

UNIVERSIDAD NACIONAL DE
INGENIERÍA
FACULTAD DE CIENCIAS



TESIS
“TRANSPORTE ÓPTIMO E INVERSIÓN DE
ONDAS”

PARA OBTENER EL TÍTULO PROFESIONAL DE:
LICENCIADO EN MATEMÁTICA

ELABORADA POR:
ENRIQUE IDAEL CHAVEZ SARMIENTO

ASESOR:
Dr. ELADIO OCAÑA

LIMA-PERÚ

2023

Dedicada a Matías

Agradecimientos

La presente tesis ha sido lograda gracias al apoyo de muchos, y me será imposible mencionar a todos aquí. Empezando por varias personas muy especiales en mi familia, a quienes les debo y deberé siempre gran parte de los logros académicos o profesionales de mi camino. Mi infinito agradecimiento a todas ellas. Les debo mucho más que el haber terminado este trabajo.

Además agradezco a Eladio Ocaña, quién fue el principal impulsor de que este proyecto de tesis se haga realidad y quién me motivó a terminarlo. Sin él, ni su apoyo, esta tesis no existiría. Del mismo modo, no quiero dejar pasar la oportunidad de agradecer a Juan Pablo Luna, quién me presentó este interesante mundo de la Optimización. Esto terminó motivando la elección del tema de esta tesis y espero que, de la misma manera, permita el desarrollo de otros proyectos en la misma línea.

Finalmente, agradezco a la Facultad de Ciencias, donde inicié mi formación matemática y pasé algunos de los mejores años que puedo recordar.

Resumen

En este trabajo desarrollamos las herramientas necesarias para poder aproximar una onda *real* por una teórica, proveniente de la solución de la ecuación de la onda. Ésto lo conseguimos mediante el uso de funciones de discrepancia, que nos ayudan a comparar las ondas reales con las teóricas. Concretamente, si d_v representa a la onda teórica, generada por un campo de velocidades v , y d_0 es la onda real que queremos aproximar, nuestro objetivo es minimizar una función del tipo $F_D(v) = D(d_v, d_0)$, donde D es una función de discrepancia.

Parte fundamental del trabajo es entender que incluso cuando D es una función suficientemente suave, el proceso de minimización de F_D , en la práctica, no es simple. Por la regla de la cadena, el cálculo del gradiente de F depende del gradiente $\nabla_v d_v$. Sin embargo, en términos prácticos este último gradiente no es simple de calcular. Entonces, para cumplir nuestro objetivo es fundamental calcular $\nabla_v F$ sin el conocimiento de $\nabla_v d_v$. Ésto lo hacemos bajo ciertas condiciones, suficientemente generales, que nos permiten trabajar incluso con discrepancias que no provienen de distancias en $\mathbb{R}^m$. En particular, dentro de las aplicaciones que realizamos está el cálculo del gradiente de una función objetivo basada en una función de discrepancia que hace uso del transporte óptimo desbalanceado, en adelante UOT, por sus siglas en inglés.

Palabras clave: Ecuación de onda, inversión de ondas, transporte óptimo desbalanceado, función de discrepancia, distancia de Wasserstein.

Abstract

In this work, we develop the necessary tools to approximate a *real* wave by a theoretic one, that comes from the solution of the wave equation. We achieve this through the use of discrepancy functions, that help us to compare the two kinds of waves. More precisely, if d_v represents the theoretic wave, generated by a velocity field v , and d_0 is the real wave, that we want to approximate, our goal is to minimize an objective function like $F_D(v) = D(d_v, d_0)$, where D is a discrepancy function.

A fundamental part of this work is to understand that, even when D is a sufficiently smooth function, in practice, the minimization process of F_D is nontrivial. From the chain rule, the gradient of F depends on the gradient $\nabla_v d_v$. However, in practical applications, last gradient do not have a simple computation. Hence, to achieve our goal is key to compute $\nabla_v F$ without the knowledge of $\nabla_v d_v$. We do this under certain conditions, general enough, to allow us to work even with discrepancies that do not come from metrics in $\mathbb{R}^m$. In particular, among the applications we do, there is the computation of the gradient of an objective function that derives from a discrepancy function that uses the unbalanced optimal transportation, UOT.

Key words: Wave equation, wave inversion, unbalanced optimal transport, discrepancy function, Wasserstein distance.

Tabla de contenido

Agradecimientos	III
Resumen	IV
Abstract	V
Tabla de contenido	VI
Prólogo	VIII
Introducción	IX
Capítulo I: Preliminares	1
Tópicos de Cálculo	1
Funciones convexas	6
Función de onda	11
Discretización de la función de onda	12
Suavización de la parte positiva	15
Casos particulares	20
Distancia de Wasserstein	21
Tópicos de probabilidad	21
El problema de Monge-Kantorovich	26
Función KL	29

Capítulo II: Transporte óptimo desbalanceado	32
Transporte óptimo balanceado	32
Transporte óptimo desbalanceado - UOT	33
Capítulo III: Inversión de ondas	38
Problema directo	38
Problema inverso	39
Cálculo del gradiente	40
Funciones de discrepancia desagregadas por instante de tiempo . .	42
Demostración del Método de la adjunta	44
Discrepancia usando una distancia L^2	47
Discrepancia usando una distancia W_2	48
Tratamiento de valores negativos	52
Discrepancia basada en la función KL	57
Discrepancia basada en el UOT	58
Discrepancia simetrizando el UOT	59
Conclusiones	60
Índice alfabético	61
Bibliografía	63

Prólogo

Esta tesis está organizada del siguiente modo:

En el capítulo **I** discutimos resultados y notaciones básicas que emplearemos en el resto del trabajo. Esto incluye la introducción de la función de onda que usaremos y nuestro método de discretización espacial de la misma. Además estudiamos un método para obtener suavizaciones de la función parte positiva y concluimos el capítulo introduciendo la función KL que será de vital importancia para entender el UOT.

En el capítulo **II** introducimos la noción de transporte óptimo desbalanceado, UOT. El principal resultado del capítulo es el cálculo de la derivada de la función D_H .

En el capítulo **III** definimos el problema de inversión de ondas. También demostramos el *Método de la adjunta* que nos permite calcular gradientes de funciones objetivo, con la finalidad de resolver el problema de inversión de onda. Finalmente presentamos una serie aplicaciones de este método.

Introducción

En la ciencia y la industria existe la necesidad de explorar un medio con métodos mínimamente invasivos. Cuando hay la posibilidad de que por este medio se propague una onda generada de forma controlada, es posible que se pueda conocer el medio a través del estudio de las ondas propagadas en este. La Geología, Arqueología, Geografía, y las industrias extractivas frecuentemente se enfrentan a este tipo de situaciones [24], [20], [19], [14].

En la presente tesis, abordamos el problema desde un punto de vista teórico, brindando las herramientas matemáticas suficientes, para una posterior implementación numérica de los métodos aquí presentados. En situaciones como la descrita, típicamente se tiene un modelo G que depende de un parámetro v , de modo que la onda producida por el modelo $u_v := G(v)$ se corresponde con una onda real propagándose en un medio asociado al parámetro v . En este contexto asumimos que existe una caracterización del medio en estudio por el parámetro v . En otras palabras, si conocemos el parámetro v , conocemos el medio de estudio. Esto significa que para conocer el medio, nos basta con conocer el parámetro v que lo caracteriza.

Supongamos, por ejemplo, que deseamos estudiar el medio Ω y que al conocer la velocidad de propagación $v(x)$ de la una onda en el punto $x \in \Omega$ somos capaces de conocer las propiedades deseadas de Ω en x (densidad, presión, composición, etc.). Asumamos además que disponemos de un modelo G que relaciona la onda con nuestro parámetro $v(\cdot)$; por ejemplo, digamos que $G(v)$ nos proporciona la solución de la ecuación

$$\frac{1}{v^2(x)} u_{tt}(x, t) = \nabla^2 u(x, t) + S(x, t) . \quad (1)$$

En estas condiciones, si somos capaces de registrar partes de una onda real u_0 , para conocer Ω , es suficiente con encontrar el $v(\cdot)$ óptimo de modo que $G(v) \simeq u_0$. En un escenario ideal, el problema consiste en encontrar el parámetro $v(\cdot)$ conociendo la solución u_v de la ecuación (1). En la realidad, no conocemos u_v , en su lugar tenemos acceso a una onda real u_0 generada, de forma controlada por nosotros y tenemos que seleccionar “el mejor” $v(\cdot)$ de modo que u_v sea los más parecido posible a u_0 . Esta es la idea principal del problema de inversión de ondas.

El abordaje técnico que seguimos en este trabajo está enfocado en la teoría de inversión de ondas [7], [2], [12], [24]. Presentamos ejemplos variados, entre los que destacan un abordaje clásico, usando la distancia L^2 de $\mathbb{R}^n$, y otro usando la Distancia de Wasserstein W_2 . La comparación entre la onda teórica u_v y la real u_0 la hacemos por medio de una función de discrepancia D . De modo que mientras más pequeño sea el valor de $D(u_v, u_0)$, más parecidas serán las ondas. Entonces,

haciendo $F_D(v) = D(u_v, u_0)$, tenemos que resolver el problema de optimización, en general no lineal, sin restricciones

$$\min_v F_D(v) . \quad (2)$$

En problemas reales, este problema no tiene solución directa, incluso si F_D tiene propiedades favorables que permiten su manipulación. Por ejemplo, en el caso en que F_D sea diferenciable, podemos pensar en usar su gradiente para resolver (2) [8], [1]. Notemos que en este caso, incluso si logramos darle sentido a la expresión

$$\nabla F_D(v) = \frac{\partial D}{\partial x}(u_v, u_0) \frac{\partial u_v}{\partial v}(v) ,$$

nos encontramos frente al problema de calcular $\frac{\partial u_v}{\partial v}$, pues la correspondencia $v \mapsto u_v$ es de difícil manejo en términos prácticos. Para superar esta dificultad, cuando F_D es diferenciable, empleamos el *método de la adjunta* [3], [21]. Utilizando este método, somos capaces de encontrar el gradiente de F_D usando información de (1), pero sin calcular directamente $\frac{\partial u_v}{\partial v}$.

Para comparar la onda registrada con la teórica recientemente se está planteando el uso de la teoría de transporte óptimo [12], [18], [23]. Este es el camino en el que nos enfocamos en la tesis. Para poder implementarlo enfrentamos la dificultad técnica de tener que transportar ondas que pueden ser positivas y negativas; mientras que el transporte óptimo clásico está desarrollado para transportar cantidades positivas y además de la misma

“masa”. En la Sección **III.6** discutimos diferentes formas de abordar este inconveniente, decantándonos esencialmente por la comparación de dos magnitudes de x, y de signo cambiante, mediante el cálculo de la expresión

$$D([x]^+, [y]^+) + D([x]^-, [y]^-) .$$

Con este abordaje somos capaces de estudiar discrepancias basadas en el transporte óptimo desbalanceado, al que nos referimos como UOT por sus siglas en inglés.

Capítulo I

Preliminares

En este capítulo definimos la función de onda que consideramos para nuestro problema de inversión de ondas. También desarrollamos una parte técnica en la que estudiamos aproximaciones suaves de la parte positiva. Esto nos es útil a la hora de definir nuestra principal función de discrepancia en el capítulo [III](#). Finalmente, en la última sección, introducimos la función KL que es interesante en sí misma y que también utilizamos en el capítulo [II](#) para definir el UOT.

I.1. Tópicos de Cálculo

En esta sección repasaremos brevemente algunos resultados fundamentales de cálculo que serán usados en el resto de este trabajo, a veces sin mención explícita. Para las demostraciones de los resultados que exponemos, sin demostración, referimos al lector a [\[9\]](#) y [\[10\]](#).

Teorema I.1.1 (Primer teorema fundamental del cálculo). Sea $f : [a, b] \rightarrow \mathbb{R}$ una función continua. Si $F : [a, b] \rightarrow \mathbb{R}$ está definida por

$$F(x) = \int_a^x f(t) dt ,$$

entonces F es uniformemente continua en $[a, b]$, diferenciable en (a, b) y además

$$F'(x) = f(x) , \quad \forall x \in (a, b) .$$

Como corolario de este resultado tenemos la siguiente versión débil del segundo teorema fundamental del cálculo.

Corolario I.1.2. Sea $f : [a, b] \rightarrow \mathbb{R}$ una función continua para la que existe una función continua $F : [a, b] \rightarrow \mathbb{R}$ tal que $F'(x) = f(x)$ para todo $x \in (a, b)$. Entonces

$$\int_a^b f(t) dt = F(b) - F(a) .$$

Demostración. Sea $G : [a, b] \rightarrow \mathbb{R}$ definida por

$$G(x) = \int_a^x f(t) dt .$$

Por el primer teorema fundamental del cálculo, tenemos que $G'(x) = f(x)$ para todo $x \in (a, b)$. Luego

$$(F - G)'(x) = f(x) - f(x) = 0 ,$$

para todo $x \in (a, b)$. Esto significa, como consecuencia del teorema del valor medio para derivadas, que $F - G$ es constante en (a, b) . De la continuidad de F y G , esto se extiende a todo el intervalo $[a, b]$. Es decir, existe $K \in \mathbb{R}$ tal que

$$F(x) - G(x) = K , \quad \forall x \in [a, b] .$$

Por tanto

$$\begin{aligned} \int_a^b f(t) dt &= G(b) - G(a) \\ &= [F(b) - K] - [F(a) - K] \\ &= F(b) - F(a) , \end{aligned}$$

como queríamos demostrar. □

También tenemos la siguiente versión de la integración por partes para integrales definidas.

Proposición I.1.3 (Integración por partes). Sean $f, g : [a, b] \rightarrow \mathbb{R}$ funciones continuas en $[a, b]$, con derivada continua en (a, b) . Entonces

$$\begin{aligned}\int_a^b f(x)g'(x) &= \left(f(x)g(x) \right) \Big|_a^b - \int_a^b f'(x)g(x) dx \\ &= f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x) dx .\end{aligned}$$

Por otro lado, para funciones de dos (o más variables) tenemos el siguiente resultado, que nos permitirá intercambiar el orden de las derivadas parciales con las integrales en todos los cálculos realizados a partir de la Sección III.3.

Proposición I.1.4 (Regla de Leibniz). Sea $f : \mathbb{R} \times [a, b] \rightarrow \mathbb{R}$ una función continua para la que existe $f_x : \mathbb{R} \times [a, b] \rightarrow \mathbb{R}$ y es continua. Entonces

$$\frac{d}{dx} \left(\int_a^b f(x, t) dt \right) = \int_a^b \frac{\partial f}{\partial x}(x, t) dt .$$

Demostración. Sean $x \in \mathbb{R}$ y $h > 0$. Entonces

$$\int_x^{x+h} \int_a^b f_x(x, t) dt dx = \int_a^b \int_x^{x+h} f_x(x, t) dx dt .$$

Luego, por el Corolario I.1.2, resulta que

$$\begin{aligned}\int_x^{x+h} \int_a^b f_x(x, t) dt dx &= \int_a^b [f(x+h, t) - f(x, t)] dt \\ &= \int_a^b f(x+h, t) dt - \int_a^b f(x, t) dt .\end{aligned}$$

Entonces, si tomamos $x_0 < x$ y definimos

$$F(u) = \int_{x_0}^u \int_a^b f_x(x, t) dt dx ,$$

resulta que

$$\begin{aligned} \frac{1}{h} \left(\int_a^b f(x+h, t) dt - \int_a^b f(x, t) dt \right) &= \frac{1}{h} \int_x^{x+h} \int_a^b f_x(x, t) dt dx \\ &= \frac{F(x+h) - F(x)}{h} . \end{aligned} \quad (\text{I.1})$$

Gracias al Teorema [I.1.1](#), sabemos que

$$F'(x) = \int_a^b f_x(x, t) dt .$$

Esto significa que el límite en [\(I.1\)](#), cuando $h \rightarrow 0^+$, existe y más aún

$$\begin{aligned} \int_a^b f_x(x, t) dt &= F'(x) \\ &= \lim_{h \rightarrow 0^+} \frac{F(x+h) - F(x)}{h} \\ &= \lim_{h \rightarrow 0^+} \frac{1}{h} \left(\int_a^b f(x+h, t) dt - \int_a^b f(x, t) dt \right) . \end{aligned}$$

El otro límite lateral se trata de forma análoga. □

I.2. Funciones convexas

En esta sección enunciaremos algunos resultados básicos sobre análisis convexo. Referimos a [16] y [6] para una presentación mucho más extensa y detallada del tema.

Definición I.2.1. *Decimos que un conjunto $C \subset \mathbb{R}^n$ es convexo cuando para cualquier $x, y \in C$*

$$tx + (1 - t)y \in C, \quad \text{para todo } t \in [0, 1].$$

Geométricamente, un conjunto es convexo cuando para cualquier par de puntos del conjunto, el segmento que une esos puntos está contenido en el conjunto. Notemos que en la Definición I.2.1 los puntos de la forma $p(t) = tx + (1 - t)y$, con $t \in [0, 1]$, recorren todos los puntos del segmento que une a x e y . Cuando $t = 1$ estamos en $p(1) = x$, cuando $t = 0$ estamos en $p(0) = y$; mientras que cuando $t = 1/2$ estamos en $p(1/2) = \frac{1}{2}(x + y)$, la mitad del segmento que une x con y . Por ejemplo cualquier bola $B_r(x_0) = \{x \in \mathbb{R}^n : \|x - x_0\| < r\}$ es un conjunto convexo. En la Figura I.1 puede verse el caso de dimensión 2. Además, en $\mathbb{R}$ los únicos conjuntos convexas son los intervalos.

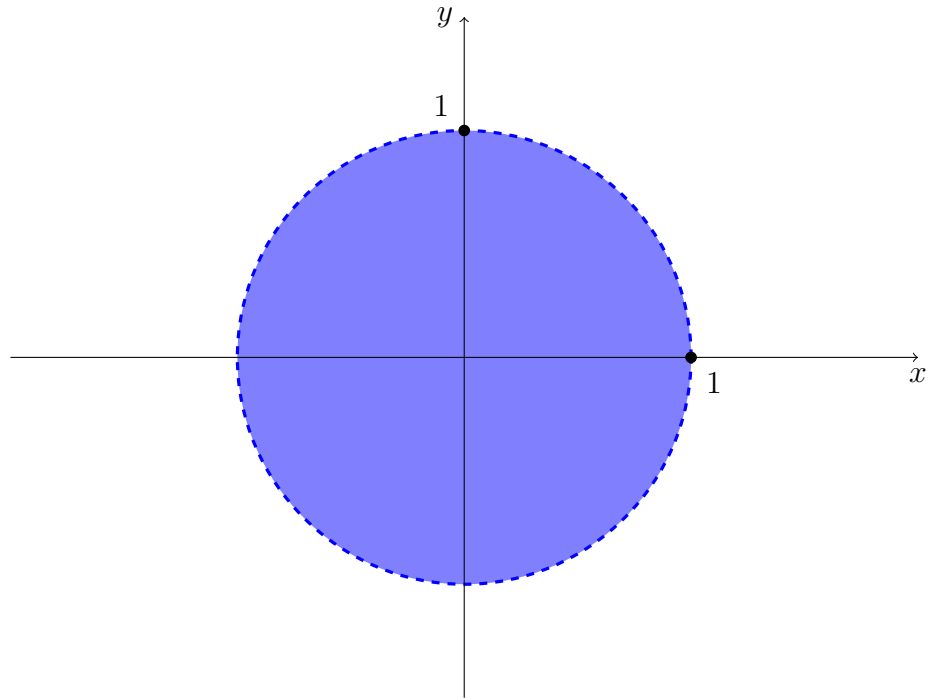


Figura I.1: El disco $B_1(0,0) \subset \mathbb{R}^2$ es convexo

Definición I.2.2 (Función convexa). *Una función $f : C \subset \mathbb{R}^n \rightarrow \mathbb{R}$, definida en un conjunto convexo C , es convexa cuando para todo $x, y \in C$ vale la desigualdad*

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y), \quad \forall t \in [0, 1].$$

Si en la anterior expresión la desigualdad es estricta, entonces decimos que f es estrictamente convexa.

Geométricamente, una función es convexa cuando al unir dos puntos cualesquiera de su gráfica, el segmento que une dichos puntos está por encima de, o se sobrepone a, la gráfica de la función. Además una función es estrictamente convexa, cuando el segmento abierto que une dos puntos de la gráfica de la función está estrictamente por encima de la gráfica de la función.

En la figura I.2 podemos ver un ejemplo de una función $f : \mathbb{R} \rightarrow \mathbb{R}$ (estrictamente) convexa.

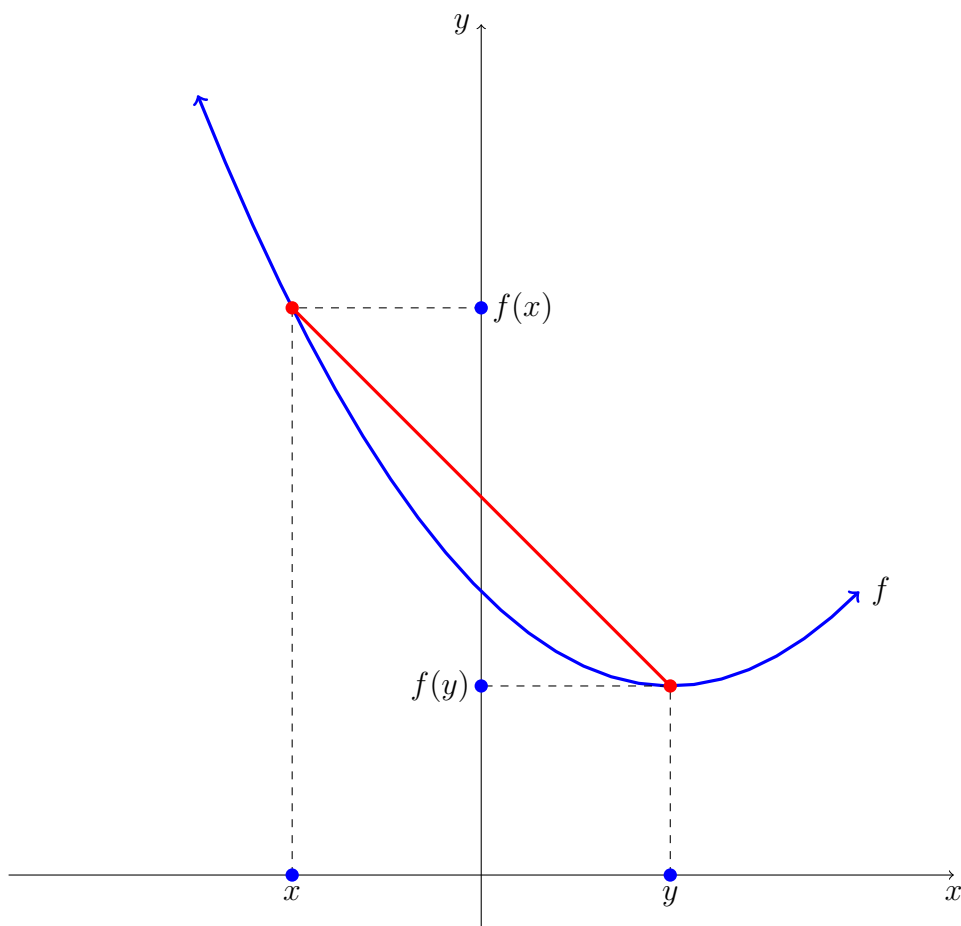


Figura I.2: La función f es (estrictamente) convexa

Observación I.2.3. Una función $f : C \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es convexa, si y solo si el conjunto

$$\mathcal{E}(f) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq y\} \subset \mathbb{R}^{n+1},$$

es convexo.

Ahora enunciaremos algunos hechos importantes sobre funciones convexas.

Para todos ellos fijamos una función $f : C \rightarrow \mathbb{R}$ convexa.

- **Diferenciabilidad.** La función f es continua y localmente Lipschitz en el interior de su dominio. Además su gradiente ∇f está bien definido en casi todo punto y es localmente acotado.
- **Gráfica y tangente.** Para todos los puntos x donde ∇f está bien definido tenemos la siguiente relación.

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle .$$

Esto expresa el hecho geométrico de que todo el gráfico de f se sitúa por encima de su hiperplano tangente en el punto $(x, f(x))$. De esto se deduce que ∇f es monótono, es decir, si $x, y \in C$ son dos puntos donde ∇f está definido, entonces

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0 .$$

Ilustramos este hecho en la Figura I.3.

A la luz de los hechos antes mencionados, queda saber qué ocurre en los puntos donde una función convexa no es diferenciable. Para abordar esto tenemos la siguiente definición que generaliza la noción de gradiente.

Definición I.2.4. Sea $f : C \rightarrow \mathbb{R}$ una función convexa. Para cada $x \in C$, el subdiferencial de f en x , denotado por $\partial f(x)$, es el conjunto de puntos $y \in \mathbb{R}^n$

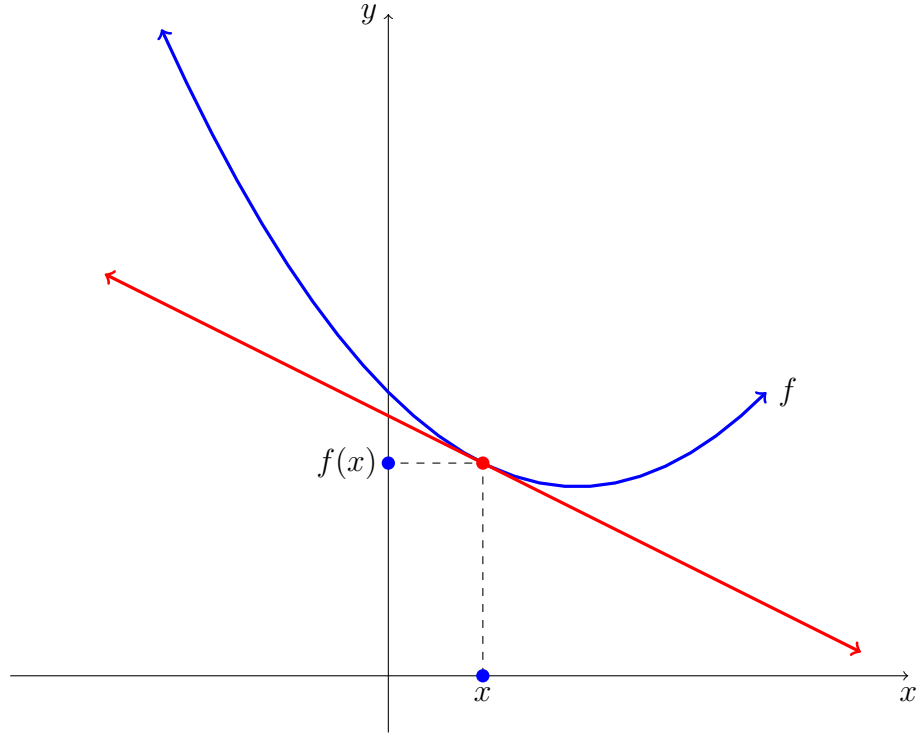


Figura I.3: La gráfica de f está sobre su tangente

tales que

$$f(z) \geq f(x) + \langle y, z - x \rangle, \quad \forall z \in \mathbb{R}^n.$$

Geométricamente, podemos pensar en el subdiferencial de una función, como el conjunto de vectores que, igual que el gradiente cuando existe, generan hiperplanos que se ubican por debajo de la gráfica de la función y tienen algún punto en común con la gráfica de la función. En la figura I.4 tenemos una representación de esta situación.

Observación I.2.5. Como el conjunto $\mathcal{E}(f)$ de la Observación I.2.3 es convexo y $\{(x, f(x))\}$ también lo es, podemos separar estos conjuntos por un hiperplano. Esto muestra que para todo x en el interior de C , el subdiferencial $\partial f(x)$ siempre

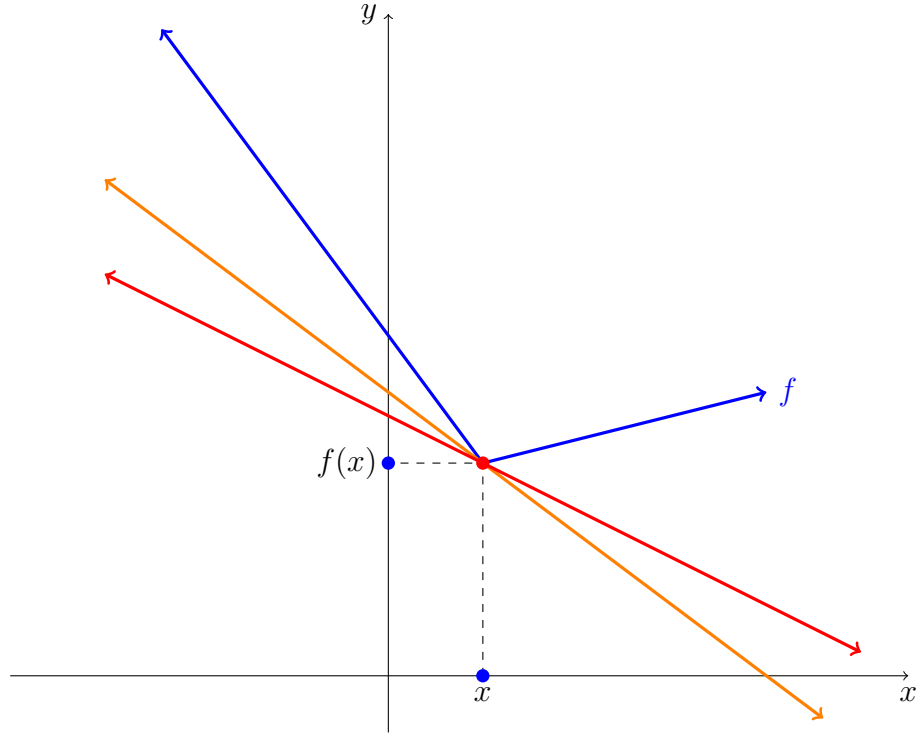


Figura I.4: Representación gráfica de $\partial f(x)$

es diferente del vacío. Más aún, se cumple que f es diferenciable en x , si y solo si $\partial f(x)$ es un conjunto unitario.

I.3. Función de onda

Para simplificar consideramos un dominio espacial unidimensional $\Omega \subset \mathbb{R}$ igual a un intervalo cerrado y acotado. Además $T > 0$ es una constante.

Definición I.3.1. El campo de onda $u : \Omega \times [0, T] \rightarrow \mathbb{R}$ generado por el modelo de velocidades $v : \Omega \rightarrow \mathbb{R} \setminus \{0\}$ y la fuente $S : \Omega \times [0, T] \rightarrow \mathbb{R}$ es la solución de la ecuación de onda

$$\frac{1}{v^2(x)} \frac{\partial^2 u}{\partial t^2}(x, t) = \frac{\partial^2 u}{\partial x^2}(x, t) + S(x, t) ,$$

junto a las condiciones de contorno

$$\begin{aligned} u(x, 0) &= 0, & \forall x \in \Omega \\ u_t(x, 0) &= 0, & \forall x \in \Omega \\ u(x, t) &= 0, & \forall x \in \partial\Omega, \quad \forall t \in [0, T]. \end{aligned}$$

En un contexto práctico, nosotros tenemos control sobre la fuente S y además tenemos acceso a la velocidad $v(x)$ de la onda, dependiendo del medio x donde se propague.

I.4. Discretización de la función de onda

En problemas reales, nuestra ecuación de onda será manejada desde el punto de vista numérico. Por esta razón discretizaremos el dominio espacial Ω considerando una *mall*a

$$\tilde{\Omega} = \{\text{mín } \Omega = x_1 < x_2 < \cdots < x_m = \text{máx } \Omega\},$$

uniformemente distribuida.

Definición I.4.1. *La discretización espacial de una función $f : \Omega \times [0, T] \rightarrow \mathbb{R}$ es $\tilde{f} = (f_1, f_2, \dots, f_m)^T : [0, T] \rightarrow \mathbb{R}$, donde*

$$f_i(t) = f(x_i, t), \quad \forall i \in \{1, 2, \dots, m\}, \quad \forall t \in [0, T].$$

Similarmente, la discretización de una función $g : \Omega \rightarrow \mathbb{R}$ es el vector $\tilde{g} = (g_1, g_2, \dots, g_m)^T$ definido por $g_i = g(x_i)$ para todo $i \in \{1, 2, \dots, m\}$

Proposición I.4.2. Sea $f : \Omega \times [0, T] \rightarrow \mathbb{R}$ una función de clase C^2 . Existen una malla suficientemente fina y una matriz K tales que $K\tilde{f}$ aproxima $\frac{\partial^2 f}{\partial x^2}$ en el siguiente sentido: Dado $\varepsilon > 0$, existe un número natural $m > 2$ tal que

$$\left| \frac{\partial^2 f}{\partial x^2}(x_i, t) - K_i \tilde{f}_i(t) \right| < \varepsilon, \quad \forall i \in \{1, 2, \dots, m\}, \quad \forall t \in [0, T]; \quad (\text{I.2})$$

donde K_i es la i -ésima fila de la matriz K , de orden m .

Demostración. Afirmamos que si $g : [a, b] \rightarrow \mathbb{R}$ es una función de clase C^2 , entonces

$$\begin{aligned} g''(x) &= \frac{g(x+h) + g(x-h) - 2g(x)}{h^2} + o(1), \quad x \in (a, b) \\ g''(a) &= \frac{g(a+2h) + g(a) - 2g(a+h)}{h^2} + o(1) \quad \text{y} \\ g''(b) &= \frac{g(b-2h) + g(b) - 2g(b-h)}{h^2} + o(1), \end{aligned}$$

aproximan $g''(x)$ de manera uniforme en $[a, b]$. En efecto. Veremos la primera aproximación; las dos últimas son aproximaciones puntuales más simples. Sea $\varepsilon > 0$, de la continuidad uniforme de g'' en $[a, b]$, existe $\delta > 0$ tal que si $x, y \in [a, b]$ son tales que $|x - y| < \delta$, entonces $|g''(x) - g''(y)| < \varepsilon$. Notemos que si $x \in (a, b)$

y $h \in (0, \delta)$, por el primer teorema fundamental del cálculo

$$\begin{aligned}
& \left| g''(x) - \frac{g(x+h) + g(x-h) - 2g(x)}{h^2} \right| \\
&= \left| g''(x) - \frac{1}{h^2} \left(\int_x^{x+h} g'(t) dt - \int_x^{x+h} g'(t-h) dt \right) \right| \\
&= \left| g''(x) - \frac{1}{h^2} \int_x^{x+h} \int_{t-h}^t g''(s) ds dt \right| \\
&= \left| \frac{1}{h^2} \int_x^{x+h} \int_{t-h}^t [g''(x) - g''(s)] ds dt \right| \\
&\leq \varepsilon ,
\end{aligned}$$

donde la última desigualdad es porque $|x - s| \leq h < \delta$.

En el contexto de la proposición, denotamos por $\Delta_m = (x_m - x_1)/(m-1)$ al espaciamiento entre puntos de la malla $\tilde{\Omega}$.

Sean $\varepsilon > 0$, $t \in [0, T]$ y $\delta_t > 0$ tales que si $|x - y| < \delta_t$, con $x, y \in \Omega$, entonces $|f_{xx}(x, t) - f_{xx}(y, t)| < \varepsilon$. Escogemos $m = m(t)$ tal que $\Delta_m < \delta_t$. Entonces, por la primera parte de la demostración

$$\left| \frac{1}{\Delta_m^2} (0, 0, \dots, 1, -2, 1, 0, \dots, 0) \tilde{f}(t) - \frac{\partial^2 f}{\partial x^2}(x_i, t) \right| < \varepsilon , \quad \forall i \in \{2, 3, \dots, m-1\} ,$$

donde -2 ocupa la posición i . Entonces, si

$$K_i = \frac{1}{\Delta_m^2} (0, 0, \dots, 1, -2, 1, 0, \dots, 0) ,$$

obtenemos (I.2) para $i \in \{2, 3, \dots, m-1\}$ y $t \in [0, T]$ fijo. Similarmente obtenemos aproximaciones para $i \in \{1, m\}$, con t fijo. Finalmente, la

compacidad del intervalo $[0, T]$ nos permite encontrar $m > 2$, que no depende de $t \in [0, T]$, de modo que (I.2) se cumple. $\square$

Observación I.4.3. *Si u es un campo de onda como en la Definición I.3.1, gracias a las condiciones de contorno, podemos lograr que la aproximación de la proposición I.4.2 se consiga con una matriz K simétrica.*

Finalmente, si definimos la matriz diagonal $M(v)$ como

$$M(v) := \text{diag} \left[\frac{1}{v^2(x_1)}, \frac{1}{v^2(x_2)}, \dots, \frac{1}{v^2(x_m)} \right],$$

obtenemos que

$$M(v)\tilde{u}''(t) = K\tilde{u}(t) + \tilde{S}(t), \quad \text{donde} \quad \tilde{u}(0) = \tilde{u}'(0) = 0, \quad (\text{I.3})$$

es la versión discreta de la ecuación de onda en la Definición I.3.1.

De ahora en adelante el análisis que realizaremos se restringirá a la versión discretizada de la ecuación de onda. Por tanto, abusando de la notación, no usaremos el símbolo “ $\sim$ ” para referirnos a las funciones discretizadas en (I.3).

I.5. Suavización de la parte positiva

La función parte positiva es un recurso importante en matemática, en donde muchos problemas pueden ser modelados usando esta función. Este recurso también es usado en otras áreas como la Ingeniería, la Economía y las ciencias de la naturaleza, que muchas veces lidian con problemas en los que las

propiedades del objeto de estudio imponen la restricción de ser no negativo. Esto ocurre, por ejemplo, en problemas involucrando minimización de precios, restauración de imágenes, reconstrucción de señales, equilibrio estocástico y otros [4]. Desafortunadamente la función parte positiva no es diferenciable. Esto significa que los métodos numéricos de optimización que usan gradiente no pueden ser empleados directamente en esta función. Para salvar este problema usamos versiones suaves de la función, como explicamos a continuación.

Definición I.5.1. *La función parte positiva $[\cdot]^+ : \mathbb{R} \rightarrow \mathbb{R}$ está definida por $[x]^+ = \max\{x, 0\}$.*

Entonces, si $\rho : \mathbb{R} \rightarrow \mathbb{R}$ está definida por

$$\rho(x) = \begin{cases} 1, & \text{si } x \geq 0 \\ 0, & \text{si } x < 0, \end{cases}$$

tenemos que

$$[x]^+ = \int_{-\infty}^x \rho(t) dt .$$

Notemos que si δ es la medida de probabilidad en $\mathbb{R}$ tal que $\delta(\{0\}) = 1$ y $\delta(\mathbb{R} \setminus \{0\}) = 0$, entonces

$$\rho(x) = \int_{(-\infty, x]} d\delta .$$

Nuestra estrategia de aproximación para la función $[\cdot]^+$ es considerar una familia de funciones $(\sigma_\tau, \tau > 0)$ convergiendo para $[\cdot]^+$ cuando $\tau \rightarrow 0$, a partir de una familia de densidades de probabilidad $(\delta_\tau, \tau > 0)$ que converge débilmente para δ , i.e., tal que

$$\lim_{\tau \rightarrow 0} \delta_\tau((-\infty, x]) = \delta((-\infty, x]) = \rho(x), \quad \forall x \neq 0. \quad (\text{I.4})$$

Las densidades δ_τ que consideramos son pares y tales que

$$\delta_\tau(x) = \frac{1}{\tau} \delta_1\left(\frac{x}{\tau}\right), \quad \forall \tau > 0. \quad (\text{I.5})$$

Por ejemplo, comenzando con una densidad par δ_1 tal que

$$\lim_{x \rightarrow +\infty} x \delta_1(x) = 0,$$

y usando (I.5) para definir el resto de la familia, tenemos que (I.4) se cumple.

Luego $\rho_\tau : \mathbb{R} \rightarrow \mathbb{R}$ definida por

$$\rho_\tau(x) = \int_{-\infty}^x \delta_\tau(t) dt,$$

converge para ρ , cuando $\tau \rightarrow 0$, para todo $x \neq 0$. Además suponemos que $\sigma_\tau :$

$\mathbb{R} \rightarrow \mathbb{R}$ dada por

$$\sigma_\tau(x) = \int_{-\infty}^x \rho_\tau(t) dt, \quad (\text{I.6})$$

está bien definida. En este caso $\rho_\tau(x) = \rho_1(x/\tau)$ para todo $x \in \mathbb{R}$ y también vale el siguiente lema.

Lema I.5.2. *Si σ_τ está definida por (I.6), entonces*

$$\sigma_\tau(x) - [x]^+ = \int_{-\infty}^{-|x|} \rho_\tau(t) dt, \quad \forall x \in \mathbb{R}. \quad (\text{I.7})$$

En particular $\sigma_\tau(x) \geq [x]^+$ para todo $x \in \mathbb{R}$.

Demostración. Tenemos que

$$\sigma_\tau(x) - [x]^+ = \int_{-\infty}^x \rho_\tau(t) - \rho(t) dt, \quad \forall x \in \mathbb{R}.$$

Luego (I.7) es clara para $x \leq 0$. Sea entonces $x > 0$. Como δ_τ es par, si $t \geq 0$

$$\begin{aligned} \rho_\tau(-t) &= \int_{-\infty}^{-t} \delta_\tau(s) ds \\ &= \int_t^{+\infty} \delta_\tau(s) ds \\ &= 1 - \rho_\tau(t). \end{aligned}$$

De esto se sigue que $1 - \rho_\tau(-t) = \rho_\tau(t)$ para todo $t \geq 0$ y por tanto

$$\begin{aligned}
\sigma_\tau(x) - [x]^+ &= \int_{-\infty}^0 \rho_\tau(t) dt + \int_0^x \rho_\tau(t) - 1 dt \\
&= \int_{-\infty}^0 \rho_\tau(t) dt + \int_0^{-x} 1 - \rho_\tau(-t) dt \\
&= \int_{-\infty}^0 \rho_\tau(t) dt + \int_0^{-x} \rho_\tau(t) dt \\
&= \int_{-\infty}^{-x} \rho_\tau(t) dt ,
\end{aligned}$$

como buscamos. □

Observación I.5.3. Del Lema I.5.2 se sigue que para todo $x \in \mathbb{R}$

$$\begin{aligned}
0 &\leq \sigma_\tau(x) - [x]^+ \\
&= \int_{-\infty}^{-|x|/\tau} \rho_1(t) dt \\
&\leq \tau \int_{-\infty}^0 \rho_1(t) dt .
\end{aligned}$$

En particular

$$\lim_{\tau \rightarrow 0} \left\{ \sup_{x \in \mathbb{R}} \sigma_\tau(x) - [x]^+ \right\} = 0 .$$

En este sentido $(\sigma_\tau, \tau > 0)$ aproxima a $[\cdot]^+$.

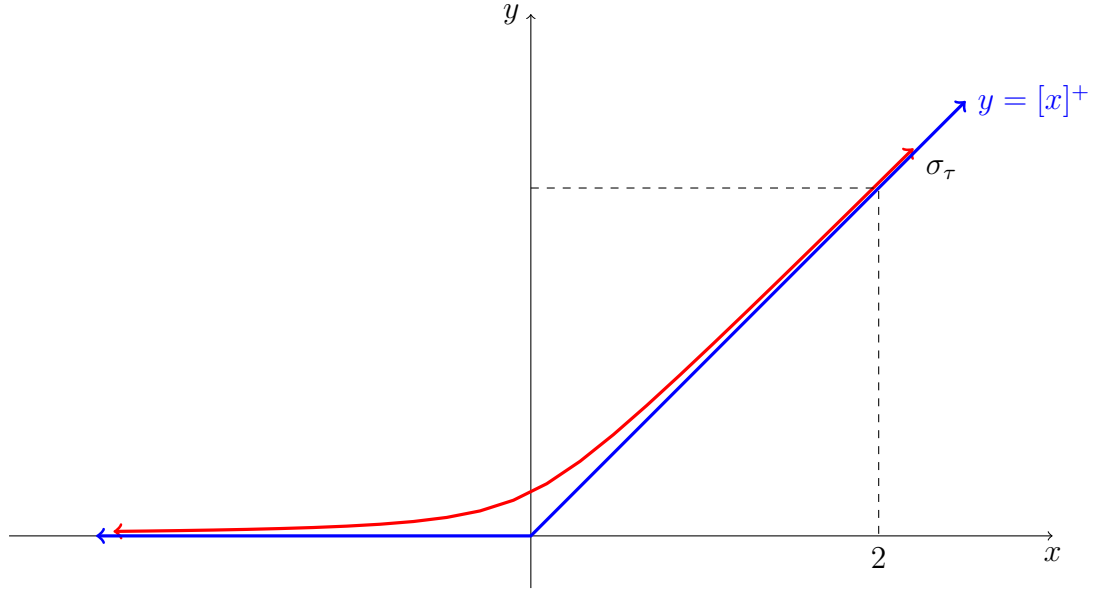


Figura I.5: Aproximación de $[\cdot]^+$ por σ_τ

I.5.1. Casos particulares

Si la familia $(\delta_\tau, \tau > 0)$ es continua, entonces las aproximaciones $(\sigma_\tau, \tau > 0)$ son dos veces diferenciables y convexas. En particular las familias de aproximaciones generadas por

$$\delta_1 = \frac{e^x}{(1 + e^x)^2} \quad \text{y} \quad \delta_1 = \frac{1}{2} \left[\frac{1}{\sqrt{x^2 + 4}} - \frac{x^2}{\sqrt{x^2 + 4}^3} \right],$$

son

$$\sigma_\tau(x) = x + \tau \log(1 + e^{-x/\tau}) \quad \text{con} \quad \sigma'_\tau(x) = \frac{1}{1 + e^{-x/\tau}} \quad \text{y} \quad (\text{I.8})$$

$$\sigma_\tau(x) = \frac{x + \sqrt{x^2 + 4\tau^2}}{2} \quad \text{con} \quad \sigma'_\tau(x) = \frac{1}{2} \left(1 + \frac{x}{\sqrt{x^2 + 4\tau^2}} \right). \quad (\text{I.9})$$

Estas son familias de aproximaciones de la parte positiva comúnmente usadas en aplicaciones prácticas [4], [11], [12], [25]. En la figura I.5 podemos ver la gráfica de la función definida por (I.9), para $\tau = 1/4$.

I.6. Distancia de Wasserstein

El objetivo de esta sección es presentar la distancia de Wasserstein como parte del problema de Monge-Kantorovich y revisar algunas de sus propiedades básicas. Más adelante, en la Sección III.5, usamos una caracterización de W_2 en dimensión 1 para dar un ejemplo de una función de discrepancia.

I.6.1. Tópicos de probabilidad

En esta sección revisaremos algunas cuestiones básicas sobre teoría de probabilidad que esencialmente serán usadas en la Sección I.6.2.

Recordemos que una *probabilidad* μ en un conjunto X dotado de una sigma-álgebra $\mathcal{F}$ es una medida tal que $\mu(X) = 1$. A la tripleta $(X, \mathcal{F}, \mu)$ es común llamarla *espacio de probabilidad*. Cuando no sea necesario, podemos obviar a $\mathcal{F}$ o μ en la mención del espacio de probabilidad. Finalmente, una función medible $F : X \rightarrow Y$ entre dos espacios de probabilidad es llamada *variable aleatoria*.

Sean (X, μ) un espacio de probabilidad, Y un espacio medible y $F : X \rightarrow Y$ una función medible. Entonces la medida inducida en Y por μ y F , denotada por

μF^{-1} , y definida por

$$\mu F^{-1}(A) = \mu(F^{-1}(A)) , \quad \text{para todo conjunto medible } A ,$$

es en realidad una probabilidad llamada *probabilidad inducida* por F en Y .

Si $X = X_1 \times X_2 \times \cdots \times X_n$ es un espacio producto, entonces las funciones $p_k : X \rightarrow X_k$ y $p_{k,l} : X \rightarrow X_k \times X_l$, con $k < l$, definidas por

$$p_k(x_1, x_2, \dots, x_k) = x_k$$

$$p_{k,l}(x_1, x_2, \dots, x_k) = (x_k, x_l) ,$$

son llamadas *proyecciones*. .

Definición I.6.1. Sean (X_k, μ_k) , $k \in \{1, 2, \dots, n\}$, espacios de probabilidad.

Decimos que la probabilidad π en $X = X_1 \times X_2 \times \cdots \times X_n$ es un acoplamiento de $\mu_1, \mu_2, \dots, \mu_n$, si

$$\pi p_k^{-1} = \mu_k \quad \text{para todo} \quad k \in \{1, 2, \dots, n\} .$$

Al espacio de todos los acoplamientos de las probabilidades μ_k lo denotamos por

$$\Pi(\mu_1, \mu_2, \dots, \mu_n).$$

En el caso especial de probabilidades definidas en $\mathbb{R}$ tenemos la siguiente definición.

Definición I.6.2. Si μ es una probabilidad en $\mathbb{R}$, la función $F : \mathbb{R} \rightarrow [0, 1]$ definida por

$$F(x) = \mu((-\infty, x]) , \quad \text{para todo } x \in \mathbb{R} ,$$

es la función de distribución acumulada de μ . En el caso en que μ sea absolutamente continua con respecto a la medida de Lebesgue λ , la densidad de μ respecto de λ es llamada función de densidad o función de distribución de μ .

Notemos que si $f : \mathbb{R} \rightarrow [0, +\infty)$ es la función de densidad de la probabilidad μ de $\mathbb{R}$, para cualquier variable aleatoria $G : \mathbb{R} \rightarrow \mathbb{R}$ μ -integrable vale la igualdad

$$\int_{\mathbb{R}} G(x) d\mu(x) = \int_{\mathbb{R}} G(t) f(t) dt$$

donde la última integral es con respecto a la medida de Lebesgue. Notemos además que si por ejemplo f es continua, el primer teorema fundamental del cálculo implica que $F'(x) = f(x)$ para todo $x \in \mathbb{R}$; donde F es la función de distribución acumulada de μ .

Por otro lado, relacionado con *probabilidad condicional* tenemos la siguiente proposición [5].

Proposición I.6.3 (Desintegración). Sea $F : (X, \mu) \rightarrow (Y, \nu)$ una variable aleatoria entre espacios métricos separables, tal que $\mu F^{-1} = \nu$. Entonces existe una única familia de probabilidades $\{\mu_y, y \in Y\}$ de X , ν -casi ciertamente

determinada tal que

$$\mu_x(X \setminus F^{-1}(y)) = 0 \quad \text{para casi todo } y \in Y ,$$

y

$$\int_X f(x) d\mu(x) = \int_Y \left(\int_{F^{-1}(y)} f(x) d\mu_y(x) \right) d\nu(y) ,$$

para toda variable aleatoria $f : X \rightarrow \mathbb{R}_+$.

En las condiciones de la anterior proposición, la familia de probabilidades $\{\mu_y, y \in Y\}$ se llama *desintegración* de μ con respecto a ν .

Observación I.6.4. En el caso en que $p_1 : (X \times Y, \mu) \rightarrow (X, \nu)$ verifique las condiciones de la proposición I.6.3, en particular $\nu = \mu p_1^{-1}$, podemos identificar canónicamente cada fibra $p_1^{-1}(x)$ con Y . De modo que la familia de probabilidades $\{\mu_x, x \in X\}$ del teorema la podemos considerar sobre Y . En esta situación, obtenemos

$$\int_{X \times Y} f(x, y) d\mu(x, y) = \int_X \left(\int_Y f(x, y) d\mu_x(y) \right) d\nu(x) , \quad (\text{I.10})$$

para toda variable aleatoria $f : X \times Y \rightarrow \mathbb{R}$. Esta situación la podemos resumir escribiendo

$$d\mu(x, y) = d\mu_x(y) d\nu(x) .$$

Similarmente, si η es una probabilidad en Y tal que $\mu p_2^{-1} = \eta$, representamos a la desintegración de μ respecto de η por medio de la igualdad

$$d\mu(x, y) = d\mu_y(x) d\eta(y) .$$

El siguiente resultado nos permite “pegar” dos acoplamientos, si estos son compatibles.

Lema I.6.5. Sean (X_i, μ_i) , $i = 1, 2, 3$, tres espacios de probabilidad. Si $\pi^1 \in \Pi(\mu_1, \mu_2)$ y $\pi^2 \in \Pi(\mu_2, \mu_3)$ son acoplamientos, entonces existe $\pi \in \Pi(\mu_1, \mu_2, \mu_3)$ tal que

$$\pi p_{1,2}^{-1} = \pi^1 \qquad y \qquad \pi p_{2,3}^{-1} = \pi^2 .$$

Demostración. Usando la notación de la Observación I.6.4, tenemos las desintegraciones

$$d\pi^1(x_1, x_2) = d\pi_{x_2}^1(x_1) d\mu_2(x_2) \quad y$$

$$d\pi^2(x_2, x_3) = d\pi_{x_2}^2(x_3) d\mu_2(x_2) .$$

Entonces, es fácil ver la probabilidad π definida por

$$d\pi(x_1, x_2, x_3) = d\pi_{x_2}^1(x_1) d\pi_{x_2}^2(x_3) d\mu_2(x_2) ,$$

satisface las propiedades deseadas. □

Observación I.6.6. *El Lema I.6.5 puede extenderse a $n > 3$ coordenadas en lugar de sólo 3.*

I.6.2. El problema de Monge-Kantorovich

Sean (X, μ) y (Y, ν) dos espacios de probabilidad y $c : Y \times Y \rightarrow \mathbb{R}_+$ una función medible. El *Problema de Monge-Kantorovich* consiste en minimizar la expresión

$$\int_{X \times Y} c(x, y) d\pi(x, y) ,$$

cuando π recorre el conjunto $\Pi(\mu, \nu)$ de todos los acoplamientos de las medidas μ y ν .

Observación I.6.7. *En el contexto del transporte óptimo, el problema de Monge-Kantorovich puede interpretarse como encontrar la forma más óptima, representada por π , de transportar la probabilidad μ en ν , teniendo fijo el costo de transporte c .*

En el caso en que X e Y sean espacios dotados de una métrica y $c : X \times Y \rightarrow \mathbb{R}$ sea una métrica en el producto $X \times Y$, diremos que el problema de Monge-Kantorovich tiene una *función de costo métrica*. En este contexto, en el caso en que $X = Y = \mathbb{R}^n$ y $\|\cdot\|$ represente a la norma euclidiana usual

$$\|(x_1, x_2, \dots, x_n)\| = (x_1^2 + x_2^2 + \dots + x_n^2)^{1/2} ,$$

tenemos la siguiente definición

Definición I.6.8. Sean $p \geq 1$ una constante y μ, ν dos probabilidades en un conjunto compacto $X \subset \mathbb{R}^n$, entonces la p -distancia de Wasserstein, entre μ y ν es

$$W_p(\mu, \nu) = \min_{\pi \in \Pi(\mu, \nu)} \left(\int_{X \times X} \|x - y\|^p d\pi(\mu, \nu) \right)^{1/p}.$$

Observación I.6.9. La distancia de Wasserstein está bien definida, es decir el mínimo en la definición I.6.8 se alcanza [22, Chapter 2], [15, Chapter 2].

Proposición I.6.10. La p -distancia de Wasserstein es una métrica en el conjunto de probabilidades de X que se define.

Demostración. La simetría de W_p es clara, pues el mapa $f : X \times X \rightarrow X \times X$ definido por $f(x, y) = (y, x)$ establece una correspondencia biunívoca entre $\Pi(\mu, \nu)$ y $\Pi(\nu, \mu)$ y además $f \circ c = c$ para la función de costo $c(x, y) = \|x - y\|^p$.

Supongamos que $W_p = 0$. Entonces existe $\gamma \in \Pi(\mu, \nu)$ tal que

$$\int_{X \times X} \|x - y\|^p d\gamma(x, y) = 0.$$

Esto implica que el soporte de γ es la diagonal $D_X = \{(x, x) \in X \times X : x \in X\}$ y por tanto ambas proyecciones de γ son iguales, es decir $\mu = \nu$.

Nos queda demostrar la desigualdad triangular. Sean μ, ν, η probabilidades en X y supongamos que $\pi \in \Pi(\mu, \nu)$ y $\gamma \in \Pi(\nu, \eta)$ sean los acoplamientos que logran $W_p(\mu, \nu)$ y $W_p(\nu, \eta)$. Gracias al Lema I.6.5, existe $\sigma \in \Pi(\mu, \nu, \eta)$ tal que $\sigma p_{1,2}^{-1} = \pi$

y $\sigma p_{2,3}^{-1} = \gamma$, luego

$$\begin{aligned}
W_p(\mu, \eta) &\leq \left(\int_{X^2} \|x - z\|^p d\sigma p_{1,3}^{-1}(x, z) \right)^{1/p} \\
&= \left(\int_{X^3} \|x - z\|^p d\sigma(x, y, z) \right)^{1/p} \\
&\leq \left(\int_{X^3} (\|x - y\| + \|y - z\|)^p d\sigma(x, y, z) \right)^{1/p} \\
&\leq \left(\int_{X^3} \|x - y\|^p d\sigma(x, y, z) \right)^{1/p} + \left(\int_{X^3} \|y - z\|^p d\sigma(x, y, z) \right)^{1/p} \\
&= W_p(\mu, \nu) + W_p(\nu, \eta) .
\end{aligned}$$

Por tanto W_p es una distancia en el espacio de probabilidades de X . $\square$

En dimensión $n = 1$ y cuando $p = 2$, cuando tenemos un *costo cuadrático*, existe la siguiente expresión explícita para W_2 [22, Teorema 2.18], [15, Ejemplo 3.2.14].

Teorema I.6.11. *Sean μ, ν dos probabilidades en un conjunto compacto $X \subset \mathbb{R}$ con F y G sus respectivas funciones de distribución acumuladas. Entonces*

$$W_2 = \left(\int_X |G^{-1}(F(x)) - x|^2 d\mu(x) \right)^{1/2} .$$

Observación I.6.12. 1. *En las condiciones del Teorema I.6.11, por abuso de notación, cuando μ y ν son absolutamente continuas respecto de la medida de Lebesgue, se suele escribir*

$$W_2(f, g) = W_2(\mu, \nu) ,$$

donde f y g son las respectivas funciones de distribución de μ y ν .

2. En la Sección III.5, para nuestra aplicación de discrepancia, únicamente usaremos la expresión explícita del Teorema I.6.11; sin ser muy importante, en los cálculos técnicos ahí realizados, de donde proviene la expresión o que corresponde a una distancia en un espacio de probabilidades.

I.7. Función KL

Definición I.7.1. Sean $a = (a_1, a_2, \dots, a_m) \in \mathbb{R}_+^m$ y $b = (b_1, b_2, \dots, b_m) \in \mathbb{R}_+^m$.

Definimos la función $\text{KL} : \mathbb{R}_+^m \times \mathbb{R}_+^m \rightarrow \mathbb{R}$ como

$$\text{KL}(a, b) = \sum_{i=1}^m \varphi\left(\frac{a_i}{b_i}\right) b_i ,$$

donde $\varphi(p) = p \log(p) - p + 1$, para todo $p > 0$ y $\varphi(0) = 1$.

Observación I.7.2. La función φ es continua, estrictamente convexa, su valor mínimo es igual a cero y se alcanza en $p = 1$.

Proposición I.7.3. La función $\text{KL} : \mathbb{R}_+^m \times \mathbb{R}_+^m \rightarrow \mathbb{R}$ es estrictamente convexa, su mínimo es cero, el que es alcanzado en el conjunto

$$\{(a, b) \in \mathbb{R}_+^m \times \mathbb{R}_+^m : a_i = b_i, i = 1, \dots, m\} .$$

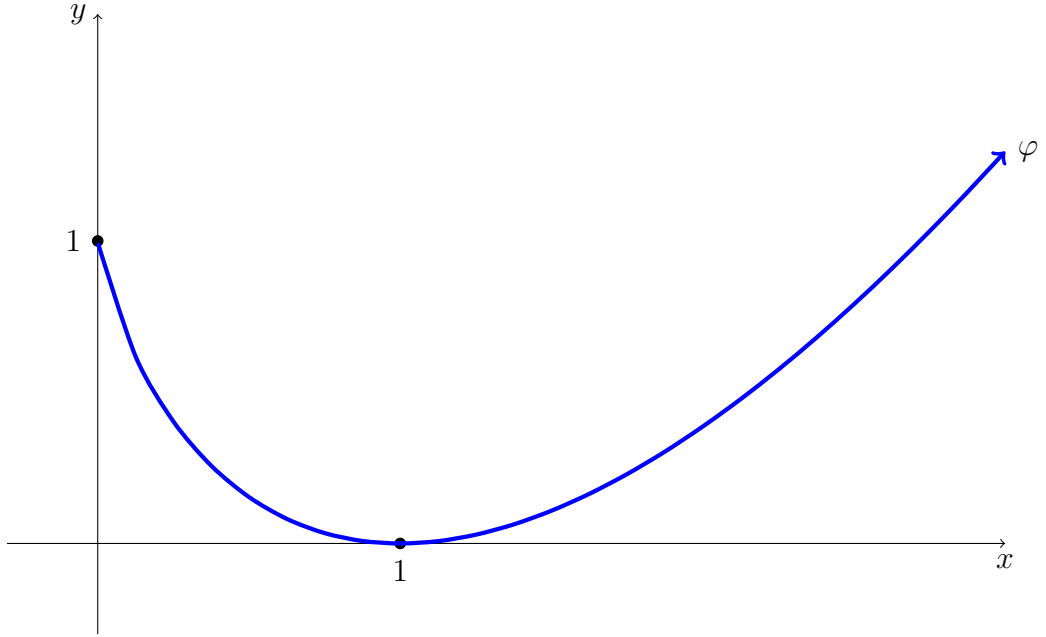


Figura I.6: Gráfica de la función φ

Demostración. Para $m = 1$

$$\begin{aligned} \langle \nabla \text{KL}(a_1, b_1), (\bar{a}_1 - a_1, \bar{b}_1 - b_1) \rangle \\ = (\bar{a}_1 - a_1) \log\left(\frac{a_1}{b_1}\right) + \bar{b}_1 - b_1 - \frac{a_1}{b_1}(\bar{b}_1 - b_1), \end{aligned}$$

luego

$$\begin{aligned} \text{KL}(\bar{a}_1, \bar{b}_1) - \text{KL}(a_1, b_1) - \langle \nabla \text{KL}(a_1, b_1), (\bar{a}_1 - a_1, \bar{b}_1 - b_1) \rangle \\ = \bar{a}_1 \left[\frac{a_1 \bar{b}_1}{b_1 \bar{a}_1} - 1 - \log\left(\frac{a_1 \bar{b}_1}{b_1 \bar{a}_1}\right) \right] > 0, \end{aligned}$$

si $(\bar{a}_1, \bar{b}_1) \neq (a_1, b_1)$. Entonces $(a_1, b_1) \mapsto \text{KL}(a_1, b_1)$ es estrictamente convexa y por tanto, en general, KL también lo es, pues es suma de funciones estrictamente convexas.

La última parte de la proposición es una consecuencia directa de la Observación I.7.2. □

Capítulo II

Transporte óptimo desbalanceado

II.1. Transporte óptimo balanceado

En su planteamiento clásico, el transporte óptimo se realiza entre dos vectores de entradas no negativas $x = (x_1, \dots, x_m), y = (y_1, \dots, y_m) \in \mathbb{R}^m$ de la misma *masa*, es decir, tales que

$$\mathfrak{m}(x) = \sum_{i=1}^m x_i = \sum_{i=1}^m y_i = \mathfrak{m}(y) .$$

En estas condiciones, el *transporte óptimo balanceado* se encarga de resolver el problema de optimización lineal

$$\min_{\Gamma \in \mathcal{S}} \langle C, \Gamma \rangle_F ,$$

donde $\langle \cdot \rangle_F$ es el producto de Frobenius, $C \in \mathbb{R}^{m \times m}$ es una matriz de costos y

$$\mathcal{S} = \{\Gamma = (\gamma_{ij}) : \sum_j \gamma_{ij} = x_i \wedge \sum_i \gamma_{ij} = y_j\}.$$

Observación II.1.1. *En el caso en que x e y sean re-escalados para convertirse en probabilidades, el conjunto $\mathcal{S}$ puede interpretarse como el conjunto de acoplamientos de las probabilidades x e y . De modo que el problema de transporte óptimo es equivalente a encontrar el acoplamiento Γ más eficiente, en el sentido que minimice $\langle C, \Gamma \rangle_F$ entre x e y .*

II.2. Transporte óptimo desbalanceado - UOT

Sin embargo, muchas veces es necesario tener una idea de transporte en el que la masa no es conservada, es decir, en el que $\mathfrak{m}(x) \neq \mathfrak{m}(y)$.

Definición II.2.1. *Sean x e y dos vectores no nulos en $\mathbb{R}^m$, con entradas no negativas y $C \in \mathbb{R}^{m \times m}$ una matriz con entradas, también no negativas. Diremos que C es la matriz de costos del problema de transporte óptimo desbalanceado, UOT,*

$$D_H(x, y) := \min_{\Gamma \geq 0} H(x, y, \Gamma), \quad (\text{II.1})$$

donde

$$H(x, y, \Gamma) := \langle C, \Gamma \rangle_F + r_m \text{KL}(\Gamma \mathbf{1}, x) + r_m \text{KL}(\Gamma^T \mathbf{1}, y) + r \sum_{i=1}^m \text{KL}(\Gamma e_i, \mathbf{1}) , \quad (\text{II.2})$$

y r_m, r son constantes positivas.

En adelante la matriz de costos C y las constantes r_m y r de la Definición II.2.1 permanecerán constantes.

Observación II.2.2. Observe que si $\Gamma = (\gamma_{ij})_m$ entonces

$$r \sum_{i=1}^m \text{KL}(\Gamma e_i, \mathbf{1}) = \sum_{i,j=1}^m \varphi(\gamma_{ij}) .$$

como φ es estrictamente convexa, se sigue que $\Gamma \mapsto r \sum_{i=1}^m \text{KL}(\Gamma e_i, \mathbf{1})$ es estrictamente convexa. Por otro lado, como fue demostrado en la Sección I.7, la función KL es convexa. Luego las aplicaciones $(\Gamma, x) \mapsto \text{KL}(\Gamma \mathbf{1}, x)$ y $(\Gamma, y) \mapsto \text{KL}(\Gamma^T \mathbf{1}, y)$ son convexas. Además $\Gamma \mapsto \langle C, \Gamma \rangle_F$ es convexa, pues es lineal. Luego H es estrictamente convexa y por tanto el problema de encontrar $D_H(x, y)$ tiene como máximo una solución.

Ahora mostraremos que, en el caso en el que la función D_H está bien definida, ella es diferenciable y calcularemos su gradiente.

Primero expresamos $H(x, y, \Gamma)$ en términos de la variable $z := (x, y) \in \mathbb{R}^{2m}$.

Considerando $\widehat{\Gamma} := \text{diag}(\Gamma, \Gamma^T) \in \mathcal{M}_+(2m)$ es fácil ver que

$$H(x, y, \Gamma) = H(z, \Gamma) = \langle C, \Gamma \rangle_F + r_m \text{KL}(\widehat{\Gamma} \mathbf{1}, z) + r \sum_{i=1}^m \text{KL}(\Gamma e_i, \mathbf{1}) . \quad (\text{II.3})$$

Con esta notación $D(x, y) = D(z)$. Más aún, tenemos la siguiente proposición.

Proposición II.2.3. *Sea Γ_z la matriz correspondiente a $D_H(z)$ para todo $z \in \mathbb{R}_+^{2m}$. Entonces D_H es convexa, diferenciable y si $\widehat{\Gamma}_z = \text{diag}(\Gamma_z, \Gamma_z^T)$,*

$$\nabla D_H(z) = \frac{\partial H}{\partial z}(z, \Gamma_z) = r_m \left(\mathbf{1} - \text{diag}\left(\frac{1}{z}\right) \widehat{\Gamma}_z \mathbf{1} \right) . \quad (\text{II.4})$$

Demostración. Sea $\Gamma \in \mathcal{M}_+(m)$. Tenemos que

$$\begin{aligned} H(z, \Gamma) &= H(z, \Gamma_z) + \left\langle \frac{\partial H}{\partial \Gamma}(z, \Gamma_z), (\Gamma - \Gamma_z) \right\rangle + o(\Gamma - \Gamma_z) \\ &\geq H(z, \Gamma_z) . \end{aligned}$$

Luego, para $\Gamma_t = (1-t)\Gamma_z + t\Gamma \in \mathcal{M}_+(m)$ en lugar de Γ en la anterior expresión obtenemos

$$\left\langle \frac{\partial H}{\partial \Gamma}(z, \Gamma_t), (t(\Gamma - \Gamma_z)) \right\rangle + o(t(\Gamma - \Gamma_z)) \geq 0 , \quad \forall t \in (0, 1) .$$

Dividiendo por t y haciendo $t \rightarrow 0$ se deduce la desigualdad

$$\left\langle \frac{\partial H}{\partial \Gamma}(z, \Gamma_z), (\Gamma - \Gamma_z) \right\rangle \geq 0 , \quad \forall \Gamma \in \mathcal{M}_+(m) . \quad (\text{II.5})$$

Por otro lado, como H es convexa

$$\begin{aligned} D_H(w) &= H(w, \Gamma_w) \\ &\geq H(z, \Gamma_z) + \left\langle \frac{\partial H}{\partial z}(z, \Gamma_z), (w - z) \right\rangle \\ &\quad + \left\langle \frac{\partial H}{\partial \Gamma}(z, \Gamma_z), (\Gamma_w - \Gamma_z) \right\rangle . \end{aligned}$$

Entonces, usando (II.5) con $\Gamma = \Gamma_w$, obtenemos que

$$\begin{aligned} D_H(w) &\geq H(z, \Gamma_z) + \left\langle \frac{\partial H}{\partial z}(z, \Gamma_z), (w - z) \right\rangle \\ &= D_H(z) + \left\langle \frac{\partial H}{\partial z}(z, \Gamma_z), (w - z) \right\rangle , \end{aligned}$$

luego D_H es convexa. Ahora supongamos que existe una aplicación $M_z : \mathbb{R}^{2m} \rightarrow \mathbb{R}^{2m}$ tal que

$$D_H(w) \geq D_H(z) + \langle M_z, (w - z) \rangle , \quad \forall w \in \mathbb{R}_+^{2m} .$$

Afirmamos que $M_z = \frac{\partial H}{\partial z}(z, \Gamma_z)$. En efecto, como $H(w, \Gamma_z) \geq D_H(w)$, se sigue que

$$H(w, \Gamma_z) \geq H(z, \Gamma_z) + \langle M_z, (w - z) \rangle , \quad \forall w \in \mathbb{R}_+^{2m} .$$

Entonces, como $w \mapsto H(w, \Gamma_z)$ es diferenciable, convexa y Γ_z es única, gracias a la Observación I.2.5, concluimos que $M_z = \frac{\partial H}{\partial z}(z, \Gamma_z)$. Por tanto D_H es diferenciable

con

$$\frac{\partial D_H}{\partial z}(z) = \frac{\partial H}{\partial z}(z, \Gamma_z) .$$

Finalmente, derivando en (II.3) con respecto a z , obtenemos (II.4). $\square$

Se sigue el siguiente corolario

Corolario II.2.4. *Sean $x, y \in \mathbb{R}_+^m$. Denotamos por Γ_x a la matriz correspondiente a $D_H(x, y)$ cuando y se mantiene fijo, y por Γ_y a la matriz correspondiente a $D_H(x, y)$ cuando x se mantiene fijo. Entonces*

$$\frac{\partial D_H}{\partial x}(x, y) = \frac{\partial H}{\partial x}(x, y, \Gamma_x) = r_m \left[\mathbf{1} - \text{diag} \left(\frac{1}{x} \right) \Gamma_x \mathbf{1} \right] \quad y \quad (\text{II.6})$$

$$\frac{\partial D_H}{\partial y}(x, y) = \frac{\partial H}{\partial y}(x, y, \Gamma_y) = r_m \left[\mathbf{1} - \text{diag} \left(\frac{1}{y} \right) \Gamma_y^T \mathbf{1} \right] . \quad (\text{II.7})$$

Capítulo III

Inversión de ondas

Los problemas que abordamos en nuestro contexto son aquellos en donde tenemos el registro de un *sismograma* (real) $d_0 : [0, T] \rightarrow \mathbb{R}^r$ el cual queremos aproximar por un sismograma teórico $d_v : [0, T] \rightarrow \mathbb{R}^r$, usando una función de discrepancia $F_D(v) = D(d_v, d_0)$.

III.1. Problema directo

En aplicaciones prácticas es imposible conocer la onda real u_0 en toda la malla $\tilde{\Omega}$. Únicamente podemos conocerla en un subconjunto de puntos $\{x_{i_1}, \dots, x_{i_r}\} \subset \tilde{\Omega}$ a los que llamaremos *receptores*. Entonces tenemos que

$$d_{0,k}(t) := u_{0,i_k}(t) \simeq u_0(x_{i_k}, t), \quad \forall k \in \{1, 2, \dots, r\}, \quad \forall t \in [0, T].$$

El *problema directo* consiste en, dado un campo de velocidades v , simular un sismograma u_v que resuelva el sistema (I.3). Esto permite registrar un sismograma

teórico d_v definido por

$$d_{v,k}(t) := u_{v,i_k}(t) \simeq u_v(x_{i_k}, t), \quad \forall k \in \{1, 2, \dots, r\}, \forall t \in [0, T].$$

De este modo, para cada sismograma real, somos capaces de generar una familia de sismogramas teóricos.

III.2. Problema inverso

En la práctica, normalmente, luego de registrar nuestro sismograma real d_0 , no contamos con más información para encontrar el campo de velocidades asociado a él.

El *problema inverso* consiste en determinar el campo de velocidades que genera un sismograma teórico d_v “suficientemente próximo” a d_0 . En la práctica, por limitaciones técnicas, es virtualmente imposible aspirar a que $d_0 = d_v$. Entonces apuntaremos a precisar nuestra idea de proximidad para resolver el problema inverso de acuerdo a ella.

La idea de proximidad entre sismogramas es expresada vía una *función de discrepancia*

$$D : C^2([0, T]; \mathbb{R}^r) \times C^2([0, T]; \mathbb{R}^r) \rightarrow \mathbb{R}_+.$$

De modo que la *discrepancia* F_D entre el sismograma teórico d_v y el real d_0 es dada por

$$F_D(v) = D(d_v, d_0) .$$

En este contexto, el problema inverso puede ser definido de forma más precisa como el problema de optimización sin restricciones

$$\min_v F_D(v) . \tag{III.1}$$

En consecuencia, la solución del problema (III.1) depende únicamente de las propiedades de la *función objetivo* F_D . Naturalmente F_D depende de la discrepancia D y de la dependencia $v \mapsto d_v$. Esta última dependencia se encuentra implícita en la solución de la ecuación de onda (I.3).

Un caso especial, en el que nos enfocamos, es cuando F_D es diferenciable. Pues en este caso es posible aplicar una amplia gama de técnicas de optimización no lineal sin restricciones [13], [1], [8], [17]. En este escenario, un elemento esencial es el cálculo eficiente del gradiente ∇F_D . Este proceso es descrito para una familia grande de funciones de discrepancia en lo que resta del capítulo.

III.3. Cálculo del gradiente

Como hemos discutido anteriormente, el cálculo de $\partial d_v / \partial v$, en la practica, no es sencillo, tanto por la complejidad del proceso de obtención del sismograma,

como por la dimensión alta de d_v .

Una alternativa para superar este problema es usar el método conocido como *método de la adjunta* [21], [3]. En el próximo teorema mostramos una versión bastante general que nos permitirá usarlo en una gran variedad de funciones de discrepancia.

Teorema III.3.1 (Método de la adjunta). *Supongamos que existen funciones $H : \mathbb{R}^m \rightarrow \mathbb{R}^m$ y $Z_v : [0, T] \rightarrow \mathbb{R}^r$ continuas, tales que para todo $i = 1, 2, \dots, m$ se cumple que*

$$\frac{\partial F_D}{\partial v_i} = \int_0^T \left\langle Z_v(t), \frac{\partial d_v}{\partial v_i}(t) \right\rangle dt + H_i(v) , \quad (\text{III.2})$$

entonces el gradiente de F_D puede calcularse por medio de la igualdad

$$\nabla_v F_D(v) = H(v) - 2 \int_0^T \left\langle \lambda_v(t), \text{diag} \left(\frac{1}{v_i^3} \right) u_v''(t) \right\rangle dt , \quad (\text{III.3})$$

donde $\lambda_v \in C^2([0, T], \mathbb{R}^m)$ es solución del sistema

$$\begin{aligned} M(v)\lambda''(t) &= K^T \lambda(t) - S_v^A \\ \lambda'(T) &= 0 \\ \lambda(T) &= 0 , \end{aligned} \quad (\text{III.4})$$

y $S_v^A : [0, T] \rightarrow \mathbb{R}^m$ está definida por

$$S_{v_i}^A(t) = \begin{cases} 0, & \text{si } i \notin \{i_1, i_2, \dots, i_r\} \\ Z_{v,k}(t), & \text{si } i = i_k. \end{cases}$$

En el anterior teorema tenemos que el campo λ_v es conocido como *campo adjunto* y la fuente S_v^A o Z_v que lo genera, normalmente es llamada *fente adjunta*.

El Teorema III.3.1 implica que no importa la naturaleza de la función de discrepancia que genera F_D , siempre se sea posible calcular una fuente adjunta Z_v que satisfaga la expresión (III.2), el gradiente de la función de discrepancia puede ser calculado por medio del campo adjunto λ_v , gracias a la ecuación (III.3).

Antes de demostrar el teorema principal de este trabajo, veremos algunas consecuencias.

III.3.1. Funciones de discrepancia desagregadas por instante de tiempo

Sea $f : \mathbb{R}^r \times \mathbb{R}^r \times \mathbb{R}^m \rightarrow \mathbb{R}_+$ una función de clase C^1 . Definimos la función de discrepancia D_f por medio de la igualdad

$$D_f(d_1, d_2) = \int_0^T f(d_1(t), d_2(t), v) dt. \quad (\text{III.5})$$

En este caso

$$F_f(v) := F_{D_f}(v) = \int_0^T f(d_v(t), d_0(t), v) dt ; \quad (\text{III.6})$$

de donde, usando la Regla de Leibniz, se sigue que

$$\frac{\partial F_f}{\partial v_i}(v) = \int_0^T \left\langle \frac{\partial f}{\partial d_1}(d_v(t), d_0(t), v), \frac{\partial d_v}{\partial v_i}(t) \right\rangle dt + \int_0^T \frac{\partial f}{\partial v_i}(d_v(t), d_0(t), v) dt .$$

Entonces, aplicando el Teorema III.3.1 con

$$\begin{aligned} Z_v(t) &= \frac{\partial f}{\partial d_1}(d_v(t), d_0(t), v) \quad \text{y} \\ H(v) &= \int_0^T \nabla_v f(d_v(t), d_0(t), v) dt , \end{aligned}$$

tenemos el siguiente corolario

Corolario III.3.2. *Si f es de clase C^1 , tenemos que F_f es diferenciable, con*

$$\nabla F_f(v) = \int_0^T \nabla_v f(d_v(t), d_0(t), v) dt - 2 \int_0^T \left\langle \lambda_v(t), \text{diag} \left(\frac{1}{v_i^3} \right) u_v''(t) \right\rangle dt ,$$

donde $\lambda_v \in C^2([0, T], \mathbb{R}^m)$ es solución del sistema

$$M(v)\lambda''(t) = K^T \lambda(t) - S_v^A$$

$$\lambda'(T) = 0$$

$$\lambda(T) = 0 ,$$

con $S_v^A : [0, T] \rightarrow \mathbb{R}^m$ definida por

$$S_{v_i}^A(t) = \begin{cases} 0, & \text{si } i \notin i_1, i_2, \dots, i_r \\ \left[\frac{\partial f}{\partial d_1}(d_v(t), d_0(t), v) \right]_k, & \text{si } i = i_k. \end{cases}$$

Observación III.3.3. Si en el Corolario III.3.2 tenemos que f no depende de v , la función objetivo está definida por $F_D(v) = D_f(d_v, d_0)$. En consecuencia, la fuente adjunta asociada está dada por

$$Z_v(t) = \frac{\partial f}{\partial d_1}(d_v(t), d_0(t)) . \quad (\text{III.7})$$

Este caso cubre una gran familia de funciones de discrepancia con especial interés en las aplicaciones y nos dedicaremos a explorar algunas de ellas.

III.3.2. Demostración del Método de la adjunta

En esta sección demostramos el Teorema III.3.1. Consideremos $v \in \mathbb{R}^m$ y u_v solución de la ecuación de onda (I.3). Definimos la función lagrangiana $L : \mathbb{R}^m \times C^2([0, T], \mathbb{R}^m) \rightarrow \mathbb{R}$ como

$$L(v, \lambda) = F_D(v) + \int_0^T \langle \lambda(t), M(v)u_v''(t) - Ku_v(t) - S(t) \rangle dt .$$

Como u_v satisface (I.3) tenemos que para cualquier v, λ se cumple que $L(v, \lambda) = F_D(v)$. De este modo F_D será diferenciable, si y solo si, L también lo sea con respecto a v .

Por otro lado, integrando por partes,

$$\begin{aligned}
\int_0^T \langle \lambda(t), M(v)u_v''(t) \rangle dt &= \int_0^T \langle M(v)\lambda(t), u_v''(t) \rangle dt \\
&= \left. \langle M(v)\lambda(t), u_v'(t) \rangle \right|_0^T - \int_0^T \langle M(v)\lambda'(t), u_v'(t) \rangle dt \\
&= \langle M(v)\lambda(T), u_v'(T) \rangle - \left[\left. \langle M(v)\lambda'(t), u_v(t) \rangle \right|_0^T \right. \\
&\quad \left. - \int_0^T \langle M(v)\lambda''(t), u_v(t) \rangle dt \right] \\
&= \langle M(v)\lambda(T), u_v'(T) \rangle - \langle M(v)\lambda'(T), u_v(T) \rangle \\
&\quad + \int_0^T \langle M(v)\lambda''(t), u_v(t) \rangle dt .
\end{aligned}$$

Entonces, para cualquier función $\lambda \in C^2([0, T], \mathbb{R}^m)$ que satisfice las condiciones de contorno

$$\lambda(T) = 0$$

$$\lambda'(T) = 0 ,$$

tenemos que

$$L(v, \lambda) = F_D(v) + \int_0^T \langle M(v)\lambda''(t) - K^T \lambda(t), u_v(t) \rangle dt - \int_0^T \langle \lambda(t), S(t) \rangle dt .$$

Por otro lado, de las definiciones de S_v^A , d_v y (III.2) se sigue que

$$\left\langle Z(t), \frac{\partial d_v}{\partial v_i}(t) \right\rangle = \left\langle S_v^A(t), \frac{\partial u_v}{\partial v_i}(t) \right\rangle , \quad \forall i \in \{1, 2, \dots, m\} .$$

Finalmente, consideramos $\lambda_v \in C^2([0, T], \mathbb{R}^m)$ solución del sistema (III.4) y tenemos que

$$L(v, \lambda_v) = F_D(v) + \int_0^T \langle M(v)\lambda_v''(t) - K^T \lambda_v(t), u_v(t) \rangle dt \\ - \int_0^T \langle \lambda_v(t), S(t) \rangle dt .$$

Tomando la derivada parcial con respecto a v_i resulta

$$\begin{aligned} \frac{\partial L}{\partial v_i}(v, \lambda_i) &= H_i(v) + \int_0^T \left\langle Z_v(t), \frac{\partial d_v}{\partial v_i}(t) \right\rangle dt \\ &\quad + \int_0^T \left\langle M(v)\lambda_v''(t) - K^T \lambda_v(t), \frac{\partial u_v}{\partial v_i}(t) \right\rangle dt + \int_0^T \left\langle \left(\frac{\partial M}{\partial v_i}(v) \right) \lambda_v''(t), u_v(t) \right\rangle dt \\ &\quad + \int_0^T \left\langle M(v) \left(\frac{\partial \lambda_v}{\partial v_i} \right)''(t) - K^T \frac{\partial \lambda_v}{\partial v_i}(t), u_v(t) \right\rangle dt - \int_0^T \left\langle \frac{\partial \lambda_v}{\partial v_i}(t), S(t) \right\rangle dt \\ &= H_i(v) + \int_0^T \left\langle \lambda_v(t), \left(\frac{\partial M}{\partial v_i}(v) \right) u_v''(t) \right\rangle dt \\ &\quad + \int_0^T \left\langle \frac{\partial \lambda_v}{\partial v_i}(t), M(v) u_v''(t) - K u_v(t) - S(t) \right\rangle dt \\ &= H_i(v) + \int_0^T \left\langle \lambda_v(t), \left(\frac{\partial M}{\partial v_i}(v) \right) u_v''(t) \right\rangle dt . \end{aligned}$$

Esto nos permite concluir que

$$\nabla_v F_D(v) = H(v) - 2 \int_0^T \left\langle \lambda_v(t), \text{diag} \left(\frac{1}{v_i^3} \right) u_v''(t) \right\rangle dt ,$$

lo que prueba el teorema.

III.4. Discrepancia usando una distancia L^2

Un caso muy frecuente, por su fácil manejo, es el de la función de discrepancia D_f definida por (III.5), cuando f está dada por

$$f(d_1, d_2) = \frac{1}{2} \|d_1 - d_2\|^2 ,$$

donde $\|\cdot\|$ es la norma L^2 en el espacio euclidiano $\mathbb{R}^r$. En este caso f no depende del campo de velocidades, de modo que, de acuerdo a la Observación III.3.3

$$\nabla F_f(v) = -2 \int_0^T \left\langle \lambda_v(t), \text{diag} \left(\frac{1}{v_i^3} \right) u_v''(t) \right\rangle ,$$

donde la fuente adjunta está dada por

$$S_v^A(t) = \begin{cases} 0 , & \text{si } i \notin \{i_1, \dots, i_r\} \\ (d_v(t) - d_0(t))_k , & \text{si } i = i_k . \end{cases}$$

Observación III.4.1. *En este y los demás casos cubiertos por el Teorema III.3.1, para lograr el cálculo del gradiente, de acuerdo a (III.3) únicamente hace falta mostrar la existencia de la fuente adjunta Z_v satisfaciendo (III.2). Por este motivo, para contribuir a la claridad de la explicación, en adelante únicamente nos limitaremos a mostrar la existencia de la fuente adjunta Z_v satisfaciendo (III.2).*

III.5. Discrepancia usando una distancia W_2

Recordemos que en el Teorema I.6.11 vimos que la distancia de Wasserstein W_2 entre dos funciones de distribución de probabilidad $f, g : [0, T] \rightarrow \mathbb{R}$ está dada por

$$W_2(f, g) = \left(\int_0^T |G^{-1}(F(t)) - t|^2 f(t) dt \right)^{1/2},$$

donde F y G son las respectivas funciones de distribución acumulada de f y g .

Supongamos ahora que $d_k : [0, T] \rightarrow \mathbb{R}$, $k \in \{1, 2, \dots, r\}$, sea una coordenada del sismograma $d : [0, T] \rightarrow \mathbb{R}^r$. Entonces elegimos constantes a_k, b_k de modo que

$$\tilde{d}_k = a_k f + b_k : [0, T] \rightarrow \mathbb{R},$$

sea una función de distribución de probabilidad. Es decir, elegimos las constantes de modo que

$$\int_0^T \tilde{d}(t) dt = 1 \quad \text{y} \quad \tilde{d}(t) \geq 0, \quad \forall t \in [0, T].$$

Esto define $\tilde{d} : [0, T] \rightarrow \mathbb{R}^r$ de modo que cada una de sus coordenadas es una función de distribución de probabilidad.

Entonces, si d_v y d_0 son los sismogramas teórico y real, respectivamente,

consideramos sus reescalamientos $\tilde{d}_v$ y $\tilde{d}_0$ de modo que

$$\tilde{d}_{v,k} = \alpha_k d_{v,k} + \beta_k \quad \text{y}$$

$$\tilde{d}_{0,k} = \gamma_k d_{0,k} + \delta_k ,$$

sean funciones de distribución de probabilidad para todo $k \in \{1, 2, \dots, r\}$, para constantes $\alpha_k, \beta_k, \gamma_k$ y δ_k adecuadas.

Definición III.5.1 (Discrepancia de Wasserstein). *Con la notación anteriormente establecida, definimos las funciones objetivo F_W y discrepancia D_W por medio de las igualdades*

$$F_W(v) = D_W(d_v, d_0) = \sum_{k=1}^r W_2^2(\tilde{d}_{v,k}, \tilde{d}_{0,k}) .$$

Observación III.5.2. *De la definición anterior y el tipo de reescalamiento hecho a los sismogramas deducimos que*

$$F_W(v) = D_W(d_v, d_0) = \sum_{k=1}^r \int_0^T |G_k^{-1}(F_{v,k}(t)) - t|^2 \tilde{d}_{v,k} dt ,$$

donde

$$F_{v,k}(t) = \int_0^t \tilde{d}_{v,k}(s) ds \quad \text{y} \quad G_k(t) = \int_0^t \tilde{d}_{0,k}(s) ds .$$

son las funciones de distribución acumulada de cada una de las coordenadas de los sismogramas reescalados $\tilde{d}_v$ y $\tilde{d}_0$, respectivamente.

Esta nueva función de discrepancia, gracias a la Observación III.5.2, cae dentro de la familia de discrepancias cubiertas por el Corolario III.3.2 y la Observación III.3.3. Gracias a esto tenemos la siguiente proposición que hace posible el cálculo del gradiente de la función objetivo F_W vía el método de la adjunta.

Proposición III.5.3. *Continuando con la notación establecida en la Observación III.5.2, el campo adjunto*

$$Z_v^W = (Z_{v,1}^W, Z_{v,2}^W, \dots, Z_{v,r}^W) : [0, T] \rightarrow \mathbb{R}^r ,$$

asociado a la función objetivo F_W está dado por

$$Z_{v,k}^W(t) = \alpha_k \left(|G_k^{-1}(F_{v,k}(t)) - t|^2 + 2 \int_t^T \frac{(G_k^{-1}(F_{v,k}(s)) - s) \tilde{d}_{v,k}(s)}{\tilde{d}_{0,k}(G_k^{-1}(F_{v,k}(s)))} ds \right) ,$$

para todo $k \in \{1, 2, \dots, r\}$.

Demostración. Sean $i \in \{1, 2, \dots, m\}$ y $k \in \{1, 2, \dots, r\}$. De la Observación

III.5.2 se sigue que

$$\begin{aligned}
\frac{\partial W_2^2(\tilde{d}_{v,k}, \tilde{d}_{0,k})}{\partial v_i} &= 2 \int_0^T (G_k^{-1}(F_{v,k}(t)) - t)(G_k^{-1})'(F_{v,k}(t)) \frac{\partial F_{v,k}}{\partial v_i}(t) \tilde{d}_{v,k}(t) dt \\
&\quad + \int_0^T |G_k^{-1}(F_{v,k}(t)) - t|^2 \alpha_k \frac{\partial d_{v,k}}{\partial v_i} dt \\
&= 2 \int_0^T \frac{(G_k^{-1}(F_{v,k}(t)) - t) \tilde{d}_{v,k}(t)}{\tilde{d}_{0,k}(G_k^{-1}(F_{v,k}(t)))} \int_0^t \alpha_k \frac{\partial d_{v,k}(s)}{\partial v_i} ds dt \\
&\quad + \int_0^T \alpha_k |G_k^{-1}(F_{v,k}(t)) - t|^2 \frac{\partial d_{v,k}}{\partial v_i} dt \\
&= 2 \int_0^T \int_0^s \frac{(G_k^{-1}(F_{v,k}(s)) - s) \tilde{d}_{v,k}(s)}{\tilde{d}_{0,k}(G_k^{-1}(F_{v,k}(s)))} \alpha_k \frac{\partial d_{v,k}(t)}{\partial v_i} dt ds \\
&\quad + \int_0^T \alpha_k |G_k^{-1}(F_{v,k}(t)) - t|^2 \frac{\partial d_{v,k}}{\partial v_i} dt \\
&= 2 \int_0^T \alpha_k \int_t^T \frac{(G_k^{-1}(F_{v,k}(s)) - s) \tilde{d}_{v,k}(s)}{\tilde{d}_{0,k}(G_k^{-1}(F_{v,k}(s)))} ds \frac{\partial d_{v,k}}{\partial v_i}(t) dt \\
&\quad + \int_0^T \alpha_k |G_k^{-1}(F_{v,k}(t)) - t|^2 \frac{\partial d_{v,k}}{\partial v_i} dt \\
&= \int_0^T Z_{v,k}^W(t) \frac{\partial d_{v,k}}{\partial v_i} dt ,
\end{aligned}$$

donde

$$Z_{v,k}^W = \alpha_k \left(|G_k^{-1}(F_{v,k}(t)) - t|^2 + 2 \int_0^T \frac{(G_k^{-1}(F_{v,k}(s)) - s) \tilde{d}_{v,k}(s)}{\tilde{d}_{0,k}(G_k^{-1}(F_{v,k}(s)))} ds \right) .$$

Además

$$\begin{aligned}
\frac{\partial F_W}{\partial v_i}(v) &= \sum_{k=1}^r \frac{\partial W_2^2(\tilde{d}_{v,k}, \tilde{d}_{0,k})}{\partial v_i} \\
&= \sum_{k=1}^r \int_0^T Z_{v,k}^W(t) \frac{\partial d_{v,k}}{\partial v_i}(t) dt \\
&= \int_0^T \left\langle Z_v^W(t), \frac{\partial d_v(t)}{\partial v_i} \right\rangle dt .
\end{aligned}$$

Esto muestra, gracias al Teorema III.3.1, que Z_v^W es la fuente adjunta de F_W . $\square$

III.6. Tratamiento de valores negativos

Muchas veces, a diferencia de la discrepancia vista en la Sección III.4, se presenta la necesidad de usar discrepancias como en la ecuación (III.5), pero con funciones f que únicamente admiten vectores con entradas no negativas, o positivas. Esto es imposible de implementar directamente en nuestro caso, porque típicamente las entradas de los sismogramas asumen valores positivos y negativos. Para solucionar este problema existen diferentes alternativas. A continuación enumeramos algunas de ellas.

I. Simplemente trasladar los sismogramas, de modo que

$$\tilde{d}_0 = d_0 + a \quad \text{y} \quad \tilde{d}_v = d_v + b ,$$

sean positivos.

II. Igual que en la aplicación hecha con la distancia W_2 , hacer un re-escalamiento, de modo que además de tener valores positivos, tengamos que

$$\tilde{d}_0 = \alpha d_0 + a \quad \text{y} \quad \tilde{d}_v = \beta d_v + b ,$$

pueden considerarse como distribuciones de probabilidad en cada coordenada. En este caso, podemos usar discrepancias que se benefician

de esto, como lo hicimos en la Sección III.5.

III. En el mismo espíritu que antes, podemos considerar ahora una transformación no lineal que convierta los sismogramas en positivos. Por ejemplo, con expresiones como las siguientes

$$\tilde{d}_{0,k} = \alpha_k \exp(a_k d_{0,k}) \quad \text{y} \quad \tilde{d}_{v,k} = \beta_k \exp(b_k d_{v,k}) , \quad \forall k \in \{1, 2, \dots, r\} ,$$

es posible, además de convertir los sismogramas en positivos, volverlos como en el tratamiento que hicimos en la Sección III.5.

IV. Finalmente, otro abordaje es, esencialmente, repartir cada sismograma en sus partes positiva y negativa, para, de alguna manera, aprovechar la positividad de cada parte para definir una discrepancia.

Todos estos abordajes son válidos y han sido puestos en práctica, pero tenemos limitaciones [11], [25]. En las opciones **I** y **II** tenemos el inconveniente de no saber, a priori, los factores de traslación. Esto genera problemas, pues valores excesivamente grandes dan lugar a malas aproximaciones. Además este tipo de transformaciones es insensible a traslaciones. Es decir, en principio, no hay cómo diferenciar dos sismogramas, donde uno sea una traslación de otro. En la opción **III** tenemos que transformaciones no lineales, no dan el mismo tratamiento a todos los valores que puede asumir un sismograma. Por ejemplo, la función exponencial expande mucho más los valores positivos, en comparación con los valores negativos. Simetrizando la transformación,

planteado algo del tipo

$$\begin{aligned}\tilde{d}_{0,k} &= \alpha_k \exp(a_k d_{0,k}) + \alpha_k \exp(-a_k d_{0,k}) & y \\ \tilde{d}_{v,k} &= \beta_k \exp(b_k d_{v,k}) + \beta_k \exp(-b_k d_{v,k}) ,\end{aligned}$$

para todo $k \in \{1, 2, \dots, r\}$, se resuelve este problema; pero tenemos algo similar, ahora comparando cantidades en una vecindad del origen, con cantidades que tienen un valor absoluto grande. Por estos motivos, para solucionar el cambio de signo en nuestros sismogramas seguiremos un camino en la línea de la opción **IV**. Este camino no está exento de dificultades, por ejemplo, desde el inicio nos enfrentamos al hecho que la función parte positiva no es diferenciable. Así que necesitamos hacer algunas manipulaciones técnicas para sobrellevar esto. Para esto necesitamos algo de notación.

Recordemos la función parte positiva $[\cdot]^+$ definida en la Sección **I.5**. Para un número real x definimos su *parte negativa*, denotada por $[x]^-$, mediante la igualdad $[x]^- = [x]^+ - x$. Dado un vector d de $\mathbb{R}^r$ definimos sus *partes positiva y negativa* como

$$\begin{aligned}[d]^+ &= ([d_1]^+, [d_2]^+, \dots, [d_r]^+) \\ [d]^- &= [-d]^+ .\end{aligned}$$

Naturalmente extendemos la anterior definición a funciones $d : [0, T] \rightarrow \mathbb{R}^r$ por

medio de las igualdades

$$[d]^+(t) = [d(t)]^+ , \quad \forall t \in [0, T]$$

$$[d]^-(t) = [d(t)]^- , \quad \forall t \in [0, T] .$$

Similarmente, para el tipo de funciones σ_τ estudiadas en la sección I.5, extendemos su uso a vectores y funciones de r coordenadas, escribiendo $\sigma_\tau(d)$ para referirnos a la aproximación de su parte positiva.

Sea $\tilde{D} : \mathbb{R}_+^r \times \mathbb{R}_+^r \rightarrow \mathbb{R}$ una función que idealmente alcanza su mínimo cuando sus dos entradas son iguales. Basados en la intuición de querer encontrar v tal que $d_v \simeq d_0$, planteamos

$$[d_v]^+ - [d_v]^- \simeq [d_0]^+ - [d_0]^- ,$$

que es equivalente a

$$[d_v]^+ + [d_0]^- \simeq [d_0]^+ + [d_v]^- .$$

Con esto en mente y queriendo utilizar el enfoque de la ecuación (III.5), nos gustaría emplear $f : \mathbb{R}^r \times \mathbb{R}^r \rightarrow \mathbb{R}$ definida por

$$f(d_1, d_2) = \tilde{D}([d_1]^+ + [d_2]^-, [d_2]^+ + [d_1]^-) .$$

Como mencionamos, la parte positiva no es diferenciable, así que teniendo

presente que

$$[d]^+ \simeq \sigma_\tau(d) \quad \text{y} \quad [d]^- \simeq \sigma_\tau(d) - d, \quad (\text{III.8})$$

tomamos en cuenta

$$f(d_1, d_2) = \tilde{D}(\sigma_\tau(d_1) + [d_2]^-, \sigma_\tau(d_1) - d_1 + [d_2]^+) .$$

Esta función, dentro de la ecuación (III.5), genera una función de discrepancia con todas las propiedades deseadas. Sin embargo, pierde sensibilidad frente a dos sismogramas iguales a menos de una traslación significativa [11].

Por este motivo, en lo que sigue, emplearemos una función $f : \mathbb{R}^r \times \mathbb{R}^r \rightarrow \mathbb{R}$ motivada por

$$f(d_1, d_2) = \tilde{D}([d_1]^+, [d_2]^+) + \tilde{D}([d_1]^-, [d_2]^-) ,$$

para poder usar (III.5) y definir nuestra función de discrepancia. Sin embargo, como ya hemos discutido, tal discrepancia, en general, no sería diferenciable, incluso si $\tilde{D}$ lo fuera, por causa de las partes positivas y negativas.

Entonces, guiados por la intuición de (III.8), definimos

$$f(d_1, d_2) = f_{\tilde{D}}(d_1, d_2) = \tilde{D}(\sigma_\tau(d_1), [d_2]^+) + \tilde{D}(\sigma_\tau(d_1) - d_1, [d_2]^-) . \quad (\text{III.9})$$

De acuerdo con (III.5) y (III.6), esta función define sus respectivas funciones de

discrepancia D_f y función objetivo F_f .

Observación III.6.1. *Con la notación establecida en esta sección y de acuerdo a la Observación III.3.3, la fuente adjunta de la función F_f generada por (III.9) es*

$$\begin{aligned} Z_v(t) = & \frac{\partial \tilde{D}}{\partial x}(\sigma_\tau(d_v(t)), d_0(t)^+) \odot \sigma'_\tau(d_v(t)) \\ & + \frac{\partial \tilde{D}}{\partial x}(\sigma_\tau(d_v(t)) - d_v(t), d_0(t)^-) \odot [\sigma'_\tau(d_v(t)) - \mathbf{1}] , \quad (\text{III.10}) \end{aligned}$$

donde $\odot$ está definido por

$$(a_1, a_2, \dots, a_r) \odot (b_1, b_2, \dots, b_r) = (a_1 b_1, a_2 b_2, \dots, a_r b_r) .$$

Gracias a la aproximación de la parte positiva por σ_τ tenemos la ventaja computacional de obtener valores siempre positivos. Esto nos permite considerar, por ejemplo, funciones de discrepancia donde la función KL esté involucrada, ya que en este caso no habrá problemas con las divisiones, ni con la evaluación de los logaritmos. Veremos dos casos de este tipo en las siguientes secciones.

III.6.1. Discrepancia basada en la función KL

Consideremos, $\tilde{D} \in \mathbb{R}_+^r \times \mathbb{R}_+^r \rightarrow \mathbb{R}$ definida por

$$\tilde{D}(x, y) = \text{KL}(x, y) = \sum_{i=1}^r x_i \log \left(\frac{x_i}{y_i} \right) - x_i + y_i .$$

Es fácil ver que, en este caso, la respectiva fuente adjunta deducida en la ecuación (III.10) es

$$\begin{aligned} Z_v(t) = & \log \left(\frac{\sigma_\tau(d_v(t))}{d_0(t)^+} \right) \odot \sigma'_\tau(d_v(t)) \\ & + \log \left(\frac{\sigma_\tau(d_v(t)) - d_v(t)}{d_0(t)^-} \right) \odot [\sigma'_\tau(d_v(t)) - \mathbf{1}] . \end{aligned}$$

III.6.2. Discrepancia basada en el UOT

En este caso consideramos $\tilde{D} = D_H$, como fue definida en la ecuación (II.1). Es decir, tenemos $\tilde{D}(x, y) = D_H(x, y) = \min_{\Gamma \geq 0} H(x, y, \Gamma)$, para todo $x, y \in \mathbb{R}_+^m$, donde

$$H(x, y, \Gamma) = \langle C, \Gamma \rangle + r_m \text{KL}(\Gamma \mathbf{1}, x) + r_m \text{KL}(\Gamma^T \mathbf{1}, y) + r \sum_{i=1}^m \text{KL}(\Gamma e_i, \mathbf{1}) .$$

Notemos que en este caso mantenemos la dimensión de los vectores igualada a m y no r únicamente para que no cause confusión con la constante r de la definición de H . En este caso la correspondiente fuente adjunta calculada en (III.10) tiene la siguiente forma

$$\begin{aligned} Z_v(t) = & r_m \left[\mathbf{1} - \text{diag} \left(\frac{1}{\sigma_\tau(d_v(t))} \right) \Gamma_{\sigma_\tau(d_v(t))} \mathbf{1} \right] \odot \sigma'_\tau(d_v(t)) \\ & + r_m \left[\mathbf{1} - \text{diag} \left(\frac{1}{\sigma_\tau(d_v(t)) - d_v(t)} \right) \Gamma_{[\sigma_\tau(d_v(t)) - d_v(t)]} \mathbf{1} \right] \\ & \odot [\sigma'_\tau(d_v(t)) - \mathbf{1}] . \end{aligned}$$

III.6.3. Discrepancia simetrizando el UOT

La que describimos aquí es otra discrepancia basada en D_H que suele considerarse a la hora de comparar los sismogramas teórico y real [11], [25].

Sea $\tilde{D} : \mathbb{R}_+^m \times \mathbb{R}_+^m \rightarrow \mathbb{R}$ definida por

$$\tilde{D}(x, y) = D_H(x, y) - \frac{1}{2}[D_H(x, x) + D_H(y, y)] + \frac{r}{2}[\mathbf{m}(x) - \mathbf{m}(y)]^2 .$$

Esta función es la *simetrización* de D_H . A pesar de ser menos manejable que D_H , tiene la ventaja de ser más eficiente [25]. En este caso la correspondiente fuente adjunta calculada en (III.10) es

$$\begin{aligned} Z_v(t) = & \left[r_m A(\sigma_\tau(d_v(t))) \mathbf{1} + r[\mathbf{m}(\sigma_\tau(d_v(t))) - \mathbf{m}(d_0(t)^+)] \mathbf{1} \right] \odot \sigma'_\tau(d_v(t)) \\ & + \left[r_m A(\sigma_\tau(d_v(t)) - d_v(t)) \mathbf{1} + r[\mathbf{m}(\sigma_\tau(d_v(t)) - d_v(t)) - \mathbf{m}(d_0(t)^-)] \mathbf{1} \right] \\ & \odot [\sigma'_\tau(d_v(t)) - \mathbf{1}] , \end{aligned}$$

donde $A(x)$ es la matriz definida por

$$A(x) = \text{diag} \left(\frac{1}{x} \right) \left[\frac{1}{2} (\Gamma_{(x,x)} + \Gamma_{(x,x)}^T) - \Gamma_x \right] .$$

En la última expresión, las matrices $\Gamma_{(x,x)}$ y Γ_x son las mencionadas en la Proposición II.2.3 y su Corolario II.2.4.

Conclusiones

- El problema de inversión de ondas presenta diversas dificultades técnicas entre las que resaltan el cálculo del gradiente de la función objetivo, en el caso en que esta sea diferenciable, y la elección de una buena discrepancia para comparar los sismogramas.
- La primera de estas dificultades es resuelta en esta tesis, en una amplia variedad de casos, con posibilidades de aplicaciones prácticas, mediante el *Método de la Adjunta* que demostramos.
- El empleo de discrepancias basadas en el UOT es viable. Hemos logrado calcular todas las fuentes adjuntas de las discrepancias propuestas, lo que en teoría permite su implementación en un potencial escenario de aplicación.
- Para una posible implementación práctica de la teoría presentada en este trabajo es necesario un estudio numérico apropiado, para que las fuentes adjuntas den los resultados esperados. Es necesario calibrar las constantes involucradas y establecer una discretización óptima.

Índice alfabético

- acoplamiento, 22
- campo adjunto, 42
- campo de onda, 11
- convexo, 6
- costo cuadrático, 28
- costo de transporte, 26
- desintegración, 24
- discrepancia, 40
- discrepancia de Wasserstein, 49
- discretización, 12
- distancia de Wasserstein, 27
- ecuación de onda, 11
- espacio de acoplamientos, 22
- espacio de probabilidad, 21
- fuelle, 11
- fuelle adjunta, 42
- función convexa, 7
- función de costo métrica, 26
- función de densidad, 23
- función de discrepancia, 39
- función de distribución, 23
- función de distribución acumulada, 23
- función estrictamente convexa, 7
- función objetivo, 40
- masa, 32
- matriz de costos, 33
- modelo de velocidades, 11
- método de la adjunta, 41
- parte negativa, 54
- parte positiva, 16, 54

probabilidad, 21	simetrización, 59
probabilidad inducida, 22	sismograma, 38
problema de Monge-Kantorovich, 26	subdiferencial, 9
problema directo, 38	teorema fundamental del cálculo, 2
problema inverso, 39	UOT, 33
proyecciones, 22	variable aleatoria, 21
receptores, 38	

Bibliografía

- [1] BONNANS, J.-F., GILBERT, J. C., LEMARÉCHAL, C., AND SAGASTIZÁBAL, C. A. *Numerical optimization: theoretical and practical aspects*, 2 ed. Universitext. Springer, Berlin Heidelberg, 2006.
- [2] BUNKS, C., SALECK, F. M., ZALESKI, S., AND CHAVENT, G. Multiscale seismic waveform inversion. *Geophysics* 60 (1995), 1457–1473.
- [3] CAO, Y., LI, S., PETZOLD, L., AND SERBAN, R. Adjoint sensitivity analysis for differential-algebraic equations: The adjoint dae system and its numerical solution. *SIAM journal on scientific computing* 24 (2003), 1076–1089.
- [4] CHEN, X. Smoothing methods for nonsmooth, nonconvex minimization. *Math. Program.* 134 (2012), 71–99.
- [5] DELLACHERIE, C., AND MEYER, P.-A. *Probabilities and Potential*, 1 ed. North-Holland Mathematics Studies. North-Holland Publishing Co., Amsterdam, 1978.

- [6] EVANS, L. C., AND GARIEPY, R. F. *Measure Theory and Fine Properties of Functions, Revised Edition*. Textbooks in Mathematics. CRC Press, 2015.
- [7] FICHTNER, A. *Full Seismic Waveform Modelling and Inversion*, 1 ed. Advances in Geophysical and Environmental Mechanics and Mathematics. Springer, Berlin Heidelberg, 2010.
- [8] IZMAILOV, A., AND SOLODOV, M. *Otimização, volume 2: Métodos Computacionais*, 3 ed. IMPA, Rio de Janeiro, 2018.
- [9] LIMA, E. L. *Curso de Análise vol.1*, 15 ed. IMPA, Rio de Janeiro, 2019.
- [10] LIMA, E. L. *Curso de Análise vol.2*, 12 ed. IMPA, Rio de Janeiro, 2020.
- [11] MÉTIVIER, L., ALLAIN, A., BROSSIER, R., MÉRIGOT, Q., OUDET, E., AND VIRIEUX, J. Optimal transport for mitigating cycle skipping in full-waveform inversion: A graph-space transform approach. *Geophysics* 83 (2018), R515–R540.
- [12] MÉTIVIER, L., BROSSIER, R., MÉRIGOT, Q., OUDET, E., AND VIRIEUX, J. Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion. *Geophysical Journal International* 205 (2016), 345–377.
- [13] NOCEDAL, J., AND WRIGHT, S. *Numerical optimization*, 2 ed. Springer Series in Operations Research and Financial Engineering. Springer, New York, NY, 2006.

- [14] PRATT, R. G. Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model. *Geophysics* 64 (1999), 888–901.
- [15] RACHEV, S. T., AND RÜSCHENDORF, L. *Mass Transportation Problems, Volume 1: Theory*, 1 ed. Probability and Its Applications. Springer Science+Business, New York, 1998.
- [16] ROCKAFELLAR, R. T. *Convex Analysis*, 1 ed. Princeton Paperbacks. Princeton University Press, Princeton, NJ, 1970.
- [17] RUSZCZYNSKI, A. *Nonlinear optimization*, 1 ed. Princeton University Press, Princeton, NJ, 2006.
- [18] SANTAMBROGIO, F. *Optimal Transport for Applied Mathematicians*, 1 ed. Progress in Nonlinear Differential Equations and Their Applications. Birkhäuser Basel, Berlin Heidelberg, 2015.
- [19] SYMES, W. W. Migration velocity analysis and waveform inversion. *Geophysical Prospecting* 56 (2008), 765 – 790.
- [20] TARANTOLA, A. Inversion of seismic reflection data in the acoustic approximation. *Geophysics* 49 (1984), 1259–1266.
- [21] TROMP, J., TAPE, C., AND LIU, Q. Tomography, adjoint methods, time-reversal, and banana-doughnut kernels. *Geophysical Journal International* 160 (2005), 195–216.

- [22] VILLANI, C. *Topics in Optimal Transportation*, 1 ed. Graduate Studies in Mathematics. American Mathematical Society, Providence, Rhode Island, 2003.
- [23] VILLANI, C. *Optimal Transport: Old and New*, 1 ed. Grundlehren der mathematischen Wissenschaften. Springer-Verlag , Berlin Heidelberg, 2009.
- [24] VIRIEUX, J., AND OPERTO, S. An overview of full-waveform inversion in exploration geophysics. *Geophysics* 74 (2009), WCC127–WCC152.
- [25] YU, M. *Entropic Unbalanced Optimal Transport: Application to Full-Waveform Inversion and Numerical Illustration*. PhD thesis, Université de Paris, 2021.