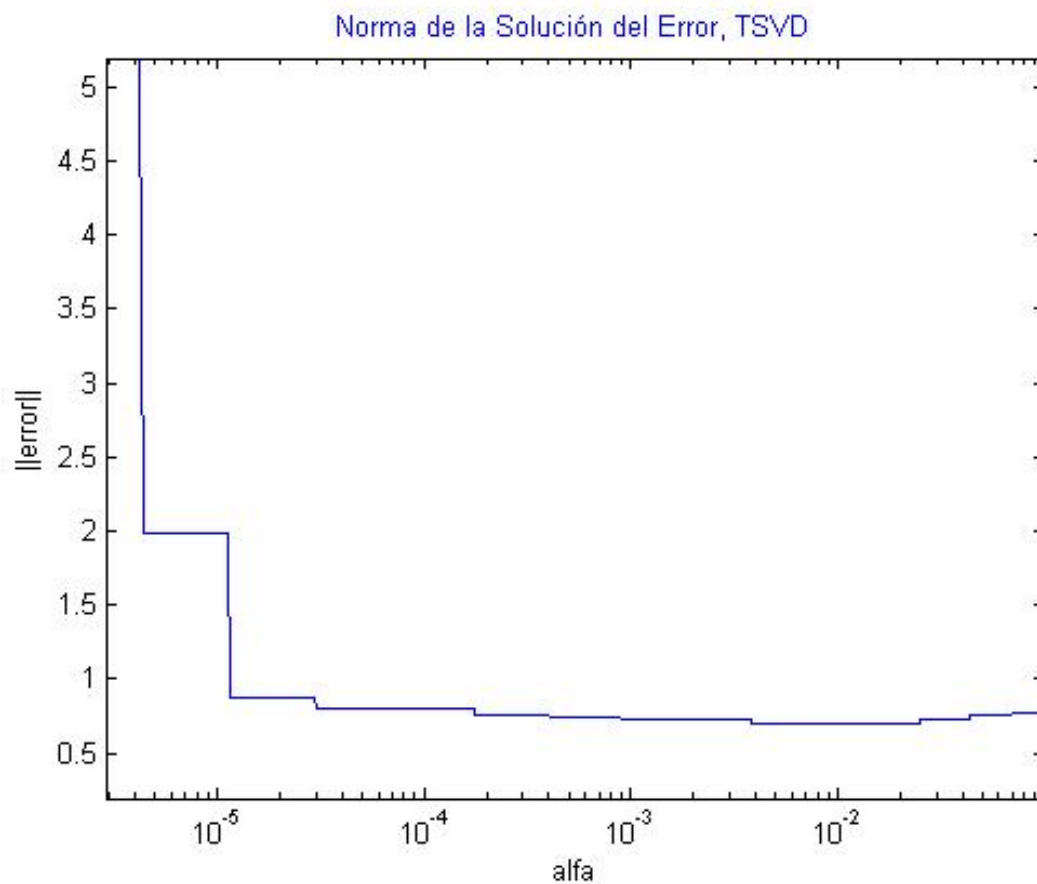


Figura 2.6: Norma del error de Solución de TSVD $\|f_\alpha - f_{true}\|$.

Resultados utilizando regularización de Tijonov

En este caso, la aproximación de la ecuación (2.10) toma la forma

$$\begin{aligned}
 f_\alpha &= \sum_{i=1}^n \frac{\sigma_i^2}{\sigma_i^2 + \alpha} \sigma_i^{-1} (u_i^T d) v_i \\
 &= (K^T K + \alpha I)^{-1} K^T d
 \end{aligned} \tag{2.12}$$

la cual depende del parámetro de regularización. A continuación demostraremos la ecuación 2.12

$$\begin{aligned}
f_\alpha &= \sum_{i=1}^n \frac{1}{\sigma_i^2 + \alpha} \sigma_i (u_i^T d) v_i \\
&= \sum_{i=1}^n \frac{1}{\sigma_i^2 + \alpha} v_i \sigma_i (u_i^T d) \\
&= \sum_{i=1}^n \frac{1}{\sigma_i^2 + \alpha} (V \Sigma U^T) d \\
&= \sum_{i=1}^n \frac{1}{\sigma_i^2 + \alpha} (U \Sigma V^T)^T d \\
&= \sum_{i=1}^n \frac{1}{\sigma_i^2 + \alpha} K^T d \\
&= \sum_{i=1}^n \frac{1}{\sigma_i^2 (v_i v_i^T) (u_i^T u_i) + \alpha} K^T d \\
&= \sum_{i=1}^n \frac{1}{\sigma_i^2 (v_i (u_i^T u_i) v_i^T) + \alpha} K^T d \\
&= \sum_{i=1}^n \frac{1}{(v_i \sigma_i u_i^T) (u_i \sigma_i v_i^T) + \alpha} K^T d \\
&= ((V \Sigma U^T) (U \Sigma V^T) + \alpha I)^{-1} K^T d \\
&= ((U \Sigma V^T)^T (U \Sigma V^T) + \alpha I)^{-1} K^T d \\
&= (K^T K + \alpha I)^{-1} K^T d
\end{aligned}$$

Al igual que en el caso anterior con $\alpha = 0$ en (2.12) se obtiene la solución sin regularización, y a medida que aumentamos el valor de α , se obtiene mayor regularización, por lo que el efecto de inestabilidad en los vectores de frecuencia alta se amortigua, obteniendo soluciones más suaves pero con mayor error. Esto se ilustra en la Figura 2.7, en donde se muestra la solución aproximada f_α considerando $n = 56$ con $\alpha = 0.8, 0.01, 0.001, 10^{-4}, 10^{-5}$ y $\times 10^{-6}$. Con el filtro de Tijonov el mínimo error se obtiene con $\alpha \approx 10^{-3}$, a diferencia del filtro TSVD en donde el mínimo error se obtiene con $\alpha \approx 10^{-2}$. Esto nos indica que el parámetro de regularización óptimo depende del método de regularización. La figura 2.8 muestra la dependencia del error con respecto al parámetro de regularización.

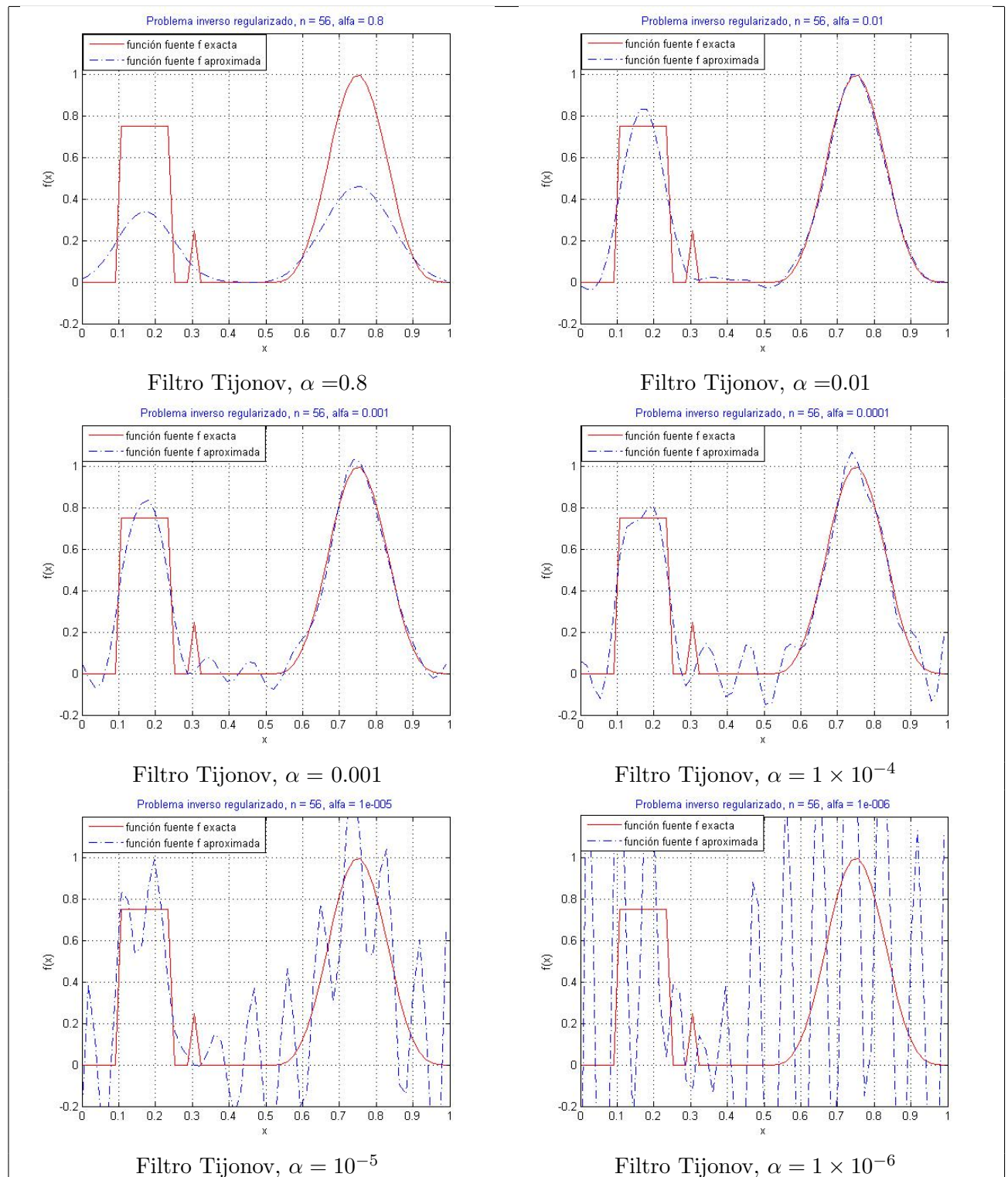


Figura 2.7: Regularización por Filtros Tijonov

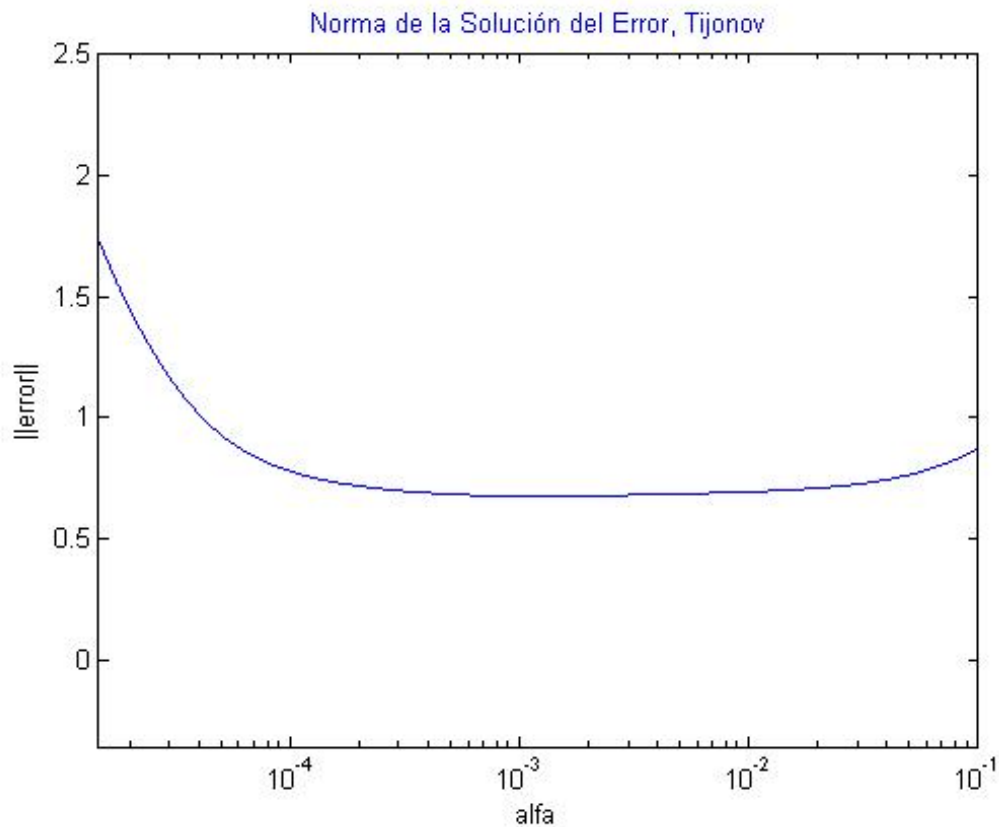


Figura 2.8: Norma del error de Solución de Tijonov $\|f_\alpha - f_{true}\|$.

Los ejemplos anteriores muestran que existen diferencias en la solución cuando aproximamos a f_α por los dos filtros. Sin embargo, debemos tener especial cuidado en el valor del parámetro de regularización, ya que tomar el más pequeño posible no siempre es la mejor opción. En el siguiente apartado llevaremos acabo un análisis para el error de solución.

2.2.1. Análisis del Error

El lado derecho de la ecuación (2.10) define un operador lineal de regularización, el cual denotamos por R_α , es decir,

$$f_\alpha = R_\alpha d = \sum_{i=1}^n w_\alpha(\sigma_i^2) \sigma_i^{-1} (u_i^T d) v_i.$$

Ahora si a $d = d_{true} + \eta$ (donde d son los datos con ruido) le aplicamos el filtro de regularización, obtenemos

$$\begin{aligned} R_\alpha d &= R_\alpha d_{true} + R_\alpha \eta \\ &= R_\alpha (K f_{true}) + R_\alpha \eta. \end{aligned}$$

Por lo tanto,

$$f_\alpha = R_\alpha(K f_{true}) + R_\alpha \eta,$$

y el error de la solución regularizada está dado por:

$$\begin{aligned} e_\alpha &= f_\alpha - f_{true} \\ &= R_\alpha(K f_{true}) + R_\alpha \eta - f_{true} \\ &= \{R_\alpha(K f_{true}) - f_{true}\} + R_\alpha \eta \\ &= e_\alpha^{trunc} + e_\alpha^{ruido}, \end{aligned} \quad (2.13)$$

donde

$$e_\alpha^{ruido} = R_\alpha \eta \quad (2.14)$$

$$= \sum_{i=1}^n w_\alpha(\sigma_i^2) \sigma_i^{-1} (u_i^T \eta) v_i, \quad (2.15)$$

y

$$\begin{aligned} e_\alpha^{trunc} &= R_\alpha(K f_{true}) - f_{true} \\ &= \sum_{i=1}^n w_\alpha(\sigma_i^2) \sigma_i^{-1} (u_i^T K f_{true}) v_i - f_{true} \\ &= \sum_{i=1}^n w_\alpha(\sigma_i^2) \sigma_i^{-1} \{(K^T u_i)^T f_{true}\} v_i - f_{true} \\ &= \sum_{i=1}^n w_\alpha(\sigma_i^2) \sigma_i^{-1} \{(\sigma_i v_i)^T f_{true}\} v_i - f_{true} \\ &= \sum_{i=1}^n w_\alpha(\sigma_i^2) \sigma_i^{-1} \{\sigma_i v_i^T f_{true}\} v_i - f_{true} \\ &= \sum_{i=1}^n w_\alpha(\sigma_i^2) (v_i^T f_{true}) v_i - f_{true}. \end{aligned}$$

Expandiendo f_{true} en $gen\{v_1, v_2, \dots, v_n\}$, donde los v_i son los vectores singulares de K por la derecha, obtenemos

$$f_{true} = \sum_{i=1}^n (v_i^T f_{true}) v_i, \quad (2.16)$$

así que

$$e_\alpha^{trunc} = \sum_{i=1}^n [w_\alpha(\sigma_i^2) - 1] (v_i^T f_{true}) v_i. \quad (2.17)$$

Llamamos e_α^{trunc} al error de truncamiento de la solución debido a la regularización, [27]. Esto cuantifica la información perdida debido al filtro de regularización, y el término e_α^{ruido} es llamado el error por amplificación del ruido. Mostramos que, tanto para el filtro TSVD como para el filtro Tijonov, el parámetro de regularización α puede ser seleccionado de manera que garantice que estos dos errores converjan a cero cuando el nivel del ruido $\delta \rightarrow 0$. Donde

$$\delta = \|\eta\|. \quad (2.18)$$

Primero consideramos el error de truncamiento. Para ambos filtros

$$w_\alpha(\sigma_i^2) = \left\{ \begin{array}{l} \frac{\sigma_i^2}{\sigma_i^2 + \alpha} \quad Tijonov \\ \left\{ \begin{array}{l} 1 \quad \text{Si } \sigma_i^2 > \alpha \\ 0 \quad \text{Si } \sigma_i^2 \leq \alpha \end{array} \right\} \quad TSVD \end{array} \right\} \rightarrow 1 \text{ cuando } \alpha \rightarrow 0$$

y por tanto a partir de la ecuación (2.17)

$$e_\alpha^{trunc} \rightarrow 0 \quad \text{cuando } \alpha \rightarrow 0.$$

Se observa que, en realidad, e_α^{trunc} no depende directamente del nivel del ruido. Ahora, para tratar con el error de amplificación del ruido, primero demostraremos que ambos filtros satisfacen

$$w_\alpha(\sigma_i^2)\sigma_i^{-1} \leq \alpha^{-1/2} \quad (2.19)$$

Para el filtro TSVD

$$w_\alpha(\sigma_i^2) = \begin{cases} 1, & \text{Si } \sigma_i^2 > \alpha, \\ 0, & \text{Si } \sigma_i^2 \leq \alpha. \end{cases}$$

Luego

$$w_\alpha(\sigma_i^2)\sigma_i^{-1} = \begin{cases} \sigma_i^{-1} & \text{Si } \sigma_i^2 > \alpha, \\ 0 & \text{Si } \sigma_i^2 \leq \alpha. \end{cases}$$

Así que

$$w_\alpha(\sigma_i^2)\sigma_i^{-1} = \sigma_i^{-1} \quad \text{si } \sigma_i^2 > \alpha.$$

Es decir

$$\sigma_i > \alpha^{1/2} \quad \text{implica} \quad w_\alpha(\sigma_i^2)\sigma_i^{-1} = \sigma_i^{-1}$$

Por lo tanto

$$\sigma_i^{-1} \leq \alpha^{-1/2} \quad \text{implica} \quad w_\alpha(\sigma_i^2)\sigma_i^{-1} \leq \alpha^{-1/2}.$$

Ahora, para el filtro Tijonov

$$w_\alpha(\sigma_i^2) = \frac{\sigma_i^2}{\sigma_i^2 + \alpha},$$

es decir

$$w_\alpha(\sigma_i^2)\sigma_i^{-1} = \frac{\sigma_i}{\sigma_i^2 + \alpha}.$$

Obsérvese que $\frac{\sigma}{\sigma^2 + \alpha} \leq \alpha^{-1/2}$ si, y solo si, $\alpha^{1/2} \leq \sigma + \frac{\alpha}{\sigma}$. Entonces definimos la siguiente función $q(\sigma) = \sigma + \frac{\alpha}{\sigma}$ y calculamos sus mínimos. Para ello hacemos

$$q'(\sigma) = 1 - \frac{\alpha}{\sigma^2} = 0,$$

y entonces

$$\sigma = \alpha^{1/2}.$$

Como

$$q''(\alpha^{1/2}) = \frac{2\alpha}{\sigma^3} \Big|_{\sigma=\alpha^{1/2}} > 0,$$

entonces $\sigma = \alpha^{1/2}$ es el mínimo, y además es un mínimo estricto. Por lo que

$$q(\alpha^{1/2}) = 2\alpha^{1/2}.$$

Por lo tanto

$$\alpha^{1/2} < 2\alpha^{1/2} \leq \sigma + \frac{\alpha}{\sigma} \quad \forall \sigma,$$

y se concluye que

$$\frac{\sigma}{\sigma^2 + \alpha} \leq \alpha^{-1/2}.$$

Entonces de (2.15), (2.18) y (2.19) llegamos a que

$$\begin{aligned} \|e_\alpha^{\text{ruido}}\|^2 &= \sum_{i=1}^n [w_\alpha(\sigma_i^2)\sigma_i^{-1}]^2 \|\eta\|^2 \\ &\leq \alpha^{-1} \|\eta\|^2 \\ &= \alpha^{-1} \delta^2. \end{aligned}$$

Por lo tanto

$$\|e_{\alpha}^{ruido}\| \leq \alpha^{-1/2}\delta. \quad (2.20)$$

Sin embargo para que $\|e_{\alpha}^{ruido}\| \rightarrow 0$ se debe buscar una relación adecuada entre α y δ . Supongamos que $\alpha = \delta^p$, entonces

$$\alpha^{-1/2}\delta = \alpha^{-1/2}\alpha^{1/p} = \alpha^{\frac{2-p}{2p}},$$

por lo que $\frac{2-p}{2p} > 0$ si $0 < p < 2$. Por lo tanto, si se escoge $\alpha(\delta) = \delta^p$ con $0 < p < 2$, entonces

$$e_{\alpha(\delta)}^{ruido} \rightarrow 0 \quad (2.21)$$

cuando $\delta \rightarrow 0$.

Ejemplo 2.2.2. En la Figura 2.9 se muestra la gráfica de la solución aproximada f_{α} usando la regularización de Tijonov y TSVD. Consideramos $p = 1$, es decir, $\alpha = \delta = \|\eta\|_2$, donde $\|\eta\|_2 = 0.042858$ y $n = 56$.

En este ejemplo, podemos verificar que la desigualdad (2.20) se cumple, ya que $\alpha^{1/2} = 0.2070$ y $\|e_{\alpha}^{ruido}\|$ para Tijonov y TSVD es 0.0381 y 0.0358 respectivamente.

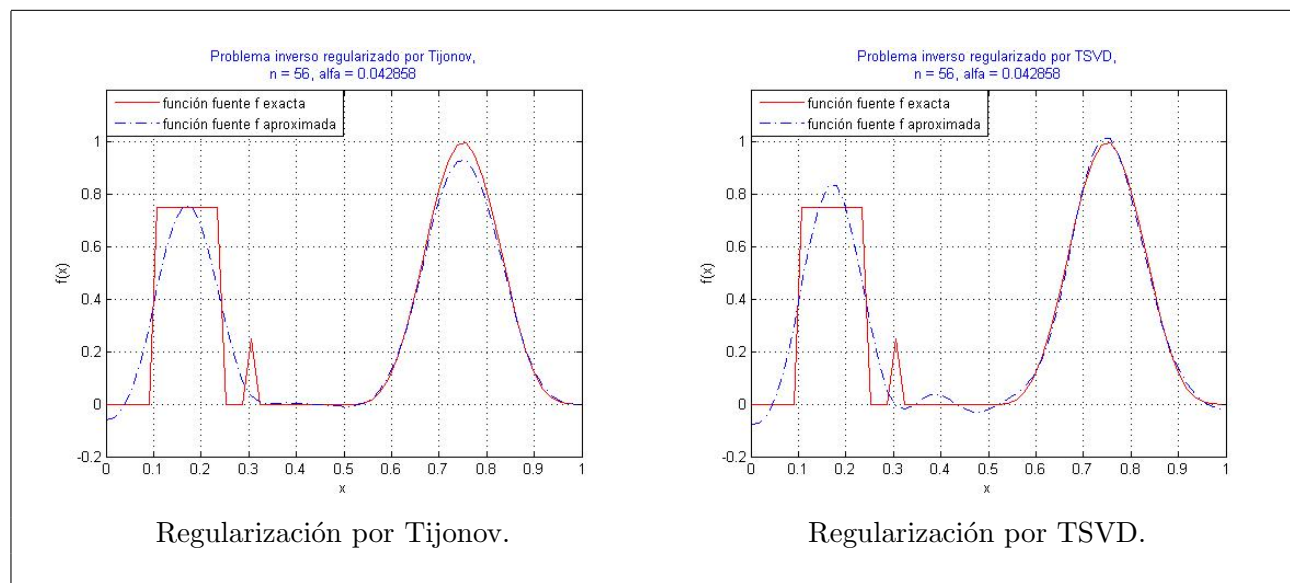


Figura 2.9: Problema Inverso Regularizado, donde $n = 56$ y $\alpha = 0.042858$

Ahora en la Figura 2.10 podemos verificar que si escogemos $\alpha = \delta^p$ con $0 < p < 2$, la desigualdad $\|e_{\alpha}^{ruido}\| \leq \alpha^{\frac{2-p}{p}}$ se cumple tanto para la regularización de Tijonov como la de TSVD.

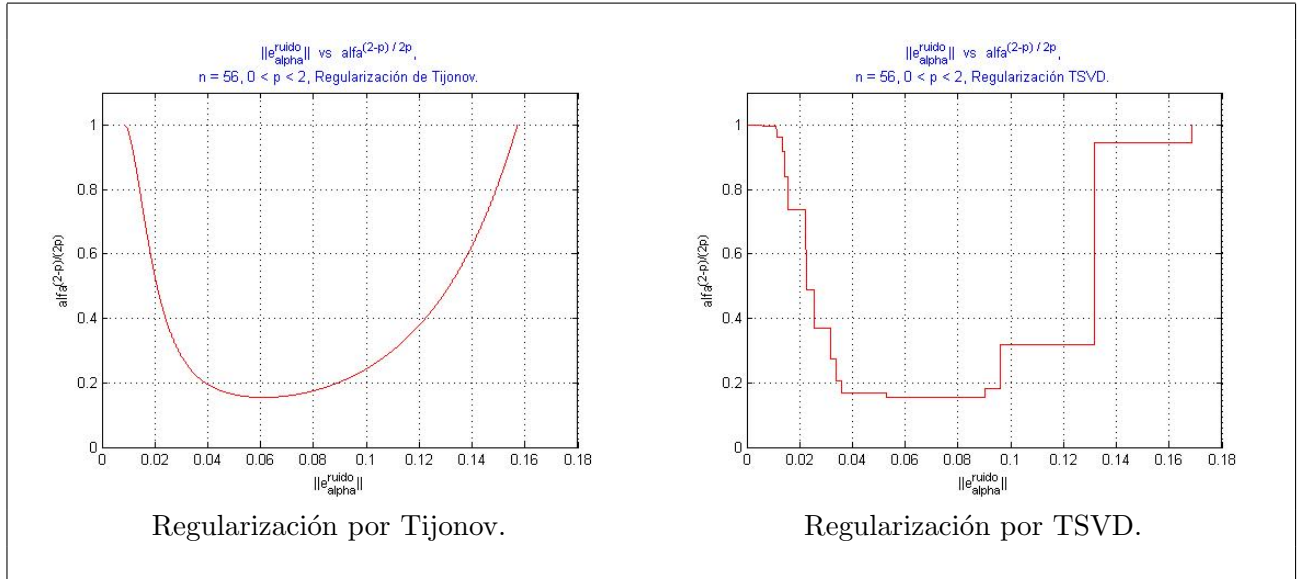


Figura 2.10: $\|e_{\alpha}^{ruido}\|$ vs $\alpha^{\frac{2-p}{p}}$, considerando $n = 56, 0 < p < 2$

2.2.2. Grado de Convergencia

Si consideramos cualquiera de los dos filtros de regularización, y asumimos $\delta \leq \sigma_n^2$, el cuadrado del valor singular más pequeño, entonces el error de amplificación del ruido es tolerable incluso sin regularización. Para obtener el grado de convergencia del error de la solución, necesitamos acotar el error de truncamiento [27]. Asumimos

$$f_{true} = K^T z, \quad z \in \mathbb{R}^n \quad (2.22)$$

lo cual tiene sentido si KK^T es invertible, y válido si K es de rango completo, ya que $Kf_{true} = KK^T z$. Ahora bien

$$\begin{aligned} f_{true} &= K^T z \\ &= \sum_{i=1}^n \sigma_i (u_i^T z) v_i. \end{aligned}$$

Luego,

$$\begin{aligned} \|f_{true}\| &= z^T K K^T z \\ \|f_{true}\|^2 &= \sum_i \sigma_i^2 \|z\|^2, \end{aligned}$$

entonces

$$\begin{aligned}
\|e_\alpha^{trunc}\|^2 &= \sum_{i=1}^n [w_\alpha(\sigma_i^2) - 1]^2 [(v_i^T f_{true})v_i][(v_i^T f_{true})v_i] \\
&= \sum_{i=1}^n [w_\alpha(\sigma_i^2) - 1]^2 \|f_{true}\|^2 \\
&= \sum_{i=1}^n [w_\alpha(\sigma_i^2) - 1]^2 \sigma_i^2 \|z\|^2 \\
&= \sum_{i=1}^n [w_\alpha(\sigma_i^2)\sigma_i - \sigma_i]^2 \|z\|^2.
\end{aligned}$$

En particular, para el filtro de Tijonov (el análisis es similar para TSVD), obtenemos

$$\begin{aligned}
\|e_\alpha^{trunc}\|^2 &= \sum_{i=1}^n [w_\alpha(\sigma_i^2)\sigma_i - \sigma_i]^2 \|z\|^2 \\
&\leq \max_{1 \leq i \leq n} [w_\alpha(\sigma_i^2)\sigma_i - \sigma_i]^2 \|z\|^2 \\
&= \left[\frac{\sigma_i^3}{\sigma_i^2 + \alpha} - \sigma_i \right]^2 \|z\|^2 \\
&= \left[\frac{\alpha\sigma_i}{\sigma_i^2 + \alpha} \right]^2 \|z\|^2 \\
&\leq \alpha \|z\|^2,
\end{aligned}$$

en donde en la última desigualdad se ha utilizado (2.19). Lo cual implica que

$$\|e_\alpha^{trunc}\| \leq \alpha^{1/2} \|z\|. \quad (2.23)$$

Ahora, tomando la norma de la ecuación (2.13), y considerando la desigualdad del triángulo y las desigualdades (2.20) y (2.23), se obtiene

$$\begin{aligned}
\|e_\alpha\| &= \|e_\alpha^{trunc} + e_\alpha^{ruido}\| \\
&\leq \|e_\alpha^{trunc}\| + \|e_\alpha^{ruido}\| \\
&\leq \alpha^{1/2} \|z\| + \alpha^{-1/2} \delta.
\end{aligned} \quad (2.24)$$

Deseamos encontrar el error óptimo total, así que buscamos el mínimo de la función

$$q(\alpha) = \|z\| \alpha^{1/2} + \delta \alpha^{-1/2},$$

por medio de

$$q'(\alpha) = \frac{1}{2} \|z\| \alpha^{-1/2} - \frac{\delta}{2} \alpha^{-3/2} = 0.$$

La solución de esta es

$$\alpha = \frac{\delta}{\|z\|}, \quad (2.25)$$

la cual es un mínimo local estricto debido a que

$$q''\left(\frac{\delta}{\|z\|}\right) = \frac{1}{4} \frac{\|z\|^{5/2}}{\delta^{3/2}} > 0.$$

Por lo tanto, tomando $\alpha = \frac{\delta}{\|z\|}$ en (2.24) se obtiene

$$\|e_\alpha\| \leq 2\|z\|^{1/2} \delta^{1/2}. \quad (2.26)$$

Por lo que el error óptimo total del método de regularización es del orden $O(\sqrt{\delta})$ siempre que $\delta \rightarrow 0$ y, dada la información $f_{true} \in \text{Rango}(K^T)$.

La ecuación (2.25) indica que se elige el parámetro de regularización a “priori”, ya que suponemos suavidad en los datos ($f_{true} = K^T z$), es decir regularidad de la fuente. Se le llama “priori” porque solo conocemos los datos d y se utiliza información desconocida: nivel del ruido (δ) y f_{true} .

Selección del parámetro α a posteriori.

Una regla de elección del parámetro de regularización se denomina a posteriori si ésta depende de los datos pero no de información previa de la solución. Una posible regla es el principio de discrepancia de Morozov [18], que nos dice que los datos regularizados (Kf_α) no deben diferir de los datos dados en una cantidad mayor al nivel del ruido, es decir,

$$\|Kf_\alpha - d\| \leq \delta. \quad (2.27)$$

Los métodos de regularización de TSVD y de Tijonov con la elección del parámetro por medio del principio de discrepancia son convergentes. Si además asumimos que se cumple la condición fuerte (2.22), estos dos métodos son de orden óptimo [6], [9], [12].

Lo que sigue a continuación es una versión finito-dimensional del análisis para el principio de discrepancia aplicada a la regularización de Tijonov. Vamos a simplificar la notación definiendo el funcional de Discrepancia,

$$D(\alpha) = \|Kf_\alpha - d\|$$

Como primer paso, establecemos condiciones bajo las cuales existe un valor del parámetro de regularización para el cual $D(\alpha) = \delta$. Utilizando la representación (2.12) tenemos,

$$\begin{aligned}
D(\alpha)^2 &= \|Kf_\alpha - d\|^2 \\
&= (Kf_\alpha - d)^T(Kf_\alpha - d) \\
&= f_\alpha^T K^T K f_\alpha - 2f_\alpha^T K^T d + d^T d \\
&= d^T K(K^T K + \alpha I)^{-1} K^T K f_\alpha - 2f_\alpha^T K^T d + d^T d \\
&= d^T K(K^T K + \alpha I)^{-1} K^T K(K^T K + \alpha I)^{-1} K^T d - 2f_\alpha^T K^T d + d^T d \\
&= d^T K(K^T K + \alpha I)^{-1} K^T K(K^T K + \alpha I)^{-1} K^T d - 2d^T K(K^T K + \alpha I)^{-1} K^T d + d^T d \\
&= \|(I - K(K^T K + \alpha I)^{-1} K^T d)\|^2.
\end{aligned}$$

Utilizando la descomposición SVD de K , obtenemos

$$D(\alpha)^2 = \sum_{i=1}^n \left(1 - \frac{\sigma_i^2}{\sigma_i^2 + \alpha}\right)^2 (u_i^T d)^2. \quad (2.28)$$

Es claro que, $D(\alpha)$ es continua respecto a α , y estrictamente creciente, además $D(0) = 0$ y

$$\lim_{\alpha \rightarrow \infty} D(\alpha) = \left(\sum_{i=1}^n (u_i^T d)^2\right)^{1/2} = \|d\|.$$

Por lo tanto, si $\delta \leq \|d\|$, existe un único α tal que $D(\alpha) = \delta$, y se denota por $\alpha(\delta)$, es decir $D(\alpha(\delta)) = \delta$.

Teorema 2.2.1. Si $\delta \leq \|d\|$ y $f_{true} = K^T z$, entonces

- i. $\|f_{\alpha(\delta)}\| \leq \|f_{true}\|$ (Estabilidad).
- ii. $\|e_{\alpha(\delta)}\| \leq 2\|z\|^{1/2}\sqrt{\delta}$, es decir, se obtiene orden óptimo.

Demostración.

- i. Como f_α se puede representar en su forma variacional, entonces

$$f_{\alpha(\delta)} = \arg \min_{f \in \mathbb{R}^n} \{\|Kf - d\|^2 + \alpha(\delta)\|f\|^2\}$$

de donde obtenemos que,

$$\|Kf_{\alpha(\delta)} - d\|^2 + \alpha(\delta)\|f_{\alpha(\delta)}\|^2 \leq \|Kf_{true} - d\|^2 + \alpha(\delta)\|f_{true}\|^2,$$

lo cual implica que

$$D(\alpha)^2 + \alpha(\delta)\|f_{\alpha(\delta)}\|^2 \leq \|Kf_{true} - d\|^2 + \alpha(\delta)\|f_{true}\|^2,$$

y entonces

$$\delta^2 + \alpha(\delta)\|f_{\alpha(\delta)}\|^2 \leq \|\eta\|^2 + \alpha(\delta)\|f_{true}\|^2,$$

y como $\|\eta\|^2 = \delta^2$, entonces se concluye que

$$\|f_{\alpha(\delta)}\| \leq \|f_{true}\|.$$

ii. Ahora, vamos a demostrar el orden óptimo,

$$\begin{aligned} \|e_{\alpha(\delta)}\|^2 &= \|f_{\alpha(\delta)} - f_{true}\|^2 \\ &= \|f_{\alpha(\delta)}\|^2 - 2f_{\alpha(\delta)}^T f_{true} + \|f_{true}\|^2 \\ &\leq 2\|f_{true}\|^2 - 2f_{\alpha(\delta)}^T f_{true}, \quad \text{ya que } \|f_{\alpha(\delta)}\| \leq \|f_{true}\| \\ &= 2f_{true}^T f_{true} - 2f_{\alpha(\delta)}^T f_{true} \\ &= 2(f_{true} - f_{\alpha(\delta)})^T f_{true} \\ &= 2(f_{true} - f_{\alpha(\delta)})^T K^T z \\ &= 2(K f_{true} - K f_{\alpha(\delta)})^T z \\ &\leq 2\|K f_{true} - K f_{\alpha(\delta)}\| \|z\|, \quad \text{por la desigualdad de Cauchy-Schwartz.} \\ &= 2\|K f_{true} - d + d - K f_{\alpha(\delta)}\| \|z\| \\ &\leq 4\|K f_{true} - d\| \|z\|, \quad \text{por la desigualdad del triángulo.} \\ &= 4D(\alpha(\delta))\|z\| \\ &= 4\delta\|z\|, \quad \text{ya que } \delta \leq \|d\|. \end{aligned}$$

Por lo tanto

$$\|e_{\alpha(\delta)}\| \leq 2\|z\|^{1/2}\sqrt{\delta}.$$

■

2.2.3. Método de Regularización Variacional

Para sistemas muy grandes que son mal condicionados usualmente es impráctico implementar regularización por filtrado directamente, dado que la factorización SVD puede resultar muy costoso y no muy preciso. Por ello es necesario notar que el método de regularización de Tijonov

$$f_{\alpha} = \sum_{i=1}^n \frac{\sigma_i}{\sigma_i^2 + \alpha} (u_i^T d) v_i = (K^T K + \alpha I)^{-1} K^T d$$

tiene una representación variacional

$$f_{\alpha} = \arg \min_{f \in \mathbb{R}^n} \{\|Kf - d\|^2 + \alpha\|f\|^2\}. \quad (2.29)$$

Ahora, si definimos la función $\varphi(f)$ como

$$\begin{aligned}\varphi(f) &= \|Kf - d\|^2 + \alpha\|f\|^2 \\ &= (Kf - d)^T(Kf - d) + \alpha f^T f \\ &= f^T K^T K f - 2f^T K^T d + d^T d + \alpha f^T f,\end{aligned}$$

entonces f_α resuelve

$$\begin{aligned}\varphi'(f) &= 2K^T K f - 2K^T d + 2\alpha f \\ &= 2(K^T K + \alpha I)f - 2K^T d \\ &= 0,\end{aligned}$$

es decir,

$$(K^T K + \alpha I)f_\alpha = K^T d. \quad (2.30)$$

Esta última ecuación proporciona un método diferente para calcular f_α . Si $\alpha > 0$, entonces la matriz $K^T K + \alpha I$ es simétrica y definida positiva, por lo que se pueden utilizar métodos eficientes, directos o iterativos, para calcular f_α . Estos métodos generalmente son mucho menos costosos que calcular la descomposición SVD de K , especialmente cuando K es una matriz grande.

La representación variacional puede tener otras ventajas. Por ejemplo en óptica la fuente de intensidad f es no-negativa. La no-negatividad puede imponerse como restricción de (2.29). Además el término de mínimos cuadrados $\|Kf - d\|^2$ puede reemplazarse por otros funcionales de ajuste de datos. En (2.29) el término $\|f\|^2$ se denomina funcional de penalización. En términos de proyecciones si $d \in \text{Im}(K)$ la ecuación (2.4) tiene solución. Sin embargo, en (2.7), $Kf_{true} \in \text{Im}(K)$ pero, η no necesariamente pertenece a $\text{Im}(K)$ luego, en general $d \notin \text{Im}(K)$. Por lo tanto, en este caso es necesario encontrar f_α por medio de técnicas de mínimos cuadrados (ver Figura 2.11).

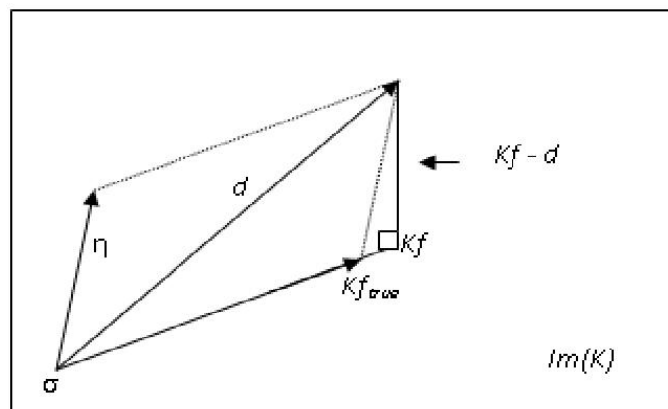


Figura 2.11: η , d no pertenecen $\text{Im}(K)$

2.2.4. Método de Regularización Iterativa

La representación variacional introducida en (2.29) nos permitió observar que f_α se puede calcular mediante la minimización de un funcional, y esto se realiza en forma práctica mediante la solución de las ecuaciones normales (2.30). El enfoque de mínimos cuadrados se puede aprovechar para formular el problema original $Kf = d$ en términos de la minimización del funcional cuadrático

$$\begin{aligned} J(f) &= \frac{1}{2} \|Kf - d\|^2 \\ &= \frac{1}{2} f^T (K^T K) f - f^T K^T d + \frac{1}{2} d^T d, \end{aligned} \quad (2.31)$$

por medio de métodos iterativos de tipo descenso en la dirección del gradiente

$$\nabla J(f) = K K^T f - K^T d, \quad (2.32)$$

en donde el número de iteraciones del método se utiliza como parámetro de regularización. Por ejemplo, en el método de descenso máximo:

Dado f_0 , generar $\{f_k\}_{k=1}^\infty$

$$f_{k+1} = f_k - \tau \nabla J(f_k), \quad k = 1, 2, \dots \quad (2.33)$$

en cada iteración el escalar τ se escoge para minimizar $J(f_k - \tau \nabla J(f_k))$. Si se fija τ con el criterio

$$0 < \tau < \frac{1}{\|K\|},$$

se obtiene el método conocido como iteración de Landweber [14]. Con esta elección para τ el método iterativo es convergente. Para verificarlo, expresamos f_{k+1} en (2.33) en la forma

$$f_{k+1} = G f_k + b$$

donde $G = I - \tau K^T K$ y $b = \tau K^T d$.

Ahora, si $f_0 = \vec{0}$, entonces, $f_1 = b$, $f_2 = Gb + b$, $f_3 = G^2b + Gb + b$, por lo que

$$f_k = (I + G + \dots + G^{k-1})b = (I - G^k)(I - G)^{-1}, \quad \text{suponiendo } \|G\| < 1$$

$$\begin{aligned} \lim_{k \rightarrow \infty} f_k &= (I - G)^{-1} b \\ &= (\tau K^T K)^{-1} \tau K^T d \\ &= K^{-1} d = f_* \\ &= K^{-1} d_{true} + K^{-1} \eta \\ &= f. \end{aligned}$$

Sin embargo, no es deseable hacer que $k \rightarrow \infty$ pues en este caso habrá inestabilidad por la amplificación del ruido $K^{-1}\eta$ debido al mal condicionamiento de K . Por otro lado, la

condición para que la matriz de iteración G tenga norma menor a 1 es que su radio espectral sea menor a 1, es decir que sus valores propios tengan magnitud menor a uno. Como G es simétrica sus valores propios son también sus valores singulares y como

$$G = I - \tau K^T K = I - \tau V \Sigma^2 V^T = V(I - \tau \Sigma^2) V^T,$$

entonces sus valores propios son $1 - \tau \sigma_i^2$, y estos tienen magnitud menor a uno, sí y solo, sí

$$-1 < 1 - \tau \sigma_i^2 < 1 \quad \Leftrightarrow \quad \tau < \frac{2}{\sigma_i},$$

por lo que el método de Landweber produce una sucesión convergente (f_k) si el parámetro τ se escoge menor $2/\|G\|$. Esta condición es más débil que la que usualmente se pide, y que es $\tau < 1/\|G\|$, como ya se mencionó anteriormente.

Terminaremos esta sección encontrando el filtro asociado al método de Landweber y presentando un ejemplo. Para encontrar el filtro debemos encontrar el operador de regularización R_k . Primero, obsérvese que

$$\begin{aligned} f_k &= (I - G^k)(I - G)^{-1}b \\ &= (I - G^k)(\tau K^T K)^{-1} \tau K^T d \\ &= (I - G^k)K^{-1}d \\ &= R_k K^{-1}d, \end{aligned}$$

ahora,

$$\begin{aligned} R_k &= I - (I - \tau K^T K)^k \\ &= I - (I - \tau V \Sigma^2 V^T)^k \\ &= V[I - (I - \tau \Sigma^2)^k]V^T, \end{aligned}$$

por SVD

$$\begin{aligned} R_k &= \sum_{i=1}^n v_i [1 - (1 - \tau \sigma_i^2)^k] v_i^T \\ &= \sum_{i=1}^n [1 - (1 - \tau \sigma_i^2)^k]. \end{aligned}$$

Por lo que el filtro en este caso es

$$w_k(\sigma^2) = 1 - (1 - \tau \sigma^2)^k. \quad (2.34)$$

Nota: Este último muestra que el número de iteraciones k funciona como parámetro de regularización: a menor k mayor regularización y viceversa. Sin embargo, también se debe tener cuidado de no utilizar un k demasiado pequeño, ya que podríamos alejarnos de la solución exacta.

Ejemplo 2.2.3. En la Figuras 2.12, 2.13 y 2.14 se muestra la regularización iterativa de la aproximación de la ecuación 2.5, con parámetro de regularización $k = 10, 1000, 10000$, respectivamente, así como sus correspondientes errores de solución.

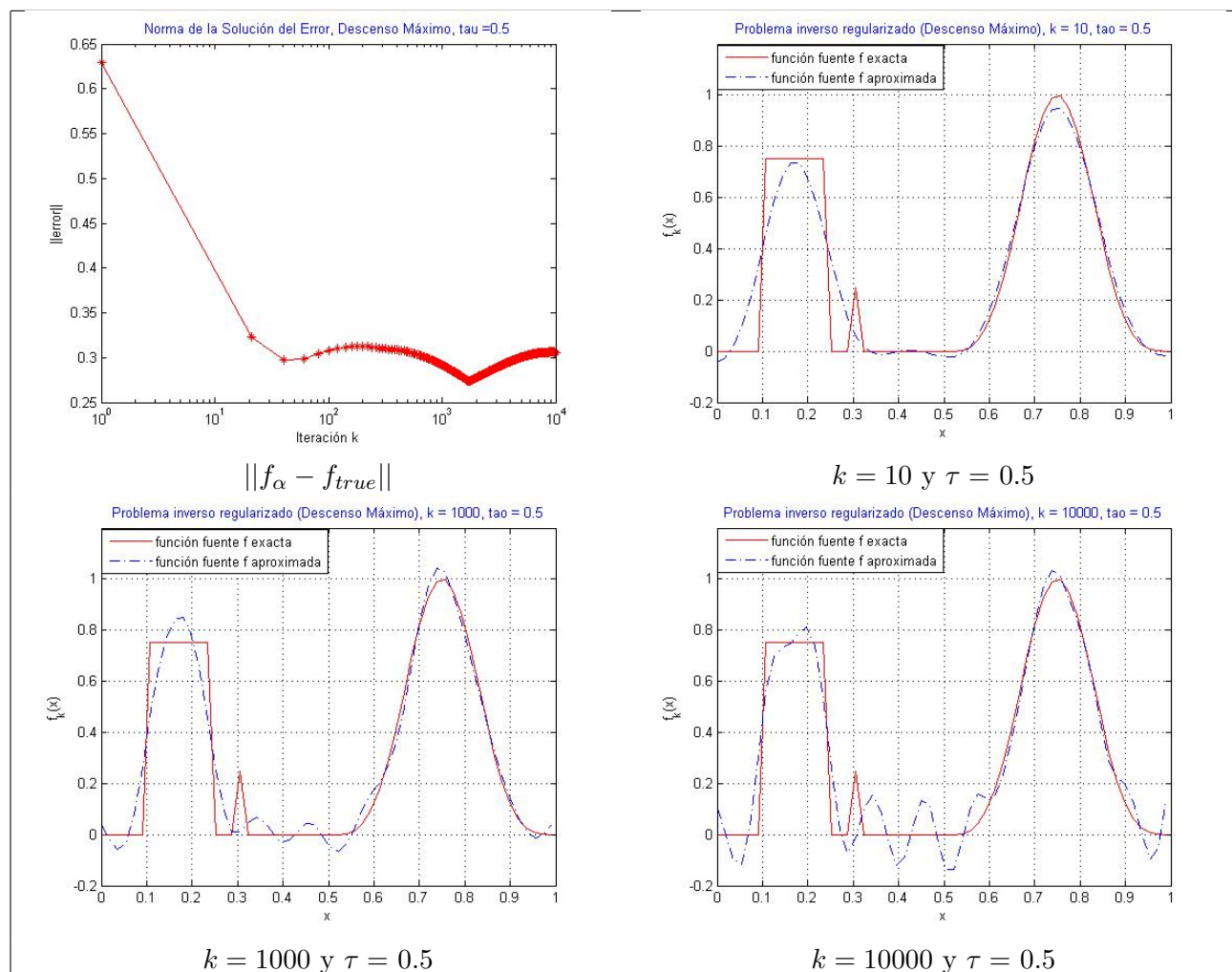


Figura 2.12: Regularización Iterativa, usando Descenso Máximo y $\tau = 0.5$, para diferentes valores de k .

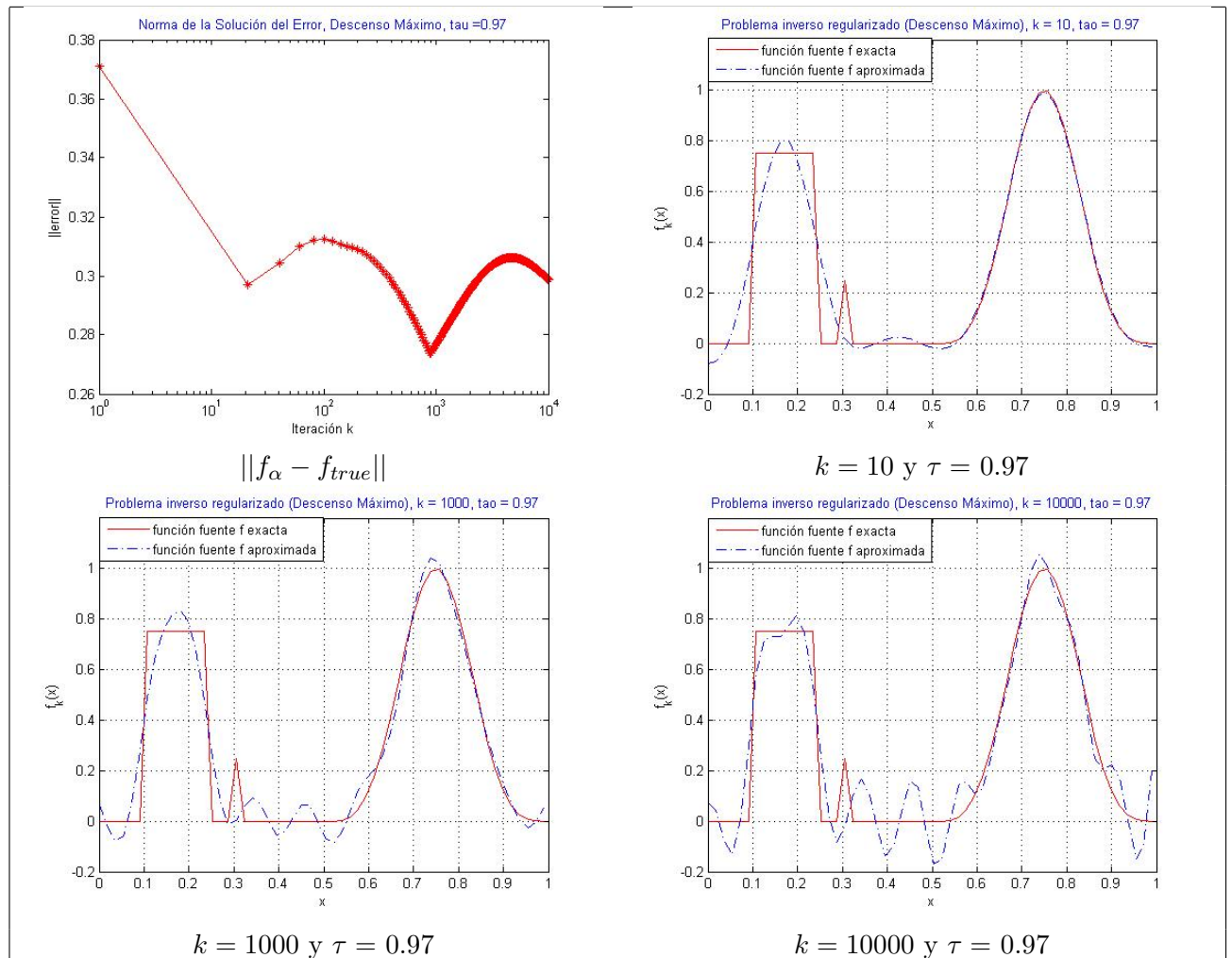


Figura 2.13: Regularización Iterativa, usando Descenso Máximo y $\tau = 0.97$, para diferentes valores de k

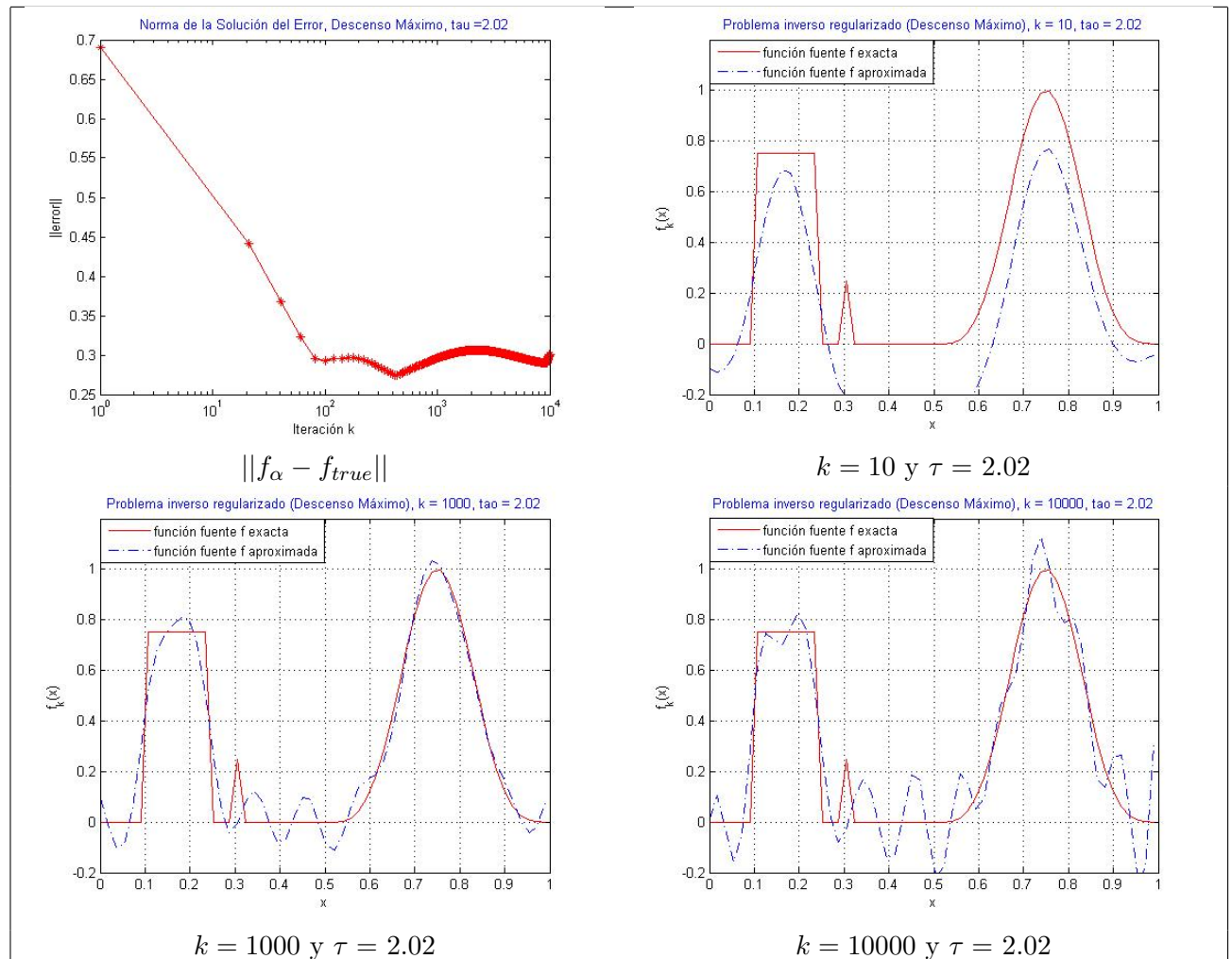


Figura 2.14: Regularización Iterativa, usando Descenso Máximo y $\tau = 2.02$, para diferentes valores de k

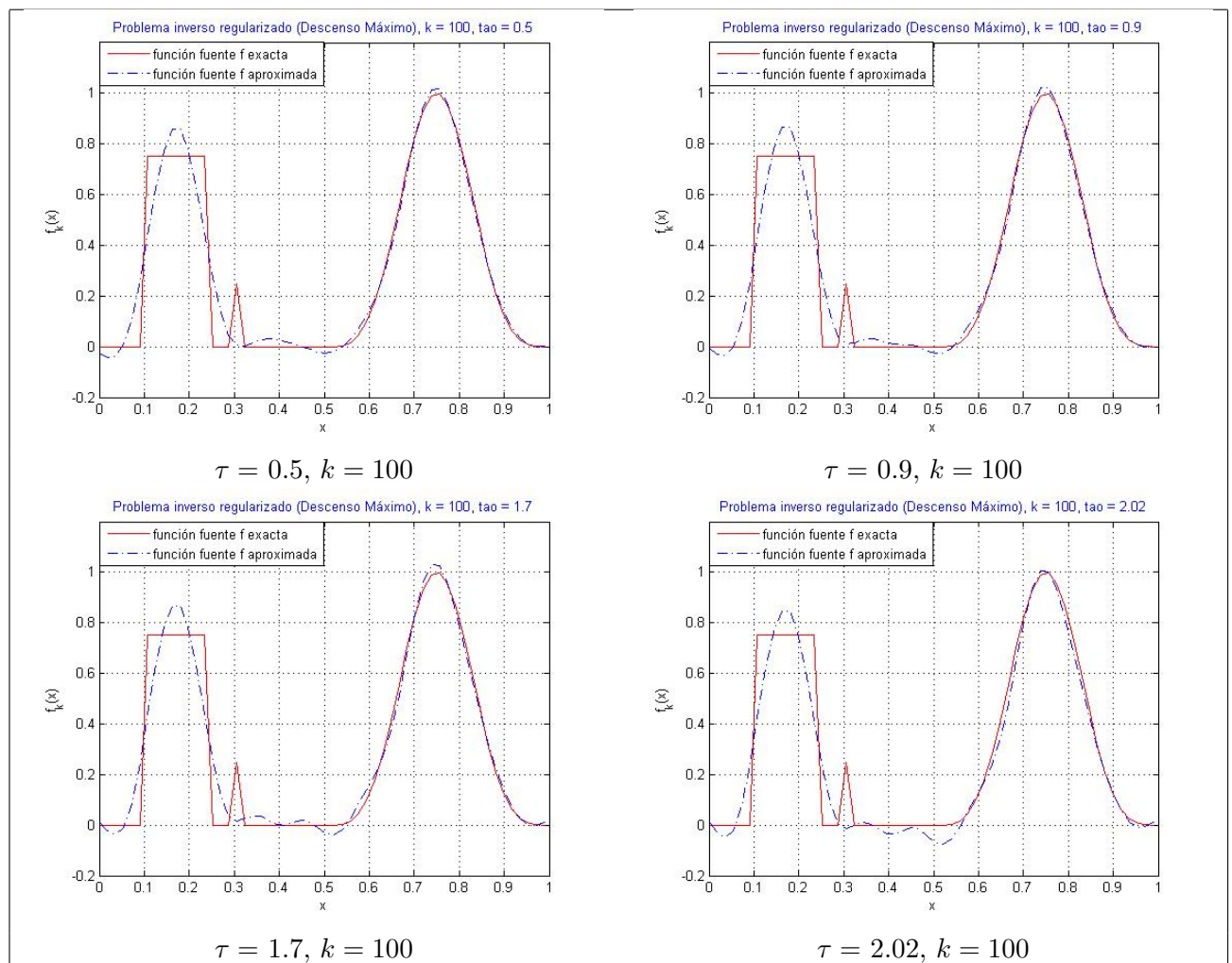


Figura 2.15: Regularización Iterativa, usando Descenso Máximo, $k = 100$ y diferentes valores de τ

En los resultados obtenidos, encontramos que, cuando $k \geq 100$ el error es relativamente el mismo para toda τ tal que $0 < \tau < 2/\|K\|$; sin embargo cuando $\tau = 0.97 \approx 1/\|K\|$ el error es menor respecto a otros valores de τ , siempre y cuando $k < 100$. Véase en la Figura 2.15, que en realidad no se observan diferencias grandes entre las aproximaciones con los distintos valores de τ y $k = 100$. En este caso se recomienda tomar $\tau = 1$ y $k = 100$.

La Inversa Generalizada, Operadores Compactos, Regularización y Métodos de Optimización.

En un contexto general, un problema inverso se puede formular como la necesidad de hallar x en una ecuación de la forma

$$Tx = y,$$

donde $T : X \rightarrow Y$ es un operador lineal acotado, X, Y son espacios de Hilbert y y es el dato. Como ya se ha mencionado anteriormente, generalmente, los problemas inversos son “mal condicionados” en el sentido de Hadamard: hay pérdida de existencia, pérdida de unicidad y/o pérdida de dependencia continua en los datos. En el estudio de problemas inversos en general, los dos primeros no constituyen problemas graves. Casi siempre puede forzarse la existencia relajando el concepto de solución y la no unicidad en ciertos casos hasta puede ser ventajosa. En la pérdida de dependencia continua, *puesta de manifiesto en la no acotación de la inversa generalizada de Moore-Penrose*, la que origina los mayores inconvenientes pues en tal caso, pequeños errores o ruidos en la medición del dato y pueden producir errores muy grandes en las soluciones del problema inverso, haciendo inestables todos los procedimientos numéricos de aproximación tradicionales.

Una solución parcial para esta pérdida de estabilidad es el uso de métodos de regularización. Sin embargo debe tenerse presente que ningún truco matemático puede hacer estable un problema que es intrínsecamente inestable. Todo lo que un método de regularización puede hacer es recuperar la mayor cantidad de información sobre la solución como sea posible sin perder la estabilidad. El “arte” de aplicar métodos de regularización consiste precisamente en hallar el compromiso justo entre exactitud y estabilidad [6].

En el presente capítulo se pretende estudiar y establecer algunos resultados sobre los métodos para resolver problemas inversos mal condicionados, realizando un estudio de la

inversa generalizada, los operadores lineales compactos y su descomposición singular, y algunos métodos de optimización que se utilizan como herramienta en la regularización de los problemas inversos. Estas herramientas nos servirán como soporte para resolver el problema inverso abordado en el Capítulo 4.

3.1. La Inversa Generalizada de Moore-Penrose

Sean X y Y espacios de Hilbert y $T : X \rightarrow Y$ un operador lineal acotado. Se dice que

y es alcanzable

si existe x en X tal que $y = T(x)$, es decir si y pertenece a $\mathcal{R}(T)$. Aquí $\mathcal{R}(T)$ denota la imagen bajo X de T . Si denotamos como $\mathcal{N}(T)$ al núcleo del operador, entonces

- i) Para todo dato admisible y existe una solución x si, y solo si, y en Y es alcanzable.
- ii) Para todo dato admisible, la solución es única si, y solo si, $\mathcal{N}(T) = \{0\}$.
Además, si se cumplen (i) y (ii), entonces T^{-1} existe y, en este caso
- iii) La solución depende continuamente de los datos si, y solo si, T^{-1} es acotada (es decir, continua).

Demostración de (iii):

Si T^{-1} es continua, entonces $x = T^{-1}y$; $x - x_1 = T^{-1}(y - y_1)$.

Así

$$\begin{aligned} \|x - x_1\|_X &= \|T^{-1}(y - y_1)\|_X \\ &\leq \|T^{-1}\| \|y - y_1\|_Y \quad \text{como } T^{-1} \text{ es continua entonces} \\ \|x - x_1\|_X &\leq M \|y - y_1\|_Y, \end{aligned}$$

lo cual muestra la dependencia continua entre los datos.

Nota:

1. El núcleo $\mathcal{N}(T)$ es un espacio vectorial cerrado.
2. En operadores compactos la inversa T^{-1} no es continua en dimensión infinita, como se demostrará más adelante.
3. Si T es un operador lineal en \mathbb{R}^n , el problema es bien planteado si y solo si alguna de (i) ó (ii) se satisface. La otra propiedad es consecuencia de la compacidad de la bola unitaria en \mathbb{R}^n .

Ejemplo 3.1.1. Sea D un operador tal que $D : \ell^2(\mathbb{R}) \rightarrow \ell^2(\mathbb{R})$ y $[Df]_j = \frac{f_j}{j}$, $j = 1, 2, \dots$

$Df = g$ es mal planteada pues $g = (1, \frac{1}{2}, \frac{1}{3}, \dots) \in \ell^2(\mathbb{R})$.

Luego

$$[Df]_j = \frac{1}{j}f_j.$$

Es decir

$$\frac{1}{j}f_j = \frac{1}{j}.$$

Por lo tanto se tiene que $f_j = 1 \forall j$. Pero $f = (1, 1, \dots) \notin \ell^2(\mathbb{R})$. Por tanto la ecuación

$$Df = g$$

es mal planteada (no tiene solución en $\ell^2(\mathbb{R})$).

La estabilidad tampoco se satisface, pues dado $e_n \in \ell^2$, donde $[e_n]_j = \delta_{nj}$ con $n \in \mathbb{N}$, se tiene

$$\|De_n\| = \frac{1}{n},$$

por lo que

$$\|De_n\| \longrightarrow 0 \quad \text{cuando} \quad n \longrightarrow \infty.$$

Pero e_n no converge a cero, por tanto, la estabilidad no se cumple.

A continuación describiremos la teoría básica sobre soluciones de mejor aproximación a ecuaciones operacionales lineales y de la inversa generalizada (Moore-Penrose) de operadores lineales en espacios de Hilbert [7], [20].

3.1.1. Soluciones de Mínimos Cuadrados y la Inversa de Moore-Penrose

Definición 3.1.1. Sea $T : X \rightarrow Y$ un operador lineal acotado

i) $x \in X$ es solución de mínimos cuadrados de $Tx = y$ si

$$\|Tx - y\| \leq \|Tz - y\|, \quad \text{para } z \in X$$

ii) $x \in X$ es la solución mejor aproximada (solución de norma mínima) de $Tx = y$, si x es una solución de mínimos cuadrados de $Tx = y$, y además

$$\|x\| < \|z\|, \quad \forall z \text{ en } X \text{ que es solución de mínimos cuadrados.}$$

Notemos que:

- La solución de mínimos cuadrados no necesariamente existe.
- Si x es una solución de mínimos cuadrados, entonces el conjunto de todas las soluciones de mínimos cuadrados es $x + \mathcal{N}(T)$.
- La solución de mínimos cuadrados de norma mínima esta íntimamente relacionada con la inversa generalizada de T , como veremos un poco más adelante.

A continuación daremos una definición de la inversa generalizada, o inversa de Moore-Penrose, para un operador lineal. La idea es restringir el dominio y rango del operador de tal manera que el operador restringido tenga inversa.

Sea $T : X \rightarrow Y$ un operador lineal acotado con X y Y espacios de Hilbert, y consideremos la restricción lineal de T a $\mathcal{N}(T)^\perp$, es decir

$$\tilde{T} : \mathcal{N}(T)^\perp \rightarrow \mathcal{R}(T).$$

Como $\mathcal{N}(\tilde{T}) = \{0\}$ y $\mathcal{R}(\tilde{T}) = \mathcal{R}(T)$, entonces \tilde{T} es lineal y biyectiva. Por lo tanto,

$$\tilde{T}^{-1} : \mathcal{R}(T) \rightarrow \mathcal{N}(T)^\perp$$

esta bien definida.

Definición 3.1.2. La Inversa Generalizada T^\dagger de T es la única extensión de \tilde{T}^{-1} a $\mathcal{D}(T^\dagger) = \mathcal{R}(T) + \mathcal{R}(T)^\perp$, es decir

$$T^\dagger : \mathcal{D}(T^\dagger) \rightarrow \mathcal{N}(T)^\perp,$$

donde

$$\mathcal{R}(T)^\perp = \mathcal{N}(T^\dagger) = \mathcal{N}(\tilde{T}^{-1}).$$

A continuación se establecen algunas propiedades de la inversa de Moore-Penrose, [6].

Propiedad 3.1.1. Sea $T : \mathbf{X} \rightarrow \mathbf{Y}$ un operador lineal, entonces

(0) $\tilde{T}^{-1} = T^\dagger$ sobre el $\mathcal{R}(T)$.

(1) $y \in \mathcal{D}(T^\dagger)$, $y = y_1 + y_2 \in \mathcal{R}(T) + \mathcal{R}(T)^\perp$

$$T^\dagger y = T^\dagger y_1 + T^\dagger y_2 = T^\dagger y_1 = \tilde{T}^{-1} y_1.$$

(2) Si Q es una proyección ortogonal sobre $\overline{\mathcal{R}(T)}$, entonces

(a) $Qy = y_1 + Qy_2$, para toda $y = y_1 + y_2 \in \mathcal{R}(T) + \mathcal{R}(T)^\perp$

(b) $T^\dagger y = \tilde{T}^{-1} Qy = T^\dagger Qy$,

(c) $\mathcal{R}(T^\dagger) = \mathcal{N}(T)^\perp$,

Demostración de la Propiedad 3.1.1 (2)

(a) Dado $y = y_1 + y_2 \in \mathcal{R}(T) + \mathcal{R}(T)^\perp$

$$\begin{aligned} Qy &= Qy_1 + Qy_2, \\ &= y_1 + Qy_2, \text{ pues } y_1 \in \mathcal{R}(T) \subset \overline{\mathcal{R}(T)} \end{aligned}$$

(b)

$$\begin{aligned} T^\dagger y &= \tilde{T}^{-1}y_1 \text{ por el inciso (1).} \\ &= \tilde{T}^{-1}Qy - \tilde{T}^{-1}Qy_2 \text{ por (a).} \end{aligned}$$

Pero $Qy_2 \in \overline{\mathcal{R}(T)}$ lo cual implica que existe $\{z_m\} \in \mathcal{R}(T)$, tal que $z_m \rightarrow Qy_2$.

Como

$$z_m \perp \mathcal{R}(T)^\perp = \mathcal{N}(T^\dagger) = \mathcal{N}(\tilde{T}^{-1})$$

entonces

$$Qy_2 \in \mathcal{N}(\tilde{T}^{-1}) \text{ y } \tilde{T}^{-1}Qy_2 = 0.$$

Por lo tanto

$$\begin{aligned} T^\dagger y &= \tilde{T}^{-1}Qy \\ &= T^\dagger Qy. \end{aligned}$$

(c) Por el inciso (1)

(\subseteq)

$$T^\dagger y = \tilde{T}^{-1}y_1 \in \mathcal{R}(\tilde{T}^{-1}) = \mathcal{N}(T)^\perp.$$

Por lo tanto

$$\mathcal{R}(T^\dagger) \subset \mathcal{N}(T)^\perp.$$

(\supseteq) Dado $x \in \mathcal{N}(T)^\perp$, por definición $\tilde{T}x = Tx$, así que

$$T^\dagger Tx = T^\dagger \tilde{T}x = \tilde{T}^{-1} \tilde{T}x = x.$$

Por lo tanto

$$\mathcal{N}(T)^\perp \subset \mathcal{R}(T^\dagger).$$

■

Proposición 3.1.1. Sean P y Q proyecciones ortogonales sobre $\mathcal{N}(T)$ y $\overline{\mathcal{R}(T)}$, respectivamente. Entonces se satisface las cuatro ecuaciones de Moore-Penrose [6]:

(a) $T^\dagger T = I - P$, i.e., proyecta ortogonalmente sobre $\mathcal{N}(T)^\perp$.

(b) $TT^\dagger = Q|_{\mathcal{D}(T^\dagger)}$, i.e., proyecta ortogonalmente sobre $\overline{\mathcal{R}(T)}$.

(c) $TT^\dagger T = T$.

$$(d) T^\dagger T T^\dagger = T^\dagger.$$

Demostración

(a)

$$\begin{aligned} T^\dagger T x &= \tilde{T}^{-1} T x \\ &= \tilde{T}^{-1} T (P x + (I - P)x) \\ &= \tilde{T}^{-1} T P x + \tilde{T}^{-1} T (I - P)x \\ &= \tilde{T}^{-1} T (I - P)x \quad \text{debido a que } P x \in \mathcal{N}(T) \\ &= \tilde{T}^{-1} \tilde{T} (I - P)x \quad \text{ya que } (I - P)x \in \mathcal{N}(T)^\perp \\ &= (I - P)x. \end{aligned}$$

(b)

$$\begin{aligned} T T^\dagger y &= T T^\dagger Q y \quad \text{por (b) del inciso (2) de la propiedad 3.1.1} \\ &= \tilde{T} \tilde{T}^{-1} Q y \\ &= Q y \quad \forall y \in \mathcal{D}(T^\dagger) \end{aligned}$$

(c) $T^\dagger T x = (I - P)x$ por (a). Entonces

$$\begin{aligned} T T^\dagger T x &= T (I - P)x \\ &= T x - T P x \\ &= T x, \quad \text{ya que } P x \in \mathcal{N}(T). \end{aligned}$$

(d) $T T^\dagger y = Q y$, $y \in \mathcal{D}(T^\dagger)$ por (b). Entonces

$$\begin{aligned} T^\dagger T T^\dagger y &= T^\dagger Q y \\ &= T^\dagger y, \quad \text{por (b) del inciso (2) de la propiedad 3.1.1.} \end{aligned}$$

■

Nota: Las ecuaciones de Moore-Penrose caracterizan en forma única T^\dagger . Hacemos énfasis en que $T^\dagger T$ se puede considerar como un proyector ortogonal sobre $\mathcal{N}(T)^\perp$ y $T T^\dagger$ como un proyector ortogonal sobre $\overline{\mathcal{R}(T)}$. Finalmente, cualquier operador T^\dagger que satisface las ecuaciones (c) ó (d) de la proposición 3.1.1 se denomina *inversa interior* ó *inversa exterior*, respectivamente.

Proposición 3.1.2. Sea $T : X \rightarrow Y$ un operador lineal,

(a) T^\dagger tiene gráfica cerrada.

(b) T^\dagger es acotada si, y solo si, $\mathcal{R}(T)$ es cerrado.

Demostración

(a)

$$\begin{aligned} gr(T^\dagger) &= \{(y, T^\dagger y) \mid y \in \mathcal{D}(T^\dagger)\} \\ &= \{(y_1 + y_2, T^\dagger y_1) \mid y_1 \in \mathcal{R}(T), y_2 \in \mathcal{R}(T)^\perp\} \\ &= \{(y_1, \tilde{T}^{-1}y_1) \mid y_1 \in \mathcal{R}(T)\} + \{(y_2, 0) \mid y_2 \in \mathcal{R}(T)^\perp\}. \end{aligned}$$

Para mostrar que $gr(T^\dagger)$ es cerrada hay que demostrar la siguiente igualdad

$$\{(y_1, \tilde{T}^{-1}y_1) \mid y_1 \in \mathcal{R}(T)\} + \{(y_2, 0) \mid y_2 \in \mathcal{R}(T)^\perp\} = \{(Tx, x) \mid x \in X\} \cap (Y \times \mathcal{N}(T)^\perp) + \mathcal{R}(T)^\perp \times \{0\}.$$

Es evidente que

$$\{(y_2, 0) \mid y_2 \in \mathcal{R}(T)^\perp\} = \mathcal{R}(T)^\perp \times \{0\},$$

así que basta con demostrar la igualdad

$$\{(y_1, \tilde{T}^{-1}y_1) \mid y_1 \in \mathcal{R}(T)\} = \{(Tx, x) \mid x \in X\} \cap (Y \times \mathcal{N}(T)^\perp).$$

Sea $y_1 \in \mathcal{R}(T)$, entonces

$$x = \tilde{T}^{-1}y_1 \in \mathcal{N}(T)^\perp$$

y además

$$y_1 = \tilde{T}x = Tx.$$

Así que

$$(y_1, \tilde{T}^{-1}y_1) = (Tx, x)$$

con $y_1 \in \mathcal{R}(T)$ y $x \in \mathcal{N}(T)^\perp$, con lo cual se obtiene la primera contención (\subseteq).

Ahora, sea $x \in \mathcal{N}(T)^\perp$, entonces

$$y_1 = Tx = \tilde{T}x \in \mathcal{R}(T).$$

Así que

$$\tilde{T}^{-1}y_1 = x,$$

por lo que

$$(Tx, x) = (y_1, \tilde{T}^{-1}y_1)$$

con $y_1 \in \mathcal{R}(T)$, con lo cual se obtiene la segunda contención de los conjuntos (\supseteq).

(b) (\Rightarrow) Sea T^\dagger acotado, $\mathcal{D}(T^\dagger) \subset Y$ entonces por el teorema de Hahn-Banach [13], T^\dagger tiene una única extensión acotada $\overline{T^\dagger}$ a todo Y . Como T es continua, entonces $T\overline{T^\dagger}$ es acotada y

$$T\overline{T^\dagger} = Q,$$

donde Q es una proyección ortogonal sobre $\overline{\mathcal{R}(T)}$. Ahora bien, si $y \in \overline{\mathcal{R}(T)}$, entonces

$$y = Qy = T\overline{T^\dagger}y,$$

es decir

$$y \in \mathcal{R}(T).$$

Por lo tanto $\overline{\mathcal{R}(T)} \subset \mathcal{R}(T)$.

(\Leftarrow) Si $\mathcal{R}(T)$ es cerrado, entonces $\mathcal{D}(T^\dagger) = \mathcal{R}(T) + \mathcal{R}(T)^\perp = \mathbf{Y}$ es cerrado. Como $gr(T^\dagger)$ es cerrada, entonces por el Teorema de la gráfica cerrada, T^\dagger es acotado. ■

El siguiente resultado establece la conexión entre soluciones de mínimos cuadrados y la inversa de Moore-Penrose.

Teorema 3.1.1. *Sea $y \in \mathcal{D}(T^\dagger)$, entonces $Tx = y$ tiene una única solución de mínimos cuadrados de norma mínima x^\dagger dada por*

$$x^\dagger = T^\dagger y.$$

El conjunto de todas las soluciones de mínimos cuadrados de $Tx = y$ es

$$x^\dagger + \mathcal{N}(T).$$

Demostración:

Supongamos que Q y P son proyecciones ortogonales sobre $\overline{\mathcal{R}(T)}$ y $\mathcal{N}(T)$ respectivamente. Sea $y \in \mathcal{D}(T^\dagger)$. Se tiene que $T\overline{T^\dagger}y = Qy$, por lo que, $Qy \in \mathcal{R}(T)$. Esto garantiza que el conjunto

$$\mathcal{S} = \{z \in X \mid Tz = Qy\}$$

es no vacío. Como la proyección ortogonal Q es también una proyección métrica, entonces $\forall z \in \mathcal{S}$

$$\|y - Tz\| = \|y - Qy\| \leq \|y - Tx\| \quad \forall x \in X,$$

por lo que toda solución en \mathcal{S} es una solución de mínimos cuadrados de $Tx = y$.

Ahora debemos verificar que toda solución de mínimos cuadrados es un elemento de \mathcal{S} : Sea z solución de mínimos cuadrados de $Tx = y$, entonces

$$\|y - Qy\| \leq \|y - Tz\| = \inf_{x \in \mathcal{D}(T)} \|y - Tx\| = \inf_{u \in \mathcal{R}(T)} \|y - u\| = \|y - Qy\|$$

pues Q es proyección ortogonal y $Qy \in \mathcal{R}(T)$. Entonces, necesariamente $Tz = Qy$. Por lo tanto, $z \in \mathcal{S}$.

Lo anterior demuestra que

$$\mathcal{S} = \{x \in X \mid x \text{ es solución de mínimos cuadrados de } Tx = y\}.$$

Sea x^* el elemento de norma mínima en \mathcal{S} , entonces $\mathcal{S} = x^* + \mathcal{N}(T)$ (pues si $x \in \mathcal{S}$, entonces claramente $x - x^* \in \mathcal{N}(T)$), y basta con demostrar que $x^* = T^\dagger y$:

Como elemento de norma mínimo en $x^* + \mathcal{N}(T)$, x^* es ortogonal a $\mathcal{N}(T)$, es decir $x^* \in \mathcal{N}(T)^\perp$. Esto implica que

$$\begin{aligned} x^* &= (I - P)x^* \\ &= T^\dagger T x^* \\ &= T^\dagger Qy \\ &= T^\dagger T T^\dagger y \\ &= T^\dagger y. \end{aligned}$$

por lo que $\mathcal{S} = x^* + \mathcal{N}(T)$. ■

La solución de mínimos cuadrados de norma mínima puede caracterizarse por medio de la solución de las ecuaciones normales Gaussianas, [6]. Antes de mostrar esto, veamos la siguiente definición, [13].

Definición 3.1.3. (Operador Adjunto T^*). Sea $T : X \rightarrow Y$, un operador lineal acotado, donde X y Y son espacios de Hilbert. Entonces, el operador adjunto T^* de T es el operador

$$T^* : X \rightarrow Y$$

tal que, para todo $x \in X$ y $y \in Y$,

$$\langle Tx, y \rangle = \langle x, T^*y \rangle.$$

El siguiente Teorema muestra la relación de la solución de mínimos cuadrados y las ecuaciones normales Gaussianas.

Teorema 3.1.2. Dado $y \in \mathcal{D}(T^\dagger)$, $x \in \mathcal{D}(T)$ es solución de mínimos cuadrados de $Tx = y$ si, y solo si

$$T^*Tx = T^*y,$$

donde T^* es el operador adjunto de T .

Demostración: x es solución de mínimos cuadrados de $Tx = y$ si, y solo si,

$$y - Tx \in \mathcal{R}(T)^\perp = \mathcal{N}(T^*),$$

pues Tx debe ser el elemento mas cercano a y en $\mathcal{R}(T)$. Esto es equivalente a que

$$T^*(y - Tx) = 0,$$

de donde se sigue el resultado. ■

Nota: De este teorema se encuentra que $T^\dagger y$ es la solución de norma mínima de $T^*Tx = T^*y$, es decir $x^\dagger = (T^*T)^\dagger T^*y$. Se deduce que

$$T^\dagger = (T^*T)^\dagger T^*.$$

Nota: $\mathcal{D}(T^\dagger) = \mathcal{R}(T) + \mathcal{R}(T)^\perp$ es el dominio natural de definición de T^\dagger en el sentido que si $y \notin \mathcal{D}(T^\dagger)$ entonces $Tx = y$ no tiene solución de mínimos cuadrados. Por lo que la solución de norma mínima, a pesar de que hace cumplir la unicidad (ver Teorema 3.1.1), no siempre nos conduce a un problema soluble. Si $\mathcal{R}(T)$ no es cerrado, entonces T^\dagger no es acotada y no hay remedio para la pérdida de dependencia continua en general.

Las ecuaciones operacionales con un operador lineal *compacto* son de especial importancia en los problemas inversos, pues generalmente su inversa generalizada no es continua como veremos en la siguiente sección.

3.2. Operadores Lineales Compactos, Sistemas Singulares y SVD

Muchos problemas mal planteados que aparecen en las aplicaciones involucran operadores compactos. De hecho el prototipo de ecuaciones mal planteadas $Kx = y$ provienen de operadores compactos. Los operadores compactos son muy importantes en las aplicaciones, juegan un papel central en la teoría de operadores integrales y en varios problemas de la física matemática. De hecho, el concepto de compacidad de un operador lineal fue sugerido del estudio de las ecuaciones integrales en la teoría de Fredholm. Las propiedades de los operadores compactos son similares a las de los operadores definidos sobre espacios de dimensión finita.

Recordemos la siguiente definición, [27]:

Definición 3.2.1. *Un operador lineal $K : X \rightarrow Y$ es compacto si, y solo si, la imagen de un conjunto acotado es un conjunto relativamente compacto (precompacto).*

Nota: Para operadores lineales acotados y compactos se utilizara el símbolo K en lugar de T .

Recordemos que un conjunto $M \subset Y$ es *precompacto* si su cerradura \overline{M} es compacta. Equivalentemente, M es precompacto sí, y solo sí, toda sucesión acotada $y_n \subset M$, contiene una subsucesión convergente en \overline{M} , [13].

Ejemplo 3.2.1. *Algunos ejemplos de operadores compactos son:*

- Un operador $K : X \rightarrow Y$ con rango $\mathcal{R}(K) < \infty$ y K lineal siempre es compacto. Por lo tanto, los operadores matriciales son compactos.
- El operador diagonal del ejemplo 3.1 es compacto sobre $\ell^2(\mathbb{R})$.
- El operador integral de Fredholm del capítulo 1 es compacto sobre $L_2[\Omega]$, como se demostrará más adelante.

El siguiente teorema muestra la conexión entre compacidad y mal planteamiento para operadores lineales en espacios de Hilbert de dimensión infinita, ver [13].

Teorema 3.2.1. *Sea $K : X \rightarrow Y$ un operador lineal compacto con X y Y de dimensión infinita*

- a) *Si el $\mathcal{R}(K)$ tiene dimensión infinita, entonces $Kx = y$ no tiene solución $\forall y \in Y$, además K^{-1} no es continua. En este caso el $\mathcal{R}(K)$ no es cerrado.*
- b) *Si $\mathcal{R}(K)$ tiene dimensión finita, entonces la solución no es única.*

El teorema nos dice que, en dimension infinita, el problema operacional en $Kx = y$ con K compacto es mal planteado. Es decir, si los espacios X y Y son infinito-dimensionales, entonces cuando $\mathcal{R}(K)$ tiene dimensión finita la solución no es única, y cuando $\mathcal{R}(K)$ tiene dimensión infinita entonces no hay solución y tampoco hay estabilidad respecto de los datos, pues en este caso K^{-1} no puede ser acotada.

Con el objeto de entender más a fondo porqué ocurre este fenómeno con operadores compactos, a continuación presentaremos la descomposición singular o descomposición en valores singulares (SVD, por sus siglas en inglés) de este tipo operadores, y veremos algunas de sus propiedades fundamentales para el estudio de los problemas mal planteados.

3.2.1. Descomposición Singular de Operadores Compactos (SVD)

La teoría espectral de los operadores lineales compactos es una generalización simple de la teoría de valores propios de las matrices y es muy parecida en muchos aspectos a la teoría en espacios de dimensión finita. Por ejemplo, un operador lineal compacto autoadjunto K puede “diagonalizarse” por medio de sus valores y funciones propias λ_n, v_n :

$$Kx = \sum_{n=1}^{\infty} \lambda_n \langle x, v_n \rangle v_n \quad \forall x \in \mathbf{X}.$$

Si K no es autoadjunto no se asegura la existencia de los valores propios y, en consecuencia de las funciones propias. Sin embargo, explotando la conexión entre $Kx = y$ y $K^*Kx = K^*y$ se puede construir un sistema singular, el cual sustituye el sistema (λ_n, v_n) como se muestra a continuación.

Para cualquier operador lineal compacto $K : X \rightarrow Y$, un sistema singular $(\sigma_n; v_n, u_n)$ es definido como, [6]:

- $\{\sigma_n^2\}_{n \in \mathbb{N}}$ son los valores propios de K^*K (y también de KK^*) con multiplicidades y en orden decreciente. Aquí K^* denota al adjunto de K definido por medio de $\langle Kx, y \rangle = \langle x, K^*y \rangle, \forall x \in X, y \in Y$.
- $\{v_n\}_{n \in \mathbb{N}}$ es un sistema completo ortonormal de vectores (funciones) propios de K^*K correspondientes a los valores propios σ_n . Estos vectores generan $\overline{\mathcal{R}(K^*)} = \overline{\mathcal{R}(K^*K)}$.
- $\{u_n\}_{n \in \mathbb{N}}$ son un sistema completo ortonormal de vectores propios de KK^* . Estos vectores generan $\overline{\mathcal{R}(K)} = \overline{\mathcal{R}(KK^*)}$.

Además se cumplen las siguientes formulas:

$$\begin{aligned} Kv_n &= \sigma_n u_n, \\ K^*u_n &= \sigma_n v_n, \\ Kx &= \sum_{n=1}^{\infty} \sigma_n \langle x, v_n \rangle u_n, \quad \forall x \in X, \\ K^*y &= \sum_{n=1}^{\infty} \sigma_n \langle y, u_n \rangle v_n, \quad \forall y \in Y, \end{aligned}$$

donde las series convergen en las normas correspondientes de los espacios de Hilbert X y Y , respectivamente. Estas series se denominan las expansiones en valores singulares, y son los análogos en dimensión infinita de la descomposición en valores singulares (SVD) de una matriz (ver capítulo 2 y apéndice B).

A continuación demostraremos que:

$\mathcal{R}(K)$ es cerrado si, y solo si, es de dimensión finita.

Demostración: Si $\mathcal{R}(K)$ es cerrado, entonces es completo por ser subespacio de un espacio de Hilbert. Luego, por el teorema del mapeo abierto, el mapeo

$$K|_{\mathcal{N}(K)^\perp} : \mathcal{N}(K)^\perp \rightarrow \mathcal{R}(K)$$

tiene inversa continua, así que

$$K(K|_{\mathcal{N}(K)^\perp})^{-1} = I|_{\mathcal{R}(K)}$$

es un mapeo compacto. Por lo tanto, $\dim \mathcal{R}(K) < \infty$.

De lo anterior se concluye que para el caso genérico en el que se tiene un número infinito de valores singulares, $\mathcal{R}(K)$ no es cerrado y, por la proposición 3.1.2, K^\dagger no es acotada. Esto demuestra la siguiente proposición:

Proposición 3.2.1. *Sea $K : X \rightarrow Y$ un operador compacto con $\dim \mathcal{R}(K) = \infty$. Entonces K^\dagger es un operador lineal acotado definido densamente con gráfica cerrada.*

Por lo tanto, para un operador lineal compacto K con rango no cerrado la solución de norma mínima de $Kx = y$ no depende continuamente del lado derecho; es decir la ecuación es mal planteada.

Usando un sistema singular se puede encontrar una representación en serie de la inversa de Moore-Penrose de un operador compacto. Cabe hacer notar que si el sistema singular tiene solo un número finito de elementos, es decir $\mathcal{R}(K)$ es de dimensión finita, entonces las series infinitas de funciones singulares en realidad son sumas finitas.

Teorema 3.2.2. *Sea $(\sigma_n; v_n, u_n)$ un sistema singular para un operador lineal compacto $K : X \rightarrow Y$. Entonces*

$$(a) \ y \in \mathcal{D}(K^\dagger) \text{ si, y solo si, } \sum_{n=1}^{\infty} \frac{|\langle y, u_n \rangle|^2}{\sigma_n^2} < \infty. \quad (\text{CRITERIO DE PICARD})$$

(b) Para $y \in \mathcal{D}(K^\dagger)$,

$$K^\dagger y = \sum_{n=1}^{\infty} \frac{\langle y, u_n \rangle}{\sigma_n} v_n.$$

Demostración:

(a) (\Rightarrow) $y \in \mathcal{D}(K^\dagger)$ implica que $x = K^\dagger y \in \mathcal{R}(K^\dagger) = \mathcal{R}(K^*) = \text{Gen}\{v_n \mid n \in \mathbb{N}\}$. Luego, $Kx = KK^\dagger y = Qy \in \mathcal{R}(K)$. La proyección ortogonal Q sobre $\overline{\mathcal{R}(K)} = \text{Gen}\{u_n \mid n \in \mathbb{N}\}$ se puede escribir como

$$Q = \sum_{n=1}^{\infty} \langle \cdot, u_n \rangle u_n,$$

así que

$$Kx = K \left(\sum_{n=1}^{\infty} \langle x, v_n \rangle v_n \right) = \sum_{n=1}^{\infty} \langle x, v_n \rangle K v_n = \sum_{n=1}^{\infty} \sigma_n \langle x, v_n \rangle u_n,$$

$$Qy = \sum_{n=1}^{\infty} \langle y, u_n \rangle u_n.$$

Como $Kx = Qy$, entonces $\langle y, u_n \rangle = \sigma_n \langle x, v_n \rangle \ \forall n \in \mathbb{N}$. Debido a que, como una sucesión de coeficientes de Fourier, $(\langle x, v_n \rangle) \in \ell^2$, entonces $(\langle y, u_n \rangle / \sigma_n) \in \ell^2$. Por lo tanto

$$\sum_{n=1}^{\infty} \frac{|\langle y, u_n \rangle|^2}{\sigma_n^2} < \infty$$

(\Leftarrow) Ahora supongamos que $\sum_{n=1}^{\infty} \frac{|\langle y, u_n \rangle|^2}{\sigma_n^2} < \infty$, y sea

$$x := \sum_{n=1}^{\infty} \frac{\langle y, u_n \rangle}{\sigma_n},$$

entonces

$$Kx = K \sum_{n=1}^{\infty} \langle y, u_n \rangle K v_n = \sum_{n=1}^{\infty} \sigma_n \langle y, u_n \rangle u_n = Qy.$$

Por lo tanto, $Qy \in \mathcal{R}(K)$, y por 2(b) de la propiedad 3.1.1 se tiene $K^\dagger Qy = K^\dagger y$, es decir $y \in \mathcal{D}(K^\dagger)$.

(b) Al comienzo de la demostración del inciso (a) se obtuvo que $y \in \mathcal{D}(K^\dagger)$ implica que $x = K^\dagger y \in \mathcal{N}(K)^\perp$ y $Kx = Qy$. Entonces $x \in \mathcal{N}(K)^\perp \cap \{z \in X \mid Kz = y\} = K^\dagger y + \mathcal{N}(K)$, es decir

$$K^\dagger y = x = \sum_{n=1}^{\infty} \langle x, v_n \rangle v_n = \sum_{n=1}^{\infty} \frac{\langle y, u_n \rangle}{\sigma_n} v_n.$$

■

Nota: La condición (a) del teorema anterior, denominada criterio de Picard, se puede interpretar de la siguiente manera:

“La solución de mínimos cuadrados de norma mínima de $Kx = y$ existe solo si los coeficientes de Fourier (generalizados) $(\langle y, u_n \rangle)_{n \in \mathbb{N}}$, asociados a las funciones generalizadas $(u_n)_{n \in \mathbb{N}}$, decaen lo suficientemente rápido con respecto a los valores singulares $(\sigma_n)_{n \in \mathbb{N}}$ ”.

Nota: El inciso (b) del teorema anterior nos dice que los errores en y afectan el resultado $K^\dagger y$, de la siguiente manera: los componentes erróneos (con respecto a las base $(u_n)_{n \in \mathbb{N}}$) correspondientes a los valores singulares grandes son inofensivos. Sin embargo, los componentes erróneos correspondiente a los valores singulares pequeños σ_n son amplificados por los factores $1/\sigma_n$, de modo que debemos tener especial cuidado con estos.

Notemos que, si $\dim \mathcal{R}(K) < \infty$, por ejemplo si K es un operador integral con kernel degenerado (ver página 37 de [6]), entonces tiene un número finito de valores singulares, de modo que los factores de amplificación están acotados, aunque aún podrían ser inaceptables.

Ahora, si $\dim \mathcal{R}(K) = \infty$, entonces $\lim_{n \rightarrow \infty} \sigma_n = 0$, por lo cual los errores de un valor fijo se pueden amplificar enormemente sin cota alguna, debido a $1/\sigma_n$. Por ejemplo, si $y_{\delta,n} := y + \delta u_n$, entonces $\|y_{\delta,n} - y\| = \delta$, del inciso (b) del Teorema 3.2.2 se sigue que

$$\begin{aligned} K^\dagger y - K^\dagger y_{\delta,n} &= K^\dagger \delta u_n \\ &= \frac{\langle \delta u_n, u_n \rangle}{\sigma_n} u_n, \end{aligned}$$

por lo tanto

$$\|K^\dagger y - K^\dagger y_{\delta,n}\| = \frac{\delta}{\sigma_n} \rightarrow \infty \quad \text{cuando } n \rightarrow \infty.$$

La inestabilidad debido a (b) del Teorema 3.2.2 (y la condición de solubilidad del criterio de Picard) hace más severo el rápido decaimiento de los valores singulares. Esto permite medir el grado del mal-planteamiento de $Kx = y$ de la siguiente forma: Un problema es llamado *modestamente mal-planteado* si $\sigma_n = \mathcal{O}(n^{-\alpha})$ para algún $\alpha \in \mathbb{R}^+$; de otra manera es llamado *severamente mal planteado*, por ejemplo si $\sigma_n = \mathcal{O}(e^{-n})$. En este último caso, el factor de amplificación de un error de datos en la n -ésima componente de Fourier con respecto a las funciones singulares aumentan tanto como e^n , por lo cual debemos asegurarnos de que los datos contienen información solo en los primeros componentes de Fourier y así filtrar los demás componentes.

3.3. Operadores de Regularización

En términos generales, a la aproximación de un problema mal-planteado por una familia cercana de problemas bien-planteados se le llama *regularización*. Ahora bien, motivaremos la definición de un operador de regularización y de un método de regularización:

Deseamos aproximar la solución de mínimos cuadrados $x^\dagger = T^\dagger y$ de

$$Tx = y \tag{3.1}$$

donde el dato exacto y no se conoce, pero si una aproximación y^δ de este con

$$\|y^\delta - y\| < \delta \tag{3.2}$$

donde y^δ es llamado el “dato con ruido” y δ el “nivel del ruido”.

En el caso de mal-planteamiento, $T^\dagger y^\delta$ no es una buena aproximación de $T^\dagger y$ debido a que T^\dagger no es acotada, aún suponiendo que exista (que en general, no es el caso, ya que $\mathcal{D}(T^\dagger)$ es un subconjunto propio de Y). Estamos buscando una aproximación x_α^δ de x^\dagger tenga las siguientes propiedades:

- (a) Que dependa continuamente de los datos con ruido y_δ , pues de esta manera se podrá calcular de manera estable, y
- (b) que a medida en que el nivel del ruido δ tienda a cero y el *parámetro de regularización* α sea elegido adecuadamente, entonces x_α^δ tienda a x^\dagger .

Nota: La construcción de x_α^δ en general involucra al operador T , sin embargo este no será el caso si consideramos $\|y^\alpha\| < \delta$. Es decir, si el nivel de ruido es mayor que el de los datos, lo mejor sería tomar $x_\alpha^\delta := 0$ independientemente del operador T , ya que en esta situación, de todas formas la parte ruidosa no contiene información. Salvo en este caso, el operador T tiene que jugar un rol en la construcción de x_α^δ .

Consideremos (3.1) como un conjunto de ecuaciones para cada $y \in \mathcal{R}(T)$ ó para cada $y \in \mathcal{D}(T^\dagger)$. De esta manera, no solo hablamos de regularizar una ecuación, sino a este conjunto de ecuaciones, es decir, hablamos de la regularización del operador de solución T^\dagger (sobre el $\mathcal{R}(T)$ si suponemos accesibilidad, o sobre $\mathcal{D}(T^\dagger)$ en otro caso). Intuitivamente, la regularización debe entonces reemplazar el operador no acotado T^\dagger por una familia $\{R_\alpha\}$ de operadores continuos. Entonces, tomamos como la aproximación de x^\dagger a $x_\alpha^\delta = R_\alpha y^\delta$, la cual puede ser calculada de manera estable (en principio, ya que R_α se considera continua). Recordemos que un requerimiento para α es que, si el nivel del ruido δ tiende a cero, entonces la solución regularizada tiende a x^\dagger . Por lo tanto, el parámetro de regularización α tiene que estar relacionado de algún modo con δ y/o y^δ y quizás con otra información acerca de T ó y . Veremos que las reglas de elección de estos parámetros deben estar vinculados ya sea con y^δ o con alguna información a-priori de los datos exactos y , así que dependerá de la ecuación. Más adelante definiremos los operadores de regularización para el operador

$$Tx = y, \quad y \in \mathcal{D}(T^\dagger). \quad (3.3)$$

Nota: Junto con la regla de elección de parámetros que depende de y y/o y^δ , un método de regularización no solo debe ser entendido para el problema (3.1), sino también es posible considerarlo para el problema más general (3.3).

Nota: Un método de regularización también depende del concepto de solución que queremos considerar: en el caso de la solución de (3.1) (todo lo visto hasta ahora y lo que veremos en esta sección) se refiere a la aproximación de la solución de norma mínima (mínimos cuadrados) $x^\dagger = T^\dagger y$.

Lo anterior nos lleva a la siguiente definición.

Definición 3.3.1. Sea $T : X \rightarrow Y$ un operador lineal acotado, donde X y Y son espacios de Hilbert, $\alpha_0 \in (0, +\infty]$. Para cada $\alpha \in (0, \alpha_0)$, sea

$$R_\alpha : Y \rightarrow X$$

un operador continuo (no necesariamente lineal). La familia $\{R_\alpha\}$ es llamada un operador de regularización (para T^\dagger), si para todo $y \in \mathcal{D}(T^\dagger)$ existe una regla de elección del parámetro $\alpha = \alpha(\delta, y^\delta)$, tal que

$$\limsup_{\delta \rightarrow 0} \left\{ \|R_{\alpha(\delta, y^\delta)} y^\delta - T^\dagger y\| \mid y^\delta \in Y, \|y^\delta - y\| \leq \delta \right\} = 0. \quad (3.4)$$

En este caso,

$$\alpha : \mathbb{R}^+ \times Y \rightarrow (0, \alpha_0) \quad (3.5)$$

es tal que

$$\limsup_{\delta \rightarrow 0} \left\{ \alpha(\delta, y^\delta) \mid y^\delta \in Y, \|y^\delta - y\| \leq \delta \right\} = 0. \quad (3.6)$$

Para un $y \in \mathcal{D}(T^\dagger)$ fijo, el par (R_α, α) es llamado un método de regularización convergente si se satisfacen (3.4) y (3.6).

Así, un método de regularización consiste de un operador de regularización y una regla de elección del parámetro que sea convergente en el sentido de que, si el parámetro de regularización se elige de acuerdo a esta regla, entonces las soluciones regularizadas convergen (en sentido de la norma) siempre que el nivel del ruido tienda a cero.

Usualmente, la regla de elección del parámetro solo se debe aplicar a las parejas (δ, y^δ) cuando (3.2) se cumple. Esto es, podemos definir la regla de elección del parámetro α en (3.5) de la siguiente manera

$$\alpha : \left\{ (\delta, y^\delta) \mid \delta > 0, \|y^\delta - y\| \leq \delta \right\} \rightarrow (0, \alpha_0).$$

Nota: Es posible extender la Definición 3.3.1 si además incluimos perturbaciones en el operador (“error de modelado”). En este caso, asumimos que en lugar de T solo tenemos una aproximación T_η tal que

$$\|T - T_\eta\| \leq \eta,$$

por lo cual, la regla de elección del parámetro dependería de δ , η , y^δ y T_η . El requerimiento natural para un método de regularización convergente sería que la condición en (3.4) se cumpla cuando $\delta, \eta \rightarrow 0$ ([26], [28]). Sin embargo, nos limitaremos al caso en el que conocemos al operador con precisión.

Nota: Hemos definido al operador de regularización $\{R_\alpha\}$ como una familia de operadores no necesariamente lineales. Si consideramos R_α lineal, entonces el método es llamado un método de regularización lineal, y la familia $\{R_\alpha\}$ es un operador de regularización lineal. Sin embargo, también tiene sentido considerar los métodos de regularización no lineales para resolver problemas lineales, como el método de Gradiente Conjugado (el cual veremos en la sección 3.4).

La regla de elección del parámetro $\alpha = \alpha(\delta, y^\delta)$ depende (hasta el momento) explícitamente del nivel del ruido δ y del dato perturbado y^δ . Asimismo, recordemos que definimos éste para un $y \in \mathcal{D}(T^\dagger)$ fijo, así que δ también depende del dato exacto y . Como, y es desconocido, entonces esta dependencia podría solo ser de alguna manera conocida a-priori sobre las propiedades suaves de y . Finalmente, aunque no este indicado explícitamente en la notación, α depende también del operador T . En la siguiente definición mencionaremos dos tipos de regla de elección del parámetro α .

Definición 3.3.2. *Sea α elegido de acuerdo a la Definición 3.3.1. Entonces, si α depende únicamente de δ la regla de elección de α es llamada a-priori y escribimos $\alpha = \alpha(\delta)$. En otra caso, se le llama a-posteriori.*

Así, la regla de elección a-priori depende únicamente del nivel del ruido, y no del dato real, y por lo tanto, tampoco de los resultados obtenidos durante el cálculo del residual $\|Tx_\alpha^\delta - y^\delta\|$, donde $x_\alpha^\delta = R_\alpha y^\delta$ es la solución regularizada. Esta regla puede ser concebida antes del cálculo, de ahí el nombre de regla de elección a-priori.

También se podría pensar que la regla de elección del parámetro depende solo de y^δ , y no del nivel del ruido δ . El siguiente resultado debido a Bakushinskii [1] muestra que tales reglas no pueden ser parte de un método de regularización convergente en sentido de la Definición 3.3.1 para un problema mal planteado:

Teorema 3.3.1. *Sea $T : X \rightarrow Y$ un operador lineal acotado, y supongamos que existe una regularización R_α para T^\dagger con una regla de elección del parámetro α que solo depende de y^δ (y no de δ) de tal manera que el método de regularización (R_α, α) es convergente para cada $y \in \mathcal{D}(T^\dagger)$. Entonces T^\dagger es acotado.*

Demostración:

Si α es independiente de δ , es decir, $\alpha = \alpha(y^\delta)$ entonces, de (3.4) se sigue que

$$\limsup_{\delta \rightarrow 0} \left\{ \|R_{\alpha(y^\delta)}y^\delta - T^\dagger y \mid y^\delta \in Y, \|y^\delta - y\| \leq \delta \right\} = 0, \quad (3.7)$$

de modo que $T^\dagger y = R_{\alpha(y)}y, \forall y \in \mathcal{D}(T^\dagger)$. Entonces, por (3.7) tenemos que para alguna sucesión $\{y_n\} \in \mathcal{D}(T^\dagger)$ que converge algún $y \in \mathcal{D}(T^\dagger)$,

$$T^\dagger y_n = R_{\alpha(y_n)}y_n \rightarrow T^\dagger y,$$

por lo tanto T^\dagger es continua en $\mathcal{D}(T^\dagger)$. Pero entonces, la proposición 3.1.2 implica que $\mathcal{D}(T^\dagger) = Y$, de modo que T^\dagger es acotado. ■

La siguiente proposición nos da una idea de como se puede construir un operador de regularización para el operador lineal T .

Proposición 3.3.1. *Supongamos que para todo $\alpha > 0$, R_α es un operador continuo (no necesariamente lineal). Entonces la familia $\{R_\alpha\}$ es una regularización para T^\dagger si*

$$R_\alpha \rightarrow T^\dagger \quad \text{puntualmente en } \mathcal{D}(T^\dagger) \quad \text{cuando } \alpha \rightarrow 0. \quad (3.8)$$

En este caso, para cada $y \in \mathcal{D}(T^\dagger)$, existe una regla de elección a-priori del parámetro α tal que (R_α, α) es un método de regularización convergente para resolver $Tx = y$.

La demostración de la proposición anterior se puede ver en [6].

Observación 3.3.1. *Si $\{R_\alpha\}$ es un método de regularización convergente, entonces se sigue de (3.6) que*

$$\lim_{\delta \rightarrow 0} R_\alpha y = T^\dagger y$$

se cumple para todo $y \in \mathcal{D}(T^\dagger)$. Si α es continua en δ , esto implica que

$$\lim_{\sigma \rightarrow 0} R_\sigma y = T^\dagger y;$$

de otra manera, esto se cumple solo sobre el conjunto de σ -valores que están en el rango de la estrategia de elección del parámetro α .

Por lo tanto, las regularizaciones son aproximaciones puntuales de la Inversa de Moore-Penrose a T . Si $\{R_\alpha\}$ es uniformemente acotada y lineal y, si $\mathcal{R}(T)$ es no cerrado, la

convergencia en (3.8) no puede ser en la norma del operador, por lo que, T^\dagger debería ser acotado. También, debido al Teorema de Banach-Steinhaus

$$\|R_\alpha\| \rightarrow +\infty \quad \text{cuando} \quad \alpha \rightarrow 0 \quad (3.9)$$

si $\mathcal{R}(T)$ es no cerrado. Por el Principio de Acotación Uniforme, (3.9) implica que debe existir $y \in Y$ tal que

$$\|R_\alpha y\| \rightarrow +\infty \quad \text{cuando} \quad \alpha \rightarrow 0. \quad (3.10)$$

De hecho, (3.10) se cumple para todo $Y \setminus M$, donde M es un conjunto de la primera categoría de Baire. Por otro lado,

$$x_\alpha := R_\alpha y \quad (3.11)$$

converge a $T^\dagger y$ sobre el conjunto denso $D(T^\dagger)$ debido a (3.8). Resulta que una condición adicional (razonable), el conjunto en (3.10) es precisamente el complemento de $\mathcal{D}(T^\dagger)$:

Proposición 3.3.2. *Sea $\{R_\alpha\}$ una regularización lineal, x_α definido como (3.11) para todo $y \in Y$. Entonces*

$$\{x_\alpha\} \text{ converge a } T^\dagger y \text{ cuando } \alpha \rightarrow 0 \text{ para } y \in \mathcal{D}(T^\dagger) \quad (3.12)$$

y si

$$\sup \{\|TR\alpha\| \mid \alpha > 0\} < \infty, \quad (3.13)$$

entonces

$$\|x_\alpha\| \rightarrow +\infty \quad \text{cuando} \quad \alpha \rightarrow 0 \quad \text{para} \quad y \notin \mathcal{D}(T^\dagger). \quad (3.14)$$

En este caso, los límites en (3.12) y (3.14) han de entenderse como se explica en la Observación 3.3.1.

Ver demostración en [6].

En la Proposición 3.3.1 hemos visto que, si se cumple (3.8), entonces existe un regla de elección a-priori de α tal que (R_α, α) es un método de regularización convergente. Esta regla de elección puede ser caracterizada como sigue:

Proposición 3.3.3. *Supongamos $\{R_\alpha\}$ una regularización lineal; para cada $y \in \mathcal{D}(T^\dagger)$, sea $\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ elegido a-priori. Entonces (R_α, α) es un método de regularización convergente si, y solo si*

$$\lim_{\delta \rightarrow 0} \alpha(\delta) = 0 \quad (3.15)$$

y

$$\lim_{\delta \rightarrow 0} \delta \|R_{\alpha(\delta)}\| = 0 \quad (3.16)$$

se cumplen.

Véase demostración en [6]

Observación 3.3.2. Si (3.16) es sustituido por

$$\limsup_{\delta \rightarrow 0} \delta \|R_{\alpha(\delta)}\| < +\infty, \quad (3.17)$$

entonces (R_{α}, α) converge débilmente, es decir, para toda sucesión $\{\delta_K\} \rightarrow 0$ y $y_k \in Y$ con

$$\|y_k - y\| \leq \delta_k,$$

$\{R_{\alpha(\delta_k)}y_k\}$ converge débilmente a $T^\dagger y$. Por el contrario (3.17) es necesario para la convergencia débil en el sentido de que, si no se cumple (3.17), entonces existen las sucesión $\{\delta_K\}$ y y_k mencionadas anteriormente tales que $\{R_{\alpha(\delta_k)}y_k\}$ diverge en una topología débil (e incluso no acotada). Ver prueba en [5].

3.4. Herramientas de la Optimización Numérica.

Sea $T : X \rightarrow Y$ un operador compacto, donde X y Y son espacios de Hilbert. Deseamos resolver el problema

$$Tx = y, \quad \text{donde } y \text{ es el dato conocido.}$$

Sin embargo, en la práctica y no es conocido de manera exacta, pero si una aproximación y^δ de éste con

$$\|y - y^\delta\| < \delta, \quad (3.18)$$

donde δ es el nivel del ruido, entonces el problema que debemos resolver es:

$$Tx = y^\delta, \quad (3.19)$$

el cual no tiene solución única. Sin embargo, la unicidad del problema se puede recobrar mediante la inversa generalidad de Moore-Penrose

$$x^\dagger = T^\dagger y$$

la cual existe si, y solo si, $Qy \in \mathcal{R}(T)$, donde Q es una proyección ortogonal sobre $\overline{\mathcal{R}(T)}$, es decir, existe si el $\mathcal{R}(T)$ es cerrado. Además, de (b) del inciso (2) de la propiedad 3.1.1 se sigue que

$$Tx^\dagger = Qy, \quad Qy \in \mathcal{R}(T),$$

por lo cual, x^\dagger define la solución de norma mínima

$$\min_{x \in X} \|Tx - y\|^2,$$

el cual es único (por el Teorema 3.1.1).

Por otro lado, la proposición 3.2.1 nos dice que T^\dagger no es acotada debido a que T es un operador compacto, esto implica que

$$\|T^\dagger y - T^\dagger y^\delta\| = \|x^\dagger - x^\delta\| \rightarrow \infty,$$

es decir, no se satisface la siguiente igualdad

$$\|x^\dagger - x^\delta\| \leq \|y - y^\delta\|.$$

Entonces el resolver el problema (3.19) sin regularización, implica resolver la siguiente ecuación (ver Teorema 3.1.2)

$$(T^*T)^{-1}T^*y^\delta = x^\delta$$

la cual no es estable (por lo visto anteriormente). Es necesario aplicar un método de regularización R_α , por ejemplo, el método de regularización de Tijonov

$$\min_{x \in X} \left\{ \|Tx - y^\delta\|^2 + \alpha \|x\|^2 \right\}$$

donde $J(x) = \|Tx - y^\delta\|^2 + \alpha \|x\|^2$ es un funcional cuadrático y $\alpha = \alpha(\delta, y^\delta)$ es el parámetro de regularización. Así el mínimo (único) satisface

$$(T^*T + \alpha I)^{-1}T^*y^\delta = x_\alpha^\delta,$$

entonces

$$\|x^\dagger - x_\alpha^\delta\| \rightarrow 0 \quad \text{cuando} \quad \delta \rightarrow 0,$$

por lo tanto, el objetivo es resolver

$$(T^*T + \alpha I)x_\alpha^\delta = T^*y^\delta.$$

En la práctica, en muchos problemas en lugar de resolver el problema en dimensión infinita, este se aproxima en espacios de dimensión finita vía algún método de discretización. Por lo que, basta con tener algún método efectivo para resolver ecuaciones lineales de la forma

$$Ax = b,$$

con A simétrica definida positiva.

A continuación veremos algunos métodos de optimización para la regularización iterativa de los problemas inversos y vamos a considerar el funcional $J : \mathbb{R}^n \rightarrow \mathbb{R}$. SDP denota simétrica definida positiva en el contexto de matrices, y asumimos que A es una matriz de $n \times n$ y sus componentes son números reales, a menos de que se especifique lo contrario.

Definición 3.4.1. *Supongamos que una sucesión $\{f_k\}$ converge a f_* cuando $n \rightarrow \infty$. La tasa de convergencia es lineal si existe una constante c , $0 \leq c < 1$, tal que*

$$\|f_{k+1} - f_*\| \leq c \|f_k - f_*\|. \quad (3.20)$$

La convergencia es superlineal si existe una sucesión $\{c_k\}$ de números reales positivos tal que $\lim_{n \rightarrow \infty} c_k = 0$, y

$$\|f_{n+1} - f_*\| \leq c_k \|f_k - f_*\|. \quad (3.21)$$

La tasa de convergencia es cuadrática si para alguna constante $C > 0$

$$\|f_{k+1} - f_*\| \leq C\|f_k - f_*\|^2. \quad (3.22)$$

Convergencia cuadrática implica convergencia superlineal, que a la vez implica convergencia lineal.

3.4.1. Método de Máximo Descenso

El método de Máximo Descenso es un algoritmo de gradiente, en donde se elige α_k de tal forma que la cantidad de avance $-\alpha_k \nabla J(f_k)$ sea máxima (en la dirección $-\nabla J(f_k)$).

De hecho es necesario encontrar el mínimo de la función

$$\phi(\alpha) = J(f_k - \alpha \nabla J(f_k)),$$

donde $\alpha = \arg \min_{\alpha > 0} \phi(\alpha)$.

Algoritmo 3.4.1. *Máximo Descenso.*

Para minimizar el funcional suave $J(f)$

f_0 := punto inicial

Para $k = 0, 1, \dots$ hasta converger

$p_k := -\nabla J(f_k)$ % calcular el gradiente negativo

$\alpha_k := \arg \min_{\alpha > 0} J(f_k + \alpha p_k)$ % búsqueda lineal

$f_{k+1} := f_k + \alpha_k p_k$ % actualización de la solución aproximada

termina

Definición 3.4.2. *El problema de minimización en una dimensión es*

$$\min_{\alpha > 0} J(f_k + \alpha p_k)$$

el cual es llamado búsqueda lineal. El vector p_k es llamado la dirección de búsqueda.

p_k es una dirección de descenso para J en f_k si existe $\delta > 0$ tal que

$$J(f_k + \alpha p_k) < J(f_k) \quad \text{siempre que } 0 < \alpha < \delta.$$

Proposición 3.4.1. *Sea $J : \mathbb{R}^n \rightarrow \mathbb{R}$ una función suave, si $\{f_k\}_{k=1}^{\infty}$ es una sucesión de máximo descenso para J y $\nabla J(f_k) \neq 0$ entonces*

$$J(f_{k+1}) < J(f_k).$$

Demostración:

$$f_{k+1} = f_k + \alpha_k p_k = f_k - \alpha_k \nabla J(f_k).$$

Si $\phi(\alpha) = J(f_k - \alpha \nabla J(f_k))$ y $\alpha_k > 0$ es el mínimo de $\phi(\alpha)$ entonces

$$\phi(\alpha_k) \leq \phi(\alpha) \quad \forall \alpha > 0.$$

Por la regla de la cadena

$$\begin{aligned} \phi'(0) &= \frac{\partial \phi(0)}{\partial \alpha} \\ &= -\nabla J(f_k)^T \nabla J(f_k) \\ &= -\|\nabla J(f_k)\|^2 \\ &< 0. \end{aligned}$$

Debido a que $\|\nabla J(f_k)\| \neq 0$, entonces implica que existe un $\bar{\alpha} > 0$ tal que

$$\begin{aligned} \phi(0) &> \phi(\alpha) \quad \forall \alpha \in (0, \bar{\alpha}] \\ J(f_k) &> \phi(\alpha) \geq \phi(\alpha_k) = J(f_{k+1}) \\ \therefore J(f_k) &> J(f_{k+1}). \end{aligned}$$

■

Para el análisis de convergencia del método de máximo descenso es necesario mencionar algunas definiciones preliminares.

Definición 3.4.3. Sea A una matriz SDP. El producto interior inducido por A está dado por

$$\langle f_1, f_2 \rangle_A := \langle Af_1, f_2 \rangle, \quad f_1, f_2 \in \mathbb{R}^n.$$

La norma inducida, $\|f\|_A = \sqrt{\langle f, f \rangle_A}$ es llamada la norma inducida por A .

La norma inducida es equivalente a la norma Euclidiana, lo cual se puede observar con la Desigualdad de Rayleig:

$$\lambda_{\min}(A)\|f\|^2 \leq \|f\|_A^2 \leq \lambda_{\max}(A)\|f\|^2 \quad \text{siempre que } f \in \mathbb{R}^n. \quad (3.23)$$

Definición 3.4.4. Si A es una matriz de $n \times n$ (no necesariamente simétrica) su número de condición se define como

$$\text{cond}(A) = \frac{\max \sigma_0(A)}{\min \sigma_0(A)}, \quad (3.24)$$

donde $\sigma_0(A)$ denota el conjunto de valores singulares no cero de A . Si A es SDP, entonces $\text{cond}(A)$ se reduce a la razón del mayor y el menor eigenvalor de A . La matriz A es llamada mal-condicionada si $\text{cond}(A)$ es grande.

Definición 3.4.5. Una función $J : \mathbb{R}^n \rightarrow \mathbb{R}$ es cuadrática si existe $c \in \mathbb{R}$, $b \in \mathbb{R}^n$ y una matriz simétrica tal que

$$J(f) = c + \langle b, f \rangle + \frac{1}{2} \langle Af, f \rangle. \quad (3.25)$$

Note que $A = \text{Hess}J(f)$, $J(f)$ está dada en forma matricial, y además $f_* = -A^{-1}b$.

A se dice que es definida positiva si $\langle Af, f \rangle > 0$; si $f \neq 0 \forall f \in \mathbb{R}^n$ es semidefinida positiva si $\langle Af, f \rangle \geq 0$.

Sea $J : \mathbb{R}^n \rightarrow \mathbb{R}$, $J(f) = c + \langle b, f \rangle + \frac{1}{2} \langle Af, f \rangle$ donde A es SDP.

Notación: $g_k = \nabla J(f_k) = Af_k + b$. Entonces dado f_k , se obtiene

$$f_{k+1} = f_k - \alpha_k g_k$$

donde α_k minimiza la función

$$\phi(\alpha) = J(f_k - \alpha_k g_k) = J(f(\alpha)),$$

donde $f(\alpha) = f_k - \alpha g_k$. Luego el punto mínimo de α_k satisface las condiciones necesarias de primer orden

$$\begin{aligned} 0 &= \frac{\partial \phi(\alpha_k)}{\partial \alpha} \\ &= \nabla J(f(\alpha_k))^T (-g_k) \\ &= -g_k^T \nabla J(f(\alpha_k)) \\ &= -g_k^T [A(f_k - \alpha_k g_k) + b] \\ &= -g_k^T [(Af_k + b) - \alpha_k Ag_k] \\ &= -g_k^T [g_k - \alpha_k Ag_k], \end{aligned}$$

por lo que la solución exacta de búsqueda lineal en el algoritmo de descenso máximo para funciones cuadráticas es

$$\alpha_k = \frac{\|g_k\|^2}{\langle Ag_k, g_k \rangle}. \quad (3.26)$$

Teorema 3.4.1. Sea J una función cuadrática positiva con representación (3.25). Entonces para cualquier punto inicial f_0 , el método de descenso máximo $f_{k+1} = f_k - \alpha_k g_k$ converge a f_* linealmente, con

$$\|f_k - f_*\|_A \leq \left(\frac{\text{cond}(A) - 1}{\text{cond}(A) + 1} \right)^k \|f_0 - f_*\|_A. \quad (3.27)$$

Un resultado similar se mantiene en el caso más general (no cuadrático) donde la matriz A es reemplazada por el Hessiano de J en f_* .

Nota: Sea $\Lambda = \left(\frac{\text{cond}(A)-1}{\text{cond}(A)+1} \right)$ y supongamos que $\Lambda \geq 1$, entonces

- Si Λ es grande la rapidez de convergencia es lenta.
- Si Λ es pequeña la rapidez de convergencia es rápida.

Sin embargo, un método de convergencia más rápido es el método de Gradiente Conjugado.

3.4.2. Método de Gradiente Conjugado

Los métodos de dirección conjugado se desarrollan para los problemas cuadráticos con representación (3.25), además se pueden aproximar problemas generales aproximando la función objetivo a una función cuadrática.

Nota: Para los problemas cuadráticos la mejor dirección de búsqueda, es la dirección A conjugada.

Definición 3.4.6. Sea A una matriz simétrica de tamaño $n \times n$. Las direcciones d^0, d^1, \dots, d^k son A -conjugadas si $\forall i \neq j$ se satisface

$$\langle d^i, d^j \rangle_A = 0.$$

Por otra parte, la convergencia lenta del método de descenso máximo para sistemas mal-condicionados limita su utilidad. En éste caso, el método de gradiente conjugado (GC), junto con el esquema de aceleración conocido como preconditionamiento, proporciona medios muy eficientes para solucionar los sistemas lineales SDP que son grandes y mal-condicionados.

Algoritmo 3.4.2. *Gradiente Conjugado para funciones cuadráticas.*

Para minimizar a $J(f) = \frac{1}{2} \langle Af, f \rangle + \langle b, f \rangle + c$, donde A es SDP, o, para resolver $Af = -b$,

```

f0 := punto inicial
g0 := Af0 + b % gradiente inicial
d0 := -g0 % dirección de búsqueda inicial
Para k = 1, 2, ... hasta converger
    rk = Adk
    αk =  $\frac{\langle g_k, d_k \rangle}{\langle d_k, r_k \rangle}$  % parámetro de búsqueda lineal
    fk+1 = fk - αkdk % actualización de la solución aproximada
    gk+1 = gk - αkrk % actualización del gradiente
    βk =  $\frac{\langle g_{k+1}, r_k \rangle}{\langle d_k, r_k \rangle}$ 
    dk+1 = -gk+1 + βkdk % actualización de la dirección de búsqueda
    k = k + 1

```

termina

Definición 3.4.7. El k -ésimo subespacio de Krylov generado por una matriz A SDP y un vector $v \in \mathbb{R}^n$ está dado por

$$\mathcal{S}_k(A, v) := \text{gen} \left(v, Av, \dots, A^{k-1}v \right) \quad (3.28)$$

$$= \left\{ p(A)v \mid p \in \Pi^{k-1} \right\}, \quad (3.29)$$

donde Π^{k-1} denota el conjunto de polinomios de grado menor o igual a $k - 1$.

Teorema 3.4.2. Para $k = 1, 2, \dots$, las iteraciones f_k del GC satisfacen

$$f_k = \arg \min_{f \in f_0 + S_k(A, g_0)} \|f - f_*\|_A. \quad (3.30)$$

El error de solución correspondiente a cada iteración $e_k = f_k - f_*$ satisface

$$\|e_k\|_A = \min_{q \in \Pi_1^k} \|q(A)e_0\|_A, \quad (3.31)$$

donde Π_1^k denota el conjunto de polinomios $q(t)$ de grado menor o igual a k con $q(0) = 1$.

Teorema 3.4.3. Si A es SDP, entonces para cualquier punto inicial f_0 , el algoritmo básico de dirección conjugada converge al único f_* (que es solución de $Af = -b$) en n pasos, es decir, $f_n = f_*$.

Demostración:

Considere $f_* - f_0 \in \mathbb{R}^n$. Debido a que las direcciones conjugadas d_i son linealmente independientes, existen constantes α_i con $i = 0, 1, \dots, n-1$ tales que

$$f_* - f_0 = \alpha_0 d_0 + \alpha_1 d_1 + \dots + \alpha_{n-1} d_{n-1}$$

lo cual implica

$$\begin{aligned} d_k^T A(f_* - f_0) &= \alpha_0 d_k^T A d_0 + \alpha_1 d_k^T A d_1 + \dots + \alpha_{n-1} d_k^T A d_{n-1} \\ &= \alpha_k d_k^T A d_k \quad \forall k = 0, 1, \dots, n-1 \end{aligned}$$

entonces

$$\alpha_k = \frac{d_k^T A(f_* - f_0)}{d_k^T A d_k}.$$

Por otro lado, siguiendo el proceso iterativo desde f_0 hasta f_k :

$$f_k = f_0 + \alpha_0 d_0 + \alpha_1 d_1 + \dots + \alpha_{k-1} d_{k-1}$$

se sigue que

$$\begin{aligned} d_k^T A(f_k - f_0) &= \alpha_0 d_k^T A d_0 + \alpha_1 d_k^T A d_1 + \dots + \alpha_{k-1} d_k^T A d_{k-1} \\ d_k^T A(f_k - f_0) &= 0. \end{aligned}$$

Si escribimos $f_* - f_0$ como $f_* - f_0 = (f_* - f_k) + (f_k - f_0)$, entonces

$$\begin{aligned} d_k^T A(f_* - f_0) &= d_k^T A(f_* - f_k) + d_k^T A(f_k - f_0) \\ &= d_k^T A(f_* - f_k) \\ &= -d_k^T b - d_k^T A f_k \\ &= -d_k^T (A f_k + b) \\ &= -d_k^T g_k, \end{aligned}$$

de donde obtenemos

$$d_k^T A(\alpha_0 d_0 + \alpha_1 d_1 + \cdots + \alpha_{n-1} d_{n-1}) = -d_k^T g_k$$

por lo cual

$$\alpha_k d_k^T A d_k - d_k^T g_k = 0,$$

y entonces

$$\alpha_k = -\frac{d_k^T g_k}{d_k^T A d_k},$$

$\forall k = 0, 1, \dots, n-1$ por lo tanto

$$f_* = f_n = f_0 + \alpha_0 d_0 + \alpha_1 d_1 + \cdots + \alpha_{n-1} d_{n-1},$$

entonces, el método de GC converge en n pasos. ■

Usando una escala del polinomio de Chebyshev $p \in \Pi_i^k$ en (3.31), podemos calcular la cota de error en cada iteración

$$\|e_k\|_A \leq 2 \left(\frac{\sqrt{\text{cond}(A)} - 1}{\sqrt{\text{cond}(A)} + 1} \right)^k \|e_0\|_A. \quad (3.32)$$

Comparando esta cota de error con la del descenso máximo (3.27), podemos ver que es considerablemente más pequeño cuando $\text{cond}(A)$ es grande. Podemos obtener convergencia fuerte si los eigenvalores de A están agrupados fuera de cero.

Ejemplo 3.4.1. En la figura 3.1 se muestra la regularización iterativa, por medio de gradiente conjugado y parámetros de regularización k y $\alpha = 0.000001$, así como una gráfica de errores de solución.

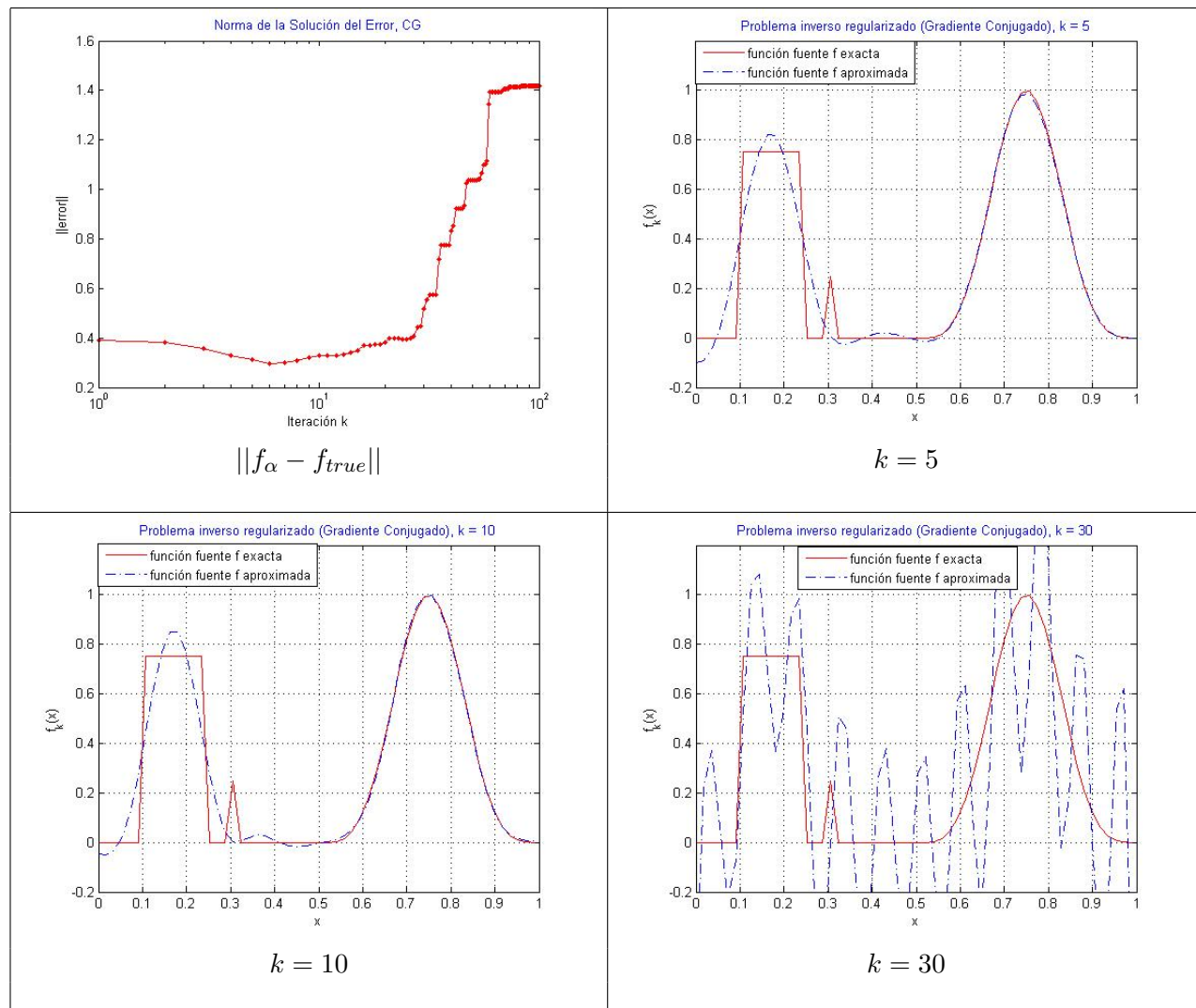


Figura 3.1: Regularización Iterativa, usando Gradiente Conjugado

Precondicionamiento

La cota de error (3.32) y los eigenvalores agrupados resultan motivadores para el concepto de precondicionamiento. Consideremos el funcional cuadrático (3.25), con la matriz A SDP. Deseamos una matriz M SDP, llamada precondicionador, para la cual $M^{-1}A$ tiene una mejor distribución de eigenvalores que la de A .

Algoritmo 3.4.3. *Gradiente Conjugado con Precondicionamiento (PGC).*

Para minimizar a $J(f) = \frac{1}{2}\langle Af, f \rangle + \langle b, f \rangle + c$, donde A es SDP, o, para resolver $Af = -b$, dada una matriz M precondicionadora SDP,

```

 $f_0 :=$  punto inicial
 $g_0 := Af_0 + b$  % gradiente inicial
 $r_0 := M^{-1}g_0$  % aplicación del precondicionador
 $d_0 := r_0$  % dirección de búsqueda inicial
 $dot_0 := \langle g_0, r_0 \rangle$ 
Para  $k = 1, 2, \dots$  hasta converger

 $w_k = Ad_k$ 
 $\alpha_k = \frac{dot_k}{\langle d_k, w_k \rangle}$  % parámetro de búsqueda lineal
 $f_{k+1} = f_k - \alpha_k d_k$  % actualización de la solución aproximada
 $g_{k+1} = g_k - \alpha_k w_k$  % actualización del gradiente
 $r_{k+1} = M^{-1}g_{k+1}$  % aplicación del precondicionador
 $dot_{k+1} = \langle g_{k+1}, r_{k+1} \rangle$ 
 $\beta_k = \frac{dot_{k+1}}{dot_k}$ 
 $d_{k+1} = r_{k+1} + \beta_k d_k$  % actualización de la dirección de búsqueda
 $k = k + 1$ 

```

termina

Nota: Una comparación del algoritmo de GC y el de PGC revela que ambos requieren de una multiplicación por la matriz Hessiana A y calcular dos productos interiores en cada iteración. Además, el algoritmo de PGC requiere precondicionamiento en cada paso, es decir, la solución del sistema lineal $Mr_k = g_k$ para obtener r_k , mientras que GC no. Por lo cual, PGC reduce en forma significativa el costo computacional si:

- (i) Es barato resolver el sistema lineal $Mr = g$, y
- (ii) La tasa de convergencia de PGC es significativamente más rápida que la tasa de GC.

Ejemplo 3.4.2. *En la figura 3.2 se muestra la regularización iterativa, por medio de gradiente conjugado con precondicionamiento y parámetros de regularización k y $\alpha = 0.000001$, así como una gráfica de errores de solución.*

En este ejemplo utilizamos el precondicionador de Jacobi $M = D$ donde D es la matriz diagonal $D = \text{diag}(K)$, es decir $d_{ii} = K_{ii}$. Note que el calculo de $M^{-1}r_{k+1}$ cuenta $\mathcal{O}(n)$ operaciones, lo cual es muy barato.

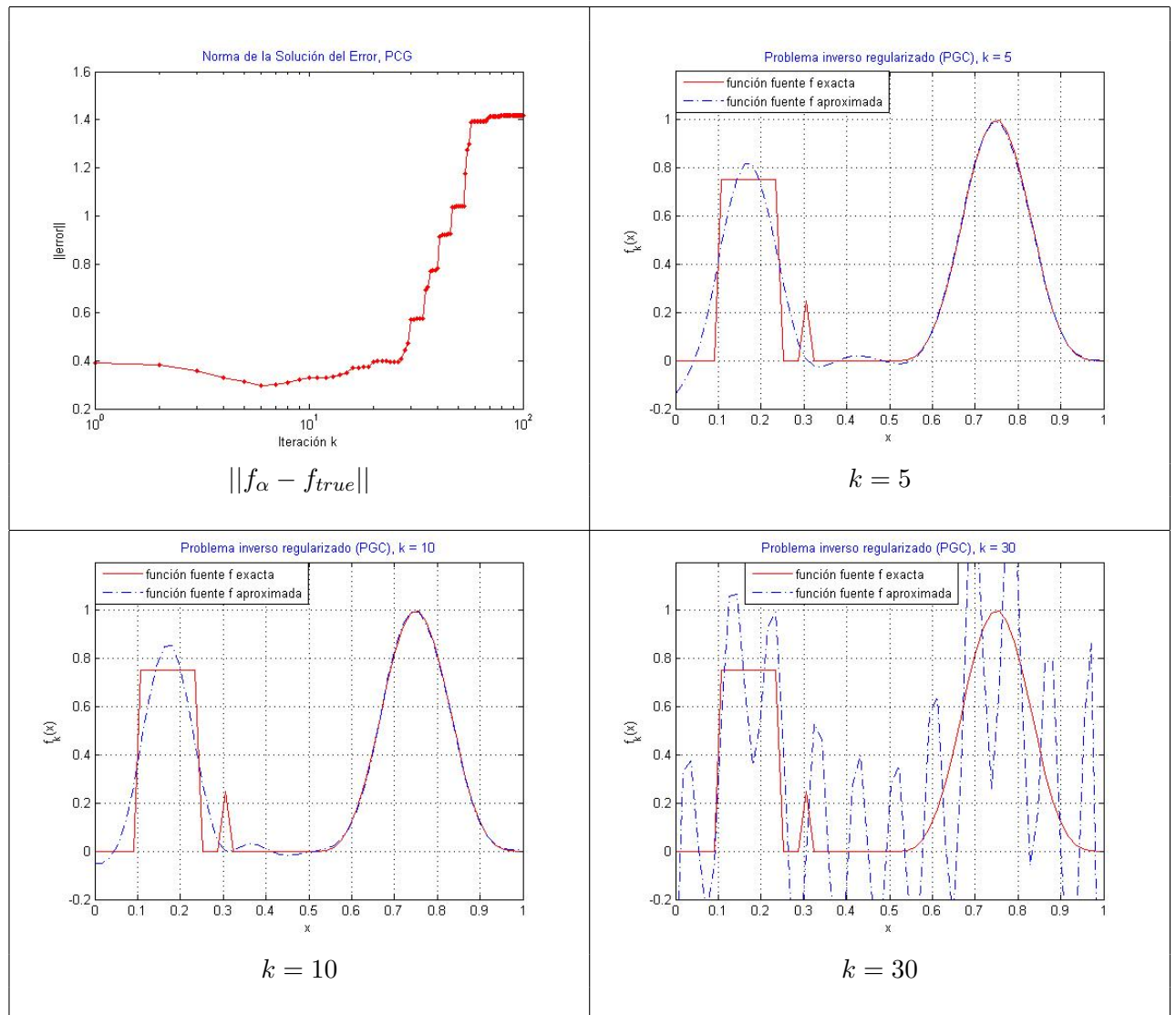


Figura 3.2: Regularización Iterativa, usando Gradiente Conjugado con preconditionamiento

Aplicación a problemas inversos para ecuaciones diferenciales parciales.

En este capítulo presentaremos la aplicación del método de Regularización de Tijonov a la resolución de problemas extremadamente mal planteados: problemas inversos en conducción de calor. Estos problemas surgen en geofísica, ecología, física, entre otras ciencias, y, por lo general, se reducen a ecuaciones integrales de Fredholm. En particular estudiaremos la ecuación de calor con retroceso en el tiempo y, mostraremos algunos ejemplos y resultados numéricos para datos con y sin ruido.

4.1. Ecuación de Calor en 1-D sobre un segmento infinito

Consideremos el problema de Cauchy para la ecuación de calor

$$\begin{aligned} u_t &= a^2 u_{xx}, & -\infty < x < \infty, t > 0 \\ u(x, 0) &= \varphi(x), & -\infty < x < \infty, \text{ Condición inicial} \end{aligned} \quad (4.1)$$

La solución del problema directo de (4.1) es

$$u(x, t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi a^2 t}} e^{-\frac{(x-\xi)^2}{4a^2 t}} \varphi(\xi) d\xi \quad (4.2)$$

4.1.1. Problema Inverso

Vamos a mencionar dos problemas inversos para la ecuación diferencial parcial (4.1). El primer problema inverso consiste en calcular la distribución inicial de temperaturas $\phi(x)$,

suponiendo que se conoce la distribución de temperaturas en un tiempo posterior $t = T > 0$; $u(x, t) = g(x)$, es decir,

Datos: $g(x) = u(x, T)$

Incógnita $\varphi(x) = u(x, 0)$

Este problema se conoce generalmente como el “problema del calor con retroceso en el tiempo”.

Evaluando (4.2) en $t = T$, obtenemos la siguiente ecuación integral de Fredholm de primer tipo

$$u(x, T) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi a^2 T}} e^{-\frac{(x-\xi)^2}{4a^2 T}} \varphi(\xi) d\xi. \quad (4.3)$$

Este problema tiene una solución única dado que la ecuación homogénea ($g(x) = 0$) satisface:

$$\int_{-\infty}^{\infty} (x - \xi)^n e^{-\frac{(x-\xi)^2}{4a^2 t}} \varphi(\xi) d\xi = 0, \quad n = 0, 1, \dots$$

El sistema $\left\{ \frac{x^n}{e^{x^2}} \right\}_{n=1}^{\infty}$ es completo en $C^{\infty}(\mathbb{R})$, esto implica que

$$\varphi(\xi) = 0, \quad -\infty < x < \infty \quad (\text{ver [2]})$$

El segundo problema inverso de (4.1) trata de calcular $\varphi(x)$ suponiendo que se conoce la evolución de la temperatura en algún punto x_0 .

Elijiendo $x_0 = 0$, el dato es: $u(0, t) = g(t)$

$$u(0, t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi a^2 t}} e^{-\frac{\xi^2}{4a^2 t}} \varphi(\xi) d\xi = g(t) \quad (4.4)$$

esta es una ecuación integral de Fredholm de primer tipo. La ecuación integral (4.4) tiene una infinidad de soluciones cuando $g(t) = 0$, debido a que cualquier función impar $\varphi(\xi)$, resuelve la ecuación homogénea.

4.2. Ecuación de Calor en 1-D sobre un segmento finito

Consideremos la ecuación de calor en un segmento finito. Tomemos en cuenta el problema con valores en la frontera.

$$\begin{aligned} u_t &= a^2 u_{xx}, & 0 < x < \ell, & t > t_0 > 0 \\ u(x, 0) &= \varphi(x), & 0 \leq x \leq \ell, & \text{Condición inicial} \\ u(0, t) &= u(\ell, t) = 0, & t \geq t_0 & \text{Condiciones de frontera} \end{aligned} \quad (4.5)$$

El problema directo se puede resolver por medio de separación de variables y series de Fourier

$$u(x, t) = \sum_{n=1}^{\infty} \varphi_n e^{-(\frac{n\pi a}{\ell})^2(t-t_0)} u_n(x)$$

donde

$$u_n(x) = \sqrt{\frac{2}{\ell}} \sin\left(\frac{n\pi x}{\ell}\right), \quad n = 1, 2, \dots$$

$$\varphi_n = \sqrt{\frac{2}{\ell}} \int_0^{\ell} \varphi(\xi) \sin\left(\frac{n\pi \xi}{\ell}\right) d\xi, \quad n = 1, 2, \dots$$

es decir,

$$u(x, t) = \int_0^{\ell} \sum_{n=1}^{\infty} \frac{2}{\ell} \varphi(\xi) \sin\left(\frac{n\pi \xi}{\ell}\right) \exp\left\{-\left(\frac{n\pi a}{\ell}\right)^2(t-t_0)\right\} \sin\left(\frac{n\pi x}{\ell}\right) d\xi \quad (4.6)$$

4.2.1. Problema Inverso

Ahora, formulemos para (4.5) el problema con retroceso en el tiempo, es decir, dado $g(x) = u(x, T)$ encontrar $\varphi(x) = u(x, 0)$.

Evaluando (4.6) en $t = T$, obtenemos

$$u(x, T) = \sum_{n=1}^{\infty} \left(\frac{2}{\ell} \int_0^{\ell} \varphi(\xi) \sin\left(\frac{n\pi \xi}{\ell}\right) d\xi \right) e^{-(\frac{n\pi a}{\ell})^2(T-t_0)} \sin\left(\frac{n\pi x}{\ell}\right) = g(x), \quad (4.7)$$

$$0 \leq x \leq \ell.$$

Notemos que el conjunto de las funciones u_n , $n = 1, 2, \dots$ es un sistema ortonormal completo en $L_2[0, \ell]$. Además, el problema tiene solución única cuando $g(x) = 0$, debido a que en este caso

$$\sum_{n=1}^{\infty} \varphi_n e^{-(\frac{n\pi a}{\ell})^2(T-t_0)} u_n(x) = 0.$$

Multiplicando esta última ecuación por $u_k(x)$ e integrando sobre $[0, \ell]$, se obtiene

$$e^{-(\frac{n\pi a}{\ell})^2(T-t_0)} \varphi_k = 0,$$

entonces

$$\varphi_k = 0,$$

para toda $k = 1, 2, \dots, \infty$. Así que

$$\int_0^\ell \varphi(\xi) \sin\left(\frac{k\pi\xi}{\ell}\right) d\xi = 0 \quad \forall k \in \mathbb{N},$$

por lo cual

$$\varphi(\xi) = 0, \quad 0 \leq \xi \leq \ell$$

de donde se sigue que (4.7) tiene solución única.

Ahora, verifiquemos las propiedades de estabilidad de la ecuación (4.7). Multiplicando (4.7) por u_k se obtiene

$$\varphi_k = e^{(\frac{k\pi a}{\ell})^2(T-t_0)} g_k,$$

con

$$g_k = \sqrt{\frac{2}{\ell}} \int_0^\ell g(\xi) \sin\left(\frac{k\pi\xi}{\ell}\right) d\xi.$$

Por la igualdad de Parseval, tenemos

$$\|\varphi\|_{L_2[0,\ell]}^2 = \sum_{k=1}^{\infty} |\varphi_k|^2 = e^{2(\frac{n\pi a}{\ell})^2(T-t_0)} \sum_{k=1}^{\infty} |g_k|^2,$$

esto implica que

$$\|\varphi\|_{L_2[0,\ell]} = e^{2(\frac{n\pi a}{\ell})^2(T-t_0)} \|g\|_{L_2[0,\ell]},$$

si tomamos $\varphi(x) = k u_k(x)$, entonces al sustituir en (4.7) obtenemos

$$k e^{-(\frac{k\pi a}{\ell})^2(T-t_0)} u_k(x) = g(x).$$

Luego

$$\|g\|_{L_2[0,\ell]} \rightarrow 0 \quad \text{cuando } k \rightarrow \infty,$$

pero

$$\|\varphi\|_{L_2[0,\ell]} = k \rightarrow \infty \quad \text{cuando } k \rightarrow \infty,$$

lo cual muestra que el problema es inestable.

La ecuación (4.7) se puede escribir como

$$\int_0^\ell \left(\sum_{n=1}^{\infty} \frac{2}{\ell} \sin\left(\frac{n\pi\xi}{\ell}\right) \sin\left(\frac{n\pi x}{\ell}\right) e^{-(\frac{n\pi a}{\ell})^2(T-t_0)} \right) \varphi(\xi) d\xi = g(x),$$

de tal forma que se reduce a una ecuación integral de Fredholm de primer tipo

$$K\varphi \equiv \int_0^\ell \mathcal{K}(x, \xi)\varphi(\xi)d\xi = g(x), \quad (4.8)$$

con el kernel

$$\mathcal{K}(x, \xi) = \sum_{n=1}^{\infty} \frac{2}{\ell} \sin\left(\frac{n\pi\xi}{\ell}\right) \sin\left(\frac{n\pi x}{\ell}\right) e^{-\left(\frac{n\pi a}{\ell}\right)^2(T-t_0)},$$

obsérvese que el problema de la ecuación (4.8) es mal-planteada, ya que la solución no depende continuamente de los datos, debido a la inestabilidad del problema (4.7).

Ahora, como los conjuntos de funciones $\{u_n\}$ y $\{\varphi_n\}$ son sistemas ortonormales en $L_2[0, \ell]$, entonces, un sistema singular para el operador integral K en (4.8) viene dado por

$$\left\{ e^{-\left(\frac{n\pi a}{\ell}\right)^2} ; \sqrt{\frac{2}{\ell}} \sin\left(\frac{n\pi x}{\ell}\right) ; \sqrt{\frac{2}{\ell}} \sin\left(\frac{n\pi x}{\ell}\right) \right\}.$$

Por lo tanto, las funciones singulares son completas en $L_2[0, \ell]$, y tenemos que $\mathcal{N}(K) = \mathcal{N}(K^*) = \{0\}$, y $\mathcal{D}(K^\dagger) = \mathcal{R}(K)$ es denso en $L_2[0, \ell]$. Entonces, del Teorema 3.2.2 se sigue que (4.8) y por tanto su problema inverso es soluble si, y solo si,

$$\sum_{n=1}^{\infty} e^{2\left(\frac{n\pi a}{\ell}\right)^2} |g_n|^2 < \infty \quad (4.9)$$

donde g_n son los coeficientes de Fourier de g . En este caso la solución es dada por

$$\varphi(x) = \sqrt{\frac{2}{\ell}} \sum_{n=1}^{\infty} e^{2\left(\frac{n\pi a}{\ell}\right)^2} g_n \sin\left(\frac{n\pi x}{\ell}\right) \quad (4.10)$$

(4.9) y (4.10) muestran que este problema inverso es extremadamente mal planteado: existe solución si los coeficiente de Fourier $\{g_n\}$ de g decaen extremadamente rápido (más rápido que $\left\{e^{-\left(\frac{n\pi a}{\ell}\right)^2}\right\}$), es decir, para g muy suave, [6].

Nota: Para $\mathcal{K}(x, \xi)$ suficiente suave, el operador K definido anteriormente es compacto en un espacio de Banach razonable, tales como un espacio L_p . Incluso si la correspondencia es biyectiva su inversa no será continua. Así pequeños errores en los datos g son amplificados en la solución φ . En este sentido el problema de inferir φ de g está mal planteado.

Por otro lado, como el kernel $\mathcal{K}(x, \xi)$ tiene derivadas parciales continuas de cualquier orden en $0 \leq x, \xi \leq \ell$, entonces podemos considerar que el operador lineal K actúa de $\mathcal{C}[0, \ell]$ en $\mathcal{C}^p[0, \ell]$, donde $p \in \mathbb{N}$. En este caso el problema de resolver (4.8) también es mal planteado.

Ahora, obtendremos una estimación de estabilidad suponiendo que tenemos información adicional sobre la solución del problema. Por simplicidad, asumimos que $a^2 = 1$, $u(x, t)$ es continua, y que tiene derivadas continuas $u_t(x, t)$, $u_{xx}(x, t)$ y satisface (4.5) para $0 \leq x \leq 1$, $t_0 - \varepsilon \leq T$ donde ε es un número fijo positivo. Además, supongamos que $u(x, t)$ no es idénticamente cero.

Para $t \in [t_0 - \varepsilon, T]$, consideremos la función de energía

$$g(t) = \int_0^\ell u^2(x, t) dt,$$

derivando dos veces, obtenemos

$$g'(t) = 2 \int_0^\ell u(x, t) u_t(x, t) dx,$$

$$g''(t) = 2 \int_0^\ell \left[(u_t(x, t))^2 + u(x, t) u_{tt}(x, t) \right] dx.$$

Como

$$(u_t(x, t))_t = (u_{xx}(x, t))_t = (u_t(x, t))_{xx} = u_{xxxx}(x, t),$$

entonces

$$g''(t) = 2 \int_0^\ell \left[(u_t(x, t))^2 + u(x, t) u_{xxxx}(x, t) \right] dx.$$

Integrando el segundo término por partes y tomando en cuenta las condiciones de frontera definidas en (4.5) obtenemos

$$\int_0^\ell u(x, t) u_{xxxx}(x, t) dx = \int_0^\ell (u_{xx}(x, t))^2 dx = \int_0^\ell (u_t(x, t))^2 dx,$$

por lo tanto,

$$g''(t) = 4 \int_0^\ell (u_t(x, t))^2 dx.$$

Ahora, consideremos la función $h(t) = \ln(g(t))$. Queremos demostrar que $h(t)$ es cóncava, es decir que $h''(t) \geq 0$ para $t \in [t_0 - \varepsilon, T]$. Notemos que

$$\begin{aligned} h''(t) &= \frac{1}{g^2(t)} \left[g''(t)g(t) - (g'(t))^2 \right] \\ &= \frac{4}{g^2(t)} \left[\int_0^\ell (u_t(x, t))^2 dx \int_0^\ell (u(x, t))^2 dx - \left(\int_0^\ell u_t(x, t)u(x, t) dx \right)^2 \right] \\ &= \frac{4}{g^2(t)} \left[\|u_t(x, t)\|^2 \|u(x, t)\|^2 - \langle u_t(x, t), u(x, t) \rangle^2 \right]. \end{aligned}$$

Por lo tanto, usando la desigualdad de Cauchy-Schwartz, obtenemos que $h''(t) \geq 0$ para $t \in [t_0 - \varepsilon, T]$. La no-negatividad de $h''(t)$ en el intervalo $[t_0 - \varepsilon, T]$ nos dice que $h(t)$ es una

función convexa, lo cual implica que

$$h(t) \leq h(t_0 - \varepsilon) \frac{T - t}{T - (t_0 - \varepsilon)} + h(T) \frac{t - (t_0 - \varepsilon)}{T - (t_0 - \varepsilon)},$$

para $t \in [t_0 - \varepsilon, T]$. De donde se sigue que

$$\ln(g(t)) \leq \frac{T - t}{T - (t_0 - \varepsilon)} \ln(g(t_0 - \varepsilon)) + \frac{t - (t_0 - \varepsilon)}{T - (t_0 - \varepsilon)} \ln(g(T)),$$

entonces

$$g(t) \leq [g(t_0 - \varepsilon)]^{\frac{T-t}{T-(t_0-\varepsilon)}} [g(T)]^{\frac{t-(t_0-\varepsilon)}{T-(t_0-\varepsilon)}}, \quad (4.11)$$

para $t \in [t_0 - \varepsilon, T]$. Esta estimación de energía se utiliza para encontrar una condición de estabilidad para la solución del problema de calor con retroceso en el tiempo. Se procede como se indica a continuación.

Consideremos las funciones $u_i(x, t)$ donde $i = 1, 2$, las cuales satisfacen condiciones similares a la de $u(x, t)$, es decir pedimos que sean solución de la ecuación de norma finita en el tiempo inicial

$$\left[\int_0^\ell (u_i(x, t_0 - \varepsilon))^2 dx \right]^{1/2} \leq C, \quad i = 1, 2, \quad (4.12)$$

donde C es una constante positiva. Denotando

$$u(x, t) = u_1(x, t) - u_2(x, t),$$

obtenemos de (4.11)

$$\int_0^\ell u^2(x, t) dx \leq \left[\int_0^\ell u^2(x, t_0 - \varepsilon) dx \right]^{\frac{T-t}{T-(t_0-\varepsilon)}} \left[\int_0^\ell u^2(x, T) dx \right]^{\frac{t-(t_0-\varepsilon)}{T-(t_0-\varepsilon)}}.$$

Tomando $t = t_0$ y usando la desigualdad (4.12) obtenemos

$$\|u(x, t_0)\|_{L_2[0, \ell]} \leq (2C)^{\frac{T-t_0}{T-(t_0-\varepsilon)}} \|u(x, T)\|_{L_2[0, \ell]}^{\frac{\varepsilon}{T-(t_0-\varepsilon)}}. \quad (4.13)$$

Esta desigualdad es el estimador de la estabilidad condicional para la ecuación de calor con retroceso en el tiempo. Esto se obtiene utilizando la hipótesis de que $u_i(x, t)$, $i = 1, 2$ satisface la desigualdad (4.12), es decir, sus normas en $L_2[0, \ell]$ para $t = t_0 - \varepsilon$ son acotadas. Esta estimación es un tanto extraña, ya que se obtuvo de una solución del problema inverso asumiendo únicamente (4.12) y esto no proporciona la compacidad del conjunto en $L_2[0, \ell]$. Veamos el siguiente ejemplo.

Ejemplo 4.2.1. Consideremos el siguiente problema de valores a la frontera

$$\begin{aligned} u_t &= u_{xx}, & 0 < x < 1, & 0 < t \leq T \\ u(0, t) &= u(1, t) = 0, & 0 \leq t \leq T. \end{aligned} \quad (4.14)$$

con condición inicial

$$u(x, 0) = \varphi(x) = \begin{cases} x, & x \leq 1/2; \\ 1 - x, & x > 1/2. \end{cases} \quad (4.15)$$

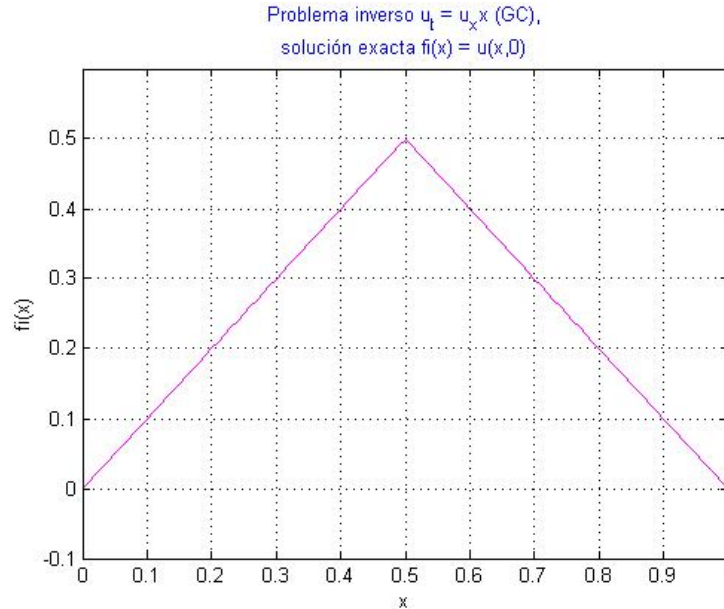


Figura 4.1: $u(x, 0) = \varphi(x)$

La solución del problema directo se puede encontrar utilizando separación de variables y series de Fourier. El resultado es

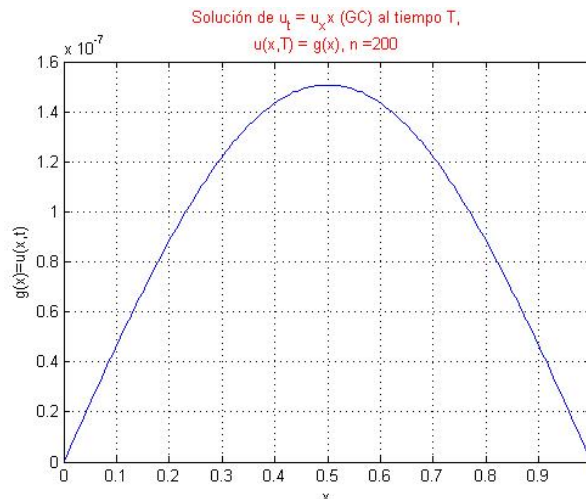
$$u(x, t) = \frac{4}{\pi^2} \sum_{n=1}^{\infty} \frac{\sin\left(\frac{n\pi}{2}\right)}{n^2} e^{-(n\pi)^2 t} \sin(n\pi x). \quad (4.16)$$

El propósito es resolver el problema inverso: suponiendo que conocemos la solución $u(x, T)$ con $T = 1.5$, queremos recuperar el valor inicial $\varphi(x) = u(x, 0)$.

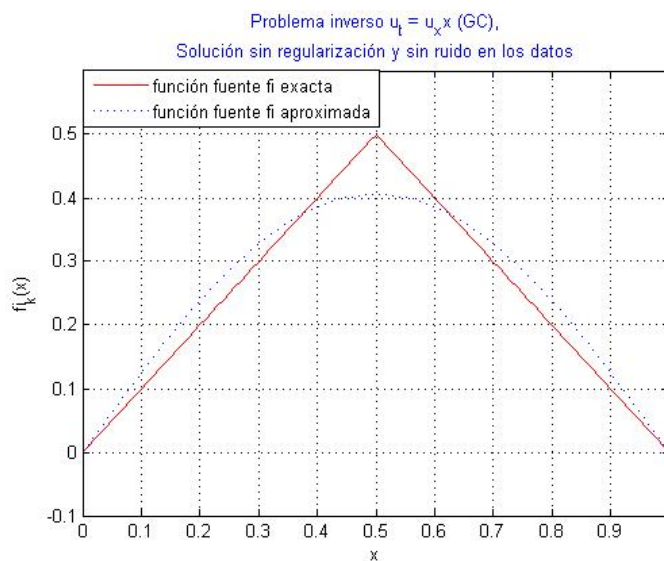
Como ya hemos discutido anteriormente el problema es extremadamente inestable. Utilizaremos el método de regularización de Tijonov aplicado al problema (4.8) tomando cuatro términos de la serie que representa al kernel $\mathcal{K}(x, \xi)$. Asimismo consideramos el valor $u(x, 1.5)$ utilizando cuatro términos de la serie (4.16). También consideraremos el caso en que a los datos le agregamos un ruido aleatorio

$$u(x, 1.5) + \delta \quad \text{con} \quad \delta \leq \|g\|.$$

Caso 1. Datos sin ruido.

Figura 4.2: $u(x, 1.5) = g(x)$

Observe en la Figura (4.3) la solución obtenida sin ruido en los datos. En este caso, obtenemos una buena solución sin realizar regularización de Tijonov al problema truncado, debido a que al truncar las series ya estamos aplicando una especie de regularización por medio de filtrado, al excluir los términos de alta frecuencia que son los más inestables.

Figura 4.3: $u(x, 1.5) = g(x)$, $\frac{\|\varphi_{exacta} - \varphi_{aprox}\|}{\|\varphi_{exacta}\|} = 0.1204$

Ahora veamos las soluciones obtenidas con la Regularización de Tijonov en la Figura (4.4), el cual utiliza el método de Gradiente Conjugado y una tolerancia de 10^{-16} (como método de paro).

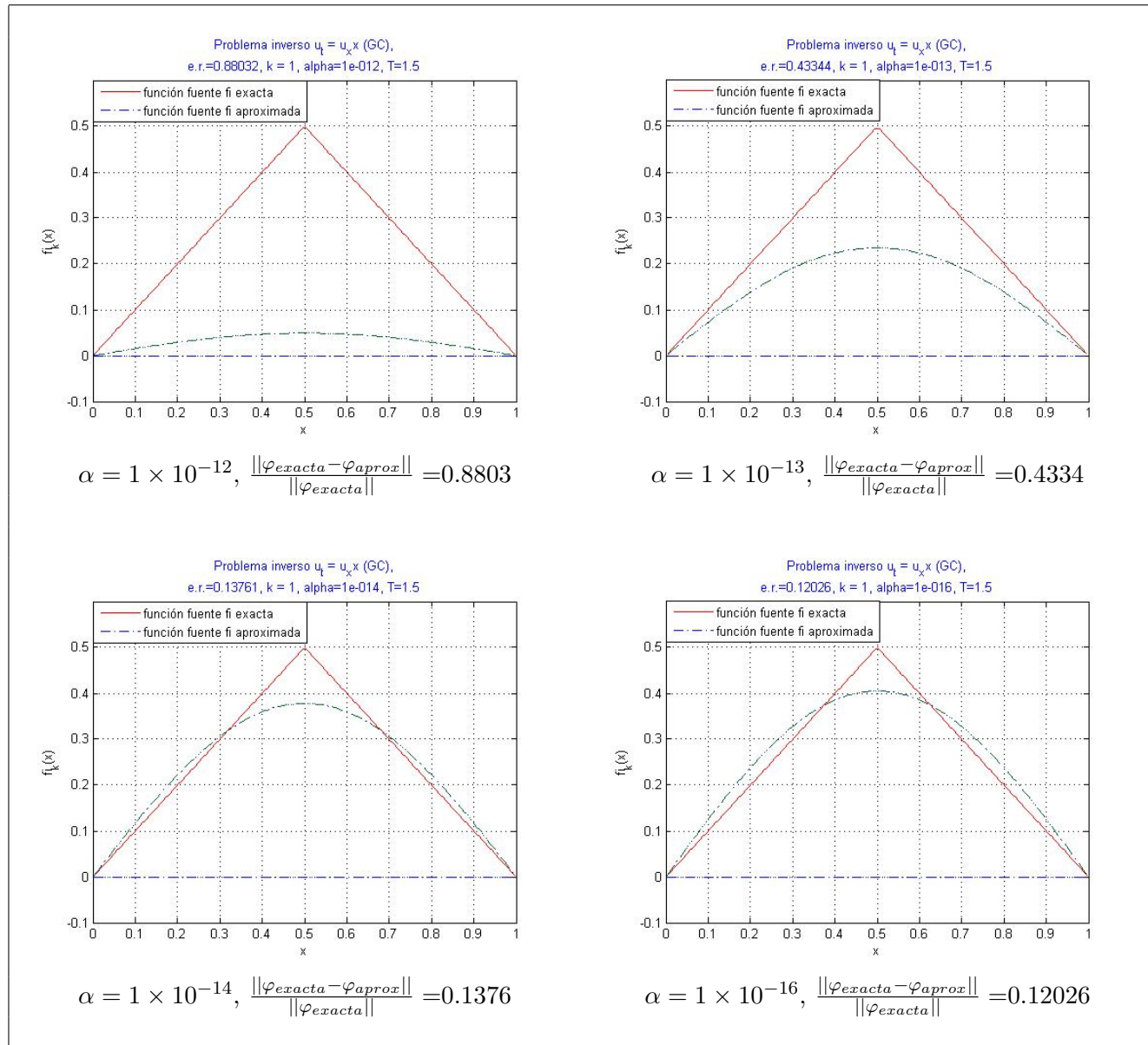


Figura 4.4: Soluciones del problema del calor con retroceso en el tiempo (4.14, sin ruido en los datos), con Regularización de Tijonov.

Cabe destacar que, cuando $\alpha \leq 1 \times 10^{-16}$ la solución ya no cambia y el error relativo es 0.12026, además esta es la mejor solución para el problema sin ruido en los datos.

Caso 2. Datos con ruido. En la siguiente Figura se muestra la función $u(x, T) = g(x)$ con ruido, la cual se construye de la siguiente forma:

$$g(x) = g(x) + \delta, \quad \text{con} \quad \|\delta\| = 1.5625 \times 10^{-6}.$$

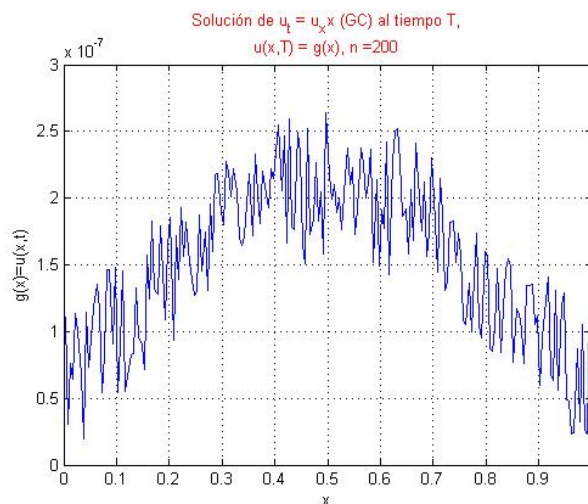


Figura 4.5: $u(x, 1.5) = g(x)$, datos con ruido.

En la Figura (4.6) podemos apreciar la solución obtenida sin utilizar un método de regularización.

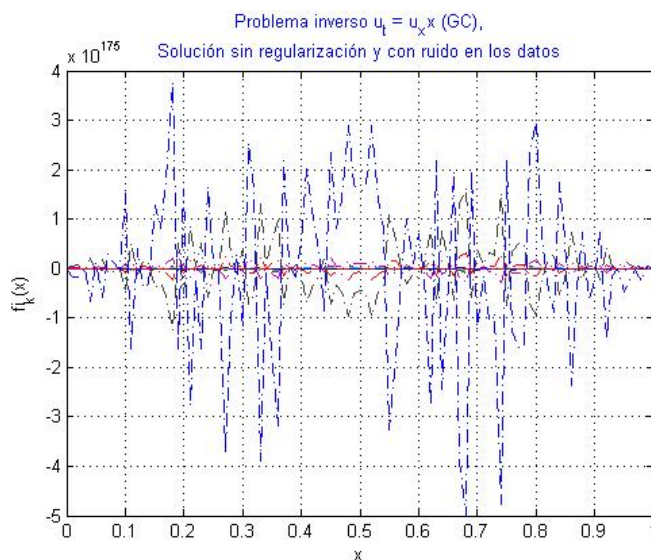


Figura 4.6: Función inicial $\phi(x) = u(x, 0)$ obtenida sin regularización.

Notemos que, la amplificación exponencial del error aparece para un tiempo corto arbitrario ($T > 0$), pero llega a ser peor a medida que aumenta T . A diferencia del caso anterior, en esta ocasión es necesario utilizar un método de regularización para obtener una solución que describa las condiciones iniciales en forma razonable. Utilizando el método de regularización de Tijonov combinado con el método de gradiente conjugado con una tolerancia de 10^{-16} , se obtienen los resultados mostrados en la Figura 4.7.

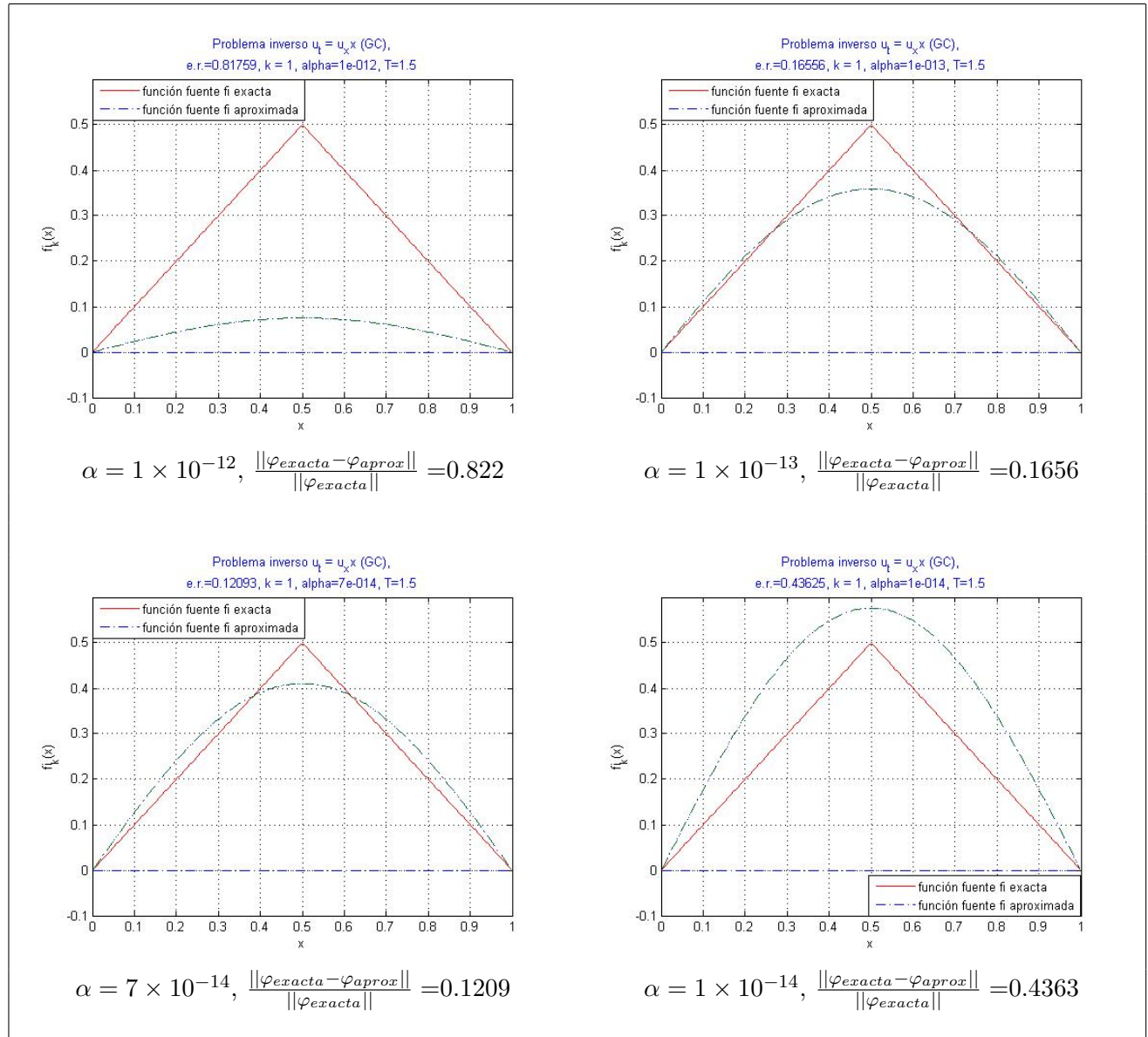


Figura 4.7: Soluciones del problema del calor con retroceso en el tiempo (4.14, datos con ruido), con Regularización de Tijonov.

Este problema inverso es extremadamente mal condicionado. A pesar de que el problema tiene solución única para g en la cual sus coeficientes de Fourier g_k decaen extremadamente rápido (mucho más que $\{e^{-(n\pi)^2}\}$), el problema es altamente inestable debido a que un error pequeño en el n -ésimo coeficiente de Fourier es amplificado por el factor $e^{(n\pi)^2}$. De hecho, un error alrededor de 10^{-8} en el quinto coeficiente de Fourier de los datos nos conduce a un error de aproximadamente 10^3 en la temperatura inicial. Debido a esta alta inestabilidad de los modos de alta frecuencia, debemos retener solamente tres primeros términos en la serie de Fourier y filtrar todos los demás, con el objeto de controlar el crecimiento del ruido en la solución del problema (4.8). Al problema filtrado resultante, el cual sigue siendo mal condicionado, se le aplica la regularización de Tijonov en combinación con el método iterativo de gradiente conjugado. Retener más términos en la serie de Fourier amplifica el ruido enormemente y hace casi imposible recobrar una buena solución a pesar de la regularización del problema resultante. Lo anterior muestra el necesario “compromiso entre precisión y estabilidad”.

El problema de calor con retroceso en el tiempo es un problema extremadamente difícil y altamente inestable como ya se ha demostrado. En el caso unidimensional se puede hacer truncamiento de la serie asociada al kernel en (4.8), además de la discretización de la integral por medio de la regla del punto medio. En el caso de los datos sin ruido, es posible recobrar una solución razonable sin realizar regularización adicional. Sin embargo, en el caso de datos con ruido, a pesar de la “prerregularización”, debido al truncamiento y discretización de la integral, el problema discreto resultante sigue siendo muy mal condicionado y la regularización es indispensable, como muestran los resultados.

4.3. Ecuación de Calor en 2-D sobre un segmento finito

Consideremos la caja rectangular $D = [0, X] \times [0, Y]$, y sea ∂D su frontera. Ahora, tomemos en cuenta el siguiente problema de valores en la frontera en dos dimensiones, para una función $u = u(x, y, t)$

$$\begin{aligned} u_t &= a^2(u_{xx} + u_{yy}), & \text{en } D, t > t_0 > 0, \\ u(x, y, 0) &= \varphi(x, y), & \text{en } D, \text{ Condición inicial,} \\ u(x, y, t) &= 0 & \text{sobre } \partial D, t \leq 0. \end{aligned} \tag{4.17}$$

El problema directo se puede resolver por medio de separación de variables y series de Fourier, de tal manera que llegamos a

$$u(x, y, t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \varphi_{mn} e^{-a^2[(\frac{m\pi}{X})^2 + (\frac{n\pi}{Y})^2](t-t_0)} u_{mn}(x, y),$$

donde

$$u_{mn}(x, y) = \sin\left(\frac{m\pi}{X}x\right) \sin\left(\frac{n\pi}{Y}y\right), \quad m, n = 1, 2, \dots,$$

y

$$\varphi_{mn} = \frac{4}{XY} \int_0^Y \int_0^X \varphi(\xi, \zeta) \sin\left(\frac{m\pi}{X}\xi\right) \sin\left(\frac{n\pi}{Y}\zeta\right) d\xi d\zeta, \quad m, n = 1, 2, \dots,$$

es decir,

$$u(x, y, t) = \int_0^Y \int_0^X \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{4}{XY} \varphi(\xi, \zeta) \sin\left(\frac{m\pi\xi}{X}\right) \sin\left(\frac{n\pi\zeta}{Y}\right) e^{-\left[\left(\frac{m\pi a}{X}\right)^2 + \left(\frac{n\pi a}{Y}\right)^2\right](t-t_0)} \sin\left(\frac{m\pi x}{X}\right) \sin\left(\frac{n\pi y}{Y}\right) d\xi d\zeta \quad (4.18)$$

4.3.1. Problema Inverso

El problema de la ecuación de calor bidimensional (4.17) con retroceso en el tiempo es: dado $g(x, y) = u(x, y, T)$ encontrar $\varphi(x, y) = u(x, y, 0)$, para $T \geq t_0 > 0$. Reescribiendo y evaluando (4.18) en $t = T$, obtenemos

$$u(x, y, T) = g(x, y) = \int_0^Y \int_0^X \left\{ \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{4}{XY} \sin\left(\frac{m\pi\xi}{X}\right) \sin\left(\frac{n\pi\zeta}{Y}\right) \sin\left(\frac{m\pi x}{X}\right) \sin\left(\frac{n\pi y}{Y}\right) e^{-\left[\left(\frac{m\pi a}{X}\right)^2 + \left(\frac{n\pi a}{Y}\right)^2\right](T-t_0)} \right\} \varphi(\xi, \zeta) d\xi d\zeta. \quad (4.19)$$

Notemos que, este problema se reduce a la siguiente ecuación de Fredholm de primer tipo

$$K\varphi \equiv \int_0^Y \int_0^X \mathcal{K}(x, y, \xi, \zeta) \varphi(\xi, \zeta) d\xi d\zeta = g(x, y), \quad (4.20)$$

con kernel

$$\mathcal{K}(x, y, \xi, \zeta) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{4}{XY} \sin\left(\frac{m\pi\xi}{X}\right) \sin\left(\frac{n\pi\zeta}{Y}\right) \sin\left(\frac{m\pi x}{X}\right) \sin\left(\frac{n\pi y}{Y}\right) e^{-\left[\left(\frac{m\pi a}{X}\right)^2 + \left(\frac{n\pi a}{Y}\right)^2\right](T-t_0)}.$$

Para dar solución a este problema inverso, vamos a discretizar la ecuación (4.20) con la cuadratura del punto medio, para así obtener un sistema lineal discreto $K\varphi = g$. Tomando $h = \frac{X}{M}$, $k = \frac{Y}{N}$, $x_i = (i - \frac{1}{2})h$ y $y_j = (j - \frac{1}{2})k$ donde $1 \leq i \leq M$ y $1 \leq j \leq N$, obtenemos

$$[K]_{ijlp} = \frac{4}{MN} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \sin\left(\frac{m\pi(l - \frac{1}{2})}{M}\right) \sin\left(\frac{n\pi(p - \frac{1}{2})}{N}\right) \sin\left(\frac{m\pi(i - \frac{1}{2})}{M}\right) \sin\left(\frac{n\pi(j - \frac{1}{2})}{N}\right) e^{-\left[\left(\frac{m\pi a}{X}\right)^2 + \left(\frac{n\pi a}{Y}\right)^2\right](T-t_0)}, \quad (4.21)$$

para resolver este sistema de ecuaciones, se hace el siguiente cambio de variables

$$(i, j) \leftrightarrow (j - 1)M + i = r,$$

$$(l, p) \leftrightarrow (p - 1)N + l = s.$$

Ahora, consideremos el siguiente ejemplo.

Ejemplo 4.3.1. *Supongamos que conocemos la solución del problema directo $u(x, y, T) = g(x, y)$ de la ecuación (4.17) al tiempo $T = 0.01$ con $a = 1$, donde $X = 1$ y $Y = 1$.*

Vamos a resolver el problema inverso usando Regularización de Tijonov y el método de Gradiente Conjugado, además de considerar 3 modos de la serie (4.21), la cual representa el kernel del problema (4.20).

Para obtener una distribución de temperaturas en el tiempo T , $g(x, y) = u(x, y, T)$, dada una distribución inicial de temperaturas $\varphi(x, y) = u(x, y, 0)$, resolvemos numéricamente (4.17) usando un método implícito de direcciones alternantes (ADI, ver apéndice D), donde se considera una malla de tamaño 160×160 , luego, usamos una malla gruesa e interpolamos, de esta forma obtenemos (Figura 4.8):

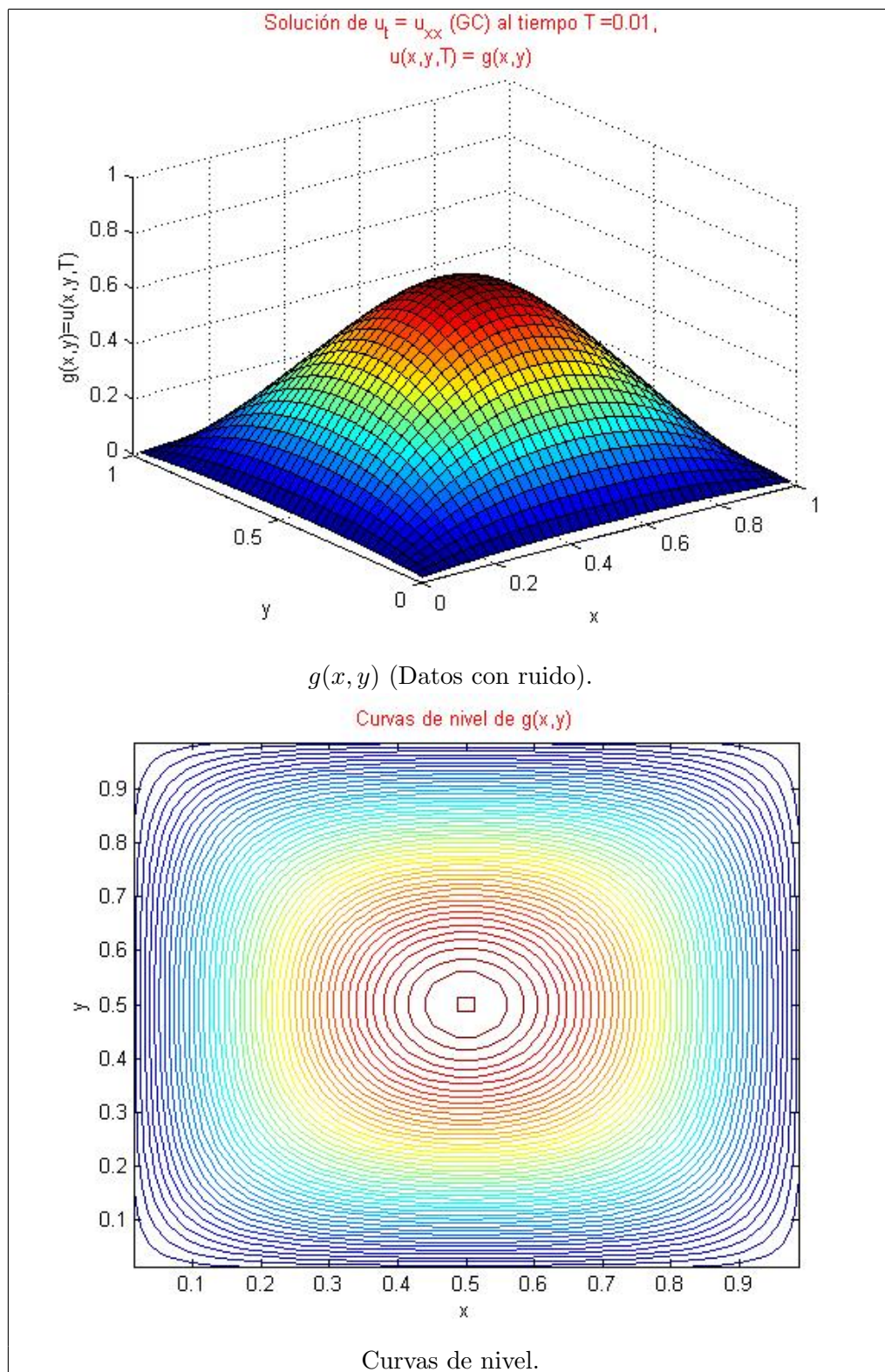


Figura 4.8: Solución del problema directo de la ecuación de calor bidimensional, $u(x,y,0.01) = g(x,y)$.

Para obtener la solución mostrada en la Figura 4.8 se utilizó la condición inicial $\varphi(x, y)$ siguiente

$$\varphi(x, y) = \begin{cases} 2y, & \text{si } x \geq y, 1 - x \geq y; \\ 2(1 - x), & \text{si } x \geq y, 1 - x \leq y; \\ 2(1 - y), & \text{si } x \leq y, 1 - x \leq y; \\ 2x, & \text{si } x \leq y, 1 - x \geq y. \end{cases} \quad (4.22)$$

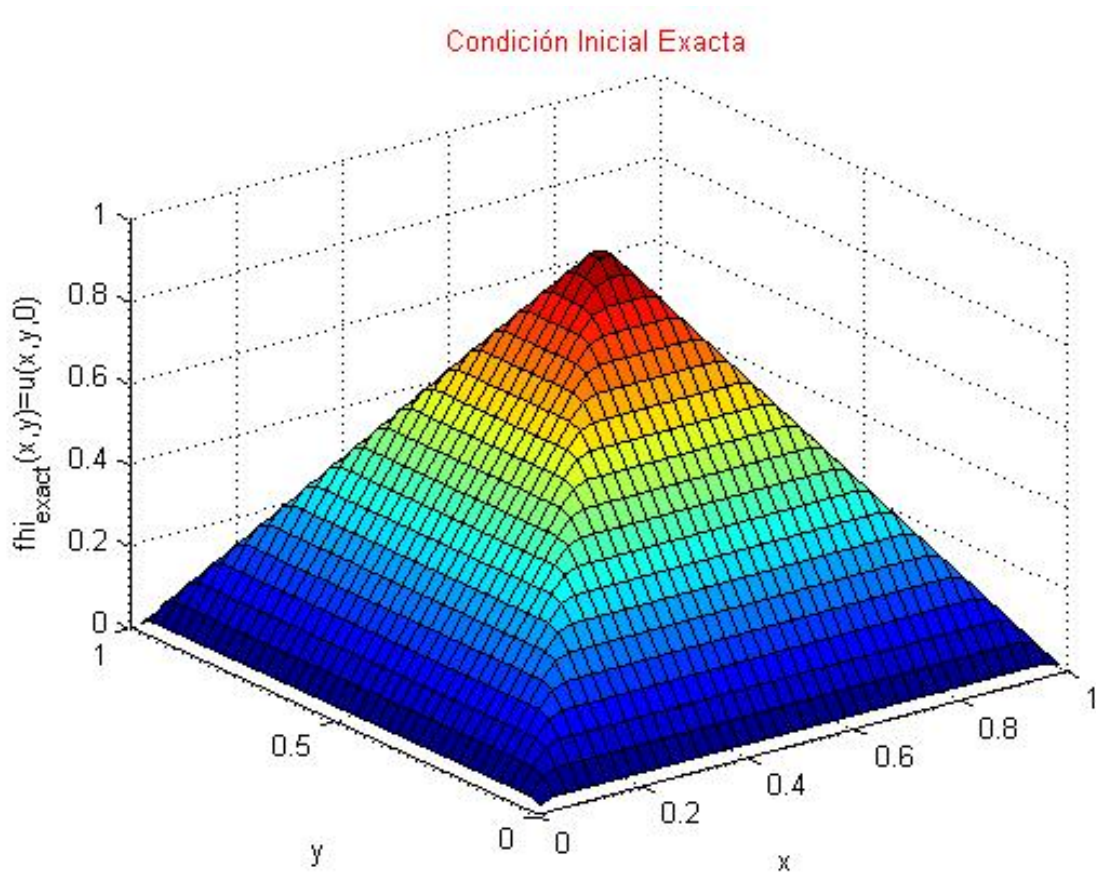


Figura 4.9: Condición inicial exacta $\varphi(x, y)$.

En la Figura 4.10 se muestra la mejor solución recuperada, comparada con la exacta. La solución recuperada difiere de la "exacta" en dos aspectos: su amplitud es menor, y no se pueden recuperar las discontinuidades de la derivada a lo largo de las líneas diagonales del dominio.

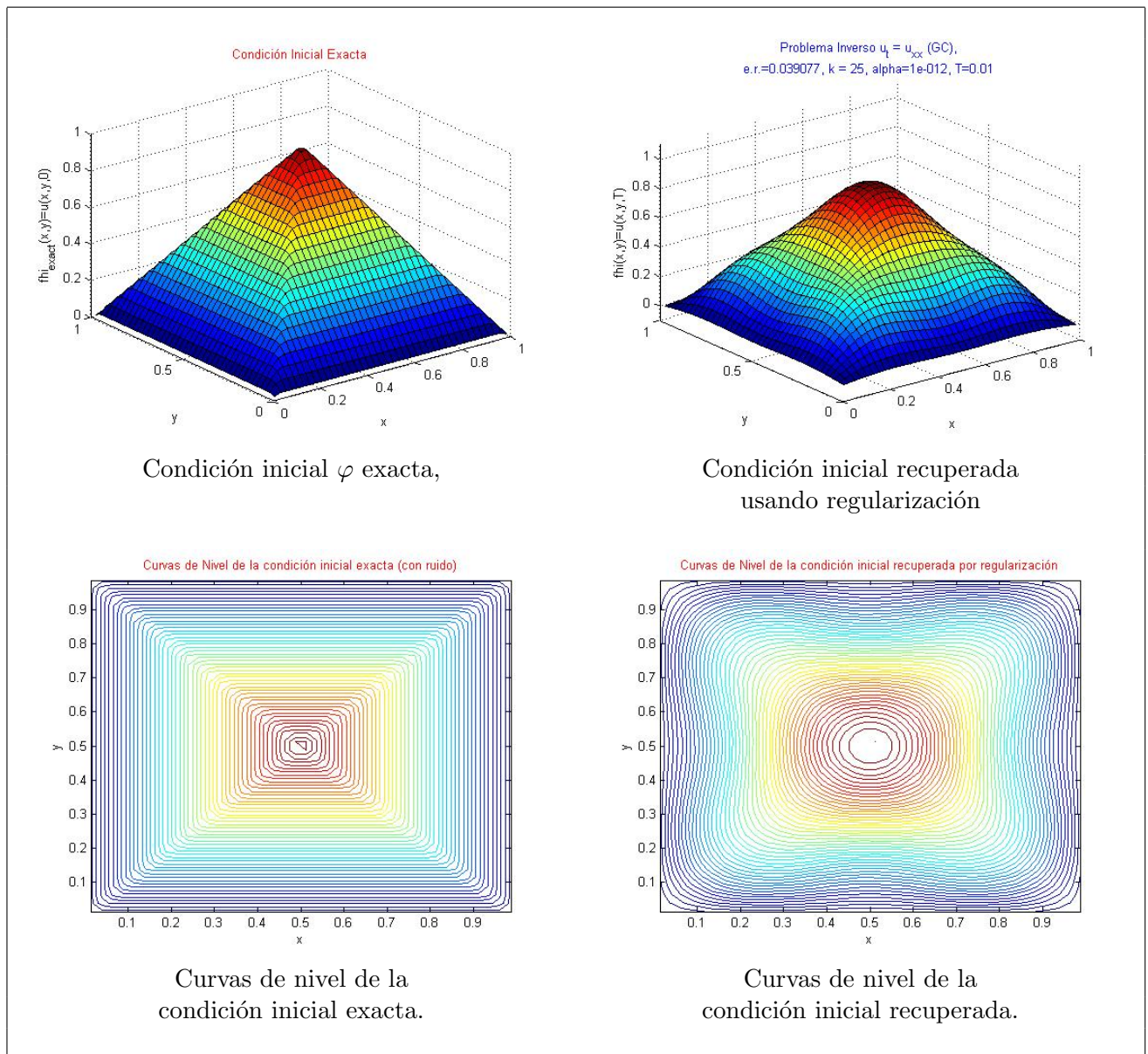


Figura 4.10: Soluciones del problema del calor bidimensional con retroceso en el tiempo.

Notemos que, la diferencia entre la condición inicial exacta y la recuperada por regularización es pequeña, es decir, la condición inicial recuperada tiene un error del 4% comparada con la exacta.

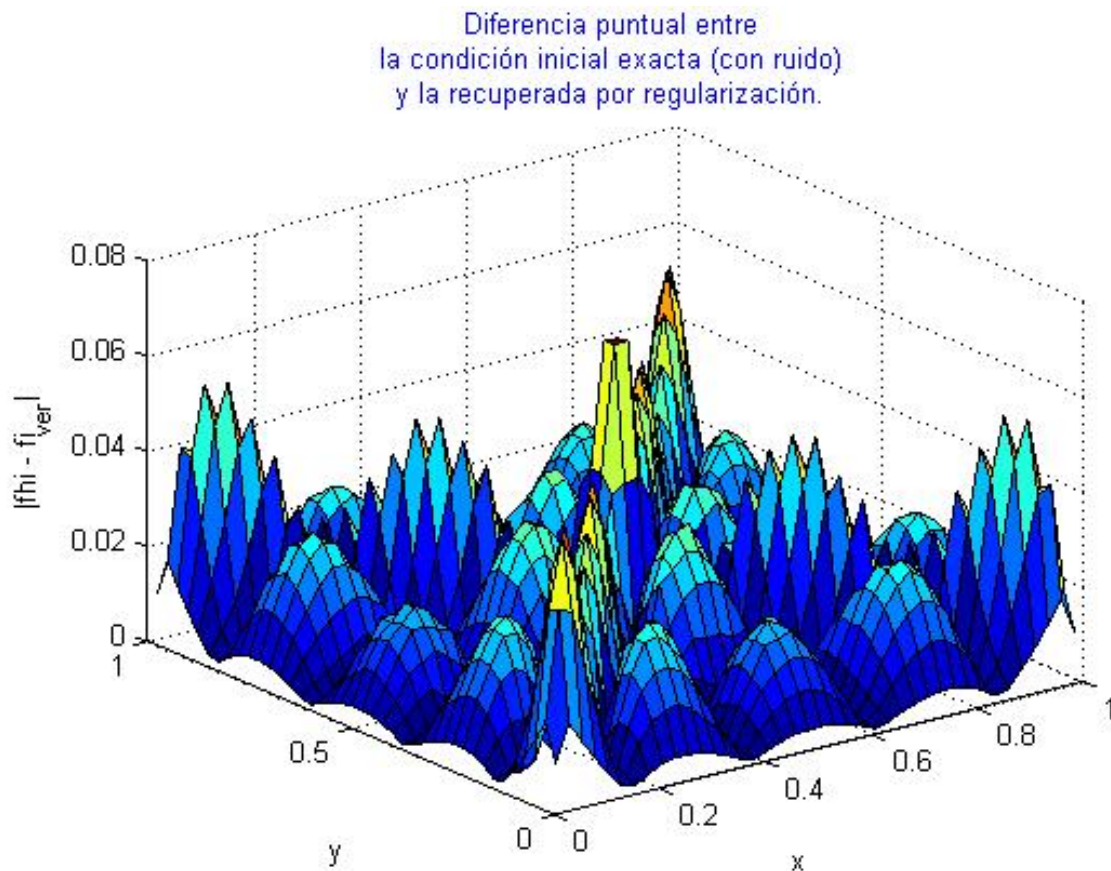


Figura 4.11: Diferencia entre la condición inicial exacta y la recuperada.

Ejemplo 4.3.2. Ahora, nuevamente consideremos $X = 1$, $Y = 1$, y a diferencia del ejemplo anterior, $T = 0.05$; en este ejemplo vamos asumir como condición inicial $u(x, y, 0) = \varphi(x, y)$ a la función "peaks" de MATLAB® 7.0. Ver Figura 4.12.

Resolvemos numéricamente el problema directo, considerando una malla de tamaño 160×160 y usando el método ADI, del cual obtenemos la función $u(x, y, T) = g(x, y)$ que se muestra en la Figura 4.13.

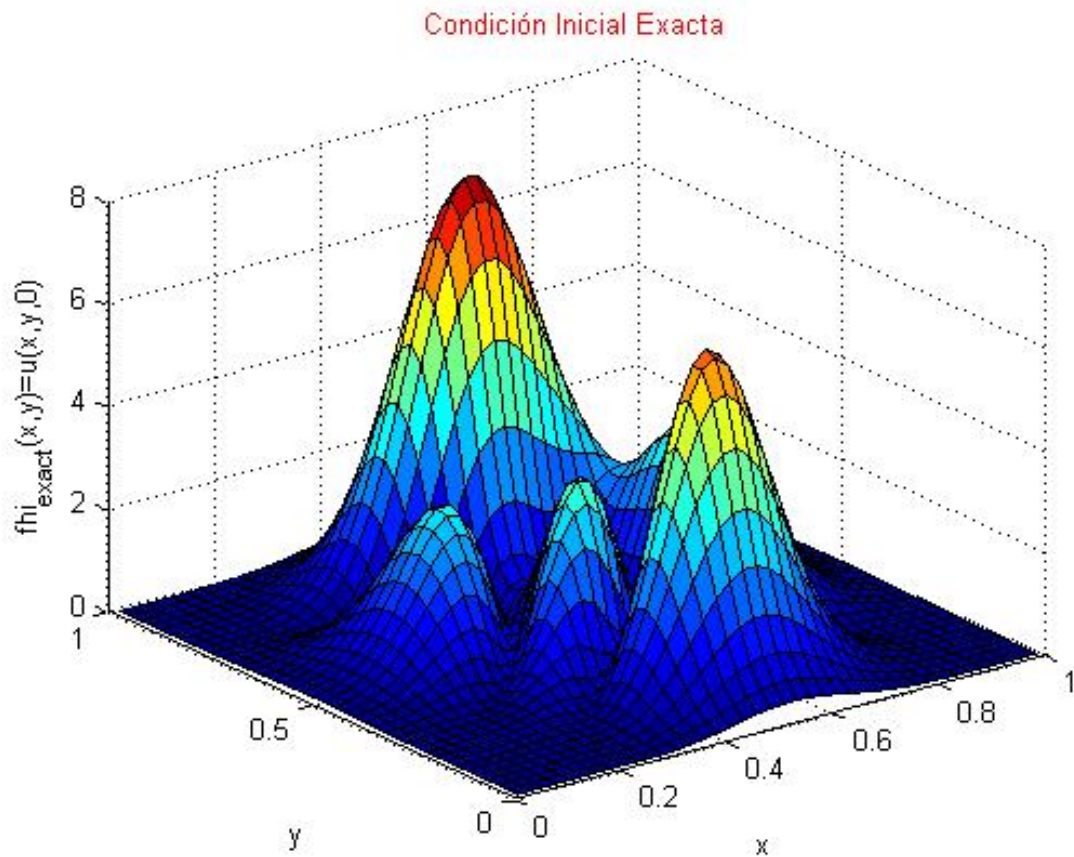


Figura 4.12: Condición inicial exacta con ruido.

Por la manera en que hemos obtenido $g(x, y)$ podemos asumir que los datos tienen ruido, además, notemos que hemos perdido las discontinuidades que se tienen en la condición inicial, como es de esperarse. Por otro lado, cabe aclarar, que usamos este tiempo final T por que de otra forma tenemos: si T es mayor no recuperamos información y, si T es más pequeño, no podremos ilustrar el método de regularización del problema inverso.

En la figura 4.13 se ilustra la función $g(x, y)$ con sus correspondientes curvas de nivel.

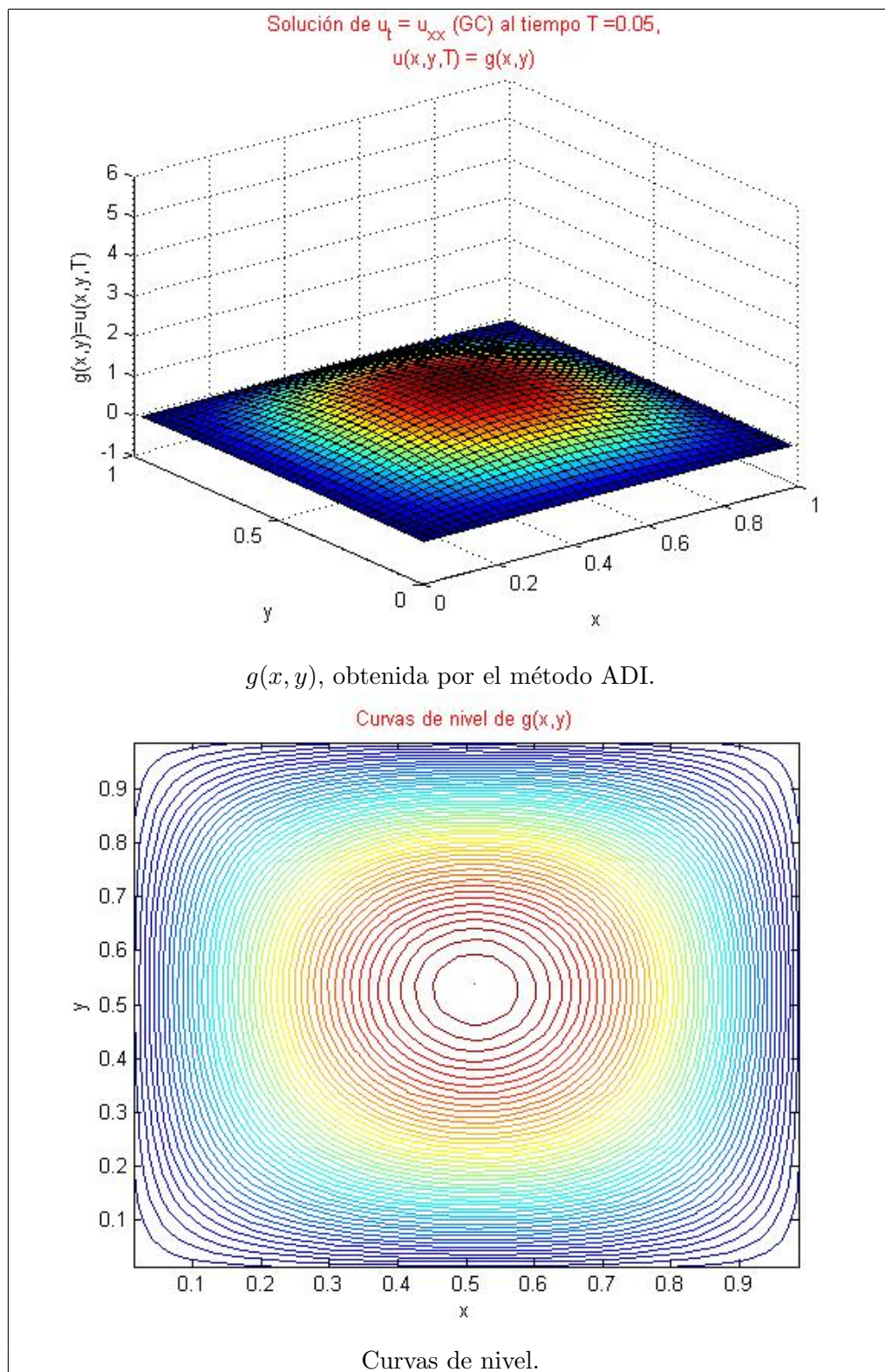


Figura 4.13: Solución del problema directo de la ecuación de calor bidimensional, $u(x,y,0.01) = g(x,y)$.

Ahora, observemos en la figura 4.14 la comparación entre la condición inicial exacta (con ruido) y la condición inicial recuperada por el método de regularización de Tijonov y Gradiente Conjugado, donde el parámetro de regularización es $\alpha = 10^{-15}$.

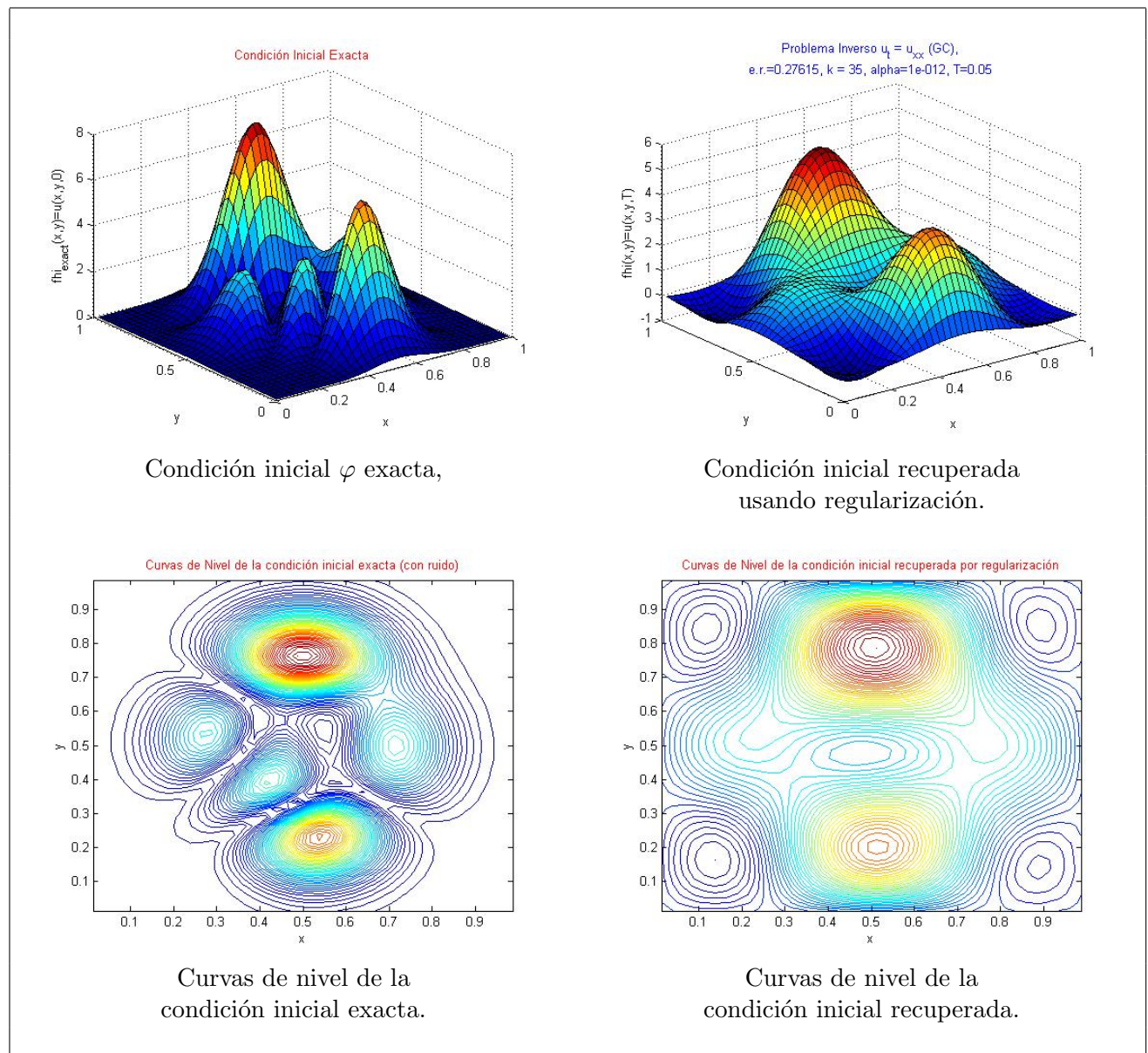


Figura 4.14: Soluciones del problema del calor bidimensional con retroceso en el tiempo.

Observe que, perdemos aproximadamente el 28% de información.

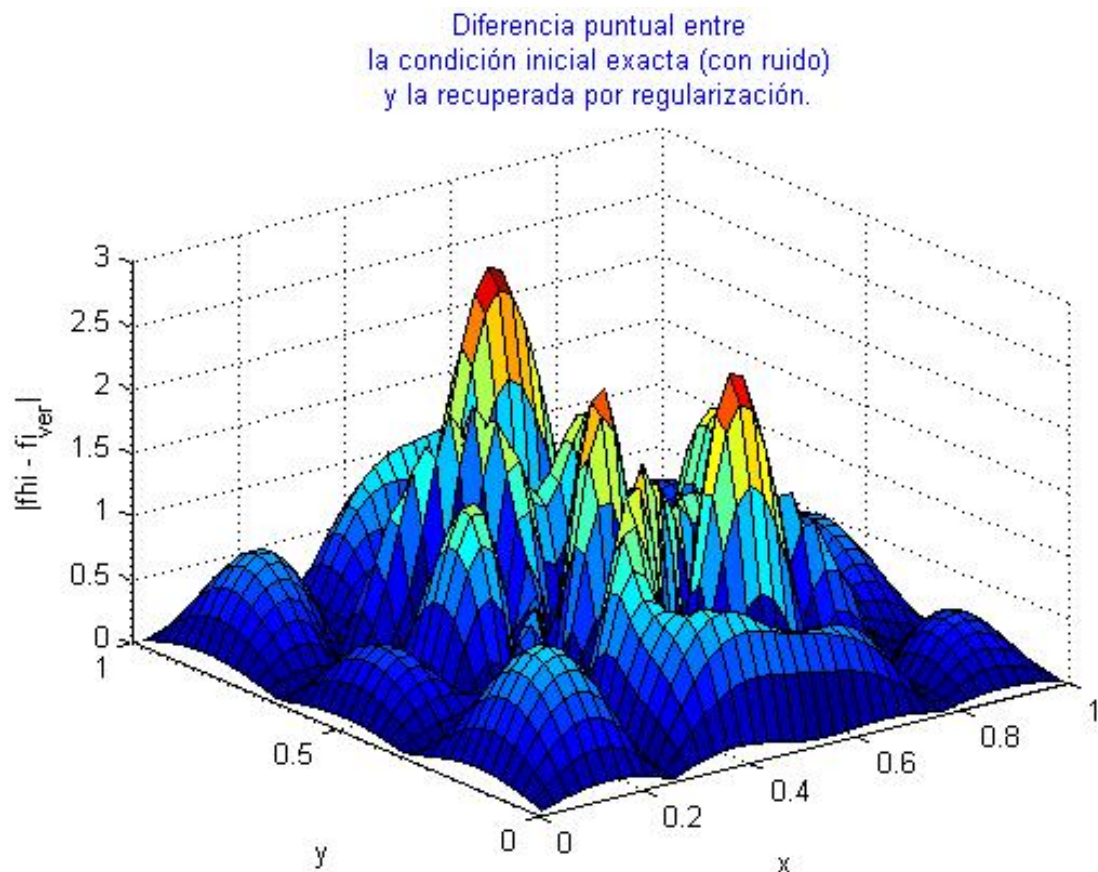


Figura 4.15: Diferencia entre la condición inicial exacta y la recuperada.

Estos problemas inversos son extremadamente mal planteados, por lo que al igual que en el caso de una dimensión, errores pequeños en el n -ésimo coeficiente de Fourier se amplifican enormemente, y debemos encontrar un equilibrio entre precisión y estabilidad. Por otro lado, el operador de difusión suaviza la solución del problema directo, por lo tanto, si la función φ no es suficientemente suave, entonces no podremos recuperar toda la información de esta. Además, en tiempos largos, las singularidades de la función φ se pierden, y resulta difícil o hasta imposible recuperar completamente la condición inicial, especialmente en regiones con gradientes altos o no continuos. Una de las razones por las que no podemos recuperar en forma razonable el dato inicial, cuando este tiene una forma compleja, es que al truncar la serie que representa el núcleo en el problema (4.20) no es posible recuperar las características fundamentales de φ con un número finito de funciones trigonométricas (tres en este caso).

Conclusiones

Los problemas inversos son de suma importancia en las aplicaciones, ya que la mayoría de estos problemas pueden ser escritos como ecuaciones operacionales de primer tipo, con operadores tanto lineales como no lineales (en este trabajo nos enfocamos únicamente en operadores lineales), por ejemplo la ecuación integral de Fredholm de primer tipo. En el presente trabajo se definió el mal planteamiento de estos problemas, y se construyeron ejemplos que ilustran el mal planteamiento de los problemas inversos.

Se formalizaron algunos conceptos básicos del mal planteamiento y regularización, mostramos un ejemplo en el cual ilustramos gráfica y analíticamente las dificultades que se presentan en la solución de los problemas inversos. Cabe resaltar, que normalmente el ruido en los datos es el que provoca el mal planteamiento en los problemas inversos, ya que un poco de “ruido” se amplifica tremendamente en el proceso de solución, debido a la división por los pequeños valores singulares del operador. El crecimiento de este error se controla mediante los métodos de regularización los cuales buscan restaurar la estabilidad de las soluciones con relación a los datos. El método de regularización más común es el de Tijonov, que al igual que otros métodos de regularización, depende de un parámetro de regularización α (escogido a priori para los ejemplos mostrados, aunque existen métodos adecuados para determinar el valor de dicho parámetro).

Hemos visto que el método de regularización de Tijonov tiene una representación variacional, por lo cual podemos utilizar métodos directos o iterativos para calcular la solución de un problema inverso, por ejemplo podemos resolver el problema inverso usando técnicas de mínimos cuadrados.

Así mismo, hemos revisado métodos de regularización iterativos los cuales tienen como parámetro de regularización el número de iteraciones k , donde a menor k mayor regularización y viceversa. No obstante, debemos tener cuidado de no utilizar un k demasiado pequeño, ya que podríamos estar lejos de la solución exacta.

Realizamos también una revisión de la teoría de Operadores Compactos, la inversa generalizada y métodos de optimización, ya que ésta teoría está íntimamente relacionada con la regularización de problemas mal planteados, y nos es de gran utilidad para resolver los problemas inversos abordados en el capítulo 4.

Por otro lado, resolvimos el problema del calor con retroceso en el tiempo en 1 y 2 dimensiones, usando el método de regularización de Tijonov en conjunto con el método de Gradiente Conjugado. Tales problemas son extremadamente mal planteados, ya que un pequeño error en la n -ésimo coeficiente de Fourier es amplificado tremendamente, de donde se concluye que: debemos encontrar un equilibrio entre precisión y estabilidad.

Hemos obtenido buenas aproximaciones de las soluciones considerando datos con y sin ruido para el problema en 1 dimensión, debido al truncamiento de la serie asociada al kernel del problema y la discretización de la integral por medio de la regla del punto medio, por lo cual es posible obtener una solución razonable sin realizar regularización cuando los datos no tienen ruido. No obstante, cuando los datos tienen un cierto nivel de ruido, a pesar del truncamiento y la discretización, el problema inverso sigue siendo muy mal condicionado y es necesario aplicar regularización para obtener resultados favorables. En el caso del problema de calor en 2 dimensiones, podemos asumir que los datos tienen ruido por la manera en que se obtuvieron. Además, las soluciones recuperadas difieren de la exacta en que su amplitud es menor y no se pueden recuperar las discontinuidades, ya que el problema directo suaviza la solución, y en tiempos largos, las singularidades de la condición inicial se pierden, debido a que ésta no es lo suficientemente suave.

Así, los resultados numéricos obtenidos permiten afirmar que el método utilizado es una técnica numérica eficiente para regularizar los problemas inversos mal planteados en este trabajo. Desde el punto de vista computacional la implementación es rápida y eficiente.

Resultados Importantes

Lema A.1.1. *Un subconjunto compacto M de un espacio métrico es cerrado y acotado.*

Teorema A.1.1. (Dimensión Finita). *Si un espacio normado \mathbf{X} tiene la propiedad que la bola unitaria cerrada es compacta, entonces $\dim(\mathbf{X}) < \infty$.*

Teorema A.1.2. (Compacidad). *Si $\dim(\mathbf{X}) < \infty$, $M \subset \mathbf{X}$ es compacto $\Leftrightarrow M$ es cerrado y acotado.*

Teorema A.1.3. (Rango y espacio nulo). *Sea T un operador lineal*

- a) El rango $\mathcal{R}(T)$ es un espacio vectorial.*
- b) Si $\dim(\mathcal{D}(T)) = n < \infty$, entonces $\dim(\mathcal{R}(T)) \leq n$.*
- c) El espacio nulo $\mathcal{N}(T)$ es un espacio vectorial.*

Teorema A.1.4. (Dimensión Finita). *Si \mathbf{X} es un espacio normado de dimensión finita, cualquier operador lineal sobre \mathbf{X} es acotado.*

Lema A.1.2. (Continuidad.) *Sean \mathbf{X} y \mathbf{Y} espacios normados. Entonces*

- a) Cualquier operador compacto $T : \mathbf{X} \rightarrow \mathbf{Y}$ es acotado (continuo).*
- b) Si $\dim(\mathbf{X}) = \infty$, el operador identidad $I : \mathbf{X} \rightarrow \mathbf{X}$ (el cual es continuo) no es compacto.*

Demostración.

a) $\|T\| = \sup_{\|x\|=1} \|Tx\|$. La bola unitaria $B = \{x \in \mathbf{X} : \|x\| = 1\}$ es acotada.

Como T es compacto, entonces $\overline{T(B)}$ es compacto sobre \mathbf{Y} (espacio métrico). Luego por el lema A.1.1 es acotado

$$\therefore \|T\| < \infty$$

b) Si el operador T fuera compacto, entonces la bola unitaria sería compacta y $\dim(\mathbf{X}) = \infty$. Esto contradice el Teorema A.1.1.

■

Teorema A.1.5. (Criterio de compacidad.) Sean \mathbf{X} y \mathbf{Y} espacios normados y $T : \mathbf{X} \rightarrow \mathbf{Y}$ un operador lineal. T es compacto \Leftrightarrow mapea cualquier sucesión acotada $\{x_n\} \in \mathbf{X}$ sobre una sucesión $\{Tx_n\} \subset \mathbf{Y}$ que tiene una subsucesión convergente.

Demostración.

\Rightarrow) Si T es compacto y $\{x_n\} \in \mathbf{X}$ acotada, entonces $\{\overline{Tx_n}\}$ es compacta, y por definición tiene una subsucesión convergente.

\Leftarrow) Sea $B \subset \mathbf{X}$ acotado, queremos demostrar que $\overline{T(B)}$ es compacto.

Sea $\{y_n\}$ en $T(B)$, $y_n = Tx_n$, $x_n \in B$. Como $\{Tx_n\}$ tiene una subsucesión convergente, entonces $\overline{T(B)}$ es compacto.

$$\therefore T \text{ es compacto.}$$

■

Teorema A.1.6. (Dominio ó rango de dimensión finita). Sean \mathbf{X} , \mathbf{Y} espacios normados y $T : \mathbf{X} \rightarrow \mathbf{Y}$ un operador lineal.

a) Si T es acotado y $\dim(T(\mathbf{X})) < \infty$, entonces T es compacto.

b) Si $\dim(\mathbf{Y}) < \infty$, el operador lineal $T : \mathbf{X} \rightarrow \mathbf{Y}$ es compacto.

Teorema A.1.7. Sea $\{T_n\}$ una sucesión de operadores lineales compactos de un espacio normado \mathbf{X} en un espacio de Banach \mathbf{Y} . Si $\|T_n - T\| \rightarrow 0$, entonces T es compacto.

Descomposición en Valores Singulares (SVD)

Sea A una matriz $m \times n$ con $m \geq n$, entonces existen matrices unitarias u ortonormal $U_{m \times m}$ y $V_{n \times n}$ tales que

$$U^T A V = \Sigma \tag{B.1}$$

donde Σ es una matriz diagonal de tamaño $m \times n$ cuyas entradas σ_i son los llamados valores singulares de la matriz A ordenados en orden decreciente, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$.

Los vectores u_i que forman las columnas de U son llamados vectores singulares de A por la izquierda, mientras los vectores v_i son llamados vectores singulares de A por la derecha.

Despejando A en la ecuación (B) obtenemos la factorización de A o SVD de A de tal forma que

$$A = U \Sigma V^T \tag{B.2}$$

la cual tiene la siguiente forma

$$\left[\begin{array}{c} \left[\begin{array}{c} A \end{array} \right]_{m \times n} = \left[\begin{array}{c|c|c|c} u_1 & u_2 & \dots & u_m \end{array} \right]_{m \times m} \left[\begin{array}{c} \sigma_1 \\ \sigma_2 \\ \dots \\ \sigma_n \end{array} \right]_{m \times n} \left[\begin{array}{c|c|c|c} v_1 & v_2 & \dots & v_n \end{array} \right]_{n \times n}^T$$

Teorema B.1.8. Toda matriz $A_{m \times n}$ admite una descomposición en valores singulares. Además, los valores singulares están determinados de forma única.

Proposición B.1.1. Los valores singulares cumplen con

i. Sea $r = \text{rango}(A)$, A tiene r valores singulares distintos de cero.

ii. Si $A \in \mathbb{C}^{n \times n}$ y sus valores singulares son $\sigma_1 \geq \dots \geq \sigma_n$

$$\Rightarrow |\det(A)| = \sigma_1 \cdot \dots \cdot \sigma_n$$

iii. Los valores singulares de $A \in \mathbb{C}^{m \times n}$ son las raíces cuadradas positivas de los valores propios de $A^T A$ y también de los de AA^T .

iv. Si $A \in \mathbb{C}^{m \times n}$ es invertible y $\sigma_1 \geq \dots \geq \sigma_n$ son sus valores singulares, los de A^{-1} serán:

$$\frac{1}{\sigma_n} \geq \dots \geq \frac{1}{\sigma_1}$$

v. Todo elemento de la diagonal σ_i cumple:

$$Av_j = \sigma_j u_j, \quad 1 \leq j \leq n.$$

vi. Si $A \in \mathbb{C}^{n \times n}$ (no necesariamente simétrica) su número de condición se define como

$$\text{cond}(A) = \frac{\sigma_1(A)}{\sigma_n(A)}. \quad (\text{B.3})$$

Esta matriz es llamada **mal-condicionada** si $\text{cond}(A)$ es grande, por ejemplo, 10^7 , 10^{16} , etc. Y decimos que, A es **bien-condicionada** si $\text{cond}(A)$ es pequeño, por ejemplo, 1, 10, 10^3 .

vii. Si $A \in \mathbb{C}^{m \times n}$ y $\sigma_1 \geq \dots \geq \sigma_r$ son sus valores singulares, entonces

$$\|A\|_F^2 = \sum_{i=1}^r \sigma_i^2 \quad (\text{B.4})$$

$$\|A\| = \sigma_1 \quad (\text{B.5})$$

Ejemplo B.1.3. Sea $A = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}$, encontrar la SVD de A .

$A^T A = \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix}$, los valores propios de $A^T A$ son $\lambda_1 = 9$ y $\lambda_2 = 4$.

Además notemos que $A^T A = V \Sigma^2 V^T$, por lo cual $\Sigma^2 = \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix} \Rightarrow \Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$

Ahora buscamos los vectores propios de $A^T A$ y los normalizamos para formar la matriz V . Para $\lambda_1 = 9$ tenemos que le corresponde el vector propio $v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ y a $\lambda_2 = 4$ le corresponde el vector propio $v_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$, de donde se obtiene $V = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

Por último, para obtener U se opera de forma análoga: $AA^T = (U\Sigma V^T)(V\Sigma U^T) = U\Sigma^2 U^T$,

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Por tanto $A = U\Sigma V^T$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Ejemplo B.1.4. Sea $A = \begin{bmatrix} 2 & 0 \\ 0 & -3 \\ 0 & 0 \end{bmatrix}$, encontrar la SVD de A .

$A^T A = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}$, los valores propios de $A^T A$ son $\lambda_1 = 9$ y $\lambda_2 = 4$.

$$\Rightarrow \Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}$$

Ahora los vectores propios de $A^T A$ normalizados correspondientes a λ_1 y λ_2 son $v_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ y $v_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, por tanto

$$V = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

análogamente para $A * A^T$

$$U = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

de esta manera

$$A = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Método Implícito

Vamos a introducir la idea del método implícito, considerando el problema de encontrar la función $u(x, t)$ la cual satisface la ecuación de calor homogénea, dada como

$$\begin{aligned} u_t &= a^2 u_{xx}, & 0 < x < \ell, & t > 0 \\ u(x, 0) &= \varphi(x), & 0 \leq x \leq \ell, & \text{Condición inicial} \\ u(0, t) &= u(\ell, t) = 0, & t \geq t_0 & \text{Condiciones de frontera} \end{aligned} \tag{C.1}$$

En este problema, asumimos que la función $\varphi(x)$ es conocida. Vamos a resolver el problema usando aproximación de las derivadas por medio de diferencias finitas de modo que este nos lleve a un esquema de tipo implícito, es decir, utilizaremos esquemas de diferencias finitas de tiempo hacia atrás. Para esto, hay que seguir los siguientes pasos:

1. Dividamos el intervalo $(0, \ell)$ en J subintervalos y el intervalo de tiempo $(0, T)$, donde T es el tiempo final, en N subintervalos

$$\begin{aligned} t_n &= n\Delta t, & n &= 0, 1, \dots, N. \\ x_j &= j\Delta x, & j &= 0, 1, \dots, J. \end{aligned}$$

donde $\Delta t = \frac{T}{N}$ y $\Delta x = \frac{\ell}{J}$. Los puntos tienen la forma (x_j, t_n) , y constituyen una malla donde deseamos encontrar los valores aproximados U_j^n de la solución exacta $u(x_j, t_n)$.

2. Evaluamos la ecuación diferencial en el punto $(x, t) = (x_j, t_n)$ y obtenemos:

$$u_{xx}(x_j, t_n) = a^2 u_t(x_j, t_n) \tag{C.2}$$

También necesitamos evaluar las condición inicial y las de frontera respecto a los puntos de la malla.

3. Vamos aproximar las derivadas en el punto (x_j, t_n) con diferencias finitas. Consideremos un esquema de diferencias implícito, es decir, utilizaremos una aproximación de diferencias finitas de tiempo hacia atrás, de esta forma obtenemos

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} = a^2 \left[\frac{U_{j+1}^{n+1} - 2U_j^{n+1} + U_{j-1}^{n+1}}{(\Delta x)^2} \right] \quad (\text{C.3})$$

$$U_j^0 = \varphi^0(x_j) \quad (\text{Condición inicial}) \quad (\text{C.4})$$

$$U_0^n = U_J^n = 0 \quad (\text{Condiciones de Frontera}) \quad (\text{C.5})$$

La molécula computacional es de la siguiente manera:

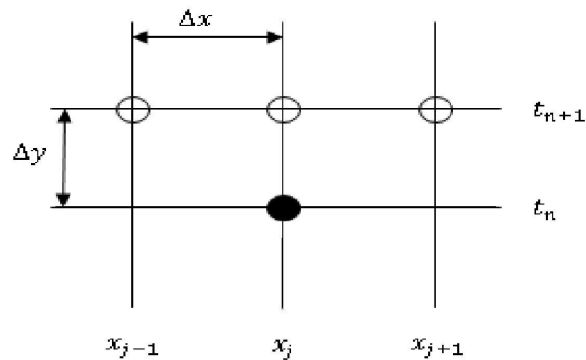


Figura C.1: Molécula del método implícito.

La ecuación (C.3) se puede reescribir de la siguiente manera:

$$-\nu U_{j-1}^{n+1} + (1 + 2\nu)U_j^{n+1} - \nu U_{j+1}^{n+1} = U_j^n \quad j = 1, \dots, J - 1, \quad (\text{C.6})$$

donde $\nu = a^2 \frac{\Delta t}{(\Delta x)^2}$.

Cabe resaltar que, el método implícito es **incondicionalmente estable**, ya que su factor de amplificación

$$\lambda(\kappa) = \frac{1}{1 + 4\nu \sin^2(\frac{1}{2}\kappa\Delta x)},$$

cumple con la condición de Von Neumann, es decir, $\lambda(\kappa)$ es menor que 1 en valor absoluto $\forall \kappa$ y para todo valor de ν . Por lo que, Δt y Δx se escogen tomando en cuenta criterios de precisión.

Este método, a diferencia de los métodos explícitos, tiene un costo computacional alto, ya que es necesario resolver en cada nivel de tiempo t_{n+1} un sistema de ecuaciones de tamaño $(J - 1) \times (J - 1)$ con una matriz constante

$$A = I + \nu \begin{pmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & 0 \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & 0 & & & & -1 \\ & & & & -1 & 2 \end{pmatrix} \quad (\text{C.7})$$

donde I es la matriz identidad de tamaño $(J-1) \times (J-1)$.

El costo computación de este método es muy alto, debido a la utilización de memoria (por la necesidad de almacenar la matriz), por el número de operaciones que hay que realizar, al tener que resolver un sistema de ecuaciones en cada paso de tiempo.

Notemos que la mayoría de los coeficientes de la matriz A del sistema son cero, por lo cual esta es una *matriz rala* (*sparse*). Además, la matriz A es una matriz *tri-diagonal* y diagonalmente dominante. Podemos aprovechar esta estructura para ahorrar memoria y resolver el sistema de manera muy eficiente. El método más eficiente en este caso es el **algoritmo de Thomas**. Este método consiste en reducir un sistema tri-diagonal a un *sistema bi-diagonal superior* por medio de la eliminación del término U_{j-1} de la ecuación (C.6), de tal manera que obtengamos

$$U_j - e_j U_{j+1} = f_j^n, \quad \text{con } e_j = \frac{\nu}{1 + 2\nu - \nu e_{j-1}}, \quad f_j^n = \frac{U_j^n + \nu f_{j-1}^n}{1 + 2\nu - \nu e_{j-1}}, \quad (\text{C.8})$$

para $j = 1, 2, \dots, J-1$ con $e_0 = f_0 = 0$. Este sistema bi-diagonal se puede resolver de manera fácil por medio de *sustitución hacia atrás*:

$$U_J = 0;$$

Para $j = J-1, \dots, 1$ hacer

$$U_j = f_j^n + e_j U_{j+1};$$

Terminar

Esta metodología es parte de los algoritmos que se conocen como resolvedores rápidos (fast solvers). La técnica de solución es un caso particular de una técnica denominada de reducción cíclica. Debido a que la matriz original es diagonalmente dominante cada coeficiente e_j en (C.8) es menor a uno en valor absoluto, lo cual garantiza que el uso de las relaciones de recurrencia como $U_j = f_j + e_j U_{j+1} = f_j$, para encontrar la solución, es numéricamente estable.

Métodos Implícitos de direcciones alternantes (ADI)

Consideremos la caja rectangular $D = [0, X] \times [0, Y]$ y denotemos por ∂D a su frontera. Sea $a > 0$ un coeficiente de difusión constante. El problema parabólico bidimensional (ecuación de calor), con condiciones de frontera Dirichlet, consiste en calcular $u = u(x, y, t)$ tal que

$$\begin{aligned} u_t &= a^2(u_{xx} + u_{yy}), & \text{en } D, t > 0, \\ u(x, y, 0) &= \varphi^0(x, y), & \text{en } D, \\ u(x, y, t) &= h(x, y, t) & \text{sobre } \partial D, t \leq 0 \end{aligned} \tag{D.1}$$

Construimos una malla uniforme sobre el dominio D de la siguiente manera: Se divide $[0, X]$ en I subintervalos, $[0, Y]$ en J subintervalos. Los puntos de la malla sobre D son

$$\begin{aligned} x_i &= i\Delta x, & i = 0, 1, \dots, I. \\ y_j &= j\Delta y, & j = 0, 1, \dots, J. \end{aligned}$$

donde $\Delta x = \frac{X}{I}$ y $\Delta y = \frac{Y}{J}$. Dividimos el intervalo de tiempo $[0, T]$ en N subintervalos, de tal manera que los niveles de tiempo discretos son $t_n = n\Delta t$, $n = 1, \dots, N$, con $\Delta t = \frac{T}{N}$.

Como los métodos implícitos son muy eficientes en una dimensión, ya que contamos con un resolvidor rápido y eficiente para sistemas tridiagonales, que es el algoritmo de Thomas. En el problema bidimensional, para cada n se debe resolver un sistema lineal de ecuaciones de tamaño $[(I - 1) \times (J - 1)] \times [(I - 1) \times (J - 1)]$. Por ejemplo, supongamos $\nu_x = \nu_y$, entonces la matriz del sistema es:

$$\mathcal{A} = I + \frac{1}{2}\nu \begin{pmatrix} A & -I & & & \\ -I & A & -I & & 0 \\ & -I & A & -I & \\ & & \ddots & \ddots & \ddots \\ 0 & & & & -I & \\ & & & & -I & A \end{pmatrix}, \quad (\text{D.2})$$

donde A es de la forma

$$A = \begin{pmatrix} 4 & -1 & & & \\ -1 & 4 & -1 & & 0 \\ & -1 & 4 & -1 & \\ & & \ddots & \ddots & \ddots \\ 0 & & & & -1 \\ & & & & -1 & 4 \end{pmatrix}$$

Obsérvese que la matriz \mathcal{A} viene dada por bloques, además no hay forma de permutar renglones y columnas, para poder resolver el sistema de forma eficiente. Por lo cual, en dos dimensiones es natural buscar métodos bidimensionales que sean implícitos en una dirección pero no en ambas. El siguiente esquema (**Paecman-Rachford(1955)**), es un ejemplo de este tipo de esquemas

$$\frac{U^{n+1/2} - U^n}{\Delta t/2} = a^2 \frac{\delta_x^2 U^{n+1/2}}{(\Delta x)^2} + a^2 \frac{\delta_y^2 U^n}{(\Delta y)^2} \quad \text{Implícito en } x, \text{ Explícito en } y. \quad (\text{D.3})$$

$$\frac{U^{n+1} - U^{n+1/2}}{\Delta t/2} = a^2 \frac{\delta_x^2 U^{n+1/2}}{(\Delta x)^2} + a^2 \frac{\delta_y^2 U^{n+1}}{(\Delta y)^2} \quad \text{Explícito en } x, \text{ Implícito en } y. \quad (\text{D.4})$$

donde

$$\begin{aligned} U &:= U_{i,j}, \\ \delta_x^2 U &:= U_{i-1,j} - 2U_{i,j} + U_{i+1,j}, \\ \delta_y^2 U &:= U_{i,j-1} - 2U_{i,j} + U_{i,j+1}. \end{aligned}$$

Sea $\nu_x = a^2 \frac{\Delta t}{(\Delta x)^2}$ y $\nu_y = a^2 \frac{\Delta t}{(\Delta y)^2}$. El esquema anterior se puede reescribir de la siguiente forma

$$\left(1 - \frac{1}{2}\nu_x \delta_x^2\right) U^{n+1/2} = \left(1 + \frac{1}{2}\nu_y \delta_y^2\right) U^n, \quad (\text{D.5})$$

$$\left(1 - \frac{1}{2}\nu_y \delta_y^2\right) U^{n+1} = \left(1 + \frac{1}{2}\nu_x \delta_x^2\right) U^{n+1/2}, \quad (\text{D.6})$$

el cual constituye un sistema de dos problemas implícitos unidimensionales. La solución de cada uno de estos sistemas involucra solo conjuntos de ecuaciones tridiagonales. El trabajo necesario, en un paso de tiempo, necesita de la solución de $J - 1$ sistemas tridiagonales de orden $J - 1$. En la siguiente figura se puede observar el orden en el que se pueden resolver estos sistemas tridiagonales. A éste tipo de esquemas se les conoce comúnmente

como **esquemas implícitos de direcciones alternantes** ó ADI (alternating direction implicit).

El método ADI es incondicionalmente estable, y su factor de amplificación es el siguiente:

$$\lambda(\kappa_1, \kappa_2) = \frac{(1 - 2\nu_x \sin^2(\frac{1}{2}\kappa_1 \Delta x))(1 - 2\nu_y \sin^2(\frac{1}{2}\kappa_2 \Delta y))}{(1 + 2\nu_x \sin^2(\frac{1}{2}\kappa_1 \Delta x))(1 + 2\nu_y \sin^2(\frac{1}{2}\kappa_2 \Delta y))}.$$

Además, este esquema implícito es de segundo orden respecto a Δt , Δx y Δy , ya que su error de truncamiento es $\mathcal{O}[(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2]$.

Códigos Computacionales

A continuación se muestran los códigos computacionales creados para los ejemplos vistos en el presente trabajo, estos fueron realizados en el ambiente de MATLAB® versión 7.0.

Ejemplo de la solución de un problema directo y Ejemplo 2.1.1, págs. 12 - 14.

problema_inverso.m

```

% Parámetros de un modelo de imagen óptica en 1-D.
gama = 0.05;
C = 1/(gama*sqrt(2*pi));

% Construcción del Kernel (Aplicando Cuadratura del punto medio).
x = 0: 0.018:1;
h=1/length(x);
for i = 1:length(x)
    for j = 1:length(x)
        K(i,j)= h*C*exp(-(((i-j)*h)^2)/(2*(gama)^2));
    end
end

% Solución exacta del sistema discreto Kf = d (función fuente).
f_ver = f(x);

% Datos (imagen borrosa) con y sin ruido
d = K*(f_ver'); % datos sin ruido
for i = 1:length(x) % construcción de los datos con ruido
    j = rand;
    if j < 0.5
        d1(i,1) = d(i,1) + rand*0.01;
    else
        d1(i,1) = d(i,1) - rand*0.01;
    end
end

end

% Gráfica el Problema directo
figure(1)
plot(x,f_ver','-r',x,d,'--g',x,d1,'ob');
grid on
axis([0 1 -0.2 1.2]);
xlabel('x')
title(['Problema directo, n = ',int2str(length(x))], 'Color', 'b')
legend('función fuente f exacta', 'd_{true} = K*f (datos sin ruido)',
       'd = d_{true} + eta (datos con ruido)')

% Resuelve el problema inverso usando SVD
[U,S,V] = svd(K);
f1 = V*(S^-1)*(U')*d1;

% Gráfica el Problema inverso sin regularización
figure(2)
plot(x,f_ver','-r',x,f1,'--g');
grid on
xlabel('x')
ylabel('f(x)')
title(['Problema inverso, n = ',int2str(length(x))], ['datos d con ruido']),
      'Color', 'b')
legend('función fuente f exacta', 'función fuente f aproximada')
    
```

f.m

```

function f_ver = f(x)
% Construcción de la función fuente f.
for i =1:length(x)
    if ( x(i) >= 0.1 & x(i) <= 0.25 )
        f_ver(i) = 0.75;
    elseif ( x(i) >= 0.3 & x(i) <= 0.32 )
        f_ver(i) = 0.25;
    elseif (x(i) >= 0.5 & x(i) <= 1)
        f_ver(i) = (sin(2*pi*x(i)))^4;
    else
        f_ver(i) = 0;
    end
end
    
```

Ejemplo 2.2.1, págs. 16 - 17.

svd_K.m

```

% Construcción del Kernel (Aplicando Cuadratura del punto medio).
gama = 0.05;
C = 1/(gama*sqrt(2*pi));
n = input('\n Número de datos \n');
x = 0: 1/n:1;
h=1/length(x);
for i = 1:length(x)
    for j= 1:length(x)
        K(i,j)= h*C*exp(-(((i-j)*h)^2)/(2*(gama)^2));
    end
end

% Descomposición en valores singulares (SVD) de la matriz K.
[U,S,V] = svd(K);

% Gráfica los valores singulares de K
i = 1:n;
figure(1)
semilogy(i,S(i,i),'*-r');
xlabel('índice i')
ylabel('i-ésimo valor singular')
title(['Valores singulares de K'],'Color','b')

% Gráfica vectores singulares v_{i}.
j = input('\n i-ésimo vector singular que desea gráfica \n');
figure(2)
plot(x,U(:,j),'-r');
grid on
xlabel('x')
ylabel({'v_{',num2str(j),'}'})
title(['Vector singular v_{',int2str(j),'}'],'Color','b')

```

Ejemplo de regularización TSVD y de regularización de Tijonov, págs. 17 - 22.

problema_inverso_regularizado.m

```

% Regulariza el problema inverso por medio de Tijonov o TSVD.
% Datos K, d, f, x, [U,S,V] = svd(K)
metodo = input('\n Qué método de regularización desea usar,
presione 1 para "TSVD" o 2 para "Tijonov" \n');
alfa=input('\n Ingrese el valor umbral
(parámetro de regularización).\n');
f1 = zeros(length(x),1);
%Método de Regularización
for i = 1:length(x)
    w = regularizacion(S(i,i), alfa, metodo);
    f1 = (w*(1/S(i,i))*U(:,i)*d1)*V(:,i) + f1;
end

% Gráfica la solución regularizada.
plot(x,f_ver,'-r',x,f1,'-b');
grid on
axis([0 1 -0.2 1.2]);
%axis(axis);
xlabel('x')
ylabel('f(x)')
title(['Problema inverso regularizado, n = ',int2str(length(x)), ',
alfa = ', num2str(alfa)],'Color','b')
legend('función fuente f exacta', 'función fuente f aproximada')

```

errores.m

```

% Regulariza el problema inverso.
metodo = input('\n Qué método de regularización desea usar,
presione 1 para "TSVD" o 2 para "Tijonov" \n');
alfa = 0.0000001:0.0000025:0.1;

% Obtiene la norma del error.
for j = 1:length(alfa)
    f1 = zeros(length(x),1);
    for i = 1:length(x)
        w = regularizacion(S(i,i), alfa(j), metodo);
        f1 = (w*(1/S(i,i))*U(:,i)*d1)*V(:,i) + f1;
    end

    err(j) = norm(f1-f_ver');
end

% Gráfica la Norma de la solución del Error
% del método de Regularización.
figure(2)
semilogx(alfa,err)
axis(axis);
xlabel('alfa')
ylabel('||error||')
title(['Norma de la Solución del Error'],'Color','b')

```

Ejemplo 2.2.2, págs. 26 - 27.

error_ruido.m

```

% Datos K, d, d1, f, x, [U,S,V] = svd(K)
% d := datos sin ruido, d1 := datos con ruido
eta = d-d1;
delta = norm(eta); % nivel del ruido
p = input('\n Cuanto vale p, 0 < p < 2 \n');
alfa = (delta)^p; % parámetro de regularización

%Regularización del problema inverso.
metodo = input('\n Qué método de regularización desea usar,
presione 1 para "TSVD" o 2 para "Tijonov" \n');
f1 = zeros(length(x),1);
e_ruido = zeros(length(x),1);
for i=1:length(x)
    w = regularizacion(S(i,i), alfa, metodo);
    e_ruido = (w*(1/S(i,i))*U(:,i)*eta)*V(:,i) + e_ruido;
    f1 = (w*(1/S(i,i))*U(:,i)*d1)*V(:,i) + f1;
end
norma_e_ruido = norm(e_ruido)

% Gráfico de la solución regularizada.
plot(x,f_ver,'-r',x,f1,'-b');
grid on
axis([0 1 -0.2 1.2]);
xlabel('x')
ylabel('f(x)')
title(['Problema inverso regularizado,'],['n = ',int2str(length(x)),
', alfa = ', num2str(alfa)],'Color','b')
legend('función fuente f exacta', 'función fuente f aproximada')

```

error_ruido_p.m

```
% Datos K, d, d1, f, x, [U,S,V] = svd(K)
% d := datos sin ruido, d1 := datos con ruido
eta = d-d1;
delta = norm(eta); %nivel del ruido
p = 0.0001:0.0001:1.9999;

% Regulariza el problema inverso.
metodo = input('\n Qué método de regularización desea usar,
presione 1 para "TSVD" o 2 para "Tijonov".\n');
for j = 1:length(p)
    alfa(j) = (delta)^p(j);
    alfa_p(j) = alfa(j)^((2-p(j))/2*p(j));
    e_ruido = zeros(length(x),1);
    for i=1:length(x)
        w = regularizacion(S(i,i), alfa(j), metodo);
        e_ruido = (w*(1/S(i,i))*(U(:,i)'*eta)*V(:,i)) + e_ruido;
    end
    norm_e_ruido(j) = norm(e_ruido); %norma del ruido
end

%Gráfica la norma del ruido vs alpha^((2-p)/p).
plot(norm_e_ruido,alfa_p,'-r');
grid on
axis([0 0.18 0 1.1]);
axis(axis);
xlabel('||e_{alpha}^{ruido}||');
ylabel('alfa^{(2-p)/(2p)}');
title(['||e_{alpha}^{ruido}|| vs alfa^{(2-p) / 2p}'], ['n = ',
int2str(length(x))', ' 0 < p < 2, Regularización.'], 'Color', 'b')
```

Ejemplo 2.2.3, págs. 35 - 38.

regularizacion_descenso_maximo.m

```
% Regularización iterativa.
% Datos K, d1, f, x
% d1:= datos con ruido.
tao = input('\n ¿Qué valor desea darle a tao, tal que
0 < tao < 1/ ||K||^2\n?');
iter = input('\n ¿Cuántas iteraciones desea? \n');
fo = zeros(length(x),1); % solución inicial.

% Resuelve por medio de Descenso Máximo.
f1 = desc_max(K,d1,fo,tao,iter);

% Gráfico de la solución regularizada
plot(x,f_ver,'-r',x,f1,'-b');
grid on
axis([0 1 -0.2 1.2]);
xlabel('x')
ylabel('f_k(x)')
title(['Problema inverso regularizado (Descenso Máximo), k = ',
int2str(iter), ', tao = ', num2str(tao)], 'Color', 'b')
legend('función fuente f exacta', 'función fuente f aproximada')
```

desc_max.m

```
function f = desc_max(K,d,fo,tao,iter)
% Método de Descenso Máximo
% gradJ := J', K := Kernel
% fo := valor inicial de f
% iter := numero de iteraciones
% tao -> f_{n+1} = f_n - tao*J'(f_n)

f = fo;
for i = 1:iter
    gradJ = (K'*K*f - K'*d);
    f = f - tao*gradJ;
end
```

errorDM.m

```
% Regularización iterativa.
% (Descenso Máximo)
% Datos K, d1, f, x
% d1 := datos con ruido.
tao = input('\n ¿Qué valor desea darle a tao, tal que
0 < tao < 2/ ||K||^2\n?');
iter = 1:20:10001;

for j = 1:length(iter)
    fo = zeros(length(x),1);
    f1 = desc_max(K,d1,fo,tao,iter(j));

    err(j) = norm(f1-f_ver,inf); % norma del error
end

% Gráfico de la norma del error vs número de iteraciones.
semilogx(iter,err,'*-r');
axis(axis);
xlabel('Iteración k')
ylabel('||error||')
title(['Norma de la Solución del Error, Descenso Máximo,
tau = ', num2str(tao) ], 'Color', 'b')
```

Ejemplo 3.4.1, pág. 66.

regularizacion_Gradiente_Conjugado.m

```
% Regularización por Gradiente Conjugado
% Datos K, d1, f, x
% d1 := datos con ruido.
fo = zeros(length(x),1); %punto inicial
iter = input('\n Número de iteraciones \n');

% Método de GC
alfa = input('\n Qué valor desea darle a alpha \n');
K1 = K'*K + alfa*eye(length(x),length(x));
d11 = K'*d1;
f1 = grad_conj2(K1,d11,fo,iter);

% Gráfico de la solución regularizada
plot(x,f_ver,'-r',x,f1,'-b');
grid on
axis([0 1 -0.2 1.2]);
xlabel('x')
ylabel('f_k(x)')
title(['Problema inverso regularizado (Gradiente Conjugado), k = ',
int2str(iter)], 'Color', 'b')
legend('función fuente f exacta', 'función fuente f aproximada')
```

grad_conj2.m

```
function f = grad_conj2(K,d,fo,ni)
% ni = número de iteraciones
go = K*fo - d; % gradiente inicial.
dir_o = -go; % dirección de búsqueda inicial.
for k=1:ni
    r = K*dir_o; % Vector auxiliar
    s = dir_o'*r; % Escalar auxiliar
    a = (go'*dir_o) / s;% parámetro de búsqueda lineal
    f = fo - a*dir_o; % actualización de la solución aproximada
    g = go - a*r; % actualización del gradiente
    b = g'*r/s; % Calculo de beta_k
    dir = g - b*dir_o; % actualización de la dirección conjugada
    go = g;
    fo = f;
    dir_o = dir;
end
```

error_GC.m

```
% Regularización iterativa
% (Gradiente Conjugado)
% Datos K, di, f, x
% di := datos con ruido.
alfa = input('\n Qué valor desea darle a alpha \n');
K1 = K'*K + alfa*eye(length(x),length(x));
d11 = K'*d1;
iter = 1:1:10;
for j = 1:length(iter)
    fo = zeros(length(x),1);
    f1 = grad_conj2(K1,d11,fo,iter(j));
    err(j) = norm(f1-f_ver',inf);
end
figure(1)
semilogx(iter,err,'.-r')
axis(axis);
xlabel('Iteración k')
ylabel('||error||')
title(['Norma de la Solución del Error, CG' ],'Color','b')
```

Ejemplo 3.4.2, págs. 67-68.

regularizacion_Precondicionador_Gradiente_Conjugado.m

```
% Regularización por Gradiente Conjugado con Precondicionador (PGC) % Método de PGC
% Datos K, di, f, x
% di := datos con ruido.
fo = zeros(length(x),1); % punto inicial
iter = input('\n Número de iteraciones \n');
K1 = K'*K + alfa*eye(length(x),length(x));
d11 = K'*d1;
M = matriz_precond(K1);
f1 = pre_grad_conj(K1,d11,fo,M,iter);
% Gráfico de la solución regularizada
plot(x,f_ver', '-r',x,f1,'-b');
grid on
axis([0 1 -0.2 1.2]);
xlabel('x')
ylabel('f_k(x)')
title(['Problema inverso regularizado (PGC), k = ',
int2str(iter)], 'Color','b')
legend('función fuente f exacta', 'función fuente f aproximada')
```

pre_grad_conj.m

```
function f = pre_grad_conj(K,d,fo,M,iter)
go = K*fo - d; % gradiente inicial.
for q = 1:length(fo)
    ro(q,1) = go(q)/M(q,q);
end
dir_o = ro; %% dirección de búsqueda inicial.
dot0 = go'*ro;
for i=1:iter
    w = K*dir_o; % Vector auxiliar
    a = (dot0) / (dir_o'*w); % parámetro de búsqueda lineal
    f = fo - a*dir_o; % actualización de la solución aproximada
    g = go - a*w; % actualización del gradiente
    for q = 1:length(fo)
        r(q,1) = g(q)/M(q,q);
    end
    dot = g'*r;
    b = dot/dot0; % Calculo de beta_k
    dir = r + b*dir_o; % actualización de la dirección conjugada
    go = g;
    fo = f;
    r0 = r;
    dir_o = dir;
    dot0 = dot;
end
```

error_PGC.m

```
% Regularización iterativa
% Gradiente Conjugado con Precondicionador (PGC)
% Datos K, di, f, x
% di:= datos con ruido.
alfa = input('\n Qué valor desea darle a alpha \n');
K1 = K'*K + alfa*eye(length(x),length(x));
d11 = K'*d1;
iter = 1:1:10;
% Matriz preconditionadora
M = matriz_precond(K1);
for j = 1:length(iter)
    fo = zeros(length(x),1);
    f1 = pre_grad_conj(K1,d11,fo,M,iter(j));
    err(j) = norm(f1-f_ver',inf);
end
figure(2)
semilogx(iter,err,'.-r')
axis(axis);
xlabel('Iteración k')
ylabel('||error||')
title(['Norma de la Solución del Error, PGC' ],'Color','b')
```


Ejemplo 4.2.1, págs. 75-80.

GC_EDP_cuadratura.m

```

n_f = input('\n Número de términos para la serie de Fourier \n');
T = input('\n Valor de la temperatura final T que desea. \n');
l = input('\n Valor final del intervalo [0, l] \n');
alfa = input('\n Que valor alfa desea para la regularización de
Tijonov (K^{T}*K + a*I)f = K^{T}*g \n');
q = input('\n Número de subdivisiones en el intervalo [0,l] \n');

% Cuadratura del punto medio
h= l/q;
for i = 1:q
    x(i) = (i - (1/2))*h;
end

%Construcción del Kernel
a = 1; % coeficiente de difusión
for i = 1:q
    for j = 1: q
        K(i,j) = 0;
        for n = 1:n_f
            K(i,j) =
(sin(n*pi*(i-(1/2))/q)*sin(n*pi*(j-(1/2))/q)*exp(-T*((n*pi*a)/l)^(2))
            end
            K(i,j) = (2/q)*K(i,j);
        end
    end
end

% Valor conocido u(x,T) = g(x)
% (La función se puede construir con ó sin ruido)
g = g1_edp(x,n_f,T);

% Condición inicial analítica
% phi(x) = u(x,0)
f_ver = phi_cond_i(x);

% Solución sin regularización de Tijonov.
fi_o = zeros(q,1); %punto inicial
iter = input('\n Número de iteraciones. \n');
f1 = grad_conj2(K,g,fi_o,iter);

% Gráfico de la solución sin regularización de Tijonov
figure(2)
plot(x,f_ver,'-r',x,f1,'b');
grid on
xlabel('x')
ylabel('fi_k(x)')
title(['Problema inverso u_t = u_xx (GC)'], ['Solución sin
regularización y con ruido en los datos.'], 'Color','b')
legend('función fuente fi exacta','función fuente fi aproximada')

% Solución con regularización de Tijonov
tol = %input('\n Que tolerancia desea? \n');
T_K = K'*K + alfa*eye(q,q);
T_g = K'*g;
[norma,iter,F] = grad_conj(T_K,T_g,fi_o,tol);

[n1,k1] = size(F);
norma = norm(f_ver-F(n1,:));
err_rel = norma/(norm(f_ver)) % error relativo de la solución aprox.

% Gráfico de la solución sin regularización de Tijonov
figure(3)
plot(x,f_ver,'-r',x,F,'-.',x,F(n1,:), 'b');
grid on
axis([0 1 -0.1 0.6]);
xlabel('x')
ylabel('fi_k(x)')
title(['Problema inverso u_t = u_xx (GC)'], ['e.r.= ',
num2str(err_rel), ', k = ', int2str(iter), ', alpha=',
num2str(alfa), ', T=', num2str(T)]], 'Color','b')
legend('función fuente fi exacta','función fuente fi aproximada')

```

g1_edp.m

```

function g = g1_edp(x,n_f,T)
for i = 1: length(x)
    g(i,1) = 0;
    for n = 1:n_f
        g(i,1)=(sin(n*pi/2)/(n^2))*exp(-T*((n*pi)^2))*sin(n*pi*x(i))
            + g(i,1);
    end
    g(i,1) = (4/(pi^2))*g(i) + rand*0.00000115; % g(x) con ruido
end

% Gráfico de u(x,t) = g(x)
figure(1)
plot(x,g);
grid on
xlabel('x')
ylabel('g(x)=u(x,t)')
title(['Solución de u_t = u_xx (GC) al tiempo T,'],
['u(x,T) = g(x), n = ',int2str(length(x))], 'Color','r')

```

phi_cond_i.m

```

function f_ver = phi_cond_i(x)
for i = 1:length(x)
    if x(i)<= 0.5
        f_ver(i,1) = x(i);
    elseif x(i)>0.5
        f_ver(i,1) = 1-x(i);
    end
end

```

grad_conj.m

```

function [norma,iter,F] = grad_conj(K,d,fo,tol)
F(1,:) = fo;
go = K*fo - d; %% gradiente inicial.
dir_o = -go; %% dirección de búsqueda inicial.
iter=0;
norma=norm(go);

while (norm(go) > tol)
    r = K*dir_o; % Vector auxiliar
    s = dir_o'*r; % Escalar auxiliar
    a = (go'*dir_o) / s; % parámetro de búsqueda lineal

    f = fo - a*dir_o; % actualización de la solución aproximada
    g = go - a*r; % actualización del gradiente
    b = g'*r/s; % Calculo de beta_k
    dir = g - b*dir_o; % actualización de la dirección conjugada
    go = g;
    fo = f;
    dir_o = dir;
    iter = iter+1;
    F(iter+1,:)= fo;
end
norma=norm(g);

```

Ejemplo 4.3.1, págs. 83-87, y Ejemplo 4.3.2, págs. 87-91

GC_EDP_cuadratura.m

```

%% % Parámetros de la doble suma (series de Fourier) %%
m_f = input('\n Número de términos para la serie de Fourier
            (parámetro x) \n');
n_f = input('\n Número de términos para la serie de Fourier
            (parámetro y) \n');

%% % Temperatura final %%
T = input('\n Que valor de la temperatura final T desea. \n');

%% % Longitud del cuadrado [0,L_1]x[0,L_2], 0<=x<=L_1, 0<=y<=L_2.
L_1 = input('\n Valor final del intervalo [0, L_1], 0<=x<=L_1. \n');
L_2 = input('\n Valor final del intervalo [0, L_2], 0<=y<=L_2. \n'); tol = input('\n Qué tolerancia desea? \n');

%% % Partición de los intervalos [0,L_1], [0,L_2].
M = 40; % subdivisiones en el intervalo [0,L_1]
N = 40; % subdivisiones en el intervalo [0,L_2]

%% % Parámetros de discretización (cuadratura del punto medio)
h = L_1/M;
k = L_2/N;
for i = 1:M
    x(i) = (i - (1/2))*h;
end
for j = 1:N
    y(j) = (j - (1/2))*k;
end

%% % Construcción del operador K, donde K*phi=g(x,y,T)
a = 1;% coeficiente de difusión
for i = 1:M;
    for j = 1:N;
        for l = 1:M;
            for p = 1:N;
                r = (j-1)*M + i ;
                s = (p-1)*M + l ;
                K(r,s) = 0;
                for m = 1:m_f
                    for n = 1:n_f
                        K(r,s) =
                            sin(m*pi*(1-(1/2))/M)*sin(n*pi*(p-(1/2))/N)
                            *sin(m*pi*(i-(1/2))/M)*sin(n*pi*(j-(1/2))/N)
                            *exp(-(a^2)*((m*pi/L_1)^2 + (n*pi/L_2)^2)*T)
                            + K(r,s);
                    end
                end
                K(r,s) = (4/(M*N))*K(r,s);
            end
        end
    end
end

%% % Temperatura final u(x,y,T) = g(x,y)
%% % CASO M = N
%% % Resuelve para la condición inicial "peaks" o la piramide.
figure(1);
[g,fi_ver_cuad] = picos_piramide(160,160,L_1,L_2,T);

%% Malla gruesa
g_int = interpola(g);
%% Interpolación lineal
for i = 1:M
    for j = 1:N
        g_cuad(i,j) = (g_int(i,j) + g_int(i+1,j) + g_int(i,j+1)
                    + g_int(i+1,j+1))/4;
    end
end

disp('Presione enter para continuar.')
```

```

pause
for i = 1:M
    for j = 1:N
        r = (j-1)*M + i;
        g_K(r,1) = g_cuad(i,j);
    end
end

%% % Solución del sistema K*fi = g_K
fi_o = ones(M*N,1); %%punto inicial
%% Regularización de Tjonov
alfa = input('\n Que valor alfa desea usar para la regularización de
            Tjonov (K^{-T}*K + a*I)f = K^{-T}*g \n');
% Resuelve (K'*K + alfa*I)*fi = K'*g_k
T_K = K'*K + alfa*eye(M*N,M*N);
T_g = K'*g_K;
% Resuelve el sistema (K'*K + alfa*I)*fi = K'*g_k usando GC
figure(3);
iter = input('\n Número de iteraciones.\n'); % 25 para pirámide,
                                           % 40 para picos.
[e_r_,fhi] = grad_conj_bid(x,y,T_K,T_g,fi_o,iter,alfa,T,fi_ver_cuad);
disp('Presione enter para continuar.')
```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Grafica de la diferencia puntual entre
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% la condición inicial exacta (con ruido)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% y la aproximada por regularización.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```

dif_puntual = abs(fhi - fi_ver_cuad);
figure(4);
surf(x,y,dif_puntual);
%axis([0 1 0 1 0 6]);
xlabel('x')
ylabel('y')
zlabel('|fhi - fi_{ver}|')
title(['Diferencia puntual entre'],'[la condición inicial exacta
      (con ruido)'],'[y la recuperada por regularización.'],'
      'Color','b')
```

```

disp('Presione enter para continuar.')
```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Curvas de nivel
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% la condición inicial exacta (con ruido)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% y la aproximada por regularización.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```

figure(5);
contour(x,y,fi_ver_cuad,50);
xlabel('x'); ylabel('y')
title(['Curvas de Nivel de la condición inicial exacta (con
      ruido)'],'Color','r')
```

```

figure(6);
contour(x,y,fhi,50);
xlabel('x'); ylabel('y')
title(['Curvas de Nivel de la condición inicial recuperada por
      regularización'],'Color','r')
```

```

figure(7);
contour(x,y,g_cuad,50);
xlabel('x'); ylabel('y')
title(['Curvas de nivel de g(x,y)'],'Color','r')
```

```

disp('Termina simulación.')
```

picos_piramide.m

```

function [u,fi_ver_cuad] =picos\_piramide(I,J,L_1,L_2,T)

% Método implícito de direcciones alternantes en 2-D
% Condiciones de frontera Dirichlet homogéneas

dt = 1e-4           % Paso del Tiempo
N_iter= T/dt;       % Número de iteraciones
dx = L_1/I;         % delta x
dy = L_2/J;         % delta y
nux = dt/dx^2;
nuy = dt/dy^2;

% Construcción de la malla bidimensional
x = (0:dx:L_1)';
y = (0:dy:L_2)';

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% CASO 1: Una condición inicial suave con valores positivos y
% negativos.
% En este caso deberá escogerse I = J
% u = abs(peaks(I+1));
% fi_ver = u ;

% CASO 2: Una condición inicial no suave para mostrar el
% principio del máximo
% I y J pueden escogerse en forma arbitraria
u = piramide(x,y,L_1,L_2);
fi_ver = u ;

% Gráfica de la condición inicial
figure(1);
surf(x,y,u)
xlabel('X')
ylabel('Y')
disp('Presione enter para continuar.')
```

```

pause

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% EMPIEZA CONDICIÓN INICIAL %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MALLA GRUESA %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Gráfico de la condición inicial en malla gruesa
% y puntos de la cuadratura.

% Cuadratura del punto medio
h= L_1/40;
k= L_2/40;
for i = 1:40
x1(i) = (i - (1/2))*h;
end
for j = 1:40
y1(j) = (j - (1/2))*k;
end

fi_ver_int = interpola(fi_ver);
fi_ver_cuad = zeros(40,40);
for i = 1:40
for j = 1:40
fi_ver_cuad(i,j) = (fi_ver_int(i,j) + fi_ver_int(i+1,j) +
fi_ver_int(i,j+1) + fi_ver_int(i+1,j+1))/4;
end
end

figure(1)
surf(x1,y1,fi_ver_cuad);
xlabel('x')
ylabel('y')
zlabel('fhi_{exact}(x,y)=u(x,y,0)')
title(['Condición Inicial Exacta'], 'Color','r')
disp('Presione enter para continuar.')
```

```

pause

u_cuad = zeros(40,40);
t_1 = 0;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% TERMINA CONDICIÓN INICIAL %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MALLA GRUESA %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Calculo de los coeficientes e(i) en el algoritmo de Thomas.
% ex(i) son los coeficientes para los sistemas tridiagonales en la
% dirección x,
% ey(j) son los coeficientes para los sistemas tridiagonales en la
% dirección y
nuxm = 0.5*nux;
ex(1) = 0;

for i = 2:I
ex(i) = nuxm/(1+nuxm*(2-ex(i-1)));
end

nuy = 0.5*nuy;
ey(1) = 0;
for j = 2:J
ey(j) = nuy/(1+nuy*(2-ey(j-1)));
end

% Arreglo auxiliar para almacenar cálculos intermedios
unew = zeros(I+1,J+1);
% figure(2);

% ITERACIONES EN EL TIEMPO
for n = 1:N_iter
% Se comienza resolviendo los SISTEMAS TRIDIAGONALES EN LA DIRECCIÓN X
for j = 2:J % Se resuelve un sistema tridiagonal para cada j
fx(1) = 0;
for i = 2:I % Algoritmo de Thomas
d = (1-nuy)*u(i,j) + 0.5*nuy*(u(i,j-1) + u(i,j+1));
fx(i) = (d + nuxm*fx(i-1))/(1+nuxm*(2-ex(i-1)));
end
unew(I+1,j) = 0; % Condición de frontera homogénea en x = L_1
for i = I:-1:2
unew(i,j) = fx(i)+ex(i)*unew(i+1,j);
end
unew(1,j) = 0; % Condición de frontera homogénea en x = 0
end
u = unew; % Actualización para ir a la otra dirección

% Se continua para resolver los
% SISTEMAS TRIDIAGONALES EN LA DIRECCIÓN Y
for i = 2:I % Se resuelve un sistema tridiagonal para cada i
fy(1) = 0;
for j = 2:J % Algoritmo de Thomas
d = (1-nux)*u(i,j) + 0.5*nux*( u(i-1,j)+u(i+1,j) );
fy(j) = (d + nuy*fy(j-1))/(1+nuy*(2-ey(j-1)));
end
unew(i,J+1) = 0; % Condición de frontera homogénea en y = L_2
for j = J:-1:2
unew(i,j) = fy(j)+ey(j)*unew(i,j+1);
end
unew(i,1) = 0; % Condición de frontera homogénea en y = 0
end
u = unew; % Actualización para avanzar en las iteraciones en el
% tiempo

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ANIMACIÓN MALLA GRUESA g(x,y,T) %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

u_int = interpola(u);
for i = 1:40
for j = 1:40
u_cuad(i,j) = (u_int(i,j) + u_int(i+1,j) + u_int(i,j+1) +
u_int(i+1,j+1))/4;
end
end

figure(2);
surf(x1,y1,u_cuad);
xlabel('x')
ylabel('y')
zlabel('g(x,y)=u(x,y,T)')
t_1 = t_1 + dt;
title(['Solución de u_t = u_{xx} (GC) al tiempo T =',
num2str(t_1), ',',[u(x,y,T) = g(x,y)']], 'Color','r')
M(n) = getframe;
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function u = condin(I,J)
u = zeros(I+1,J+1);
i1 = I/4;
j1 = J/4;
i = i1-2:i1+3;
j = j1-2:j1+3;
u(i,j) = 1;
i = 3*i1-2:3*i1+3;
j = 3*j1-2:3*j1+3;
u(i,j) = 1;
return

```

interpola.m

```
function u = interpola(v)
k = 0;
h = 1/40;
u = zeros(41,41);
for i = 1:4:161
    k = k + 1;
    l = 0;
    for j = 1:4:161
        l = l + 1;
        u(k,l) = v(i,j);
    end
end
x = 0:h:1; y = x;
```

grad_conj_bid.m

```
function [err_rel,fi] = grad_conj_bid(x,y,K,d,fo,n_it,alfa,T,fi_ver)
fi=zeros(length(x),length(y));
go = K*fo - d; %% gradiente inicial.
dir_o = -go; %% dirección de búsqueda inicial.
iter=0;
norma=1;

for k=1:n_it
    r = K*dir_o; %% Vector auxiliar
    s = dir_o'*r; %% Escalar auxiliar
    a = (go'*dir_o) / s; %% parámetro de búsqueda lineal
    f = fo - a*dir_o; %% actualización de la solución aproximada
    g = go - a*r; %% actualización del gradiente
    b = g'*r/s; %% Calculo de beta_k
    dir = g - b*dir_o; %% actualización de la dirección conjugada
    go = g;
    fo = f;
    dir_o = dir;
    iter = iter+1;
    norma = norm(go);
end

%% ANIMACIÓN
for l = 1:length(x)
    for p = 1:length(y)
        s = (p-1)*length(x) + 1;
        fi(l,p) = f(s);
    end
end
norma = norm(fi_ver - fi);
err_rel = norma/(norm(fi_ver));
surf(x,y,fi);
xlabel('x')
ylabel('y')
zlabel('phi(x,y)=u(x,y,T)')
title(['Problema Inverso u_t = u_{xx} (GC)', '],[^e.r.=',
    num2str(err_rel), ', k = ',int2str(iter), ', alpha=',
    num2str(alfa), ', T=', num2str(T)]), 'Color','b');
M(k) = getframe;
pause(.5);
end
norma=norm(g);
```

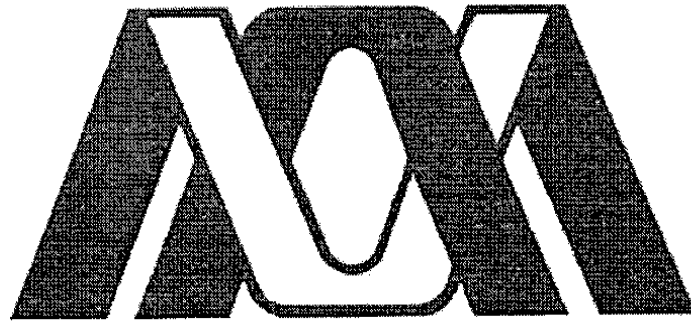
piramide.m

```
function z = piramide(x,y,L_1,L_2)
m = length(x);
n = length(y);
z=zeros(m,n);
for j = 1:(n+1)/2
    for i = 1:(m+1)/2
        if (x(i)>=y(j)) & ((1-x(i)) >= y(j))
            z(j,i) = 2*y(j);
            z(j,m+1-i) = z(j,i);
            z(n+1-j,i) = z(j,i);
            z(n+1-j,m+1-i) = z(j,i);
        elseif (x(i)>=y(j)) & ((1-x(i)) <= y(j))
            z(j,i) = 2*(1-x(i));
            z(n+1-j,i) = z(j,i);
            z(j,m+1-i) = z(j,i);
            z(n+1-j,m+1-i) = z(j,i);
        elseif (x(i)<=y(j)) & ((1-x(i)) <= y(j))
            z(n+1-j,m+1-i) = z(j,i);
            z(j,i) = 2*(1-y(j));
            z(n+1-j,i) = z(j,i);
            z(j,m+1-i) = z(j,i);
            z(n+1-j,m+1-i) = z(j,i);
            z(n+1-j,i) = z(j,i);
            z(n+1-j,m+1-i) = z(j,i);
        end
    end
end
```

Bibliografía

- [1] A. B. Bakushinskii, *Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion*, Comput. Math. Math. Phys. 32 (1992), 1353-1359.
- [2] A.M. Denisov. Inverse and Ill-Posed Problems Series. *Elements of the Theory of Inverse Problems*, VSP, 1999.
- [3] Heinz W. Engl and C. Groetsch. *Inverse and Ill-posed Problems*, Academic Press, 1987.
- [4] Heinz W. Engl. *Inverse Problems*, Sociedad Matematica Mexicana, 1995.
- [5] Heinz W. Engl, *Necessary and sufficient conditions for convergence of regularization methods for solving linear operator equations of the first kind*, Numer. Funct. Anal. Optim. 3 (1981), 201-222.
- [6] Heinz W. Engl, Martin Hanke, Andreas Neubauer. *Regularization of Inverse Problems*, Kluwer Academic Publishers, 1996.
- [7] C.W. Groetsch. *Generalized Inverses of Lineal Operators: Representación and Approximation*, Dekker, 1977.
- [8] C.W. Groetsch. *Inverse Problems Activities for Undergraduates*, The Mathematical Association of Americana, 1999.
- [9] C.W. Groetsch. *The theory of Tikhovov Regularization for Fredholm Equatinons of the First Kind*, Pitman, 1984.
- [10] M. Hanke and P-C Hansen. *Regularization methods for large scale inverse problems*, Surveys of Mathematics in Industry 3, p.p. 253-315, 1993.
- [11] L. Héctor Juárez V. *Diferencias Finitas y Ecuaciones Diferenciales Parciales en Notas de Modelación y Métodos Numéricos*, Memorias del Taller de Métodos Numéricos en Ingeniería y Ciencias Aplicadas. , 3-46, Editores Miguel A. Morelos Vázquez, Salvador Botelló Rionda, 2007
- [12] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*, Springer-Verlag, 1996

- [13] Erwin Kreyszig. *Introductory Functional Analysis with Applications*, John Wiley & Sons. Inc., 1978.
- [14] L. Landweber. *An iteration formula for Fredholm integral equation of the first kind*, American Journal of Mathematics. 73 (1951), pp. 615-624.
- [15] Lavrentiev M, Romanov V and Shishatskii S. *Ill-posed Problems of Mathematical Physics and Analysis*, American Mathematical Society Translations of Mathematical Monographs, vol. 64, (Providence: American Mathematical Society), 1986.
- [16] L. Lebedev, I. Vorovich, G. Gladwell. *Functional Analysis. Applications in Mechanics and Inverse Problems*, Kluwer Academic Publishers. Solid Mechanics and its Applications Volume 100, 2003.
- [17] Khalid Masood. *Investigation of the Initial Inverse Problem in the Heat Equation*, Journal of Heat Transfer, Vol.126, No. 2, pp. 294-296, ASME, 2004.
- [18] V. A. Morozov. *On the solution of large quadratic programming problems with bound constraints*, SIAM Journal on Optimization, 1 (1991), pp. 93-113.
- [19] V.A. Morozov. *Methods for Solving Incorrectly Posed Problems*, Springer, 1984.
- [20] M.Z. Nashed. *Generalized Inverses and Applications*, Academic Press, 1976.
- [21] A. Neumaier. *Solving ill-conditioned and singular linear systems: a tutorial on regularization*, SIAM Review 40 p.p 636-666, 1998.
- [22] Rubén D. Spies, Karina G. Temperini. *Un Método de Molificación para Resolver Problemas Inversos Mal Condicionados: Aplicaciones*, Mecánica Computacional Vol. XXIV, A. Larreteguy (Editor), 2005.
- [23] Erkki Somersalo, Jari Kaipio. *Statistical and Computational Inverse Problems*, Applied Mathematical Sciences Vol. 160, Springer, 2005.
- [24] Tikhonov A and Arsenin V. *Solutions of Ill-posed Problems*, Washington: Winston and Sons, 1977.
- [25] G. Uhlmann. *Inside Out: Inverse Problems and Applications*, Cambridge University Press, 2003.
- [26] G.M. Vainikko & A.Y. Veretennikov, *Iteration Procedures in Ill Posed Problems*, Nauka, 1986.
- [27] Curtis R. Vogel. *Computational methods for inverse problems*, SIAM, 2002.
- [28] *Methods for Solving Incorrectly Posed Problems*, Springer, 1984.



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

MAESTRÍA EN CIENCIAS
(MATEMÁTICAS APLICADAS E INDUSTRIALES)

Regularización de Problemas Mal Planteados

Tesis para la obtención del grado de:

MAESTRA EN CIENCIAS
(MATEMÁTICAS APLICADAS E INDUSTRIALES)

Alumna: Diana Assaely León Velasco

Asesor:

Dr. L. Héctor Juárez Valencia.
UAM-I.

Coasesor:
Dr. J. Jacobo Oliveros Oliveros,
BUAP.