

Introducción a la teoría del transporte óptimo para economía y finanzas

Marcelo Gallardo

&

Carlos Cosentino

Prefacio

El objetivo del presente documento es presentar algunas de las aplicaciones de la teoría del transporte óptimo en economía y finanzas, con énfasis en los métodos computacionales. En los últimos años, esta teoría ha sido utilizada en numerosas investigaciones en áreas como matching theory, teoría financiera y econometría ([Glasserman and Yang, 2015](#); [Galichon, 2016](#); [Victor Chernozhukov and Galichon, 2022](#); [Federico Echenique, Fedor Sandomirskiy and Joseph Root, 2024](#)). En este trabajo, presentamos problemas de emparejamiento y la estimación de estructuras de costos o beneficios en situaciones específicas. Además, abordamos el problema del acotamiento del *Credit Value Adjustment* y de medidas de riesgo, una aplicación relevante en finanzas. Para establecer el contexto de estudio, introducimos las herramientas matemáticas esenciales de la teoría del transporte óptimo en un anexo.

Iniciamos este documento presentando el problema de Kantorovich como base fundamental para los capítulos posteriores. Luego, nos enfocamos en el problema de la regularización entrópica para el caso discreto. Enseguida, en línea con el problema de regularización entrópica, introducimos el novedoso algoritmo SISTA ([Dupuy et al., 2021](#)) y sus aplicaciones. Continuamos con los problemas de emparejamiento en el mercado laboral ([Dupuy and Galichon, 2022](#)) y matrimonial ([Dupuy and Galichon, 2014](#)). Pasamos luego a una aplicación reciente presentada en [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#), la cual involucra la noción de estabilidad del matching. Finalmente, en los capítulos finales, nos enfocamos en las aplicaciones en finanzas: el problema del acotamiento del *Credit Value Adjustment* siguiendo a [Glasserman and Yang \(2015\)](#) y la regularización entrópica en problemas financieros, relacionados con medidas de riesgo dependientes de dos factores de riesgo ([Memartoluie, 2017](#)) y el riesgo de modelo ([Glasserman and Xu, 2014](#)).

Dado que los temas abordados en este trabajo pertenecen al estado del arte, nos enfocamos en desarrollar los conceptos teóricos y sus implicaciones en lugar de incluir ejercicios. Las demostraciones que se dejan para el lector u omiten son aquellas que, o bien son altamente técnicas y extensas, o bien no son constructivas ni especialmente didácticas.

En el anexo incluimos resultados fundamentales de análisis convexo, análisis funcional y teoría de la medida, buscando que el documento sea lo más auto-suficiente

posible. Sin embargo, es importante destacar que se requiere un conocimiento previo en análisis real, álgebra lineal y optimización en espacios vectoriales de dimensión finita, con la finalidad de comprender adecuadamente el material. Recomendamos a los lectores fortalecer estos conceptos con, por ejemplo, [Abbott \(2015\)](#), [Axler \(2014\)](#), [Billingsley \(1995\)](#) y [Aliprantis and Border \(2006\)](#).

Este documento está dirigido a alumnos de economía, finanzas o matemáticas con interés en la investigación en teoría microeconómica, específicamente en matching theory, o teoría financiera avanzada. Presentamos desarrollos recientes en la teoría del transporte óptimo y referencias a artículos clave, con el objetivo de que estudiantes interesados puedan replicar proyectos basados en este material. La importancia de esta metodología radica en su creciente aplicación en economía y teoría económica, con trabajos pioneros de economistas como por ejemplo Alfred Galichon, Arnaud Dupuy, Bernard Salanié, Federico Echenique y Fedor Sandomirskiy. En particular, la teoría del transporte óptimo ha sido utilizada en el diseño de modelos de emparejamiento, la inferencia estructural en econometría y el análisis de precios de activos en mercados financieros.

La contribución de este documento no es presentar en sí nuevos resultados, si no presentar de manera más accesible resultados ya existentes en la literatura que son estado del arte.

Finalmente, esperamos que estas notas sean de interés y utilidad para estudiantes de la Pontificia Universidad Católica del Perú y, en general, para investigadores y estudiantes de países hispanohablantes que deseen explorar estas herramientas en economía y finanzas.

Marcelo Gallardo

Asistente de investigación y de docencia

y

Carlos Cosentino

Asistente de docencia

Agradecimientos

En primer lugar, expresamos nuestro agradecimiento al profesor Johel Beltrán Ramírez, quien dictó el curso Temas de Análisis (Optimal Transport). Este trabajo tuvo su origen como proyecto final de dicho curso, y gran parte del contenido relacionado con la teoría de Kantorovich está basado en sus clases.

Asimismo, extendemos nuestro reconocimiento al profesor Percy Fernández, quien nos introdujo a numerosos resultados fundamentales del análisis funcional, incluyendo el Teorema de Riesz-Fréchet y los Teoremas de Hahn-Banach, los cuales son referenciados a lo largo de este documento. El Anexo de Análisis Funcional se basa esencialmente en sus notas de clase.

Finalmente, agradecemos al profesor Jorge Chávez, cuyas enseñanzas en el curso Investigación de Operaciones sirvieron de base para varios desarrollos en el apéndice de análisis convexo.

Índice general

Índice general	4
1. Introducción a la teoría del transporte óptimo	1
2. Regularización entrópica	11
2.1. El problema de regularización entrópica	12
2.2. Solución vía dualidad del Lagrangiano	17
2.3. Algoritmo Sinkhorn-Knopp	20
3. Estimación de costos y algoritmo SISTA	25
3.1. El modelo	25
3.2. La estructura de costos paramétrica	27
3.3. SISTA	29
3.3.1. Convergencia del algoritmo SISTA	30
3.4. Estudio de la migración	41
4. Aplicaciones en matching	43
4.1. Marriage market	45
4.2. Estimación y parametrización	52
4.3. Estimación de beneficios laborales y productividad laboral	56
4.4. Mismatching ocupacional y el mercado laboral peruano	60
5. Estabilidad del matching vía Optimal Transport	63
5.1. Emparejamiento en la recta	66
5.2. Estabilidad, equidad y transporte óptimo	69

5.3. Bienestar y equidad en emparejamientos estables	78
6. Acotamiento del CVA	82
6.1. Acotamiento por aproximaciones empíricas	86
6.2. Regularización entrópica del CVA	90
6.3. Existencia y unicidad de la solución	93
6.4. Algoritmo de Sinkhorn para la regularización entrópica del CVA	96
6.4.1. Convergencia del método	97
7. Acotamiento del riesgo de un portafolio	99
7.1. Medidas de riesgo, VaR y CVaR	99
7.2. Value at Risk (VaR)	100
7.3. Conditional Value at Risk (CVaR)	101
7.4. Medidas que dependen de dos factores de riesgo	102
7.5. Riesgo de modelo	104
8. Conclusiones	109
Anexo	111
A. Fundamentos de topología y análisis real	112
B. Teoría de la medida	116
B.1. Teoría de la probabilidad	123
C. Elementos de análisis convexo	128
D. Elementos de análisis funcional	133
E. Subgradientes y subdiferencial	140
Bibliografía	148

Capítulo 1

Introducción a la teoría del transporte óptimo

Dedicamos este primer capítulo a un breve resumen de la teoría de transporte óptimo. Para abordar esta sección, es altamente recomendable consultar el Anexo de este documento. En efecto, es necesario dominar las definiciones básicas de Análisis Real, Topología y Teoría de la Medida.

La teoría del transporte óptimo nace como un problema de redistribución de masa. Suponga que tenemos dos espacios de medida¹ (\mathcal{X}, μ) y (\mathcal{Y}, ν) . El primer espacio de medida indica la distribución inicial de una cierta masa, mientras que el segundo espacio dicta la distribución final que debe seguir. A su vez, transportar masa de un punto $x \in \mathcal{X}$ a un punto $y \in \mathcal{Y}$ viene con un costo $c(x, y)$. Por motivos obvios, llamamos a $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$ la función costo. Lógicamente, queremos minimizar el costo total del proceso de redistribución. ¿Por qué este problema es de interés y tan importante? Dejando de lado que el resto de este trabajo se dedica a presentar aplicaciones en economía y finanzas, brindamos dos aplicaciones bastante intuitivas: una en el caso en el cual la masa es continua y otra en el caso en el cual la masa es discreta. En ambos casos considere que la función costo es proporcional a la distancia, como usualmente ocurre con los costos de transporte.

1. El problema de redistribución de tierra, originalmente estudiado por Monge en

¹Dejaremos siempre el σ -álgebra implícito. En el caso de espacios métricos consideraremos el σ -álgebra de Borel.

el siglo XVIII y considerado el primer problema de transporte óptimo, consiste en la siguiente situación. Se tiene una pila de tierra con una forma dada $\mathcal{X} \subset \mathbb{R}^3$ y, en otra locación, un hueco en el suelo del mismo volumen que ocupa el espacio $\mathcal{Y} \subset \mathbb{R}^3$. Monge se pregunta por la mejor manera de transportar la pila de tierra y así, llenar el hueco. La tierra está uniformemente distribuida antes y después del transporte². Capturamos esta información con las medidas $\mu = \text{Uniforme}(\mathcal{X})$ y $\nu = \text{Uniforme}(\mathcal{Y})$.

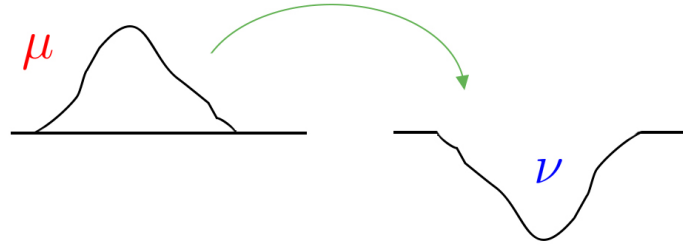


Figura 1.1 El problema original de Monge.

2. El problema de reasignación de soldados. Consideremos las posiciones (ciudades) $X = \{x_1, x_2, x_3, \dots, x_m\}$, cada una con una cierta cantidad de soldados, μ_1, \dots, μ_m respectivamente. Supongamos que deseamos transportarlos a las ciudades $Y = \{y_1, y_2, y_3, y_4, \dots, y_n\}$, de forma que en las ciudades de Y cuenten con ν_1, \dots, ν_n soldados, respectivamente. Ciertamente la cantidad de soldados en las ciudades de X debe coincidir, con la cantidad de soldados que se necesitan en las ciudades de Y . Esto es, $\sum_{i=1}^m \mu_i = \sum_{j=1}^n \nu_j$. Podemos también definir una matriz de costos que representa el costo de enviar un soldado de x_i a y_j (justamente esta es la función de costos implícita del problema):

²La densidad de tierra es igual en todos lados en la pila y en el hueco una vez que lo llenamos.

(x, y)	y_1	y_2	\cdots	y_n
x_1	c_{11}	c_{12}	\cdots	c_{1n}
x_2	c_{21}	c_{22}		\vdots
\vdots			\ddots	
x_m		\cdots		c_{mn}

Un «plan de transporte», en este contexto, es una matriz cuyas entradas representan la cantidad de soldados asignados para ir de x_i a y_j (entrada i, j):

(x, y)	y_1	y_2	\cdots	y_n
x_1	π_{11}	π_{12}	\cdots	π_{1n}
x_2	π_{21}	π_{22}		\vdots
\vdots			\ddots	
x_m		\cdots		π_{mn}

Luego de normalizar por la masa de individuos total $M = \sum_{i=1}^m \mu_i$, el problema de minimización que compete a las autoridades planificadores es

$$\min_{\pi \in \mathcal{M}_{m \times n}(\mathbb{R}_+)} \left\{ \sum_{i=1}^m \sum_{j=1}^n c_{ij} \pi_{ij} \right\},$$

sujeto a que la cantidad adecuada de soldados sea transportada y $\pi \in [0, 1]$.

Concretamente

$$\begin{aligned} & \min_{\pi_{ij} \geq 0} \sum_{i=1}^m \sum_{j=1}^n c_{ij} \pi_{ij} \\ & \text{s.a} \sum_{j=1}^n \pi_{ij} = \mu_i / M, \quad \forall i \\ & \sum_{i=1}^m \pi_{ij} = \nu_j / M, \quad \forall j. \end{aligned}$$

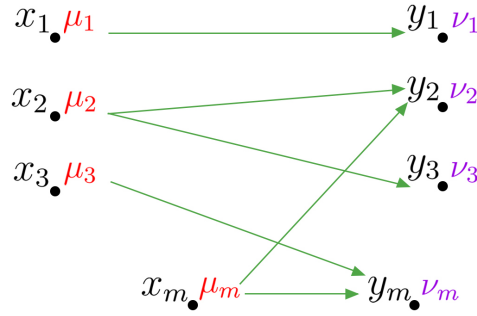


Figura 1.2 Ejemplo de plan de transporte para el problema de distribución de tropas. Una flecha verde representa que se manda algún soldado desde esa locación a la ciudad indicada.

Note que, dado que π_{ij} representa un número de personas, deberíamos considerar que $\pi \in \mathbb{Z}_+^{mn}$. Este es el mismo problema que surge en el problema de maximización de la utilidad, en el cual se asumen bienes infinitamente divisibles. Afortunadamente, resulta que dada la estructura del problema en el caso discreto, $\pi \in \mathcal{M}_{m \times n}(\mathbb{Z}_+)$. No obstante, esto no es el caso por lo general para las extensiones, como la regularización cuadrática.

En general, la redistribución puede no ser tan sencilla como una asignación determinista $T : \mathcal{X} \rightarrow \mathcal{Y}$. Si bien este enfoque podría ser adecuado en el problema de Monge, donde cada partícula de tierra es trasladada a un punto específico del destino, no es aplicable en nuestro segundo ejemplo: los soldados de un mismo campamento pueden ser enviados a distintas ubicaciones.

Por lo tanto, es necesario recurrir a una formulación más general, basada en distribuciones de probabilidad. Denotamos por $\mathcal{P}(S)$ el conjunto de todas las medidas de probabilidad sobre un espacio medible S .

Definición 1.0.1. Sean \mathcal{X}, \mathcal{Y} espacios medibles y $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ medidas de probabilidad. Decimos que $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ es un *acoplamiento* de μ y ν cuando

$$\pi(A \times Y) = \mu(A) \quad \wedge \quad \pi(X \times B) = \nu(B)$$

para cada $A \subset \mathcal{X}, B \subset \mathcal{Y}$ medibles. En este caso, decimos también que π es un transport plan entre μ y ν , o que tiene como marginales a μ en \mathcal{X} y a ν en \mathcal{Y} . Desde

ahora, denotamos el conjunto de acoplamientos por

$$\Pi(\mu, \nu) := \{\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : \pi \text{ es acoplamiento de } \mu, \nu\}.$$

Trabajaremos con espacios de probabilidad para normalizar las masas, de modo que $\mu(\mathcal{X}) = \nu(\mathcal{Y}) = 1$. Por otro lado, pensamos en $\pi(A \times B)$ como la cantidad de masa transportada de $A \subset \mathcal{X}$ hasta $B \subset \mathcal{Y}$. La idea de usar acoplamientos para caracterizar una redistribución de masa se justifica por las siguientes dos intuiciones:

1. Para cada $A \subset \mathcal{X}$, el total de masa transportada desde A hacia \mathcal{Y} es la cantidad de masa que se encontraba en A inicialmente. O sea $\pi(A \times \mathcal{Y}) = \mu(A)$.
2. Para cada $B \subset \mathcal{Y}$, el total de masa transportada desde \mathcal{X} hasta B es el total de masa que se encuentra en B luego de la redistribución. Esto es $\pi(\mathcal{X} \times B) = \nu(B)$.

Observación. El conjunto de acoplamientos $\Pi(\mu, \nu)$ siempre es no vacío, pues la medida producto $\mu \times \nu$ tiene marginales μ, ν . Por otro lado, el lector puede verificar con facilidad la convexidad de este conjunto.

Teorema 1.0.2. Sean \mathcal{X}, \mathcal{Y} espacios polacos y sean $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ y $S = \mathcal{X} \times \mathcal{Y}$. Entonces $\Pi(\mu, \nu)$ es compacto en $\mathcal{P}(S)$ respecto a la topología inducida por la métrica

$$d_P(\mu, \nu) = \inf \{ \varepsilon > 0 : \forall A \in \mathcal{B}_S : \mu(A) \leq \nu(A^\varepsilon) + \varepsilon, \nu(A) \leq \mu(A^\varepsilon) + \varepsilon \}$$

donde

$$A^\varepsilon = \{x \in S : d(x, A) < \varepsilon\}$$

y \mathcal{B}_S denota los borelianos sobre S .

La compacidad es un resultado más complicado de demostrar que nos desviaría del propósito de este documento, invitamos al lector interesado revisar la prueba en [Ambrosio et al. \(2021\)](#).

Lema 1.0.3. Sea $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ tal que $f(x, y) = g(x)$ (solo depende de $x \in \mathcal{X}$) para todo $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Entonces, para todo $\pi \in \Pi(\mu, \nu)$ tenemos

$$\int_{\mathcal{X} \times \mathcal{Y}} f(x, y) d\pi = \int_{\mathcal{X}} g(x) d\mu.$$

Análogamente, cuando f solamente depende de $y \in \mathcal{Y}$ se integra sobre (\mathcal{Y}, ν) y se obtiene el resultado análogo.

Definición 1.0.4. Sean \mathcal{X}, \mathcal{Y} espacios medibles, $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ medidas de probabilidad y $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$ una función costo. Para cada transport plan $\pi \in \Pi(\mu, \nu)$ podemos definir el costo total de transporte por el valor esperado de la función costo sobre la medida π :

$$\mathbb{E}_\pi[c] := \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi.$$

El problema de Kantorovich es encontrar π^* tal que

$$\mathbb{E}_{\pi^*}[c] = \inf_{\pi \in \Pi(\mu, \nu)} \mathbb{E}_\pi[c].$$

Diremos en este caso que $\pi^* \in \Pi(\mu, \nu)$ es un transport plan óptimo.

Observación. En el caso de conjuntos \mathcal{X}, \mathcal{Y} finitos, digamos con $|\mathcal{X}| = N$ y $|\mathcal{Y}| = M$, las medidas de probabilidad $\mu \in \mathcal{P}(\mathcal{X})$ y $\nu \in \mathcal{P}(\mathcal{Y})$ están únicamente determinadas por sus valores en cada punto $x \in X$ o $y \in Y$. Así, podemos identificarlas con vectores $\mu = (\mu_x)_{x \in \mathcal{X}} \in \mathbb{R}^N$ y $\nu = (\nu_y)_{y \in \mathcal{Y}} \in \mathbb{R}^M$ con entradas no-negativas sumando 1. Como $|\mathcal{X} \times \mathcal{Y}| = N \times M$, por el mismo argumento, una medida de probabilidad $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ puede ser vista como una matriz $\pi = (\pi_{xy})_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \in \mathcal{M}_{N \times M}(\mathbb{R})$ con entradas no-negativas y tales que sus columnas y filas sumen 1 (doblemente estocásticas). Más aún, el conjunto de acoplamientos se puede caracterizar por:

$$\Pi(\mu, \nu) = \left\{ \pi : \pi_{xy} \geq 0, \sum_{y \in \mathcal{Y}} \pi_{xy} = \mu_x, \forall x \in \mathcal{X} \wedge \sum_{x \in \mathcal{X}} \pi_{xy} = \nu_y, \forall y \in \mathcal{Y} \right\}. \quad (1.1)$$

En el caso discreto, el problema de Kantorovich se vuelve un problema de programación lineal:

$$\mathcal{P}_L : \begin{cases} \inf & \sum_{x,y} \pi_{x,y} c(x, y) \\ \text{s.a} & \sum_{y \in \mathcal{Y}} \pi_{x,y} = \mu_x, \forall x \in \mathcal{X} \\ & \sum_{x \in \mathcal{X}} \pi_{x,y} = \nu_y \forall y \in \mathcal{Y} \\ & \pi_{xy} \geq 0. \end{cases} \quad (1.2)$$

Dado que

$$\Pi(\mu, \nu) = \left\{ \pi_{xy} \in \mathbb{R}_+ : \sum_{y \in \mathcal{Y}} \pi_{x,y} = \mu_x \wedge \sum_{x \in \mathcal{X}} \pi_{x,y} = \nu_y \right\}$$

es un conjunto compacto en \mathbb{R}_+^{NM} , y la función objetivo es continua, podemos asegurar una solución al problema \mathcal{P}_L . Usualmente, para resolver \mathcal{P}_L , se emplean técnicas de programación lineal.

Observación. Note que se verifica la convexidad del conjunto de acoplamientos óptimos gracias a la linealidad del funcional $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \rightarrow \mathbb{E}_\pi(c) \in [-\infty, +\infty]$. Sin embargo, no podemos asegurar que dicho conjunto es no vacío sin imponer alguna condición sobre la función costo. El siguiente teorema garantiza su no vacuidad para una familia *grande* de funciones costo. La definición de función semicontinua superior/inferior, que se usa en el siguiente teorema, se encuentran en el anexo.

Teorema 1.0.5. Sean (\mathcal{X}, μ) y (\mathcal{Y}, ν) espacios de probabilidad polacos, y sean $a : \mathcal{X} \rightarrow \mathbb{R} \cup \{-\infty\}$ y $b : \mathcal{Y} \rightarrow \mathbb{R} \cup \{-\infty\}$ funciones semicontinuas superiores, además de ser μ -integrable y ν -integrable, respectivamente. Sea $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$ una función de costo semicontinua inferior que satisface la desigualdad $c(x, y) \geq a(x) + b(y)$ para todo $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Entonces, existe un plan de transporte óptimo $\hat{\pi} \in \Pi(\mu, \nu)$ que minimiza el costo de transporte.

Demostración. Ver el teorema 4.1 en Villani (2009). \square

Observación. La cota inferior para c nos permite asegurar, gracias al lema 1.0.3 que todo plan de transporte $\pi \in \Pi(\mu, \nu)$ nos da un costo total $\mathbb{E}_\pi[c] \in \mathbb{R} \cup \{+\infty\}$:

$$\begin{aligned} \mathbb{E}_\pi[c] &= \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi \\ &> \int_{\mathcal{X} \times \mathcal{Y}} (a(x) + b(y)) d\pi \\ &= \int_{\mathcal{X}} a(x) d\mu + \int_{\mathcal{Y}} b(y) d\nu \\ &> -\infty. \end{aligned}$$

Sin embargo, aún puede ocurrir que el costo de transporte óptimo sea $+\infty$. Es decir, que cada transport plan nos de costo infinito. Para descartar esta posibilidad, podríamos acotar superiormente c de la misma manera. Otra manera de garantizar que el valor óptimo en el problema de Kantorovich sea finito es imponer ciertas condiciones en los marginales μ y ν . Cuando $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$, en muchos casos resulta útil para nuestros propósitos imponer condiciones de cuadrado-integrabilidad. Esto es, imponer que $\int_{\mathbb{R}^d} \|x\|^2 d\mu < +\infty$ y/o $\int_{\mathbb{R}^d} \|y\|^2 d\nu < \infty$.

Definición 1.0.6. Definimos el espacio $\mathcal{P}_2(\mathbb{R}^d)$ como el conjunto de las medidas de probabilidad cuadrado-integrables en \mathbb{R}^d . Esto es:

$$\mathcal{P}_2(\mathbb{R}^d) = \{\mu \in \mathcal{P}(\mathbb{R}^d) : \mathbb{E}_\mu[\|x\|^2] < +\infty\}.$$

El siguiente resultado nos será útil en los capítulos dedicados a las aplicaciones en finanzas. Presentamos una versión bastante reducida del Teorema de Brenier, adaptado de [Ambrosio et al. \(2021\)](#). Este teorema concierne al problema de Kantorovich en el espacio euclideo, considerando un costo cuadrático. Para consultar el enunciado completo, así como la prueba, ver el teorema 5.2 de dicho libro. A continuación, denotamos por \mathcal{L} a la medida de Lebesgue sobre \mathbb{R}^d .

Teorema 1.0.7. Sean $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$, la función costo $c(x, y) = \frac{1}{2}|x - y|^2$ y $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ tal que $\mu \ll \mathcal{L}$. Entonces existe un único transport plan que minimiza el costo de transporte $\hat{\pi} \in \Pi(\mu, \nu)$.

Observación. En el teorema anterior, la existencia de un transport plan óptimo se deduce del teorema 1.0.5 tomando en cuenta que $c(x, y) \geq 0$. Además, dicho plan óptimo da un costo total finito, pues $c(x, y) = \frac{1}{2}|x - y|^2 \leq |x|^2 + |y|^2$ y μ, ν son cuadrado-integrables. La novedad está en la unicidad de esta solución.

En ciertos contextos, no tenemos control directo sobre el plan de transporte π , por lo que nos interesa conocer el costo total $\mathbb{E}_\pi[c]$ en el peor de los casos (el ínfimo). En otros escenarios, la función c representa una utilidad, en cuyo caso el objetivo es maximizarla. Ambos problemas pueden reformularse de manera equivalente.

Por ejemplo, si c corresponde a una utilidad, denotada por Φ , el problema consiste en encontrar

$$\sup_{\pi \in \Pi(\mu, \nu)} \mathbb{E}_\pi[\Phi],$$

así como, en caso de existir, un plan óptimo π^* que alcance dicho supremo.

A continuación, presentamos un resultado análogo al teorema 1.0.5, estableciendo condiciones bajo las cuales el conjunto de planes óptimos no es vacío.

Teorema 1.0.8. Sean $(\mathcal{X}, \mu), (\mathcal{Y}, \nu)$ espacios de probabilidad polacos, y a, b dos funciones semicontinuas inferior, $a : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ μ -integrable y $b : \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$ ν -integrable. Sea $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{-\infty\}$ una función costo semicontinua superior tal que $c(x, y) \leq a(x) + b(y)$ para todo $x \in \mathcal{X}, y \in \mathcal{Y}$. Entonces existe un plan $\hat{\pi} \in \Pi(\mu, \nu)$ que maximiza el costo de transporte.

Demostración. Considere $C = -c, A = -a, B = -b$ y aplique el teorema 1.0.5 para

obtener un $\hat{\pi} \in \Pi(\mu, \nu)$ tal que

$$\begin{aligned}\mathbb{E}_{\hat{\pi}}[C] &= \inf_{\pi \in \Pi(\mu, \nu)} \mathbb{E}_{\pi}[C] \\ -\mathbb{E}_{\hat{\pi}}[c] &= \inf_{\pi \in \Pi(\mu, \nu)} -\mathbb{E}_{\pi}[c] = - \sup_{\pi \in \Pi(\mu, \nu)} \mathbb{E}_{\pi}[c].\end{aligned}$$

Finalmente, multiplicando por -1 se obtiene el resultado. \square

Iniciamos nuestra discusión presentando el problema de transporte de masas, un enfoque determinístico que dio origen a la teoría del transporte óptimo. No obstante, el problema de Kantorovich admite también una interpretación probabilística, como veremos a continuación.

Lema 1.0.9. Considere un espacio de probabilidad (Ω, \mathbb{P}) y dos variables aleatorias $X : \Omega \rightarrow \mathcal{X}$ y $Y : \Omega \rightarrow \mathcal{Y}$ con distribuciones de probabilidad $\mu \in \mathcal{P}(\mathcal{X})$ y $\nu \in \mathcal{P}(\mathcal{Y})$, respectivamente. Considere el par $(X, Y) : \Omega \rightarrow \mathcal{X} \times \mathcal{Y}$ como un vector aleatorio. Entonces:

1. El par (X, Y) sigue una distribución $\pi \in \Pi(\mu, \nu)$, es decir, su distribución conjunta tiene marginales μ y ν .
2. X y Y son independientes si y solo si $\pi = \mu \times \nu$.

Demostración. Para todo $A \subset \mathcal{X}$ y $B \subset \mathcal{Y}$ medibles, se tiene que:

$$\begin{aligned}\pi(A \times \mathcal{Y}) &= \mathbb{P}\{(X, Y) \in A \times \mathcal{Y}\} = \mathbb{P}\{X \in A\} = \mu(A), \\ \pi(\mathcal{X} \times B) &= \mathbb{P}\{(X, Y) \in \mathcal{X} \times B\} = \mathbb{P}\{Y \in B\} = \nu(B).\end{aligned}$$

Esto prueba la primera afirmación.

Por definición, X y Y son independientes si y solo si, para todo $A \subset \mathcal{X}$ y $B \subset \mathcal{Y}$ medibles, se cumple que:

$$\pi(A \times B) = \mathbb{P}\{(X, Y) \in A \times B\} = \mathbb{P}\{X \in A\} \mathbb{P}\{Y \in B\} = \mu(A) \nu(B).$$

Es decir, si y solo si π es la medida producto $\mu \times \nu$. Esto concluye la demostración. \square

Las variables aleatorias X y Y representan factores con distribuciones individuales conocidas, pero con una distribución conjunta $\pi \in \Pi(\mu, \nu)$ desconocida. En otras palabras, no conocemos su grado de dependencia ni su estructura de correlación

exacta. Para realizar cálculos con estas variables, consideramos la variable aleatoria $c(X, Y) : \Omega \rightarrow \mathbb{R}$, donde $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ es una función medible. Nos interesa su valor esperado:

$$\mathbb{E}_{\mathbb{P}}[c(X, Y)] = \int_{\Omega} c(X, Y) d\mathbb{P} = \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi = \mathbb{E}_{\pi}[c].$$

Más específicamente, la teoría del transporte óptimo proporciona herramientas para acotar este resultado, es decir, encontrar:

$$\inf_{\pi \in \Pi(\mu, \nu)} \mathbb{E}_{\pi}[c] \quad \text{o} \quad \sup_{\pi \in \Pi(\mu, \nu)} \mathbb{E}_{\pi}[c],$$

según el contexto del problema. Pensar en el problema de Kantorovich bajo esta interpretación será útil en los capítulos 6 y 7, donde exploramos aplicaciones en finanzas. En particular, las variables aleatorias X y Y representarán factores de riesgo con distribuciones marginales conocidas, pero distribución conjunta desconocida. Nuestro objetivo será establecer cotas para métricas financieras ampliamente utilizadas que dependen simultáneamente de estos factores.

Con esto concluimos nuestra breve presentación de la teoría del transporte óptimo. Como se mencionó previamente, no es nuestro objetivo desarrollar en detalle la teoría del transporte óptimo, sino, más bien, enfocarnos en sus aplicaciones en economía y finanzas. Comenzamos con aplicaciones en economía, específicamente con una versión modificada del problema de transporte óptimo conocida como *regularización entrópica*. Esta formulación, bajo ciertos supuestos, puede abordarse mediante herramientas de optimización estática y convexa clásica, lo que facilita su interpretación y análisis.

Capítulo 2

Regularización entrópica

En este capítulo vamos a considerar por lo general conjuntos finitos \mathcal{X}, \mathcal{Y} , donde $|\mathcal{X}| = |\mathcal{Y}| = N$ ¹ y establecemos las medidas de probabilidad sobre \mathcal{X} y \mathcal{Y} respectivamente

$$\mu = \sum_{x \in \mathcal{X}} \mu_x \delta_x \text{ y } \nu = \sum_{y \in \mathcal{Y}} \nu_y \delta_y. \quad (2.1)$$

En (2.1), δ_a denota el Delta de Dirac, $\delta_a: \Theta \rightarrow \mathbb{R}$, con regla de correspondencia

$$\delta_a(\theta) = \begin{cases} 1, & \text{si } \theta = a \\ 0, & \text{caso contrario.} \end{cases} \quad (2.2)$$

En este caso, el conjunto de acoplamientos viene dado por (1.1).

El problema clásico consiste en resolver

$$\min_{\pi \in \Pi(\mu, \nu)} \sum_{x, y} c_{xy} \pi_{xy}, \quad (2.3)$$

donde $c_{xy} = c(x, y)$.

Para resolver el problema (2.3), generalmente se requieren métodos numéricos. Por ejemplo, el algoritmo de subastas de Bertsekas tiene una complejidad algorítmica de $\mathcal{O}(N^3)$ (Merigot and Thibert, 2020). Además, es sabido que la solución se encuentra en un vértice de $\Pi(\mu, \nu)$ (Tardella, 2010). Por lo tanto, no es viable asumir soluciones interiores ni emplear métodos de búsqueda diseñados para tales casos.

¹Estamos asumiendo que tienen la misma cardinalidad por simplicidad. De manera más general, podemos suponer $|\mathcal{X}| = N$ y $|\mathcal{Y}| = M$.

Una situación que puede surgir en esta formulación es que el plan de transporte sea concentrado; es decir, existen subconjuntos $I = \{i_1, \dots, i_k\}$ y $J = \{j_1, \dots, j_k\}$ tales que π_{xy} está concentrada en $I \times J$ (es decir, $\pi_{xy} = 0$ para todo $(x, y) \notin I \times J$). Esta concentración puede presentar desventajas desde una perspectiva práctica y teórica (Peyré and Cuturi, 2019).

2.1. El problema de regularización entrópica

Con el objetivo de encontrar soluciones más dispersas, se han introducido métodos de regularización, tanto cuadrática (Nutz, 2024) como entrópica. A continuación, estudiaremos este último en detalle.

Definición 2.1.1. Definimos la función de entropía (relativa) $H(\pi)$ de la siguiente manera

$$H(\pi) = - \sum_{x,y} h(\pi_{x,y}) \quad (2.4)$$

con

$$h(r) = \begin{cases} r(\ln(r) - 1), & \text{si } r > 0 \\ 0, & \text{si } r = 0 \\ +\infty, & \text{si } r < 0. \end{cases}$$

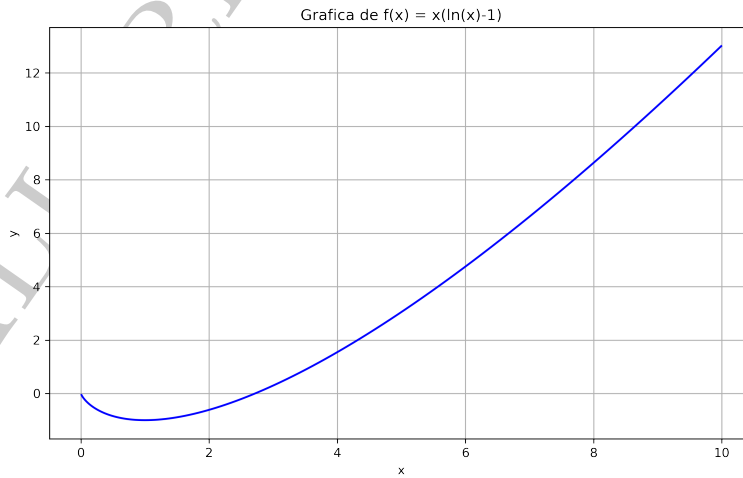


Figura 2.1 $h(x) = x(\ln x - 1)$

Observación. Esta definición (2.4) difiere de la entropía usual de Shannon:

$$-\sum_{i,j} p_{ij} \ln(p_{ij}).$$

$-H(\pi)$ se puede expresar como

$$1 - \sum_{x,y} \pi_{xy} \ln(\pi_{xy}) = - \sum_{x,y} \int_0^{\pi_{xy}} \ln t dt \geq 0 \quad (2.5)$$

Observación. La entropía es un concepto vinculado al desorden, a la incertidumbre de la información. En concreto, dado un proceso con posibles resultados $\{a_1, \dots, a_N\}$ y probabilidades de ocurrencia $\{p_1, \dots, p_N\}$, la entropía se máxima cuando $p_i = 1/N$, o sea, cuando los eventos son equiprobables. En efecto, al resolver

$$\begin{cases} \text{máx} & -\sum_{i=1}^N p_i (\ln p_i - 1) \\ \text{s.a.} & \sum_{i=1}^N p_i = 1 \end{cases}$$

por el método de los multiplicadores de Lagrange (dada la concavidad de la función objetivo y que solo hay una restricción lineal, es condición suficiente),

$$\mathcal{L}(p_1, \dots, p_N, \lambda) = - \sum_{i=1}^N p_i (\ln p_i - 1) + \lambda \left(1 - \sum_{i=1}^N p_i \right).$$

$$\frac{\partial \mathcal{L}}{\partial p_i} = -\ln p_i - \lambda = 0 \quad \forall i \implies p_i = p_j = \frac{1}{N}, \quad \forall i, j.$$

De este modo, mientras menos concentración de la probabilidad de los estados, la entropía será mayor. Esto permitirá entonces obtener soluciones *menos concentradas*.

Definición 2.1.2. El problema de regularización entrópica corresponde al siguiente problema de optimización:

$$\mathcal{P}_\varepsilon : \min_{\pi \in \Pi(\mu, \nu)} \sum_{xy} \pi_{x,y} c(x, y) - \varepsilon H(\pi), \quad (2.6)$$

con $\varepsilon > 0$.

Algunos comentarios:

- Ahora, se busca minimizar el costo sumado con

$$1 - \sum_{x,y} \pi_{xy} \ln \pi_{xy}.$$

Es decir, se busca un transport plan que simultáneamente maximice la entropía y minimice el costo de transporte.

- La función objetivo es continua sobre $\Pi(\mu, \nu)$, que es compacto y no vacío. Por lo tanto, se escribe mín y no ínf.
- El problema de regularización es convexo. Esto se establece en el siguiente lema.

Lema 2.1.3. La función $-H : \pi \in (\mathbb{R}_{+*})^{N \times N} \rightarrow \sum_{x,y} h(\pi_{xy})$ es estrictamente convexa.

Demostración. Como $\pi_{xy} > 0$, $h''(t) = \frac{1}{t} > 0$. Luego, la matriz Hessiana es

$$H(-H(\pi)) = \begin{pmatrix} \frac{1}{\pi_{11}} & & \\ & \ddots & \\ & & \ddots & \\ & & & \frac{1}{\pi_{NN}} \end{pmatrix},$$

la cual es definida positiva. □

Ejemplo 2.1.4. A continuación presentamos la gráfica de una función estrictamente convexa a modo ilustrativo. En la figura 2.2 se aprecia que f posee un único mínimo global.

Function: $f(x, y) = x^2 + y^2 + \exp(x^2 + y^2) + x + y$

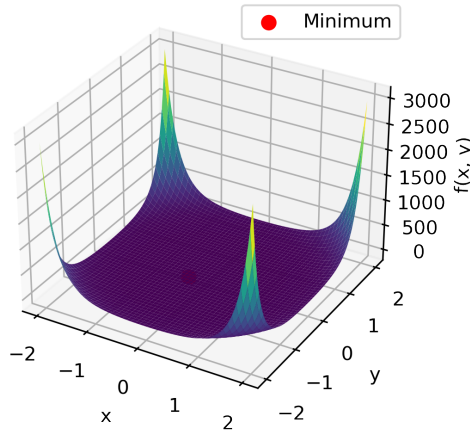


Figura 2.2 Función estrictamente convexa: $f(x, y) = x^2 + y^2 + e^{x^2+y^2} + x + y$.

Teorema 2.1.5. El problema \mathcal{P}_ε tiene una única solución $\pi^* \in \Pi(\mu, \nu)$ y, si $\min\{\min_x \mu_x, \min_y \nu_y\} > 0$, entonces $\pi_{xy}^* > 0, \forall x \in \mathcal{X}, y \in \mathcal{Y}$.

Demostración. Aseguramos directamente la existencia de la solución por la continuidad sobre el compacto $\Pi(\mu, \nu)$. Veamos entonces que $\pi_{xy}^* > 0$. Supongamos que no sea el caso, es decir, que

$$Z = \{(x, y) : \pi_{xy}^* = 0\} \neq \emptyset.$$

Tomemos²

$$\gamma^\theta \in \Gamma(\mu, \nu) : \quad \gamma^\theta = (1 - \theta)\pi^* + \theta(\mu \times \nu), \quad \forall \theta \in (0, 1).$$

Usando la convexidad de $h : r \rightarrow r \ln(r) - r$, se obtiene:

$$h(\gamma_{xy}^\theta) \leq (1 - \theta)h(\pi_{xy}^*) + \theta h(\mu_x \nu_y) \leq h(\pi_{xy}^*) + \underbrace{O(\theta)}_{\theta \sum_{x,y} \mu_x \nu_y (\ln \mu_x + \ln \nu_y - 1)}. \quad (2.7)$$

Entonces, dado $(x, y) \in Z$ y

$$C = \min\{\min_x \mu_x, \min_y \nu_y\} > 0,$$

se tiene la siguiente situación:

$$h(\gamma_{xy}^\theta) = h(\theta \mu_x \nu_y) = \theta \mu_x \nu_y (\ln \theta + \ln \mu_x \nu_y - 1) \leq C \theta \ln \theta + O(\theta). \quad (2.8)$$

Sumando adecuadamente (2.7) y (2.8), obtenemos:

$$\underbrace{\sum_{x,y} h(\gamma_{xy}^\theta)}_{-H(\gamma^\theta)} \leq \underbrace{\sum_{x,y} h(\pi_{xy}^*)}_{-H(\pi^*)} + nC\theta \ln \theta + O(\theta),$$

donde $|Z| = n$. Luego, como $1 - \theta < 1$, se tiene que:

$$\sum_{x,y} \gamma_{xy}^\theta c_{xy} \leq \sum_{x,y} \pi_{xy}^* c_{xy} + \theta \sum_{x,y} c_{xy} \mu_x \nu_y.$$

Aprovechando la linealidad de la estructura de costos y la optimalidad de π^* , se sigue que:

$$\begin{aligned} \sum_{x,y} \pi_{xy}^* c_{xy} - \varepsilon H(\pi^*) &\leq \sum_{x,y} \gamma_{xy}^\theta c_{xy} - \varepsilon H(\gamma^\theta) \\ &\leq \sum_{x,y} \pi_{xy}^* c_{xy} - \varepsilon H(\pi_{xy}^*) + O(\theta) + nC\theta \ln \theta. \end{aligned}$$

²Recordemos que la medida producto es $\mu \times \nu = \sum_{x,y} \mu_x \nu_y \delta_{x,y}$.

Por lo tanto,

$$0 \leq \theta \left(\underbrace{\sum_{x,y} (c_{xy} \mu_x \nu_y + h(\mu_x \nu_y))}_{\text{cte}} + nC \underbrace{\ln \theta}_{<0} \right), \quad \forall \theta \in (0, 1).$$

Esto solo es posible para cualquier configuración si $n = 0$, lo que implica que $\pi_{xy}^* > 0$ (es decir, $Z = \emptyset$).

Finalmente, esto asegura la unicidad de la solución por el lema 2.1.3. En efecto, supongamos que existen dos soluciones π^* y π^{**} , y consideremos la combinación convexa:

$$\pi^{***} = \theta \pi^* + (1 - \theta) \pi^{**}.$$

- π^{***} satisface las restricciones del problema.
- Por la convexidad estricta:

$$\begin{aligned} \sum_{x,y} (\pi_{xy}^{***} c(x, y)) - \varepsilon H(\pi^{***}) &< \theta \sum_{x,y} (\pi_{xy}^* c(x, y)) - \varepsilon H(\pi^*) \\ &+ (1 - \theta) \sum_{x,y} (\pi_{xy}^{**} c(x, y)) - \varepsilon H(\pi^{**}) \\ &= \min(\mathcal{P}_\varepsilon), \end{aligned}$$

lo cual es una contradicción. Esto concluye la prueba. \square

Teorema 2.1.6. La única solución π_ε a \mathcal{P}_ε converge a la solución de entropía minimal en el conjunto de soluciones óptimas al problema original (2.3). Esto es,

$$\lim_{\varepsilon \rightarrow 0} \pi_\varepsilon \rightarrow \operatorname{argmin} \{ -H(\pi) : \pi \in \Pi(\mu, \nu) \text{ plan óptimo} \}. \quad (2.9)$$

Demostración. Sea $\{\varepsilon_k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+$ una sucesión tal que $\varepsilon_k \rightarrow 0$. Sea π_k la solución asociada a $\mathcal{P}_{\varepsilon_k}$. Como $\Pi(\mu, \nu)$ es acotado, podemos extraer una subsucesión convergente, digamos $\pi_{k_j} \rightarrow \pi^*$. Al ser $\Pi(\mu, \nu)$ cerrado, $\pi^* \in \Pi(\mu, \nu)$. Sea ahora π la solución al problema desregularizado. Por la optimalidad de π_{k_j} y π

$$\begin{aligned} 0 &\leq \sum_{x,y} \pi_{k_j}(x, y) c(x, y) - \sum_{x,y} \pi_{xy} c(x, y) \\ &\leq \varepsilon_{k_j} (H(\pi_{k_j}) - H(\pi)). \end{aligned}$$

Dado que $-H(\cdot)$ es una función continua y acotada por $N \ln N$, tomando $\lim_{k_j \rightarrow \infty}$, se llega a que

$$\sum_{x,y} \pi_{xy}^* c(x, y) = \sum_{x,y} \pi_{xy} c(x, y).$$

Más aún, dividiendo por ε_k y tomando límite, $-H(\pi) \geq -H(\pi^*)$. Esto demuestra que π^* resuelve (2.3). El conjunto de acoplamientos óptimos es convexo, y por la convexidad estricta de $-H(\cdot)$, esta solución es única y, finalmente, tenemos por la unicidad del límite que toda la sucesión converge a π^* . \square

2.2. Solución vía dualidad del Lagrangiano

El contexto hasta el momento es el siguiente

$$\begin{cases} \text{mín} & \sum_{x,y} (\pi_{xy} c(x, y)) - \varepsilon H(\pi) \\ \text{s.a} & \sum_{x \in \mathcal{X}} \pi_{xy} = \nu_y \quad \forall y \in \mathcal{Y} \\ & \sum_{y \in \mathcal{Y}} \pi_{xy} = \mu_x, \quad \forall x \in \mathcal{X} \\ & \pi_{xy} \geq 0. \end{cases} \quad (2.10)$$

A continuación, un resultado de optimización convexa que nos permite caracterizar la solución al problema (2.10).

Proposición 2.2.1. Condiciones de Slater. En relación al siguiente problema de optimización:

$$\begin{cases} \text{mín}_x & f_0(x) \\ \text{s.a.} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{cases} \quad (2.11)$$

si f_0 es convexa, las restricciones son lineales y se cumple que $\exists x : f_i(x) < 0, Ax = b$ (se cumplen las restricciones lineales), entonces

$$\max_{\lambda, \eta} \min_x \mathcal{L}(x, \lambda, \eta) = \min_{x \in S} f_0(x)$$

donde S es el conjunto determinado por las restricciones y

$$\mathcal{L}(x, \lambda, \eta) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \eta_i h_i(x).$$

Demostración. Ver [Luneberger and Ye \(2021\)](#). □

Observación. El resultado (proposición 2.2.1) es análogo para máximo - concavidad.

En el caso del problema de transporte considerado, $h_i(x) = 0$ son las restricciones que debe cumplir π para que tenga marginales μ y ν , y $f_i(x) = -\pi_{xy} \leq 0$ es equivalente a $\pi_{xy} \geq 0$. Ciertamente existe un plan factible que cumple $-\pi_{xy} < 0$ (para $\min\{\mu_x, \nu_y\} > 0$), que es la medida producto $\pi_{xy} = \mu_x \nu_y > 0, \forall (x, y)$.

Ahora bien, con respecto a \mathcal{P}_ε , el Lagrangiano asociado es \mathcal{L} :

$$\begin{aligned} \mathcal{L}(\pi, \varphi, \psi) = & \sum_{x,y} \pi_{xy} (c(x, y) + \varepsilon (\ln \pi_{xy} - 1)) \\ & + \sum_{x \in \mathcal{X}} \varphi(x) \left(\mu_x - \sum_{y \in \mathcal{Y}} \pi_{xy} \right) + \sum_{y \in \mathcal{Y}} \psi(y) \left(\nu_y - \sum_{x \in \mathcal{X}} \pi_{xy} \right). \end{aligned}$$

Las variables $\{\varphi(x)\}_{x \in \mathcal{X}}$ y $\{\psi(y)\}_{y \in \mathcal{Y}}$ son los multiplicadores de Lagrange.

Teorema 2.2.2. La solución al problema \mathcal{P}_ε es de la forma

$$\pi_{xy} = \exp \left(\frac{\varphi(x) + \psi(y) - c(x, y)}{\varepsilon} \right),$$

donde φ y ψ son a determinar.

Demostración. Aplicando condiciones de primer orden al Lagrangiano

$$\begin{aligned} \mathcal{L}(\pi, \varphi, \psi) = & \sum_{x,y} \pi_{xy} \left[c(x, y) - \psi(y) - \varphi(x) + \varepsilon \ln \left(\frac{\pi_{xy}}{e} \right) \right] \\ & + \sum_{x \in \mathcal{X}} \varphi(x) \mu_x + \sum_{y \in \mathcal{Y}} \psi(y) \nu_y \end{aligned}$$

respecto a π_{xy} , obtenemos

$$\begin{aligned} c(x, y) - \varphi(x) - \psi(y) + \varepsilon (\ln \pi_{x,y} - 1) + \varepsilon &= 0 \\ \varepsilon \ln \pi_{x,y} &= \varphi(x) + \psi(y) - c(x, y) \\ \pi_{xy} &= \exp \left(\frac{\varphi(x) + \psi(y) - c(x, y)}{\varepsilon} \right). \end{aligned}$$

□

Observación. A partir de

$$\pi_{xy} = \exp \left(\frac{\varphi(x) + \psi(y) - c(x, y)}{\varepsilon} \right),$$

escribimos

$$\pi = [\pi_{xy}] = D_{\varphi(x)} \underbrace{e^{-\frac{c(x,y)}{\varepsilon}}}_{K_\varepsilon} D_{\psi(y)}, \quad (2.12)$$

donde $D_{\varphi(x)}$ una matriz diagonal cuyos elementos en la diagonal son $e^{\varphi(x)/\varepsilon}$, $x \in \mathcal{X}$. Análogo para $D_{\psi(y)}$.

Observación. Una vez π_{xy} determinado en función de φ_x y ψ_y , podemos determinar $\min_\pi \mathcal{L}(\varphi, \psi)$ reemplazando (2.12) en $\mathcal{L}(\pi, \varphi, \psi)$. Esto provee

$$\mathcal{L}(\varphi, \psi) = \sum_{x \in \mathcal{X}} \varphi_x \mu_x + \sum_{y \in \mathcal{Y}} \psi_y \mu_y - \varepsilon \sum_{x, y} \exp \left(\frac{\varphi(x) + \psi(y) - c(x, y)}{\varepsilon} \right).$$

Finalmente, $\max_{\varphi, \psi} \mathcal{L}(\varphi, \psi)$ conlleva, por condiciones de primer orden³, al siguiente sistema de ecuaciones no lineales

$$\mu_x = \sum_{y \in \mathcal{Y}} \exp \left(\frac{\varphi(x) + \psi(y) - c(x, y)}{\varepsilon} \right) \quad (2.13)$$

$$\nu_y = \sum_{x \in \mathcal{X}} \exp \left(\frac{\varphi(x) + \psi(y) - c(x, y)}{\varepsilon} \right). \quad (2.14)$$

Algebraicamente no es viable resolver el sistema (2.13), (2.14) no lineal para el caso general. Por lo tanto, la resolución se hace por métodos numérico.

Hasta el momento, hemos presentado el problema de regularización entrópica en el caso de espacios finitos⁴ y derivado diversas propiedades clave, tales como la existencia y unicidad de la solución, así como su caracterización. En particular, hemos expresado el óptimo en términos de la relación entre φ y ψ , junto con los parámetros $\{\mu_x\}_{x \in \mathcal{X}}$, $\{\nu_y\}_{y \in \mathcal{Y}}$, $\{c_{xy}\}_{(x,y) \in \mathcal{X} \times \mathcal{Y}}$. A diferencia del problema clásico de transporte óptimo, hemos obtenido una expresión explícita para π_{xy} en función de los multiplicadores y del costo c_{xy} . El siguiente paso es abordar el problema de encontrar una solución aproximada, utilizando la relación establecida en (2.12).

³Suficientes dada la concavidad estricta

⁴El caso general:

$$\inf_{\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})} \begin{cases} \text{s.a.} & \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi + \varepsilon \int_{\mathcal{X} \times \mathcal{Y}} \pi (\log \pi - 1) \, d(x_1, x_2) \\ & (P_1)_\# \pi = \mu, \quad (P_2)_\# \pi = \nu. \end{cases}$$

es estudiado en, por ejemplo, Clason et al. (2021).

2.3. Algoritmo Sinkhorn-Knopp

La utilidad de la representación (2.12) es que nos permite abordar el problema como uno de re-escalamiento de matrices. El algoritmo que presentamos en esta sección, se aprovecha de dicha estructura.

Teorema 2.3.1. Dada una matriz $N \times N$ con entradas estrictamente positivas, digamos A , le corresponde una matriz T_A doblemente estocástica tal que

$$T_A = D_1 A D_2,$$

donde D_1, D_2 son matrices diagonales con entradas positivas. Dichas matrices son únicas salvo un factor multiplicativo $c > 0$ (Sinkhorn, 1964).

En nuestro caso, el objetivo es hallar las matrices D_φ y D_ψ , de forma que las marginales coincidan con μ y ν . Es decir, se debe cumplir que

$$D_\varphi K_\varepsilon D_\psi \mathbf{1}_N = \mu \wedge D_\psi K_\varepsilon^T D_\varphi \mathbf{1}_N = \nu.$$

donde $\mathbf{1}_N = (1 \ \dots \ 1)^T \in \mathbb{R}^N$. Expresado de otra forma

$$e^{\varphi/\varepsilon} \odot (K_\varepsilon e^{\psi/\varepsilon}) = \mu, \quad e^{\psi/\varepsilon} \odot (K_\varepsilon^T e^{\varphi/\varepsilon}) = \nu. \quad (2.15)$$

En (2.15), $e^{\varphi/\varepsilon}, e^{\psi/\varepsilon}$ son las respectivas diagonales de D_φ y D_ψ vectorizados y \odot la multiplicación entrada por entrada. Usando estas ecuaciones, derivamos las siguientes ecuaciones:

$$e^{\varphi^{(k+1)}/\varepsilon} \doteq \frac{\mu}{K_\varepsilon e^{\psi^{(k)}/\varepsilon}}, \quad e^{\psi^{(k+1)}/\varepsilon} \doteq \frac{\nu}{K_\varepsilon^T e^{\varphi^{(k+1)}/\varepsilon}} \quad (2.16)$$

$e^{\psi^{(0)}/\varepsilon} = e^{\varphi^{(0)}/\varepsilon} = \mathbf{1}_N$. En (2.16), la división es componente a componente.

A continuación un ejemplo que permite entender mejor la notación.

Ejemplo 2.3.2. Para $N = 2$, el producto

$$\begin{bmatrix} e^{\varphi_1/\varepsilon} & 0 \\ 0 & e^{\varphi_2/\varepsilon} \end{bmatrix} \begin{bmatrix} e^{-\frac{c_{11}}{\varepsilon}} & e^{-\frac{c_{12}}{\varepsilon}} \\ e^{-\frac{c_{21}}{\varepsilon}} & e^{-\frac{c_{22}}{\varepsilon}} \end{bmatrix} \begin{bmatrix} e^{\psi_1/\varepsilon} & 0 \\ 0 & e^{\psi_2/\varepsilon} \end{bmatrix} = \begin{bmatrix} e^{\frac{\varphi_1+\psi_1-c_{11}}{\varepsilon}} & e^{\frac{\varphi_1+\psi_2-c_{12}}{\varepsilon}} \\ e^{\frac{\varphi_2+\psi_1-c_{21}}{\varepsilon}} & e^{\frac{\varphi_2+\psi_2-c_{22}}{\varepsilon}} \end{bmatrix}.$$

Entonces, por ejemplo, para la fila 1:

$$e^{\frac{\varphi_1+\psi_1-c_{11}}{\varepsilon}} + e^{\frac{\varphi_1+\psi_2-c_{12}}{\varepsilon}} = \mu_1.$$

Esto es,

$$e^{\varphi_1/\varepsilon} = \frac{\mu_1}{e^{\frac{\psi_1 - c_{11}}{\varepsilon}} + e^{\frac{\psi_2 - c_{12}}{\varepsilon}}}.$$

O sea, $e^{\varphi_1/\varepsilon}$ es la primera componente del vector que se obtiene dividiendo las entradas de μ por las entradas del vector

$$\begin{bmatrix} e^{-\frac{c_{11}}{\varepsilon}} & e^{-\frac{c_{12}}{\varepsilon}} \\ e^{-\frac{c_{21}}{\varepsilon}} & e^{-\frac{c_{22}}{\varepsilon}} \end{bmatrix} \begin{bmatrix} e^{\psi_1/\varepsilon} \\ e^{\psi_2/\varepsilon} \end{bmatrix}.$$

Para resolver el problema planteado en (2.16), se aplica el siguiente algoritmo, conocido como algoritmo de Sinkhorn-Knopp.

Algorithm 1 Sinkhorn-Knopp: algoritmo para el problema de transporte óptimo regularizado, caso discreto.

1: **function** SINKHORN-KNOPP(K_ε, μ, ν)

2: $e^{\varphi^{(0)}/\varepsilon} \leftarrow \mathbf{1}_N$

3: $e^{\psi^{(0)}/\varepsilon} \leftarrow \mathbf{1}_N$

4: **for** $0 \leq k < k_{\max}$ **do**

5: $e^{\varphi^{(k+1)}/\varepsilon} \leftarrow \frac{\mu}{K_\varepsilon e^{\psi^{(k)}/\varepsilon}}$

6: $e^{\psi^{(k+1)}/\varepsilon} \leftarrow \frac{\nu}{K_\varepsilon^T e^{\varphi^{(k+1)}/\varepsilon}}$

7: **end for**

8: **end function**

El objetivo a continuación es demostrar que el Algoritmo 1 converge, es decir, que para $k \rightarrow \infty$, la solución numérica se aproxima de la solución analítica teórica. Seguimos fundamentalmente los desarrollos hechos en [Even and Merad \(2020\)](#).

Proposición 2.3.3. La aplicación definida por:

$$\forall x, y \in \mathbb{R}_{+*}^N : d_{\mathcal{H}}(x, y) = \ln \max_{i,j} \left(\frac{x_i y_j}{x_j y_i} \right)$$

es una métrica sobre el cono proyectivo \mathbb{R}_{+*}^N / \sim donde $x \sim y$ si existe $r > 0$ tal que $x = ry$. Más aún, el cono proyectivo es completo con la métrica $d_{\mathcal{H}}$.

Demostración. Ciertamente, $d_{\mathcal{H}}$ es simétrica. Luego, si $x = ry$,

$$d_{\mathcal{H}}(x, y) = d_{\mathcal{H}}(ry, y) = \ln \max_{i,j} \left(\frac{ry_i y_j}{y_j r y_i} \right) = \ln 1 = 0.$$

Por otro lado, si $\frac{x_i y_j}{x_j y_i} < 1$, entonces $\frac{x_j y_i}{x_i y_j} > 1$. Así, $d_{\mathcal{H}} \geq 0$. Luego, si $d_{\mathcal{H}}(x, y) = 0$, entonces $\frac{x_i y_j}{x_j y_i} = 1$ para todo i, j . Esto es, $x_i/y_i = x_j/y_j = r$: $x = ry$, y entonces $[x] = [y]$. Finalmente, dados $x, y, z \in \mathbb{R}_{+*}^N$, existen $\ell, k \in \{1, \dots, N\}$ tales que

$$\begin{aligned} d_{\mathcal{H}}(x, z) &= \ln \left(\frac{x_k z_{\ell}}{x_{\ell} z_k} \right) \\ &= \ln \left(\frac{x_k y_{\ell}}{x_{\ell} y_k} \cdot \frac{y_k z_{\ell}}{y_{\ell} z_k} \right) \\ &= \ln \frac{x_k y_{\ell}}{x_{\ell} y_k} + \ln \frac{y_k z_{\ell}}{y_{\ell} z_k} \\ &\leq d_{\mathcal{H}}(x, y) + d_{\mathcal{H}}(y, z). \end{aligned}$$

□

Lema 2.3.4. Sea $M \in \mathcal{M}_{N \times N}^{*+}$. Entonces, dados $x, y \in \mathbb{R}_{+*}^N$

$$d_{\mathcal{H}}(Mx, My) \leq \lambda(M) d_{\mathcal{H}}(x, y)$$

donde

$$\begin{cases} \lambda(M) &= \frac{\sqrt{\eta(M)}-1}{\sqrt{\eta(M)}+1} < 1 \\ \eta(M) &= \max_{i,j,k,\ell} \left(\frac{M_{i,k} M_{j,\ell}}{M_{j,k} M_{i,\ell}} \right). \end{cases}$$

Demostración. Ver [Birkhoff \(1957\)](#).

□

Para simplificar la notación, en adelante escribiremos $e^{\varphi/\varepsilon}$ simplemente como φ y $e^{\psi/\varepsilon}$ como ψ . Note que, en general, para la inicialización del algoritmo, basta que los vectores tengan entradas positivas. En particular, podemos utilizar los vectores $\mathbf{1}_N$ como punto de partida. Finalmente, es importante mencionar que cualquier re-escalamiento de la forma $c\varphi$ y $(1/c)\psi$, con $c > 0$, sigue siendo una solución válida.

Proposición 2.3.5. Convergencia del Algoritmo Sinkhorn. Se cumple que

1. $(\varphi^{(k)}, \psi^{(k)}) \rightarrow (\varphi^*, \psi^*)$, donde (φ^*, ψ^*) es la solución óptima. Más aún, los ratios de convergencia son

$$\begin{aligned} d_{\mathcal{H}}(\varphi^{(k)}, \varphi^*) &= O(\lambda(K_{\varepsilon})^{2k}) \\ d_{\mathcal{H}}(\psi^{(k)}, \psi^*) &= O(\lambda(K_{\varepsilon})^{2k}). \end{aligned}$$

2. Por otro lado, denotando $\Pi^{(k)} = D_{\varphi^{(k)}} K_\varepsilon D_{\psi^{(k)}}$

$$d_{\mathcal{H}}(\varphi^{(k)}, \varphi^*) \leq \frac{d_{\mathcal{H}}(\Pi^{(k)} \mathbf{1}_N, \mu)}{1 - \lambda(K_\varepsilon)}, \quad d_{\mathcal{H}}(\psi^{(k)}, \psi^*) \leq \frac{d_{\mathcal{H}}(\Pi^{(k)} \mathbf{1}_N, \nu)}{1 - \lambda(K_\varepsilon)}.$$

Demostración. Para cualesquiera $x, y \in \mathbb{R}_{+*}^N$ tenemos

$$d_{\mathcal{H}}(x, y) = d_{\mathcal{H}}(x/y, \mathbf{1}_N) = d_{\mathcal{H}}(\mathbf{1}_N/x, \mathbf{1}_N/y)$$

donde $/$ representa la división entrada por entrada. Luego, por la definición del algoritmo y el lema (2.3.4)

$$\begin{aligned} d_{\mathcal{H}}(\varphi^{(k+1)}, \varphi^*) &= d_{\mathcal{H}}\left(\frac{\mu}{K_\varepsilon \psi^{(k)}}, \frac{\mu}{K_\varepsilon \psi^*}\right) \\ &= d_{\mathcal{H}}(K_\varepsilon \psi^{(k)}, K_\varepsilon \psi^*). \\ &\leq \lambda(K_\varepsilon) d_{\mathcal{H}}(\psi^{(k)}, \psi^*). \end{aligned}$$

De manera análoga

$$\begin{aligned} d_{\mathcal{H}}(\psi^{(k)}, \psi^*) &= d_{\mathcal{H}}\left(\frac{\nu}{K_\varepsilon^T \varphi^{(k)}}, \frac{\nu}{K_\varepsilon^T \varphi^*}\right) \\ &= d_{\mathcal{H}}(K_\varepsilon^T \varphi^{(k)}, K_\varepsilon^T \varphi^*) \\ &\leq \lambda(K_\varepsilon^T) d_{\mathcal{H}}(\varphi^{(k)}, \varphi^*). \end{aligned}$$

Como $\lambda(K_\varepsilon) = \lambda(K_\varepsilon^T)$

$$\begin{aligned} d_{\mathcal{H}}(\varphi^{(k+1)}, \varphi^*) &\leq \lambda(K_\varepsilon)^2 d_{\mathcal{H}}(\varphi^{(k)}, \varphi^*) \\ d_{\mathcal{H}}(\psi^{(k)}, \psi^*) &\leq \lambda(K_\varepsilon)^2 d_{\mathcal{H}}(\psi^{(k-1)}, \psi^*). \end{aligned}$$

Dado que $\lambda(K)^2 < 1$ concluimos que $\varphi^{(k)} \rightarrow \varphi^*$ y $\psi^{(k)} \rightarrow \psi^*$. Finalmente, para el segundo inciso,

$$\begin{aligned} d_{\mathcal{H}}(\varphi^{(k)}, \varphi^*) &\leq d_{\mathcal{H}}(\varphi^{(k+1)}, \varphi^{(k)}) + d_{\mathcal{H}}(\varphi^{(k+1)}, \varphi^*) \\ &\leq d_{\mathcal{H}}\left(\frac{\mu}{K_\varepsilon \psi^{(k)}}, \varphi^{(k)}\right) + \lambda(K_\varepsilon) d_{\mathcal{H}}(\varphi^{(k)}, \varphi^*) \\ &= d_{\mathcal{H}}(\mu, \varphi^{(k)} \odot (K_\varepsilon \psi^{(k)})) + \lambda(K_\varepsilon) d_{\mathcal{H}}(\varphi^{(k)}, \varphi^*) \end{aligned}$$

y, dado que $\varphi^{(k)} \odot (K_\varepsilon \psi^{(k)}) = \Pi^{(k)} \mathbf{1}_N$, se obtiene lo deseado. Para el caso de $d_{\mathcal{H}}(\psi^{(k)}, \psi^*)$ es totalmente análogo. \square

En esta última sección de este capítulo, hemos demostrado la convergencia de los multiplicadores de Lagrange respecto a la métrica de Hilbert hacia la solución del problema \mathcal{P}_ε , proporcionando además estimaciones sobre la velocidad de convergencia. En particular, la convergencia es exponencial y la complejidad computacional del algoritmo es $O(N^2)$ [Merigot and Thibert \(2020\)](#). En la demostración de la proposición [2.3.5](#), omitimos los detalles del argumento inductivo, ya que su verificación es sencilla. Para la prueba de la completitud del espacio $(\mathbb{R}_+^n / \sim, d_{\mathcal{H}})$, remitimos al lector a [Nussabaum \(1987\)](#).

Además, dado un $\varepsilon > 0$, es suficiente tomar $k \in \mathbb{N}$ tal que

$$k \geq \left\lfloor \frac{1}{2} \ln_\lambda \frac{\varepsilon}{C} \right\rfloor + 1$$

para garantizar una distancia menor que ε . La constante C en esta expresión depende de las medidas μ, ν y de la función de costo c .

En la siguiente sección, comenzamos con las aplicaciones de la teoría del transporte óptimo en economía. Es importante mencionar que el desarrollo presentado es, en esencia, una transcripción detallada de [Dupuy et al. \(2021\)](#). Nuestro principal aporte radica en la exposición detallada de los resultados fundamentales, incluyendo la demostración explícita de muchas proposiciones y lemas que en el trabajo original solo se mencionan sin mayor desarrollo.

Capítulo 3

Estimación de costos y algoritmo SISTA

En este capítulo, estudiamos el problema de estimación de costos presentado en [Dupuy et al. \(2021\)](#). En dicho artículo se introduce un método iterativo innovador en la literatura: SISTA, cuyo objetivo es *aprender* los costos en un problema de transporte óptimo, es decir, recuperar la función de costo $c(x, y)$. Este método combina el algoritmo de Sinkhorn con un esquema de *gradient descent*. La aplicación cuantitativa en el artículo original se centra en un modelo de migración. Nuestro propósito es exponer en detalle los desarrollos presentados en dicho trabajo.

Para facilitar la comprensión, hemos introducido en los anexos de este documento, así como en los capítulos anteriores, todos los resultados matemáticos necesarios.

3.1. El modelo

Nos situamos nuevamente en el contexto de conjuntos finitos cuya cardinalidad es N ¹ y se busca resolver

$$\min_{\pi \in \Pi(\mu, \nu)} \left\{ \sum_{x, y} \pi_{xy} c_{xy} + \varepsilon \pi_{xy} (\ln \pi_{xy} - 1) \right\}.$$

¹Nuevamente, sin pérdida de generalidad se asume misma cardinalidad de los conjuntos \mathcal{X} y \mathcal{Y} .

Para simplificar la notación, re-escribimos usando (i, j) en vez de (x, y) :

$$\min_{\pi \in \Pi(\mu, \nu)} \left\{ \sum_{1 \leq i, j \leq N} \pi_{ij} c_{ij} + \varepsilon \pi_{ij} (\ln \pi_{ij} - 1) \right\}. \quad (3.1)$$

El problema (3.1) busca minimizar el costo del emparejamiento, buscando soluciones *sparse* es interiores (véase el capítulo 2).

En la práctica, uno no conoce c pero sí observa un matching $\hat{\pi}$ y más bien, se busca c tal que la solución óptima al problema (3.1) con dicha función de costo sea $\hat{\pi}$.

En ese sentido, se trata del problema inverso, dado $\hat{\pi}$ *aprender* c . Este tipo de problemas es de particular interés en teoría económica [Chambers et al. \(2021, 2023\)](#).

En [Dupuy et al. \(2021\)](#), se propone una estructura de costos:

$$c_{ij} = c_{ij}^\beta = \sum_{k=1}^K \beta_k d_{ij}^k,$$

donde el parámetro a ser estimado es β . Las variables d_{ij} son observadas. La estimación del vector β se hace vía la minimización de una función convexa con una penalización de tipo Lasso que introduciremos más adelante.

Dada la estructura del problema, sabemos que $\pi_{ij}^* > 0$ y que las condiciones de primer orden sobre el Lagrangiano proveen:

$$\pi_{ij} = \exp \left(\frac{\varphi_i + \psi_j - c_{ij}}{\varepsilon} \right).$$

Desde ahora, sin pérdida de generalidad, tomamos $\varepsilon = 1$. El algoritmo Sinkhorn, adaptado a este contexto, es el siguiente:

$$\begin{aligned} \exp(\varphi_i^{(k+1)}) &= \frac{\mu_i}{\sum_{j=1}^N \exp(\psi_j^{(k)} - c_{ij})} \\ \exp(\psi_j^{(k+1)}) &= \frac{\nu_j}{\sum_{i=1}^N \exp(\varphi_i^{(k+1)} - c_{ij})}. \end{aligned}$$

Observación. Por un argumento de dualidad, resolver

$$\min_{\pi \in \Pi} \left\{ \sum_{1 \leq i, j \leq N} \pi_{ij} c_{ij} + \varepsilon \pi_{ij} (\ln \pi_{ij} - 1) \right\}$$

es equivalente a resolver

$$\max_{\varphi, \psi} \left\{ \sum_{i=1}^N \varphi_i \mu_i + \sum_{j=1}^N \psi_j \nu_j - \varepsilon \sum_{1 \leq i, j \leq N} \exp \left(\frac{\mu_i + \nu_j - c_{ij}}{\varepsilon} \right) \right\}.$$

3.2. La estructura de costos paramétrica

La estructura de costos está completamente identificada por un vector de parámetros $\beta \in \mathbb{R}^K$, de forma que

$$c_{ij}^\beta = \sum_{k=1}^K \beta_k d_{ij}^k.$$

Matricialmente:

$$\begin{pmatrix} c_{11}^\beta & c_{12}^\beta & & \\ c_{21}^\beta & \ddots & & \\ & & \ddots & \\ & & & c_{NN}^\beta \end{pmatrix} = \sum_{k=1}^K \beta_k \begin{pmatrix} d_{11}^k & d_{12}^k & & \\ d_{21}^k & \ddots & & \\ & & \ddots & \\ & & & d_{NN}^k \end{pmatrix}. \quad (3.2)$$

En (3.2) d_{ij} mide la *disimilitud* entre i y j . Por ejemplo, en caso de problemas espaciales, puede ser una distancia, diferencias de PBI, diferencias de población etc. Una situación estándar es considerar lo siguiente:

1. Si x_i e y_j guardan las características de i y j , definir $d_{ij}^k = (x_i^k - y_j^k)^2$.
2. Considerar siempre las mismas unidades: variables uniformizadas.

Observación. No se tiene necesariamente que $d_{ij}^k = d_{ji}^k$. Pensemos por ejemplo en d_{ij}^k como la fracción de individuos en el país i que no dominan el idioma el país j .

Proposición 3.2.1. Para estimar β , junto a los multiplicadores de Lagrange φ, ψ , teniendo en cuenta las siguientes restricciones:

$$\sum_j \pi_{ij}^\beta = \mu_i \quad \wedge \quad \sum_i \pi_{ij}^\beta = \nu_j$$

$$\sum_{1 \leq i, j \leq N} \pi_{ij}^\beta d_{ij}^k = \sum_{1 \leq i, j \leq N} \hat{\pi}_{ij} d_{ij}^k, \quad k = 1, \dots, K$$

donde, $\pi_{ij}^\beta = \exp(\varphi_i + \psi_j - c_{ij}^\beta)$, debe resolverse el siguiente problema de optimización

$$\min_{\varphi, \psi, \beta} \underbrace{\left\{ \sum_{1 \leq i, j \leq N} \exp(\varphi_i + \psi_j - c_{ij}^\beta) + \sum_{1 \leq i, j \leq N} \hat{\pi}_{ij} (c_{ij}^\beta - \varphi_i - \psi_j) \right\}}_{F \text{ convexa}}. \quad (3.3)$$

En (3.3), $\hat{\pi}_{ij}$ es el emparejamiento observado entre i y j .

Demostración. Intuitivamente, a modo de sketch, notemos que

$$\begin{aligned}\frac{\partial F}{\partial \varphi_i} &= \sum_{j=1}^N \exp(\varphi_i + \psi_j - c_{ij}^\beta) - \underbrace{\sum_{j=1}^N \hat{\pi}_{ij}}_{=\mu_i} = 0, \quad \forall i \\ \frac{\partial F}{\partial \psi_j} &= \sum_{i=1}^N \exp(\varphi_i + \psi_j - c_{ij}^\beta) - \underbrace{\sum_{i=1}^N \hat{\pi}_{ij}}_{=\nu_j} = 0, \quad \forall j \\ \frac{\partial F}{\partial \beta_k} &= - \sum_{1 \leq i, j \leq N} d_{ij}^k \exp(\varphi_i + \psi_j - c_{ij}^\beta) + \sum_{1 \leq i, j \leq N} \hat{\pi}_{ij} d_{ij}^k.\end{aligned}$$

Es decir, de las condiciones de primer orden aplicadas a $F(\cdot)$, se obtienen las restricciones. Ver un argumento más detallado en [Galichon and Salanié \(2022a\)](#). \square

Con el objetivo de capturar solo los efectos importantes, se agrega una penalización tipo Lasso ([Murphy, 2022](#)):

$$\min_{\varphi, \psi, \beta} \underbrace{\sum_{1 \leq i, j \leq N} \exp(\varphi_i + \psi_j - c_{ij}^\beta) - \hat{\pi}_{ij}(\varphi_i + \psi_j - c_{ij}^\beta)}_{=\Phi(\varphi, \psi, \beta)} + \gamma \|\beta\|_1. \quad (3.4)$$

Para continuar con la lectura, recomendamos fuertemente consultar al anexo de análisis convexo y el de subgradientes y subdiferenciales. Estos introducen definiciones y resultados indispensables para abordar el algoritmo SISTA y probar su convergencia. La siguiente figura ilustra como en \mathbb{R}^2 , la penalización Lasso restringe el dominio de elección de los β . Esto, tal y como se explica en [James et al. \(2021\)](#), nos permite recuperar solo las regresoras más importantes y medir dicha importancia por el valor numérico del β^k . Si tiende a cero, menor será la importancia de la interacción d_{ij}^k

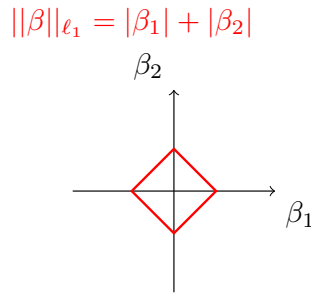


Figura 3.1 Restricción LASSO.

3.3. SISTA

El algoritmo SISTA - Sinkhorn (S) y Proximal Gradient Descent (ISTA) - consiste en minimizar (3.4) en tres etapas:

1. Iterar φ , manteniendo β y ψ constantes.
2. Iterar ψ , manteniendo β y φ constantes.
3. Descenso de gradiente proximal respecto a β manteniendo φ y ψ constantes.

Definición 3.3.1. Función prox $\|\cdot\|_1$.

$$\text{prox}_{\rho, \gamma, \|\cdot\|_1}(z) = \begin{cases} z - \rho\gamma, & \text{si } z > \rho\gamma \\ 0, & \text{si } |z| \leq \rho\gamma \\ z + \rho\gamma, & \text{si } z < -\rho\gamma. \end{cases}$$

Algorithm 2 SISTA.

1: $\beta^{(0)}, \rho, \mu, \nu, d_{ij}^k, \hat{\pi}_{ij}, \varphi^{(0)}, \psi^{(0)}$

2:

3: **while** not converged **do**

4: Planteamos $c_{ij}^{\beta^{(t)}} = \sum_{k=1}^K \beta_k^{(t)} d_{ij}^k$

$$\begin{cases} \exp(\varphi_i^{(t+1)}) &= \frac{\mu_i}{\sum_{j=1}^N \exp(\psi_j^{(t)} - c_{ij}^{\beta^{(t)}})} \\ \exp(\psi_j^{(t+1)}) &= \frac{\nu_j}{\sum_{i=1}^N \exp(\varphi_i^{(t+1)} - c_{ij}^{\beta^{(t)}})} \end{cases}$$

5: Sea $\pi_{ij}^{\beta^{(t)}} = \exp(\varphi_i^{(t+1)} + \psi_j^{(t+1)} - c_{ij}^{\beta^{(t)}})$. Para $k = 1, \dots, K$

$$\beta_k^{(t+1)} = \text{prox}_{\rho, \gamma, \|\cdot\|_1} \left(\beta_k^{(t)} - \rho \underbrace{\sum_{1 \leq j, i \leq N} (\hat{\pi}_{ij} - \pi_{ij}^{\beta^{(t)}}) d_{ij}^{(k)}}_{\nabla_{\beta} F} \right).$$

6: **end while**

7: **return** β

3.3.1. Convergencia del algoritmo SISTA

En esta sub-sección presentamos la prueba de la convergencia del algoritmo. Seguimos nuevamente Dupuy et al. (2021). Se introducen tres supuestos sobre las disimilitudes $d_{ij}^k; \forall k = 1, \dots, K$

$$\sum_{i=1}^N d_{ij}^k = 0, \forall j = 1, \dots, N \quad (3.5)$$

$$\sum_{j=1}^N d_{ij}^k = 0, \forall i = 1, \dots, N \quad (3.6)$$

$$\{d_1, \dots, d^K\} \text{ es un conjunto l.i.} \quad (3.7)$$

y, un supuesto respecto al plan de transporte observado:

$$\hat{\pi}_{ij} > 0. \quad (3.8)$$

Observación. Siempre podemos obtener $\sum_{i,j} d_{ij}^k = 0$ y las condiciones (3.5), (3.5) re-definiendo

$$\tilde{d}_{ij}^k = d_{ij}^k - a_i^k - b_j^k$$

con $a_i^k = \frac{1}{N} \sum_{j=1}^N d_{ij}^k$ y $b_j^k = \frac{1}{N} \sum_{i=1}^N d_{ij}^k - \frac{1}{N^2} \sum_{1 \leq p, q \leq N} d_{pq}^k$. Más aún, en este contexto $\sum_{1 \leq i, j \leq N} \tilde{d}_{ij}^k = 0$.

Demostración. Sumando:

$$\begin{aligned} \sum_{1 \leq i, j \leq N} \tilde{d}_{ij}^k &= \sum_{1 \leq i, j \leq N} (d_{ij}^k - a_i^k - b_j^k) \\ &= \sum_{1 \leq i, j \leq N} d_{ij}^k - \sum_{1 \leq i, j \leq N} \left[\frac{1}{N} \sum_{j=1}^N d_{ij}^k \right] - \sum_{1 \leq i, j \leq N} \left[\frac{1}{N} \sum_{i=1}^N d_{ij}^k - \frac{1}{N^2} \sum_{1 \leq p, q \leq N} d_{pq}^k \right] \\ &= \sum_{1 \leq i, j \leq N} d_{ij}^k - \sum_{1 \leq i \leq N} \sum_{1 \leq j \leq N} N \cdot \frac{1}{N} d_{ij}^k - \sum_{1 \leq j \leq N} \sum_{1 \leq i \leq N} N \cdot \frac{1}{N} d_{ij}^k \\ &\quad + N^2 \frac{1}{N^2} \sum_{1 \leq p, q \leq N} d_{pq}^k \\ &= 2 \sum_{1 \leq i, j \leq N} d_{ij}^k - 2 \sum_{1 \leq i, j \leq N} d_{ij}^k \\ &= 0. \end{aligned}$$

Luego, uno puede ver fácilmente que $\sum_{i=1}^N \tilde{d}_{ij}^k = 0$ y $\sum_{j=1}^N \tilde{d}_{ij}^k = 0$. □

Teorema 3.3.2. Supongamos que tenemos 3.5, 3.5 3.7 y 3.8. Entonces, la sucesión $x^{(t)} = (\varphi^{(t)}, \psi^{(t)}, \beta^{(t)})$ generada por el algoritmo SISTA converge a la solución del problema de optimización, $\min_{(\varphi, \psi, \beta) \in K} \Phi(\varphi, \psi, \beta)$, cuando $t \rightarrow \infty$ (para un ρ suficientemente chico que será dado), con $\varphi(0), \psi(0) \in \mathbb{R}_+^N$ y $\beta^{(0)}$ un *guess*. Más aún, existe $\delta > 0$ tal que

$$\Phi(x^{(t)}) - \Phi(x^*) \leq \frac{\Phi(x^{(0)}) - \Phi(x^*)}{(1 + \delta)^t}. \quad (3.9)$$

La prueba del teorema 3.3.2 se descompone en las siguientes partes. Primero, se estudia la función Φ . Luego, se establece la existencia de una solución y como caracterizarla. Finalmente, se prueba que el algoritmo SISTA permite converger a dicha solución de forma que (3.9) se satisface.

Nuestro objetivo es resolver

$$\inf_{(\varphi, \psi, \beta) \in \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^K} \Phi(\varphi, \psi, \beta) = G \circ \Lambda(\varphi, \psi, \beta) + \gamma \|\beta\|_1 \quad (3.10)$$

donde $\Lambda : \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^K \rightarrow \mathcal{M}_{N \times N}$ es una aplicación lineal definida de la siguiente forma:

$$(\Lambda(\varphi, \psi, \beta))_{ij} = \varphi_i + \psi_j - c_{ij}^\beta, \quad \forall (i, j)$$

y G una función C^∞ y convexa, con regla de correspondencia

$$G(\lambda) = \sum_{1 \leq i, j \leq N} (\exp(\lambda_{ij}) - \hat{\pi}_{ij} \lambda_{ij}), \quad \forall \lambda \in \mathcal{M}_{N \times N}.$$

Luego,

$$\begin{aligned} \nabla G(\lambda) &= [e^{\lambda_{ij}} - \hat{\pi}_{ij}] \in \mathcal{M}_{N \times N} \\ \mathbf{H}(G(\lambda)) &= \text{Diag}(\exp(\lambda)) = \begin{pmatrix} e^{\lambda_{11}} & & & \\ & e^{\lambda_{12}} & & \\ & & \ddots & \\ & & & \ddots & \\ & & & & e^{\lambda_{NN}} \end{pmatrix} \in \mathcal{M}_{N^2 \times N^2}. \end{aligned}$$

Denotando $x = (\varphi, \psi, \beta)$

$$\nabla F(x) = \underbrace{\underbrace{\Lambda^T}_{(N \times N \times K) \times (N \times N)} \underbrace{(\exp \Lambda(x) - \hat{\pi})}_{\sim \mathbb{R}^{N^2}}}_{(N \times N \times K) \times (N \times N)} \quad (3.11)$$

$$\mathbf{H}(F(x)) = \Lambda^T \text{Diag}(\exp(\Lambda(x))) \Lambda. \quad (3.12)$$

En efecto,

$$\nabla F(x) = (dG_{\Lambda(x)} \circ d\Lambda_x)^T = \Lambda^T(\exp \Lambda(x) - \hat{\pi}).$$

Definamos ahora

$$\tilde{c}^\beta = \sum_{k=1}^K \beta_k \tilde{d}^k, \quad \forall \beta \in \mathbb{R}^K$$

y

$$\tilde{\Lambda}(\tilde{\varphi}, \tilde{\psi}, \beta)_{ij} = \tilde{\varphi}_{ij} + \tilde{\psi}_{ij} - \tilde{c}_{ij}^\beta.$$

Entonces, notamos que

$$\Lambda(\varphi, \psi, \beta) = \tilde{\Lambda}\left(\varphi - \sum_{k=1}^K \beta_k a^k, \psi - \sum_{k=1}^K \beta_k b^k, \beta\right).$$

En otras palabras, sin pérdida de generalidad, podemos considerar \tilde{d} en vez de d . Ahora, por otro lado, observemos que, para cualquier vector fijo $m \in \mathbb{R}^N$

$$\Lambda(\varphi + m, \psi - m, \beta) = \Lambda(\varphi, \psi, \beta).$$

Si se impone la restricción $\varphi_1 = 0$, se remueve esta invarianza. Así, llegamos al siguiente resultado.

Lema 3.3.3. Bajo los supuestos del teorema 3.3.2, Λ es inyectiva de

$$E = \{(\varphi, \psi, \beta) \in \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^K : \varphi_1 = 0\} \simeq \mathbb{R}^{2N-1+K}.$$

a $\mathcal{M}_{N \times N}$.

Demostración. Sea $(\varphi, \psi, \beta) \in E$, perteneciendo al núcleo de Λ , entonces:

$$\forall i, j : \varphi_i + \psi_j = \sum_{k=1}^K \beta_k d_{ij}^k.$$

Luego, por los supuestos

$$\begin{aligned} \sum_{j=1}^N (\varphi_i + \psi_j) &= \sum_{j=1}^N \sum_{k=1}^K \beta_k d_{ij}^k \\ &= \sum_{k=1}^K \underbrace{\left[\sum_{j=1}^N \beta_k d_{ij}^k \right]}_{=0}. \end{aligned}$$

Esto es, $N\varphi_i + \sum_{j=1}^N \psi_j = 0$, para todo i . En particular, para $i = 1$

$$\sum_{j=1}^N \psi_j = 0.$$

Así, $\varphi = 0$. Pero entonces, análogamente, $\psi = 0$ también. Con lo cual,

$$\sum_{k=1}^K \beta_k d^k = 0.$$

Por la independencia lineal de $\{d^k\}_{k=1}^K$, $\beta = \mathbf{0}$. Así, $\{\mathbf{0}_{\mathbb{R}^N}, \mathbf{0}_{\mathbb{R}^N}, \mathbf{0}_{\mathbb{R}^K}\} = \text{Ker}(\Lambda)$. \square

Lema 3.3.4. Bajo los supuestos del teorema 3.3.2, para $x, y \in E$ existen $\alpha = \alpha(M)$ y $v = v(M)$, con

$$\max\{\|\Lambda(x)\|_\infty, \|\Lambda(y)\|_\infty\} \leq M,$$

tales que

$$F(x) \geq F(y) + \nabla F(y)(x - y) + \frac{v}{2}\|x - y\|_2^2$$

$$\|\nabla F(x) - \nabla F(y)\|_2 \leq \alpha\|x - y\|_2.$$

Demostración. Proponemos $v = e^{-M}\sigma_{\min}$ donde σ_{\min} es el menor autovalor de la matriz simétrica $\Lambda^T \Lambda$. Por el Teorema de Taylor, existe $c \in [x, y]$ tal que

$$F(y) = F(x) + \nabla F(x)^T(y - x) + \frac{1}{2}(y - x)^T \mathbf{H}(F(c))(y - x),$$

con $c = \theta x + (1 - \theta)y$, $\theta \in [0, 1]$. Luego,

$$\begin{aligned} (y - x)^T \mathbf{H}(F(c))(y - x) &= (y - x)^T \Lambda^T \text{Diag}(\exp(\Lambda(\theta x + (1 - \theta)y))) \Lambda(y - x)^T \\ &\stackrel{\text{linealidad de } \Lambda}{=} (y - x)^T \Lambda^T \text{Diag}(\exp(\theta \Lambda(x) + (1 - \theta)\Lambda(y))) \Lambda(y - x) \\ &\geq (y - x)^T \Lambda^T \text{Diag}(\exp(-M)) \Lambda(y - x) \\ &\geq e^{-M}\sigma_{\min}\|y - x\|_2^2, \end{aligned}$$

donde la última desigualdad es consecuencia del Teorema Espectral ([Botelho et al.](#),

2023):

$$\begin{aligned}
 \min_{\|n\|=1} \{n^T A n\} &= \min_{\|n\|=1} \{n^T (U D U^T) n\} \\
 &= \min_{\|n\|=1} \{(U^T n)^T D (U^T n)\} \\
 &= \min_{\|U^T n\|=1} \{(U^T n)^T D (U^T n)\} \\
 &= \min_{\|z\|=1} \{z^T D z\} \\
 &= \min_{\|z\|=1} \sum_i z_i^2 D_{ii} \\
 &= \min_{\|z\|=1} \sum_i z_i^2 \lambda_i \\
 &\geq \min_i \{\lambda_i\} \underbrace{\sum_i z_i^2}_{=1}.
 \end{aligned}$$

Así,

$$\frac{1}{\|y - x\|_2^2} (y - x)^T (\Lambda^T \Lambda) (y - x) \geq \min_i \lambda_i.$$

Con respecto a la segunda desigualdad: $\|\nabla F(x) - \nabla F(y)\|_2 \leq \alpha \|x - y\|_2$, esto es consecuencia de que

$$\|g(x) - g(y)\| \leq \sup_{z \in (x, y)} \|g'(z)\| \cdot \|x - y\|$$

y

$$\begin{aligned}
 \sup_{z \in (x, y)} \|g'(z)\| &= \sup_{z \in (x, y)} \|J(\nabla F(z))\| \\
 &= \sup_{z \in (x, y)} \|\mathbf{H}(F(z))\| \\
 &= \sup_{z \in (x, y)} \|\Lambda^T \text{diag}(\exp(\Lambda(z))) \Lambda\| \\
 &\leq e^M \|\Lambda^T \Lambda\| \\
 &\leq e^M \|\Lambda^T\| \cdot \|\Lambda\|.
 \end{aligned}$$

□

Proposición 3.3.5. Bajo los supuestos del teorema 3.3.2

1. Φ es coerciva en \mathbb{R}^{2N-1+K} , i.e., los conjuntos $\{x \in \mathbb{R}^{2N-1+K} : f(x) \leq c\}$ son acotados.

2. El problema (3.10) admite una única solución $x^* = (\varphi^*, \psi^*, \beta^*)$.
3. El óptimo x^* viene caracterizado por

$$\nabla_{\varphi} F(x^*) = \mathbf{0}, \nabla_{\psi} F(x^*) = \mathbf{0}, -\nabla_{\beta} F(x^*) \in \gamma \partial \|\cdot\|_1(\beta^*). \quad (3.13)$$

Lema 3.3.6. Toda función coerciva continua $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}$, definida en un conjunto cerrado, posee un mínimo.

Demostración. Escojamos $x_0 \in X$ arbitrario. Como f es coerciva, existe $k > 0$ tal que, si $\|x\| \geq k$, $f(x) \geq 1 + f(x_0)$. Luego, tomemos el compacto $K = X \cap \overline{B}(0, k)$. Como f es continua, existe x^* tal que $f(x^*)$ es el mínimo en K . Luego, si $x_0 \notin K$, $\|x_0\| > k$ por lo que $x_0 \in K$. Por ende, $f(x^*) \geq f(x_0)$. Ahora, si $x \in X - K$, $\|x\| > k$, por lo que $f(x) > f(x_0) \geq f(x^*)$. De este modo, x^* es minimizador global. \square

Demostración. Probamos ahora la proposición 3.3.5. Tenemos las siguientes 3 desigualdades:

$$\begin{aligned} e^t - pt &\geq p - \ln p, \quad \forall p > 0 \\ e^{\lambda} - \hat{\pi}_{ij}\lambda &\geq \hat{\pi}_{ij}|\lambda|, \quad \lambda < 0 \\ e^{\lambda} - \hat{\pi}_{ij}\lambda &= \hat{\pi}_{ij}|\lambda| + e^{\lambda} - 2\hat{\pi}_{ij}\lambda \\ e^{\lambda} - \hat{\pi}_{ij}\lambda &\geq \hat{\pi}_{ij}\lambda - 2\ln 2. \end{aligned}$$

Combinando esto y sumando, llegamos a

$$\Phi(\varphi, \psi, \beta) \geq \sum_{1 \leq i, j \leq N} \hat{\pi}_{ij} |\Lambda_{ij}(\varphi, \psi, \beta)| - 2N^2 \ln 2 + \gamma \|\beta\|_1. \quad (3.14)$$

Dado que Λ es inyectiva sobre E y $\hat{\pi}_{ij} > 0$, podemos concluir la coercividad de Φ . Luego, dada la continuidad de Φ (pues F es suave y $\|\cdot\|_1$ continua), y el hecho que E es cerrado en \mathbb{R}^{2N-1+K} , podemos asegurar la existencia de un minimizador global x^* (lema 3.3.6). Pero entonces, usando la proposición E.0.10, concluimos que

$$\mathbf{0} \in \partial \Phi(\varphi^*, \psi^*, \beta^*). \quad (3.15)$$

Dado que (Boyd, 2022)²

$$\begin{aligned} \partial(\alpha f(x)) &= \alpha \partial f(x) \\ \partial \left(\sum_{i=1}^n f_i(x) \right) &= \sum_{i=1}^n \partial f_i(x), \end{aligned}$$

²A esto se le conoce como la propiedad de Moreau-Rockafellar.

concluimos que (3.15) es equivalente a (3.13). \square

Prueba del teorema 3.3.2 (es la misma de Dupuy et al. (2021) solo que reparando en algunos detalles).

Demostración. Empezamos con un vector de inicialización $(\varphi^{(0)}, \psi^{(0)}, \beta^{(0)}) \in \mathbb{R}^{2N+K}$ (previamente se ha comentado sobre esta inicialización), y se define recursivamente la secuencia $(\varphi^{(t)}, \psi^{(t)}, \beta^{(t)})$

$$\begin{aligned}\varphi^{(t+1)} &\in \operatorname{argmin} F(\cdot, \psi^{(t)}, \beta^{(t)}) \\ \psi^{(t+1)} &\in \operatorname{argmin} F(\varphi^{(t+1)}, \cdot, \beta^{(t)})\end{aligned}$$

construidos de manera explícita vía

$$\begin{cases} \exp(\varphi_i^{(t+1)}) &= \frac{\mu_i}{\sum_{j=1}^N \exp(\psi_j^{(t)} - c_{ij}^{\beta^{(t)}})} \\ \exp(\psi_j^{(t+1)}) &= \frac{\nu_j}{\sum_{i=1}^N \exp(\varphi_i^{(t+1)} - c_{ij}^{\beta^{(t)}})} \end{cases}$$

Por otro lado, teníamos

$$\beta^{(t+1)} = \operatorname{prox}_{\rho, \gamma, \|\cdot\|} \left(\beta_k^{(t)} - \rho \underbrace{\sum_{1 \leq j, i \leq N} (\hat{\pi}_{ij} - \pi_{ij}^{\beta^{(t)}}) d_{ij}^{(k)}}_{\nabla_{\beta} F} \right).$$

Esto es lo mismo que minimizar respecto a β

$$\rho \gamma \|\beta\|_1 + \frac{1}{2} \|\beta - (\beta^{(t)} - \rho \nabla_{\beta} F(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t)}))\|_2^2. \quad (3.16)$$

El primer elemento de la demostración es escoger ρ adecuadamente. Sea $C = \Phi(\varphi^{(0)}, \psi^{(0)}, \beta^{(0)})$. Gracias a la cota (3.14)

$$\Phi(\varphi, \psi, \beta) \leq C \implies \|\beta\|_1 \leq \frac{C + 2N^2 \ln 2}{\gamma},$$

definiendo $A = C + 2N^2 \ln 2 + 1$ tendremos que

$$A > C + 2N^2 \ln 2, \quad F \leq A \text{ cuando } \Phi \leq C.$$

Sea ahora $\theta \in C^\infty(\mathbb{R}, \mathbb{R})$, función no decreciente tal que

$$\theta(t) = \begin{cases} t, & \text{si } t \leq A \\ \theta(t) \geq t, & \text{si } t \in [A, 2A] \\ 2A, & \text{si } t \geq 2A. \end{cases}$$

A modo de ejemplo, para C^1 podríamos tener

$$\theta(t) = \begin{cases} t, & \text{si } t \leq A \\ \frac{1}{A^3}(t - A)^3(2A - t) + t \geq t, & \text{si } t \in [A, 2A] \\ 2A, & \text{si } t \geq 2A. \end{cases}$$

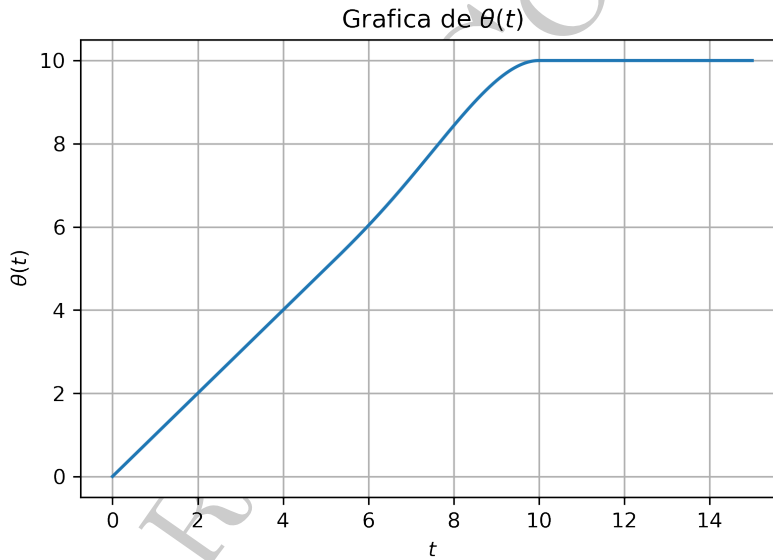


Figura 3.2 $\theta(t) \in C^1(\mathbb{R}, \mathbb{R})$ para $t \in [0, 3A]$, $A = 5$.

Luego, $\tilde{F} = \theta \circ F$ tiene gradiente globalmente Lipschitz; basta combinar la suavidad de θ con el lema 3.3.4. Sea α dicha constante de Lipschitz. Entonces, afirmamos que $\rho \in (0, \alpha^{-1}]$. Como $\nabla \tilde{F}(x)$ es α -Lipschitz continua,

$$\tilde{F}(x) \leq \tilde{F}(y) + \nabla \tilde{F}(y)(x - y) + \frac{\alpha}{2} \|x - y\|_2^2.$$

En efecto, si consideramos $\phi(t) = \tilde{F}(y + t(x - y))$

$$\begin{aligned}
 \tilde{F}(x) - \tilde{F}(y) - \langle \nabla \tilde{F}(y), x - y \rangle &= \int_0^1 \langle \nabla \tilde{F}(y + t(x - y)), x - y \rangle dt - \langle \nabla \tilde{F}(y), x - y \rangle \\
 &= \int_0^1 \langle \nabla \tilde{F}(y + t(x - y)) - \nabla \tilde{F}(y), x - y \rangle dt \\
 &\leq \int_0^1 \|\nabla \tilde{F}(y + t(x - y)) - \nabla \tilde{F}(y)\|_2 \|x - y\|_2 dt \\
 &\leq \int_0^1 \alpha t \|x - y\|_2^2 dt \\
 &= \frac{\alpha}{2} \|x - y\|_2^2.
 \end{aligned}$$

Ahora, probemos por inducción que el ρ escogido hace que $\Phi \leq C$ para futuras iteraciones. Ciertamente por construcción

$$C \geq \Phi(\varphi^{(t)}, \psi^{(t)}, \beta^{(t)}) \geq \Phi(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t)}).$$

Probar sin embargo que $\Phi(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t+1)}) \leq C$ requiere más pasos. Primero, cuando $\Phi \leq A$, $F \leq A$ y por ende, $F = \tilde{F}$ y $\nabla F = \nabla \tilde{F}$, de (3.16)

$$\beta^t - \beta^{t+1} - \rho \nabla_{\beta} F(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t)}) \in \rho \gamma \partial(\|\cdot\|_1)(\beta^{t+1}).$$

Esto es, usando la definición de subdiferencial:

$$\gamma \|\beta^{(t)}\|_1 \geq \gamma \|\beta^{(t+1)}\|_1 + \frac{1}{\rho} \|\beta^{(t+1)} - \beta^{(t)}\|_2^2 - \nabla_{\beta} F(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t)})(\beta^{(t)} - \beta^{(t+1)}). \quad (3.17)$$

Por la α -Lipschitz continuidad

$$F(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t)}) \geq \tilde{F}(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t+1)}) \quad (3.18)$$

$$+ \nabla_{\beta} F(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t)})(\beta^{(t)} - \beta^{(t+1)}) \quad (3.19)$$

$$- \frac{\alpha}{2} \|\beta^{(t+1)} - \beta^{(t)}\|_2^2. \quad (3.20)$$

Sumando (3.17) con (3.18) y tomando en cuenta que $\rho \leq 1/\alpha$

$$\Phi(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t)}) \geq \tilde{F}(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t+1)}) + \gamma \|\beta^{(t+1)}\|_1 + \frac{\alpha}{2} \|\beta^{(t+1)} - \beta^{(t)}\|_2^2. \quad (3.21)$$

Si $F(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t+1)}) \leq 2A$, entonces

$$\tilde{F}(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t+1)}) \geq F(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t+1)})$$

y así

$$\Phi(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t+1)}) \leq C.$$

Si $F(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t+1)}) > 2A$, de (3.21) tenemos que

$$2A \leq F(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t)}) + \gamma \|\beta^{(t)}\|_1 \leq 2(C + 2N^2 \ln 2).$$

Pero esto es una contradicción por la definición de A . Así,

$$\Phi(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t)}) \geq \Phi(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t+1)}) + \frac{\alpha}{2} \|\beta^{(t+1)} - \beta^{(t)}\|_2^2, \quad (3.22)$$

y entonces $\Phi(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t+1)}) \leq C$. De este modo, Φ_t es una sucesión decreciente, acotada superior e inferiormente. Por lo tanto, converge monótonamente.

Nos falta únicamente establecer (3.9). Por el lema 3.3.4 y las condiciones $\nabla_{\varphi} F(\varphi^{(t+1)}, \psi^{(t)}, \beta^{(t)}) = 0$, $\nabla_{\psi} F(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t)}) = 0$ tenemos que

$$\begin{aligned} F(\varphi^{(t)}, \psi^{(t)}, \beta^{(t)}) - F(\varphi^{(t+1)}, \psi^{(t)}, \beta^{(t)}) &\geq \frac{v}{2} \|\varphi^{(t+1)} - \varphi^{(t)}\|_2^2 \\ F(\varphi^{(t+1)}, \psi^{(t)}, \beta^{(t)}) - F(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t)}) &\geq \frac{v}{2} \|\psi^{(t+1)} - \psi^{(t)}\|_2^2. \end{aligned}$$

Sumando y teniendo en cuenta que

$$F(\varphi^{(t)}, \psi^{(t)}, \beta^{(t)}) - F(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t)}) = \Phi(\varphi^{(t)}, \psi^{(t)}, \beta^{(t)}) - \Phi(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t)})$$

llegamos a la desigualdad

$$\Phi(\varphi^{(t)}, \psi^{(t)}, \beta^{(t)}) - \Phi(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t)}) \geq \frac{v}{2} (\|\varphi^{(t+1)} - \varphi^{(t)}\|_2^2 + \|\psi^{(t+1)} - \psi^{(t)}\|_2^2). \quad (3.23)$$

Juntando (3.22) y (3.23)

$$\begin{aligned} \Phi(\varphi^{(t)}, \psi^{(t)}, \beta^{(t)}) - \Phi(\varphi^{(t+1)}, \psi^{(t+1)}, \beta^{(t+1)}) &\geq \frac{v}{2} (\|\varphi^{(t+1)} - \varphi^{(t)}\|_2^2 + \|\psi^{(t+1)} - \psi^{(t)}\|_2^2) \\ &\quad + \frac{\alpha}{2} \|\beta^{(t+1)} - \beta^{(t)}\|_2^2. \end{aligned} \quad (3.24)$$

$$+ \frac{\alpha}{2} \|\beta^{(t+1)} - \beta^{(t)}\|_2^2. \quad (3.25)$$

Dado que $\alpha \geq v$ (ver lema 3.3.4), definiendo $A_t = \Phi(x_t) - \Phi(x^*)$

$$A_{t-1} - A_t \geq \frac{v}{2} \|x_t - x_{t-1}\|_2^2, \quad \forall t \geq 1.$$

Queda por acotar A_t superiormente. Por la construcción de $\beta^{(t)}$

$$q_t = \frac{\beta^{(t-1)} - \beta^{(t)}}{\rho} - \nabla_{\beta} F(\varphi^{(t)}, \psi^{(t)}, \beta^{(t-1)}) \in \gamma \partial \|\cdot\|_1(\beta^{(t)}).$$

Así,

$$\Phi(x^*) \geq \Phi(x_t) + \nabla_{\varphi} F(x_t)(\varphi^* - \varphi^{(t)}) + \frac{v}{2} \|\varphi^{(t)} - \varphi^*\|_2^2 \quad (3.26)$$

$$+ \nabla_{\psi} F(x_t)(\psi^* - \psi^{(t)}) + \frac{\psi}{2} \|\psi^{(t)} - \psi^*\|_2^2 \quad (3.27)$$

$$+ (\nabla_{\beta} F(x_t) + q_t)(\beta^* - \beta^{(t)}) + \frac{v}{2} \|\beta^{(t)} - \beta^*\|_2^2. \quad (3.28)$$

En (3.26) se hace uso del lema 3.3.4 y la definición E.0.5. Luego, por la Desigualdad de Young:

$$|q \cdot z| \leq \frac{v}{2} \|z\|_2^2 + \frac{1}{2v} \|q\|_2^2. \quad (3.29)$$

Usando (3.29) en (3.26),

$$A_t \leq \frac{1}{2v} (\|\nabla_{\varphi} F(x_t)\|_2^2 + \|\nabla_{\psi} F(x_t)\|_2^2 + \|\nabla_{\beta} F(x_t) + q_t\|_2^2).$$

Como $\nabla_{\varphi} F(\varphi^{(t)}, \psi^{(t-1)}, \beta^{(t-1)}) = 0$, $\nabla_{\psi} F(\varphi^{(t)}, \psi^{(t)}, \beta^{(t-1)}) = 0$,

$$\|\nabla_{\varphi} F(\varphi^{(t)}, \psi^{(t)}, \beta^{(t)}) - \nabla_{\varphi} F(\varphi^{(t)}, \psi^{(t-1)}, \beta^{(t-1)})\|_2 \leq \alpha \|x_t - x_{t-1}\|_2^2$$

$$\|\nabla_{\varphi} F(\varphi^{(t)}, \psi^{(t)}, \beta^{(t)}) - \nabla_{\varphi} F(\varphi^{(t)}, \psi^{(t)}, \beta^{(t-1)})\|_2 \leq \alpha \|x_t - x_{t-1}\|_2^2$$

$$\|\nabla_{\varphi} F(x_t)\|_2^2 + \|\nabla_{\psi} F(x_t)\|_2^2 \leq 2\alpha^2 \|x_t - x_{t-1}\|_2^2.$$

Usando la bilinealidad del producto interno y el hecho que $\nabla_{\beta} F(\varphi^{(t)}, \psi^{(t)}, \cdot)$ es decreciente:

$$\begin{aligned} \|\nabla_{\beta} F(x_t) + q_t\|_2^2 &= \left\| \frac{\beta^{(t-1)} - \beta^{(t)}}{\rho} + \nabla_{\beta} F(x_t) - \nabla_{\beta} F(\varphi^{(t)}, \psi^{(t)}, \beta^{(t-1)}) \right\|_2^2 \\ &= \frac{1}{\rho^2} \|\beta^{(t-1)} - \beta^{(t)} + \rho(\nabla_{\beta} F(x_t) - \nabla_{\beta} F(\varphi^{(t)}, \psi^{(t)}, \beta^{(t-1)}))\|_2^2 \\ &= \frac{1}{\rho^2} \|\beta^{(t-1)} - \beta^{(t)}\|_2^2 + \|\nabla_{\beta} F(x_t) - \nabla_{\beta} F(\varphi^{(t)}, \psi^{(t)}, \beta^{(t-1)})\|_2^2 \\ &\quad + \frac{2}{\rho} (\beta^{(t-1)} - \beta^{(t)}) \cdot (\nabla_{\beta} F(x_t) - \nabla_{\beta} F(\varphi^{(t)}, \psi^{(t)}, \beta^{(t-1)})) \\ &\leq \left(\frac{1}{\rho^2} + \alpha^2 \right) \|\beta^{(t-1)} - \beta^{(t)}\|_2^2 \\ &\quad + \frac{2}{\rho} (\beta^{(t-1)} - \beta^{(t)}) \cdot (\nabla_{\beta} F(x_t) - \nabla_{\beta} F(\varphi^{(t)}, \psi^{(t)}, \beta^{(t-1)})) \\ &\leq \left(\frac{1}{\rho^2} + \alpha^2 \right) \|\beta^{(t-1)} - \beta^{(t)}\|_2^2. \end{aligned}$$

De este modo,

$$A_t \frac{1}{2v} \left(\frac{1}{\rho^2} + \alpha^2 + 2\alpha^2 \right) \|x_t - x_{t-1}\|_2^2 = \frac{3\alpha^2 + \rho^{-2}}{2v} \|x_t - x_{t-1}\|_2^2.$$

Definiendo

$$\delta = \frac{v^2 \rho^2}{2\alpha^2 \rho^2 + 1},$$

como

$$A_t \leq A_{t-1} - \frac{v}{2} \|x_t - x_{t-1}\|_2^2,$$

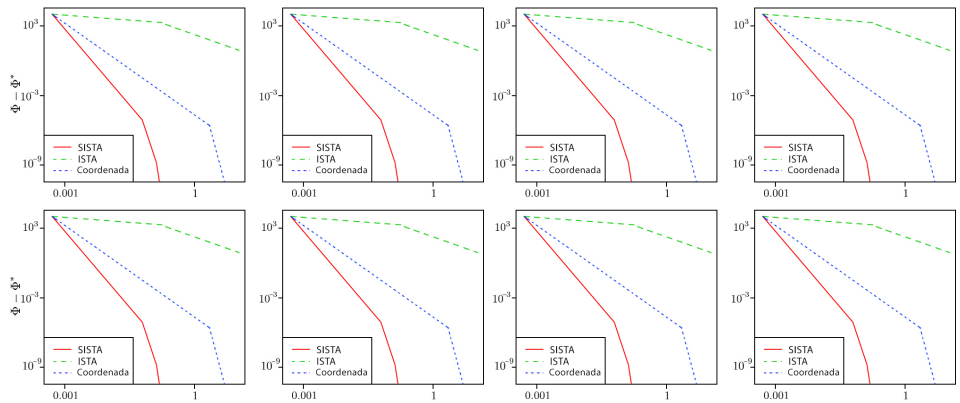
se sigue que

$$A_t \leq \frac{A_{t-1}}{1 + \delta}. \quad (3.30)$$

Finalmente, por inducción, concluimos que (3.30) implica (3.9). \square

3.4. Estudio de la migración

En Dupuy et al. (2021), los autores realizan una aplicación del algoritmo SISTA en el contexto del estudio de la migración. Los autores acceden a la probabilidad de que un migrante (escogido aleatoriamente) del país i se encuentre en el país j : $\hat{\pi}_{ij}$. Luego, recuperan a partir de las bases de datos del World Bank, Centre d'Études Prospectives de l'Information Internacionales y la Freedom House, las variables a usar en la matriz d . Aplicando el algoritmo SISTA obtienen que las variables más importantes en el emparejamiento son: el idioma, la relación colonial, la distancia geográfica, una dummy del estado de los servicios públicos, la interacción entre la superficie geográfica de los países, y una variable vinculada la esperanza de vida de las mujeres en el país de destino.



(A) $K = 100, N = 100$ (B) $K = 100, N = 200$ (C) $K = 500, N = 100$ (D) $K = 500, N = 200$

Figura 3.3 Comparación: traducido y editado del artículo original Dupuy et al. (2021).

Esto concluye nuestro estudio sobre el artículo [Dupuy et al. \(2021\)](#). Empezamos el siguiente capítulo presentando brevemente el modelo básico de elección discreta que será necesario para comprender las aplicaciones de la teoría de transporte óptimo en matching.

Capítulo 4

Aplicaciones en matching

Empezamos este capítulo con una breve introducción a la teoría de la elección discreta en microeconomía. Estas nociones serán necesarias para poder presentar unos ejemplos donde se plantean problemas de emparejamiento bajo la formulación de un problema de transporte. Seguimos a [Train \(2012\)](#).

Un agente de decisión n debe elegir entre $\{1, \dots, J\}$ opciones que le generan una utilidad $U_{nj} = V_{nj} + \varepsilon_{nj}$, $\forall j = 1, \dots, J$. El término V_{nj} es conocido, mientras que ε_{nj} es un término estocástico. Suponemos que los términos $\{\varepsilon_{nj}\}_j$ son independientes y se distribuyen según una Extreme Value de Tipo 1 ([de Haan and Ferreira, 2006](#)):

$$F(s) = e^{-e^{-s}}$$
$$f(s) = e^{-s} e^{-e^{-s}},$$

siendo $F(s)$ la función de distribución y $f(s)$ la función de densidad respecto a la medida de Lebesgue. Luego, siguiendo a [McFadden \(1974\)](#) y [Echenique and Chambers \(2016\)](#) la probabilidad de que el agente n escoja la alternativa i es

$$\mathbb{P}_{ni} = \mathbb{P}(V_{ni} + \varepsilon_{ni} \geq V_{nj} + \varepsilon_{nj}, \forall j \neq i).$$

Esto es,

$$\mathbb{P}(\varepsilon_{nj} \leq V_{ni} - V_{nj} + \varepsilon_{ni}, \forall j \neq i).$$

Usando la independencia,

$$\mathbb{P}(\varepsilon_{nj} \leq V_{ni} - V_{nj} + \varepsilon_{ni}, \forall j \neq i | \varepsilon_{ni}) = \prod_{j \neq i} e^{-e^{-(\varepsilon_{ni} + V_{ni} - V_{nj})}}.$$

Así,

$$\mathbb{P}_{ni} = \int_{\mathbb{R}} \underbrace{\left(\prod_{j \neq i} e^{-e^{-(s+V_{ni}-V_{nj})}} \right)}_{=\mathbb{P}(\epsilon_{nj} \leq V_{ni}-V_{nj}+\epsilon_{ni}, \forall j \neq i | \epsilon_{ni}=s)} \underbrace{e^{-s} e^{-e^{-s}}}_{f(s)} ds.$$

Como $e^{-e^{-s}} = e^{-e^{-(s+V_{ni}-V_{ni})}}$

$$\mathbb{P}_{ni} = \int_{\mathbb{R}} \left(\prod_j e^{-e^{-(s+V_{ni}-V_{nj})}} \right) e^{-s} ds.$$

Luego,

$$\prod_j e^{-e^{-(s+V_{ni}-V_{nj})}} = \exp \left\{ - \sum_j e^{-(V_{ni}-V_{nj}+s)} \right\}$$

y, por lo tanto,

$$\mathbb{P}_{ni} = \int_{\mathbb{R}} \exp \left(- \exp(s) \sum_j e^{-(V_{ni}-V_{nj})} \right) e^{-s} ds. \quad (4.1)$$

Sea $t = -e^{-s}$, $dt = e^{-s} ds$. Podemos entonces re-escribir 4.1 como sigue

$$\mathbb{P}_{ni} = \int_{-\infty}^0 \exp \left(t \sum_j e^{-(V_{ni}-V_{nj})} \right) dt.$$

De ahí,

$$\begin{aligned} \mathbb{P}_{ni} &= \frac{\exp \left(t \sum_j e^{-(V_{ni}-V_{nj})} \right) \Big|_0}{\sum_{j=1}^J e^{V_{nj}-V_{ni}}} \Big|_{-\infty}^0 \\ &= \frac{1}{\sum_{j=1}^J e^{V_{nj}-V_{ni}}} \\ &= \frac{e^{V_{ni}}}{\sum_{j=1}^J e^{V_{nj}}}. \end{aligned}$$

De este modo, la probabilidad de que n escoja a i es

$$\frac{e^{V_{ni}}}{\sum_{j=1}^J e^{V_{nj}}}.$$

Notación. A continuación, no distinguiremos explícitamente la suma discreta de la continua. Es decir, a veces denotamos

$$\sum_{x \in \mathcal{X}} \pi_{xy} = \int_{\mathcal{X}} \pi(x, y) dx$$

y

$$\sum_{y \in \mathcal{Y}} \pi_{xy} = \int_{\mathcal{Y}} \pi(x, y) dy.$$

Note que eso significa que las marginales son atomizadas. Más aún, por lo general, los conjuntos \mathcal{X} y \mathcal{Y} son finitos.

4.1. Marriage market

El siguiente modelo fue desarrollado en [Dupuy and Galichon \(2014\)](#) y [Dupuy et al. \(2017\)](#). En los problemas de transporte óptimo, se cuenta con una función de costos $c(x, y)$ o beneficios $\Phi(x, y)$, y se busca encontrar el plan de transporte $\pi(x, y)$ que resuelva

$$\min_{\pi \in \Pi(\mu, \nu)} \mathbb{E}_{\pi}[c(x, y)]$$

o bien

$$\max_{\pi \in \Pi(\mu, \nu)} \mathbb{E}_{\pi}[\Phi(x, y)], \quad (4.2)$$

donde $\Pi(\mu, \nu)$ es el conjunto de acoplamientos. Sin embargo (ver capítulo 3), se suele resolver el problema inverso: dado un emparejamiento o plan de transport óptimo (observado) $\hat{\pi}$, determinar la función Φ (un estimado $\hat{\Phi}$) - análogo para el costo c (\hat{c}).

En el modelo que consideramos, el objetivo es determinar la función de afinidad que rige el emparejamiento entre hombres y mujeres. Sean \mathcal{X} y \mathcal{Y} subconjuntos cerrados¹ de \mathbb{R}^L , donde cada individuo está representado por un vector $x \in \mathcal{X}$ y cada individua por un vector $y \in \mathcal{Y}$. Estos vectores encapsulan diversas características, tales como nivel educativo, edad, indicadores de personalidad, salario, entre otras.

Debido a la presencia de *shocks de simpatía*, que modelan las variables no observables en el proceso de emparejamiento, el problema clásico de transporte óptimo (4.2) se transforma en un problema de transporte regularizado (ver teorema 4.1.2)²:

$$\max_{\pi(x, y) \in \Pi(\mu, \nu)} \mathbb{E}_{\pi}[\Phi(x, y) - \sigma(\ln \pi(x, y) - 1)]. \quad (4.3)$$

A continuación, derivamos en detalle el modelo presentado en [Dupuy and Galichon \(2014\)](#).

Observación. Los artículos originales trabajan con un modelo de *logit continuo*. No obstante, en esta exposición nos limitamos al caso discreto, así como a conjuntos de igual cardinalidad. Estas suposiciones simplifican significativamente las derivaciones y facilitan la interpretación de los resultados.

¹Y, por lo tanto, medibles; ver el anexo B.

²Obsérvese que ahora la entropía se suma directamente en la función objetivo, mientras que en la minimización del costo se restaba. Se ha agregado un término constante que, a diferencia del modelo original, permite ciertas simplificaciones algebraicas.

Sean μ y ν las distribuciones de las características individuales, con densidades f y g respecto a la medida de Lebesgue:

$$\mu(A) = \int_A f(x)dx, \quad \nu(A) = \int_A g(x)dx.$$

Sin pérdida de generalidad, asumimos que $\mathbb{E}_\mu[X] = \mathbb{E}_\nu[Y] = 0$, donde X representa la variable aleatoria asociada a las características de los hombres y Y la de las mujeres. En este contexto, un *emparejamiento* se define como una probabilidad $\pi(x, y)$ que describe la ocurrencia de la pareja (x, y) . Dicho emparejamiento debe satisfacer las siguientes restricciones para ser un acoplamiento de μ y ν :

$$\Pi(\mu, \nu) = \left\{ \pi : \pi(x, y) \geq 0, \quad \int_Y \pi(x, y)dy = f(x), \quad \int_X \pi(x, y)dx = g(y) \right\}.$$

La función objetivo, conocida como *función de afinidad total* o *afinidad conjunta*, mide la calidad del emparejamiento. Así, el problema de optimización consiste en resolver:

$$\max_{\pi(x, y) \in \Pi(\mu, \nu)} \left\{ \int_{X \times Y} \Phi(x, y)\pi(x, y)dxdy \right\}.$$

Sin embargo, para capturar la heterogeneidad en las preferencias, se introducen los *shocks de simpatía*. Esto lleva a la maximización de la expresión:

$$\Phi(x, y) + \varepsilon_m(y) + \eta_w(x),$$

donde m y w representan hombre y mujer, respectivamente, y la función de afinidad conjunta se descompone como $\Phi(x, y) = U(x, y) + V(x, y)$ ³. Bajo esta formulación:

$$U(x, y) + \varepsilon_m(y)$$

representa la utilidad de un hombre con características x al emparejarse con una mujer de características y , mientras que

$$V(x, y) + \eta_w(x)$$

representa la utilidad de una mujer con características y al emparejarse con un hombre de características x . Finalmente, el problema de optimización consiste en maximizar

³Formalmente, esta descomposición es posible, pero determinar explícitamente U y V requiere el uso del algoritmo de Sinkhorn.

la suma de las utilidades aleatorias de ambos géneros, reflejando las preferencias individuales y las interacciones en el mercado de emparejamiento.

Se asume (Dupuy and Galichon, 2014) que el número de puntos $(y_k, \varepsilon_k) \in A$, denotado por N_A , sigue una distribución de Poisson $\text{Pois}(\theta(A))$, donde la medida θ está definida por $d\theta = e^{-\varepsilon} d\varepsilon dy$. Además, las variables N_A y N_B son independientes siempre que $A \cap B = \emptyset$. Este modelo para N_A es conocido como un *Poisson Point Process* (Karatzas and Shreve, 1991).

Cada individuo masculino m maximiza su utilidad estocástica, lo que nos permite definir la variable aleatoria

$$Z = \max_{y \in \mathcal{Y}} \{U(x, y_j) + \varepsilon_m(y_j)\}.$$

Para esta variable, se tiene la relación probabilística

$$\mathbb{P}(Z \leq c) = \underbrace{\mathbb{P}(U(x, y_j) + \varepsilon_m(y_j) \leq c, \forall y \in \mathcal{Y})}_{\#\{(y_j, \varepsilon_m(y_j)) \in \{(y, \varepsilon): U(x, y) + \varepsilon > c\}\} = 0}.$$

Esto equivale a

$$\mathbb{P}(Z \leq c) = \exp \left(\iint_{\mathcal{Y} \times \mathbb{R}} \mathbf{1}_{\{U(x, y) + \varepsilon > c\}} e^{-\varepsilon} d\varepsilon dy \right).$$

Desarrollando la integral, obtenemos

$$\begin{aligned} \ln \mathbb{P}(Z \leq c) &= - \iint_{\mathcal{Y} \times \mathbb{R}} \mathbf{1}_{\{U(x, y) + \varepsilon > c\}} e^{-\varepsilon} d\varepsilon dy \\ &= - \int_{\mathcal{Y}} \int_{c - U(x, y)}^{\infty} e^{-\varepsilon} d\varepsilon dy \\ &= - \exp \left(-c + \ln \int_{\mathcal{Y}} \exp(U(x, y')) dy' \right). \end{aligned}$$

Por lo tanto,

$$\mathbb{P}(Z \leq c) = \exp \left(- \exp \left(-c + \ln \int_{\mathcal{Y}} \exp(U(x, y')) dy' \right) \right).$$

Esto implica que $Z \sim \text{Gumbel}(\ln \int_{\mathcal{Y}} \exp(U(x, y')) dy', 1)$, lo que permite obtener la siguiente expresión para su valor esperado:

$$\mathbb{E}[Z] = \ln \left(\int_{\mathcal{Y}} \exp(U(x, y')) dy \right). \quad (4.4)$$

Por otro lado, la probabilidad condicional de emparejamiento está dada por

$$\pi(y|x) = \frac{e^{U(x,y)}}{\int_{\mathcal{Y}} \exp(U(x,y')) dy'}.$$

Finalmente, utilizando la notación establecida, obtenemos las siguientes expresiones⁴:

$$\pi(y|x) = \exp \left\{ U(x,y) - \ln \sum_{y' \in \mathcal{Y}} \exp(U(x,y')) \right\}, \quad (4.5)$$

$$\pi(x|y) = \exp \left\{ U(x,y) - \ln \sum_{x' \in \mathcal{X}} \exp(U(x',y)) \right\}. \quad (4.6)$$

Observación. Si se realiza un reescalamiento de la forma $\varepsilon \rightarrow \frac{\varepsilon\sigma}{2}$, donde σ es un parámetro positivo, se obtiene la siguiente forma de las probabilidades condicionales:

$$\begin{aligned} \pi(y|x) &= \frac{\exp \left[\frac{U(x,y)}{\sigma/2} \right]}{\int_{\mathcal{Y}} \exp \left[\frac{U(x,y')}{\sigma/2} \right] dy'}, \\ \pi(x|y) &= \frac{\exp \left[\frac{U(x,y)}{\sigma/2} \right]}{\int_{\mathcal{X}} \exp \left[\frac{U(x',y)}{\sigma/2} \right] dx'}. \end{aligned}$$

Proposición 4.1.1. El plan de transporte óptimo es tal que

$$\ln \pi(x,y) = \frac{\Phi(x,y) - a(x) - b(y)}{\sigma}, \quad (4.7)$$

donde a y b se determinan de forma que $\pi \in \Pi(\mu, \nu)$.

Demostración. Hacemos uso de (4.5) para obtener (4.7). Por otro lado, la segunda afirmación será consecuencia de lo derivado en el capítulo 2 y el teorema 4.1.2. Como $\pi(y|x)f(x) = \pi(x,y)$,

$$\pi(x,y) = \frac{\exp \left[\frac{U(x,y)}{\sigma/2} \right] f(x)}{\int_{\mathcal{Y}} \exp \left[\frac{U(x,y')}{\sigma/2} \right] dy'}.$$

Luego,

$$\frac{\sigma}{2} \ln \pi(x,y) = U(x,y) + \frac{\sigma}{2} \left\{ \ln f(x) - \ln \left(\int_{\mathcal{Y}} \exp \left[\frac{U(x,y')}{\sigma/2} \right] dy' \right) \right\}.$$

⁴Para la segunda expresión, basta reemplazar y por x : el problema es simétrico.

Análogamente,

$$\frac{\sigma}{2} \ln \pi(x, y) = V(x, y) + \frac{\sigma}{2} \left\{ \ln g(y) - \ln \left(\int_{\mathcal{X}} \exp \left[\frac{V(x', y)}{\sigma/2} \right] dx' \right) \right\}.$$

Así, sumando,

$$\ln \pi(x, y) = \frac{\underbrace{U(x, y) + V(x, y)}_{=\Phi(x, y)} - a(x) - b(y)}{\sigma}$$

con

$$a(x) = \frac{\sigma}{2} \ln \int_{\mathcal{Y}} \frac{e^{\frac{U(x, y')}{\sigma/2}}}{f(y')} dy'$$

$$b(y) = \frac{\sigma}{2} \ln \int_{\mathcal{X}} \frac{e^{\frac{U(x', y)}{\sigma/2}}}{g(x')} dx'.$$

□

Teorema 4.1.2. El problema de maximización, relativo a los shocks aleatorios, es el siguiente

$$\mathcal{W} = \max_{\pi(x, y) \in \Pi(\mu, \nu)} \left\{ \iint_{\mathcal{X} \times \mathcal{Y}} \Phi(x, y) \pi(x, y) dx dy - \sigma \iint_{\mathcal{X} \times \mathcal{Y}} \ln \pi(x, y) \pi(x, y) dx dy \right\}.$$

Demostración. Seguimos la prueba hecha en [Dupuy and Galichon \(2014\)](#). Los resultados de dualidad [Ambrosio et al. \(2021\)](#) permiten establecer que

$$\mathcal{W} = \inf_{u_m + v_w \geq \Phi(x_m, y_w) + \frac{\sigma}{2} \varepsilon_m(y) + \frac{\sigma}{2} \eta_w(x)} \left\{ \int u_m dm + \int v_w dw \right\}. \quad (4.8)$$

Luego, la restricción puede escribirse

$$U(x, y) + V(x, y) \geq \Phi(x, y)$$

con

$$U(x, y) = \inf_m \left\{ u_m - \frac{\sigma}{2} \varepsilon_m(y) \right\}$$

$$V(x, y) = \inf_w \left\{ v_w - \frac{\sigma}{2} \eta_w(x) \right\}.$$

A su vez, esto implica que

$$u_m = \sup_{y \in \mathcal{Y}} \left\{ U(x, y) + \frac{\sigma}{2} \varepsilon_m(y) \right\}$$

$$v_w = \sup_{x \in \mathcal{X}} \left\{ V(x, y) + \frac{\sigma}{2} \eta_w(x) \right\}.$$

De ahí, denotando

$$\begin{aligned} G_x(U(x, \cdot)) &= \mathbb{E} \left[\max_k \left\{ U(x, y_k^m) + \frac{\sigma}{2} \varepsilon_k^m \right\} \right] \\ H_y(V(\cdot, y)) &= \mathbb{E} \left[\max_l \left\{ V(x_l^m, y) + \frac{\sigma}{2} \eta_l^w \right\} \right] \end{aligned}$$

el problema (4.8) puede escribirse de la siguiente manera

$$\min_{U, V: U(x) + V(y) \geq \phi(x, y)} \left\{ \int_{\mathcal{X}} G_x(U(x, \cdot)) f(x) dx + \int_{\mathcal{Y}} H_y(V(\cdot, y)) g(y) dy \right\}. \quad (4.9)$$

A partir de (4.9), podemos plantear el problema desde la perspectiva del Lagrangiano:

$$\mathcal{W} = \inf_{U, V} \sup_{\pi} \int_{\mathcal{X}} G_x(U(x, \cdot)) f(x) dx + \int_{\mathcal{Y}} H_y(V(\cdot, y)) g(y) dy \quad (4.10)$$

$$+ \iint_{\mathcal{X} \times \mathcal{Y}} (\Phi(x, y) - U(x, y) - V(x, y)) \pi(x, y) dx dy. \quad (4.11)$$

Esta situación puede re-escribirse de la siguiente manera:

$$\sup_{\pi} \int_{\mathcal{X} \times \mathcal{Y}} \Phi(x, y) \pi(x, y) dx dy - \mathcal{I}(\pi) \quad (4.12)$$

con

$$\mathcal{I}(\pi) = \sup_U (*) + \sup_V (**)$$

$$\begin{aligned} * &= \int_{\mathcal{X} \times \mathcal{Y}} U(x, y) \pi(x, y) dx dy - \int_{\mathcal{X}} G_x(U(x, \cdot)) f(x) dx \\ ** &= \int_{\mathcal{X} \times \mathcal{Y}} V(x, y) \pi(x, y) dx dy - \int_{\mathcal{Y}} H_x(V(\cdot, y)) g(y) dy. \end{aligned}$$

Recordemos de (4.4) que

$$\begin{aligned} G_x(U(x, \cdot)) &= \frac{\sigma}{2} \ln \int_{\mathcal{Y}} \exp \left(\frac{U(x, y)}{\sigma/2} \right) dy \\ H_x(V(\cdot, y)) &= \frac{\sigma}{2} \ln \int_{\mathcal{X}} \exp \left(\frac{V(x, y)}{\sigma/2} \right) dx. \end{aligned}$$

Aplicando condiciones de primer orden a⁵

$$\iint_{\mathcal{X} \times \mathcal{Y}} U(x, y) \pi(x, y) dx dy - \int_{\mathcal{X}} G_x(U(x, \cdot)) f(x) dx \quad (4.13)$$

⁵Recordemos que, haciendo abuso de notación, seguimos en el caso discreto en el cual $U(x, y) = U_{x,y}$ y $\pi(x, y) = \pi_{x,y}$. O sea, $\sum_A = \int_A$.

respecto a $U(x, y)$:

$$\pi(x, y) = \frac{f(x) \exp\left(\frac{U(x, y)}{\sigma/2}\right)}{\int_{\mathcal{Y}} \exp\left(\frac{U(x, y)}{\sigma/2}\right) dy} dy.$$

Esto pues,

$$\frac{d}{dU(x, y)} \left\{ \frac{\sigma}{2} \ln \int_{\mathcal{Y}} \exp\left(\frac{U(x, y)}{\sigma/2}\right) dy \right\} = \frac{\sigma}{2} \frac{1}{\sigma/2} \frac{\exp\left(\frac{U(x, y)}{\sigma/2}\right)}{\int_{\mathcal{Y}} \exp\left(\frac{U(x, y)}{\sigma/2}\right) dy}.$$

Así, podemos re-escribir (4.13)⁶ como sigue

$$\frac{\sigma}{2} \iint_{\mathcal{X} \times \mathcal{Y}} \pi(x, y) \ln \left(\frac{\pi(x, y)}{f(x)} \right) dx dy.$$

Procediendo de manera análoga para la expresión en $V(x, y)$,

$$\mathcal{I}(\pi) = \sigma \iint_{\mathcal{X} \times \mathcal{Y}} \ln \pi(x, y) \pi(x, y) dx dy - \frac{\sigma}{2} \left[\int_{\mathcal{X}} \ln f(x) f(x) dx + \int_{\mathcal{Y}} \ln g(y) g(y) dy \right].$$

Pero entonces, (4.12) es equivalente a

$$\sup_{\pi} \left\{ \int_{\mathcal{X} \times \mathcal{Y}} \Phi(x, y) \pi(x, y) dx dy - \sigma \iint_{\mathcal{X} \times \mathcal{Y}} \ln \pi(x, y) \pi(x, y) dx dy \right\}.$$

Además, usando el hecho que $\pi(x, y) \in \Pi(\mu, \nu)$ y μ, ν son dadas, esto es lo mismo que resolver

$$\sup_{\pi} \left\{ \int_{\mathcal{X} \times \mathcal{Y}} \Phi(x, y) \pi(x, y) dx dy - \sigma \iint_{\mathcal{X} \times \mathcal{Y}} (\ln \pi(x, y) - 1) \pi(x, y) dx dy \right\}.$$

□

Corolario 4.1.3. El plan de transporte es tal que

$$\pi(x, y) = \exp \left[\frac{\Phi(x, y) - a(x) - b(y)}{\sigma} \right] \quad (4.14)$$

con

$$a(x) = -\frac{\sigma}{2} \ln \left(\frac{f(x)}{\int_{\mathcal{Y}} \exp U(x, y) dy} \right), \quad b(y) = -\frac{\sigma}{2} \ln \left(\frac{g(y)}{\int_{\mathcal{X}} \exp V(x, y) dx} \right).$$

El modelamiento que incorpora shocks de simpatía (en el contexto del mercado del matrimonio), conlleva a una formulación de regularización entrópica y su resolución, puede efectuarse vía los algoritmos presentados en secciones anteriores. A continuación presentamos brevemente la parametrización escogida en [Dupuy and Galichon \(2014\)](#).

⁶Recordando que $\int_{\mathcal{Y}} \pi(x, y) dy = f(x)$.

4.2. Estimación y parametrización

A continuación, presentamos un método para estimar la función de beneficios Φ en un caso particular que, aunque específico, tiene aplicaciones en una amplia variedad de contextos. Recordemos que el problema de optimización a resolver es el siguiente:

$$\max_{\pi \in \Pi} \mathbb{E}_{\pi}[\Phi(x, y) - \sigma(\ln \pi(x, y) - 1)].$$

En el caso discreto, esta expresión se re-escribe como:

$$\max_{\pi \in \Pi} \left\{ \sum_{x, y} \pi_{xy} \Phi_{xy} - \sigma \pi_{xy} (\ln \pi_{xy} - 1) \right\}. \quad (4.15)$$

Este problema de optimización⁷ en las variables $\{\pi_{xy}\}$ es convexo y posee una solución única. No obstante, al igual que en el capítulo 3, el objetivo es, dado un emparejamiento observable $\hat{\pi}$, inferir la función de beneficios Φ implícita en los datos.

Siguiendo a Dupuy et al. (2017), se asume que Φ adopta la forma

$$\Phi_A(x, y) = x^T A y.$$

Además, se normaliza $\sigma = 1$. En este caso, como

$$\frac{\partial^2 \Phi_A}{\partial x_i \partial y_j} = A_{ij} = \frac{\partial^2 \ln \pi^A}{\partial x_i \partial y_j},$$

el signo de A_{ij} determina la naturaleza de la interacción entre las características x e y :

- Si $A_{ij} > 0$, la interacción es positiva.
- Si $A_{ij} = 0$, no existe interacción.
- Si $A_{ij} < 0$, la interacción es negativa o repulsiva.

El problema ahora consiste en estimar la matriz A utilizando una muestra de N observaciones $(x^k, y^k) \in \mathbb{R}^L \times \mathbb{R}^L$, para $k = 1, \dots, N$. La medida empírica correspondiente es

$$\hat{\pi}(x, y) = \frac{1}{N} \sum_{k=1}^N \delta_{x^k}(x) \delta_{y^k}(y),$$

con marginales $\mu_1(x) = \sum_{k=1}^N \delta_{x^k}(x)$ y $\mu_2(y) = \sum_{k=1}^N \delta_{y^k}(y)$. Aquí, δ denota la medida de Dirac (2.2).

⁷Se trata de un problema de regularización entrópica.

Proposición 4.2.1. Dado un muestreo aleatorio independiente, la función de verosimilitud $\ell(A; \hat{\pi})$ está dada por

$$\ell(A; \hat{\pi}) = N \mathbb{E}_{\hat{\pi}}[\ln \pi^A(x, y)] = N \{ \mathbb{E}_{\hat{\pi}}[\Phi_A(x, y)] - \mathcal{W}(A) \},$$

donde

$$\mathcal{W}(A) = \max_{\pi \in \Pi} \mathbb{E}_{\pi}[\Phi_A(x, y) - \sigma(\ln \pi(x, y) - 1)]. \quad (4.16)$$

Demostración. Usando la independencia de las observaciones:

$$\begin{aligned} \ell(A; \hat{\pi}) &= \ln \left(\prod_{k=1}^N \pi^A(x^k, y^k) \right) \\ &= \sum_{k=1}^N \ln \pi^A(x^k, y^k) \\ &= N \mathbb{E}_{\hat{\pi}}[\ln \pi^A(x, y)]. \end{aligned}$$

Luego, utilizando la definición de $\pi^A(x, y)$,

$$\ln \pi^A(x^k, y^k) = \ln a(x^k) + \ln b(y^k) + \Phi_A(x^k, y^k).$$

Finalmente, como $\pi^A, \hat{\pi} \in \Pi(\mu, \nu)$, se obtiene

$$\mathbb{E}_{\hat{\pi}}[\ln \pi^A(x, y)] = \mathbb{E}_{\hat{\pi}}[\Phi_A(x, y)] - \mathcal{W}(A).$$

□

Proposición 4.2.2. La función $\ell(A; \hat{\pi})$ es cóncava en A y su gradiente está dado por

$$\nabla_A \ell(A; \hat{\pi})_{i,j} = N \{ \mathbb{E}_{\hat{\pi}}[x_i y_j] - \mathbb{E}_{\pi^A}[x_i y_j] \}.$$

Corolario 4.2.3. Si \hat{A} es la solución óptima, entonces

$$\mathbb{E}_{\hat{\pi}}[x_i y_j] - \mathbb{E}_{\pi^{\hat{A}}}[x_i y_j] = 0.$$

Observación. Si las variables X e Y están centradas (por ejemplo, normalizadas), la condición dada por el corolario 4.2.3 se reduce a:

$$\text{Cov}_{\pi^A}(x, y) = \text{Cov}_{\hat{\pi}}(x, y).$$

El problema de optimización final se expresa como

$$\begin{aligned} & \max_A \ell(A; \hat{\pi}) \\ & \text{s.a. } \text{Rango}(A) \leq r. \end{aligned}$$

Dada su complejidad computacional, este problema se reformula como:

$$\min_A \{ \mathcal{W}(A) - \mathbb{E}_{\hat{\pi}}[\Phi_A(x, y)] + \lambda \|A\|_* \}, \quad (4.17)$$

donde $\|A\|_*$ representa la suma de los valores singulares de A (Gentle, 2017). La solución numérica se obtiene mediante un proximal gradient descent en combinación con el algoritmo de Sinkhorn. Este método, similar a SISTA, garantiza convergencia gracias a los resultados en Toh and Yun (2010).

Algorithm 3 Estimación de la matriz A

- 1: **Entrada:** Matriz inicial A , tamaño de paso θ , emparejamiento observado $\{(x^k, y^k)\}_{k=1, \dots, N}$, parámetro de regularización λ .
- 2: Recursive:
- 3: **while** el criterio de convergencia no se cumple **do** Aplicar el algoritmo de Sinkhorn para estimar π^A .
- 4: Actualizar la matriz A :

$$A \leftarrow A - \theta \sum_{i,j=1}^N (\pi_{ij}^A - \hat{\pi}_{ij}) x^i (y^j)^T.$$

- 5: Descomposición en valores singulares (SVD):

$$[U, D(s_1, \dots, s_d), V] = \text{SVD}(A).$$

- 6: Aplicar la penalización nuclear:

$$A \leftarrow UD((s_1 - t\lambda)^+, \dots, (s_d - t\lambda)^+)V^T.$$

- 7: **end while**

- 8: **return** A .
-

La descomposición en valores singulares (SVD) se refiere a la factorización de la matriz A en la forma:

$$A = USV^T,$$

donde U y V son matrices ortogonales y S es una matriz diagonal con valores no negativos. El rango de A está acotado por $r \leq \min\{\dim(x), \dim(y)\}$. Para más detalles, véase [Gentle \(2017\)](#). Por otro lado, $x^+ = \max\{0, x\}$.

Al igual que en el algoritmo SISTA, la convergencia del procedimiento propuesto en [Dupuy et al. \(2017\)](#) está garantizada por los resultados de [Toh and Yun \(2010\)](#). El objetivo del estudio en [Dupuy and Galichon \(2014\)](#) y [Dupuy et al. \(2017\)](#) es determinar los factores que influyen en la elección de pareja entre hombres y mujeres.

En [Dupuy et al. \(2017\)](#), se implementa este marco teórico para analizar datos de la encuesta Dutch Household Survey (DHS), incorporando variables como educación, peso, índice de masa corporal (IMC) y rasgos de personalidad basados en el 16 PF Test. En total, se consideran 26 dimensiones, por lo que la matriz A tiene dimensiones 26×26 (i.e., $A \in \mathcal{M}_{26 \times 26} \simeq \mathbb{R}^{676}$). Posteriormente, se realiza una reducción de dimensionalidad mediante una serie de indicadores.

Los resultados principales del estudio indican que la característica más relevante en la elección de pareja para las mujeres es la tendencia a ser *fácilmente ofendido*, mientras que para los hombres es el rasgo de *disciplinada*.

Para ilustrar cómo se puede aplicar esta metodología en otro contexto, consideremos el estudio del emparejamiento en el mercado educativo universitario. El primer paso consiste en identificar las variables clave para la investigación y generar indicadores basados en sumas ponderadas del tipo:

$$Y = \sum_{j=1}^k \theta_j X_j.$$

Posteriormente, utilizando técnicas de reducción de dimensionalidad provenientes del aprendizaje automático ([Murphy, 2022](#); [James et al., 2021](#)), se pueden agrupar los individuos en *clusters* según sus características. De este modo, cada grupo de individuos estará representado por un vector x , mientras que cada universidad estará representada por un vector y . Además, se obtiene una matriz de transporte $\hat{\pi}$ tal que $\hat{\pi}_{xy}$ indica la proporción de individuos con características x que terminan matriculados en una universidad con características y .

Siguiendo un enfoque similar al de [Dupuy et al. \(2017\)](#), se puede plantear una especificación funcional para el costo del emparejamiento:

$$c(x, y) = x^T A y. \quad (4.18)$$

A partir de esta formulación, la matriz A puede estimarse utilizando el procedimiento descrito en esta sección.

4.3. Estimación de beneficios laborales y productividad laboral

Pasamos del *mercado del matrimonio* al mercado laboral. El artículo [Dupuy and Galichon \(2022\)](#) introduce un estimador de máxima verosimilitud para el valor de comodidades⁸ laborales y productividad laboral en un mercado de emparejamiento único, basado en los equilibrios de emparejamiento y salarios observados. Los autores, si bien mencionan que su estimador se acomoda a varias situaciones, ajustan el modelo al caso de compensaciones por accidente en el trabajo (usan datos para Estados Unidos 2017). El set-up del modelo es similar al que se presentó para [Dupuy and Galichon \(2014\)](#). Las modificaciones siguen fundamentalmente las ideas presentadas en [Echenique et al. \(2023\)](#) (capítulo 26).

Demanda y oferta. Se define el salario $w(x, y)$ del trabajador de tipo x en la firma y . No obstante, la decisión del trabajador no depende únicamente del salario, sino también de ciertos factores adicionales de comodidad. Estos se descomponen en un componente sistemático $\alpha(x, y)$ y un componente estocástico $\varepsilon(y)$. La utilidad total percibida por el trabajador es:

$$\alpha(x, y) + w(x, y) + \sigma_1 \varepsilon(y).$$

Se asume que ε sigue una distribución de Gumbel (Extreme Value Tipo 1). A continuación, se describe el proceso de emparejamiento:

1. Los trabajadores seleccionan aleatoriamente un subconjunto finito de firmas dentro de \mathcal{Y} , denominado *random pool*. Para simplificar la exposición y aplicar la teoría de elección discreta, consideramos \mathcal{Y} como un conjunto finito, aunque en casos más generales podría ser continuo.
2. El conjunto de firmas elegibles se modela mediante un Poisson Point Process, como se explicó en la sección 4.1.

⁸Se puede entender como beneficios a los trabajadores.

3. Se asume que este proceso está definido sobre $\mathcal{Y} \times \mathbb{R}$, con intensidad $dye^{-\varepsilon}d\varepsilon$.

De manera análoga, las firmas perciben un beneficio neto dado por:

$$\gamma(x, y) - w(x, y) + \sigma_2\eta(x),$$

donde $\gamma(x, y)$ representa los beneficios productivos de la firma, $w(x, y)$ es el salario y $\eta(x)$ es un shock de productividad.

Definición 4.3.1. Se define la utilidad indirecta aleatoria como:

$$\tilde{U} = \max_{y \in \mathcal{Y}} \left\{ \underbrace{U(x, y)}_{=\alpha(x, y) + w(x, y)} + \sigma_1\varepsilon(y) \right\}.$$

A partir de la sección 4.1, se obtiene que:

$$\begin{aligned} \pi(y|x) &= \frac{\exp\left(\frac{\alpha(x, y) + w(x, y)}{\sigma_1}\right)}{\int_{\mathcal{Y}} \exp\left(\frac{\alpha(x, y') + w(x, y')}{\sigma_1}\right) dy'}, \\ \pi(x|y) &= \frac{\exp\left(\frac{\gamma(x, y) - w(x, y)}{\sigma_2}\right)}{\int_{\mathcal{X}} \exp\left(\frac{\gamma(x', y) - w(x', y)}{\sigma_2}\right) dx'}. \end{aligned}$$

De manera más compacta:

$$\begin{aligned} \pi(y|x) &= \exp\left(\frac{\alpha(x, y) + w(x, y) - u(x)}{\sigma_1}\right), \\ \pi(x|y) &= \exp\left(\frac{\gamma(x, y) - w(x, y) - v(y)}{\sigma_2}\right), \end{aligned}$$

donde

$$\begin{aligned} u(x) &= \sigma_1 \ln \int_{\mathcal{Y}} \exp\left(\frac{\alpha(x, y') + w(x, y')}{\sigma_1}\right) dy', \\ v(y) &= \sigma_2 \ln \int_{\mathcal{X}} \exp\left(\frac{\gamma(x', y) - w(x', y)}{\sigma_2}\right) dx'. \end{aligned}$$

Definición 4.3.2. Equilibrio. Un salario de equilibrio $w(x, y)$ es aquel para el cual la densidad $\pi(x, y)$ obtenida del problema de los trabajadores coincide con la densidad resultante del problema de las firmas. Esto implica:

$$\exp\left(\frac{\alpha(x, y) + w(x, y) - a(x)}{\sigma_1}\right) = \pi(x, y) = \exp\left(\frac{\gamma(x, y) - w(x, y) - b(y)}{\sigma_2}\right), \quad (4.19)$$

con

$$\begin{cases} a(x) &= u(x) - \sigma_1 \ln f(x), \\ b(y) &= v(y) - \sigma_2 \ln g(y). \end{cases}$$

Corolario 4.3.3. Denotando $\sigma = \sigma_1 + \sigma_2$, en equilibrio el salario satisface:

$$w(x, y) = \frac{\sigma_1}{\sigma}(\gamma(x, y) - b(y)) + \frac{\sigma_2}{\sigma}(a(x) - \alpha(x, y)). \quad (4.20)$$

Corolario 4.3.4. Se tiene que:

$$\pi(x, y) = \exp\left(\frac{\phi(x, y) - a(x) - b(y)}{\sigma}\right),$$

donde $\phi = \gamma + \alpha$.

Observación. Dado que a priori no se conocen $\alpha(x, y)$ y $\gamma(x, y)$, es necesario determinar a y b para satisfacer:

$$\begin{cases} \int_{\mathcal{Y}} \exp\left(\frac{\phi(x, y) - a(x) - b(y)}{\sigma}\right) dy = f(x), \\ \int_{\mathcal{X}} \exp\left(\frac{\phi(x, y) - a(x) - b(y)}{\sigma}\right) dx = g(y). \end{cases}$$

Una opción para ello es el algoritmo de Sinkhorn.

Estimación paramétrica. El conjunto de observaciones está compuesto por las ternas $\{X_i, Y_i, W_i\}_{i=1}^n$, donde X_i y Y_i representan los vectores de características de empleados y firmas, respectivamente, mientras que $W_i = w(X_i, Y_i) + \epsilon_i$ es el salario observado, con un término de error $\epsilon_i \sim \mathcal{N}(0, s^2)$. El objetivo es estimar las funciones $\alpha(x, y)$ (beneficio del trabajador) y $\gamma(x, y)$ (productividad de la firma). Recordemos que $\phi(x, y)$ representa la ganancia conjunta del emparejamiento. A partir de los corolarios 4.3.3 y 4.3.4, se obtiene:

$$\alpha(x, y) = \sigma_1 \ln \pi(x, y) - w(x, y) + c(x),$$

$$\gamma(x, y) = \sigma_2 \ln \pi(x, y) + w(x, y) - d(y).$$

El emparejamiento $\pi(x, y)$ se modela como:

$$\pi(x, y) = \exp(\phi(x, y) - a(x) - b(y)),$$

donde las funciones $a(x)$ y $b(y)$ deben satisfacer las siguientes ecuaciones de restricción:

$$\begin{cases} \int_Y \exp(\phi(x, y) - a(x) - b(y)) dy = f(x), \\ \int_X \exp(\phi(x, y) - a(x) - b(y)) dx = g(y). \end{cases} \quad (4.21)$$

Además, se definen las siguientes funciones auxiliares:

$$\begin{aligned} u(x) &= \sigma a(x) + \sigma_1 \ln f(x) + t, \\ v(y) &= \sigma b(y) + \sigma_2 \ln g(y) - t. \end{aligned}$$

Utilizando estas expresiones, el salario en equilibrio se puede escribir como:

$$w(x, y) = \sigma_1(\gamma(x, y) - b(y)) + \sigma_2(a(x) - \alpha(x, y)) + t.$$

Para estimar los parámetros, se asume la siguiente parametrización para $\alpha(x, y)$ y $\gamma(x, y)$, en términos de una base de funciones $\{\varphi_k\}_{k=1}^K$ linealmente independientes:

$$\begin{aligned} \alpha(x, y; A) &= \sum_{k=1}^K A_k \varphi_k(x, y), \\ \gamma(x, y; \Gamma) &= \sum_{k=1}^K \Gamma_k \varphi_k(x, y). \end{aligned}$$

De esta manera, la función de ganancia conjunta se expresa como:

$$\phi(x, y; \Phi) = \sum_{k=1}^K \Phi_k \varphi_k(x, y), \quad \text{con } \Phi_k = \Gamma_k + A_k.$$

El modelo queda entonces completamente parametrizado por el vector de parámetros:

$$\theta = (A, \Gamma, \sigma_1, \sigma_2, t, s^2).$$

Dado que $\pi(x, y)$ está modelado en términos de Φ , se tiene:

$$\pi(x, y; \Phi) = \exp(\phi(x, y; \Phi) - a(x; \Phi) - b(y; \Phi)),$$

donde $a(x; \Phi)$ y $b(y; \Phi)$ se determinan resolviendo el sistema (4.21). Finalmente, la función de log-verosimilitud de la muestra es:

$$\ln L(\theta) = n \mathbb{E}_{\hat{\pi}} \left[\phi(X, Y; \Phi) - a(X; \Phi) - b(Y; \Phi) - \frac{(W - w(X, Y; \theta))^2}{2s^2} \right] - \frac{n}{2} \ln s^2.$$

Aquí, $\hat{\pi}$ representa la densidad empírica observada. Al maximizar esta log-verosimilitud, se obtiene una estimación para los parámetros π_{ij}^Φ . Para una exposición más detallada, se recomienda consultar directamente [Dupuy and Galichon \(2022\)](#). Si bien las ideas principales y más importantes ya fueron expuestas en este documentos, los detalles del artículo son necesarios si se quiere profundizar e investigar en la misma línea.

4.4. Mismatching ocupacional y el mercado laboral peruano

En esta última sección, exploramos una aplicación de la teoría revisada a lo largo de este documento en el contexto del mercado laboral peruano. Siguiendo la línea de [Burga and Moreno \(2001\)](#), introducimos una idea de índole metodológica para modelar el emparejamiento ocupacional mediante técnicas de transporte óptimo regularizado. Esta aproximación permite un análisis detallado de los factores que determinan la asignación de trabajadores a ocupaciones, aprovechando recientes avances en teoría económica ([Galichon, 2016](#)).

Desde mediados del siglo XX, los países en desarrollo, incluida América Latina, experimentaron una expansión educativa sin precedentes. En el Perú, las tasas de matrícula crecieron de manera significativa, con aumentos en la educación primaria (61.1 % a 85.5 %), secundaria (13.3 % a 48 %) y superior (1.8 % a 11.7 %). Si bien la inversión en educación suele estar asociada con mayores ingresos y mejores oportunidades laborales, en muchos casos ha generado una sobreoferta de trabajadores calificados para empleos que no requieren dichas habilidades. Esto ha resultado en altos niveles de subempleo y desempleo en los sectores más educados. En este contexto, el mismatching ocupacional, definido como la discordancia entre la formación profesional y el empleo desempeñado, se ha convertido en un problema clave en la economía laboral peruana.

El Perú presenta una alta incidencia de mismatching ocupacional, reflejada en la inserción laboral de individuos en puestos que no corresponden a su formación. A diferencia del enfoque de [Burga and Moreno \(2001\)](#), en lugar de medir el grado de discordancia en términos de escolaridad excedente (*overeducation*), proponemos modelar directamente la asignación de trabajadores a ocupaciones mediante una estructura de costos paramétrica. En este sentido, empleamos una formulación basada en el criterio de disimilitud [Dupuy et al. \(2021\)](#), en la que el costo de emparejamiento se define como

$$c_{ij}^{\beta} = \sum_{k=1}^K \beta_k d_{ij}^k,$$

donde d_{ij}^k captura diferencias clave entre la ocupación j y el perfil del trabajador i .

Las matrices de disimilitud d_{ij}^k pueden construirse a partir de variables relevantes extraídas de encuestas como ENAHO y ENDES. Por ejemplo, entre los factores que pueden influir en el emparejamiento laboral destacan:

- Diferencias en la edad promedio de los trabajadores en cada ocupación.
- Brechas en los años de educación requeridos y alcanzados.
- Niveles de habilidades técnicas y blandas demandadas por ocupaciones específicas.
- Localización geográfica y accesibilidad al empleo (movilidad laboral).
- Condiciones de formalidad e informalidad en el mercado laboral.

Para la estimación de los costos de emparejamiento, se emplea el algoritmo SISTA. Esto permite identificar los factores más determinantes en la asignación laboral y cuantificar el impacto de la sobreeducación dentro de la estructura de costos. A diferencia de [Burga and Moreno \(2001\)](#), donde se definía a un trabajador sobreeducado si su nivel de escolaridad superaba la media más una desviación estándar ($\mu + \sigma$), en nuestro enfoque, la sobreeducación es una característica endógena dentro de la función de costos.

El desafío clave en este estudio es la construcción de las matrices $(d_{ij})^k$ y la estimación del parámetro $\beta \in \mathbb{R}^K$. Para ello, se incorpora una penalización LASSO ($\|\beta\|_1$), lo que permite seleccionar automáticamente las variables más relevantes. Un ejemplo concreto de construcción de estas matrices es el análisis de la localización geográfica:

- Sea d_i^k una medida de conectividad de la región donde reside el trabajador i .
- Sea d_j^k un indicador de la necesidad de conectividad de la ocupación j (e.g., médicos e informáticos requieren mayor conectividad que taxistas o pintores).
- La disimilitud entre i y j se modela como $d_{ij}^k = (d_i^k - d_j^k)^2$.

Este análisis permite capturar las fricciones en la movilidad laboral y su impacto en la asignación ocupacional. En general, la metodología propuesta ofrece un marco más flexible y granular para estudiar el mercado laboral peruano, considerando dimensiones

que van más allá de la educación y permitiendo una mejor identificación de los determinantes del mismatching ocupacional.

Con esto concluimos la sección de aplicaciones en economía. Cabe resaltar que las herramientas presentadas en este documento tienen un alcance más amplio que el abordado aquí.

Capítulo 5

Estabilidad del matching vía Optimal Transport

El trabajo de [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#) estudia mercados de emparejamiento con preferencias alineadas ([Niederle and Yariv, 2009](#))¹, estableciendo conexiones entre estos mercados y la teoría del transporte óptimo. El resultado principal de [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#) relaciona los emparejamientos estables, el bienestar utilitario y los emparejamientos igualitarios con un problema de transporte óptimo que depende de un parámetro $\alpha \in \mathbb{R}$. [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#) demuestran que si, en lugar de maximizar la suma de las utilidades, se maximiza una transformación convexa de estas², la solución al problema de transporte óptimo es aproximadamente estable. Más específicamente, los autores prueban que cuando $\alpha > 0$, la solución del problema de transporte óptimo genera un emparejamiento aproximadamente estable, donde ninguna pareja bloqueante puede generar más de $\ln 2/\alpha$ de utilidad adicional. Cuando $\alpha = 0$, se recupera el problema de bienestar utilitario. Esto no se obtiene simplemente sustituyendo $\alpha = 0$ en $f_\alpha(u(x, y))$ (donde f_α es la transformación convexa), sino tomando el límite cuando $\alpha \rightarrow 0$. Si en cambio tomamos $\alpha < 0$, se genera una transformación cóncava de las utilidades de los agentes, y

¹Un caso particular ocurre cuando la utilidad obtenida por cada individuo en una pareja emparejada (x, y) es la misma para x y y .

²Capturan el grado de convexidad de la función objetivo (la transformación convexa de las utilidades) mediante el parámetro α .

la solución al problema de transporte óptimo asociado es aproximadamente igualitaria, con un factor de aproximación de $|\alpha|^{-1} \max(1, \ln |\alpha|)$. Como se menciona en [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#), el resultado principal de su trabajo se resume de la siguiente manera:

El mismo problema de transporte óptimo puede generar emparejamientos aproximadamente estables, maximizadores del bienestar o igualitarios, dependiendo de cómo se ajuste un solo parámetro (α).

Haciendo uso de la teoría del transporte óptimo, [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#) permite abarcar mercados finitos e infinitos, incluyendo aquellos con agentes contables e incontables, cuya importancia ha sido resaltada en trabajos como [Azevedo and Leshno \(2019\)](#) y [Choi \(2020\)](#). Los resultados de [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#) son independientes de la cardinalidad del conjunto de agentes. Sin embargo, es importante notar que el marco teórico asume que las preferencias están alineadas, una hipótesis clave explicada en la sección 6 de dicho artículo. Cuando las preferencias están alineadas y el mercado es finito, siempre existirá un emparejamiento estable: la pareja con mayor utilidad será emparejada primero, seguida de la siguiente, y así sucesivamente [Niederle and Yariv \(2009\)](#). El caso infinito hace uso de teoremas y resultados *más pesados*.

El concepto de preferencias alineadas, fundamental para garantizar la factibilidad de los emparejamientos estables en diversos mercados, ha sido poco explorado en la literatura existente, como se menciona en [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#). Una excepción importante es el estudio de [Ferdowsian et al. \(2023\)](#), que examina el uso de preferencias alineadas para analizar interacciones estratégicas en mercados de emparejamiento descentralizados. Este trabajo demuestra que los emparejamientos estables pueden ser implementados bajo condiciones estrictas, incluyendo información de preferencias completa, ausencia de fricciones temporales o una riqueza suficiente de escenarios de mercado. Los mercados descentralizados presentan desafíos particulares, como fricciones informativas y temporales, que pueden afectar la eficiencia de los resultados y representar obstáculos críticos para la estabilidad. [Ferdowsian et al. \(2023\)](#) introduce un modelo de juegos en mercados descentralizados y examina cómo las fricciones temporales y la información incompleta influyen en la posibilidad de alcanzar resultados estables de manera eficiente. Si bien

la hipótesis de preferencias alineadas puede parecer fuerte, permite describir muchos escenarios con cierto grado de precisión. Muchos mercados en los que la calidad es un factor determinante presentan simetría. Otra variable detrás de esta simetría es la distancia. Por ejemplo, en el contexto de la elección de escuelas, (Walters, 2014; Agarwal and Somaini, 2019; Laverde, 2023; Alba-Vivar, 2025) destacan que una variable clave en el emparejamiento es la distancia entre la escuela y el hogar del estudiante. En particular, Laverde (2023) analiza las limitaciones de las políticas de elección de escuelas en el contexto de la segregación residencial, usando datos del sistema de escuelas públicas de Boston para revelar disparidades raciales en el acceso a educación de calidad. El estudio encuentra que los estudiantes blancos de pre-escolar tienen mayor probabilidad de ser asignados a escuelas de mejor calidad que los estudiantes negros e hispanos, con brechas raciales comparables a aquellas bajo un sistema de asignación basado en vecindarios. La investigación identifica la distancia a las escuelas de alta calidad como el principal factor explicativo de la mitad de la brecha racial en la calidad de las escuelas asignadas. Este hallazgo desafía que tan efectivos son los sistemas de elección para proporcionar acceso equitativo a los recursos educativos, destacando la necesidad de abordar las barreras de distancia y re-evaluar las reglas de asignación.

Otro punto importante señalado en Federico Echenique, Fedor Sandomirskiy and Joseph Root (2024) sobre el emparejamiento con preferencias alineadas es que los mercados de emparejamiento con utilidad transferible, también conocidos como ³ *emparejamiento con dinero*, suelen generar naturalmente preferencias alineadas. Por ejemplo, en el mercado matrimonial, estudiado en Becker (1973) (basado en Shapley and Shubik (1971)), hay dos conjuntos de agentes X, Y , ambos finitos, y cada par $(x, y) \in X \times Y$ genera un excedente $\Phi(x, y)$. Luego, los agentes en la pareja dividen el excedente. En este contexto, la negociación de la división del excedente suele ocurrir después de que la pareja se forme. Si esta negociación sigue un protocolo de negociación de Nash con opciones externas fijas, las preferencias de los agentes se alinean naturalmente, ya que cada uno recibe una fracción constante del excedente total

³Aparecen en varios contextos económicos, como plataformas que conectan compradores y vendedores (por ejemplo, Freelancer o Mercado Libre), mercados bilaterales con bienes heterogéneos Rochet and Tirole (2003), y el mercado matrimonial Galichon and Salanié (2022b).

$\Phi(x, y)$. Este mecanismo ilustra cómo las preferencias alineadas emergen naturalmente en problemas de emparejamiento; ver [Niederle and Yariv \(2009\)](#) para una situación similar en el contexto del emparejamiento entre empresas y trabajadores.

En lo que sigue, estudiamos las secciones del artículo de [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#), el cual empieza abordando el *matching en la recta*. En seguida se escala a espacios más generales y se presentan los resultados centrales del artículo.

5.1. Emparejamiento en la recta

[Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#) inicia su análisis con un estudio del emparejamiento en la recta real. En este marco conceptual, los agentes son representados como puntos dentro del espacio continuo \mathbb{R} , mostrando una preferencia por parejas cercanas (interpretado en sentido topológico, con \mathbb{R} equipado con la métrica estándar $|\cdot|$). Este modelo facilita la visualización de emparejamientos estables, proporcionando una comprensión intuitiva de la estabilidad en estos contextos. Los autores presentan este modelo para proporcionar una intuición que podría ser bastante útil en resultados posteriores, donde la situación ya no es representable visualmente.

Los participantes del mercado, designados como $X \subset \mathbb{R}$ y $Y \subset \mathbb{R}$, se modelan utilizando medidas no atómicas sobre \mathbb{R} , representadas por μ y ν , respectivamente. Se asume que $\mu(X) = \nu(Y)$, es decir, que ambos lados tienen la misma masa. Podemos pensar, por ejemplo, en un mercado matrimonial con un continuo de agentes. Cada agente es caracterizado por un tipo $x \in X$. De manera análoga, cada mujer es caracterizada por su tipo $y \in Y$. El tipo de un agente podría estar determinado por variables como ingresos, altura, peso o nivel educativo. Se asume que $X, Y \subset [a, b] \subset \mathbb{R}$. Cuando dos agentes con tipos x e y se emparejan, obtienen una utilidad dada por:

$$u(x, y) = -|x - y| \in [-(b - a), 0]. \quad (5.1)$$

Por lo tanto, si los agentes están muy cerca con respecto a la topología usual de \mathbb{R} inducida por la norma $|\cdot|$, su utilidad aumenta.

Definición 5.1.1. Un emparejamiento π es estable si, para cualquier par de elementos (x_1, y_1) y (x_2, y_2) en el soporte⁴ de π :

$$|x_1 - y_2| \geq \min\{|x_1 - y_1|, |x_2 - y_2|\} \quad \text{y} \quad |x_2 - y_1| \geq \min\{|x_1 - y_1|, |x_2 - y_2|\}.$$

Esto significa que un emparejamiento es estable si no es bloqueado⁵ por algún par.

La definición 5.1.1 es ligeramente diferente a la de Federico Echenique, Fedor Sandomirskiy and Joseph Root (2024): aquí consideramos ambos pares *bloqueadores* (x_1, y_2) y (x_2, y_1) . En la siguiente figura mostramos algunos emparejamientos inestables.

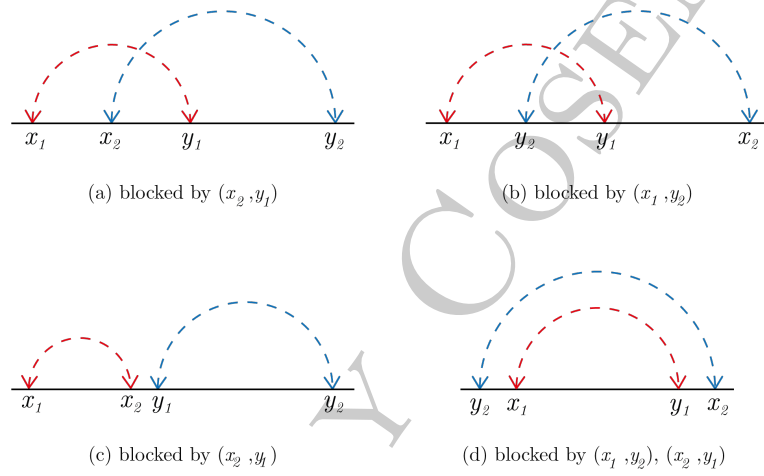


Figura 5.1 Emparejamientos inestables en la línea. Adaptado de Federico Echenique, Fedor Sandomirskiy and Joseph Root (2024).

Una observación importante es que cualquier emparejamiento estable agotara todas las coincidencias posibles tales que $x = y$. En efecto, $0 \leq |\theta|$ para cualquier $\theta \in \mathbb{R}$. Por lo tanto, un emparejamiento estable agota toda la masa común a μ y ν . Para entender como se empareja el resto de la población, Federico Echenique, Fedor Sandomirskiy and Joseph Root (2024) establecen una propiedad estructural que deben cumplir los emparejamientos estables. Veamos.

Definición 5.1.2. Dados $z_1, z_2 \in \mathbb{R}$, sea $\mathcal{O}(z_1, z_2)$ el círculo más pequeño en \mathbb{R}^2 que contiene los puntos $(z_1, 0)$ y $(z_2, 0)$.

⁴El menor conjunto cerrado en $X \times Y$ con medida completa.

⁵Aquí la relación de preferencia implícita \succeq está representada por la función de utilidad (5.1).

Definición 5.1.3. Un emparejamiento π satisface «no-cruce» si, para cualquier par de elementos (x_1, y_1) y (x_2, y_2) en el soporte de π , los círculos $\mathcal{O}(x_1, y_1)$ y $\mathcal{O}(x_2, y_2)$ no se intersecan a menos que $x_1 = x_2$ o $y_1 = y_2$.

Lema 5.1.4. Todo emparejamiento estable satisface la propiedad de no-cruce.

Demostración. Consideremos $(x_1, y_1), (x_2, y_2) \in \text{supp } \pi$, con $x_1 \neq x_2$ y $y_1 \neq y_2$. Supongamos que $\mathcal{O}(x_1, y_1) \cap \mathcal{O}(x_2, y_2) \neq \emptyset$. Entonces, hay dos casos posibles.

El primer caso es que

$$x_1 < x_2 < y_1 < y_2.$$

En este caso, $|x_2 - y_1| < |x_1 - y_1|$ y $|x_2 - y_1| < |x_2 - y_2|$. Por lo tanto, π es bloqueado por (x_2, y_1) .

El segundo caso es cuando se cumple

$$x_1 < y_2 < y_1 < x_2.$$

Si esto ocurre, entonces $|x_1 - y_2| < |x_1 - y_1|$ y $|x_2 - y_1| < |x_2 - y_2|$. Si π fuera estable, entonces

$$|x_1 - y_1| \leq |x_2 - y_1|,$$

$$|x_2 - y_2| \leq |x_1 - y_2|.$$

Sin embargo, por transitividad,

$$|x_1 - y_1| < |x_2 - y_2|$$

y

$$|x_2 - y_2| < |x_1 - y_1|.$$

Esto es una contradicción, por lo que π no es estable. \square

Observación. Como se menciona en [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#), los emparejamientos estables tienen mas estructura que solo la propiedad de no-cruce. No obstante, la propiedad de no-cruce por si misma es lo suficientemente fuerte como para reducir la búsqueda de un emparejamiento estable. La siguiente proposición [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#) es el resultado principal sobre emparejamientos en \mathbb{R} .

Proposición 5.1.5. Si $\rho = \mu - \nu$ es una medida no-atómica⁶ que cambia de signo un número finito de veces, entonces existe un único emparejamiento estable π ⁷. Además, para cada $x_1 \in X$, existen a lo más dos elementos $y_1, y_2 \in Y$ tales que (x_1, y_1) y (x_2, y_2) pertenecen al soporte de π . Finalmente, si μ y ν tienen densidad constante por partes con a lo más m intervalos de constancia, el algoritmo se ejecuta en $O(m^2)$.

Demostración. Ver el anexo de [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#). \square

5.2. Estabilidad, equidad y transporte óptimo

Consideremos ahora el marco general, estudiado en la sección 3 de [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#). Sea X, Y dos espacios métricos separables ([Aliprantis and Border, 2006](#)) dotados de la σ -álgebra de Borel (ver anexo de medida y probabilidad). Para un espacio Z , denotemos por $\mathcal{M}_+(Z)$ el conjunto de medidas positivas en Z con masa total finita. Entonces, la distribución de los tipos de agentes en X y en Y está representada por $\mu \in \mathcal{M}_+(X)$ y $\nu \in \mathcal{M}_+(Y)$ respectivamente. Se asume que $\mu(X) = \nu(Y)$. Como antes, una distribución $\pi \in \mathcal{M}_+(X \times Y)$ es un emparejamiento si tiene μ y ν como distribuciones marginales. Denotamos nuevamente por $\Pi(\mu, \nu)$ el conjunto de todos los emparejamientos.

Supongamos que las preferencias de los individuos están representadas por funciones de utilidad, $u(x, y)$ para $x \in X$ y $v(x, y)$ para $y \in Y$. Entonces, en particular, si $u(x, y) = v(x, y)$, las preferencias están alineadas. La siguiente definición, propuesta en [Greinecker and Kah \(2021\)](#), extiende la noción clásica de estabilidad introducida por [Gale and Shapley \(1962\)](#) a espacios métricos separables generales.

Definición 5.2.1. Un emparejamiento π es ε -estable con un parámetro $\varepsilon \geq 0$ si, para

⁶Una medida no-atómica es una medida en la que ningún punto (o átomo) tiene medida positiva. En otros términos, una medida μ en una sigma-álgebra \mathcal{F} es llamada no-atómica si para todo conjunto $A \in \mathcal{F}$ con $\mu(A) > 0$, existe un subconjunto $B \subset A$, $B \in \mathcal{F}$, tal que $0 < \mu(B) < \mu(A)$.

⁷Que puede obtenerse mediante un algoritmo, ver el apéndice A en [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#). Si μ y ν tienen densidad constante por partes con a lo más m intervalos de constancia, el algoritmo se ejecuta en tiempo del orden de m^2 .

$\pi \times \pi$ casi todos los pares $(x_1, y_1), (x_2, y_2)$, $(5.2 \vee 5.3) \wedge (5.4 \vee 5.5)$, donde:

$$u(x_1, y_1) - u(x_1, y_2) \geq -\varepsilon \quad (5.2)$$

$$v(x_2, y_2) - v(x_1, y_2) \geq -\varepsilon \quad (5.3)$$

$$u(x_2, y_2) - u(x_2, y_1) \geq -\varepsilon \quad (5.4)$$

$$v(x_1, y_1) - v(x_2, y_1) \geq -\varepsilon. \quad (5.5)$$

Si 5.2 o 5.3 se satisface, el par (x_1, y_2) no puede bloquear el emparejamiento. De manera similar, si 5.4 o 5.5 se cumple, el emparejamiento no puede ser bloqueado por (x_2, y_1) .

Definición 5.2.2. Si, en la definición 5.2.1, ambas desigualdades 5.2 y 5.3 se violan, entonces (x_1, y_2) se llama un ε -par bloqueador. Análogamente, si 5.4 y 5.5 se violan, (x_2, y_1) se llama un ε -par bloqueador.

En la definición 5.2.1, cuando $\varepsilon = 0$, el emparejamiento π es estable.

Lema 5.2.3. Sean u y v funciones continuas. Un emparejamiento π es ε -estable si y solo si, para cualquier $(x_1, y_1), (x_2, y_2) \in \text{supp}(\pi)$, se cumple $(5.2 \vee 5.3) \wedge (5.4 \vee 5.5)$.

Demostración. Una implicación es directa; la propiedad puntual (pares en el soporte) implica $\pi \times \pi$ en casi todo punto. Por lo tanto, solo queda probar la dirección opuesta. Sea π un emparejamiento ε -estable de poblaciones $\mu \in \mathcal{M}_+(X)$ y $\nu \in \mathcal{M}_+(Y)$, con u y v continuas. El objetivo es mostrar que para cualquier $(x_1, y_1), (x_2, y_2) \in \text{supp}(\pi)$, se cumple $(5.2 \vee 5.3) \wedge (5.4 \vee 5.5)$. Se procede por contradicción. Supongamos que existen $(x_1^*, y_1^*), (x_2^*, y_2^*) \in \text{supp } \pi$ tales que se cumple $(\neg 5.2 \wedge \neg 5.3) \vee (\neg 5.4 \wedge \neg 5.5)$. Nos enfocamos en $(\neg 5.2 \wedge \neg 5.3)$ ya que el caso $(\neg 5.4 \wedge \neg 5.5)$ es análogo.

Dado que u y v son continuas, existen vecindades $U, V \subset X \times Y$ de (x_1^*, y_1^*) y (x_2^*, y_2^*) , respectivamente, tales que ambas desigualdades (5.2) y (5.3) se violan para $(x_1, y_1), (x_2, y_2) \in U \times V$. Como $(x_1^*, y_1^*), (x_2^*, y_2^*) \in \text{supp}(\pi)$, se cumple $\pi(U), \pi(V) > 0$.

En efecto, por un lado $U \cap \text{supp}(\pi) \neq \emptyset$. Si asumimos que $\pi(U) = 0$, automáticamente $\pi(U^c) = 1$. Dado que U^c es cerrado, $U^c \cap \text{supp}(\pi)$ también es cerrado. Finalmente,

$$\pi(U^c \cap \text{supp}(\pi)) = 1 \quad \wedge \quad U^c \cap \text{supp } \pi \subset \text{supp}(\pi),$$

lo cual contradice la definición de $\text{supp}(\pi)$. El mismo argumento aplica para V . Por lo tanto, $U \times V$ tiene medida positiva bajo $\pi \times \pi$, lo que contradice la ε -estabilidad de π . \square

Observación. Si las preferencias de los agentes en X y Y están alineadas, entonces (5.2), (5.3), (5.4) y (5.5) se pueden reescribir en términos de dos desigualdades:

$$u(x_1, y_2) \leq \max\{u(x_1, y_1), u(x_2, y_2)\} + \varepsilon \quad (5.6)$$

y

$$u(x_2, y_1) \leq \max\{u(x_1, y_1), u(x_2, y_2)\} + \varepsilon. \quad (5.7)$$

Por lo tanto, un emparejamiento es estable si (5.6) y (5.7) se cumplen $\pi \times \pi$ en casi todo punto.

La siguiente noción importante estudiada en [Echenique et al. \(2023\)](#) es la noción de equidad. Dado un emparejamiento $\pi \in \Pi(\mu, \nu)$, definimos el siguiente número real $U_{\min}(\pi)$:

$$U_{\min}(\pi) = \min_{(x,y) \in \text{supp}(\pi)} u(x, y). \quad (5.8)$$

Cuando X y Y son compactos y $u(\cdot, \cdot)$ es continua, (5.8) está bien definida ya que $\text{supp } \pi$ es un subconjunto cerrado de un conjunto compacto $X \times Y$, por lo que también es compacto.

Además, utilizando (5.8), definimos el «mejor» $U_{\min}(\pi)$ sobre todos los emparejamientos. Es decir:

$$U_{\min}^*(\mu, \nu) = \max_{\pi \in \Pi(\mu, \nu)} U_{\min}(\pi).$$

Note que $U_{\min}^*(\mu, \nu)$ es una especie de mejor cota inferior (cota inferior igualitaria en el espíritu de Rawls ([Rawls, 1971](#))).

Cuando X y Y son compactos y u es continua, tanto U_{\min} como U_{\min}^* están bien definidas⁸. Cuando no se puede tomar el mínimo, este se reemplaza por el ínfimo esencial.

Definición 5.2.4. Un emparejamiento $\pi \in \Pi(\mu, \nu)$ es ε -igualitario si existe un subconjunto $S \subset X \times Y$ con $\pi(S) \geq (1 - \varepsilon)\pi(X \times Y)$ tal que

$$u(x, y) \geq U_{\min}^*(\mu, \nu) - \varepsilon, \quad \forall (x, y) \in S.$$

⁸La compacidad de $\Pi(\mu, \nu)$ se sigue de las propiedades impuestas, [Villani \(2009\)](#).

Para $\varepsilon = 0$, el emparejamiento es igualitario.

Note que la definición 5.2.4 tiene una interpretación clara. S representa un conjunto que es, conforme ε se hace más chico, *casi igual de grande* que $X \times Y$. En paralelo, para ε cada vez más chico, la utilidad de los individuos en S es mayor igual que la *cota inferior* U_{\min}^* .

Definición 5.2.5. Dada una función de costo $c : X \times Y \rightarrow \mathbb{R}$, un conjunto $\Gamma \subset X \times Y$ es llamado c -cíclicamente monótono si

$$\sum_{i=1}^n c(x_i, y_i) \leq \sum_{i=1}^n c(x_i, y_{i+1}), \quad \forall n \geq 2,$$

donde $y_{n+1} = y_1$, y $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$.

Observación. Se sabe que las soluciones a un problema de transporte óptimo están soportadas en un conjunto c -cíclicamente monótono, véase Villani (2009) o Ambrosio et al. (2021).

Ahora presentamos el resultado principal en Federico Echenique, Fedor Sandomirskiy and Joseph Root (2024), que describe la conexión entre estabilidad, equidad y transporte óptimo. Para un mercado de emparejamiento con preferencias alineadas, donde la función de utilidad es $u : X \times Y \rightarrow \mathbb{R}$, los autores consideran una familia paramétrica de funciones de costo, parametrizada por α :

$$c_\alpha(x, y) = \frac{1 - \exp(\alpha \cdot u(x, y))}{\alpha}.$$

Naturalmente, existe una tensión (o compensación) entre estabilidad y equidad. De hecho, Federico Echenique, Fedor Sandomirskiy and Joseph Root (2024) demuestran que la estabilidad y la equidad resultan de tomar valores extremos opuestos para el parámetro α . Veamos.

Teorema 5.2.6. Sea π^* una solución del siguiente problema de transporte óptimo:

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{X \times Y} c_\alpha(x, y) d\pi(x, y). \quad (5.9)$$

Entonces,

1. Si $\alpha > 0$, π^* es ε -estable, con $\varepsilon = \ln 2 / \alpha$.

2. Si $\alpha < 0$, π^* es ε -igualitario, con $\varepsilon = (\max\{1, \ln |\alpha|\})/|\alpha|$.

Observación. Si tomamos el límite $\alpha \rightarrow 0$ en c_α , obtenemos $-u(x, y)$. Por lo tanto, definimos $c_0(x, y) = -u(x, y)$. De esta manera, el problema de transporte para $\alpha = 0$ corresponde a la maximización del bienestar social utilitario:

$$W(\pi) = \int_{X \times Y} u(x, y) d\pi(x, y).$$

Así, variando α de $-\infty$ a ∞ , es posible interpolar entre equidad, bienestar y estabilidad.

Observación. Según Villani (2009), se sabe que solo se requiere que u sea continua y acotada⁹ para garantizar la existencia de una solución a (5.9). No se necesitan supuestos de compacidad sobre X y Y , basta con que sean espacios polacos.

La demostración se descompone en los casos $\alpha > 0$ y $\alpha < 0$. Primero, se analiza el caso $\alpha > 0$.

Demostración. Siguiendo el apéndice B en Federico Echenique, Fedor Sandomirskiy and Joseph Root (2024). Consideramos preferencias alineadas. Sea u la función de utilidad que representa dichas preferencias y tome valores en un intervalo abierto $I \subset \mathbb{R}$, posiblemente infinito. Sea $h : I \rightarrow \mathbb{R}_{++}$ una función diferenciable y supongamos que existe $\alpha > 0$ tal que

$$\frac{h'(t)}{h(t)} \geq \alpha, \quad \forall t \in I.$$

El propósito es demostrar que cualquier solución al problema de transporte óptimo con función de costo $-h(u(x, y))$ es ε -estable, con $\varepsilon = \ln 2/\alpha$. De hecho, $h(t) = \exp(\alpha t)$ cumple las condiciones de h .

Se requiere un resultado de Beiglböck et al. (2009): Dado un problema de transporte óptimo

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{X \times Y} c(x, y) d\pi(x, y),$$

donde X y Y son espacios polacos¹⁰ y c un costo medible, supongamos que el valor de este problema de transporte es finito y se alcanza en algún π^* . Entonces, existe un conjunto c -cíclicamente monótono Γ tal que $\pi^*(\Gamma) = 1$. Así, considerando $n = 2$,

$$h(u(x_1, y_2)) + h(u(x_2, y_1)) \leq h(u(x_1, y_1)) + h(u(x_2, y_2)).$$

⁹ c semicontinua inferiormente y acotada.

¹⁰Espacios topológicos separables completamente metrizable; es decir, espacios homeomorfos a un espacio métrico completo que tiene un subconjunto denso numerable.

Dado que h es positiva,

$$h(u(x_1, y_2)) \leq 2 \max\{h(u(x_1, y_1)), h(u(x_2, y_2))\}.$$

Ahora, consideremos $t, t' \in I$. Mediante integración simple:

$$\ln \left(\frac{h'(t)}{h(t)} \right) \geq \alpha(t' - t), \quad t \leq t'.$$

Sea $t = \max\{u(x_1, y_1), u(x_2, y_2)\}$ y $t' = u(x_1, y_2)$. Si $t \leq t'$, entonces

$$\ln 2 \geq \alpha(u(x_1, y_2) - \max\{u(x_1, y_1), u(x_2, y_2)\}).$$

Si $t > t'$, entonces, dado que $\alpha(u(x_1, y_2) - \max\{u(x_1, y_1), u(x_2, y_2)\}) < 0$, trivialmente

$$\ln 2 \geq \alpha(u(x_1, y_2) - \max\{u(x_1, y_1), u(x_2, y_2)\}).$$

Por lo tanto, en ambos casos, π^* es ε -estable con $\varepsilon = \ln 2/\alpha$.¹¹ □

Ahora, considere el caso $\alpha < 0$.

Demostración. Nuevamente, en un mercado con preferencias alineadas, supongamos que u toma valores en un intervalo abierto $I \subset \mathbb{R}$, posiblemente infinito. Sea $h : I \rightarrow (-\infty, 0)$ una función diferenciable creciente y supongamos que existe $\alpha < 0$ tal que

$$\frac{h'(t)}{h(t)} \geq \alpha, \quad \forall t \in I. \quad (5.10)$$

El objetivo es demostrar que cualquier solución al problema de transporte óptimo con costo $c(x, y) = -h(u(x, y))$ es ε -igualitaria con

$$\varepsilon = \frac{\max\{1, \ln |\alpha|\}}{|\alpha|}.$$

Un ejemplo de función que satisface (5.10) es $h(t) = -\exp(-|\alpha|t)$. Sin pérdida de generalidad, los autores consideran medidas de probabilidad. Recordemos que

$$U_{\min}(\pi) = \inf\{\lambda \in \mathbb{R} : \pi(\{(x, y) : u(x, y) < \lambda\}) > 0\}.$$

Entonces,

$$U_{\min}^*(\mu, \nu) = \sup_{\pi \in \Pi(\mu, \nu)} U_{\min}(\pi).$$

¹¹La situación es análoga para el otro posible par bloqueador.

Luego, considere el conjunto de todas las parejas hipotéticas cuya utilidad está por debajo de la cota inferior igualitaria en más de ε

$$C = \{(x, y) \in X \times Y : u(x, y) < U_{\min}^*(\mu, \nu) - \varepsilon\}.$$

Sea π^* una solución al problema de transporte óptimo con costo $-h(u(x, y))$. El objetivo es establecer que $\pi^*(C)$ *no puede ser demasiado grande*. Sea $\delta > 0$ y $\pi' \in \Pi(\mu, \nu)$ tal que

$$U_{\min}(\pi') \geq U_{\min}^*(\mu, \nu) - \varepsilon.$$

Cabe resaltar que puede no existir tal π' para $\delta = 0$ ya que el supremo podría no alcanzarse. Ahora, dado que π^* es óptimo para $-h(x, y)$,

$$\int_{X \times Y} h(u(x, y)) d\pi'(x, y) \leq \int_{X \times Y} h(u(x, y)) d\pi^*(x, y).$$

Dado que $h(\cdot)$ es creciente,

$$h(U_{\min}^*(\mu, \nu) - \delta) \leq h(u(x, y))$$

para casi todos los pares (x, y) bajo π' . Así,

$$h(U_{\min}^*(\mu, \nu) - \delta) \underbrace{\pi(X \times Y)}_{=1} \leq \int_{X \times Y} h(u(x, y)) d\pi^*(x, y).$$

Dado que u es continua y $\delta > 0$ es arbitrario,

$$h(U_{\min}^*(\mu, \nu)) \underbrace{\pi(X \times Y)}_{=1} \leq \int_{X \times Y} h(u(x, y)) d\pi^*(x, y).$$

Ahora, descomponiendo la integral:

$$\begin{aligned} \int_{X \times Y} h(u(x, y)) d\pi^*(x, y) &= \int_C h(u(x, y)) d\pi^*(x, y) + \int_{X \times Y - C} h(u(x, y)) d\pi^*(x, y) \\ &\leq h(U_{\min}^* - \varepsilon) \pi^*(C). \end{aligned}$$

Por lo tanto, dado que h toma valores en \mathbb{R}_{--} ,

$$h(U_{\min}^*(\mu, \nu)) \leq h(U_{\min}^* - \varepsilon) \pi^*(C) \implies \pi^*(C) \leq \frac{h(U_{\min}^*(\mu, \nu))}{h(U_{\min}^*(\mu, \nu) - \varepsilon)}.$$

Por otro lado, integrando (5.10) obtenemos:

$$\ln \left(\frac{|h(t')|}{|h(t)|} \right) \leq |\alpha| \cdot (t' - t), \quad t \leq t'.$$

Por lo tanto,

$$\pi^*(C) \leq \exp(-\alpha\varepsilon).$$

Tomando $t = U_{\min}^*(\mu, \nu)$ y $t' = U_{\min}^*(\mu, \nu) - \varepsilon$, y sustituyendo con

$$\varepsilon = \frac{\max\{1, \ln |\alpha|\}}{|\alpha|},$$

se obtiene

$$\pi^*(C) \leq \min\{e^{-1}, |\alpha|^{-1}\}.$$

Así,

$$\pi^*(C) \leq \varepsilon.$$

□

Notación. Se denota por $\Pi_{+\infty}^u(\mu, \nu)$ al conjunto de emparejamientos que pueden obtenerse como el límite débil $\pi = \lim_{n \rightarrow \infty} \pi_{\alpha_n}$ de una sucesión de soluciones $\{\pi_{\alpha_n}\}_{n \in \mathbb{N}}$ al problema de transporte óptimo donde $c = c_{\alpha_n}$, para alguna sucesión $\alpha_n \rightarrow \infty$. De manera análoga, se define $\Pi_{-\infty}^u(\mu, \nu)$ cuando $\alpha_n \rightarrow -\infty$.

Corolario 5.2.7. Si u es continua y acotada, entonces los conjuntos $\Pi_{+\infty}^u(\mu, \nu)$ y $\Pi_{-\infty}^u(\mu, \nu)$ son no vacíos, convexos y débilmente cerrados¹². Además, todos los emparejamientos en $\Pi_{+\infty}^u(\mu, \nu)$ son estables y todos los emparejamientos en $\Pi_{-\infty}^u(\mu, \nu)$ son igualitarios.

Demostración. El primer paso consiste en demostrar que $\Pi_{+\infty}^u(\mu, \nu)$ y $\Pi_{-\infty}^u(\mu, \nu)$, correspondientes a $\alpha \rightarrow \pm\infty$, son no vacíos. Dado que el argumento en cada caso es idéntico, basta probarlo para $\Pi_{+\infty}^u(\mu, \nu)$. Considérese $\alpha_n \rightarrow \infty$ y π_n como la solución del problema de transporte óptimo con $c = c_{\alpha_n}$. La existencia de tal solución está garantizada por las condiciones impuestas sobre u y los espacios X, Y . Por otro lado, el conjunto $\Pi(\mu, \nu)$ ¹³ es secuencialmente compacto en la topología de convergencia débil Villani (2009). Así, posiblemente pasando a una subsucesión, se obtiene que $\pi_n \rightarrow \pi_{+\infty} \in \Pi(\mu, \nu)$. Por lo tanto, $\Pi_{+\infty}^u(\mu, \nu)$ y $\Pi_{-\infty}^u(\mu, \nu)$ son no vacíos. La convexidad de $\Pi_{+\infty}^u(\mu, \nu)$ y $\Pi_{-\infty}^u(\mu, \nu)$ se sigue de la convexidad del conjunto de soluciones al problema de transporte óptimo. La misma idea se aplica para la cerradura débil.

¹²Cerrados en la topología débil.

¹³Se asume que X, Y son espacios polacos y que $\mu \in \mathcal{M}_+(X)$, $\nu \in \mathcal{M}_+(Y)$.

Finalmente, se prueba que los elementos de $\Pi_{\infty}^u(\mu, \nu)$ son emparejamientos estables y que los elementos de $\Pi_{-\infty}^u(\mu, \nu)$ son igualitarios. Considere la función continua $f : (X \times Y) \times (X \times Y) \rightarrow \mathbb{R}$ definida como

$$f(x_1, y_1, x_2, y_2) = \max\{0, u(x_1, y_2) - \max\{u(x_1, y_1), u(x_2, y_2)\}\}.$$

Un emparejamiento π es estable si y solo si

$$\int f(d\pi \times d\pi) = 0. \quad (5.11)$$

Para un emparejamiento ε -estable, (5.11) está acotado por $\varepsilon \cdot \mu(X) \cdot \nu(Y)$. Luego, usando la convergencia débil,

$$\int f(d\pi_{+\infty} \times d\pi_{+\infty}) = \lim_n \int f d(\pi_n \times d\pi_n) \leq \lim_n \frac{\ln 2}{\alpha_n} \mu(X) \nu(Y) = 0.$$

Dado que $\int f(d\pi_{+\infty} \times d\pi_{+\infty}) \geq 0$, se concluye que

$$\int f(d\pi_{+\infty} \times d\pi_{+\infty}) = 0.$$

Por lo tanto, π_{∞} es estable. Para demostrar que los elementos de $\Pi_{-\infty}^u(\mu, \nu)$ son igualitarios, considérese un límite débil $\pi_{-\infty}$ de una sucesión π_{α_n} , tal que $\alpha_n \rightarrow -\infty$. Fijado $\varepsilon > 0$, considérese el conjunto de parejas hipotéticas cuya utilidad está por debajo de la cota inferior igualitaria en más de ε :

$$C_{\varepsilon} = \{(x, y) \in X \times Y : u(x, y) < U_{\min}^*(\mu, \nu) - \varepsilon\}.$$

Dado que u es continua, el conjunto C_{ε} es abierto, por lo que se tiene

$$\pi_{-\infty}(C_{\varepsilon}) \leq \lim_{n \rightarrow \infty} \inf \pi_{\alpha_n}(C_{\varepsilon}).$$

Por el teorema 5.2.6, se obtiene que $\lim_n \inf \pi_{\alpha_n}(C_{\varepsilon}) = 0$. En consecuencia, $\pi_{-\infty}(C_{\varepsilon}) = 0$ para cualquier $\varepsilon > 0$. Dado que $C_{\varepsilon} \downarrow C_0$ cuando $\varepsilon \rightarrow 0$,

$$\pi_{-\infty}(C_0) = \pi_{-\infty}\left(\bigcup_{\varepsilon > 0} C_{\varepsilon}\right) = \lim_{\varepsilon \rightarrow 0} \pi_{-\infty}(C_{\varepsilon}) = 0.$$

Por lo tanto, dado que $\pi_{-\infty}(C_0) = 0$, para casi todas las parejas (x, y) bajo $\pi_{-\infty}$ la utilidad $u(x, y)$ es al menos $U_{\min}^*(\mu, \nu)$, lo que implica que $\pi_{-\infty}$ es un emparejamiento igualitario. Así, todos los elementos de $\Pi_{-\infty}^u(\mu, \nu)$ son emparejamientos igualitarios. \square

Observación. El corolario 5.2.7 garantiza la existencia de emparejamientos estables e igualitarios y proporciona un método computacionalmente viable para encontrarlos.

Cuando el mercado es finito, para α suficientemente grande, se obtiene estabilidad. En efecto, si X y Y son finitos, y se define

$$\delta = \min \left\{ \min_{x_1, y_1 \neq y_2} |u(x_1, y_1) - u(x_1, y_2)|, \min_{y_1, x_1 \neq x_2} |u(x_1, y_1) - u(x_2, y_1)| \right\}$$

como el cambio mínimo en la utilidad que un agente puede experimentar si cambia de pareja, entonces cualquier emparejamiento ε -estable con $\varepsilon < \delta$ es automáticamente estable. Por lo tanto, utilizando el teorema 5.2.6, cualquier solución a un problema de transporte óptimo en el caso finito con $\alpha > \delta / \ln 2$ es estable.

Observación. A partir de la demostración del teorema 5.2.6, es evidente que el resultado no está limitado a c_α .

5.3. Bienestar y equidad en emparejamientos estables

El teorema 5.2.6 establece que la equidad y la estabilidad corresponden a los extremos del espectro de α , específicamente $+\infty$ y $-\infty$. Esto indica que la estabilidad no garantiza inherentemente la equidad, y que la posible pérdida en equidad y bienestar, aunque potencialmente significativa, permanece acotada.

Para comprender por qué las pérdidas en bienestar y equidad son finitas, se reconsidera el concepto de estabilidad. Sea u una función de utilidad continua, y considérese un emparejamiento $\pi \in M^+(X \times Y)$ con marginales μ y ν . El emparejamiento π es ϵ -estable, para $\epsilon \geq 0$, si satisface la siguiente condición para todos (x_1, y_1) y (x_2, y_2) en el soporte de π :

$$u(x_1, y_2) \leq \max\{u(x_1, y_1), u(x_2, y_2)\} + \epsilon.$$

Este enfoque permite interpretar (x_1, y_2) como un elemento representativo del espacio producto $X \times Y$. El criterio de estabilidad puede expresarse de la siguiente manera:

Un emparejamiento π es ϵ -estable si, para una pareja genérica (x, y) , la utilidad de al menos uno de los integrantes en π no es menor que $u(x, y) - \epsilon$.

Por lo tanto, la utilidad de una pareja hipotética (x, y) sirve como una cota inferior para la utilidad de al menos uno de los integrantes en cualquier emparejamiento estable. Esta observación impone un límite crucial en la pérdida de bienestar y equidad que puede producirse al buscar estabilidad.

Teorema 5.3.1. Sea un mercado con poblaciones unitarias¹⁴ y preferencias alineadas representadas por una función de utilidad $u \geq 0$, continua y acotada. Entonces, cualquier emparejamiento ε -estable π satisface

$$W(\pi) \geq \frac{1}{2}(W^*(\mu, \nu) - \varepsilon),$$

donde

$$W(\pi) = \int_{X \times Y} u(x, y) d\pi(x, y),$$

y $W^*(\mu, \nu) = \max_{\pi \in \Pi(\mu, \nu)} W(\pi)$. Además, π es ε' -igualitario con

$$\varepsilon' = \max \left\{ \frac{1}{2}, \varepsilon \right\}.$$

Demostración. Sea π un emparejamiento ϖ -estable con marginales μ y ν . Dado que u es continua, para cualquier $(x_1, y_1), (x_2, y_2) \in \text{supp}(\pi)$ se cumple:

$$u(x_1, y_2) \leq \max\{u(x_1, y_1), u(x_2, y_2)\} + \varepsilon.$$

Por la no negatividad de u , se obtiene

$$u(x_1, y_2) \leq u(x_1, y_1) + u(x_2, y_2) + \varepsilon.$$

Sea π' otro emparejamiento con marginales μ y ν . Considérese $\lambda \in \mathcal{M}_+((X \times Y) \times (X \times Y))$ tal que sus marginales en (x_1, y_1) y (x_2, y_2) coincidan con π y su marginal en (x_1, y_2) sea π' . Se tiene entonces:

$$\begin{aligned} W(\pi') &= \int_{X \times Y} u(x_1, y_2) d\pi'(x_1, y_2) \\ &= \int_{(X \times Y) \times (X \times Y)} u(x_1, y_2) d\lambda(x_1, y_1, x_2, y_2) \\ &\leq \int_{(X \times Y) \times (X \times Y)} (u(x_1, y_1) + u(x_2, y_2) + \varepsilon) d\lambda(x_1, y_1, x_2, y_2) \\ &= \int_{X \times Y} u(x_1, y_1) d\pi(x_1, y_1) + \int_{X \times Y} u(x_2, y_2) d\pi(x_2, y_2) + \varepsilon \\ &= 2W(\pi) + \varepsilon. \end{aligned}$$

¹⁴ $\mu(X) = \nu(Y) = 1$.

Por lo tanto,

$$W(\pi) \geq \frac{1}{2}[W(\pi') - \varepsilon],$$

para cualquier emparejamiento π' . Así, se concluye que

$$W(\pi) \geq \frac{1}{2}(W^*(\mu, \nu) - \varepsilon).$$

□

Corolario 5.3.2. Considérese un mercado donde $X = Y = \mathbb{R}^d$ y ambas medidas μ y ν tienen soporte compacto y son absolutamente continuas respecto a la medida de Lebesgue. Si la función de utilidad se define como $u(x, y) = -\|x - y\|$, el problema de transporte óptimo con función de costo

$$c_\alpha(x, y) = \frac{1 - \exp(-\alpha\|x - y\|)}{\alpha}$$

admite una solución óptima π^* con las siguientes propiedades:

- Para $\alpha < 0$, π^* es única y determinista.
- Para $\alpha > 0$, π^* es única y representa una combinación convexa entre un emparejamiento determinista y un emparejamiento diagonal.
- Existe un emparejamiento determinista π^* que maximiza el bienestar.

Además, la condición de absoluta continuidad con respecto a la medida de Lebesgue puede relajarse requiriendo únicamente que μ y ν asignen masa nula a superficies de dimensión $(d - 1)$. El emparejamiento igualitario obtenido en el límite $\alpha \rightarrow -\infty$ corresponde al problema de transporte L^∞ ; véase, por ejemplo, [Champion et al. \(208\)](#).

Observación. En la asignación escolar, las preferencias de los estudiantes y las prioridades de las escuelas suelen centrarse en la proximidad. La investigación muestra que la distancia influye significativamente en las preferencias de los estudiantes por las escuelas, con alumnos locales frecuentemente priorizados por las políticas escolares. El emparejamiento estable en \mathbb{R}^d proporciona un marco para analizar este aspecto, contrastando el objetivo de estabilidad, que valora altamente la proximidad, con medidas tradicionales de bienestar que enfatizan la equidad y minimizan funciones convexas de las utilidades. Estudios como los de [Walters \(2014\)](#) y [Angrist et al.](#)

(2022) destacan cómo el emparejamiento estable puede generar mayores tiempos de desplazamiento y costos de transporte, evidenciando un claro intercambio entre eficiencia y equidad en la asignación escolar.

El análisis presentado en [Federico Echenique, Fedor Sandomirskiy and Joseph Root \(2024\)](#) no solo establece la conexión entre estabilidad, equidad y transporte óptimo, sino que también aborda extensiones relevantes como el emparejamiento con múltiples parejas y los fundamentos de las preferencias alineadas. En particular, se argumenta que la alineación de preferencias es natural en entornos con transferencias, siguiendo a [Pycia \(2012\)](#), y se establecen condiciones suficientes para la existencia de una función de utilidad común, destacando la importancia de una propiedad de acíclicidad. Estos resultados refuerzan la aplicabilidad del marco propuesto y sugieren direcciones para futuras investigaciones en mercados con estructuras más generales.

Capítulo 6

Acotamiento del CVA

Siguiendo a [Hull \(2006\)](#), un derivado se define como un instrumento financiero cuyo valor se deriva de variables subyacentes, comúnmente de activos como acciones o divisas. Aunque dichos instrumentos pueden ser bastante complejos, nos centraremos a continuación en contratos entre dos entidades financieras. Tomaremos el punto de vista de una de ellas, y llamaremos a la otra la *contraparte*. Brindamos tres ejemplos con la finalidad de aterrizar estas ideas:

1. Un contrato tipo *forward*: es un acuerdo para comprar o vender a la contraparte un activo a un precio y en un tiempo específico. Por ejemplo, si se piensa que el precio de una acción caerá, se buscará entrar en un contrato en el que se venda dicha acción en el futuro a un precio similar al actual. Por otro lado, la contraparte espera que el precio suba, y así obtener la acción a un precio menor.
2. Una *opción*: es un contrato donde una de las partes paga un precio por el derecho a comprar o vender un activo a la contraparte, a un valor y en un plazo previamente acordados. La diferencia con un contrato forward radica en que no se está obligado a ejercer este derecho, pero si no se ejerce se asume el precio del contrato como pérdida.
3. Un contrato *swap*: es un acuerdo entre dos entidades para intercambiar pagos en el futuro. El acuerdo especifica las fechas donde deben llevarse a cabo los pagos, y el método para calcularlos ([Hull \(2006\)](#)). Este cálculo involucra el valor futuro de algún activo u otras variables del mercado.

Considere un portafolio compuesto por contratos de derivados, incluyendo contratos swap, celebrados con una única contraparte¹. Como es bien sabido, los precios de los activos en el mercado están sujetos a fluctuaciones aleatorias, lo que impacta directamente en el valor de los derivados. En este contexto, el *riesgo de mercado* influye en la variabilidad del valor total del portafolio.

Adicionalmente, dicho valor está expuesto al *riesgo de crédito*, es decir, la posibilidad de que la contraparte incumpla sus compromisos de pago, lo que se conoce como *default*. La exposición en caso de default corresponde a la pérdida potencial que enfrenta la entidad que mantiene el portafolio, y equivale al valor del portafolio en el momento del incumplimiento, siempre que este valor sea positivo. En caso contrario, si el valor del portafolio es negativo, la entidad se libera de una deuda, por lo que la exposición se considera nula.

En resumen, el riesgo de mercado determina la magnitud de la exposición a la contraparte, mientras que el riesgo de crédito establece la probabilidad de que dicha exposición se traduzca en una pérdida efectiva (Glasserman and Yang, 2015).

El *Credit Value Adjustment* (ajuste por riesgo de crédito) o CVA es una métrica que cuantifica las pérdidas potenciales de un portafolio de derivados debido al incumplimiento de la contraparte. Se define como la diferencia entre el valor del portafolio considerando la posibilidad de default y su valor en ausencia de dicha eventualidad².

El CVA aumenta en presencia de *wrong-way risk*, es decir, cuando el riesgo de crédito y el riesgo de mercado están positivamente correlacionados. Nuestro interés radica en determinar el CVA en el peor escenario posible, proporcionando así una cota superior para las pérdidas ocasionadas por default. Este valor depende de la relación entre el riesgo de crédito y el riesgo de mercado. Aunque asumimos que las distribuciones individuales de ambos factores son conocidas, la distribución conjunta sigue siendo incierta. Como discutimos en el capítulo 2, la teoría del transporte óptimo ofrece una herramienta eficaz para abordar este tipo de problemas.

¹En la práctica, un portafolio puede contener miles de posiciones en contratos con múltiples contrapartes.

²Esto convierte al CVA en un precio y no en una medida de riesgo. Puede utilizarse para determinar un cargo adicional que los bancos imponen por asumir este riesgo. Para una discusión más detallada, ver (Rosen and Saunders, 2012).

A partir de este punto, trabajaremos sobre un espacio de probabilidad (Ω, \mathbb{P}) , el cual puede interpretarse como el conjunto de todas las posibles realizaciones de la economía, dotado con una medida de probabilidad que asigna la ocurrencia de cada evento. Siguiendo la formulación de [Glasserman and Yang \(2015\)](#), definimos la variable aleatoria $\tau : \Omega \rightarrow [0, +\infty]$ como el tiempo de default de la contraparte, y denotamos por $V(\tau)$ el valor del portafolio en dicho instante. Como se mencionó previamente, la pérdida en el momento del default corresponde a la parte positiva de $V(\tau)$, que representamos como $V^+(\tau)$.

Definimos el CVA en un horizonte de tiempo $T > 0$ como el valor esperado de la exposición en el instante del default:

$$CVA := \mathbb{E}_{\mathbb{P}} [V^+(\tau) \mathbf{1}_{\{\tau \leq T\}}].$$

Comúnmente, el cálculo de la esperanza en la expresión del CVA se realiza sobre una serie de tiempos intermedios $t_0 = 0 < t_1 < \dots < t_d = T$, los cuales pueden corresponder a las fechas de pago previamente acordadas en los contratos de derivados. Consideremos los valores aleatorios $V^+(t_1), \dots, V^+(t_d)$, donde cada $V^+(t_i)$ representa la pérdida en caso de que la contraparte incurra en default durante el intervalo $(t_{i-1}, t_i]$. La evaluación de estos valores requiere la valoración de cada activo del portafolio en todos los escenarios posibles y en cada instante de tiempo, un proceso computacionalmente intensivo en el que no nos enfocaremos aquí.

Definimos el vector aleatorio $X : \Omega \rightarrow \mathbb{R}^d$ como:

$$X = (V^+(t_1), \dots, V^+(t_d)),$$

cuyo comportamiento probabilístico sigue una distribución de probabilidad conocida $\mu \in \mathcal{P}(\mathbb{R}^d)$, la cual está vinculada al riesgo de mercado. Suponemos que X es un vector aleatorio continuo, es decir, que μ no posee átomos.

Por otro lado, el riesgo de crédito determina las probabilidades de default en cada periodo, lo que da lugar a una distribución de probabilidad $\nu \in \mathcal{P}(\mathbb{R}^d)$ para el vector aleatorio $Y : \Omega \rightarrow \mathbb{R}^d$, definido como:

$$Y = (\mathbf{1}_{\{\tau \leq t_1\}}, \mathbf{1}_{\{t_1 < \tau \leq t_2\}}, \dots, \mathbf{1}_{\{t_{d-1} < \tau \leq t_d\}}).$$

Este vector indica en qué periodo (si es que en alguno) ocurre el default y, por lo tanto, qué nivel de pérdida se asume. Notemos que el soporte de Y está contenido en el conjunto finito $\{y_1, \dots, y_d, y_{d+1}\}$, donde:

$$y_1 = (1, 0, \dots, 0), \quad \dots, \quad y_d = (0, 0, \dots, 1), \quad y_{d+1} = (0, 0, \dots, 0).$$

Dado que Y toma valores en un conjunto discreto, la medida ν es discreta. La pérdida en el instante del default se expresa como:

$$V^+(\tau)\mathbf{1}_{\{\tau \leq T\}} = \langle X, Y \rangle = \sum_{i=1}^d V^+(t_i)\mathbf{1}_{\{t_{i-1} < \tau \leq t_i\}}.$$

Es importante destacar que, aunque conocemos las distribuciones marginales μ y ν , la distribución conjunta $\pi = (X, Y)_\# \mathbb{P} \in \Pi(\mu, \nu)$ es desconocida. Mediante un cambio de variable, podemos reformular el CVA como:

$$CVA = \int_{\Omega} \langle X, Y \rangle d\mathbb{P} = \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi.$$

Deseamos evaluar el peor escenario posible para el CVA, lo que nos conduce a un problema de transporte óptimo³.

$$CVA_{\text{sup}} = \sup_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi. \quad (6.1)$$

Bajo condiciones razonables, este problema admite una única solución, es decir, existe un único plan de transporte óptimo.

Teorema 6.0.1. Si $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ ⁴ y $\mu \ll \mathcal{L}$ ⁵, entonces existe un único plan de transporte óptimo $\hat{\pi} \in \Pi(\mu, \nu)$ tal que:

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\hat{\pi} = \sup_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi.$$

Demostración. Consideremos la identidad:

$$\langle x, y \rangle = \frac{\|x\|^2 + \|y\|^2 - \|x - y\|^2}{2}.$$

³Recuerde la interpretación probabilística del problema de Kantorovich, comentada en el capítulo

1

⁴Ver la definición 1.0.6

⁵Recordar que \mathcal{L} representa la medida de Lebesgue en \mathbb{R}^d

Entonces, el problema de maximización del CVA se re-escribe como:

$$\sup_{\pi \in \Pi(\mu, \nu)} \left[\int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{\|x\|^2}{2} d\pi + \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{\|y\|^2}{2} d\pi - \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{\|x - y\|^2}{2} d\pi \right].$$

Aplicando el lema 1.0.3, podemos reformularlo como:

$$\int_{\mathbb{R}^d} \frac{\|x\|^2}{2} d\mu + \int_{\mathbb{R}^d} \frac{\|y\|^2}{2} d\nu - \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{\|x - y\|^2}{2} d\pi. \quad (6.2)$$

Dado que $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, los dos primeros términos son finitos, y se cumplen las condiciones del teorema 1.0.7. Como consecuencia, el ínfimo en (6.2) admite una única solución $\hat{\pi}$, lo que garantiza la existencia y unicidad del plan de transporte óptimo para el problema (6.1). \square

Observación. En esta formulación, hemos definido el *CVA unilateral*, que supone que solo una de las partes puede entrar en default. En contraste, el *CVA bilateral* considera que ambas contrapartes pueden incumplir los pagos, por lo que es necesario evaluar la exposición de cada una de ellas por separado. Además, esta formulación no incorpora el valor temporal del dinero, es decir, no considera el descuento de flujos futuros. Para una formulación más completa, que integre estos factores y el CVA bilateral, puede consultarse (Memartoluie, 2017).

6.1. Acotamiento por aproximaciones empíricas

El cálculo del CVA, formulado como un problema de transporte óptimo en el caso continuo, no admite una solución analítica en general. Por ello, recurrimos a una discretización del problema a través de un número finito de simulaciones, lo que permite su resolución mediante métodos de programación lineal. Posteriormente, el problema continuo se recupera en el límite cuando el número de simulaciones tiende a infinito.

Para ello, utilizamos el concepto de *medida empírica*, el cual formaliza la intuición de una distribución basada en observaciones experimentales. Supongamos que queremos estimar la probabilidad de que un vector aleatorio $X : \Omega \rightarrow \mathbb{R}^d$ tome valores en un conjunto $A \subset \mathbb{R}^d$. Dado un conjunto de copias independientes X_1, X_2, \dots de X , la realización de N experimentos independientes equivale a evaluar los valores $X_1(\omega), \dots, X_N(\omega)$ para algún $\omega \in \Omega$. La estimación de la probabilidad de pertenencia a

A se obtiene contando la fracción de veces que los X_i caen en A , dividiendo el total entre N . A medida que $N \rightarrow \infty$, se espera que esta estimación converja a la probabilidad real $\mathbb{P}(X \in A)$.

Definición 6.1.1. Sea (Ω, \mathbb{P}) un espacio de probabilidad, $X : \Omega \rightarrow \mathbb{R}^d$ un vector aleatorio con distribución $\mu \in \mathcal{P}(\mathbb{R}^d)$ y sea X_1, X_2, \dots una secuencia de copias independientes de X . Definimos la *medida empírica de X* como:

$$\mu_N(A) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{X_i \in A\}},$$

para todo conjunto medible $A \in \mathcal{B}(\mathbb{R}^d)$.

Observación. En rigor, la medida empírica μ_N es una variable aleatoria que toma valores en el conjunto de medidas de probabilidad $\mathcal{P}(\mathbb{R}^d)$. Es decir, $\mu_N(\omega) \in \mathcal{P}(\mathbb{R}^d)$ y se define por:

$$\mu_N(\omega)(A) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{X_i(\omega) \in A\}},$$

para cada $\omega \in \Omega$ y $A \in \mathcal{B}(\mathbb{R}^d)$. No obstante, omitimos la dependencia en ω en la notación para simplificar la presentación, ya que las propiedades fundamentales de la medida empírica se satisfacen casi seguramente con respecto a \mathbb{P} .

Lema 6.1.2. En el contexto de la definición anterior, se cumple que $\mu_N \rightarrow \mu$ puntualmente, casi seguramente con respecto a \mathbb{P} , cuando $N \rightarrow \infty$.

Demostración. Para todo conjunto medible $A \in \mathcal{B}(\mathbb{R}^d)$, definimos $Z_i = \mathbf{1}_{\{X_i \in A\}}$ para cada $i \in \mathbb{N}$. Como los X_i son independientes e idénticamente distribuidos, se tiene que:

$$\mathbb{E}_{\mathbb{P}}[Z_i] = \mathbb{P}(X_i \in A) = \mu(A), \quad \forall i \in \mathbb{N}.$$

Aplicando la ley fuerte de los grandes números, obtenemos:

$$\mu_N(A) = \frac{1}{N} \sum_{i=1}^N Z_i \rightarrow \mu(A), \quad \text{casi seguramente.}$$

Como A es arbitrario, se concluye la convergencia puntual de μ_N hacia μ , casi seguramente con respecto a \mathbb{P} . \square

Corolario 6.1.3. Como consecuencia directa del lema anterior, se cumple que $\mu_N \rightarrow \mu$ en la topología débil ($\mu_N \xrightarrow{w} \mu$) y en la métrica de Wasserstein-2 ($\mu_N \xrightarrow{\mathcal{P}_2} \mu$), casi seguramente con respecto a \mathbb{P} .

Sean ahora X, Y como en la sección anterior y consideremos las secuencias de copias independientes X_1, X_2, \dots y Y_1, Y_2, \dots . Bajo el supuesto de que μ no tiene átomos, los valores X_1, \dots, X_N son distintos casi seguramente, lo que nos permite interpretar μ_N como la distribución uniforme sobre $\{X_1, \dots, X_N\}$. En general, las medidas empíricas μ_N y ν_N intentan aproximar las distribuciones reales μ y ν . De esta manera, podemos iterar la simulación numérica de valores $(X_i)_{i \in \mathbb{N}}$ mientras resolvemos el problema de transporte óptimo sobre las medidas discretizadas μ_N y ν_N .

Dado que cada $\pi \in \Pi(\mu_N, \nu_N)$ está concentrada en el conjunto finito $\{(X_i, y_j) : i = 1, \dots, N, j = 1, \dots, d+1\}$, denotamos $\pi\{(X_i, y_j)\} = \pi_{ij}$ y buscamos resolver el problema:

$$CVA_{\text{sup}}^N := \max_{\pi \in \Pi(\mu_N, \nu_N)} \sum_{i=1}^N \sum_{j=1}^{d+1} \pi_{ij} \langle X_i, y_j \rangle.$$

En el caso de distribuciones discretas, el problema de Kantorovich se reduce a un problema de programación lineal. Así, debemos resolver:

$$\begin{cases} \text{máx} & \sum_{i=1}^N \sum_{j=1}^{d+1} \pi_{ij} \langle X_i, y_j \rangle, \\ \text{s.a} & \sum_{i=1}^N \pi_{ij} = \nu_N(y_j), \quad \forall j = 1, \dots, d+1, \\ & \sum_{j=1}^{d+1} \pi_{ij} = \frac{1}{N}, \quad \forall i = 1, \dots, N, \\ & \pi_{ij} \geq 0. \end{cases} \quad (6.3)$$

Este planteamiento permite aproximar el CVA en un marco computacional eficiente. En la siguiente sección, analizamos la convergencia de esta aproximación al problema continuo.

Estos valores dependen de los resultados de las simulaciones numéricas $(X_i(\omega))_{i \in \mathbb{N}}$ y $(Y_i(\omega))_{i \in \mathbb{N}}$, es decir, del evento $\omega \in \Omega$ sobre el cual estamos evaluando todo de manera implícita⁶. No obstante, podemos garantizar con certeza que los valores CVA_{sup}^N convergen a la solución del problema general CVA_{sup} . Esto se justifica mediante el siguiente resultado, cuya demostración requiere los conceptos desarrollados en el Apéndice C.

Teorema 6.1.4. Sean X, Y vectores aleatorios en \mathbb{R}^d con distribuciones de probabili-

⁶Para evitar sobrecargar la notación, omitimos la referencia explícita a ω en las ecuaciones.

dad $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, respectivamente. Entonces, se cumple que:

$$\lim_{N \rightarrow \infty} \sup_{\pi \in \Pi(\mu_N, \nu_N)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi = \sup_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi,$$

casi seguramente con respecto a \mathbb{P} .

Demostración. Por definición de la métrica de Wasserstein W_2 (ver sección B.1) en $\mathcal{P}_2(\mathbb{R}^d)$, se tiene la expresión:

$$\begin{aligned} W_2^2(\mu, \nu) &= \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 d\pi \\ &= \int_{\mathbb{R}^d} \|x\|^2 d\mu + \int_{\mathbb{R}^d} \|y\|^2 d\nu - 2 \sup_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi. \end{aligned}$$

Por el lema 6.1.2, se garantiza la convergencia puntual $\mu_N \rightarrow \mu$ y $\nu_N \rightarrow \nu$ casi seguramente con respecto a \mathbb{P} . Esto implica que:

$$\begin{aligned} \int_{\mathbb{R}^d} \|x\|^2 d\mu_N &\rightarrow \int_{\mathbb{R}^d} \|x\|^2 d\mu, \quad \text{casi seguramente con respecto a } \mathbb{P}, \\ \int_{\mathbb{R}^d} \|y\|^2 d\nu_N &\rightarrow \int_{\mathbb{R}^d} \|y\|^2 d\nu, \quad \text{casi seguramente con respecto a } \mathbb{P}. \end{aligned}$$

Además, la convergencia débil en $\mathcal{P}_2(\mathbb{R}^d)$ implica que $\mu_N \rightarrow \mu$ y $\nu_N \rightarrow \nu$ en la métrica de Wasserstein. Es decir,

$$W(\mu_N, \mu) \rightarrow 0, \quad W(\nu_N, \nu) \rightarrow 0, \quad \text{casi seguramente con respecto a } \mathbb{P}.$$

Aplicando la desigualdad triangular en el espacio de Wasserstein, se obtiene:

$$0 \leq W(\mu_N, \nu_N) \leq W(\mu_N, \mu) + W(\mu, \nu) + W(\nu, \nu_N), \quad \text{casi seguramente.}$$

Luego, por el teorema del sandwich, se concluye que:

$$W(\mu_N, \nu_N) \rightarrow W(\mu, \nu), \quad \text{casi seguramente con respecto a } \mathbb{P}.$$

Finalmente, utilizando estos resultados, obtenemos:

$$-2 \sup_{\pi \in \Pi(\mu_N, \nu_N)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 d\pi \rightarrow -2 \sup_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 d\pi,$$

casi seguramente. Multiplicando por $-\frac{1}{2}$ en ambos lados, se obtiene el resultado deseado. \square

Observación. En la práctica, el cálculo del CVA máximo siempre se realiza mediante discretización y simulaciones de Monte Carlo. Sin embargo, este procedimiento puede ser computacionalmente costoso, especialmente en portafolios que contienen decenas de miles de contratos de derivados.

6.2. Regularización entrópica del CVA

Si bien el método previo resulta computacionalmente eficiente, la cota CVA_{sup} puede ser demasiado pesimista, ya que en muchos casos es poco probable que se alcance. Para mitigar este problema, introduciremos una penalización basada en la entropía, siguiendo el enfoque de regularización entrópica presentado en el capítulo 2. La entropía en ese contexto se interpreta como una medida de concentración de una distribución discreta. No obstante, en esta sección utilizaremos un concepto distinto: la entropía relativa de Kullback-Leibler.

Este tipo de entropía mide la divergencia entre una distribución Q y una distribución de referencia P , evaluando cuánta información se gana al pasar de P a Q . Como se explica en (Glasserman and Xu, 2014), esta métrica permite cuantificar desviaciones con respecto a un modelo base, lo que la hace útil para imponer restricciones en problemas de optimización. Para una discusión más detallada, remitimos al lector a (Glasserman and Xu, 2014).

Definición 6.2.1. Entropía de Kullback-Leibler. Sea Ω un espacio medible y sean $P, Q \in \mathcal{P}(\Omega)$ tales que $Q \ll P$. Definimos la entropía relativa de Q con respecto a P como:

$$D(Q|P) = \mathbb{E}_P \left[h \left(\frac{dQ}{dP} \right) \right] = \mathbb{E}_Q \left[\ln \left(\frac{dQ}{dP} \right) \right],$$

donde la función $h : [0, +\infty) \rightarrow \mathbb{R}$ se define como:

$$h(x) = \begin{cases} 0, & \text{si } x = 0, \\ x \ln(x), & \text{si } x > 0. \end{cases}$$

La función h es continua y convexa. Su gráfico se presenta en la figura 2.1.

Observación. En el caso discreto, cuando Ω es finito, la entropía relativa se expresa como:

$$D(Q|P) = \sum_{\omega \in \Omega} Q(\omega) \ln \left(\frac{Q(\omega)}{P(\omega)} \right).$$

Aunque la entropía de Kullback-Leibler no define una métrica en $\mathcal{P}(\Omega)$ (pues no es simétrica ni satisface la desigualdad triangular), posee algunas propiedades clave que la hacen útil en problemas de optimización.

Lema 6.2.2. En el contexto de la definición 6.2.1, se cumplen las siguientes propiedades:

1. $D(Q|P) \geq 0$ para toda Q .
2. $D(Q|P) = 0$ si y sólo si $Q = P$.
3. $D(Q|P)$ es convexa con respecto a Q .

Demostración. Para demostrar (1) y (2), utilizamos la convexidad estricta de $h(x)$ y aplicamos la desigualdad de Jensen:

$$D(Q|P) = \mathbb{E}_P \left[h \left(\frac{dQ}{dP} \right) \right] \geq h \left(\mathbb{E}_P \left[\frac{dQ}{dP} \right] \right) = h(\mathbb{E}_Q[1]) = h(1) = 0.$$

La igualdad se alcanza si y sólo si dQ/dP es constante, lo que implica que $Q = P$.

Para (3), sean $Q_1, Q_2 \in \mathcal{P}(\Omega)$ absolutamente continuas con respecto a P y sea $t \in [0,1]$. Dado que $h(x)$ es convexa, se cumple:

$$\begin{aligned} D(tQ_1 + (1-t)Q_2|P) &= \mathbb{E}_P \left[h \left(t \frac{dQ_1}{dP} + (1-t) \frac{dQ_2}{dP} \right) \right] \\ &\leq \mathbb{E}_P \left[t h \left(\frac{dQ_1}{dP} \right) + (1-t) h \left(\frac{dQ_2}{dP} \right) \right] \\ &= tD(Q_1|P) + (1-t)D(Q_2|P). \end{aligned}$$

Esto prueba la convexidad de $D(Q|P)$. □

Nuestro modelo de referencia será la medida producto $P = \mu \times \nu$, que supone independencia entre el riesgo de mercado y el riesgo de crédito. Esta elección es arbitraria, aunque se puede justificar argumentando que el default de la contraparte suele depender mayormente de factores internos⁷ y no exclusivamente de la evolución del mercado⁸. De esta forma, regularizamos desviaciones con respecto a un modelo que asume independencia total entre el riesgo de mercado y el riesgo de crédito.

A continuación, presentamos dos enfoques para regularizar el problema: uno basado en restricciones y otro basado en penalización.

⁷Si la contraparte es una empresa o un individuo, el default puede deberse a una mala gestión financiera y no necesariamente a condiciones de mercado adversas.

⁸Sin embargo, en algunos casos el default está altamente correlacionado con el mercado. Por ejemplo, si una entidad sufre pérdidas significativas en sus posiciones de mercado, podría enfrentar problemas de liquidez que la lleven a incumplir sus obligaciones.

Definición 6.2.3. Sean $\eta, \theta > 0$. Definimos los siguientes problemas de optimización:

■ **Optimización restringida:**

$$\mathcal{P}_{\text{res},\eta} : \sup_{\pi \in \Pi(\mu, \nu), D(\pi|P) \leq \eta} \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi.$$

■ **Optimización penalizada:**

$$\mathcal{P}_{\text{pen},\theta} : \sup_{\pi \in \Pi(\mu, \nu)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi - \frac{1}{\theta} D(\pi|P) \right\}.$$

donde $P = \mu \times \nu$ es la medida de referencia que asume independencia.

En la formulación restringida, cuando $\eta = 0$, la única distribución permitida es $\pi = \mu \times \nu$ (por el lema 6.2.2). En el otro extremo, cuando $\eta \rightarrow \infty$, no se impone ninguna restricción y el problema coincide con la cota pesimista $CV A_{\text{sup}}$.

Por otro lado, en la formulación penalizada, si $\theta = 0$, la penalización por entropía es infinita a menos que $\pi = \mu \times \nu$, que se convierte en la solución única. En el caso opuesto, cuando $\theta \rightarrow \infty$, la penalización desaparece y el problema vuelve a ser $CV A_{\text{sup}}$.

Observación. En la práctica, la formulación penalizada es preferida debido a su mejor comportamiento numérico y su relación con la regularización entrópica, lo que permite resolver el problema mediante métodos eficientes como el algoritmo de Sinkhorn.

La siguiente figura ilustra cómo varía la solución del problema penalizado según el parámetro $\epsilon = \frac{1}{\theta}$. Cuando ϵ es grande (θ pequeño), la penalización induce una solución muy próxima a la medida producto. Conforme ϵ disminuye, la solución se aleja de la independencia, lo que permite capturar dependencias entre los riesgos de mercado y crédito.

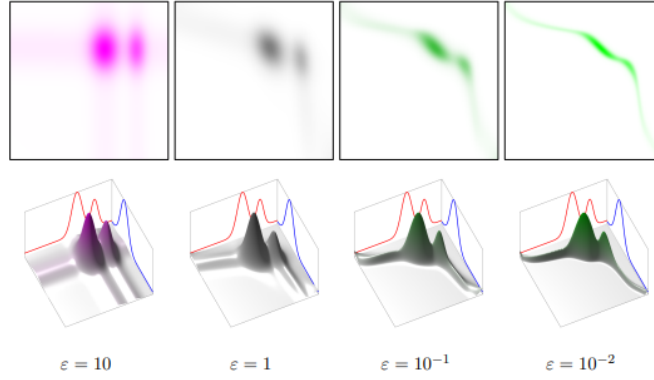


Figura 6.1 Tomado de (Peyre and Cuturi, 2019). El impacto de $\epsilon = \frac{1}{\theta}$ en las soluciones del problema penalizado. La segunda fila muestra las densidades marginales (en rojo y azul) y la solución óptima.

6.3. Existencia y unicidad de la solución

Se puede demostrar que ambos enfoques de regularización son, en cierto sentido, equivalentes. Específicamente, para cada $\theta > 0$, existe un $\eta > 0$ tal que los problemas $\mathcal{P}_{\text{pen},\theta}$ y $\mathcal{P}_{\text{res},\eta}$ tienen el mismo optimizador.

Para establecer este resultado, primero probamos que el problema penalizado $\mathcal{P}_{\text{pen},\theta}$ tiene una única solución. (Glasserman and Yang, 2015) hace referencia al teorema 3 de (Rüschendorf and Thomsen, 1993) para justificar la existencia de un óptimo único. En este documento, en lugar de depender de ese resultado, imponemos una condición adicional:

$$\mathbb{E}_{\mu \times \nu}[e^{\theta \langle x, y \rangle}] < +\infty,$$

y proporcionamos una demostración rigurosa basada en esta hipótesis. El siguiente lema es una versión adaptada del resultado en (Rüschendorf and Thomsen, 1993).

Lema 6.3.1. Sean $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ y $P \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. Si $D(\pi|P) < +\infty$ para algún $\pi \in \Pi(\mu, \nu)$, entonces existe una única medida $\hat{\pi} \in \Pi(\mu, \nu)$ tal que:

$$D(\hat{\pi}|P) = \inf_{\pi \in \Pi(\mu, \nu)} D(\pi|P).$$

Teorema 6.3.2. Sean $\theta > 0$ y $P = \mu \times \nu$. Si $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ y $\mathbb{E}_P[e^{\theta\langle x, y \rangle}] < +\infty$, entonces el problema $\mathcal{P}_{\text{pen}, \theta}$ tiene una única solución $\pi^\theta \in \Pi(\mu, \nu)$.

Demostración. Definamos la medida de probabilidad $\hat{P} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ con $\hat{P} \approx P = \mu \times \nu$ dada por

$$d\hat{P} = \frac{e^{\theta\langle x, y \rangle}}{\mathbb{E}_P[e^{\theta\langle x, y \rangle}]} dP.$$

Entonces, para cada $\pi \in \Pi(\mu, \nu)$ con $\pi \ll P$, tenemos $\pi \ll \hat{P}$. Más aún, las derivadas de Radon-Nikodym se comportan de manera deseable:

$$\frac{d\pi}{d\hat{P}} = \frac{d\pi}{dP} \frac{dP}{d\hat{P}} \quad (6.4)$$

$$= \frac{\mathbb{E}_P[e^{\theta\langle x, y \rangle}] d\pi}{e^{\theta\langle x, y \rangle} dP} \quad (6.5)$$

El problema $\mathcal{P}_{\text{pen}, \theta}$ implica encontrar

$$\sup_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi - \frac{1}{\theta} \int_{\mathbb{R}^d \times \mathbb{R}^d} \ln \left(\frac{d\pi}{d\hat{P}} \right) d\pi \quad (6.6)$$

$$= -\frac{1}{\theta} \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \ln \left(\frac{d\pi}{e^{\theta\langle x, y \rangle} dP} \right) d\pi \quad (6.7)$$

$$= -\frac{1}{\theta} \inf_{\pi \in \Pi(\mu, \nu)} \left[\int_{\mathbb{R}^d \times \mathbb{R}^d} \ln \left(\frac{\mathbb{E}_P[e^{\theta\langle x, y \rangle}] d\pi}{e^{\theta\langle x, y \rangle} dP} \right) d\pi - \ln(\mathbb{E}_P[e^{\theta\langle x, y \rangle}]) \right] \quad (6.8)$$

$$= -\frac{1}{\theta} \inf_{\pi \in \Pi(\mu, \nu)} \left[-\ln(\mathbb{E}_P[e^{\theta\langle x, y \rangle}]) + \int_{\mathbb{R}^d \times \mathbb{R}^d} \ln \left(\frac{d\pi}{d\hat{P}} \right) d\pi \right] \quad (6.9)$$

$$= \frac{1}{\theta} \left[\ln(\mathbb{E}_P[e^{\theta\langle x, y \rangle}]) - \inf_{\pi \in \Pi(\mu, \nu)} D(\pi | \hat{P}) \right]. \quad (6.10)$$

Verificamos que se cumple la condición del lema 6.3.1: el mismo $P = \mu \times \nu$ da entropía finita.

$$D(P | \hat{P}) = \mathbb{E}_P \left[\ln \left(\frac{dP}{d\hat{P}} \right) \right] \quad (6.11)$$

$$= \mathbb{E}_P \left[\ln \left(\frac{\mathbb{E}_P[e^{\theta\langle x, y \rangle}]}{e^{\theta\langle x, y \rangle}} \right) \right] \quad (6.12)$$

$$= \ln(\mathbb{E}_P[e^{\theta\langle x, y \rangle}]) - \theta \mathbb{E}_P[\langle x, y \rangle]. \quad (6.13)$$

Por hipótesis, $0 < \mathbb{E}_P[e^{\theta\langle x, y \rangle}] < +\infty$, por lo que $\ln(\mathbb{E}_P[e^{\theta\langle x, y \rangle}]) \in \mathbb{R}$. Además, como μ, ν son cuadrado-integrables, podemos acotar

$$\mathbb{E}_P[|\langle x, y \rangle|] \leq \mathbb{E}_P \left[\frac{||x||^2 + ||y||^2}{2} \right] = \frac{1}{2} \mathbb{E}_\mu[||x||^2] + \frac{1}{2} \mathbb{E}_\nu[||y||^2] < +\infty.$$

Por ende $\mathbb{E}_P[\langle x, y \rangle] \in \mathbb{R}$ y luego $D(P|\hat{P}) \in \mathbb{R}$. Por último, el lema 6.3.1 garantiza que el ínfimo en 6.10, y por ende el problema $\mathcal{P}_{\text{pen},\theta}$, tienen un único optimizador $\pi^\theta \in \Pi(\mu, \nu)$. \square

Observación. El resultado anterior justifica el uso de regularización entrópica para el problema del CVA. La existencia y unicidad del óptimo permiten resolver el problema mediante métodos computacionales eficientes, como el algoritmo de Sinkhorn.

Para concluir esta sección, presentamos el resultado de Glasserman and Yang (2015) sobre la equivalencia entre el problema penalizado y el problema restringido.

Teorema 6.3.3. Sea $\theta > 0$. La solución π^θ a $\mathcal{P}_{\text{pen},\theta}$ es la solución al problema $\mathcal{P}_{\text{res},\eta(\theta)}$ con

$$\eta(\theta) = D(\pi^\theta|P).$$

La asignación $\theta \rightarrow \eta(\theta)$ es creciente.

Demostración. Ya vimos que $\mathcal{P}_{\text{pen},\theta}$ tiene una única solución π^θ . Suponga que π^θ no es óptimo en el problema $\mathcal{P}_{\text{res},\eta(\theta)}$. Entonces existe un $\pi \in \Pi(\mu, \nu)$ tal que $D(\pi|P) \leq D(\pi^\theta|P)$ y

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi > \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi^\theta$$

Esto implica que

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi - \frac{1}{\theta} D(\pi|P) > \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi^\theta - \frac{1}{\theta} D(\pi^\theta|P),$$

lo cual contradice la optimalidad de π^θ para el problema penalizado $\mathcal{P}_{\text{pen},\theta}$. Ahora probamos que la asignación $\theta \rightarrow \eta(\theta)$ es creciente. Sean $\theta_2 > \theta_1 > 0$ y $\pi^{\theta_1}, \pi^{\theta_2}$ las soluciones a los problemas penalizados $\mathcal{P}_{\text{pen},\theta_1}, \mathcal{P}_{\text{pen},\theta_2}$, respectivamente. Si $\pi^{\theta_1} = \pi^{\theta_2}$ entonces sus divergencias son iguales, o sea $\eta(\theta_1) = \eta(\theta_2)$. Caso contrario, dado que π^{θ_2} es la solución única al problema $\mathcal{P}_{\text{pen},\theta_2}$, entonces

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi^{\theta_2} - \frac{1}{\theta_2} D(\pi^{\theta_2}|P) > \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi^{\theta_1} - \frac{1}{\theta_2} D(\pi^{\theta_1}|P). \quad (6.14)$$

Si $\int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi^{\theta_2} \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi^{\theta_1}$, entonces $D(\pi^{\theta_2}|P) < D(\pi^{\theta_1}|P)$, o sea

$$\left(\frac{1}{\theta_2} - \frac{1}{\theta_1} \right) D(\pi^{\theta_2}|P) > \left(\frac{1}{\theta_2} - \frac{1}{\theta_1} \right) D(\pi^{\theta_1}|P). \quad (6.15)$$

Sumando las inecuaciones 6.14 y 6.15 obtenemos

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi^{\theta_2} - \frac{1}{\theta_1} D(\pi^{\theta_2} | P) > \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi^{\theta_1} - \frac{1}{\theta_1} D(\pi^{\theta_1} | P), \quad (6.16)$$

lo que contradice la optimalidad de π^{θ_1} . Luego, solo puede ocurrir

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi^{\theta_2} > \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi^{\theta_1}. \quad (6.17)$$

En caso $D(\pi^{\theta_2} | P) \leq D(\pi^{\theta_1} | P)$, podemos sumar $-\frac{1}{\theta_1} D(\pi^{\theta_2} | P) \geq -\frac{1}{\theta_1} D(\pi^{\theta_1} | P)$ para obtener la ecuación 6.16: otra vez llegamos a una contradicción. Así, $\eta(\theta_2) > \eta(\theta_1)$. \square

Observación. No podemos asegurar la converso. Es decir, no garantizamos que para cada $\eta > 0$ existe un $\theta(\eta) > 0$ tal que la solución a $\mathcal{P}_{\text{res}, \eta}$ sea solución a $\mathcal{P}_{\text{pen}, \theta(\eta)}$.

6.4. Algoritmo de Sinkhorn para la regularización entrópica del CVA

Rescatando el contexto del capítulo 6 y la expresión para la divergencia de Kullback-Leibler en espacios finitos, discretizamos el problema penalizado. Mostramos que esta formulación es equivalente al problema de regularización entrópica descrito en el capítulo 2, lo que permite resolverlo mediante el algoritmo de Sinkhorn.

Para cada $\theta > 0$ y $N \in \mathbb{N}$, consideramos el problema penalizado bajo el modelo de referencia $P = \mu_N \times \nu_N$:

$$\mathcal{P}_{\theta}^N : \max_{\pi^N \in \Pi(\mu_N, \nu_N)} \sum_{i=1}^N \sum_{j=1}^{d+1} \left\{ \pi_{ij}^N \langle X_i, y_j \rangle - \frac{1}{\theta} \pi_{ij} \ln \left(\frac{\pi_{ij}}{\mu_N(X_i) \nu_N(y_j)} \right) \right\}.$$

Reordenando términos, los sumandos pueden reescribirse como:

$$\pi_{ij}^N \left(\langle X_i, y_j \rangle + \frac{1}{\theta} \ln(\mu_N(X_i) \nu_N(y_j)) \right) - \frac{1}{\theta} \pi_{ij} \ln(\pi_{ij}).$$

Definiendo la función de costo

$$c(X_i, y_j) = - \left(\langle X_i, y_j \rangle + \frac{1}{\theta} \ln(\mu_N(X_i) \nu_N(y_j)) \right),$$

el problema se reformula como:

$$\mathcal{P}_\theta^N : - \min_{\pi \in \Pi(\mu_N, \nu_N)} \sum_{i=1}^N \sum_{j=1}^{d+1} \pi_{ij} c(X_i, y_j) + \frac{1}{\theta} \pi_{ij} \ln(\pi_{ij}).$$

Este problema es equivalente a la regularización entrópica del problema de transporte óptimo, lo que permite su resolución mediante el algoritmo de Sinkhorn. En este contexto, identificamos $\mathcal{P}_{\frac{1}{\theta}}$ con los conjuntos discretos $\mathcal{X} = \{X_1, \dots, X_N\}$ y $\mathcal{Y} = \{y_1, \dots, y_{d+1}\}$, y la función de costo previamente definida.

Por la proposición 2.3.5, el algoritmo de Sinkhorn converge a la solución óptima de \mathcal{P}_θ^N . Además, la función objetivo es convexa en los $N \times (d+1)$ parámetros de optimización, lo que garantiza la optimalidad de la solución obtenida vía Lagrange.

A diferencia del caso estándar, aquí no se impone que los conjuntos tengan la misma cardinalidad. En este caso, definimos $\varepsilon = 1/\theta$, y notamos que la constante $1/\theta$ no se ha incorporado explícitamente en la formulación del problema.

6.4.1. Convergencia del método

Presentamos ahora un resultado de convergencia cuando el número de simulaciones tiende a infinito. Para ello, definimos la función de ganancia regularizada para dos medidas de probabilidad arbitrarias $\pi_1, \pi_2 \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$:

$$G(\pi_1, \pi_2) := \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi_1 - \frac{1}{\theta} D(\pi_1 | \pi_2).$$

Teorema 6.4.1. Sean X, Y vectores aleatorios en \mathbb{R}^d con distribuciones $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. Si además $\mathbb{E}_{\mu \times \nu} [e^{\theta \langle x, y \rangle}] < +\infty$ y Y tiene soporte finito, entonces:

1. El valor óptimo en \mathcal{P}_θ^N converge al valor óptimo en el problema $\mathcal{P}_{\text{pen}, \theta}$:

$$\lim_{N \rightarrow \infty} \max_{\pi \in \Pi(\mu_N, \nu_N)} G(\pi, \mu_N \times \nu_N) = \sup_{\pi \in \Pi(\mu, \nu)} G(\pi, \mu \times \nu).$$

2. El maximizador $\pi_N^* \in \Pi(\mu_N, \nu_N)$ de $G(\cdot, \mu_N \times \nu_N)$ converge débilmente al maximizador $\pi^* \in \Pi(\mu, \nu)$ de $G(\cdot, \mu \times \nu)$.
3. El CVA máximo en \mathcal{P}_θ^N converge al CVA máximo en \mathcal{P}_θ casi seguramente:

$$\lim_{N \rightarrow \infty} \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi_N^* = \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle d\pi^* \quad \mathbb{P}\text{-c.s.}$$

Demostración. Ver el anexo C de (Glasserman and Yang, 2015). \square

En este capítulo hemos formulado un problema de transporte óptimo para calcular una cota superior del CVA. Además, hemos demostrado la existencia y unicidad de la solución óptima bajo regularización entrópica. Posteriormente, adaptamos el modelo para permitir su resolución numérica mediante métodos computacionales eficientes, probando la convergencia del algoritmo de Sinkhorn en este contexto.

Concretamente, hemos presentado dos formulaciones para la regularización entrópica y establecimos un marco riguroso para analizar la convergencia del método. Finalmente, hemos desarrollado un resultado que, en gran medida, es una contribución original de este trabajo: el teorema 6.3.2, que garantiza la existencia y unicidad del problema penalizado en un contexto generalizado.

Estos resultados no solo mejoran la comprensión teórica del problema, sino que también proporcionan herramientas prácticas para su implementación en la gestión del riesgo de crédito en mercados financieros complejos.

Capítulo 7

Acotamiento del riesgo de un portafolio

En este capítulo, exploramos cómo acotar el riesgo de un portafolio utilizando herramientas de optimización y teoría de transporte óptimo. La estructura es la siguiente: primero, presentamos el concepto de medida de riesgo, junto con dos métricas fundamentales, el *Value at Risk* (VaR) y el *Conditional Value at Risk* (CVaR). En segundo lugar, siguiendo a [Memartoluie \(2017\)](#), formulamos un problema de transporte óptimo que permite acotar medidas de riesgo que dependen de dos factores de riesgo X, Y , similar al enfoque de la sección anterior. Finalmente, basándonos en [Glasserman and Yang \(2015\)](#), aplicamos regularización entrópica para obtener cotas robustas del riesgo de modelo.

7.1. Medidas de riesgo, VaR y CVaR

Siguiendo a [Memartoluie \(2017\)](#), consideramos una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$, que representa el valor futuro de un portafolio. Nuestro objetivo es cuantificar su riesgo mediante una función $\rho(X)$, definida sobre el espacio vectorial de variables aleatorias χ en (Ω, \mathbb{P}) .

Definición 7.1.1. Una *medida de riesgo* es una función $\rho : \chi \rightarrow \mathbb{R}$ que satisface los siguientes axiomas:

1. **Monotonicidad:** si $X, Y \in \chi$ y $X \leq Y$ casi seguramente, entonces $\rho(X) \leq \rho(Y)$.

2. **Invarianza por traslación:** si $X \in \chi$ y $m \in \mathbb{R}$, entonces $\rho(X+m) = \rho(X) + m$.

La medida de riesgo es *convexa* si cumple:

$$\rho(tX + (1-t)Y) \leq t\rho(X) + (1-t)\rho(Y), \quad \forall X, Y \in \chi, \forall t \in [0,1].$$

Además, es *coherente* si satisface la homogeneidad positiva:

$$\rho(\lambda X) = \lambda\rho(X), \quad \forall X \in \chi, \lambda \geq 0.$$

Observación. Algunas propiedades clave de las medidas de riesgo coherentes incluyen:

1. La **subaditividad**, es decir, que el riesgo total de dos posiciones no es mayor que la suma de sus riesgos individuales:

$$\rho(X + Y) \leq \rho(X) + \rho(Y), \quad \forall X, Y \in \chi.$$

2. La **convexidad**, que refleja la idea de que la diversificación reduce el riesgo. Un inversor que distribuye su capital entre varias posiciones espera que su riesgo total no aumente.
3. Para una discusión más detallada, ver la sección 4.1 en [Follmer and Schied \(2002\)](#).

7.2. Value at Risk (VaR)

Una medida ampliamente utilizada en la práctica financiera es el *Value at Risk* (VaR). Se define como un cuantil de la distribución de la pérdida L de un portafolio en un horizonte de tiempo fijo $T > 0$, que dejaremos implícito.

Definición 7.2.1. Dada una variable aleatoria $L : \Omega \rightarrow \mathbb{R}$ que representa la pérdida de un portafolio, el VaR con nivel de confianza $\alpha \in [0,1)$ se define como:

$$VaR_\alpha(L) = \inf \{ \ell \in \mathbb{R} \mid \mathbb{P}(L \geq \ell) \leq 1 - \alpha \}.$$

Este valor indica que la probabilidad de que las pérdidas superen $VaR_\alpha(L)$ es como máximo $1 - \alpha$. En la práctica, los valores más utilizados son $\alpha = 0,999$ con un horizonte de un año ([Memartoluie, 2017](#)). Sin embargo, el VaR presenta dos deficiencias significativas:

1. No es una medida de riesgo coherente, ya que no satisface la propiedad de subaditividad. En otras palabras, no garantiza que la diversificación reduzca el riesgo.
2. No ofrece información sobre la magnitud de las pérdidas más allá del umbral $VaR_\alpha(L)$. Esto significa que ignora eventos de cola extrema en la distribución de pérdidas.

7.3. Conditional Value at Risk (CVaR)

Para superar las limitaciones del VaR, se introdujo el *Conditional Value at Risk* (CVaR), que es una medida coherente y proporciona una mejor caracterización del riesgo.

Definición 7.3.1. Dada una variable aleatoria $L : \Omega \rightarrow \mathbb{R}$ que representa la pérdida de un portafolio, el CVaR con nivel de confianza $\alpha \in [0,1)$ se define como:

$$CVaR_\alpha(L) = \frac{1}{1-\alpha} \int_\alpha^1 VaR_\xi(L) d\xi.$$

Para cada $\alpha \in [0,1)$, el $CVaR_\alpha$ es una medida de riesgo coherente¹. A diferencia del VaR, el CVaR considera no solo la probabilidad de pérdidas extremas, sino también su magnitud.

Para ilustrar estas propiedades, consideremos pérdidas que siguen una distribución exponencial con parámetro $\lambda > 0$, cuya función de distribución es:

$$F_L(x) = 1 - e^{-\lambda x}.$$

Dado que la distribución es continua, podemos calcular el VaR resolviendo la ecuación $F_L(x) = \alpha$, obteniendo:

$$VaR_\alpha(L) = \frac{1}{\lambda} \ln \left(\frac{1}{1-\alpha} \right).$$

A partir de esto, el CVaR se obtiene integrando:

¹Dado que es una medida convexa, es más fácil de manejar computacionalmente. Además, su optimización en un portafolio con múltiples activos es más eficiente que la del VaR.

$$CVaR_\alpha(L) = \frac{1}{\lambda(1-\alpha)} \int_\alpha^1 \ln\left(\frac{1}{1-\xi}\right) d\xi.$$

Resolviendo la integral, obtenemos:

$$CVaR_\alpha(L) = \frac{1 + \ln\left(\frac{1}{1-\alpha}\right)}{\lambda}.$$

Este resultado ilustra cómo el CVaR no solo captura el umbral de pérdidas extremas, sino que también cuantifica la gravedad de dichas pérdidas.

7.4. Medidas que dependen de dos factores de riesgo

Consideremos dos vectores aleatorios X, Y que representan factores de riesgo. Como es común en la práctica, conocemos sus distribuciones marginales μ y ν , pero no su estructura de dependencia, es decir, la distribución conjunta $\pi \in \Pi(\mu, \nu)$. Si la pérdida futura del portafolio está modelada por una función $L(X, Y)$ y se evalúa mediante una medida de riesgo ρ , nuestro objetivo es determinar el riesgo máximo al que estamos expuestos. Formalmente, deseamos resolver:

$$\sup_{\pi \in \Pi(\mu, \nu)} \rho(L(X, Y)).$$

Para abordar este problema desde la teoría del transporte óptimo, resulta esencial la siguiente caracterización del $CVaR$, que nos servirá como fundamento en el planteamiento del problema de optimización. Presentamos una adaptación² del Teorema 4.47 de [Follmer and Schied \(2002\)](#) y remitimos al lector a dicho texto para la prueba.

Teorema 7.4.1. Para todo $\alpha \in [0, 1)$, el $CVaR_\alpha$ es una medida de riesgo coherente y

²Las definiciones de medida de riesgo, VaR y $CVaR$ en [Follmer and Schied \(2002\)](#) difieren ligeramente de las de [Memartoluie \(2017\)](#), pero son equivalentes. En particular, si ρ es una medida de riesgo en [Memartoluie \(2017\)](#), entonces $-\rho$ cumple la definición de medida de riesgo en [Follmer and Schied \(2002\)](#). Además, en [Follmer and Schied \(2002\)](#) se emplea el parámetro $\lambda = 1 - \alpha$ tanto para el VaR como para el $CVaR$.

admite la representación:

$$CVaR_\alpha(L) = \sup_{Q \in \mathcal{G}_\alpha} \mathbb{E}_Q[L], \quad \forall L \in \chi,$$

donde el conjunto \mathcal{G}_α está dado por:

$$\mathcal{G}_\alpha = \left\{ Q \in \mathcal{P}(\Omega) \mid Q \ll \mathbb{P} \text{ y } \frac{dQ}{d\mathbb{P}} \leq \frac{1}{1-\alpha}, \text{ } \mathbb{P}\text{-c.s.} \right\}.$$

En nuestro contexto, la pérdida $L(X,Y)$ depende de los factores de riesgo X e Y , por lo que el problema a resolver se expresa como:

$$\sup_{\pi \in \Pi(\mu, \nu)} CVaR_\alpha(L(X,Y)) = \sup_{\pi \in \Pi(\mu, \nu), Q \in \mathcal{G}_\alpha} \mathbb{E}_Q[L(X,Y)], \quad \forall L \in \chi.$$

En la práctica, se suele trabajar con distribuciones marginales discretas. Por ejemplo, como vimos anteriormente, μ y ν pueden aproximarse³ mediante sus medidas empíricas $\{\mu_N\}$ y $\{\nu_N\}$. Suponiendo que X y Y tienen soportes finitos $\{x_1, \dots, x_n\}$ y $\{y_1, \dots, y_m\}$, respectivamente, escribimos para simplificar la notación:

$$\mu_i = \mu(x_i), \quad \nu_j = \nu(y_j).$$

Asimismo, denotamos $\pi = (X,Y)_\# \mathbb{P}$ y $\gamma = (X,Y)_\# Q$ como las distribuciones conjuntas de (X,Y) bajo \mathbb{P} y Q , respectivamente. Esto nos permite formular el problema como un programa de optimización lineal:

$$\begin{cases} \text{máx} & \sum_{i=1}^n \sum_{j=1}^m \gamma_{ij} L(x_i, y_j) \\ \text{s.a} & \sum_{i=1}^n \pi_{ij} = \nu_j, \quad \forall j = 1, \dots, m, \\ & \sum_{j=1}^m \pi_{ij} = \mu_i, \quad \forall i = 1, \dots, n, \\ & 0 \leq \gamma_{ij} \leq \frac{1}{1-\alpha} \pi_{ij}. \end{cases} \quad (7.1)$$

Cabe destacar que la última restricción introduce una complejidad adicional en comparación con el problema estándar de transporte óptimo. En particular, la cota superior de γ_{ij} depende de π_{ij} , lo que hace que la resolución de este problema sea computacionalmente más costosa que el problema básico (1.2).

³En el sentido de la convergencia débil.

7.5. Riesgo de modelo

En esta sección nos enfocamos en el control del llamado *riesgo de modelo*, siguiendo la formulación presentada en [Glasserman and Xu \(2014\)](#) y manteniendo la notación establecida previamente. La evaluación del riesgo financiero requiere asumir supuestos sobre los modelos utilizados. Sin embargo, los errores en estos supuestos se traducen en errores en la medida del riesgo. El objetivo de esta sección es desarrollar herramientas para acotar dichos errores, imponiendo una desviación máxima con respecto a un modelo de referencia.

Consideremos un vector aleatorio $X : \Omega \rightarrow \mathbb{R}^d$ que representa los factores de riesgo o, en general, los elementos estocásticos del modelo. Trabajamos con un indicador de riesgo de la forma $\rho(X) = \mathbb{E}_{\mathbb{P}}[V(X)]$, donde $V : \mathbb{R}^d \rightarrow \mathbb{R}$ es una función medible. Suponemos que no conocemos la distribución real del vector aleatorio X , pero sí una distribución de referencia $P \in \mathcal{P}(\mathbb{R}^d)$. Para modelar desviaciones del modelo, consideramos distribuciones de probabilidad Q que no se alejen demasiado de P , midiendo esta desviación mediante la entropía de Kullback-Leibler $D(Q\|P)$, sujeta a una cota máxima η . En este contexto, la entropía de Kullback-Leibler cuantifica la información adicional requerida para justificar una desviación del modelo de referencia. Formulamos el problema restringido $\mathcal{M}_{\text{pen},\eta}$ como:

$$\sup_{D(Q\|P) < \eta} \mathbb{E}_Q[V(X)]. \quad (7.2)$$

Más aún, podemos extender el teorema 6.3.3 sobre la equivalencia entre el problema penalizado y el problema restringido⁴. Para una discusión más detallada, remitimos al lector a [Glasserman and Xu \(2014\)](#).

Observación. Si $P, Q \in \mathcal{P}(\mathbb{R}^d)$ son absolutamente continuas respecto a la medida de Lebesgue \mathcal{L} , con funciones de densidad respectivas f y g , y además $Q \ll P$, entonces:

$$\frac{dQ}{dP} = \frac{g}{f}.$$

En consecuencia, capturamos la desviación del modelo Q respecto a P a través del *likelihood ratio* $m := g/f$. La entropía de Kullback-Leibler se puede expresar en términos de m como:

⁴Mostraremos el caso particular $V : \mathbb{R}^{d+1} \times \mathbb{R}^{d+1} \rightarrow \mathbb{R}$ dado por $V(x, y) = \langle x, y \rangle$.

$$D(Q|P) = \mathbb{E}_P[m \ln m].$$

Observación. Dado que $g(x) = m(x)f(x)$, podemos reconstruir el modelo Q a partir de P y m . Además, para cualquier función medible $h : \mathbb{R}^d \rightarrow \mathbb{R}$, se cumple:

$$\mathbb{E}_P[m(x)h(x)] = \mathbb{E}_Q[h(x)].$$

Definición 7.5.1. Definimos el conjunto de todas las funciones de densidad respecto a P como:

$$\mathcal{M} = \{m : \mathbb{R}^d \rightarrow \mathbb{R} \text{ medible}, m \geq 0, \mathbb{E}_P[m] = 1\}.$$

Para cada $\eta > 0$, consideramos el subconjunto de \mathcal{M} dado por:

$$\mathcal{M}_\eta = \{m \in \mathcal{M} \mid \mathbb{E}_P[m \ln m] < \eta\}.$$

Reformulamos entonces (7.2) como:

$$\sup_{m \in \mathcal{M}_\eta} \mathbb{E}_P[m(X)V(X)]. \quad (7.3)$$

El problema dual correspondiente (ver [Florenzano and le Van \(2001\)](#)) es:

$$\inf_{\theta > 0} \sup_{m \in \mathcal{M}} \mathbb{E}_P \left[mV(X) - \frac{1}{\theta}(m \ln m - \eta) \right]. \quad (7.4)$$

Dado que el problema interno de maximización

$$\sup_{m \in \mathcal{M}} \mathbb{E}_P \left[mV(X) - \frac{1}{\theta}m \ln m \right]$$

es análogo al problema penalizado $\mathcal{M}_{\text{pen},\theta}$, su solución está dada por:

$$m_\theta(x) = \frac{\exp(\theta V(x))}{\mathbb{E}_P[\exp(\theta V(X))]}, \quad (7.5)$$

siempre que el denominador sea finito ([Glasserman and Xu, 2014](#)).

Definición 7.5.2. Considere el espacio de probabilidad (Ω, \mathbb{P}) . Sea $X : \Omega \rightarrow \mathbb{R}^d$ un vector aleatorio que representa los retornos de los activos de un portafolio con pesos $a = (a_1, \dots, a_d)^T$. Si la media es $\mathbb{E}_\mathbb{P}[X] = \mu$, definimos la *varianza del portafolio* como:

$$\mathbb{E}_\mathbb{P} [a^T (X - \mu)(X - \mu)^T a].$$

Dado que queremos modelar la peor desviación posible respecto a un modelo de referencia P , consideramos el problema de optimización:

$$\sup_{m \in \mathcal{M}_\eta} \mathbb{E}_P [m(X) a^T (X - \mu)(X - \mu)^T a].$$

Aquí, la función objetivo supone que conocemos con certeza la media $\mathbb{E}_Q[X] = \mu$, pero no la distribución completa. Sin embargo, cuando el modelo de referencia es normal multivariado, la solución del problema sin restricciones sobre la media sigue manteniendo $\mathbb{E}_Q[X] = \mu$. Para demostrar esto, recordemos que el optimizador de m es de la forma:

$$m_\theta(x) = C \exp(\theta a^T (x - \mu)(x - \mu)^T a), \quad (7.6)$$

para alguna constante de normalización $C > 0$. Por lo tanto, la distribución Q_θ óptima para (7.2) tiene densidad:

$$g_\theta(x) = C \exp(\theta a^T (x - \mu)(x - \mu)^T a) f(x), \quad (7.7)$$

donde $f(x)$ es la densidad del modelo de referencia P .

Un caso interesante surge cuando asumimos que el modelo de referencia sigue una distribución normal multivariada, ya que la solución óptima también resulta ser normal multivariada. Nos apoyamos en el siguiente resultado:

Teorema 7.5.3. Sea $\mu \in \mathbb{R}^d$ y $\Sigma \in \mathcal{M}_{d \times d}(\mathbb{R})$ una matriz semidefinida positiva y simétrica. Entonces, la distribución normal multivariada $\mathcal{N}(\mu, \Sigma)$ tiene densidad respecto a la medida de Lebesgue dada por:

$$f(x) = \hat{C} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right),$$

para alguna constante de normalización $\hat{C} > 0$.

Usamos el modelo de referencia $P = \mathcal{N}(\mu, \Sigma)$ con μ, Σ como en el teorema 7.5.3. La solución del problema restringido (7.2) tiene entonces la densidad:

$$g_\theta(x) = C \exp(\theta a^T (x - \mu)(x - \mu)^T a) \times \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right).$$

Observemos que los productos escalares en el exponente pueden reorganizarse como:

$$\theta a^T (x - \mu)(x - \mu)^T a = (x - \mu)^T (\theta a^T a I)(x - \mu),$$

donde I es la matriz identidad de tamaño $d \times d$. Por lo tanto, podemos escribir:

$$g_\theta(x) = C \exp \left((x - \mu)^T (\theta a^T a I) (x - \mu) - \frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right) \quad (7.8)$$

$$= C \exp \left(-\frac{1}{2} (x - \mu)^T (\Sigma^{-1} - 2\theta a^T a I) (x - \mu) \right). \quad (7.9)$$

Si el determinante de $\Sigma^{-1} - 2\theta a^T a I$ no se anula para un θ suficientemente pequeño, podemos definir:

$$\hat{\Sigma} := (\Sigma^{-1} - 2\theta a^T a I)^{-1},$$

lo que implica que $Q_\theta \sim \mathcal{N}(\mu, \hat{\Sigma})$, es decir, la distribución óptima sigue siendo normal multivariada. Este es un caso notable, ya que en general no se espera que el modelo óptimo pertenezca a la misma familia de distribuciones del modelo de referencia.

Este fenómeno se mantiene cuando consideramos una función cuadrática arbitraria de la forma $V(x) = x^T A x$, donde A es una matriz simétrica y definida positiva. Aplicando nuevamente la ecuación (7.5), obtenemos la solución:

$$m_\theta(x) = C \exp(\theta x^T A x). \quad (7.10)$$

Si el modelo de referencia sigue una distribución normal multivariada $\mathcal{N}(\mu, \Sigma)$, la densidad $g_\theta(x)$ se expresa como:

$$g_\theta(x) = C \exp \left(\theta x^T A x - \frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right).$$

Tras una reordenación algebraica, se obtiene que la distribución óptima Q_θ también es normal multivariada con:

$$\hat{\Sigma} = (\Sigma^{-1} - 2\theta A)^{-1}, \quad \hat{\mu} = \hat{\Sigma} \Sigma^{-1} \mu.$$

Este resultado es sorprendente, ya que implica que la distribución que maximiza la incertidumbre respecto al modelo sigue perteneciendo a la familia normal. Según [Glasserman and Xu \(2014\)](#), este método es sumamente poderoso, ya que evalúa una gama infinita de distribuciones en lugar de simplemente analizar la variación de parámetros dentro de una familia específica.

En este capítulo hemos formulado un problema de transporte óptimo para acotar la medida de riesgo de un portafolio. Posteriormente, mostramos que el problema de acotar el $CVaR$ se reduce a un problema de programación lineal en el caso discreto.

A partir de ello, recurrimos a la regularización entrópica para controlar el riesgo de modelo, extendiendo el enfoque más allá de los espacios de transporte óptimo tradicionales. En particular, esta metodología permite evaluar la sensibilidad del riesgo con respecto a errores de modelado sin restringir nuestro conjunto de elección a $\Pi(\mu, \nu)$.

Además, hemos explorado un caso particular en el que el modelo de referencia es una distribución normal multivariada. Sorprendentemente, en este caso, la distribución óptima que maximiza la varianza del portafolio también resulta ser normal multivariada, lo que refuerza la aplicabilidad del enfoque en contextos financieros. Este resultado es destacable, ya que en general, no se espera que el modelo óptimo pertenezca a la misma familia de distribuciones que el modelo de referencia.

Finalmente, este análisis destaca la potencia del enfoque basado en la regularización entrópica en la gestión del riesgo financiero. Como se menciona en [Glasserman and Xu \(2014\)](#), este método va más allá de simples variaciones paramétricas, permitiendo evaluar de manera robusta los efectos de desviaciones generales del modelo de referencia. Con esto concluimos la aplicación de la teoría del transporte óptimo en finanzas y el desarrollo del presente documento.

Capítulo 8

Conclusiones

En este documento, se ha realizado un análisis de las aplicaciones del transporte óptimo en economía y finanzas, enfocándonos en una selección de artículos relevantes. Se comenzó estudiando el problema de la regularización entrópica, fundamental para comprender el algoritmo de Sinkhorn-Knopp y su aplicación en problemas de transporte. Posteriormente, se exploró cómo esta técnica puede emplearse en la estimación de costos, destacando su utilidad en contextos donde la asignación eficiente de recursos es clave.

A continuación, se estudió el problema de emparejamiento tanto en el mercado matrimonial como en el mercado laboral, mostrando cómo el propio modelo conlleva naturalmente a una formulación de regularización entrópica. Además, se abordó el problema de estabilidad en el matching. Esto refuerza el objetivo central de este documento: demostrar cómo el transporte óptimo surge como una herramienta poderosa en la teoría económica.

Aunado a ello, se ha propuesto una línea de investigación que, basándose en los desarrollos previos, permita extender estos modelos al análisis del mercado laboral peruano. Se espera que este documento sirva como punto de partida para futuros proyectos en los que se apliquen técnicas de transporte óptimo a problemas económicos concretos. Es importante enfatizar que aún quedan muchas aplicaciones por explorar, como el análisis del transporte urbano. En este ámbito, luego de una estimación económica convencional de costos de movilidad, se puede formular un problema de transporte tradicional para optimizar asignaciones en infraestructuras viales y redes de transporte público.

En el ámbito financiero, se formuló un problema de transporte óptimo para acotar el *Credit Value Adjustment* (CVA) y otras medidas de riesgo. Esta metodología puede aplicarse a cualquier métrica financiera dependiente de múltiples factores de riesgo, siempre y cuando se conozcan sus distribuciones individuales. Asimismo, se abordó el problema de regularización entrópica en el caso continuo, tanto en su forma penalizada como restringida, y su capacidad para controlar desviaciones respecto a un modelo de referencia. A su vez, se mostró cómo discretizar el problema de regularización con el fin de permitir la implementación del algoritmo de Sinkhorn.

Permítanos enfatizar una vez más que la teoría del transporte óptimo sigue siendo un área de investigación muy activa, con aplicaciones tanto en economía como en finanzas. Este trabajo representa una adaptación y revisión exhaustiva de la literatura reciente en el campo, proporcionando un marco riguroso para futuras aplicaciones.

Además, a partir de estos desarrollos teóricos, pueden surgir diversas líneas de investigación, tanto empíricas como teóricas. Desde una perspectiva aplicada, la teoría del transporte óptimo puede utilizarse en problemas de emparejamiento en distintos sectores del Perú, tales como la educación, la salud y el mercado laboral. Por otro lado, en el ámbito teórico, estos métodos pueden extenderse a la formulación de modelos más generales o flexibles de asignación, como por ejemplo, el desarrollado en [Gallardo et al. \(2024\)](#), o la cartera de artículos relacionados a la regularización cuadrática.

Anexo

Apéndice A

Fundamentos de topología y análisis real

Empezamos con una breve recordatorio de algunas definiciones básicas del análisis real. Un libro de referencia de Análisis Real con aplicaciones en Economía es, por ejemplo, [Ok \(2007\)](#). Nos limitamos a definir brevemente los conceptos a los cuales aludimos, sin pasar por los resultados que suponemos conocidos para el lector (convergencia de sucesiones, continuidad, compacidad de un conjunto etc).

Definición A.0.1. Dado $A \subset \mathbb{R}$, no vacío, decimos que $\theta \in \mathbb{R}$ es **cota superior** de A si $x \leq \theta, \forall x \in A$.

Ejemplo A.0.2. Dado el conjunto $\{x \in \mathbb{R}_+ : x^2 < 2\}$, 10 es una cota superior.

Definición A.0.3. Dado $A \subset \mathbb{R}$, no vacío, decimos que $\theta \in \mathbb{R}$ es **cota inferior** de A si $\theta \leq x, \forall x \in A$.

Definición A.0.4. Dado $A \subset \mathbb{R}$, definimos $\beta = \sup A$ como la menor cota superior, es decir,

1. β es cota superior.
2. Si β' es otra cota superior, $\beta \leq \beta'$.

Ejemplo A.0.5. Consideremos nuevamente el conjunto $A = \{x \in \mathbb{R}_+ : x^2 < 2\}$. Es sencillo notar que $\sqrt{2} = \sup A$.

Definición A.0.6. Dado $A \subset \overline{\mathbb{R}}$, definimos $\alpha = \inf A$ como la mayor cota inferior, es decir,

1. α es cota inferior.
2. Si α' es otra cota superior, $\alpha' \leq \alpha$.

Notación. Sea A un conjunto y $f : A \rightarrow \overline{\mathbb{R}}$. Denotamos

$$\inf_{x \in A} f(x) = \inf f(A) \text{ y } \sup_{x \in A} f(x) = \sup f(A).$$

Proposición A.0.7. Sea A un conjunto y $f : A \rightarrow \overline{\mathbb{R}}$, entonces

$$\inf_{x \in A} -f(x) = -\sup_{x \in A} f(x).$$

A continuación brindamos la definición de distancia y espacio métrico.

Definición A.0.8. Dado un conjunto M , una métrica sobre M es una función $d : M \times M \rightarrow \mathbb{R}_+$ que cumple con las siguientes propiedades:

1. $d(x, y) = 0 \Leftrightarrow x = y$.
2. $d(x, y) = d(y, x)$.
3. $d(x, z) \leq d(x, y) + d(y, z)$ (desigualdad triangular).

Ejemplo A.0.9. En \mathbb{R}^n , $d(x, y) = \|x - y\|_2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$ define una métrica. Más aún, se conoce como la métrica inducida por la norma $\|\cdot\|_2$.

Definición A.0.10. Un espacio métrico es un par (M, d) . Es decir, un espacio métrico es un conjunto M dotado con una función distancia.

Definición A.0.11. Sea X un conjunto no vacío. Una topología sobre X , denotada \mathcal{T} , es una familia de subconjuntos de X tales que:

- a) $X, \emptyset \in \mathcal{T}$.
- b) Dado \mathcal{I} un conjunto arbitrario de índices, si $\forall i \in \mathcal{I}, A_i \in \mathcal{T}$, entonces:

$$\left(\bigcup_{i \in \mathcal{I}} A_i \right) \in \mathcal{T}.$$

c) Si para todo $i = 1, \dots, m$, $A_i \in \mathcal{T}$, entonces:

$$\left(\bigcap_{i=1}^m A_i \right) \in \mathcal{T}.$$

Llamamos *abiertos* a los elementos de la topología.

Definición A.0.12. Diremos que una familia S de subconjuntos de X genera la topología \mathcal{T} cuando

$$\mathcal{T} = \left\{ \bigcup_{B \in A} B : A \subset S \right\}.$$

Es decir, cuando la topología está formada por todas las posibles uniones en S .

Definición A.0.13. Dado un espacio métrico (M, d) , la topología inducida por d es aquella generada por los abiertos respecto a la métrica d , esto es, los conjuntos

$$B(x, \epsilon) = \{y \in M : d(x, y) < \epsilon\}.$$

Definición A.0.14. Dada $(x_n) \subset \mathbb{R}$, definimos

$$\limsup x_n = \inf_{k \geq 1} \sup\{x_n : n \geq k\} \in \overline{\mathbb{R}}$$

$$\liminf x_n = \sup_{k \geq 1} \inf\{x_n : n \geq k\} \in \overline{\mathbb{R}}.$$

Alternativamente, podemos definir

$$\limsup x_n = \lim_{k \rightarrow \infty} \sup\{x_n : n \geq k\} = \inf_{k \geq 1} \sup\{x_n : n \geq k\}$$

$$\liminf x_n = \lim_{k \rightarrow \infty} \inf\{x_n : n \geq k\} = \sup_{k \geq 1} \inf\{x_n : n \geq k\}.$$

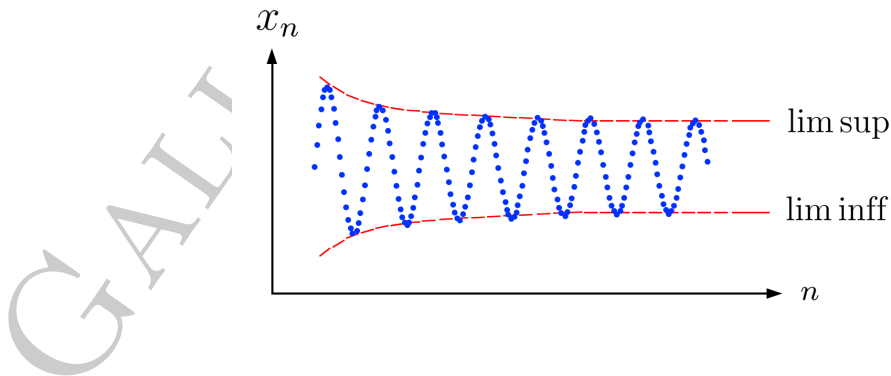


Figura A.1 Lim sup y lim inf.

Ejemplo A.0.15. Sea $x_n = \frac{(-1)^n(n+4)}{n}$, $n \in \mathbb{N}$. Entonces,

$$\begin{aligned} \sup\{x_k : k \geq n\} &= \sup\{(-1)^n(n+4)/n : (-1)^{n+1}(n+5)/(n+1), \dots\} \\ &= (n+4)/n \text{ para } n \text{ par y } (n+5)/(n+1) \text{ para } n \text{ impar} \\ &= 1. \end{aligned}$$

Así, $\limsup x_n = 1$. Análogamente, el lector puede verificar que $\liminf x_n = -1$.

Proposición A.0.16. La sucesión $(x_n) \subset \mathbb{R}$ es convergente si y sólo si $\limsup x_n = \liminf x_n$. Dicho valor común coincide con $\lim x_n$.

Definición A.0.17. Sea M un espacio métrico. Decimos que una función $f : M \rightarrow \overline{\mathbb{R}}$ es semicontinua inferior en $x_0 \in M$ cuando para toda sucesión (x_n) en M tal que $x_n \rightarrow x_0$, se tiene $f(x_0) \leq \liminf f(x_n)$.

Definición A.0.18. Sea M un espacio métrico. Decimos que una función $f : M \rightarrow \overline{\mathbb{R}}$ es semicontinua superior en $x_0 \in M$ cuando para toda sucesión (x_n) en M tal que $x_n \rightarrow x_0$, se tiene $\limsup f(x_n) \leq f(x_0)$.

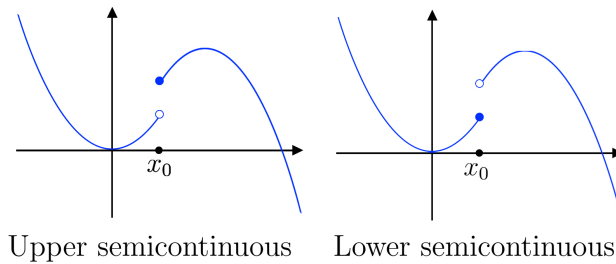


Figura A.2 Upper y lower semicontinuous.

Proposición A.0.19. Sea M un espacio métrico. La función $f : M \rightarrow \mathbb{R}$ es continua si y sólo si es semicontinua inferior y semicontinua superior.

Apéndice B

Teoría de la medida

Este anexo se basa en el curso de teoría de la medida y probabilidad de la especialidad de Matemáticas en la Pontificia Universidad Católica del Perú. Para un tratamiento más detallado del contenido del curso de probabilidad, véase [Leon and Gallardo \(2025\)](#). Por otro lado, referencias clásicas sobre teoría de la medida y probabilidad incluyen [Folland \(1984\)](#) y [Billingsley \(1995\)](#).

Dado un universo \mathcal{X} no vacío, una medida μ se encarga de asignar a subconjuntos A de \mathcal{X} , algún valor cuantitativo $\mu(A) \in [0, +\infty]$ de su tamaño.

Notación. Dado un conjunto \mathcal{X} , para todo sub-conjunto $A \subset \mathcal{X}$, definimos la función indicatriz de A de la siguiente manera:

$$\mathbf{1}_A(x) = \begin{cases} 1, & \text{si } x \in A \\ 0, & \text{caso contrario.} \end{cases}$$

Definición B.0.1. Dado un universo \mathcal{X} no vacío decimos que una colección de conjuntos $\mathcal{F} \subset \mathcal{P}(\mathcal{X})$ es σ -álgebra cuando

1. $X \in \mathcal{F}$
2. Si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$.
3. Si $\{A_j\}_{j \in \mathcal{J}} \in \mathcal{F}$, \mathcal{J} enumerable, entonces

$$\bigcup_{j \in \mathcal{J}} A_j \in \mathcal{F}.$$

Definición B.0.2. Dado un universo \mathcal{X} , no vacío, y \mathcal{F} un σ -álgebra de sub-conjuntos de \mathcal{X} , decimos que $\mu : \mathcal{F} \rightarrow [0, +\infty]$ es una medida cuando

1. $\mu(\emptyset) = 0$.
2. Si $\{A_j\}_{j \in \mathcal{J}} \subset \mathcal{F}$, siendo \mathcal{J} enumerable, tal que $A_j \cap A_i = \emptyset$, $\forall j \neq i$ (disjuntos dos a dos), entonces

$$\mu\left(\bigcup_{j \in \mathcal{J}} A_j\right) = \sum_{j \in \mathcal{J}} \mu(A_j).$$

Esta condición se conoce como σ -aditividad.

Definición B.0.3. Se conoce a $(\mathcal{X}, \mathcal{F})$ como espacio medible, mientras que, $(\mathcal{X}, \mathcal{F}, \mu)$ se conoce como espacio de medida.

Definición B.0.4. Decimos que la medida μ en el espacio medible $(\mathcal{X}, \mathcal{F})$ es una **medida finita** cuando $\mu(\mathcal{X}) < +\infty$.

Ejemplo B.0.5. Dado un espacio medible (\mathcal{X}, Σ) y un punto $x \in \mathcal{X}$, el delta de Dirac

$$\delta_x(A) = \begin{cases} 1, & \text{si } x \in A \\ 0, & \text{caso contrario} \end{cases}, \quad \forall A \in \Sigma$$

define una medida finita sobre dicho espacio. En efecto, por definición $\delta(A) \geq 0$ para todo $A \in \Sigma$. Luego, $\delta_x(\emptyset) = 0$ pues $x \notin \emptyset$. Finalmente, sea $\{A_n\}_{n \in \mathbb{N}}$ una sucesión de conjuntos en Σ disjuntos dos a dos. Si $x \in A_m$ para algún $m \in \mathbb{N}$, entonces $x \notin A_n$ para todo $n \neq m$ y además, $x \in \bigcup_{n \in \mathbb{N}} A_n$. De este modo,

$$\delta_x\left(\bigcup_{n \in \mathbb{N}} A_n\right) = 1 = \sum_{n \in \mathbb{N}} \delta_x(A_n),$$

pues $\delta_x(A_n) = 0$ para todo $n \neq m$. En caso $x \notin A_n$, $\forall n \in \mathbb{N}$, tenemos $x \notin \bigcup_{n \in \mathbb{N}} A_n$ y así

$$\delta_x\left(\bigcup_{n \in \mathbb{N}} A_n\right) = 0 = \sum_{n \in \mathbb{N}} \delta_x(A_n).$$

En ambos casos se satisface la σ -aditividad. Dado $\delta_x(X) = 1$, tenemos que la medida es finita.

Definición B.0.6. Dado \mathcal{X} no vacío y $\mathcal{C} \subset \mathcal{P}(\mathcal{X})$. Definimos el σ -álgebra generado por \mathcal{C} , $\sigma(\mathcal{C})$, como el menor σ -álgebra que contiene a \mathcal{C} , o sea

1. $\sigma(\mathcal{C})$ es σ -álgebra
2. $\mathcal{C} \subset \sigma(\mathcal{C})$
3. Si \mathcal{G} es un σ -álgebra, $\mathcal{C} \subset \mathcal{G}$, $\sigma(\mathcal{C}) \subset \mathcal{G}$.

Ejemplo B.0.7. Para cada $A \subset \mathcal{X}$, tenemos $\sigma(\{A\}) = \{\emptyset, A, A^c, \mathcal{X}\}$

Definición B.0.8. Sobre \mathbb{R}^d definimos el σ -álgebra de los borelianos por

$$\mathcal{B}_{\mathbb{R}^d} = \sigma\{\mathcal{O} \in \mathbb{R}^d : \mathcal{O} \text{ abierto}\}.$$

Este es el σ -álgebra implícito que consideramos al definir medidas los espacios euclidianos.

Proposición B.0.9. Considere el siguiente conjunto:

$$\mathcal{C} = \left\{ \prod_{i=1}^d [a_i, b_i] : a_i, b_i \in \mathbb{R}, i = 1, \dots, d \right\}.$$

Entonces, $\mathcal{B}_{\mathbb{R}^d} = \sigma(\mathcal{C})$.

Las medidas más intuitivas son la de longitud en \mathbb{R} , área en \mathbb{R}^2 y volumen en \mathbb{R}^3 . Quizás el lector desconoce, pero esta es justamente la medida de Lebesgue, la cual a cada caja le otorga como medida el producto de las longitudes de sus lados.

Teorema B.0.10. Existe una única medida \mathcal{L} en $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ tal que

$$\mathcal{L}\left(\prod_{i=1}^d [a_i, b_i]\right) = \prod_{i=1}^d (b_i - a_i).$$

Definición B.0.11. Llamamos a la medida \mathcal{L} del teorema anterior la medida de Lebesgue en \mathbb{R}^d .

Ya habiendo definido lo que es un σ -álgebra y una medida, podemos ampliar el horizonte y pensar en la integración. De los cursos de cálculo, es sabido que la integral permite medir un área o volumen. Sin embargo, la teoría de la integración se enriquece mucho cuando se construye a partir de los conceptos que se acaban de definir en este anexo. La integral de Riemman permite integrar una clase muy reducida de funciones. Por ejemplo, la función $f = \mathbf{1}_{\mathbb{Q}}$ no es Riemman-integrable. La integral de Lebesgue nace por la motivación de extender el universo de funciones que podemos integrar.

Empezamos con algunos elementos básicos de la teoría de la integración. Para más detalles, consultar [Folland \(1984\)](#).

Dado $(\mathcal{X}, \mathcal{F})$ espacio medible, definimos el conjunto de las funciones *simples* como aquellas que tienen rango finito. Denotamos

$$S(\mathcal{X}, \mathcal{F}) = \left\{ f : \mathcal{X} \rightarrow \mathbb{R} : f \doteq \sum_{i=1}^m a_i \mathbf{1}_{A_i}, A_i \in \mathcal{F} \right\}.$$

Asimismo, definimos

$$S_+(\mathcal{X}, \mathcal{F}) = \{f \in S(\mathcal{X}, \mathcal{F}) : f \geq 0\}.$$

Definición B.0.12. Dado un espacio de medida $(\mathcal{X}, \mathcal{F}, \mu)$, definimos para cada $f \in S_+$, con $f \doteq \sum_{i=1}^m a_i \mathbf{1}_{A_i}$

$$\int_{\mathcal{X}} f d\mu = \sum_{i=1}^m a_i \mu(A_i).$$

Definición B.0.13. Sean $(\mathcal{X}, \mathcal{F})$ y $(\mathcal{Y}, \mathcal{H})$ espacios medibles. Decimos que $f : \mathcal{X} \rightarrow \mathcal{Y}$ es una función medible si

$$f^{-1}(B) \in \mathcal{F}, \forall B \in \mathcal{H}.$$

Además, si $(\mathcal{Y}, \mathcal{H}) = (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ diremos que la función es Borel-medible.

Proposición B.0.14. Sean $(\mathcal{X}, \mathcal{F}), (\mathcal{Y}, \mathcal{G}), (\mathcal{Z}, \mathcal{H})$ espacios medibles. Si $f : \mathcal{X} \rightarrow \mathcal{Y}$ y $g : \mathcal{Y} \rightarrow \mathcal{Z}$ son medibles, entonces la composición $g \circ f : \mathcal{X} \rightarrow \mathcal{Z}$ es medible.

Proposición B.0.15. Sea $(\mathcal{X}, \mathcal{F})$ un espacio medible. Entonces, $\forall f : \mathcal{X} \rightarrow \mathbb{R}$, función Borel-medible positiva, existe una sucesión (f_n) en $S_+(\mathcal{X}, \mathcal{F})$ tal que $f_n(x) \uparrow f(x)$, $\forall x \in \mathcal{X}$.

Demostración. Definimos $\varphi_n : [0, +\infty) \rightarrow \mathbb{R}$, $n \in \mathbb{N}$, como

$$\varphi_n = \sum_{i=0}^{n2^n-1} i2^{-n} \mathbf{1}_{[i/2^n, (i+1)/2^n)} + n \mathbf{1}_{[n, +\infty)}.$$

Observemos que $\varphi_n(x) \uparrow x$, $\forall x \geq 0$ y $\varphi_n : [0, \infty) \rightarrow \mathbb{R}$ es Borel medible positiva. Por lo tanto, definiendo $f_n = \varphi_n \circ f$, $f_n \uparrow f$, $\forall x \in \mathbb{R}$. En efecto,

$$\underbrace{\varphi_n(f(x))}_{=f_n(x)} \leq \underbrace{\varphi_{n+1}(f(x))}_{=f_{n+1}(x)}.$$

Finalmente, $f_n \in S_+(\mathcal{X}, \mathcal{F})$ pues es composición de medibles. □

Definición B.0.16. Para $f : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ medible positiva, definimos

$$\int_{\mathcal{X}} f d\mu = \sup \left\{ \int_{\mathcal{X}} g : g \in S_+(\mathcal{X}, \mathcal{F}), 0 \leq g \leq f \right\}.$$

Para definir la integral sobre toda la clase de funciones medibles, separamos a cada función en sus partes positiva y negativa.

Definición B.0.17. Sea $f : \mathcal{X} \rightarrow \overline{\mathbb{R}}$. Definimos su parte positiva f^+ y parte negativa f^- por

$$f^+(x) = \max\{f(x), 0\}, \quad f^-(x) = \max\{-f(x), 0\}.$$

Observación. El lector puede verificar las descomposiciones $f = f^+ - f^-$ y $|f| = f^+ + f^-$.

Observación. Si f es medible, entonces f^+ y f^- también lo son.

Definición B.0.18. Para $f : X \rightarrow \overline{\mathbb{R}}$ medibles, definimos

$$\int_{\mathcal{X}} f d\mu = \int_{\mathcal{X}} f^+ d\mu - \int_{\mathcal{X}} f^- d\mu,$$

donde las integrales en el lado derecho son sobre funciones medibles positivas, y por ende ya están definidas.

Definición B.0.19. Dado un espacio de medida $(\mathcal{X}, \mathcal{F}, \mu)$, decimos que una función medible $f : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ es μ -integrable cuando

$$\int_{\mathcal{X}} |f| d\mu < +\infty.$$

A continuación exhibimos ciertas propiedades útiles de la integral.

Proposición B.0.20. Sea $(\mathcal{X}, \mathcal{F}, \mu)$ espacio de medida y $f, g : \mathcal{X} \rightarrow \mathbb{R}$ funciones μ -integrables.

1. Monotonicidad: Si $f \leq g$ entonces $\int_{\mathcal{X}} f d\mu \leq \int_{\mathcal{X}} g d\mu$
2. Linealidad: $\int_{\mathcal{X}} (af + bg) d\mu = a \int_{\mathcal{X}} f d\mu + b \int_{\mathcal{X}} g d\mu$ para todo $a, b \in \mathbb{R}$
3. $\left| \int_{\mathcal{X}} f d\mu \right| \leq \int_{\mathcal{X}} |f| d\mu.$

Observación. La integrabilidad, como bien dice el término, garantiza que la integral de f está bien comportada. Como $-|f| \leq f \leq |f|$, entonces

$$\begin{aligned} \int_{\mathcal{X}} -|f|d\mu &\leq \int_{\mathcal{X}} fd\mu \leq \int_{\mathcal{X}} |f|d\mu \\ -\int_{\mathcal{X}} |f|d\mu &\leq \int_{\mathcal{X}} fd\mu \leq \int_{\mathcal{X}} |f|d\mu \end{aligned}$$

y por ende $\int_{\mathcal{X}} fd\mu \in \mathbb{R}$

A parte de poder definir la integral sobre ellas, una gran ventaja de las funciones medibles es que podemos inducir medidas en el espacio de llegada a través de ellas.

Definición B.0.21. Push-forward de medidas. Sean $(\mathcal{X}, \mathcal{F}, \mu)$ espacio de medida y $(\mathcal{Y}, \mathcal{G})$ espacio medible y $T : \mathcal{X} \rightarrow \mathcal{Y}$ medible. Definimos la medida push-forward en $(\mathcal{Y}, \mathcal{G})$, denotada $T_{\#}\mu$, de la siguiente forma:

$$(T_{\#}\mu)(A) = \mu(T^{-1}(A)), \quad \forall A \in \mathcal{G}.$$

Ese concepto nos permite establecer un resultado análogo al famoso teorema de cambio de variables del cálculo integral.

Teorema B.0.22. Sean $(\mathcal{X}, \mathcal{F})$ y $(\mathcal{Y}, \mathcal{G})$ espacios medibles. Sean μ medida en $(\mathcal{X}, \mathcal{F})$ y $T : \mathcal{X} \rightarrow \mathcal{Y}$ medible, y defina $\nu = T_{\#}\mu$ medida en $(\mathcal{Y}, \mathcal{G})$. Si $f : \mathcal{Y} \rightarrow \overline{\mathbb{R}}$ es ν -integrable, entonces $f \circ T$ es μ -integrable y se cumple

$$\int_{\mathcal{X}} f \circ T d\mu = \int_{\mathcal{Y}} f d\nu.$$

Terminamos los fundamentos de teoría de la medida definiendo los conceptos de medida producto, medida absolutamente continua respecto a otra, y enunciando el Teorema de Radon-Nikodym. Cabe mencionar que hemos omitido en esta presentación de teoría de la medida varios resultados clásicos como el Teorema de la Convergencia Monótona, Dominada o el Lema de Fatou, ([Gall, 2022](#)).

Definición B.0.23. Dados $(\mathcal{X}_1, \mathcal{F}_1)$, $(\mathcal{X}_2, \mathcal{F}_2)$, ..., $(\mathcal{X}_K, \mathcal{F}_K)$, definimos sobre el producto cartesiano $X_1 \times \dots \times X_k$ el σ -álgebra producto $\mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_k$ como el σ -álgebra generado por las cajas.

$$\mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_k = \sigma(\{A_1 \times \dots \times A_k : A_1 \in \mathcal{F}_1, \dots, A_k \in \mathcal{F}_k\}).$$

Definición B.0.24. Diremos que una medida ν sobre $(\mathcal{X}_1 \times \dots \times \mathcal{X}_k, \mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_k)$ es una medida producto de μ_1, \dots, μ_k si

$$\nu(A_1 \times \dots \times A_k) = \mu_1(A_1) \cdots \mu_k(A_k)$$

Las medidas producto siempre existen, pues las podemos construir a partir de la condición B.0.21 haciendo uso de herramientas de teoría de la medida que no mencionaremos. Sin embargo, bajo la siguiente condición podemos asegurar que existe una única medida producto.

Definición B.0.25. Dado $(\mathcal{X}, \mathcal{F}, \mu)$ decimos que μ es σ -finita cuando existe $(A_n)_{n \in \mathbb{N}}$ tal que

$$\mathcal{X} = \bigcup_{n \in \mathbb{N}} A_n,$$

con $\mu(A_n) < \infty, \forall n \in \mathbb{N}$.

Observación. Las medidas finitas son σ -finitas.

Proposición B.0.26. Si μ_1, \dots, μ_k son σ -finitas, entonces existe una única medida producto $\mu_1 \times \dots \times \mu_k$.

Definición B.0.27. Dadas dos medidas μ, ν en un espacio medible $(\mathcal{X}, \mathcal{F})$, decimos que ν es absolutamente continua respecto a μ , denotado por $\nu \ll \mu$, cuando

$$\mu(A) = 0 \implies \nu(A) = 0, \forall A \in \mathcal{F}.$$

Teorema B.0.28. Radon-Nikodym para medidas positivas. Dado un espacio medible positiva $(\mathcal{X}, \mathcal{F})$, dos medidas σ -finitas $\mu, \nu : \mathcal{F} \rightarrow \mathbb{R}_+$ con $\nu \ll \mu$, existe una función medible f sobre $(\mathcal{X}, \mathcal{F})$ que satisface

$$\nu(A) = \int_{\mathcal{X}} \mathbf{1}_A f d\mu, \forall A \in \mathcal{F}. \quad (\text{B.1})$$

Definición B.0.29. Derivada de Radon-Nikodym. En (B.1), la función f se conoce como derivada de Radon-Nikodym y escribimos $f = d\nu/d\mu$.

Proposición B.0.30. Sean μ, ν, γ medidas en $(\mathcal{X}, \mathcal{F})$. Se cumplen:

1. Linealidad: Si $\nu, \gamma \ll \mu$ entonces $\nu + \gamma \ll \mu$ y se cumple

$$\frac{d(a\nu + b\gamma)}{d\mu} = a \frac{d\nu}{d\mu} + b \frac{d\gamma}{d\mu}.$$

2. Conmutatividad: Si $\gamma \ll \nu$ y $\nu \ll \mu$ entonces $\gamma \ll \mu$ y se cumple

$$\frac{d\gamma}{d\mu} = \frac{d\gamma}{d\nu} \frac{d\nu}{d\mu}, \quad \forall a, b \in \mathbb{R}.$$

Habiendo establecido las bases de la teoría de la medida, nos enfocamos ahora en brindar los resultados básicos de teoría de la probabilidad a los que se ha hecho referencia y han sido usados en diferentes partes de este documento.

B.1. Teoría de la probabilidad

Definición B.1.1. Sea μ una medida en el espacio medible (Ω, \mathcal{F}) . Decimos que \mathbb{P} es una medida de probabilidad si $\mathbb{P}(\Omega) = 1$. En este contexto, decimos que $(\Omega, \mathcal{F}, \mathbb{P})$ es un espacio de probabilidad.

Ejemplo B.1.2. Para cada $\omega \in \Omega$, como ya se vio, δ_ω es una medida de probabilidad

Definición B.1.3. Sea Ω un conjunto finito con $|\Omega| = N$ y \mathcal{F} la familia (finita) de todos los subconjuntos de X . Definimos la distribución uniforme $\mu = \text{Unif}(\Omega)$ por

$$\mu(A) = \frac{|A|}{N} \quad \forall A \subset \Omega.$$

Ejemplo B.1.4. Si $\mu = \text{Unif}\{1, 2, 3, 4, 5\}$, entonces $\mu(\{1, 3, 4\}) = 3/5$.

Definición B.1.5. Para cada espacio medible (Ω, \mathcal{F}) , definimos $\mathcal{P}(\Omega)$ como el conjunto de todas sus medidas de probabilidad.

Definición B.1.6. Para cada $\mathcal{X} \in \mathcal{B}_{\mathbb{R}^d}$, definimos el σ -álgebra de los borelianos en X por

$$\mathcal{B}_{\mathcal{X}} := \{A \in \mathcal{B}_{\mathbb{R}^d} : A \subset \mathcal{X}\}.$$

Sobre este, si además $\mathcal{L}(\mathcal{X}) < +\infty$, definimos la medida $\mu = \text{Unif}(\mathcal{X})$ por

$$\mu(A) = \frac{\mathcal{L}(A)}{\mathcal{L}(\mathcal{X})}, \quad \forall A \in \mathcal{B}_{\mathcal{X}}.$$

Definición B.1.7. Dado un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, una variable aleatoria es una función Borel-medible $X : \Omega \rightarrow \mathbb{R}$. De forma más general, un vector aleatorio es una función Borel-medible $X : \Omega \rightarrow \mathbb{R}^d$ en algún espacio euclideo.

Definición B.1.8. Sea $(\Omega, \mathcal{F}, \mathbb{P})$. Decimos que el vector aleatorio $X : \Omega \rightarrow \mathbb{R}$ tiene distribución o ley $\mu \in \mathcal{P}(\mathbb{R})$ cuando $\mu = X_{\#}\mathbb{P}$. Del mismo modo, la ley de un vector aleatorio es su medida push-forward $\mu \in \mathcal{P}(\mathbb{R}^d)$

Observación. La ley de X nos dice cómo se distribuyen los valores que toma dicho vector. Por definición, para cada $A \in \mathcal{B}_{\mathbb{R}^d}$ tenemos $\mu(A) = \mathbb{P}\{X \in A\}$, donde denotamos $\{X \in A\} = X^{-1}(A)$. Por ejemplo, en el caso $d = 1$, $\mu([0,1])$ indica con qué probabilidad ocurre $0 \leq X \leq 1$.

Definición B.1.9. Sea $(\Omega, \mathcal{F}, \mathbb{P})$ espacio de probabilidad y $X : \Omega \rightarrow \mathbb{R}$ una variable aleatoria con ley $\mu \in \mathcal{P}(\mathbb{R})$. Definimos la función de distribución $F_X : \mathbb{R} \rightarrow [0,1]$ por

$$F_X(s) = \mathbb{P}\{X \leq s\} \quad \forall s \in \mathbb{R}.$$

Esto es equivalente a $F_X(s) = \mu((-\infty, s])$, por lo que la función de distribución depende de sólomente de la ley de X .

Definición B.1.10. Dado el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ integrable, definimos el valor esperado como la integral de X . Escribimos

$$\mathbb{E}_{\mathbb{P}}[X] = \int_{\Omega} X d\mathbb{P}.$$

El siguiente resultado es el teorema B.0.22 expresado en el lenguaje de probabilidad.

Teorema B.1.11. Sean $(\Omega, \mathcal{F}, \mathbb{P})$ espacio de probabilidad y $X : \Omega \rightarrow \mathbb{R}^d$ vector aleatorio con ley $\mu = X_{\#}\mathbb{P}$. Si $f : \mathbb{R}^d \rightarrow \mathbb{R}$ función Borel-medible. Entonces $f(X) : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria integrable si y solo si f es μ -integrable, y se cumple

$$\mathbb{E}_{\mathbb{P}}[f(X)] = \mathbb{E}_{\mu}[f].$$

Definición B.1.12. Sean μ y ν medidas en un espacio medible $(\mathcal{X}, \mathcal{F})$. Cuando $\nu \ll \mu$, decimos que la derivada de Radon-Nykodim $f = d\nu/d\mu$ es la densidad de ν respecto a μ .

Proposición B.1.13. En el contexto de la definición anterior, para cada $h : \mathcal{X} \rightarrow \mathbb{R}$ medible y μ -integrable tenemos

$$\mathbb{E}_{\mu}[h(x)] = \mathbb{E}_{\nu}[h(x)f(x)].$$

Proposición B.1.14. Sea $(\Omega, \mathcal{F}, \mu)$ espacio de probabilidad. Para cada función Borel-medible positiva $f : \Omega \rightarrow \mathbb{R}^+$ con $\mathbb{E}_\mu[f] = 1$, la fórmula

$$\nu(A) = \int_{\Omega} \mathbf{1}_A f d\mu, \quad \forall A \in \mathcal{F}$$

define una medida de probabilidad $\nu \in \mathcal{P}(\Omega)$ que cumple $\nu \ll \mu$.

Definición B.1.15. Sea $(\Omega, \mathcal{F}, \mathbb{P})$ espacio de probabilidad y considere un evento $A \in \mathcal{F}$. Decimos que A ocurre \mathbb{P} -casi seguramente (abreviado \mathbb{P} -c.s.) cuando $\mathbb{P}(A) = 1$. Sin embargo, puede ocurrir que dicho conjunto no sea medible. En ese caso, pedimos que exista un $B \in \mathcal{F}$ de probabilidad uno tal que $B \subset A$.

En la práctica A es el conjunto donde se cumple una cierta condición que depende de $\omega \in \Omega$, y decimos que la condición ocurre \mathbb{P} -casi seguramente. Por ejemplo, si tenemos la variable aleatoria X , decimos que $X > 0$ \mathbb{P} -c.s. cuando $\mathbb{P}\{X > 0\} = 1$.

Teorema B.1.16. Desigualdad de Jensen. Sea $(\Omega, \mathcal{F}, \mathbb{P})$ espacio de probabilidad, X variable aleatoria integrable y $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ convexa. Entonces

$$\varphi(\mathbb{E}_{\mathbb{P}}[X]) \leq \mathbb{E}_{\mathbb{P}}[\varphi(X)]$$

con igualdad si y sólo si φ es una función afín o X es una constante \mathbb{P} -casi seguramente.

Definición B.1.17. Considere dos vectores aleatorios $X, Y : \Omega \rightarrow \mathbb{R}^d$ en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Decimos que X, Y son independientes cuando

$$\mathbb{P}\{X \in A, Y \in B\} = \mathbb{P}\{X \in A\}\mathbb{P}\{Y \in B\}, \quad \forall A, B \in \mathcal{B}_{\mathbb{R}^d}.$$

De forma más general, considere una colección $(X_i)_{i \in \mathcal{J}}$ de vectores aleatorios en \mathbb{R}^d , con \mathcal{J} finito o infinito. Decimos que los $(X_i)_{i \in \mathcal{J}}$ son independientes cuando:

$$\mathbb{P}\{X_{i_1} \in A_1, \dots, X_{i_n} \in A_n\} = \mathbb{P}\{X_{i_1} \in A_1\} \times \dots \times \mathbb{P}\{X_{i_n} \in A_n\},$$

para todo subconjunto de índices finito $\{i_1, \dots, i_n\} \subset \mathcal{J}$ y para todo $A_i \in \mathcal{B}_{\mathbb{R}^d}$.

Definición B.1.18. Decimos que los vectores aleatorios $\{X_i\}_{i \in \mathcal{J}}$ son independientes e idénticamente distribuidos (i.i.d) cuando son independientes y todos tienen la misma ley $\mu \in \mathcal{P}(\mathbb{R}^d)$.

Teorema B.1.19. Ley Fuerte de los Grandes Números. Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y $(X_i)_{i \in \mathbb{N}}$ una sucesión de variables aleatorias i.i.d. e integrables. Entonces

$$\frac{1}{n}(X_1 + \dots + X_n) \rightarrow \mathbb{E}_{\mathbb{P}}[X_1]$$

\mathbb{P} -casi seguramente.

Observación. Ser idénticamente distribuidos implica que $\mathbb{E}_{\mathbb{P}}[X_1] = \mathbb{E}_{\mathbb{P}}[X_i]$ para cada $i \in \mathbb{N}$. Además, la convergencia \mathbb{P} -casi seguramente indica que el conjunto $\{\omega \in \Omega : \frac{1}{n}(X_1(\omega) + \dots + X_n(\omega)) \rightarrow \mathbb{E}_{\mathbb{P}}[X_1]\}$ es o contiene a un conjunto medible de probabilidad uno.

Así como existen distintos tipos de convergencia de funciones, como la convergencia puntual o la convergencia uniforme, también las hay para medidas de probabilidad. Nos centramos en convergencia de probabilidades en $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$.

Definición B.1.20. Sea $\mu \in \mathcal{P}(\mathbb{R}^d)$ y $(\mu_n)_{n \in \mathbb{N}}$ una sucesión en $\mathcal{P}(\mathbb{R}^d)$. Decimos que $(\mu_n)_{n \in \mathbb{N}}$ converge puntualmente¹ cuando

$$\mu_n(A) \rightarrow \mu(A), \quad \forall A \in \mathcal{B}_{\mathbb{R}^d}.$$

Definición B.1.21. Sea $\mu \in \mathcal{P}(\mathbb{R}^d)$ y $(\mu_n)_{n \in \mathbb{N}}$ una sucesión en $\mathcal{P}(\mathbb{R}^d)$. Decimos que $(\mu_n)_{n \in \mathbb{N}}$ converge debilmente a μ cuando

$$\int_{\mathbb{R}^d} \varphi d\mu_n \rightarrow \int_{\mathbb{R}^d} \varphi d\mu, \quad \forall \varphi \in C_b(\mathbb{R}^d)$$

donde $C_b(\mathbb{R}^d)$ es el conjunto de todas las funciones $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ continuas y acotadas. Denotamos esta convergencia por $\mu_n \xrightarrow{w} \mu$

Proposición B.1.22. $\mu_n \rightarrow \mu$ puntualmente implica $\mu_n \xrightarrow{w} \mu$.

Pasamos ahora a conceptos vinculados a la teoría del transporte óptimo, altamente vinculados con las nociones expuestas en medida y teoría de la probabilidad. Comenzamos por el concepto de espacio polaco.

Definición B.1.23. Decimos que un espacio métrico (\mathcal{X}, d) es *polaco* cuando es completo y separable.

¹En inglés *setwise convergence*, traducido como convergencia conjunto a conjunto, es un nombre más apropiado.

Para las definiciones posteriores seguimos a [Villani \(2009\)](#).

Definición B.1.24. Sea (\mathcal{X}, d) un espacio polaco y $p \in [1, +\infty)$. Para cada $\mu, \nu \in \mathcal{P}(\mathcal{X})$, definimos la distancia de Wasserstein de orden p entre μ y ν por

$$\mathcal{W}_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X}} d(x, y)^p d\pi \right)^{\frac{1}{p}}.$$

Esta no es una métrica todavía, pues puede tomar el valor $+\infty$. Para que lo sea, debemos tomarla en el siguiente espacio.

Definición B.1.25. Sea (\mathcal{X}, d) un espacio polaco y $p \in [1, +\infty)$. Definimos el espacio de Wasserstein de orden p por

$$\mathcal{P}_p(\mathcal{X}) = \left\{ \mu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} d(x_0, x)^p d\mu < +\infty \right\},$$

donde $x_0 \in \mathcal{X}$ es arbitrario y el espacio $\mathcal{W}_p(\mathcal{X})$ no depende de la elección del x_0 .

Definición B.1.26. Sea (\mathcal{X}, d) un espacio polaco, $p \in [1, +\infty)$, $\mu \in \mathcal{P}(\mathcal{X})$ y $(\mu_n)_{n \in \mathbb{N}}$ sucesión en $\mathcal{P}(\mathcal{X})$. Decimos que μ_n converge debilmente en $\mathcal{P}_p(\mathcal{X})$ cuando $\mu_k \xrightarrow{w} \mu$ y

$$\int_{\mathcal{X}} d(x_0, x)^p d\mu_n \rightarrow \int_{\mathcal{X}} d(x_0, x)^p d\mu$$

para algún $x_0 \in \mathcal{X}$. Denotamos $\mu_n \xrightarrow{\mathcal{P}_p} \mu$. De hecho, si se cumple para un $x_0 \in \mathcal{X}$ se cumple para cualquier elección de x_0 .

Teorema B.1.27. Sea (\mathcal{X}, d) espacio polaco y $p \in [1, +\infty)$. Entonces la topología inducida por la métrica \mathcal{W}_p en \mathcal{P}_p es la topología de la convergencia débil en $\mathcal{P}_p(\mathcal{X})$. Esto es,

$$\mathcal{W}(\mu_n, \mu) \rightarrow 0 \Leftrightarrow \mu_n \xrightarrow{\mathcal{P}_p} \mu.$$

Apéndice C

Elementos de análisis convexo

A continuación, recordamos algunos elementos fundamentales del análisis convexo que serán de utilidad para probar un lema necesario para la prueba de la convergencia del algoritmo SISTA.

Definición C.0.1. Una función $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$, definida sobre un conjunto S convexo no vacío, es convexa si para cualesquiera $x, y \in S$

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y), \quad \theta \in [0, 1].$$

Teorema C.0.2. Sea $S \subset \mathbb{R}^n$ un conjunto no vacío, abierto y convexo. Entonces, $f : S \rightarrow \mathbb{R}$ es convexa si y solamente si

$$f(y) \geq f(x) + \nabla f(x)(y - x), \quad \forall x, y \in S. \quad (\text{C.1})$$

Demostración. \implies Si f es estrictamente convexa sobre S , por la Definición C.0.1

$$f(\theta y + (1 - \theta)x) \leq \theta f(y) + (1 - \theta)f(x)$$

$$f(x + \theta(y - x)) \leq f(x) + \theta(f(y) - f(x))$$

$$f(x + \theta(y - x)) - f(x) \leq \theta(f(y) - f(x))$$

$$f(x) + \frac{f(x + \theta(y - x)) - f(x)}{\theta} \leq f(y).$$

Luego, como f es diferenciable, haciendo $\theta \rightarrow 0$

$$f(x) + \underbrace{\lim_{\theta \rightarrow 0^+} \frac{f(x + \theta(y - x)) - f(x)}{\theta}}_{df_x(v) = \nabla f(x)^T (y - x)} \leq f(y).$$

\Leftarrow Como (C.1) vale para cualesquiera x, y , definamos $z = \theta x + (1 - \theta)y$

$$f(x) \geq f(z) + \nabla f(z)^T(x - z)$$

$$f(y) \geq f(z) + \nabla f(z)^T(y - z).$$

Luego,

$$\theta f(x) \geq \theta f(z) + \theta \nabla f(z)^T(x - z)$$

$$(1 - \theta)f(y) \geq (1 - \theta)f(z) + (1 - \theta)\nabla f(z)^T(y - z).$$

Sumando,

$$\begin{aligned} \theta f(x) + (1 - \theta)f(y) &\geq f(z) + \nabla f(z)^T(\theta x + (1 - \theta)y - z) \\ &= f(z). \end{aligned}$$

□

Corolario C.0.3. Se cumple que, dada f en el contexto del teorema C.0.2, si f es convexa, entonces

$$(\nabla f(x) - \nabla f(y))^T(x - y) \geq 0.$$

Demostración. Tenemos

$$f(y) \geq f(x) + \nabla f(x)^T(y - x)$$

$$f(x) \geq f(y) + \nabla f(y)^T(x - y).$$

Con lo cual,

$$f(y) + f(x) \geq f(x) + f(y) + \nabla f(x)^T(y - x) + \nabla f(y)^T(x - y)$$

$$0 \geq \nabla f(x)^T(y - x) + \nabla f(y)^T(x - y)$$

$$0 \geq -\nabla f(x)^T(x - y) + \nabla f(y)^T(x - y).$$

Así

$$(\nabla f(x) - \nabla f(y))^T(x - y) \geq 0.$$

□

Definición C.0.4. Una función f , en el contexto del teorema C.0.2, es fuertemente convexa si para cualesquiera $x, y \in S$ y algún $\theta > 0$

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{\theta}{2}\|x - y\|^2.$$

Observación. Trabajamos con $\|\cdot\|_2$.

Observación. Intuitivamente, la convexidad fuerte nos dice que existe una cota inferior cuadrática. Esto implica directamente la convexidad de la función. Notemos además que esto implica directamente la convexidad de la función (4).

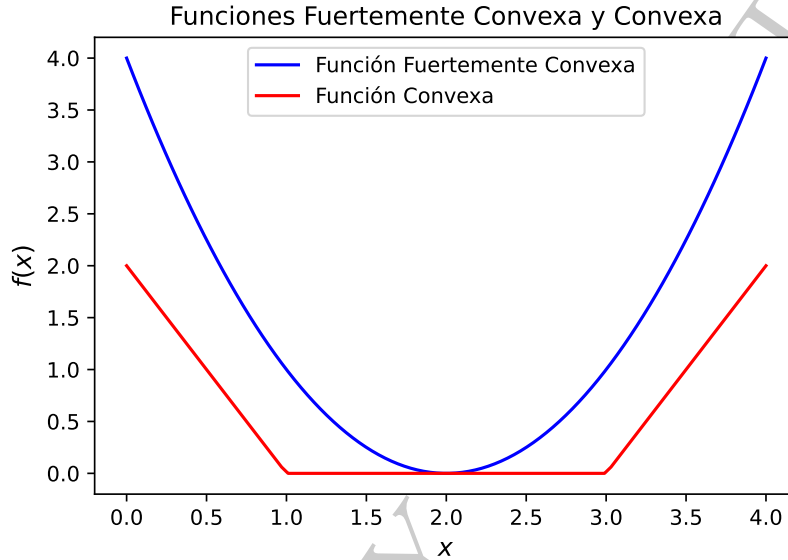


Figura C.1 Fuertemente convexa y convexa.

Lema C.0.5. Las siguientes condiciones son equivalentes para una función diferenciable f fuertemente convexa, con $\theta > 0$

- a) $f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{\theta}{2}\|y - x\|^2$.
- b) $g(x) = f(x) - \frac{\theta}{2}\|x\|^2$ es convexa.
- c) $(\nabla f(x) - \nabla f(y))^T(x - y) \geq \theta\|x - y\|^2$.
- d) $f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) - \frac{\alpha(1 - \alpha)\theta}{2}\|x - y\|^2$, $\alpha \in (0, 1)$.

Demostración. Vamos a probar $(a) \Leftrightarrow (b)$, $(b) \Leftrightarrow (c)$ y $(b) \Leftrightarrow (d)$.

1. (a) \Leftrightarrow (b) Si g es convexa,

$$g(y) \geq g(x) + \nabla g(x)^T(y - x).$$

Luego

$$f(y) - \frac{\theta}{2}\|y\|^2 \geq f(x) - \frac{\theta}{2}\|x\|^2 + \nabla^T f(x)(y - x) - \theta x(y - x).$$

Esto pues, $\frac{\theta}{2}\|x\|^2 = \frac{\theta}{2} \sum_{i=1}^n x_i^2$. Luego, agrupando términos,

$$\begin{aligned} f(y) &\geq f(x) + \frac{\theta}{2}(\|y\|^2 - \|x\|^2) + \nabla^T f(x)(y - x) - \theta x(y - x) \\ f(x) &+ \frac{\theta}{2}(\|y\|^2 - \|x\|^2) + \nabla^T f(x)(y - x) + \theta\|x\|^2 - \theta xy \\ &= f(x) + \frac{\theta}{2}(\|y\|^2 + \|x\|^2) + \nabla^T f(x)(y - x) - \theta xy \\ &= f(x) + \nabla^T f(x)(y - x) + \underbrace{\left[\frac{\theta}{2}(\|y\|^2 + \|x\|^2) - \theta xy \right]}_{= \frac{\theta}{2}\|x - y\|^2} \\ &= f(x) + \nabla^T f(x)(y - x) + \frac{\theta}{2}\|x - y\|^2. \end{aligned}$$

2. (b) \Leftrightarrow (c) Como g es convexa diferenciable, por el corolario C.0.3

$$(\nabla f(x) - \theta x - \nabla f(y) + \theta y)^T(x - y) \geq 0.$$

Entonces,

$$(\nabla f(x) - \nabla f(y))^T(x - y) + \theta(y - x)^T(x - y) \geq 0.$$

Así,

$$(\nabla f(x) - \nabla f(y))^T(x - y) \geq \underbrace{\theta(x - y)^T(x - y)}_{= \|x - y\|^2}.$$

3. (b) \Leftrightarrow (d) De la convexidad de g

$$\begin{aligned} g(\alpha x + (1 - \alpha)y) &\leq f(\alpha x + (1 - \alpha)y) - \frac{\theta}{2}\|\alpha x + (1 - \alpha)y\|^2 \\ &\leq \alpha f(x) - \frac{\alpha\theta}{2}\|x\|^2 + (1 - \alpha)f(y) - \frac{(1 - \alpha)\theta}{2}\|y\|^2. \end{aligned}$$

Ahora bien, estas desigualdades implican que

$$f(\alpha x + (1 - \alpha)y) \leq \frac{\theta}{2}\|\alpha x + (1 - \alpha)y\|^2 + \alpha f(x) - \frac{\alpha\theta}{2}\|x\|^2 + (1 - \alpha)f(y) - \frac{(1 - \alpha)\theta}{2}\|y\|^2.$$

Analicemos el término

$$\frac{\theta}{2} \|\alpha x + (1 - \alpha)y\|^2 - \frac{\alpha\theta}{2} \|x\|^2 - \frac{(1 - \alpha)\theta}{2} \|y\|^2.$$

Por la convexidad de la norma

$$\frac{\theta}{2} (\alpha \|x\| + (1 - \alpha)\|y\|)^2 - \frac{\alpha\theta}{2} \|x\|^2 - \frac{(1 - \alpha)\theta}{2} \|y\|^2.$$

Desarrollando,

$$\frac{\theta}{2} (\alpha^2 \|x\|^2 + 2\alpha(1 - \alpha)\|x\| \cdot \|y\| + (1 - \alpha)^2 \|y\|^2) - \frac{\alpha\theta}{2} \|x\|^2 - \frac{(1 - \alpha)\theta}{2} \|y\|^2.$$

Factorizamos todo por $\theta/2$

$$\frac{\theta}{2} [\alpha(\alpha - 1)(\|x\|^2 + \|y\|^2 - 2\|x\| \cdot \|y\|)] \leq \frac{\theta}{2} \alpha(\alpha - 1) \|x - y\|^2.$$

Con esto concluimos.

□

La teoría del análisis convexo es bastante extensa y útil en economía. No ahondamos más en este tema pero recomendamos al lector consultar [Ok \(2007\)](#) o [Florenzano and le Van \(2001\)](#) para profundizar en estos temas. Solo hemos presentado aquello estrictamente necesario, es decir, aquello a lo que se ha aludido en el texto, o está directamente vinculado. Continuamos con el apéndice de análisis funcional que, en cierta forma, es un preámbulo al último apéndice. Este último se encuentra fuertemente vinculado con los desarrollos del capítulo 3 y presenta resultados no tan estándares en la literatura.

Apéndice D

Elementos de análisis funcional

A continuación algunos elementos de análisis funcional, necesarios para poder llegar a la 1era forma geométrica del Teorema de Hahn-Banach, de la cual se hace uso indirectamente en el capítulo 3 (permite probar un resultado sobre el subdiferencial de una función convexa: esto es abordado en nuestro último anexo). Los siguientes enunciados y pruebas vienen fundamentalmente de [Botelho et al. \(2023\)](#) y las notas de clase del profesor de la especialidad de matemáticas en la PUCP, Percy Fernandez. No se han incorporado modificaciones sustanciales y recomendamos consultar [Botelho et al. \(2023\)](#) para, por ejemplo, revisar conceptos como la norma de un operador lineal o continuidad de un operador lineal.

Definición D.0.1. Sea V un espacio vectorial no nulo. Un hiperplano V es un subespacio $W \neq V$ tal que, si W_1 es subespacio de V , $W \subset W_1$, entonces $W_1 = V$ o $W = W_1$.

Proposición D.0.2. Sea $V \neq \{0\}$ espacio vectorial, W subespacio de V . Entonces, W es un hiperplano de V si y solo si existe $\varphi : V \rightarrow \mathbb{R}$ funcional lineal tal que $W = \text{Ker}(\varphi)$.

Demostración. \Rightarrow Sea $W \subset V$ hiperplano y sea $v_0 \in V - W$ y $W_1 = W + [v_0] \neq W$. Luego, $V = W_1$. Así, tenemos

$$\begin{aligned}\varphi : W_1 &\rightarrow \mathbb{R} \\ w + av_0 &\rightarrow a\end{aligned}$$

que es lineal, $\varphi \neq 0$ y $\text{Ker}(\varphi) = W$. \Leftarrow Ahora, sea $\varphi : V \rightarrow \mathbb{R}$ funcional lineal no nulo y $W = \text{Ker}(\varphi)$. Sea W_1 subespacio de V tal que $W \subsetneq W_1$ pero $W_1 \neq V$ y sea

$v_0 \in W_1 - W$. Dado $v \in V$, considere

$$u = v - \frac{\varphi(v)}{\varphi(v_0)}v_0 \in \text{Ker}(\varphi) = W.$$

De este modo, $v \in W_1$ pues $W \subset W_1$. Así, $W_1 = V$. □

Definición D.0.3. Si H es un hiperplano de V y $v_0 \in V$, el conjunto

$$H + v_0 = \{v + v_0 : v \in H\}$$

es llamado hiperplano afín. Por la proposición (D.0.2), los hiperplanos afines son de la forma

$$[\varphi = a] = \{v \in V : \varphi(v) = a\}$$

donde φ es un funcional lineal.

Proposición D.0.4. Un hiperplano $[\varphi = a]$ es cerrado si y solamente si φ es continuo.

Demostración. \Leftarrow Como φ es continua, entonces $[\varphi = a] = \varphi^{-1}\{a\}$ es cerrado. \Rightarrow Sea $\varphi : V \rightarrow \mathbb{R}$ funcional lineal no nulo y $H = [\varphi = a]$ cerrado. Como $V - H$ es abierto y no vacío, si $x_0 \in V - H$, existe $r > 0$ tal que $B(x_0, r) \subset V - H$. Tenemos que $\varphi(x_0) \neq a$. S.p.d.g. digamos que $\varphi(x_0) < a$. Probaremos que $\varphi(x) < a$ para todo $x \in B(x_0, r)$. Supongamos que no, y que existe x_1 tal que $\varphi(x_1) > a$ con $x_1 \in B(x_0, r)$. Tomemos

$$t = \frac{\varphi(x_1) - a}{\varphi(x_1) - \varphi(x_0)} < 1.$$

Luego,

$$tx_0 + (1 - t)x_1 \in B(x_0, r)$$

y

$$\begin{aligned} \varphi(tx_0 + (1 - t)x_1) &= t\varphi(x_0) + (1 - t)\varphi(x_1) \\ &= \varphi(x_1) + t(\varphi(x_0) - \varphi(x_1)) \\ &= \varphi(x_1) - \varphi(x_1) + a = a, \end{aligned}$$

lo cual es una contradicción: $B(x_0, r) \subset V - H$. De ahí,

$$\varphi(x_0) + r\varphi(z) = \varphi(x_0 + rz) < a, \quad \forall z \in V, \|z\| < 1.$$

O sea,

$$\frac{\varphi(x_0) - a}{r} < \varphi(z) < \frac{a - \varphi(x_0)}{r}.$$

Haciendo $z \rightarrow -z$ ($\| -z \| < 1$)

$$\|\varphi\| < \frac{a - \varphi(x_0)}{r},$$

y así, φ es continua. □

Definición D.0.5. Funcional de Minkowski. Sea C un subconjunto convexo y abierto con $0 \in C$, de un espacio normado. La aplicación

$$P_C : E \rightarrow \mathbb{R}, \quad x \rightarrow \inf \left\{ a > 0 : \frac{x}{a} \in C \right\}$$

es llamada funcional de Minkowski de C .

Proposición D.0.6. Tenemos las siguientes propiedades:

1. $P_C(bx) = bP_C(x)$, $\forall b > 0$ y $x \in E$.
2. $C = \{x \in E : P_C(x) < 1\}$.
3. Existe $M > 0$ tal que $0 \leq P_C(x) \leq M\|x\|$, $\forall x \in E$.
4. $P_C(x + y) \leq P_C(x) + P_C(y)$, $\forall x, y \in E$.

Demostración. Procedemos inciso por inciso:

1. Por definición,

$$\begin{aligned} P_C(bx) &= \inf \left\{ a > 0 : \frac{bx}{a} \in C \right\} \\ &= \inf \left\{ ba > 0 : \frac{x}{a} \in C \right\} \\ &= bP_C(x). \end{aligned}$$

2. Como C es abierto, para todo $x \in C$, $\exists \varepsilon > 0$ tal que $(1 + \varepsilon)x \in B[x, \varepsilon\|x\|] \subset C$ (bola cerrada). Luego, $x(1 + \varepsilon) \in C$. Por ende $P_C(x) \leq (1 + \varepsilon)^{-1} < 1$. Recíprocamente, si $P_C(x) < 1$, entonces existe $0 < a < 1$ tal que $\frac{x}{a} \in C$:

$$x = a \left(\frac{x}{a} \right) + (1 - a)0 \in C.$$

3. Sea $r > 0$ tal que $B(0, r) \subset C$. Para $0 < s < r$ tenemos

$$s \left(\frac{x}{\|x\|} \right) \in C, \forall x \in E - \{0\}.$$

Luego,

$$\left(\frac{s}{\|x\|} \right) P_C(x) = P_C \left(\frac{s}{\|x\|} x \right) < 1.$$

Entonces, $P_C(x) < (1/s)\|x\|$, $1/s = M$.

4. Dado $\varepsilon > 0$ veamos que $\frac{x}{P_C(x) + \varepsilon} \in C$. Por el inciso (a)

$$P_C \left(\frac{x}{P_C(x) + \varepsilon} \right) = \left(\frac{1}{P_C(x) + \varepsilon} \right) P_C(x) < 1.$$

Por ende, de (b) se tiene que $\frac{x}{P_C(x) + \varepsilon} \in C$. Análogamente $\frac{y}{P_C(y) + \varepsilon} \in C$. Sea

$$0 < t = \frac{P_C(x) + \varepsilon}{P_C(x) + P_C(y) + \varepsilon} < 1.$$

Tenemos

$$\frac{x + y}{P_C(x) + P_C(y) + 2\varepsilon} = t \left(\frac{x}{P_C(x) + \varepsilon} \right) + (1 - t) \left(\frac{y}{P_C(y) + \varepsilon} \right) \in C.$$

Luego,

$$\left(\frac{1}{P_C(x) + P_C(y) + 2\varepsilon} \right) P_C(x + y) = P_C \left(\frac{x + y}{P_C(x) + P_C(y) + 2\varepsilon} \right) < 1.$$

Así

$$P_C(x + y) < P_C(x) + P_C(y) + 2\varepsilon, \forall \varepsilon > 0.$$

Por lo tanto,

$$P_C(x + y) \leq P_C(x) + P_C(y), \forall x, y \in E.$$

□

Lema D.0.7. Sea C convexo, abierto, propio y no vacío de un espacio normado E y $x_0 \in E - C$. Entonces, $\exists \varphi \in E'$ tal que $\varphi(x) < \varphi(x_0)$, $\forall x \in C$.

Demostración. Supongamos que $0 \notin C$, sea $z_0 \in C$ y considere $0 \in D = C - z_0$ y $y_0 = x_0 - z_0$. Tenemos que $y_0 \notin D$ (pues $x_0 \notin C$), es convexo, no vacío y abierto. Podemos entonces considerar el funcional de Minkowski P_D . Definimos $G = [y_0]$ y

$$g : G \rightarrow \mathbb{R}$$

$$ty_0 \rightarrow tP_D(y_0)$$

es un funcional lineal. Como

$$\forall t > 0 : g(ty_0) = tP_D(y_0) = P_D(ty_0)$$

$$t \leq 0 : g(ty_0) = tP_D(y_0) \leq 0 \leq P_D(ty_0)$$

tenemos

$$g(x) \leq P_D(x), \forall x \in G.$$

Por el Teorema de Hahn-Banach (versión base) existe $\varphi : E \rightarrow \mathbb{R}$ $\varphi|_G = g$ y $\varphi(x) \leq P_D(x)$, $\forall x \in E$. Luego, sabemos que existe $M > 0$ tal que $\varphi(x) \leq P_D(x) \leq M\|x\|$. Esto es, φ es continua. Para todo $x \in D$ se tiene $\varphi(x) \leq P_D(x) < 1$. Como $P_D(y_0) \geq 1$ luego

$$\varphi(y) < 1 \leq P_D(y_0) = g(y_0) = \varphi(y_0) = \varphi(x_0 - z_0), \forall y \in D = C - z_0$$

$$\varphi(x) = \varphi(x - z_0) + \varphi(z_0) < \varphi(x_0 - z_0) + \varphi(z_0) = \varphi(x_0), \forall x \in C.$$

En caso $0 \in C$, tomamos $z_0 = 0$. □

Lema D.0.8. Sean E un espacio normado, $\varphi \in E' - \{0\}$ continuo, y $C \subset E$ convexo, abierto y no vacío. Entonces, $\varphi(C)$ es un intervalo abierto.

Demostración. Como $\varphi \neq 0$ entonces φ es sobreyectiva y $\varphi(C)$ es un intervalo (pues debe ser conexo al ser C convexo, i.e., en particular conexo). Suponga que $\varphi(C)$ esté limitado superiormente por $a \in \varphi(C)$. Sea entonces $x \in C$ tal que $\varphi(x) = a$ y $\varphi(y) \leq a$ para todo $y \in C$. Como C es abierto, $\exists \varepsilon > 0$ tal que $B(x, \varepsilon) \subset C$. Sea $z \in E - \{0\}$. Entonces

$$x + \frac{\varepsilon}{2\|z\|}z \in B(x, \varepsilon) \subset C.$$

Luego,

$$a \geq \varphi\left(x + \frac{\varepsilon}{2\|z\|}z\right) = \varphi(x) + \frac{\varepsilon}{2\|z\|}\varphi(z) = a + \frac{\varepsilon}{2\|z\|}\varphi(z)$$

por lo que $\varphi(z) \leq 0$, $\forall z \in E$ (contradicción pues φ es sobreyectiva). □

Teorema D.0.9. Primera forma geométrica del Teorema de Hahn-Banach.

Sean A y B subconjuntos convexos, no vacíos y disjuntos del espacio normado E . Si A es abierto entonces existen un funcional $\varphi \in E'$ y $a \in \mathbb{R}$ tal que

$$\varphi(x) < a \leq \varphi(y), \forall x \in A, y \in B$$

se dice que el hiperplano cerrado $[\varphi = a]$ separa A y B .

Demostración. Consideremos el conjunto $C = A - B$:

1. C es abierto pues $C = \bigcup_{b \in B} (A - \{b\})$.
2. C es convexo: para todo $x_1, x_2 \in A$ y $y_1, y_2 \in B$ y $0 < t < 1$

$$t(x_1 - y_1) + (1 - t)(x_2 - y_2) = (tx_1 + (1 - t)x_2) - (ty_1 + (1 - t)y_2) \in A - B$$

3. $\emptyset \neq C$, $0 \notin C$ pues $A \cap B = \emptyset$, luego $C \neq E$.

Por el lema (D.0.7) existe $\varphi \in E'$ tal que

$$\varphi(z) < \varphi(0) = 0, \forall z \in C.$$

Luego,

$$\varphi(x) + \varphi(x - y) + \varphi(y) < \varphi(y), \forall x \in A, \forall y \in B \quad (\text{D.1})$$

y como $B \neq \emptyset$, de esta desigualdad $\varphi(A)$ es acotado y así $a = \sup \varphi(A) \in \mathbb{R}$. Por ende, $\varphi(A) \subset (-\infty, a]$. De (D.1)

$$a \leq \varphi(y), \forall y \in B.$$

Finalmente, por el lema (D.0.8)

$$\varphi(x) < a, \forall x \in A.$$

□

Antes de concluir con esta sección, es necesario hacer mención del siguiente Teorema pues, es usado en la prueba de la proposición E.0.6. En nuestro contexto, nos hemos limitado a $H = \mathbb{R}^d$.

Teorema D.0.10. Teorema de Riesz-Fréchet. Sea H un espacio de Hilbert y $\varphi : H \rightarrow \mathbb{K}$ un funcional lineal continuo. Entonces, existe un único $y_0 \in H$ tal que

$$\varphi(x) = \langle x, y_0 \rangle, \forall x \in H.$$

Además de ello, $\|\varphi\| = \|y_0\|$.

Antes de pasar al último apéndice, unas breves definiciones que, implícitamente se usaron en los preliminares de la prueba del teorema 3.3.2.

Definición D.0.11. Sean X, Y espacios de Banach normados y $f : X \rightarrow Y$. Diremos que f es Fréchet diferenciable en $x_0 \in X$ si existe una transformación lineal $T : X \rightarrow Y$ tal que

$$\lim_{h \rightarrow 0} \frac{\|f(x_0 + h) - f(x_0) - T(h)\|_Y}{\|h\|_X} = 0.$$

En dicho caso, $f'(x_0)$ es la derivada de Fréchet en x_0 .

Corolario D.0.12. Si X es un espacio de Hilbert sobre \mathbb{R} y $f : X \rightarrow \mathbb{R}$, entonces, si la derivada de Fréchet en $x \in X$ existe, es un funcional lineal continuo. Por el teorema D.0.10, existe $u \in X$ tal que $f'(x)(h) = \langle h, u \rangle$ para todo $h \in X$. Dicho vector u se conoce como el vector gradiente $\nabla f(x)$. De este modo,

$$f'(x)(h) = \langle h, \nabla f(x) \rangle, \quad \forall h \in X.$$

La teoría expuesta en esta sección es de gran importancia en las aplicaciones. Para el caso $V = \mathbb{R}^n$ podemos prescindir de estos desarrollos y aludir a resultados geométricos (la separación de conjuntos convexos en \mathbb{R}^n es un resultado que es consecuencia directa del Teorema de Proyección para conjuntos cerrados y aspectos topológicos de los conjuntos convexos). Sin embargo, para abordar el caso más general posible, en el que el espacio vectorial considerado tiene dimensión infinita, este apéndice y la teoría del análisis funcional resultan fundamentales. Estos resultados no solo son aplicables al transporte óptimo, sino que también tienen implicaciones en el equilibrio general (Aliprantis et al., 1990) y la teoría de juegos (Fudenberg and Tirole, 1991). Terminamos este trabajo con el anexo sobre los subgradiientes y subdiferenciales.

Apéndice E

Subgradientes y subdiferencial

Dada $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$ convexa y diferenciable, se cumple que

$$f(y) \geq f(x) + \nabla f(x)^T(y - x). \quad (\text{E.1})$$

¿Y si f no es diferenciable en x ? A continuación estudiamos la teoría de los subdiferenciales que, de cierta forma, permiten extender y generalizar (E.1).

Definición E.0.1. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$. El epigrafo de f es el conjunto

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

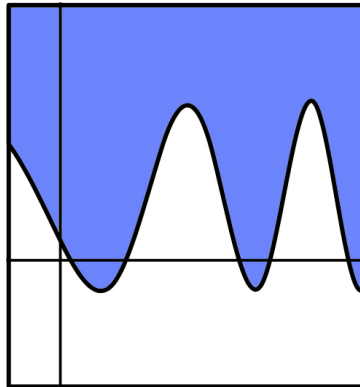


Figura E.1 Epigrafo.

Definición E.0.2. Diremos que z soporta el epígrafo de f en $(x_0, f(x_0))$ si, $\forall (y, t) \in \mathcal{E}(f)$

$$z^T \begin{bmatrix} \underbrace{p}_{(y,t)} - \underbrace{p_0}_{(x_0, f(x_0))} \end{bmatrix} \leq 0.$$

Corolario E.0.3. Si f es convexa y diferenciable en x_0 , el vector $(\nabla f(x_0), -1)$ soporta el epigrafo de f en $(x_0, f(x_0))$.

Demostración. Simplemente verificamos que, dado $(y, t) \in \mathcal{E}(f)$,

$$\begin{bmatrix} \nabla f(x_0) \\ -1 \end{bmatrix}^T \begin{pmatrix} y - x_0 \\ t - f(x_0) \end{pmatrix} = \nabla f(x_0)(y - x_0) - t + f(x_0) \leq 0.$$

En efecto, como $(y, t) \in \mathcal{E}(f)$, inmediatamente $f(y) \leq t$ y, de (E.1) se tiene la otra desigualdad. \square

Definición E.0.4. $g \in \mathbb{R}^n$ es un subgradiente de $f: \mathbb{R}^n \rightarrow \mathbb{R}$ en x si

$$f(y) \geq f(x) + g^T(y - x), \quad \forall y.$$

Observación. Por definición:

1. $(g, -1)$ soporta el epígrafo de f en $(x, f(x))$.
2. Dada f convexa, $\nabla f(x)$ es subgradiente (si es que f es diferenciable en x) en x .

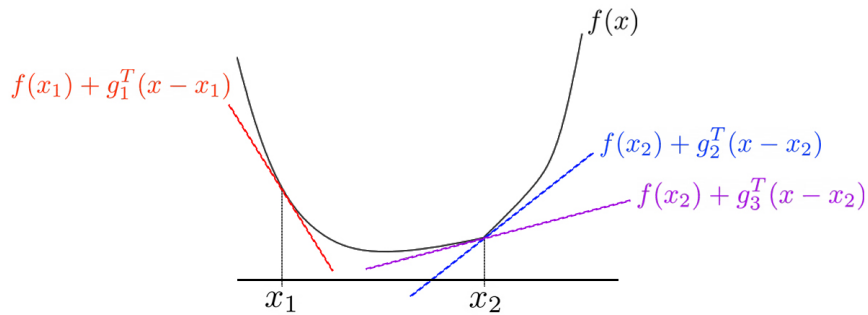


Figura E.2 Sub-gradients.

Definición E.0.5. El subdiferencial de $f : \mathbb{R}^d \rightarrow \mathbb{R}$ en x , denotado $\partial f(x)$ es el conjunto de subgradiientes en x . Esto es,

$$\partial f(x) = \{y \in \mathbb{R}^d : f(z) \geq f(x) + \langle y, z - x \rangle, \forall z \in \mathbb{R}^d\}.$$

Proposición E.0.6. El subdiferencial en un punto interior del dominio de finitud de f es no vacío cuando f es convexa.

Demostración. Sea $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ una función convexa. Sea $x_0 \in D^\circ$, donde D es el conjunto de finitud de la función f . Definamos el abierto y convexo (pues f lo es)

$$A = \{(x, y) \in D^\circ \times \mathbb{R} : y > f(x)\}.$$

Dado que $(x_0, f(x_0)) \notin A$, por los teoremas D.0.9 y D.0.10 existe un hiperplano que separa A de $(x_0, f(x_0))$. Ahora, veamos que no es de la forma

$$\{(x, y) \in \mathbb{R}^d \times \mathbb{R} : \langle w, x \rangle_{\mathbb{R}^d} = c\}.$$

O sea, que no es vertical. Acá $c \in \mathbb{R}$ y $w \in \mathbb{R}^d - \{0\}$ (fijo). Supongamos por contradicción que lo es. En dicho caso, como D° es un abierto, existe $r > 0$ tal que $B(x_0, r) \subset D^\circ$. Entonces, existen $x_1, x_2 \in B(x_0, r)$ tales que $\langle x_1, w \rangle < c < \langle x_2, w \rangle$. Pero entonces, $(x_1, f(x_1) + 1)$ y $(x_2, f(x_2) + 1)$ son separados por el hiperplano y ambos pertenecen a A . De este modo, el hiperplano es de la forma

$$\{(x, y) \in \mathbb{R}^d \times \mathbb{R} : \langle w, x \rangle_{\mathbb{R}^d} + y = c\}.$$

Luego, como $(x_0, f(x_0)) \in \bar{A}$,

$$\langle w, x_0 \rangle + f(x_0) = c.$$

Notando que $\langle w, x \rangle + y \geq c$ en A , para todo $x \in D^\circ$

$$\langle w, x \rangle + f(x) + \varepsilon \geq c, \forall \varepsilon > 0 \implies \langle w, x \rangle + f(x) \geq c.$$

De este modo,

$$f(x) \geq f(x_0) = \langle w, x - x_0 \rangle \implies w \in \partial f(x_0).$$

Por ende $\partial f(x_0)$ es no vacío. □

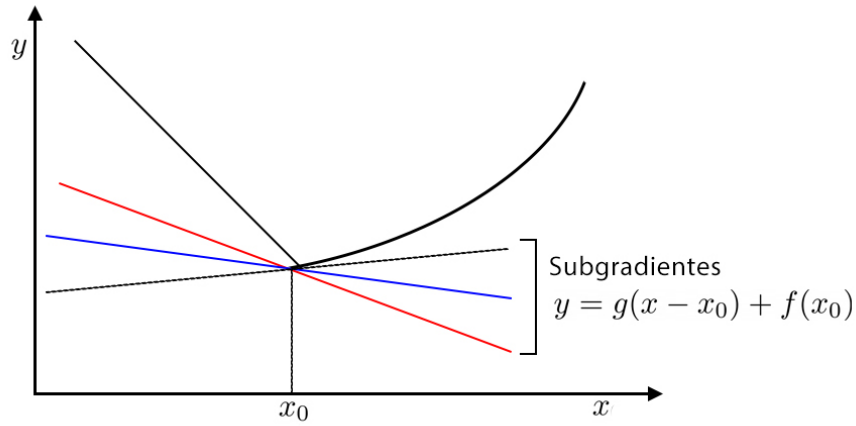


Figura E.3 Sub-gradiientes.

Proposición E.0.7. El subdiferencial de una función convexa f en x en un conjunto convexo y cerrado.

Demostración. Sean $y, z \in \partial f(x)$ y $\theta \in [0, 1]$. Entonces, para todo $x' \in \mathbb{R}^d$ tenemos

$$\begin{aligned} f(x') &= (1 - \theta)f(x') + \theta f(x') \geq (1 - \theta)(f(x) + \langle y, x' - x \rangle) + \theta(f(x) + \langle z, x' - x \rangle) \\ &= f(x) + \langle (1 - \theta)y + \theta z, x' - x \rangle. \end{aligned}$$

Así, $\partial f(x)$ es convexo. Veamos ahora que se trata de un cerrado:

$$\partial f(x) = \underbrace{\bigcap_{x' \in \mathbb{R}^d} \underbrace{\{y : f(x') \geq f(x) + \langle y, x' - x \rangle\}}_{\text{cerrado}}}_{\text{cerrado}}.$$

□

Proposición E.0.8. Si $f : X \rightarrow \mathbb{R}$ es convexa y C^1 , entonces $\partial f(x) = \{\nabla f(x)\}$.

Demostración. Ciertamente, por el teorema C.1, $\nabla f(x) \in \partial f(x)$. Ahora, supongamos que existe otro elemento, digamos $y \in \partial f(x)$. Por definición $f(x') \geq f(x) + \langle y, x' - x \rangle$, $x' \in X$. Tomamos $x' = x + tz$ para $z \in X$ y $t > 0$. Entonces,

$$\frac{f(x + tz) - f(x)}{t} \geq \langle y, z \rangle.$$

Haciendo $t \rightarrow 0$,

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

Pero entonces, $\langle \nabla f(x) - y, z \rangle \geq 0$. Al ser z arbitrario, tomamos

$$z = -(\nabla f(x) - y).$$

Esto nos permite obtener

$$\|\nabla f(x) - y\| \leq 0 \implies y = \nabla f(x).$$

□

Ejemplo E.0.9. Consideremos $f(x) = \|x\|_2$. Entonces,

$$\partial f(x) = \begin{cases} \left\{ \left(\frac{x}{\sqrt{x^2 + y^2}}, \frac{y}{\sqrt{x^2 + y^2}} \right) \right\}, & \text{si } (x, y) \neq (0, 0) \\ B_{\|\cdot\|_2}(0, 1), & \text{si } (x, y) = (0, 0). \end{cases}$$

Esto es consecuencia del lema E.0.8 y de lo siguiente. Si $g \in \partial f((0, 0))$, entonces,

$$\langle g, z \rangle \leq \|z\|_2, \quad z = (x, y).$$

Luego, si $\|g\| \leq 1$, por Cauchy-Schwarz

$$\langle g, z \rangle \leq \|g\|_2 \|z\|_2 \leq \|z\|_2.$$

O sea, g es subgradiente si $\|g\|_2 \leq 1$. Ahora, si g es un subgradiente, $\langle g, z \rangle \leq \|z\|_2$.

Esto debe cumplirse para $\|y\| \leq 1$. Entonces,

$$\sup_{y: \|y\| \leq 1} \langle g, y \rangle \leq \|y\| \implies \|g\|_2 \leq 1.$$

Note que se puede generalizar este resultado a \mathbb{R}^n .

Proposición E.0.10. x^* minimiza f si y solamente si $0 \in \partial f(x^*)$.

Demostración. Tenemos:

$$\begin{aligned}
 f(x^*) &= \min_x f(x) \\
 &\Leftrightarrow \\
 f(x^*) &\leq f(y), \forall y \\
 &\Leftrightarrow \\
 f(y) &\geq f(x^*) + \mathbf{0}^T(y - x^*), \forall y \\
 &\Leftrightarrow \\
 \mathbf{0} &\in \partial f(x^*).
 \end{aligned}$$

□

Lema E.0.11. Sea $f(x) = \max\{f_1(x), f_2(x)\}$, con f_1, f_2 diferenciables. Entonces,

$$\partial f(x) = \begin{cases} \nabla f_1(x), & \text{si } f_1(x) > f_2(x) \\ \nabla f_2(x), & \text{si } f_2(x) > f_1(x) \\ [\nabla_1 f(x), \nabla_2 f(x)] & \text{si } f_1(x) = f_2(x). \end{cases}$$

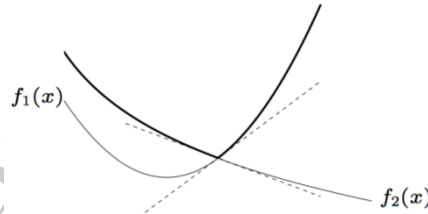


Figura E.4 Sub-diferencial $\max\{f_1, f_2\}$.

Definición E.0.12. La función prox. La función proximidad de una función convexa $h(\cdot)$ es, para $t > 0$

$$\text{prox}_{h,t}(x) = \underset{u}{\text{argmin}} \left(h(u) + \frac{1}{2t} \|u - x\|_2^2 \right). \quad (\text{E.2})$$

Observación. Un resultado crucial del cuál se ha hecho uso es el Teorema de Moreau-Rockafellar [Lain \(1988\)](#). Este resultado nos dice que, si $f, g : \mathbb{R}^n \rightarrow (-\infty, \infty]$ son

funciones convexas tales que la intersección de sus dominios de finitud Ω es no vacía, entonces para todo $x_0 \in \Omega$

$$\partial f(x_0) + \partial g(x_0) = \partial(f + g)(x_0).$$

Ejemplo E.0.13. Si $h(x) = \lambda \|x\|_1$, $t = 1$

$$\text{prox}_h(x)_i = \begin{cases} x_i - \lambda, & \text{si } x_i \geq \lambda \\ 0, & \text{si } |x_i| \leq \lambda \\ x_i + \lambda, & \text{si } x_i \leq -\lambda. \end{cases} \quad (\text{E.3})$$

En efecto, $|x| = \max\{x, -x\}$. Así pues, de (E.2), observando que el problema es separable, basta tener

$$0 \in \partial f(u_i^*).$$

Luego, por Moreau-Rockafellar,

$$\partial f(u_i^*) = \{u_i^* - x_i + \lambda \partial |\cdot|(u_i^*)\}.$$

Finalmente, dado que

$$\partial |\cdot|(u_i^*) = \begin{cases} -1, & \text{si } u_i < 0 \\ [-1, 1], & \text{si } u_i = 0 \\ 1 & \text{si } u_i > 0 \end{cases}$$

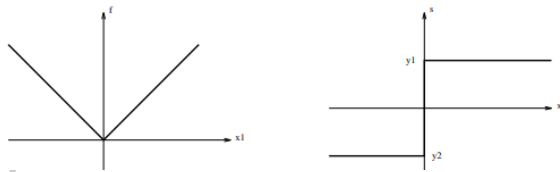


Figura E.5 Sub-diferencial $|\cdot|$.

concluimos que

$$u_i^* = \begin{cases} x_i - \lambda, & \text{si } x_i \geq \lambda \\ 0, & \text{si } |x_i| \leq \lambda \\ x_i + \lambda, & \text{si } x_i \leq -\lambda. \end{cases}$$

Ahora bien, en el caso de tener

$$f(x) = g(x) + h(x) \quad (\text{E.4})$$

con g convexa y diferenciable sobre su dominio y h convexa pero eventualmente no diferenciable, se ejecuta el siguiente algoritmo con la finalidad de minimizar (E.4):

$$x^{(k)} = \text{prox}_{t_k, h}(x^{(k-1)} - t_k \nabla g(x^{(k-1)})).$$

¿Cuál es la intuición?

$$\begin{aligned} x^{(k)} &= \underset{u}{\operatorname{argmin}} \left(h(u) + \frac{1}{2t_k} \|u - x^{(k-1)} - t_k \nabla g(x^{(k-1)})\|_2^2 \right) \\ &= \underset{u}{\operatorname{argmin}} \left(h(u) + g(x) + \nabla g(x)^T (u - x) + \frac{1}{2t} \|u - x\|_2^2 \right). \end{aligned}$$

Lo que minimizamos es $h(u)$ más la diferencia de u y el clásico gradient-descent.

Observación. La definición 3.3.1 es entonces en realidad consecuencia de (E.2) y (E.3).

Bibliografía

- Abbott, S. (2015). *Understanding Analysis*. Springer.
- Agarwal, N. and Somaini, P. (2019). Revealed preference analysis of school choice models. *NBER Working Paper*, (w26505).
- Alba-Vivar, F. M. (2025). Opportunity bound: Transport and access to college in a megacity. *Job Market Paper*. Department of Economics - Wake Forest University.
- Aliprantis, C. and Border, K. (2006). *Infinite dimensional analysis*. Springer Verlag.
- Aliprantis, C. D., Brown, D. J., and Burkinshaw, O. (1990). *Existence and Optimality of Competitive Equilibria*. Economic Theory, Econometrics, and Mathematical Economics. Springer.
- Ambrosio, L., Brué, E., and Semola, D. (2021). *Lectures on Optimal Transport*. Springer Verlag.
- Angrist, J., Gray-Lobe, G., Idoux, C. M., and Pathak, P. A. (2022). Still worth the trip? school busing effect in boston and new york. *NBER*.
- Axler, S. (2014). *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Springer International Publishing.
- Azevedo, E. M. and Leshno, J. D. (2019). A supply and demand framework for two-sided matching markets. *Journal of Political Economy*, 124(5).
- Becker, G. (1973). A theory of marriage: Part i. *Journal of Political Economy*, 81(4):813–846.

- Beiglböck, M., Goldstern, M., Maresch, G., and Schachermayer, W. (2009). Optimal and better transport plans. *arXiv preprint*.
- Billingsley, P. (1995). *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, 3rd edition.
- Birkhoff, G. (1957). Extensions of jentzsch theorem. *American Mathematical Society*, 85(1):219–227.
- Botelho, G., Pellegrino, D., and Teixeira, E. (2023). *Fundamentos de Analise Funcional*. SBM.
- Boyd, S. (2022). Subgradients. *Lectures Notes*.
- Burga, C. and Moreno, M. (2001). ¿existe subempleo profesional en el Perú urbano? *Investigaciones Breves* 17.
- Chambers, C. P., Echenique, F., and Lambert, N. S. (2021). Recovering preferences from finite data. *Econometrica*, 89(4):1633–1664.
- Chambers, C. P., Echenique, F., and Lambert, N. S. (2023). Recovering utility. *arXiv preprint arXiv:2301.11492*.
- Champion, T., de Pascale, L., and Juutinen, P. (2008). The infinity wasserstein distance: Local solutions and existence of optimal transport maps. *SIAM Journal on Mathematical Analysis*, 40(1):1–20.
- Choi, Y. (2020). On two-sided matching in infinite markets. *Lecture Notes*.
- Clason, C., Lorenz, D. A., Mahler, H., and Wirth, B. (2021). Entropic regularization of continuous optimal transport problems. *Journal of Mathematical Analysis and Applications*, 494(2):124432.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer.
- Dupuy, A., Carlier, G., Galichon, A., and Sun, Y. (2021). Sista learning optimal transport costs under sparsity constraints. *Discussion Paper Series*.

- Dupuy, A. and Galichon, A. (2014). Personality traits and the marriage market. *Journal of Political Economy*, 122(6):1271–1319.
- Dupuy, A. and Galichon, A. (2022). A note on the estimation of job amenities and labor productivity. *Quantitative economics*, 13(1):157–177.
- Dupuy, A., Galichon, A., and Sun, Y. (2017). Estimating matching affinity matrix under low-rank constraints. *Information and Inference: A Journal of the IMA*, 8(4):677–689.
- Echenique, F. and Chambers, C. (2016). *Revealed Preference Theory*. Cambridge University Press.
- Echenique, F., Immorlica, N., and Vazirani, V. (2023). *Online and Matching-Based Market Design*. Cambridge University Press.
- Even, M. and Merad, I. (2020). Optimal transport and stochastic optimization. *ENS*.
- Federico Echenique, Fedor Sandomirskiy and Joseph Root (2024). Stable matching as transportation. *arXiv preprint arXiv:2402.13378*.
- Ferdowsian, A., Niederle, M., and Yariv, L. (2023). Strategic decentralized matching: The effects of information frictions. *Working Paper*.
- Florenzano, M. and le Van, C. (2001). *Finite Dimensional Convexity and Optimization*. Springer Verlag.
- Folland, G. (1984). *Real Analysis: Modern Techniques and Their Application*. Wiley.
- Follmer, H. and Schied, A. (2002). *Stochastic Finance: An Introduction in Discrete Time*. De Gruyter studies in mathematics. Walter de Gruyter.
- Fudenberg, D. and Tirole, J. (1991). *Game Theory*. MIT Press.
- Gale, D. and Shapley, L. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.
- Galichon, A. (2016). *Optimal Transport Methods in Economics*. Princeton University Press.

- Galichon, A. and Salanié, B. (2022a). Cupid's invisible hand: Social surplus and identification in matching models. *The Review of Economic Studies*, 89(5):2600–2629.
- Galichon, A. and Salanié, B. (2022b). Cupid's invisible hand: Social surplus and identification in matching models. *The Review of Economic Studies*, 89(5):2600–2629.
- Gall, J.-F. L. (2022). *Measure Theory, Probability, and Stochastic Processes*. Springer Verlag.
- Gallardo, M. M., Loaiza, M., and Chávez, J. (2024). Congestion and penalization in optimal transport. Available at SSRN: <https://ssrn.com/abstract=5025651> or <http://dx.doi.org/10.2139/ssrn.5025651>.
- Gentle, J. (2017). *Matrix Algebra Theory, Computations and Applications in Statistics*. Springer, 2 edition.
- Glasserman, P. and Xu, X. (2014). Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58.
- Glasserman, P. and Yang, L. (2015). Bounding wrong-way risk in cva calculation.
- Greinecker, M. and Kah, C. (2021). Pairwise stable matching in large economies. *Econometrica*, 89(6):2929–2974.
- Hull, J. (2006). *Options, futures, and other derivatives*. Pearson Prentice Hall, 8. ed., pearson internat. ed edition.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). *An Introduction to Statistical Learning*. Springer Verlag.
- Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2nd edition.
- Lain, H. C. (1988). Moreau-rockafellar type theorem for convex set functions. *Journal of Mathematical Analysis and Applications*, 132:558–571.

- Laverde, M. (2023). Distance to schools and equal access in school choice systems. *Working Paper*.
- Leon, J. and Gallardo, M. (2025). Fundamentos de econometría. Material de enseñanza #13, DECON.
- Luneberger, D. and Ye, Y. (2021). *Linear and Nonlinear Programming*. Springer Verlag.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. pages 105–142.
- Memartoluie, A. (2017). *Computational Methods in Finance Related to Distributions with Known Marginals*. PhD thesis.
- Merigot, Q. and Thibert, B. (2020). Optimal transport, discretization and algorithms.
- Murphy, K. P. (2022). *Probabilistic Machine Learning: An Introduction*. MIT Press.
- Niederle, M. and Yariv, L. (2009). Decentralized matching with aligned preferences. *Working Paper*.
- Nussabaum, R. (1987). Iterated nonlinear maps and hilbert’s projective metric: A summary. *Dynamics of Infinite Dimensional Systems*.
- Nutz, M. (2024). Quadratically regularized optimal transport: Existence and multiplicity of potentials. *arXiv preprint arXiv:2404.06847*.
- Ok, E. A. (2007). *Real Analysis with Economic Applications*. Princeton University Press.
- Peyre, G. and Cuturi, M. (2019). Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607.
- Peyré, G. and Cuturi, M. (2019). *Computational Optimal Transport: With Applications to Data Science*, volume 11. Foundations and Trends in Machine Learning.
- Pycia, M. (2012). Matching with transfers: The economics of marriage revisited. *American Economic Review*, 102(3):317–322.

- Rawls, J. (1971). *A theory of justice*. Cambridge University Press.
- Rochet, J.-C. and Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4):990–1029.
- Rosen, D. and Saunders, D. (2012). Cva the wrong way. *Journal of Risk Management in Financial Institutions*, 5:252–272.
- Rüschendorf, L. and Thomsen, W. (1993). Note on the schrödinger equation and i-projections. *Statistics & Probability Letters*, 17(5):369–375.
- Shapley, L. and Shubik, M. (1971). The assignment game i: The core. *International Journal of Game Theory*, 1:111–130.
- Sinkhorn, R. (1964). A relationship between arbitrary positive matrices and doubly stochastic matrices. *Annales of Mathematics and Statistics*.
- Tardella, F. (2010). The fundamental theorem of linear programming: Extensions and applications. *Optimization*, 59(3):283–301.
- Toh, K.-C. and Yun, S. (2010). An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific Journal of Optimization*, 6(3):615–640.
- Train, K. E. (2012). *Discrete Choice Methods with Simulation*. Cambridge University Press, 2nd edition.
- Victor Chernozhukov, Guillaume Carlier, G. D. B. and Galichon, A. (2022). Vector quantile regression and optimal transport, from theory to numerics. *Empirical economics*, 62(6):35–62.
- Villani, C. (2009). *Optimal Transport Old and New*. Springer Verlag.
- Walters, C. (2014). The demand for effective charter schools. *Working Paper*.