



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Aproximación discreta del transporte óptimo

Tesis presentada para optar al título de Doctor de la
Universidad de Buenos Aires en el área Ciencias Matemáticas

Maximiliano Omar Frungillo

Director de tesis: Pablo De Nápoli
Consejero de estudios: Pablo Amster
Lugar de trabajo: Departamento de Matemática
Fecha de defensa: 24/06/2024

Buenos Aires, 2024

Aproximación Discreta del Transporte Óptimo

Resumen

En esta tesis estudiamos la aproximación discreta de problemas de transporte óptimo en espacios métricos compactos. Precisamente, consideramos la clase de problemas de Monge–Kantorovich que pueden definirse como una 5-upla (X, Y, μ, ν, c) , donde X e Y son espacios métricos compactos, μ y ν son medidas borelianas de probabilidad en X e Y respectivamente y $c : X \times Y \rightarrow \mathbb{R}$ es una función de costo continua. Para dicha clase de problemas estudiamos la aproximación discreta de los tres objetos más relevantes: el transporte óptimo π_* en el sentido de Kantorovich, el transporte óptimo T_* en el sentido de Monge y el valor óptimo del problema I_* .

Nuestros objetos de estudio son las discretizaciones inducidas por particiones punteadas finitas de diámetro h para los espacios X e Y , que resultan equivalentes a un problema de programación lineal en dimensión finita. En primer lugar introducimos métodos efectivos para construir aproximaciones T_h de T_* a partir de las soluciones discretas π_h de dichos problemas, extendiendo la idea de la *proyección baricéntrica* a espacios sin estructura lineal. Luego probamos, para $h \rightarrow 0$, la convergencia cuantitativa de valores óptimos $I[\pi_h] \rightarrow I_*$ junto con la convergencia de los aproximantes $\pi_h \rightarrow \pi_*$ en W_p y $T_h \rightarrow T_*$ en $L^p(\mu, Y)$. Finalmente, asumiendo una forma de regularidad Hölder a trozos para T_* , garantizamos órdenes de convergencia para $T_h \rightarrow T_*$ en términos del exponente de Hölder de T_* y del orden de crecimiento del operador de costo de Monge en torno a T_* .

Palabras clave: *Transporte Óptimo, Problema de Monge–Kantorovich, Aproximación Discreta, Espacios Compactos, Convergencia en L^p .*

MSC 2020: 65J99, 65K99, 49Q22, 90C08, 35J96.

Página en blanco.

Discrete Approximation of Optimal Transport

Abstract

In this thesis we study the discrete approximation of optimal transport problems on compact metric spaces. Precisely, we consider the class of Monge–Kantorovich problems that can be defined by a 5-tuple (X, Y, μ, ν, c) , where X and Y are compact metric spaces, μ and ν are Borel probability measures on X and Y respectively and $c : X \times Y \rightarrow \mathbb{R}$ is a continuous cost function. For that class of problems we study the discrete approximation of the three most relevant objects: the optimal transport π_* in the sense of Kantorovich, the optimal transport T_* in the sense of Monge and the optimal value of the problem I_* .

Our objects of study are the discretizations induced by finite pointed partitions with diameter h for the spaces X and Y , that turn out to be equivalent to a linear programming problem in finite dimension. First, we introduce effective methods for constructing approximations T_h for T_* from the discrete solutions π_h of said problems, extending the idea of the *barycentric projection* to spaces without linear structure. Then, for $h \rightarrow 0$, we prove the quantitative convergence of optimal values $I[\pi_h] \rightarrow I_*$ and the convergence of approximants $\pi_h \rightarrow \pi_*$ in W_p and $T_h \rightarrow T_*$ in $L^p(\mu, Y)$. Finally, assuming a form of piecewise Hölder regularity for T_* , we guarantee convergence rates for $T_h \rightarrow T_*$ in terms of the Hölder exponent of T_* and the growth order of the Monge cost operator around T_* .

Keywords: *Optimal Transport, Monge–Kantorovich Problem, Discrete Approximation, Compact Spaces, Convergence in L^p .*

MSC 2020: 65J99, 65K99, 49Q22, 90C08, 35J96.

Página en blanco.

Contenidos

Resumen	i
Abstract	iii
Introducción	1
1. Preliminares	11
1.1. Los problemas de Monge y Kantorovich	11
1.2. Minimizantes para el problema de Kantorovich	20
1.2.1. Distancias de Wasserstein	22
1.3. Dualidad de Kantorovich	23
1.4. Minimizantes para el problema de Monge	26
1.5. Las hipótesis (KU) y (SU)	28
1.6. Regularidad	32
2. Análisis de estabilidad para el problema de Monge	37
2.1. Existencia de órdenes de estabilidad	37
2.2. El caso cuadrático en $\mathbb{R}^n$	40
3. Aproximación discreta de (X, Y, μ, ν, c)	45
3.1. Discretización de (X, Y, μ, ν, c)	45
3.2. Planes discretos y proyecciones	48
3.2.1. Medianas geométricas pesadas y baricentros	48
3.2.2. GM proyecciones y B proyecciones	50
3.3. Cómputo de proyecciones	51
4. Convergencia de planes discretos y valores óptimos	53
4.1. Versión continua de un h -plan	53
4.2. Demostración del Teorema A	55
4.3. Discusión	56

5. Convergencia de proyecciones	59
5.1. Versión semidiscreta de un h -plan	59
5.2. Demostración del Teorema B	63
5.3. Discusión	66
6. Órdenes de convergencia para proyecciones	69
6.1. Regularidad Hölder a trozos	69
6.2. Transportes compatibles	71
6.3. Demostración del Teorema C	79
6.4. Discusión	82
6.5. Demostración del Teorema D	85
Apéndices	89
A. Medida en espacios métricos	89
A.1. Medidas regulares	89
A.2. Convergencia débil de medidas	91
A.3. Rigidez y Teorema de Prokhorov	94
A.4. Desintegración de medidas	95
Bibliografía	99

Introducción

En las últimas décadas la teoría de transporte óptimo se consolidó como una teoría matemática en sí misma, situándose en la intersección entre la teoría de probabilidades, el análisis convexo, la optimización, el cálculo variacional y el estudio de ecuaciones en derivadas parciales. Su relevancia se debe no sólo a la generalidad y aplicabilidad de sus resultados dentro de la matemática sino también a sus diversas aplicaciones en áreas como la mecánica de fluidos [5, 24], la meteorología [48, 49], el transporte [40, 41, 115, 128], la economía y las finanzas cuantitativas [13, 47, 53, 72, 73], el procesamiento de imágenes [79, 110, 113] y más recientemente la gran área de análisis de datos, estadística computacional y machine learning [10, 36, 46, 98, 99, 121, 122, 129].

En esta tesis estudiamos la aproximación discreta de problemas de transporte óptimo en espacios métricos compactos, obteniendo resultados de convergencia y estabilidad bajo hipótesis generales. A continuación presentamos dichos resultados, introduciendo primero las definiciones necesarias para formularlos y darles contexto. Por razones de claridad y brevedad omitimos algunos tecnicismos que podrían oscurecer la presentación de las ideas fundamentales. Las definiciones precisas de los objetos discutidos en esta introducción serán dadas en los Capítulos 1, 2, 3 y en los Apéndices.

El problema del Transporte Óptimo

Considerar espacios métricos X, Y dotados de distribuciones de masa μ, ν junto con una función $c(x, y)$ que indique el costo de mover una unidad de masa desde $x \in X$ hasta $y \in Y$. El *problema del transporte óptimo de masa* es el de redistribuir la masa de X en Y de modo que μ se transforme en ν , minimizando el costo de mover la totalidad de la masa de X . Notar que la factibilidad del problema requiere $\mu(X) = \nu(Y)$, con lo cual siempre puede asumirse que μ y ν son medidas de probabilidad.

Para formalizar la idea de redistribuir μ en ν decimos que una función $T : X \rightarrow Y$ es un *transporte* de μ en ν si verifica

$$\mu[T^{-1}(F)] = \nu[F] \text{ para todo } F \subset Y \text{ medible.}$$

Cuando esto sucede lo notamos $T_{\#}\mu = \nu$ y decimos que ν es el *pushforward* de μ por T , reservando la notación $\mathcal{T}(\mu, \nu)$ para el conjunto de los transportes de μ en ν . Cada transporte T de μ en ν tiene un costo total asociado, que resulta de acumular los costos de mover cada diferencial de masa $d\mu(x)$ desde su posición inicial x hasta su posición final $T(x)$. Así, el costo total de T resulta ser

$$C[T] := \int_X c(x, T(x)) d\mu(x) \quad (1)$$

y llamamos *operador de costo de Monge* al operador C definido por (1). El *problema de Monge* inducido por (X, Y, μ, ν, c) es el de decidir si existe un transporte óptimo T_* de μ en ν que minimice el valor de C .

Así como una función $T : X \rightarrow Y$ transporta medidas de X en Y , la aplicación producto asociada $\text{Id}_X \otimes T := x \mapsto (x, T(x))$ transporta medidas de X en $X \times Y$. Notando π_T a la medida $(\text{Id}_X \otimes T)_{\#}\mu$ y notando $\mathcal{A}(\mu, \nu)$ al conjunto de medidas en $X \times Y$ con marginales μ y ν , resulta la equivalencia $T \in \mathcal{T}(\mu, \nu) \iff \pi_T \in \mathcal{A}(\mu, \nu)$. Más aun, se tiene

$$C[T] = \int_X c(x, T(x)) d\mu(x) = \int_{X \times Y} c(x, y) d\pi_T(x, y) = I[\pi_T], \quad (2)$$

donde I es el *operador de costo de Kantorovich* definido por

$$I[\pi] = \int_{X \times Y} c(x, y) d\pi(x, y).$$

La igualdad (2) permite interpretar el problema de minimizar I sobre $\mathcal{A}(\mu, \nu)$ como una relajación del problema de Monge. Llamaremos *problema de Kantorovich* inducido por (X, Y, μ, ν, c) a dicho problema.

Un caso particular relevante para nuestros propósitos se da cuando μ y ν son medidas finitamente soportadas. En tal caso puede asumirse que X e Y son finitos: digamos que para conjuntos finitos de índices I, J se tiene $X = \{x_i\}_{i \in I}$, $Y = \{y_j\}_{j \in J}$ y por lo tanto

$$\mu = \sum_{i \in I} \mu_i \delta_{x_i} \quad y \quad \nu = \sum_{j \in J} \nu_j \delta_{y_j}. \quad (3)$$

Luego de identificar las medidas finitamente soportadas en $X \times Y$ con sucesiones finitas doblemente indexadas, el problema de Kantorovich toma la forma de un problema de programación lineal en dimensión finita que puede escribirse como

$$\begin{aligned} \min_{(\pi_{ij})_{ij} \geq 0} \quad & \sum_{i \in I} \sum_{j \in J} \pi_{ij} c_{ij} \\ \text{s. a.} \quad & \sum_{j \in J} \pi_{ij} = \mu_i \quad \forall i \in I, \quad \sum_{i \in I} \pi_{ij} = \nu_j \quad \forall j \in J, \end{aligned} \quad (\text{FP})$$

donde $c_{ij} = c(x_i, y_j)$ y las restricciones en (FP) representan la condición de que las medidas marginales de $\pi = (\pi_{ij})_{ij}$ sean μ y ν . Es fácil ver que no siempre es factible el problema de Monge en el caso discreto, aunque sí lo es en algunos casos particulares de interés, como por ejemplo cuando $|I| = |J| = N$ y $\mu_i = \nu_j = \frac{1}{N}$ para todo $i \in I$ y todo $j \in J$.

Métodos numéricos para (X, Y, μ, ν, c)

El diseño de estrategias numéricas para aproximar el transporte óptimo depende tanto de la estructura disponible en X, Y como de la regularidad de las medidas μ, ν y de la función de costo c . Puede decirse que existen tres clases de métodos: los *continuos*, que asumen medidas μ, ν absolutamente continuas; los *discretos*, aplicables cuando μ, ν tienen soportes finitos; y los *semidiscretos*, que asumen una medida absolutamente continua y otra discreta. Teóricamente siempre es posible discretizar μ y ν de algún modo, por lo que en principio todo problema (X, Y, μ, ν, c) puede ser tratado con métodos discretos.

Una buena propiedad del caso discreto es que se conocen algoritmos eficientes [51, 68, 71, 81, 119] para aproximar minimizantes de (FP) desde mediados del siglo pasado. Entre ellos destacamos el método (network) simplex [2, 21, 52, 92], una variedad de algoritmos combinatorios [2, 19, 20, 26, 86, 118] o el más reciente método de regularización entrópica [50, 105]. Generalmente sucede que la estructura adicional de (X, Y, μ, ν, c) permite diseñar métodos específicos con mejor performance que en el caso general, como sucede para ciertos problemas unidimensionales [27, 58, 83, 127] en los que puede explotarse la monotonía de c o como cuando (FP) proviene de discretizar un problema continuo regular, donde resultan eficientes los métodos [102, 117] fundamentados en la idea de que $(\pi_{ij})_{ij}$ se concentra en torno al gráfico de un transporte óptimo T_* para el problema continuo original.

Por otra parte, los métodos semidiscretos [11, 84, 88, 101] y continuos [9, 14, 15, 16, 45, 69, 70, 90, 103, 109] en general resultan de aplicar ideas provenientes del cálculo variacional y el análisis de PDEs, de modo que requieren alguna estructura diferenciable y usualmente se limitan a problemas en espacios euclídeos con costo cuadrático. Dicha limitación surge generalmente de la formulación misma del método, de las dificultades computacionales de algún problema intermedio o bien de la falta de garantías de convergencia. Para una discusión de algunos de estos métodos ver [116, Cap. 6] o [105, Caps. 5, 7].

Aproximación discreta de (X, Y, μ, ν, c)

En esta tesis consideramos una generalización del esquema propuesto por Oberman y Ruan en [102], permitiendo aproximar problemas (X, Y, μ, ν, c) generales. El esquema consiste en particionar X, Y con resolución espacial h para construir medidas discretas μ_h, ν_h que inducen un problema discreto de la forma (FP). Resolviendo (FP) obtendremos un plan de transporte discreto π_h de costo $I[\pi_h]$, sirviendo como aproximantes de π_* e

I_* respectivamente. También construiremos aproximantes T_h de T_* a partir π_h mediante cálculos efectivos y sin asumir estructura lineal en (Y, d_Y) . Nuestro principal objetivo será probar rigurosamente la convergencia de estos tres aproximantes para $h \rightarrow 0^+$, con hipótesis de regularidad mínimas sobre (X, Y, μ, ν, c) .

Discretización de (X, Y, μ, ν, c)

Un método simple para discretizar medidas en un espacio Z consiste en distribuir puntos en Z y pesarlos según el comportamiento local de la medida en cuestión. Esta idea se conoce como *cuantización* en el caso general y como *clustering* para el caso en el que la medida original ya es discreta. Notar que aun si μ y ν son discretas tiene sentido aproximarlas por nuevas medidas discretas con menor cantidad de átomos a fin de reducir la dimensión de (FP). En ciertos contextos tiene sentido llevar a cabo el procedimiento descrito anteriormente con algún criterio de calidad, sirviendo como ejemplos el método de Lloyd [91] en el caso continuo y el método k -means [82, 94] en el caso discreto.

En nuestro caso vamos a aplicar el método antes descrito a medidas en espacios métricos generales, ignorando toda noción de calidad de la cuantización/clustering. Para esto consideramos particiones punteadas finitas en X e Y , digamos

$$\mathcal{C}_X = \{(E_i, x_i)\}_{i \in I} \text{ y } \mathcal{C}_Y = \{(F_j, y_j)\}_{j \in J},$$

donde sólo asumimos que los conjuntos E_i, F_j son borelianos y verifican $x_i \in E_i, y_j \in F_j$. Al par de particiones $\mathcal{C}_X, \mathcal{C}_Y$ le asociamos un parámetro de resolución

$$h := \max \left\{ \max_{i \in I} \text{diám}(E_i), \max_{j \in J} \text{diám}(F_j) \right\}$$

y medidas finitamente soportadas

$$\mu_h = \sum_{i \in I} \mu[E_i] \cdot \delta_{x_i} \quad \text{y} \quad \nu_h = \sum_{j \in J} \nu[F_j] \cdot \delta_{y_j}.$$

Escribiendo $\mu_i = \mu[E_i]$ y $\nu_j = \nu[F_j]$ las medidas μ_h, ν_h toman la forma (3), de modo que las particiones $\mathcal{C}_X, \mathcal{C}_Y$ inducen inmediatamente un problema de la forma (FP). Dicho problema es el problema de Kantorovich asociado a (X, Y, μ_h, ν_h, c) , que resulta ser una aproximación discreta del problema de Kantorovich para (X, Y, μ, ν, c) .

Construcción de aproximantes para el Transporte Óptimo

Para aproximar el transporte óptimo en el sentido de Kantorovich basta tomar un plan de transporte óptimo para (X, Y, μ_h, ν_h, c) o, equivalentemente, un minimizante π_h de

(FP). Para h pequeño es esperable que π_h sea una buena aproximación discreta de algún plan de transporte óptimo π_* para (X, Y, μ, ν, c) . La construcción de un aproximante T_h para el transporte óptimo T_* en el sentido de Monge no es igual de inmediata, dado que π_h no necesariamente se concentra sobre un gráfico. En general sucede que para $i \in I$ existe más de un $j \in J$ con $\pi_{ij} > 0$, de modo que no existe un candidato inmediato para $T_h(x_i)$ y por lo tanto es necesaria alguna construcción para producir T_h a partir de π_h .

La estrategia usual para obtener un aproximante T_h de T_* a partir de π_h consiste en construir T_h a trozos del siguiente modo: para todo $i \in I$ y todo $x \in E_i$ tomar

$$T_h(x) = T_h|_{E_i}(x) = y_i := \sum_{j \in J} \frac{\pi_{ij}}{\mu_i} \cdot y_j. \quad (4)$$

Esta construcción se conoce como *proyección baricéntrica* y carece de sentido si Y no tiene estructura lineal, por lo que propondremos una construcción alternativa disponible para Y general. Para esto notar que y_i puede interpretarse como el valor esperado $\mathbb{E}[Y_i]$ de una distribución Y -valuada Y_i que toma cada valor y_j con probabilidad π_{ij}/μ_i . Luego bastaría considerar una noción de tendencia central que dependa sólo de la estructura métrica de (Y, d_Y) en lugar de $\mathbb{E}$. Como se explicará en el Capítulo 3, seguiremos esta estrategia utilizando medianas geométricas pesadas y llamaremos *proyección* a todo aproximante T_h de T_* construido a partir de π_h por alguno de estos dos métodos.

La construcción propuesta no puede ser calculada efectivamente dado que, a diferencia del baricentro, la mediana geométrica no resulta de evaluar una expresión efectiva. Esto es un defecto desde el punto de vista computacional, pero lo solucionaremos asignando a toda proyección T_h un parámetro de calidad $Q \geq 0$ que indica qué tan exacto es su cómputo. Como se explicará en el Capítulo 3, en nuestro contexto siempre será posible construir efectivamente proyecciones T_h de calidad $Q = 1$, lo que bastará para computar aproximantes arbitrariamente buenos de T_* a partir de π_h .

Resultados obtenidos

En lo que sigue supondremos que (X, d_X) e (Y, d_Y) son espacios métricos compactos y que $c : X \times Y \rightarrow \mathbb{R}$ es una función continua, llamando *compacto* a un problema (X, Y, μ, ν, c) que verifica dichas hipótesis. En tal caso resultan finitos los diámetros $D_X := \text{diám}(X)$ y $D_Y := \text{diám}(Y)$, a la vez que la métrica producto

$$d_X \times d_Y((x_1, y_1), (x_2, y_2)) = \max \{d_X(x_1, x_2), d_Y(y_1, y_2)\}$$

hace de $(X \times Y, d_X \times d_Y)$ un espacio compacto. Así, c resulta uniformemente continua con módulo de continuidad ω_c dado por

$$\omega_c(h) = \max \left\{ |c(z_1) - c(z_2)| : z_1, z_2 \in X \times Y \text{ con } d_X \times d_Y(z_1, z_2) \leq h \right\}.$$

Vamos a metrizar el conjunto $\mathcal{P}(Z)$ de medidas de probabilidad en un espacio métrico (Z, d_Z) con las *distancias de Wasserstein* W_p , definidas como

$$W_p(\pi_1, \pi_2) = \left(\min_{\sigma \in \mathcal{A}(\pi_1, \pi_2)} \int_{Z \times Z} d_Z^p(z_1, z_2) d\sigma(z_1, z_2) \right)^{1/p} \text{ para } 1 \leq p < \infty.$$

En nuestro caso particular tomaremos $Z = X \times Y$ y $d_Z = d_X \times d_Y$, con lo cual d_Z resultará acotada y por lo tanto W_p metrizará la topología de la convergencia débil en $\mathcal{P}(X \times Y)$. Por otra parte, metrizaremos el conjunto de funciones borelianas $\mathcal{B}(X, Y)$ con las distancias d_p de los espacios $L^p(\mu, Y)$, definidas por

$$d_p(T_1, T_2) = \left(\int_X d_Y^p(T_1(x), T_2(x)) d\mu(x) \right)^{1/p} \text{ para } 1 \leq p < \infty.$$

Notar que, luego de identificar las funciones de $\mathcal{B}(X, Y)$ que coinciden μ -a.e., la buena definición de d_p es consecuencia de la acotación de Y y de la finitud de μ .

Para la convergencia de los valores óptimos $I[\pi_h]$ requeriremos sólo la compacidad de (X, Y, μ, ν, c) , mientras que para la convergencia de los aproximantes π_h y T_h serán necesarias hipótesis de unicidad adicionales. Requeriremos, respectivamente, una hipótesis de unicidad (KU) para el problema de Kantorovich y una hipótesis de unicidad fuerte (SU) que vincula los problema de Monge y Kantorovich:

(KU) Existe un único plan de transporte óptimo $\pi_* \in \mathcal{A}(\mu, \nu)$.

(SU) Existe un único plan de transporte óptimo $\pi_* \in \mathcal{A}(\mu, \nu)$ y está inducido por un transporte $T_* \in \mathcal{T}(\mu, \nu)$, con lo cual T_* es único y $C[T_*] = I[\pi_*]$.

Es claro que (SU) implica (KU) y también es fácil ver que la implicación recíproca no es cierta en general. Ambas hipótesis son no vacías dado que (SU) se verifica para una variedad de problemas de transporte, entre los que podemos mencionar problemas en $\mathbb{R}^d$ con costos cóncavos o convexos [33, 76, 77, 106], problemas en $\mathbb{R}^d$ con costos generales [38, 54, 87, 93], problemas en variedades riemannianas con costos generales [17, 60, 63, 96, 125] y problemas en espacios métricos geodésicos que satisfacen ciertas condiciones de curvatura [18, 42, 78, 111]. Por otra parte, parece ser un problema difícil el de decidir clases no triviales de problemas que impliquen (KU) sin forzar también (SU). Para una discusión sobre este punto ver [1, en particular Theorem 5.1 y discusión posterior].

Vamos a considerar discretizaciones dadas por sucesiones de particiones con diámetros $(h_k)_k$ verificando $h_k \rightarrow 0$. Para cada discretización tomamos un plan discreto óptimo π_k del

cual extraemos una proyección T_k . Llamamos *sucesiones aproximantes de planes discretos* a la sucesiones $(\pi_k)_k$ y *sucesiones aproximantes de proyecciones* a las sucesiones $(T_k)_k$.

Teorema A Sea (X, Y, μ, ν, c) un problema compacto con valor óptimo I_* para su problema de Kantorovich. Entonces todo h -plan óptimo π_h satisface

$$|I[\pi_h] - I_*| \leq \omega_c(h), \quad (5)$$

de modo que si $(\pi_k)_k$ es una sucesión aproximante de planes discretos resulta $I[\pi_k] \xrightarrow{k} I_*$. Más aun, si (X, Y, μ, ν, c) satisface adicionalmente la hipótesis (KU) con plan de transporte óptimo π_* entonces se tiene $W_p(\pi_k, \pi_*) \xrightarrow{k} 0$ para todo $1 \leq p < \infty$.

Teorema B Sea (X, Y, μ, ν, c) un problema compacto satisfaciendo la hipótesis (SU) con transporte óptimo T_* . Luego, si $(T_k)_k$ es una sucesión aproximante de proyecciones, para todo $1 \leq p < \infty$ se tiene $d_p(T_k, T_*) \xrightarrow{k} 0$.

Notar que los Teoremas A y B sólo cuantifican la convergencia de valores óptimos, por lo que restaría estudiar la posibilidad de cuantificar las convergencias de π_h y T_h . Para esto requeriremos regularidad adicional para el problema (X, Y, μ, ν, c) y nos limitaremos a cuantificar la convergencia $T_h \rightarrow T_*$. El motivo de esta decisión es la falta de resultados generales no triviales que permitan concluir (KU) sin concluir también (SU). Peor aun, puede fallar (KU) incluso bajo hipótesis lo suficientemente fuertes como para concluir la regularidad del soporte de planes de transporte óptimos [97, Remark 3.3].

Para obtener una versión cuantitativa del Teorema B vamos a requerir una forma de regularidad α -Hölder a trozos para T_* , motivada por los resultados de regularidad parcial disponibles para el transporte óptimo en $\mathbb{R}^d$ y en variedades riemannianas [55, 64, 66]. Asumiendo dicha regularidad para T_* cuantificaremos $d_p(T_h, T_*)$ en términos de h , del exponente α y del orden de estabilidad del operador de costo de Monge C.

Cuando el problema compacto (X, Y, μ, ν, c) verifica la hipótesis (SU) con transporte óptimo T_* , el operador de costo de Monge C admite un p -orden de estabilidad alrededor de T_* para todo $1 \leq p < \infty$. Esto quiere decir que existe una función $s_p : [0, +\infty) \rightarrow [0, +\infty)$ no decreciente, continua en 0, satisfaciendo $s_p(0) = 0$ y tal que

$$d_p(T, T_*) \leq s_p(C[T] - C[T_*])$$

para todo $T \in \mathcal{B}(X, Y)$ con $T_{\#}\mu = \nu$, siendo s_p la mínima función con estas propiedades.

Por otra parte, decimos que una función $T \in \mathcal{B}(X, Y)$ es α -Hölder a trozos si, luego de remover un conjunto cerrado $\Sigma \subset X$ con $\mu[\Sigma] = 0$, el conjunto $X \setminus \Sigma$ puede ser particionado por abiertos $(X_k)_k$ de modo tal que la restricción de T a cada uno de ellos es α -Hölder con constante C_α uniforme en k . Notamos $C_{PW}^{0, \alpha}(X, Y)$ al conjunto de dichas funciones.

Teorema C Sea (X, Y, μ, ν, c) un problema compacto satisfaciendo la hipótesis (SU) con transporte óptimo T_* . Adicionalmente, supongamos que μ es no atómica y que T_* está en

$C_{PW}^{0,\alpha}(X, Y)$ con constante C_α . Luego, dados $Q \geq 0$ y $1 \leq p < \infty$, existen constantes A, B dependientes sólo de $p, Q, \alpha, C_\alpha, D_X$ y D_Y tales que, si $0 < h \leq \min\{D_X, D_Y\}$, vale

$$d_p(T_h, T_*) \leq Ah^\alpha + B\sqrt{s_p(4\omega_c(h))} \quad (6)$$

para toda proyección T_h de calidad Q extraída de un h -plan óptimo.

Como aplicación del Teorema C, consideremos el caso del transporte óptimo cuadrático entre convexos de $\mathbb{R}^d$. Asumiendo regularidad de las medidas μ y ν , la teoría de regularidad de Caffarelli garantiza en este caso que T_* es un difeomorfismo Lipschitz, lo que a su vez implica que $s_2(t) \leq K\sqrt{t}$ para una constante K . Aplicando el Teorema C, resulta que $d_2(T_h, T_*)$ es $O(h^{1/4})$. Un análisis más fino, específico para el transporte cuadrático en $\mathbb{R}^d$, muestra que $d_2(T_h, T_*)$ es $O(h^{1/2})$ bajo estas hipótesis [89]. La experiencia empírica sugiere, bajo estas mismas hipótesis, que el orden de convergencia es $O(h)$ [102].

Un defecto del Teorema C es que su aplicación práctica depende de que exista una teoría de regularidad para el problema (X, Y, μ, ν, c) . Por otra parte, también llama la atención la raíz cuadrada en el lado derecho de (6), que como se verá en la prueba del Teorema C, aparece al controlar la dispersión de $\text{sop}(\pi_h)$ respecto del gráfico de T_* . El siguiente y último resultado que presentamos es un intento de aliviar estos defectos.

Para medir la dispersión de $\text{sop}(\pi_h)$ vamos a calcular la norma L^p de la “oscilación puntual” de π_h visto como función multivaluada, de modo que una dispersión chica para $\text{sop}(\pi_h)$ indique que π_h está casi concentrada en un gráfico. Precisamente, dado un h -plan $\pi = (\pi_{ij})_{ij}$, consideramos $D(\pi_h) : I \rightarrow \mathbb{R}_0^+$ dada por

$$D(\pi_h)(i) = \max \left\{ d_Y(y_{j_1}, y_{j_2}) : j_1, j_2 \text{ tales que } \pi_{ij_1} \neq 0 \neq \pi_{ij_2} \right\} \cup \{0\}$$

y definimos

$$\begin{aligned} \|D(\pi_h)\|_p &= \left(\sum_{i \in I} (D(\pi_h)(i))^p \mu_i \right)^{1/p} \quad \text{para } 1 \leq p < \infty, \\ \|D(\pi_h)\|_\infty &= \max \{ D(\pi_h)(i) : \text{para } i \in I \text{ con } \mu_i > 0 \}. \end{aligned}$$

Notar que la medida de dispersión $\|D(\pi_h)\|_p$ depende de $\mathcal{C}_X$ y de μ . Notar también que $\|D(\pi_h)\|_p = 0$ si y sólo si $\text{sop}(\pi_h)$ está μ -a.e. contenido en el gráfico de una función $X \rightarrow Y$ constante sobre cada conjunto E_i de la partición $\mathcal{C}_X$.

Teorema D Sea (X, Y, μ, ν, c) un problema compacto. Supongamos, adicionalmente, que el problema satisface (SU) con transporte óptimo T_* y que μ es no atómica. Luego, dados $Q \geq 0$ y $1 \leq p < \infty$, existen constantes A, B dependientes sólo de p y Q tales que

$$d_p(T_h, T_*) \leq Ah + B\|D(\pi_h)\|_p + s_p(4\omega_c(h)) \quad (7)$$

para toda proyección T_h de calidad Q extraída de un h -plan óptimo π_h .

El Teorema D es, en principio, de aplicación empírica. Consideremos un problema con función de costo Lipschitz, asumiendo que $s_p(h)$ es $O(h^\beta)$ para algún $\beta \in (0, 1]$. Si al tomar particiones cada vez más finas resulta $\|D(\pi_h)\|_p \leq Kh$, entonces (7) implica que $d_p(T_h, T_*)$ es $O(h^\beta)$, indicando que las aproximaciones T_h convergen con orden al menos $s_p(h)$. Los experimentos numéricos [102] sugieren que es razonable esperar $\|D(\pi_h)\|_p \leq Kh$ para alguna constante K independiente de h cuando se toman particiones suficientemente regulares, al menos para costos convexos y dominios “buenos” en $\mathbb{R}^d$.

Contribución y trabajos relacionados

Desde el punto de vista de los objetivos y resultados obtenidos puede decirse que los trabajos [12, 56, 57, 89, 102] están entre los más cercanos al trabajo realizado en esta tesis, cuyos aportes pueden resumirse en tres ítems principales:

- i)* Un método efectivo para aproximar T_* cuando el espacio de llegada (Y, d_Y) no posee estructura lineal, generalizando la idea de la proyección baricéntrica.
- ii)* Los Teoremas A y B que garantizan la convergencia de aproximaciones discretas para problemas compactos generales, siendo resultados novedosos aun para problemas cuadráticos compactos en $\mathbb{R}^d$.
- iii)* Un framework para obtener resultados cuantitativos de convergencia: si la teoría de regularidad disponible permite estimar $s_p(h)$ para una clase $\mathfrak{C}$ de problemas, los Teoremas C y D garantizan órdenes de convergencia de T_h para problemas en $\mathfrak{C}$.

Como antecedente de *i)* podemos mencionar el estudio [6] de la proyección baricéntrica como herramienta teórica y el estudio [56, 89, 102] de su aplicación práctica al cómputo de aproximaciones de T_* . Con respecto a *ii)* y *iii)* podemos mencionar como antecedentes el estudio [12, 57] de convergencia de valores óptimos para esquemas de discretización similares al aquí propuesto y el estudio tanto experimental [102] como teórico [56, 89] de las propiedades de convergencia de aproximantes T_h construidos vía proyección baricéntrica.

A diferencia del Teorema A, la aplicabilidad de los resultados de convergencia de valores óptimos obtenidos en [12, 57] se limita a problemas en $\mathbb{R}^d$. En [57] los autores estudian un esquema de discretización similar al presentado aquí aplicando la teoría de análisis de intervalos [80, 100], obteniendo resultados de convergencia para problemas en $\mathbb{R}^1$ con medidas absolutamente continuas y $c \in C(\mathbb{R} \times \mathbb{R})$. Por otra parte, en [12] se asume que X e Y son dominios triangulables en $\mathbb{R}^d$ y su análisis permite probar convergencia cuantitativa en el sentido de (5), obteniendo orden $O(h^{1+\alpha})$ para $c \in C^{1,\alpha}(\mathbb{R} \times \mathbb{R})$.

Con respecto a la convergencia de T_h , los trabajos [56, 89] se limitan a problemas cuadráticos en $\mathbb{R}^d$ con T_* satisfaciendo una condición de Lipschitz, obteniendo cotas para $\|T_h - T_*\|_{L^2}$ en ambos casos. Mientras que en [89] se concluye la convergencia de T_h

con orden $O(\sqrt{h})$ para esquemas discretos como el aquí estudiado, en [56] se obtiene un resultado de convergencia estocástico para medidas empíricas μ_h, ν_h . Finalmente señalamos que la estrategia propuesta en [102, Remark 2.2] para deducir la convergencia de T_h a T_* a partir de resultados generales sobre proyecciones baricéntricas permitiría, en principio, obtener resultados como el Teorema B sólo para (Y, d_Y) lineal y c convexo.

Estructura de la Tesis

En el Capítulo 1 presentamos los resultados fundamentales de la teoría de transporte óptimo en espacios polacos. Junto con la presentación de la teoría general discutimos algunos ejemplos y resultados concretos que contextualizan los resultados presentados en este trabajo. Luego discutimos la hipótesis (SU) y presentamos los resultados de regularidad parcial que motivan la noción de regularidad Hölder a trozos.

En el Capítulo 2 estudiamos la estabilidad del operador de costo de Monge C para problemas compactos (X, Y, μ, ν, c) generales. Probamos la existencia de s_p y a modo de ejemplo estimamos $s_2(h)$ para el caso del transporte cuadrático en $\mathbb{R}^d$, obteniendo una demostración alternativa para el resultado de estabilidad probado por Li y Nochetto [89].

En el Capítulo 3 fijamos la notación que utilizaremos para trabajar con los h -planes y definimos los distintos tipos de proyecciones que vamos a considerar. Presentamos una clase de proyecciones análogas a la proyección baricéntrica, las *GM proyecciones*, disponibles en espacios métricos compactos generales. Finalizamos el capítulo mostrando que dichas proyecciones pueden ser efectivamente calculadas.

En los Capítulos 4 y 5 estudiamos la convergencia de valores óptimos, planes discretos y proyecciones. Probamos los Teoremas A y B y mostramos que ambos resultados son sharp. Finalmente, en el Capítulo 6 introducimos la noción de regularidad Hölder a trozos para luego probar y discutir los Teoremas C y D.

Capítulo 1

Preliminares

En este capítulo presentamos los resultados fundamentales de la teoría de transporte óptimo en espacios generales. Pruebas completas de los resultados presentados en este capítulo pueden leerse en la bibliografía estándar del área [3, 4, 6, 65, 116, 124, 125] y en los trabajos oportunamente citados a lo largo del capítulo.

Empezamos formulando los problemas de Monge y Kantorovich en el contexto de espacios métricos polacos en la Sección 1.1. Probamos la existencia de minimizantes para el problema de Kantorovich en la Sección 1.2 y presentamos su formulación dual en la Sección 1.3. A continuación estudiamos la existencia de minimizantes para el problema de Monge en la Sección 1.4 y su relación con la hipótesis (SU) en la Sección 1.5. Finalmente, en la Sección 1.6 presentamos resultados de regularidad y regularidad parcial en espacios euclídeos y variedades riemannianas.

1.1. Los problemas de Monge y Kantorovich

1.1.1. Formulación del problema de Monge

Definición 1.1.1 Sean X e Y dos espacios métricos, $T : X \rightarrow Y$ una función boreliana y $\mu \in \mathcal{P}(X)$ una medida boreliana de probabilidad. Se define el **pushforward** de μ por T como la medida de probabilidad $T_{\#}\mu \in \mathcal{P}(Y)$ dada por

$$T_{\#}\mu[E] = \mu[T^{-1}(E)]$$

para todo $E \subseteq Y$ boreliano. Si para $\nu \in \mathcal{P}(Y)$ se verifica $T_{\#}\mu = \nu$, decimos que T es un **transporte** de μ en ν . Notamos $\mathcal{T}(\mu, \nu)$ al conjunto de los transportes de μ en ν .

Lema 1.1.2 *Dada una aplicación boreliana entre espacios métricos $T : X \rightarrow Y$ y dos*

medidas de probabilidad $\mu \in \mathcal{P}(X)$ y $\nu \in \mathcal{P}(Y)$, se tiene $T_{\#}\mu = \nu$ si y sólo si

$$\int_Y \varphi d\nu = \int_X \varphi \circ T d\mu$$

para toda $\varphi \in L^1(\nu)$.

Consideramos el problema de transportar masa, inicialmente distribuida en X según μ , de modo que finalmente se distribuya en Y según ν . Además nos interesa hacerlo de forma óptima, minimizando el costo total de transporte. Para dar forma a esta idea consideramos una función medible $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ de modo que $c(x, y)$ se interprete como el costo de transportar masa desde la posición $x \in X$ hasta la posición $y \in Y$. En términos de *transportes* en el sentido de la definición 1.1.1, el problema se formula como sigue.

Problema de transporte óptimo de Monge

Dados espacios métricos X, Y junto con medidas de probabilidad $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y un costo medible $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$, minimizar

$$C[T] = \int_X c(x, T(x)) d\mu(x) \quad (\text{MP})$$

para T en el conjunto $\mathcal{T}(\mu, \nu)$ de los transportes de μ en ν .

Siguiendo la terminología usual para problemas de optimización, decimos que (MP) es **factible** si $\mathcal{T}(\mu, \nu)$ es no vacío y decimos que es **infactible** en caso contrario. Llamamos **conjunto factible** al conjunto $\mathcal{T}(\mu, \nu)$ y **soluciones factibles** a sus elementos. Una **solución óptima** o **transporte óptimo** para (MP) es un elemento $T_* \in \mathcal{T}(\mu, \nu)$ que minimiza C sobre $\mathcal{T}(\mu, \nu)$ y, en tal caso, llamamos **valor óptimo** a $C[T_*]$.

La principal dificultad para probar la existencia de minimizantes para el problema de Monge son las malas propiedades del conjunto factible $\mathcal{T}(\mu, \nu)$. Precisamente:

- i) cuando μ tiene átomos $\mathcal{T}(\mu, \nu)$ puede ser vacío;
- ii) el conjunto factible $\mathcal{T}(\mu, \nu)$ no es compacto.

La consecuencia de i) es que el problema (MP) puede ser infactible. La consecuencia de ii) es que, aun siendo (MP) factible, C podría no alcanzar un mínimo sobre $\mathcal{T}(\mu, \nu)$. En particular, ii) impide aplicar el método directo del cálculo de variaciones para probar que (MP) admite un minimizante.

Lema 1.1.3 *Para la factibilidad del problema (MP) es necesaria la condición*

$$(\forall x \in X) : (\exists y \in Y) : \mu[\{x\}] \leq \nu[\{y\}]. \quad (8)$$

En particular, (MP) no es factible si μ tiene al menos un átomo y ν es no atómica.

Demostración. Supongamos que no se verifica (8) y que existe $T \in \mathcal{T}(\mu, \nu)$. Entonces existe $x_0 \in X$ verificando $\nu[\{y\}] < \mu[\{x_0\}]$ para todo $y \in Y$ y por lo tanto

$$\mu[\{x_0\}] \leq \mu[T^{-1}(T(x_0))] = \mu[T^{-1}(\{T(x_0)\})] = \nu[\{T(x_0)\}] < \mu[\{x_0\}]. \quad \square$$

Es fácil ver que la condición (8) no es suficiente para la existencia de transportes. Por ejemplo, tomando $X = Y = \{z_1, z_2\}$ y definiendo $\mu = \frac{1}{2}\delta_{z_1} + \frac{1}{2}\delta_{z_2}$ y $\nu = \frac{2}{3}\delta_{z_1} + \frac{1}{3}\delta_{z_2}$, es claro que vale (8) pero $\mathcal{T}(\mu, \nu) = \emptyset$. Notar que la condición (8) se satisface trivialmente si μ es no atómica. Más aun, cuando μ es no atómica los teoremas clásicos de isomorfismo de espacios de medida implican que $\mathcal{T}(\mu, \nu)$ es no vacío.

Teorema (75, 108) Sean X e Y son espacios métricos polacos, $\mu \in \mathcal{P}(X)$ y $\nu \in \mathcal{P}(Y)$. Si μ es no atómica, entonces existe $T \in \mathcal{B}(X, Y)$ que transporta μ en ν .

A continuación damos dos ejemplos que ilustran las dificultades provocadas por el hecho de que $\mathcal{T}(\mu, \nu)$ no es cerrado.

Ejemplo 1.1.4 Sea $\phi : \mathbb{R} \rightarrow \mathbb{R}$ la extensión 1-periódica de la función dada por

$$[0; 1) \ni x \mapsto \begin{cases} 1 & , \text{ si } x \in [0, \frac{1}{2}) \\ -1 & , \text{ si } x \in [\frac{1}{2}, 1) \end{cases}$$

y sea $(T_n)_n$ la sucesión de funciones de $[0, 1]$ en $[0, 1]$ definida por $T_n(x) = \phi(nx)$ para $n \in \mathbb{N}$ y $x \in [0, 1]$. Sean $X = Y = [0, 1]$, $\mu = \mathcal{L}^1|_{[0,1]}$ y $\nu = (\delta_{-1} + \delta_1)/2$. Es inmediato que

$$(T_n)_\# \mu = \nu \quad (9)$$

para todo $n \in \mathbb{N}$. Por el lema de Riemann–Lebesgue $(T_n)_n$ converge débilmente a $T \equiv 0$, con lo cual también converge en la topología débil de $L^p(\mu)$ para $1 \leq p < \infty$. Esto muestra que la igualdad (9) no se preserva por límites débiles, pues $T_\# \mu = \delta_0 \neq \nu$. En particular, se ve que $\mathcal{T}(\mu, \nu)$ no es cerrado por límites débiles. $\triangle$

Ejemplo 1.1.5 Veamos ahora un ejemplo para el cual (MP) es factible pero no alcanza mínimo. Para esto consideramos $X = Y = \mathbb{R}^2$, el costo $c(x, y) = |x - y|$ y las medidas

$$\mu = \mathcal{H}^1|_{\{0\} \times I} \quad \text{y} \quad \nu = \frac{1}{2}\mathcal{H}^1|_{\{-1\} \times I} + \frac{1}{2}\mathcal{H}^1|_{\{1\} \times I},$$

donde $I = [0, 1]$ y $\mathcal{H}^1$ es la medida de Hausdorff 1-dimensional. Es fácil ver que $C[T] \geq 1$ para $T \in \mathcal{T}(\mu, \nu)$, pues $\text{dist}(\text{sop}(\mu), \text{sop}(\nu)) = 1$. Por otra parte, es posible construir transportes con costo tan cercano a 1 como se desee. Para esto, tomar $n \in \mathbb{N}$, partir el segmento $\{0\} \times I$ en $2n$ segmentos contiguos

$$\{0\} \times I = S_1 \sqcup S_2 \sqcup \cdots \sqcup S_{2n-1} \sqcup S_{2n}$$

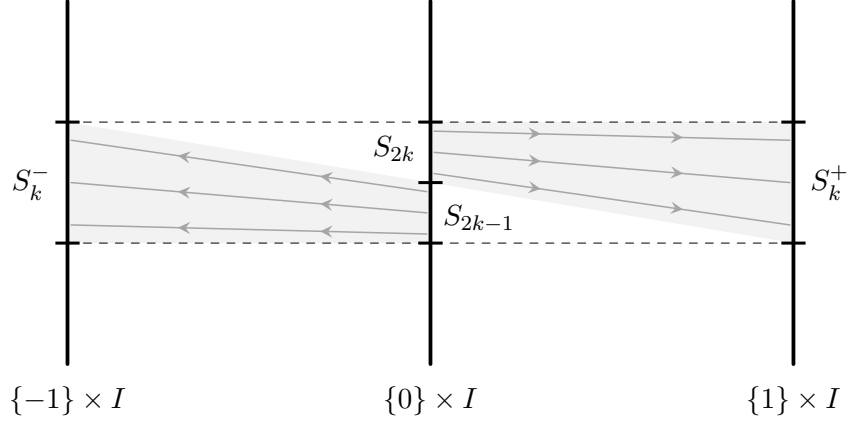


Figura 1.1: visualización del transporte T_n , las flechas grises indican el sentido de los rayos de transporte $[x, T_n(x)]$.

y partir los segmentos $\{-1\} \times I$ y $\{1\} \times I$ en n segmentos contiguos

$$\{-1\} \times I = S_1^- \sqcup S_2^- \sqcup \cdots \sqcup S_n^- \quad \text{y} \quad \{1\} \times I = S_1^+ \sqcup S_2^+ \sqcup \cdots \sqcup S_n^+.$$

Definiendo T_n de modo que

$$(T_n)_\#(\mu|_{S_{2k-1}^-}) = \nu|_{S_k^-} \quad \text{y} \quad (T_n)_\#(\mu|_{S_{2k}^+}) = \nu|_{S_k^+} \quad \forall k = 1, \dots, n$$

como se ve en la fig. 1.1, resulta

$$1 \leq C[T_n] \leq \sqrt{1 + (1/2n)^2} \xrightarrow{n} 1. \quad (10)$$

Notar que si C alcanzara mínimo sobre $\mathcal{T}(\mu, \nu)$ resultaría a partir de (10) que $C[T] = 1$ para algún transporte $T \in \mathcal{T}(\mu, \nu)$. Probaremos que $C[T] > 1$ para todo $T \in \mathcal{T}(\mu, \nu)$. Para esto notar que, si existe $T \in \mathcal{T}(\mu, \nu)$ con $C[T] = 1$, necesariamente verifica la condición

$$|T(x) - x| = 1 \text{ para } \mathcal{H}^1 \text{ casi todo } x \in \{0\} \times I.$$

Luego existe una partición $I = E_- \sqcup E_+ \sqcup Z$ verificando $\mathcal{H}^1[\{0\} \times Z] = 0$,

$$\{0\} \times E_- = T^{-1}(\{-1\} \times I) \quad \text{y} \quad \{0\} \times E_+ = T^{-1}(\{1\} \times I)$$

de modo tal que $T|_{\{0\} \times E_-}$ y $T|_{\{0\} \times E_+}$ son traslaciones horizontales. Así, resulta

$$\frac{1}{2} \mathcal{H}^1[E_-] = \nu[\{-1\} \times E_-] = \mu[T^{-1}(\{-1\} \times E_-)] = \mu[\{0\} \times E_-] = \mathcal{H}^1[E_-]$$

y razonando análogamente para E_+ se llega al absurdo $\mathcal{H}^1[I] = 2 \cdot \mathcal{H}^1[I]$. $\triangle$

1.1.2. Formulación del problema de Kantorovich

Una característica del problema de Monge es que no permite “dividir masa” para transportar μ en ν : para cada $x \in X$ toda la masa en x va a parar a $T(x)$. Una forma de relajar dicha restricción es permitiendo dividir masa, considerando un conjunto factible de medidas en lugar de un conjunto factible de funciones de transporte.

Definición 1.1.6 Sean X e Y dos espacios métricos, $\mu \in \mathcal{P}(X)$ y $\nu \in \mathcal{P}(Y)$. Decimos que una medida $\pi \in \mathcal{P}(X \times Y)$ es un **plan de transporte** de μ en ν si las medidas marginales de π son μ y ν , es decir, si para todo $E \in \mathcal{B}(X)$ y todo $F \in \mathcal{B}(Y)$ verifica

$$\mu[E] = \pi[E \times Y] \quad \text{y} \quad \nu[F] = \pi[X \times F].$$

Notamos $\mathcal{A}(\mu, \nu)$ al conjunto de todos los planes de transporte de μ en ν .

Lema 1.1.7 *Dados dos espacios métricos X e Y , notemos*

$$P^X : X \times Y \rightarrow X \quad \text{y} \quad P^Y : X \times Y \rightarrow Y$$

a las proyecciones canónicas y consideremos medidas de probabilidad $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y $\pi \in \mathcal{P}(X \times Y)$. Son equivalentes:

- i) la medida π es un elemento de $\mathcal{A}(\mu, \nu)$;*
- ii) valen las igualdades $P^X_\# \pi = \mu$ y $P^Y_\# \pi = \nu$;*
- iii) se verifica*

$$\int_X \varphi(x) d\mu(x) = \int_{X \times Y} \varphi(x) d\pi(x, y) \quad \text{y} \quad \int_Y \psi(y) d\nu(y) = \int_{X \times Y} \psi(y) d\pi(x, y)$$

para todo par de funciones $\varphi \in L^1(\mu)$, $\psi \in L^1(\nu)$.

Es fácil ver que dadas $\mu \in \mathcal{P}(X)$ y $\nu \in \mathcal{P}(Y)$ la medida producto $\mu \otimes \nu$ es siempre un elemento de $\mathcal{A}(\mu, \nu)$. Más aun, cada transporte $T \in \mathcal{T}(\mu, \nu)$ se corresponde con un plan de transporte $\pi_T \in \mathcal{A}(\mu, \nu)$ dado por $\pi_T = (I \otimes T)_\# \mu$ y concentrado sobre el gráfico de T .

Lema 1.1.8 *Sean X e Y espacios métricos y $\mu \in \mathcal{P}(X)$. Si $T : X \rightarrow Y$ es boreliana, entonces la medida π_T dada por $(I \otimes T)_\# \mu$ está concentrada en*

$$\Gamma(T) = \{(x, T(x)) \in X \times Y : x \in X\}.$$

Para toda $f : X \times Y \rightarrow \overline{\mathbb{R}}$ boreliana resulta $f \in L^1(\pi_T) \iff f \circ (I \otimes T) \in L^1(\mu)$ y vale

$$\int_X f(x, T(x)) d\mu(x) = \int_{X \times Y} f(x, y) d\pi_T(x, y).$$

Más aun, dada $\nu \in \mathcal{P}(Y)$ se tiene:

i) si $T \in \mathcal{T}(\mu, \nu)$, entonces $\pi_T \in \mathcal{A}(\mu, \nu)$;

ii) si $\pi \in \mathcal{A}(\mu, \nu)$ está concentrada en $\Gamma(T)$, entonces $T \in \mathcal{T}(\mu, \nu)$ y $\pi = \pi_T$.

En particular, para $T_1, T_2 \in \mathcal{B}(X, Y)$ resulta $\pi_{T_1} = \pi_{T_2}$ si y sólo si $T_1 = T_2$ $\mu - a.e.$

Una consecuencia inmediata del lema anterior es que para todo $T \in \mathcal{B}(X, Y)$ se tiene

$$C[T] = \int_X c(x, T(x)) d\mu(x) = \int_{X \times Y} c(x, y) d\pi_T(x, y). \quad (11)$$

Notar que el lado derecho de (11) tiene sentido para cualquier medida $\pi \in \mathcal{P}(X \times Y)$, mientras que el lado izquierdo tiene sentido sólo para funciones $T \in \mathcal{B}(X, Y)$. De este modo, el problema de Monge puede relajarse a un problema de minimización sobre $\mathcal{A}(\mu, \nu)$.

Problema de transporte óptimo de Kantorovich

Dados espacios métricos X, Y junto con medidas de probabilidad $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y un costo medible $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$, minimizar

$$I[\pi] = \int_{X \times Y} c(x, y) d\pi(x, y) \quad (\text{KP})$$

para π en el conjunto $\mathcal{A}(\mu, \nu)$ de los planes de transporte de μ en ν .

Como veremos en la próxima sección, el problema de Kantorovich resuelve todas las dificultades del problema de Monge: es un problema de minimización de un funcional lineal sobre un conjunto convexo y cerrado, es factible y siempre alcanza su valor óptimo.

1.1.3. El caso finito

Cuando μ y ν tienen soportes finitos el problema de Kantorovich para (X, Y, μ, ν, c) es un problema de programación lineal finita. Precisamente, si I y J son conjuntos finitos,

$$\mu = \sum_{i \in I} \mu_i \delta_{x_i} \quad \text{y} \quad \nu = \sum_{j \in J} \nu_j \delta_{y_j},$$

es fácil ver que una medida $\pi \in \mathcal{P}(X \times Y)$ está en $\mathcal{A}(\mu, \nu)$ si y sólo si tiene la forma

$$\pi = \sum_{i \in I} \sum_{j \in J} \pi_{ij} \delta_{(x_i, y_j)},$$

con pesos no negativos $(\pi_{ij})_{ij}$ satisfaciendo

$$\sum_{j \in J} \pi_{ij} = \mu_i \quad \text{para } i \in I \quad \text{y} \quad \sum_{i \in I} \pi_{ij} = \nu_j \quad \text{para } j \in J. \quad (12)$$

Así, identificando π con $(\pi_{ij})_{ij}$, el problema de Kantorovich toma la forma

$$\begin{aligned} \min_{(\pi_{ij})_{ij} \geq 0} \quad & \sum_{i \in I} \sum_{j \in J} \pi_{ij} c_{ij} \\ \text{s. a.} \quad & \sum_{j \in J} \pi_{ij} = \mu_i \quad \forall i \in I, \quad \sum_{i \in I} \pi_{ij} = \nu_j \quad \forall j \in J, \end{aligned} \tag{FP}$$

donde $c_{ij} = c(x_i, y_j)$ para todo $(i, j) \in I \times J$. La medida producto

$$\mu \otimes \nu = (\mu_i \nu_j)_{ij} = \sum_{ij} \mu_i \nu_j \delta_{(x_i, y_j)} \in \mathcal{A}(\mu, \nu)$$

muestra que el problema (FP) es factible. Más aun, $\mathcal{A}(\mu, \nu)$ es un subconjunto convexo y compacto de $[0, 1]^{I \times J}$, lo que inmediatamente implica la existencia de un minimizante π_* .

Con respecto al problema de Monge para (X, Y, μ, ν, c) , es evidente a partir del lema 1.1.3 que no siempre resultará factible. Como veremos a continuación, un caso particular en el que sí es factible se da cuando $|I| = |J| = N$ y $\mu_i = \nu_j = \frac{1}{N}$ para $(i, j) \in I \times J$. Más aún, en tal caso el problema de Monge admite un minimizante T_* y su valor óptimo coincide con el de Kantorovich.

Notar que así como identificamos $\mathcal{A}(\mu, \nu)$ con sucesiones doblemente indexadas y no negativas satisfaciendo la condición (12), podemos identificar $\mathcal{T}(\mu, \nu)$ con el subconjunto de las sucesiones que, adicionalmente, tienen un soporte con forma de gráfico:

$$\mathcal{T}(\mu, \nu) = \{ (\pi_{ij})_{ij} \in \mathcal{A}(\mu, \nu) : \text{para todo } i \in I \text{ existe a lo sumo un } j \in J \text{ con } \pi_{ij} > 0 \}.$$

Teorema 1.1.9 (Birkhoff) *Sea (X, Y, μ, ν, c) un problema de Monge–Kantorovich con la propiedad de que para algún $N \in \mathbb{N}$ existen $\{x_i\}_{i=1}^N \subset X$ e $\{y_j\}_{j=1}^N \subset Y$ tales que*

$$\mu = \sum_{i=1}^N \frac{1}{N} \delta_{x_i} \quad \text{y} \quad \nu = \sum_{j=1}^N \frac{1}{N} \delta_{y_j}.$$

Entonces $\mathcal{A}(\mu, \nu)$ es la cápsula convexa de $\mathcal{T}(\mu, \nu)$ y se verifica que:

- i) el conjunto $\mathcal{T}(\mu, \nu)$ sólo contiene permutaciones, esto es, medidas $(\pi_{ij})_{ij}$ tales que $\pi_{ij} = \frac{1}{N} \delta_{\sigma(i)j}$ para alguna biyección $\sigma : \llbracket 1, N \rrbracket \rightarrow \llbracket 1, N \rrbracket$;*
- ii) el problema (FP) alcanza un mínimo en $\mathcal{T}(\mu, \nu)$ y por lo tanto el problema de Monge asociado a (X, Y, μ, ν, c) admite un minimizante T_* definido μ - a.e. en la forma $T_*(x_i) = y_{\sigma(i)}$ para alguna biyección $\sigma : \llbracket 1, N \rrbracket \rightarrow \llbracket 1, N \rrbracket$.*

Es evidente que en el caso finito la dificultad no es teórica sino práctica, pues en las aplicaciones interesa calcular minimizantes de (FP). La bibliografía es extensa y se conocen algoritmos eficientes desde mediados del siglo pasado. Entre ellos podemos mencionar el

método simplex y su especialización network simplex [21, 92], una variedad de algoritmos combinatorios [2, 26] y el método de regularización entrópica [50, 105]. Cuando (FP) proviene de discretizar un problema continuo con cierta regularidad, resultan eficientes los métodos [102, 117] fundamentados en la idea de que el soporte de $(\pi_{ij})_{ij}$ se concentra en torno al gráfico de un transporte óptimo T_* para el problema continuo original.

1.1.4. La recta real

Cuando μ y ν son medidas en $\mathbb{R}$ es posible caracterizar de manera sencilla qué sucede para costos *convexos* de la forma $c(x, y) = \phi(|x - y|)$, donde $\phi : [0, \infty) \rightarrow [0, \infty)$ es una función convexa y no decreciente. A continuación presentamos los resultados estudiados en [124, Theorem 2.18, Remarks 2.19] y [4, Theorem 1.11, Remark 1.12].

Notando $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, la *función de distribución* de $\mu \in \mathcal{P}(\mathbb{R})$ es la función no decreciente $F : \overline{\mathbb{R}} \rightarrow [0, 1]$ dada por $F(x) = \mu[(-\infty, x]]$ y su *inversa* se define como la función $F^{-1} : [0, 1] \rightarrow \overline{\mathbb{R}}$ dada por $F^{-1}(t) = \inf \{x \in \mathbb{R} : F(x) > t\}$. Para $\pi \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$, la *función de distribución* de π es la función $F : \overline{\mathbb{R}} \times \overline{\mathbb{R}} \rightarrow [0, 1]$ dada por

$$F(x, y) = \pi[(-\infty, x] \times (-\infty, y]].$$

Teorema 1.1.10 *Consideremos un costo $c : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ de la forma $c(x, y) = \phi(|x - y|)$, con $\phi : [0, \infty) \rightarrow [0, \infty)$ convexa y no decreciente. Sean $\mu, \nu \in \mathcal{P}(\mathbb{R})$ medidas con funciones de distribución F, G y sea $\pi \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$ la única medida con función de distribución*

$$H(x, y) = \min \{F(x), G(y)\}.$$

Entonces:

i) *La medida π minimiza el problema de Kantorovich y vale*

$$I[\pi] = \int_0^1 c(F^{-1}(t), G^{-1}(t)) dt = \int_0^1 \phi(|F^{-1}(t) - G^{-1}(t)|) dt.$$

ii) *Si μ es no atómica existe una μ -a.e. única $T : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ no decreciente con $T_{\#}\mu = \nu$. El transporte T resulta ser un minimizante para problema de Monge si $C[T] < \infty$ y es el único minimizante si, adicionalmente, ϕ es estrictamente convexa. Más aun, en tal caso $\pi = \pi_T$ es el único minimizante para el problema de Kantorovich.*

Ejemplo 1.1.11 Cuando ϕ no es estrictamente convexa en general no existe un único transporte óptimo. Para verlo, considerar las medidas $\mu = \mathcal{L}^1|_{[0,1]}$, $\nu = \mathcal{L}^1|_{[1,2]}$ y notar que $T(x) = x + 1$ y $S(x) = 2 - x$ transportan μ en ν . Si se toma $\phi(t) = t$ y por lo tanto $c(x, y) = |x - y|$, entonces tanto T como S minimizan (MP): T es minimizante por el teorema anterior y S lo es porque $C[T] = 1 = C[S]$. $\triangle$

1.1.5. El caso cuadrático

Para problemas en $\mathbb{R}^n$ con costo cuadrático $c(x, y) = \frac{1}{2}|x - y|^2$ el transporte óptimo está caracterizado por potenciales convexos [25, 85, 95]. Precisamente, McCann probó en [95] que el transporte óptimo de μ en ν en el sentido de Kantorovich está inducido por el gradiente ∇P de un potencial convexo P si μ no da masa a *conjuntos pequeños* (i.e. con dimensión de Hausdorff a lo sumo $n - 1$), siendo P único salvo constantes aditivas. Más aún, cuando μ y ν tienen segundos momentos finitos, i.e.

$$\int_{\mathbb{R}^n} |x|^2 d\mu(x) < \infty \quad \text{y} \quad \int_{\mathbb{R}^n} |y|^2 d\nu(y) < \infty,$$

puede darse una caracterización precisa de la optimalidad en términos de subdiferenciales de funciones convexas. A continuación introducimos brevemente las definiciones necesarias para enunciar dichos resultados. Las reformulaciones de los resultados de Knott–Smith [85] y Brenier [25] fueron tomadas de [124]. Las pruebas omitidas y detalles adicionales pueden leerse en [4, 23, 116, 124] y en las referencias allí citadas.

Una función $P : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ se dice *propia* si no es idénticamente infinita, es decir, si su *dominio* $\text{dom}(P) = P^{-1}(\mathbb{R})$ es no vacío. Cuando P es convexa también es convexo $\text{dom}(P)$ y en su interior P resulta continua, localmente lipschitz y por lo tanto diferenciable en casi todo punto.

Para $P : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ definimos su *subdiferencial* en $x \in \mathbb{R}^n$ como el conjunto

$$\partial P(x) = \{ y \in \mathbb{R}^n : P(z) \geq P(x) + \langle y, z - x \rangle \quad \forall z \in \mathbb{R}^n \} \quad (13)$$

y llamamos *subdiferencial* de P a la aplicación multivaluada $\partial P : \mathbb{R}^n \rightarrow \mathbb{R}^n$ dada por (13). Para P propia y convexa se tiene $\partial P(x) \neq \emptyset$ sii $x \in \text{dom}(P)$ y además vale

$$P \text{ es diferenciable en } x \iff \#(\partial P(x)) = 1 \iff \partial P(x) = \{\nabla P(x)\}.$$

La *conjugada de Legendre* de P es la función $P^* : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ dada por

$$P^*(y) = \sup_{x \in \mathbb{R}^n} \langle x, y \rangle - P(x),$$

que resulta convexa y semicontinua inferiormente. Dos funciones convexas P, Q se dicen *conjugadas* entre sí si verifican $P^* = Q$ y $Q^* = P$. Cuando P es propia y convexa también lo es P^* y ambas verifican la *desigualdad de Young*

$$P(x) + P^*(y) \geq \langle x, y \rangle \quad \text{para todo } (x, y) \text{ en } \mathbb{R}^n \times \mathbb{R}^n.$$

Si además P es semicontinua inferiormente, se verifica

$$P(x) + P^*(y) = \langle x, y \rangle \iff y \in \partial P(x) \iff x \in \partial P^*(y)$$

para todo par (x, y) en $\mathbb{R}^n \times \mathbb{R}^n$. Más aun, el *teorema de biconjugación de Fenchel* asegura que, si es P propia, son equivalentes:

- i) P es convexa y semicontinua inferiormente;
- ii) P^* es propia y vale $P = P^{**}$.

Teorema 1.1.12 (Knott–Smith) Para $X = Y = \mathbb{R}^n$, sea $c(x, y) = \frac{1}{2}|x - y|^2$ el costo cuadrático y sean $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ medidas con segundos momentos finitos. Entonces, para el problema (X, Y, μ, ν, c) vale que:

- i) Una medida $\pi \in \mathcal{A}(\mu, \nu)$ minimiza el problema de Kantorovich sii existe una función P convexa y semicontinua inferiormente verificando $\text{sop}(\pi) \subset \Gamma(\partial P)$.
- ii) Si P es convexa, semicontinua inferiormente y verifica $\nu = (\nabla P)_\# \mu$, entonces el problema de Kantorovich admite un único minimizante y está inducido por ∇P , el único minimizante para el problema de Monge.

Teorema 1.1.13 (Brenier) Con las hipótesis del teorema anterior, supongamos que μ no da masa a conjuntos pequeños. Entonces el problema de Kantorovich para (X, Y, μ, ν, c) admite un único minimizante y está inducido por el gradiente de una función P convexa y semicontinua inferiormente.

Si, adicionalmente, ν tampoco da masa a conjuntos pequeños, entonces ∇P^* induce el único minimizante para el problema de Kantorovich (Y, X, ν, μ, c) y se verifica

$$\nabla P^* \circ \nabla P(x) = x \quad \mu\text{-a.e.} \quad y \quad \nabla P \circ \nabla P^*(y) = y \quad \nu\text{-a.e.}$$

1.2. Minimizantes para el problema de Kantorovich

Formulamos el problema de Kantorovich como una relajación del problema de Monge y vimos que siempre admite soluciones factibles. Probaremos ahora que siempre existen minimizantes para el problema de Kantorovich bajo hipótesis generales. Las definiciones y resultados previos necesarios para esta sección pueden leerse en el apéndice A.1.

Lema 1.2.1 Sean Z y W espacios métricos y $f : Z \rightarrow W$ una función continua. Si $(\sigma_n)_n \subset \mathcal{P}(Z)$ converge débilmente a $\sigma \in \mathcal{P}(Z)$, entonces $(f_\# \sigma_n)_n \subset \mathcal{P}(W)$ converge débilmente a $f_\# \sigma \in \mathcal{P}(W)$.

Demostración. Sea $\phi \in C_b(W)$. Por el lema 1.1.2, se tiene

$$\int_W \phi(w) d(f_\# \sigma_n)(w) = \int_Z (\phi \circ f)(z) d\sigma_n(z), \quad (14)$$

$$\int_W \phi(w) d(f_\# \sigma)(w) = \int_Z (\phi \circ f)(z) d\sigma(z). \quad (15)$$

Basta ver que el lado derecho de (14) converge al lado derecho de (15), lo cual sucede por hipótesis al ser $\phi \circ f$ un elemento de $C_b(Z)$. $\square$

Lema 1.2.2 Sean X e Y espacios métricos, $\mu \in \mathcal{P}(X)$ y $\nu \in \mathcal{P}(Y)$. Si las familias de medidas $\mathcal{F}_1 \subset \mathcal{P}(X)$ y $\mathcal{F}_2 \subset \mathcal{P}(Y)$ son rígidas, entonces también es rígida

$$\mathcal{A}(\mathcal{F}_1, \mathcal{F}_2) = \{ \pi \in \mathcal{P}(X \times Y) : P_{\#}^X \pi \in \mathcal{F}_1 \text{ y } P_{\#}^Y \pi \in \mathcal{F}_2 \}.$$

Demostración. Sean $\varepsilon > 0$ y $K_1 \subset X$, $K_2 \subset Y$ compactos tales que $\mu[X \setminus K_1] < \frac{\varepsilon}{2}$ y $\nu[Y \setminus K_2] < \frac{\varepsilon}{2}$ para toda $\mu \in \mathcal{F}_1$ y toda $\nu \in \mathcal{F}_2$. A partir de la inclusión

$$(X \times Y) \setminus (K_1 \times K_2) \subset (X \setminus K_1 \times Y) \cup (X \times Y \setminus K_2),$$

para toda $\pi \in \mathcal{A}(\mathcal{F}_1, \mathcal{F}_2)$ resulta

$$\pi[X \times Y \setminus K_1 \times K_2] \leq \pi[X \setminus K_1 \times Y] + \pi[X \times Y \setminus K_2] = \mu[X \setminus K_1] + \nu[Y \setminus K_2] < \varepsilon. \quad \square$$

Lema 1.2.3 Sean X e Y espacios métricos y $(\pi_n)_n \subset \mathcal{P}(X \times Y)$ tal que $\pi_n \in \mathcal{A}(\mu_n, \nu_n)$ para todo $n \in \mathbb{N}$. Valen las siguientes afirmaciones:

- i) Si $\pi_n \rightharpoonup \pi \in \mathcal{P}(X \times Y)$, entonces $\pi \in \mathcal{A}(\mu, \nu)$ sii $\mu_n \rightharpoonup \mu$ y $\nu_n \rightharpoonup \nu$.
- ii) Si X e Y son polacos y $(\mu_n)_n, (\nu_n)_n$ convergen débilmente a μ, ν respectivamente, entonces existe una subsucesión $(\pi_{n_k})_k$ débilmente convergente a $\pi \in \mathcal{A}(\mu, \nu)$.

Demostración. La continuidad de las proyecciones canónicas y el lema 1.2.1 implican

$$\mu_n = P_{\#}^X \pi_n \rightharpoonup P_{\#}^X \pi \quad \text{y} \quad \nu_n = P_{\#}^Y \pi_n \rightharpoonup P_{\#}^Y \pi.$$

Este hecho, junto con la unicidad de los límites débiles, prueba i). Para ii), notar que

$$\mathcal{F}_1 = \{\mu\} \cup \{\mu_n : n \in \mathbb{N}\} \quad \text{y} \quad \mathcal{F}_2 = \{\nu\} \cup \{\nu_n : n \in \mathbb{N}\}$$

son familias compactas por las convergencias de $(\mu_n)_n$ a μ y de $(\nu_n)_n$ a ν . Luego, $\mathcal{F}_1$ y $\mathcal{F}_2$ son rígidas por el teorema de Prokhorov A.3.11 y por lo tanto también es rígida $\mathcal{A}(\mathcal{F}_1, \mathcal{F}_2)$ por el lema 1.2.2. Dado que $(\pi_n)_n$ está contenida en $\mathcal{A}(\mathcal{F}_1, \mathcal{F}_2)$, el teorema de Prokhorov implica la existencia de una subsucesión $(\pi_{n_k})_k$ débilmente convergente a π . Como $\mu_{n_k} \rightharpoonup \mu$ y $\nu_{n_k} \rightharpoonup \nu$ por hipótesis, i) implica $\pi \in \mathcal{A}(\mu, \nu)$. $\square$

Proposición 1.2.4 Sean X e Y espacios métricos polacos y sea $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ acotada por abajo y semicontinua inferiormente. Dadas $\mu \in \mathcal{P}(X)$ y $\nu \in \mathcal{P}(Y)$, existe un minimizante para el problema de Kantorovich.

Demostración. Basta ver que $\mathcal{A}(\mu, \nu)$ es compacto para la topología de la convergencia débil en $\mathcal{P}(X \times Y)$ y que el operador I es semicontinuo inferiormente. Para ver que $\mathcal{A}(\mu, \nu)$

es compacto notar que, para toda $(\pi_n)_n \subset \mathcal{A}(\mu, \nu)$, la parte *ii*) del lema 1.2.3 garantiza la existencia de una subsucesión $(\pi_{n_k})_k$ convergente a un elemento de $\mathcal{A}(\mu, \nu)$.

Para ver que I es semicontinuo inferiormente considerar $(\pi_n)_n \subset \mathcal{A}(\mu, \nu)$ con $\pi_n \rightarrow \pi$. Por ser c semicontinua y acotada inferiormente existe sucesión $(c_k)_k$ de funciones continuas y acotadas tales que $c_k \nearrow c$. Para cada n tenemos

$$\int_{X \times Y} c_k d\pi_n \leq \int_{X \times Y} c d\pi_n$$

y, tomando $\liminf_n$ a ambos lados, $\pi_n \rightarrow \pi$ implica

$$\int_{X \times Y} c_k d\pi \leq \liminf_n \int_{X \times Y} c d\pi_n.$$

Como $\pi[X \times Y]$ es finita, cada c_k está en $L^1(\pi)$ y por convergencia monótona resulta

$$I[\pi] = \int_{X \times Y} c d\pi \leq \liminf_n \int_{X \times Y} c d\pi_n = \liminf_n I[\pi_n]. \quad \square$$

1.2.1. Distancias de Wasserstein

La existencia de minimizantes para el problema de Kantorovich permite metrizar la topología de la convergencia débil de medidas con una métrica definida en términos del transporte óptimo. Para esto, dado un espacio métrico polaco (Z, d_Z) , consideramos el conjunto de medidas en $\mathcal{P}(Z)$ con p -momentos finitos

$$\mathcal{P}_p(Z) = \left\{ \pi \in \mathcal{P}(Z) : \int_Z d_Z^p(z, z_0) d\pi(z) < \infty \text{ para algún } z_0 \in Z \right\}, \quad (16)$$

donde $p \in [1, \infty)$. Es fácil ver que (16) es independiente de la elección del punto $z_0 \in Z$, que $\mathcal{P}_{p_2}(Z) \subset \mathcal{P}_{p_1}(Z)$ si $p_1 \leq p_2$ y que $\mathcal{P}_p(Z) = \mathcal{P}(Z)$ si d_Z es acotada.

Definición 1.2.5 Para $p \in [1, \infty)$ se define la *distancia de Wasserstein* W_p por

$$W_p(\pi_1, \pi_2) = \left(\min_{\sigma \in \mathcal{A}(\pi_1, \pi_2)} \int_{Z \times Z} d_Z^p(z_1, z_2) d\sigma(z_1, z_2) \right)^{1/p} \quad (17)$$

para $\pi_1, \pi_2 \in \mathcal{P}(Z)$.

Notar que la función W_p definida por (17) no es necesariamente finita, pero resulta a partir de (16) que W_p sí es finita si se la restringe a $\mathcal{P}_p(Z) \times \mathcal{P}_p(Z)$. Se verifica además que W_p define una métrica en $\mathcal{P}_p(Z)$ y que ésta metriza la topología de la convergencia débil en $\mathcal{P}(Z)$ si d_Z es acotada. En la siguiente proposición registramos todos estos hechos, cuyas pruebas pueden leerse en [124, Cap. 7].

Proposición 1.2.6 Sea (Z, d_Z) un espacio métrico polaco y sea $p \in [1, \infty)$. Entonces:

- i) La función W_p dada por (17) es finita en $\mathcal{P}_p(Z)$ y define allí una métrica.
- ii) Si $(\pi_k)_k \subset \mathcal{P}_p(Z)$ y $\pi \in \mathcal{P}(Z)$ verifican $W_p(\pi_k, \pi) \xrightarrow{k} 0$, entonces $\pi \in \mathcal{P}_p(Z)$ y la sucesión $(\pi_k)_k$ converge débilmente a π .
- iii) Si d_Z es una métrica acotada, resulta $\mathcal{P}(Z) = \mathcal{P}_p(Z)$ y W_p define una métrica en $\mathcal{P}(Z)$ que metriza la topología de la convergencia débil.

En particular, se deduce a partir de iii) que si (Z, d_Z) es compacto entonces

$$W_p(\pi_k, \pi) \xrightarrow{k} 0 \iff \pi_k \xrightarrow{k} \pi$$

para toda $(\pi_k)_k \subset \mathcal{P}(Z)$, toda $\pi \in \mathcal{P}(Z)$ y todo $p \in [1, \infty)$.

1.3. Dualidad de Kantorovich

Una de las buenas propiedades de la formulación de Kantorovich para el problema de transporte óptimo es que admite una formulación dual. Más aun, dicha formulación resulta útil para estudiar el problema de Monge asociado. A continuación presentamos el problema dual de Kantorovich y definimos la noción de c -concavidad, que debe entenderse como una abstracción formal de la noción de convexidad usual para espacios lineales.

Definimos el funcional $J : L^1(\mu) \times L^1(\nu) \rightarrow \mathbb{R}$ por

$$J[\varphi, \psi] = \int_X \varphi d\mu + \int_Y \psi d\nu.$$

Notamos $\mathcal{F}_c$ al conjunto de pares $(\varphi, \psi) \in L^1(\mu) \times L^1(\nu)$ tales que

$$\varphi(x) + \psi(y) \leq c(x, y) \tag{18}$$

para μ -c.t. $x \in X$ y ν -c.t. $y \in Y$. Notamos $\mathcal{F}_c \cap C_b$ al conjunto de pares (φ, ψ) de $\mathcal{F}_c$ con φ y ψ continuas y acotadas.

Llamamos **problema primal de Kantorovich** al problema

$$\begin{array}{ll} \text{minimizar} & I[\pi] \\ \text{sujeto a} & \pi \in \mathcal{A}(\mu, \nu) \end{array} \tag{KP}$$

y **problema dual de Kantorovich** al problema

$$\begin{array}{ll} \text{maximizar} & J[\varphi, \psi] \\ \text{sujeto a} & (\varphi, \psi) \in \mathcal{F}_c. \end{array} \tag{DP}$$

Notar que, en principio, el mínimo y el máximo de los problemas primal y dual podrían ser sólo ínfimo y supremo respectivamente. Es fácil ver que si c es una función de costo medible entonces (18) implica inmediatamente

$$J[\varphi, \psi] \leq I[\pi]$$

para todo $(\varphi, \psi) \in \mathcal{F}_c$ y toda $\pi \in \mathcal{A}(\mu, \nu)$. Este tipo de desigualdad se conoce con el nombre de *principio débil de dualidad*, pues implica que los valores óptimos verifican $J_* \leq I_*$, permitiendo en principio la posibilidad de que el *gap de dualidad* $I_* - J_*$ sea estrictamente positivo. A continuación enunciamos un *principio fuerte de dualidad* para costos semicontinuos.

Teorema 1.3.1 (Dualidad de Kantorovich I) Sean X e Y espacios polacos dotados de medidas borelianas de probabilidad $\mu \in \mathcal{P}(X)$ y $\nu \in \mathcal{P}(Y)$. Sea $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ una función de costo semicontinua inferiormente y acotada inferiormente. Entonces

$$\sup_{(\varphi, \psi) \in \mathcal{F}_c} J[\varphi, \psi] = \inf_{\pi \in \mathcal{A}(\mu, \nu)} I[\pi]. \quad (19)$$

Adicionalmente, el ínfimo de (19) es un mínimo y el supremo permanece inalterado si se toma sobre $\mathcal{F}_c \cap C_b$.

Para refinar el enunciado del teorema 1.3.1 vamos a definir la noción de c -conjugación, la cual será útil para seleccionar maximizantes de J con buenas propiedades.

Definición 1.3.2 Dada una función $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$, para cada $\varphi : X \rightarrow \mathbb{R} \cup \{-\infty\}$ no constantemente $-\infty$ definimos su **c -conjugada** (o **c -transformada**) como

$$\varphi^c(y) = \inf_{x \in X} c(x, y) - \varphi(x) \quad , \quad \forall y \in Y.$$

Similarmente, para cada función $\psi : Y \rightarrow \mathbb{R} \cup \{-\infty\}$ no constantemente infinita definimos su **c -conjugada** (o **c -transformada**) como

$$\psi^c(x) = \inf_{y \in Y} c(x, y) - \psi(y) \quad , \quad \forall x \in X.$$

Notar que cuando la suma $\varphi(x) + \varphi^c(y)$ no se indefine tenemos

$$\varphi(x) + \varphi^c(y) \leq c(x, y) \quad (20)$$

y que, si inicialmente se tiene $\varphi(x) + \psi(y) \leq c(x, y)$ para todo (x, y) , entonces

$$\psi \leq \varphi^c \quad \text{y} \quad \varphi \leq \psi^c.$$

Definición 1.3.3 Una función $\varphi : X \rightarrow \mathbb{R} \cup \{-\infty\}$ se dice **c -cóncava** si existe alguna función $\psi : Y \rightarrow \mathbb{R} \cup \{-\infty\}$, $\psi \not\equiv -\infty$, tal que $\varphi = \psi^c$. Análogamente, una función

$\psi : Y \rightarrow \mathbb{R} \cup \{-\infty\}$ se dice ***c*-cóncava** si existe $\varphi : X \rightarrow \mathbb{R} \cup \{-\infty\}$, $\varphi \not\equiv -\infty$, tal que $\psi = \varphi^c$. Un par de funciones ***c*-cóncavas conjugadas** es un par (φ, ψ) de funciones *c*-cóncavas como antes tales que $\varphi = \psi^c$ y $\psi = \varphi^c$.

Definición 1.3.4 Decimos que una función *c*-cóncava φ es un **potencial *c*-cóncavo** para un problema de Kantorovich si (φ, φ^c) es un maximizante para su problema dual. Otros nombres posibles son **potencial de Kantorovich** o simplemente **potencial**, cuando no hay ambigüedad sobre el costo *c*.

A continuación enunciamos dos resultados que ilustran la utilidad de las nociones de *c*-concavidad y *c*-conjugación. El primer resultado es un refinamiento del teorema de dualidad 1.3.1, garantizando que el problema dual de Kantorovich alcanza máximo en un par *c*-conjugado. El segundo resultado es una caracterización de la optimalidad para soluciones factibles del problema primal de Kantorovich. Para ambos resultados es necesario que el costo *c* cumpla una cierta condición de integrabilidad que definimos a continuación.

Definición 1.3.5 Dado un problema (X, Y, μ, ν, c) , decimos que *c* verifica la **hipótesis de integrabilidad** si existen funciones no negativas $c_X \in L^1(\mu)$ y $c_Y \in L^1(\nu)$ tales que

$$c(x, y) \leq c_X(x) + c_Y(y) \quad (21)$$

para todo (x, y) en $X \times Y$.

Observación 1.3.6 Es inmediato que la desigualdad (21) implica $I[\pi] < +\infty$ para todo plan $\pi \in \mathcal{A}(\mu, \nu)$ y que toda función de costo *c* acotada superiormente verifica la hipótesis de integrabilidad. En particular, *c* verifica la hipótesis de integrabilidad si es continuo y *X* e *Y* son compactos. Un caso importante es el de los costos $c_p(x, y) = \frac{1}{p}|x - y|^p$ para problemas con $X, Y \subset \mathbb{R}^n$. Dado que para todo $p > 0$ existe una constante $A(p)$ tal que

$$c_p(x, y) = \frac{1}{p}|x - y|^p \leq A(p)[|x|^p + |y|^p] \text{ para } x, y \in \mathbb{R}^n,$$

resulta que c_p verifica la hipótesis de integrabilidad si μ, ν tienen *p*-momentos finitos, i.e.

$$\int_X |x|^p d\mu(x) < \infty \quad \text{y} \quad \int_Y |y|^p d\nu(y) < \infty.$$

Teorema 1.3.7 (Dualidad de Kantorovich II) Sean *X, Y* espacios polacos dotados de medidas borelianas de probabilidad $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y supongamos que el costo $c : X \times Y \rightarrow \mathbb{R}$ es semicontinuo inferiormente, acotado inferiormente y verifica la hipótesis de integrabilidad. Entonces

$$\min_{\pi \in \mathcal{A}(\mu, \nu)} I[\pi] = \max \left\{ J[\varphi, \varphi^c] : (\varphi, \varphi^c) \in L^1(\mu) \times L^1(\nu) \text{ y } \varphi \text{ } c\text{-cóncava} \right\}. \quad (22)$$

En particular, existen un par de funciones *c*-conjugadas $(\varphi_*, \psi_*) \in L^1(\mu) \times L^1(\nu)$ y una

medida boreliana de probabilidad $\pi_* \in \mathcal{A}(\mu, \nu)$ tales que

$$I[\pi_*] = J[\varphi_*, \psi_*].$$

Proposición 1.3.8 (Caracterización de soluciones óptimas) Sean X, Y espacios polacos dotados de medidas borelianas de probabilidad $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ y supongamos que $c : X \times Y \rightarrow \mathbb{R}$ es semicontinua inferiormente, acotada inferiormente y verifica la hipótesis de integrabilidad. Si $(\varphi, \psi) \in \mathcal{F}_c$ es un maximizante para el problema dual de Kantorovich, entonces $\pi \in \mathcal{A}(\mu, \nu)$ es un minimizante para el problema primal de Kantorovich si y sólo si π está concentrada en

$$\left\{ (x, y) \in X \times Y : \varphi(x) + \psi(y) = c(x, y) \right\}. \quad (23)$$

Las referencias para esta sección son [39, 116, 125]. Para una demostración del teorema de dualidad 1.3.1 pueden verse las pruebas de [116, Theorem 1.42] o [125, Theorem 5.10 (i)]. Para una demostración del teorema de dualidad 1.3.7 puede verse la prueba de [125, Theorem 5.10 (iii)]. Las definiciones de c -concavidad y c -transformada elegidas para formular 1.3.7 son levemente distintas a las elegidas en [125] e idénticas a las elegidas en [116, Definition 1.10], aunque aquí tomamos la simplificación notacional propuesta en [39, Section 1.4].

1.4. Minimizantes para el problema de Monge

El ejemplo 1.1.5 muestra que no siempre existen transportes óptimos para el problema de Monge asociado a (X, Y, μ, ν, c) , aun si éste es factible. Por otra parte, en la sección 1.1 se discutieron diferentes clases de problemas para las que sí existen minimizantes en el sentido de Monge. En esta sección repasamos algunos resultados generales en esa dirección. Mencionamos resultados de existencia y unicidad para costos cóncavos y convexos en $\mathbb{R}^n$ [33, 76, 77, 106], para costos generales en $\mathbb{R}^n$ [38, 54, 87, 93] y para costos generales en variedades riemannianas [17, 60, 63, 96, 125]. Finalmente presentamos resultados de existencia para el caso de costos dados por una norma en $\mathbb{R}^n$ o por la distancia geodésica en una variedad riemanniana [7, 8, 35, 37, 43, 44, 59, 62, 120, 123]

Empezamos enunciando resultados para costos convexos y costos cóncavos en $\mathbb{R}^n$. Por costos (*estrictamente*) *convexos* nos referimos a costos de la forma $c(x, y) = h(x - y)$ con h (*estrictamente*) convexa, definiendo de manera análoga el caso cóncavo. Por ejemplo, los costos de la forma $c(x, y) = |x - y|^p$ son convexos para $p \geq 1$, cóncavos para $0 < p \leq 1$ y lo son estrictamente si $p \neq 1$.

Teorema 1.4.1 (116, Theorem 1.17 + Theorem 3.26) Sean $X, Y \subset \mathbb{R}^n$ compactos y sea un costo $c(x, y) = h(x - y)$ con $h : \mathbb{R}^n \rightarrow \mathbb{R}$. Sean $\nu \in \mathcal{P}(Y)$ arbitraria y $\mu \in \mathcal{P}(X)$ absolutamente continua. Supongamos, adicionalmente, alguna de estas dos hipótesis:

i) La función h es estrictamente convexa y de clase C^1 en $\mathbb{R}^n$.

ii) Se verifica $\text{sop}(\mu) \cap \text{sop}(\nu) = \emptyset$ y la función h es de la forma $h(z) = l(|z|)$ con l estrictamente cóncava y creciente en $[0, +\infty)$ siendo de clase C^1 en $(0, +\infty)$.

Entonces existe un único minimizante π_* para el problema de Kantorovich y está dado por un transporte T_* . Más aun, existe un potencial de Kantorovich φ tal que μ -a.e. vale

$$T_*(x) = x - (\nabla h)^{-1}(\nabla \varphi(x)).$$

La idea central de la demostración del teorema 1.4.1 consiste en tomar un maximizante c -conjugado (φ, ψ) para (DP) y probar que todo minimizante π de (KP) verifica $\pi = \pi_T$ para un cierto transporte T expresable en términos de φ e independiente de π . Para esto, notar que todo punto $(x_0, y_0) \in \text{sop}(\pi)$ verifica

$$\begin{aligned} \psi(y_0) &= c(x_0, y_0) - \phi(x_0), \\ \psi(y_0) &\leq c(x, y_0) - \phi(x) \quad \text{para todo } x \in X. \end{aligned} \tag{24}$$

Dejando de lado detalles de regularidad, se concluye a partir de (24) que

$$\nabla_x c(x_0, y_0) = \nabla \phi(x_0),$$

con lo cual $T(x_0) := y_0$ queda unívocamente determinado si $\nabla_x c(x_0, \cdot)$ es inyectiva. Al ser $(x_0, y_0) \in \text{sop}(\pi)$ arbitrario, esto define la T candidata sobre $\text{sop}(\mu)$.

La idea original de la demostración se remonta a Brenier [25] y fue luego generalizada por Caffarelli [33], Gangbo y McCann [76, 77] y otros [38, 87, 93, 123]. En el caso del teorema 1.4.1, al suponer $c(x, y) = h(x - y)$, la inyectividad de $\nabla_x c(x_0, \cdot) = \nabla h(x_0 - \cdot)$ es consecuencia de la convexidad o concavidad estricta de h . En particular, cuando c es el costo cuadrático, $h(x) = \frac{1}{2}|x|^2$ y por lo tanto $\nabla h = \text{Id}_{\mathbb{R}^n}$. En el caso general debe requerirse como hipótesis que $\nabla_x c(x_0, \cdot)$ sea inyectiva. Vale aclarar que si X e Y son dominios buenos de $\mathbb{R}^n$ dicha hipótesis es consecuencia de la condición $\det \left(\frac{\partial^2 c}{\partial y_j \partial x_i} \right) \neq 0$, que en algunos contextos se conoce como condición de Spence-Mirrlees.

Teorema 1.4.2 (38, Theorem 1) Sean $X \subset \mathbb{R}^n$ un subconjunto compacto, Y un espacio métrico compacto y $c : X \times Y \rightarrow \mathbb{R}$ un costo lipschitz. Sean $\nu \in \mathcal{P}(Y)$ arbitraria y $\mu \in \mathcal{P}(X)$ absolutamente continua con $\mu[\partial X] = 0$. Supongamos que $c(\cdot, y)$ es diferenciable en X° para todo $y \in Y$ y que $\nabla_x c(x, \cdot)$ es inyectiva para $x \in X^\circ$, precisamente,

$$(\forall x \in X^\circ) : (\forall y_1, y_2 \in Y) : \nabla_x c(x, y_1) = \nabla_x c(x, y_2) \implies y_1 = y_2.$$

Entonces existe un único minimizante π_* para el problema de Kantorovich y está dado por

un transporte T_* . Más aun, existe un potencial de Kantorovich φ tal que μ -a.e. vale

$$T_*(x) = (\nabla_x c(x, \cdot))^{-1}(\nabla \varphi(x)).$$

Teorema 1.4.3 (96, Theorem 9) Sea (M, g) una variedad riemanniana suave, conexa, compacta y sin borde. Sean $X = Y = M$ y sea el costo cuadrático $c(x, y) = \frac{1}{2}d^2(x, y)$, donde d es la distancia geodésica en M . Sean $\nu \in \mathcal{P}(Y)$ arbitraria y $\mu \in \mathcal{P}(X)$ absolutamente continua respecto de la medida de volumen vol_M . Entonces existe un único minimizante π_* para el problema de Kantorovich y está dado por un transporte T_* . Más aun, existe un potencial de Kantorovich φ tal que μ -a.e. vale

$$T_*(x) = \exp_x(-\nabla \varphi(x)),$$

donde $\exp_x : T_x X \rightarrow X$ es la aplicación exponencial usual.

Finalizamos esta sección presentando algunos resultados para problemas de Monge con costo dado por una distancia. Bajo hipótesis adecuadas es posible probar la existencia de transportes óptimos para estos casos, aunque en general falla la unicidad al igual que en el ejemplo 1.1.11. En espacios euclídeos se sabe que existen transportes óptimos si c está dado por una norma. Fue probado primero para la norma euclídea [59, 120, 123] y luego para normas uniformemente convexas [8, 35], cristalinas [7], estrictamente convexas [37, 43] y, finalmente, arbitrarias [44]. Para el caso de una variedad riemanniana con el costo c dado por la distancia geodésica la existencia fue probada en [62].

Teorema 1.4.4 (44, Theorem 1.1) Sean $X = Y = \mathbb{R}^n$ y sean $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ medidas de soportes compactos con μ absolutamente continua. Si $c(x, y) = \|x - y\|$ está dado por una norma general $\|\cdot\|$, entonces el problema de Kantorovich admite un minimizante π_* inducido por un transporte T_* que minimiza el problema de Monge.

Teorema 1.4.5 (62, Theorem 1) Sea (M, g) una variedad riemanniana suave, conexa y geodésicamente completa. Sean $X = Y = M$ y sea el costo $c(x, y) = d(x, y)$ dado por la distancia geodésica en M . Sean $\mu, \nu \in \mathcal{P}(M)$ medidas con soportes compactos y absolutamente continuas respecto de la medida de volumen vol_M . Entonces existe un minimizante π_* para el problema de Kantorovich dado por un transporte T_* que minimiza el problema de Monge.

1.5. Las hipótesis (KU) y (SU)

Existen resultados razonablemente generales que garantizan la hipótesis (SU), de la cual se deduce inmediatamente (KU). Sirven como ejemplos aquéllos presentados en la sección anterior, a excepción de los últimos dos. Como veremos en esta sección la hipótesis (SU) es más fuerte que la existencia minimizantes únicos para los problemas de Monge

y Kantorovich. Para estudiar las relaciones entre todas estas propiedades de unicidad definimos las hipótesis auxiliares:

(MM) Existe un minimizante para el problema de Monge.

(MU) Existe un único minimizante para el problema de Monge.

Es fácil ver que la conjunción de (KU) y (MU) implica (SU) bajo la hipótesis adicional de que los valores óptimos coinciden, es decir, bajo la hipótesis

$$(EQ) \quad \inf_{T \in \mathcal{T}(\mu, \nu)} C[T] = \min_{\pi \in \mathcal{A}(\mu, \nu)} I[\pi].$$

En general, la forma de probar que un problema satisface la hipótesis (SU) es probando primero que todo minimizante π del problema de Kantorovich está concentrado sobre el gráfico de alguna función boreliana T , de modo que $\pi = \pi_T$ y por lo tanto $\pi[\Gamma(T)] = 1$. Notamos (CG) a dicha hipótesis. Precisamente:

(CG) Para todo minimizante π de (KP) existe $T \in \mathcal{B}(X, Y)$ con $\pi = \pi_T$.

Resumimos las relaciones entre todas estas propiedades en el siguiente lema.

Lema 1.5.1 *Sea el problema (X, Y, μ, ν, c) con X e Y polacos y $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ semicontinua inferiormente y acotada inferiormente. Entonces valen las equivalencias*

$$(KU) + (MM) + (EQ) \iff (KU) + (MU) + (EQ) \iff (SU) \iff (CG).$$

Demostración. Toda $T \in \mathcal{T}(\mu, \nu)$ induce $\pi_T \in \mathcal{A}(\mu, \nu)$ que satisface $C[\pi_T] = I[\pi]$. Esto, junto con la observación de que $\pi_{T_1} = \pi_{T_2} \iff T_1 = T_2$ μ -a.e., prueba las dos primeras equivalencias. Por otra parte, si se verifica (SU), el único minimizante π para (KP) es de la forma π_T para alguna $T \in \mathcal{T}(\mu, \nu)$. En tal caso, dado cualquier minimizante $\tilde{\pi} \in \mathcal{A}(\mu, \nu)$, (KU) implica $\tilde{\pi} = \pi = \pi_T$ y por lo tanto se verifica (CG).

Finalmente, asumamos que vale (CG) y veamos que necesariamente vale (SU). Para esto, notar que la proposición 1.2.4 implica la existencia de un minimizante $\pi \in \mathcal{A}(\mu, \nu)$ para el problema de Kantorovich y la hipótesis (CG) implica que existe T con $\pi = \pi_T$. Observar que esto implica las hipótesis (EQ) y (MM), con lo cual sólo resta verificar (KU). Sean $\pi_1, \pi_2 \in \mathcal{A}(\mu, \nu)$ minimizantes para (KP). Por (CG) existen T_1, T_2 tales que $\pi_1 = \pi_{T_1}$ y $\pi_2 = \pi_{T_2}$. La linealidad de I implica que $\tilde{\pi} = \frac{1}{2}\pi_1 + \frac{1}{2}\pi_2$ es un minimizante de (KP) y por (CG) debe existir $\tilde{T}$ con $\tilde{\pi} = \pi_{\tilde{T}}$. Luego

$$\begin{aligned} 1 &= \tilde{\pi}[\Gamma(\tilde{T})] = \frac{1}{2}\pi_1[\Gamma(\tilde{T})] + \frac{1}{2}\pi_2[\Gamma(\tilde{T})] \\ &= \frac{1}{2}\mu[\{x \in X : T_1(x) = \tilde{T}(x)\}] + \frac{1}{2}\mu[\{x \in X : T_2(x) = \tilde{T}(x)\}] \end{aligned}$$

y por lo tanto $T_1 = \tilde{T} = T_2$ μ -a.e., con lo cual $\pi_1 = \pi_2$. □

En el contexto de espacios métricos polacos generales puede garantizarse la hipótesis (EQ) requiriendo que μ sea no atómica y c sea continua, como lo prueba el siguiente resultado debido a Pratelli [108].

Teorema 1.5.2 (Pratelli) Sean X e Y espacios métricos polacos, sea $\nu \in \mathcal{P}(Y)$ y sea $\mu \in \mathcal{P}(X)$ una medida no atómica. Para un costo $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ continuo y acotado inferiormente, considerar el problema (X, Y, μ, ν, c) con valor óptimo I_* para el problema de Kantorovich. Entonces, para todo $\varepsilon > 0$ existe un transporte $T_\varepsilon \in \mathcal{T}(\mu, \nu)$ tal que $C[T_\varepsilon] \leq \varepsilon + I_*$ y por lo tanto vale

$$-\infty < \min_{\pi \in \mathcal{A}(\mu, \nu)} I[\pi] = \inf_{T \in \mathcal{T}(\mu, \nu)} C[T].$$

Finalizamos esta sección mostrando que las implicaciones del lema 1.5.1 son sharp. Más precisamente, para la clase de problemas que no satisfacen la hipótesis (EQ) se tiene la no implicación

$$(KU) + (MU) \not\Rightarrow (EQ),$$

como lo prueba el ejemplo 1.5.3 tomado de [108]. En general, se tienen las no implicaciones

$$(KU) + (EQ) \not\Rightarrow (MM) \quad \text{y} \quad (MU) + (EQ) \not\Rightarrow (KU),$$

aun bajo la hipótesis de que μ sea no atómica y c sea continua. Para la primera de ellas, un contraejemplo está dado por el problema del ejemplo 1.1.5. Para la segunda, los problemas discutidos en los ejemplos 1.5.4 y 1.5.5 proveen contraejemplos.

Ejemplo 1.5.3 Tomando $X = Y = \mathbb{R}^2$, considerar bolas $B_1 = B_r(c_1)$ y $B_2 = B_r(c_1)$ de área $1/2$ y puntos $x_1, x_2 \in \mathbb{R}^2$ tales que

$$r < d := \text{dist}(x_1, c_2) = \text{dist}(x_2, c_1) \ll \text{dist}(c_1, c_2) =: D,$$

como en la fig. 1.2. Sea $c(x, y) = |x - y|^2$ el costo cuadrático y sean las medidas

$$\mu = \frac{1}{2} \delta_{x_1} + \mathcal{L}^2|_{B_1} \quad \text{y} \quad \nu = \frac{1}{2} \delta_{x_2} + \mathcal{L}^2|_{B_2}.$$



Figura 1.2: Soportes de μ y ν .

Es claro que el único transporte óptimo $T_* \in \mathcal{T}(\mu, \nu)$ es el que manda x_1 en x_2 y $\mu|_{B_1}$

en $\mu|_{B_2}$ de forma óptima, de modo que $C[T_*] \approx D$. Observar que T_* está unívocamente determinada, pues necesariamente $T_*(x_1) = x_2$ y el teorema 1.1.13 implica que la condición $(T_*)_{\#}(\mu|_{B_1}) = \nu|_{B_2}$ determina T_* unívocamente sobre B_1 . Por otra parte, al ser $d \ll D$, es claro que el único plan de transporte óptimo $\pi_* \in \mathcal{A}(\mu, \nu)$ desintegra el átomo en x_1 para cubrir B_2 y concentra la masa de B_1 en x_2 , de modo que $I[\pi_*] \approx d$. Se verifican las hipótesis (KU) y (MU), pero al ser $d \ll D$ no se verifica la hipótesis (EQ). $\triangle$

Ejemplo 1.5.4 Tomando $X = Y = \mathbb{R}^2$, sean x_1, y_1, x_2, y_2 los vértices de un cuadrado unitario en $\mathbb{R}^2$ recorridos en orden. Considerar el costo $c(x, y) = |x - y|$ y las medidas

$$\mu = \frac{1}{3} \delta_{x_1} + \frac{2}{3} \delta_{x_2} \quad \text{y} \quad \nu = \frac{1}{3} \delta_{y_1} + \frac{2}{3} \delta_{y_2}.$$

Notar que vale

$$|x - y| = 1 \quad \forall x \in \text{sop}(\mu), \forall y \in \text{sop}(\nu),$$

con lo cual son óptimos y de costo 1 todos los planes de transporte y todos los transportes. Existe un único transporte $T \in \mathcal{T}(\mu, \nu)$ y está dado por $T(x_1) = y_1$ y $T(x_2) = y_2$, con lo cual se verifican (MU) y (EQ). Por otra parte, la medida

$$\pi = \frac{1}{3} [\delta_{x_1} \otimes \delta_{y_2} + \delta_{x_2} \otimes \delta_{y_1} + \delta_{x_2} \otimes \delta_{y_2}] \in \mathcal{A}(\mu, \nu)$$

verifica $\pi \neq \pi_T = \frac{1}{3} \delta_{x_1} \otimes \delta_{y_1} + \frac{2}{3} \delta_{x_2} \otimes \delta_{y_2}$. Esto muestra que no se verifica (KU). $\triangle$

Ejemplo 1.5.5 Ahora refinamos el ejemplo anterior para mostrar que (MU) y (EQ) no implican (KU) aun si μ es no atómica. Para esto construimos un problema en $\mathbb{R}^3$ que al ser “colapsado” sobre el plano $\mathbb{R}^2 \times \{0\}$ coincide con el ejemplo anterior.

Tomando $X = Y = \mathbb{R}^3$, sean $x_1, y_1, x_2, y_2 \in \mathbb{R}^2 = \mathbb{R}^2 \times \{0\} \subset \mathbb{R}^3$ elegidos como en el ejemplo anterior y sea $I = [0, 1]$. Considerar el costo $c(x, y) = |x - y|$ y las medidas

$$\mu = \frac{1}{3} \mathcal{H}^1|_{\{x_1\} \times I} + \frac{2}{3} \mathcal{H}^1|_{\{x_2\} \times I} = \frac{1}{3} \delta_{x_1} \otimes \mathcal{H}^1|_I + \frac{2}{3} \delta_{x_2} \otimes \mathcal{H}^1|_I,$$

$$\nu = \frac{1}{3} \mathcal{H}^1|_{\{y_1\} \times I} + \frac{2}{3} \mathcal{H}^1|_{\{y_2\} \times I} = \frac{1}{3} \delta_{y_1} \otimes \mathcal{H}^1|_I + \frac{2}{3} \delta_{y_2} \otimes \mathcal{H}^1|_I.$$

Como se ve en la fig. 1.3, existe un único transporte óptimo T_* y está dado por

$$T_*(x_1, s) = (y_1, s) \quad \text{y} \quad T_*(x_2, s) = (y_2, s) \quad \text{para todo } s \in I.$$

Al ser $C[T_*] = 1$, la optimalidad de T_* resulta de que $\text{dist}(\text{sop}(\mu), \text{sop}(\nu)) = 1$, pues esto implica $C[T] \geq 1$ para todo $T \in \mathcal{T}(\mu, \nu)$. Para ver la unicidad, suponer que T es un transporte óptimo y notar que necesariamente verifica $|T(x) - x| = 1$ μ -a.e. Esto dice que para $\mathcal{H}^1$ casi todo $s \in I$ existe $j \in \{1, 2\}$ tal que $T(x_2, s) = (y_j, s)$. Si para un conjunto $S \subset I$ con $\mathcal{H}^1[S] > 0$ se tiene $T(x_2, s) = (y_1, s)$, entonces

$$\frac{1}{3} \mathcal{H}^1[S] = \nu[\{y_1\} \times S] = \mu[T^{-1}(\{y_1\} \times S)] \geq \mu[\{x_2\} \times S] = \frac{2}{3} \mathcal{H}^1[S] > 0,$$

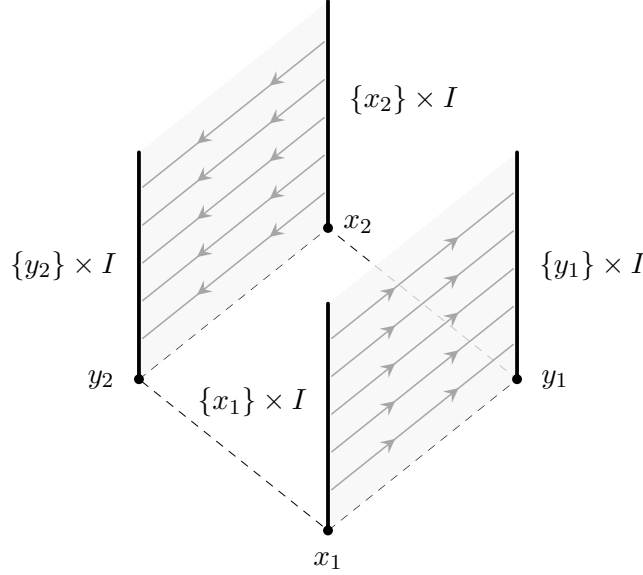


Figura 1.3: Visualización del transporte óptimo T_* , las flechas grises indican el sentido de los rayos de transporte $[x, T_*(x)]$.

lo que es un absurdo. Luego debe ser $T(x_2, s) = (y_2, s)$ para $\mathcal{H}^1$ casi todo $s \in I$ y por lo tanto $T(x_1, s) = (y_1, s)$ para $\mathcal{H}^1$ casi todo $s \in I$, de modo que $T = T_*$ μ -a.e. Esto prueba que se verifica (MU) y, al ser μ no atómica, el teorema 1.5.2 implica (EQ). Para probar que no se verifica (KU) basta notar que, si π_0 es un plan de transporte para el problema del ejemplo anterior, entonces $\pi = \pi_0 \otimes \mathcal{H}^1|_I$ es un plan de transporte para este problema. $\triangle$

1.6. Regularidad

En esta sección presentamos resultados de regularidad y regularidad parcial para el transporte óptimo en espacios euclídeos y en variedades riemannianas. Presentamos la teoría de regularidad desarrollada por Caffarelli [28, 29, 30, 31, 32, 34] y la teoría de regularidad parcial desarrollada por Figalli, Kim y De Philippis [55, 64, 66].

Sean $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ dadas por densidades f y g , de modo que

$$\mu[E] = \int_{\mathbb{R}^n} f(x) dx \quad \text{y} \quad \nu[E] = \int_{\mathbb{R}^n} g(y) dy$$

para todo boreliano $E \subset \mathbb{R}^n$. Si $T \in \mathcal{T}(\mu, \nu)$ es Lipschitz e inyectiva, el teorema de cambio de variables implica

$$\nu[T(E)] = \int_{T(E)} g(y) dy = \int_E (g \circ T)(x) |\det DT(x)| dx$$

para todo boreliano $E \subset \mathbb{R}^n$. Como T transporta μ en ν , tenemos también

$$\nu[T(E)] = \mu[T^{-1}(T(E))] = \mu[E] = \int_E f(x) dx.$$

Luego T está sujeta a la condición no lineal

$$f(x) = g(T(x)) |\det DT(x)|, \quad (25)$$

conocida como **ecuación del jacobiano**.

En el caso de que T sea el transporte óptimo cuadrático de μ en ν , i.e. para el costo cuadrático $c(x, y) = \frac{1}{2}|x - y|^2$, existe un potencial convexo P para T y (25) toma la forma de una **ecuación de Monge–Ampère**

$$\det D^2P(x) = \frac{f(x)}{g(\nabla P(x))}.$$

Asumiendo que son abiertos acotados los conjuntos

$$X = \{x \in \mathbb{R}^n : f(x) > 0\} \quad \text{e} \quad Y = \{y \in \mathbb{R}^n : g(y) > 0\}$$

y asumiendo que f y g están despegadas de cero e infinito en X e Y respectivamente, existen constantes λ_1, λ_2 tales que P es una solución débil de

$$0 < \lambda_1 \leq \det D^2P \leq \lambda_2 < \infty \quad (26)$$

en X satisfaciendo $\nabla P(X) \subset Y$.

La teoría de regularidad para soluciones débiles de (26) fue desarrollada por Caffarelli en [28, 29, 30], obteniendo luego una teoría de regularidad para potenciales de Brenier [31, 32, 34]. Resumimos dichos resultados en el siguiente teorema.

Teorema 1.6.1 Sean $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ dadas por densidades $f, g \in L^1(\mathbb{R}^n)$ tales que

$$X = \{x \in \mathbb{R}^n : f(x) > 0\} \quad \text{e} \quad Y = \{y \in \mathbb{R}^n : g(y) > 0\}$$

son abiertos acotados y supongamos que $0 < \lambda \leq f, g \leq 1/\lambda$ sobre X e Y para alguna constante $\lambda > 0$. Sea T el transporte óptimo cuadrático de μ en ν .

Si Y es convexo, entonces:

- i) Para algún $\beta \in (0, 1)$ se tiene $T \in C_{loc}^{0, \beta}(X)$.
- ii) Si $f \in C(X)$ y $g \in C(Y)$, resulta $T \in W_{loc}^{1, p}(X)$ para todo $1 \leq p \leq \infty$.
- iii) Si para algún $k \in \mathbb{N}_0$ y algún $\alpha \in (0, 1)$ se tiene $f \in C_{loc}^{k, \alpha}(X)$ y $g \in C_{loc}^{k, \alpha}(Y)$, resulta $T \in C_{loc}^{k+1, \alpha}(X)$.

Si X e Y son convexos, entonces:

iv) Para algún $\beta \in (0, 1)$ se tiene $T \in C^{0,\beta}(\overline{X})$.

Si X e Y son uniformemente convexos, entonces:

v) Si para algún $k \in \mathbb{N}_0$ y algún $\alpha \in (0, 1)$ se tiene $f \in C^{k,\alpha}(\overline{X})$, $g \in C^{k,\alpha}(\overline{Y})$ y las fronteras de X e Y son de clase C^{k+2} , resulta $T \in C^{k+1,\alpha}(\overline{X})$.

La hipótesis de convexidad para Y en el teorema 1.6.1 es necesaria: sin dicha hipótesis T es en general discontinua, aun en ausencia de impedimentos topológicos (e.g. X conexo e Y disconexo). Para ilustrar este fenómeno presentamos un ejemplo dado por Caffarelli [32] que, como veremos luego, es en cierto sentido representativo del caso general.

Ejemplo 1.6.2 Sea X la bola unitaria de $\mathbb{R}^n$ y Y la unión de dos medias bolas centradas en $-e_1$ y e_1 con partes chatas paralelas a $x_1 = 0$ (ver fig. 1.4). Tomando μ y ν como las medidas de Lebesgue normalizadas, el transporte óptimo cuadrático T_* está dado por el gradiente del potencial convexo $P(x_1, \dots, x_n) = |x_1| + \frac{1}{2}|x|^2$:

$$T_*(x) = \nabla P(x) = \mathbb{1}_{\{x_1 < 0\}}(x) \cdot (x - e_1) + \mathbb{1}_{\{x_1 > 0\}}(x) \cdot (x + e_1).$$

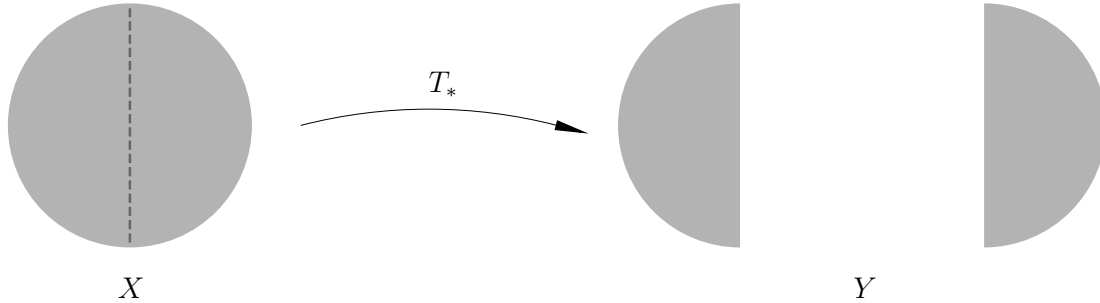


Figura 1.4: El transporte óptimo T_* divide la bola X a lo largo del segmento punteado $\Sigma = \{0\} \times (-1, 1)$. Notar que T_* es una isometría cuando se lo restringe a las medias bolas $X_- = X \cap \{x_1 < 0\}$, $X_+ = X \cap \{x_1 > 0\}$ que particionan $X \setminus \Sigma$.

Notar que $T_* = \nabla P$ está indefinida en $\{0\} \times (-1, 1)$, aunque T_* es Lipschitz cuando se la restringe a las medias bolas $X \cap \{x_1 < 0\}$ y $X \cap \{x_1 > 0\}$. El hecho de que Y sea disconexo no es esencial: como se explica en [32], uniendo las dos medias bolas con un franja horizontal de ancho ε y tomando el límite para $\varepsilon \rightarrow 0$ se construye un ejemplo con T_* discontinua e Y conexo (ver [125, Theorem 12.3] para una prueba completa). $\triangle$

El transporte óptimo T_* del ejemplo 1.6.2 no es regular en todo el abierto X pero verifica una propiedad de regularidad parcial: existen cerrados relativos

$$\Sigma_X := \{0\} \times (-1, 1) \subset X \quad \text{y} \quad \Sigma_Y := \emptyset \subset Y$$

de medida nula tales que $T_* : X \setminus \Sigma_X \rightarrow Y \setminus \Sigma_Y$ es un homeomorfismo bi-Lipschitz. Luego del trabajo de Figalli, Kim y De Philippis [55, 64, 66] se sabe que este tipo de regularidad parcial es un fenómeno general. Más aun, lo es incluso para costos generales en $\mathbb{R}^n$ y para el costo cuadrático en variedades riemannianas. Enunciamos a continuación estos resultados.

Sean X e Y abiertos acotados en $\mathbb{R}^n$ y $c : X \times Y \rightarrow \mathbb{R}$ una función de costo general. Para c definimos las siguientes hipótesis:

(C0) El costo c es de clase $C^{2,\alpha}$ para algún $\alpha \in (0, 1)$ y $\|c\|_{C^{2,\alpha}(X \times Y)} < \infty$.

(C1) Para todo $x \in X$, la aplicación $\nabla_x c(x, \cdot)$ es inyectiva en Y .

(C2) Para todo $y \in Y$, la aplicación $\nabla_y c(\cdot, y)$ es inyectiva en X .

(C3) Se verifica $\det(D_{xy}^2 c)(x, y) \neq 0$ para todo $(x, y) \in X \times Y$.

Notar que el costo cuadrático $c(x, y) = \frac{1}{2}|x - y|^2$ satisface todas las hipótesis anteriores.

Teorema 1.6.3 (55, Theorem 1.3) Sean $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ dadas por un par de densidades continuas $f, g \in L^1(\mathbb{R}^n)$ tales que

$$X = \{x \in \mathbb{R}^n : f(x) > 0\} \quad e \quad Y = \{y \in \mathbb{R}^n : g(y) > 0\}$$

son abiertos acotados y supongamos que $0 < \lambda \leq f, g \leq 1/\lambda$ sobre X e Y para alguna constante $\lambda > 0$. Sea T el transporte óptimo de μ en ν para un costo $c : X \times Y \rightarrow \mathbb{R}$ que verifica (C0) - (C3). Entonces, existen cerrados relativos $\Sigma_X \subset X$ y $\Sigma_Y \subset Y$ de medida nula tales que:

- i) La restricción $T : X \setminus \Sigma_X \rightarrow Y \setminus \Sigma_Y$ resulta ser un homeomorfismo de clase $C_{loc}^{0,\beta}$ para todo $\beta \in (0, 1)$.
- ii) Si para algún $k \in \mathbb{N}_0$ y algún $\alpha \in (0, 1)$ se tiene $c \in C_{loc}^{k+2,\alpha}(X \times Y)$, $f \in C_{loc}^{k,\alpha}(X)$ y $g \in C_{loc}^{k,\alpha}(Y)$, entonces $T : X \setminus \Sigma_X \rightarrow Y \setminus \Sigma_Y$ es un difeomorfismo de clase $C_{loc}^{k+1,\alpha}$.

Teorema 1.6.4 (55, Theorem 1.4) Sea M una variedad riemanniana suave, conexa, compacta y sin borde con distancia geodésica d . Sean $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ dadas por densidades f, g continuas y positivas. Sea $T : M \rightarrow M$ el transporte óptimo de μ en ν para el costo cuadrático $c = d^2/2$. Entonces existen cerrados $\Sigma_X, \Sigma_Y \subset M$ de medida nula tales que:

- i) La restricción $T : M \setminus \Sigma_X \rightarrow M \setminus \Sigma_Y$ resulta ser un homeomorfismo de clase $C_{loc}^{0,\beta}$ para todo $\beta \in (0, 1)$.
- ii) Si para algún $k \in \mathbb{N}_0$ y algún $\alpha \in (0, 1)$ se tiene $f, g \in C^{k,\alpha}(M)$, entonces resulta ser $T : X \setminus \Sigma_X \rightarrow Y \setminus \Sigma_Y$ un difeomorfismo de clase $C_{loc}^{k+1,\alpha}$.

Página en blanco.

Capítulo 2

Análisis de estabilidad para el problema de Monge

A continuación probamos la existencia de un p -orden de estabilidad s_p alrededor de T_* para un problema compacto (X, Y, μ, ν, c) que verifica la hipótesis (SU). Luego discutimos algunas propiedades de s_p y estudiamos el caso del transporte óptimo cuadrático en $\mathbb{R}^n$ a modo de ejemplo. En particular, mostramos que $s_2(h)$ es de orden $\sqrt{h}$ dando una prueba alternativa a la obtenida por Li y Nochetto [89, Remark 3.3]. Para problemas con espacios y costos más generales, asumiendo cierta regularidad para T_* pueden deducirse estimaciones de s_p aplicando los resultados obtenidos en [74, Proposition 8 y Corollary 9].

2.1. Existencia de órdenes de estabilidad

Cuando un problema compacto (X, Y, μ, ν, c) verifica la hipótesis (SU), el operador de costo de Monge C admite un p -orden de estabilidad alrededor del único transporte óptimo T_* para todo $1 \leq p < \infty$. Esto es, sin asumir regularidad adicional, existe una función no decreciente $s_p : [0, +\infty) \rightarrow [0, +\infty)$ continua en 0, satisfaciendo $s_p(0) = 0$ y verificando

$$d_p(T, T_*) \leq s_p(C[T] - C[T_*])$$

para todo $T \in \mathcal{B}(X, Y)$ con $T_{\#}\mu = \nu$, siendo s_p la mínima función con estas propiedades.

Para una demostración de este hecho necesitamos un resultado de estabilidad que probamos a continuación. Su demostración, que es una variación sobre la prueba de un resultado de estabilidad bien conocido [125, Corollary 5.23], depende de la siguiente versión del teorema de Lusin [61, Theorem 2.3.5].

Teorema 2.1.1 (Lusin) *Sea $\mathcal{X}$ un espacio métrico localmente compacto dotado de una medida boreliana de probabilidad μ y sea $\mathcal{Y}$ un espacio métrico separable. Si $f : \mathcal{X} \rightarrow \mathcal{Y}$ es*

una función boreliana y $A \subset \mathcal{X}$ es boreliano, entonces para todo $\varepsilon > 0$ existe un cerrado $K \subset A$ tal que $\mu[A \setminus K] < \varepsilon$ y $f|_K$ es continua.

Proposición 2.1.2 Sea (X, Y, μ, ν, c) un problema compacto con valor óptimo I_* para el problema de Kantorovich. Sea $(\pi_k)_k \subset \mathcal{P}(X \times Y)$ tal que $I[\pi_k] \xrightarrow{k} I_*$ y $\pi_k \in \mathcal{A}(\mu_k, \nu_k)$ para todo k , donde $\mu_k, \nu_k \xrightarrow{k} \mu, \nu$. Entonces valen las siguientes afirmaciones:

- i) Si el problema (X, Y, μ, ν, c) satisface la hipótesis (KU) con plan de transporte óptimo π_* , entonces se tiene $\pi_k \xrightarrow{k} \pi_*$.
- ii) Si $\mu_k = \mu$ para todo k y el problema (X, Y, μ, ν, c) satisface la hipótesis (SU) con transporte óptimo T_* , entonces para todo $\delta > 0$ vale

$$\pi_k[\{(x, y) \in X \times Y : d_Y(T_*(x), y) \geq \delta\}] \xrightarrow{k} 0. \quad (27)$$

Demostración. Dado que $\mu_k, \nu_k \xrightarrow{k} \mu, \nu$ y $\pi_k \in \mathcal{A}(\mu_k, \nu_k)$, la familia de medidas $\{\pi_k\}_k$ es un subconjunto rígido de $\mathcal{P}(X \times Y)$ y por lo tanto existe una subsucesión $(\pi_{k_j})_j$ débilmente convergente a $\pi \in \mathcal{A}(\mu, \nu)$. Se deduce $I[\pi] = I_*$ a partir de que $I[\pi_k] \xrightarrow{k} I_*$, con lo cual π es un plan de transporte óptimo. Asumiendo (KU) resulta $\pi = \pi_*$ y repitiendo el argumento para una subsucesión arbitraria de $(\pi_k)_k$ resulta $\pi_k \xrightarrow{k} \pi_*$. Esto prueba i) y asumiendo (SU) se tiene, adicionalmente, $\pi_{T_*} = \pi_*$.

Fijar $\delta > 0$ y tomar $\varepsilon > 0$ arbitrario. Por el teorema de Lusin 2.1.1 existe un conjunto cerrado $K \subset X$ con $\mu[X \setminus K] \leq \varepsilon$ y $T_*|_K$ continua. Luego

$$B_K(\delta) = \{(x, y) \in K \times Y : d_Y(T_*(x), y) \geq \delta\}$$

es cerrado en $K \times Y$ y $\pi_*[B_K(\delta)] = 0$, dado que π_* está concentrada en el gráfico de T_* . Luego, la convergencia débil de $(\pi_k)_k$ a π_* implica

$$\begin{aligned} 0 = \pi_*[B_K(\delta)] &\geq \limsup_k \pi_k[B_K(\delta)] \\ &= \limsup_k \pi_k[\{(x, y) \in K \times Y : d_Y(T_*(x), y) \geq \delta\}] \\ &\geq \limsup_k \pi_k[\{(x, y) \in X \times Y : d_Y(T_*(x), y) \geq \delta\}] - \varepsilon, \end{aligned}$$

donde la última desigualdad resulta de que $\pi_k \in \mathcal{A}(\mu_k, \nu_k)$ y $\mu[X \setminus K] \leq \varepsilon$. Tomando $\varepsilon \rightarrow 0$ obtenemos (27) y eso completa la prueba de ii). $\square$

Proposición 2.1.3 Sea (X, Y, μ, ν, c) un problema compacto con transporte óptimo T_* y satisfaciendo la hipótesis (SU). Entonces, si $(T_k)_k \subset \mathcal{T}(\mu, \nu)$ satisface $C[T_k] \xrightarrow{k} C[T_*]$, resulta $d_p(T_k, T_*) \xrightarrow{k} 0$ para todo $p \in [1, +\infty)$.

Demostración. Sea $\pi_k := \pi_{T_k} \in \mathcal{A}(\mu, \nu)$ para todo k . Dado que $I[\pi_k] = C[T_k]$, resulta $I[\pi_k] \rightarrow C[T_*] = I_*$, siendo I_* el valor óptimo del problema de Kantorovich asociado.

Tomando $\delta > 0$ y $\nu_k = \nu$, la proposición 2.1.2 implica

$$\mu[x \in X : d_Y(T_*(x), T_k(x)) > \delta] = \pi_k[(x, y) \in X \times Y : d_Y(T_*(x), y) > \delta] \xrightarrow{k} 0,$$

con lo cual $(T_k)_k$ converge en medida a T_* . Fijar $p \in [1, \infty)$ y notar que

$$d_p^p(T_k, T_*) = \int_X d_Y^p(T_k(x), T_*(x)) d\mu(x)$$

por definición de d_p . Resulta $\text{diám}(Y) < \infty$ por compacidad y para $\delta > 0$ se tiene

$$\begin{aligned} \int_{d_Y(T_k, T_*) \geq \delta} d_Y^p(T_k(x), T_*(x)) d\mu(x) &\leq \text{diám}(Y)^p \mu[\{d_Y(T_k, T_*) \geq \delta\}], \\ \int_{d_Y(T_k, T_*) < \delta} d_Y^p(T_k(x), T_*(x)) d\mu(x) &\leq \delta^p \mu[X] = \delta^p. \end{aligned} \quad (28)$$

Dado $\varepsilon > 0$, tomando $\delta = (\frac{\varepsilon}{2})^{1/p}$ y k_0 tal que

$$\mu[\{d_Y(T_k, T_*) \geq \delta\}] < \frac{\varepsilon}{2 \text{diám}(Y)^p + 1} \quad \text{para } k \geq k_0,$$

se obtiene $d_p^p(T_k, T_*) < \varepsilon$ para $k \geq k_0$ a partir de (28), probando $d_p(T_k, T_*) \xrightarrow{k} 0$. $\square$

Definición 2.1.4 Cuando (X, Y, μ, ν, c) verifica las hipótesis de la proposición 2.1.3 para todo exponente $p \in [1, \infty)$ definimos $s_p : [0, +\infty) \rightarrow [0, +\infty)$ como

$$s_p(h) = \sup \left\{ d_p(T, T_*) : \text{para } T \in \mathcal{T}(\mu, \nu) \text{ con } C[T] - C[T_*] \leq h \right\}$$

para todo $h \geq 0$. La buena definición de s_p resulta de la existencia de un transporte óptimo y del hecho de que Y es acotado. Llamamos *p-orden de estabilidad* de C alrededor de T_* a la función s_p .

Proposición 2.1.5 *Bajo las hipótesis de la proposición 2.1.3, para todo $p \in [1, \infty)$ valen las siguientes propiedades:*

i) s_p es no-decreciente y

$$d_p(T, T_*) \leq s_p(C[T] - C[T_*])$$

para toda $T \in \mathcal{T}(\mu, \nu)$;

ii) $0 \leq s_p \leq \text{diám}(Y)$;

iii) $s_p(0) = 0$ y s_p es continua en 0;

iv) s_p es la mínima función satisfaciendo i).

Demostración. Las propiedades i) y ii) son inmediatas a partir de la definición de s_p . Para probar iii), primero notar que $s_p(0) = 0$ resulta de la unicidad del transporte óptimo.

Luego, para cualquier sucesión $h_k \searrow 0$, tomar $T_k \in \mathcal{T}(\mu, \nu)$ verificando

$$C[T_k] - C[T_*] \leq h_k \quad \text{y} \quad d_p(T_k, T_*) > s_p(h_k) - \frac{1}{k}. \quad (29)$$

Así, (29) implica $C[T_k] \rightarrow C[T_*]$ y la proposición 2.1.3 implica $d_p(T_k, T_*) \rightarrow 0$. Resulta de (29) que $s_p(0) = 0 \geq \limsup_k s_p(h_k)$.

Para *iv*), asumir que $\tilde{s}_p : [0, +\infty) \rightarrow [0, +\infty)$ satisface *i*). Fijar $h \geq 0$ y considerar un transporte $T \in \mathcal{T}(\mu, \nu)$ con $C[T] - C[T_*] \leq h$. Luego *i*) para $\tilde{s}_p$ implica

$$d_p(T, T_*) \leq \tilde{s}_p(C[T] - C[T_*]) \leq \tilde{s}_p(h)$$

y tomando supremo sobre T resulta $s_p(h) \leq \tilde{s}_p(h)$. $\square$

Observación 2.1.6 La proposición 2.1.3 es falsa para $p = \infty$ y por lo tanto carece de sentido definir s_∞ . Para verlo, tomar $X = Y = [0, 1]$ y $\mu = \nu = \mathcal{L}^1$ la medida de Lebesgue 1-dimensional en el intervalo unitario. Considerar el costo cuadrático $c(x, y) = \frac{1}{2}|x - y|^2$. Notar que T_* es la identidad de $[0, 1]$ y por lo tanto $C[T_*] = 0$. Definiendo

$$T_\varepsilon(x) = \begin{cases} x & , \text{ si } \varepsilon < x < 1 - \varepsilon \\ 1 - x & , \text{ si no} \end{cases}$$

para $\varepsilon \in (0, \frac{1}{2})$, es claro que $T_\varepsilon \in \mathcal{T}(\mu, \nu)$ y que $d_\infty(T_\varepsilon, T_*) = 1$, pero un simple cálculo muestra que $C[T_\varepsilon] = \varepsilon - 2\varepsilon^2 + \frac{4}{3}\varepsilon^3 \rightarrow 0 = C[T_*]$.

2.2. El caso cuadrático en $\mathbb{R}^n$

Definición 2.2.1 Decimos que (X, Y, μ, ν, c) es un *problema cuadrático en $\mathbb{R}^n$* si X e Y son compactos en $\mathbb{R}^n$, $c(x, y) = \frac{1}{2}|x - y|^2$ y $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ son absolutamente continuas con densidades f, g tales que $X_\circ := \{f > 0\} \subset X$ e $Y_\circ := \{g > 0\} \subset Y$ son abiertos.

Observación 2.2.2 La definición 2.2.1 no requiere que se verifique $X_\circ = X^\circ$ ni $Y_\circ = Y^\circ$ y tampoco requiere regularidad de las fronteras $\partial X_\circ, \partial Y_\circ$. Notar también que $(X, d|_{X \times X})$ e $(Y, d|_{Y \times Y})$ son espacios polacos si d es la distancia usual en $\mathbb{R}^n$. Finalmente, notar que en todo problema cuadrático el costo verifica la hipótesis de integrabilidad.

Sea (X, Y, μ, ν, c) un problema cuadrático en $\mathbb{R}^n$ y sea (φ, ψ) un par c -conjugado que maximice el problema dual de Kantorovich (DP). De acuerdo con la proposición 1.3.8, un plan de transporte $\pi \in \mathcal{A}(\mu, \nu)$ es óptimo si y sólo si π está concentrado en el conjunto

$$\left\{ (x, y) \in X \times Y : \varphi(x) + \psi(y) = \frac{1}{2}|x - y|^2 \right\}. \quad (30)$$

Al ser $\varphi(x) + \psi(y) \leq \frac{1}{2}|x - y|^2$ en $X \times Y$, para $x \in X$ definimos $\varepsilon_x : Y \rightarrow [0, +\infty]$ por

$$\varepsilon_x(y) := \frac{1}{2}|x - y|^2 - \varphi(x) - \psi(y). \quad (31)$$

Asumiendo que el único transporte óptimo T_* para el problema de Monge es continuo, se tiene $\varepsilon_x(T_*(x)) = 0$ para todo $x \in X_\circ$ por la optimalidad de T_* . Recíprocamente, dada $T \in \mathcal{T}(\mu, \nu)$ subóptimo, puede interpretarse informalmente que estimar $\varepsilon_x(T(x))$ por debajo sobre $X_\circ$ es una manera de cuantificar la no inclusión $\Gamma(T) \not\subset \text{sop}(\pi_{T_*})$. Se puede ver que $P(x) := \frac{1}{2}|x|^2 - \varphi(x)$ verifica $\nabla P = T_*$ y que, si T_* es un difeomorfismo, tomando $Q := P^*$ se tiene $\nabla Q = (T_*)^{-1}$. Más aun, dado que

$$\varepsilon_x(y) = \left[\frac{1}{2}|x|^2 - \varphi(x) \right] - xy + \left[\frac{1}{2}|y|^2 - \psi(y) \right] = P(x) - xy + Q(y), \quad (32)$$

la convexidad de ε_x puede ser cuantificada notando que $D^2\varepsilon_x = D^2Q = D(T_*^{-1})$.

Las observaciones previas conducen a la estimación (33) de $d_2(T, T_*)$, que a su vez implica la estimación $s_2(h) \leq K\sqrt{h}$ para una constante $K > 0$ dependiente del problema (X, Y, μ, ν, c) . Esta misma estimación fue obtenida por Li y Nochetto [89, Remark 3.3]. A diferencia del razonamiento local presentado aquí, en [89] el resultado es obtenido con un razonamiento global a partir la noción de λ -smoothness para funciones convexas.

Teorema 2.2.3 *Sea (X, Y, μ, ν, c) un problema cuadrático en $\mathbb{R}^n$ con $Y_\circ$ convexo y tal que su transporte óptimo $T_* : X \rightarrow Y$ es un difeomorfismo en $D^1(X_\circ, Y_\circ)$. Si los autovalores de DT_* están acotados por arriba en $X_\circ$, i.e. $\sup_{X_\circ} \lambda_n(DT_*) < \infty$, entonces existe una constante $K = K(X, Y, \mu, \nu)$ tal que para toda $T \in \mathcal{T}(\mu, \nu)$ vale*

$$\|T - T_*\|_{L^2(\mu)} \leq K \left| C[T] - C[T_*] \right|^{1/2}. \quad (33)$$

Demostración. Con las notaciones introducidas previamente, al ser T_* un difeomorfismo en $D^1(X_\circ, Y_\circ)$, el potencial convexo Q tal que $\nabla Q = T_*^{-1}$ está en $C^2(Y_\circ)$. Notar que los autovalores de $D^2\varepsilon_x = D^2Q = D(T_*^{-1})$ están despegados de cero porque los autovalores de DT_* están acotados por arriba, precisamente:

$$\lambda_m := \inf_{Y_\circ} \lambda_1(D^2Q) = \inf_{X_\circ} \frac{1}{\lambda_n(DT_*)} = \frac{1}{\sup_{X_\circ} \lambda_n(DT_*)} > 0. \quad (34)$$

Fijando $x \in X_\circ$, notar que $\varepsilon_x(T_*(x)) = 0$ y $\nabla \varepsilon_x(T_*(x)) = 0$ por la optimalidad de T_* . Un desarrollo de Taylor de ε_x en torno a $T_*(x)$ muestra que, dado $y \in Y_\circ$ arbitrario, gracias a la convexidad de $Y_\circ$ existe $\theta_{x,y} \in [T_*(x), y] \subset Y_\circ$ con

$$\varepsilon_x(y) = \frac{1}{2}(y - T_*(x)) \cdot D^2\varepsilon_x(\theta_{x,y}) \cdot (y - T_*(x))^T \geq \frac{1}{2}\lambda_m|y - T_*(x)|^2. \quad (35)$$

La desigualdad (35) implica

$$\begin{aligned} \int_{X \times Y} \varepsilon_x(y) d\pi_T(x, y) &\geq \frac{1}{2} \lambda_m \int_{X \times Y} |y - T_*(x)|^2 d\pi_T(x, y) \\ &= \frac{1}{2} \lambda_m \int_X |T(x) - T_*(x)|^2 d\mu(x) = \frac{1}{2} \lambda_m \|T - T_*\|_{L^2(\mu)}^2. \end{aligned} \quad (36)$$

Por otra parte, dada $T \in \mathcal{T}(\mu, \nu)$, se tiene $\pi_T \in \mathcal{A}(\mu, \nu)$ y por lo tanto

$$\begin{aligned} \int_{X \times Y} \varepsilon_x(y) d\pi_T(x, y) &\stackrel{(31)}{=} \int_{X \times Y} c(x, y) d\pi_T(x, y) - \left[\int_X \varphi d\mu + \int_Y \psi d\nu \right] \\ &= I[\pi_T] - J[\varphi, \psi] = C[T] - C[T_*]. \end{aligned} \quad (37)$$

Luego (33) resulta de (37) y (36) tomando $K = \sqrt{\frac{2}{\lambda_m}}$. Notar que K depende sólo de $D^2Q = D(T_*^{-1})$, que a su vez depende sólo de X, Y, μ y ν . $\square$

El teorema anterior da una cota superior para $\|T - T_*\|_{L^2(\mu)}$ en el caso de que los autovalores de DT_* estén acotados por arriba. Asumiendo que dichos autovalores también están despegados de cero, i.e.

$$0 < \inf_{X_o} \lambda_1(DT_*) \leq \sup_{X_o} \lambda_n(DT_*) < +\infty, \quad (38)$$

puede probarse que $s_2(h)$ es de orden exactamente $\sqrt{h}$ para $h \rightarrow 0^+$. Aunque no lo probaremos aquí, tanto las constantes como los órdenes obtenidos son óptimos.

Teorema 2.2.4 *Sea (X, Y, μ, ν, c) un problema cuadrático en $\mathbb{R}^n$ con Y_o convexo y tal que su transporte óptimo $T_* : X \rightarrow Y$ es un difeomorfismo en $D^1(X_o, Y_o)$. Si los autovalores de DT_* están acotados en X_o como en (38), entonces vale*

$$\sqrt{\frac{2}{\lambda_M}} |C[T] - C[T_*]|^{1/2} \leq \|T - T_*\|_{L^2(\mu)} \leq \sqrt{\frac{2}{\lambda_m}} |C[T] - C[T_*]|^{1/2} \quad (39)$$

para toda $T \in \mathcal{T}(\mu, \nu)$, donde

$$\lambda_m = \inf_{Y_o} \lambda_1(D^2Q) > 0 \quad y \quad \lambda_M = \sup_{Y_o} \lambda_n(D^2Q) < +\infty. \quad (40)$$

Demostración. Al igual que en la demostración del teorema anterior basta acotar $\varepsilon_x(y)$ como en (35), pero esta vez también por arriba. Se obtienen así las cotas

$$\frac{1}{2} \lambda_m |y - T_*(x)|^2 \leq \varepsilon_x(y) \leq \frac{1}{2} \lambda_M |y - T_*(x)|^2 \quad (41)$$

para $(x, y) \in X_o \times Y_o$. Notar que se obtiene la cota $\lambda_M < +\infty$ razonando como en (34).

Luego de integrar (41) como en (37) y (36) resulta

$$\frac{1}{2}\lambda_m\|T - T_*\|_{L^2(\mu)}^2 \leq C[T] - C[T_*] \leq \frac{1}{2}\lambda_M\|T - T_*\|_{L^2(\mu)}^2$$

y se obtiene (39) inmediatamente. □

Página en blanco.

Capítulo 3

Aproximación discreta de (X, Y, μ, ν, c)

Empezamos definiendo h -aproximaciones de (X, Y, μ, ν, c) e introduciendo la notación necesaria para trabajar con ellas. Luego, probamos la existencia de sucesiones aproximantes de planes discretos. Definimos medianas geométricas en espacios métricos generales y comentamos sus propiedades. En particular, probamos el lema 3.2.6, que será fundamental para controlar las distancias $d_p(T_h, T_*)$. Finalmente, definimos distintos tipos de proyecciones y terminamos con una breve discusión sobre sus aspectos computacionales.

3.1. Discretización de (X, Y, μ, ν, c)

Definición 3.1.1 Una h -partición de un espacio métrico (Z, d_Z) es una colección finita $\{(Z_k, z_k)\}_{k \in K}$ donde $\{Z_k\}_{k \in K}$ es una partición de Z verificando $Z_k \in \mathcal{B}(Z)$, $\text{diám}(Z_k) \leq h$ y $z_k \in Z_k$ para todo $k \in K$. Dada una medida finita $\sigma \in \mathcal{P}(Z)$, una h -aproximación de σ es una medida $\sigma_h \in \mathcal{P}(Z)$ de la forma $\sum_{k \in K} \sigma_k \delta_{z_k}$, donde $\sigma_k = \sigma[Z_k]$ y $\{(Z_k, z_k)\}_{k \in K}$ es una h -partición de Z .

Observación 3.1.2 En general vamos a considerar h -particiones de espacios dotados de una medida $\sigma \in \mathcal{P}(Z)$ fija. Los resultados obtenidos no cambiarían si definiésemos las h -particiones requiriendo que los conjuntos $\{Z_k\}_{k \in K}$ sean sólo σ -disjuntos, en el sentido de que $\sigma[Z_{k_1} \cap Z_{k_2}] = 0$ para $z_1 \neq z_2$. Evitamos esto para prevenir dificultades técnicas no esenciales en algunas definiciones y demostraciones, pero consideraremos particiones σ -disjuntas en algunos ejemplos por la misma razón.

Definición 3.1.3 Una h -aproximación del problema (X, Y, μ, ν, c) es cualquier problema de la forma (X, Y, μ_h, ν_h, c) donde μ_h y ν_h son h -aproximaciones de μ y ν respectivamente.

En adelante, los nombres E y F denotarán conjuntos borelianos genéricos en X e Y

respectivamente. Los índices i serán usados para objetos relacionados con X y los índices j para los relacionados con Y . Las h -particiones genéricas para X e Y serán notadas

$$\mathcal{C}_X = \{(E_i, x_i)\}_{i \in I} \text{ y } \mathcal{C}_Y = \{(F_j, y_j)\}_{j \in J}, \quad (42)$$

y usaremos también las notaciones

$$\mu_i = \mu[E_i], \quad \nu_j = \nu[F_j] \quad \text{y} \quad c_{ij} = c(x_i, y_j) \quad \text{para } i \in I, j \in J.$$

Con estas notaciones tenemos

$$\mu_h = \sum_{i \in I} \mu[E_i] \cdot \delta_{x_i} = \sum_{i \in I} \mu_i \delta_{x_i}, \quad \nu_h = \sum_{j \in J} \nu[F_j] \cdot \delta_{y_j} = \sum_{j \in J} \nu_j \delta_{y_j} \quad (43)$$

y el problema de Kantorovich asociado a (X, Y, μ_h, ν_h, c) toma la forma

$$\begin{aligned} \min_{(\pi_{ij})_{ij} \geq 0} \quad & \sum_{i \in I} \sum_{j \in J} \pi_{ij} c_{ij} \\ \text{s. a.} \quad & \sum_{j \in J} \pi_{ij} = \mu_i \quad \forall i \in I, \quad \sum_{i \in I} \pi_{ij} = \nu_j \quad \forall j \in J. \end{aligned} \quad (\text{FP})$$

Notaremos $\pi_h = (\pi_{ij})_{ij}$ a las soluciones factibles de (FP) y definimos su *soporte* como

$$\text{sop}(\pi_h) = \{ (i, j) \in I \times J : \pi_{ij} > 0 \}. \quad (44)$$

Definimos también

$$I_+(j) = \{ i \in I : \pi_{ij} > 0 \} \quad (45)$$

para todo $j \in J$ e

$$I_+ = \{ i \in I : \exists j \in J \text{ con } \pi_{ij} > 0 \} = \bigcup_{j \in J} I_+(j). \quad (46)$$

Los conjuntos $J_+(i)$ para $i \in I$ y J_+ se definen análogamente. Observar que

$$\text{sop}(\pi_h) = \bigsqcup_{j \in J_+} I_+(j) \times \{j\} = \bigsqcup_{i \in I_+} \{i\} \times J_+(i). \quad (47)$$

Cometeremos el abuso de notar π_h tanto a las medidas del conjunto $\mathcal{A}(\mu_h, \nu_h)$ como a las sucesiones finitas doblemente indexadas y no negativas que definan soluciones factibles de (FP). Esto implica que la definición (44) comparte notación con el soporte de π_h visto como medida, es decir,

$$\text{sop}(\pi_h) = \{ (x_i, y_j) \in I \times J : \pi_{ij} > 0 \}. \quad (48)$$

Definición 3.1.4 Decimos que π_h es un h -plan óptimo (factible) para (X, Y, μ, ν, c) si π_h es un plan óptimo (factible) para alguna h -aproximación de (X, Y, μ, ν, c) .

Definición 3.1.5 Una sucesión aproximante de planes discretos para (X, Y, μ, ν, c) es una sucesión $(\pi_k)_k$ tal que todo π_k es un h_k -plan óptimo para (X, Y, μ, ν, c) y $h_k \xrightarrow{k} 0$.

Lema 3.1.6 Sea (X, Y, μ, ν, c) un problema de transporte con (X, d_X) e (Y, d_Y) espacios métricos polacos. Son equivalentes:

- i) Los espacios (X, d_X) e (Y, d_Y) son compactos.
- ii) Para todo $h > 0$ el problema (X, Y, μ, ν, c) admite un h -plan óptimo.
- iii) Existen sucesiones aproximantes de planes discretos para (X, Y, μ, ν, c) .

Demostración. La implicación ii) $\implies$ iii) es inmediata a partir de la definición 3.1.5. Para la implicación iii) $\implies$ i) basta probar que iii) implica que X e Y son totalmente acotados, pues ambos son completos por hipótesis. Para esto, asumir iii) y notar que las definiciones 3.1.5 y 3.1.1 implican inmediatamente que X e Y pueden ser particionados por finitos conjuntos de diámetro arbitrariamente pequeño. Luego X e Y son totalmente acotados y por lo tanto vale i).

Probemos ahora la implicación i) $\implies$ ii). Asumiendo i), la compacidad de X e Y implica la existencia de h -particiones. Para verlo, fijar $h > 0$, considerar el cubrimiento $\{B_{h/2}(\tilde{x})\}_{\tilde{x} \in X}$ de X y extraer un subcubrimiento finito minimal $\{B_{h/2}(\tilde{x}_i)\}_{i \in I}$. La minimalidad implica que para cada $i \in I$ existe $x_i \in B_{h/2}(\tilde{x}_i)$ no incluido en ninguna de las otras bolas del subcubrimiento. Luego, si $I = \{i_1, i_2, \dots, i_n\}$, la disyunción

$$\begin{aligned} E_{i_1} &:= B_{h/2}(\tilde{x}_{i_1}), \\ E_{i_k} &:= B_{h/2}(\tilde{x}_{i_k}) \setminus (E_{i_1} \cup \dots \cup E_{i_{k-1}}) \quad \text{para } 2 \leq k \leq n, \end{aligned}$$

induce la h -partición $\{(E_i, x_i)\}_{i \in I}$ para X . Tomando ahora una h -partición para Y y definiendo μ_h y ν_h como en (43), el problema (FP) resulta factible (ver sección 1.1.3) y por lo tanto existe un h -plan óptimo para (X, Y, μ, ν, c) . $\square$

Observación 3.1.7 El lema 3.1.6 implica que, si (X, Y, μ, ν, c) es compacto y $h_k \searrow 0$, existe una sucesión aproximante de planes discretos $(\pi_k)_k$ para (X, Y, μ, ν, c) con π_k un h_k -plan para todo k . Notando μ_k y ν_k a las medidas marginales de π_k , se tiene

$$\mu_k \xrightarrow{k} \mu \quad \text{y} \quad \nu_k \xrightarrow{k} \nu.$$

En efecto, si $\phi \in C_b(X)$ y notamos ω_ϕ a su módulo de continuidad, la compacidad de X implica que $\omega_\phi(h) \rightarrow 0$ para $h \rightarrow 0^+$. Fijando k y razonando con la notación genérica (42)

para las particiones, se obtiene

$$\left| \int_{E_i} \phi(x) d\mu(x) - \int_{E_i} \phi(x) d\mu_k(x) \right| = \left| \int_{E_i} \phi d\mu - \int_{E_i} \phi(x_i) d\mu \right| \leq \mu[E_i] \omega_\phi(h_k)$$

y sumando sobre $i \in I$ resulta la cota

$$\left| \int_X \phi d\mu - \int_X \phi d\mu_k \right| \leq \sum_i \mu[E_i] \cdot \omega_\phi(h_k) = \omega_\phi(h_k). \quad (49)$$

Dado que $h_k \xrightarrow{k} 0$, (49) implica $\mu_k \xrightarrow{k} \mu$. La prueba para $(\nu_k)_k$ es análoga.

3.2. Planes discretos y proyecciones

Para aproximar T_* nos interesa extraer una función $T_h : X \rightarrow Y$ de cada h -plan. Buscamos una T_h constante sobre cada conjunto de la h -partición $\mathcal{C}_X$, digamos

$$T_h|_{E_i} \equiv y_i \in Y \quad \text{para todo } i \text{ en } I.$$

En el caso de que π_h tenga un soporte “con forma de gráfico de función”, i.e. $|J_+(i)| = 1$ para todo i , entonces es intuitivamente claro cómo elegir cada y_i para construir T_h . En general, π_h es tal que $|J_+(i)| > 1$ para algún i . En este caso, para cada $i \in I$ tenemos una distribución de probabilidad sobre Y dada por

$$\sum_{j \in J} \frac{\pi_{ij}}{\mu_i} \delta_{y_j},$$

y tomaremos y_i como una tendencia central de dicha distribución, dependiendo de la estructura disponible en (Y, d_Y) .

Definición 3.2.1 Un espacio métrico (Y, d_Y) es *normado* si Y es un subconjunto convexo de un $\mathbb{R}$ -espacio vectorial $\mathbb{V}$ y d_Y está inducida en la forma usual por una norma $\|\cdot\|$ en $\mathbb{V}$.

Para (Y, d_Y) general extraeremos funciones T_h considerando medianas geométricas pesadas para elegir los y_i y llamaremos *GM proyecciones* a las funciones T_h construidas de esta manera. Para (Y, d_Y) normado consideraremos, además, funciones T_h construidas calculando baricentros, llamando a estas funciones *B proyecciones*.

3.2.1. Medianas geométricas pesadas y baricentros

Sea (Y, d_Y) un espacio métrico general. Dados $y_1, \dots, y_m$ en Y y pesos no negativos $w_1, \dots, w_m$ con $\sum_j w_j = 1$, considerar la función $V(y) = \sum_j w_j d_Y(y, y_j)$ para y en Y .

Definición 3.2.2 Un punto $\hat{y}$ en Y es una *mediana geométrica* para $(y_j)_j$ con pesos $(w_j)_j$ si $\hat{y}$ es un minimizante para V .

Definición 3.2.3 Si (Y, d_Y) es normado, el *baricentro* de $(y_j)_j$ con pesos $(w_j)_j$ es el punto $\sum_j w_j y_j \in Y$ que se obtiene operando en el espacio vectorial $\mathbb{V} \supset Y$.

Las medianas geométricas siempre existen cuando (Y, d_Y) es compacto, aunque pueden no ser únicas. Por ejemplo, cuando (Y, d_Y) es una variedad riemanniana, la unicidad se verifica si los puntos $(y_j)_j$ no yacen todos sobre una misma geodésica [130, Theorem 3.1]. Notar también que los baricentros están bien definidos: al ser Y convexo por definición de (Y, d_Y) normado, la suma $\sum_j w_j y_j$ yace en Y ; más aun, la cápsula convexa K de los puntos $(y_j)_j$ está contenida en Y , con lo cual el baricentro y todas las medianas geométricas pesadas de $(y_j)_j$ yacen también en Y por ser elementos de K .

Aclaración 3.2.4 Las medianas geométricas pesadas permiten “extender la idea” del baricentro a espacios métricos sin estructura lineal. Notar que no generalizan la noción de baricentro en un sentido estricto, pues una mediana geométrica pesada no necesariamente coincide con el baricentro cuando (Y, d_Y) es normado.

En general no existe una expresión explícita para las medianas geométricas pesadas, y por lo tanto deben ser aproximadas en la práctica. Para tratar esta dificultad definimos la noción de mediana geométrica aproximada y posponemos la discusión de sus aspectos computacionales hasta 3.3.

Definición 3.2.5 Un punto $\hat{y}$ en Y es una *mediana geométrica ε -aproximada* para $(y_j)_j$ con pesos $(w_j)_j$ si $V(\hat{y}) \leq V(y) + \varepsilon$ para todo y en Y . En particular, toda mediana geométrica 0-aproximada es una mediana geométrica.

Lema 3.2.6 (Lema de cercanía) Sean (Y, d_Y) un espacio métrico e $y_1, \dots, y_m$ puntos en Y con pesos no negativos $w_1, \dots, w_m$ satisfaciendo $\sum_j w_j = 1$.

- i) Si $\hat{y}_\varepsilon$ es una mediana geométrica ε -aproximada para $(y_j)_j$ con pesos $(w_j)_j$ e y es un punto cualquiera en Y , entonces

$$d_Y(y, \hat{y}_\varepsilon) \leq \varepsilon + 2 \sum_j w_j d_Y(y, y_j).$$

- ii) Si (Y, d_Y) es normado, $\hat{y}$ es el baricentro de $(y_j)_j$ con pesos $(w_j)_j$ e y es un punto cualquiera en Y , entonces

$$d_Y(y, \hat{y}) \leq \sum_j w_j d_Y(y, y_j).$$

La motivación para el nombre de lema 3.2.6 es la siguiente: cuando todos los puntos $(y_j)_j$ están cerca de un punto y , toda mediana geométrica o baricentro de los puntos $(y_j)_j$ también estará cerca de y . Esta idea será fundamental para probar la convergencia de las proyecciones T_h a T_* en el Capítulo 5.

Demostración del Lema 3.2.6. Para i), notar que

$$d_Y(y, \hat{y}_\varepsilon) \leq d_Y(y, y_j) + d_Y(y_j, \hat{y}_\varepsilon) \quad \text{para } 1 \leq j \leq m.$$

Multiplicando dicha desigualdad por w_j y sumando sobre j resulta

$$d_Y(y, \hat{y}_\varepsilon) \leq V(y) + V(\hat{y}_\varepsilon) \leq \varepsilon + 2V(y).$$

Para ii), d_Y viene de la norma $\|\cdot\|$ y operando en $\mathbb{V} \supset Y$ tenemos

$$\begin{aligned} d_Y(y, \hat{y}) &= \left\| y - \left(\sum_j w_j y_j \right) \right\| = \left\| \sum_j w_j (y - y_j) \right\| \\ &\leq \sum_j w_j \|y - y_j\| = V(y). \end{aligned} \quad \square$$

3.2.2. GM proyecciones y B proyecciones

Definición 3.2.7 Sea $\pi_h = (\pi_{ij})_{ij}$ un h -plan factible para (X, Y, μ, ν, c) y sean $(y_i)_{i \in I} \subset Y$ y $T \in \mathcal{B}(X, Y)$ tales que $T|_{E_i} \equiv y_i$ para todo $i \in I$.

- i) Decimos que T es una *GM proyección* si existe $Q \geq 0$ tal que, para todo $i \in I_+$, es y_i una mediana geométrica Q - h -aproximada para $(y_j)_{j \in J}$ con pesos $(\pi_{ij}/\mu_i)_{j \in J}$.
- ii) Si (Y, d_Y) es normado, decimos que T es una *B proyección* si para todo $i \in I_+$ es y_i el baricentro de $(y_j)_{j \in J}$ con pesos $(\pi_{ij}/\mu_i)_{j \in J}$.
- iii) Decimos que T es una *proyección extraída de π_h* si T verifica i) o ii).

Notamos $\mathcal{T}_{GM}^Q[\pi_h]$ al conjunto de las T construidas de acuerdo a i) y $\mathcal{T}_B[\pi_h]$ al formado por las construidas de acuerdo a ii), reservando la notación $\mathcal{T}_{GM}[\pi_h]$ para $\mathcal{T}_{GM}^0[\pi_h]$. Finalmente, decimos que T es de *calidad Q* si $T \in \mathcal{T}_{GM}^Q[\pi_h]$ o si $T \in \mathcal{T}_B[\pi_h]$.

Aclaración 3.2.8 Los conjuntos $\mathcal{T}_{GM}^Q[\pi_h]$, $\mathcal{T}_{GM}[\pi_h]$ y $\mathcal{T}_B[\pi_h]$ dependen de h -particiones $\mathcal{C}_X$ y $\mathcal{C}_Y$, del parámetro h y de un h -plan π_h . Omitimos explicitar la dependencia de dichos conjuntos respecto de $\mathcal{C}_X$, $\mathcal{C}_Y$ y h para simplificar la notación. Por otra parte, se verifica que $\mathcal{T}_{GM}^Q[\pi_h]$ es no vacío si (Y, d_Y) es compacto y que $\mathcal{T}_B[\pi_h]$ es no vacío si (Y, d_Y) es normado. Observar que $\mathcal{T}_{GM}^{Q_1}[\pi_h] \subset \mathcal{T}_{GM}^{Q_2}[\pi_h]$ si $0 \leq Q_1 \leq Q_2$, con lo cual toda proyección de calidad Q_1 es también de calidad Q_2 .

Definición 3.2.9 Una *sucesión aproximante de proyecciones* para (X, Y, μ, ν, c) es una sucesión $(T_k)_k \subset \mathcal{B}(X, Y)$ para la que existen $Q \geq 0$ y una sucesión aproximante de planes discretos $(\pi_k)_k$ tales que cada T_k es una proyección de calidad Q extraída de π_k .

Aclaración 3.2.10 Por definición, un espacio métrico (Y, d_Y) no puede ser normado si Y es un subconjunto no convexo del espacio normado $\mathbb{V}$. No obstante, en un caso así sigue teniendo sentido intentar extraer una B proyección $T_h : X \rightarrow \mathbb{V}$ de π_h si puede aceptarse

que T_h tome valores en $\text{conv}(Y)$, la cápsula convexa de Y . Este caso está incluido en nuestra teoría: dado que no asumimos $\text{sop}(\mu) = X$ ni $\text{sop}(\nu) = Y$, puede cambiarse (X, Y, μ, ν, c) por $(X, \text{conv}(Y), \mu, \nu, c)$ y trabajar con dicho problema.

3.3. Cómputo de proyecciones

Extraer B proyecciones de un h -plan π_h requiere calcular combinaciones lineales en un espacio vectorial $\mathbb{V}$. Precisamente, $O(|J_+(i)|)$ operaciones de espacio vectorial para cada $i \in I_+$. En el peor caso esto implica $O(|I||J|)$ operaciones de espacio vectorial para calcular una proyección baricéntrica. Dicha estimación mejora a $O(|I| + |J|)$ cuando el soporte de π_h es de tamaño $O(|I| + |J|)$, lo cual puede conseguirse e.g. tomando π_h como una solución básica para el problema (FP).

El cálculo de una GM proyección es comparativamente más difícil dado que no existe una expresión explícita para las medianas geométricas, con lo cual deben ser aproximadas. En espacios euclídeos y en variedades riemannianas dichas aproximaciones pueden ser calculadas con el algoritmo de Weiszfeld y sus variantes [67, 107, 126]. Para espacios métricos generales (Y, d_Y) , en principio, no es claro cómo aproximar medianas geométricas. Afortunadamente, como veremos a continuación, en nuestro contexto siempre es posible calcular aproximaciones suficientemente buenas de las medianas geométricas necesarias para calcular GM proyecciones.

Desde un punto de vista práctico, para encontrarnos con el problema de calcular una GM proyección debemos previamente haber computado una h -partición $\mathcal{C}_Y$. Dicha tarea requiere evaluar d_Y (para conocer la cota de diámetro h) y producir una colección de puntos parejamente distribuidos en Y (los puntos y_j elegidos en los conjuntos F_j). Como veremos, esto será suficiente para poder computar aproximaciones de las medianas geométricas necesarias para el cálculo de una GM proyección.

Decimos que $A \subset Y$ es una ε -red si para todo $y \in Y$ existe $a \in A$ con $d_Y(a, y) \leq \varepsilon$. Por ejemplo, los puntos $(y_j)_j$ de una h -partición $\{(F_j, y_j)\}_j$ de Y definen una h -red en Y .

Lema 3.3.1 Sean (Y, d_Y) un espacio métrico compacto, $A = (y_j)_{1 \leq j \leq m} \subset Y$ una ε -red y $(w_j)_{1 \leq j \leq m}$ pesos no negativos con $\sum_j w_j = 1$. Luego

$$\hat{y}_\varepsilon := \operatorname{argmin}_{y \in A} V(y) = \operatorname{argmin}_{y \in A} \sum_{j=1}^m w_j d_Y(y, y_j) \quad (50)$$

es una mediana geométrica ε -aproximada para $(y_j)_j$ con pesos $(w_j)_j$.

Demostración. Por ser Y compacto, existe $\hat{y} \in Y$ mediana geométrica para $(y_j)_j$ con pesos

$(w_j)_j$. Sea $y_{j_0} \in A$ con $d_Y(\hat{y}, y_{j_0}) \leq \varepsilon$. Luego, para cada $y \in Y$, se obtiene

$$\begin{aligned} \sum_j w_j d_Y(y_{j_0}, y_j) &\leq \sum_j w_j [d_Y(y_{j_0}, \hat{y}) + d_Y(\hat{y}, y_j)] \leq \sum_j w_j [\varepsilon + d_Y(\hat{y}, y_j)] \\ &= \varepsilon + \sum_j w_j d_Y(\hat{y}, y_j) \leq \varepsilon + \sum_j w_j d_Y(y, y_j). \end{aligned}$$

Tomando $\hat{y}_\varepsilon$ como en (50), para todo $y \in Y$ resulta

$$\begin{aligned} V(\hat{y}_\varepsilon) &= \sum_j w_j d_Y(\hat{y}_\varepsilon, y_j) \leq \sum_j w_j d_Y(y_{j_0}, y_j) \\ &\leq \varepsilon + \sum_j w_j d_Y(y, y_j) = \varepsilon + V(y). \end{aligned} \quad \square$$

Notar que (50) es un problema de minimización sobre un dominio finito, con lo cual el lema 3.3.1 proporciona un método efectivo para computar medianas geométricas aproximadas en Y . Aplicado a la h -red inducida por $\mathcal{C}_Y$, permite computar las medianas geométricas h -aproximadas necesarias para construir una T_h en $\mathcal{T}_{GM}^Q[\pi_h]$ con $Q = 1$.

Computar (50) por fuerza bruta requiere $O(|A|^2)$ operaciones, contando evaluaciones de d_Y y también sumas, productos y comparaciones de números. Notar que una cota de complejidad más justa es $O(|A||A_+|)$ operaciones, donde $A_+ = \{y_j \in A : w_j > 0\}$.

Aplicando este algoritmo al cómputo de una GM proyección, para cada $i \in I_+$ tenemos $A = \{y_j : j \in J\}$, $A_+ = \{y_j : j \in J_+(i)\}$ y $O(|J||J_+(i)|)$ operaciones. Sumando sobre $i \in I_+$ obtenemos

$$\sum_{i \in I_+} O(|J||J_+(i)|) = O\left(|J| \sum_{i \in I_+} |J_+(i)|\right) = O(|J| \text{sop}(\pi_h))$$

operaciones, que en el peor caso son $O(|I||J|^2)$ operaciones. Nuevamente, dicha estimación mejora a $O(|I||J| + |J|^2)$ cuando el soporte de π_h es de tamaño $O(|I| + |J|)$.

En la práctica pueden obtenerse algoritmos con mejores cotas de complejidad para (familias de) problemas concretos. Además, como regla general, el algoritmo de fuerza bruta sugerido por lema 3.3.1 puede ser mejorado achicando $|A|$ y $|A_+|$ apropiadamente. Para h pequeño (i.e. $|I|$ y $|J|$ grandes) los elementos de $\{y_j : j \in J_+(i)\}$ tienden a acumularse en torno a $T_*(x_i)$. Esto hace pequeño al conjunto $J_+(i)$, lo que reduce el factor $|A_+|$ en la cota $O(|A||A_+|)$. Adicionalmente, cuando $\{y_j : j \in J_+(i)\}$ está concentrado en un entorno de un punto, gran parte de los puntos en $\{y_j : j \in J\}$ puede ser descartada en la búsqueda de $\hat{y}_\varepsilon$. Esto reduce el factor $|A|$ en la cota $O(|A||A_+|)$.

Capítulo 4

Convergencia de planes discretos y valores óptimos

La acotación de c implica la continuidad del operador I con respecto a la convergencia débil de medidas en $\mathcal{P}(X \times Y)$. Es decir, $I[\pi_k] \xrightarrow{k} I[\pi]$ siempre que $\pi_k \xrightarrow{k} \pi$. Además, es fácil ver que las h -aproximaciones μ_h, ν_h de μ, ν convergen débilmente a μ, ν cuando $h \rightarrow 0$. Luego, dada cualquier sucesión aproximante de planes discretos $(\pi_k)_k$, la estabilidad del transporte óptimo [125, Theorem 5.20] implica la existencia de una subsucesión $(\pi_{k_l})_l$ convergente a un plan óptimo π_* . Finalmente, la continuidad de I implica $I[\pi_{k_l}] \xrightarrow{l} I_*$.

Veremos que, en nuestro contexto, la continuidad uniforme de c elimina la necesidad de pasar a una subsucesión para obtener la convergencia de $I[\pi_k]$ a I_* . Más aun, probaremos que la velocidad a la que $I[\pi_k]$ converge a I_* está controlada por ω_c .

4.1. Versión continua de un h -plan

Definición 4.1.1 Para una solución factible $\pi_h = (\pi_{ij})_{ij}$ de (FP), se define la *versión continua de π_h* como la medida de probabilidad $\pi_h^c \in \mathcal{P}(X \times Y)$ dada por

$$\pi_h^c = \sum_{(i,j) \in \text{sop}(\pi_h)} \frac{\pi_{ij}}{\mu_i \nu_j} \cdot \mu \otimes \nu|_{E_i \times F_j}. \quad (51)$$

Observación 4.1.2 La medida de probabilidad π_h^c está bien definida: se tiene $\mu_i, \nu_j > 0$ para todo $(i, j) \in \text{sop}(\pi_h)$, con lo cual π_h^c es una medida bien definida en $\mathcal{M}^+(X \times Y)$ y

$$\pi_h^c[X \times Y] = \sum_{ij} \frac{\pi_{ij}}{\mu_i \nu_j} \mu|_{E_i}[X] \cdot \nu|_{F_j}[Y] = \sum_{ij} \frac{\pi_{ij}}{\mu_i \nu_j} \mu[E_i] \nu[F_j] = \sum_{ij} \pi_{ij} = 1.$$

La motivación para la definición 4.1.1 es obtener un plan factible π_h^c para (KP) similar

a un h -plan discreto π_h dado. La siguiente proposición formaliza el hecho de que π_h^c es similar a π_h .

Proposición 4.1.3 *Sea $\pi_h = (\pi_{ij})_{ij}$ una solución factible de (FP). Luego π_h^c tiene las siguientes propiedades.*

- i) Es un elemento de $\mathcal{A}(\mu, \nu)$.*
- ii) Distribuye masa de forma similar a π_h , en el sentido de que*

$$d_{\mathcal{H}}(\text{sop}(\pi_h), \text{sop}(\pi_h^c)) \leq h \quad \text{y} \quad \pi_h^c[E_i \times F_j] = \pi_{ij} = \pi_h[E_i \times F_j] \quad \forall i, j. \quad (52)$$

- iii) Su costo es similar al de π_h si c es uniformemente continua, pues*

$$|I[\pi_h^c] - I[\pi_h]| \leq \omega_c(h).$$

Demostración. Para probar *i)* basta chequear las igualdades

$$\pi_h^c[E \times Y] = \mu[E] \quad \text{y} \quad \pi_h^c[X \times F] = \nu[F]$$

para $E \in \mathcal{B}(X)$ y $F \in \mathcal{B}(Y)$. En efecto, para $E \in \mathcal{B}(X)$ tenemos

$$\begin{aligned} \pi_h^c[E \times Y] &= \sum_{ij} \frac{\pi_{ij}}{\mu_i \nu_j} \cdot \mu \otimes \nu|_{E_i \times F_j}[E \times Y] \\ &= \sum_{ij} \frac{\pi_{ij}}{\mu_i \nu_j} \cdot \mu \otimes \nu[(E \cap E_i) \times F_j] \\ &= \sum_{ij} \frac{\pi_{ij}}{\mu_i \nu_j} \cdot \mu[E \cap E_i] \cdot \nu[F_j] = \sum_{ij} \frac{\pi_{ij}}{\mu_i} \cdot \mu[E \cap E_i] \\ &= \sum_{i \in I_+} \left(\sum_{j \in J_+(i)} \pi_{ij} \right) \cdot \frac{1}{\mu_i} \cdot \mu[E \cap E_i] \\ &= \sum_{i \in I_+} \mu_i \cdot \frac{1}{\mu_i} \cdot \mu[E \cap E_i] = \sum_{i \in I_+} \mu[E \cap E_i] = \mu[E], \end{aligned}$$

con lo cual la primera marginal de π_h^c es μ . Un argumento análogo muestra que ν es la segunda marginal de π_h^c , probando así que $\pi_h^c \in \mathcal{A}(\mu, \nu)$.

Ahora probamos *ii)*. Primero notar que $\pi_h^c[E_i \times F_j] = \pi_{ij}$ para todo (i, j) , lo cual es inmediato a partir de la definición de π_h^c . Ahora fijamos $(x_i, y_j) \in \text{sop}(\pi_h)$. Luego,

$$\text{sop}(\mu|_{E_i} \otimes \nu|_{F_j}) \subset \overline{E_i} \times \overline{F_j} \subset B_h[(x_i, y_j)] \subset B_h[\text{sop}(\pi_h)]. \quad (53)$$

Tenemos $\text{sop}(\pi_h^c) \cap (E_i \times F_j) \neq \emptyset$ porque $\pi_h^c[E_i \times F_j] = \pi_{ij} > 0$. Luego, tomando $(\hat{x}_i, \hat{y}_j) \in$

$\text{sop}(\pi_h^c) \cap (E_i \times F_j)$, resulta

$$(x_i, y_j) \in E_i \times F_j \subset B_h[(\hat{x}_i, \hat{y}_j)] \subset B_h[\text{sop}(\pi_h^c)]. \quad (54)$$

Las inclusiones (53) y (54) para todo (x_i, y_j) en $\text{sop}(\pi_h)$ implican las inclusiones

$$\text{sop}(\pi_h^c) \subset B_h[\text{sop}(\pi_h)] \quad \text{y} \quad \text{sop}(\pi_h) \subset B_h[\text{sop}(\pi_h^c)],$$

que juntas implican $d_{\mathcal{H}}(\text{sop}(\pi_h), \text{sop}(\pi_h^c)) \leq h$.

Finalmente, como $\text{diám}(E_i \times F_j) \leq h$ para todo (i, j) , resulta

$$\left| \int_{E_i \times F_j} c \, d\pi_h^c - \pi_{ij} c_{ij} \right| = \left| \int_{E_i \times F_j} [c - c_{ij}] \, d\pi_h^c \right| \leq \pi_{ij} \omega_c(h)$$

y *iii*) es inmediato

$$|I[\pi_h^c] - I[\pi_h]| \leq \left| \sum_{ij} \int_{E_i \times F_j} c \, d\pi_h^c - \sum_{ij} \pi_{ij} c_{ij} \right| \leq \sum_{ij} \pi_{ij} \omega_c(h) = \omega_c(h). \quad \square$$

4.2. Demostración del Teorema A

Teorema A Sea (X, Y, μ, ν, c) un problema compacto y sea I_* el valor óptimo para su problema de Kantorovich. Entonces todo h -plan óptimo π_h satisface

$$|I[\pi_h] - I_*| \leq \omega_c(h), \quad (55)$$

de modo que si $(\pi_k)_k$ es una sucesión aproximante de planes discretos resulta $I[\pi_k] \xrightarrow{k} I_*$. Más aun, si (X, Y, μ, ν, c) satisface adicionalmente la hipótesis (KU) con plan de transporte óptimo π_* entonces se tiene $W_p(\pi_k, \pi_*) \xrightarrow{k} 0$ para todo $1 \leq p < \infty$.

Demostración. Sea π_h un h -plan óptimo para el problema compacto (X, Y, μ, ν, c) . Por el ítem *i*) de la proposición 4.1.3, π_h^c es un plan factible para (KP) y el ítem *iii*) de esta misma proposición implica

$$I_* \leq I[\pi_h^c] \leq I[\pi_h] + \omega_c(h).$$

Para terminar de probar (55), sea π_* un plan óptimo para (KP) y consideremos la discretización $\tilde{\pi}_h = (\tilde{\pi}_{ij})_{ij}$ de π_* dada por

$$\tilde{\pi}_{ij} = \pi_*[E_i \times F_j] \quad \text{para } (i, j) \in I \times J. \quad (56)$$

Al ser $\tilde{\pi}_h$ una solución factible de (FP) resulta

$$\begin{aligned}
 I[\pi_h] &\leq I[\tilde{\pi}_h] = \sum_{ij} \tilde{\pi}_{ij} c_{ij} = \sum_{ij} c_{ij} \pi_*[E_i \times F_j] \\
 &= \sum_{ij} \int_{E_i \times F_j} c_{ij} d\pi_* \stackrel{(*)}{\leq} \sum_{ij} \int_{E_i \times F_j} [c + \omega_c(h)] d\pi_* \\
 &= \sum_{ij} \int_{E_i \times F_j} c d\pi_* + \sum_{ij} \int_{E_i \times F_j} \omega_c(h) d\pi_* \\
 &= \int_{X \times Y} c d\pi_* + \int_{X \times Y} \omega_c(h) d\pi_* = I[\pi_*] + \omega_c(h) = I_* + \omega_c(h),
 \end{aligned}$$

donde la desigualdad $(*)$ es consecuencia de que $\text{diám}(E_i \times F_j) \leq h$ para todo (i, j) . Esto completa la prueba de (55). Si $(\pi_k)_k$ es una sucesión aproximante de planes discretos resulta $I[\pi_k] \xrightarrow{k} I_*$ inmediatamente a partir de (55) y de la definición 3.1.5. Asumiendo adicionalmente la hipótesis (KU), el ítem i) de la proposición 2.1.2 prueba $\pi_k \xrightarrow{k} \pi_*$. Al ser $X \times Y$ compacto, W_p metriza la convergencia débil y por lo tanto $W_p(\pi_k, \pi_*) \xrightarrow{k} 0$. $\square$

El Teorema A garantiza que los valores óptimos aproximados $I[\pi_h]$ convergen a I_* con orden al menos $\omega_c(h)$. Más aun, dicha garantía de orden de convergencia es también válida para las versiones continuas de los h -planes. Registramos este hecho en el corolario enunciado a continuación, que será de fundamental importancia en el capítulo 6 para garantizar un orden en la convergencia $T_h \rightarrow T_*$.

Corolario 4.2.1 *Sea (X, Y, μ, ν, c) un problema compacto y sea I_* el valor óptimo para su problema de Kantorovich. Entonces para todo h -plan óptimo π_h se tiene*

$$|I[\pi_h^c] - I_*| \leq 2\omega_c(h).$$

Demostración. Por (55) y iii) de la proposición 4.1.3, es inmediato que

$$|I[\pi_h^c] - I_*| \leq |I[\pi_h^c] - I[\pi_h]| + |I[\pi_h] - I_*| \leq \omega_c(h) + \omega_c(h). \quad \square$$

4.3. Discusión

Finalizamos este capítulo probando que el orden de convergencia $\omega_c(h)$ es óptimo a este nivel de generalidad. Más precisamente, el orden de convergencia $\omega_c(h)$ para $I[\pi_h] \rightarrow I_*$ no puede ser mejorado sin imponer restricciones sobre (X, Y, μ, ν, c) y sobre las h -particiones.

Proposición 4.3.1 *Sean X, Y espacios métricos compactos y $c : X \times Y \rightarrow \mathbb{R}$ una función de costo continua. Supongamos que existe una función $\phi : (0, +\infty) \rightarrow (0, +\infty)$ dependiendo sólo de X, Y y c tal que*

$$|I[\pi_h] - I_*| \leq \phi(h)$$

siempre que $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y π_h sea un h -plan óptimo para (X, Y, μ, ν, c) . Entonces $\phi \geq \omega_c$.

Demostración. Fijamos $h > 0$ y tomamos $\rho > 0$. Por definición de ω_c , se tiene que existen $(x_1, y_1), (x_2, y_2)$ en $X \times Y$ tales que

$$|c(x_1, y_1) - c(x_2, y_2)| \geq (1 - \rho) \omega_c(h)$$

y

$$h \geq d_{X \times Y}((x_1, y_1), (x_2, y_2)) = \max\{d_X(x_1, x_2), d_Y(y_1, y_2)\}.$$

Definiendo las medidas $\mu = \delta_{x_1}$ y $\nu = \delta_{y_1}$, observar que $I_* = c(x_1, y_1)$ para el problema de transporte (X, Y, μ, ν, c) . Tomamos h -particiones $\mathcal{C}_X, \mathcal{C}_Y$ conteniendo, respectivamente, los conjuntos punteados $(\{x_1, x_2\}, x_2)$ y $(\{y_1, y_2\}, y_2)$. Observar que $\mathcal{C}_X$ y $\mathcal{C}_Y$ inducen un único h -plan óptimo π_h con $I[\pi_h] = c(x_2, y_2)$. Esto implica que

$$\phi(h) \geq |I[\pi_h] - I_*| = |c(x_1, y_1) - c(x_2, y_2)| \geq (1 - \rho) \omega_c(h)$$

y por lo tanto resulta $\phi(h) \geq \omega_c(h)$ tomando $\rho \rightarrow 0^+$. $\square$

Notar que la proposición 4.3.1 no implica que, el orden de convergencia de $I[\pi_k] \rightarrow I_*$ sea siempre exactamente $O(\omega_c(h))$. Aunque no daremos una prueba de este hecho, dicho orden puede mejorarse a $O(h^2)$ en espacios euclídeos si tanto el problema (X, Y, μ, ν, c) como las h -particiones son lo suficientemente regulares. Para enfatizar la importancia de las particiones a los fines de mejorar el orden de convergencia de los valores óptimos, en el siguiente ejemplo exhibimos una familia de problemas con espacios, medidas y funciones de costo regulares para los cuales la convergencia es de orden exactamente $O(\omega_c(h))$.

Ejemplo 4.3.2 Sean $X = Y = [0, 1]$ y $\mu = \nu = \mathcal{L}^1$ la medida de Lebesgue 1-dimensional en el intervalo unitario. Para $\alpha \in (0, 1]$, considerara el costo $c(x, y) = |x - y|^\alpha$ y notar que $\omega_c(h) = 2^\alpha h^\alpha$. Es fácil ver que (X, Y, μ, ν, c) admite un único transporte óptimo T_* , la identidad de $[0, 1]$, con $C[T_*] = 0$.

Para $k \in \mathbb{N}$ y $h_k = \frac{1}{k}$, considerar las h_k -particiones $\mathcal{C}_X^k, \mathcal{C}_Y^k$ dadas por

$$E_i^k = \left[\frac{i-1}{k}, \frac{i}{k} \right], \quad x_i^k = \frac{2i-1}{2k} + \frac{1}{3k}, \quad F_j^k = \left[\frac{j-1}{k}, \frac{j}{k} \right], \quad y_j^k = \frac{2j-1}{2k},$$

para $1 \leq i, j \leq k$. Notar que los elementos de $\mathcal{C}_Y^k$ están punteados en sus centros, pero no los elementos de $\mathcal{C}_X^k$. Consideramos ahora la familia de h_k -aproximaciones de (X, Y, μ, ν, c) definidas por $\mathcal{C}_X^k$ y $\mathcal{C}_Y^k$.

Veamos que, para todo k , existe un único h_k -plan óptimo $\pi_k = (\pi_{ij}^k)_{ij}$ dado por $\pi_{ij}^k = \frac{1}{k} \delta_{ij}$. Primero, notar que π_k es factible para (FP), dado que $\mu_i = \nu_j = \frac{1}{k}$ para

$1 \leq i, j \leq k$. Observar que

$$\begin{aligned} |x_i^k - y_j^k| &= \left| \frac{i-j}{k} + \frac{1}{3k} \right| \geq \frac{1}{k} - \frac{1}{3k} \geq \frac{1}{2k} \quad \text{si } i \neq j, \\ |x_i^k - y_j^k| &= \frac{1}{3k} \quad \text{si } i = j. \end{aligned}$$

Luego,

$$I[\pi_k] = \sum_{ij} \frac{1}{k} \delta_{ij} c_{ij} = \sum_{i=1}^k \frac{1}{k} \left(\frac{1}{3k} \right)^\alpha = \left(\frac{1}{3} \right)^\alpha h_k^\alpha$$

Para ver que π_k es el único h_k -plan óptimo, dado cualquier h_k -plan $\tilde{\pi}_k$ notamos m a la masa fuera de la diagonal, es decir, $m = \sum_{i \neq j} \tilde{\pi}_{ij}$. Luego, si $m > 0$,

$$\begin{aligned} I[\tilde{\pi}_k] &= \sum_{ij} \tilde{\pi}_{ij} c_{ij} = \sum_{i=j} \tilde{\pi}_{ij} c_{ij} + \sum_{i \neq j} \tilde{\pi}_{ij} c_{ij} \geq \left(\frac{1}{3k} \right)^\alpha \sum_{i=j} \tilde{\pi}_{ij} + \left(\frac{1}{2k} \right)^\alpha \sum_{i \neq j} \tilde{\pi}_{ij} \\ &= \left(\frac{1}{3k} \right)^\alpha (1 - m) + \left(\frac{1}{2k} \right)^\alpha m = \left[\left(\frac{1}{2k} \right)^\alpha - \left(\frac{1}{3k} \right)^\alpha \right] m + \left(\frac{1}{3k} \right)^\alpha > \left(\frac{1}{3k} \right)^\alpha = I[\pi_k] \end{aligned}$$

probando que π_k es el único h_k -plan óptimo. Finalmente, observar que

$$\omega_c(h_k) \geq |I[\pi_k] - I_*| = I[\pi_k] = \left(\frac{1}{3} \right)^\alpha h_k^\alpha = \left(\frac{1}{6} \right)^\alpha \omega_c(h_k),$$

con lo cual la convergencia $I[\pi_k] \rightarrow I_*$ es de orden exactamente $\omega_c(h_k)$. △

Capítulo 5

Convergencia de proyecciones

En este capítulo demostraremos la convergencia de las proyecciones T_h al transporte óptimo T_* en las métricas d_p para $1 \leq p < \infty$, probando así el Teorema B. Empezaremos definiendo la versión semidiscreta π_h^s de un h -plan π_h , lo que nos permitirá controlar la medida dada por π_h a los conjuntos “malos”

$$B_{T_*}(\delta) = \{ (x, y) \in X \times Y : d_Y(T_*(x), y) \geq \delta \}$$

para $h \rightarrow 0^+$. Luego, en la proposición 5.2.1, controlaremos $d_p(T_h, T_*)$ con la medida π_h^s de $B_{T_*}(\delta)$ y la convergencia de T_h a T_* se deducirá fácilmente a partir de este hecho. Finalmente, daremos algunos ejemplos para ilustrar la optimalidad del Teorema B.

5.1. Versión semidiscreta de un h -plan

Empezamos definiendo las medidas semidiscretas π_h^s y probando sus propiedades en la proposición 5.1.3. Luego aplicamos el resultado de estabilidad 2.1.2 para probar la proposición 5.1.4, demostrando que las medidas $\pi_h^s[B(\delta)]$ tienden a cero para $h \rightarrow 0^+$.

Definición 5.1.1 Para una solución factible $\pi_h = (\pi_{ij})_{ij}$ de (FP), se define la *versión semidiscreta de π_h* como la medida de probabilidad $\pi_h^s \in \mathcal{P}(X \times Y)$ dada por

$$\pi_h^s = \sum_{(i,j) \in \text{sop}(\pi_h)} \frac{\pi_{ij}}{\mu_i} \mu|_{E_i} \otimes \delta_{y_j}. \quad (57)$$

Observación 5.1.2 La medida de probabilidad π_h^s está bien definida: se tiene $\mu_i > 0$ para todo $(i, j) \in \text{sop}(\pi_h)$, con lo cual π_h^s es una medida bien definida en $\mathcal{M}^+(X \times Y)$ y

$$\pi_h^s[X \times Y] = \sum_{ij} \frac{\pi_{ij}}{\mu_i} \mu|_{E_i}[X] \cdot \delta_{y_j}[Y] = \sum_{ij} \frac{\pi_{ij}}{\mu_i} \mu[E_i] = \sum_{ij} \pi_{ij} = 1.$$

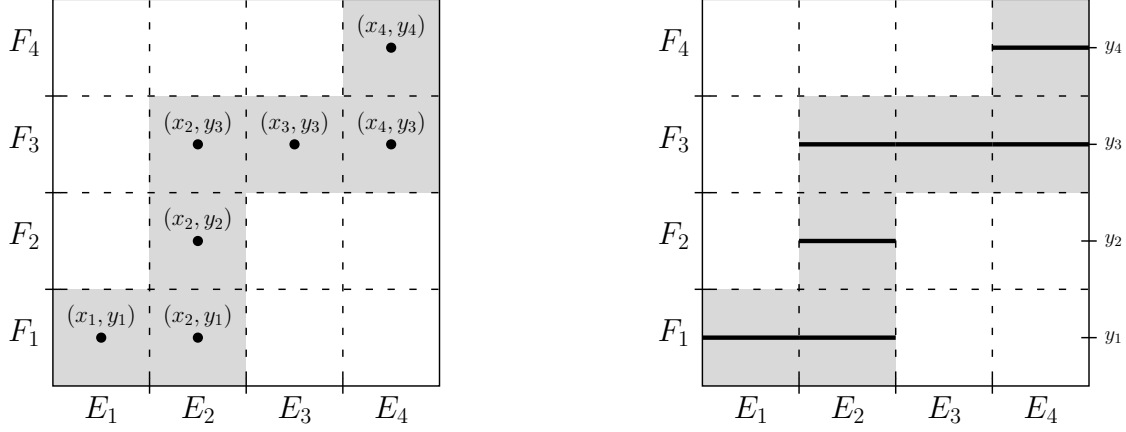


Figura 5.1: Los soportes de π_h (izquierda) y π_h^s (derecha). Es claro que ambos soportes distan a lo sumo h en distancia de Hausdorff, pues cada producto $E_i \times F_j$ tiene diámetro acotado por h (ver proposición 5.1.3).

La idea detrás de la definición 5.1.1 es adaptar las medidas π_h para poder aplicar la proposición 2.1.2, modificándolas lo menos posible con el fin de que su primera marginal sea μ . Sobre cada E_i , $\mu|_{E_i}$ se reescribe como una combinación convexa de copias de $\mu|_{E_i}$ “levantadas a alturas y_j ” (i.e. $\mu|_{E_i} \otimes \delta_{y_j}$), dándole a cada copia el peso correspondiente a la proporción de masa que x_i manda a y_j en el plan discreto π_h (i.e. π_{ij}/μ_i).

Las medidas π_h^s preservan su segunda marginal ν_h y preservan aproximadamente sus soportes (e.g. en distancia de Hausdorff, ver fig. 5.1). Más aun, la expresión explícita de π_h^s da una forma relativamente sencilla de computar medidas de conjuntos en $\mathcal{B}(X \times Y)$ (ver fig. 5.2). La siguiente proposición formaliza todas estas observaciones.

Proposición 5.1.3 *Sea $\pi_h = (\pi_{ij})_{ij}$ una solución factible de (FP). Luego π_h^s tiene las siguientes propiedades.*

- i) *Es un elemento de $\mathcal{A}(\mu, \nu_h)$.*
- ii) *Distribuye masa de forma similar a π_h , en el sentido de que*

$$d_{\mathcal{H}}(\text{sop}(\pi_h), \text{sop}(\pi_h^s)) \leq h \quad \text{y} \quad \pi_h^s[E_i \times F_j] = \pi_{ij} = \pi_h[E_i \times F_j] \quad \forall i, j. \quad (58)$$

- iii) *Admite la desintegración*

$$\pi_h^s = \sum_i \int_{E_i} \delta_x \otimes \left(\sum_j \frac{\pi_{ij}}{\mu_i} \delta_{y_j} \right) d\mu(x).$$

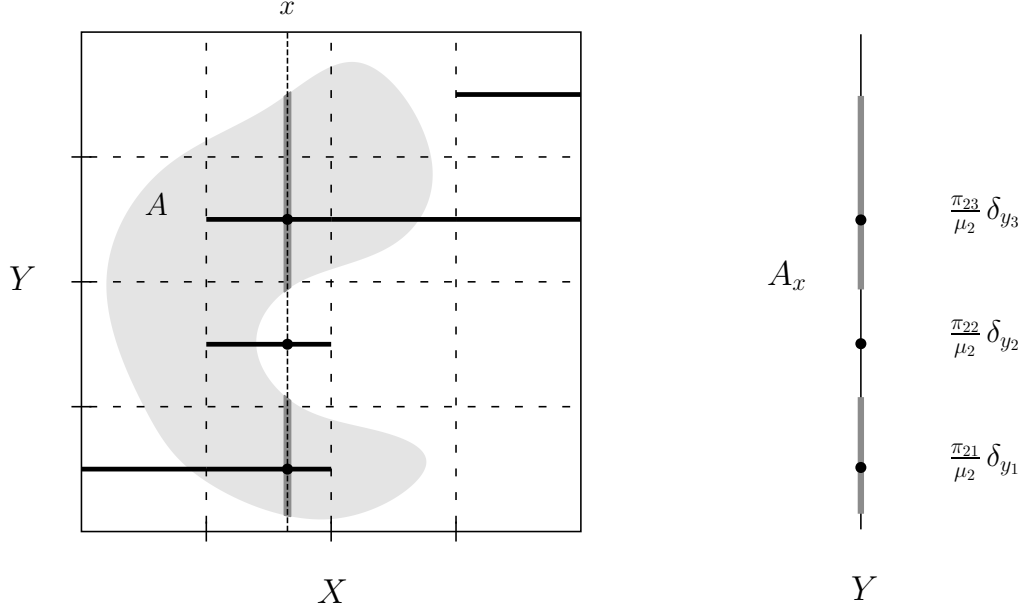


Figura 5.2: La medida π_h^s se desintegra con respecto a μ como una familia de medidas puramente atómicas. Las secciones A_x son pesadas por medidas finitamente soportadas (ver proposición 5.1.3).

Es decir, para cada boreliano $A \in \mathcal{B}(X \times Y)$ vale

$$\pi_h^s[A] = \sum_i \int_{E_i} \left(\sum_j \frac{\pi_{ij}}{\mu_i} \delta_{y_j}[A_x] \right) d\mu(x), \quad (59)$$

donde $A_x = \{y \in Y : (x, y) \in A\}$.

Demostración. Omitimos las pruebas de *i)* y *ii)* dado que son similares a las de *i)* y *ii)* de la proposición 4.1.3. Para probar *iii)* es suficiente chequear que la función $\eta : \mathcal{B}(X \times Y) \rightarrow \mathbb{R}$ dada por

$$\eta[A] = \sum_i \int_{E_i} \left(\sum_j \frac{\pi_{ij}}{\mu_i} \delta_{y_j}[A_x] \right) d\mu(x) \quad (60)$$

es una medida boreliana bien definida con $\eta = \pi_h^s$, lo que prueba (59).

Fijamos $A \in \mathcal{B}(X \times Y)$ y definimos $f_{ij} : E_i \rightarrow \mathbb{R}$ como $f_{ij}(x) = \delta_{y_j}[A_x]$. Notar que f_{ij} toma valores en $\{0, 1\}$, con lo cual la medibilidad de f_{ij} es equivalente a la de $f_{ij}^{-1}(1)$. Definiendo la función boreliana $\phi_{ij} : E_i \rightarrow X \times Y$ por $\phi_{ij}(x) = (x, y_j)$, observar que

$$\begin{aligned} f_{ij}^{-1}(1) &= \{x \in E_i : \delta_{y_j}[A_x] = 1\} = \{x \in E_i : (x, y_j) \in A\} \\ &= \{x \in E_i : (x, y_j) \in A \cap (E_i \times \{y_j\})\} \end{aligned}$$

$$= \phi_{ij}^{-1}(A \cap (E_i \times \{y_j\})),$$

con lo cual $f_{ij}^{-1}(1)$ es boreliano y entonces lo es f_{ij} . Luego, las integrales

$$\int_{E_i} \sum_j \frac{\pi_{ij}}{\mu_i} \delta_{y_j}[A_x] d\mu(x) = \int_{E_i} \sum_j \frac{\pi_{ij}}{\mu_i} f_{ij}(x) d\mu(x)$$

están bien definidas. Como ambas son finitas y no negativas, está bien definida η como función según (60) y también resulta finita y no negativa.

Ahora consideramos una familia disjunta $(A_k)_{k \in \mathbb{N}} \subset \mathcal{B}(X \times Y)$. Para $y \in Y$ y $x \in X$ es fácil chequear que $\sum_k \delta_y((A_k)_x) \leq 1$ y que

$$\delta_y \left[\left(\bigcup_k A_k \right)_x \right] = 1 \iff \sum_k \delta_y[(A_k)_x] = 1.$$

Esto implica que η es σ -aditiva y completa la prueba de que η es una medida boreliana no negativa y finita. Es inmediato que $\eta(X \times Y) = 1$.

Para probar que $\pi_h^s = \eta$ notar que

$$\Pi = \{E \times F : E \in \mathcal{B}(X), F \in \mathcal{B}(Y)\}$$

es un π -sistema que genera $\mathcal{B}(X \times Y)$ y que

$$\Lambda = \{A \in \mathcal{B}(X \times Y) : \pi_h^s[A] = \eta[A]\}$$

es un λ -sistema. Probando primero la contención $\Pi \subset \Lambda$, el Teorema $\pi - \lambda$ A.1.1 permite concluir $\sigma(\Pi) \subset \Lambda$, demostrando así que π_h^s y η coinciden sobre $\mathcal{B}(X \times Y) = \sigma(\Pi)$. Luego basta verificar que π_h^s y η coinciden sobre Π . Para esto, fijar $E \times F \in \Pi$ y notar que

$$\mathbb{1}_E(x) \cdot \left(\sum_j \frac{\pi_{ij}}{\mu_i} \delta_{y_j}[F] \right) = \sum_j \frac{\pi_{ij}}{\mu_i} \delta_{y_j}[(E \times F)_x] \quad (61)$$

para $i \in I$ y $x \in X$. Luego, se tiene

$$\begin{aligned} \eta[E \times F] &\stackrel{(60)}{=} \sum_i \int_{E_i} \left(\sum_j \frac{\pi_{ij}}{\mu_i} \delta_{y_j}[(E \times F)_x] \right) d\mu(x) \\ &\stackrel{(61)}{=} \sum_i \int_{E_i} \left(\sum_j \frac{\pi_{ij}}{\mu_i} \delta_{y_j}[F] \right) \mathbb{1}_E(x) d\mu(x) \\ &= \sum_i \int_{E \cap E_i} \left(\sum_j \frac{\pi_{ij}}{\mu_i} \delta_{y_j}[F] \right) d\mu(x) \\ &= \sum_i \left(\sum_j \frac{\pi_{ij}}{\mu_i} \delta_{y_j}[F] \right) \mu[E \cap E_i] \end{aligned}$$

$$= \sum_{ij} \frac{\pi_{ij}}{\mu_i} \delta_{y_j}[F] \mu[E \cap E_i] = \sum_{ij} \frac{\pi_{ij}}{\mu_i} \mu|_{E_i}[E] \delta_{y_j}[F] \stackrel{(57)}{=} \pi_h^s[E \times F]$$

y esto completa la demostración de *iii*). $\square$

Proposición 5.1.4 *Sea (X, Y, μ, ν, c) un problema compacto satisfaciendo la hipótesis (SU) con transporte óptimo T_* . Si $(\pi_k)_k$ es una sucesión aproximante de planes discretos para (X, Y, μ, ν, c) , entonces para todo $\delta > 0$ vale*

$$\pi_k^s[(x, y) \in X \times Y : d_Y(T_*(x), y) \geq \delta] \xrightarrow{k} 0. \quad (62)$$

Demostración. Tenemos $\pi_k^s \in \mathcal{A}(\mu, \nu_k)$ por la proposición 5.1.3 y $\nu_k \xrightarrow{k} \nu$ por la observación 3.1.7. Notando I_* al valor óptimo del problema de Kantorovich asociado a (X, Y, μ, ν, c) , basta probar $I[\pi_k^s] \xrightarrow{k} I_*$ para obtener (62) aplicando la proposición 2.1.2 a $(\pi_k^s)_k$. Para esto, notar primero que

$$|I[\pi_k^s] - I_*| \leq |I[\pi_k^s] - I[\pi_k]| + |I[\pi_k] - I_*|.$$

Dado que $I[\pi_k] \xrightarrow{k} I_*$ por el Teorema A, es suficiente verificar que $|I[\pi_k^s] - I[\pi_k]| \xrightarrow{k} 0$ para obtener $I[\pi_k^s] \xrightarrow{k} I_*$. Para esto, notar que $\omega_c(h) \rightarrow 0$ para $h \rightarrow 0^+$ por la compacidad de (X, Y, μ, ν, c) , con lo cual

$$\begin{aligned} |I[\pi_k^s] - I[\pi_k]| &= \left| \sum_{ij} \int_{E_i \times F_j} c(x, y) d\pi_k^s(x, y) - c_{ij} \pi_{ij} \right| \\ &\stackrel{(58)}{\leq} \sum_{ij} \int_{E_i \times F_j} |c(x, y) - c_{ij}| d\pi_k^s(x, y) \leq \sum_{ij} \omega_c(h_k) \pi_{ij} \\ &= \omega_c(h_k) \xrightarrow{k} 0 \end{aligned}$$

y esto completa la demostración. $\square$

5.2. Demostración del Teorema B

Proposición 5.2.1 *Dado un problema compacto (X, Y, μ, ν, c) , sean $T \in \mathcal{B}(X, Y)$ y π_h un h -plan factible para (X, Y, μ, ν, c) . Para cada $\delta > 0$, sea*

$$B_T(\delta) = \{(x, y) \in X \times Y : d_Y(T(x), y) \geq \delta\}. \quad (63)$$

Luego, dados $p \in [1, +\infty)$ y $Q \geq 0$, existen constantes A, B y C dependientes únicamente de p, D_Y y Q tales que

$$d_p(T_h, T) \leq A \cdot h + B \cdot \delta + C \cdot (\pi_h^s[B_T(\delta)])^{\frac{1}{p}} \quad (64)$$

para toda proyección T_h de calidad Q extraída de π_h . Más aun, las constantes A , B y C pueden tomarse como sigue:

i) para $T_h \in \mathcal{T}_{GM}^Q[\pi_h]$,

$$A = 3^{\frac{p-1}{p}} \cdot Q, \quad B = 2 \cdot 3^{\frac{p-1}{p}} \quad y \quad C = 2 \cdot 3^{\frac{p-1}{p}} \cdot D_Y; \quad (65)$$

ii) para $T_h \in \mathcal{T}_{GM}[\pi_h]$,

$$A = 0, \quad B = 2 \cdot 2^{\frac{p-1}{p}} \quad y \quad C = 2 \cdot 2^{\frac{p-1}{p}} \cdot D_Y; \quad (66)$$

iii) para (Y, d_Y) normado y $T_h \in \mathcal{T}_B[\pi_h]$,

$$A = 0, \quad B = 2^{\frac{p-1}{p}} \quad y \quad C = 2^{\frac{p-1}{p}} \cdot D_Y; \quad (67)$$

Demostración. Empezamos asumiendo $T_h \in \mathcal{T}_{GM}^Q[\pi_h]$ y consideramos, para cada $x \in X$, la partición $J = B_{h,\delta}(x) \sqcup G_{h,\delta}(x)$ dada por

$$B_{h,\delta}(x) = \{ j \in J : d_Y(y_j, T(x)) \geq \delta \} \quad y \quad G_{h,\delta}(x) = \{ j \in J : d_Y(y_j, T(x)) < \delta \}.$$

Fijando $i \in I_+$, se tiene $\mu_i > 0$ y $T_h(x) = y_i$ para $x \in E_i$, donde y_i es una mediana geométrica $Q \cdot h$ -aproximada para $(y_j)_j$ con pesos $(\pi_{ij}/\mu_i)_j$. Fijando ahora $x \in E_i$, el lema de cercanía 3.2.6 implica

$$d_Y(T_h(x), T(x)) \leq Qh + 2 \sum_{j \in J} \frac{\pi_{ij}}{\mu_i} d_Y(y_j, T(x)). \quad (68)$$

La partición $J = B_{h,\delta}(x) \sqcup G_{h,\delta}(x)$ permite escribir

$$\begin{aligned} \sum_j \frac{\pi_{ij}}{\mu_i} d_Y(y_j, T(x)) &= \sum_{j \in G_{h,\delta}(x)} \frac{\pi_{ij}}{\mu_i} d_Y(y_j, T(x)) + \sum_{j \in B_{h,\delta}(x)} \frac{\pi_{ij}}{\mu_i} d_Y(y_j, T(x)) \\ &\leq \delta \sum_{j \in G_{h,\delta}(x)} \frac{\pi_{ij}}{\mu_i} + D_Y \sum_{j \in B_{h,\delta}(x)} \frac{\pi_{ij}}{\mu_i} \\ &\leq \delta + D_Y \sum_{j \in B_{h,\delta}(x)} \frac{\pi_{ij}}{\mu_i}, \end{aligned}$$

que junto con (68) implica

$$d_Y(T_h(x), T(x)) \leq Qh + 2\delta + 2D_Y \sum_{j \in B_{h,\delta}(x)} \frac{\pi_{ij}}{\mu_i}. \quad (69)$$

A partir de (69), aplicando la desigualdad de Jensen se obtiene

$$\begin{aligned} d_Y^p(T_h(x), T(x)) &\leq \left(Qh + 2\delta + 2D_Y \sum_{j \in B_{h,\delta}(x)} \frac{\pi_{ij}}{\mu_i} \right)^p \\ &\leq 3^{p-1} \left((Qh)^p + (2\delta)^p + (2D_Y)^p \left(\sum_{j \in B_{h,\delta}(x)} \frac{\pi_{ij}}{\mu_i} \right) \right), \end{aligned} \quad (70)$$

donde la segunda desigualdad resulta de observar que $\left(\sum_{j \in B_{h,\delta}(x)} \frac{\pi_{ij}}{\mu_i} \right) \leq 1$. Al ser $x \in E_i$ arbitrario, (70) implica

$$\begin{aligned} &\int_{E_i} d_Y^p(T_h(x), T(x)) d\mu(x) \\ &\leq 3^{p-1} \mu_i ((Qh)^p + (2\delta)^p) + 3^{p-1} (2D_Y)^p \int_{E_i} \left(\sum_{j \in B_{h,\delta}(x)} \frac{\pi_{ij}}{\mu_i} \right) d\mu(x). \end{aligned} \quad (71)$$

Como μ está concentrada en $\bigcup_{i \in I_+} E_i$, sumando (71) sobre $i \in I_+$ obtenemos

$$\begin{aligned} &\int_X d_Y^p(T_h(x), T(x)) d\mu(x) \\ &\leq 3^{p-1} ((Qh)^p + (2\delta)^p) + 3^{p-1} (2D_Y)^p \sum_{i \in I_+} \int_{E_i} \left(\sum_{j \in B_{h,\delta}(x)} \frac{\pi_{ij}}{\mu_i} \right) d\mu(x). \end{aligned} \quad (72)$$

Observar que para $i \in I_+$ y $x \in E_i$ tenemos

$$\sum_{j \in B_{h,\delta}(x)} \frac{\pi_{ij}}{\mu_i} = \sum_{j \in J} \frac{\pi_{ij}}{\mu_i} \mathbb{1}_{B_{h,\delta}(x)}(j) = \sum_{j \in J} \frac{\pi_{ij}}{\mu_i} \delta_{y_j} [B_T(\delta)_x],$$

por lo que la identidad de desintegración (59) implica

$$\sum_{i \in I_+} \int_{E_i} \left(\sum_{j \in B_{h,\delta}(x)} \frac{\pi_{ij}}{\mu_i} \right) d\mu(x) = \sum_{i \in I_+} \int_{E_i} \left(\sum_{j \in J} \frac{\pi_{ij}}{\mu_i} \delta_{y_j} [B_T(\delta)_x] \right) d\mu(x) = \pi_h^s[B_T(\delta)].$$

Finalmente, esta última igualdad junto con (72) implica

$$\begin{aligned} d_p(T_h, T) &\leq \left(3^{p-1} ((Qh)^p + (2\delta)^p) + 3^{p-1} (2D_Y)^p \pi_h^s[B_T(\delta)] \right)^{\frac{1}{p}} \\ &\leq 3^{\frac{p-1}{p}} ((Qh) + (2\delta)) + 3^{\frac{p-1}{p}} (2D_Y) (\pi_h^s[B_T(\delta)])^{\frac{1}{p}}, \end{aligned}$$

lo que prueba (65) y completa la demostración de *i*).

Para probar *ii*), notar que en (69) podemos tomar $Q = 0$ y por lo tanto el lado derecho de (69) tiene sólo dos términos. Luego, al aplicar la desigualdad de Jensen como en (70), el factor 3^{p-1} cambia a 2^{p-1} . Repitiendo la demostración de *i*) con esta modificación menor se obtiene una prueba de *ii*).

Ahora probamos *iii*). En este caso, (Y, d_Y) es normado y T_h se obtiene calculando un baricentro sobre cada E_i con $i \in I_+$. Nuevamente, el lema de cercanía 3.2.6 implica

$$d_Y(T_h(x), T(x)) \leq \sum_j \frac{\pi_{ij}}{\mu_i} d_Y(T(x), y_j)$$

para $i \in I_+$ y $x \in E_i$, lo que mejora (69) a

$$d_Y(T_h(x), T(x)) \leq \delta + D_Y \sum_{j \in B_{h,\delta}(x)} \frac{\pi_{ij}}{\mu_i}.$$

Aplicando la desigualdad de Jensen y teniendo en cuenta las observaciones hechas para la prueba de *ii*), se obtiene una prueba de *iii*) repitiendo los pasos de la prueba de *i*). $\square$

Teorema B Sea (X, Y, μ, ν, c) un problema compacto satisfaciendo la hipótesis (SU) con transporte óptimo T_* . Luego, si $(T_k)_k$ es una sucesión aproximante de proyecciones, para todo $1 \leq p < \infty$ se tiene $d_p(T_k, T_*) \xrightarrow{k} 0$.

Demostración. Por definición de sucesión aproximante de proyecciones, existen $Q \geq 0$ y una sucesión aproximante de planes discretos $(\pi_k)_k$ tales que π_k es un h_k -plan, T_k es una proyección de calidad Q extraída de π_k y $h_k \xrightarrow{k} 0$. Fijando $1 \leq p < \infty$, se deduce a partir de la proposición 5.2.1 la desigualdad

$$d_p(T_k, T_*) \leq A \cdot h_k + B \cdot \delta + C \cdot \pi_k^s[B_{T_*}(\delta)]^{\frac{1}{p}} \quad (73)$$

para todo k y todo $\delta > 0$, donde $B_{T_*}(\delta)$ se define como en (63) tomando $T = T_*$. Notar que las mismas constantes A , B y C funcionan para todo k y todo δ , dado que dependen únicamente de p , D_Y , Q .

Dado $\varepsilon > 0$, tomamos $\delta > 0$ tal que $\delta < \varepsilon/(3B)$. Por la proposición 5.1.4, existe k_1 tal que $\pi_k^s[B_{T_*}(\delta)] < (\varepsilon/(3C))^p$ para $k \geq k_1$. Como $h_k \xrightarrow{k} 0$, existe k_2 tal que $h_k < \varepsilon/(3A)$ para $k \geq k_2$. Tomando $k \geq \max\{k_1, k_2\}$ se deduce $d_p(T_k, T_*) < \varepsilon$ a partir de (73). $\square$

5.3. Discusión

A continuación mostramos que las hipótesis del Teorema B son óptimas. Precisamente, mostramos que el Teorema B falla para $p = \infty$ en el ejemplo 5.3.1, que no vale eliminando la hipótesis (SU) en el ejemplo 5.3.2 y que tampoco vale relajando la continuidad de c a semicontinuidad inferior en el ejemplo 5.3.3.

Ejemplo 5.3.1 El Teorema B es falso para $p = \infty$. Para verlo, tomar $X = Y = [0, 1]$ y $\mu = \nu = \mathcal{L}^1$ la medida de Lebesgue 1-dimensional en el intervalo unitario. Considerar el costo $c(x, y) = (x - y)^2(x - \frac{1}{2})^2(y - \frac{1}{2})^2$. Notar que el problema compacto (X, Y, μ, ν, c)

verifica la hipótesis (SU), siendo la identidad del $[0, 1]$ el único transporte T_* que verifica $C[T_*] = 0$ y por lo tanto el único óptimo.

Para $k \in \mathbb{N}$, considerar las sucesiones de particiones “regulares” $\mathcal{C}_X^k, \mathcal{C}_Y^k$ dadas por

$$E_i^k = \left[\frac{i-1}{2k+1}, \frac{i}{2k+1} \right], \quad x_i^k = \frac{2i-1}{2(2k+1)}, \quad F_j^k = \left[\frac{j-1}{2k+1}, \frac{j}{2k+1} \right] \quad \text{e} \quad y_j^k = \frac{2j-1}{2(2k+1)}$$

para $1 \leq i, j \leq 2k+1$. Esto es, las particiones $\mathcal{C}_X^k, \mathcal{C}_Y^k$ están formadas por $2k+1$ intervalos $\mathcal{L}^1$ -disjuntos de longitud $h_k = \frac{1}{2k+1}$ punteados en sus centros. Para cada h_k -aproximación (X, Y, μ_k, ν_k, c) inducida por $\mathcal{C}_X^k$ y $\mathcal{C}_Y^k$ considerar, como se ve en la fig. 5.3, la solución óptima $\pi_k = (\pi_{ij})_{ij}$ de (FP) dada por

$$\pi_{ij} = \begin{cases} \frac{1}{2k+1} & , \text{ para } i = j \text{ con } i, j \neq 1, k+1, \\ \frac{1}{2k+1} & , \text{ para } (i, j) = (1, k+1) \text{ o } (i, j) = (k+1, 1), \\ 0 & , \text{ si no.} \end{cases}$$

Es inmediato a partir de las definiciones que $I[\pi_k] = 0$, con lo cual que π_k es óptima dado que $c \geq 0$. Por otra parte, como $\text{sop}(\pi_k)$ tiene forma de gráfico de función (ver fig. 5.3), toda proyección T_k de π_k está dada por

$$T_k(x) = \begin{cases} x_i & , \text{ para } x \in E_i, \quad i \neq 1, k+1, \\ x_{k+1} & , \text{ para } x \in E_1, \\ x_1 & , \text{ para } x \in E_{k+1}. \end{cases}$$

Notar que esta expresión para T_k está bien definida sólo μ -a.e., pues los elementos de $\mathcal{C}_X^k$ no son verdaderamente disjuntos. Finalmente, observar que

$$d_\infty(T_k, T_*) \geq \text{ess sup}_{E_1} |T_k - T_*| = \text{ess sup}_{0 \leq x \leq h_k} \left| \frac{1}{2} - x \right| = \frac{1}{2}. \quad \triangle$$

Notar que si (X, Y, μ, ν, c) admitiera más de un transporte óptimo sería posible, en principio, intentar generalizar el enunciado del Teorema B reemplazando $d_p(T_h, T_*)$ por $d_p(T_h, \mathcal{O})$, donde $\mathcal{O} \subset \mathcal{B}(X, Y)$ es el conjunto de transportes óptimos. El problema es que, en tal caso, las proyecciones podrían no estar cerca de ningún transporte de μ en ν , como lo muestra el siguiente ejemplo.

Ejemplo 5.3.2 Considerar los mismos espacios, medidas y particiones del ejemplo 5.3.1, pero cambiando el costo por $c(x, y) = (x - y)^2(1 - x - y)^2$. Es fácil ver que (X, Y, μ, ν, c) es un problema compacto que no verifica la hipótesis (SU). En efecto, es un transporte óptimo toda función boreliana $T : X \rightarrow Y$ con $T_\# \mu = \nu$ y su gráfico contenido en

$$Z_c = \{(x, x) : x \in [0, 1]\} \cup \{(x, 1 - x) : x \in [0, 1]\} \subset [0, 1] \times [0, 1].$$

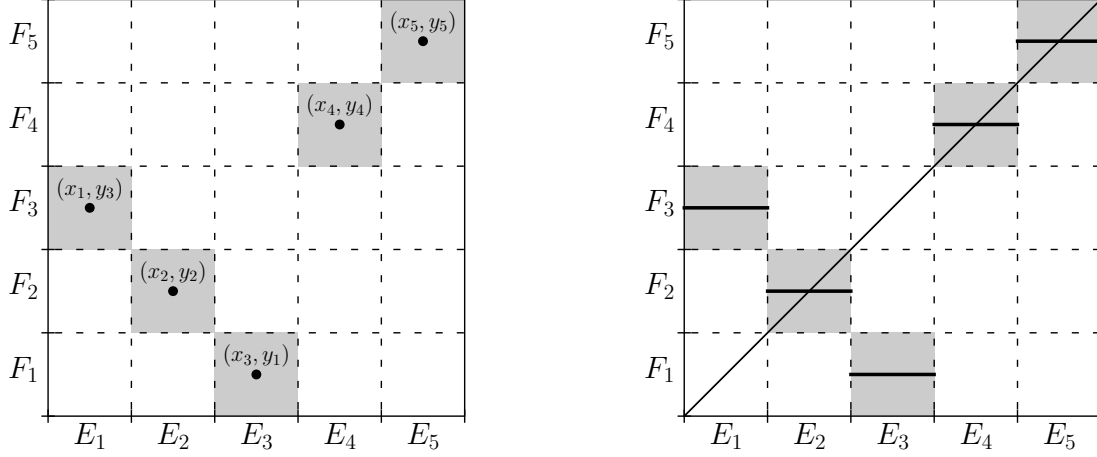


Figura 5.3: A la izquierda, el soporte de π_k . A la derecha, los gráficos del transporte óptimo T_* (sobre la diagonal) y de la proyección T_k (constante a trozos).

Notar que Z_c es el conjunto de ceros de c y que todo h_k -plan concentrado en Z_c es óptimo para (FP). En particular, el h_k -plan óptimo $\pi_k = (\pi_{ij})_{ij}$ dado por

$$\pi_{ij} = \begin{cases} \frac{1}{2(2k+1)} & , \text{ si } x_i^k \neq \frac{1}{2} \neq y_j^k \text{ y } x_i^k = y_j^k \text{ o } x_i^k + y_j^k = 1, \\ \frac{1}{2k+1} & , \text{ si } x_i^k = y_j^k = \frac{1}{2}, \\ 0 & , \text{ si no,} \end{cases}$$

tiene proyección baricéntrica $T_k \equiv \frac{1}{2}$ y por lo tanto $d_p(T_k, \mathcal{O}) \not\rightarrow 0$ para todo p . Esta última afirmación resulta de observar que todo transporte óptimo T para (X, Y, μ, ν, c) debe tener un gráfico esencialmente contenido en Z_c , con lo cual

$$d_p(T_k, T) = \|T_k - T\|_{L^p[0,1]} \geq \|1/2 - T\|_{L^p[3/4,1]} \geq \frac{1}{16}. \quad \triangle$$

Ejemplo 5.3.3 Considerar $X = Y = [0, 1]$ con la métrica usual y tomar $\mu = \nu = \mathcal{L}^1$, siendo $\mathcal{L}^1$ la medida de Lebesgue 1-dimensional en $[0, 1]$. Considerar la función de costo semicontinua inferiormente $c : X \times Y \rightarrow \mathbb{R}$ que toma el valor 0 sobre la diagonal

$$D = \{(x, y) \in X \times Y : x = y\}$$

y el valor 1 en su complemento. Notar que $T_*(x) = x$ es el único transporte óptimo para el problema (X, Y, μ, ν, c) , que sí satisface (SU) pero no es compacto porque c es discontinua. Considerar particiones $\mathcal{C}_X^k, \mathcal{C}_Y^k$ como en el ejemplo 5.3.1, pero moviendo los puntos x_i^k para lograr $(x_i^k, y_j^k) \notin D$. Notar que cualquier solución factible π_k de (FP) es un h_k -plan óptimo con costo $1 > 0 = C[T_*]$. Esto muestra que $I[\pi_k] \not\rightarrow I_*$ y que $d_p(T_k, T_*) \not\rightarrow 0$ si π_k y T_k son como en el ejemplo 5.3.2. $\triangle$

Capítulo 6

Órdenes de convergencia para proyecciones

Como se anticipó en la introducción, en este capítulo probaremos de forma cuantitativa la convergencia de las proyecciones T_h al transporte óptimo T_* . Probaremos primero el Teorema C cuantificando $d_p(T_h, T_*)$ en términos del orden de estabilidad del problema de transporte y de la regularidad de T_* , expresando el primero en términos de s_p y la segunda en términos de la constante y el exponente de Hölder a trozos de T_* . Luego probaremos el Teorema D, cuantificando $d_p(T_h, T_*)$ en términos s_p y de la dispersión del h -plan π_h del cual se extrae T_h .

6.1. Regularidad Hölder a trozos

Definición 6.1.1 Sea (X, d_X) un espacio métrico dotado de una medida de probabilidad $\mu \in \mathcal{P}(X)$ y sea (Y, d_Y) otro espacio métrico. Dados $\alpha \in (0, 1]$, una constante C_α y una función $T \in \mathcal{B}(X, Y)$, decimos que T es α -Hölder a trozos con constante C_α si existe un conjunto cerrado $\Sigma \subset X$ con $\mu[\Sigma] = 0$ y tal que

- i) $X \setminus \Sigma$ es una unión numerable de abiertos disjuntos $\{X_k\}_k$;
- ii) cada restricción $T|_{X_k}$ es α -Hölder con constante C_α , es decir,

$$\sup_k \sup_{\substack{x_1, x_2 \in X_k \\ x_1 \neq x_2}} \frac{d_Y(T(x_1), T(x_2))}{d_X^\alpha(x_1, x_2)} \leq C_\alpha.$$

Decimos que $T \in \mathcal{B}(X, Y)$ tiene *regularidad α -Hölder a trozos* si T es α -Hölder a trozos con constante C_α para alguna constante C_α . Notamos $C_{PW}^{0, \alpha}(X, Y)$ al conjunto de todas las funciones $T \in \mathcal{B}(X, Y)$ con regularidad α -Hölder a trozos.

Partición de $X \setminus \Sigma$ Dado que cada X_k en $i)$ es abierto, toda partición $X \setminus \Sigma = \bigsqcup_k X_k$ por abiertos es numerable si X es separable. Notar también que $i)$ no requiere que los conjuntos X_k sean conexos. Para un ejemplo considerar la función signo $sg : X \rightarrow \mathbb{R}$, donde X es el conjunto totalmente desconexo $[-1, 1] \setminus \mathbb{Q}$ equipado con la métrica de subespacio y la medida de Lebesgue heredada de $\mathbb{R}$. Es claro que sg es un elemento de $C_{PW}^{0,\alpha}(X, \mathbb{R})$ para todo $\alpha \in (0, 1]$, como puede verse tomando $\Sigma = \emptyset$,

$$X \setminus \Sigma = X = X_1 \sqcup X_2, \quad X_1 = (-1, 0) \setminus \mathbb{Q} \quad \text{y} \quad X_2 = (0, 1) \setminus \mathbb{Q}.$$

Notar que particionar $X \setminus \Sigma$ en componentes conexas requeriría una cantidad no numerable de componentes para cualquier elección de Σ con $\mu[\Sigma] = 0$.

Constante de Hölder a trozos En general, la constante de Hölder a trozos C_α en $ii)$ depende de la elección de Σ . Para un ejemplo trivial basta considerar $X = [-1, 1]$ con la métrica usual, tomar la restricción de la medida de Lebesgue a $[-1, 0]$ como μ y definir $T : X \rightarrow \mathbb{R}$ según $T(x) = \mathbb{1}_{(0,1]} \cdot x^\alpha$. Tomando $\Sigma = [0, 1]$ funciona $C_\alpha = 0$, pero tomando $\Sigma = \emptyset$ se requiere $C_\alpha \geq 1$.

Para un ejemplo menos trivial, considerar $X = [-1, 1] \times [-1, 1] \subset \mathbb{R}^2$ con la métrica y la medida heredadas de $\mathbb{R}^2$ y fijando $\alpha \in (0, 1]$ definir $T : X \rightarrow \mathbb{R}$ por

$$T(x_1, x_2) = \mathbb{1}_{(0,1] \times (0,1]}(x_1, x_2) \cdot x_1^\alpha.$$

Para $0 < \theta < \frac{\pi}{4}$, definir Σ_θ como la unión de $[0, 1] \times \{0\}$ y el rayo $\{re^{i\theta} : r \geq 0\} \subset \mathbb{C} \simeq \mathbb{R}^2$. Considerar la partición $X \setminus \Sigma_\theta = X_1 \sqcup X_2$ donde X_1 es la componente conexa más pequeña de $X \setminus \Sigma_\theta$ (ver fig. 6.1). Notar que para $t \approx 0^+$, $x^0 = (t, 0)$ y $x^1 = (t, t \cdot \tan(\theta))$ están en X y por lo tanto vale

$$\frac{|T(x^1) - T(x^0)|}{|x^1 - x^0|^\alpha} = \frac{t^\alpha}{t^\alpha \tan^\alpha(\theta)} = \frac{1}{\tan^\alpha(\theta)}. \quad (74)$$

Aproximando los puntos x^0, x^1 desde dentro de X_2 se tiene a partir de (74) que $T|_{X_2}$ es α -Hölder con constante $\tan^{-\alpha}(\theta)$. Como $T|_{X_1}$ es α -Hölder con constante 1, se deduce que T es α -Hölder a trozos y para $\Sigma = \Sigma_\theta$ la constante óptima es $C_\alpha = \tan^{-\alpha}(\theta) \geq 1$.

¿Por qué considerar regularidad Hölder a trozos? Los resultados de regularidad parcial presentados al final de la sección 1.6 garantizan, bajo ciertas condiciones, que existen cerrados de medida nula $\Sigma_X \subset X$, $\Sigma_Y \subset Y$ con la propiedad de que el transporte óptimo $T_* : X \setminus \Sigma_X \rightarrow Y \setminus \Sigma_Y$ es un homeomorfismo bi- α -Hölder local. En particular, se tiene $T_* \in C_{PW}^{0,\alpha}(X, Y)$ en caso de que T_* sea globalmente α -Hölder en cada componente conexa de $X \setminus \Sigma_X$ con una constante uniformemente acotada. Notar que el clásico ejemplo 1.6.2 dado por Caffarelli presenta este tipo de regularidad para $\alpha = 1$.

La definición 6.1.1 captura el tipo de regularidad exhibido por el transporte óptimo

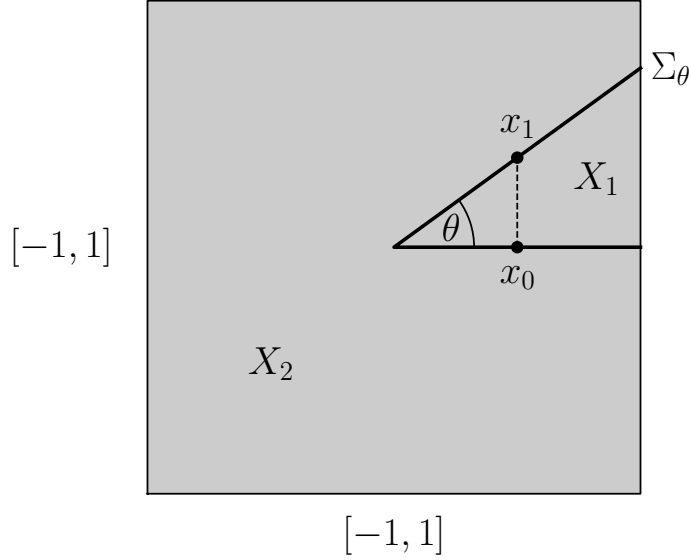


Figura 6.1: Partición de $X \setminus \Sigma_\theta$ en componentes conexas. La variación relativa de T al saltar de un lado a otro de X_1 está controlada por el ángulo θ .

T_* en un caso como el del ejemplo 1.6.2, generalizándola a espacios métricos arbitrarios para un exponente $\alpha \in (0, 1]$. Notar que la constante de Hölder a trozos para T_* es global en los abiertos que particionan $X \setminus \Sigma_X$, lo cual es un requerimiento más fuerte que la regularidad Hölder local de T_* en $X \setminus \Sigma_X$ pero más débil que la regularidad Hölder global de T_* en $X \setminus \Sigma_X$. Por otra parte, T_* no necesita ser un homeomorfismo para ser un elemento de $C_{PW}^{0,\alpha}(X, Y)$, lo que permite considerar casos en los que T_* transporta una medida no atómica μ en una medida atómica ν .

6.2. Transportes compatibles

Definición 6.2.1 Dado un h -plan $\pi_h = (\pi_{ij})_{ij}$ para (X, Y, μ, ν, c) , se dice que $T \in \mathcal{T}(\mu, \nu)$ es *compatible con π_h* si existe una colección finita de conjuntos disjuntos $\{E_i^j\}_{ij} \subset \mathcal{B}(X)$, indexada por $\text{sop}(\pi_h)$, con las siguientes propiedades:

- i) Para $i \in I_+$ fijo, $\{E_i^j\}_{j \in J_+(i)}$ particiona E_i y $\mu[E_i^j] = \pi_{ij}$ para todo $j \in J_+(i)$.
- ii) La inclusión $T(E_i^j) \subset F_j$ vale para todo $(i, j) \in \text{sop}(\pi_h)$.

La colección $\{E_i^j\}_{ij}$ se llama la *partición para T* .

Aclaración 6.2.2 En la figura 6.2 se muestra cómo funciona la partición $\{E_i^j\}_{ij}$ para un transporte compatible T . Es importante notar que la colección $\{E_i^j\}_{ij}$ está unívocamente

determinada por T . En efecto, para $(i, j) \in \text{sop}(\pi_h)$ las condiciones $i)$ y $ii)$ implican

$$E_i^j \subset E_i \cap T^{-1}(F_j) = \bigsqcup_{l \in J_+(i)} E_i^l \cap T^{-1}(F_j) \stackrel{(*)}{=} E_i^j \cap T^{-1}(F_j) \subset E_i^j,$$

con lo cual $E_i^j = E_i \cap T^{-1}(F_j)$. La igualdad $(*)$ resulta de que $E_i^l \cap T^{-1}(F_j) = \emptyset$ si $j \neq l \in J_+(i)$. En caso contrario, se tendría $T(E_i^l) \cap F_j \neq \emptyset$ y la condición $ii)$ implicaría el absurdo $F_l \cap F_j \neq \emptyset$.

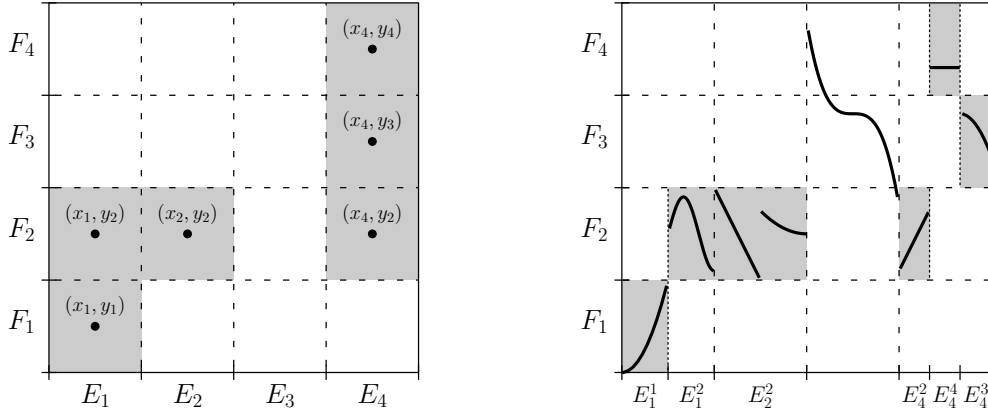


Figura 6.2: El soporte de un h -plan π_h (izquierda) y el gráfico de un transporte T compatible con π_h (derecha). Notar que $\{E_i^j\}_j$ particiona E_i y que $T(E_i^j) \subset F_j$. No hay conjuntos E_i^j contenidos en E_3 porque $\mu[E_3] = 0$.

Lema 6.2.3 Sea π_h un h -plan para un problema compacto (X, Y, μ, ν, c) con μ no atómica. Sean $\rho > 0$ y $\{X_k\}_k \subset \mathcal{B}(X)$ una partición numerable de X . Entonces existe un transporte $T_\rho \in \mathcal{T}(\mu, \nu)$ compatible con π_h satisfaciendo

$$C[T_\rho] \leq I[\pi_h^c] + \rho \quad (75)$$

y tal que su partición $\{E_i^j\}_{ij}$ satisface

$$\mu[E_i^j] \mu[E_{ik}] = \mu[E_{ik}^j] \mu[E_i] \text{ para todo } i, j, k, \quad (76)$$

donde $E_{ik} = E_i \cap X_k$ y $E_{ik}^j = E_i^j \cap X_k$.

Observación 6.2.4 El lema 6.2.3 garantiza la existencia de transportes compatibles. Más aun, el lema establece que, dado un h -plan π_h , existen transportes compatibles con π_h con un costo casi tan bueno como el de π_h (75) y con una partición que verifica las igualdades (76). Dichas igualdades expresan una relación de independendencia entre T_ρ y $\{X_k\}$ desde el punto de vista de las particiones $\mathcal{C}_X$ y $\mathcal{C}_Y$. Precisamente, fijando $i \in I_+$ y definiendo

$P_i = \frac{1}{\mu_i} \mu|_{E_i} \in \mathcal{P}(E_i)$, las igualdades (76) pueden reescribirse como

$$P_i[E_i^j] P_i[E_{ik}] = P_i[E_i^j \cap E_{ik}]$$

luego de dividir ambos lados por μ_i^2 , dado que $E_{ik}^j = E_i^j \cap E_{ik}$. Así, (76) equivale a la relación de independencia

$$P_i[\{T_\rho \in F_j\}] P_i[X_k] = P_i[\{T_\rho \in F_j\} \cap X_k].$$

para $(i, j) \in I_+ \times J_+$ y para todo k .

Demostración del lema 6.2.3. Sea K el conjunto de índices de $\{X_k\}_k$. Para (i, k) en el conjunto numerable $I_+ \times K$, definir $E_{ik} = E_i \cap X_k$ y considerar $\varepsilon_{ik} > 0$ tal que $\sum_{ik} \varepsilon_{ik} = \rho$. Notar que $\{E_{ik}\}_k$ particiona E_i para $i \in I_+$. Definir $F_i = \bigsqcup_{j \in J_+(i)} F_j$ para $i \in I_+$.

Fijar (i, k) en $I_+ \times K$ y definir las medidas

$$\mu_{ik} = \mu|_{E_{ik}} \in \mathcal{M}_+(E_{ik}) \hookrightarrow \mathcal{M}_+(\overline{E_{ik}}), \quad (77)$$

$$\nu_{ik} = \frac{\mu[E_{ik}]}{\mu_i} \sum_{j \in J_+(i)} \frac{\pi_{ij}}{\nu_j} \nu|_{F_j} \in \mathcal{M}_+(F_i) \hookrightarrow \mathcal{M}_+(\overline{F_i}). \quad (78)$$

Notar que $\overline{E_{ik}}$ y $\overline{F_i}$ resultan ser espacios métricos polacos con la métrica de subespacio y, por definición de μ_{ik} y ν_{ik} ,

$$\mu_{ik}[\overline{E_{ik}}] = \mu_{ik}[E_{ik}] = \mu[E_{ik}] = \nu_{ik}[F_i] = \nu_{ik}[\overline{F_i}]. \quad (79)$$

Resulta de (79) que todo transporte $T : \overline{E_{ik}} \rightarrow \overline{F_i}$ para $(\overline{E_{ik}}, \overline{F_i}, \mu_{ik}, \nu_{ik}, c)$ puede modificarse en conjunto μ -nulo para obtener un transporte boreliano de μ_{ik} en ν_{ik} con dominio E_{ik} y codominio F_i : basta redefinirlo en el conjunto μ -nulo $T^{-1}(\overline{F_i} \setminus F_i)$ para conseguir $T(\overline{E_{ik}}) \subset F_i$ y luego restringirlo al conjunto E_{ik} .

Luego, al ser μ_{ik} no atómica, aplicando el teorema 1.5.2 a $(\overline{E_{ik}}, \overline{F_i}, \mu_{ik}, \nu_{ik}, c)$ se obtiene una función boreliana $T_{ik} : E_{ik} \rightarrow F_i$ que transporta μ_{ik} en ν_{ik} con

$$C_{ik}[T_{ik}] \leq I_{ik}^* + \varepsilon_{ik}, \quad (80)$$

donde I_{ik}^* es el valor óptimo del problema de Kantorovich para $(\overline{E_{ik}}, \overline{F_i}, \mu_{ik}, \nu_{ik}, c)$ y C_{ik} es el operador de costo de Monge de (X, Y, μ, ν, c) restringido a E_{ik}

$$C_{ik}[T] = \int_{\overline{E_{ik}}} c(x, T(x)) d\mu_{ik}(x) \stackrel{(77)}{=} \int_{E_{ik}} c(x, T(x)) d\mu(x). \quad (81)$$

Definiendo $X_0 = \bigcup_{i \in I \setminus I_+} E_i$, notar que $\mu[X_0] = 0$ y que $\{E_{ik}\}_{ik}$ es una partición de $X \setminus X_0$ por borelianos. Luego, la función boreliana $T_\rho : X \setminus X_0 \rightarrow Y$ dada por $T_\rho|_{E_{ik}} = T_{ik}$

para $(i, k) \in I_+ \times K$ se extiende a una función boreliana T_ρ en X , bien definida en $X \setminus X_0$ y definida μ -a.e. en X . Definimos $E_i^j = E_i \cap T_\rho^{-1}(F_j)$ para (i, j) en $\text{sop}(\pi_h)$.

Resulta que $T_\rho : X \rightarrow Y$ transporta μ en ν , es compatible con π_h y su partición está dada por $\{E_i^j\}_{ij}$. Más aun, T_ρ y su partición $\{E_i^j\}_{ij}$ satisfacen (75) y (76). Las siguientes afirmaciones prueban estos hechos.

Afirmación #1: La función T_ρ transporta μ en ν .

Para i en I_+ definir las funciones auxiliares $T_i \in \mathcal{B}(E_i, F_i)$ dadas por $T_i = T_\rho|_{E_i}$. Las funciones T_i están bien definidas dado que $\{E_{ik}\}_k$ particiona E_i y, para $k \in K$, se tiene $T_i|_{E_{ik}} = T_\rho|_{E_{ik}} = T_{ik}$ y $T_{ik}(E_{ik}) \subset F_i$.

Para $F \in \mathcal{B}(Y)$ e $i \in I_+$ se obtiene

$$\begin{aligned} \mu[T_i^{-1}(F)] &= \sum_k \mu[T_{ik}^{-1}(F)] = \sum_k \mu_{ik}[T_{ik}^{-1}(F)] = \sum_k \nu_{ik}[F] \\ &\stackrel{(78)}{=} \sum_k \left[\frac{\mu[E_{ik}]}{\mu_i} \sum_{j \in J_+(i)} \frac{\pi_{ij}}{\nu_j} \nu[F_j \cap F] \right] \\ &= \frac{1}{\mu_i} \left(\sum_k \mu[E_{ik}] \right) \sum_{j \in J_+(i)} \frac{\pi_{ij}}{\nu_j} \nu[F_j \cap F] \\ &= \sum_{j \in J_+(i)} \frac{\pi_{ij}}{\nu_j} \nu[F_j \cap F]. \end{aligned} \tag{82}$$

Luego se tiene

$$\begin{aligned} \mu[T_\rho^{-1}(F)] &= \sum_{i \in I_+} \mu[T_i^{-1}(F)] \stackrel{(82)}{=} \sum_{i \in I_+} \sum_{j \in J_+(i)} \frac{\pi_{ij}}{\nu_j} \nu[F_j \cap F] \\ &= \sum_{j \in J_+} \left[\frac{\nu[F_j \cap F]}{\nu_j} \sum_{i \in I_+(j)} \pi_{ij} \right] = \sum_{j \in J_+} \nu[F_j \cap F] = \nu[F] \end{aligned}$$

para todo $F \in \mathcal{B}(Y)$ y por lo tanto $(T_\rho)_\# \mu = \nu$.

Afirmación #2: El transporte T_ρ es compatible con π_h y tiene partición $\{E_i^j\}_{ij}$.

Al ser $E_i^j = E_i \cap T_\rho^{-1}(F_j)$ por definición, es inmediato que los conjuntos $\{E_i^j\}_{ij}$ son mutuamente disjuntos. Para $i \in I_+$ se tiene $T_\rho(E_i) = T_i(E_i) \subset F_i$ y

$$\begin{aligned} E_i &= E_i \cap T_\rho^{-1}(F_i) = E_i \cap T_\rho^{-1} \left(\bigsqcup_{j \in J_+(i)} F_j \right) \\ &= \bigsqcup_{j \in J_+(i)} E_i \cap T_\rho^{-1}(F_j) = \bigsqcup_{j \in J_+(i)} E_i^j, \end{aligned}$$

con lo cual $\{E_i^j\}_j$ particiona E_i . Además, para (i, j) en $\text{sop}(\pi_h)$, se tiene

$$\mu[E_i^j] = \mu[T_\rho^{-1}(F_j) \cap E_i] = \mu[T_i^{-1}(F_j)] \stackrel{(82)}{=} \sum_{l \in J_+(i)} \frac{\pi_{il}}{\nu_l} \nu[F_l \cap F_j] = \pi_{ij},$$

$$T_\rho(E_i^j) = T_\rho(T_\rho^{-1}(F_j) \cap E_i) \subset T_\rho(T_\rho^{-1}(F_j)) \subset F_j,$$

lo que prueba la afirmación.

Afirmación #3: Se verifica la relación de independendencia (76).

Notar que

$$E_{ik}^j = E_i^j \cap X_k = (T_\rho^{-1}(F_j) \cap E_i) \cap X_k = T_\rho^{-1}(F_j) \cap E_{ik} = T_{ik}^{-1}(F_j),$$

con lo cual

$$\begin{aligned} \mu[E_{ik}^j] &= \mu_{ik}[E_{ik}^j] = \mu_{ik}[T_{ik}^{-1}(F_j)] = \nu_{ik}[F_j] \\ &\stackrel{(78)}{=} \frac{\mu[E_{ik}]}{\mu_i} \sum_{l \in J_+(i)} \frac{\pi_{il}}{\nu_l} \nu[F_l \cap F_j] = \frac{\mu[E_{ik}]}{\mu_i} \pi_{ij} \end{aligned}$$

y por lo tanto vale (76).

Afirmación #4: Vale la desigualdad $\sum_{ik} I_{ik}^* \leq I[\pi_h^c]$.

Para $(i, k) \in I_+ \times K$ considerar las medidas

$$\pi_{ik} = \sum_{j \in J_+(i)} \frac{\pi_{ij}}{\mu_i \nu_j} \mu|_{E_{ik}} \otimes \nu|_{F_j} \in \mathcal{M}_+(\overline{E_{ik}} \times \overline{F_i}) \hookrightarrow \mathcal{M}_+(X \times Y).$$

Un simple cálculo muestra que $\pi_{ik} \in \mathcal{A}(\mu_{ik}, \nu_{ik})$, con lo cual $I_{ik}^* \leq I[\pi_{ik}]$. A partir de la definición 4.1.1 es fácil chequear la igualdad $\sum_{ik} \pi_{ik} = \pi_h^c$. Luego, la linealidad del operador de costo I y el teorema de convergencia dominada implican

$$\sum_{ik} I_{ik}^* \leq \sum_{ik} I[\pi_{ik}] = I\left[\sum_{ik} \pi_{ik}\right] = I[\pi_h^c].$$

Afirmación #5: El transporte T_ρ satisface la desigualdad de costo (75).

Observar que

$$\begin{aligned} C[T_\rho] &= \int_X c(x, T_\rho(x)) d\mu(x) = \sum_{ik} \int_{E_{ik}} c(x, T_\rho(x)) d\mu(x) \\ &= \sum_{ik} \int_{E_{ik}} c(x, T_{ik}(x)) d\mu(x) \stackrel{(81)}{=} \sum_{ik} C_{ik}[T_{ik}] \\ &\stackrel{(80)}{\leq} \sum_{ik} I_{ik}^* + \varepsilon_{ik} = \rho + \sum_{ik} I_{ik}^* \leq \rho + I[\pi_h^c], \end{aligned}$$

donde la última desigualdad resulta de la afirmación anterior. $\square$

Lema 6.2.5 Sea (X, Y, μ, ν, c) un problema compacto satisfaciendo la hipótesis (SU), con μ no atómica y transporte óptimo $T_* \in C_{PW}^{0, \alpha}(X, Y)$ con constante C_α . Entonces, dados

$\rho > 0$ y un h -plan π_h para (X, Y, μ, ν, c) , existe un transporte $T_\rho \in \mathcal{T}(\mu, \nu)$ compatible con π_h satisfaciendo

$$C[T_\rho] \leq I[\pi_h^c] + \rho \quad (83)$$

y tal que, para todo $\varepsilon > 0$,

$$\begin{aligned} \pi_h^s[(x, y) \in X \times Y : d_Y(T_*(x), y) \geq \varepsilon + h + C_\alpha h^\alpha] \\ \leq \mu[x \in X : d_Y(T_*(x), T_\rho(x)) \geq \varepsilon]. \end{aligned} \quad (84)$$

Demostración. Existe un cerrado $\Sigma \subset X$ con $\mu[\Sigma] = 0$ y tal que $X \setminus \Sigma$ es particionado por una colección numerable de abiertos $\{X_k\}_{k \in K}$, con la propiedad de que $T_*|_{X_k}$ es α -Hölder con constante C_α para todo $k \in K$. Fijando un h -plan π_h y $\rho > 0$, el lema 6.2.3 provee un transporte T_ρ compatible con π_h , con partición $\{E_i^j\}_{ij}$ y satisfaciendo (75) y (76). Fijamos $\varepsilon > 0$ y terminamos la prueba mostrando que vale (84).

Definimos los conjuntos de puntos

$$\begin{aligned} A &= \{ (x, y) \in X \times Y : d_Y(T_*(x), y) \geq \varepsilon + h + C_\alpha h^\alpha \}, \\ B &= \{ x \in X : d_Y(T_*(x), T_\rho(x)) \geq \varepsilon \}, \end{aligned}$$

el conjunto numerable de índices

$$\Lambda = \{ (i, j, k) : (i, j) \in \text{sop}(\pi_h) \text{ y } k \in K \},$$

y, para $(i, j, k) \in \Lambda$, los conjuntos de puntos

$$A_{ijk} = \{ (x, y) \in E_{ik} \times \{y_j\} : d_Y(T_*(x), y) \geq \varepsilon + h + C_\alpha h^\alpha \}, \quad (85)$$

$$B_{ijk} = \{ x \in E_{ik}^j : d_Y(T_*(x), T_\rho(x)) \geq \varepsilon \}, \quad (86)$$

donde $E_{ik} = E_i \cap X_k$ y $E_{ik}^j = E_i^j \cap X_k$. Dado que T_ρ satisface las relaciones de independencia (76), para $(i, j, k) \in \Lambda$ se obtiene

$$\frac{\pi_{ij}}{\mu_i} \mu[E_{ik}] = \frac{\mu[E_i^j] \mu[E_{ik}]}{\mu[E_i]} = \mu[E_{ik}^j]. \quad (87)$$

Finalmente, notar que (84) se reescribe como la desigualdad $\pi_h^s[A] \leq \mu[B]$, cuya validez es consecuencia de las siguientes afirmaciones.

Afirmación #1: Vale la igualdad $\pi_h^s[A] = \sum_{(i,j,k) \in \Lambda} \pi_h^s[A_{ijk}]$.

Notar que π_h^s está concentrada en $\bigsqcup_{ij} E_i \times \{y_j\}$, donde la unión está indexada por

$\text{sop}(\pi_h)$. Esto implica que, para $M \in \mathcal{B}(X \times Y)$, se tiene

$$\pi_h^s[M] = \sum_{(i,j) \in \text{sop}(\pi_h)} \pi_h^s[M \cap (E_i \times \{y_j\})]. \quad (88)$$

También se tiene

$$\pi_h^s[M \cap (\Sigma \times Y)] \leq \pi_h^s[\Sigma \times Y] = \mu[\Sigma] = 0 \quad (89)$$

y dado que $\{X_k\}_k$ particiona $X \setminus \Sigma$, vale

$$M \cap ((X \setminus \Sigma) \times Y) = \bigsqcup_k M \cap (X_k \times Y). \quad (90)$$

Se deduce a partir de (89) y (90) que

$$\pi_h^s[M] = \pi_h^s[M \cap ((X \setminus \Sigma) \times Y)] = \sum_k \pi_h^s[M \cap (X_k \times Y)]. \quad (91)$$

Dado que M es arbitrario en (88) y (91), en conjunto implican

$$\pi_h^s[M] = \sum_{(i,j,k) \in \Lambda} \pi_h^s[M \cap (E_i \times \{y_j\}) \cap (X_k \times Y)]$$

para $M \in \mathcal{B}(X \times Y)$. La afirmación resulta de tomar $M = A$ y observar que

$$A_{ijk} \stackrel{(85)}{=} A \cap (E_i \times \{y_j\}) \cap (X_k \times Y).$$

Afirmación #2: Vale la igualdad $\mu[B] = \sum_{(i,j,k) \in \Lambda} \mu[B_{ijk}]$.

Definiendo $X_0 = \bigcup_{i \in I \setminus I_+} E_i$, notar que $\mu[X_0] = 0$ y que $\{E_i^j\}_{ij}$ es una partición de $X \setminus X_0$ por borelianos. Como $\{E_{ik}^j\}_k$ particiona $E_i^j \setminus \Sigma$, se obtiene

$$\mu[x \in E_i^j \setminus \Sigma : d_Y(T_*(x), T_\rho(x)) \geq \varepsilon] = \sum_k \mu[x \in E_{ik}^j : d_Y(T_*(x), T_\rho(x)) \geq \varepsilon] \quad (92)$$

para $(i, j) \in \text{sop}(\pi_h)$. Luego, al ser $\mu[X_0] = 0 = \mu[\Sigma]$, resulta

$$\begin{aligned} \mu[B] &= \mu[x \in X : d_Y(T_*(x), T_\rho(x)) \geq \varepsilon] \\ &= \mu[x \in X \setminus X_0 : d_Y(T_*(x), T_\rho(x)) \geq \varepsilon] \\ &= \sum_{(i,j) \in \text{sop}(\pi_h)} \mu[x \in E_i^j : d_Y(T_*(x), T_\rho(x)) \geq \varepsilon] \\ &= \sum_{(i,j) \in \text{sop}(\pi_h)} \mu[x \in E_i^j \setminus \Sigma : d_Y(T_*(x), T_\rho(x)) \geq \varepsilon] \\ &\stackrel{(92)}{=} \sum_{(i,j,k) \in \Lambda} \mu[x \in E_{ik}^j : d_Y(T_*(x), T_\rho(x)) \geq \varepsilon] \stackrel{(86)}{=} \sum_{(i,j,k) \in \Lambda} \mu[B_{ijk}]. \end{aligned}$$

Afirmación #3: La igualdad $\pi_h^s[A_{ijk}] \leq \mu[B_{ijk}]$ vale para $(i, j, k) \in \Lambda$.

Como la afirmación es trivial si $\pi_h^s[A_{ijk}] = 0$, fijar $(i, j, k) \in \Lambda$ con $\pi_h^s[A_{ijk}] > 0$. Por definición de A_{ijk} , esto implica la existencia de $x_{ik} \in E_{ik}$ tal que

$$d_Y(T_*(x_{ik}), y_j) \geq \varepsilon + h + C_\alpha h^\alpha. \quad (93)$$

Fijar $x \in E_{ik}^j$ y observar que (93) implica

$$\varepsilon + h + C_\alpha h^\alpha \leq d_Y(T_*(x_{ik}), T_*(x)) + d_Y(T_*(x), T_\rho(x)) + d_Y(T_\rho(x), y_j). \quad (94)$$

Notar que $x, x_{ik} \in E_{ik} = E_i \cap X_k$. Luego, como $T_*|_{X_k}$ es α -Hölder con constante C_α y $\text{diám}(E_{ik}) \leq \text{diám}(E_i) \leq h$, se obtiene

$$d_Y(T_*(x_{ik}), T_*(x)) \leq C_\alpha h^\alpha. \quad (95)$$

Notar también que $x \in E_{ik}^j \subset E_i^j$ y, al ser $\{E_i^j\}_{ij}$ la partición de T_ρ , resulta a partir de la definición 6.2.1 que $T_\rho(x) \in T(E_{ik}^j) \subset T(E_i^j) \subset F_j$, con lo cual

$$d_Y(T_\rho(x), y_j) \leq \text{diám}(F_j) \leq h. \quad (96)$$

Acotando el lado derecho de (94) por (95) y (96) se obtiene

$$\varepsilon + h + C_\alpha h^\alpha \leq C_\alpha h^\alpha + d_Y(T_*(x), T_\rho(x)) + h,$$

lo que implica

$$\varepsilon \leq d_Y(T_*(x), T_\rho(x)). \quad (97)$$

Como $x \in E_{ik}^j$ es arbitrario, la desigualdad (97) vale en E_{ik}^j y por lo tanto

$$B_{ijk} = \{x \in E_{ik}^j : d_Y(T_*(x), T_\rho(x)) \geq \varepsilon\} = E_{ik}^j. \quad (98)$$

Finalmente,

$$\pi_h^s[A_{ijk}] \underset{(85)}{\leq} \pi_h^s[E_{ik} \times \{y_j\}] \underset{(57)}{=} \frac{\pi_{ij}}{\mu_i} \mu[E_{ik}] \underset{(87)}{=} \mu[E_{ik}^j] \underset{(98)}{=} \mu[B_{ijk}].$$

Afirmación #4: Vale la desigualdad $\pi_h^s[A] \leq \mu[B]$.

A partir de las afirmaciones anteriores se obtiene

$$\pi_h^s[A] = \sum_{(i,j,k) \in \Lambda} \pi_h^s[A_{ijk}] \leq \sum_{(i,j,k) \in \Lambda} \mu[B_{ijk}] = \mu[B]. \quad \square$$

6.3. Demostración del Teorema C

Proposición 6.3.1 *Sea (X, Y, μ, ν, c) un problema compacto satisfaciendo la hipótesis (SU) con transporte óptimo T_* y valor óptimo $I_* = C[T_*]$. Adicionalmente, supongamos que μ es no atómica y que T_* está en $C_{PW}^{0,\alpha}(X, Y)$ con constante C_α . Luego, dados $Q \geq 0$ y $1 \leq p < \infty$, existen constantes A , B y C dependientes sólo de p , C_α , D_Y y Q tales que*

$$d_p(T_h, T_*) \leq A \cdot h + B \cdot h^\alpha + C \cdot \sqrt{s_p(2(I[\pi_h^c] - I_*))} \quad (99)$$

para toda proyección T_h de calidad Q extraída de un h -plan π_h . Más aun, las constantes A , B y C pueden tomarse como sigue:

i) Para $T_h \in \mathcal{T}_{GM}^Q[\pi_h]$,

$$A = 3^{\frac{p-1}{p}} \cdot (2 + Q), \quad B = 2 \cdot 3^{\frac{p-1}{p}} \cdot C_\alpha, \quad C = 4 \cdot 3^{\frac{p-1}{p}} \cdot \sqrt{D_Y}.$$

ii) Para $T_h \in \mathcal{T}_{GM}[\pi_h]$,

$$A = 2^{\frac{p-1}{p}} \cdot 2, \quad B = 2 \cdot 2^{\frac{p-1}{p}} \cdot C_\alpha, \quad C = 4 \cdot 2^{\frac{p-1}{p}} \cdot \sqrt{D_Y}.$$

iii) Para (Y, d_Y) normado y $T_h \in \mathcal{T}_B[\pi_h]$,

$$A = 2^{\frac{p-1}{p}}, \quad B = 2^{\frac{p-1}{p}} \cdot C_\alpha, \quad C = 2 \cdot 2^{\frac{p-1}{p}} \cdot \sqrt{D_Y}.$$

Demostración. Probamos el resultado sólo para el caso i) dado que las pruebas para los casos ii) y iii) son completamente análogas. Para esto, consideramos un h -plan π_h y $T_h \in \mathcal{T}_{GM}^Q[\pi_h]$. Por la proposición 5.2.1, para todo $1 \leq p < \infty$ y todo $\delta > 0$ se tiene

$$d_p(T_h, T_*) \leq 3^{\frac{p-1}{p}} \cdot Q \cdot h + 2 \cdot 3^{\frac{p-1}{p}} \cdot \delta + 2 \cdot 3^{\frac{p-1}{p}} \cdot D_Y \cdot (\pi_h^s[B(\delta)])^{\frac{1}{p}}, \quad (100)$$

donde

$$B(\delta) = \{ (x, y) \in X \times Y : d_Y(T_*(x), y) \geq \delta \}.$$

Por otra parte, fijamos $\rho > 0$ y tomamos T_ρ como en el lema 6.2.5. En particular, para $\varepsilon > 0$ fijo y $\delta = \delta_{h,\varepsilon} := \varepsilon + h + C_\alpha h^\alpha$ la elección de T_ρ implica

$$\begin{aligned} \pi_h^s[B(\delta_{h,\varepsilon})] &\stackrel{(84)}{\leq} \mu[x \in X : d_Y(T_*(x), T_\rho(x)) \geq \varepsilon] \\ &= \mu[x \in X : \varepsilon^{-p} d_Y^p(T_*(x), T_\rho(x)) \geq 1] \\ &\leq \int_X \varepsilon^{-p} d_Y^p(T_*(x), T_\rho(x)) d\mu(x) = \frac{d_p^p(T_*, T_\rho)}{\varepsilon^p}, \end{aligned}$$

con lo cual

$$\pi_h^s[B(\delta_{h,\varepsilon})] \leq \frac{d_p^p(T_*, T_\rho)}{\varepsilon^p}. \quad (101)$$

Tomando $\delta = \delta_{h,\varepsilon}$ en (100) y haciendo uso de la cota (101), se obtiene

$$d_p(T_h, T_*) \leq 3^{\frac{p-1}{p}} \cdot Q \cdot h + 2 \cdot 3^{\frac{p-1}{p}} \cdot \delta_{h,\varepsilon} + 2 \cdot 3^{\frac{p-1}{p}} \cdot D_Y \cdot \frac{d_p(T_*, T_\rho)}{\varepsilon},$$

que puede reescribirse como

$$d_p(T_h, T_*) \leq A \cdot h + B \cdot h^\alpha + 2 \cdot 3^{\frac{p-1}{p}} \cdot \left(\varepsilon + D_Y \cdot \frac{d_p(T_*, T_\rho)}{\varepsilon} \right), \quad (102)$$

donde

$$A = 3^{\frac{p-1}{p}} \cdot (2 + Q) \quad \text{y} \quad B = 2 \cdot 3^{\frac{p-1}{p}} \cdot C_\alpha.$$

Minimizando el lado derecho de (102) sobre $\varepsilon > 0$ se llega a la cota

$$d_p(T_h, T_*) \leq A \cdot h + B \cdot h^\alpha + 4 \cdot 3^{\frac{p-1}{p}} \cdot \sqrt{D_Y \cdot d_p(T_*, T_\rho)}, \quad (103)$$

que resulta de tomar $\varepsilon = \sqrt{D_Y \cdot d_p(T_*, T_\rho)}$ en (102). Dado que $C[T_*] = I_*$ y T_ρ verifica la cota de costo $C[T_\rho] \leq I[\pi_h^c] + \rho$, se tiene

$$C[T_\rho] - C[T_*] = C[T_\rho] - I_* \leq I[\pi_h^c] - I_* + \rho,$$

con lo cual la monotonía de s_p implica

$$d_p(T_*, T_\rho) \leq s_p(C[T_\rho] - C[T_*]) \leq s_p(I[\pi_h^c] - I_* + \rho). \quad (104)$$

Aplicando la cota (104) al lado derecho de (103) se llega a

$$d_p(T_h, T_*) \leq A \cdot h + B \cdot h^\alpha + C \cdot \sqrt{s_p(I[\pi_h^c] - I_* + \rho)}, \quad (105)$$

donde

$$A = 3^{\frac{p-1}{p}} \cdot (2 + Q), \quad B = 2 \cdot 3^{\frac{p-1}{p}} \cdot C_\alpha \quad \text{y} \quad C = 4 \cdot 3^{\frac{p-1}{p}} \cdot \sqrt{D_Y}.$$

Finalizamos la prueba de *i*) eliminando la dependencia de (105) respecto de ρ como sigue. Si $I[\pi_h^c] - I_* > 0$, tomando $\rho = I[\pi_h^c] - I_*$ se obtiene

$$d_p(T_h, T_*) \leq A \cdot h + B \cdot h^\alpha + C \cdot \sqrt{s_p(2(I[\pi_h^c] - I_*))}.$$

En caso contrario, $I[\pi_h^c] - I_* = 0$ y (105) se reescribe como

$$d_p(T_h, T_*) \leq A \cdot h + B \cdot h^\alpha + C \cdot \sqrt{s_p(\rho)}, \quad \forall \rho > 0.$$

Tomando límite para $\rho \rightarrow 0^+$, la continuidad de s_p en 0 implica

$$\begin{aligned} d_p(T_h, T_*) &\leq A \cdot h + B \cdot h^\alpha + C \cdot \sqrt{s_p(0)} \\ &= A \cdot h + B \cdot h^\alpha + C \cdot \sqrt{s_p(2(I[\pi_h^c] - I_*))}. \end{aligned} \quad \square$$

Teorema C Sea (X, Y, μ, ν, c) un problema compacto satisfaciendo la hipótesis (SU) con transporte óptimo T_* . Adicionalmente, supongamos que μ es no atómica y que T_* está en $C_{PW}^{0,\alpha}(X, Y)$ con constante C_α . Luego, dados $Q \geq 0$ y $1 \leq p < \infty$, existen constantes A, B dependientes sólo de $p, Q, \alpha, C_\alpha, D_X$ y D_Y tales que, si $0 < h \leq \min\{D_X, D_Y\}$, vale

$$d_p(T_h, T_*) \leq A h^\alpha + B \sqrt{s_p(4\omega_c(h))}$$

para toda proyección T_h de calidad Q extraída de un h -plan óptimo.

Demostración del Teorema C. Al igual que en la demostración de la proposición 6.3.1, probamos el resultado sólo para el caso $T_h \in \mathcal{T}_{GM}^Q[\pi_h]$, donde π_h es algún h -plan óptimo para (X, Y, μ, ν, c) . Sea $1 \leq p < \infty$ y asumamos $0 < h \leq D := \min\{D_X, D_Y\}$. Por el corolario 4.2.1 sabemos que

$$|I[\pi_h^c] - I_*| \leq 2\omega_c(h),$$

lo que junto con la proposición 6.3.1 y la monotonía de s_p implica

$$\begin{aligned} d_p(T_h, T_*) &\leq A \cdot h + B \cdot h^\alpha + C \cdot \sqrt{s_p(2(I[\pi_h^c] - I_*))} \\ &\leq A \cdot h + B \cdot h^\alpha + C \cdot \sqrt{s_p(4\omega_c(h))}, \end{aligned} \quad (106)$$

donde

$$A = 3^{\frac{p-1}{p}} \cdot (2 + Q), \quad B = 2 \cdot 3^{\frac{p-1}{p}} \cdot C_\alpha \quad \text{y} \quad C = 4 \cdot 3^{\frac{p-1}{p}} \cdot \sqrt{D_Y}.$$

Queremos constantes $\tilde{A}$ y $\tilde{B}$, independientes de h y T_h , tales que

$$d_p(T_h, T_*) \leq \tilde{A} h^\alpha + \tilde{B} \sqrt{s_p(4\omega_c(h))}. \quad (107)$$

Para esto notar que $\phi(t) = t^{1-\alpha}$ es no decreciente en $[0, D]$, pues $\alpha \in (0, 1]$. Luego, vale $h \leq D^{1-\alpha} \cdot h^\alpha$ para todo $h \in [0, D]$ y por lo tanto

$$A \cdot h + B \cdot h^\alpha \leq A \cdot D^{1-\alpha} \cdot h^\alpha + B \cdot h^\alpha = [A \cdot D^{1-\alpha} + B] \cdot h^\alpha.$$

Aplicando esta cota a (106) se ve que vale (107) tomando

$$\begin{aligned} \tilde{B} &= C = 4 \cdot 3^{\frac{p-1}{p}} \cdot \sqrt{D_Y}, \\ \tilde{A} &= [A \cdot D^{1-\alpha} + B] = 3^{\frac{p-1}{p}} \cdot [(2 + Q) \cdot D^{1-\alpha} + 2 \cdot C_\alpha]. \end{aligned}$$

Notar que $\tilde{A}$ y $\tilde{B}$ dependen sólo de $p, \alpha, C_\alpha, D_X, D_Y$ y Q . Las pruebas para los otros casos son análogas y, obviamente, $\tilde{A}$ y $\tilde{B}$ son independientes de Q en tales casos. $\square$

6.4. Discusión

La cota superior para $d_p(T_h, T_*)$ provista por el Teorema C tiene dos términos: uno de orden h^α y otro con un orden controlado por s_p y ω_c . El término de orden h^α es una consecuencia de la regularidad α -Hölder a trozos asumida para T_* . De hecho, para una función T_* general con regularidad α -Hölder no es posible obtener $d_p(T_h, T_*)$ de orden h^β para $\beta > \alpha$, independientemente del método utilizado para construir las funciones simples T_h (ver ejemplo 6.4.1).

Por otra parte, el término controlado por s_p y ω_c es una consecuencia de las propiedades de estabilidad de (X, Y, μ, ν, c) y del método utilizado para construir las aproximaciones T_h . Se puede ver que es necesario contar con un término dependiente de s_p para controlar $d_p(T_h, T_*)$, pues la convergencia $d_p(T_h, T_*) \rightarrow 0$ puede hacerse arbitrariamente lenta aun en el caso de que T_* y c posean regularidad Lipschitz (ver ejemplo 6.4.3).

Ejemplo 6.4.1 Considerar $X = Y = [0, 1]$ y una medida absolutamente continua μ con densidad $f(x)$ coincidiendo con $\frac{1}{x \log^2(x)}$ en un entorno de 0. Sea $T_* : X \rightarrow Y$ dada por $T_*(x) = x^\alpha$ y $\nu = (T_*)_\# \mu$. Para el costo $c(x, y) = |x - y|^2$, la función T_* es el único transporte óptimo para (X, Y, μ, ν, c) .

Para cada $k \in \mathbb{N}$ considerar la partición $\mathcal{P}_k$ de X en k intervalos de longitud $h_k = \frac{1}{k}$. Considerar una sucesión $(T_k)_k$ de funciones simples a valores en Y , de modo tal que cada T_k sea constante en cada intervalo de $\mathcal{P}_k$ y se verifique $d_p(T_k, T_*) \leq C h_k^\beta$ para algún $1 \leq p < \infty$ y todo $k \in \mathbb{N}$, donde C es una constante independiente de k . Veremos que, necesariamente, se tiene $\beta \leq \alpha$.

Llamando y_k al valor que toma T_k en $[0, h_k]$, notar que

$$d_p^p(T_k, T_*) \geq \int_0^{h_k} |T_*(x) - T_k(x)|^p f(x) dx = \int_0^{h_k} \frac{|x^\alpha - y_k|^p}{x \log^2(x)} dx \quad (108)$$

para k suficientemente grande. Fijando $\varepsilon > 0$ y aplicando el lema 6.4.2 que probaremos luego, para k grande se obtiene a partir de (108) la desigualdad

$$C^p h_k^{\beta p} \geq d_p^p(T_k, T_*) \geq c(\alpha, \varepsilon, p) h_k^{\alpha p + \varepsilon},$$

donde $c(\alpha, \varepsilon, p)$ es una constante positiva que depende únicamente de α, ε y p . Luego, se tiene $h_k^{p(\beta - \alpha) - \varepsilon} \geq \tilde{C}$ para alguna constante positiva $\tilde{C}$ independiente de k . Como $h_k \rightarrow 0$, se deduce $p(\beta - \alpha) \leq \varepsilon$ y por lo tanto resulta $\beta \leq \alpha$ tomando $\varepsilon \rightarrow 0^+$. $\triangle$

Lema 6.4.2 Sean $\alpha \in (0, 1)$ y $p \geq 1$. Para todo $\varepsilon > 0$, existe una constante positiva

$c(\alpha, \varepsilon, p)$ dependiente sólo de α , ε y p tal que

$$\int_0^h \frac{|x^\alpha - y|^p}{x \log^2(x)} dx \geq c(\alpha, \varepsilon, p) h^{\alpha p + \varepsilon}$$

para h suficientemente pequeño y para todo $y \in [0, 1]$.

Demostración. Fijar α , p y ε . Notar que $x^\varepsilon \log^2(x) \leq 1$ para x suficientemente pequeño, con lo cual existe $h_0 > 0$ tal que $\frac{1}{x \log^2(x)} \geq \frac{1}{x^{1-\varepsilon}}$ para $x \in (0, h_0]$.

Tomar $0 < h \leq h_0$. Si $0 \leq y < (\frac{h}{2})^\alpha$, para $\frac{3h}{4} \leq x \leq h$ se obtiene

$$\frac{|x^\alpha - y|^p}{x \log^2(x)} \geq c(\alpha, p) \frac{h^{\alpha p}}{x^{1-\varepsilon}}$$

y por lo tanto

$$\int_0^h \frac{|x^\alpha - y|^p}{x \log^2(x)} dx \geq \int_{\frac{3h}{4}}^h \frac{|x^\alpha - y|^p}{x \log^2(x)} dx \geq c(\alpha, \varepsilon, p) h^{\alpha p + \varepsilon}.$$

Las pruebas para los casos $(\frac{h}{2})^\alpha \leq y < h^\alpha$ y $h^\alpha \leq y \leq 1$ son análogas. $\square$

Ejemplo 6.4.3 Consideramos una modificación del ejemplo 5.3.1, preservando X , Y y el costo c pero tomando $\mu = \nu = \sigma$, con σ una medida con $\text{sop}(\sigma) = [0, 1]$ que concentra masa alrededor de 0 y $1/2$. Precisamente, fijamos $\theta > -1$ y tomamos σ con densidad

$$f(x) = \begin{cases} A|x - \frac{1}{2}|^\theta & , \text{ si } 0 < |x - \frac{1}{2}| < \frac{1}{4}, \\ B(\frac{1}{2} - |x - \frac{1}{2}|)^\theta & , \text{ si } \frac{1}{4} < |x - \frac{1}{2}| < \frac{1}{2}, \end{cases}$$

donde A y B son tales que σ es una medida de probabilidad en $[0, 1]$ que verifica

$$\sigma([0, h]) = \sigma([\frac{1}{2} - \frac{h}{2}, \frac{1}{2} + \frac{h}{2}]) \quad \text{para } 0 \leq h \leq \frac{1}{4}. \quad (109)$$

Un simple cálculo muestra que A y B están dadas por

$$A = 2^\theta B = \frac{(1 + \theta) \cdot 2^{1+3\theta}}{1 + 2^\theta}$$

y que resulta

$$\mu([0, h]) = \sigma([0, h]) = \frac{2^{1+2\theta}}{1 + 2^\theta} \cdot h^{1+\theta} \quad \text{para } 0 \leq h \leq \frac{1}{4}.$$

Notar que $T_*(x) = x$ es el único transporte óptimo dado que $c \geq 0$, $C[T_*] = 0$ y ningún otro transporte $T \in \mathcal{T}(\mu, \nu)$ verifica $C[T] = 0$. Por otra parte, para $k \geq 2$ tomar $\mathcal{C}_X^k$, $\mathcal{C}_Y^k$, π_k y T_k como en el ejemplo 5.3.1, observando que $h_k \leq 1/5$ y que la condición (109) garantiza

que π_k es una solución factible de (FP). Luego, para todo $p \in [1, \infty)$ se tiene

$$\begin{aligned} d_p^p(T_k, T_*) &\geq \int_0^{h_k} |T_*(x) - T_k(x)|^p d\mu(x) \geq \left(\inf_{0 \leq x \leq h_k} \left| \frac{1}{2} - x \right| \right) \mu([0, h_k]) \\ &= \left(\frac{1}{2} - h_k \right) \cdot \frac{2^{1+2\theta}}{1+2\theta} \cdot h_k^{1+\theta} \geq \frac{3}{10} \cdot \frac{1}{3} \cdot h_k^{1+\theta} = \frac{1}{10} \cdot h_k^{1+\theta}. \end{aligned} \quad (110)$$

Notar que para $k \geq 2$ se tiene

$$\frac{1}{10} \cdot h_k^{1+\theta} \leq \frac{1}{10} \cdot \left(\frac{1}{5} \right)^{1+\theta} \leq \frac{1}{10} \leq 1,$$

con lo cual (110) implica $d_p(T_k, T_*) \geq \frac{1}{10} \cdot h_k^{1+\theta}$ para $k \geq 2$. Así, dado $\beta > 0$ arbitrario y tomando $\theta = \beta - 1$, se tiene que $(T_k)_k$ converge a T_* en d_p con orden a lo sumo h_k^β . $\triangle$

Terminamos esta sección mostrando un problema compacto, con μ, ν no atómicas y verificando la hipótesis (SU), con la propiedad adicional de que su único transporte óptimo T_* no es α -Hölder a trozos. Por este motivo no puede aplicarse el Teorema C a dicho problema, aunque sí puede aplicársele el Teorema D que probaremos en la próxima sección.

Ejemplo 6.4.4 Construimos una variante del ejemplo 1.6.2, que esencialmente consiste en “pesar” el conjunto singular de T_* . Para esto, consideramos

$$X = \{x \in \mathbb{R}^n : |x| \leq 1\} = X_- \cup X_0 \cup X_+ \quad \text{e} \quad Y = Y_- \cup Y_0 \cup Y_+,$$

donde $X_0 = \{0\} \times \{x \in \mathbb{R}^{n-1} : |x| \leq 1\}$, los conjuntos X_- , X_+ son las semiesferas

$$X_- = \{x \in \mathbb{R}^n : |x| \leq 1 \wedge x_1 \leq 0\} \quad \text{y} \quad X_+ = \{x \in \mathbb{R}^n : |x| \leq 1 \wedge x_1 \geq 0\},$$

y los conjuntos Y_- , Y_0 e Y_+ se definen como

$$Y_0 = X_0, \quad Y_- = \{x - e_1 : x \in X_-\} \quad \text{e} \quad Y_+ = \{x + e_1 : x \in X_+\}.$$

Luego tomamos $c(x, y)$ como el costo cuadrático y μ, ν como

$$\mu = \frac{1}{2} \mathcal{H}^n|_X + \frac{1}{2} \mathcal{H}^{n-1}|_{X_0} \quad \text{y} \quad \nu = \frac{1}{2} \mathcal{H}^n|_Y + \frac{1}{2} \mathcal{H}^{n-1}|_{Y_0},$$

donde $\mathcal{H}^n$ y $\mathcal{H}^{n-1}$ son las medidas de Hausdorff normalizadas para que den medida 1 a los conjuntos X y X_0 respectivamente. El problema (X, Y, μ, ν, c) resulta compacto y es fácil ver que satisface la hipótesis (SU) con transporte óptimo

$$T_*(x) = \mathbb{1}_{\{x_1 < 0\}}(x) \cdot (x - e_1) + \mathbb{1}_{\{x_1 = 0\}}(x) \cdot (x) + \mathbb{1}_{\{x_1 > 0\}}(x) \cdot (x + e_1).$$

Es claro que T_* resulta discontinua en X_0 y se deduce que T_* no posee regularidad α -Hölder a trozos, pues el conjunto Σ debería contener a X_0 pero $\mu[X_0] = \frac{1}{2} > 0$. $\triangle$

6.5. Demostración del Teorema D

Proposición 6.5.1 Sea (X, Y, μ, ν, c) un problema compacto verificando la hipótesis (SU) con transporte óptimo T_* y valor óptimo $I_* = C[T_*]$. Supongamos, adicionalmente, que μ es no atómica. Luego, dados $Q \geq 0$ y $1 \leq p < \infty$, existen constantes A y B dependientes sólo de p y Q tales que

$$d_p(T_h, T_*) \leq A \cdot h + B \cdot \|D(\pi_h)\|_p + s_p(2(I[\pi_h^c] - I_*)) \quad (111)$$

para toda proyección T_h de calidad Q extraída de un h -plan π_h . Más aun, las constantes A y B pueden tomarse como sigue:

i) Para $T_h \in \mathcal{T}_{GM}^Q[\pi_h]$,

$$A = 2^{1-\frac{1}{p}} \cdot (Q+1) \quad y \quad B = 2^{2-\frac{1}{p}}.$$

ii) Para $T_h \in \mathcal{T}_{GM}[\pi_h]$,

$$A = 2^{1-\frac{1}{p}} \quad y \quad B = 2^{2-\frac{1}{p}}.$$

iii) Para (Y, d_Y) normado y $T_h \in \mathcal{T}_B[\pi_h]$,

$$A = 2^{1-\frac{1}{p}} \quad y \quad B = 2^{1-\frac{1}{p}}.$$

Demostración. Probamos el resultado sólo para el caso i) dado que las pruebas para los casos ii) y iii) son completamente análogas. Para esto, consideramos un h -plan π_h y $T_h \in \mathcal{T}_{GM}^Q[\pi_h]$.

Sea $\rho > 0$. Por el lema 6.2.3, existe un transporte compatible T_ρ con partición $\{E_i^j\}_{ij}$ y tal que $C[T_\rho] \leq I[\pi_h^c] + \rho$.

Sean $i \in I_+$ y $x \in E_i$. Luego $x \in E_i^j$ para algún $j \in J_+(i)$ y

$$d_Y(T_h(x), T_\rho(x)) \leq d_Y(T_h(x), y_j) + d_Y(y_j, T_\rho(x)) \leq d_Y(T_h(x), y_j) + h, \quad (112)$$

dado que $T_\rho(x) \in F_j$. Por el lema de cercanía 3.2.6 se tiene

$$d_Y(T_h(x), y_j) \leq Q \cdot h + 2 \sum_{k \in J_+(i)} \frac{\pi_{ik}}{\mu_i} d_Y(y_k, y_j) \leq Q \cdot h + 2 \cdot D(\pi_h)(i),$$

y aplicando esta cota al lado derecho de (112) resulta

$$d_Y(T_h(x), T_\rho(x)) \leq (Q+1) \cdot h + 2 \cdot D(\pi_h)(i).$$

Como esto vale para $i \in I_+$ y $x \in E_i$, se deducen las desigualdades

$$\begin{aligned} \int_{E_i} d_Y^p(T_h(x), T_\rho(x)) d\mu(x) &\leq [(Q+1) \cdot h + 2 \cdot D(\pi_h)(i)]^p \mu_i \\ &\leq 2^{p-1} [(Q+1)^p \cdot h^p + 2^p \cdot (D(\pi_h)(i))^p] \mu_i \\ &\leq 2^{p-1} \cdot (Q+1)^p \cdot h^p \cdot \mu_i + 2^{2p-1} \cdot (D(\pi_h)(i))^p \cdot \mu_i \end{aligned}$$

para $i \in I_+$. Observando que las desigualdades anteriores valen trivialmente para $i \in I \setminus I_+$, sumando sobre $i \in I$ se obtiene

$$\int_X d_Y^p(T_h(x), T_\rho(x)) d\mu(x) \leq 2^{p-1} \cdot (Q+1)^p \cdot h^p + 2^{2p-1} \cdot \|D(\pi_h)\|_p^p,$$

y entonces

$$d_p(T_h, T_\rho) \leq 2^{1-\frac{1}{p}} \cdot (Q+1) \cdot h + 2^{2-\frac{1}{p}} \cdot \|D(\pi_h)\|_p. \quad (113)$$

Como en la prueba de la proposición 6.3.1, se tiene

$$d_p(T_\rho, T_*) \leq s_p(I[\pi_h^c] - I_* + \rho). \quad (114)$$

Dado que $d_p(T_h, T_*) \leq d_p(T_h, T_\rho) + d_p(T_\rho, T_*)$, las desigualdades (113) y (114) implican

$$d_p(T_h, T_*) \leq A \cdot h + B \cdot \|D(\pi_h)\|_p + s_p(I[\pi_h^c] - I_* + \rho),$$

donde

$$A = 2^{1-\frac{1}{p}} \cdot (Q+1) \quad \text{y} \quad B = 2^{2-\frac{1}{p}}.$$

Eliminando la dependencia de ρ como en la prueba de la proposición 6.3.1 resulta

$$d_p(T_h, T_*) \leq A \cdot h + B \cdot \|D(\pi_h)\|_p + s_p(2(I[\pi_h^c] - I_*)). \quad \square$$

Teorema D Sea (X, Y, μ, ν, c) un problema compacto. Supongamos, adicionalmente, que el problema satisface (SU) con transporte óptimo T_* y que μ es no atómica. Luego, dados $Q \geq 0$ y $1 \leq p < \infty$, existen constantes A, B dependientes sólo de p y Q tales que

$$d_p(T_h, T_*) \leq Ah + B \|D(\pi_h)\|_p + s_p(4\omega_c(h))$$

para toda proyección T_h de calidad Q extraída de un h -plan óptimo π_h .

Demostración del Teorema D. La prueba es análoga a la dada para el Teorema C, pero aplicando la desigualdad (111) provista por la proposición 6.5.1 en lugar de la desigualdad análoga provista por la proposición 6.3.1. Omitimos los detalles. $\square$

Apéndices

Apéndice A

Medida en espacios métricos

En este apéndice resumimos las definiciones y resultados fundamentales sobre medidas en espacios métricos y topológicos. Gran parte de los resultados estudiados en esta tesis se formulan en el contexto de espacios métricos polacos, principalmente porque algunos resultados citados en este apéndice requieren esa hipótesis para poder probarse. Pueden leerse pruebas completas de todos ellos en [22, 104, 112, 114].

A.1. Medidas regulares

En lo que sigue X es siempre un espacio métrico o topológico y notamos $\mathcal{B}(X)$ a su σ -álgebra de Borel. Llamamos *medida boreliana* en X a toda medida, positiva o con signo, definida en el espacio medible $(X, \mathcal{B}(X))$. Una medida boreliana μ en X es *finita* si $\sup_{E \in \mathcal{B}(X)} |\mu[E]| < \infty$, que se reduce a la condición $\mu[X] < \infty$ cuando μ es positiva. Notamos $\mathcal{M}(X)$ al conjunto de las medidas borelianas con signo y finitas en X , $\mathcal{M}^+(X)$ al subconjunto de $\mathcal{M}(X)$ formado por las medidas positivas y $\mathcal{P}(X)$ al subconjunto de $\mathcal{M}^+(X)$ formado por las medidas de probabilidad.

El conjunto $\mathcal{M}(X)$ admite estructura de espacio vectorial con las operaciones

$$(\mu + \nu)[E] = \mu[E] + \nu[E] \quad \text{y} \quad (\lambda \cdot \mu)[E] = \lambda \cdot \mu[E]$$

definidas para $\mu, \nu \in \mathcal{M}(X)$, $\lambda \in \mathbb{R}$ y $E \in \mathcal{B}(X)$. Con esta estructura lineal $\mathcal{M}^+(X)$ es un cono convexo y $\mathcal{P}(X)$ un conjunto convexo. Toda medida $\mu \in \mathcal{M}(X)$ se descompone de manera única como la diferencia de dos medidas μ^+ y μ^- en $\mathcal{M}^+(X)$. La unicidad debe entenderse en el sentido de que si $\mu^+ - \mu^- = \mu = \mu_1 - \mu_2$ para μ_1 y μ_2 en $\mathcal{M}^+(X)$, entonces $\mu_1 \geq \mu^+$ y $\mu_2 \geq \mu^-$. Esta descomposición se conoce como la *descomposición de Hahn–Jordan* de μ . Para cada medida $\mu \in \mathcal{M}(X)$ se define la *medida variación total* de μ como la medida positiva $|\mu| = \mu^+ + \mu^-$ y se llama *variación total* de μ al número no

negativo

$$|\mu|[X] = \sup \left\{ \sum_i |\mu[E_i]| : (E_i)_i \text{ partición boreliana y numerable de } X \right\}.$$

Operaciones con medidas

Para producir nuevas medidas a partir de medidas ya conocidas consideramos tres operaciones elementales: combinaciones lineales en $\mathcal{M}(X)$, restricciones y productos. Las combinaciones lineales no necesitan ninguna explicación. La *restricción* de una medida $\mu \in \mathcal{M}(X)$ a un boreliano $E \in \mathcal{B}(X)$ es la medida $\mu|_E \in \mathcal{M}(X)$ definida por la igualdad $\mu|_E[A] = \mu[A \cap E]$ para todo $A \in \mathcal{B}(X)$. Notar que $\mu|_E$ es también un elemento de $\mathcal{M}(E)$. Finalmente, dados espacios topológicos X y Y junto con medidas $\mu \in \mathcal{M}(X)$ y $\nu \in \mathcal{M}(Y)$, la *medida producto* de μ y ν es la única medida $\mu \otimes \nu \in \mathcal{M}(X \times Y)$ que verifica

$$(\mu \otimes \nu)[E \times F] = \mu[E]\nu[F] \quad \text{para } E \in \mathcal{B}(X) \text{ y } F \in \mathcal{B}(Y).$$

Soporte y concentración de medidas

Dada una medida positiva μ en X , el *soporte* de μ es el conjunto cerrado

$$\text{sop}(\mu) = \{x \in X : \text{para todo abierto } U \ni x \text{ vale } \mu[U] > 0\}.$$

Para $\mu \in \mathcal{M}(X)$ su *soporte* es el conjunto cerrado $\text{sop}(\mu) = \text{sop}(\mu^+) \cup \text{sop}(\mu^-)$, donde μ^+ y μ^- son las medidas de la descomposición de Hahn–Jordan de μ . Una medida $\mu \in \mathcal{M}(X)$ está *concentrada* en un conjunto $E \in \mathcal{B}(X)$ si vale $\mu|_E = \mu$. Dicha condición equivale a $|\mu|[X \setminus E] = 0$, que en el caso de μ positiva y finita se simplifica a $\mu[E] = \mu[X]$.

Teorema $\pi - \lambda$ de Dynkin

No siempre es inmediato concluir la igualdad de dos medidas coincidentes sobre una σ -álgebra $\Sigma \subset 2^X$. En general basta probar la igualdad sobre una familia de conjuntos cerrada por intersecciones y suficientemente grande como para generar la σ -álgebra Σ . Ese es el contenido del Teorema $\pi - \lambda$ que enunciamos a continuación.

Un subconjunto $\Pi \subset 2^X$ es un π -sistema si es cerrado por intersecciones finitas, es decir, si verifica $A \cap B \in \Pi$ cada vez que $A \in \Pi$ y $B \in \Pi$. Por otra parte, un subconjunto $\Lambda \subset 2^X$ es un λ -sistema si verifica las siguientes tres propiedades:

- i) Contiene al conjunto vacío: $\emptyset \in \Lambda$.
- ii) Es cerrado por complementos: si $A \in \Lambda$ entonces $X \setminus A \in \Lambda$.

iii) Es cerrado por uniones numerables disjuntas: si los conjuntos $A_1, A_2, \dots \in \Lambda$ son disjuntos dos a dos, entonces $\bigcup_i A_i \in \Lambda$.

Teorema A.1.1 (Dynkin) Sean $\Pi \subset \Lambda \subset 2^X$ tales que Π es un π -sistema y Λ es un λ -sistema. Entonces $\sigma(\Pi) \subset \Lambda$, es decir, Λ contiene a la σ -álgebra generada por Π .

Corolario A.1.2 Sean μ, ν medidas finitas definidas sobre una σ -álgebra $\Sigma \subset 2^X$ y tales que $\mu[X] = \nu[X]$. Se tiene entonces que $\{A \in \Sigma : \mu[A] = \nu[A]\}$ es un λ -sistema, de modo que si μ y ν coinciden sobre un π -sistema $\Pi \subset 2^X$ con $\sigma(\Pi) = \Sigma$ resulta $\mu = \nu$.

Medidas regulares

Para poder hacer análisis en $\mathcal{M}(X)$ interesan las medidas con buenas propiedades de aproximación. Una primera noción en este sentido es la de *medidas regulares*. Para un espacio topológico X las medidas de $\mathcal{M}(X)$ no son en general regulares, por lo que es necesario requerir axiomas de separación. En nuestro caso no será un problema porque trabajaremos con espacios métricos.

Definición A.1.3 Una medida positiva μ en un espacio topológico X es *regular* si

$$\mu[E] = \sup \{ \mu[F] : F \subset E \text{ cerrado} \} = \inf \{ \mu[U] : U \supset E \text{ abierto} \}$$

para todo $E \in \mathcal{B}(X)$. Cuando μ es una medida con signo, decimos que μ es *regular* si lo son las medidas μ^+ y μ^- de su descomposición de Hahn–Jordan.

Lema A.1.4 Sea X un espacio topológico y sean $\mu, \nu \in \mathcal{M}(X)$ medidas regulares. Luego, son equivalentes:

- i) $\mu = \nu$,
- ii) μ y ν coinciden sobre los cerrados,
- iii) μ y ν coinciden sobre los abiertos.

Teorema A.1.5 Sea X un espacio métrico. Luego, toda medida boreliana y finita en X es regular. En particular, para $\mu, \nu \in \mathcal{M}(X)$ las proposiciones i), ii) y iii) del lema anterior son también equivalentes a

- iv) $\int_X f d\mu = \int_X f d\nu$ para toda $f : X \rightarrow \mathbb{R}$ acotada y uniformemente continua.

A.2. Convergencia débil de medidas

Se define la norma $\|\mu\| := |\mu|[X]$ que hace de $\mathcal{M}(X)$ un espacio de Banach. Ésta induce una *topología fuerte* en $\mathcal{M}(X)$, pero para el estudio de medidas en $\mathcal{P}(X)$ resulta más

fructífero trabajar con una *topología débil* que definimos vía funciones de test. Para esto, dado un espacio topológico X , definimos los conjuntos

$$\begin{aligned} C(X) &= \{ f : X \rightarrow \mathbb{R} : f \text{ continua} \}, & C_b(X) &= \{ f \in C(X) : f \text{ acotada} \}, \\ C_0(X) &= \{ f \in C(X) : \lim_{x \rightarrow \infty} f(x) = 0 \}, & C_c(X) &= \{ f \in C(X) : \text{sop}(f) \text{ compacto} \}, \end{aligned}$$

donde $\lim_{x \rightarrow \infty} f(x) = 0$ significa que para todo $\varepsilon > 0$ existe un compacto $K \subset X$ tal que $|f| < \varepsilon$ en $X \setminus K$. Notar que valen las inclusiones

$$C_c(X) \subset C_0(X) \subset C_b(X) \subset C(X).$$

En general, la norma $\|\cdot\|_\infty$ está bien definida en $C_c(X)$, $C_b(X)$ y $C_0(X)$ pero no en $C(X)$. Dicha norma hace de $C_b(X)$ y $C_0(X)$ espacios de Banach y de $C_c(X)$ un subespacio denso de $C_0(X)$. En general, $C_c(X)$ no es un espacio de Banach con $\|\cdot\|_\infty$.

Teorema A.2.1 (Riesz–Markov) *Sea X un espacio topológico localmente compacto y Hausdorff. Dado un funcional continuo $\psi \in C_0(X)^*$, existe una única medida con signo boreliana regular μ en X tal que*

$$\psi(f) = \int_X f d\mu$$

para toda f en $C_0(X)$. Más aun, la norma de ψ como funcional lineal coincide con la variación total de μ .

La dualidad del teorema de Riesz–Markov sugiere considerar en $\mathcal{M}(X)$ una topología débil interpretando dicho espacio como $C_0(X)^*$. Esta topología, que en términos precisos es la topología débil-* en $\mathcal{M}(X)$, es más débil que la topología que vamos a considerar, pues en lugar de tomar funciones de test en $C_0(X)$ vamos a hacerlo en $C_b(X)$.

Definición A.2.2 Sea X un espacio topológico. Dadas una sucesión $(\mu_n)_n \subset \mathcal{M}(X)$ y una medida $\mu \in \mathcal{M}(X)$, decimos que $(\mu_n)_n$ *converge débilmente* a μ si para toda $f \in C_b(X)$ tenemos $\int_X f d\mu_n \xrightarrow{n} \int_X f d\mu$. Cuando esto sucede, lo notamos $\mu_n \longrightarrow \mu$.

Ejemplo A.2.3 Sea $X = \mathbb{R}$ con la topología usual, $\mu_n = \delta_n$ y $\mu = 0$. Notar que $(\mu_n)_n$ no converge débilmente a μ . En efecto, tomando la función de test $f = 1 \in C_b(X)$ se obtiene

$$\int_X f d\mu_n = f(n) = 1 \not\rightarrow 0 = \int_X f d\mu.$$

Si utilizáramos los elementos de $C_0(X)$ como funciones de test para definir la noción de convergencia débil entonces $(\mu_n)_n$ sí convergería a μ , pues

$$\int_X f d\mu_n = f(n) \longrightarrow 0 = \int_X f d\mu \quad \text{para } f \in C_0(X).$$

Notar que $\mathcal{P}(X)$ no resultaría cerrado definiendo así la convergencia débil. El “problema” está en que las funciones de $C_0(X)$ no detectan la masa que escapa al infinito. $\triangle$

Aclaración A.2.4 La noción de convergencia débil considerada, i.e. la que resulta al utilizar los elementos de $C_b(X)$ como funciones de test, se conoce en la bibliografía como *narrow convergence*. La que resulta de utilizar los elementos de $C_0(X)$ como funciones de test, se conoce como *vague convergence*.

Lema A.2.5 Si X es un espacio métrico entonces los límites débiles en $\mathcal{M}(X)$ son únicos y tanto $\mathcal{M}^+(X)$ como $\mathcal{P}(X)$ son cerrados para la topología de la convergencia débil.

Ejemplo A.2.6 La convergencia débil de medidas en espacios topológicos no posee las buenas propiedades del lema anterior. Para ilustrar este hecho, consideramos el espacio topológico de Sierpiński $X = \{0, 1\}$, donde los abiertos son los conjuntos $\emptyset$, $\{1\}$ y X . Notar que $\mathcal{B}(X)$ coincide con el conjunto de partes de X y que las únicas funciones continuas en $C_b(X)$ son las constantes.

Consideramos la sucesión $(\mu_n)_n$ definida por $\mu_n = \delta_0$ y tomamos $\mu = 2\delta_0 - \delta_1$. Dada $f \in C_b(X)$, resulta $f \equiv c$ para alguna constante c y por lo tanto

$$\int_X f d\mu_n = f(0) = c = 2 \cdot c - c = 2 \cdot f(0) - f(1) = \int_X f d\mu,$$

con lo cual $\mu_n \longrightarrow \mu$. Como $(\mu_n)_n \subset \mathcal{P}(X)$ y $\mu \notin \mathcal{M}^+(X)$, esto prueba que ni $\mathcal{M}^+(X)$ ni $\mathcal{P}(X)$ son cerrados para la topología débil de $\mathcal{M}(X)$. Dado que $(\mu_n)_n$ obviamente converge a $\delta_0 \neq \mu$, este mismo ejemplo muestra que la unicidad de límites débiles en $\mathcal{M}(X)$ también falla cuando X es un espacio topológico arbitrario. $\triangle$

Convergencia débil en $\mathcal{P}(X)$

El siguiente teorema da algunas condiciones equivalentes a la convergencia débil de medidas en $\mathcal{P}(X) \subset \mathcal{M}(X)$.

Teorema A.2.7 Sea X un espacio métrico, $(\mu_n)_n$ una sucesión de medidas de probabilidad en X y $\mu \in \mathcal{P}(X)$. Son equivalentes:

- i) $\mu_n \longrightarrow \mu$;
- ii) $\int_X f d\mu_n \xrightarrow{n} \int_X f d\mu$ para toda f uniformemente continua y acotada en X ;
- iii) $\limsup_n \mu_n[F] \leq \mu[F]$ para todo cerrado $F \subset X$;
- iv) $\liminf_n \mu_n[U] \geq \mu[U]$ para todo abierto $U \subset X$;
- v) $\mu_n[B] \xrightarrow{n} \mu[B]$ para todo boreliano $B \in \mathcal{B}(X)$ con $\mu[\partial B] = 0$.

Dado un espacio métrico separable (X, d) es posible definir una métrica d_P en $\mathcal{P}(X)$ de modo que toda sucesión de medidas de probabilidad $(\mu_n)_n$ converja a μ si y sólo si $d_P(\mu_n, \mu) \xrightarrow{n} 0$. En otras palabras, es posible metrizar la convergencia débil en $\mathcal{P}(X)$.

Definición A.2.8 Sea (X, d) un espacio métrico. Dado un subconjunto no vacío A de X y $\alpha > 0$, definimos

$$A_\alpha = \{x \in X : d(x, A) < \alpha\},$$

donde $d(x, A) = \inf \{d(x, a) : a \in A\}$. Definimos también $\emptyset_\alpha = \emptyset$ para todo $\alpha > 0$. Para cada par de medidas $\mu, \nu \in \mathcal{P}(X)$ definimos

$$d_P(\mu, \nu) = \inf \left\{ \alpha > 0 : \mu[B] \leq \nu[B_\alpha] + \alpha \text{ y } \nu[B] \leq \mu[B_\alpha] + \alpha \quad \forall B \in \mathcal{B}(X) \right\}.$$

El siguiente teorema muestra que tenemos un espacio métrico $(\mathcal{P}(X), d_P)$ y que éste metriza la convergencia débil en $\mathcal{P}(X)$ cuando (X, d) es separable. La métrica d_P se llama *métrica de Prokhorov* inducida por d .

Teorema A.2.9 Sea (X, d) un espacio métrico. Entonces:

- i) La función d_P es una métrica en $\mathcal{P}(X)$.
- ii) Para $(\mu_n)_n \subset \mathcal{P}(X)$ y $\mu \in \mathcal{P}(X)$ vale que $d_P(\mu_n, \mu) \xrightarrow{n} 0$ implica $\mu_n \longrightarrow \mu$.
- iii) Si (X, d) es separable entonces $(\mathcal{P}(X), d_P)$ también lo es y, en ese caso, para toda sucesión $(\mu_n)_n$ y toda $\mu \in \mathcal{P}(X)$ resulta $d_P(\mu_n, \mu) \xrightarrow{n} 0$ si y sólo si $\mu_n \longrightarrow \mu$.

A.3. Rigidez y Teorema de Prokhorov

Para hacer análisis siempre es útil contar con teoremas de compacidad que permitan construir subsucesiones convergentes a puntos límite de interés. El teorema de Prokhorov que enunciamos más abajo da condiciones necesarias y suficientes para que una familia $\mathcal{F}$ de medidas en $(\mathcal{P}(X), d_P)$ tenga clausura compacta o, equivalentemente, para que toda sucesión de $\mathcal{F}$ tenga alguna subsucesión convergente.

Una situación fácil de solucionar es la que se presenta cuando $(\mathcal{P}(X), d_P)$ es compacto. En este caso toda sucesión de medidas de probabilidad tendrá subsucesión convergente. La siguiente proposición dice exactamente cuándo sucede esto.

Proposición A.3.1 Un espacio métrico (X, d) es compacto si y sólo si $(\mathcal{P}(X), d_P)$ lo es.

Sin ninguna hipótesis sobre el espacio métrico (X, d) no es posible dar un teorema de compacidad. Afortunadamente, exigiendo completitud y separabilidad a (X, d) es posible obtener un teorema de compacidad usando la noción de *rigidez*.

Definición A.3.2 Una medida $\mu \in \mathcal{P}(X)$ es *rígida* (*tight*) si para todo $\varepsilon > 0$ existe $K \subset X$ compacto tal que $\mu[X \setminus K] < \varepsilon$ o, equivalentemente, tal que $\mu[K] \geq 1 - \varepsilon$.

Observación A.3.3 En el caso de trabajar con medidas con signo, en la definición de rigidez debe sustituirse la condición $\mu[X \setminus K] < \varepsilon$ por $|\mu|[X \setminus K] < \varepsilon$. Notar que ambas

condiciones son equivalentes entre sí cuando μ es una medida positiva. Notar también que cuando X es compacto toda medida boreliana y finita en X es rígida.

Lema A.3.4 *Si X es un espacio métrico y μ una medida boreliana, finita y rígida en él, entonces μ está concentrada en un subconjunto σ -compacto de X .*

Teorema A.3.5 *Si μ es una medida boreliana, finita y rígida en X , entonces para cada $B \in \mathcal{B}(X)$ vale que*

$$\mu(B) = \sup \{ \mu[K] : K \subset B \text{ compacto} \}.$$

Para hacer una buena parte de la teoría de transporte óptimo es útil contar con la hipótesis de que las medidas que aparecen son rígidas. Una forma de garantizar esta hipótesis es trabajar en espacios polacos, dado que estos definen una clase razonablemente amplia de espacios en los que cualquier medida boreliana y finita es rígida.

Definición A.3.6 Un *espacio polaco* es un espacio métrico completo y separable.

Teorema A.3.7 *Toda medida boreliana y finita en un espacio polaco es rígida.*

Definición A.3.8 Si X es un espacio métrico, una familia de medidas de probabilidad $\mathcal{F} \subset \mathcal{P}(X)$ se dice *rígida* (*tight*) si para todo $\varepsilon > 0$ existe un subconjunto compacto K de X tal que $\mu[K] \geq 1 - \varepsilon$ para toda μ en $\mathcal{F}$.

Observación A.3.9 Cuando $\mathcal{F}$ es un singleton $\{\mu\}$, $\mathcal{F}$ es rígida si y sólo si μ lo es.

Observación A.3.10 La rigidez de una familia de medidas elimina la posibilidad de que la masa se escape al infinito, como sucede en el ejemplo A.2.3. De hecho, vale que si $(\mu_n)_n$ es rígida, entonces $(\mu_n)_n$ converge a μ en la dualidad con $C_0(X)$ si y sólo si lo hace en la dualidad con $C_b(X)$. Con las definiciones de la aclaración A.2.4, una sucesión rígida es vaguely convergent si y sólo si es narrowly convergent.

Teorema A.3.11 (Prokhorov) *Sea (X, d) un espacio métrico polaco y consideremos en $\mathcal{P}(X)$ la métrica de Prokhorov d_P . Para toda familia $\mathcal{F} \subset \mathcal{P}(X)$ son equivalentes:*

- i) $\overline{\mathcal{F}}$ es compacta en $(\mathcal{P}(X), d_P)$.
- ii) $\mathcal{F}$ es rígida.

A.4. Desintegración de medidas

El teorema de desintegración formaliza la intuición de que una medida debería poder descomponerse como una familia de medidas definidas sobre las fibras de una función medible. Más precisamente, dados espacios Z y X , una medida $\pi \in \mathcal{P}(Z)$ y una función Boreliana $P : Z \rightarrow X$, es posible definir medidas $(\pi_x)_{x \in X}$ sobre las fibras $(P^{-1}(x))_{x \in X}$ de modo que π pueda reconstruirse a partir de las medidas $(\pi_x)_{x \in X}$ y el *pushforward* de π por P . El *pushforward* de π por P se define como la medida $P_{\#}\pi \in \mathcal{P}(X)$ dada por $P_{\#}\pi[E] = \pi[P^{-1}(E)]$ para $E \in \mathcal{B}(X)$.

Definición A.4.1 Un espacio topológico X es un *espacio de Radon* si verifica que toda medida $\mu \in \mathcal{M}(X)$ es rígida. En particular, todo espacio polaco es un espacio de Radon.

Teorema A.4.2 *Dados espacios de Radon Z y X , $P \in \mathcal{B}(Z, X)$ y $\pi \in \mathcal{P}(Z)$, definimos las notaciones $\mu := P_{\#}\pi$ y $E_x := P^{-1}(x) \cap E$ para $x \in X$ y $E \in \mathcal{B}(Z)$. Entonces existen medidas $(\pi_x)_{x \in X}$ en $\mathcal{P}(Z)$, unívocamente determinadas μ -a.e., que desintegran π respecto de μ . Precisamente:*

i) *La aplicación $x \mapsto \pi_x$ es boreliana, en el sentido de que para todo $B \in \mathcal{B}(Z)$ resulta boreliana la aplicación $x \mapsto \pi_x[B]$.*

ii) *Las medidas $(\pi_x)_{x \in X}$ se concentran sobre las fibras $(Z_x)_{x \in X}$: μ -a.e. valen*

$$\pi_x[Z \setminus Z_x] = 0 \quad y \quad \pi_x[E] = \pi_x[E_x] = \pi_x[E \cap Z_x] \quad \text{para todo } E \in \mathcal{B}(Z).$$

iii) *Para toda función boreliana $f : Z \rightarrow [0, +\infty]$ vale*

$$\int_Z f(z) d\pi(z) = \int_X \left(\int_{Z_x} f(z) d\pi_x(z) \right) d\mu(x).$$

Observación A.4.3 El teorema A.4.2 puede extenderse a medidas $\pi \in \mathcal{M}(X)$ tomando la descomposición de Hahn–Jordan $\pi = \pi^+ - \pi^-$ y aplicando luego el teorema a las normalizaciones de las medidas π^+ y π^- . Notar que la identidad del ítem iii) vale para funciones $f \in L(\pi)$ y que para $E \in \mathcal{B}(Z)$ implica las igualdades

$$\pi[E] = \int_X \pi_x[E_x] d\mu(x) = \int_X \pi_x[E] d\mu(x).$$

Desintegración en un espacio producto

Dados espacios de Radon X e Y , considerar $Z = X \times Y$ y $P : Z \rightarrow X$ la proyección canónica sobre X . Las fibras de P toman la forma $Z_x = \{x\} \times Y$ y por lo tanto pueden ser identificadas con Y . Más aun, para cada medible $E \in \mathcal{B}(Z)$ puede pensarse la sección E_x como el elemento de $\mathcal{B}(Y)$ definido por

$$E_x = \{ y \in Y : (x, y) \in E \}. \quad (115)$$

Con esta idea en mente puede tomarse cada medida π_x en el teorema A.4.2 como un elemento de $\mathcal{P}(Y)$ y reconstruir π a partir de las medidas producto $\delta_x \otimes \pi_x$ que sí están concentradas sobre las fibras $\{x\} \times Y$. Así, el teorema A.4.2 toma la siguiente forma.

Corolario A.4.4 *Sean X e Y espacios de Radon y sea $\pi \in \mathcal{P}(X \times Y)$ una medida con primera marginal $\mu \in \mathcal{P}(X)$. Entonces existen medidas $(\pi_x)_{x \in X}$ en $\mathcal{P}(Y)$, unívocamente determinadas μ -a.e., que desintegran π respecto de μ . Precisamente:*

- i) La aplicación $x \mapsto \pi_x$ es boreliana, en el sentido de que para todo $B \in \mathcal{B}(Y)$ resulta boreliana la aplicación $x \mapsto \pi_x[B]$.
- ii) Para toda función boreliana $f : X \times Y \rightarrow [0, +\infty]$ vale

$$\int_{X \times Y} f(x, y) d\pi(x, y) = \int_X \left(\int_Y f(x, y) d\pi_x(y) \right) d\mu(x).$$

En particular, para todo $E \in \mathcal{B}(X \times Y)$ vale

$$\pi[E] = \int_X (\delta_x \otimes \pi_x)[E] d\mu(x) = \int_X \pi_x[E_x] d\mu(x),$$

donde se definen las secciones E_x como en (115).

Página en blanco.

Bibliografía

- [1] N. Ahmad, H. K. Kim, y R. J. McCann. Optimal transportation, topology and uniqueness. *Bull. Math. Sci.*, 1(1):13–32, 2011.
- [2] R. K. Ahuja, T. L. Magnanti, y J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice–Hall, 1993.
- [3] L. Ambrosio. Lecture Notes on Optimal Transport Problems. En *Mathematical Aspects of Evolving Interfaces: Lectures given at the C.I.M.-C.I.M.E. joint Euro-Summer School held in Madeira, Funchal, Portugal, July 3-9, 2000*, páginas 1–52. Springer Berlin Heidelberg, 2003.
- [4] L. Ambrosio, E. Brué, y D. Semola. *Lectures on Optimal Transport*. Springer, 2021.
- [5] L. Ambrosio y A. Figalli. Lecture notes on variational models for incompressible Euler equations. En *Optimal transport. Theory and applications. Proceedings of the summer school ‘Optimal transportation: theory and applications’, Fourier Institute, Grenoble, France*, páginas 58–71. Cambridge: Cambridge University Press, 2014.
- [6] L. Ambrosio, N. Gigli, y G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Basel: Birkhäuser, 2005.
- [7] L. Ambrosio, B. Kirchheim, y A. Pratelli. Existence of optimal transport maps for crystalline norms. *Duke Math. J.*, 125(2):207–241, 2004.
- [8] L. Ambrosio y A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation. En *Optimal transportation and applications. Lectures given at the C. I. M. E. summer school, Martina Franca, Italy, September 2–8, 2001*, páginas 123–160. Berlin: Springer, 2003.
- [9] S. Angenent, S. Haker, y A. Tannenbaum. Minimizing flows for the Monge-Kantorovich problem. *SIAM J. Math. Anal.*, 35(1):61–97, 2003.
- [10] M. Arjovsky, S. Chintala, y L. Bottou. Wasserstein generative adversarial networks. En *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, Australia, 7-9 August, 2017*, 2017.

- [11] F. Aurenhammer, F. Hoffmann, y B. Aronov. Minkowski-type theorems and least-squares clustering. *Algorithmica*, 20(1):61–76, 1998.
- [12] S. Bartels y S. Hertzog. *Error Bounds for Discretized Optimal Transport and Its Reliable Efficient Numerical Solution*, páginas 1–20. Springer International Publishing, 2022.
- [13] M. Beiglböck, P. Henry-Labordère, y F. Penkner. Model-independent bounds for option prices – a mass transport approach. *Finance Stoch.*, 17(3):477–501, 2013.
- [14] J.-D. Benamou y Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.
- [15] J.-D. Benamou y V. Duval. Minimal convex extensions and finite difference discretisation of the quadratic Monge-Kantorovich problem. *Eur. J. Appl. Math.*, 30(6):1041–1078, 2019.
- [16] J.-D. Benamou, B. D. Froese, y A. M. Oberman. Numerical solution of the optimal transportation problem using the Monge-Ampère equation. *J. Comput. Phys.*, 260:107–126, 2014.
- [17] P. Bernard y B. Buffoni. Optimal mass transportation and Mather theory. *J. Eur. Math. Soc. (JEMS)*, 9(1):85–121, 2007.
- [18] J. Bertrand. Existence and uniqueness of optimal maps on Alexandrov spaces. *Adv. Math.*, 219(3):838–851, 2008.
- [19] D. P. Bertsekas. A distributed algorithm for the assignment problem. *Laboratory for Information and Decision Sciences Working Paper, MIT*, 1979.
- [20] D. P. Bertsekas y D. A. Castanon. The auction algorithm for the transportation problem. *Ann. Oper. Res.*, 20:67–96, 1989.
- [21] D. Bertsimas y J. Tsitsiklis. *Introduction to linear optimization*. Athena Scientific Belmont, MA, 1997.
- [22] P. Billingsley. *Convergence of Probability Measures*. Wiley, 2nd edition, 1999.
- [23] J. Borwein y A. Lewis. *Convex analysis and nonlinear optimization: theory and examples*. Springer Science & Business Media, 2010.
- [24] Y. Brenier. The least action principle and the related concept of generalized flows for incompressible perfect fluids. *J. Am. Math. Soc.*, 2(2):225–255, 1989.
- [25] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

-
- [26] R. Burkard, M. Dell’Amico, y S. Martello. *Assignment Problems*. SIAM, 2012.
- [27] C. A. Cabrelli y U. M. Molter. A linear time algorithm for a matching problem on the circle. *Inf. Process. Lett.*, 66(3):161–164, 1998.
- [28] L. A. Caffarelli. Interior $W^{2,p}$ estimates for solutions of the Monge-Ampère equation. *Ann. Math. (2)*, 131(1):135–150, 1990.
- [29] L. A. Caffarelli. A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity. *Ann. of Math. (2)*, 131(1):129–134, 1990.
- [30] L. A. Caffarelli. Some regularity properties of solutions of Monge Ampère equation. *Commun. Pure Appl. Math.*, 44(8-9):965–969, 1991.
- [31] L. A. Caffarelli. Boundary regularity of maps with convex potentials. *Comm. Pure Appl. Math.*, 45(9):1141–1151, 1992.
- [32] L. A. Caffarelli. The regularity of mappings with a convex potential. *J. Amer. Math. Soc.*, 5(1):99–104, 1992.
- [33] L. A. Caffarelli. Allocation maps with general cost functions. En *Partial differential equations and applications. Collected papers in honor of Carlo Pucci on the occasion of his 70th birthday*, páginas 29–35. New York, NY: Marcel Dekker, 1996.
- [34] L. A. Caffarelli. Boundary regularity of maps with convex potentials. II. *Ann. Math. (2)*, 144(3):453–496, 1996.
- [35] L. A. Caffarelli, M. Feldman, y R. J. McCann. Constructing optimal maps for Monge’s transport problem as a limit of strictly convex costs. *J. Am. Math. Soc.*, 15(1):1–26, 2002.
- [36] L. Caicedo Torres, L. Manella Pereira, y M. Hadi Amini. A Survey on Optimal Transport for Machine Learning: Theory and Applications. *arXiv:2106.01963*, June 2021.
- [37] L. Caravenna. A proof of Sudakov theorem with strictly convex norms. *Math. Z.*, 268(1-2):371–407, 2011.
- [38] G. Carlier. Duality and existence for a class of mass transportation problems and economic applications. En *Advances in Mathematical Economics. Vol. 5*, páginas 1–21. Tokyo: Springer, 2003.
- [39] G. Carlier. Optimal Transportation and Economic Applications. <https://www.ceremade.dauphine.fr/~carlier/IMA-transport-Lecture-Notes.pdf>, 2010.
- [40] G. Carlier, C. Jimenez, y F. Santambrogio. Optimal transportation with traffic congestion and Wardrop equilibria. *SIAM Journal on Control and Optimization*, 47(3):1330–1350, 2008.

- [41] G. Carlier y F. Santambrogio. A variational model for urban planning with traffic congestion. *ESAIM, Control Optim. Calc. Var.*, 11:595–613, 2005.
- [42] F. Cavalletti. Monge problem in metric measure spaces with Riemannian curvature-dimension condition. *Nonlinear Anal., Theory Methods Appl., Ser. A, Theory Methods*, 99:136–151, 2014.
- [43] T. Champion y L. De Pascale. The Monge problem for strictly convex norms in $\mathbb{R}^d$. *J. Eur. Math. Soc. (JEMS)*, 12(6):1355–1369, 2010.
- [44] T. Champion y L. De Pascale. The Monge problem in $\mathbb{R}^d$. *Duke Math. J.*, 157(3):551–572, 2011.
- [45] R. Chartrand, B. Wohlberg, K. Vixie, y E. Bollt. A gradient descent solution to the Monge-Kantorovich problem. *Applied Mathematical Sciences*, 3(22):1071–1080, 2009.
- [46] N. Chehebar. Transporte Óptimo y Baricentro con distancia de Fermat. Master’s thesis, FCEyN - UBA, 2021. Disponible en https://cms.dm.uba.ar/academico/carreras/licenciatura/tesis/2021/Nicolas_Chehebar.pdf.
- [47] P.-A. Chiappori, R. J. McCann, y L. P. Nesheim. Hedonic price equilibria, stable matching, and optimal transport: Equivalence, topology, and uniqueness. *Econ. Theory*, 42(2):317–354, 2010.
- [48] M. Cullen y H. Maroofi. The fully compressible semi-geostrophic system from meteorology. *Arch. Ration. Mech. Anal.*, 167(4):309–336, 2003.
- [49] M. J. P. Cullen y R. J. Douglas. Applications of the Monge-Ampère equation and Monge transport problem to meteorology and oceanography. En *Monge Ampère equation: applications to geometry and optimization. Proceedings of the NSF-CBMS conference, Deerfield Beach, FL, USA, July 9–13, 1997*, páginas 33–53. Providence, RI: American Mathematical Society, 1999.
- [50] M. Cuturi. Sinkhorn distances: lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.
- [51] G. B. Dantzig. Application of the simplex method to a transportation problem. *Activity Analysis of Production and Allocation*, Cap. XXIII, 359-373, 1951.
- [52] G. B. Dantzig. Linear programming and extensions. A RAND Corporation Research Study. Princeton University Press, 1963.
- [53] C. Daskalakis, A. Deckelbaum, y C. Tzamos. Strong duality for a multiple-good monopolist. *Econometrica*, 85(3):735–767, 2017.
- [54] G. De Philippis y A. Figalli. The Monge-Ampère equation and its link to optimal transportation. *Bull. Am. Math. Soc., New Ser.*, 51(4):527–580, 2014.

-
- [55] G. De Philippis y A. Figalli. Partial regularity for optimal transport maps. *Publ. Math. Inst. Hautes Études Sci.*, 121:81–112, 2015.
- [56] N. Deb, P. Ghosal, y B. Sen. Rates of estimation of optimal transport maps using plug-in estimators via barycentric projections. *Advances in Neural Information Processing Systems*, 34:29736–29753, 2021.
- [57] N. Delanoue, M. Lhommeau, y P. Lucidarme. Numerical enclosures of the optimal cost of the Kantorovitch’s mass transportation problem. *Computational Optimization and Applications*, 63(3):855–873, Apr 2016.
- [58] J. Delon, J. Salomon, y A. Sobolevski. Fast transport optimization for monge costs on the circle. *SIAM Journal on Applied Mathematics*, 70(7):2239–2258, 2010.
- [59] L. C. Evans y W. Gangbo. *Differential equations methods for the Monge-Kantorovich mass transfer problem*, volume 653 of *Mem. Am. Math. Soc.* Providence, RI: American Mathematical Society (AMS), 1999.
- [60] A. Fathi y A. Figalli. Optimal transportation on non-compact manifolds. *Isr. J. Math.*, 175:1–59, 2010.
- [61] H. Federer. *Geometric Measure Theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.
- [62] M. Feldman y R. J. McCann. Monge’s transport problem on a Riemannian manifold. *Trans. Am. Math. Soc.*, 354(4):1667–1697, 2002.
- [63] A. Figalli. Existence, uniqueness, and regularity of optimal transport maps. *SIAM J. Math. Anal.*, 39(1):126–137, 2007.
- [64] A. Figalli. Regularity properties of optimal maps between nonconvex domains in the plane. *Comm. Partial Differential Equations*, 35(3):465–479, 2010.
- [65] A. Figalli y F. Glaudo. *An Invitation to Optimal Transport, Wasserstein Distances, and Gradient Flows*. European Mathematical Society (EMS), 2021.
- [66] A. Figalli y Y. Kim. Partial regularity of Brenier solutions of the Monge-Ampère equation. *Discrete and Continuous Dynamical Systems*, 28(2):559–565, 2010.
- [67] P. T. Fletcher, S. Venkatasubramanian, y S. Joshi. Robust statistics on Riemannian manifolds via the geometric median. En *2008 IEEE Conference on Computer Vision and Pattern Recognition*, páginas 1–8. IEEE, 2008.
- [68] L. R. Ford Jr. y D. R. Fulkerson. *Flows in networks*. Princeton University Press, 1962.

- [69] B. D. Froese. A numerical method for the elliptic Monge-Ampère equation with transport boundary conditions. *SIAM J. Sci. Comput.*, 34(3):a1432–a1459, 2012.
- [70] B. Froese Hamfeldt. Convergence framework for the second boundary value problem for the Monge-Ampère equation. *SIAM J. Numer. Anal.*, 57(2):945–971, 2019.
- [71] D. R. Fulkerson. *Hitchcock transportation problem*. Rand Corporation, 1956.
- [72] A. Galichon. A survey of some recent applications of optimal transport methods to econometrics. *Econom. J.*, 20(2):c1–c11, 2017.
- [73] A. Galichon. *Optimal transport methods in economics*. Princeton University Press, 2018.
- [74] A. Gallouët, Q. Mérigot, y B. Thibert. Strong c -concavity and stability in optimal transport. *arXiv:2207.11042*, July 2022.
- [75] W. Gangbo. The Monge mass transfer problem and its applications. En L. Caffarelli y M. Milman, editores, *Monge Ampère equation: applications to geometry and optimization. Proceedings of the NSF-CBMS conference, Deerfield Beach, FL, USA, July 9–13, 1997*, número 226 en Contemporary Mathematics, páginas 79–104. American Mathematical Society, Providence RI, 1999.
- [76] W. Gangbo y R. J. McCann. Optimal maps in Monge’s mass transport problem. *C. R. Acad. Sci., Paris, Sér. I*, 321(12):1653–1658, 1995.
- [77] W. Gangbo y R. J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.
- [78] N. Gigli. Optimal maps in non branching spaces with Ricci curvature bounded from below. *Geom. Funct. Anal.*, 22(4):990–999, 2012.
- [79] S. Haker, L. Zhu, A. Tannenbaum, y S. Angenent. Optimal mass transport for registration and warping. *Int. J. Comput. Vis.*, 60(3):225–240, 2004.
- [80] E. Hansen y G. Walster. *Global Optimization Using Interval Analysis: Revised And Expanded*. CRC Press, 2003.
- [81] F. L. Hitchcock. The distribution of a product from several sources to numerous localities. *J. Math. Phys., Mass. Inst. Techn.*, 20:224–230, 1941.
- [82] X. Jin y J. Han. K-Means Clustering. En *Encyclopedia of Machine Learning*, páginas 563–564. Springer US, 2010.
- [83] R. M. Karp y S.-Y. R. Li. Two special cases of the assignment problem. *Discrete Mathematics*, 13(2):129–142, 1975.

-
- [84] J. Kitagawa, Q. Mérigot, y B. Thibert. Convergence of a Newton algorithm for semi-discrete optimal transport. *J. Eur. Math. Soc. (JEMS)*, 21(9):2603–2651, 2019.
- [85] M. Knott y C. S. Smith. On the optimal mapping of distributions. *Journal of Optimization Theory and Applications*, 43:39–49, 1984.
- [86] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955.
- [87] V. Levin. Abstract cyclical monotonicity and Monge solutions for the general Monge-Kantorovich problem. *Set-Valued Anal.*, 7(1):7–32, 1999.
- [88] B. Lévy. A numerical algorithm for L_2 semi-discrete optimal transport in 3d. *ESAIM, Math. Model. Numer. Anal.*, 49(6):1693–1715, 2015.
- [89] W. Li y R. H. Nochetto. Quantitative stability and error estimates for optimal transport plans. *IMA J. Numer. Anal.*, 41(3):1941–1965, 2021.
- [90] M. Lindsey y Y. A. Rubinstein. Optimal transport via a Monge-Ampère optimization problem. *SIAM J. Math. Anal.*, 49(4):3073–3124, 2017.
- [91] S. P. Lloyd. Least squares quantization in PCM. *IEEE Trans. Inf. Theory*, 28:129–137, 1982.
- [92] D. G. Luenberger y Y. Ye. *Linear and Nonlinear Programming*. Springer, 2008.
- [93] X.-N. Ma, N. S. Trudinger, y X.-J. Wang. Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.*, 177(2):151–183, 2005.
- [94] J. MacQueen. Some methods for classification and analysis of multivariate observations. Proc. 5th Berkeley Symp. Math. Stat. Probab., Univ. Calif. 1965/66, 1, 281–297 (1967).
- [95] R. J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995.
- [96] R. J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11(3):589–608, 2001.
- [97] R. J. McCann, B. Pass, y M. Warren. Rectifiability of optimal transportation plans. *Can. J. Math.*, 64(4):924–933, 2012.
- [98] G. Montavon, K.-R. Müller, y M. Cuturi. Wasserstein Training of Restricted Boltzmann Machines. En *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS’16, página 3718–3726, Red Hook, NY, USA, 2016. Curran Associates Inc.

- [99] E. F. Montesuma, F. N. Mboula, y A. Souloumiac. Recent advances in optimal transport for machine learning. *arXiv:2306.16156*, June 2023.
- [100] R. E. Moore. *Methods and Applications of Interval Analysis*. Society for Industrial and Applied Mathematics, 1979.
- [101] Q. Mérigot. A multiscale approach to optimal transport. *Computer Graphics Forum*, 30(5):1583–1592, 2011.
- [102] A. M. Oberman y Y. Ruan. Solution of optimal transportation problems using a multigrid linear programming approach. *J. Comput. Math.*, 38(6):937–956, 2020.
- [103] N. Papadakis, G. Peyré, y E. Oudet. Optimal transport with proximal splitting. *SIAM J. Imaging Sci.*, 7(1):212–238, 2014.
- [104] K. R. Parthasarathy. *Probability Measures on Metric Spaces*. Academic Press, 1967.
- [105] G. Peyré y M. Cuturi. Computational optimal transport. *Found. Trends Mach. Learn.*, 11(5-6):355–607, 2019.
- [106] D. Piazzoli, F. Santambrogio, y P. Pegon. Full characterization of optimal transport plans for concave costs. *Discrete Contin. Dyn. Syst.*, 35(12):6113–6132, 2015.
- [107] F. Plastria. The Weiszfeld algorithm: proof, amendments, and extensions. En *Foundations of location analysis*, páginas 357–389. Springer, 2011.
- [108] A. Pratelli. On the equality between Monge’s infimum and Kantorovich’s minimum in optimal mass transportation. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 43(1):1–13, 2007.
- [109] C. R. Prins, R. Beltman, J. H. M. ten Thije Boonkkamp, W. L. IJzerman, y T. W. Tukker. A least-squares method for optimal transport using the monge–ampère equation. *SIAM Journal on Scientific Computing*, 37(6):B937–B961, 2015.
- [110] J. Rabin, J. Delon, y Y. Gousseau. Regularization of transportation maps for color and contrast transfer. En *2010 IEEE International Conference on Image Processing*, páginas 1933–1936, 2010.
- [111] T. Rajala y T. Schultz. Optimal transport maps on Alexandrov spaces revisited. *Manuscr. Math.*, 169(1-2):1–18, 2022.
- [112] H. L. Royden. *Real Analysis*. Macmillan Publishing Company, 3rd edition, 1988.
- [113] Y. Rubner, C. Tomasi, y L. J. Guibas. The earth mover’s distance as a metric for image retrieval. *Int. J. Comput. Vis.*, 40(2):99–121, 2000.
- [114] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 3rd edition, 1987.

- [115] F. Santambrogio. Models and applications of optimal transport in economics, traffic, and urban planning. En *Optimal transport. Theory and applications. Proceedings of the summer school ‘Optimal transportation: theory and applications’, Fourier Institute, Grenoble, France*, páginas 22–40. Cambridge: Cambridge University Press, 2014.
- [116] F. Santambrogio. *Optimal Transport for Applied Mathematicians*. Progress in Nonlinear Differential Equations and their applications. Springer, 2015.
- [117] B. Schmitzer. A sparse multiscale algorithm for dense optimal transport. *Journal of Mathematical Imaging and Vision*, 56(2):238–259, Oct 2016.
- [118] B. Schmitzer y C. Schnörr. *A Hierarchical Approach to Optimal Transport*, páginas 452–464. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [119] A. Schrijver. On the history of the transportation and maximum flow problems. *Math. Program.*, 91(3 (B)):437–445, 2002.
- [120] V. N. Sudakov. Geometric problems in the theory of infinite-dimensional probability distributions. *Proc. Steklov Inst. Math.* 141, 178 p. (1979); translation from *Tr. Mat. Inst. Steklov* 141, 191 p. (1976)., 1976.
- [121] E. G. Tabak, G. Trigila, y W. Zhao. Conditional density estimation and simulation through optimal transport. *Mach. Learn.*, 109(4):665–688, 2020.
- [122] G. Trigila y E. G. Tabak. Data-driven optimal transport. *Commun. Pure Appl. Math.*, 69(4):613–648, 2016.
- [123] N. S. Trudinger y X.-J. Wang. On the Monge mass transfer problem. *Calc. Var. Partial Differ. Equ.*, 13(1):19–31, 2001.
- [124] C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.
- [125] C. Villani. *Optimal Transport: Old and New*. Número 338 en Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2009.
- [126] E. Weiszfeld. Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tôhoku Math. J.*, 43:355–386, 1937.
- [127] M. Werman, S. Peleg, R. Meller, y T. Kong. Bipartite graph matching for points on a line or a circle. *Journal of Algorithms*, 7(2):277–284, 1986.
- [128] Q. Xia. Optimal paths related to transport problems. *Commun. Contemp. Math.*, 5(2):251–279, 2003.
- [129] H. Yang y E. G. Tabak. Clustering, factor discovery and optimal transport. *Inf. Inference*, 10(4):1353–1387, 2021.

Bibliografía

- [130] L. Yang. Riemannian median and its estimation. *LMS Journal of Computation and Mathematics*, 13:461–479, 2010.