



Universidad de Sevilla  
Facultad de Matemáticas

TRABAJO FIN DE GRADO  
DOBLE GRADO EN FÍSICA Y MATEMÁTICAS

**Algunas técnicas de aprendizaje automático para  
problemas inversos  
ligados a ecuaciones diferenciales**

Jaime Bohoyo Barrado  
Sevilla, Julio 2023

**Tutora**

Anna Doubova

Dpto. Ecuaciones Diferenciales y Análisis Numérico



Most people, if you describe a train of events to them, will tell you what the result would be. They can put those events together in their minds and argue from them that something will come to pass. There are few people, however, who, if you told them a result, would be able to evolve from their own inner consciousness what the steps were which led up to that result. This power is what I mean when I talk of reasoning backward or analytically.

—Arthur Conan Doyle, *A Study in Scarlet*

## **Abstract**

The study of inverse problems is a branch of mathematics that has developed a lot in recent years. Deducing and recomposing the causes that have given rise to the phenomena we observe is common and necessary in our daily lives. Many inverse problems are related to differential equations. The main obstacle presented by solving this kind of problems is that they are ill-posed problems (it is not known if the problem has a unique solution, and if it does, it is usually unstable). It is not known either how to solve them analytically. This fact makes it necessary to use numerical analysis combined with machine learning tools to help us in this task. The analysis of inverse problems and machine learning algorithms as an alternative to solve them will be the subject of this Final Degree Project.

## Resumen

El estudio de los problemas inversos es una rama de las matemáticas que se ha desarrollado mucho en los últimos años. Deducir y recomponer las causas que han dado lugar a los fenómenos que observamos es común y necesario en nuestra vida cotidiana. Muchos problemas inversos están ligados a las ecuaciones diferenciales. El obstáculo principal que presenta la resolución de este tipo de problemas es que son problemas mal planteados (no se sabe si el problema tiene solución única, y en caso de tenerla suele ser inestable). Tampoco se sabe resolver de forma analítica dichos problemas. Esto hace necesario emplear herramientas de análisis numérico combinado con técnicas de aprendizaje automático (*machine learning*) para ayudarnos en esta tarea. El análisis de los problemas inversos y del aprendizaje automático como alternativa para resolverlos será el objeto de este Trabajo Fin de Grado.

# Índice

<b>Introducción</b>	<b>3</b>
<b>1. Conceptos previos</b>	<b>6</b>
<b>2. Ejemplos de problemas mal planteados</b>	<b>13</b>
<b>3. Regularización</b>	<b>16</b>
3.1. El teorema No Free Lunch . . . . .	17
3.2. El funcional de regularización de Tikhonov . . . . .	18
<b>4. Elección del parámetro de regularización</b>	<b>24</b>
4.1. Overfitting y underfitting . . . . .	24
4.2. Elección <i>a priori</i> . . . . .	26
4.3. Elección <i>a posteriori</i> . . . . .	26
4.3.1. Principio de discrepancia de Morozov . . . . .	28
4.3.2. Principio de balance . . . . .	31
<b>5. Aproximación numérica para problemas inversos</b>	<b>34</b>
5.1. El algoritmo de descenso del gradiente estocástico . . . . .	34
5.2. Método de Elementos Finitos aplicado a Problemas Inversos . . . . .	38
5.2.1. Fundamento teórico . . . . .	38
5.2.2. Aplicación del método de elementos finitos . . . . .	39
<b>6. Un problema inverso para la ecuación de ondas</b>	<b>43</b>
<b>Bibliografía</b>	<b>50</b>

# Índice de figuras

1.	Dos tipos de accidente cerebrovascular: isquémico y hemorrágico. . . . .	3
3.1.	Efecto de la regularización $L^2$ . . . . .	20
4.1.	Solución del algoritmo de regularización de Tikhonov en 1D para $\alpha = 1$ . . . . .	25
4.2.	Solución del algoritmo de regularización de Tikhonov en 1D para $\alpha = 0.01$ . . . . .	26
5.1.	Triangulación de un dominio de $\mathbb{R}^2$ . . . . .	40
6.1.	Perfil de la onda reconstruida para $D$ calculado. . . . .	48
6.2.	Amplitud de la onda reconstruida para $D$ calculado. . . . .	49

# Introducción

Los Ángeles (California), diciembre de 2001. La actriz estadounidense Sharon Stone comienza repentinamente a sentirse de manera extraña. Empieza a sufrir disartria (tartamudez), nota cómo pierde parte del oído izquierdo, se le paraliza el rostro y siente que su memoria comienza a desvanecerse. Está sufriendo un ictus. ¿Pero de qué clase de ictus se trata? La respuesta a esta pregunta es crucial para la supervivencia de la actriz, ya que hay dos tipos fundamentales de ictus: el ictus hemorrágico y el ictus isquémico (véase Figura 1). En el primero, se da la rotura de uno o varios vasos sanguíneos y en el segundo; un coágulo impide el flujo de sangre pudiendo ocasionar la pérdida de la región cerebral afectada o incluso la muerte. Para tratar el segundo, es indispensable emplear un anticoagulante, mientras que para tratar el primero no se debe usar bajo ningún concepto por razones evidentes. Por tanto, es necesario reconstruir una imagen lo más exacta posible del interior de la corteza craneal a partir de ciertas mediciones efectuadas sobre la cabeza (obtenidas por ejemplo mediante una resonancia magnética craneal). De manera informal, podemos decir que la acción general de reconstruir las causas a partir de los efectos se le conoce en matemáticas como *problemas inversos* y es hoy en día una rama muy importante de las ciencias aplicadas. Gracias a los medios disponibles y a la eficiencia de los profesionales sanitarios, se identificó qué clase de ictus era (uno hemorrágico) y la actriz se acabó recuperando. Aunque quizá ella no lo sepa, un problema inverso le salvó la vida.

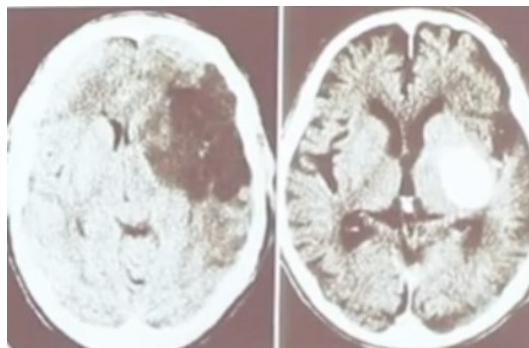


Figura 1: Dos tipos de accidente cerebrovascular: isquémico (izquierda) y hemorrágico (derecha). Las imágenes reconstruidas son sustancialmente distintas en cada caso, lo que permite actuar en consecuencia y salvar al paciente.



Por ello, cuando se escucha el término de *problema inverso*, no se debe pensar que se trata de una oscura e inaccesible rama de las matemáticas. La nomenclatura de problema inverso sólo designa un tipo de razonamiento que el ser humano lleva haciendo desde el principio de los tiempos: deducir las causas a partir de los efectos. Resulta incluso rutinario: al levantarnos por la mañana y asomarnos por la ventana, podemos deducir si ha llovido la noche anterior. Otro ejemplo de problema inverso ampliamente conocido fue la determinación de la altura de las pirámides de Egipto por Tales de Mileto (624 a.C. - 546 a.C.), quien en lugar de subirse a la pirámide (procedimiento bastante inútil e infructuoso), midió la altura de la misma atendiendo a la proyección de su sombra y a la sombra de un bastón.

Sin embargo, a pesar de ser una rama tan útil e interesante, el estudio de los problemas inversos no fue redescubierto hasta mediados del siglo XX, tras la Segunda Guerra Mundial. En aquel entonces, un grupo de investigación sueco encontró las publicaciones de un joven ruso-armenio llamado Viktor Ambartsumian. Los artículos de Ambartsumian, publicados en 1929 en la revista alemana *Zeitschrift für Physik* y redescubiertos unos veinte años después, trataban de la determinación de las ecuaciones de una cuerda vibrante a partir del conocido problema de Sturm-Liouville. Desde entonces, esta rama de las matemáticas adquirió cada vez mayor popularidad. Cabe destacar sobre todo grandes grupos de investigación de la antigua Unión Soviética, como I.M.Gelfand (1913-2009) y B.Levitan (1914-2004), quienes resolvieron el problema inverso de dispersión (*scattering*), así como los trabajos publicados por grandes matemáticos como A.N.Tikhonov [2] y Lavrent'ev [16].

El concepto de un problema inverso es relativamente sencillo: se parte de la base de que los modelos matemáticos que gobiernan los procesos (físicos, químicos, biológicos...) que se van a estudiar son conocidos. Estos modelos dependen de cierto/s parámetro/s desconocido/s, siendo el objetivo principal reconstruirlos a través de ciertas mediciones que podemos realizar de la solución del modelo. Los parámetros que interesa determinar pueden ser:

- Los coeficientes que aparecen en las ecuaciones que gobiernan el modelo. En este caso, se suele hablar de CIP (*Coefficient Inverse Problems*).
- Las condiciones iniciales y/o de contorno.
- Una parte del dominio donde se satisfacen las ecuaciones.

Son, por tanto, problemas que se pueden atajar desde un punto de vista tanto teórico como numérico, pero también de gran relevancia para aplicaciones de la Física y otras ciencias.

Las observaciones que podemos realizar de la solución dan una información parcial e indirecta de los parámetros desconocidos. Se enumeran a continuación una serie de procesos físicos que pueden servir de ejemplos de aplicaciones: detección de explosivos en un campo de batalla (donde muchas veces sólo es posible tomar medidas en una región reducida de la frontera del campo), geofísica (localización de pozos de petróleo), reconstrucción médica de imágenes a

partir de mediciones sobre la frontera de un órgano... Los fenómenos de interés pueden describirse mediante modelos gobernados por (sistemas de) ecuaciones algebraicas, diferenciales, integrales... En este trabajo, sólo nos centraremos en modelos descritos mediante ecuaciones o sistemas diferenciales.

Desde el punto de vista matemático, el análisis de los Problemas Inversos pasa por determinar la unicidad, la estabilidad de los mismos y diseñar algoritmos de reconstrucción de la solución.

El escollo principal de los Problemas Inversos es que son mal definidos/mal planteados (*ill-posed*) en el sentido de Hadamard (1865-1963), ver [8]: es decir, un pequeño cambio en las medidas observadas suele generar un cambio mayúsculo en la solución (son inestables).

Curiosamente, y por fortuna para el desarrollo de la teoría de problemas inversos, se estaba desarrollando de forma paralela otro área de investigación que sería de gran ayuda: el aprendizaje automático (*machine learning*). Una definición de *machine learning* podría ser “la rama de la Inteligencia Artificial (*Artificial Intelligence, AI*) que tiene por objetivo lograr que los ordenadores adquieran la capacidad de aprender sin ser expresamente programados”. Fue precisamente el hecho de entender el *machine learning* de esta manera lo que contribuyó al desarrollo de esta disciplina: hasta los años 70, se creía que lo máximo a lo que se podía aspirar era que los expertos en cierto problema programasen explícitamente la máquina para que esta pudiera resolverlo. Este tipo de programación *if-then* (si se cumplen ciertas condiciones, **entonces** el programa debe actuar de cierta manera) era muy árida y altamente ineficiente, por lo que el *machine learning* acabó cayendo en el olvido. No fue hasta el año 1983 cuando se publicaron unos trabajos que marcaron un antes y un después en esta disciplina. *Machine Learning: The AI Approach* [19] fue un compendio de artículos de investigadores que ilustraban con ejemplos que la programación indirecta (i.e. “dejar aprender a la máquina”) no sólo era posible, sino lo más eficiente. Una vez cambiado el enfoque, no sólo se planteó la pregunta de *cómo* aprender, sino también *qué* aprender. Para los investigadores en problemas inversos, la respuesta a qué aprender estuvo bastante clara.

Por ello, el objetivo central de este trabajo será abordar el estudio de dichos problemas haciendo uso de la teoría del cálculo variacional de Tikhonov y de herramientas de aprendizaje automático. En el Capítulo 1, se verán una serie de conceptos previos necesarios para entender los aspectos teóricos que subyacen tras los problemas inversos. En el Capítulo 2 se proporcionarán algunos ejemplos de este tipo de problemas. En el Capítulo 3, se introducirá la técnica de *machine learning* de regularización, ampliamente utilizada por Tikhonov y que nos ayudará a simplificar este tipo de problemas difíciles e inestables, convirtiéndolos en problemas de optimización. En el Capítulo 4, se abordará el problema de la elección del parámetro de regularización y en el Capítulo 5, se definirán métodos numéricos para la resolución de problemas inversos. Por último, en el Capítulo 6 se empleará todo lo aprendido para llevar a cabo una experiencia numérica para un problema inverso relativo a la ecuación de ondas.

# Capítulo 1

## Conceptos previos

De aquí en adelante, el problema que nos va a ocupar se puede expresar de la forma general siguiente:

$$Kx = y, \tag{1.1}$$

donde  $K : X \rightarrow Y$  es un operador diferencial (no necesariamente lineal),  $X$  e  $Y$  son dos espacios de Banach. Dado  $y \in Y$ , se pretende encontrar  $x \in X$  tal que se tiene (1.1). La razón de elegir espacios de Banach es que los mismos son "naturales" para las ecuaciones en derivadas parciales que surgen en el ámbito de la física matemática. Consideraremos la resolución de la ecuación (1.1) cuando el operador  $K^{-1}$  no exista, basándonos en una serie de resultados debidos a Lavrent'ev [16] y a Tikhonov [2].

Observemos que, en realidad, basta con considerar  $X$  e  $Y$  como subespacios de dichos espacios de Banach, que a su vez también son espacios de Banach.

Una vez hemos establecido la ecuación anterior, recordemos el significado de un problema bien planteado.

**Definición 1.1** Se dice que la ecuación (1.1) representa un problema bien planteado (*well-posed*) en el sentido de Hadamard [22] si se verifica lo siguiente:

- 1) Para todo  $y \in Y$ , existe  $x = x(y) \in X$  (existencia de solución).
- 2) Para cada  $y \in Y$ , existe un único  $x \in X$  satisfaciendo (1.1) (unicidad de solución), es decir si  $Kx_1 = Kx_2$  con  $x_i \in X$ , entonces  $x_1 = x_2$ .
- 3) El operador  $K^{-1} : Y \rightarrow X$  es continuo. Dicho de otro modo, la solución  $x = x(y)$  depende de  $y$  de forma continua.

**Definición 1.2** Si al menos una de las tres condiciones anteriores no se verifica para el problema (1.1), diremos que dicho problema está mal planteado (*ill-posed*).

**Observación 1.3** Nótese que las condiciones de la Definición 1.1 no son independientes. Para operadores  $K$  lineales entre espacios de Banach, las condiciones 1) y 2) implican la condición 3) por el teorema del grafo cerrado de Banach [21], según el cual “si la gráfica de un operador lineal entre espacios de Banach es un cerrado en el espacio producto, entonces el operador es continuo”. De hecho, si existe un operador  $K$  entre un abierto  $X_1$  de un subespacio de Banach  $X$  y un abierto  $Y_1$  de un subespacio de Banach  $Y$  ( $\text{codim}X_1 + \text{codim}Y_1 < \infty$ ), entonces  $K$  lleva  $X$  en  $Y$  y, por 1) y 2) es inyectivo. Recuérdese que la codimensión de un subespacio vectorial es la diferencia de dimensión entre este y el espacio en el cual está contenido. En nuestro caso,  $\text{codim}X_1 = \dim(X) - \dim(X_1)$  y  $\text{codim}Y_1 = \dim(Y) - \dim(Y_1)$ . Por consiguiente, y aplicando el teorema del grafo cerrado, el inverso  $K^{-1} : Y \rightarrow X$  es continuo con respecto a las normas en  $X_1$  e  $Y_1$ , y tenemos la propiedad 3).

Como los datos tomados en la práctica adolecen por lo general de error, debemos incluir la noción de error en los datos. Este error proviene del ruido aleatorio, así como de las discrepancias entre el modelo matemático y la realidad, puesto que en la práctica, las mediciones que realizamos no son exactas. En lo sucesivo, denotaremos como  $y_\delta$  las medidas que vienen afectadas de error. Sin embargo,  $y_\delta$  no es completamente aleatorio. Más concretamente, tenemos

**Definición 1.4** Sea  $\delta > 0$  suficientemente pequeño. Diremos que  $y_\delta \in Y$  tiene un error de nivel  $\delta$  si  $\|y_\delta - y^*\|_Y \leq \delta$ , donde  $y^* \in Y$  es el valor exacto (ideal) de  $y_\delta$ , que no está afectado de error.

Procedamos a definir ahora la noción de *corrección condicional* de Tikhonov.

**Definición 1.5** Sean  $X$  e  $Y$  dos espacios de Banach. Sea  $G \subset X$  un conjunto elegido a priori de la forma  $G = \bar{G}_1$ , siendo  $G_1$  un abierto de  $X$ . Sea  $K : G \rightarrow Y$  continuo. Supongamos también que  $y_\delta$  tiene un error de nivel  $\delta$ . Diremos que el problema (1.1) está *condicionalmente bien planteado* o *bien planteado en el sentido de Tikhonov* si se verifica lo siguiente:

- 1) Se supone *a priori* que existe una solución ideal  $x^* = x^*(y^*)$  del problema (1.1) para un dato ideal sin ruido  $y^*$ .
- 2) Dados  $x, x^* \in G$  y  $Kx = Ky^*$ , se tiene que  $x = x^*$ . En otras palabras, el operador  $K : G \rightarrow Y$  es inyectivo (unicidad de solución en  $G$ ).
- 3) El operador  $K^{-1}$  es continuo en el conjunto  $K(G)$  (estabilidad de solución, es decir  $\|x - x^*\|_G \rightarrow 0$  si  $\|Kx - Ky^*\|_Y \rightarrow 0$ ).

**Definición 1.6** En las condiciones anteriores, se denomina *conjunto de corrección* al conjunto  $G$  de la Definición 1.5.

La idea clave de Tikhonov era, pues, conseguir condiciones de estabilidad para un problema inestable. Su idea es resolver el problema en un conjunto compacto, como demostraremos a

continuación. La cuestión es ver si es razonable asumir que la solución pertenece a un conjunto compacto determinado. Veamos que sí. Por ejemplo, un compacto en  $C(\Omega)$  podría ser el conjunto de todas las funciones de  $C^1(\Omega)$  acotadas y con derivada primera acotada, en valor absoluto, por una constante elegida a priori. Argumentos similares pueden emplearse en el caso de otros espacios de funciones, como  $C^k(\Omega)$  y  $H$  (espacios de Hilbert). Veamos, pues, un resultado interesante en este sentido.

**Teorema 1.7** Sean  $X$  e  $Y$  dos espacios de Banach. Sea  $U \subset X$  un conjunto compacto y  $F : U \rightarrow Y$  un operador continuo. Denotemos  $V = F(U)$ . Supongamos que el operador  $F$  es inyectivo. Entonces, el operador inverso  $F^{-1} : V \rightarrow U$  es continuo.

*Demostración.* Procedamos por reducción al absurdo y supongamos que el operador  $F^{-1}$  no es continuo en  $V$ . Es decir,  $\exists y_0 \in V$  y  $\exists \epsilon > 0$  tales que  $\forall \delta > 0, \exists y_\delta$  tal que aunque  $\|y_\delta - y_0\|_Y < \delta$ , se tiene  $\|F^{-1}(y_\delta) - F^{-1}(y_0)\|_X \geq \epsilon$ . Por tanto, existe una sucesión  $\{\delta_n\}$ , con  $\lim_{n \rightarrow \infty} \delta_n = 0^+$  y una correspondiente sucesión  $\{y_n\} \subset V$  tales que

$$\|y_{\delta_n} - y_0\|_Y < \delta_n, \quad \|F^{-1}(y_n) - F^{-1}(y_0)\|_X \geq \epsilon.$$

Pongamos

$$x_n = F^{-1}(y_n), \quad x_0 = F^{-1}(y_0).$$

Entonces, se tiene que

$$\|x_n - x_0\|_X \geq \epsilon.$$

Como  $U$  es compacto y todo punto  $x_n \in U$ , podemos extraer una subsucesión convergente  $\{x_{n_k}\} \subseteq \{x_n\}$ . Sea  $\lim_{k \rightarrow \infty} x_{n_k} = \bar{x}$ , entonces  $\bar{x} \in U$ . Como  $F(x_{n_k}) = y_{n_k}$  y el operador  $F$  es continuo, deducimos de lo anterior que  $F(\bar{x}) = y_0 = F(x_0)$ . Como  $F$  es inyectivo, deberíamos tener  $\bar{x} = x_0$ , pero  $\|\bar{x} - x_0\|_X \geq \epsilon$ . Hemos llegado a una contradicción. Esto termina la demostración del teorema.  $\square$

Antes de ilustrar con ejemplos algunos problemas inversos, resulta preciso recordar algunas definiciones previas.

**Definición 1.8** Se define  $C^0(\Omega) \equiv C(\Omega)$  como el espacio de funciones continuas en  $\Omega$ , que se puede dotar con la norma:

$$\|u\|_{C(\Omega)} := \sup_{x \in \Omega} |u(x)|.$$

**Definición 1.9** Sea  $(X, \mu)$  un espacio métrico. Se define el espacio vectorial  $L^p(X)$  para  $p \in [1, \infty)$  como el espacio de clases de funciones medibles  $f$  que cumplen:

$$\int_X |f|^p d\mu < \infty.$$

Una norma natural en estos espacios es:

$$\|f\|_p = \left( \int_X |f|^p d\mu \right)^{1/p}.$$

La definición anterior puede extenderse al caso en que  $p = \infty$ , pero no nos va a interesar para nuestro análisis.

**Definición 1.10** Se define el *espacio de Hilbert*  $H$  como un espacio de producto interior que es completo con respecto a la norma definida por este producto. Dicha norma queda definida como sigue:

$$\|u\|_H = \sqrt{\langle u(x), u(x) \rangle},$$

donde  $\langle \cdot, \cdot \rangle$  denota el producto interior.

**Observación 1.11** Todo espacio de Hilbert es un espacio de Banach. El recíproco no es cierto en general.

Veamos un caso particular donde el recíproco sí es cierto, lo cual nos permitirá definir un algoritmo de optimización más adelante.

**Teorema 1.12** Para todo espacio métrico  $(X, \mu)$ , el espacio  $L^2(X)$  es un espacio de Hilbert con producto escalar definido por:

$$\langle f, g \rangle = \int_X fg d\mu.$$

*Demostración.* Basta observar que la norma asociada al producto escalar es, de hecho, la norma en  $L^2(X)$ :

$$\|f\| = \sqrt{\langle f, f \rangle} = \sqrt{\int_X f^2 d\mu} := \|f\|_2.$$

Por el teorema de Fischer-Riesz [6],  $L^2(X)$  es un espacio de Banach, por lo que es completo con respecto a esta norma y, por tanto, es un espacio de Hilbert.

Para finalizar este capítulo, se definirá el concepto de problema de optimización de un funcional, ya que los problemas inversos pueden ser enfocados de esta manera, como se verá más adelante.

**Definición 1.13** Sea  $V$  un espacio normado. Dados  $X \subset V$  un conjunto,  $J : X \rightarrow \mathbb{R}$  un funcional y el problema

$$\begin{cases} \text{Hallar } x \in X \text{ tal que} \\ J(x) \leq J(v), \quad \forall v \in X, \end{cases}$$

se dirá que el problema anterior es un problema de optimización sin restricciones si  $X \equiv V$ .

**Definición 1.14 (Semicontinuidad inferior)** Sea  $V$  un espacio normado y  $J : X \rightarrow \mathbb{R}$  un funcional, con  $X \subset V$ . Diremos que  $J$  es semicontinuo inferiormente en un punto  $x \in X$  si para toda sucesión  $\{x_n\}_{n \geq 1} \subset X$  con  $x_n \rightarrow x$ , se tiene

$$J(x) \leq \liminf J(x_n).$$

Se dice que  $J$  es semicontinuo inferiormente en  $X$  si lo es en todo punto  $x \in X$ .

**Definición 1.15 (Coercividad)** Sea  $V$  un espacio normado,  $X \subset V$  un subconjunto no acotado y  $J : X \rightarrow \mathbb{R}$  un funcional. Se dice que  $J$  es coercivo en  $X$  si

$$\lim_{\|u\| \rightarrow \infty} J(u) = +\infty$$

Ahora se definirá la noción de semicontinuidad secuencial débil, para lo cual es preciso recordar qué es la convergencia débil:

**Definición 1.16 (Convergencia débil)** Sea  $\{x_n\}_{n \geq 1} \subset X$  una sucesión y  $x \in X$  un elemento de  $X$ . Se dice que  $x_n$  converge débilmente a  $x$  (y se escribirá  $x_n \rightharpoonup x$ ) si para cualquier  $x' \in X'$  se tiene

$$\langle x', x_n \rangle_{X', X} \rightarrow \langle x', x \rangle_{X', X},$$

donde  $\langle \cdot, \cdot \rangle_{X', X}$  representa el producto dual entre el espacio normado  $X$  y su espacio dual  $X'$ .

**Definición 1.17 (Semicontinuidad secuencial débil)** Sea  $V$  un espacio normado y  $J : X \rightarrow \mathbb{R}$  un funcional, con  $X \subset V$ . Diremos que  $J$  es secuencialmente débilmente semicontinuo inferiormente en un punto  $x \in X$  si para toda sucesión  $\{x_n\}_{n \geq 1} \subset X$  con  $x_n \rightharpoonup x$ , se tiene

$$J(x) \leq \liminf J(x_n).$$

Se dice que  $J$  es secuencialmente débilmente semicontinuo inferiormente en  $X$  si lo es en todo punto  $x \in X$ .

A continuación, se definirá la noción de mínimo de un funcional y se demostrará un teorema (Weierstrass) que garantiza la existencia de dicho mínimo.

**Definición 1.18** Sea  $V$  un espacio normado,  $X \subset V$  un subconjunto y  $J : X \rightarrow \mathbb{R}$  un funcional. Se dice que  $x$  es un mínimo local (o mínimo relativo) de  $J$  en el conjunto  $X$  si  $x \in X$  y existe  $\epsilon > 0$  tal que

$$J(x) \leq J(v) \quad \forall v \in X \cap \bar{B}(x; \epsilon),$$

donde  $\bar{B}(x; \epsilon)$  denota la bola cerrada de centro  $x$  y radio  $\epsilon$ . Por otro lado, se dirá que  $x$  es un mínimo global (o absoluto) de  $J$  en el conjunto  $X$  si  $x \in X$  y se satisface

$$J(x) \leq J(v), \quad \forall v \in X.$$

**Teorema 1.19 (Teorema de Weierstrass)** *Supongamos que  $X$  es un espacio métrico no vacío y compacto (i.e. para todo recubrimiento por abiertos de  $X$ , existe un subrecubrimiento finito) y  $J : X \rightarrow \mathbb{R}$  es un funcional semicontinuo inferiormente en  $X$ . Entonces,  $J$  admite un mínimo global en  $X$ .*

*Demostración.* Para probar el teorema, se tomará una sucesión minimizante  $\{x_n\}_{n \geq 1}$  y probaremos que admite una subsucesión convergente a un mínimo global de  $J$  en  $X$ . Por ser  $X$  compacto, se puede extraer una subsucesión convergente y también se usará la semicontinuidad inferior de  $J$  en  $X$  para ver que el límite es, en efecto, un mínimo global de  $J$  en  $X$ . Si llamamos  $\mu = \inf_{v \in X} J(v) \in [-\infty, \infty)$ , entonces, existe una sucesión  $\{x_n\}_{n \geq 1} \subset X$  (sucesión minimizante) tal que

$$\lim_{n \rightarrow \infty} J(x_n) = \mu.$$

Por ser  $X$  compacto, existen  $\bar{x} \in X$  y una subsucesión  $\{x_{n_k}\}_{k \geq 1}$  de  $\{x_n\}_{n \geq 1}$  tales que  $x_{n_k} \rightarrow \bar{x}$  en  $X$ . Al ser  $J$  un funcional semicontinuo inferiormente en  $X$ , se tiene

$$J(\bar{x}) \leq \liminf_{k \rightarrow \infty} J(x_{n_k}) = \lim_{n \rightarrow \infty} J(x_n) = \mu.$$

Deducimos por tanto que  $\mu \in \mathbb{R}$  y  $J(\bar{x}) = \mu = \inf_{x \in X} J(x)$  y hemos terminado.  $\square$

Por último, una vez conocidas las definiciones necesarias para entender el problema de minimización de un funcional, se introducirán las derivadas de Fréchet y de Gateaux, que nos serán de utilidad en la Sección 5.2 para definir el Método de Elementos Finitos.

**Definición 1.20 (Derivada de Gateaux)** Sean  $X$  e  $Y$  dos espacios vectoriales topológicos localmente convexos,  $U \subset X$  abierto, y  $F : X \rightarrow Y$ . El diferencial de Gateaux  $dF(u, \psi)$  de  $F$  en  $u \in U$  en la dirección  $\psi \in X$  se define como sigue:

$$dF(u, \psi) = \lim_{\tau \rightarrow 0} \frac{F(u + \tau\psi) - F(u)}{\tau} = \left. \frac{d}{d\tau} F(u + \tau\psi) \right|_{\tau=0} \quad (1.2)$$

Si el límite anterior existe  $\forall \psi \in X$ , se dice que  $F$  es *Gateaux diferenciable* en  $u$ .

Se observa que el límite que aparece en la expresión (1.2) se toma relativo a la topología de  $Y$ . Si además  $X$  e  $Y$  son espacios topológicos reales, el límite se toma para  $\tau \in \mathbb{R}$ .

Los espacios de Banach son espacios vectoriales topológicos localmente convexos (i.e. espacios vectoriales cuya topología está generada por conjuntos convexos), por lo que el cálculo de la derivada de Gateaux para el funcional de Tikhonov, definido con respecto a la norma en dichos espacios de Banach, está bien definido.

A continuación se definirá la derivada de Fréchet, que es otra definición importante que permite formalizar la noción de derivada de un funcional en el marco del cálculo variacional.



**Definición 1.21 (Derivada de Fréchet)** Sean  $V$  y  $W$  espacios vectoriales normados, y  $U \subset V$  abierto. Se dice que una función  $f : U \rightarrow W$  es *Fréchet diferenciable* en  $x \in U$  si existe un operador lineal acotado  $A : V \rightarrow W$  que satisface

$$\lim_{\|h\| \rightarrow 0} \frac{\|f(x+h) - f(x) - Ah\|_W}{\|h\|_V} = 0 \quad (1.3)$$

Se tiene que toda función (funcional) Fréchet diferenciable es, a su vez, Gateaux diferenciable. El recíproco no es cierto en general. La razón es que la existencia de todas las derivadas direccionales en un punto no garantiza diferenciabilidad total (ni siquiera continuidad) en ese punto.

## Capítulo 2

# Ejemplos de problemas mal planteados

Veamos en este capítulo algunos ejemplos de problemas mal planteados.

EJEMPLO 2.1. (ECUACIÓN DEL CALOR CON TIEMPO INVERTIDO / DIFUSIÓN INVERSA) Sean  $\Omega = (0, \pi)$ ,  $T > 0$ ; consideramos la ecuación del calor en una dimensión con condiciones de contorno de tipo Dirichlet homogéneas:

$$\begin{cases} \partial_t u - \partial_{xx}^2 u = 0 & \text{en } \Omega \times (0, T), \\ u = 0 & \text{sobre } \partial\Omega \times (0, T), \\ u(x, 0) = u_0 & \text{en } \Omega. \end{cases} \quad (2.1)$$

El problema inverso, en este caso, es encontrar el dato inicial a partir del dato final. Más concretamente, dado  $u_T \in L^2(\Omega)$ , hallar  $u_0 \in L^2(\Omega)$  y después  $u = u(x, t)$  que sea solución de (2.1). O sea, que además de las condiciones de contorno anteriores, se debe satisfacer

$$u(x, T) = u_T(x).$$

Sabemos que una solución del problema, si existe, viene dada por el principio de superposición, sumando para todo  $n$  las funciones que son solución de esta ecuación  $u_n(x, t) = e^{-n^2 t} \text{sen } nx$ , es decir

$$u(x, t) = \sum_n e^{-n^2 t} \text{sen } nx$$

Se observa que las funciones  $u_n$  verifican la ecuación del calor y las condiciones de contorno de (2.1). Los datos iniciales son, por tanto,  $u_n(x, 0) = \text{sen } nx$ . Estas funciones tienen  $\|\cdot\|_{C^0} = 1$  y  $\|\cdot\|_{L^2} = \sqrt{1/2}$ . Los datos finales son  $u_n(x, T) = e^{-n^2 T} \text{sen } nx$ . Estos tienen  $\|\cdot\|_{C^0} = e^{-n^2 T}$  y  $\|\cdot\|_{L^2} = e^{-n^2 T}(\sqrt{1/2})$ . Si definimos  $u_T = Ku_0$ , entonces la cota  $\|u_0\|_X \leq C\|u_T\|_Y$  es imposible para ningún espacio clásico de funciones ( $C^0, L^2$ ) de los que estamos considerando: las normas de las funciones  $u_n(x, T)$  tienden a cero exponencialmente, mientras que las normas de las  $u_n(x, 0)$  son mayores que  $1/2$ . Por consiguiente, el problema de hallar los datos iniciales a partir de los datos finales es exponencialmente inestable en dichos espacios de funciones. Por tanto, no se

tiene la condición 3) de la Definición 1.1, y el problema está *mal planteado*.

Los problemas mal planteados (*ill-posed*) no tienen por qué ser necesariamente problemas inversos, como se muestra a continuación.

**EJEMPLO 2.2.** (DIFERENCIACIÓN DE UNA FUNCIÓN CON RUIDO) Vamos a demostrar que el problema de diferenciación de funciones con ruido es inestable. Supongamos que la función  $g(x), x \in [0, 1]$  adolece de ruido. Es decir, en lugar de  $g(x) \in C^1[0, 1]$ , la función que consideraremos será  $g_\delta(x)$  definida como sigue:

$$g_\delta(x) = g(x) + \delta g(x), \quad x \in [0, 1],$$

donde  $\delta g(x)$  es la componente de ruido. Sea  $\delta > 0$  el parámetro característico del ruido, suficientemente pequeño. Sin pérdida de generalidad, supongamos que la componente de ruido también es pequeña, esto es,  $\|\delta g(x)\|_{C^0[0,1]} \leq \delta$ . Veamos que el problema del cálculo de la derivada,  $g'_\delta(x)$ , es inestable. En efecto, consideramos por ejemplo que una posible componente ruidosa que satisfaga lo anterior bien podría ser:

$$\delta g(x) = \frac{\text{sen } n^2 x}{n},$$

con  $n \gg 0$  suficientemente grande ( $n$  tal que  $1/n \leq \delta$ ). Vemos que

$$\|\delta g\|_{C^0[0,1]} = 1/n \leq \delta.$$

Sin embargo, la diferencia entre la derivada de la función con ruido y la ideal da como resultado:

$$g'_\delta(x) - g'(x) = n \cos n^2 x,$$

que no está acotado en ninguna norma de espacios clásicos de funciones ( $C^0, L^2$ , espacios de Hilbert).

**EJEMPLO 2.3.** (TOMOGRAFÍA DE IMPEDANCIA ELÉCTRICA / PROBLEMA DE CALDERÓN)

En el siguiente problema inverso, lo que se busca es reconstruir la conductividad eléctrica  $\sigma = \sigma(x)$  en un abierto  $\Omega \subset \mathbb{R}^n$ , conocido el potencial eléctrico  $u$  sobre la frontera  $\partial\Omega$  y/o su corriente asociada.

Este problema surge en el ámbito médico cuando se desea conocer la conductividad de alguna parte del cuerpo a partir de mediciones sobre el contorno (la piel). En la práctica, se sitúan unos electrodos sobre el paciente y se trata de reconstruir un mapa de  $\sigma$  en la zona de interés mediante la aplicación de pequeñas corrientes (resp. potenciales) en dichos electrodos. Esto resulta muy útil para la detección de tumores, ya que la conductividad del tejido tumoral es mucho mayor que la del tejido sano. Veamos el problema desde un punto de vista más formal. Este problema surge de la conocida ecuación de continuidad electromagnética, que se deriva, a su vez, de las ecuaciones de Maxwell. Dicha ecuación es

$$\nabla \cdot J + \frac{\partial \rho}{\partial t} = 0, \tag{2.2}$$

es decir, la divergencia de la densidad de corriente  $J$  en un medio y la derivada temporal de la densidad de carga  $\rho$  en ese mismo medio se compensan. Generalmente, en la práctica, en los medios donde se aborda este problema se tiene  $\rho \neq \rho(t)$ , por lo que la derivada parcial se anula y queda

$$\nabla \cdot J = 0. \quad (2.3)$$

Si recordamos ahora la fórmula para la densidad de corriente  $J = \sigma E$ , donde  $\sigma$  es la conductividad eléctrica y  $E$ , el campo eléctrico; y la expresión para el campo eléctrico  $E = -\nabla u$ , siendo  $u$  el potencial eléctrico, la ecuación anterior se modifica. Si se efectúan adicionalmente medidas del potencial eléctrico sobre la frontera, obtenemos un problema de contorno:

$$\begin{cases} \nabla \cdot (\sigma(x)\nabla u) = 0, & x \in \Omega, \\ u = f(x), & x \in \partial\Omega. \end{cases}$$

El problema anterior es la expresión matemática del problema de Calderón. Veamos que es un problema mal planteado. Por simplicidad, consideraremos el problema en una dimensión. El caso multidimensional  $n > 1$  es análogo. Ahora, pues, imponemos  $\Omega = (0, L)$ . El problema directo sería, conocida la conductividad  $\sigma = \sigma(x) > 0$ , hallar  $u = u(x)$  satisfaciendo

$$\begin{cases} \frac{d}{dx} \left( \sigma(x) \frac{du}{dx} \right) = 0, & x \in (0, L), \\ u(0) = u_0, & u(L) = u_L \end{cases}$$

y después, determinar las densidades de corriente  $j_0 := (\sigma \frac{du}{dx})(0)$  y  $j_L := (\sigma \frac{du}{dx})(L)$ . Para el problema directo, encontrar la solución en función de  $u_0, u_L$  y  $\sigma$  es fácil:

$$u(x) = \frac{u_L - u_0}{\int_0^L \sigma^{-1}(s) ds} \int_0^x \sigma^{-1}(s) ds + u_0. \quad (2.4)$$

Ahora, reformulemos el problema inverso en base a la aplicación Dirichlet-a-Neumann, que no es más que la aplicación que transforma el potencial aplicado en densidad de corriente:  $\Lambda_\sigma : f \mapsto g$ , donde  $f = u(x)|_{\partial\Omega}$  y

$$g := \sigma(x) \sum_{i=1}^n \partial_i u n_i(x),$$

donde  $n_i(x)$  es la  $i$ -ésima componente del vector normal a  $\Omega$  en el punto  $x \in \partial\Omega$ . Por tanto, el problema de Calderón puede entenderse como la reconstrucción de  $\sigma$ , conocida la aplicación Dirichlet-a-Neumann  $\Lambda_\sigma$ . Por tanto, retomando la ecuación (2.4), podemos obtener valores de la aplicación  $\Lambda_\sigma$ :

$$\Lambda_\sigma|_{x=0} = \Lambda_\sigma|_{x=L} \Rightarrow \sigma(0) \frac{du}{dx}(0) = \sigma(L) \frac{du}{dx}(L) = \frac{u_L - u_0}{\int_0^L \sigma^{-1}(s) ds}.$$

Por tanto, conocemos  $\Lambda_\sigma$ , pero cualquier función  $\sigma$  que cumpla la igualdad anterior es una solución válida del problema inverso. La consecuencia es que no tenemos unicidad de solución para el problema y está, por tanto, *mal planteado*.

## Capítulo 3

# Regularización

Como ya se ha comentado anteriormente, un algoritmo de *machine learning* es un algoritmo capaz de aprender de los datos que le son suministrados. El concepto de “aprender” es muy amplio, pero una definición informal podría ser la siguiente (véase [7]): “se dice que un programa informático aprende de experiencias  $E$  con respecto a alguna clase de tareas  $T$  y a una medida de rendimiento  $P$ , si su rendimiento en tareas de  $T$ , medido por  $P$ , mejora con las experiencias  $E$ ”. A este conjunto de experiencias  $E$  se le conoce como *entrenamiento*, si el algoritmo es sometido a dichas experiencias en numerosas ocasiones; o bien como *test*, si es la primera vez que se enfrenta a ello. Esto da lugar a dos tipos de errores:

1. **Error de entrenamiento.** Es el error promedio que resulta de utilizar un algoritmo de aprendizaje automático para predecir una solución dentro del *conjunto de entrenamiento*, i.e. dentro del conjunto de medidas y observaciones proporcionadas al sistema para entrenarlo.
2. **Error de test.** Se trata del error promedio que surge al emplear el algoritmo de *machine learning* para predecir la solución de una observación nueva, i.e. una observación que no se utilizó para entrenar el sistema.

En un sentido amplio, el término regularización se usa para denotar toda modificación realizada a un algoritmo de aprendizaje automático cuyo fin es minimizar su error de test, pero no su error de entrenamiento. Si trasladamos este concepto al ámbito de los problemas inversos, obtenemos la idea principal de Tikhonov para resolver (1.1): reemplazar dicha ecuación por otra ecuación “parecida” que involucra un parámetro  $\alpha$  suficientemente pequeño. Veremos a continuación cómo el problema de resolución de un problema inverso se reduce a un problema de minimización de un funcional, y la solución de dicho problema de minimización se aproxima a la solución del problema inverso original (1.1) para una elección adecuada de  $\alpha$ .

### 3.1. El teorema No Free Lunch

El *modus operandi* tradicional del aprendizaje automático se basa en la premisa de que un algoritmo puede generalizar («aprender») a partir de un conjunto finito de ejemplos. Esto parece algo contraintuitivo: para poder resolver cualquier problema mediante un algoritmo, este debería entrenarse con el mayor número de ejemplos posible. Sin embargo, el desempeño de un algoritmo debe evaluarse en términos probabilísticos: un algoritmo inferirá soluciones que son probablemente correctas para la mayoría de problemas del mismo tipo.

Más aún, el teorema No Free Lunch de *machine learning*, introducido en 1997 por William MacReady y David Wolpert en [4], establece que, una vez entrenado un algoritmo con todas las distribuciones generadoras de datos (datos generados en base a un modelo, sin necesidad de recurrir a algún archivo o base de datos), todo algoritmo tiene el mismo margen de error a la hora de resolver un problema al que no se ha enfrentado previamente. El nombre del teorema hace referencia a una cita del propio Wolpert: “*there ain’t no such thing as a free lunch*”, o sea, que no existen almuerzos gratis ni atajos para el éxito. Vamos a definir esto de manera más rigurosa.

**Teorema 3.1 (No Free Lunch)** Sean  $A$  y  $B$  dos algoritmos de optimización, y sean  $P$  y  $Q$  dos clases de problemas. Sea  $f_A(P)$  la mejor solución encontrada por el algoritmo  $A$  en los problemas de la clase  $P$ , y sea  $f_B(Q)$  la mejor solución encontrada por el algoritmo  $B$  en los problemas de la clase  $Q$ . Entonces, se tiene:

$$\mathbb{E}[f_A(P)] = \mathbb{E}[f_B(Q)],$$

donde  $\mathbb{E}$  denota el valor esperado de una función  $f$  respecto a una determinada distribución de probabilidad  $D$ :

$$\mathbb{E}[f] = \int f(x)D(x)dx.$$

La demostración del teorema anterior se basa en que, independientemente del desempeño de un algoritmo  $A$  en los problemas de la clase  $P$ , siempre existe una distribución de probabilidad equivalente entre  $P$  y  $Q$  tal que el algoritmo  $B$  sea igualmente bueno en promedio. Dicho de otro modo, con respecto a esta distribución de probabilidad equivalente  $D^*$ , cualquier mejora en el rendimiento del algoritmo  $A$  sobre el algoritmo  $B$  en los problemas de clase  $P$  se compensa exactamente con una disminución en el rendimiento en los problemas de clase  $Q$ , y viceversa. En definitiva, no existe ningún algoritmo de *machine learning* que sea universalmente mejor que algún otro.

### 3.2. El funcional de regularización de Tikhonov

En la discusión anterior se ha presentado el concepto de regularización de una manera un tanto abstracta. En lo sucesivo, se particularizará dicho concepto de forma más rigurosa para el caso de problemas inversos ligados a las ecuaciones diferenciales.

**Definición 3.2** Sea  $\delta_0$  suficientemente pequeño. Para cada  $\delta \in (0, \delta_0)$  denotemos por  $E_\delta(y^*)$  un entorno de  $y^*$ , es decir  $E_\delta(y^*) = \{z \in Y : \|z - y^*\|_Y \leq \delta\}$ . Sea  $\alpha > 0$ . Una familia de operadores continuos  $R_\alpha : E_\delta \rightarrow G$  (con  $G \subset X$  el conjunto de corrección del problema inverso) se llama *regularizador* para la ecuación (1.1) en  $G$  si:

$$\lim_{\delta \rightarrow 0} \|R_{\alpha(\delta)}(y_\delta) - x^*\|_X = 0. \quad (3.1)$$

Se observa que en la definición anterior, existe un parámetro  $\alpha > 0$  que se denominará *parámetro de regularización*. El procedimiento de construcción de la solución aproximada  $x_{\alpha(\delta)} := R_{\alpha(\delta)}(y_\delta)$  se llama regularización.

A la luz de lo anterior para encontrar una solución aproximada de (1.1), surge la duda de cómo construir un funcional de regularización para el problema inverso (1.1). Veámoslo a continuación.

**Definición 3.3** Diremos que  $\psi$  es un *funcional de estabilización* para el conjunto de corrección  $G \subset X$  si verifica lo siguiente:

- (i)  $\psi$  es un funcional semicontinuo inferiormente no negativo definido en  $G$ .
- (ii) El conjunto  $G_\tau := \{x \in G : \psi(x) \leq \tau\}$  está acotado en  $X$  para todo  $\tau$ .

**Definición 3.4** El *funcional de regularización de Tikhonov*  $J_\alpha : G \rightarrow \mathbb{R}$  en un espacio de Hilbert  $Y$  para el problema (1.1) se define como sigue:

$$J_\alpha(x) := \frac{1}{2} \|Kx - y\|_Y^2 + \alpha\psi(x), \quad (3.2)$$

donde  $\alpha > 0$  es el parámetro de regularización y  $\psi$  es un funcional de estabilización.

En la definición anterior, al sumando  $\frac{1}{2} \|Kx - y\|_Y^2$  se le denomina *término de fidelidad* y al sumando  $\alpha\psi(x)$ , *término de regularización*.

Por tanto, para encontrar una solución aproximada de (1.1), se plantea el problema de minimizar el funcional de Tikhonov, es decir

$$x_\alpha = \arg \min_{x \in G} J_\alpha(x).$$

Se observa también que si  $\alpha = 0$ , obtenemos un problema habitual en el sentido de mínimos cuadrados para aproximar el problema (1.1).

El término de regularización  $\alpha\psi(x)$  encierra información disponible a priori sobre la solución desconocida, como suavidad o monotonía de la misma. Desde el punto de vista de *machine learning*, esto es de especial relevancia, ya que añadir este término supone aumentar o disminuir la capacidad representativa del modelo. Subliminalmente, este término añade o excluye (considérese «exclusión» en el sentido de una preferencia severa en contra de esa solución) funciones del espacio hipotético de soluciones donde debe buscar nuestro algoritmo. En virtud el teorema No Free Lunch, no existe un término de regularización óptimo para todos los casos, sino que se debe encontrar el término más adecuado en función del problema inverso que se considere. Elecciones comunes para este término se ilustran en la Tabla 3.1:

Información <i>a priori</i>	$\psi(x)$
Espacio $L^p$	$\ x\ _{L^p}^p$ , con $1 \leq p \leq 2$
BV (funciones de variación acotada)	$\ x\ _{BV}$
TV (total variation)	$\ \nabla x\ _{L^1}$
Espacio de Hilbert, $H^1$	$\frac{1}{2}\ x\ _{H^1}^2$
Combinación lineal	$\ x\ _{L^1} + \ x\ _{L^2}^2$

Tabla 3.1: Posibles elecciones del término de regularización.

Tradicionalmente, el funcional estándar de regularización de Tikhonov  $J_\alpha$  se supone definido tomando las normas en  $L^2$ , pues así se asume implícitamente que el ruido sigue una distribución gaussiana. Dicho funcional sería de la forma:

$$J_\alpha(x) := \frac{1}{2}\|Kx - y\|_2^2 + \frac{\alpha}{2}\|x\|_2^2.$$

Una motivación intuitiva de lo que supone la regularización en  $L^2$  puede apreciarse visualmente en la Figura 3.1. La representación gráfica se basa en la siguiente aproximación del funcional:

$$\hat{J}_\alpha(\omega) = J_\alpha(\omega^*) + \frac{1}{2}(\omega - \omega^*)^T H(\omega - \omega^*), \tag{3.3}$$

donde  $\omega^*$  es la solución no regularizada, y  $H$  representa la matriz hessiana. Se aprecia que la solución óptima se halla a medio camino entre el ajuste estricto a los datos y la suavidad impuesta por el término de regularización, siempre y cuando se haga una elección óptima de  $\alpha$ , lo cual se verá en el Capítulo 4.

Sin embargo, considerar el término de regularización en  $L^2$  no siempre es lo más adecuado. Por ejemplo, para la reconstrucción de una imagen tomada en la oscuridad, tenemos que restablecer la solución de un problema basado en la toma de fotones, proceso de emisión regular cuantificado, lo cual está distribuido por la distribución de Poisson. Una elección adecuada de este término de regularización sería la correspondiente a TV en la Tabla 3.1.



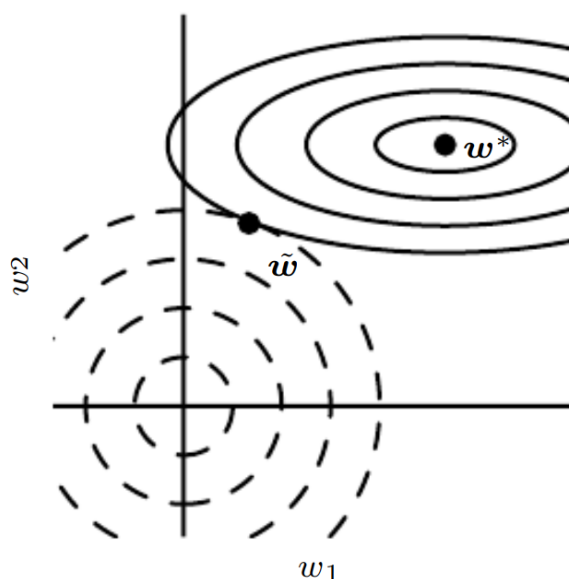


Figura 3.1: Efecto de la regularización  $L^2$ . Tras la aproximación del funcional dada por la ecuación 3.3, vemos que  $\hat{J}_\alpha(\omega)$  es mínimo cuando  $H(\omega - \omega^*) = 0$ . Tras un cálculo sofisticado [7], vemos que el efecto de la regularización es reescalar la solución no regularizada  $\omega^*$  a lo largo de los ejes que representan los autovectores (autofunciones, en este caso) de  $H$ :  $\omega_1$  y  $\omega_2$ . En este espacio de fases, las elipses continuas representan curvas de nivel de  $\omega^*$ . Los círculos discontinuos representan curvas de nivel del regularizador en  $L^2$ . En  $\bar{\omega}$  (solución regularizada), ambos términos alcanzan un equilibrio.

Veamos ahora un resultado de existencia de solución para el funcional de regularización dado en la Definición 3.4, véase [12].

**Teorema 3.5** *Sea  $G \subset X$  convexo y cerrado, siendo  $G$  el conjunto de corrección para el problema (1.1). Supongamos que los funcionales  $\frac{1}{2}\|Kx - y\|_Y^2$  y  $\psi(x)$  satisfacen:*

- a) *El funcional  $J_\alpha$  es coercivo, es decir, para toda sucesión  $\{x_n\} \subset X$  tal que el valor del funcional  $\{J_\alpha(x_n)\}$  está uniformemente acotado, entonces la sucesión  $\{x_n\}$  está uniformemente acotada en  $X$ .*
- b) *El funcional  $\psi$  es secuencialmente débilmente semicontinuo inferiormente.*
- c) *El operador  $K$  es acotado.*

*Entonces, para todo  $\alpha > 0$ , existe al menos un mínimo  $x_\alpha \in X$  para el funcional  $J_\alpha$  de la Definición 3.4.*

*Demostración.* Como los funcionales  $\frac{1}{2}\|Kx - y\|_Y^2$  y  $\psi$  son estrictamente positivos, existe una sucesión minimizante  $\{x_n\} \subset G$ , tal que

$$\lim_{n \rightarrow \infty} J_\alpha(x_n) = \lim_{n \rightarrow \infty} \left( \frac{1}{2}\|Kx_n - y\|_Y^2 + \alpha\psi(x_n) \right) = \inf_{x \in G} \left( \frac{1}{2}\|Kx - y\|_Y^2 + \alpha\psi(x) \right) := \eta.$$

Por tanto, la sucesión de valores del funcional  $\{J_\alpha(x_n)\}$  está uniformemente acotada. También, gracias a la hipótesis a) del teorema, deducimos que existe una subsucesión, que seguiremos denotando  $\{x_n\}$ , y un  $x^* \in X$  tal que  $x_n^* \rightharpoonup x^*$  débilmente. Como el conjunto de corrección  $G$  es cerrado y convexo, deducimos que  $x^* \in G$ . Ahora como  $\psi$  es semicontinua inferiormente por hipótesis b) y el primer término también lo es, ya que las normas son semicontinuas inferiormente, se deduce:

$$\begin{aligned} \frac{1}{2}\|Kx^* - y\|_Y^2 + \alpha\psi(x^*) &\leq \liminf_{n \rightarrow \infty} \frac{1}{2}\|Kx_n - y\|_Y^2 + \alpha \liminf_{n \rightarrow \infty} \psi(x_n) \\ &\leq \liminf_{n \rightarrow \infty} \left( \frac{1}{2}\|Kx_n - y\|_Y^2 + \alpha\psi(x_n) \right) = \eta, \end{aligned}$$

i.e.,  $x^*$  es un mínimo para  $J_\alpha$ . Esto termina la demostración.  $\square$

El modelo de Tikhonov es estable, esto es, el mínimo  $x_\alpha$  depende de forma continua de los datos  $y$ , como se verá a continuación.

**Teorema 3.6** *Supongamos que se verifican las hipótesis del Teorema 3.5. Sea la sucesión  $\{y_n\} \subset Y$  convergente a  $y \in Y$ , y sea  $x_n$  un mínimo para el funcional  $J_\alpha$  con  $y_n$  en lugar de  $y$ . Entonces, la sucesión  $\{x_n\}$  contiene una subsucesión convergente a un mínimo de  $J_\alpha$ . Más aún, si el mínimo de  $J_\alpha$  es único, entonces la sucesión entera converge.*

*Demostración.* Sea  $x_\alpha$  un mínimo para  $J_\alpha$ . Como  $x_n$  minimiza el funcional, se tiene:

$$\frac{1}{2}\|Kx_n - y_n\|^2 + \alpha\psi(x_n) \leq \frac{1}{2}\|Kx_\alpha - y_n\|^2 + \alpha\psi(x_\alpha).$$

Por tanto las sucesiones  $\{\|Kx_n - y_n\|^2\}$  y  $\{\psi(x_n)\}$  están uniformemente acotadas. Ahora, haciendo uso de la desigualdad triangular

$$\frac{1}{2}\|Kx_n - y\|^2 \leq 2 \left( \frac{1}{2}\|Kx_n - y_n\|^2 + \frac{1}{2}\|y_n - y\|^2 \right),$$

se deduce que la sucesión  $\{\|Kx_n - y\|^2\}$  está también uniformemente acotada. Entonces, por la hipótesis a), la sucesión  $\{x_n\}$  está uniformemente acotada en  $X$ , y existe una subsucesión, que por abuso de notación denotaremos también por  $\{x_n\}$ , convergente en  $X$  a algún  $x^* \in G$ . Por la continuidad de  $K$  y la convergencia de  $y_n$  a  $y$  en  $Y$ , se tiene que  $Kx_n - y_n \rightharpoonup Kx^* - y$  débilmente. Ahora, la propiedad de la secuencialmente débil semicontinuidad inferior de las normas y la hipótesis b) implican que

$$\begin{aligned} \frac{1}{2}\|Kx^* - y\|_Y^2 + \alpha\psi(x^*) &\leq \liminf_{n \rightarrow \infty} \frac{1}{2}\|Kx_n - y_n\|^2 + \alpha \liminf_{n \rightarrow \infty} \psi(x_n) \\ &\leq \liminf_{n \rightarrow \infty} \left( \frac{1}{2}\|Kx_n - y_n\|^2 + \alpha\psi(x_n) \right) \\ &\leq \liminf_{n \rightarrow \infty} \left( \frac{1}{2}\|Kx_\alpha - y_n\|^2 + \alpha\psi(x_\alpha) \right) \\ &= \left( \frac{1}{2}\|Kx_\alpha - y\|^2 + \alpha\psi(x_\alpha) \right). \end{aligned}$$

Por tanto, el límite  $x^*$  de la subsucesión es un mínimo para  $J_\alpha$ . Si el mínimo  $x_\alpha$  es único, entonces trivialmente  $x^* = x_\alpha$ , y toda subsucesión contiene una subsucesión convergente a  $x^*$ . Por tanto, toda la sucesión converge, lo que termina la demostración del teorema.  $\square$

**Nota 3.7** Los resultados anteriores también se tienen para un término de fidelidad más general  $\|Kx - y\|^p$ , sin necesidad de trabajar en un espacio de Hilbert.

Ahora nos centraremos en el comportamiento del mínimo  $x_\alpha$  cuando el nivel de ruido  $\delta$  tiende a cero. Una cuestión fundamental en el marco del método de regularización de Tikhonov es ver cuándo la solución aproximada  $x_\alpha$  converge a la solución verdadera, sin ruido, lo que denominaremos solución  $\psi$ -minimizante.

**Definición 3.8** Un elemento  $x^\dagger$  de  $X$  es una solución  $\psi$ -minimizante si satisface lo siguiente:

$$Kx^\dagger = y^\dagger \quad y \quad \psi(x^\dagger) \leq \psi(x), \quad \forall x \in \{x \in G : Kx = y^\dagger\},$$

donde  $y^\dagger$  representa los datos ideales, no afectados por error.

**Teorema 3.9** *Supongamos que se verifican las hipótesis del Teorema 3.5. Entonces, existe al menos una solución  $\psi$ -minimizante para el problema (1.1).*

*Demostración.* Procedamos por reducción al absurdo y supongamos que no existe una solución  $\psi$ -minimizante en  $G$ . Sabemos que existe una sucesión  $\{x_n\} \subset G$  de soluciones para (1.1) tales que  $\psi(x_n) \rightarrow \beta$  y

$$\beta < \psi(x), \quad \forall x \in \{x \in G : Kx = y^\dagger\}.$$

Por consiguiente, el funcional  $J_\alpha(x_n)$  (para cualquier  $\alpha > 0$  fijo) con  $y^\dagger$  en lugar de  $y$  está uniformemente acotado, y por la coercividad de  $J_\alpha$  (hipótesis a)), la sucesión  $\{x_n\}$  contiene una subsucesión, que por abuso de notación también notaremos  $\{x_n\}$ , que converge débilmente a un  $x^* \in G$ . Luego  $Kx^* = y^\dagger$ , y por la hipótesis b), se tiene que  $\psi(x^*) = \lim_{n \rightarrow \infty} \psi(x_n) = \beta$ . Esto es una contradicción, lo que completa la prueba.  $\square$

Para el siguiente resultado, asumiremos que no estamos trabajando con espacios de Hilbert y  $J_\alpha$  viene dado por

$$J_\alpha(x) = \|Kx - y\|_Y^p + \alpha\psi(x).$$

**Teorema 3.10** *En las condiciones del Teorema 3.5, sea  $\{y^n\}_n \subset Y$  una sucesión de datos afectados de ruido, con nivel de ruido  $\delta_n = \|y^\dagger - y^n\| \rightarrow 0$ . Si el parámetro de regularización  $\alpha_n = \alpha(\delta_n)$  satisface:*

$$\lim_{n \rightarrow \infty} \frac{\delta_n^p}{\alpha_n} = 0 \quad y \quad \lim_{n \rightarrow \infty} \alpha_n = 0,$$

entonces la sucesión de mínimos  $\{x_{\alpha_n}\}$  tiene una subsucesión convergente a una sucesión  $\psi$ -minimizante  $x^\dagger$ . Además, si la solución  $\psi$ -minimizante  $x^\dagger$  es única, entonces la sucesión completa converge.

*Demostración.* Como  $x_{\alpha_n}$  minimiza el funcional, tenemos

$$\|Kx_{\alpha_n} - y^n\|^p + \alpha_n \psi(x_{\alpha_n}) \leq \|Kx^\dagger - y^n\|^p + \alpha_n \psi(x^\dagger) \leq \delta_n^p + \alpha_n \psi(x^\dagger).$$

Por la elección de la sucesión  $\{\alpha_n\}$ , las sucesiones  $\{\|x_{\alpha_n} - y^n\|\}$  y  $\{\psi(x_{\alpha_n})\}$  están ambas uniformemente acotadas. Además (desigualdad triangular), podemos escribir

$$\|Kx_{\alpha_n} - y^\dagger\|^p \leq 2^{p-1}(\|Kx_{\alpha_n} - y^n\|^p + \|y^n - y^\dagger\|^p).$$

Por consiguiente, la sucesión  $\{\|Kx_{\alpha_n} - y^\dagger\|^p\}$  está también uniformemente acotada. Por la hipótesis a), la sucesión  $\{x_{\alpha_n}\}_n$  está uniformemente acotada, y contiene una subsucesión que denotaremos igual y converge débil a un  $x^*$ . Como  $G$  es cerrado y convexo,  $x^* \in G$ .

De la semicontinuidad inferior de las normas, deducimos

$$\|Kx^* - y^\dagger\|^p \leq \liminf_{n \rightarrow \infty} \|Kx_{\alpha_n} - y^n\|^p \leq \limsup_{n \rightarrow \infty} \|Kx_{\alpha_n} - y^n\|^p + \alpha_n \psi(x_{\alpha_n}) = 0,$$

i.e.,  $Kx^* = y^\dagger$ . A continuación probaremos que  $x^*$  es una solución  $\psi$ -minimizante. De la propiedad minimizante de  $u_{\alpha_n}$  y de la semicontinuidad inferior del funcional  $\psi$ , se sigue que

$$\psi(x^*) \leq \liminf_{n \rightarrow \infty} \psi(u_{\alpha_n}) \leq \limsup_{n \rightarrow \infty} \psi(u_{\alpha_n}) \leq \lim_{n \rightarrow \infty} \left( \frac{\delta_n^p}{\alpha_n} + \psi(x^\dagger) \right) = \psi(x^\dagger).$$

La primera parte queda demostrada con lo anterior y sabiendo que  $x^\dagger$  es una solución  $\psi$ -minimizante. Si la solución  $\psi$ -minimizante  $x^\dagger$  es única, entonces  $x^* = x^\dagger$ . Claramente, toda subsucesión de  $\{u_{\alpha_n}\}$  contiene una subsucesión convergente a  $x^\dagger$ , por ello la sucesión completa converge. Esto demuestra el teorema.  $\square$

Los resultados anteriores muestran propiedades de existencia de una solución  $x_\alpha$  y estabilidad para el funcional de regularización de Tikhonov  $J_\alpha$ , pero para poder aplicar todos estos resultados de minimización del funcional, debemos hacer una elección adecuada del parámetro  $\alpha$ . En el capítulo siguiente veremos cómo podría estimarse dicho parámetro.

## Capítulo 4

# Elección del parámetro de regularización

Para elegir el parámetro de regularización, debemos recordar la Definición 3.2, de la cual podemos intuir que cualquier elección de  $\alpha$  debe tener en cuenta que  $\alpha$  tiene que ser un parámetro suficientemente pequeño. Veamos en este capítulo cómo hacer una elección adecuada del mismo.

### 4.1. Overfitting y underfitting

A modo de motivación de la importancia de la elección del parámetro  $\alpha$ , en esta sección serán presentados los dos retos más importantes del *machine learning*: el sobreajuste (en inglés *overfitting*) y el subajuste (o *underfitting*, que en el caso que nos compete también puede ser designado como *oversmoothing*). Estos dos conceptos pueden definirse como:

- **Underfitting.** Problema que surge cuando el algoritmo no es capaz de obtener un error suficientemente bajo en el conjunto de entrenamiento. En este caso, la respuesta del algoritmo no se ajusta a los datos del modelo y es necesario reprogramar el algoritmo para que tenga en cuenta las exigencias del mismo.
- **Overfitting.** Brecha significativa entre el error de entrenamiento y el error de test, esto es, cuando el algoritmo está tan específicamente programado que sólo está en condiciones de extraer resultados dentro del conjunto de entrenamiento. Para una observación nueva con un modelo ligeramente distinto, el algoritmo no dará una solución fiable.

Para el caso de un problema inverso descrito por la ecuación (1.1) y regularizado por (3.2), ambos problemas surgen por una mala elección del parámetro  $\alpha$ : para  $\alpha \gg$ , la norma del término de regularización del funcional (3.2) debe hacerse pequeña para poder minimizar el funcional, lo que implica que nuestra solución será una función extremadamente suave que además  $x_\alpha = \arg \min_{x \in G} J_\alpha(x) \rightarrow 0$ , lo que no es la solución buscada en ningún caso. Estamos ante un caso de *underfitting*.

Por contra, para el caso en que  $\alpha \ll$ , el término de regularización no tendrá peso alguno y el ajuste vendrá determinado exclusivamente por el término de fidelidad, lo que nos lleva a un problema de *overfitting*. De hecho, el escenario es incluso peor, ya que las medidas tomadas  $y = y_\delta$  tienen un error de nivel  $\delta > 0$ , lo que implica que ni siquiera un ajuste estricto al modelo sería fiable por la amplificación de la componente ruidosa, como se vio en el Ejemplo 2.2.

Las Figuras 4.1 y 4.2 representan la solución regularizada de Tikhonov para el problema inverso clásico de deconvolución en 1D. Conviene recordar aquí la definición de convolución.

**Definición 4.1 (Convolución)** Dadas dos funciones  $f, g$ , se define la convolución de  $f$  y  $g$ , y se denotará  $f * g$ , como la integral del producto de ambas funciones después de haber desplazado una de ellas una distancia  $t$ :

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\eta)g(t - \eta)d\eta. \quad (4.1)$$

Por tanto, el problema inverso consistiría en, dada una señal convolucionada  $f * g$ , obtener la señal  $f$  (resp.  $g$ ) previa a la convolución. A este proceso se le denomina deconvolución, que se puede apreciar en las Figuras 4.1 y 4.2 para dos señales (funciones) empleando la regularización de Tikhonov. En ambas figuras se pone de manifiesto una mala elección del parámetro  $\alpha$ . Para  $\alpha = 1$  tenemos un caso claro de *oversmoothing*: para la señal suave (a la izquierda), la recompuesta dista mucho de la verdadera y para la señal puntiaguda (derecha), los picos no están suficientemente bien estimados. Por otro lado, para  $\alpha = 0.01$  se amplifica en gran medida la señal ruidosa, que no se parece a la solución libre de ruido para ninguna señal. Tenemos *overfitting* en este caso.

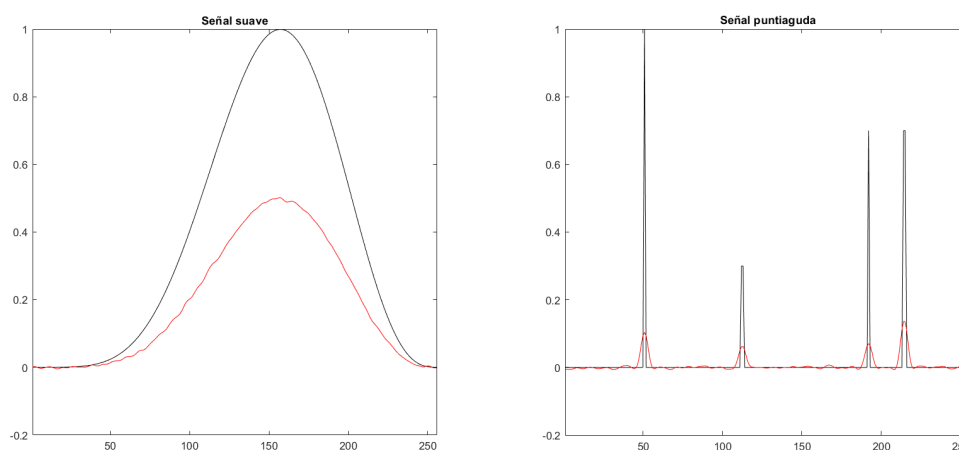


Figura 4.1: Solución del algoritmo de regularización de Tikhonov en 1D para  $\alpha = 1$ . En rojo, la señal deconvolucionada. En negro, la solución analítica verdadera, libre de ruido.

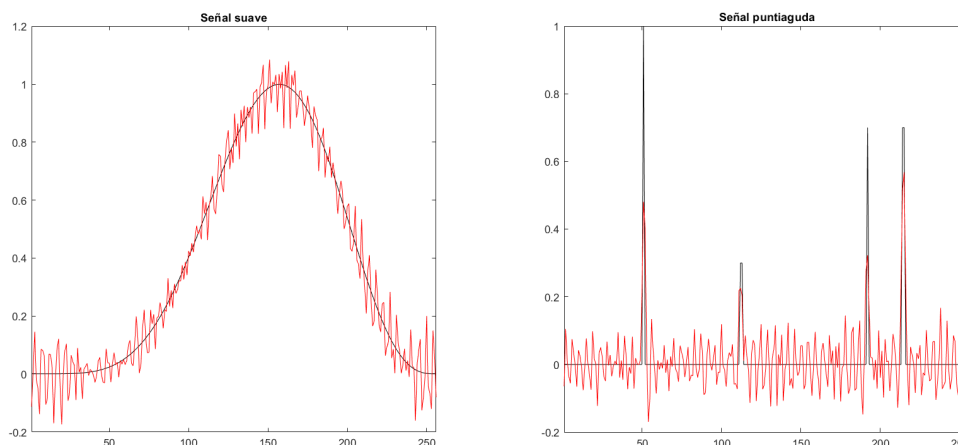


Figura 4.2: Solución del algoritmo de regularización de Tikhonov en 1D para  $\alpha = 0.01$ .

## 4.2. Elección a priori

Ya consideramos en el apartado anterior condiciones necesarias para conseguir que la solución aproximada del problema inverso se acercase a la solución real. Una de esas condiciones se expone en el Teorema 3.10, del cual podemos inferir:

- $\alpha(\delta) \rightarrow 0$  si  $\delta \rightarrow 0$ .
- $\frac{\delta^p}{\alpha(\delta)} \rightarrow 0$ . En un espacio de Hilbert, podemos particularizar el parámetro  $p$  tomando  $p = 2$ , ya que sabemos que el término de fidelidad es cuadrático, por lo que tenemos que  $\frac{\delta^2}{\alpha(\delta)} \rightarrow 0$ . Una elección posible sería  $\alpha(\delta) = C\delta^\mu$ ,  $\mu \in (0, 2)$ , con  $C$  una constante positiva.

La elección *a priori* del parámetro  $\alpha$  en función del nivel de ruido  $\delta$  ha sido utilizada habitualmente, si bien no es muy útil a efectos prácticos. Una elección más conveniente del parámetro debe llevarse a cabo *a posteriori*, en función de los datos obtenidos y del nivel de ruido.

## 4.3. Elección a posteriori

La elección del parámetro de regularización  $\alpha$  *a posteriori* es un problema altamente no trivial. En el siguiente apartado veremos dos métodos importantes para llevar a cabo dicha elección, a saber: el principio de discrepancia (que se debe a Morozov) y el principio de balance (véase [13], [12]). Para todos los desarrollos posteriores estaremos suponiendo trabajar en un espacio de Hilbert  $Y$ , con un término de fidelidad cuadrático en el funcional de regularización, esto es  $\frac{1}{2}\|Kx - y\|^2$ .

Comenzaremos por algunas definiciones y resultados básicos, como puede ser el cálculo de la función valor, importante para la elección del parámetro y para la implementación numérica de la regularización de Tikhonov.

**Definición 4.2** Dado el funcional de regularización de Tikhonov  $J_\alpha(x)$ , se define la función valor  $F(\alpha)$  como sigue:

$$\begin{aligned} F(\alpha) &: \mathbb{R}^+ \rightarrow \mathbb{R}, \\ F(\alpha) &= \inf_{x \in G} J_\alpha(x). \end{aligned} \quad (4.2)$$

Veamos a continuación resultados relevantes acerca de la función valor.

**Teorema 4.3** *La función valor  $F(\alpha)$  es cóncava no decreciente.*

*Demostración.* Dado  $\bar{\alpha} < \alpha$  y sabiendo que el funcional  $\psi(x)$  es no negativo para todo  $x \in X$ , se tiene:

$$F(\bar{\alpha}) = J_{\bar{\alpha}}(x) = \phi(x, y) + \bar{\alpha}\psi(x) \leq \phi(x, y) + \alpha\psi(x),$$

donde  $\phi(x, y) = \frac{1}{2}\|Kx - y\|_Y^2$  denota el término de fidelidad del funcional de Tikhonov. Tomando ínfimo en lo anterior con respecto a  $x \in G$ , obtenemos  $F(\bar{\alpha}) \leq F(\alpha)$ . Esto demuestra la monotonía de la función valor  $F(\alpha)$ .

Demostremos a continuación la concavidad de la función  $F$ . Sean  $\alpha_1 > 0$  y  $\alpha_2 > 0$  fijos, y consideremos  $\alpha_t = (1 - t)\alpha_1 + t\alpha_2$  para algún  $t \in [0, 1]$ . Entonces

$$\begin{aligned} F((1 - t)\alpha_1 + t\alpha_2) &= \inf_{x \in G} J_{\alpha_t}(x) = \inf_{x \in G} \{\phi(x, y) + ((1 - t)\alpha_1 + t\alpha_2)\psi(x)\} \\ &\geq (1 - t) \inf_{x \in G} J_{\alpha_1}(x) + t \inf_{x \in G} J_{\alpha_2}(x) = (1 - t)F(\alpha_1) + tF(\alpha_2). \end{aligned}$$

Por tanto, la función  $F(\alpha)$  es cóncava.

**Corolario 4.4** *La función valor  $F(\alpha)$  es continua.*

*Demostración.* Se deduce trivialmente de la concavidad de  $F$ . Ahora estudiaremos la diferenciabilidad de la función  $F$ . Con este fin, procedemos a definir:

**Definición 4.5** Para la función valor  $F(\alpha)$ , las derivadas de Dini (o *derivadas unilaterales*) de  $F$  se definen como sigue:

$$\begin{aligned} D^- F(\alpha) &= \lim_{h \rightarrow 0^+} \frac{F(\alpha) - F(\alpha - h)}{h}, \\ D^+ F(\alpha) &= \lim_{h \rightarrow 0^+} \frac{F(\alpha + h) - F(\alpha)}{h}. \end{aligned}$$

**Lema 4.6** *Las derivadas unilaterales  $D^- F(\alpha)$  y  $D^+ F(\alpha)$  de la función valor  $F(\alpha)$  existen para todo  $\alpha > 0$ . Más aún,  $D^\pm F(\alpha) \geq 0$ .*



*Demostración.* Para un  $\alpha > 0$  dado, tomemos  $h_1$  y  $h_2$  tales que  $0 < h_1 < h_2 < \alpha$ , y definamos  $t = 1 - \frac{h_1}{h_2} < 1$ . (Esto es,  $h_1 = h_2 - h_2 t$ .) Entonces vemos que  $\alpha - h_1 = t\alpha + (1 - t)(\alpha - h_2)$ . De la concavidad de  $F$  se deduce:

$$F(\alpha - h_1) \geq tF(\alpha) + (1 - t)F(\alpha - h_2) = \left(1 - \frac{h_1}{h_2}\right) F(\alpha) + \frac{h_1}{h_2} F(\alpha - h_2).$$

Despejando en la expresión anterior, obtenemos

$$\frac{F(\alpha) - F(\alpha - h_1)}{h_1} \leq \frac{F(\alpha) - F(\alpha - h_2)}{h_2}.$$

Por consiguiente, la sucesión  $\left\{\frac{F(\alpha) - F(\alpha - h)}{h}\right\}_{h>0}$  es monótona decreciente cuando  $h$  tiende a cero y acotada inferiormente por cero, lo que implica que  $\lim_{h \rightarrow 0^+} \frac{F(\alpha) - F(\alpha - h)}{h}$  existe, i.e., la derivada de Dini  $D^-F(\alpha)$  existe. Un razonamiento análogo prueba la existencia de la derivada lateral  $D^+F(\alpha)$ .  $\square$

Se observa que los resultados anteriores sobre la función valor  $F(\alpha)$  no requieren la existencia de un mínimo  $x_\alpha$  del funcional de regularización  $J_\alpha$ , y son válidos en cualquier espacio de funciones. De hecho, son válidos para operadores  $K$  no lineales. Con lo anterior, estamos en condiciones de introducir el principio de discrepancia de Morozov. Esta elección *a posteriori* del parámetro de regularización es la más utilizada hoy en día, pues sólo requiere el conocimiento del nivel de ruido  $\delta$  y resulta sencilla de implementar a nivel numérico.

### 4.3.1. Principio de discrepancia de Morozov

**Definición 4.7** Según el principio de discrepancia, la elección óptima del parámetro  $\alpha = \alpha(\delta)$  es aquella que verifica:

$$\|Kx_\alpha - y\| = C\delta, \tag{4.3}$$

donde  $C \geq 1$  es una constante fija. Una versión menos restrictiva del mismo establece una elección de  $\alpha$  tal que:

$$C_1\delta \leq \|Kx_\alpha - y\| \leq C_2\delta,$$

para ciertas constantes  $1 \leq C_1 < C_2$ .

Nótese que la motivación de definir el principio de esta manera reside en que no podemos esperar una solución del problema inverso  $x_\alpha$  más exacta que los datos  $y$  en términos del nivel de ruido  $\delta$ . Nótese también que dicha solución  $x_\alpha$  sólo ha sido definida de manera implícita a través del funcional  $J_\alpha$ , lo que quiere decir que la aplicación de este principio suele requerir la resolución de la ecuación (3.2), altamente no lineal.

**Lema 4.8** *El principio de discrepancia se puede entender en términos de la función valor  $F(\alpha)$  como el punto crítico de la función  $\Gamma(\alpha)$  definida por:*

$$\Gamma(\alpha) := \frac{F(\alpha) - \frac{\delta^2}{2}}{\alpha^{\frac{1}{1-\gamma}}}, \quad \delta \in [0, 1)$$

cuando  $\gamma = 0$ .

*Demostración.* El punto crítico de la función  $\Gamma(\alpha)$  se obtiene imponiendo:

$$\Gamma'(\alpha) = \frac{F'(\alpha)}{\alpha^{\frac{1}{1-\gamma}}} - \frac{F(\alpha) - \frac{\delta^2}{2}}{(1-\gamma)\alpha^{\frac{1}{1-\gamma}+1}} = 0,$$

donde  $'$  denota derivada con respecto a  $\alpha$ . Despejando, se obtiene

$$F(\alpha) - \alpha F'(\alpha) + \alpha\gamma F'(\alpha) = \frac{\delta^2}{2}. \quad (4.4)$$

Por otro lado, si existe  $F'(\alpha)$  con  $\alpha > 0$ , de (3.2) y (4.2) se sigue que

$$F(\alpha) = \inf_{x \in G} J_\alpha(x) = \phi'(x, y) + \alpha\psi'(x).$$

Renombrando los sumandos de la expresión anterior como función de  $\alpha$ , tenemos:

$$\bar{\phi}(\alpha) = \phi'(x, y),$$

$$\bar{\psi}(\alpha) = \psi'(x).$$

Es inmediato por la definición de  $F(\alpha)$  verificar lo siguiente:

$$F'(\alpha) = \psi'(x) = \bar{\psi}(\alpha),$$

$$F(\alpha) - \alpha F'(\alpha) = \phi'(x, y) = \bar{\phi}(\alpha).$$

Y sustituyendo en (4.4), la expresión queda:

$$\bar{\phi}(\alpha) + \alpha\gamma\bar{\psi}(\alpha) = \frac{\delta^2}{2},$$

que para  $\gamma = 0$  y fidelidad cuadrática tenemos que

$$\bar{\phi}(\alpha) = \frac{1}{2}\|Kx_\alpha - y\|_Y^2 = F(\alpha) - \alpha F'(\alpha) = \frac{\delta^2}{2}, \quad (4.5)$$

de donde es evidente ver que

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

definido salvo una constante  $C \geq 1$ , es decir, obtenemos (4.3), que es el principio de discrepancia, y hemos terminado la demostración del lema.  $\square$

**Nota 4.9** A la función  $\Gamma(\alpha)$  se le suele denotar como *principio de discrepancia generalizado* o *principio de discrepancia amortiguado*.

De la demostración anterior podemos intuir que una buena forma de inferir el parámetro  $\alpha$  a partir del principio de discrepancia es resolviendo la ecuación (4.5). Sin embargo, la función valor  $F(\alpha)$  es altamente no lineal. La estrategia en este caso será aproximar  $F(\alpha)$  por una función racional, que refinaremos mediante un método numérico iterativo.

**Definición 4.10** Llamaremos *función modelo* o *aproximación de Padé* para  $F(\alpha)$  a la función  $m_k(\alpha)$  definida como sigue:

$$F_k(\alpha) \approx m_k(\alpha) = r + \frac{s_k}{t_k + \alpha},$$

donde  $r$ ,  $s_k$  y  $t_k$  son constantes a determinar. El subíndice  $k$  hace referencia a las iteraciones.

Normalmente, la constante  $r$  suele determinarse mediante cálculo asintótico en  $\alpha$ , esto es, evaluando  $m(0^+)$  o bien  $m(+\infty)$ . Sin pérdida de generalidad, consideraremos

$$r = \lim_{\alpha \rightarrow \infty} F(\alpha). \quad (4.6)$$

Imponiendo condiciones de interpolación de Hermite para  $\alpha_k$ , obtenemos las relaciones:

$$m_k(\alpha_k) = F(\alpha_k), \quad m'_k(\alpha_k) = F'(\alpha_k),$$

lo que nos lleva a

$$m_k(\alpha) = r + \frac{s_k}{t_k + \alpha_k} = F(\alpha_k) \Rightarrow s_k = (F(\alpha_k) - r)(t_k + \alpha_k),$$

$$m'_k(\alpha) = \frac{-s_k}{(t_k + \alpha_k)^2} = F'(\alpha_k) \Rightarrow F'(\alpha_k) = \frac{(F(\alpha_k) - r)(t_k + \alpha_k)}{(t_k + \alpha_k)^2}.$$

De la segunda ecuación es fácil ver que

$$t_k + \alpha_k = -\frac{(F(\alpha_k) - r)}{F'(\alpha_k)}.$$

Sustituyendo para  $s_k$  y recordando que  $\bar{\psi}(\alpha_k) = F'(\alpha_k)$ , llegamos a

$$s_k = -\frac{(F(\alpha_k) - r)^2}{\bar{\psi}(\alpha_k)} \quad (4.7)$$

y también

$$t_k = \frac{(r - F(\alpha_k))}{\bar{\psi}(\alpha_k)} - \alpha_k. \quad (4.8)$$

En resumen, la ecuación (4.5) dada por el principio de discrepancia puede aproximarse por

$$m_k(\alpha) - \alpha m'_k(\alpha) = \frac{\delta^2}{2}.$$

La ecuación anterior también es no lineal, pero puede resolverse por el método de Newton, notando que

$$g(\alpha) = m_k(\alpha) - \alpha m'_k(\alpha) - \frac{\delta^2}{2} = 0$$

y el método para resolver  $g(\alpha) = 0$  sería

$$\alpha_{k+1} = \alpha_k - \frac{g(\alpha_k)}{g'(\alpha_k)}, \quad (4.9)$$

con  $k \geq 0$  y partiendo de un  $\alpha_0$  conocido. Sabemos que

$$g(\alpha_k) = m_k(\alpha_k) - \alpha_k m'_k(\alpha_k) - \frac{\delta^2}{2}, \quad (4.10)$$

$$g'(\alpha_k) = -\alpha_k m''_k(\alpha_k) \quad (4.11)$$

y la función modelo  $m_k(\alpha_k)$  y todas sus derivadas son conocidas.

En resumen, el análisis anterior nos lleva a generar un algoritmo de elección del parámetro y de minimización del funcional:

---

**Algoritmo 4.1:** Elección de  $\alpha$  por el método de la función modelo (principio de discrepancia)

---

1. Tomar  $\alpha_0$  (grande en virtud de la ecuación 4.6) y un nivel de tolerancia  $\theta \in (0, 1)$ .
  2. Calcular  $r = F(\alpha_0)$ , siendo  $F(\alpha_0) = \inf J_{\alpha_0}(x)$ .
  3. **Dados**  $\alpha_k, k \geq 0$ :
    - (3.a) Calcular  $x_{\alpha_k}$ , y actualizar los valores de  $s_k$  y  $t_k$  según las relaciones (4.7) y (4.8), respectivamente.
    - (3.b) Calcular la función modelo  $m(\alpha_k)$ .
    - (3.c) Actualizar el parámetro de regularización  $\alpha := \alpha_{k+1}$  via método de Newton (4.9), siendo necesarias las expresiones (4.10) y (4.11).
    - (3.d) Si  $|\alpha_k - \alpha_{k+1}| \leq \theta$ , **parar** y devolver los valores definitivos  $\alpha_k$  y  $x_{\alpha_k}$ .  
En otro caso, hacer  $k := k + 1$  y repetir el paso 3.
- 

### 4.3.2. Principio de balance

El segundo principio que consideraremos para una elección óptima del parámetro  $\alpha$  es el principio de balance (véase [13] y [12]).

**Definición 4.11** Para el funcional de Tikhonov definido en (3.2) y las relaciones

$$\bar{\psi}(\alpha) = F'(\alpha), \tag{4.12}$$

$$\bar{\phi}(\alpha) = F(\alpha) - \alpha F'(\alpha), \tag{4.13}$$

ya usadas en la demostración del Lema 4.8, el principio de balance establece que la elección óptima del parámetro  $\alpha > 0$  es aquella que verifica:

$$\bar{\phi}(\alpha) = \alpha \gamma \bar{\psi}(\alpha),$$

donde el parámetro  $\gamma$  denota el parámetro de forma de la distribución de probabilidad que gobierna el nivel de ruido.

Recuérdese que el parámetro de forma es cualquier parámetro de una familia de distribuciones de probabilidad que no es un parámetro de localización (i.e. uno que determina el desplazamiento de la distribución, como la media  $\mu$  en la distribución gaussiana) ni de escala (i.e. uno que determina la expansión de la distribución, como la desviación típica  $\sigma$  en la distribución normal), ni una función de ellos.

**Lema 4.12** *El principio de balance es equivalente a minimizar la función*

$$\Phi_\gamma(\alpha) := \frac{F^{\gamma+1}(\alpha)}{\alpha}. \tag{4.14}$$

*Demostración.* Para demostrar que el principio y el problema de minimización son equivalentes, procederemos a deducir uno a partir del otro y viceversa. Más precisamente,

⇒) Del principio de balance y las relaciones (4.12) y (4.13) se deduce que

$$0 = \bar{\phi}(\alpha) - \alpha\gamma\bar{\psi}(\alpha) = F(\alpha) - \alpha F'(\alpha) - \alpha\gamma F'(\alpha) = F(\alpha) - \alpha F'(\alpha)(1 + \gamma)$$

o también

$$F(\alpha) = \alpha F'(\alpha)(1 + \gamma).$$

Podemos reescribir esta expresión como

$$\frac{1}{\alpha} = \frac{F'(\alpha)}{F(\alpha)}(1 + \gamma) = \frac{dF/d\alpha}{F(\alpha)}(1 + \gamma),$$

de donde

$$\frac{d\alpha}{\alpha} = \frac{dF}{F(\alpha)}(1 + \gamma).$$

Integrando en la ecuación anterior, tenemos que

$$\ln \alpha + D = (1 + \gamma) \ln F(\alpha) + E,$$

donde  $D$  y  $E$  son constantes de integración. Sin pérdida de generalidad podemos suponer  $D = E$  y llegamos a

$$\alpha = \exp(1 + \gamma) \ln F(\alpha) = F(\alpha)^{1+\gamma},$$

que se puede reescribir como

$$\Phi_\gamma(\alpha) := \frac{F^{1+\gamma}(\alpha)}{\alpha} = 1,$$

que es la función a minimizar.

⇐) Ahora procedamos a minimizar  $\Phi_\gamma(\alpha)$ . Vemos que el mínimo se alcanza cuando

$$0 = \Phi'_\gamma(\alpha) = \frac{(1 + \gamma)F'(\alpha)F^\gamma(\alpha)\alpha - F^{1+\gamma}(\alpha)}{\alpha^2}.$$

De la ecuación anterior deducimos

$$(1 + \gamma)F'(\alpha)F^\gamma(\alpha)\alpha = F^{1+\gamma}(\alpha) \Rightarrow (1 + \gamma)F'(\alpha)\alpha = F(\alpha)$$

y sustituyendo (4.12) y (4.13), llegamos a

$$\bar{\phi}(\alpha) = \alpha\gamma\bar{\psi}(\alpha),$$

que es precisamente el principio de balance. Del procedimiento anterior también podemos extraer un método numérico para el cálculo del parámetro de regularización.

---

**Algoritmo 4.2:** Algoritmo de punto fijo (principio de balance)

---

1. Tomar  $\alpha_0$  (por ejemplo,  $\alpha_0 = \delta^\mu$ ,  $\mu \in (0, 2)$ , como se vio en el apartado de elección *a priori*) y un nivel de tolerancia  $\theta \in (0, 1)$  para el cálculo de  $\alpha_k$ .
  2. **Dados**  $\alpha_k$ , calcular  $x_{\alpha_k}$  (minimizando el funcional de Tikhonov).
  3. Actualizar el parámetro de regularización  $\alpha := \alpha_{k+1}$  según el principio de balance, esto es 
$$\alpha_{k+1} = \frac{1}{\alpha} \frac{\bar{\phi}(x_{\alpha_k})}{\psi(x_{\alpha_k})}.$$
  4. Si  $|\alpha_k - \alpha_{k+1}| \leq \theta$ , **parar** y devolver los valores definitivos  $\alpha_k$  y  $x_{\alpha_k}$ . En otro caso, hacer  $k := k + 1$  y volver al paso 2.
-

## Capítulo 5

# Aproximación numérica para problemas inversos

El *machine learning* suele requerir habitualmente una gran cantidad de algoritmos numéricos. Bajo este término se engloban todos aquellos algoritmos que buscan resolver problemas (inversos, en este caso) mediante procesos iterativos que encuentran soluciones aproximadas del problema que se van actualizando en cada iteración. La razón de emplear algoritmos numéricos para abordar este tipo de problemas es la imposibilidad de encontrar una solución analítica para los mismos, por lo ya expuesto anteriormente. Los problemas inversos están, por definición, mal planteados y la inestabilidad de la solución hace que encontrarla de forma analítica sea inviable; además de la imposibilidad, en general, de calcular soluciones analíticas de problemas diferenciales.

### 5.1. El algoritmo de descenso del gradiente estocástico

El algoritmo de descenso del gradiente estocástico (también denotado como SGD por sus siglas en inglés: *Stochastic Gradient Descent*) y todas sus variantes son probablemente los algoritmos de optimización más utilizados dentro de la rama de *machine learning*. La motivación de denominar a este algoritmo como “estocástico” surge de constatar el hecho de que el gradiente del funcional objetivo (el funcional de regularización de Tikhonov visto en la Sección 3.2) con respecto a las variables de entrada puede presentar cierto ruido (que proviene del ruido de los datos registrados), cuya magnitud y distribución estadística dependerá del problema inverso que estemos considerando.

Por otra parte, si bien es sabido cómo llevar a cabo el gradiente de una función, aún no se ha definido rigurosamente cómo calcular el gradiente de un funcional [5]. Para ello, se definirán los espacios de Hilbert con núcleo reproductor (también conocidos como RKHS por sus siglas en inglés: *Reproducing Kernel Hilbert Space*), espacios de funciones que subyacen tras numerosos

problemas inversos y que pueden estar dotados de la norma  $L^2$ , de acuerdo con el funcional de Tikhonov.

**Definición 5.1 (RKHS)** Sea  $\Omega \subset \mathbb{R}^n$  abierto. Se denomina espacio de Hilbert con núcleo reproductor  $G(x_i, \cdot) : \Omega \times \Omega \rightarrow \mathbb{R}$  y se denotará por  $\mathcal{H}_k$  (donde  $k$  no es un subíndice, sino que denota un *kernel*, un núcleo reproductor) al espacio de funciones  $u(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$  que pueden expresarse de la siguiente manera:

$$u(\cdot) = \int_{\Omega} f(x_i)G(x_i, \cdot)dx_i, \quad \forall f \in \mathcal{H}_k, i = 1, \dots, n$$

y equipado con el siguiente producto escalar:

$$\langle u, v \rangle = \int_{\Omega} f(x_i)g(x_j)G(x_i, x_j)dx_idx_j \quad i, j = 1, \dots, n$$

**Observación 5.2** Nótese que el producto escalar anterior satisface, por definición, la propiedad de linealidad:

- $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle \quad \forall \lambda \in \mathbb{R}$
- $\langle u_1 + u_2, v \rangle = \langle u_1, v \rangle + \langle u_2, v \rangle \quad \forall u_1, u_2 \in \mathcal{H}_k.$

**Observación 5.3** Nótese que en un RKHS, la propiedad reproductiva del núcleo garantiza que el funcional evaluación  $E_{x^0}[u]$  es equivalente a tomar el producto escalar con un elemento  $G(x_i, \cdot) \in \mathcal{H}_k$ . En efecto:

$$E_{x^0}[u] = u(x^0) = \int_{\Omega} f(x_i)G(x_i, x^0) = \int_{\Omega} f(x_i)G(x_i, \cdot)G(\cdot, x^0) = \langle u, G(x^0, \cdot) \rangle.$$

Procedamos ahora a definir el gradiente de un funcional:

**Definición 5.4** Sea  $\epsilon > 0$ . Sea  $X$  un espacio de funciones y  $u, v \in X$ . Se define de manera implícita el gradiente de un funcional  $\nabla J(u)$  como el término lineal de variación de una función cuando es sometida a una pequeña perturbación de entrada  $\epsilon$ :

$$J(u + \epsilon v) = J(u) + \epsilon \langle \nabla J(u), v \rangle + O(\epsilon^2).$$

Con la noción de gradiente de un funcional y suponiendo que trabajamos en un RKHS, podemos demostrar el siguiente lema.

**Lema 5.5** *El gradiente del funcional evaluación  $E_{x^0}[u] = u(x_0)$  en un RKHS es el núcleo reproductor de dicho espacio.*



*Demostración.* Por definición del funcional evaluación, se tiene:

$$\begin{aligned} E_{x^0}[u + \epsilon v] &= u(x^0) + \epsilon v(x^0) + 0 \\ &= u(x^0) + \epsilon \langle v, G(x^0, \cdot) \rangle + 0 \\ &= E_{x^0}[u] + \epsilon \langle v, \nabla E_{x^0}[u] \rangle + O(\epsilon^2). \end{aligned} \quad (5.1)$$

El objetivo de este desarrollo no es sino encontrar el gradiente del funcional de regularización de Tikhonov. Para ello, recordemos la expresión de dicho funcional de acuerdo con la Definición 3.4:

$$J_\alpha(u) = \frac{1}{2} \|Ku - y\|_2^2 + \frac{\alpha}{2} \|u\|_2^2, \quad (5.2)$$

definido sobre todas las funciones  $u \in X$ , siendo  $X$  un RKHS. Queremos encontrar la mejor función  $u \in X$  que minimiza dicho funcional, que obtendremos “moviéndonos” en la dirección negativa del gradiente:  $u - \eta_k \nabla J$ . (En esta expresión  $\eta_k$  es la tasa de aprendizaje, concepto que se verá a continuación.) Vamos a encontrar una expresión para  $\nabla J$ . El siguiente resultado resulta especialmente útil para problemas inversos de reconstrucción de imágenes: resonancias magnéticas, microscopía fluorescente, refinamiento de imágenes borrosas. . .

**Lema 5.6 (Gradiente del funcional de regularización)** *Sea  $X$  un RKHS y  $\Omega \subset \mathbb{R}^n$ . Sea  $K : X \rightarrow L^2(\Omega)$  un operador lineal con núcleo reproductor  $k(x, \cdot)$ . Entonces, el gradiente del funcional de regularización de Tikhonov (5.2) viene dado por la siguiente expresión:*

$$\nabla J_\alpha(u) = \|\sqrt{Ku - y}\|_2^2 k(x, \cdot) + \alpha u \quad (5.3)$$

*Demostración.* Tenemos  $J_\alpha(u) = \phi(u) + \psi(u)$ . Analicemos cada sumando del funcional por separado.

Para el **término de fidelidad** se tiene

$$\phi(u) = \frac{1}{2} \|Ku - y\|_2^2.$$

Haciendo uso de la regla de la cadena, obtenemos:

$$\begin{aligned} \nabla \phi(u) &= \nabla \left( \frac{1}{2} \int_{\Omega} (Ku(x) - y)^2 dx \right) \\ &= \frac{1}{2} \int_{\Omega} \nabla (Ku(x) - y)^2 dx \\ &= \left( \int_{\Omega} Ku(x) - y dx \right) \nabla (Ku(x)) \\ &= \left( \int_{\Omega} Ku(x) - y dx \right) \nabla (E_x[Ku]) \\ &= \left( \int_{\Omega} Ku(x) - y dx \right) k(x, \cdot), \end{aligned} \quad (5.4)$$

donde se ha empleado el Lema 5.5 para sustituir el gradiente del funcional evaluación.

Para el **término de regularización** se tiene

$$\psi(u) = \frac{\alpha}{2} \|u\|_2^2.$$

Podemos desarrollar la norma  $L^2$  de  $u$  usando una expansión de tipo Taylor:

$$\begin{aligned}
 \|u + \epsilon v\|_2^2 &= \int_{\Omega} (u + \epsilon v)^2 dx \\
 &= \langle u + \epsilon v \rangle_{L^2} \\
 &= \|u\|_2^2 + 2\langle u, \epsilon v \rangle + \epsilon^2 \|v\|_2^2 \\
 &= \|u\|_2^2 + \epsilon \langle 2u, v \rangle + O(\epsilon^2).
 \end{aligned} \tag{5.5}$$

Identificando el término lineal en  $\epsilon$  de la expresión anterior con el gradiente según la Definición 5.4, se observa

$$\nabla \psi(u) = \frac{\alpha}{2} \nabla \|u\|_2^2 = \frac{\alpha}{2} 2u = \alpha u.$$

Con lo que  $\nabla J_{\alpha}(u) = \nabla \phi(u) + \nabla \psi(u)$  y hemos terminado.  $\square$

De manera análoga al gradiente de una función, el gradiente del funcional de Tikhonov nos da la “dirección” en la que  $J_{\alpha}$  crece más rápido. Por tanto, podemos minimizar  $J_{\alpha}$  desplazándonos en la dirección negativa del gradiente. De manera intuitiva, el algoritmo de descenso del gradiente nos proporcionará en cada iteración un punto nuevo de nuestro espacio de funciones:

$$u_{k+1} = u_k - \eta_k \nabla J_{\alpha},$$

donde  $\eta_k$  es la tasa de aprendizaje, un escalar positivo que determina el tamaño del paso. Podemos elegir  $\eta_k$  de muchas maneras distintas. Un enfoque común es fijar  $\eta_k = \eta = cte$ . Sin embargo, lo más habitual en la práctica es reducir la tasa de aprendizaje en cada iteración, de ahí la notación de  $\eta_k$ . Dicho esto, a continuación se detalla un procedimiento formal para el algoritmo SGD.

**Algoritmo 5.1:** Descenso del gradiente estocástico (SGD) para el funcional de Tikhonov

1. Fijar una sucesión de tasas de aprendizaje  $\eta_k: \eta_1, \eta_2, \dots$ . Sin pérdida de generalidad, se puede elegir  $\eta_k$  linealmente decreciente hasta cierta iteración  $\tau: \eta_k = (1 - \beta)\eta_0 + \beta\eta_{\tau}$  para  $\beta = \frac{k}{\tau} \leq 1$ ,  $\eta_k = \eta_{\tau}$  para  $\beta > 1$ .
2. Elegir  $\eta_0, \eta_{\tau}, \tau$  en la expresión anterior (generalmente  $\eta_0$  arbitrario,  $\eta_{\tau} = \frac{\eta_0}{100}$ ,  $\tau$  del orden de  $10^2$ ) y fijar una tolerancia  $\theta \in (0, 1)$ .
3. **Dados**  $k > 0, u_k$ :
  - (3.a) Tomar una serie (ruidosa) de datos  $y_k = y_k^{\delta}$  correspondiente a  $u_k$ .
  - (3.b) Calcular el gradiente del funcional de Tikhonov:  $\nabla J_{\alpha}(u_k) = \|\sqrt{K}u_k - y_k\|_2^2 k(x, \cdot) + \alpha u_k$
  - (3.c) Tomar  $u_{k+1} = u_k - \eta_k \nabla J_{\alpha}(u_k)$
  - (3.d) Si  $\|u_{k+1} - u_k\| \leq \theta$ , **fin**. Devolver  $u_{k+1}$ . Si no, tomar  $k = k + 1$  y repetir el paso 3.

**Fin**

La convergencia del algoritmo SGD viene asegurada por el siguiente teorema que no se demostrará aquí, véase [7].

**Teorema 5.7 (Condición suficiente para convergencia de SGD)** *La sucesión  $\{u_k\}_{k \geq 1}$  definida en el Algoritmo 5.1. converge cuando  $k \rightarrow \infty$  si:*

$$\sum_{k=1}^{\infty} \eta_k = \infty, \quad y \quad \sum_{k=1}^{\infty} \eta_k^2 < \infty.$$

Si bien el algoritmo SGD es ampliamente utilizado en el ámbito de *machine learning*, sólo es posible definirlo de esta manera cuando nos encontremos en un RKHS, que podría no ser el caso. Por ello, en la siguiente sección se recordará otro algoritmo clásico más general, muy usado en la resolución de ecuaciones en derivadas parciales: el método de elementos finitos.

## 5.2. Método de Elementos Finitos aplicado a Problemas Inversos

### 5.2.1. Fundamento teórico

El objetivo de esta sección será describir brevemente la aplicación del Método de Elementos Finitos, ampliamente empleado para la resolución de ecuaciones en derivadas parciales, a la minimización del funcional de Tikhonov. Para ello, se usará la noción de derivada de Fréchet definida en el Capítulo 1, que nos será de gran utilidad ya que es la derivada empleada en espacios normados por antonomasia. Se probará una expresión de la derivada de Fréchet del funcional de Tikhonov y dicho resultado será útil, a su vez, en la aplicación del Método de Elementos Finitos [18].

Recordando las definiciones de las derivadas de Gateaux y de Fréchet introducidas en el Capítulo 1, el siguiente lema nos dará una expresión para la derivada de Fréchet del funcional de Tikhonov.

**Lema 5.8** *Sean  $X$  e  $Y$  dos espacios de Hilbert reales, y  $K : X \rightarrow Y$  un operador lineal continuo. Entonces, la derivada de Fréchet del funcional de regularización de Tikhonov*

$$J_{\alpha}(x) := \frac{1}{2} \|Kx - y\|_Y^2 + \frac{\alpha}{2} \|x\|_X^2 \quad (5.6)$$

viene dada por

$$J'_{\alpha}(x) = K^*Kx - K^*y + \alpha x, \quad (5.7)$$

donde  $K^*$  es el operador adjunto de  $K$ .

*Demostración.* Por simplicidad en el cálculo, se hallará la derivada de Gateaux del funcional de Tikhonov. Como el operador  $K$  es continuo por hipótesis, el funcional  $J_{\alpha}$  es Fréchet diferenciable y entonces las expresiones para las derivadas coinciden.

Calculemos, pues la derivada de  $2J_\alpha$  en  $x \in X$ , donde multiplicaremos toda la expresión del funcional por 2 por comodidad. Para algún  $h \in X$  se tiene

$$\begin{aligned}
 2J_\alpha(x+h) &= \|K(x+h) - y\|_Y^2 + \alpha\|x+h\|_X^2 \\
 &= \|Kx - y\|_Y^2 + \|Kh\|_Y^2 + 2\langle Kx - y, Kh \rangle_Y + \alpha\|x\|_X^2 + \alpha\|h\|_X^2 + 2\alpha\langle x, h \rangle_X \\
 &= 2J_\alpha(x) + 2\langle Kx - y, Kh \rangle_Y + 2\alpha\langle x, h \rangle_X + \|Kh\|_Y^2 + \alpha\|h\|_X^2 \\
 &= 2J_\alpha(x) + 2\langle K^*(Kx - y), h \rangle_X + 2\alpha\langle x, h \rangle_X + \|Kh\|_Y^2 + \alpha\|h\|_X^2 \\
 &= 2J_\alpha(x) + 2\langle \alpha x + K^*(Kx - y), h \rangle_X + \|Kh\|_Y^2 + \alpha\|h\|_X^2.
 \end{aligned} \tag{5.8}$$

Por tanto, la derivada de Gateaux de  $2J_\alpha$  en  $x$  en dirección  $h$  es

$$\begin{aligned}
 D_G 2J_\alpha(x)h &= \lim_{\tau \rightarrow 0} \frac{2J_\alpha(x + \tau h) - 2J_\alpha(x)}{\tau} \\
 &= \lim_{\tau \rightarrow 0} (2\langle \alpha x + K^*(Kx - y), h \rangle_X + \tau\|Kh\|_Y^2 + \tau\alpha\|h\|_X^2) \\
 &= 2\langle \alpha x + K^*Kx - K^*y, h \rangle_X \quad \forall h \in X.
 \end{aligned} \tag{5.9}$$

La derivada de Gateaux  $D_G 2J_\alpha(x)$  es un elemento de  $X^*$  que puede ser identificado con  $X$  por medio del producto escalar. Así pues, de (5.9) se tiene que

$$D_G 2J_\alpha(x) = 2(\alpha x + K^*Kx - K^*y),$$

y dividiendo por 2 y viendo que esta expresión es, efectivamente, la derivada de Fréchet por lo expuesto anteriormente, se tiene que

$$D_G J_\alpha(x) = J'_\alpha(x) = K^*Kx - K^*y + \alpha x,$$

como queríamos demostrar.  $\square$

### 5.2.2. Aplicación del método de elementos finitos

En lo sucesivo, denotaremos como  $\Omega \subset \mathbb{R}^d$ ,  $d = 2, 3$  el dominio de la solución con frontera  $\partial\Omega$ . Por simplicidad supondremos que  $\Omega$  es poligonal.

Para implementar el método de elementos finitos en este dominio se discretizará  $\Omega$  mediante el mallado (triangulación)  $T_h$ . Este mallado consiste en una familia finita de  $d$ -símplices  $T$  (triángulos cerrados si  $d = 2$ , tetraedros cerrados si  $d = 3$ ) con las propiedades siguientes:

- (i)  $\bar{\Omega} = \bigcup_{T \in T_h} T$ .
- (ii) Si  $T, T' \in T_h$  y  $T \neq T'$ , entonces  $T \cap T'$  sólo puede ser el vacío o un vértice, lado o cara común a  $T$  y  $T'$ .

La Figura 5.1 muestra una representación de dicha triangulación en un dominio  $\Omega \subset \mathbb{R}^2$ .

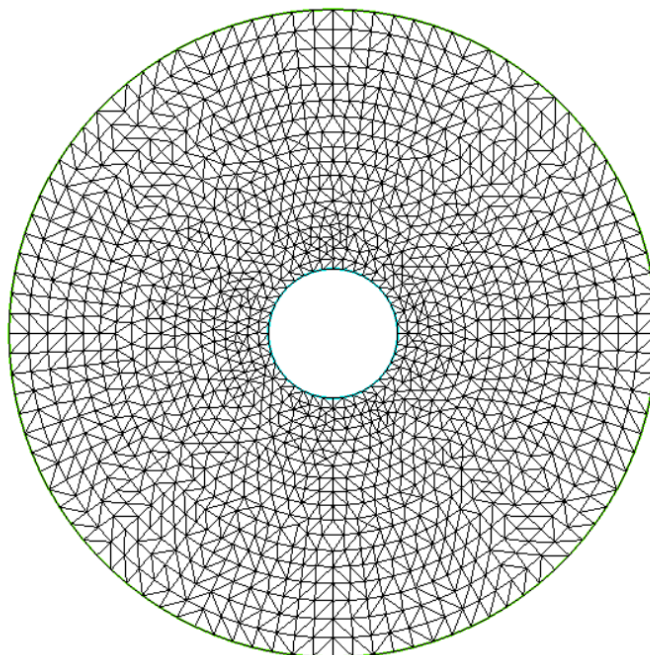


Figura 5.1: Triangulación de un dominio de  $\mathbb{R}^2$ .

A continuación, se define el espacio de elementos finitos  $V_h$  como

$$V_h = \{v \in C^0(\Omega) : v|_T = P_1(T), \forall T \in T_h\}, \tag{5.10}$$

donde para cada  $T$ ,  $P_1(T)$  es el espacio vectorial de las funciones polinómicas de grado  $\leq 1$  en las variables  $x_i, i = 1, \dots, d$ . La dimensión de  $P_1(T)$  es  $d + 1$ . Veamos un resultado interesante para este espacio  $V_h$  [18].

**Teorema 5.9**  $V_h \subset H^1(\Omega)$ . Más aún, la dimensión de  $V_h$  coincide con el número de vértices de  $T_h$ . De hecho, para todo  $a^i, i = 1, \dots, s$ , donde  $a^i$  denotan los vértices de  $T_h$ , y fijados  $\gamma^i \in \mathbb{R}$ , existe una única  $v_h \in V_h$  que cumple

$$v_h(a^i) = \gamma^i \quad \forall i = 1, \dots, s. \tag{5.11}$$

*Demostración.* Supongamos  $d = 3$ , los casos  $d = 1, 2$  se prueban de manera análoga. Sea  $v \in V_h$ . Consideremos la función  $w$  definida c.p.d. en  $\Omega$  como sigue:

$$w = \partial_1(v|_T) \quad \text{en } \overset{\circ}{T}, \quad \forall T \in T_h.$$

Veamos que  $w$  es la derivada generalizada de  $v$  respecto de  $x_1$ . En efecto, si  $\phi \in D(\Omega)$ , se tiene:

$$\int_{\Omega} w \phi dx = \sum_{T \in T_h} \int_T w \phi dx = \sum_{T \in T_h} \int_T \partial_1(v|_T) \phi dx \tag{5.12}$$

$$= \sum_{T \in T_h} \left( - \int_T v \partial_1 \phi dx + \int_{\partial T} v \phi n_1^T d\Gamma \right) \tag{5.13}$$

$$= - \int_{\Omega} v \partial_1 \phi dx + \sum_{T \in T_h} \int_{\partial T} v \phi n_1^T d\Gamma, \tag{5.14}$$

donde  $n_1^T = n_1^T(x)$  es, para cada  $T \in T_h$ , la primera componente del vector normal exterior a  $T$  en los puntos de  $\partial T$ . El segundo sumando en la expresión anterior es una suma de integrales sobre las caras de los tetraedros  $T \in T_h$ . Las integrales correspondientes a las caras que se encuentran en  $\partial\Omega$  se anulan, ya que  $\phi = 0$  sobre  $\partial\Omega$ . Como las otras caras son siempre comunes a dos tetraedros  $T \cap T'$ , se tiene que  $n_1^T = -n_1^{T'}$  (i.e. se suman dos veces, pero con signo contrario) y tenemos

$$\sum_{T \in T_h} \int_{\partial T} v \phi n_1^T d\Gamma = 0.$$

Por tanto

$$\int_{\Omega} w \phi dx = - \int_{\Omega} v \partial \phi dx,$$

y como  $\phi$  es arbitraria en  $D(\Omega)$ , tenemos lo anunciado. Procediendo de forma análoga para el resto de  $\partial_i v$ , concluimos que  $v \in H^1(\Omega)$ . Probemos ahora la segunda parte del teorema. Sea  $\{a^1, \dots, a^s\}$  el conjunto de los vértices de  $T_h$  y consideremos la aplicación lineal  $g : V_h \mapsto \mathbb{R}^s$  definida como sigue:

$$g v_h = (v_h(a^1), \dots, v_h(a^s)) \quad \forall v_h \in V_h.$$

Si probamos que  $g$  es biyectiva, quedará demostrado el teorema. Por una parte, si  $g v_h = 0$ , entonces en cada tetraedro  $T \in T_h$ ,  $v_h|_T$  es una función polinómica de grado  $\leq 1$  que se anula en los cuatro vértices. Trivialmente, esto implica  $v_h|_T \equiv 0$ . Por tanto,  $v_h = 0$  y concluimos que  $g$  es inyectiva. Por otra parte, sea  $A = (\gamma^1, \dots, \gamma^s)$  en  $\mathbb{R}^s$  dado. Es inmediato comprobar que, fijado un tetraedro  $T \in T_h$  de vértices  $a^t, a^u, a^v, a^w$ , existe una única función  $p^T$  en  $P_1(T)$  cuya imagen de dichos vértices es precisamente  $\gamma^t, \gamma^u, \gamma^v, \gamma^w \in A$ . Por tanto,  $v_h = p^T$  para dicho  $T$ . Repitiendo este procedimiento para cada tetraedro, se puede definir  $v_h$  en todo  $\bar{\Omega}$ . Nótese que en las caras comunes a dos tetraedros  $T, T'$ ,  $p^T$  y  $p^{T'}$  coinciden, pues son dos funciones polinómicas en dos variables idénticas en tres puntos no alineados. Por construcción,  $v_h \in V_h$  y  $g v_h = (\gamma^1, \dots, \gamma^s)$ . Luego  $g$  es sobreyectiva, y hemos terminado.  $\square$

Al espacio  $V_h$  definido anteriormente se le denomina *espacio de elementos finitos*, donde las ternas  $\{T, P_1(T), \Sigma_T\}$  (con  $\Sigma_T$  el conjunto de los vértices de  $T$ ) son elementos finitos de tipo  $P_1$ -Lagrange. Por tanto, el espacio  $V_h$  nos permite introducir el método de elementos finitos para (5.7). Este consiste en aproximar un problema planteado en un espacio de dimensión infinita por un problema en dimensión finita. Su aplicación está basada en una formulación débil (véase [6], [14]) que se puede expresar en términos de la derivada de  $J_\alpha$ , i.e. encontrar  $x_h \in V_h$  tal que  $\forall v \in V_h$  se tiene

$$J'_\alpha(x_h)(v) = \langle K^* K x_h - K^* y, v \rangle + \alpha \langle x_h, v \rangle = 0 \quad (5.15)$$

Por tanto, la solución real  $x$  del problema se aproxima por  $x_h \in V_h$  de manera que

$$x_h = \sum_{i=1}^N c_i \phi_i, \quad (5.16)$$

donde  $\{\phi_i\}_{i=1}^N$  son funciones lineales continuas a trozos y  $c_i$  denotan los valores de las funciones discretas desconocidas en el punto  $a_i \in T_h$ , i.e.  $x_h(a_i)$ . Introduciendo (5.16) en (5.15) con  $y = \phi_j$  y tomando valores discretos de  $y_i$  en el punto del mallado  $a_i \in T_{h_k}$ , obtenemos el siguiente sistema de ecuaciones discreto:

$$\sum_{i,j=1}^N \langle K\phi_i, K\phi_j \rangle x_i - \sum_{i,j=1}^N \langle y_i, K\phi_j \rangle + \alpha \sum_{i,j=1}^N \langle \phi_i, \phi_j \rangle x_i = 0. \quad (5.17)$$

Dicho sistema puede ser reescrito de la siguiente manera:

$$\sum_{i,j=1}^N \langle K\phi_i, K\phi_j \rangle x_i + \alpha \sum_{i,j=1}^N \langle \phi_i, \phi_j \rangle x_i = \sum_{i,j=1}^N \langle y_i, K\phi_j \rangle, \quad (5.18)$$

que es equivalente al siguiente sistema de ecuaciones lineal:

$$(C + \alpha A)\mathbf{x} = \mathbf{z}. \quad (5.19)$$

En el sistema (5.19),  $C$  y  $A$  son las matrices correspondientes a los dos sumandos del término de la izquierda en (5.18),  $\mathbf{x}$  denota el vector de valores nodales (i.e. los valores en los nodos del mallado del dominio) de la aproximación por elementos finitos  $x_h$  y  $\mathbf{z}$  es el correspondiente término de la derecha en (5.18).

Por último, ya que este método consiste en una aproximación numérica de  $x$  por  $x_h$ , solución por el método de elementos finitos de la ecuación (5.15), tenemos que ver que dichas  $x_h$  convergen hacia la solución  $x$  de (1.1). Se finalizará la sección con el siguiente resultado.

**Teorema 5.10** *Existe una constante  $C > 0$  tal que, para todo  $h > 0$ , se tiene:*

$$\|x - x_h\|_{H^1} \leq C \inf_{v_h \in V_h} \|x - v_h\|_{H^1}. \quad (5.20)$$

*Demostración.* Para cada  $v_h \in V_h$  tenemos que

$$\begin{aligned} \beta \|x - x_h\|_{H^1}^2 &\leq J_\alpha(x - x_h)(x - x_h) = \langle K(x - x_h) - y, x - x_h \rangle + \alpha \langle x - x_h, x - x_h \rangle \\ &= J_\alpha(x - x_h)(x - v_h) + J_\alpha(x - x_h)(v_h - x_h) \\ &= J_\alpha(x - x_h)(x - v_h) \\ &\leq M \|x - x_h\|_{H^1} \|x - v_h\|_{H^1}, \end{aligned}$$

donde la primera desigualdad se tiene para una constante  $\beta$  por ser  $J_\alpha$  coercivo, y la segunda desigualdad se tiene por Cauchy-Schwarz. Por consiguiente,

$$\|x - x_h\|_{H^1} \leq \frac{M}{\beta} \|x - v_h\|_{H^1} \quad \forall v_h \in V_h,$$

de donde se deduce (5.20). □

La interpretación de la cota anterior es que la aproximación en el espacio de elementos finitos es convergente a la solución real del problema.

## Capítulo 6

# Un problema inverso para la ecuación de ondas

A modo ilustrativo, podemos considerar un problema inverso ampliamente presente en electromagnetismo, que es la ecuación de ondas con permitividad eléctrica desconocida.

**PROBLEMA INVERSO 6.1. ECUACIÓN DE ONDAS INVERSA.** Consideremos el modelo ampliamente conocido de la propagación de una onda electromagnética en un medio no dispersivo y no magnético. Por simplicidad, consideramos el caso unidimensional. Sea  $\Omega \subset \mathbb{R}, \Omega = (-L, L)$ . El problema directo a considerar es de la forma:

$$\begin{cases} \epsilon_r u_{tt} = u_{xx} & \text{en } \Omega \times (0, T), \\ u(-L, t) = u(L, t) = 0 & \forall t \in (0, T), \\ u(x, 0) = \mu_0 \left(1 - \left(\frac{x}{L}\right)^2\right)^2, u_t(x, 0) = 0 & \text{en } \Omega, \end{cases} \quad (6.1)$$

donde  $\epsilon_r$  es la permitividad dieléctrica relativa del medio que estemos considerando. Su expresión se deduce teniendo en cuenta que la velocidad de una onda electromagnética en un medio arbitrario se expresa como  $c = \frac{1}{\sqrt{\epsilon_r \mu_r}}$ , que en un medio comúnmente no magnético se simplifica teniendo en cuenta que  $\mu_r \approx 1$ , por lo que  $c \approx \frac{1}{\sqrt{\epsilon_r}}$ . De ahí el coeficiente de la derivada temporal en el problema anterior.

En (6.1),  $u = u(x, t)$  es la amplitud de la onda y  $u(x, 0), u_t(x, 0)$  son la posición y la velocidad de la onda iniciales, respectivamente, que se suponen conocidas. El **problema inverso** que se plantea es hallar el coeficiente  $\epsilon_r$ , desconocido en el dominio  $\Omega$ , suponiendo que conocemos el dato adicional de la derivada normal sobre  $\partial\Omega \times (0, T)$ , esto es

$$\partial_x u(L, t) = h(x, t).$$



Para resolver este problema, se minimizará el funcional de regularización de Tikhonov (3.2). En el caso particular de la ecuación de ondas inversa y empleando el Método de Elementos Finitos (FEM), se trabaja bajo la suposición de que la solución de (6.1) pertenece a un espacio de dimensión finita (el espacio  $V_h$  introducido en la Sección 5.2.2) equipado con la norma  $\|\cdot\|_{L_2(\Omega)}$ . Nótese que la minimización del funcional de regularización con dicha norma en un espacio discreto se corresponde con un problema clásico de mínimos cuadrados, con la penalización adicional dada por el término de regularización. Por tanto, en este caso el funcional a minimizar es

$$J_\alpha(u, \epsilon) = \frac{1}{2} \frac{T}{n_t} \sum_{n=1}^{n_t} (\partial_x u^\epsilon - h)^2 + \frac{1}{2} \alpha (\epsilon - \epsilon_0)^2,$$

donde  $n_t$  es el número de puntos tomados para la discretización en tiempo,  $\epsilon_0$  es una estimación *a priori* de la permitividad dieléctrica y  $\alpha$  es el parámetro de regularización.

Por tanto, una vez planteado el problema sólo faltaría diseñar un algoritmo numérico para la resolución del mismo. En este algoritmo, resulta esencial llevar a cabo una elección adecuada del parámetro  $\alpha$ , para la cual se empleará el principio de balance. Dado  $k > 0$ , número de iteraciones del algoritmo, el parámetro  $\alpha$  resulta de la siguiente expresión simplificada del principio de balance (véase [1]):

$$\alpha_k = \frac{\alpha_0}{(k+1)^p}, \quad p \in (0, 1].$$

En nuestro caso, con  $k = 6$  y  $p = 0,5$ , se tiene que  $\alpha = 0,378$ . Vamos a realizar la minimización del funcional usando *Matlab* y la función *fmincon* del *Optimization Toolbox*. Veamos, pues, los detalles de la implementación del algoritmo. Primero, se programa un algoritmo que calcule la derivada normal que estamos considerando en función del coeficiente  $D := \frac{1}{\epsilon_r}$ , resolviendo el problema directo. La manera de resolver la ecuación de ondas en *Matlab* usando la función *pdepe* es llevar a cabo un cambio de variable  $u_1 = u, u_2 = u_t$ , transformando el problema en encontrar  $u = (u_1, u_2)$  tal que

$$\left\{ \begin{array}{l} u_{1t} = u_2, \\ \epsilon_r u_{2t} = u_{1xx} \quad \text{en } \Omega \times (0, T), \\ u_1(-L, t) = u_1(L, t) = 0 \quad \forall t \in (0, T), \\ u_2(-L, t) = u_2(L, t) = 0 \quad \forall t \in (0, T), \\ u_1(x, 0) = \mu_0 \left(1 - \left(\frac{x}{L}\right)^2\right)^2 \quad \text{en } \Omega, \\ u_2(x, 0) = 0 \quad \text{en } \Omega. \end{array} \right.$$

Por tanto, el algoritmo que calcula la derivada normal es:

```

function [uxL,u] = Wave1D.M.NormDer(D)

% Calculo de la derivada normal u_x(L,t)
%
% Solucion de la ecuacion de ondas monodimensional
%  $u_{tt} - D u_{xx} = 0 \quad (x,t) \in (-L,L) \times (0,T)$ 
%  $u(-L,t) = u(L,t) = 0 \quad \text{forall } t \in (0,T)$ 
%  $u(x,0) = \mu_0 (1 - (x/L)^2)^2 \quad \text{forall } x \in (-L,L)$ 
%  $u_t(x,0) = 0 \quad \text{forall } x \in (-L,L)$ 
%
% El cambio de variable u_1 = u, u_2 = u_t
% transforma el problema en:
% | Encontrar u = (u_1, u_2) tal que
% |  $(u_1)_t = u_2$ 
% |  $(u_2)_t = D (u_1)_{xx}$ 
% |  $u_1(-L,t) = u_2(-L,t) = u_1(L,t) = u_2(L,t) = 0 \quad t \in (0,T)$ 
% |  $u_1(x,0) = \mu_0 (1 - (x/L)^2)^2 \quad x \in (-L,L)$ 
% |  $u_2(x,0) = 0 \quad x \in (-L,L)$ 

% Datos del problema
%
T = 12;
data.L = 1;
data.D = D;
data.mu0 = 0.1;

% Datos de mallado, para calculos posteriores
%
np = 21;
nt = 41;
xmesh = linspace(-data.L, data.L, np);
tmesh = linspace(0,T, nt);
delx = xmesh(end)-xmesh(end-1);
undelx = 1./delx;
% ht = T/nt;

```

```

% Solucion
%
pdefun = @(x, t, u, dudx) pdeparam(x,t,u,dudx, data);
icfun = @(x) icparam(x, data);
sol = pdepe(0, pdefun, icfun, @bcfun, xmesh, tmesh);

uxL = undelx*(sol(:,end,1)-sol(:,end-1,1));
u = sol(:, :, 1);
end

%% Funciones auxiliares
%
function [c, f, s] = pdeparam(~, ~, u, dudx, data)
% funcion que proporciona los coeficientes de la EDP
c = [1; 1];
f = [0; data.D*dudx(1)];
s = [u(2); 0];
end

function [u0] = icparam(x, data)
% funcion con los datos iniciales
u0 = [data.mu0*(1-(x/data.L)^2)^2 ; 0];
end

function [pa, qa, pb, qb] = bcfun(~, ua, ~, ub, ~)
% funcion con las condiciones de contorno
pa = ua;
qa = [0;0];
pb = ub;
qb = [0;0];
end

```

El código anterior proporciona un programa para el cálculo de la derivada normal para cada coeficiente  $D$  considerado. El siguiente código es un algoritmo de optimización que emplea la regularización de Tikhonov para la reconstrucción del coeficiente. Además de calcularlo, se representa la solución de la ecuación de ondas para este coeficiente. El algoritmo en cuestión es:

```
function IP
```

```
% Problema inverso del calculo de  $D = \frac{1}{\varepsilon_r}$  en  
% la ecuacion de ondas a partir de  $u_x(L, t)$   
%  $[uxL] = Wave1D\_M\_NormDer(0.25)$ 
```

```
options = optimset('Algorithm','interior-point','Display','iter',...  
'TolFun',1.e-15);
```

```
[Dcal, cost]= fmincon(@COSTfun,0.35,[],[],[],[],0,2,[], options);
```

```
eps = 1/Dcal;
```

```
disp(eps)
```

```
[~, u] = Wave1D\_M\_NormDer(Dcal);
```

```
% Solucion calculada
```

```
figure(1)
```

```
np = 21;
```

```
nt = 41;
```

```
T = 12;
```

```
xmesh = linspace(-1, 1, np);
```

```
tmesh = linspace(0,T, nt);
```

```
surf(xmesh, tmesh, u)
```

```
title('Perfil de la onda en el mallado')
```

```
xlabel('Espacio: x')
```

```
ylabel('Tiempo: t')
```

```
zlabel('u(x, t): Amplitud de la onda')
```

```
view(290, 14)
```

```
figure(2)
```

```
surf(xmesh, tmesh, u) % Misma representacion, pero con otra perspectiva
```

```
view(2)
```

```
colorbar
```

```
xlabel('Espacio: x')
```

```
ylabel('Tiempo: t')
```

```
zlabel('u(x, t)')
```

```
title('Amplitud de la onda en el mallado')
```

```
end
```

```

function val = COSTfun(D)
%
%  $J = J_{\alpha}(D)$ 

% Regularizacion a posteriori
niter = 6;
p = 1/2;
alpha0 = 1;
for n = 1:niter
    alpha = alpha0/((n+1)^p);
end
disp(alpha)

uxL = Wave1D_M_NormDer(0.25);
uxD = Wave1D_M_NormDer(D);
dif = uxL -uxD;
val = 0.5*12/41*sum(dif.*dif)+ alpha/2*(D-0.3)^2;
end

```

Las primera gráfica que proporciona la solución se puede apreciar en la Figura 6.1.

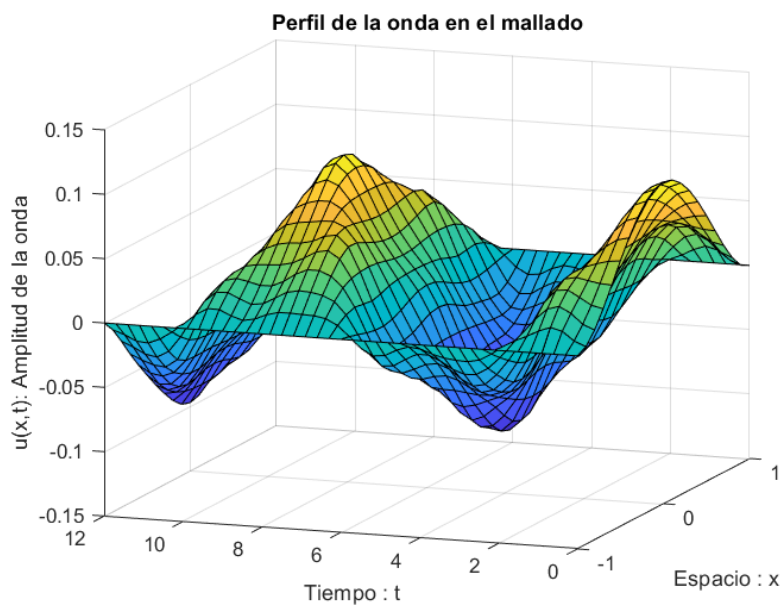


Figura 6.1: Perfil de la onda reconstruida para  $D$  calculado.

La segunda gráfica se puede ver en la Figura 6.2.

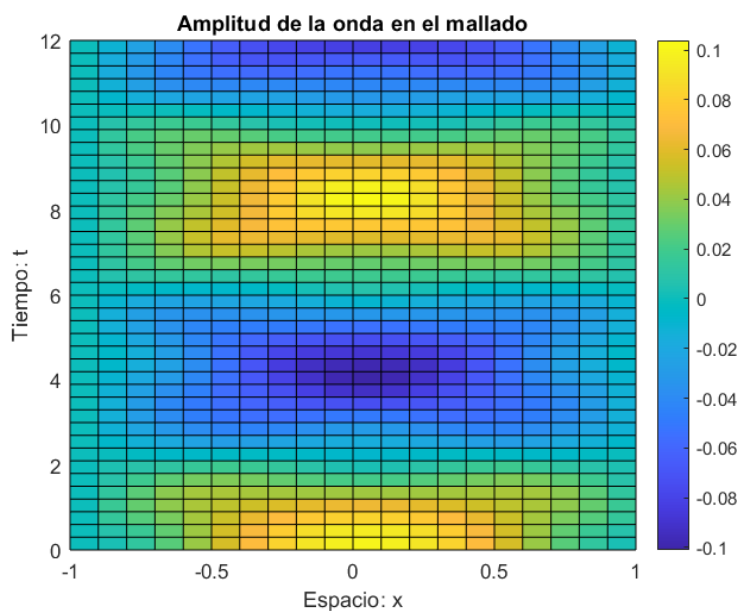


Figura 6.2: Amplitud de la onda reconstruida para  $D$  calculado.

Finalmente, el algoritmo nos proporciona con el comando *disp(alpha)* el parámetro de regularización detallado anteriormente; y con el comando *disp(eps)*, el coeficiente  $\epsilon_r$  reconstruido. Para este problema en particular, se ha supuesto la propagación de la onda en un filamento de celuloide (material empleado para la fabricación de las antiguas películas de cine) cuya permitividad dieléctrica ideal sería  $\epsilon_r^* = 4$ , o bien coeficiente  $D = 0,25$ . La reconstrucción de dicho coeficiente nos da los valores  $D_{calc} = 0,2504$ ,  $\epsilon_r = 3,9939$ , con lo que obtenemos un error absoluto  $\approx 0,15\%$ , margen suficientemente satisfactorio para nuestra experiencia numérica, y que corrobora empíricamente la regularización empleada.

# Bibliografía

- [1] A. B. Bakushinsky, M. Y. Kokurin, *Iterative Methods for Approximate Solution of Inverse Problems*. Math. Appl. (New York) 577, Springer, Dordrecht, 2004.
- [2] A.N. Tikhonov, A.V. Goncharsky, V.V. Stepanov, A.G. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems*. Springer Science+Business Media, 1995.
- [3] A. Doubova, R.Echevarría, *La ecuación de ondas en 2D*. <http://departamento.us.es/edan/php/asig/GRAMAT/GMANED/FreeFEM/Wave2D.html>
- [4] D.H. Wolpert, W.G. MacReady, *No Free Lunch theorems for optimization*. IEEE Transactions on Evolutionary Computation, 1997.
- [5] Functional Gradient Descent. (2022) En D. Bagnell, *16-831: Statistical Techniques in Robotics. Machine Learning and Probabilistic Approaches to Uncertainty and Adaptation*. Carnegie Mellon School of Computer Science.
- [6] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer Science+Business Media, 2010.
- [7] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*. MIT Press, 2016.
- [8] J. Hadamard, *Lectures on Cauchy's problem in linear partial differential equations*. New Haven Yale University Press, 1923.
- [9] J. B. Malmberg, L. Beilina, *An Adaptive Finite Element Method in Quantitative Reconstruction of Small Inclusions from Limited Observations*. Appl. Math. Inf. Sci., 2018.
- [10] J. Torres, *Python Deep Learning*. Grupo Editor Alfaomega S.A., 2020.
- [11] Triangulations, finite elements and finite volumes. (2017) En J. Führmann, *Scientific computing*. Weierstrass Institut für Angewandte Analysis und Stochastik, Berlin.
- [12] K. Ito, B. Jin, *Inverse Problems. Tikhonov Theory and Algorithms*. World Scientific Publishing Company, Ltd., 2014.
- [13] L. Beilina, M. V. Klibanov, *Approximate Global Convergence and Adaptivity for Coefficient Inverse Problems*. Springer Science+Business Media, LLC 2012.

- [14] M. Bergonioux, A. Litman, A. Da Silva, M. Cristofol, L. Beilina, *Mathematical and Numerical Approaches for Multi-Wave Problems*. Springer Proceedings in Mathematics and Statistics, 2020.
- [15] M. Burger, A. C. G. Menucci, S. Osher, M. Rumpf, *Level set and PDE based reconstruction methods in imaging*. Springer International Publishing, 2013.
- [16] M.A. Lavrent'ev, V.G. Romanov, S.P. Shishatskii, *Ill-posed Problems of Mathematical Physics and Analysis*. American Mathematical Society, 1986.
- [17] M. Kubat, *An Introduction to Machine Learning*. Springer Nature Switzerland, 2021.
- [18] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems* Society for Industrial and Applied Mathematics, 1978.
- [19] R. S. Michalski, J. G. Carbonell, T. M. Mitchell, *Machine Learning: an AI approach* Springer Berlin, Heidelberg, 1983.
- [20] Regularization for Inverse Problems. (2018) En S. Siltanen, *Inverse Problems*. Helsinki University.
- [21] S. Banach, *Theory of Linear Operations*. Elsevier Science B.V., 1987.
- [22] V. Isakov, *Inverse Problems for Partial Differential Equations*. Springer International Publishing, 2017