

UNIVERSIDAD AUTÓNOMA METROPOLITANA UNIDAD
IZTAPALAPA

Maestría en Ciencias Matemáticas Aplicadas e industriales CIENCIAS BÁSICAS

E INGENIERÍA

*Métodos deterministas y bayesianos para
resolver el problema de Cauchy para la ecuación
de Laplace*

TESIS PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS (MATEMÁTICAS)
QUE PRESENTA:

JOSÉ ALBERTO VILLEDA TREJO

supervisado por

*Dr. Lorenzo Héctor Juárez Valencia y Dr. Marcos Aurelio Capistrán
Ocampo.*

Índice general

1	Introducción y antecedentes	5
1.1	Motivación	5
1.2	El problema de Cauchy para la ecuación de Laplace	5
1.3	Objetivos	7
2	Método determinista variacional	9
2.1	Formulación como problema de control	9
2.2	Formulación variacional. Regularización.	10
2.3	La derivada del funcional cuadrático	11
2.4	Algoritmo de gradiente conjugado	12
2.5	Discretización de los subproblemas elípticos	13
2.6	Estrategia de discretizar y después optimizar	14
3	Inversión Estadística	17
3.1	Problemas inversos y fórmula de Bayes	18
3.2	Estimadores puntuales	20
3.3	Construcción del <i>modelo a priori</i>	21
3.4	Construcción de la función de verosimilitud	24
3.5	Formulación del modelo Bayesiano completo	25
3.5.1	La distribución de la posterior es Gaussiana	26
3.5.2	Necesidad de explorar la densidad posterior	27
3.6	Relación entre regularización de Tijonov y la distribución a priori	27
3.6.1	Regularizadores suavizantes	28
3.6.2	Términos de regularización no suavizantes	30
3.7	Métodos de Monte Carlo para Cadenas de Markov	31
3.7.1	Integración de Monte Carlo	33
3.7.2	Algoritmo de Metropolis-Hastings	34
3.7.3	Algoritmo t-walk	37
3.8	Aproximación de rango bajo	38
3.8.1	Aproximación óptima de la matriz de covarianza de la distribución posterior	39
3.8.2	Clases de aproximación	40

3.8.3	Función de costos	41
3.8.4	Resultados de optimalidad	42
4	Resultados numéricos	45
4.1	Ejemplo 1. Región anular circular simple	46
4.2	Ejemplo 2. Región anular con frontera exterior circular y frontera interior elíptica.	49
4.3	Ejemplo 3. Región anular compleja	52
4.4	Aproximación de rango bajo	57
5	Conclusiones	61

Capítulo 1

Introducción y antecedentes

1.1 Motivación

En el campo de la neurobiología el estudio del sistema nervioso del cerebro es importante y altamente complejo, ya que el cerebro está formado por millones neuronas que transmiten información proveniente de estímulos del entorno o señales del mismo cuerpo para después procesar y generar respuestas acordes con necesidades específicas.

Por otro lado, también existen diversos problemas relacionados con el sistema nervioso denominados patologías como son la epilepsia, las alteraciones de estado de conciencia y las infecciones, entre otras. Para entender como funciona el sistema nervioso ayuda mucho aplicar técnicas de exploración no invasivas, como el registro de actividad eléctrica en el cerebro que se realiza colocando electrodos de superficie adheridos al cuero cabelludo por un gel conductor con los cuales se puede medir las fluctuaciones eléctricas. Esta técnica fue utilizada por primera vez en 1929 por Hans Berger [24] y ha dado origen a un modelo electrostático de medio continuo que describe este fenómeno denominado el *problema inverso electroencefalográfico*.

Dicho modelo pretende ampliar el conocimiento en el estudio del sistema nervioso para proponer diagnósticos a partir del registro de actividad eléctrica y ha sido estudiado recientemente en [9], [15] y [3], así como en algunas de las referencias citadas en esos trabajos.

1.2 El problema de Cauchy para la ecuación de Laplace

Un subproblema del problema inverso electroencefalográfico es el problema de Cauchy, el cual puede consultarse por ejemplo en [9], [15] y más recientemente en [10], El problema de Cauchy para la ecuación de Laplace consiste en determinar una función armónica definida en una región anular a partir de mediciones en su frontera o parte de la misma (datos de Cauchy). Por ejemplo, dada Ω una región anular acotada en \mathbb{R}^2 con una frontera interior

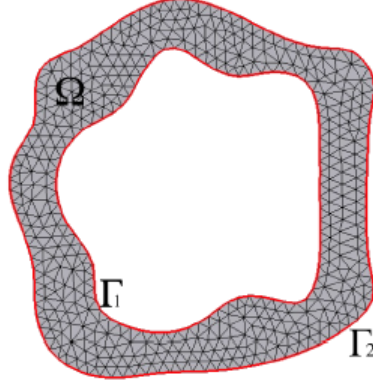


Figura 1.1: Representación de la región anular Ω

acotada Γ_1 suficientemente suave y frontera exterior Γ_2 acotada como se muestra en la figura 1.1, el problema de Cauchy consiste en encontrar una función ω definida en Ω tal que

$$\begin{aligned} -\sigma\Delta\omega &= 0 \text{ en } \Omega, \\ \omega &= \bar{V} \text{ sobre } \Gamma_2, \\ \sigma\frac{\partial\omega}{\partial\mathbf{n}} &= \bar{h} \text{ sobre } \Gamma_2, \end{aligned} \tag{1.1}$$

en donde \bar{V} , \bar{h} son funciones conocidas, denominadas *datos de Cauchy*, que en problemas de electrostática representan la medida del potencial y la corriente eléctrica en la frontera exterior Γ_2 , respectivamente, mientras que la constante σ es la conductividad eléctrica en el medio anular Ω . En esta ecuación \mathbf{n} denota al vector normal unitario exterior sobre la frontera de Ω y $\partial\omega/\partial\mathbf{n}$ es la derivada normal exterior de la función ω sobre Γ_2 . Obsérvese que en este problema se dan dos condiciones de frontera sobre la frontera exterior Γ_2 y ninguna sobre la frontera interior Γ_1 y es bien conocido que no tiene solución única, es decir es un problema mal planteado en el sentido de Hadamard. La formulación del problema en el caso tridimensional es exactamente la misma. Hemos formulado el problema en dos dimensiones a propósito y debido a que es el caso que estudiaremos en este trabajo.

El problema de Cauchy no solo es importante en el problema electroencefalográfico, es importante en muchas otras aplicaciones, como en estimar el deterioro de una tubería, calcular una solución o *potencial* en algunas regiones o en límites donde no hay acceso directo, y encontrar grietas en las placas, ([4], [8], [12], [18]). También es importante para aplicaciones en algunas áreas tales como geofísica, tomografía de procesos, electroencefalografía inversa y electrocardiografía inversa ([7], [13], [23]).

Sin pérdida de generalidad se puede considerar $\bar{h} = 0$ en (1.1), dado que ese problema puede reducirse haciendo $\omega = \omega_1 + v$, en donde ω_1 es una función armónica que satisface $\sigma\partial\omega_1/\partial\mathbf{n} = \bar{h}$ sobre Γ_2 y $\omega_1 = 0$ sobre Γ_1 , y v es solución del siguiente problema

$$\begin{aligned}
-\sigma \Delta v &= 0 \text{ en } \Omega, \\
v &= V \text{ sobre } \Gamma_2 \\
\sigma \frac{\partial v}{\partial \mathbf{n}} &= 0 \text{ sobre } \Gamma_2,
\end{aligned} \tag{1.2}$$

donde $V = \bar{V} - \omega_1$ sobre Γ_2 . Por lo tanto, de ahora en adelante le llamaremos *problema de Cauchy* a este último problema. De la misma forma que en 1.1, este problema es mal planteado en el sentido de Hadamard no sólo porque tiene múltiples soluciones, sino también porque pequeñas variaciones de V (los datos de Cauchy) pueden producir grandes variaciones de la solución. Es decir, el problema presenta inestabilidad y son necesarias técnicas de regularización cuando se utilizan métodos de aproximación numérica para resolverlo. Además, para garantizar la existencia de una solución al problema de Cauchy, es necesario imponer algunas condiciones de suavidad a los datos, como se verá más adelante.

A la fecha, existen diferentes enfoques y algoritmos para resolver el problema de Cauchy. Por ejemplo, en [19] se empleó una descomposición de valores singulares para encontrar soluciones en una región anular circular, y se usó truncamiento espectral del método pseudo-inverso como técnica de regularización. Sin embargo, este método está limitado a geometrías simples.

En [27] se utilizaron potenciales de capa simple y doble, con densidades definidas en la frontera de la región anular, para obtener un sistema equivalente de ecuaciones integrales, obteniendo un método que puede aplicarse a regiones anulares más generales. Cuando se discretizan estas ecuaciones integrales la matriz resultante de colocación es mal condicionada y se requieren métodos de regularización como se muestra en [4], [7], [27], [5], [20] y [28].

1.3 Objetivos

En este trabajo se pretende complementar [10], [15] y [9]. Por un lado, la complementaridad se establece por medio la conexión entre los métodos de inversión determinista utilizados en los trabajos mencionados, y por el otro, en la introducción de los métodos y procedimientos de la inversión estadística utilizados en este texto. Esta tesis muestra que ambos métodos más que competir, se complementan. Cada enfoque contiene aspectos importantes y novedosos, y a veces poderosos, que permiten enriquecer los métodos de solución y análisis, no solo del problema de Cauchy, sino también de otros problemas inversos o de control muy frecuentes en las aplicaciones.

Se sabe que las soluciones encontradas numéricamente con un enfoque determinista tienen ciertos inconvenientes, como la falta de un procedimiento claro para medir la calidad de la solución encontrada, además de no contemplar en el método los errores de medición

en forma natural. La información obtenida en estos casos es limitada, debido a que generalmente sólo proporciona una solución que es óptima en cierto sentido. Por el contrario, los métodos de inversión estadística al encontrar una distribución de probabilidad como solución del problema, permite extraer mayor información, sobre todo asociada a la incertidumbre de las soluciones, a los momentos y estimadores de la distribución como la media y la varianza. Sin embargo, se debe invertir mucho mayor costo computacional para obtener dicha información.

Más específicamente en la tesis se revisa el método determinista en el capítulo 2, profundizando en el estudio de la discretización y su conexión con la optimización del modelo variacional, lo cual permite establecer una relación natural con el enfoque estadístico, estudiado ampliamente en el capítulo 3. En particular, se muestran diferentes estrategias de regularización y su relación con la distribución a priori en el modelo estadístico.

La inversión estadística se introduce y estudia en el capítulo 3. Se propone un modelo estadístico con distribuciones Gaussianas. En particular, el modelo a priori se construye utilizando campos aleatorios de Markov Gaussianos, y se muestra las diferentes posibilidades en el problema de Cauchy. Se muestra la relación entre la regularización de Tijonov y la distribución a priori, profundizando en los diferentes tipos de regularizadores suavizantes o no suavizantes. Además de explorar la distribución posterior con el método T-walk, un algoritmo del tipo Metropolis-Hastings y que es parte de la familia de métodos de Markov Chain Monte Carlo (MCMC).

Para los cálculos numéricos se utilizará el software FEniCS, el cual está especializado en resolución de ecuaciones diferenciales parciales (EDPs) por medio del método de los elementos finitos. Para el trabajo de programación se utilizará el lenguaje de programación Python y el algoritmo T-walk como explorador de la distribución posterior.

Capítulo 2

Método determinista variacional

2.1 Formulación como problema de control

Más recientemente, en [10], se presenta un enfoque variacional basada en la reformulación del problema de Cauchy como un problema de control, en donde el control es la función potencial definida sobre la frontera interior de la región anular, y cuyo funcional de costo incorpora un término de penalización con los datos de entrada V . En ese trabajo también se muestra que cuando los datos de entrada tienen ruido, o errores de medición, el parámetro de penalización se puede escoger como el inverso de la magnitud del ruido. Los resultados numéricos que se presentan en regiones simples y complejas ejemplifican que esta metodología produce soluciones estables y precisas. Por estas razones, este último artículo es una de las principales referencias para el presente trabajo.

Por lo tanto, partimos de la formulación del problema de Cauchy como un problema de control, como se indica en [10]. Primero se introduce el problema auxiliar siguiente, cuya ecuación (2.1) se denomina *ecuación de estado*:

Dada una función φ definida sobre Γ_1 , encontrar la función u definida en Ω tal que

$$\begin{aligned} -\sigma\Delta u &= 0 \text{ en } \Omega, \\ u &= \varphi \text{ sobre } \Gamma_1, \\ \sigma\frac{\partial u}{\partial \mathbf{n}} &= 0 \text{ sobre } \Gamma_2. \end{aligned} \tag{2.1}$$

Este problema es bien planteado en el sentido de Hadamard si se escogen los espacios apropiados para las funciones φ y u , y si el dominio Ω es suficientemente regular. Es posible reformular (1.2) como un problema inverso de manera informal se puede decir que consiste en encontrar una función φ definida sobre Γ_1 tal que u , solución de (2.1), es tal que $u|_{\Gamma_2} = V$, suponiendo V conocida. Para precisar estos conceptos, a continuación incluimos algunas definiciones y resultados presentados en [10].

Definición 1 (Formulación débil) Una función $u \in H^1(\Omega)$ con $u = \varphi$ sobre Γ_1 es una solución débil de la ecuación de estado (2.1) si

$$\int_{\Omega} \nabla u \cdot \nabla v \, d\Omega = 0 \text{ para toda } v \in U_0, \quad (2.2)$$

en donde $U_0 = \{v \in H^1(\Omega) : v = 0 \text{ sobre } \Gamma_1\}$ es el espacio de funciones de prueba.

Suponiendo que el dominio Ω es suficientemente suave se demuestra el siguiente teorema.

Teorema 1 Dado $\varphi \in H^{-1/2}(\Gamma_1)$ con $m \geq 1$, fija que no depende de φ , la solución débil u de (2.1) existe y es única. Además $\|u\|_{H^m(\Omega)} \leq C \|\varphi\|_{H^{m-1/2}(\Gamma_1)}$ con C constante que no depende de φ .

Este resultado permite definir el siguiente operador

$$K : H^{m-1/2}(\Gamma_1) \rightarrow L^2(\Gamma_2), \quad K(\varphi) = u|_{\Gamma_2} \text{ con } u \text{ solución de (2.1),}$$

el cual relaciona de manera única voltajes en la frontera interior (causa) con el voltaje resultante en la frontera exterior (efecto). Este operador es lineal, inyectivo y compacto: es lineal debido a que el problema (2.1) es lineal, inyectivo debido al teorema 1, y compacto por el operador de traza al restringir la solución u a la frontera exterior Γ_2 . Entonces, la relación entre el problema de Cauchy (1.2) y la ecuación de estado (2.1) se puede describir mediante el operador K en la forma siguiente:

Una solución del problema (2.1) es una solución del problema (1.2) si se considera φ en Γ_1 tal que

$$K\varphi := u|_{\Gamma_2} = V, \quad (2.3)$$

donde u denota la solución de la ecuación de estado (2.1), V la medición conocida sobre Γ_2 en el problema de Cauchy (1.2). Entonces podemos escribir $\varphi = K^{-1}(V)$.

2.2 Formulación variacional. Regularización.

La ecuación (2.3) no tiene solución para toda función $V \in L^2(\Gamma_2)$, sin embargo en [10] se garantiza que $Im(K)$ es denso en $L^2(\Gamma_2)$, lo cual permite reformular el problema mediante la minimización del funcional cuadrático convexo:

$$J_k(\varphi) = \frac{1}{2} \|\varphi\|_{L^2(\Gamma_1)}^2 + \frac{k}{2} \|K\varphi - V\|_{L^2(\Gamma_2)}^2. \quad (2.4)$$

La importancia de este procedimiento es que garantiza la existencia de una solución única para toda $V \in L^2(\Gamma_2)$, la cual puede encontrarse en forma estable debido a la penalización introducida en el segundo término mediante la constante k , la cual usualmente se considera mucho mayor a 1. El mínimo del funcional de costo (2.4) es igual al del siguiente funcional:

$$J_\alpha(\varphi) = \frac{\alpha}{2} \|\varphi\|_{L^2(\Gamma_1)}^2 + \frac{1}{2} \|K\varphi - V\|_{L^2(\Gamma_2)}^2 \quad \text{con } \alpha = \frac{1}{k}, \quad (2.5)$$

en donde $0 < \alpha < 1$, ya que $\alpha = \frac{1}{k}$, y k es considerado mayor que 1, es el parámetro de regularización. El primer término corresponde a la regularización usual de Tijonov.

En este trabajo se utiliza un funcional cuadrático adicional, el cual no es considerado en [10]:

$$J_{\alpha s}(\varphi) = \frac{\alpha}{2} \left\| \frac{\partial \varphi}{\partial s} \right\|_{L^2(\Gamma_1)}^2 + \frac{1}{2} \|K\varphi - V\|_{L^2(\Gamma_2)}^2. \quad (2.6)$$

Este funcional cuadrático difiere del anterior en el término de regularización, ya que se ha incluido $\partial\varphi/\partial s$ en lugar de φ . Más adelante se muestra que este tipo de regularización produce un modelo más cercano al construido en el capítulo 3. Por lo pronto, podemos adelantar que este tipo de regularización producirá resultados más suaves, ya que ahora se requiere que tanto φ como $\partial\varphi/\partial s$ sean cuadrado integrables sobre Γ_1 .

2.3 La derivada del funcional cuadrático

El mínimo del funcional cuadrático convexo J_k (respectivamente J_α y $J_{\alpha s}$) se puede obtener mediante un metodo iterativo de descenso, en donde se utiliza la primera derivada (o variación) de J_k .

En [10] se utiliza el método de la ecuación adjunta para calcular dicha derivada.

Se define el operador adjunto $K^* : L^2(\Gamma_2) \rightarrow L^2(\Gamma_1)$ como

$$K^*v = \sigma \frac{\partial p}{\partial \mathbf{n}} \Big|_{\Gamma_1},$$

donde p es la solución de

$$\begin{aligned} -\sigma \Delta p &= 0 \text{ en } \Omega, \\ p &= 0 \text{ sobre } \Gamma_1, \\ \sigma \frac{\partial p}{\partial \mathbf{n}} &= v \text{ sobre } \Gamma_2. \end{aligned} \quad (2.7)$$

La derivada del funcional de costo J_k es formalmente (ver [10]):

$$DJ_k(\varphi) = \varphi + k K^*(K\varphi - V).$$

Más precisamente, dado $\varphi \in L^2(\Gamma_1)$ primero se resuelve la ecuación de estado (2.1) obteniendo la solución $u = u(\varphi)$, a continuación se le calcula la traza de u sobre Γ_1 y, finalmente se resuelve la ecuación adjunta (2.7) con $v = K\varphi - V = u|_{\Gamma_2} - V$. Al final, se obtiene la siguiente expresión para la derivada de J_k

$$DJ_k(\varphi) = \varphi + k \left(\sigma \frac{\partial p}{\partial \mathbf{n}} \Big|_{\Gamma_1} \right). \quad (2.8)$$

De manera similar se calculan las derivadas de los otros dos funcionales:

$$DJ_\alpha(\varphi) = \alpha\varphi + K^*(K\varphi - V) = \alpha\varphi + \sigma \frac{\partial p}{\partial \mathbf{n}} \Big|_{\Gamma_1}, \quad (2.9)$$

$$DJ_{\alpha s}(\varphi) = -\alpha \frac{\partial^2 \varphi}{\partial s^2} + K^*(K\varphi - V) = -\alpha \frac{\partial^2 \varphi}{\partial s^2} + \sigma \frac{\partial p}{\partial \mathbf{n}} \Big|_{\Gamma_1}. \quad (2.10)$$

En (2.10), φ se supone suficientemente suave.

2.4 Algoritmo de gradiente conjugado

Para minimizar el funcional J_k . A continuación se muestra el algoritmo de gradiente conjugado para el funcional cuadrático convexo (2.4), en donde en cada iteración se utiliza el cálculo de la derivada (2.8). Se presenta en forma detallada, ya que se usará en este trabajo con el fin de comparar los resultados obtenidos con los métodos de inversión estadística.

Inicialización. Escoger una función inicial $\varphi^0 \in L^2(\Gamma_1)$.

1. Resolver para u^0 el problema

$$-\sigma \Delta u^0 = 0 \text{ en } \Omega, \quad u^0 = \varphi^0 \text{ sobre } \Gamma_1, \quad \sigma \frac{\partial u^0}{\partial \mathbf{n}} = 0 \text{ sobre } \Gamma_2,$$

y para p^0 el siguiente problema

$$-\sigma \Delta p^0 = 0 \text{ en } \Omega, \quad p^0 = 0 \text{ sobre } \Gamma_1, \quad \sigma \frac{\partial p^0}{\partial \mathbf{n}} = u^0 - V \text{ sobre } \Gamma_2.$$

2. Evaluar la derivada: $g^0 = \varphi^0 + k \left(\sigma \frac{\partial p^0}{\partial \mathbf{n}} \right)$ sobre Γ_1 .

3. Hacer $d^0 = -g^0$ sobre Γ_1 .

Descenso mediante iteraciones. Para $n \geq 0$, considerar φ^n , d^n y g^n conocidos y φ^{n+1} , d^{n+1} y g^{n+1} respectivamente de la siguiente manera:

4. Resolver para \bar{u}^n el problema

$$-\sigma \Delta \bar{u}^n = 0 \text{ en } \Omega, \quad \bar{u}^n = d^n \text{ sobre } \Gamma_1, \quad \sigma \frac{\partial \bar{u}^n}{\partial \mathbf{n}} = 0 \text{ sobre } \Gamma_2,$$

y para \bar{p}^n el siguiente problema

$$-\sigma \Delta \bar{p}^n = 0 \text{ en } \Omega, \quad \bar{p}^n = 0 \text{ sobre } \Gamma_1, \quad \sigma \frac{\partial \bar{p}^n}{\partial \mathbf{n}} = \bar{u}^n \text{ sobre } \Gamma_2.$$

5. Evaluar $\bar{g}^n = d^n + k \left(\sigma \frac{\partial \bar{p}^n}{\partial \mathbf{n}} \right)$ sobre Γ_1 .
6. Calcular el paso de descenso $\rho_n = \frac{\langle g^n, d^n \rangle_{L^2(\Gamma_1)}}{\langle \bar{g}^n, d^n \rangle_{L^2(\Gamma_1)}}$.
7. Actualizar $\varphi^{n+1} = \varphi^n + \rho_n d^n$.
8. Actualizar $g^{n+1} = g^n + \rho_n \bar{g}^n$.

Prueba de convergencia. Si $\langle g^{n+1}, g^{n+1} \rangle_{L^2(\Gamma_1)} / \max \{1, \langle \varphi^0, \varphi^0 \rangle_{L^2(\Gamma_1)}\} \leq \varepsilon$, tomar como solución $\varphi = \varphi^{n+1}$, parar las iteraciones y salir. En caso contrario hacer lo siguiente:

9. Calcular la constante $\beta_n = \frac{\langle g^{n+1}, g^{n+1} \rangle_{L^2(\Gamma_1)}}{\langle g^n, g^n \rangle_{L^2(\Gamma_1)}}$.
10. Calcular la **nueva dirección de descenso** $d^{n+1} = -g^{n+1} + \beta_n d_n$.
11. Hacer $n + 1 = n$ y regresar al paso 4.

En el algoritmo anterior, los pasos en donde se realizan los cálculos para el cálculo de la derivada son claramente los pasos 1 y 2, en la etapa de inicialización, y en los pasos 4 y 5, en la etapa de iteraciones para el descenso. El algoritmo es básicamente el mismo cuando se quiere minimizar los funcionales cuadráticos J_α y $J_{\alpha s}$, pues basta modificar el paso 2 y 5 en el algoritmo de gradiente conjugado de la siguiente manera:

- Para el funcional J_α definido en (2.5) se evalúa:

$$\text{En el paso 2: } g^0 = \alpha \varphi^0 + \sigma \frac{\partial p^0}{\partial \mathbf{n}}.$$

$$\text{En el paso 5: } \bar{g}^n = \alpha d^n + \sigma \frac{\partial \bar{p}^n}{\partial \mathbf{n}}.$$

- Para el funcional $J_{\alpha s}$ definido en (2.6) se evalúa:

$$\text{En el paso 2: } g^0 = -\alpha \frac{\partial^2 \varphi^0}{\partial s^2} + \sigma \frac{\partial p^0}{\partial \mathbf{n}}.$$

$$\text{En el paso 5: } \bar{g}^n = -\alpha \frac{\partial^2 \varphi^n}{\partial s^2} + \sigma \frac{\partial \bar{p}^n}{\partial \mathbf{n}}.$$

2.5 Discretización de los subproblemas elípticos

La realización computacional del algoritmo anterior requiere de la solución de dos problemas elípticos en la inicialización y otros dos en cada iteración. En este trabajo, se emplea el método de los elementos finitos para la discretización de esos problemas, no sólo porque

es adecuado para la formulación variacional del problema de Cauchy, ya que además necesitamos resolver numéricamente la ecuación de estado (2.1) cuando se utilicen los métodos de inversión estadística. Por tal motivo se considerará el siguiente modelo elíptico clásico:

$$\begin{aligned} -\sigma \Delta u &= f & \text{en } \Omega, \\ u &= g_0 & \text{sobre } \Gamma_1, \\ \sigma \frac{\partial u}{\partial \mathbf{n}} &= g_1 & \text{sobre } \Gamma_2. \end{aligned} \quad (2.11)$$

cuya formulación variacional es la siguiente:

Encontrar la función escalar u definida en Ω con $u - g_0 \in U_0$, tal que

$$\sigma \int_{\Omega} \nabla u \cdot \nabla v \, d\Omega = \int_{\Omega} f v \, d\Omega + \int_{\Gamma_2} g_1 v \, d\Gamma, \quad \forall v \in U_0, \quad (2.12)$$

donde U_0 se define como en (2.2).

Estos problemas elípticos se resuelven utilizando el método de los elementos finitos con elementos triangulares, como se muestra la Figura (1.1). Elementos lineales de Lagrange bastan para obtener una buena aproximación, de tal manera que el orden de convergencia es proporcional a h^2 , en donde h es el tamaño característico de los elementos triangulares. Más precisamente, la solución discreta u_h es un polinomio lineal por tramos (en cada elemento) y continuo que satisface

$$\|u - u_h\|_{L^2(\Omega)} \leq c h^2,$$

donde c es una constante que no depende de la solución exacta u ni del parámetro de discretización h .

2.6 Estrategia de discretizar y después optimizar

El enfoque determinista presentado en este capítulo, se basa en encontrar el modelo de optimización en los espacios de funciones de Hilbert adecuados, con el fin de poder definir productos internos y normas que nos permitan utilizar los métodos variacionales. Se formula el modelo cuadrático de optimización, como (2.4), y se propone un algoritmo iterativo como el método de gradiente conjugado en la sección 2.4. Al final se discretizan los subproblemas elípticos (ecuaciones diferenciales) que aparecen en el algoritmo de optimización.

Una alternativa, dentro del enfoque determinista, es proceder en el sentido opuesto. Es decir, primero discretizar el problema operacional que se quiere resolver, obteniendo un problema en dimensión finita. Posteriormente, se formula el problema de optimización y se resuelve con métodos tradicionales para funciones multidimensionales.

Con el objeto de fijar la notación y ciertas ideas, se definen los siguientes vectores y las relaciones entre ellos.

- $\mathbf{x} \in \mathbb{R}^n$, el vector con coordenadas $x_i = \varphi_h(P_i)$, $i = 1, \dots, n$, en donde φ_h es la discretización del potencial φ sobre la frontera interior Γ_1 y $\{P_i\}_{i=1}^n$ es el conjunto de nodos de la malla de elemento finito sobre Γ_1 .
- $\mathbf{y} \in \mathbb{R}^m$, el vector con coordenadas $y_j = V_h(Q_j)$, $j = 1, \dots, m$, en donde V_h representa la discretización del potencial V sobre la frontera exterior Γ_2 y $\{Q_j\}_{j=1}^m$ los nodos de la malla de elemento finito sobre Γ_2 .
- El análogo discreto a la relación operacional (2.3) es:

$$K_h(\varphi_h) = u_h|_{\Gamma_2} = V_h. \quad (2.13)$$

Debido a que la discretización se obtiene con elementos finitos lineales, la relación queda determinada de manera única por los valores nodales de φ_h y V_h sobre las fronteras interior y exterior, respectivamente. Por lo tanto, utilizando los valores nodales \mathbf{x} y \mathbf{y} anteriores, se define la función nodal $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ asociada a (2.13), dada por

$$f(\mathbf{x}) = \mathbf{y}. \quad (2.14)$$

En tal problema la función es lineal ya que representa la versión nodal de la relación lineal (2.13). Por lo tanto, existe una matriz (obtenida por discretización) $A \in \mathbb{R}^{m \times n}$ tal que $f(\mathbf{x}) = A\mathbf{x}$ y la relación (2.14) también puede escribirse como

$$A\mathbf{x} = \mathbf{y}. \quad (2.15)$$

Esta última ecuación, o bien (2.14), es el modelo discreto en espacios de dimensión finita de (2.3). Por supuesto, si se suponen conocidos la matriz A y el vector con los datos \mathbf{y} , el sistema algebraico lineal no tiene solución cuando $\mathbf{y} \notin \mathcal{R}(A)$ (generalmente cuando $m > n$). En caso de que tenga una solución, ésta no será única cuando $\text{Ker}(A) \neq \{\mathbf{0}\}$ (generalmente cuando $m < n$). Además, la matriz A frecuentemente es mal condicionada y las posibles perturbaciones en los datos o las introducidas por los métodos de aproximación numérica, se amplificarán, obteniendo resultados muy poco confiables. Sin embargo, cuando el problema es mal planteado (en el sentido de Hadamard) se puede estabilizar utilizando regularización. Por ejemplo, el modelo de optimización con regularización de Tijonov análogo al funcional J_α en (2.5) es la función cuadrática

$$F_\alpha(\mathbf{x}) = \frac{\alpha}{2} \|\mathbf{x}\|_{\mathbb{R}^n}^2 + \frac{1}{2} \|A\mathbf{x} - \mathbf{y}\|_{\mathbb{R}^m}^2, \quad (2.16)$$

y el análogo discreto de la derivada (2.9) es:

$$DF_\alpha(\mathbf{x}) = \alpha\mathbf{x} + A^T(A\mathbf{x} - \mathbf{y}). \quad (2.17)$$

Los análogos discretos del funcional J_k y su derivada son similares. Finalmente, los análogos discretos del funcional J_{α_s} y su derivada DJ_{α_s} dadas en (2.6) y (2.10), respectivamente, son

$$F_{\alpha_s}(\mathbf{x}) = \frac{\alpha}{2} \|D\mathbf{x}\|_{\mathbb{R}^n}^2 + \frac{1}{2} \|A\mathbf{x} - \mathbf{y}\|_{\mathbb{R}^m}^2 \quad (2.18)$$

y

$$DF_{\alpha s}(\mathbf{x}) = \alpha D^T D \mathbf{x} + A^T (A \mathbf{x} - \mathbf{y}). \quad (2.19)$$

respectivamente. En estas últimas dos expresiones, D es una matriz de discretización de $n \times n$ de la derivada $\partial\varphi/\partial s$ sobre Γ_1 , de tal manera que $L = D^T D$ proporciona una matriz de discretización de $-\partial^2\varphi/\partial s^2$.

En esta tesis no se seguirá el procedimiento de discretizar y luego optimizar con el enfoque determinista. Sin embargo, en el siguiente capítulo, se observa que esta estrategia se puede relacionar con el enfoque de inversión estadística basado en métodos Bayesianos.

Capítulo 3

Inversión Estadística

Este capítulo se introducen y discuten distintos métodos de aproximación, utilizando estadística bayesiana, para el estudio y solución de problemas inversos. Además, se analizan algunos métodos numéricos y su implementación computacional, los cuales se comparan con los resultados obtenidos mediante el enfoque determinista. Los conceptos fundamentales presentados se basan principalmente en el enfoque seguido en [16], el donde se estudian los temas de inversión estadística de una forma clara y concisa. La filosofía detrás de los métodos Bayesianos es reformular el problema inverso en forma de búsqueda estadística de información, tomando en cuenta que hay cantidades directamente observables y otras que no se pueden observar.

En problemas inversos, algunas de las cantidades no observables son de interés primordial. Las diferentes cantidades dependen unas de otras a través de modelos. Uno de los objetivos de los métodos estadísticos es evaluar la incertidumbre sobre las variables considerando el conocimiento que se tiene del problema tanto antes como después de procesar mediciones. La inversión estadística está basada fundamentalmente en los siguientes principios:

- Todas las variables en el modelo son variables aleatorias.
- La aleatoriedad describe nuestro grado de información con respecto a sus realizaciones.
- El grado de información sobre estos valores se codifica en las distribuciones de probabilidad.
- La solución de un problema inverso es la distribución de probabilidad posterior.

El último punto es el que marca mayor diferencia entre la aproximación tradicional y el enfoque Bayesiano, ya que en los métodos tradicionales se utilizan métodos analíticos y numéricos para obtener estimaciones únicas o puntuales de las cantidades desconocidas, mientras que en el método estadístico se obtienen distribuciones que proporcionan no sólo estimaciones puntuales sino también la respectiva incertidumbre asociada a los datos observados y a la información a priori disponible. Por lo tanto, el método estadístico aporta

más información y esto permite mejorar la toma de decisiones; claro, estos métodos tienen ciertos inconvenientes que se discutirán a medida que avance el desarrollo del capítulo.

3.1 Problemas inversos y fórmula de Bayes

En los problemas inversos clásicos se considera que la información que se puede medir, y que denotamos por $\mathbf{y} \in \mathbb{R}^m$, para obtener información sobre una cantidad desconocida inaccesible o difícil de medir y denotada por $\mathbf{x} \in \mathbb{R}^n$, están relacionadas por medio de un modelo $\mathbf{y} = f(\mathbf{x})$, como en (2.14) para el problema que nos interesa. Los modelos suelen ser inexactos y además las mediciones generalmente contienen ruido asociado a errores de medición o errores de discretización y redondeo por el procesamiento numérico. Por lo tanto, en el estudio de los problemas inversos generalmente se incluye una variable adicional para modelar dicho ruido, por lo que el modelo tradicional, en problemas inversos se expresa de la siguiente manera:

$$\mathbf{y} = f(\mathbf{x}, \mathbf{e}), \quad (3.1)$$

donde $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ denota al modelo que relaciona las cantidades medidas \mathbf{y} y los datos desconocidos \mathbf{x} , considerando un error \mathbf{e} .

Por ejemplo, en el problema de Cauchy para la ecuación de Laplace el modelo es el problema (2.1), las mediciones son los voltajes V en la frontera exterior Γ_2 y la cantidad desconocida es el potencial eléctrico φ en la frontera interior Γ_1 . El problema discretizado (2.13) da origen a la relación entre los valores nodales de esas funciones, asociados a una malla, que se denota por la relación (2.14). En la relación (2.14) no se consideraron los posibles errores de medición o discretización del problema o incertidumbres de otro tipo, por lo que (3.1) es una generalización de la misma, en donde se han incorporado dichas incertidumbres. Para resolver estos problemas suele utilizarse regularización para evitar la inestabilidad asociada a la amplificación del ruido causado por los errores, como ya se ha explicado anteriormente el introducir los funcionales cuadráticos en la sección 2.2.

Por otro lado, en problemas de inversión estadística todos los parámetros son considerados como variables aleatorias, que se denotan con letras mayúsculas, y su relación se puede escribir como

$$Y = f(X, E). \quad (3.2)$$

Esta relación se establece ahora entre variables aleatorias $X = (X_1, \dots, X_n)$, $Y = (Y_1, \dots, Y_m)$, $E = (E_1, \dots, E_m)$ y consecuentemente las distribuciones de probabilidad dependen entre sí. Sin embargo, en inversión estadística ni siquiera se requiere considerar un modelo de la forma (3.2), ya que el enfoque se basa en la relación entre las distintas distribuciones de probabilidad. Para fijar ideas se introduce la siguiente nomenclatura: la variable aleatoria directamente observable es Y y una realización u observación es $Y = \mathbf{y}_{obs}$, es decir, los datos de medición del proceso; la variable aleatoria no observable o desconocida es X , incluye todas las cantidades que se quieren obtener en el problema inverso; por último, las variables que no son observables ni de interés principal se denominan parámetros o ruido,

según el contexto y se incluyen en la variable aleatoria E .

Una característica importante de la inversión estadística es que, antes de realizar la medición de Y , se supone que se tiene alguna información previa de la variable aleatoria X y que esa información está codificada dentro de la densidad de probabilidad $\mathbf{x} \rightarrow \pi_{pr}(\mathbf{x})$, conocida como *densidad a priori*. Esta densidad incorpora todo el conocimiento que se tenga de la variable aleatoria X antes de la medición. Cabe mencionar que es posible realizar el método prácticamente sin información a priori, pero ello incrementará la incertidumbre, es decir a menor información mayor la incertidumbre.

Considerando las configuraciones y analizando la información previa de la variable aleatoria X se supone que la densidad de probabilidad conjunta denotada por $\pi(\mathbf{x}, \mathbf{y})$ debe satisfacer

$$\int_{\mathbb{R}^m} \pi(\mathbf{x}, \mathbf{y}) d\mathbf{y} = \pi_{pr}(\mathbf{x}), \quad (3.3)$$

Es decir, la densidad marginal de X es la densidad a priori. Por otro lado, si se fija el valor de $X = \mathbf{x}$ la densidad de probabilidad condicional de Y debe ser

$$\pi(\mathbf{x} | \mathbf{y}) = \frac{\pi(\mathbf{x}, \mathbf{y})}{\pi_{pr}(\mathbf{x})}, \quad \text{si } \pi_{pr}(\mathbf{x}) \neq 0. \quad (3.4)$$

A esta probabilidad condicional se le conoce como función de verosimilitud, ya que expresa la probabilidad de distintas mediciones sujetas a un valor fijo de X . Además, a la distribución condicional de probabilidad

$$\pi(\mathbf{x} | \mathbf{y}_{obs}) = \frac{\pi(\mathbf{x}, \mathbf{y}_{obs})}{\pi(\mathbf{y}_{obs})}, \quad \text{si } \pi(\mathbf{y}_{obs}) \neq 0, \quad (3.5)$$

se le conoce como *distribución posterior* de X . Esta distribución expresa la información de X una vez realizada la observación $Y = \mathbf{y}_{obs}$.

Por lo tanto, para problemas inversos el problema se puede plantear de la siguiente manera: *dados los datos $Y = \mathbf{y}_{obs}$, encontrar la distribución posterior de la variable X .*

Resumiendo lo anterior se puede plantear el teorema de Bayes:

Teorema 2 (Teorema de Bayes) *Sea la variable aleatoria $X \in \mathbb{R}^n$ con densidad de probabilidad a priori $\pi_{pr}(\mathbf{x})$ y datos que consisten en observaciones \mathbf{y}_{obs} de una variable aleatoria $Y \in \mathbb{R}^m$ tal que $\pi(\mathbf{y}_{obs}) \neq 0$, entonces la distribución posterior de X dado \mathbf{y}_{obs} es:*

$$\pi_{post}(\mathbf{x}) = \pi(\mathbf{x} | \mathbf{y}_{obs}) = \frac{\pi_{pr}(\mathbf{x}) \pi(\mathbf{y}_{obs} | \mathbf{x})}{\pi(\mathbf{y}_{obs})}. \quad (3.6)$$

Para simplificar la notación se considerará $\mathbf{y}_{obs} = \mathbf{y}$ sin olvidar que se trata de observaciones. Entonces la densidad marginal de Y es

$$\pi(\mathbf{y}) = \int_{\mathbb{R}^n} \pi(\mathbf{x}, \mathbf{y}) d\mathbf{x} = \int_{\mathbb{R}^n} \pi(\mathbf{x} | \mathbf{y}) \pi_{pr}(\mathbf{x}) d\mathbf{x}, \quad (3.7)$$

y este valor puede utilizarse como constante de normalización ya que no depende de la variable X . Nótese que no puede ocurrir que $\pi(\bar{\mathbf{y}}) = 0$, ya que ello significaría que la probabilidad de que se obtengan los datos $\bar{\mathbf{y}}$ en una medición es cero, alertando que nuestro modelo no es consistente con la realidad.

En resumen, la inversión estadística se basa en el teorema de Bayes para resolver el problema inverso, y este proceso se puede dividir en tres tareas:

- Encontrar una densidad a priori $\pi_{pr}(\mathbf{x})$ adecuada que refleje la información que se conozca de la incógnita (variable aleatoria X).
- Encontrar una función de verosimilitud $\pi(\mathbf{y}_{obs}|\mathbf{x})$ adecuada que describa la interrelación entre las variables aleatorias que describen las observaciones y las incógnitas.
- Desarrollar métodos que permitan explorar la densidad de probabilidad posterior.

3.2 Estimadores puntuales

En la sección anterior se ha mostrado que la solución obtenida por inversión estadística es una distribución de probabilidad (la distribución posterior). Si la incógnita es una variable aleatoria con pocos componentes, es posible visualizar la densidad de probabilidad posterior como una función no negativa de estas variables. Sin embargo, en la mayoría de los problemas inversos, la dimensionalidad del problema inverso puede ser muy grande y la distribución posterior vive en un espacio de dimension muy alta, en el que la visualización directa es imposible. Por ejemplo, en el problema de Cauchy la dimensionalidad de la posterior es el número de nodos n de la malla de elemento finito en la frontera interior Γ_1 , y en los problemas de interés n será grande. La alternativa en estos casos es entonces calcular diferentes estimadores, que pueden ser puntuales o bien intervalos de confianza, utilizando la distribución posterior.

Las técnicas de inversión clásica, como la desarrollada en el capítulo 2, pueden verse como metodologías para producir estimaciones puntuales, sin necesidad de ligarlas a un modelo estadístico. El modelo estadístico Bayesiano también es útil para analizar esas estimaciones y tiene la ventaja de ser más versátil, ya que es posible extraer información adicional de las distribuciones de probabilidad encontradas. En nuestro caso solo consideraremos estimadores puntuales para conocer el valor más probable de la incógnita X , conocidos los datos \mathbf{y}_{obs} y la distribución a priori.

Un estimador comúnmente utilizado es el estimador *máximo a posteriori* (MAP por sus siglas en inglés) y se obtiene de la densidad de probabilidad a posteriori $\pi(\mathbf{x}|\mathbf{y})$ de la siguiente manera:

$$\mathbf{x}_{MAP} = \arg \max_{\mathbf{x} \in \mathbb{R}^n} \pi(\mathbf{x}|\mathbf{y}), \quad (3.8)$$

suponiendo que tal máximo exista. Notar que aún y cuando exista el máximo, éste puede ser no único. De cualquier manera, el problema de encontrar el MAP requiere de la solución

de un problema de optimización, y la forma más común del buscar el máximo es mediante métodos iterativos basados en gradientes. Como veremos más adelante con el problema de Cauchy, esta técnica lleva al mismo problema computacional que el obtenido con métodos de regularización clásica. Sin embargo, es esencial no mezclar los dos enfoques, ya que con el enfoque estadístico el estimador representa sólo una parte de la información desconocida que se puede extraer de la distribución posterior.

Otro estimador es *la media condicional* (CM por sus siglas en inglés), el cual se define de la siguiente manera:

$$\mathbf{x}_{CM} = E(\pi(\mathbf{x} | \mathbf{y})) = \int_{\mathbb{R}^n} \mathbf{x} \pi(\mathbf{x} | \mathbf{y}) d\mathbf{x}, \quad (3.9)$$

suponiendo que la integral converge. El problema de la estimación del CM equivale a resolver un problema de integración, en donde las propiedades de suavidad de la distribución a posterior no son tan importantes como en el problema de la estimación del MAP para resolver el problema de optimización con métodos de tipo gradiente. El principal problema técnico en la estimación del CM es que la integración típicamente se debe realizar sobre un espacio de dimensión muy alta en donde los métodos comunes de cuadratura no son aplicables. Posteriormente se mencionará como abordar este problema con los métodos de tipo Markov Chain Monte Carlo.

Existen otros estimadores como la covarianza condicional, así como el estimador de máxima verosimilitud. Éste último, aunque es el más popular estimador puntual, no un estimador Bayesiano y tampoco es útil en problemas inversos, debido a que frecuentemente equivale a resolver el problema inverso clásico sin introducir regularización.

3.3 Construcción del *modelo a priori*

Siguiendo con la discusión para resolver problemas inversos con métodos Bayesianos, en esta sección profundizaremos en la construcción del modelo a priori. En la teoría estadística de problemas inversos, la determinación de una densidad a priori es un paso muy importante, y a menudo desafiante, debido a que es deseable incorporar información del modelo a resolver en la distribución a priori. Generalmente la información que se conoce de la incógnita en la etapa previa es de naturaleza cualitativa, por lo que el problema es expresarla en forma cuantitativa para que pueda ser codificada en una densidad a priori.

Por ejemplo, en el problema de Cauchy la variable aleatoria X a estimar es una función discretizada sobre una malla que define un voltaje φ en la frontera interior Γ_1 del dominio anular Ω (ver figura 1.1 y ecuación (2.3)). Dicha información se quiere obtener a partir de mediciones de potencial eléctrico V sobre la frontera exterior (electroencefalograma), cuyos valores en los nodos de la malla en la frontera exterior Γ_2 definen otra variable Y . El modelo (3.2) que relaciona X y Y , suponiendo errores de medición E se puede considerar como el asociado al problema discreto (2.13) y expresado por (3.1) al agregar los

posibles errores de medición y discretización. Por lo tanto, se pueden suponer las siguientes propiedades de las variables X y Y :

- Cada nodo en la frontera interior está relacionado únicamente con sus vecinos inmediatos, es decir cada elemento x_i de X depende de x_{i+1} y x_{i-1} . La misma propiedad se cumple para la variable Y .
- Ambas fronteras son curvas cerradas, de tal manera que las funciones o variables definidas sobre ellas deben ser periódicas, es decir $x_{n+1} = x_0$ y $y_{m+1} = y_0$, con n y m la cantidad de nodos en la frontera interior y en la frontera exterior, respectivamente.

Esta información se puede utilizar para generar una distribución a priori, tomando en cuenta conceptos de estadística espacial, utilizando los denominados campos aleatorios de Markov Gaussianos o GMRF (Gaussian Markov Random Fields por sus siglas en inglés), ver [1]. Es decir, se genera un modelo estadístico de un vector $\mathbf{x} = (x_1, \dots, x_n)$ de parámetros distribuidos espacialmente. En este enfoque, se asignan distribuciones de probabilidad a las condicionales totales $x_i | \mathbf{x}_{-i}$, con $\mathbf{x}_{-i} = (x_1, \dots, x_{i-1}, \dots, x_n)$, obteniendo una densidad conjunta ('joint density') para el vector aleatorio \mathbf{x} (campo aleatorio de Markov). Un caso especial, es cuando las condicionales totales son Gaussianas:

$$x_i | \mathbf{x}_{-i} \sim N \left(\sum_{j \neq i} \beta_{ij} x_j, k_i^{-1} \right), \quad (3.10)$$

donde β_{ij} y k_i satisfacen $\beta_{ij} k_i = \beta_{ji} k_j$ para todo i y j . Estos modelos son adecuados para nuestro problema, pues los valores (parámetros) que queremos estimar son intensidades de voltaje sobre puntos distribuidos espacialmente en forma regular sobre una frontera interior. Si se fijan todos los parámetros exceptuando a uno, es razonable suponer que la distribución de la intensidad resultante es normal. Además, las condicionales totales tienen la distribución (3.10) si y sólo si la función de densidad conjunta para \mathbf{x} es Gaussiana y de la forma (ver [25]):

$$\pi_{pr}(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |Q|^{\frac{1}{2}} \exp \left(-\frac{1}{2} \mathbf{x}^T Q \mathbf{x} \right), \quad (3.11)$$

donde

$$[Q]_{ij} = \begin{cases} k_i, & \text{si } i = j \\ -k_i \beta_{ij}, & \text{si } i \neq j \end{cases}$$

Los GMRFs más comunes, se definen sobre mallas computacionales uniformes y corresponden a la discretización del negativo del operador Laplaciano, de donde se contruye la matriz de regularización Q . En nuestro caso, hay dos variantes que debemos tomar en cuenta para construir dicha matriz:

1. La malla corresponde a una distribución no uniforme de puntos sobre la frontera interior Γ_1 .
2. La frontera interior Γ_1 es una curva unidimensional cerrada, que podemos suponer suave.

Por lo tanto, proponemos construir la *matriz de precisión* Q a partir de la discretización del siguiente operador de segundas derivadas de arco sobre Γ_1 :

$$-\frac{\partial^2 \varphi}{\partial s^2}, \quad (3.12)$$

en donde la función φ se evalúa sobre una parametrización adecuada de Γ_1 .

Consideramos primero los nodos de la malla de elemento finito, P_i , $i = 1, \dots, n$, que se encuentran sobre Γ_1 . Con el ánimo de evitar confusión, aclaramos que los parámetros x_i en (3.10) y (3.11) corresponden a los valores $x_i = \varphi(P_i)$, que se quiere estimar. Denotemos por $P_{1+\frac{1}{2}}$ a los puntos medios entre P_i y P_{i+1} , y por h_i a las diferencias entre esos puntos. La distancia entre los puntos intermedios $P_{i-\frac{1}{2}}$ y $P_{i+\frac{1}{2}}$ es $\frac{h_{i-1}+h_i}{2}$, así que

$$\begin{aligned} \frac{\partial \varphi}{\partial s}(P_i) &\approx \frac{\varphi_{i+\frac{1}{2}} - \varphi_{i-\frac{1}{2}}}{(h_i + h_{i-1})/2}, \\ \frac{\partial^2 \varphi}{\partial s^2}(P_i) &\approx \frac{\frac{\partial \varphi}{\partial s}(P_{i+\frac{1}{2}}) - \frac{\partial \varphi}{\partial s}(P_{i-\frac{1}{2}})}{(h_i + h_{i-1})/2} \approx \frac{\frac{\varphi_{i+1} - \varphi_i}{h_i} - \frac{\varphi_i - \varphi_{i-1}}{h_{i-1}}}{(h_i + h_{i-1})/2}. \end{aligned}$$

Por lo tanto

$$-\frac{\partial^2 \varphi}{\partial s^2}(P_i) \approx -\frac{2\varphi_{i-1}}{h_{i-1}(h_{i-1} + h_i)} + \frac{2\varphi_i}{h_{i-1}h_i} - \frac{2\varphi_{i+1}}{(h_{i-1} + h_i)h_i}, \quad i = 1, \dots, n.$$

Finalmente, debido a que la frontera interior Γ_1 es una curva cerrada se introducen condiciones periódicas: $P_0 = P_n$ y $h_0 = h_n$. Se obtiene la siguiente matriz:

$$Q_1 = \alpha \begin{bmatrix} \frac{2}{h_n h_1} & \frac{-2}{(h_n + h_1)h_1} & \cdots & 0 & \frac{-2}{h_n(h_n + h_1)} \\ \frac{-2}{h_1(h_1 + h_2)} & \frac{2}{h_1 h_2} & \cdots & 0 & 0 \\ 0 & \frac{-2}{h_2(h_2 + h_3)} & \vdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \frac{-2}{(h_{n-1} + h_n)h_n} & 0 & \cdots & \frac{-2}{h_{n-1}(h_{n-1} + h_n)} & \frac{2}{h_{n-1}h_n} \end{bmatrix} = \delta L, \quad (3.13)$$

en donde α es un parámetro de regularización y L el operador negativo de segundas derivadas discretizado.

En caso de que así convenga, la matriz anterior se puede simetrizar, obteniendo la matriz

$$Q_2 = \alpha \begin{bmatrix} \frac{2}{h_n h_1} & -\frac{1}{h_1^2} & 0 & \cdots & -\frac{1}{h_n^2} \\ -\frac{1}{h_1^2} & \frac{2}{h_1 h_2} & -\frac{1}{h_2^2} & \cdots & 0 \\ 0 & -\frac{1}{h_2^2} & \frac{2}{h_2 h_3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ -\frac{1}{h_n^2} & 0 & \cdots & -\frac{1}{h_{n-1}^2} & \frac{2}{h_{n-1} h_n} \end{bmatrix}. \quad (3.14)$$

Cuando se utiliza una malla regular con tamaño de paso constante ($h_i = h$ para toda i), en un dominio periódico, se obtiene la matriz de difusión:

$$Q_3 = \frac{\alpha}{h^2} \begin{bmatrix} 2 & -1 & 0 & \cdots & \cdots & -1 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & -1 & 2 & -1 \\ -1 & 0 & \cdots & 0 & -1 & 2 \end{bmatrix}. \quad (3.15)$$

Es importante mencionar que las matrices de precisión anteriores no definen una distribución de probabilidad en forma estricta, ya que no son matrices definidas positivas. Las matrices Q_1 y Q_3 tienen vector propio $\mathbf{1}$ (vector de unos) con valor propio $\lambda = 0$, es decir $Q_1\mathbf{1} = Q_3\mathbf{1} = \mathbf{0}$; la matriz Q_2 satisface $\mathbf{h}^T Q_2 \mathbf{h} = 0$ con $\mathbf{h} = \{h_i\}_{i=1}^n \neq \mathbf{0}$. Sin embargo, dichas matrices son semidefinidas positivas, una propiedad ligada tanto a las características de difusión como a la periodicidad en las condiciones de frontera, siendo esta última propiedad la que provoca la pérdida de la positividad estricta. El correspondiente GMRF se denomina intrínseco cuando ocurre lo anterior y se escribe IGMRF, por sus siglas en inglés. Es claro entonces que la correspondiente función en (3.11) no define técnicamente una función de densidad de probabilidad, pero a pesar de ello, usualmente se sigue considerando como distribución a priori y a su correspondiente matriz se le sigue llamando matriz de precisión (ver [1]). Las muestras de estas densidades impropias se pueden generar utilizando la pseudo inversa de Moore-Penrose L^\dagger de L . Por ejemplo, para el caso en que L es simétrica, si $L = V D V^T$ es la descomposición de valores propios de L , entonces $\mathbf{w} = V(D)^{1/2}\mathbf{v}$, con $\mathbf{v} \sim N(\mathbf{0}, I)$, es una muestra de $N(\mathbf{0}, L^\dagger)$. En los experimentos numéricos mostraremos resultados cuando se utilizan estas matrices de precisión.

3.4 Construcción de la función de verosimilitud

La construcción de la función de verosimilitud suele ser la parte más directa en el proceso de inversión estadística, ya que involucra al modelo asociado al problema directo (forward problem). En nuestro caso el problema directo está asociado a la ecuación de estado (2.1), el cual se discretiza por medio del método de los elementos finitos, dando lugar a la función nodal lineal $\mathbf{y} = f(\mathbf{x}) = A\mathbf{x}$ descrita en (2.14)-(2.15). La función de verosimilitud contiene dicha información en forma implícita, además del ruido numérico, los errores de medición de los datos y otras posibles incertidumbres.

En muchos problemas el ruido se puede considerar como aditivo y mutuamente independiente de la incógnita X , lo cual se puede constatar en el estudio de los métodos clásicos de inversión, por ejemplo. En nuestro caso es razonable asumir estas suposiciones, debido a que los datos Y se obtienen de mediciones sobre la frontera exterior del dominio anular, y las incógnitas X se obtienen sobre la frontera interior, a la vez que no interesa medir la discrepancia entre los datos y los valores obtenidos por medio de la solución del modelo matemático sobre la frontera exterior, la cual se incluye en E . Por lo tanto, el modelo estocástico

se puede escribir en la forma siguiente:

$$Y = f(X) + E, \quad X \in \mathbb{R}^n, Y \in \mathbb{R}^m, E \in \mathbb{R}^m, \quad (3.16)$$

en donde X y E son mutuamente independientes. Suponemos además que la distribución del error E es conocida e igual a

$$\mu_E(B) = P(E \in B) = \int_B \pi_{noise}(\mathbf{e}) d\mathbf{e}. \quad (3.17)$$

Si se fija $X = \mathbf{x}$, la suposición de independencia mutua de X y E asegura que la densidad de probabilidad de E permanece inalterada cuando se condiciona sobre $X = \mathbf{x}$. Por lo tanto, se deduce que Y condicionada sobre $X = \mathbf{x}$ se distribuye como E , cuya densidad de probabilidad es trasladada por $f(\mathbf{x})$, es decir, la función de verosimilitud es

$$\pi(\mathbf{x} | \mathbf{y}) \propto \pi_{noise}(\mathbf{e}) = \pi_{noise}(\mathbf{y} - f(\mathbf{x})). \quad (3.18)$$

Considerando una distribución Gaussiana para el ruido $E \sim N(0, \sigma_{noise}^2 I)$, en la relación anterior se obtiene:

$$\pi(\mathbf{x} | \mathbf{y}) \propto \exp\left(-\frac{1}{2\sigma_{noise}^2}(\mathbf{y} - f(\mathbf{x}))^T(\mathbf{y} - f(\mathbf{x}))\right) \quad (3.19)$$

$$= \exp\left(-\frac{1}{2\sigma_{noise}^2}\|\mathbf{y} - A\mathbf{x}\|_{\mathbb{R}^m}^2\right), \quad (3.20)$$

en donde $f(\mathbf{x}) = A\mathbf{x}$ y la constante σ_{noise} expresa la relación de señal a ruido estándar. Este valor usualmente se calcula como se hace en muchos problemas de medición, y satisface $\frac{\max(\mathbf{y})}{\sigma_{noise}} \approx 100$. En otras palabras σ_{noise} es aproximadamente el 1% del máximo de las mediciones captadas.

3.5 Formulación del modelo Bayesiano completo

Por el teorema de Bayes, la distribución posterior satisface:

$$\pi_{post}(\mathbf{x}) = \frac{\pi_{pr}(\mathbf{x}) \pi(\mathbf{y}_{obs} | \mathbf{x})}{\pi(\mathbf{y}_{obs})} \propto \pi_{pr}(\mathbf{x}) \pi(\mathbf{y}_{obs} | \mathbf{x}),$$

y sustituyendo las expresiones encontradas en las dos secciones previas, se obtiene:

$$\pi_{post}(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\mathbf{x}^T Q_2 \mathbf{x}\right) \exp\left(-\frac{1}{2\sigma_{noise}^2}\|\mathbf{y} - A\mathbf{x}\|_{\mathbb{R}^m}^2\right).$$

Recordando que la matriz de precisión en el modelo a priori, definida en (3.14), es $Q_2 = (1/\gamma^2)L$, se obtiene la distribución de probabilidad a posterior:

$$\pi_{post}(\mathbf{x}) \propto \exp\left(-\frac{1}{2\gamma^2}\mathbf{x}^T L \mathbf{x} - \frac{1}{2\sigma_{noise}^2}\|\mathbf{y} - A\mathbf{x}\|_{\mathbb{R}^m}^2\right). \quad (3.21)$$

La matriz de difusión L es simétrica, y de su decomposición $L = U\Lambda U^T$, con U matriz ortogonal y $\Lambda = \text{diag}(\lambda_i)$ la matriz diagonal con los valores propios $\lambda_i \geq 0$, se puede expresar como $L = D^T D$ con $D = \Lambda^{1/2} U^T$. Por lo tanto (3.21) se puede escribir como

$$\pi_{post}(\mathbf{x}) \propto \exp\left(-\frac{1}{2\gamma^2}\|D\mathbf{x}\|_{\mathbb{R}^n}^2 - \frac{1}{2\sigma_{noise}^2}\|\mathbf{y} - A\mathbf{x}\|_{\mathbb{R}^m}^2\right). \quad (3.22)$$

La matriz D se puede interpretar como una matriz de discretización de la derivada $\partial\varphi/\partial s$ sobre Γ_1 , como ya se ha indicado al final del capítulo anterior.

3.5.1 La distribución de la posterior es Gaussiana

A pesar de que $L = D^T D$ es una matriz semidefinida positiva, la expresión (3.22) define una distribución Gaussiana con matriz de covarianza:

$$\Sigma = \left(\frac{1}{\gamma^2}D^T D + \frac{1}{\sigma_{noise}^2}A^T A\right)^{-1} \quad (3.23)$$

y centro

$$\bar{\mathbf{x}} = \Sigma A^T \mathbf{y}. \quad (3.24)$$

Para verificarlo, primero debemos mostrar que la matriz simétrica

$$S = \frac{1}{\gamma^2}D^T D + \frac{1}{\sigma_{noise}^2}A^T A \quad (3.25)$$

es invertible, y para ello basta verificar que es definida positiva.

Sea $\mathbf{x} \in \mathbb{R}^n$ tal que

$$\mathbf{0} = \mathbf{x}^T S \mathbf{x} = \frac{1}{2\gamma^2}\|D\mathbf{x}\|_{\mathbb{R}^n}^2 + \frac{1}{2\sigma_{noise}^2}\|A\mathbf{x}\|_{\mathbb{R}^m}^2,$$

entonces \mathbf{x} satisface $D\mathbf{x} = A\mathbf{x} = \mathbf{0}$, es decir $\mathbf{x} \in \text{Ker}(D) \cap \text{Ker}(A)$. La condición $D\mathbf{x} = \mathbf{0}$ indica que el potencial es constante a lo largo de la frontera interior del dominio anular y la condición $A\mathbf{x} = \mathbf{0}$ indica que el potencial de salida sobre la frontera exterior también es cero. Utilizando el formalismo matemático, la matriz A es la versión nodal del operador discreto K_h , el cual es lineal e inyectivo ya que el operador continuo K tiene esa propiedad, (teorema 1 y relación (2.3)). La inyectividad de A implica $\mathbf{x} = \mathbf{0}$, y se concluye que la matriz Σ es definida positiva e invertible.

Por otro lado, se sabe que cuando ambas densidades, la prior y la verosimilitud son Gaussianas entonces el MAP = CM define el centro. El MAP se obtiene minimizando el potencial posterior

$$V(\mathbf{x} | \mathbf{y}) = \frac{1}{2\gamma^2}\|D\mathbf{x}\|_{\mathbb{R}^n}^2 + \frac{1}{2\sigma_{noise}^2}\|\mathbf{y} - A\mathbf{x}\|_{\mathbb{R}^m}^2. \quad (3.26)$$

3.6. RELACIÓN ENTRE REGULARIZACIÓN DE TIJONOV Y LA DISTRIBUCIÓN A PRIORI 27

Como la función es cuadrática y su Hessiana S es definida positiva, su centro se encuentra precisamente en el punto donde se obtiene el mínimo, el cual coincide con el punto que satisface $\nabla V(\mathbf{x}|\mathbf{y}) = \mathbf{0}$, es decir,

$$\frac{1}{\gamma^2} D^T D \mathbf{x} + \frac{1}{\sigma_{noise}^2} A^T (A \mathbf{x} - \mathbf{y}) = \mathbf{0}. \quad (3.27)$$

Por lo tanto,

$$\bar{\mathbf{x}} = \left(\frac{1}{\gamma^2} D^T D + \frac{1}{\sigma_{noise}^2} A^T A \right)^{-1} A^T \mathbf{y} = S^{-1} A^T \mathbf{y} = \Sigma A^T \mathbf{y}.$$

3.5.2 Necesidad de explorar la densidad posterior

Quisieramos enfatizar que la distribución de probabilidad posterior es la solución del problema inverso cuando se utiliza inversión estadística Bayesiana. Por lo tanto, es necesario explorarla para extraer la información de interés. Anteriormente se introdujeron estimadores basados en la distribución a posterior. Se mostró por ejemplo que la estimación del máximo a posterior implica un problema de optimización de la forma (3.8), mientras que la media condicional requiere de integración sobre el espacio \mathbb{R}^n donde se define la probabilidad a posterior. Está claro que si la dimensión n del espacio de parámetros \mathbb{R}^n es grande, el uso de reglas de cuadratura numéricas no es posible en forma práctica ya que eso requeriría m^n puntos de integración, lo cual excede la capacidad de la mayoría de las computadoras. Otro problema con las reglas de cuadratura es que se requiere de un buen conocimiento del soporte de la distribución a posterior, pero éste es desconocido ya que es parte de la información que se busca.

Una forma alternativa de mirar el problema es la siguiente. En lugar de evaluar la densidad de probabilidad en puntos dados, se deja que la densidad misma determine un conjunto de puntos que soporte la distribución. Estos puntos pueden ser entonces utilizados para la integración aproximada. Los métodos MCMC son (al menos en el nivel conceptual) algoritmos relativamente simples para generar conjuntos de puntos del soporte que permiten realizar integración de tipo Monte Carlo en forma razonable.

Antes de abordar los métodos MCMC, queremos volver al tema del cálculo de la distribución a priori, para agotar algunos aspectos importantes. Aprovechando la relación que hemos observado entre los funcionales de Tijonov y el potencial posterior en el caso gaussiano puro.

3.6 Relación entre regularización de Tijonov y la distribución a priori

La matriz Hessiana en S (3.25), asociada al potencial $V(\mathbf{x}|\mathbf{y})$ en (3.26) está relacionada con la derivada del funcional (3.30) y su derivada (2.19) si se escoge el parámetro de regularización como

$$\alpha = \sigma_{noise}^2 / \gamma^2. \quad (3.28)$$

Además, la solución del problema inverso en el caso determinista resuelve las ecuaciones normales (3.27). Por lo tanto, en el caso puramente Gaussiano, el potencial posterior $V(\mathbf{x} | \mathbf{y})$ es igual al funcional de Tijonov $F_{\alpha s}$, salvo una constante multiplicativa.

Por otro lado, el punto central de la distribución posterior es la solución regularizada de Tijonov para la ecuación $A\mathbf{x} = \mathbf{y}$, con el parámetro de regularización dado por (3.28). Este parámetro de regularización, de acuerdo a (3.28) es proporcional al nivel de ruido en los datos σ_{noise}^2 . Esta propiedad coincide con el resultado en [10] en donde se demuestra que los mejores resultados con regularización de Tijonov se encuentran cuando el parámetro de regularización α es proporcional al nivel del ruido en los datos. Adicionalmente, se demuestra en [10] que para obtener soluciones numéricas convergentes, el tamaño de la malla h para discretizar el modelo continuo con elementos finitos lineales, es suficiente con escoger $h^{3/2} \leq \sigma_{noise}^2$.

3.6.1 Regularizadores suavizantes

Anteriormente derivamos la matriz de precisión Q_3 en (3.15) discretizando directamente la segunda derivada (3.12) sobre Γ_1 , siguiendo la construcción de GMRFs. A continuación queremos aprovechar la analogía entre la regularización de Tijonov y las distribuciones a priori con el objeto de reconstruir las matrices de precisión comunes en el modelo (3.11), pero cuando las mallas son arbitrarias, posiblemente no uniformes y no necesariamente de la discretización de elemento finito. Partimos de la propuesta de distribución posterior

$$\pi_{post}(\mathbf{x}) \propto \exp\left(-\frac{1}{2\gamma^2}\|D\mathbf{x}\|_{\mathbb{R}^n}^2 - \frac{1}{2\sigma_{noise}^2}\|\mathbf{y} - A\mathbf{x}\|_{\mathbb{R}^m}^2\right) \quad (3.29)$$

en donde D es una matriz asociada a la discretización de la derivada $\partial\varphi/\partial s$, el cual sería el análogo discreto del funcional $J_{\alpha s}$ definido en (2.6) con función cuadrática

$$F_{\alpha s}(\mathbf{x}) = \frac{\alpha}{2}\|D\mathbf{x}\|_{\mathbb{R}^n}^2 + \frac{1}{2}\|\mathbf{y} - A\mathbf{x}\|_{\mathbb{R}^m}^2, \quad (3.30)$$

con $\alpha = \sigma_{noise}^2/\gamma^2$. Una discretización natural sobre los nodos de la malla sobre la frontera interior Γ_1 sería:

$$\frac{\partial\varphi}{\partial s}(P_i) \approx \frac{\varphi_{i+1} - \varphi_i}{h_i}, \quad i = 1, \dots, n.$$

La matriz D en este caso, considerando periodicidad ($\varphi_{n+1} = \varphi_1$) es

$$D = \begin{bmatrix} -\frac{1}{h_1} & \frac{1}{h_1} & 0 & \dots & 0 & 0 \\ 0 & -\frac{1}{h_2} & \frac{1}{h_2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\frac{1}{h_{n-1}} & \frac{1}{h_{n-1}} \\ \frac{1}{h_n} & 0 & 0 & \dots & 0 & -\frac{1}{h_n} \end{bmatrix}. \quad (3.31)$$

3.6. RELACIÓN ENTRE REGULARIZACIÓN DE TIJONOV Y LA DISTRIBUCIÓN A PRIORI 29

En este caso no utilizamos diferencias finitas centrales. Se propone la matriz de precisión

$$Q = \alpha D^T D = \alpha \begin{bmatrix} \left(\frac{1}{h_1^2} + \frac{1}{h_n^2}\right) & -\frac{1}{h_1^2} & 0 & \dots & 0 & -\frac{1}{h_n^2} \\ -\frac{1}{h_1^2} & \left(\frac{1}{h_1^2} + \frac{1}{h_2^2}\right) & -\frac{1}{h_2^2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \left(\frac{1}{h_{n-2}^2} + \frac{1}{h_{n-1}^2}\right) & -\frac{1}{h_{n-1}^2} \\ -\frac{1}{h_n^2} & 0 & 0 & \dots & -\frac{1}{h_{n-1}^2} & \left(\frac{1}{h_{n-1}^2} + \frac{1}{h_n^2}\right) \end{bmatrix}. \quad (3.32)$$

Esta matriz es una alternativa a la matriz que define a Q_2 en (3.14). Claramente la matriz $L = D^T D$ es simétrica y tiene vector propio $\mathbf{h} = (h_i)_{i=1}^n$ con valor propio $\lambda = 0$, y es el análogo discreto de (3.12).

Obsérvese que podemos interpretar D^T como el análogo discreto del operador diferencial $-\partial/\partial s$, el cual es el adjunto a $\partial/\partial s$ bajo la norma en $L^2(\Gamma_1)$. Además, haciendo integración por partes

$$\left\langle -\frac{\partial^2 \varphi}{\partial s^2}, \eta \right\rangle_{L^2(\Gamma_1)} = \int_{\Gamma_1} -\frac{\partial^2 \phi}{\partial s^2} \eta ds = \int_{\Gamma_1} \frac{\partial \phi}{\partial s} \frac{\partial \eta}{\partial s} ds = \left\langle \frac{\partial \varphi}{\partial s}, \frac{\partial \eta}{\partial s} \right\rangle_{L^2(\Gamma_1)}, \quad (3.33)$$

para toda φ y η en $H^1(\Gamma_1)$. En la expresión anterior no hay términos frontera debido a Γ_1 es una curva cerrada. El análogo discreto es

$$\langle D^T D \mathbf{x}, \mathbf{z} \rangle_{\mathbb{R}^n} = (D^T D \mathbf{x})^T \mathbf{z} = (D \mathbf{x})^T D \mathbf{z} = \langle D \mathbf{x}, D \mathbf{z} \rangle_{\mathbb{R}^n}, \quad (3.34)$$

para toda \mathbf{x} y \mathbf{z} en \mathbb{R}^n .

Algunos autores utilizan la norma cuadrada del Laplaciano como término regularizante el cual es todavía más suavizante que la norma cuadrada del gradiente. En nuestro caso el término es

$$\frac{\alpha}{2} \left\| \frac{\partial^2 \varphi}{\partial s^2} \right\|_{L^2(\Gamma_1)}^2 \quad \text{en lugar de} \quad \frac{\alpha}{2} \left\| \frac{\partial \varphi}{\partial s} \right\|_{L^2(\Gamma_1)}^2.$$

En este caso la matriz de precisión para el GMRF correspondiente se puede tomar como $Q = \alpha L^T L$, siendo $L = D^T D$. Es decir, la matriz de precisión Q ahora es una discretización del operador biarmónico y se puede construir a partir de la matriz básica simple D que representa una discretización del gradiente. Más aún, es posible obtener diferentes opciones utilizando diferentes esquemas de discretización del gradiente y diferentes opciones de aproximación como diferencias finitas, esquemas compactos, esquemas espectrales, elemento finito, entre muchas otras.

Concluimos con una tabla para enlistar las posibles regularizaciones cuadráticas de Tjonov y las correspondientes matrices de precisión para los GMRF definidos en mallas no

regulares, en donde I es la matriz identidad de $n \times n$:

Caso	Regularización de Tijonov	Matriz del GMRF
Simple	$\frac{\alpha}{2} \ \varphi\ ^2$	$Q = \alpha I$
Suavizante	$\frac{\alpha}{2} \left\ \frac{\partial \varphi}{\partial s} \right\ ^2$	$Q = \alpha L = \alpha D^T D$
Más suavizante	$\frac{\alpha}{2} \left\ \frac{\partial^2 \varphi}{\partial s^2} \right\ ^2$	$Q = \alpha L^T L$

3.6.2 Términos de regularización no suavizantes

Las anteriores estrategias de regularización, especialmente la segunda y tercera son suavizantes. Sin embargo, en muchos problemas la solución φ puede ser discontinua en ciertas regiones de su dominio, o bien se pueden presentar cambios muy pronunciados. Por ejemplo, en la recuperación de imágenes existen cambios de intensidad muy fuertes alrededor de ciertos píxeles que definen contornos de cambio de contraste, en donde los gradientes son muy altos. Una opción para ese tipo de problemas es utilizar TV-regularización:

$$J_{TV}(\varphi) = \frac{\alpha}{2} \left\| \frac{\partial \varphi}{\partial s} \right\|_{L^1(\Gamma_1)} + \frac{1}{2} \|K\varphi - V\|_{L^2(\Gamma_2)}^2. \quad (3.35)$$

cuya derivada es

$$DJ_{TV}(\varphi) = -\alpha \frac{\partial}{\partial s} \left(\omega(s) \frac{\partial \varphi}{\partial s} \right) + K^*(K\varphi - V), \quad (3.36)$$

donde $\omega(s) = |\partial \varphi / \partial s|^{-1/2}$ y condiciones de frontera periódicas. Por lo tanto, la correspondiente matriz de precisión Q asociada al GMRF es de la forma:

$$Q = \alpha D^T W D, \quad (3.37)$$

donde $W = \text{diag}(w_1, \dots, w_n)$, con w_i una discretización de $\omega(s_i)$ sobre Γ_i , por ejemplo $w_i = |(D\mathbf{x})_i|^{-1/2}$. Para evitar dividir por cantidades cercanas a cero, se puede definir $\omega(s) = |\epsilon + \partial \varphi / \partial s|^{-1/2}$ con $0 < \epsilon$ pequeño.

Este último caso de regularización, proporciona un ejemplo particular de IGMRF para operadores de difusión con coeficientes de difusión variables espacialmente, discutidos en [1]. Desde el punto de vista estadístico, se permite que la varianza se incremente más en ciertas regiones que en otras: un incremento grande de la varianza implica menor precisión y viceversa. En TV-regularización, la difusión $w(s)$ crece cuando el gradiente decrece y viceversa.

Terminamos recordando que los regularizadores anteriores y sus correspondientes matrices de precisión (modelos y distribuciones a priori) tienen generalizaciones directas para modelos con EDP en dimensiones mayores a uno, ya sea con discretizaciones que utilizan tanto mallas regulares como mallas no uniformes.

3.7 Métodos de Monte Carlo para Cadenas de Markov

Los métodos conocidos como MCMC por sus siglas en inglés (Markov Chain Monte Carlo) constituyen una herramienta extraordinariamente poderosa para el muestreo de densidades de probabilidad muy complicadas y de alta dimensión, y en consecuencia brindan la posibilidad de calcular en forma indirecta la esperanza matemática de las correspondientes distribuciones, la cual generalmente es imposible de calcular en la forma tradicional.

Ya en 1777 George Louis Leclerc Comte de Buffon estableció un método de Monte Carlo para aproximar una densidad de probabilidad π arrojando una aguja al azar sobre una cuadrícula de líneas paralelas y registrando la frecuencia de aparición de la aguja sobre las mismas líneas. Actualmente, el objetivo primordial de los métodos MCMC, como ya se mencionó, es el muestreo estadístico de alguna distribución de probabilidad que generalmente es muy complicada, que denominaremos π , o bien calcular el valor esperado $Ef(\mathbf{x})$ de una función complicada $f(\mathbf{x})$, en donde X es una variable aleatoria con distribución π (hecho que denotaremos por $X \approx \pi$) en dimensión alta.

En la época moderna los métodos se desarrollaron a partir de 1940, cuando algunos científicos (entre ellos, Nicolas C. Metropolis, Richard Feynman y John von Neumann) que trabajaban en el diseño de la bomba atómica, introdujeron los principios para el desarrollo del algoritmo de Metropolis-Hastings, dando inicio a una de las grandes ramas de la estadística contemporánea: los métodos de Monte Carlo. Más adelante, en 1949 Metropolis junto con el matemático Ulam en [22] presentó la idea del método a la comunidad estadística por medio del artículo *The Monte Carlo Method*, ahí se presentaron dos ideas: una, para calcular la tasa de éxito siguiendo una estrategia en el juego solitario mediante muchos ensayos y considerando la proporción de aciertos; otra, para encontrar el volumen de una región de 20 dimensiones contenida en un cubo unitario realizando simulaciones de variables aleatorias dentro del cubo y considerando la proporción de puntos dentro de la región dada. Esta técnica a la postre fue de mucha importancia debido a que proporcionaba un método práctico para tratar con integrales de dimensiones altas evitando el problema del cálculo de integración iterada aproximada, lo cual hasta ese momento había sido en general un problema bastante complicado. Éste método se le conoce hoy en día integración de Monte Carlo. Por supuesto, la estimación se ajustará, con gran probabilidad al valor real, mientras que el número de ensayos sea lo suficientemente grande.

En 1953 Metropolis publicó más contribuciones, ahora con la colaboración de dos parejas de casados, Marshall y Arianna Rosenbluth, así como Edward y Augusta Teller, en una revista de fisicoquímica, [21]. Se presenta el cálculo de una integral multidimensional de orden $2N$ (que define una función de energía) por medio de la utilización de N partículas en un cuadrado, para encontrar diferentes configuraciones de estado mediante el movimiento aleatorio (con distribución uniforme) de cada partícula. La configuración se acepta o rechaza dependiendo de una probabilidad en función de la energía del sistema. Un resultado

central es la ergodicidad del método, y que el sistema tiende a una distribución proporcional a $\exp(-E_r/kT)$ para cada estado r , además de que las probabilidades de transición de estado son simétricas, es decir, la probabilidad de ir de un estado r a otro s es la misma que la de ir de s a r . El trabajo, realizado desde el punto de vista físico, no llamó mucho la atención de la comunidad estadística, pero en 1964 el método fue reconocido como un método para resolver problemas en mecánica estadística de equilibrio. El paso decisivo en el desarrollo del método de Metropolis-Hastings(M-H) lo dió en 1970 precisamente el profesor de la U. de Toronto, W. Keith Hastings, generalizando en método con la introducción de la matriz de transición de una cadena de Markov[14] . A pesar de ello, estos métodos no fueran utilizados regularmente en la estadística. Hubo que esperar hasta 1997, cuando Robert Kass presenta los algoritmos conocidos como Gibbs sampling y el algoritmo M-H en [17]. Vale la pena destacar que en esa época ya había un desarrollo muy importante en el poder de cómputo, lo cual permitió que se retomaran y se difundieran ampliamente estos métodos.

Para comprender mejor los métodos MCMC, y sus resultados, se introducen primero los conceptos generales de cadenas de Markov.

Definición 2 (Cadena de Markov) *Una cadena de Markov es una sucesión ordenada de variables aleatorias $X_0, X_1, \dots, X_n, \dots$ en \mathbb{R}^n denotada por $\{X_j\}_{j=1}^{\infty}$ con la propiedad de que la distribución condicional de $X_n = \mathbf{x}_n$ dado $X_0 = \mathbf{x}_0, X_1 = \mathbf{x}_1, \dots, X_{n-1} = \mathbf{x}_{n-1}$, es igual la distribución condicional dada únicamente $X_{n-1} = \mathbf{x}_{n-1}$ es decir*

$$P(\mathbf{x}_n | \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{n-1}) = P(\mathbf{x}_n | \mathbf{x}_{n-1}),$$

Esta propiedad puede interpretarse como que la cadena tiene memoria corta.

Definición 3 *Una cadena de Markov se dice homogénea en el tiempo si $P(\mathbf{x}_{i+1} = \mathbf{b} | \mathbf{x}_i = \mathbf{a}) = P(\mathbf{x}_1 = \mathbf{b} | X_0 = \mathbf{a})$ para toda $i \geq 0$ y para toda \mathbf{a}, \mathbf{b} .*

Es decir, la cadena no depende del tiempo en ningún estado.

Definición 4 (Kernel de transición) *Una función K , definida en el producto cruz de un espacio de probabilidad \mathcal{X} y los borelianos $\mathcal{B}(\mathcal{X})$, Se dice que es un kernel de transición si cumple lo siguiente*

- Para todo $\mathbf{x} \in \mathcal{X}$, la función $K(\mathbf{x}, \cdot)$ es una medida de probabilidad.
- Para todo subconjunto $A \in \mathcal{B}(\mathcal{X})$ $K(\cdot, A)$ es una función medible.

El concepto se simplifica cuando \mathcal{X} es un espacio de probabilidad discreto. En este caso el kernel de transición es una matriz P , denominada matriz de transición, en donde las entradas p_{ij} representan la probabilidad de ir del estado \mathbf{x}_i al estado \mathbf{x}_j es decir

$$p_{ij} = P(\mathbf{x}_n = \mathbf{x}_j | X_{n-1} = \mathbf{x}_i) \quad \text{con } \mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}.$$

Definición 5 Una medida μ es invariante con respecto al kernel K si cumple que $\mu K = \mu$, es decir

$$\mu_{X_{j+1}}(B_{j+1}) = \int_{X_j} K(\mathbf{x}_j, B_{j+1}) \mu_{X_j}(d\mathbf{x}_j),$$

es decir, la distribución de a variable aleatoria X_j antes del paso $j \rightarrow j + 1$ es igual a la distribución de la variable aleatoria X_{j+1} después del paso.

Definición 6 Una cadena de Markov se dice que es irreducible en términos del Kernel K con respecto a μ si para todo $\mathbf{x} \in \mathcal{X}$ y todo $B \in \mathcal{B}(\mathcal{X})$ tal que $\mu(B) > 0$ se cumple que $K^j(\mathbf{x}, B) > 0$ para algún $j \in \mathbb{N}$.

Es decir, de la definición anterior se deduce que, no importando que estado \mathbf{x} comience la cadena de Markov dicha cadena visitará con probabilidad positiva todo conjunto de medida mayor que cero.

Definición 7 Una cadena de Markov (\mathbf{x}_n) se dice aperiódica si

$$\text{mcd}\{n \geq 1 : P(\mathbf{x}_n = \mathbf{x} | X_0 = \mathbf{x}) > 0\} = 1 \quad \text{para todo } \mathbf{x} \in \mathcal{X}.$$

Es decir, no existe un subconjunto $A \subset \mathcal{X}$ tal que la cadena de Markov se mantenga dentro de ese conjunto sin importar el número de pasos. En otras palabras no existe subconjunto $A \subset \mathcal{X}$ tal que $K^n(\mathbf{x}, A) = 1$ con $\mathbf{x} \in A$ para todo $n \in \mathbb{N}$

3.7.1 Integración de Monte Carlo

Sea μ una medida de probabilidad sobre \mathbb{R}^n y f una función medible e integrable sobre \mathbb{R}^n con respecto a la medida μ , es decir, $f \in L^1(\mu(d\mathbf{x}))$. Consideremos que el objetivo es estimar la integral de f con respecto a la medida μ . Cuando se utilizan métodos de integración numérica tradicionales, generalmente se definen un conjunto de puntos de soporte $\mathbf{x}_j \in \mathbb{R}^n$ $1 \leq j \leq N$ y los pesos correspondientes w_j para obtener la aproximación de la integral de f como sigue

$$\int_{\mathbb{R}^n} f(\mathbf{x}) \mu(d\mathbf{x}) \approx \sum_{j=1}^N w_j f(\mathbf{x}_j).$$

Típicamente estos métodos de cuadratura se diseñan procurando que sean precisos o exactos para las funciones que pertenecen al espacio finito dimensional generado por funciones base como los polinomios de un grado menor o igual a un valor dado (restringido). Por otro lado, en la integración por Monte Carlo, los puntos de soporte \mathbf{x}_j son generados aleatoriamente a partir de una densidad de probabilidad, y los pesos w_j son entonces determinados por la distribución μ . Idealmente los puntos de soporte se obtienen de la distribución de probabilidad misma. De hecho, sea $X \in \mathbb{R}^n$ una variable aleatoria cuya distribución de probabilidad es μ . Si tuviésemos un generador que pudiera producir repetidas realizaciones de X , se podrían generar un conjunto de puntos distribuidos de acuerdo a μ .

Suponiendo que $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subset \mathbb{R}^n$ es tal grupo de muestras distribuidas de acuerdo a μ , entonces podríamos aproximar la integral de f por medio del promedio ergódico

$$\int_{\mathbb{R}^n} f(\mathbf{x})\mu(d\mathbf{x}) \approx E\{f(\mathbf{x})\} \approx \frac{1}{N} \sum_{j=1}^N f(\mathbf{x}_j).$$

Los métodos MCMC son formas sistemáticas de generar conjuntos de agregados de puntos tales que la anterior aproximación se cumpla.

Teorema 3 *Sea μ una medida de probabilidad en \mathbb{R}^n y $\{X_j\}$ una cadena de Markov homogénea en tiempo con kernel de transición K . Se considera que μ es una medida invariante del kernel de transición K , y que K irreducible y aperiódica. Entonces para todo $\mathbf{x} \in \mathbb{R}^n$*

$$\lim_{N \rightarrow \infty} K^{(N)}(\mathbf{x}, B) = \mu(B) \quad \forall B \in \mathcal{B}, \quad (3.38)$$

y para todo $f \in L^1(\mu(d\mathbf{x}))$,

$$\frac{1}{N} \sum_{j=1}^N f(\mathbf{x}_j) = \int_{\mathbb{R}^n} f(\mathbf{x})\mu(d\mathbf{x}), \quad (3.39)$$

casi seguramente (es decir, con probabilidad 1).

El resultado anterior proporciona una indicación clara de como explorar una distribución de probabilidad dada: uno necesita construir un kernel de transición invariante, aperiódico e irreducible, y generar una sucesión de puntos muestra $\mathbf{x}_1, \mathbf{x}_2, \dots$ utilizando este kernel, es decir, debemos calcular una realización de la cadena de Markov.

En la siguiente parte se discutirá como construir un kernel de transición con las propiedades antes mencionadas, en este caso nos interesará construir el kernel llamado de Metropolis-Hastings el cual es utilizado en el algoritmo que tiene el mismo nombre.

3.7.2 Algoritmo de Metropolis-Hastings

Como ya se mencionó anteriormente el algoritmo de Metropolis-Hastings (M-H) es uno de los algoritmos más usados actualmente en estadística, ya que es uno de los algoritmos más sencillos dentro de la familia de métodos MCMC. Este algoritmo tiene como base principal construir una cadena de Markov cuya densidad de equilibrio sea precisamente la densidad buscada o de interés, que en nuestro caso es la posterior salvo una constante.

Sea P cualquier kernel transicional. Dado un punto $\mathbf{x} \in \mathbb{R}^n$, queremos postular que el kernel propone un salto a otro punto $\mathbf{y} \in \mathbb{R}^n$ o de otra manera no se mueve de \mathbf{x} . Es decir, el kernel se puede descomponer en dos partes:

$$P(\mathbf{x}, B) = \int_B K(\mathbf{x}, \mathbf{y}) d\mathbf{y} + r(\mathbf{x}) \chi_B(\mathbf{x}), \quad \text{con } \chi_B(\mathbf{x}) = \begin{cases} 1, & \text{si } \mathbf{x} \in B \\ 0, & \text{si } \mathbf{x} \notin B \end{cases}$$

Aunque $K(\mathbf{x}, \mathbf{y})$ es realmente una densidad, se puede pensar a $K(\mathbf{x}, \mathbf{y}) d\mathbf{y}$ como la probabilidad del movimiento desde \mathbf{x} al conjunto infinitesimal $d\mathbf{y}$ en \mathbf{y} , mientras que $r(\mathbf{x}) \geq 0$ es la probabilidad de \mathbf{x} permanezca inmóvil. La función característica se introduce porque si $\mathbf{x} \notin B$, la única forma de alcanzar B es por medio de un movimiento. La condición $P(\mathbf{x}, \mathbb{R}^n) = 1$ implica que

$$r(\mathbf{x}) = 1 - \int_{\mathbb{R}^n} K(\mathbf{x}, \mathbf{y}) d\mathbf{y}. \quad (3.40)$$

Para que $\pi d\mathbf{x}$ sea una medida invariante de P se debe satisfacer la identidad

$$\begin{aligned} \mu P(B) &= \int_{\mathbb{R}^n} \left(\int_B K(\mathbf{x}, \mathbf{y}) d\mathbf{y} + r(\mathbf{x}) \chi_B(\mathbf{x}) \right) \pi(\mathbf{x}) d\mathbf{x} \\ &= \int_B \left(\int_{\mathbb{R}^n} \pi(\mathbf{x}) K(\mathbf{x}, \mathbf{y}) d\mathbf{x} + r(\mathbf{y}) \pi(\mathbf{y}) \right) d\mathbf{y} = \int_B \pi(\mathbf{y}) d\mathbf{y}, \end{aligned}$$

para todo conjunto de Borel B . Luego

$$\pi(\mathbf{y})(1 - r(\mathbf{y})) = \int_{\mathbb{R}^n} \pi(\mathbf{x}) K(\mathbf{x}, \mathbf{y}) d\mathbf{x}.$$

Por la fórmula anterior para $r(\mathbf{x})$, se obtiene

$$\int_{\mathbb{R}^n} \pi(\mathbf{y}) K(\mathbf{y}, \mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^n} \pi(\mathbf{x}) K(\mathbf{x}, \mathbf{y}) d\mathbf{x}. \quad (3.41)$$

Esta condición se conoce como la *ecuación de balance*. En particular, si K satisface la *ecuación de balance detallado*:

$$\pi(\mathbf{y}) K(\mathbf{y}, \mathbf{x}) = \pi(\mathbf{x}) K(\mathbf{x}, \mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \quad (3.42)$$

entonces la ecuación de balance se satisface forzosamente.

Las dos condiciones anteriores, son el punto de partida para la construcción de kernels de transición de cadenas de Markov utilizadas en muestreo estocástico. De hecho, en el algoritmo Metropolis-Hasting, el objetivo es construir un kernel de transición que satisface la ecuación de balance detallado (3.42).

A una función $q : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}_+$ con la propiedad $\int_{\mathbb{R}^n} q(\mathbf{x}, \mathbf{y}) d\mathbf{y} = 1$, se le llama *distribución propuesta* o *kernel candidato-generador* y define el kernel transicional:

$$Q(\mathbf{x}, A) = \int_A q(\mathbf{x}, \mathbf{y}) d\mathbf{y}.$$

Si ocurre que Q satisface la ecuación de balance detallado, entonces simplemente se toma $K(\mathbf{x}, \mathbf{y}) = q(\mathbf{x}, \mathbf{y})$ y $r(\mathbf{x}) = 0$ y hemos terminado. De otra manera, se corrige el kernel por un factor multiplicativo $\alpha(\mathbf{x}, \mathbf{y})$ a determinar y que satisface

$$K(\mathbf{x}, \mathbf{y}) = \alpha(\mathbf{x}, \mathbf{y}) q(\mathbf{x}, \mathbf{y}). \quad (3.43)$$

Suponiendo que para un $\mathbf{y} \in \mathbb{R}^n$, en lugar de la ecuación de balance detallado se tiene

$$\pi(\mathbf{y}) q(\mathbf{y}, \mathbf{x}) < \pi(\mathbf{x}) q(\mathbf{x}, \mathbf{y}),$$

el objetivo es entonces escoger α tal que

$$\pi(\mathbf{y}) \alpha(\mathbf{y}, \mathbf{x}) q(\mathbf{y}, \mathbf{x}) = \pi(\mathbf{x}) \alpha(\mathbf{x}, \mathbf{y}) q(\mathbf{x}, \mathbf{y}),$$

lo cual se logra si

$$\alpha(\mathbf{y}, \mathbf{x}) = 1, \quad \alpha(\mathbf{x}, \mathbf{y}) = \frac{\pi(\mathbf{y}) q(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x}) q(\mathbf{x}, \mathbf{y})} < 1.$$

Intercambiando \mathbf{x} con \mathbf{y} vemos que el kernel definido por (3.43) satisface la ecuación de balance detallado (3.42) si se define

$$\alpha(\mathbf{x}, \mathbf{y}) = \min \left\{ 1, \frac{\pi(\mathbf{y}) q(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x}) q(\mathbf{x}, \mathbf{y})} \right\}.$$

Este kernel de transición se denomina *kernel Metropolis-Hastings*. El kernel de transición determina la probabilidad de que la cadena se mueva desde un estado \mathbf{x} al estado \mathbf{y} , también el kernel representa la probabilidad de rechazar el salto hacia el estado \mathbf{y} , además que el kernel está diseñado para que satisfaga la condición de balance detallado.

En términos generales el método parte de un estado inicial \mathbf{x}_0 y a continuación se proponen nuevos estados a partir del kernel de transición, aceptándolos con cierta probabilidad. El método se diseña para que la cadena de Markov $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T$ tenga como densidad estacionaria la distribución posterior (o la distribución que se desee), es decir, para que el histograma de dicha cadena tenga distribución similar a la de la distribución posterior (o la distribución deseada).

El algoritmo M-H usualmente se lleva a cabo en la práctica realizando los siguientes pasos:

1. Elegir un valor inicial $\mathbf{x}_1 \in \mathbb{R}^n$ e iniciar el contador a $k = 1$.
2. Extraer un $\mathbf{y} \in \mathbb{R}^n$ de la distribución propuesta $q(\mathbf{x}_k, \mathbf{y})$ y calcular la razón de aceptación en el kernel de transición

$$\alpha(\mathbf{x}_k, \mathbf{y}) = \min \left\{ 1, \frac{\pi(\mathbf{y}) q(\mathbf{y}, \mathbf{x}_k)}{\pi(\mathbf{x}_k) q(\mathbf{x}_k, \mathbf{y})} \right\}. \quad (3.44)$$

3. Extraer un valor aleatorio t de la distribución uniforme en $[0, 1]$.
4. Si $\alpha(\mathbf{x}_k, \mathbf{y}) \geq t$ establecer $\mathbf{x}_{k+1} = \mathbf{y}$, sino hacer $\mathbf{x}_{k+1} = \mathbf{x}_k$ e incrementar $k = k + 1$. Regresar al paso 2 mientras k nos sobrepase al valor límite T .

3.7.3 Algoritmo t-walk

El *algoritmo t-walk* es un algoritmo de muestreo MCMC basado en M-H [6]. Actualmente hay diversas variantes de los algoritmos de muestreo basados en M-H; entre las más conocidas se encuentran los métodos adaptivos, que como su nombre lo indican, intentan aprender o adaptarse a una escala en función de la estructura de correlación de las distribuciones generadas en la misma cadena de Markov. Estos métodos suelen ser particulares y su construcción depende de la escala y de la estructura de correlación del problema. Por el contrario, el método t-walk es un algoritmo de muestreo MCMC que no considera adaptabilidad o afinación de parámetros pero que puede muestrear distribuciones objetivo con escala arbitraria y estructura de correlación, procurando aprender la escala y correlación en el proceso de muestreo, pues t-walk se construye como algoritmo M-H sobre el espacio producto. El algoritmo t-walk puede muestrear densidades continuas con escala y estructura de correlación desconocida. Una de sus ventajas, gracias a que no es adaptivo, es la habilidad para muestrear eficientemente distribuciones con estructura de correlación local que difieren en diferentes partes del espacio de estado. Otra ventaja importante es que el algoritmo supera problemas de escala y correlación, lo que da lugar a que sea un algoritmo genérico para densidades continuas.

Gracias al creciente interés en el uso de los métodos bayesianos en aplicaciones científicas y de ingeniería, se requiere de métodos de muestreo eficientes y fáciles de utilizar como el método t-walk, ya que los investigadores interesados a utilizar la inferencia bayesiana generalmente no son probabilistas o estadísticos. El método t-walk puede utilizarse como una herramienta *de caja negra*, pues las entradas requeridas son el registro de la distribución objetivo y dos puntos iniciales en el espacio de parámetros, evitando ciertas dificultades como calcular las distribuciones condicionales. Esto ayuda a los investigadores a centrarse en el análisis de datos de los respectivos problemas, en lugar de ciertos detalles de los algoritmos MCMC, con la certeza de que para problemas de dimensión hasta 200 el algoritmo t-walk tiene un comportamiento bastante bueno. El algoritmo se ha utilizado en distintas ramas, como la datación por radio-carbono con dimensión $n \in [3, 6]$, un modelo bacteriano de transferencia horizontal de genes con $n \geq 50$, la tomografía de capacitancia eléctrica usando representaciones poligonales con $n \in [32, 128]$, e imágenes en impedancia basadas en píxeles con $n = 576$, entre otras aplicaciones.

A continuación se presenta, sin ser exhaustivos, aspectos para el diseño del algoritmo t-walk. Para profundizar en el tema se puede consultar [6] y [2].

Diseño

Para una función objetivo $\pi(\mathbf{x})$ con $\mathbf{x} \in \mathcal{X}$ y $\mathcal{X} \subset \mathbb{R}^n$ con dimensión n , se crea una nueva función objetivo como $f(\mathbf{x}, \mathbf{x}') = \pi(\mathbf{x})\pi(\mathbf{x}')$ en el correspondiente espacio producto $\mathcal{X} \times \mathcal{X}$, y la densidad condicionada de probabilidad de paso de la manera general $q\{(\mathbf{y}, \mathbf{y}') | (\mathbf{x}, \mathbf{x}')\}$

acompañada de dos propuestas restringidas:

$$\text{con } (\mathbf{y}, \mathbf{y}') = \begin{cases} (\mathbf{x}, h(\mathbf{x}', \mathbf{x})), & \text{con probabilidad } 0.5 \\ (h(\mathbf{x}, \mathbf{x}'), \mathbf{x}'), & \text{con probabilidad } 0.5 \end{cases} \quad (3.45)$$

donde $h(\mathbf{x}, \mathbf{x}')$ es una variable aleatoria utilizada para proponer nuevos pasos. Es decir, se cambia solo una de x o x' en cada paso. Es importante notar que no se consideran dos cadenas paralelas independientes en cada \mathcal{X} sino más bien el proceso completo reside totalmente en $\mathcal{X} \times \mathcal{X}$.

Siguiendo el esquema Metropolis-Hasting, se necesita calcular la razón de aceptación. Si se denota la función de densidad de la variable aleatoria $h(\mathbf{x}, \mathbf{x}')$ por $g(\cdot | \mathbf{x}, \mathbf{x}')$, la razón de aceptación que se obtiene es:

$$\frac{\pi(\mathbf{y}') g(\mathbf{x}' | \mathbf{y}', \mathbf{x})}{\pi(\mathbf{x}') g(\mathbf{y}' | \mathbf{x}', \mathbf{x})} \quad \text{y} \quad \frac{\pi(\mathbf{y}) g(\mathbf{x} | \mathbf{y}, \mathbf{x}')}{\pi(\mathbf{x}) g(\mathbf{y} | \mathbf{x}, \mathbf{x}')}, \quad (3.46)$$

en el primer y segundo caso en las ecuaciones (3.45), respectivamente. Los autores demuestran que si la variable aleatoria h es invariante bajo transformaciones afines, entonces también lo son las propuestas (3.45) y el muestreador MCMC que se obtiene. Se demuestra que cuatro elecciones de h : 1) *de caminata*, 2) *movimiento transversal*, 3) *movimientos de salto* y 4) *de golpe*, dan una mezcla adecuada en una amplia gama de distribuciones objetivo de dimensión moderada.

El *movimiento de caminata* se refiere a caminatas aleatorias escaladas, las cuales permiten alcanzar una mezcla adecuada de la cadena en muchas aplicaciones con correlación débil. Para densidades con una fuerte correlación entre algunas o varias de las variables, especialmente con distribuciones donde la estructura de correlación cambia a través del espacio de estados, la eficiencia del muestreador aumenta en gran medida con el *movimiento transversal*. Sin embargo, los movimientos de caminata y transversales, a su vez, no son suficientes para garantizar la irreducibilidad de la cadena sobre distribuciones de blancos arbitrarios. Para superar esta dificultad, y la de la posible mezcla extremadamente lenta para distribuciones con correlación muy alta, se introducen los *movimientos de salto y golpe* para evitar que \mathbf{x} y \mathbf{x}' colapsen (*i.e.* $\mathbf{x} \approx \mathbf{x}'$); estos movimientos se eligen con probabilidad baja, por ejemplo en razón de 60:60:1:1 para caminata, transversal, salto y golpe (relación de probabilidades, 0.49:0.49:01:01, aproximadamente).

Con la configuración anterior, de acuerdo a los autores del algoritmo, se garantiza la convergencia y el kernel formado satisface balance detallado.

3.8 Aproximación de rango bajo

La discusión general sobre inversión estadística gira alrededor de la relación entre la distribución a priori, la función de verosimilitud y la distribución posterior, por medio del teorema de Bayes. El principal obstáculo, para implantar la inferencia bayesiana en aplicaciones de alta dimensión, es computacional. Si bien la formulación bayesiana es bastante

general, las características esenciales de un modelo estadístico particular pueden aportar una estructura adicional a la actualización bayesiana. Por ejemplo, la distribución a priori a menudo codifica algún tipo de suavidad o correlación entre los parámetros de inversión; las observaciones pueden ser finitas, locales, escasas y corrompidas por el ruido; las observaciones también pueden ser indirectas, relacionadas con los parámetros por la acción del operador asociado al problema directo que filtra cierta información. Una consecuencia clave de estas características es que los datos proporcionan información por medio de la priori sólo en un *subespacio de baja dimensión* del espacio de parámetros. Con el fin de diseñar aproximaciones de la distribución posterior y los estimadores de Bayes relacionados, identificar y explotar este subespacio de baja dimensión puede generar ahorros computacionales sustanciales.

Los teoremas de existencia y los que garantizan la convergencia son útiles para tener certeza que se obtendrán buenos resultados de estimación teniendo en cuenta que estos resultados son asintóticos. Sin embargo, al trabajar con aproximación e inferencia de problemas como el problema de Cauchy, no se tiene claro cuantas mediciones se requieren en la frontera exterior Γ_2 para tener buenas estimaciones del potencial interior, o bien también es posible que sólo se tenga un número fijo de mediciones. Por estos motivos es necesario hacer un análisis de la aproximación de los resultados con rangos bajos, y para este análisis se utilizarán los resultados del artículo de Spatini et. al. [26].

3.8.1 Aproximación óptima de la matriz de covarianza de la distribución posterior

Recordemos que para el problema de Cauchy se supone que las distribuciones de la prior, la verosimilitud (y el error), se suponen normales. Siguiendo el procedimiento de la sección 3.5, para un problema general lineal $A\mathbf{x} = \mathbf{y} + \mathbf{e}$ con distribuciones Gaussianas $X \sim N(0, \Gamma_{pr})$, $Y|X \sim N(A\mathbf{x}, \Gamma_{obs})$ (es decir, $E \sim N(0, \Gamma_{obs})$), se tiene:

$$\begin{aligned} \pi_{pos}(\mathbf{x}|\mathbf{y}) &\propto \pi_{pr}(\mathbf{x}) \pi(\mathbf{y}_{obs} | \mathbf{x}) \\ &\propto \exp\left(-\frac{1}{2}\mathbf{x}^T \Gamma_{pr}^{-1} \mathbf{x}\right) \exp\left(-\frac{1}{2}(\mathbf{y} - A\mathbf{x})^T \Gamma_{obs}^{-1} (\mathbf{y} - A\mathbf{x})\right) \\ &= \exp\left(-\frac{1}{2}\mathbf{x}^T \Gamma_{pr}^{-1} \mathbf{x} - \frac{1}{2}\mathbf{x}^T A^T \Gamma_{obs}^{-1} A \mathbf{x} + \mathbf{x}^T A^T \Gamma_{obs}^{-1} \mathbf{y} - \frac{1}{2}\mathbf{y}^T \Gamma_{obs}^{-1} \mathbf{y}\right) \\ &= \exp\left(-\frac{1}{2}\mathbf{x}^T [\Gamma_{pr}^{-1} + A^T \Gamma_{obs}^{-1} A] \mathbf{x} + \mathbf{x}^T A^T \Gamma_{obs}^{-1} \mathbf{y}\right) \exp\left(-\frac{1}{2}\mathbf{y}^T \Gamma_{obs}^{-1} \mathbf{y}\right) \\ &\propto \exp\left(-\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^T \Gamma_{pos}^{-1} (\mathbf{x} - \bar{\mathbf{x}})\right), \end{aligned}$$

en donde

$$\Gamma_{pos} = [\Gamma_{pr}^{-1} + A^T \Gamma_{obs}^{-1} A]^{-1} \quad y \quad \bar{\mathbf{x}} = \mu_{pos}(\mathbf{y}) = \Gamma_{pos}^{-1} A^T \Gamma_{obs}^{-1} \mathbf{y}. \quad (3.47)$$

Por lo tanto,

$$X|Y \sim N(\mu_{pos}(\mathbf{y}), \Gamma_{pos}). \quad (3.48)$$

En Spatini, [26], se considera la Hessiana del negativo de la función log-verosimilitud: $\frac{1}{2}(\mathbf{y} - A\mathbf{x})^T \Gamma_{obs}^{-1}(\mathbf{y} - A\mathbf{x})$, la cual es claramente

$$H = A^T \Gamma_{obs}^{-1} A \quad (\text{Matriz de información de Fisher}). \quad (3.49)$$

Con (3.47) y (3.49) se obtiene la siguiente relación entre las covarianzas de la prior y la posterior:

$$\Gamma_{pos}^{-1} = \Gamma_{pr}^{-1} + H. \quad (3.50)$$

Para el caso del problema de Cauchy: (ver sección 3.5) propusimos

$$\Gamma_{pr}^{-1} = \frac{1}{\gamma^2} D^T D \quad \text{y} \quad \Gamma_{obs}^{-1} = \frac{1}{\sigma_{noise}^2} I,$$

por lo que

$$\Gamma_{pos}^{-1} = \frac{1}{\gamma^2} D^T D + H \quad \text{con} \quad H = \frac{1}{\sigma_{noise}^2} A^T A, \quad (3.51)$$

recuperando la expresión (3.23) en donde $\Sigma = \Gamma_{pos}$.

3.8.2 Clases de aproximación

En Spatini et. al. [26] se busca una aproximación, $\hat{\Gamma}_{pos}$, de rango bajo a Γ_{pos} que es óptima en una clase de matrices que se introducirá más abajo, en (3.53).

La idea proviene de la relación (3.50) en donde se observa que Γ_{pos}^{-1} es una actualización nonegativa de la matriz de precisión a priori Γ_{pr}^{-1} . Es decir,

$$\Gamma_{post}^{-1} = \Gamma_{pr}^{-1} + ZZ^T \quad \text{en donde} \quad ZZ^T = H.$$

Similarmente, utilizando la fórmula de inversión de Sherman-Morrison-Woodbury, Γ_{post} de puede expresar como una actualización no positiva de Γ_{pr} :

$$\Gamma_{post} = \Gamma_{pr} - KK^T, \quad (3.52)$$

donde

$$KK^T = \Gamma_{pr} A^T \Gamma_y^{-1} A \Gamma_{pr} \quad \text{y} \quad \Gamma_y = \Gamma_{obs} + A \Gamma_{pr} A^T,$$

siendo Γ_y la matriz de covarianza de la distribución marginal de Y . Esta actualización de Γ_{pr} es semidefinida negativa debido a que los datos agregan información: la varianza posterior en cualquier dirección siempre es más pequeña que la correspondiente varianza de la prior. Además, la actualización usualmente es de rango bajo ya que hay direcciones en el espacio de parámetros a lo largo de las cuales los datos no informan mucho, relativos a la prior. Por ejemplo, H puede tener un espectro que decae rápidamente. Notése sin embargo que la Γ_{post} misma podría no ser de rango bajo. La estructura rango-bajo, cuando está presente, radica

en la actualización de Γ_{pr} que lleva a Γ_{post} . Por lo tanto, una clase natural de matrices para aproximar Γ_{post} es el conjunto de actualizaciones semidefinidas negativas de Γ_{pr} , con un rango máximo fijo, y que resulta en matrices definidas positivas:

$$\mathcal{M}_r = \{ \Gamma \succ 0 \mid \Gamma = \Gamma_{pr} - KK^T : \text{rango}(K) \leq r \}. \quad (3.53)$$

Esta clase de aproximaciones de Γ_{post} toma ventaja de la estructura de la actualización prior-a-posterior.

3.8.3 Función de costos

Para hablar de una aproximación de una matriz en términos de optimización es necesario introducir un criterio de distancia entre matrices simétricas definidas positivas (SDP). Se utiliza la métrica introducida por Förstner y Moonen [11] en donde la distancia $d_{\mathcal{F}}(A, B)$ entre un par de matrices SDP, A y B , se define como:

$$d_{\mathcal{F}}^2(A, B) = \text{tr} \left[\ln^2 \left(A^{-1/2} B A^{-1/2} \right) \right] = \sum_i \ln^2(\sigma_i), \quad (3.54)$$

donde (σ_i) es la sucesión de valores propios generalizados del Pencil (A, B) (i.e. $(A - \sigma_i B)\mathbf{x} = \mathbf{0}$ o $\det(A - \sigma_i B) = 0$). Dos propiedades muy importantes de esta métrica son: 1) es invariante bajo la inversa de las matrices y bajo transformaciones de similitud de matrices:

$$d_{\mathcal{F}}(A^{-1}, B^{-1}) = d_{\mathcal{F}}(A, B) \quad \text{y} \quad d_{\mathcal{F}}(MAM^T, MBM^T) = d_{\mathcal{F}}(A, B),$$

y 2) trata de manera similar las sub- y sobreaproximaciones en el sentido:

$$d_{\mathcal{F}}(\Gamma_{pos}, \alpha \hat{\Gamma}_{pos}) \rightarrow \infty \quad \text{cuando} \quad \alpha \rightarrow 0 \quad \text{y} \quad \text{cuando} \quad \alpha \rightarrow \infty.$$

Ninguna de las propiedades anteriores ocurre con la norma de Frobenius (muy utilizada en procesos de optimización con matrices).

En [26] se muestra que la aproximación $\hat{\Gamma}_{pos}$ a Γ_{pos} es óptimal no sólo en la norma de Förstner sino también en la siguiente clase más general, \mathcal{L} , de funciones de pérdida para matrices SDP:

Definición 8 (funciones de pérdida) *La clase \mathcal{L} se define como la colección de funciones de la forma*

$$L(A, B) = \sum_{i=1}^n f(\sigma_i), \quad (3.55)$$

donde A, B son matrices SDP y (σ_i) son los valores propios generalizados del pencil (A, B) , y donde

$$f \in \mathcal{U} = \left\{ g \in \mathcal{C}^1(\mathbb{R}_+) : g'(x)(x-1) < 0 \text{ para } x \neq 1, \text{ y } \lim_{x \rightarrow \infty} g(x) = \infty \right\}.$$

Los elementos de \mathcal{U} son funciones diferenciables de valores reales, definidas sobre el eje positivo que decrecen para $x < 1$, crecen para $x > 1$, y tienden a infinito cuando $x \rightarrow \infty$. La métrica cuadrada de Förstner pertenece a la clase de funciones de pérdida definida por (3.55), mientras que la distancia inducida por la métrica de Frobenius no.

El siguiente lema justifica la importancia de la clase \mathcal{L} . En particular, muestra que la optimalidad de la aproximación $\hat{\Gamma}_{pos}$ con respecto a cualquier función de pérdida $L \in \mathcal{L}$ lleva a una aproximación óptima de la distribución posterior usando una Gaussiana (con la misma media) en términos de otros criterios familiares usados para comparar medidas de probabilidad, tales como la distancia de Hellinger y la divergencia K-L [70]¹. Más precisamente, se tiene el siguiente resultado:

Lema 1 (Equivalencia de aproximaciones) *Si $L \in \mathcal{L}$, entonces una matriz $\hat{\Gamma}_{pos} \in \mathcal{M}_r$ minimiza la distancia de Hellinger y la divergencia K-L entre $N(\mu_{post}, \Gamma_{post})$ y la aproximación $N(\mu_{pos}, \hat{\Gamma}_{post})$ si y solo si minimiza $L(\Gamma_{post}, \hat{\Gamma}_{post})$*

Nota. Se observa que ni la distancia de Hellinger ni la divergencia K-L entre las distribuciones $N(\mu_{post}, \Gamma_{post})$ y $N(\mu_{pos}, \hat{\Gamma}_{post})$ dependen de los datos \mathbf{y} . La optimalidad en la distribución no necesariamente se cumple cuando las medias posteriores son diferentes.

3.8.4 Resultados de optimalidad

El principal resultado en [26] es:

Teorema 4 (Aproximación óptima de la matriz de covarianza) *sea $(\delta_i^2, \hat{\omega}_i)$ el par de valores y vectores propios generalizados a partir del pencil*

$$(H, \Gamma_{pr}^{-1}), \quad (3.56)$$

con los valores propios ordenados en forma decreciente y $H = A^T \Gamma_{obs}^{-1} A$. Sea $L \in \mathcal{L}$ definida en (3.55). Entonces, se cumple lo siguiente:

- Un minimizador $\hat{\Gamma}_{pos}$, de la pérdida L , entre Γ_{pos} y un elemento de \mathcal{M}_r está dado por:

$$\hat{\Gamma}_{pos} = \Gamma_{pr} - K K^T, \quad K K^T = \sum_{i=1}^r \frac{\delta_i^2}{1 + \delta^2} \hat{\omega}_i \hat{\omega}_i^T. \quad (3.57)$$

La correspondiente pérdida mínima está dada por

$$L(\hat{\Gamma}_{pos}, \Gamma_{pos}) = f(1)r + \sum_{i>r} f\left(\frac{1}{1 + \delta^2}\right). \quad (3.58)$$

¹La distancia de Hellinger se utiliza en estadística para medir las distancias entre dos distribuciones de probabilidad π_1, π_2 denotada como $H^2(\pi_1, \pi_2) = \int_{\mathbb{R}^n} (\sqrt{\pi_1(\mathbf{x})} - \sqrt{\pi_2(\mathbf{x})})^2 d\mathbf{x}$ y la divergencia Kullback-Leibler(K-L) es una medida no simétrica de la similitud de dos distribuciones de probabilidad denotada por un valor esperado de la siguiente manera $D_{KL}(\pi_1, \pi_2) = E_{x \sim \pi_1} [\log(\frac{\pi_2}{\pi_1})]$.

- *El minimizador $\hat{\Gamma}_{pos}$ es único si los primeros valores propios r del pencil (H, Γ_{pr}^{-1}) son diferentes.*

Este teorema proporciona una forma de calcular la mejor aproximación de Γ_{pos} por matrices en \mathcal{M}_r : sólo se deben calcular los pares de valores/vectores propios correspondientes a la sucesión decreciente de (δ_i^2) del pencil (H, Γ_{pr}^{-1}) hasta satisfacer un criterio de paro, que se puede basar en la pérdida mínima. Notamos que la fórmula anterior y el procedimiento recuerda el cálculo de la descomposición SVD de una matriz.

En resumen, con el resultado anterior se puede estimar la dimensión óptima o efectiva de la matriz de covarianza de la posterior, y en consecuencia nos puede ayuda a decidir el número de datos para obtener una estimación adecuada de los parámetros en el problema inverso.

Capítulo 4

Resultados numéricos

En este capítulo se presentan resultados numéricos sobre la solución del problema de Cauchy, utilizando los métodos de inversión introducidos en los capítulos 2 y 3. Recordemos que el problema inverso de Cauchy consiste en encontrar φ sobre la frontera interior Γ_1 de una región anular bidimensional Ω , en donde se satisface la ecuación de estado para u :

$$\begin{aligned} -\sigma\Delta u &= 0, \text{ en } \Omega, \\ u &= \varphi, \text{ sobre } \Gamma_1 \\ \sigma \frac{\partial u}{\partial n} &= 0, \text{ sobre } \Gamma_2. \end{aligned} \tag{4.1}$$

a partir de datos V_δ en la frontera exterior, i.e. $V_\delta \approx u|_{\Gamma_2}$, con distintos errores de medición δ . En todos los casos, se considera una conductividad constante σ .

A continuación se mostrarán resultados en tres regiones anulares Ω distintas:

1. Una región con frontera exterior circular Γ_2 con radio 1.2 y frontera interior circular Γ_1 de radio 1. A esta región se le llamará región circular simple.
2. Una segunda región con frontera exterior circular Γ_2 de radio 1.2 de nuevo, pero con una frontera interior Γ_1 elíptica con centro en la posición $(-0.4, 0.3)$, semiejes de longitud 0.7 y 0.3 y rotada un ángulo $\pi/4$. A esta región se le llamará región circular con frontera interior elíptica.
3. La tercer región es una región compleja, bidimensional que asemeja el corte transversal de una cabeza humana. Se le llamará simplemente región compleja.

Cada malla en las distintas regiones de cada problema fueron creadas mediante el programa gmsh el cual está disponible para el ambiente linux.

De acuerdo a lo discutido sobre los distintos de modelos a priori que podemos utilizar, en este caso se utilizará la matriz de la forma (3.32) ya que es la que proporciona soluciones

más suavizantes con lo cual se pretende obtener resultados más exactos.

Para resolver estos problemas se considerarán *datos sintéticos* V en la región Γ_2 , los cuales se obtienen resolviendo numéricamente el modelo (4.1) con el software FeniCS sobre cada una de las regiones anulares descritas anteriormente, y utilizando la *solución exacta* dada por $u(x, y) = e^x \sin(y)$ para generar las condiciones de frontera φ sobre Γ_1 . A los *datos sintéticos* V se les agrega *ruido blanco* (Gaussiano); así, para la región circular y la región con frontera elíptica se consideró un nivel de ruido sin embargo, para la región compleja se considera dos diferentes niveles o desviaciones: $\sigma_1 = \frac{\max(V_i)}{100}$, $\sigma_2 = \frac{\max(V_i)}{10}$, es decir los datos con ruido son de la forma $V_r = V + Norm(0, \sigma_r)$. Estas desviaciones representan el 1% y 10% del valor máximo que alcanza la solución en la frontera exterior. Cabe destacar que el caso del 1% describe la relación señal-a-ruido utilizada en análisis de señales. Los otros casos, se incluyen con el objeto de analizar la sensibilidad de los modelos y algoritmos con respecto al aumento del ruido en los datos.

Todos los resultados numéricos fueron obtenidos en una computadora Laptop personal de la marca HP con sistema operativo Ubuntu 16.04 con 4GB de ram y procesador AMD A8-7410.

4.1 Ejemplo 1. Región anular circular simple

Se considera una región anular circular centrada en el origen con radios interno y externo 1 y 1.2, respectivamente. Esta región anular es uno de los ejemplos considerados en el artículo [10] y la intención es comparar nuestros resultados numéricos con los obtenidos en ese trabajo. Para este caso solo mostramos resultados con ruido del 5%.

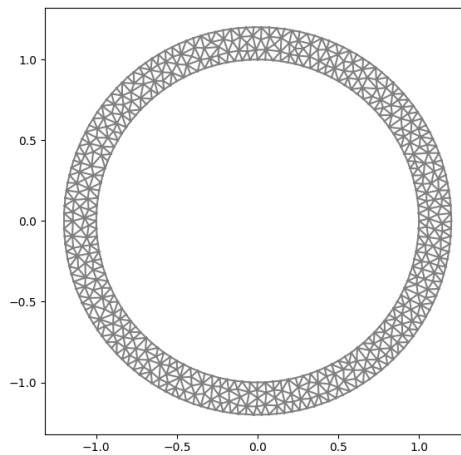


Figura 4.1: Región anular y malla para la región anular circular.

La figura 4.1 muestra la región y la malla utilizada, la cual consta de 452 elementos

triangulares y 224 nodos. Del total de nodos, 65 se encuentran sobre la frontera interior Γ_1 y 77 sobre la frontera exterior Γ_2 .

En la estimación numérica de φ , utilizando el método de inversión Bayesiana, se realizaron 820,000 iteraciones en el algoritmo de muestreo MCMC utilizando el algoritmo t-walk. Los primeros 20,000 estados de la cadena de Markov se toman como ‘*periodo de calentamiento*’ (en inglés se le denomina *burn-in*) antes de entrar a la región de alta probabilidad, los cuales se desechan del muestreo para estimar la solución. Cabe aclarar que no se utiliza un procedimiento general para decidir el tamaño del mencionado periodo de calentamiento, solo se elige de manera que la gráfica de la energía¹ no presente comportamientos de monotonía, es decir, que la energía siempre crezca o siempre decrezca, como se muestra en la figura 4.2.

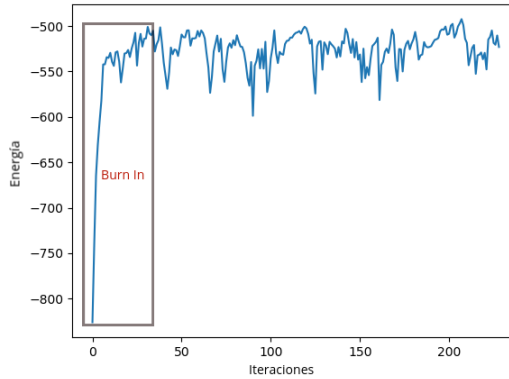


Figura 4.2: Elección del periodo de calentamiento.

Además de establecer el periodo de calentamiento es conveniente considerar un tamaño de paso entre iteraciones para garantizar que la muestra generada sea independiente. Una de las ventajas del algoritmo t-walk es que estima el tiempo de autocorrelación integrado (IAT por sus siglas en inglés). El IAT es un diagnóstico de MCMC que estima el número promedio de iteraciones para obtener una muestra independiente, dada una cadena de Markov. Dicho de otra manera, IAT es el número de muestras correlacionadas con el mismo poder de reducción de varianza que una muestra independiente. Un valor más bajo de IAT es mejor, pues indica que se requieren menos iteraciones para generar una muestra independiente. Para este ejemplo el modelo t-walk arrojó un IAT de 997, para que al final terminaríamos con una muestra de tamaño 800.

En la figura 4.3 se muestra la energía de los elementos generados por el método de t-walk. Más precisamente, se muestran los elementos, quitándole el *burn-in*, y considerando el IAT como tamaño de paso. Como se mencionó anteriormente, esta gráfica nos ayuda a verificar que la cadena haya llegado a un estado de equilibrio, es decir, que no haya

¹A el valor de la distribución posterior evaluada en los parámetros propuestos por cada iteración x se le conoce como energía

tendencias crecientes o decrecientes para poder garantizar que se puede obtener realmente una muestra de la distribución posterior adecuada. En la gráfica claramente se nota que hay tendencias crecientes o decrecientes.

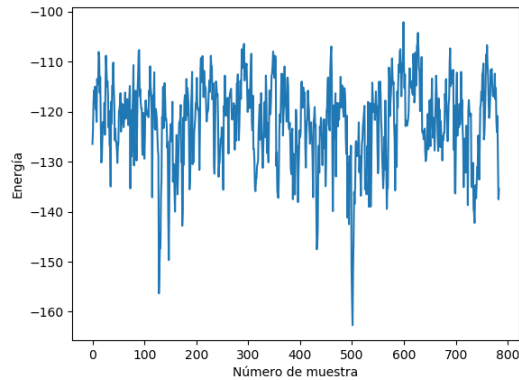


Figura 4.3: Gráfica energía del problema en el ejemplo 1

Los datos del ejemplo 1 se muestran en la figura 4.4. La línea roja representa el potencial exacto $V = \varphi|_{\Gamma_2}$, los puntos indican los datos con ruido V_r que se utilizan para resolver el problema inverso.

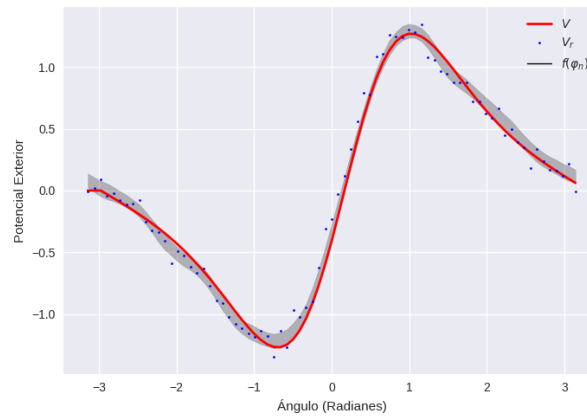


Figura 4.4: Potencial en la frontera externa Γ_2 con respecto al ángulo para la región anular circular. En rojo se muestra el valor exacto, los puntos azules indican las ‘mediciones con ruido’ y en gris la imagen de las muestras generadas.

Los resultados numéricos obtenidos al resolver el problema inverso se muestran en la figura 4.5. Junto a la solución exacta φ (en línea roja) se muestra la solución obtenida con el

4.2. EJEMPLO 2. REGIÓN ANULAR CON FRONTERA EXTERIOR CIRCULAR Y FRONTERA INTERNA ELÍPTICA

método determinista $\hat{\varphi}$ en línea verde. Las estimaciones obtenidas con inversión estadística son: la media condicional (CM) $\tilde{\varphi}$ en línea azul, el máximo a posteriori (MAP) φ_{MAP} en línea negra y los potenciales generados por la muestra con la distribución posterior φ_n en gris.

Claramente los resultados, tanto los obtenidos con el método determinista como con el bayesiano, son muy satisfactorios en este caso. Por supuesto, la franja gris indica que hay una amplificación del error de las muestra en los resultados, pero dicho error de salida se mantiene acotado, gracias a los términos regularizantes en los modelos computacionales.

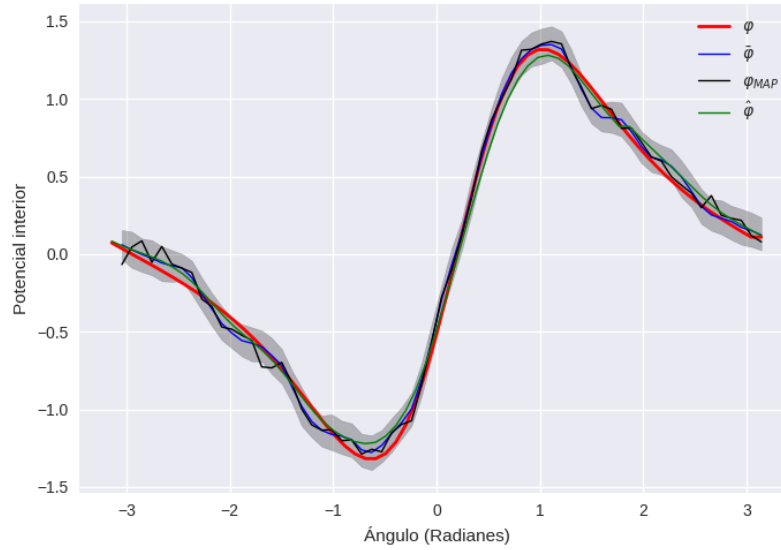


Figura 4.5: Potencial en la frontera interna Γ_1 con respecto al ángulo para la región anular circular. En rojo se muestra el valor exacto, en verde el calculado con el método determinista. Los resultados con el método bayesiano se muestran en azul para el CM, en negro para el MAP y en gris los obtenidos con distribución a posteriori.

4.2 Ejemplo 2. Región anular con frontera exterior circular y frontera interior elíptica.

Como ya se mencionó al inicio de este capítulo, esta región anular tiene una frontera exterior circular Γ_2 y una frontera interior Γ_1 elíptica, con las dimensiones mencionadas anteriormente. La región junto con la malla utilizada para los cálculos numéricos se muestra en la Figura 4.6. La malla contiene 939 elementos triangulares y 1878 nodos, de los cuales 45 son nodos sobre la frontera interior Γ_1 y 97 son nodos en la frontera exterior Γ_2 .

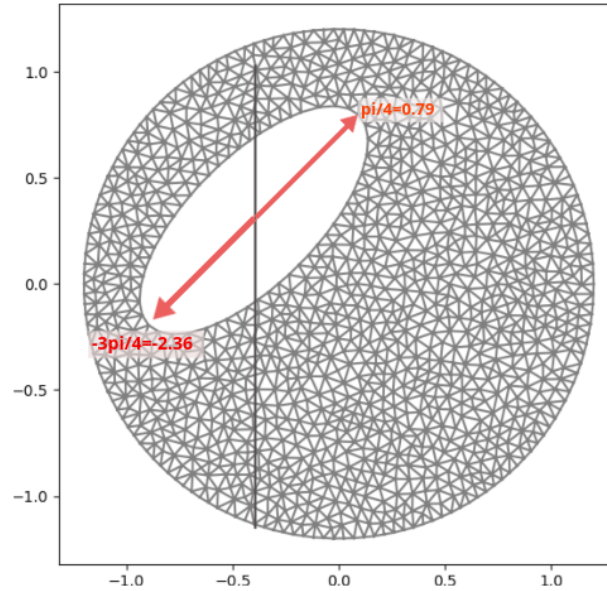


Figura 4.6: Región anular y malla para la región anular con frontera exterior circular y frontera interior elíptica.

Queremos ver el comportamiento de los algoritmos en una región en donde se pierde simétrica:

Al igual que en el ejemplo anterior, se realizaron 820,000 iteraciones en el algoritmo *t-wlak*, considerando 10,000 de periodo de *'calentamiento'*. La gráfica de energía se muestra en la figura 4.7, después de eliminar el periodo de calentamiento. En este caso el modelo arroja un IAT con valor de 922, es decir, que se tiene una muestra de tamaño 890.

En la figura (4.8) se muestran los datos utilizados para este ejemplo. En rojo se muestra el potencial exacto V en la frontera exterior Γ_2 , en puntos azules los datos perturbados V_r con 3% de ruido agregado los cuales se utilizan para resolver el problema inverso.

Los resultados numéricos para el potencial en la frontera interior (elíptica) se muestran en la figura 4.9. De nuevo, en rojo se muestra el potencial exacto φ , en verde el resultado $\hat{\varphi}$ obtenido con el método determinista variacional. los resultados por medio de inversión bayesiana incluyen: en azul la media condicional $\bar{\varphi}$, en negro el MAP φ_{MAP} , y sobre la franja gris muestras obtenidas con la distribución a posteriori.

La discrepancia entre el potencial exacto y el numérico sobre la frontera interior Γ_1 es relativamente mayor que en ejemplo anterior. En este caso, la forma elíptica de la frontera interior y su ubicación asimétrica influyen en el resultado. Creemos que la forma de la frontera interior influye más, debido a que la mayor discrepancia entre la solución exacta y numérica ocurre alrededor de los extremos de la elipse localizados sobre los semiejes mayores. Menor variabilidad ocurre en la parte más suave de la frontera elíptica. Los resultados en

4.2. EJEMPLO 2. REGIÓN ANULAR CON FRONTERA EXTERIOR CIRCULAR Y FRONTERA INTERNA

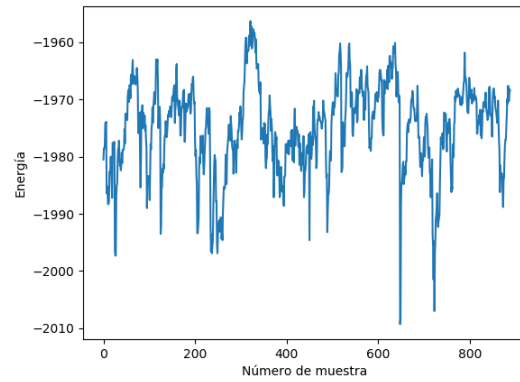


Figura 4.7: Gráfica de energía obtenida para el problema del ejemplo 2

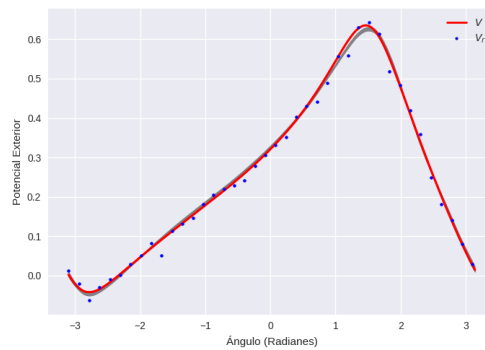


Figura 4.8: Potencial en la frontera externa Γ_2 con respecto al ángulo para la región del ejemplo 2. En rojo se muestra el valor exacto, los puntos azules indican las 'mediciones con ruido' y en gris la franja con resultados de las muestras.

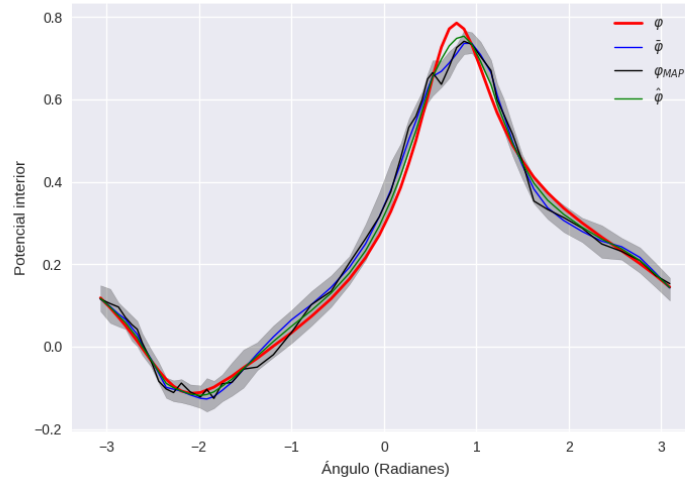


Figura 4.9: Potencial en la frontera interna Γ_1 con respecto al ángulo para la región del ejemplo 2. En rojo se muestra el valor exacto, en verde el calculado con el método determinista. Los resultados con el método bayesiano se muestran en azul para el CM, en negro para el MAP y en gris los obtenidos con distribución a posteriori.

su conjunto siguen siendo muy satisfactorios para esta región, un tanto más compleja que en el ejemplo anterior, por la asimetría de la región anular.

4.3 Ejemplo 3. Región anular compleja

En este caso se considera una región compleja que asemeja un corte lateral de la cabeza humana. La región y su discretización se muestran en la figura 4.10. La malla en esta ocasión consiste de 2908 elementos triangulares y 1454 nodos, con 98 nodos sobre la frontera interior Γ_1 y 156 nodos sobre la frontera exterior Γ_2 .

En este caso decidimos hacer un estudio un poco más amplio y presentamos resultados cuando los datos de entrada se les agrega ruido al 1% y 10%. Es importante señalar que la dimensión del problema es mayor en este caso, ya que se deben estimar 98 parámetros, aumentando considerablemente el tiempo de cómputo (más de un día completo en la computadora personal mencionada arriba), sin tener certeza de que la memoria disponible sea suficiente para poder generar una muestra significativa. El IAT que arroja el algoritmo es de 1,400, lo cual implica que necesitaríamos al menos 400 mil iteraciones para obtener una muestra adecuada. Tomando en cuenta la experiencia de los dos ejemplos anteriores, en donde se realizaron alrededor de 200 mil iteraciones para obtener una muestra adecuada, se decidió realizar interpolación a 50 datos en la frontera exterior Γ_2 y estimar igualmente 50 parámetros en la frontera interior. De esta manera fue posible bajar el tiempo de cómputo

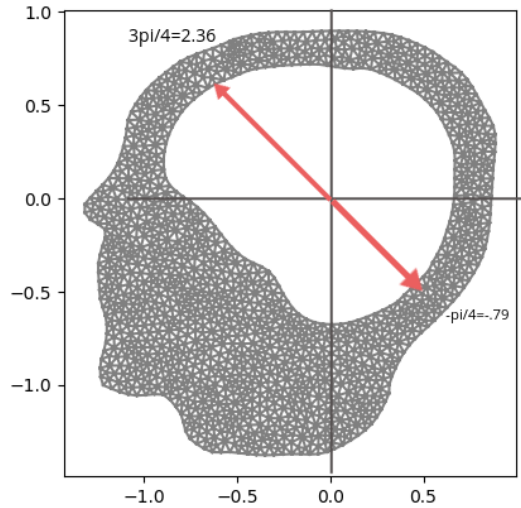


Figura 4.10: Región anular y malla para la región anular compleja.

a un tercio del total original.

Consideramos primero el caso con 1% de perturbación en los datos. La figura 4.11 muestra la gráfica con los datos exactos (en línea roja), los datos perturbados con 1% de ruido blanco mostrada en puntos negros y los datos del mapeo. La figura muestra que visualmente este nivel de perturbación es pequeño.

Los resultados mostrados en la figura 4.12 muestran una amplificación del error en la identificación del potencial en la frontera interior Γ_1 . Como antes, en línea roja se muestra el potencial exacto φ , en línea verde se muestra el estimado $\hat{\varphi}$ con en método determinista variacional. Los resultados con los métodos bayesianos incluyen: la media condicional $\bar{\varphi}$ en línea azul, el máximo a posteriori φ_{MAP} en línea negra y las funciones generadas φ_n por la muestra aleatoria en gris.

Se observa una amplificación del error en los resultados, pues la franja gris es más ancha en la figura de resultados que la mostrada en la figura de datos. Hay que recordar que en este caso además de la perturbación en los datos, se debe considerar la agregada por el proceso de interpolación en los datos y resultados. Además, en la gráfica es claro que la mayor amplificación del ruido ocurre en la región en donde se encuentra la mayor separación entre las fronteras exterior e interior (entre $-\pi$ y $-\pi/3$ radianes), además de ser la parte en donde tanto Γ_1 como Γ_2 son menos suaves. En este caso el método determinista proporciona una mejor estimación del potencial en la frontera interior debido a que no es necesario realizar el proceso de interpolación.

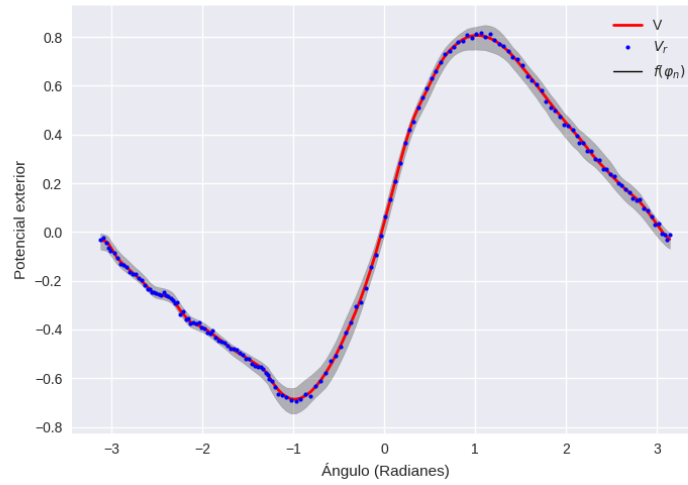


Figura 4.11: Datos en la frontera exterior Γ_2 de la región compleja con error de 1%

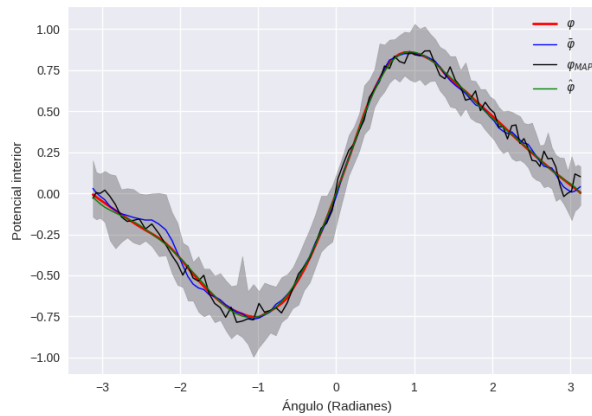


Figura 4.12: Estimaciones obtenidas para el potencial en la frontera interior Γ_1 de la región compleja, correspondiente a un error de 1% en las mediciones.

Por último, se muestra el ejemplo con un nivel de 10% de ruido agregado a los datos. La figura 4.13 muestra la gráfica con los datos. La nomenclatura es la misma que en los casos anteriores. Los correspondientes resultados numéricos obtenidos para el potencial en la frontera interior se muestran en la figura 4.14. Estos resultados son cualitativamente similares a los dos casos anteriores, con la diferencia de que ahora se muestra mayor error, principalmente en la regiones de $(-3, -1)$ radianes y $(1, 2)$ radianes, que corresponden a la parte de la región en donde las frontera son menos suaves.

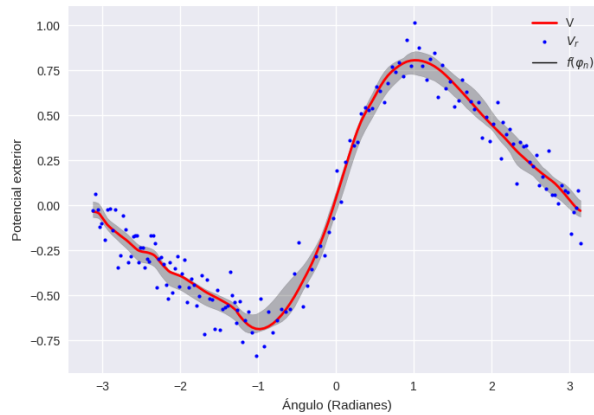


Figura 4.13: Datos en la frontera exterior Γ_2 de la región compleja con error de 10%

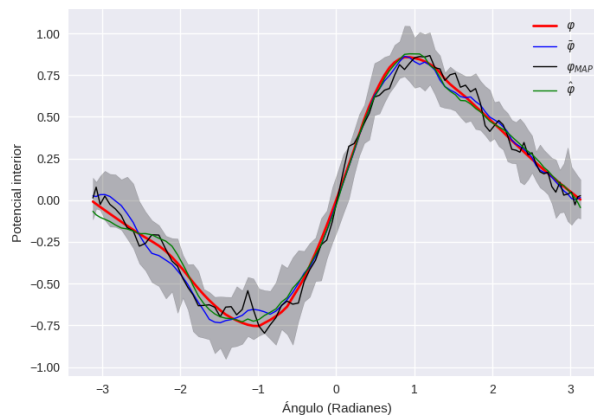


Figura 4.14: Estimaciones obtenidas para el potencial en la frontera interior Γ_1 de la región compleja, correspondiente a un error de 10%.

Hasta ahora, se han mostrado los resultados en forma gráfica solamente. A continuación se realiza un resumen de los errores en forma cuantitativa. Utilizaremos la norma $L^2(\Gamma_1) = \left(\int_{\Gamma_1} |\varphi - \varphi^*|^2\right)^{1/2}$ sobre la frontera interior Γ_1 y la análoga sobre la frontera exterior. La cuantificación del error se realiza para los dos estimadores puntuales obtenidos con inversión estadística, $\bar{\varphi}$, φ_{MAP} , y también para la solución numérica obtenida con el método variacional determinista $\hat{\varphi}$. Como solo se dispone de valores numéricos sobre los nodos de la malla, para calcular las integrales asociadas a las normas se utiliza la regla del trapecio. Es suficiente esta regla para estimar las normas adecuadamente, ya que tienen el mismo orden de convergencia que el método de los elementos finitos para resolver numéricamente los problemas elípticos. Para precisar, utilizaremos la siguiente notación para los errores relativos:

$$ER(V_r, V)_{\Gamma_2} = \frac{\|V_r - V\|_{L^2(\Gamma_2)}}{\|V\|_{L^2(\Gamma_2)}},$$

$$ER(\hat{\varphi}, \varphi)_{\Gamma_1} = \frac{\|\hat{\varphi} - \varphi\|_{L^2(\Gamma_1)}}{\|\varphi\|_{L^2(\Gamma_1)}}, \quad ER(f(\hat{\varphi}), V)_{\Gamma_2} = \frac{\|f(\hat{\varphi}) - V\|_{L^2(\Gamma_2)}}{\|V\|_{L^2(\Gamma_2)}}.$$

En la última expresión, $f(\hat{\varphi})$ denota al potencial numérico en la frontera exterior Γ_2 que se obtiene al resolver el problema inverso con el método determinista variacional. De manera análoga se definen errores relativos asociados a $\bar{\varphi}$ y φ_{MAP} sobre Γ_1 y sus correspondientes potenciales numéricos $f(\bar{\varphi})$ y $f(\varphi_{MAP})$ sobre la frontera exterior Γ_2 .

El resumen de los errores relativos obtenidos por solución numérica se muestra en la tabla 4.1. Esta tabla nos permite analizar como se propaga el ruido en la mediciones sobre los resultados numéricos.

Nivel de ruido r	1 %	10 %
$ER(V_r, V)_{\Gamma_2}$	0.0170	0.1815
$ER(\hat{\varphi}, \varphi)_{\Gamma_1}$	0.0164	0.0614
$ER(\bar{\varphi}, \varphi)_{\Gamma_1}$	0.0519	0.0912
$ER(\varphi_{MAP}, \varphi)_{\Gamma_1}$	0.0806	0.1282
$ER(f(\hat{\varphi}), V)_{\Gamma_2}$	0.0073	0.0323
$ER(f(\bar{\varphi}), V)_{\Gamma_2}$	0.0135	0.0364
$ER(f(\varphi_{MAP}), V)_{\Gamma_2}$	0.0168	0.0506

Cuadro 4.1: Errores de los datos y su propagación en los resultados numéricos sobre la frontera interior Γ_1 y la frontera exterior Γ_2 de la región compleja.

Antes de realizar las comparaciones es necesario aclarar que para cada nivel de ruido r , la desviación relativa de los datos $ER(V_r, V)_{\Gamma_2}$ en el segundo renglón nos indican el verdadero error en las mediciones. Se observa que para los niveles de ruido 1 % y 10 % las

desviaciones reales de los datos son de 1.8 % y 18 % respectivamente.

Se observa que el error en los datos influye con mayor fuerza en el cálculo de φ sobre la frontera interior que sobre la frontera exterior, en todos los casos. Es decir, en todos los casos $ER(\varphi_{approx}, \varphi)_{\Gamma_1}$ es menor que el correspondiente $ER(f(\varphi_{approx}), V)_{\Gamma_2}$. De hecho, este fenómeno no es casual y es muy común en la solución de problemas inversos, en particular es consistente con los resultados obtenidos en [10]. Las gráficas mostradas anteriormente también muestran este comportamiento. Además, en ningún caso se amplifica el error de los datos sobre el resultado numérico sobre Γ_2 , ya que todos los errores relativos mostrados en los últimos dos renglones son menores a 1.8 % y 18 % en cada columna. Las gráficas también ilustran este comportamiento, ya que se ve más dispersión en los puntos que indican mediciones que en las franjas en gris.

Los resultados sobre la frontera interior Γ_1 son más variados. En algunos casos el error es menor al de las mediciones, en otros es mayor. Los casos más significativos, por tener el mayor nivel de error en los datos, son los mostrados en la segunda y tercera columnas. En esos dos casos los errores relativos en los datos son del 10 % y 18 %, mientras que los errores relativos $ER(\hat{\varphi}, \varphi)_{\Gamma_1}$ $ER(\bar{\varphi}, \varphi)_{\Gamma_1}$ son de menor magnitud. Por el contrario, el error relativo con el MAP es mayor, con un exceso del 6.6 % en el primer caso y 4.4 % en el segundo. La conclusión general es que tanto el método determinista como el bayesiano estiman el potencial de manera muy satisfactoria, obtenido la mayor discrepancia con el MAP. En particular, las estimaciones obtenidas con el método bayesiano (basado en el t-walk) muestra que la media $\bar{\varphi}$ es un mejor estimador puntual que el MAP en nuestro problema. Por supuesto una análisis más profundo de los resultados numéricos implicaría generar muestras de mayor tamaño, lo cual implica utilizar mayor poder de cómputo.

4.4 Aproximación de rango bajo

Ahora bien, aunque ya se ha visto que se obtienen buenas estimaciones numéricas, es de particular interés estimar la dimensión óptima de la matriz de covarianza de la posterior en los ejemplos resueltos anteriormente, dado que hemos considerado diferentes regiones con sus respectivas mallas. Como en cada simulación se generó una muestra entonces con dicha muestra es posible calcular la matriz de varianzas y covarianzas o al menos una aproximación de ellas.

Entonces, dada la aproximación a Γ_{post}^{-1} , basta encontrar la matriz H para utilizar el resultado del Teorema 4. Como la distribución a priori Γ_{pr}^{-1} es conocida y dada por (3.32), sabemos que $H = \Gamma_{post}^{-1} - \Gamma_{pr}^{-1}$ por (3.51). Sabemos que Γ_{pr}^{-1} no es definida positiva por tener un valor propio nulo, y se considera su descomposición en valores singulares para poder calcular los valores y vectores propios generalizados del pencil (3.56) con el objeto de poder calcular la función de pérdida (3.58) y gráficarla para cada rango r . De esta manera podemos estimar la dimensión óptima para la matriz de covarianza.

En la figura 4.15, la gráfica del lado izquierdo muestra el valor de la función de costo contra el valor de la dimensión r , es decir se gráfica la función de pérdida (3.58) con respecto a r en el caso con dominio anular circular. En dicha gráfica de puede visualizar un intervalo en donde ocurre el mínimo costo. La gráfica de la derecha es una ampliación (zoom) de la izquierda con el objeto de visualizar mejor el intervalo donde ocurre el mínimo. Por lo tanto, en este caso, podemos afirmar que la dimensión informada por los datos se encuentra entre 20 y 30.

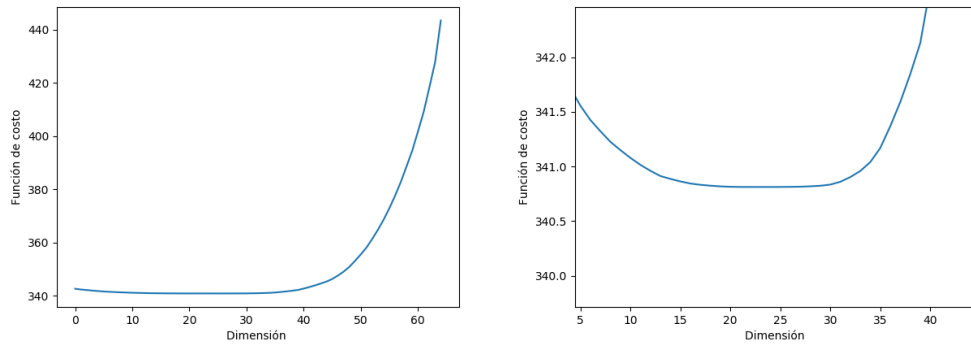


Figura 4.15: Gráfica de la aproximación de rango bajo. Caso con dominio anular circular.

Del mismo modo, en la figura (4.16) se muestra la función de costo contra la dimensión para el ejemplo con frontera interior elíptica. La gráfica ampliada en la derecha nos indica que la dimensión informada e este caso ocurre en el intervalo de 15 a 20 aproximadamente.

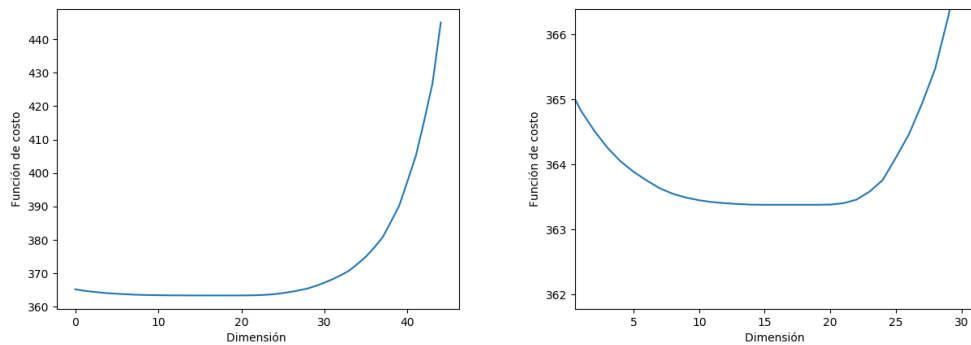


Figura 4.16: Gráfica de la aproximación de rango bajo. Caso con frontera interior elíptica

Las siguientes tres figuras muestran las gráficas análogas para el ejemplo con la región

anular compleja. La figura 4.17 muestra el caso con nivel de ruido de 1% y la figura 4.18 el caso con nivel de ruido de 10%. Se obtuvieron dimensiones informadas de 40 a 50, de 40 a 45 y de 40 a 45, respectivamente. En los tres casos se puede observar que la dimensión informada por los datos es menor a la dimensión de los datos utilizados en la solución del problema inversos. Recordemos que la dimensión de los datos que se utilizó fue de 50, así como la dimensión de los parámetros a estimar.

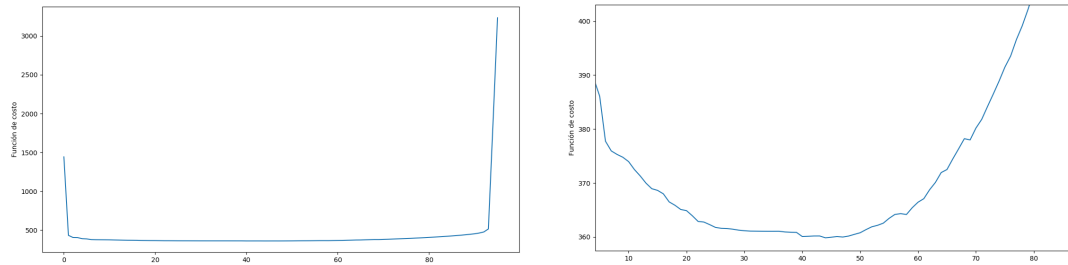


Figura 4.17: Gráfica de la aproximación de rango bajo. Dominio anular complejo con 1% de error

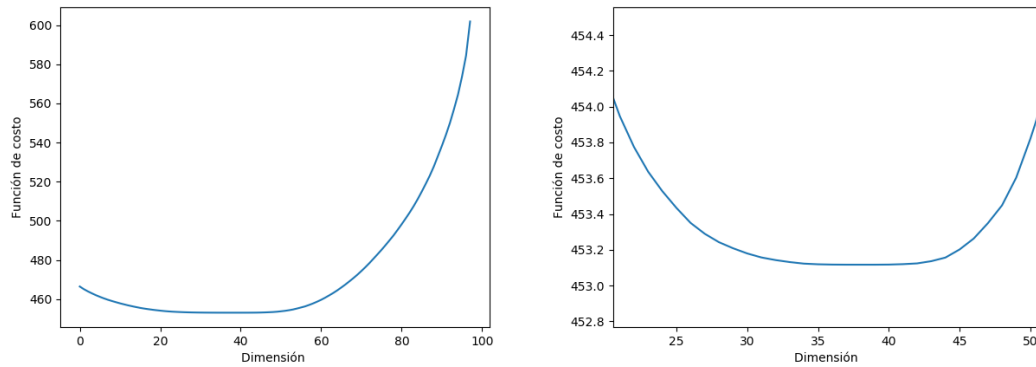


Figura 4.18: Gráfica de la aproximación de rango bajo. Dominio anular complejo con 10% de error

Capítulo 5

Conclusiones

En este trabajo se ha realizado un estudio numérico del problema de Cauchy para la ecuación de Laplace. En particular se ha ampliado el trabajo realizado en el artículo [10], el cual está basado en métodos variacionales y regularización de Tijonov de corte tradicional.

La extensión más importante ha sido la utilización de métodos de inversión Bayesiana, con algoritmos basados en MCMC. La solución con este enfoque es una distribución de probabilidad, lo cual permite extraer mayor información como es la incertidumbre en los resultados, además de las estimaciones puntuales obtenidas con los métodos deterministas.

Además de obtener soluciones del problema de Cauchy para la ecuación de Laplace de manera exitosa y satisfactoria con ambos enfoques, se han resaltado las conexiones entre ambos. Por ejemplo, el término de regularización en el método determinista está asociado con la distribución a priori en el método Bayesiano.

Las conexiones entre ambos enfoques nos han permitido agregar otros esquemas de regularización y las correspondientes distribuciones a priori que pueden ser más suavizantes o menos suavizantes. En particular se toman en cuenta conceptos de estadística espacial, utilizando los denominados campos aleatorios de Markov Gaussianos o GMRF para construir diferentes distribuciones Gaussianas por medio de diferentes matrices de precisión, generalizando las propuestas dadas en el trabajo de Bardsley et. al. [1] al caso de mallas no uniformes.

Dichas conexiones también permiten identificar el parámetro de regularización en el método determinista como la razón de la varianza de la distribución del error entre la varianza de la distribución a priori es decir $\alpha = \sigma_{noise}^2/\gamma^2$. Esta propiedad es consistente con el resultado en [10], en donde se demuestra que los mejores resultados con regularización de Tijonov se encuentran cuando el parámetro de regularización es proporcional al nivel del ruido en los datos.

En la sección 3.6 se muestra que a partir de la matriz D asociada a la discretización de $\partial\varphi/\partial s$ sobre los nodos de la malla se pueden construir en forma muy simple diferentes matrices de precisión Q asociadas a la distribución a priori. La malla no necesariamente es uniforme y los nodos no tienen una separación constante. Toda matriz de precisión Q también genera un correspondiente término de regularización de Tijonov en el método determinista. Esta propiedad es muy útil y puede servir para proponer diferentes distribuciones con GMRF definidos sobre esas mallas no uniformes.

Al final, se realiza un análisis de rango bajo, siguiendo en enfoque de Spantini et. al. [26], mostrando que no es necesario un gran número de datos para estimar el potencial en la frontera interior Γ_1 . Sin embargo, es posible refinar el mallado de la región para monitorear la calidad de los resultados.

Todos los resultados obtenidos con ambos métodos (determinista y Bayesiano) son muy satisfactorios y cercanos entre sí. En todos los casos no se observa una amplificación del ruido mayor a la introducida por los errores en los datos, principalmente cuando el ruido en los datos es significativo.

Cabe destacar que el objetivo de presentar ambos enfoques (Determinista y Bayesiano) es más que nada para puntualizar la relación que hay entre ellos ya que son de distinta naturaleza y garantizan cosas distintas debido al planteamiento que tienen.

Las herramientas utilizadas en las simulaciones como, el algoritmo t-walk en el método Bayesiano, y el software FEniCS para la solución del problema directo (forward map) han sido de gran ayuda y nos han permitido a simplificar la complejidad del código en el ambiente Python.

Posibles opciones para investigación futura pueden ser:

- Estudiar el caso en que el potencial sobre Γ_1 sea no suave, discontinuo o con focos puntuales.
- Hacer un análisis del error agregando la influencia de la discretización del modelo.
- Extender la aplicación de la metodología al problema inverso electroencefalográfico.
- Considerar el caso en que el número de datos sea pequeño y escenarios realistas.
- Estudiar y resolver el problema en tres dimensiones.

Bibliografía

- [1] J. M. Bardsley and J. Kaipio. Gaussian markov random field priors for inverse problems. *Inverse Problems & Imaging*, 7(2), 2013.
- [2] P. Bautista Otero. *Experimentación con un algoritmo de MCMC multiescala y autoajustable*. PhD thesis, Centro de Investigación en Matemáticas(CIMAT), 2007.
- [3] G. Y. T. Bello. Métodos variacionales en espacios de hilbert para la identificación de fuentes bioeléctricas. Master's thesis, UAM Iztapalapa, Matemáticas Aplicadas e Industriales), 2018.
- [4] F. Berntsson and L. Eldén. Numerical solution of a cauchy problem for the laplace equation. *Inverse Problems*, 17(4):839, 2001.
- [5] J.-R. Chang, W.-C. Yeih, and M.-H. Shieh. On the modified tikhonov. *Journal of Marine Science and Technology*, 9(2):113–121, 2001.
- [6] J. A. Christen, C. Fox, et al. A general purpose sampling algorithm for continuous distributions (the t-walk). *Bayesian Analysis*, 5(2):263–281, 2010.
- [7] M. Clerc and J. Kybic. Cortical mapping by laplace–cauchy transmission using a boundary element method. *Inverse Problems*, 23(6):2589, 2007.
- [8] D. Colton and R. Kress. *Integral equation methods in scattering theory*, volume 72. SIAM, 2013.
- [9] J. J. Conde Mones. *El problema inverso electroencefalográfico para fuentes volumétricas y su desarrollo numérico en geometrías simples*. PhD thesis, Facultad de Ciencias Físico Matemáticas de la Benemérita Universidad Autónoma de Puebla, 2013.
- [10] J. J. Conde Mones, L. H. Juárez Valencia, J. J. Oliveros Oliveros, and D. A. León Velasco. Stable numerical solution of the cauchy problem for the laplace equation in irregular annular regions. *Numerical Methods for Partial Differential Equations*, 33(6):1799–1822, 2017.
- [11] W. Förstner and B. Moonen. *A metric for covariance matrices*. Springer, 2003.

- [12] A. Fraguela, J. Oliveros, L. Cervantes, M. Morin, and S. Gomez. A non-iterative algorithm for electrical capacitance tomography. *REVISTA MEXICANA DE FISICA*, 51(3):236–242, 2005.
- [13] A. Fraguela, J. Oliveros, M. Morín, and L. Cervantes. Inverse electroencephalography for cortical sources. *Applied Numerical Mathematics*, 55(2):191–203, 2005.
- [14] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.
- [15] R. R. B. Isabel. Problema inverso electroencefalográfico: estudio numérico del sub-problema de detección de fuentes volumétricas. Master’s thesis, UAM Iztapalapa, (Matemáticas Aplicadas e Industriales), 2015.
- [16] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media, 2006.
- [17] R. E. Kass. Markov chain monte carlo in practice. *Journal of the American Statistical Association*, 92(440):1645–1647, 1997.
- [18] R. Kress. Inverse dirichlet problem and conformal mapping. *Mathematics and Computers in Simulation*, 66(4-5):255–265, 2004.
- [19] J.-Y. Lee and J.-R. Yoon. A numerical method for cauchy problem using singular value decomposition. *COMMUNICATIONS-KOREAN MATHEMATICAL SOCIETY*, 16(3):487–508, 2001.
- [20] C.-S. Liu and S. N. Atluri. Numerical solution of the laplacian cauchy problem by using a better postconditioning collocation trefftz method. *Engineering Analysis with Boundary Elements*, 37(1):74–83, 2013.
- [21] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.
- [22] N. Metropolis and S. Ulam. The monte carlo method. *Journal of the American statistical association*, 44(247):335–341, 1949.
- [23] M. Morín-Castillo, J. Oliveros-Oliveros, J. Conde-Mones, A. Fraguela-Collar, E. Gutiérrez-Arias, and E. Flores-Mena. Simplificación del problema inverso electroencefalográfico a una sola región homogénea con condición de neumann nula. *Revista mexicana de ingeniería biomédica*, 34(1):41–51, 2013.
- [24] F. Ramos-Argüelles, G. Morales, S. Egozcue, R. Pabón, and M. Alonso. Técnicas básicas de electroencefalografía: principios y aplicaciones clínicas. In *Anales del sistema sanitario de Navarra*, volume 32, pages 69–82. SciELO Espana, 2009.

- [25] H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC, 2005.
- [26] A. Spantini, A. Solonen, T. Cui, J. Martin, L. Tenorio, and Y. Marzouk. Optimal low-rank approximations of bayesian linear inverse problems. *SIAM Journal on Scientific Computing*, 37(6):A2451–A2487, 2015.
- [27] M. A. C. Sumano, A. F. Collar, A. Grebennikov, M. M. M. Castillo, and J. J. O. Oliveros. Solución estable del problema de cauchy para la ecuación de laplace usando potenciales de superficie. *Lecturas Matemáticas*, 32:61–77, 2011.
- [28] D. Young, C. Tsai, C. Chen, and C.-M. Fan. The method of fundamental solutions and condition number analysis for inverse problems of laplace equation. *Computers & Mathematics with Applications*, 55(6):1189–1200, 2008.