

# Introducción a problemas inversos

Cristian Camilo Barreto Bejarano cbarreto@unal.edu.co

Juan Diego Murcia Porras jmurciap@unal.edu.co

Carlos Enrique Nosa Guzmán cnosa@unal.edu.co

Douglas Leonardo Velasquez Rodriguez dovelasquezr@unal.edu.co

---

## Resumen

En el presente documento, se realiza un acercamiento introductorio a los problemas inversos, con un enfoque sobre las ecuaciones integrales. En primer lugar, se muestra una serie de ejemplos que pretenden evidenciar las extensas aplicaciones de los problemas inversos en la ciencia. Posteriormente, se realiza un breve estudio sobre los principales resultados en el campo de los espacios de Hilbert y la teoría de operadores, necesarios para mostrar los principales aspectos y complicaciones que surgen al intentar solucionar un problema inverso. Por último, se lleva a cabo un análisis sobre los métodos más sencillos para su resolución. Cada método viene acompañado de una implementación numérica ilustrativa que pretende resolver alguna ecuación integral.

*Palabras clave:* Regularización, operadores, ecuaciones integrales, problemas inversos, discretización.

---

## 1. Introducción

Considere una aplicación  $T : X \rightarrow Y$ , que toma una entrada  $f$ , y la transforma en una observación  $g$ . Es decir,

$$Tf = g$$

Un problema directo se nos presenta cuando, dados  $T$  y  $f$ , calculamos cuál será la observación  $g$ . Sin embargo, en muchos problemas matemáticos con aplicaciones en las ciencias e ingeniería, un elemento del miembro izquierdo de nuestra ecuación se convierte en lo que queremos hallar. En un primer caso, nos son dados  $g$  y  $T$ , e intentamos resolver la ecuación para nuestra entrada  $f$ . O, alternativamente, se nos son dados  $f$  y  $g$ , y quisiéramos averiguar qué tipo de proceso  $T$  nos condujo a este resultado. Ambas situaciones se enmarcan dentro de lo que denominamos un *problema inverso* (Groetsch, 2014). En este texto, nos enfocaremos en el primer caso: obtenemos datos medidos indirectamente de  $g$  mediante un proceso bien conocido  $T$ , e intentamos reconstruir el  $f$  que pudo haber conducido a  $g$ .

Como ejemplo, considere la medición de la temperatura del aire, que puede encontrar en [4]. La temperatura como tal es una magnitud que no podemos observar directamente. Para obtener información acerca de esta cantidad, podemos recurrir a un termómetro de mercurio, cuya utilidad reposa en lo predecible que se expande este material en función de la temperatura del aire. Un problema directo, en este caso, sería relacionar el volumen del mercurio en función de la temperatura. Un problema inverso relacionado es dar una estimación de la temperatura que condujo a una expansión específica del mercurio.

No obstante este problema es bastante simple, y no representa ningún reto interesante. Suponga ahora que queremos medir la temperatura de una fragua. En este contexto, nuestro termómetro de mercurio resulta inútil, y debemos recurrir a otros métodos. Una opción es usar el ultrasonido. En este punto, una nueva complicación aparece: las altas temperaturas hacen que los gases dentro de la fragua presenten una mayor turbulencia, afectando sus propiedades acústicas, y por consiguiente, obtenemos cierto “eco” en nuestras mediciones de ultrasonido. Ahora, el problema directo consiste en la tarea no tan fácil de determinar la turbulencia que se genera dentro de la fragua en función de su temperatura, y, a partir de allí, dar una estimación de las posibles mediciones acústicas que obtendremos. El problema inverso asociado es aún más retador: determinar la temperatura dentro de la fragua, dadas ciertas observaciones acústicas.

Dado que la mayoría de problemas inversos no pueden ser resueltos analíticamente, los métodos computacionales toman aquí un rol fundamental. No obstante, la complicación de los problemas inversos radica en que una misma observación pudo haber resultado de distintos tipos de entradas y, debido a ello, requerimos de la mayor cantidad de información posible sobre nuestro sistema. Adicionalmente, una pequeña variación en nuestras observaciones puede afectar de manera drástica nuestra estimación; es decir, estos problemas son *mal condicionados*. Por ello, técnicas especiales como los *métodos de regularización* deben ser usadas para obtener aproximaciones estables de nuestra solución (Vogel, 2002).

El presente trabajo pretende que un lector estudiante de pregrado en Matemáticas o áreas afines, pueda familiarizarse con los principales conceptos, necesarios para el entendimiento de los procedimientos para la resolución de problemas inversos.

## 2. Ejemplos de motivación

Muchos de los problemas inversos que surgen en las distintas ciencias, vienen de situaciones donde el modelo que relaciona los datos que se observan y los que se quieren estimar representa un sistema dinámico, esto es, el modelo en estos casos suele ser una ecuación diferencial, o un sistema de ecuaciones diferenciales incluso, como también, las conocidas como *ecuaciones integrales* (Groetsch, 2004). Estas últimas serán el principal objetivo a tratar en este texto.

Una ecuación integral es aquella en la que la función incógnita aparece dentro de una integral. Los tipos de ecuaciones integrales que mas comunmente surgen en las aplicaciones practicas son las conocidas como *ecuaciones integrales de Fredholm*.<sup>1</sup>

### Ecuaciones integrales de Fredholm

- De primer tipo

$$g(x) = \int_a^b k(x, y)f(y)dy.$$

- De segundo tipo

$$f(x) = g(x) + \lambda \int_a^b k(x, y)f(y)dy.$$

Donde  $k$  y  $g$  son funciones conocidas,  $\lambda$  es un escalar no nulo y  $f$  es la función incógnita. La función  $k$  se denomina el núcleo o kernel de la ecuación integral y en las aplicaciones prácticas suele representar algunas condiciones de los experimentos que se estudian.

Estas ecuaciones se pueden ver como operadores, expresandose de la siguiente manera

$$(Tf)(x) = \int_a^b k(x, y)f(y)dy.$$

$$(T_\lambda f)(x) = f(x) - \lambda \int_a^b k(x, y)f(y)dy.$$

Un tipo especial de kernel que puede aparecer en algunas ecuaciones integrales es el que se conoce como kernel *degenerado* o *separable*, porque se puede expresar de la siguiente forma

$$k(x, y) = \sum_{i=0}^n \alpha_i(x)\beta_i(y),$$

donde  $\{\alpha_i(x)\}_{i=0}^n$  y  $\{\beta_i(y)\}_{i=0}^n$  pueden ser considerados cada uno un conjunto de funciones linealmente independientes. Observe que la diferencia entre las ecuaciones de primer y segundo tipo consiste en que en las primeras la función incógnita aparece únicamente dentro de la integral, mientras que en las segundas, la función incógnita aparece también fuera de la integral.

Otra cosa que vale resaltar es que en estas ecuaciones los límites de integración son fijos, sin embargo, existen situaciones que dan lugar a ecuaciones en las que los límites de integración son variables. Las mas comunes son las conocidas como ecuaciones integrales de Volterra (para mas detalles puede consultar [3]).

### Ecuaciones integrales de Volterra

- De primer tipo

$$g(x) = \int_a^x k(x, y)f(y)dy.$$

- De segundo tipo

$$f(x) = g(x) + \lambda \int_a^x k(x, y)f(y)dy.$$

<sup>1</sup>Los conceptos presentados se encuentran en [3], donde se puede consultar un poco mas a fondo sobre las ecuaciones integrales, sus aspectos teóricos y aplicaciones en las ciencias.

Dado que estos tipos de ecuaciones integrales representan muchas de las situaciones practicas que se estudian en los diferentes campos de investigación, resulta de interés conocer ciertas propiedades de estas, relacionadas con qué tan bien se “comportan”, o como se dice formalmente, su *buen condicionamiento*.

Un problema se dice *bien condicionado* en el sentido de Hadamard si satisface las siguientes condiciones

- el problema tiene solución
- esta solución es única
- la solución depende continuamente de los datos proporcionados.

Desafortunadamente, las ecuaciones integrales en general suelen fallar en alguna de estas tres condiciones, ya sea en la existencia de la solución, como sucede con una ecuación de Fredholm de primer tipo con  $k(x, y) = 1$ , la cual no tiene solución cuando  $g(x)$  no es constante, o bien, en la unicidad de esta, como sucede con esa misma ecuación cuando  $g(x)$  es constante, ya que hay distintas funciones cuyas integrales sobre un mismo intervalo tienen el mismo valor. A medida que avancemos en los ejemplos, veremos que en particular la tercera condición es la que mas suele fallar, en especial en las ecuaciones de Fredholm de primer tipo, y de peor manera, ya que este fallo no es debido a alguna característica en específico de  $k$  o  $g$ , sino que es algo intrínseco a las ecuaciones de este tipo, consecuencia del *Lema de Riemann-Lebesgue* que de manera simple, asegura que cuando  $k \in L^2([a, b]^2)$ , se pueden causar grandes variaciones de  $f(x)$  que tengan como consecuencia variaciones prácticamente imperceptibles en  $g(x)$ , o rigurosamente hablando, si  $Tf = g$ , se puede encontrar  $\hat{f}$  tal que  $T\hat{f} = \hat{g}$  y  $\|\hat{g} - g\|_\infty < \varepsilon$ , aun cuando  $\|\hat{f} - f\|_\infty > M$ , para  $\varepsilon, M > 0$  arbitrarios (Groetsch, 2004).

A continuación presentamos algunos ejemplos que ilustran las dificultades que pueden aparecer al enfrentarse a los problemas inversos. Estos y otros ejemplos relacionados con ecuaciones integrales y aplicaciones de las ciencias matemáticas pueden encontrarse mas detalladamente en [1]. Una selección de ejemplos mas generales se encuentra en [8].

## 2.1. Arqueología térmica

Una de las cuestiones que mas aparece en el estudio de diversos sistemas físicos está relacionada con cómo se comporta la temperatura del sistema, ya sea con respecto a la variación del tiempo, o bien, respecto a otras propiedades de dicho sistema.

Como ejemplo sencillo, veamos el estudio del comportamiento de la distribución de calor en un objeto uniforme, conforme varia el tiempo: consideremos una vara uniforme de longitud 1, aislada en su superficie lateral, de tal manera que el calor fluye unicamente en dirección del eje  $x$ . La función de distribución de calor  $u(x, t)$ , satisface la ecuación diferencial parcial

$$u_t = u_{xx}, \quad 0 \leq x \leq 1.$$

Suponiendo que los extremos de la vara siempre están a temperatura 0, y que  $f(x)$  es la distribución inicial de calor, tenemos las siguientes condiciones iniciales y de frontera

$$u(0, t) = 0, \quad u(1, t) = 0, \quad u(x, 0) = f(x).$$

En estas condiciones, el problema directo que se podria considerar, es hallar la distribución de calor en la vara en un determinado instante, como por ejemplo,  $t=1$ , hallando una función  $g(x)$  tal que  $g(x) = u(x, 1)$ . En nuestro caso, el objetivo es el opuesto, esto es, con la distribución de calor  $g(x)$  de la vara en el instante  $t=1$ , encontrar una aproximación razonable de la distribución inicial  $f(x) = u(x, 0)$ .

Usando el método de separación de variables,  $g(x)$  se puede escribir en forma de serie de la siguiente manera

$$g(x) = \sum_{n=1}^{\infty} a_n \sin(nx), \quad a_n = e^{-n^2} \frac{2}{\pi} \int_0^{\pi} f(\omega) \sin(n\omega) d\omega.$$

Combinando estas dos expresiones y reorganizando, se observa que  $g(x)$  se puede expresar como sigue

$$g(x) = \int_0^{\pi} k(x, \omega) f(\omega) d\omega, \quad k(x, \omega) = \frac{2}{\pi} \sum_{n=1}^{\infty} e^{-n^2} \sin(nx) \sin(n\omega).$$

## 2.2. Prospección geológica

Un problema de interés, principalmente en geología, es el de determinar la localización, forma y constitución de cuerpos subterráneos a partir de medidas realizadas en la superficie. Para hacer esto, una manera es medir la componente vertical de la fuerza de atracción gravitacional que ejercen los cuerpos subterráneos.

El modelo que presentamos como ejemplo es una versión muy simplificada de lo que se hace en realidad, pero que basta para comprender la idea que hay detrás de este tipo de métodos.

La situación que consideramos es la siguiente: realizamos mediciones de la componente vertical de la fuerza gravitacional causada por una densidad de masa variable ubicada una unidad por debajo de la superficie. Así, si la densidad de masa en la posición  $t$ , con  $0 \leq t \leq 1$  está dada por  $x(t)$  y la medida de la componente vertical en la fuerza gravitacional en la posición  $s$ , con  $0 \leq s \leq 1$  es  $y(s)$ , estas se pueden relacionar de la siguiente manera.

- Primero observemos la que fuerza que genera una porción del cuerpo subterráneo ubicada en la posición  $\Delta t$  está dada por la siguiente ecuación ( $\gamma$  es la constante de gravitación universal),

$$\Delta y(s) = \gamma \frac{x(t)\Delta t}{(s-t)^2 + 1} \cos \theta = \gamma \frac{x(t)\Delta t}{((s-t)^2 + 1)^{3/2}}.$$

- Luego, como la fuerza total corresponde a la suma de las fuerzas generadas por cada una de las porciones, se tiene que  $y(s)$  corresponde a esta integral,

$$y(s) = \gamma \int_0^1 \frac{x(t)}{((s-t)^2 + 1)^{3/2}} dt.$$

Así, el problema se reduce a resolver esta ecuación integral.

### 2.2.1. La forma de un cuerpo

Otro problema que se puede considerar y resulta muy parecido al anterior es el de determinar la forma de una colina a partir de mediciones en la componente vertical de la fuerza de atracción gravitacional generada por esta.

De nuevo, considerando un modelo muy simplificado y una situación igualmente sencilla, el problema es el siguiente: en un avión que vuela a una altura de una milla, se realizan mediciones de la componente vertical de la fuerza de atracción gravitacional ejercida por una colina de densidad uniforme. Si denotamos por  $f(s)$  la función que representa estas mediciones, y  $\varphi(x)$  la función que representa la forma de la colina, se tienen las siguientes ecuaciones.

- La fuerza que genera una porción de la colina ubicada en la posición  $\Delta x \Delta y$  es ( $\rho$  es la densidad de la colina)

$$\Delta f(s) = \frac{\gamma \rho \Delta y \Delta x}{(s-x)^2 + (1-y)^2} \cos \theta = \frac{\gamma \rho (1-y) \Delta y \Delta x}{((s-x)^2 + (1-y)^2)^{3/2}}.$$

- La fuerza total que resulta de la suma de las fuerzas de cada una de las porciones de la colina es

$$f(s) = \gamma \rho \int_0^1 \int_0^{\varphi(x)} \frac{(1-y)}{((s-x)^2 + (1-y)^2)^{3/2}} dy dx.$$

En este caso, la ecuación integral que obtuvimos no entra en ninguna de las clasificaciones mencionadas anteriormente, debido a que no es lineal respecto a la función incógnita, sin embargo se puede ver que es una ecuación de Fredholm (límites de integración constantes) escribiéndola de manera adecuada

$$f(s) = \gamma \rho \int_0^1 k(s, x, \varphi(x)) dx, \quad k(s, x, z) = \int_0^z \frac{(1-y)}{((s-x)^2 + (1-y)^2)^{3/2}} dy.$$

## 2.3. Tomografía simplificada

Uno de los avances mas significativos en la medicina del siglo pasado fue debido al desarrollo de métodos y técnicas que permitieran realizar la reconstrucción de imágenes a partir de varias proyecciones de estas, principalmente utilizadas en las tomografías.

El modelo que veremos a continuación ilustra de manera simple cómo se realiza esta reconstrucción a partir de

proyecciones. Para esto, consideremos como objeto a analizar un círculo de radio  $R$  en el plano. Este objeto es iluminado con un rayo de radiación de intensidad  $I_0$  en dirección paralela al eje  $y$ , y en el lado opuesto se mide la intensidad  $I_x$  que tiene el rayo luego de atravesar el objeto. Suponiendo que  $f(x, y)$  representa el coeficiente de absorción de radiación del objeto en el punto  $(x, y)$ , la ley de Bouguer<sup>2</sup> indica que se tiene la siguiente relación

$$\frac{dI}{dy} = -fI.$$

Así, la intensidad de la radiación del rayo luego de atravesar el objeto es

$$I_x = I_0 e^{\int_{-y(x)}^{y(x)} f(x, y) dy}, \quad y(x) = \sqrt{R^2 - x^2}.$$

Definiendo  $p(x) = \ln(I_x/I_0)$  y suponiendo que  $f$  es circularmente simétrica ( $f(x, y) = f(r)$ , con  $r = \sqrt{x^2 + y^2}$ ), podemos reescribir la ecuación anterior como una ecuación integral de primer tipo

$$p(x) = \int_x^R \frac{2r}{\sqrt{r^2 - x^2}} f(r) dr.$$

## 2.4. Modelo de crecimiento exponencial

Como se mencionó al inicio de la sección, muchos problemas inversos se pueden modelar mediante ecuaciones diferenciales, que si bien, no son el objetivo principal a tratar en el texto, pueden ser de utilidad para ilustrar algunas de las características de los problemas inversos mencionadas en la introducción.

El modelo mas simple para considerar quizá sea el de crecimiento exponencial, cuya ecuación que lo representa es

$$\frac{du}{dt} = ru,$$

donde  $r$  es la razón de crecimiento del modelo.

Uno de los problemas inversos a considerar teniendo en cuenta este modelo puede ser el de determinar  $r$  a partir de ciertas mediciones de  $u$ , cuestión que resulta sencilla cuando  $r$  es un parámetro constante, pero cuando  $r$  es variable, ya sea por ejemplo, siendo una función de  $t$  ( $r = r(t)$ ), este problema que en principio aparenta ser igual de sencillo que el anterior, resulta ser en realidad complicado debido a diversos factores como algunos de los que se mencionan a continuación.

Primero, observemos  $r(t)$  se escribe en función de  $u$  de la siguiente manera

$$r(t) = \frac{d}{dt} \ln(u).$$

Teniendo en cuenta que en situaciones reales es casi inevitable la presencia de errores en la medida de  $u$ , debido a como  $r(t)$  se relaciona con  $u$ , es posible que estos errores se magnifiquen a la hora de estimar  $r(t)$ . Un ejemplo de esto es una situación en que el error en la medida de  $u$  cause una perturbación en  $\ln(u)$  de la siguiente manera

$$\ln_\epsilon(u) = \ln(u) + \epsilon \sin\left(\frac{t}{\epsilon^2}\right).$$

Esta perturbación puede resultar despreciable para  $\ln(u)$ , sin embargo, el efecto que tiene en  $r(t)$  puede ser bastante grande

$$r_\epsilon(t) = \frac{d}{dt} \ln(u) + \frac{t}{\epsilon} \cos\left(\frac{1}{\epsilon^2}\right),$$

$$\|r - r_\epsilon\|_\infty = \left\| \frac{1}{\epsilon} \cos\left(\frac{t}{\epsilon^2}\right) \right\|_\infty \rightarrow \infty \text{ cuando } \epsilon \rightarrow 0.$$

Este problema de inestabilidad es algo frecuente en los problemas que involucran estimación de parámetros variables.

<sup>2</sup>La ley de Bouger establece que la intensidad del rayo de radiación decrece proporcionalmente al coeficiente de absorción del objeto que atraviesa.

### 3. Conceptos previos

Antes de introducir formalmente las nociones de mal condicionamiento y regularización, comenzaremos definiendo conceptos y enunciando teoremas que resultan fundamentales para desarrollar la teoría de problemas inversos. Queremos aclarar que, para no extendernos demasiado, no haremos las pruebas de una buena cantidad de teoremas y propiedades. Además, asumiremos que el lector ya tiene cierto conocimiento en álgebra lineal.

#### 3.1. Espacios de Hilbert

Como punto de partida, necesitamos definir los espacios sobre los que trabajaremos. Los espacios de Hilbert resultan siendo bastante adecuados para ello.

**Definición 3.1.1** (Producto interno). Sea  $V$  un espacio vectorial real. Decimos que  $\langle \cdot, \cdot \rangle : V \times V \longrightarrow \mathbb{R}$  es un producto interno si

- (i)  $\langle f, f \rangle \geq 0$  para todo  $f \in V$ ,  $\langle f, f \rangle = 0$  si y sólo si  $f = 0$ .
- (ii)  $\langle f, g \rangle = \langle g, f \rangle$  para todo  $f, g \in V$ .
- (iii)  $\langle \alpha f + \beta g, h \rangle = \alpha \langle f, h \rangle + \beta \langle g, h \rangle$  para todo  $f, g, h \in V$  y  $\alpha, \beta \in \mathbb{R}$ .

**Definición 3.1.2** (Norma). Sea  $V$  un espacio vectorial. Se dice que  $\| \cdot \| : V \longrightarrow \mathbb{R}$  es una norma, si, dados  $f, g \in V$ ,

- (i)  $\|f\| \geq 0$ ,  $\|f\| = 0$  si y sólo si  $f = 0$ .
- (ii)  $\|\alpha f\| = |\alpha| \|f\|$  para todo  $\alpha \in \mathbb{R}$
- (iii)  $\|f + g\| \leq \|f\| + \|g\|$

**Teorema 3.1.1** (Desigualdad de Cauchy-Schwarz). Sea  $V$  un espacio vectorial con producto interno  $\langle \cdot, \cdot \rangle$ . Entonces,  $f, g \in V$  cualesquiera, dados

$$|\langle f, g \rangle|^2 \leq \langle f, f \rangle \langle g, g \rangle.$$

Un producto interno sobre un espacio vectorial induce una norma, como veremos a continuación.

**Definición 3.1.3** (Norma inducida). Sea  $H$  un espacio vectorial con producto interno  $\langle \cdot, \cdot \rangle$ . Definimos la norma inducida por  $\langle \cdot, \cdot \rangle$  a la función  $\| \cdot \| : H \longrightarrow \mathbb{R}$  tal que  $\|x\| = \sqrt{\langle x, x \rangle}$ .

De las propiedades del valor absoluto y de la definición de norma inducida, no es difícil observar que en un espacio vectorial  $S$  con norma inducida  $\|f + g\|^2 + \|f - g\|^2 = 2\|f\|^2 + 2\|g\|^2$  para todo  $f, g \in S$ .

**Ejemplo 3.1.1.** Sea  $V = C[a, b]$  el espacio lineal de las funciones continuas en  $[a, b]$  con la suma componente a componente y el producto por escalar usual. Entonces, dados  $f, g \in V$ ,

$$\|f\|_\infty = \max_{a \leq x \leq b} |f(x)|$$

y,

$$\langle f, g \rangle_c = \int_a^b f(x)g(x)dx.$$

Tenemos que  $\| \cdot \|$  es una norma y  $\langle \cdot, \cdot \rangle$  es un producto interno en  $V$ . Definiremos a la norma inducida por  $\langle \cdot, \cdot \rangle_c$  como  $\| \cdot \|_c$ .

**Definición 3.1.4** (Convergencia de sucesiones). Sea  $V$  un espacio vectorial con norma  $\| \cdot \|$  y producto interno  $\langle \cdot, \cdot \rangle$ , y tome  $(f_n)_{n \in \mathbb{N}}$  una sucesión en  $V$ . Decimos que  $(f_n)_{n \in \mathbb{N}}$  converge a  $f \in V$  si,

$$\lim_{n \rightarrow \infty} \|f_n - f\| = 0,$$

y se denota por  $f_n \rightarrow f$ . Además, si  $\| \cdot \|$  es la norma inducida por  $\langle \cdot, \cdot \rangle$  decimos que  $(f_n)_{n \in \mathbb{N}}$  converge débilmente a  $f$ , si  $\langle f_n, g \rangle \rightarrow \langle f, g \rangle$  para todo  $g \in V$ , y se denota por  $f_n \rightharpoonup f$ .

Observemos que la convergencia de una sucesión implica su convergencia débil, por la desigualdad de Cauchy-Schwarz.

**Definición 3.1.5** (Sucesión de Cauchy). Sea  $V$  un espacio vectorial con norma  $\| \cdot \|$ , y  $(f_n)_{n \in \mathbb{N}}$  una sucesión en  $V$ . Si para todo  $\epsilon > 0$  existe  $N \in \mathbb{N}$  tal que  $n, m \geq N$  implica  $\|f_n - f_m\| < \epsilon$ , decimos que  $(f_n)_{n \in \mathbb{N}}$  es una sucesión de Cauchy.

**Definición 3.1.6** (Espacio completo). Sea  $V$  un espacio vectorial normado.  $V$  se dice un espacio vectorial completo, si toda sucesión de Cauchy converge.

De estas definiciones, lo inmediato es pensar que toda sucesión de Cauchy converge. Sin embargo, hay una variedad de ejemplos en donde esto no se cumple. Lo que sí es cierto, es que a través de clases de equivalencia dadas por sucesiones de Cauchy podemos completar cualquier espacio vectorial normado<sup>3</sup>.

**Definición 3.1.7** (Espacios de Hilbert). Un espacio de Hilbert se define como un espacio vectorial  $H$  con producto interno  $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ , que es completo con la norma inducida por  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$

**Ejemplo 3.1.2.**  $\mathbb{R}^n$  con el producto punto usual que induce la norma euclidiana es un espacio de Hilbert.

**Ejemplo 3.1.3.** Definimos el espacio  $L^2[a, b]$ , respecto al producto interno  $\langle \cdot, \cdot \rangle_c$ , correspondiente a las clases de equivalencia (de las funciones iguales en casi toda parte bajo la medida de Lebesgue) de las funciones medibles  $f$  tales que la integral de  $|f|$  en  $[a, b]$  es finita.  $(L^2[a, b], \langle \cdot, \cdot \rangle_c)$  es un espacio de Hilbert<sup>4</sup>.

De hecho, se puede demostrar que  $L^2[a, b]$  es la completación de  $C[a, b]$ .

**Ejemplo 3.1.4.** El espacio  $H^n[a, b]$ , que corresponde a la completación del espacio de todas las funciones con  $n$  derivadas continuas tales que la norma inducida por

$$\langle f, g \rangle_n = \sum_{i=0}^n \langle f^{(i)}, g^{(i)} \rangle_c$$

es finita, es un espacio de Hilbert.

**Definición 3.1.8** (Conjunto cerrado y débilmente cerrado). Sea  $V$  un espacio vectorial normado, y  $S, N \subseteq V$ .  $S$  se dice cerrado si contiene todos los límites de las sucesiones contenidas en  $S$ ; es decir, si  $(f_n)_{n \in \mathbb{N}} \subseteq S$  y  $f_n \rightarrow f$ , entonces,  $f \in S$ . Si, además, la norma de  $V$  es una norma inducida, decimos que  $N$  es débilmente cerrado, si,  $(g_n)_{n \in \mathbb{N}} \subseteq N$  y  $g_n \rightharpoonup g$ , implican  $g \in N$

A pesar de no ser esta la definición más común de un conjunto cerrado, es una equivalencia bastante útil que nos permite ver más fácilmente cómo definir un concepto análogo con convergencia débil. Es curioso que, además, los espacios de Hilbert cumplen una propiedad similar a la de Bolzano-Weierstrass.

**Teorema 3.1.2** (Teorema de la sucesión débilmente convergente). Toda sucesión acotada en un espacio de Hilbert tiene una subsucesión débilmente convergente.

De este teorema deduce la siguiente propiedad de aproximación.

**Teorema 3.1.3.** Sea  $\mathcal{H}$  un espacio de Hilbert. Si  $C \subseteq \mathcal{H}$  un conjunto débilmente cerrado y  $x_0 \in \mathcal{H}$ , entonces existe  $y \in C$  tal que

$$\|y - x_0\| = \inf\{\|w - x_0\| : w \in C\}.$$

Veamos un poco por qué sucede esto. Sea  $x_0 \in C$  cualquiera. Como  $0 \leq \|w - x_0\|$  para todo  $w \in C$ , entonces, existe  $d \in \mathbb{R}$  tal que  $d = \inf\{\|w - x_0\| : w \in C\}$ . Tome  $(y_n)_{n \in \mathbb{N}}$  una sucesión en  $C$  tal que  $\|y_n\| \rightarrow d$  (está sucesión existe por la definición de ínfimo). Entonces, por el **teorema de la sucesión débilmente convergente**, existen una subsucesión  $(y_{n_k})_{k \in \mathbb{N}}$  de  $(y_n)_{n \in \mathbb{N}}$  y  $y \in C$  tal que  $y_{n_k} \rightharpoonup y$ . Por la desigualdad de Cauchy-Schwarz,

$$\|y - x_0\|^2 = \lim_{k \rightarrow \infty} \langle y_{n_k} - x_0, y - x_0 \rangle \leq \|y - x_0\| \lim_{k \rightarrow \infty} \|y_{n_k} - x_0\| = \|y - x_0\|d$$

y así,  $\|y - x_0\| = d$  (si  $x_0 \in C$ ,  $d = 0$ ). Por lo tanto, como  $C$  es débilmente cerrado,  $y \in C$  y  $\|y - x_0\| = \inf\{\|w - x_0\| : w \in C\}$ .

**Definición 3.1.9.** Sea  $\mathcal{H}$  un espacio de Hilbert. Decimos que  $f, g \in \mathcal{H}$  son ortogonales, si  $\langle f, g \rangle = 0$ . Además, dado  $S \subseteq \mathcal{H}$ , el complemento ortogonal de  $S$ , denotado por  $S^\perp$ , está dado por,

$$S^\perp = \{f \in \mathcal{H} : \langle f, g \rangle = 0 \text{ para todo } g \in S\}.$$

<sup>3</sup>Si al lector le interesa, puede encontrar esta construcción en [13].

<sup>4</sup>Puede encontrar más información sobre la teoría de Lebesgue y sobre los espacios  $L^p$  en [10]

El complemento ortonormal de un conjunto es un subespacio vectorial cerrado. Además, se puede demostrar que todo subespacio cerrado  $S$  de un espacio de Hilbert  $\mathcal{H}$  induce una descomposición ortonormal,  $\mathcal{H} = S \oplus S^\perp$ . Es decir que, para todo  $f \in \mathcal{H}$ , existen  $f_1 \in S$  y  $f_2 \in S^\perp$  tal que  $f = f_1 + f_2$ . En efecto, podemos tomar a  $f_1$  como la proyección de  $f$  en  $S$ , que satisface

$$\|f - f_1\| = \inf\{\|f - g\| : g \in S\}$$

y luego mostrar que  $f_2 = f - f_1 \in S^\perp$ .

**Definición 3.1.10** (Conjunto ortonormal). Sea  $\{f_\alpha\}_{\alpha \in A}$  un conjunto de vectores en un espacio de Hilbert  $\mathcal{H}$ .  $\{f_\alpha\}_{\alpha \in A}$  se llama conjunto o sistema ortogonal si  $\langle f_\alpha, f_\beta \rangle = 0$  para todo  $\alpha, \beta \in A$  con  $\alpha \neq \beta$ . Si, además,  $\|f_\alpha\| = 1$  para todo  $\alpha \in A$ ,  $\{f_\alpha\}_{\alpha \in A}$  se dice conjunto o sistema ortonormal.

Note que, en esta definición,  $A$  no tiene que ser necesariamente finito ni contable<sup>5</sup>.

**Teorema 3.1.4.** Sea  $\{f_\alpha\}_{\alpha \in A}$  un sistema ortonormal en un espacio de Hilbert  $\mathcal{H}$ . Entonces, para todo  $f \in \mathcal{H}$ , a lo sumo contables<sup>6</sup>  $\langle f, f_\alpha \rangle$  son distintos de cero. Además, la desigualdad de Bessel se cumple, esto es,

$$\sum_{\alpha \in A} |\langle f, f_\alpha \rangle|^2 \leq \|f\|^2$$

Veamos por qué la Desigualdad de Bessel se cumple en el caso en que  $A$  es finito.

Sean  $\{f_k\}_{1 \leq k \leq n} \subseteq \mathcal{H}$  un sistema ortonormal y  $f \in \mathcal{H}$ . Entonces, por la definición de producto interno,

$$0 \leq \langle f - \sum_{k=1}^n \langle f, f_k \rangle f_k, f - \sum_{k=1}^n \langle f, f_k \rangle f_k \rangle \quad (1)$$

$$= \|f\|^2 - 2 \sum_{k=1}^n \langle f, f_k \rangle^2 + \sum_{k=1}^n \langle f, f_k \rangle^2 \|f_k\|^2 \quad (2)$$

$$= \|f\|^2 - 2 \sum_{k=1}^n \langle f, f_k \rangle^2 + \sum_{k=1}^n \langle f, f_k \rangle^2 \quad (3)$$

$$= \|f\|^2 - \sum_{k=1}^n \langle f, f_k \rangle^2 \quad (4)$$

luego, si  $A$  es finito, observamos que la desigualdad se cumple para todo  $E \subseteq A$  finito, y por consiguiente, se cumple para el supremo.

**Definición 3.1.11.** Sea  $\{f_\alpha\}_{\alpha \in A}$  un sistema ortonormal en un espacio de Hilbert  $\mathcal{H}$ . Decimos que  $\{f_\alpha\}_{\alpha \in A}$  es un sistema ortonormal completo si  $\langle f, f_\alpha \rangle = 0$  para todo  $\alpha \in A$  si y sólo si  $f = 0$

**Teorema 3.1.5.** Sea  $\{f_\alpha\}_{\alpha \in A}$  un sistema ortonormal completo en  $\mathcal{H}$ . Entonces, para todo  $f \in \mathcal{H}$  existe una única expresión de Fourier,

$$f = \sum_{\alpha \in A} \langle f, f_\alpha \rangle f_\alpha,$$

y la identidad de Parseval se cumple. Esto es,

$$\|f\|^2 = \sum_{\alpha \in A} |\langle f, f_\alpha \rangle|^2.$$

Los coeficientes  $\langle f, f_\alpha \rangle$  son también llamados coeficientes de Fourier. Note que, utilizando la expresión de Fourier, la desigualdad en (1) se vuelve una igualdad si el sistema ortonormal finito es completo.

**Definición 3.1.12** (Espacio separable). Un espacio de Hilbert se dice separable si contiene un sistema ortonormal completo a lo sumo contable.

**Ejemplo 3.1.5.** El conjunto  $\{\sin n\pi x\}_{n \in \mathbb{N}}$  es un sistema ortonormal completo del espacio de Hilbert  $L^2[0, 1]$ , y, por consiguiente,  $L^2[0, 1]$  es separable.

<sup>5</sup>En este trabajo, un conjunto es contable si es equipotente al conjunto de los naturales

<sup>6</sup>Un conjunto es a lo sumo contable si existe una función inyectiva del conjunto a  $\mathbb{N}$



Para finalizar esta subsección, daremos una muy breve introducción de lo que significa un conjunto compacto.

**Definición 3.1.13** (Compacidad). *Un conjunto  $S \subseteq V$ , con  $V$  un espacio vectorial normado, se dice compacto, si todo cubrimiento abierto  $\{O_\alpha\}_{\alpha \in A}$  de  $S$  admite un subcubrimiento finito  $\{O_{\alpha_n}\}_{1 \leq n \leq k}$ .*<sup>7</sup>

De esta definición se desprenden varias caracterizaciones de lo que significa ser un conjunto compacto. Por ejemplo, en  $\mathbb{R}^n$ , los conjuntos compactos son conjuntos cerrados y acotados. Sin embargo, en esta sección recalcaremos la siguiente.

**Teorema 3.1.6.** *Sea  $S \subseteq V$ , con  $V$  un espacio vectorial normado. entonces,  $S$  es compacto si y solo si toda sucesión en  $S$  tiene una subsucesión que converge en  $S$ .*

Por último, para enunciar el **teorema de Arzela-Ascoli**, que caracteriza a los conjuntos compactos en  $C[a, b]$ , definiremos primero qué es una familia **equicontinua** de funciones.

**Definición 3.1.14** (Familia equicontinua de funciones). *Sea  $S \subseteq C[a, b]$ . entonces,  $S$  es llamada una familia equicontinua, o simplemente equicontinua de funciones, si para todo  $\epsilon > 0$ , existe  $\delta > 0$  tal que*

$$\text{si } |x - y| < \delta, \text{ entonces } |f(x) - f(y)| < \epsilon,$$

para todo  $x, y \in [a, b]$ .

**Ejemplo 3.1.6.**  $\{\frac{x^n}{n}\}_{n=1}^\infty$ , con  $x \in [0, 1]$  es una familia equicontinua de funciones.

Observe que toda familia de funciones Lipschitz<sup>8</sup> es equicontinua.

**Teorema 3.1.7** (Teorema de Arzela-Ascoli). *Sea  $S \subseteq C[a, b]$ . Entonces,  $S$  es compacto si y solo si  $S$  es acotada y equicontinua.*

**Ejemplo 3.1.7.** Veamos que, como caso particular del teorema de inmersión de Sobolev,

$$S := \{f \in H^1[0, 1] : \|f\|_1 \leq C\}$$

para algún  $C \in \mathbb{R}$ , es compacto en  $C[0, 1]$ . Sea  $f \in S$ . Entonces,

$$\begin{aligned} f(x) &= f(t) + f(x) - f(t) \\ &= f(t) + \int_t^x f'(s) ds. \end{aligned}$$

Así,

$$\begin{aligned} \int_0^1 f(x) dx &= f(x) \\ &= \int_0^1 f(t) dt + \int_0^1 \int_t^x f'(s) ds dt \\ &= \int_0^1 f(t) dt + \int_0^x \int_t^x f'(s) ds dt - \int_x^1 \int_x^t f'(s) ds dt \\ &= \int_0^1 f(t) dt + \int_0^x \int_0^s f'(s) dt ds - \int_x^1 \int_s^1 f'(s) dt ds \\ &= \int_0^1 f(t) dt + \int_0^x s f'(s) ds - \int_x^1 (1-s) f'(s) ds \\ &= \int_0^1 f(t) dt + \int_0^1 (2s f'(s) - f'(s)) ds. \end{aligned}$$

Ahora bien, por la desigualdad de Cauchy-Schwarz en  $C[0, 1]$  y  $\mathbb{R}$ , y, por la desigualdad triangular,

$$\begin{aligned} |f(x)| &\leq \left| \int_0^1 f(t) dt \right| + \left| \int_0^1 (2s f'(s) - f'(s)) ds \right| \\ &\leq \|f\|_0 + \frac{1}{3} \|f'\|_0 \\ &\leq \|f\|_0 + \|f'\|_0 \\ &\leq \sqrt{2} \|f\|_1 \\ &\leq \sqrt{2} C. \end{aligned}$$

<sup>7</sup>Recuerde que  $\{O_\alpha\}_{\alpha \in A}$  se dice un cubrimiento abierto de  $S$  si cada  $O_\alpha$  es abierto en  $V$ , y  $S \subseteq \cup_{\alpha \in A} O_\alpha$ . Un subcubrimiento finito de  $\{O_\alpha\}_{\alpha \in A}$  se define como una familia finita  $\{O_{\alpha_n}\}_{1 \leq n \leq k}$  contenida en  $\{O_\alpha\}_{\alpha \in A}$  tal que  $S \subseteq \cup_{1 \leq n \leq k} O_{\alpha_n}$ .

<sup>8</sup>Se dice que  $f$  es una función Lipschitz, si existe  $C > 0$  tal que,  $|f(x) - f(y)| < C|x - y|$  para todo  $x, y \in \text{dom}(f)$

Entonces, por la definición de supremo,  $\|f\|_\infty \leq \sqrt{2}C$ . Por consiguiente, por la desigualdad de Cauchy-Schwarz en  $C[t, x]$  o  $C[x, t]$ ,  $y$ , como  $0 \leq \|f\|_0$ ,

$$\begin{aligned} |f(x) - f(t)| &\leq \int_t^x f'(s) ds \\ &\leq |x - t|^{1/2} \|f'\|_0 \\ &\leq |x - t|^{1/2} \|f\|_1 \\ &\leq |x - t|^{1/2} C. \end{aligned}$$

Ello implica que  $S$  es equicontinuo y acotado. Por el teorema de Arzela-Ascoli, concluimos que  $S$  es compacto.

### 3.2. Operadores

Las funciones definidas entre espacios vectoriales se definen formalmente como operadores. Al trabajar con espacios de Hilbert y problemas inversos, es fundamental comprender las principales características de ciertos tipos de operadores (sobre todo, los lineales).

**Definición 3.2.1** (Operador). Sean  $V_1$  y  $V_2$  espacios vectoriales, y,  $S \subseteq V_1$ . Entonces, un **operador** es una función  $T : S \rightarrow V_2$ . Además, denotaremos  $Tf := T(f)$ .

**Ejemplo 3.2.1.** Sea  $D : C^1[a, b] \rightarrow C[a, b]$  definido como  $D(f) = f'$ . Entonces,  $D$  es un operador.

De ahora en adelante, si no se especifica otra cosa, el campo de escalares de un espacio vectorial  $V$  se denotará por  $\mathbb{F}$

**Definición 3.2.2** (Operador lineal). Sean  $V_1$  y  $V_2$  espacios lineales con el mismo campo de escalares  $\mathbb{F}$ , y  $T : V_1 \rightarrow V_2$  un operador. Entonces,  $T$  se llama un **operador lineal**, si,

$$T(\alpha f + \beta g) = \alpha T(f) + \beta T(g)$$

cualesquiera  $f, g \in V_1$  y  $\alpha, \beta \in \mathbb{F}$ .

**Ejemplo 3.2.2.** Sea  $k \in C([a, b] \times [c, d])$ . Entonces, el operador

$$(Kf)(s) = \int_a^g k(s, t) f(t) dt$$

con  $s \in [c, d]$ , es un operador lineal.

Note que el operador derivada  $D$ , definido en el Ejemplo 3.2.1, es un operador lineal.

**Definición 3.2.3** (Operador funcional). Si  $V$  un espacio lineal, y  $T : V \rightarrow \mathbb{F}$  un operador lineal, decimos que  $T$  es un **operador funcional**.

**Ejemplo 3.2.3.** Sea  $\phi_0 : C[a, b] \rightarrow \mathbb{R}$  definido como  $\phi_0(f) = f(0)$ . Entonces,  $\phi_0$  es un operador funcional. Este operador se llama el operador evaluación.

**Definición 3.2.4** (Operador acotado). Sean  $V_1$  y  $V_2$  espacios vectoriales normados. Sean  $T : V_1 \rightarrow V_2$  un operador. Entonces,  $T$  se dice acotado si existe  $M \in \mathbb{R}$  tal que

$$\|T(f)\|_{V_2} \leq M \|f\|_{V_1}$$

para todo  $f \in V_1$ .

**Definición 3.2.5** (Norma de operadores). Sean  $V_1$  y  $V_2$  espacios vectoriales normados, y  $T : V_1 \rightarrow V_2$  un operador acotado. Se define la norma de  $T$  como

$$\|T\| = \sup_{f \neq 0} \frac{\|Tf\|_{V_2}}{\|f\|_{V_1}}$$

Una equivalencia de esta definición que nos será útil más adelante es definir, para  $T$  un operador acotado,  $\|T\| = \sup\{|\langle Tf, g \rangle| : \|f\| \leq 1 \text{ y } \|g\| \leq 1\}$ . En efecto, si  $s$  se define como tal supremo, considere  $f \in \mathcal{H}_1$  y  $g \in \mathcal{H}_2$  tales que  $\|f\| \leq 1$  y  $\|g\| \leq 1$ . Entonces, por la desigualdad de Cauchy-Schwarz,

$$|\langle Tf, g \rangle| \leq \frac{\|Tf\|}{\|f\|} \|g\| \leq \frac{\|Tf\|}{\|f\|}.$$

Es decir,  $s \leq \|T\|$ . Recíprocamente, sea  $f \in \mathcal{H}_1$  tal que  $\|f\| \neq 0$ . Si  $Tf = 0$ ,  $\frac{\|Tf\|}{\|f\|} = 0 \leq s$ . Supongamos ahora que  $\|Tf\| \neq 0$ . Si definimos  $x := \frac{f}{\|f\|}$  y  $y := \frac{Tf}{\|Tf\|}$ , luego  $|\langle Tx, y \rangle| = \frac{\|Tf\|}{\|f\|}$ , y por lo tanto,  $s \leq \|T\|$ . Así,  $\|T\| = s$ .

Además, si tomamos el conjunto de todos los operadores lineales acotados entre dos espacios de Hilbert  $\mathcal{H}_\infty$  y  $\mathcal{H}_\infty$ , con la suma punto a punto y el producto por escalar y la norma definida anteriormente, este es un espacio vectorial normado. La siguiente propiedad caracteriza los operadores funcionales.

**Teorema 3.2.1** (Teorema de representación de Riesz). *Sea  $\mathcal{H}$  un espacio de Hilbert. Entonces, para todo  $T : \mathcal{H} \rightarrow \mathbb{R}$  operador funcional, existe un único  $g \in \mathcal{H}$  tal que*

$$Tf = \langle f, g \rangle$$

para todo  $f \in \mathcal{H}$ . Además,  $\|T\| = \|g\|$ .

Un concepto muy importante con el que trabajaremos mucho de aquí en adelante es el de operador adjunto, del cual discutiremos a continuación.

**Definición 3.2.6** (Operador adjunto). *Sean  $\mathcal{H}_1$  y  $\mathcal{H}_2$  espacios de Hilbert,  $y, T : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  un operador lineal acotado. Entonces, el operador adjunto de  $T$ , si existe, es el operador  $T^* : \mathcal{H}_2 \rightarrow \mathcal{H}_1$  tal que,*

$$\langle Tf, g \rangle = \langle f, T^*g \rangle$$

para todo  $f \in \mathcal{H}_1$  y para todo  $g \in \mathcal{H}_2$ .

**Proposición 3.2.1.** *Sean  $\mathcal{H}_1$  y  $\mathcal{H}_2$  espacios de Hilbert,  $y, T : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  y  $S : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  operadores lineales acotados. Entonces,  $T^*$  existe,  $\|T^*\| = \|T\|$ ,  $T^{**} = T$ , y,  $(TS)^* = S^*T^*$ .*

**Demostración.** Probaremos solo la existencia del operador adjunto y que  $\|T\| = \|T^*\|$ . Las otras dos propiedades se deducen directamente de la definición. Para ver la existencia, sea  $y \in \mathcal{H}_2$ , definamos  $G_y(x) = \langle Tx, y \rangle$  para todo  $x \in \mathcal{H}_1$ . Entonces,  $G$  es un operador funcional continuo, y por el teorema de representación de Riesz, existe un único  $z_y \in \mathcal{H}_2$  tal que  $G_y(x) = \langle x, z_y \rangle$  para todo  $x \in \mathcal{H}_1$ . Así, sea  $T^* : \mathcal{H}_2 \rightarrow \mathcal{H}_1$ , entonces,  $T^*$  es el operador adjunto de  $T$ . Ahora, veamos que  $\|T\| = \|T^*\|$ . Entonces, por definición de operador adjunto,

$$\begin{aligned} \|T\| &= \sup\{|\langle Tf, g \rangle| : \|f\| \leq 1 \text{ y } \|g\| \leq 1\} \\ &= \sup\{|\langle f, T^*g \rangle| : \|f\| \leq 1 \text{ y } \|g\| \leq 1\} \\ &= \sup\{|\langle T^*g, f \rangle| : \|f\| \leq 1 \text{ y } \|g\| \leq 1\} \\ &= \|T^*\|. \end{aligned}$$

□

**Definición 3.2.7.** *Dado  $T : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  un operador, el espacio nulo de  $T$  se define como*

$$N(T) := \{x \in \mathcal{H}_1 : Tx = 0\}$$

y el rango de  $T$  se define como

$$\text{Ran}(T) := \{y \in \mathcal{H}_2 : y = Tx \text{ para algún } x \in \mathcal{H}_1\}.$$

Para la siguiente proposición será útil saber que si  $T$  es un operador lineal acotado, dada  $(f_n)_{n \in \mathbb{N}}$  una sucesión en  $\mathcal{H}_1$  tal que  $f_n \rightarrow f$  con  $f \in \mathcal{H}_1$ , entonces,  $Tf_n \rightarrow Tf$ . Esto es fácil de ver utilizando que, por la definición de norma en operadores,  $\|Tf\| \leq \|T\| \cdot \|f\|$ .

**Proposición 3.2.2.** *Sean  $\mathcal{H}_1$  y  $\mathcal{H}_2$  espacios de Hilbert,  $y, T : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  un operador lineal acotado. Entonces,  $N(T)$  es un espacio vectorial cerrado,  $N(T) = \text{Ran}(T^*)^\perp$ , y,  $N(T^*)^\perp = \text{Ran}(T)$ .*

**Demostración.**

- a. Como  $T$  es un operador lineal, entonces, sean  $x, y \in N(T)^\perp$  y  $\alpha, \beta \in \mathbb{F}$  cualesquiera, por la definición de espacio nulo,  $T(\alpha x + \beta y) = \alpha Tx + \beta Ty = \alpha \vec{0} + \beta \vec{0} = \vec{0}$ , y por consiguiente,  $N(T)^\perp$  es un espacio vectorial. Veamos ahora que  $N(T)^\perp$  es cerrado. Sea  $(f_n)_{n \in \mathbb{N}}$  una sucesión en  $N(T)^\perp$  tal que  $f_n \rightarrow f$  para algún  $f \in \mathcal{H}_1$ , entonces, como  $T$  es un operador lineal acotado, por la definición de  $N(T)^\perp$

$$0 = \lim_{n \rightarrow \infty} Tf_n = T \lim_{n \rightarrow \infty} f_n = Tf$$

así,  $f \in N(T)^\perp$ , y por lo tanto,  $N(T)^\perp$  es cerrado.

- b. Sea  $x \in N(T^*)$ . sea  $y \in \text{Ran}(T)$ . Entonces, por la definición de rango, existe  $a \in \mathcal{H}_1$  tal que  $Ta = y$ . Así,  $\langle x, y \rangle = \langle x, Ta \rangle = \langle T^*x, a \rangle = \langle \vec{0}, a \rangle = 0$ , es decir,  $x \in \text{Ran}(T)^\perp$ . Sea, ahora,  $x \in \text{Ran}(T)^\perp$ . Entonces,  $\langle x, Ta \rangle = 0$  para todo  $a \in \mathcal{H}_1$ , en particular,  $\langle T^*x, T^*x \rangle = \langle x, TT^*x \rangle = 0$ . Por consiguiente,  $\|Tx\| = 0$ , y por la definición de norma,  $T^*x = 0$ , es decir,  $x \in N(T^*)$ . Por lo tanto,  $N(T^*) = \text{Ran}(T)^\perp$ .
- c. Veamos, primero, que  $\overline{\text{Ran}(T)}^\perp = \text{Ran}(T)^\perp$ . Como  $\text{Ran}(T) \subseteq \overline{\text{Ran}(T)}$ , entonces,  $\overline{\text{Ran}(T)}^\perp \subseteq \text{Ran}(T)^\perp$ . Luego, sea  $Z \in \text{Ran}(T)^\perp$ , para todo  $x \in \overline{\text{Ran}(T)}$  existe  $(x_n)_{n \in \mathbb{N}}$ , una sucesión en  $\text{Ran}(T)$  tal que  $x_n \rightarrow x$ . Por consiguiente,  $\langle z, x \rangle = \langle \lim_{n \rightarrow \infty} x_n, z \rangle = \lim_{n \rightarrow \infty} \langle x_n, z \rangle = 0$ . Por lo tanto,  $\overline{\text{Ran}(T)}^\perp = \text{Ran}(T)^\perp$ . Así, como  $\overline{\text{Ran}(T)}$  es un subconjunto cerrado en  $\mathcal{H}_2$ ,

$$N(T)^\perp = \text{Ran}(T^*)^{\perp\perp} = \overline{\text{Ran}(T^*)}^{\perp\perp} = \overline{\text{Ran}(T^*)}$$

**Definición 3.2.8** (Operador compacto). *Un operador lineal acotado<sup>9</sup>  $K : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  es **compacto** si y solo si la imagen de cualquier conjunto acotado es un conjunto relativamente compacto (es decir, su adherencia es un conjunto compacto en  $\mathcal{H}_2$ ).*

**Ejemplo 3.2.4.** Sea  $K : H^1[0, 1] \rightarrow C[0, 1]$  definido como  $Kf = f$ . Utilizando el caso particular del teorema de inmersión de Sobolev vista en el **ejemplo 3.1.7**,  $K$  es un operador compacto:

Se puede demostrar también que operadores lineales acotados con rango de dimensión finita son compactos.

Si  $T : H \rightarrow H$  es un operador lineal sobre un espacio de Hilbert  $H$ , el *espectro* de  $T$ ,  $\sigma(T)$ , se define como el conjunto de valores complejos  $\lambda$  para los cuales el operador  $(T - \lambda I)$  no tiene inverso acotado. Si  $T$  es además auto-adjunto, se tiene que  $\sigma(T) \subseteq \mathbb{R}$ . Cuando  $T$  es compacto y auto-adjunto,  $\sigma(T)$  resulta ser el conjunto de valores propios de  $T$ . Los valores no nulos en  $\sigma(T)$  puede ser ordenados en una sucesión decreciente  $|\lambda_1| > |\lambda_2| > \dots$  que es finita cuando  $\text{Ran}(T)$  es de dimensión finita, o que satisface que  $\lambda_n \rightarrow 0$  cuando  $\text{Ran}(T)$  es de dimensión infinita. Si tomamos un conjunto  $\{v_j\}$  formado por vectores propios, cada uno asociado a un valor propio distinto, este conjunto resulta ser ortogonal y completo en  $\overline{\text{Ran}(T)} = N(T)^\perp$ <sup>10</sup>. Además, para todo  $f \in H$ ,

$$f = Pf + \sum_{i=1}^{\infty} \langle f, v_i \rangle v_i$$

donde  $Pf$  es la proyección ortogonal de  $f$  en  $N(T)$ . Por lo tanto,

$$Tf = \sum_{i=1}^{\infty} \lambda_i \langle f, v_i \rangle v_i.$$

Considere la ecuación de Fredholm de segundo tipo de  $L^2[0, 1]$ ,

$$-\lambda f(s) + \int_0^1 k(s, t) f(t) dt = g(s),$$

donde  $\lambda$  es un parámetro no nulo conocido. Esta ecuación puede ser transformada en la siguiente ecuación de operadores

$$(K - \lambda I)f = g \tag{5}$$

donde  $K$  es un operador integral por  $k(\cdot, \cdot)$ . Si la función  $k$  es simétrica y cuadrado integrable, entonces  $K$  es compacto y autoadjunto y por lo tanto es válido aplicar la teoría espectral. En particular, si  $\lambda$  no es un valor propio de  $K$  entonces  $(K - \lambda I)^{-1}$  es un operador lineal acotado. Por lo tanto, la anterior ecuación tiene una única solución en  $L^2[0, 1]$  dada por  $f = (K - \lambda I)^{-1}g$  para cada  $g \in L^2[0, 1]$ , solución que depende continuamente de  $g$ . Por la anterior característica de la solución a este problema, tenemos que (5) es un problema bien condicionado siempre y cuando  $\lambda$  no sea un valor propio de  $K$ .

Por otra parte, cuando  $\lambda$  es valor propio de  $K$  existe una solución de (5) si y solo si

$$g \in \text{Ran}(K - \lambda I) = N(K - \lambda I)^\perp$$

<sup>9</sup>Recuerde que un operador entre dos espacios vectoriales es **lineal** si y solamente si preserva la suma y el producto por escalar. Además, el operador es **acotado** si la imagen de todo el dominio del operador es acotada bajo la norma inducida por el producto interno definido en el codominio de la función.

<sup>10</sup>El desarrollo de estos resultados puede encontrarse mas detalladamente en [9].

es decir, si y solo si  $g$  es ortogonal a todos los vectores propios asociados al valor propio  $\lambda$ . Todas las soluciones para este caso pueden ser expresadas de la siguiente manera

$$f = \sum_j \frac{\langle g, v_j \rangle}{\lambda_j - \lambda} v_j - \lambda^{-1} P g + \Psi,$$

donde  $\Psi$  es una función del espacio propio  $N(K - \lambda I)$ ,  $Pg$  es la proyección ortogonal de  $g$  sobre  $N(K)$  y la sumatoria  $\sum_j$  se hace sobre todos los índices  $j$  tales que  $\lambda_j \neq \lambda$ . Note que para esta solución general existe una única solución que tiene norma mínima, la cual representaremos por

$$f = \sum_j \frac{\langle g, v_j \rangle}{\lambda_j - \lambda} v_j - \lambda^{-1} P g,$$

esta solución depende continuamente de  $g$ . En este sentido, la ecuación integral de Fredholm de segundo grado con operador autoadjunto y compacto  $K$  es un problema bien condicionado independientemente del valor de  $\lambda$  ( $\lambda \neq 0$ ). No obstante, no podemos asegurar que siempre todos los operadores que se tienen para ecuaciones de este tipo son autoadjuntos, para arreglar este percance a partir de estos operadores dados, podemos definir otros operadores que sí resultan auto-adjuntos, y de los cuales podemos obtener ciertos valores que permiten dar una representación especial del operador original.

Sea  $K : H_1 \rightarrow H_2$  un operador lineal compacto del espacio de Hilbert  $H_1$  en el espacio de Hilbert  $H_2$ . Consideremos los operadores  $K^*K : H_1 \rightarrow H_1$  y  $KK^* : H_2 \rightarrow H_2$ . Estos operadores tienen los mismos valores propios, además son compactos y auto-adjuntos, con lo cual, sabemos que sus valores propios son reales, y además en este caso, los valores propios no nulos son todos positivos. Podemos ordenar estos valores propios no nulos en una sucesión decreciente  $\lambda_1 \geq \lambda_2 \geq \dots$ , y podemos asociarle una sucesión de vectores propios ortonormales de  $K^*K$ ,  $v_1, v_2, \dots$ . El conjunto  $\{v_n\}$  es completo en  $\overline{Ran(K^*K)} = N(K)^\perp$ . Sean  $\mu_j = \sqrt{\lambda_j}$  y  $u_j = \mu_j^{-1} K v_j$ , entonces se tienen las siguientes igualdades

$$\begin{aligned} K^* u_j &= \mu_j v_j, \\ K v_j &= \mu_j u_j, \\ KK^* u_j &= \mu_j K v_j = \mu_j^2 u_j = \lambda_j u_j. \end{aligned}$$

El conjunto de vectores propios ortonormales  $\{u_j\}$  del operador  $KK^*$  es completo en  $\overline{Ran(KK^*)} = N(K^*)^\perp$ . El sistema  $\{v_j, u_j; \mu_j\}$  se denomina *sistema singular* del operador  $K$ , y los valores  $\mu_j$  se llaman *valores singulares* de  $K$ . Todo elemento  $f \in H_1$  se puede representar de la siguiente manera

$$f = P f + \sum_{i=1}^{\infty} \langle f, v_j \rangle v_j,$$

donde  $P$  es la proyección ortogonal de  $H_1$  sobre  $N(K)$ , con lo cual se tiene la siguiente expresión para  $K$

$$K f = \sum_{i=1}^{\infty} \mu_j \langle f, v_j \rangle u_j.$$

Esta representación de  $K$  se conoce como la *descomposición en valores singulares*.

Como comentamos al principio, esta forma de expresar el operador  $K$  es útil en cuanto nos permite observar más fácilmente algunas de las propiedades que posee. Sin embargo, esa no es la única utilidad de esta representación, ya que entre otras cosas, esta permite establecer un criterio para determinar cuándo la ecuación  $Kx = y$  tiene solución, como veremos a continuación.

Si la ecuación  $Kf = g$  tiene solución, entonces  $g \in Ran(K)$  y

$$\lambda_j^{-1} |\langle g, u_j \rangle|^2 = \lambda_j^{-1} |\langle K f, \mu_j^{-1} K v_j \rangle|^2 = |\langle f, v_j \rangle|^2,$$

y de la desigualdad de Bessel, se tiene que

$$\sum_{k=1}^{\infty} \lambda_j^{-1} |\langle g, u_j \rangle|^2 = \sum_{j=1}^{\infty} |\langle f, v_j \rangle|^2 \leq \|f\|^2 < \infty.$$

Recíprocamente, si  $g \in \overline{\text{Ran}(K)} = N(K^*)^\perp$ , satisface

$$\sum_{k=1}^{\infty} \lambda_j^{-1} |\langle g, u_j \rangle|^2 < \infty,$$

entonces cualquier función de la forma

$$f = \sum_{j=1}^{\infty} \frac{\langle g, u_j \rangle}{\mu_j} v_j + \varphi,$$

donde  $\varphi \in N(K)$  es solución de esta ecuación.

De lo anterior tenemos el que se conoce como *criterio de existencia de Picard*, el cual dice que la ecuación  $Kf = g$  tiene solución si y sólo si  $g \in \overline{\text{Ran}(K)}$  y satisface la condición mencionada arriba.<sup>11</sup>

### Descomposición en valores singulares de una matriz

La descomposición en valores singulares de una matriz es útil tanto para la teoría como para la práctica. A continuación se presenta un breve tratamiento del tema.

**Definición 3.2.9** (SVF). *La factorización en valores singulares de una matriz  $A_{m \times n}$  es una factorización de la forma*

$$A = U \Sigma V^*$$

donde  $U_{m \times m}$  y  $V_{n \times n}$  son matrices unitarias<sup>1213</sup> y  $\Sigma_{m \times n}$  es una matriz diagonal no negativa, es decir, todo elemento que no se encuentre en la diagonal de  $\Sigma$  es igual a cero y que los elementos de la diagonal son no negativos. Los elementos  $\sigma_{i=1}^{\min(m,n)}$  de la diagonal de  $\Sigma$  son llamados **valores singulares** de la matriz y los vectores columnas de las matrices  $U$  y  $V$  son llamados **vectores singulares**.

Podemos considerar que en la factorización SVF de una matriz  $A$ , la matriz  $\Sigma$  cumple que sus elementos diagonales cumplen que  $\sigma_i \geq \sigma_{i+1}$  para  $i = 1, \dots, \min(m, n) - 1$ . Cabe destacar que la matriz  $\Sigma$  generaliza el concepto de una matriz diagonal puesto que no necesariamente es una matriz es cuadrada y puede considerarse de las siguientes formas

$$\Sigma = \begin{bmatrix} \text{diag}((\sigma_i)_{1 \leq i \leq n}) \\ (0)_{(m-n) \times n} \end{bmatrix} \quad \Sigma = \begin{bmatrix} \text{diag}((\sigma_i)_{1 \leq i \leq m}) & (0)_{m \times (n-m)} \end{bmatrix}$$

Un ejemplo de la existencia de esta factorización para una matriz de tamaño pequeño es el siguiente

**Ejemplo 3.2.5.** *Dada la matriz*

$$A = \frac{1}{15} \begin{bmatrix} 14 & 2 \\ 4 & 22 \\ 16 & 13 \end{bmatrix}$$

su factorización en valores singulares es<sup>14</sup>

$$A = U \Sigma V^* = \frac{1}{3} \begin{bmatrix} 1 & 2 & 2 \\ 2 & -2 & 1 \\ 2 & 1 & -2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \frac{1}{5} \begin{bmatrix} 3 & 4 \\ 4 & -3 \end{bmatrix}$$

Un resultado importante es la existencia de esta descomposición para cualquier matriz perteneciente al conjunto  $\mathbb{C}^{m \times n}$ , lo cual se enuncia a continuación

**Teorema 3.2.2** (Existencia de SVD). <sup>15</sup> *Suponga que para  $m, n, r \in \mathbb{N}$  se tiene que  $A \in \mathbb{C}^{m \times n}$  con  $\text{ran}(A) = r$ , y que  $(\lambda_j), \mathbf{v}_j$  son pares propios ortonormales para  $A^*A$  (matriz de tamaño  $n \times n$ ) con  $\lambda_1 \geq \dots \geq \lambda_r > 0 = \lambda_{r+1} = \dots = \lambda_n$ . Defina las matrices*

a.  $\mathbf{V} := [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n] \in \mathbb{C}^{n \times n}$

b.  $\Sigma \in \mathbb{R}^{m \times n}$  es la matriz diagonal con elementos  $\sigma_{jj} := \sqrt{\lambda_j}$  para  $j = 1, \dots, \min(m, n)$ .

c.  $\mathbf{U} := [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_m] \in \mathbb{C}^{m \times m}$  donde  $\mathbf{u}_j := \sigma_{jj}^{-1} A \mathbf{v}_j$  para  $j = 1, \dots, r$  y  $\mathbf{u}_{r+1}, \dots, \mathbf{u}_m$  son tomados de tal manera que las columnas de  $\mathbf{U}$  forman una base ortonormal de  $\mathbb{C}^m$ .

Entonces  $A = \mathbf{U} \Sigma \mathbf{V}^*$ .

<sup>11</sup>El esbozo de la teoría y resultados anteriores se encuentra en [1].

<sup>12</sup>La matriz  $V^*$  significa la matriz transpuesta conjugada de  $V$ .

<sup>13</sup>Recordemos que una matriz es **unitaria** es equivalente a afirmar que esta matriz multiplicada por su transpuesta conjugada es igual a la identidad.

<sup>14</sup>Es un ejercicio ilustrativo comprobar que en realidad la igualdad es cierta.

<sup>15</sup>Puede revisar en [4]

### 3.3. Operadores mal condicionados

<sup>16</sup> Como observamos anteriormente en la sección de ejemplos, muchos de los problemas inversos que se consideran frecuentemente suelen ser mal condicionados, esto es, bien pueden no tener solución, o en caso de tenerla puede que esta no sea única, y en muchos casos a pesar de existir y ser única, esta es bastante susceptible de sufrir grandes perturbaciones cuando los datos de entrada sufren pequeñas perturbaciones.

Por esto, se hace necesario buscar maneras de tratar con el mal condicionamiento de estos problemas. Para este propósito consideraremos un enfoque un poco mas abstracto, viendo los problemas inversos como ecuaciones de operadores lineales entre espacios de Hilbert, esto es, ecuaciones de la forma

$$Kx = y,$$

donde  $K : H_1 \rightarrow H_2$  es un operador lineal acotado entre espacios de Hilbert reales.

La dificultad de la estabilidad en estas ecuaciones en ciertos casos puede ser solventada, cuando el problema considerado permite imponer ciertas restricciones sobre las soluciones que admitimos, como se enuncia en el siguiente resultado:

**Teorema 3.3.1** (Tikhonov). *Si  $K : D(K) \rightarrow Y$  es un operador continuo e inyectivo, y  $C \subseteq D(K)$  es compacto, entonces el operador inverso  $(K|_C)^{-1}$  es continuo.*

Lo que nos dice este resultado, es que en los problemas en los que podamos permitirnos considerar como posibles soluciones únicamente aquellas que cumplan con cierta restricción que haga del conjunto de estas un conjunto compacto, será un problema estable en el sentido de que pequeñas perturbaciones en los datos iniciales no generarán grandes perturbaciones en las estimaciones realizadas.

Sobre la existencia de soluciones de estas ecuaciones, si bien no siempre es posible hallar una solución exacta al problema puesto que estos surgen de aplicaciones prácticas, muchas veces será suficiente encontrar una aproximación razonable a la solución que estamos buscando. Este tipo de *soluciones generalizadas* se conocen como *soluciones de mínimos cuadrados*, y se definen de la siguiente manera,

**Definición 3.3.1.** *Una función  $x_0$  se dice que es una solución de mínimos cuadrados del problema  $Kx = y$  si satisface la siguiente condición*

$$\|Kx_0 - y\| = \inf\{\|ku - y\| : u \in H_1\}.$$

Esta condición es equivalente a decir que  $Py \in \text{Ran}(K)$ , donde  $P$  es la proyección ortogonal de  $H_2$  sobre  $\text{Ran}(K)$ , y esta última condición resulta equivalente a que

$$y = Py + (I - P)y \in \text{Ran}(K) + \text{Ran}(K)^\perp.$$

Con esto garantizamos la existencia de una solución de mínimos cuadrados para cualquier elemento del subespacio denso de  $H_2$ ,  $\text{Ran}(K) + \text{Ran}(K)^\perp$ .

Otra condición equivalente a las anteriores es la siguiente

$$Kx - y \in \text{Ran}(K)^\perp = N(K^*),$$

o bien,

$$K^*Kx = K^*y,$$

de donde se observa que existe una única solución de mínimos cuadrados si y solo si

$$N(K^*K) = N(K) = \{0\}.$$

Además, el conjunto de todas las soluciones de minimos cuadrados es cerrado y convexo, y por tanto, existe una única solución de minimos cuadrados cuya norma es menor que la de las demás soluciones. Considerando esta solución podemos definir un operador  $K^\dagger : D(K^\dagger) \rightarrow H_1$  que a cada elemento de  $D(K^\dagger) = \text{Ran}(K) + \text{Ran}(K)^\perp$  le asigne su correspondiente solución por mínimos cuadrados de norma mínima. Este operador se conoce como *inverso generalizado de Moore-Penrose* del operador  $K$ .

Con lo anterior hemos dado una solución razonable al problema de la existencia y unicidad de las soluciones a las ecuaciones de operadores. Sin embargo, la que puede ser la mayor dificultad (la estabilidad de las soluciones), sigue presente y es de interés conocer bajo qué condiciones podemos asegurar que el problema considerado es estable en este sentido, esto es, cuándo el operador  $K^\dagger$  es continuo. La respuesta a esto se encuentra en el siguiente resultado:

<sup>16</sup>La teoría y resultados mencionados a continuación se encuentran en [1].

**Teorema 3.3.2.**  $K^\dagger : D(K^\dagger) \rightarrow H_1$  es un operador cerrado densamente definido que es acotado si y sólo si  $\text{Ran}(K)$  es cerrado.

Muchos de los problemas que se pueden considerar en las ciencias aplicadas dan lugar a ecuaciones integrales cuyo núcleo es cuadrado-integrable, esto es, que  $k(x, y) \in L^2([a, b]^2)$ , lo cual hace que estos operadores sean compactos, y para que  $\text{Ran}(K)$  sea cerrado cuando  $K$  es compacto, necesariamente  $\text{Ran}(K)$  debe ser de dimensión finita. Esta condición nos dice que las únicas ecuaciones integrales de Fredholm de primer tipo para las cuales  $K^\dagger$  es continuo, son aquellas cuyo kernel es degenerado.

Cuando el operador  $K$  es compacto se puede encontrar una representación bastante útil de  $K^\dagger$  en términos de su sistema de valores singulares  $\{v_j, u_j; \mu_j\}$

Si  $y \in D(K^\dagger)$ , entonces  $y = y_1 + y_2$ , con  $y_1 \in \text{Ran}(K)$  y  $y_2 \in \text{Ran}(K)^\perp$ . Como  $u_j \in \text{Ran}(K)$ , para todo  $j \in \mathbb{N}$  se tiene que  $\langle y, u_j \rangle = \langle y_1, u_j \rangle$  y por tanto, el vector

$$x = \sum_{i=0}^{\infty} \frac{\langle y_1, u_k \rangle}{\mu_k} v_k = \sum_{i=0}^{\infty} \frac{\langle y, u_k \rangle}{\mu_k} v_k,$$

existe por el criterio de Picard. Este vector satisface que  $Kx = y_1$ , y  $x \in N(K)^\perp$ , esto es, el vector es la solución de mínimos cuadrados de norma mínima, esto es,  $K^\dagger$  se puede representar como

$$K^\dagger y = \sum_{i=0}^{\infty} \frac{\langle y, u_k \rangle}{\mu_k} v_k.$$

Una de las cosas que se pueden notar a primera vista con esta representación de  $K^\dagger$  es por qué es necesario que  $\text{Ran}(K)$  tenga dimensión finita para que  $K^\dagger$  sea continuo.

Si consideramos  $y^\varepsilon = y + \varepsilon u_n$ , se tiene que  $\|y - y^\varepsilon\| = \varepsilon$  pero

$$\|K^\dagger y - K^\dagger y^\varepsilon\| = \frac{\varepsilon}{\mu_n} \rightarrow \infty \text{ cuando } n \rightarrow \infty.$$



## 4. Algunos métodos *generales* para resolver problemas inversos

Hasta el momento, el lector ha podido evidenciar la importancia de los problemas inversos en diversos campos de estudio. Hemos introducido los conceptos claves para su desarrollo, mostrando las dificultades a la hora de buscar soluciones. El presente capítulo pretende dar una justificación teórica a los métodos más simples para su resolución<sup>17</sup>. Cada uno de los métodos propuestos viene acompañado de una implementación numérica, realizada para la resolución de algunas ecuaciones integrales.

Existe una gran cantidad de literatura que habla sobre estos métodos. Recomendamos al lector que quiera profundizar en estos temas ver [1],[3] y [4]. Siendo este un trabajo introductorio, es imposible introducir los métodos que se utilizan actualmente. En el presente trabajo, discutiremos primero las dificultades en la discretización de problemas inversos. Posteriormente, estudiaremos tres métodos para su resolución: Regularización (con enfoque en la Regularización de Tikhonov), los métodos iterativos, y la descomposición en valores singulares truncada (TSVD por sus siglas en inglés).

### 4.1. Discretización

Como es de esperarse, al implementar computacionalmente un problema inverso, es necesario *discretizar* el método. Es decir, convertir nuestro problema en uno en el que tengamos que hallar finitas variables. Considere, por ejemplo, la ecuación de Fredholm de primer tipo

$$\int_a^b k(s, t)x(t)dt = y(s).$$

La manera más inmediata para discretizar este problema es aplicar alguna técnica de cuadratura, como la regla del trapecio o la de Simpson, a la integral<sup>18</sup>, y obtener el problema aproximado

$$\sum_{j=1}^n w_j k(s, t_j)x_j = y(s), \quad (6)$$

donde queremos hallar  $x_j = x(t_j)$ . Es claro que todavía 6 representa un problema infinito. Debemos entonces realizar lo que se denomina *colocación*: pedir que la condición se cumpla en una cantidad  $m$  de puntos  $\{s_i\}_{i=1}^m$ . Es decir

$$\sum_{j=1}^n w_j k(s_i, t_j)x_j = y(s_i), i = 1, 2, \dots, m.$$

Este problema se convierte entonces en un sistema de ecuaciones de tamaño  $m \times n$  de la forma

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

donde

$$\mathbf{A} = [w_j k(s_i, t_j)], \mathbf{x} = [x(t_1), x(t_2), \dots, x(t_n)]^T, \mathbf{b} = [y(s_1), y(s_2), \dots, y(s_m)]^T$$

Idealmente, entre mayor sea el número de particiones que consideremos, mejor será nuestra aproximación del problema. Sin embargo, para problemas mal condicionados, a medida que queramos aumentar tanto  $m$  como  $n$  en nuestra discretización, el mal condicionamiento del problema continuo se heredará, y nuestra matriz  $\mathbf{A}$  será cada vez peor condicionada ( $\kappa(\mathbf{A})$  aumentará considerablemente) (Groetsch, 2014).

Como ejemplo consideremos la siguiente aplicación al problema de desenfoque de imagen: Un modelo *natural*<sup>19</sup> para resolver el problema de determinar una imagen nítida conociendo su imagen desenfocada y algunos fenómenos que puedan causar este fenómeno se reduce en la ecuación integral de Fredholm planteada sobre un espacio de dos dimensiones, donde el espacio de dos dimensiones es un rectángulo de  $\mathbb{R}^2$  que representa el espacio de píxeles de la imagen. Para dar una pequeña introducción a este problema inverso, solo vamos a considerar este modelo en una dimensión. Así, dada la ecuación

$$g(x) = \int_0^1 K(x - x')f(x')dx' \stackrel{def}{=} \mathcal{K}f \quad (7)$$

<sup>17</sup>Algunos autores, como en [3], utilizan el término “resolución” a la hora de dar una respuesta para un problema inverso, en vez de “solución” o “solución aproximada”. Ello con el objetivo de resaltar las dificultades inusuales que se presentan. En el presente trabajo, utilizaremos todos estos términos sin distinción por motivos lingüísticos.

<sup>18</sup>El alcance de este trabajo no nos permite cubrir el tema de integración numérica. Si el lector quiere profundizar en estos temas, recomendamos ver la Sección 7 de [6].

<sup>19</sup>La razón de proponer este modelo a el problema inverso es desarrollado en [7]. Las funciones  $f$  y  $g$  son funciones de valor real definidas sobre un subconjunto de  $\mathbb{R}^2$  que representan cierto valor para cada pixel en el espacio de la imagen y el kernel  $k$  es una comparación entre la imagen nítida y la imagen desenfocada basandose en herramientas físicas para el estudio de fenómenos como el de difracción de la luz.

queremos hallar la función  $f$ , la cual representa la imagen nítida conociendo los fenómenos de dispersión representados por el kernel  $K$  y la imagen desenfocada representada por la función  $g$ .

Para dar una noción del mal condicionamiento de este problema consideremos las funciones conocidas y definidas sobre  $[0, 1]$

$$g(x) = x \quad K(y) = e^{y^2}$$

donde  $y = x - x'$ . Se construye una partición equiespaciada de tamaño  $n$  del intervalo  $[0, 1]$  y se aproxima la integral de (1) de la siguiente manera

$$g(x) \approx \sum_{k=0}^n K(x - x'_k) f(x'_k) \frac{1}{n} \quad \text{donde } x'_k = \frac{k}{n}$$

Para aproximar la función  $f$ , consideramos los puntos pertenecientes a la partición equiespaciada para que la variable  $x$  tome tales valores, esto resulta en el sistema de ecuaciones lineales

$$g(x_i) \approx \sum_{k=0}^n K(x_i - x'_k) f(x'_k) \frac{1}{n} \quad \text{para } 0 \leq i \leq n$$

El resultado de resolver este sistema de ecuaciones lineales debería resultar en una restricción de la función  $f$ , por tanto, es de esperar que haciendo el tamaño de la partición cada vez mayor, se obtenga una aproximación mejor a la función  $f$ . Sin embargo, el número de condición de cualquier matriz de tamaño  $n$  asociada al problema es demasiado grande como para asegurar una buena aproximación a la función  $f$ .

Una manera de visualizar<sup>20</sup> el mal condicionamiento de tal proceso es presentada en la Figura 1

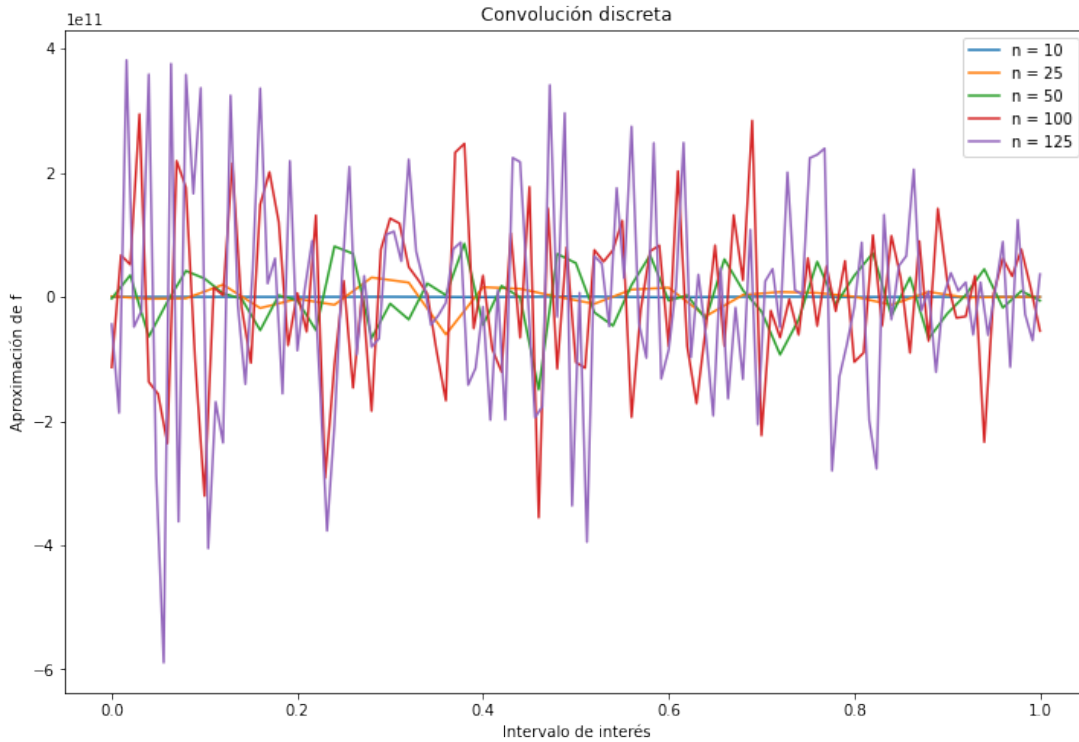


Figura 1: Realización propia. Mal condicionamiento de la matriz del sistema lineal asociado a la discretización del problema de la ecuación (7). Para aplicar el método de discretización se toma el tamaño de la partición en el conjunto  $\{10, 15, 50, 100, 125\}$ .

A pesar de que el tamaño de la partición se va incrementando, la solución no parece tener alguna función límite. Para poder hacer una comparación entre la solución esperada y la solución dada por cada uno de los métodos tratados en este documento, consideremos la ecuación

$$\int_0^1 (s^2 + t^3)^2 f(t) dt = \frac{1}{10} + \frac{2}{7}s^2 + \frac{1}{4}s^4. \quad (8)$$

<sup>20</sup>Todas las imágenes y tablas presentadas en este documento se encuentran desarrolladas en un script anexo que usa el lenguaje de programación *Python*.

donde  $K(t, s) = (s^2 + t^3)^2$  y  $g(s) = \frac{1}{10} + \frac{2}{7}s^2 + \frac{1}{4}s^4$ . Puede comprobarse que la solución a esta ecuación es  $f(t) = t^3$ . Al realizar discretización sobre este problema y posteriormente resolver, obtenemos lo expuesto en la Figura 2.

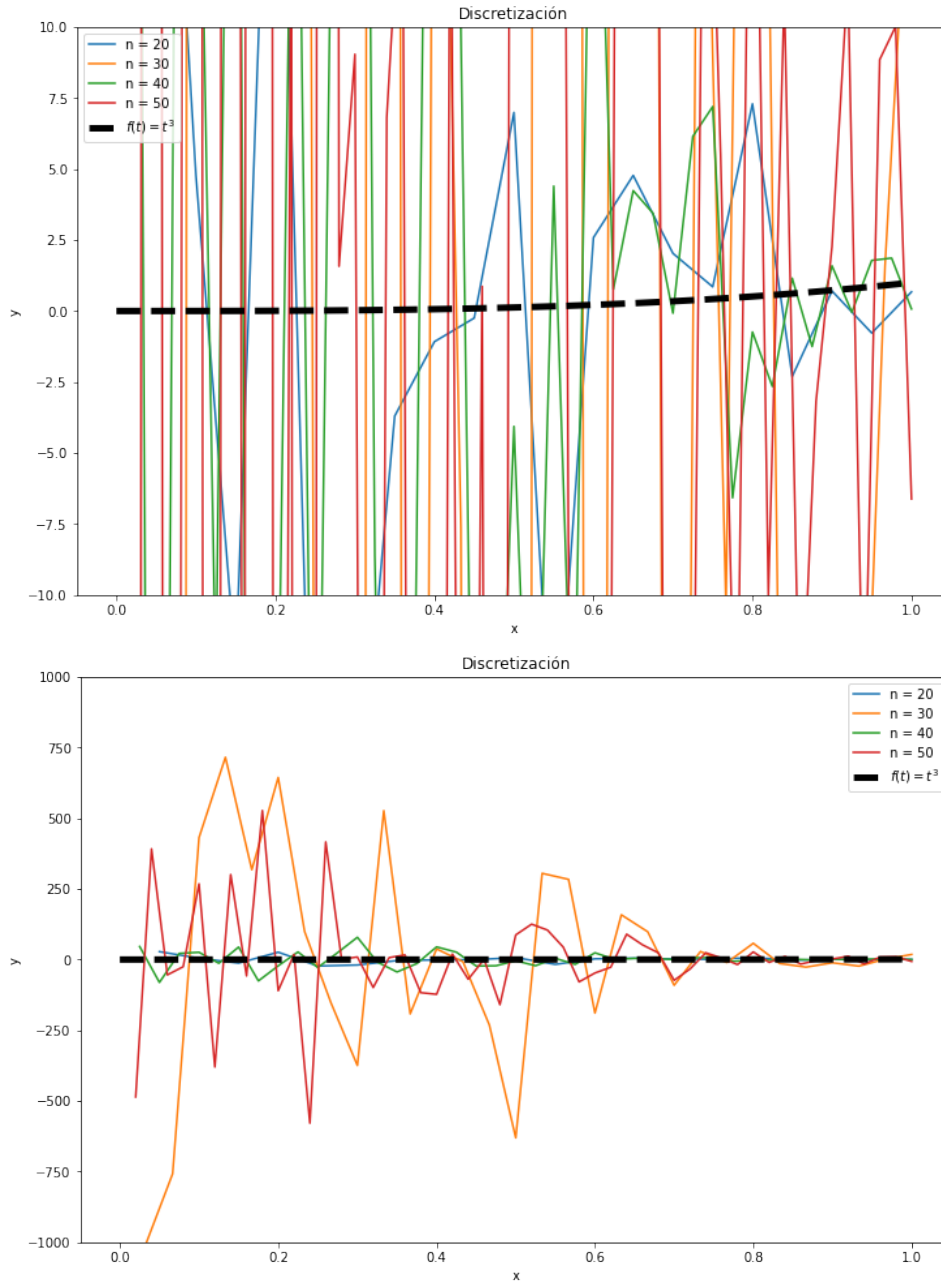


Figura 2: Realización propia. Resultado de discretización de la ecuación (8). En la figura inferior se muestra la *aproximación* de  $f$  por medio del método de discretización cuando  $y$  varía en el intervalo  $[-1000, 1000]$  y en la imagen superior cuando  $y$  varía en  $[-10, 10]$ . El tamaño de la partición  $n$  para el proceso de discretización varía en el conjunto de valores  $\{20, 30, 40, 50\}$ .

En esta figura se comprueba que al aplicar discretización a la ecuación (8) se tiene un problema de mal condicionamiento al resolver por medio de técnicas de álgebra lineal numérica. Otra forma de evidenciar este mal condicionamiento es el número de condición de la matriz  $A$  mostrado en el Cuadro 1.

Con un número de condición tan alto es razonable esperar una aproximación de la solución algorítmica a la solución teórica  $f(t) = t^3$  de una manera *pobre*, como se evidenció en las gráficas.

Tamaño de la matriz	Número de condición
5	1.05896e+17
25	5.87007e+18
45	1.55566e+19
65	1.97658e+20
85	6.27645e+19
105	9.00252e+20
125	5.07849e+19

Cuadro 1: Número de condición de la matriz de discretización con diferentes tamaños de la partición.

#### 4.1.1. Solución de la ecuación de Fredholm de segundo tipo con un núcleo degenerado

En el anterior capítulo, hemos hablado de que, teóricamente y bajo ciertas condiciones, las ecuaciones de Fredholm de primer tipo con un kernel degenerado, y cualquier ecuación de Fredholm de segundo tipo, son problemas bien condicionados. En particular, resolver una ecuación de Fredholm de segundo tipo con un kernel degenerado es un problema que, además de estar bien condicionado, es bastante sencillo de implementar. Lo desarrollado en este apartado nos será útil para realizar la implementación de la Regularización de Tikhonov en el siguiente capítulo. El lector puede consultar más acerca de los métodos para resolver ecuaciones integrales en [3].

Considere una ecuación de Fredholm de segundo tipo

$$f(x) = g(x) + \gamma \int_a^b K(x, y) f(y) dy, \quad \gamma \neq 0 \quad (9)$$

y suponga que el kernel  $K(x, y)$  es *degenerado* (o separable). Es decir,

$$K(x, y) = \sum_{i=1}^n \alpha_i(x) \beta_i(y)$$

Note entonces que

$$f(x) - g(x) = \gamma \sum_{i=1}^n \alpha_i(x) \int_a^b \beta_i(y) f(y) dy := \sum_{i=1}^n h_i \alpha_i(x) \quad (10)$$

$$h_i = \gamma \int_a^b \beta_i(y) f(y) dy \quad (11)$$

Por ende,

$$f(x) = g(x) + \sum_{j=1}^n h_j \alpha_j(x) \quad (12)$$

Calculemos entonces los coeficientes  $h_p$ .

$$\begin{aligned}
g(x) + \sum_{i=1}^n h_i \alpha_i(x) &= g(x) + \gamma \sum_{i=1}^n \alpha_i(x) \int_a^b \beta_i(y) \left[ g(y) + \sum_{j=1}^n h_j \alpha_j(y) \right] dy \\
&= g(x) + \sum_{i=1}^n \alpha_i(x) \left[ \gamma \int_a^b g(y) \beta_i(y) dy + \gamma \sum_{j=1}^n h_j \int_a^b \alpha_j(y) \beta_i(y) dy \right] \\
&:= g(x) + \sum_{i=1}^n \alpha_i(x) \left[ g_i + \gamma \sum_{j=1}^n \gamma_{ij} h_j \right]
\end{aligned}$$

donde

$$g_i = \gamma \int_a^b g(y) \beta_i(y) dy, \quad \gamma_{ij} = \int_a^b \alpha_j(y) \beta_i(y) dy.$$

Entonces,

$$\sum_{i=1}^n h_i \alpha_i(x) = \sum_{i=1}^n \alpha_i(x) \left[ g_i + \gamma \sum_{j=1}^n \gamma_{ij} h_j \right].$$

Es decir que, para  $i = 1, 2, \dots, n$ ,

$$h_i = g_i + \gamma \sum_{j=1}^n \gamma_{ij} h_j.$$

O en forma matricial,

$$\mathbf{h} = \mathbf{g} + \gamma \mathbf{\Gamma} \mathbf{h}. \quad (13)$$

donde

$$\mathbf{\Gamma} = [\gamma_{ij}], \quad \mathbf{h} = [h_i], \quad \mathbf{g} = [g_j].$$

La ecuación (13) puede ser reescrita como

$$(\mathbf{I} - \gamma \mathbf{\Gamma}) \mathbf{h} = \mathbf{g}. \quad (14)$$

donde  $\mathbf{I}$  es la matriz identidad. Este sistema de ecuaciones tiene una única solución, siempre y cuando  $\gamma$  no sea el inverso multiplicativo de algún autovalor de  $\mathbf{\Gamma}$ . Ello produce la *única* solución  $f$  para nuestra ecuación (9). Bajo ciertas condiciones, (14) puede ser resuelto, así  $\gamma$  sea el inverso multiplicativo de algún autovalor de  $\mathbf{\Gamma}$ . En dado caso,  $f$  no es única (Wing, 1991). A continuación se muestra la aplicación de este método a un problema para mostrar su efectividad. Considere el núcleo de la ecuación definido como  $K(s, t) = st$ , y el resto de parámetros del método  $g(s) = s^2 - s$ ,  $[a, b] = [0, 1]$  y  $\gamma = 4$ , entonces la solución la ecuación de Fredholm de segundo tipo es  $f(s) = s^2$ . Al hallar  $f$  por medio del método de núcleo degenerado se obtiene la Figura 3, en la cual la solución teórica y la solución algorítmica se sobrepone.

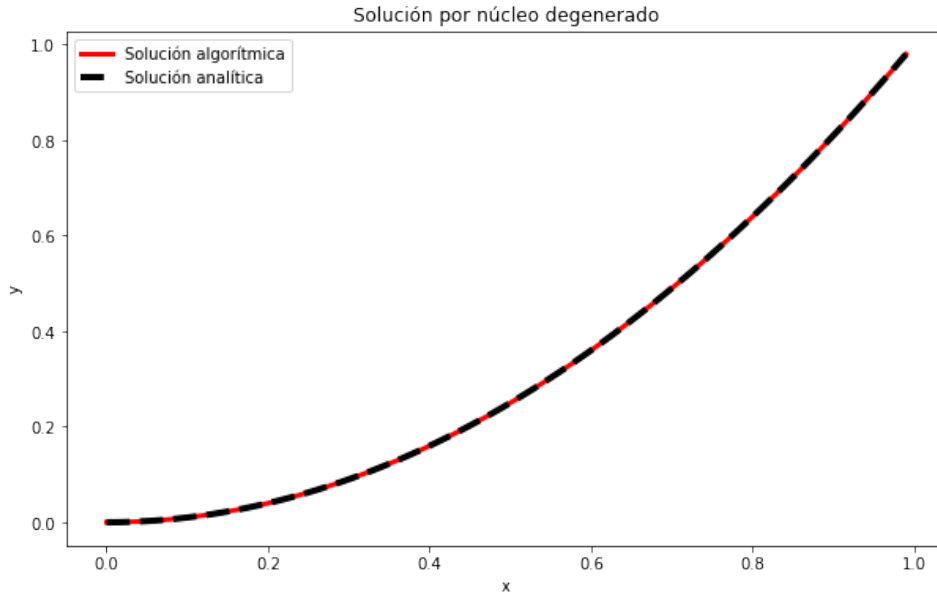


Figura 3: Realización propia. Validación del método de núcleo degenerado para el problema de hallar  $f$  sabiendo que  $f(s) = (s^2 - s) + 4 \int_0^1 (st)f(t)dt$ .

## 4.2. El método de regularización

En general, la regularización trata de métodos utilizados para evitar una variación grande de las aproximaciones obtenidas, controlando el mal condicionamiento de algunos problemas. En el caso de las ecuaciones integrales, sabemos que las ecuaciones que mayor cantidad de problemas nos plantean son las de primer tipo mal condicionadas (con un kernel no degenerado)

$$\int_0^1 k(s, t)x(t)dt = y(s). \quad (15)$$

Podríamos intentar aproximar esta ecuación por una ecuación de Fredholm de segundo tipo (que sabemos que se comportan mejor). Utilizaremos la notación usada en el anterior capítulo, representando la ecuación (15) de manera abstracta mediante

$$Kx = y, \quad (16)$$

donde  $K : H_1 \rightarrow H_2$  (siendo  $H_1$  y  $H_2$  espacios de Hilbert) es un operador que supondremos compacto. Hemos visto que, en general, (16) no tiene solución única. Supondremos que  $y \in D(K^\dagger)$ , y nuestro propósito será aproximar  $K^\dagger y$ . Sabemos que, ignorando el caso en el que  $k(s, t)$  es degenerado,  $K^\dagger y$  depende de manera discontinua de  $y$ . Sin embargo, quisiéramos que nuestras aproximaciones lo hicieran de manera continua. Es decir, cambiaremos un problema mal condicionado en la búsqueda de una solución exacta, por uno bien condicionado en la búsqueda de una solución aproximada.

Recordemos que la solución generalizada  $x = K^\dagger y$  de (16) de mínimos cuadrados satisface

$$K^* Kx = K^* y, \quad (17)$$

siendo  $K^*$  el operador adjunto de  $K$ . Sabemos además que el operador  $K^* K$  tiene autovalores reales no negativos  $\{\lambda_j\}$ , y, por ende, para todo  $\alpha \in \mathbb{R}$ , el operador  $K^* K + \alpha I$  tiene autovalores estrictamente positivos, gracias al Teorema del Mapeo Espectral. En particular, este operador tiene una inversa acotada. Es decir, el problema

$$(K^* K + \alpha I) x_\alpha = K^* y. \quad (18)$$

es bien condicionado. Llamamos a (18) la forma *regularizada* de la ecuación (17), y su única solución

$$x_\alpha = (K^* K + \alpha I)^{-1} K^* y.$$

es llamada la *Aproximación de Tikhonov* para  $K^\dagger y$ . Al definir esa aproximación, sería importante saber si  $x_\alpha \rightarrow K^\dagger y$  cuando  $\alpha \rightarrow 0$ . Esto puede ser logrado de manera sencilla utilizando convenientemente la SVD  $\{v_j, u_j; \mu_j\}$  para  $K$  (Groetsch, 2014). Recuerde que:

- $\{v_j\}$  es un conjunto ortonormal completo para  $N(K)^\perp$
- $\{u_j\}$  es un conjunto ortonormal completo para  $\overline{Ran(K)}$ .
- $\mu_j \rightarrow 0$ ,  $Kv_j = \mu_j u_j$  y  $K^* u_j = \mu_j v_j$ .

De (18), tenemos que  $\alpha x_\alpha = K^* y - K^* Kx_\alpha$ , y <sup>21</sup>

$$x_\alpha \in Ran(K^*) \subseteq N(K)^\perp$$

Por ende, podemos expandir  $x_\alpha$  en términos de  $\{v_j\}$  como

$$x_\alpha = \sum_{j=1}^{\infty} \langle x_\alpha, v_j \rangle v_j.$$

De manera similar,

$$K^* y = \sum_{j=1}^{\infty} \langle K^* y, v_j \rangle v_j = \sum_{j=1}^{\infty} \langle y, Kv_j \rangle v_j = \sum_{j=1}^{\infty} \mu_j \langle y, u_j \rangle v_j$$

Además,

$$K^* Kx_\alpha = \sum_{j=1}^{\infty} \lambda_j \langle x_\alpha, v_j \rangle v_j = \sum_{j=1}^{\infty} \mu_j^2 \langle x_\alpha, v_j \rangle v_j.$$

Sustituyendo en (18)

$$(K^* K + \alpha I) x_\alpha = K^* y$$

$$\sum_{j=1}^{\infty} \lambda_j \langle x_\alpha, v_j \rangle v_j = \sum_{j=1}^{\infty} \mu_j^2 \langle x_\alpha, v_j \rangle v_j + \alpha \sum_{j=1}^{\infty} \langle x_\alpha, v_j \rangle v_j = \sum_{j=1}^{\infty} \mu_j \langle y, u_j \rangle v_j.$$

<sup>21</sup>Recuerde que  $N(K^*) = Ran(K)^\perp$  y  $N(T)^\perp = Ran(T)$ .

Y, por ende,

$$\sum_{j=1}^{\infty} (\mu_j^2 + \alpha) \langle x_\alpha, v_j \rangle v_j = \sum_{j=1}^{\infty} \mu_j \langle y, u_j \rangle v_j.$$

Es decir,

$$\langle x_\alpha, v_j \rangle = \frac{\mu_j}{\mu_j^2 + \alpha} \langle y, u_j \rangle.$$

Concluimos entonces que

$$x_\alpha = \sum_{j=1}^{\infty} \frac{\mu_j}{\mu_j^2 + \alpha} \langle y, u_j \rangle v_j. \quad (19)$$

Y, por lo visto en la anterior sección,

$$K^\dagger y = \sum_{j=1}^{\infty} \frac{\langle y, u_j \rangle}{\mu_j} v_j.$$

Entonces,

$$\|x_\alpha - K^\dagger y\|^2 = \sum_{j=1}^{\infty} \left( \frac{\mu_j}{\mu_j^2 + \alpha} - \frac{1}{\mu_j} \right)^2 \langle y, u_j \rangle^2 \quad (20)$$

$$= \sum_{j=1}^{\infty} \left( \frac{\alpha}{\mu_j(\mu_j^2 + \alpha)} \right)^2 \langle y, u_j \rangle^2. \quad (21)$$

Note que, si tomamos el límite cuando  $\alpha$  tiende a 0 a ambos lados de la igualdad, obtenemos lo requerido. No obstante, ello se cumple siempre y cuando podamos intercambiar el límite con la suma infinita de la derecha. Ello se puede realizar, debido a que:

$$\left( \frac{\alpha}{\mu_j(\mu_j^2 + \alpha)} \right)^2 \langle y, u_j \rangle^2 \leq \frac{1}{\mu_j^2} \langle y, u_j \rangle^2$$

y

$$\sum_{j=1}^{\infty} \frac{1}{\mu_j^2} \langle y, u_j \rangle^2 = \|K^\dagger y\|^2 < \infty$$

Hemos demostrado entonces que

$$\lim_{\alpha \rightarrow 0} \|x_\alpha - K^\dagger y\|^2 = 0 \quad (22)$$

Y, por lo tanto, nuestras aproximaciones  $x_\alpha$  tienden hacia nuestra solución de mínimos cuadrados. Adicional a ello, note que  $x_\alpha$  depende de manera continua de  $y$  para cada  $\alpha > 0$  fijo (el problema inverso asociado a resolver  $x_\alpha$  está bien condicionado). El valor  $\alpha$  es comunmente denominado *parámetro de regularización*.

Con lo visto hasta el momento, pareciera que entre menor sea nuestro  $\alpha$  en la aproximación de nuestra ecuación original 15, mejor será nuestra aproximación. No obstante, desde el principio de este trabajo se remarcó que, en la práctica, las mediciones que se realizan no son exactas y que, en vez de tener el valor exacto de  $y$  en (16), medimos  $y^\delta$ , que satisface:

$$\|y - y^\delta\| \leq \delta, \quad (23)$$

donde  $\delta$  es una cota conocida para el error de medición. En este caso, usando la misma aproximación que en (18), estamos midiendo

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

Sabemos que  $x_\alpha \rightarrow K^\dagger y$  cuando medimos datos "limpios". Veamos qué sucede con  $\{x_\alpha^\delta\}$ . Para ello, basta con ver su comportamiento con respecto a  $x_\alpha$ . Note que  $(K^* K + \alpha I)^{-1} K^* = K^* (K^* K + \alpha I)^{-1}$ , y

$$\begin{aligned} \|x_\alpha - x_\alpha^\delta\| &= \langle K^* (K^* K + \alpha I)^{-1} (y^\delta - y), K^* (K^* K + \alpha I)^{-1} (y^\delta - y) \rangle \\ &= \langle (K^* K + \alpha I)^{-1} (y^\delta - y), K K^* (K^* K + \alpha I)^{-1} (y^\delta - y) \rangle. \end{aligned}$$

Como tanto  $KK^*(K^*K + \alpha I)^{-1}$  como  $(K^*K + \alpha I)^{-1}$  son autoadjuntos, por el Teorema de Mapeo Espectral sabemos que

$$\begin{aligned}
\|KK^*(K^*K + \alpha I)^{-1}\| &= \max\{|\theta| : \theta \in \sigma(KK^*(K^*K + \alpha I)^{-1})\} \\
&= \max\left\{\left|\frac{\lambda}{\lambda + \alpha}\right| : \lambda \in \sigma(KK^*)\right\} \\
&\leq \max\left\{\left|\frac{\lambda}{\lambda}\right| : \lambda \in \sigma(KK^*)\right\} \quad (\alpha > 0 \text{ y } \lambda > 0 \text{ para todo } \lambda \in \sigma(KK^*)) \\
&\leq 1.
\end{aligned}$$

y, de manera similar,

$$\begin{aligned}
\|(K^*K + \alpha I)^{-1}\| &= \max\{|\theta| : \theta \in \sigma((K^*K + \alpha I)^{-1})\} \\
&= \max\left\{\left|\frac{1}{\lambda + \alpha}\right| : \lambda \in \sigma(K^*K)\right\} \\
&\leq \frac{1}{\alpha}.
\end{aligned}$$

Es decir,

$$\|x_\alpha - x_\alpha^\delta\|^2 \leq \frac{\delta^2}{\alpha}$$

y,

$$\|x_\alpha - x_\alpha^\delta\| \leq \frac{\delta}{\sqrt{\alpha}} \quad (24)$$

En las ecuaciones (22) y (24) vemos condensada la esencia del estudio de problemas inversos y la regularización de Tikhonov. Ilustran un dilema clásico para los problemas mal condicionados: entre más cercano escojamos nuestro parámetro de regularización a cero, obtendremos teóricamente una mejor aproximación, que sin embargo será más inestable respecto a ligeros cambios en nuestra medición.

Escoger un parámetro de regularización adecuado se convierte entonces en nuestro objetivo principal. En el sentido de Tikhonov, una selección  $\alpha = \alpha(\delta)$  nos lleva a un algoritmo *regular* si

$$\alpha(\delta) \rightarrow 0 \text{ y } x_\alpha(\delta)^\delta \rightarrow K^\dagger y \text{ cuando } \delta \rightarrow 0$$

Note que:

$$\|x_{\alpha(\delta)}^\delta - K^\dagger y\| \leq \|x_\alpha^\delta - x_{\alpha(\delta)}\| + \|x_{\alpha(\delta)} - K^\dagger y\| \leq \frac{\delta}{\sqrt{\alpha(\delta)}} + \|x_{\alpha(\delta)} - K^\dagger y\|,$$

y, como ya mostramos que  $x_\alpha(\delta) \rightarrow K^\dagger y$  cuando  $\alpha(\delta) \rightarrow 0$ , vemos que la condición

$$\frac{\delta^2}{\alpha(\delta)} \rightarrow 0 \text{ con } \delta \rightarrow 0$$

es suficiente para garantizar que  $\alpha(\delta)$  resulta en un algoritmo regular para resolver nuestro problema inverso (Groetsch, 2014).

Muchos métodos para escoger  $\alpha(\delta)$  han sido propuestos en la literatura. Uno de los más reconocidos es denominado el *Principio de Discrepancia de Morozov*: una selección *a posteriori* de nuestro parámetro  $\alpha$  que satisface

$$\|Kx_\alpha - y^\delta\| = \delta \quad (25)$$

Es decir, la solución regularizada no debería intentar satisfacer los datos originales de manera más precisa que el error (o “ruido”) en la medición. Asumiendo  $\|y^\delta\| > \delta$  y  $y \in \text{Ran}(K)$ , no es difícil ver que, en efecto, existe un único  $\alpha$  que satisface (25). En efecto, de manera muy similar a como se llegó a (21):

$$\|Kx_\alpha^\delta - y^\delta\|^2 = \sum_{j=1}^{\infty} \left( \frac{\alpha}{\mu_j(\mu_j^2 + \alpha)} \right)^2 \langle y^\delta, u_j \rangle^2 + \|Py^\delta\|^2,$$

donde  $P$  es la proyección ortogonal de  $H_2$  en  $\text{Ran}(K)^\perp$ , y, si definimos

$$f(\alpha) = \|Kx_\alpha^\delta - y^\delta\|,$$



ella es una función estrictamente creciente y continua. Como  $Py = 0$ , ella satisface:

$$\lim_{\alpha \rightarrow 0^+} f(\alpha) = \|Py^\delta - Py\| \leq \|y^\delta - y\| \leq \delta$$

y

$$\lim_{\alpha \rightarrow \infty} f(\alpha) = \|y^\delta\| > \delta$$

Y, por el Teorema del Valor Intermedio, concluimos que existe un único  $\alpha = \alpha(\delta)$  que satisface (25). Una demostración de que una selección del parámetro basada en el Principio de Discrepancia de Morozov realmente conduce a un algoritmo regular en el sentido de Tikhonov puede ser vista en [1].

**Nota:** La suposición de que  $\|Py^\delta - Py\| \leq \|y^\delta - y\|$  es natural, en el sentido de que cualquier componente que en  $y^\delta$  que sea ortogonal a  $Ran(K)$  existe debido al ruido en la medición. Por otro lado, la condición de que  $\|y^\delta\| > \delta$  puede ser entendida como que el error no debería exceder el nivel de la señal. En efecto, si se diera que  $\|y^\delta\| < \delta$ , podríamos argumentar que, por el principio de Discrepancia,  $x = 0$  es una solución aceptable (Kaipio & Somersalo, 2005).

### Interpretación variacional de la regularización de Tikhonov.

En muchos textos, como en [4], se presenta la regularización de Tikhonov como un problema de minimización. A saber, definimos  $x_\alpha$  como el valor que minimiza el funcional  $F_\alpha$ , dado por

$$F_\alpha(x) = \|Kx - y^\delta\|^2 + \alpha\|x\|^2. \quad (26)$$

Note que el primer término de la suma corresponde a qué tan cercano está nuestro  $x$  de una solución de mínimos cuadrados, mientras el segundo término castiga grandes inestabilidades. Para ver que  $x_\alpha$  coincide con la definición dada anteriormente, podemos definir la norma  $|\cdot, \cdot|$  sobre  $H^1 \times H^2$  dada por

$$|(u, v)|^2 = \|v\|^2 + \alpha\|u\|^2.$$

En este caso, (26) mide la distancia (al cuadrado) de  $(x, Kx)$  a  $(0, y^\delta)$ . Al conjunto  $\{(x, Kx) \in H_1 \times H_2 : x \in H_1\}$  se le denomina el *grafo* de  $K$ . Si suponemos al operador  $K$  compacto, resulta que el grafo de  $K$  también lo es, y, por ende, existe un valor allí que minimiza  $F_\alpha$ .

Un valor  $z$  que minimice (26) debe satisfacer

$$\left. \frac{d}{dt} \left\{ \|K(z + tw) - y^\delta\|^2 + \alpha\|z + tw\|^2 \right\} \right|_{t=0} = 0. \quad (27)$$

para todo  $w \in H_1$ . Note que, para  $x, y \in H_1$ ,

$$\|x + y\|^2 = \|x\|^2 + 2\langle x, y \rangle + \|y\|^2.$$

Y, por lo tanto, (27) puede ser vista como

$$\left. \frac{dt}{t} \left\{ \|Kz - y^\delta\|^2 + t^2\|Kw\|^2 + \alpha\|z\|^2 + t^2\|w\|^2 + 2t(\langle Kz - y^\delta, Kw \rangle + \alpha\langle z, w \rangle) \right\} \right|_{t=0} = 2(\langle Kz - y^\delta, Kw \rangle + \alpha\langle z, w \rangle) = 0.$$

Es decir,

$$\begin{aligned} \langle Kz - y^\delta, Kw \rangle + \alpha\langle z, w \rangle &= 0 \\ \langle K^*Kz - K^*y^\delta, w \rangle + \langle \alpha z, w \rangle &= 0 \\ \langle (K^*K + \alpha I)z - K^*y^\delta, w \rangle &= 0, \end{aligned}$$

para todo  $w \in H_1$ . Por ende,

$$(K^*K + \alpha I)z - K^*y^\delta = 0.$$

Entonces, el único argumento que minimiza  $F_\alpha$  está dado por:

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

que coincide con la definición dada al principio de la sección. Este enfoque para la definición de  $x_\alpha$  tiene importantes implicaciones, tanto teóricas como prácticas[1].

### Implementaciones numéricas

Se hace la regularización de Tikhonov sobre el problema presentado al principio de este capítulo. Se presentarán dos enfoques. En el primero, se realiza una discretización del problema, para después aplicar regularización. Por otra parte, en el segundo enfoque primero realizamos regularización, para posteriormente discretizar. El resultado de aplicar estos dos enfoques al ejemplo principal de este capítulo se puede evidenciar en la gráficas presentadas en la Figura 4

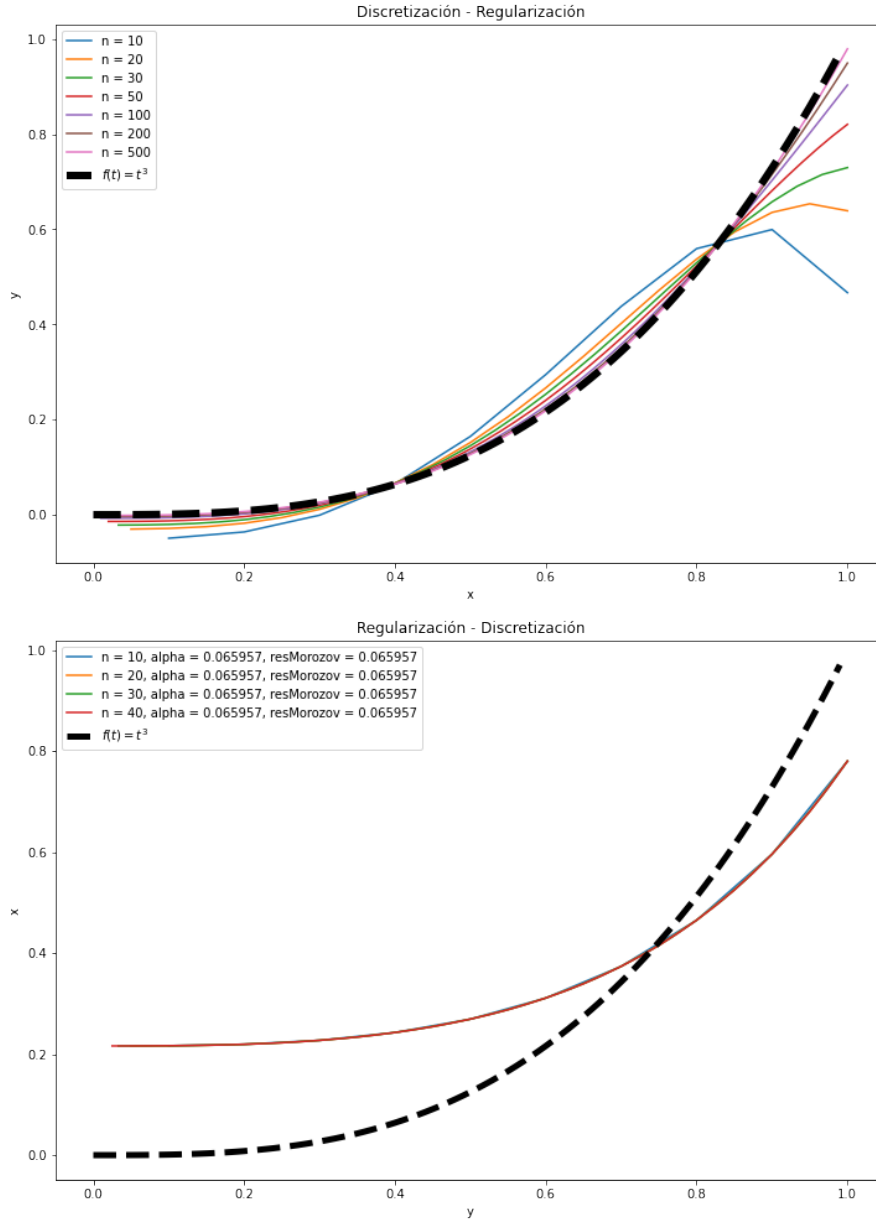


Figura 4: Realización propia. Métodos de regularización. La gráfica superior es la aplicación del primer enfoque al problema (8), el tamaño de la partición  $n$  para la discretización se toma en el conjunto  $\{10, 20, 30, 50, 100, 200, 500\}$ . En la figura inferior se presenta el resultado de ejecutar el segundo enfoque en el problema (8); la leyenda de esta gráfica indica el tamaño de partición  $n$  del núcleo, el parámetro  $\alpha$  óptimo dado un  $\delta$  y el residuo de Morozov bajo los parámetros  $\delta$  y  $\alpha$ . En ambos métodos se toma el rango del ruido en el intervalo  $[-10^4, 10^4]$ , intervalo controlado por el parámetro  $\delta$ .

El primer enfoque se realiza una discretización del problema como se describió en la sección 4.1.1, luego, para evitar el

mal condicionamiento de la matriz asociada al sistema finito aplicamos regularización de Tikhonov aleatorizando el sistema para mejorar el error que se supone que poseen los datos tomados en la discretización. Para aproximarse a satisfacer el Principio de Discrepancia de Morozov se realiza una optimización sobre el parámetro  $\alpha$  de tal forma que la expresión

$$|||Kx_\alpha - y^\delta|| - \delta|$$

se vuelva lo más pequeña posible conociendo en principio el valor del parámetro  $\delta$ . Este enfoque presenta un acercamiento notable a la solución a medida que el tamaño de la partición aumenta, en comparación con la gráfica obtenida cuando únicamente se aplica discretización y se resuelve sin pasar por un proceso de regularización.

Por otra parte, en el segundo enfoque primero se regulariza el problema obteniendo un ecuación de Fredholm de segundo grado; para poder aplicar el método de núcleo degenerado lo que hacemos es discretizar el kernel del término conocido de la ecuación, el proceso en detalle será mostrado a continuación.

El problema a resolver es

$$y(s) = \int_a^b K(s, t)x(t)dt = Kx(s)$$

A esta integral aplicamos regularización de Tikhonov, obteniendo el siguiente sistema

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

donde  $K^*Kx(s) = \int_a^b \check{K}(s, t)x(t)dt = \int_a^b (\int_a^b K(u, s)K(u, t)du)x(t)dt$ , esta igualdad se tiene gracias al siguiente resultado.

**Teorema 4.2.1.** *Si  $K$  es una función de valor real, entonces*

$$K^*Kx(s) = \int_a^b \check{K}(s, t)x(t)dt = \int_a^b \left( \int_a^b K(u, s)K(u, t)du \right) x(t)dt$$

#### **Demostración.**

- En primer lugar veamos como es la forma del operador  $K^*x$ .

$$\begin{aligned} \langle Kx, y \rangle &= \int_a^b Kx(s)y(s)ds \\ &= \int_a^b \left( \int_a^b K(s, t)x(t)dt \right) y(s)ds \\ &= \int_a^b \left( \int_a^b K(s, t)x(t)y(s)ds \right) dt \\ &= \int_a^b x(t) \left( \int_a^b K(s, t)y(s)ds \right) dt \\ &= \langle x, K^*y \rangle, \end{aligned}$$

por lo tanto  $K^*x(t) = \int_a^b K(s, t)y(s)ds$ .

- En segundo lugar, veamos la forma del operador  $K^*K$ ,

$$\begin{aligned} K^*Kx(s) &= \int_a^b K(u, s)Kx(u)du \\ &= \int_a^b K(u, s) \left( \int_a^b K(u, t)x(t)dt \right) du \\ &= \int_a^b \left( \int_a^b K(u, s)K(u, t)du \right) x(t)dt, \end{aligned}$$

definiendo  $\check{K}(s, t) := \int_a^b K(u, s)K(u, t)du$ , obtenemos la igualdad buscada.

Al nuevo kernel  $\check{K}$  obtenido luego de hacer regularización lo representamos como una sumatoria de Riemann finita, para que así nos quede un kernel degenerado, así

$$\check{K}(s, t) \approx \sum_{i=1}^n K(u_i, s) K(u_i, t) \frac{1}{n} \quad \text{donde } u_i \text{ pertenece a una partición de } [a, b].$$

De esta forma obtenemos la ecuación

$$\alpha x_\alpha^\delta = K^* y^\delta - \frac{1}{\alpha} K^* K(\alpha x_\alpha^\delta)$$

luego, aplicando el método de núcleo degenerado a la anterior ecuación obtenemos la función  $\alpha x_\alpha^\delta$ , para finalmente dividir por el parámetro  $\alpha$  para obtener la solución al problema inicial.

En la implementación de este enfoque también se busca utilizar el valor relacionado con el Principio de Discrepancia de Morozov para obtener un ajuste óptimo a la solución. La gráfica presenta la solución hallada por este método para diferentes tamaños de discretización del kernel y valores de  $\alpha$  que optimizan el residuo de Morozov dado  $\delta = 10^{-5}$ . En comparación con el método anterior, el costo computacional aumenta y la proximidad a la solución teórica disminuye. El algoritmo en este caso presenta una cantidad lo suficientemente grande de cálculo de integrales numéricamente<sup>22</sup>, ocasionando, por ejemplo, que el proceso de optimización tome un tiempo más grande en comparación con la optimización realizada en el primer enfoque.

### 4.3. Métodos iterativos

Si consideramos una ecuación matricial  $Ax = y$ , el análisis numérico presenta una gran variedad de métodos iterativos para resolver esta ecuación. Resulta que algunos de estos métodos resultan útiles para la resolución de problemas inversos, y terminan siendo herramientas muy usadas en este campo. Por el nivel de complejidad de estos métodos, desarrollaremos el más simple de todos: la *iteración de Landweber* (también conocida como la *iteración de Landweber-Fridman*).

Suponga que  $K$  es un operador lineal compacto, y suponga que  $y \in D(K^\dagger)$ . Recuerde que la solución generalizada de un problema inverso  $Kx = u$  es el único  $x \in \text{Ran}(K^\perp)$  que satisface

$$K^* Kx = K^* y. \quad (28)$$

O, de otra forma,

$$x = x + \beta(K^* y - K^* Kx) = T(x),$$

donde  $\beta \in \mathbb{R}$  es un parámetro que dentro de poco estudiaremos. Ello sugiere un método iterativo de punto fijo sobre la función  $T(x)$ , dado por

$$x_{n+1} = x_n + \beta(K^* y - K^* Kx_n). \quad (29)$$

Note que si tomamos  $x_0 = 0$ , luego  $x_n \in \text{Ran}(K) \subseteq N(K)^\perp$  para todo  $n \in \mathbb{N}$ , y, por ende, si  $x_n \rightarrow x$ , luego  $x \in N(K)^\perp$  (al ser  $N(K)^\perp$  cerrado), y satisface 28. Es decir,  $x = K^\dagger y$ . El siguiente resultado se encuentra en [4], e ilustra qué valores de  $\beta$  se deben tomar para garantizar que nuestra sucesión  $x_n$  converge.

**Teorema 4.3.1.** *Sea  $K : H_1 \rightarrow H_2$  un operador lineal compacto. Dado un problema inverso  $Kx = y$ , tome en 29  $0 < \beta < 2/\lambda$ , donde  $\lambda$  es el mayor autovalor de  $K^* K$ . Luego, la iteración de punto de fijo  $(x_n)$  dada en 29 converge a  $x \in N(K)^\perp$ , tomando  $x_0 = 0$ .*

**Demostración.** Si tomamos  $S = N(K)^\perp = \overline{\text{Ran}(K^*)}$ , es claro que  $S$  es invariante bajo  $T$ . Es decir,  $T(S) \subseteq S$ . Veamos ahora que  $T$  representa una contracción sobre  $S$ . Es decir que existe  $0 \leq \kappa < 1$  tal que, para cualesquiera  $x, z \in S$ ,

$$\|T(x) - T(z)\| < \kappa \|x - z\|.$$

En efecto, si  $(v_n, u_n, \lambda_n)$  es la *SVD* de  $K$ , tenemos que

$$\begin{aligned} \|T(x) - T(z)\| &= \|(1 - \beta K^* K)(x - z)\|^2 \\ &= \sum_{j=1}^{\infty} (1 - \beta \lambda_j)^2 |\langle x - z, v_j \rangle|^2 \\ &\leq \kappa^2 \|x - z\|^2, \end{aligned}$$

<sup>22</sup>A lo largo de los algoritmos anexos a este documento se trabaja con métodos de integración numérica como Simpson 1/3, trapecio y cuadratura.

donde

$$\kappa^2 = \max_{1 \leq j \leq N} (1 - \beta \lambda_j)^2.$$

Note que  $\kappa < 1$ , dado que  $0 < \beta \lambda_j < 2$  para cualquier  $1 \leq j \leq N$ . □

Definamos ahora

$$e_n = x_n - x.$$

Note que, por la ecuación (29), tenemos que

$$e_n + 1 = (I - BK^*K).$$

Por consiguiente,

$$e_n = (I - BK^*K)^n e_0.$$

Como  $e_0 = -x \in N(K)^\perp$ , podemos expandir  $\|e_0\|$  en términos de la *SVD* de  $K$  y, usando la notación del Teorema 4.3.1,

$$\|e_n\|^2 = \sum_{j=1}^{\infty} (1 - \beta \lambda_j)^{2n} |\langle e_j, v_j \rangle|^2. \quad (30)$$

Dado que los autovalores  $\lambda_j$  de  $K^*K$  convergen a cero, a medida que aumenta  $j$  en la sumatoria, los factores  $(1 - \beta \lambda_j)^{2n}$  son cada vez más cercanos a 1. Por lo tanto, en las primeras iteraciones los cambios visibles se darán gracias a los primeros autovectores  $v_j$  de la *SVD* (debido a que, para el resto, los cambios respecto a la suma correspondiente a  $\|e_0\|^2$  son muy pequeños). Entonces, se requiere de bastantes iteraciones para lograr visualizar el efecto de autovectores de un orden grande en nuestra iteración de Landweber, y ello limita su aplicabilidad (Vogel, 2002). Veamos ahora cómo afecta una perturbación en  $y$  a nuestro método iterativo. Tome  $y^\delta$  que satisface  $\|y - y^\delta\| \leq \delta$  para algún  $\delta > 0$ . Resolvemos ahora nuestro método iterativo, tomando  $x_0 = 1$ , y

$$x_{n+1}^\delta = x_n^\delta + \beta(K^*y^\delta - K^*Kx_n^\delta). \quad (31)$$

Como en la anterior sección, comparemos la sucesión  $(x_n^\delta)$  con  $(x_n)$ . Definimos

$$d_n^\delta = x_n^\delta - x_n.$$

Tenemos entonces que  $d_n^\delta = 0$  y, por (29) y (31)

$$\begin{aligned} d_{n+1}^\delta &= x_{n+1}^\delta - x_{n+1} \\ &= x_n^\delta + \beta(K^*y^\delta - K^*Kx_n^\delta) - x_n - \beta(K^*y - K^*Kx_n) \\ &= (x_n^\delta - x_n) - \beta K^*K(x_n^{\delta\text{delta}} - x_n) + \beta K^*(y^\delta - y) \\ &= (I - \beta K^*K)d_n^\delta + \beta K^*(y^\delta - y). \end{aligned}$$

Por la selección de nuestro  $\beta$ , es claro que  $\|(I - \beta K^*K)\| \leq 1$ , y, por ende,

$$\|\delta\| \leq \|d_n^\delta\| + \beta\|K\|y^{23}.$$

Concluimos entonces que

$$\|d_n^\delta\| \leq n\beta\|K\|\delta.$$

Y, por lo tanto,

$$\begin{aligned} \|x_n^\delta - K^\dagger y\| &\leq \|x_n^\delta - x_n\| + \|x_n - K^\dagger y\| \\ &\leq \|x_n^\delta - x_n\| + O(n\delta). \end{aligned}$$

Pero, como ya mostramos que  $x_n \rightarrow K^\dagger y$ , obtenemos que una condición suficiente de regularidad para nuestro  $x_n^\delta$  es que el número de iteraciones  $n = n(\delta)$  satisfaga que  $n\delta \rightarrow 0$  con  $\delta \rightarrow 0$ .

Podemos entonces notar que, cuando presentamos una perturbación en la observación  $y$ , un mayor número de iteraciones  $n$  no garantiza que se obtenga una mejor aproximación a la observación original. En este caso, el número de iteraciones  $n$  toma el papel de parámetro de regularización (Groetsch, 2014). Después de cierto número de iteraciones, continuar con el algoritmo produce vectores cada vez más oscilantes, que se alejan cada vez más de la solución. A este fenómeno lo denominamos *semiconvergencia* (Vogel, 2002).

Algunas selecciones del número de iteraciones  $n$  que se deben realizar han sido propuestas en la literatura. Por ejemplo, se puede imitar el método de discrepancia de Mozorov, buscando  $n$  tal que

$$\|Kx_n^\delta - y^\delta\| = \delta.$$

Otra opción que surge es tomar  $n = \lceil \delta^{-1/2} \rceil$ . En este caso, note que, gracias a lo desarrollado anteriormente,

$$\|x_n^\delta - K^\dagger y\| \leq \|x_n^\delta - x_n\| + O(\sqrt{\delta}).$$

Continuando con el ejemplo principal tratado en la implementación de los anteriores métodos, la iteración de Landweber es un método sencillo de implementar y que en este caso en particular parece aproximar la solución teórica de manera óptima, como se evidencia en la figura 5. Para la implementación algorítmica solo basta tener en cuenta la iteración, un número máximo de iteraciones y un criterio de parada del algoritmo. Para escoger el parámetro  $\beta$  que es necesario en la iteración tomamos un valor pseudoaleatorio en el rango  $\left(0, \frac{2}{\rho(A)}\right)$  donde  $A$  es la matriz obtenida en la discretización y  $\rho(\cdot)$  es el radio espectral de la matriz.

Según lo visto en la gráfica de la Figura 5 mientras aumentamos el tamaño de la partición  $n$  en la discretización el resultado del algoritmo parece acercarse a la solución esperada de manera *lenta* como se mencionó anteriormente. En una sección posterior se mostrará el comportamiento de esta aproximación de forma numérica.

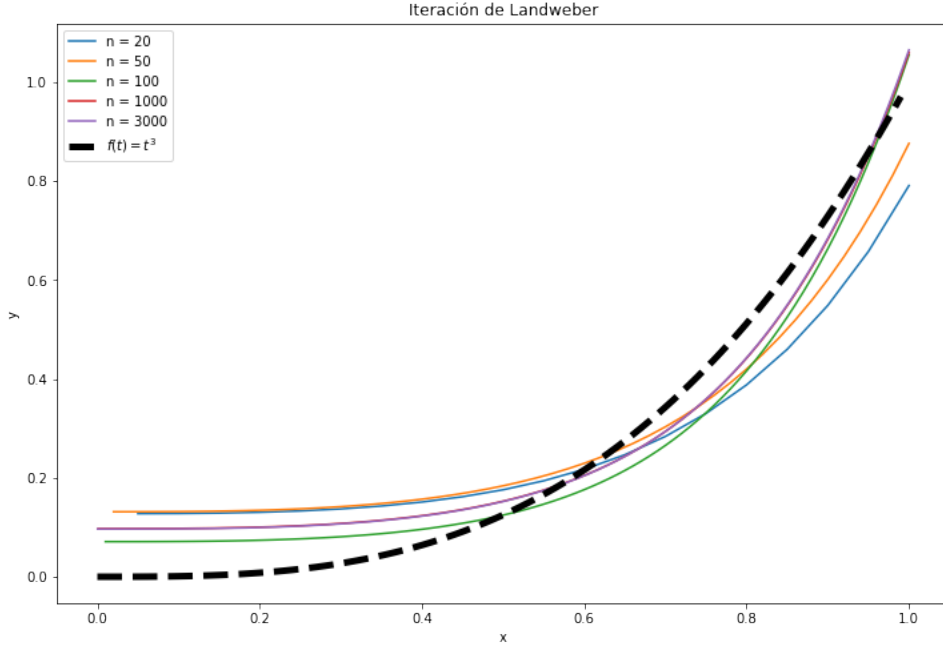


Figura 5: Realización propia. Iteración de Landweber aplicada a (8). El tamaño de la partición  $n$  en la discretización es tomado del conjunto  $\{20, 50, 100, 1000, 3000\}$ . Para los demás parámetros se toma  $\delta = 10^{-4}$  que controla el ruido blanco añadido, el máximo número de iteraciones se fijó en 100000, el criterio de parada fue la norma infinito de la expresión  $Ax - b$  que surge de la discretización y se propuso que este valor fuese menor que  $\epsilon = 0,1$ .

#### 4.4. TSVD

En la anterior sección, vimos que nuestra solución  $K^\dagger y$  de un problema inverso dado por la ecuación  $Kx = y$ , tiene una descomposición en valores singulares (*SVD*)

$$K^\dagger y = \sum_{j=1}^{\infty} \frac{\langle y, u_j \rangle}{\mu_j} v_j, \quad (32)$$

donde  $\{u_j\}$  y  $\{v_j\}$  son conjuntos ortonormales completos para  $\overline{\text{Ran}(K)}$  y  $N(K)^\perp$ , respectivamente. Además,  $\{\mu_j^2\}$  es el conjunto decreciente de autovalores del operador  $K^*K$  con autovectores asociados  $\{v_j\}$ ; y a la vez son los autovalores de  $KK^*$ , con autovectores asociados  $\{u_j\}$ . Ellos satisfacen

$$Kv_j = \mu_j u_j \text{ y } Ku_j = \mu_j v_j$$

Una idea natural sería truncar esta suma hasta cierto valor, para obtener una aproximación de la forma

$$x_n = \sum_{j=1}^n \frac{\langle y, u_j \rangle}{\mu_j},$$

donde  $x_n \rightarrow K^\dagger y$  cuando  $n \rightarrow \infty$ . A esta aproximación la llamamos el método de *la descomposición en valores singulares truncada* (o *TSVD*, por sus siglas en inglés). Como es de esperarse, a medida que aumenta nuestro valor  $n$ , nos acercamos cada vez más a un problema continuo mal condicionado, de manera que nuestras soluciones comienzan a diverger si presentamos errores de medición. En efecto, tomemos de nuevo  $y^\delta$  que satisface  $\|y - y^\delta\| \leq \delta$ . Tenemos que, si definimos

$$x_n^\delta = \sum_{j=1}^n \frac{\langle y^\delta, u_j \rangle}{\mu_j},$$

entonces

$$x_n - x_n^\delta = \sum_{j=1}^n \frac{\langle y - y^\delta, u_j \rangle}{\mu_j}.$$

Si queremos analizar el error de estabilidad, obtenemos que

$$\begin{aligned} \|x_n - x_n^\delta\|^2 &= \left\langle \sum_{j=1}^n \langle y - y^\delta, u_j \rangle \mu_j^{-1}, \sum_{j=1}^n \langle y - y^\delta, u_j \rangle \mu_j^{-1} \right\rangle \\ &= \sum_{j=1}^n \langle y - y^\delta, u_j \rangle^2 \mu_j^{-2} \\ &\leq \mu_n^{-2} \sum_{j=1}^n \langle y - y^\delta, u_j \rangle^2 \\ &\leq \mu_n^{-2} \|y - y^\delta\|^2 \leq \delta^2 \mu_n^{-2}. \end{aligned}$$

Por ende, la estabilidad de nuestra solución al truncar la suma (32) depende del tamaño del valor singular  $\mu_n$ . Entre más pequeño sea este valor, más inestable será nuestra solución. A saber,

$$\|x_n - K^\dagger y\| \leq \|x_n - x_n^\delta\| + \delta \mu_n. \quad (33)$$

Por consiguiente, si  $n = n(\delta)$  es seleccionado de tal forma que  $\delta \mu_n^{-1} \rightarrow 0$  cuando  $\delta \rightarrow 0$ , luego  $x_n \rightarrow K^\dagger y$ . Es decir, el método de la descomposición en valores singulares truncada es un método de regularización (Groetsch, 2014). De nuevo, la selección del parámetro  $n$  a utilizar es de vital importancia. En [4] recomienda, de manera similar al resto de métodos, tomar  $n$  tal que

$$\|Kx_n - y^\delta\| \approx \delta$$

### Breve análisis para la discretización:

En la práctica, como ya se vió en la Sección 4.1, es necesario discretizar un problema inverso  $Kx = y$ , y trabajar con un sistema de ecuaciones lineales  $Ax = b$ , en donde  $A_{m \times n}$  es una matriz peor condicionada a medida que aumenta en dimensión. En la anterior sección vimos que  $A$  tiene una descomposición en valores singulares

$$A = U \Sigma V^*,$$

donde  $U_{m \times m}$  y  $V_{n \times n}$  son matrices unitarias, y  $\Sigma_{m \times n}$  es una matriz diagonal no negativa, en la que los valores de la diagonal  $\{\sigma_i\}$  son dados en orden decreciente. Tomemos  $p$  siendo el mayor índice para el cual se satisface  $\sigma_p > 0$ . Definimos la *seudoinversa* o *inversa generalizada de Moore-Penrose* de  $A$ , como la matriz

$$A^\dagger := V \Sigma^\dagger U^*,$$

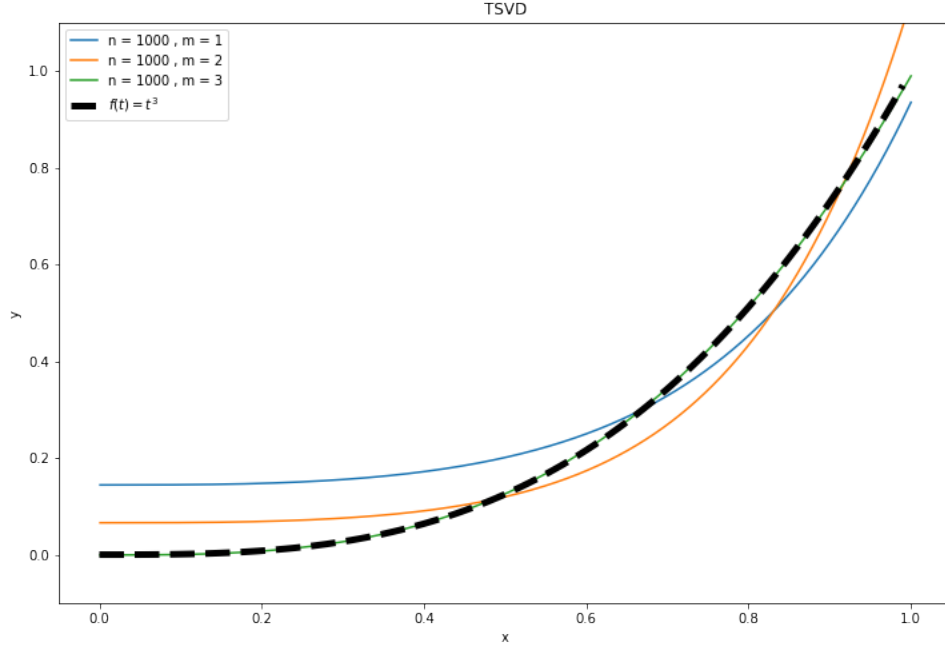


Figura 6: Realización propia. Método TSVD aplicado al problema (8). En la leyenda de la gráfica,  $n$  representa el tamaño de discretización y  $m$  es el valor donde se trunca la serie. El ruido blanco añadido a este método está en el intervalo  $[-10^4, 10^4]$ .

donde

$$\Sigma^\dagger = \begin{pmatrix} 1/\sigma_1 & 0 & \cdots & \cdots & 0 \\ 0 & 1/\sigma_2 & 0 & \cdots & 0 \\ & & \ddots & & \\ \vdots & & & 1/\sigma_p & \vdots \\ & & & 0 & \\ 0 & & \cdots & & \ddots & 0 \end{pmatrix}.$$

Llamamos a  $A^\dagger x$  la solución de mínimos cuadrados del problema  $Ax = y$ . No obstante, si  $A$  está mal condicionada, esta expresión es poco útil, debido a que los valores  $\{\sigma_i\}_{1 \leq i \leq p}$  son cada vez más cercanos a cero, y, por ende, más sensibles a un ligero cambio en nuestra observación. Debemos entonces escoger nuestro índice de truncamiento como un valor  $k < p$  (Kaipio & Somersalo, 2005).

### Implementación numérica

El método TSVD es sencillo de implementar algorítmicamente conociendo como hacer la descomposición de una matriz en valores singulares. Para aplicar el método al problema (8), realizamos la discretización fijando cierto tamaño  $n$  para el tamaño de la partición, calculamos la forma  $SVD$  de la matriz de discretización y luego hacemos la serie truncada hasta otro valor  $m \leq n$ . En la Figura 6 se presenta el resultado de la ejecución.

Respecto al tamaño de discretización podemos asegurar que mientras aumenta este valor también aumenta el ajuste de la solución algorítmica a la solución teórica, sin embargo, algo para destacar en este método es la sensibilidad en el valor donde se trunca la serie puesto que después de cierto valor, la solución pierde validez porque toma valores demasiado lejanos de la solución esperada. Esto se puede evidenciar en la figura 7.

## 4.5. Comparación entre resultados en la implementación de métodos

A continuación realizamos la presentación de la comparación entre los métodos tratados en las anteriores secciones. Nuestros criterios de calificación de métodos van a ser dos: **Calidad de ajuste a la solución analítica** y **Costo computacional**.



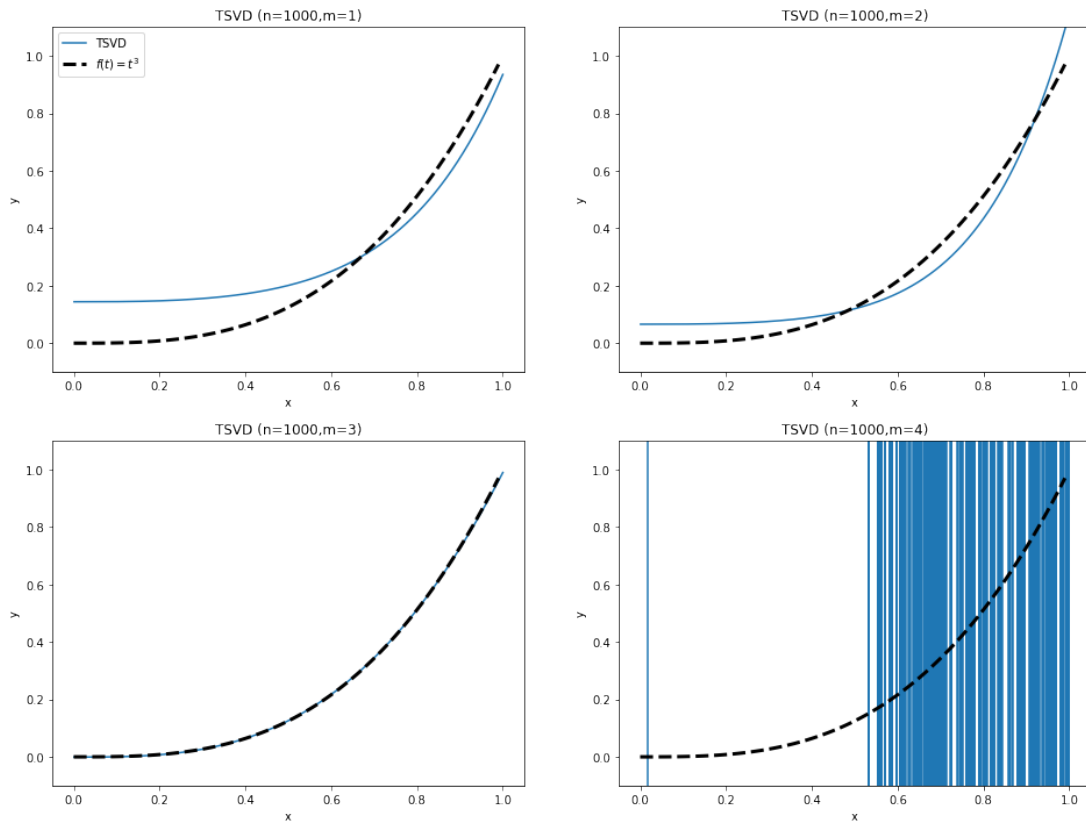


Figura 7: Realización propia. Método TSVD aplicado al problema expresado en la ecuación (8). Se presentan diferentes soluciones algorítmicas para valores distintos de  $m$ . Es notable la falta de ajuste cuando  $m = 4$ , contrario cuando  $m = 3$ .

$n$	Tiempo de ejecución(s)	Residuo( $\ \cdot\ _\infty$ )
10	0.433623	0.533233
20	1.04975	0.3611
30	0.951015	0.270379
50	1.47205	0.179465
100	3.34854	0.0968652
200	10.9744	0.0506083
500	91.6211	0.0207057

Cuadro 2: Discretización - Regularización (primer enfoque de regularización).

$n$	Tiempo de ejecución(s)	Residuo( $\ \cdot\ _\infty$ )
10	44.4944	0.216147
20	66.6651	0.216146
30	90.0799	0.216146
40	144.44	0.216146

Cuadro 3: Regularización - Discretización (segundo enfoque de regularización).

Los cuadros 2, 3 y 4 resumen la información sobre la aplicación de los métodos al problema representado en la ecuación (8). El tamaño de la partición de la discretización es  $n$  y, en el método TSVD  $m$  representa el valor de truncamiento de la serie. Como se mencionó en las secciones anteriores tenemos que los dos enfoques relacionados con el método de regularización son los que tardan más tiempo en ejecutarse en relación con el valor de  $n$ . Tenemos que

- En el primer enfoque de regularización mientras más aumenta el tamaño de la partición el residuo se vuelve menor esto con el costo de que el tiempo de ejecución aumenta.
- En el segundo enfoque de regularización a pesar de que el valor  $n$  aumenta, el residuo de las soluciones algorítmicas

$n$	Tiempo de ejecución(s)	Residuo( $\ \cdot\ _\infty$ )
20	0.433623	0.533233
50	1.04975	0.3611
100	0.951015	0.270379
1000	1.47205	0.179465
3000	3.34854	0.0968652

Cuadro 4: Iteración de Landweber.

$n$	$m$	Tiempo de ejecución(s)	Residuo( $\ \cdot\ _\infty$ )
10	1	0.00061655	0.32563
10	2	0.000407696	0.355271
10	3	0.000472784	0.533619
10	4	0.000358582	1.31508e+11
100	1	0.0168912	0.141029
100	2	0.0235028	0.0910022
100	3	0.0286801	0.0970359
100	4	0.0228519	2.33791e+10
1000	1	1.95932	0.144081
1000	2	1.82244	0.146261
1000	3	2.04617	0.0103106
1000	4	2.80026	2.3008e+09

Cuadro 5: TSVD.

permanece constante y alto en comparación con los residuos de los otros métodos. Tengase en cuenta que la imagen de la solución esperada está en el intervalo  $[0, 1]$  y un error mayor a 0,1 representa un desajuste considerable.

- Respecto a la iteración de Landweber, el tiempo de ejecución no varía mucho respecto al aumento de  $n$ , no obstante, es destacable que pese al aumento de  $n$  no hay una convergencia rápida de la iteración.
- Finalmente, en el método TSVD se tiene un buen ajuste cuando  $m \in \{1, 2, 3\}$ , no obstante, cuando  $m = 4$  el residuo supera el orden de  $10^9$ .

## 5. Conclusiones

En el presente trabajo, se pretendió dar una breve introducción a los problemas inversos, pasando por las principales dificultades a la hora de resolverlos, y analizando algunos métodos sencillos para sobrepasarlos. Desde el principio de la sección 4, se indicó que los métodos aquí estudiados no son los que se utilizan actualmente, aunque sí constituyen el corazón de la resolución de problemas inversos.

Existe una gama bastante amplia de métodos para resolver problemas inversos más allá de lo visto en este escrito, como el método de Máxima Entropía, o el de Backus-Gilbert, que el lector puede profundizar en [1]. Además, varios métodos iterativos bien conocidos en análisis numérico pueden ser usados en la resolución de problemas inversos, como lo son el método del Gradiente Conjugado, el método de Newton, o el método de Descenso más Pronunciado, que el lector puede consultar en [2].

Una vez terminado el estudio y el análisis de los métodos trabajados en este escrito, se propone continuar el tratamiento de soluciones generales de problemas inversos para ecuaciones que involucren dos o más variables, como sucede para el caso de análisis de imagen. Otra exploración posible en la cual se usan estos métodos es la estimación paramétrica y no paramétrica de funciones de distribución de variables aleatorias continuas.

Un enfoque para resolver problemas inversos que viene tomando fuerza en los últimos años es el estadístico o no determinista, en el que, en vez de intentar estimar los parámetros de un problema inverso, se asume que ellos son variables aleatorias, y se pretende determinar su densidad. Una de las tantas ventajas de este enfoque es que el ruido se puede ver también como una variable aleatoria con cierta función de densidad (Rodríguez, 2018).

## 6. Referencias

- [1] Groetsch, C. W. (2014). Inverse problems in the mathematical sciences. Vieweg & Teubner.
- [2] Vogel, C. R. (2002). Computational methods for inverse problems. Society for Industrial and Applied Mathematics.
- [3] Wing G. (1991) A primer on integral equations of the first kind. The problem of deconvolution and unfolding. Siam
- [4] Kaipio, J., & Somersalo, E. (2005). Statistical and computational inverse problems. Springer.
- [5] Lyche, T. (2020). Numerical linear algebra and matrix factorizations. Springer.
- [6] Kincaid D., Cheney W. (1991) Numerical Analysis Mathematics for Scientific Computing. Brooks/Cole Publishing Company.
- [7] Bertero, M., & Boccacci, P. (1998). Introduction to inverse problems in imaging. Institute of Physics Pub.
- [8] Groetsch, C. W. (1999). Inverse problems: activities for undergraduates. The Mathematical Association of America.
- [9] Riesz, F., & Sz.-Nagy B. (1955). Functional Analysis. Blackie & Son Limited.
- [10] Stein M., & Rami S. (2005). Real Analysis. Princeton.
- [11] Abuabara, T., & Lesmes, J. (2010). Elementos de Análisis Funcional. Ediciones Uniandes.
- [12] Rodriguez, O. (2018) Sobre la solución numérica de problemas inversos. Universidad Nacional de Colombia.
- [13] Brent. N. The completion of a metric space.