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

Tesis de Licenciatura

Transporte Óptimo y Aplicaciones

Maximiliano Omar Frungillo

Director: Pablo De Nápoli

14 de diciembre de 2016

Contenidos

Introducción

En 1781 Gaspard Monge formuló [22] un problema que en su versión más simple puede entenderse como sigue: dada una montaña de tierra y un conjunto de pozos que deben rellenarse con esa misma tierra, ¿cuál es la forma óptima de transportarla para rellenar los pozos? Podemos imponer la restricción de que el volumen de tierra coincida con el volumen total de los pozos y entonces suponer que ambos volúmenes son unitarios. Así, el problema de Monge se transforma en el de redistribuir masa para transportar una medida de probabilidad en otra, minimizando algún costo relacionado con los movimientos de masa que se hagan para conseguirlo.

Siguiendo la práctica habitual en la bibliografía, en este trabajo se estudia el problema de Monge en el contexto de espacios métricos polacos, es decir, espacios métricos completos y separables. Dados dos espacios polacos X e Y y medidas de probabilidad μ y ν en ellos, el problema de Monge consiste en minimizar el funcional

$$I(T) = \int_X c(x, T(x))\, d\mu(x) \ ,$$

donde $c : X \times Y \to \mathbb{R}$ es una función de costo dada y la minimización se hace sobre el conjunto de las $T : X \to Y$ que transforman μ en ν. El conjunto de las T que verifican esta propiedad no tiene una buena estructura y por lo tanto resultará útil atacar el problema de Monge estudiando primero una relajación lineal de éste debida a Kantorovich.

Esta tesis está dividida en dos partes. La primera, destinada a establecer algunos resultados fundamentales de la teoría de transporte óptimo. La segunda, a estudiar unas pocas pero importantes aplicaciones de esta teoría a distintos problemas. Al final se incluye un apéndice en el que pueden leerse los resultados previos necesarios para desarrollar la teoría y estudiar los problemas propuestos en las aplicaciones.

En el primer capítulo formularemos precisamente el problema de Monge y la relajación de Kantorovich. Probaremos también la existencia de minimizantes para el problema de Kantorovich y estudiaremos algunos ejemplos. Al final, estudiaremos el problema en el caso de distribuciones dadas por finitas masas puntuales.

En el segundo y el tercer capítulo estudiaremos la teoría de dualidad para el problema de Kantorovich. Revisaremos primero los teoremas de dualidad en el caso finito y daremos un ejemplo que mostrará la necesidad de hipótesis adicionales más sutiles para obtener teoremas de dualidad en el caso general. En el tercer capítulo se aplicará la teoría de dualidad a fin de obtener criterios de optimalidad para soluciones factibles del problema de Kantorovich. Para esto será necesario desarrollar la teoría de c-convexidad, útil para generalizar a espacios sin estructura lineal algunas nociones de la teoría de convexidad.

En el cuarto capítulo se apreciará la utilidad de estudiar el problema de Kantorovich como paso previo a atacar el problema de Monge. Obtendremos resultados de existencia de minimizantes para el problema de Monge explotando la existencia de maximizantes para el problema dual de Kantorovich. Los resultados que probaremos en este capítulo se deben a Brenier [11], a McCann y Gangbo [20] y a Sudakov [35].

Los tres capítulos restantes conforman la segunda parte de la tesis. En el primero de ellos estudiaremos la aplicación de la teoría de transporte óptimo a la demostración de desigualdades. Probaremos primero la desigualdad isoperimétrica y luego la desigualdad de Sobolev–Gagliardo–Nirenberg. En el capítulo siguiente estudiaremos dos modelos de transporte. Empezaremos por uno simple debido a Beckmann [5] y seguiremos con otro modelo más sofisticado debido a Carlier, Jimenez y Santambrogio [14] que incorpora efectos de congestión. En el último capítulo estudiaremos las distancias de Wasserstein y una variante conocida como earth mover's distance útil en las aplicaciones.

Si bien la bibliografía consultada es extensa, las principales referencias para este trabajo fueron las exposiciones de la teoría de transporte óptimo que pueden verse en [37], [33], [2] y [13].

Parte I

Teoría de Transporte Óptimo

Capítulo 1

El problema de Monge–Kantorovich

El objetivo de este capítulo es enunciar precisamente los problemas de Monge y de Kantorovich en el contexto de espacios métricos polacos. Luego de mostrar la necesidad de hipótesis fuertes para probar la existencia de minimizantes para el problema de Monge, probamos que para el problema de Kantorovich sí existen soluciones óptimas con hipótesis mínimas. Finalizamos el capítulo con algunos resultados de existencia para el problema de Monge y estudiando el problema de Kantorovich en el caso finito.

1.1. Formulación del problema de Monge–Kantorovich

Sean X e Y dos espacios métricos, $T : X \to Y$ una aplicación boreliana y $\mu \in \mathcal{P}(X)$ una medida boreliana de probabilidad. Se define el **push forward de $\boldsymbol{\mu}$ por $\boldsymbol{T}$** 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. Dada una medida $\nu \in \mathcal{P}(Y)$, diremos que **$\boldsymbol{T}$ transporta $\boldsymbol{\mu}$ en $\boldsymbol{\nu}$** si $T_\#\mu = \nu$.

Lema 1.1.1. *Dada una aplicación boreliana entre espacios métricos $T : X \to Y$ 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 $\varphi \in L^1(\nu)$.

Demostración. Si $T_{\#}\mu = \nu$, entonces para E boreliano de Y se tiene

$$\int_Y \chi_E(y)\, d\nu(y) = \nu(E) = \mu(T^{-1}(E)) = \int_X \chi_{T^{-1}(E)}(x)\, d\mu(x) = \int_X \chi_E \circ T(x)\, d\mu(x)$$

donde la última igualdad es consecuencia de que $\chi_{T^{-1}(E)}(x) = \chi_E(T(x))$ para todo x en X. Esto prueba la igualdad para características y, por linealidad, para funciones simples. Por definición de integral resulta la igualdad para φ no negativa y por linealidad para φ en $L^1(\nu)$. La otra implicación es inmediata dado que $\chi_E \in L^1(\nu)$ para todo E boreliano de Y. ■

Nos interesa transportar la masa de X — distribuida según μ — hasta Y, de modo que se distribuya según ν. Además nos interesa hacerlo de forma óptima, minimizando un costo. El costo de transportar masa estará dado por una función $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ medible. Así, transportar masa desde la posición $x \in X$ hasta la posición $y \in Y$ tendrá costo $c(x, y)$. Notar que permitiremos costos negativos pero no infinitos negativos.

Problema del transporte óptimo de Monge

Dados dos espacios métricos X e Y junto con dos medidas $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y una función medible $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$, minimizar la aplicación

$$T \mapsto \int_X c(x, T(x)) d\mu(x)$$

entre todas las T que transportan μ en ν.

En muchos casos el problema de Monge es **factible**, en el sentido de que existe al menos una $T : X \to Y$ que realiza el transporte de μ a ν. En el caso general la factibilidad no está garantizada. El problema aparece cuando μ tiene átomos, dado que éstos no pueden ser deshechos por una aplicación $T : X \to Y$. Más precisamente, si $x_0 \in X$ es un átomo de μ entonces $T(x_0)$ es un átomo de $T_{\#}\mu$:

$$T_{\#}\mu(\{T(x_0)\}) = \mu(T^{-1}(\{T(x_0)\})) = \mu(T^{-1}(T(x_0))) \geq \mu(\{x_0\}) > 0.$$

Esta observación prueba el siguiente lema.

Lema 1.1.2. *El problema de Monge no es factible si* μ *tiene al menos un átomo y* ν *es no atómica.*

La imposibilidad de dividir la masa de un punto puede impedir la factibilidad también en el caso de que ν tenga átomos.

Ejemplo 1.1.3. Sean los conjuntos $X = \{x_0\}$, $Y = \{y_0, y_1\}$ y consideremos las medidas $\mu = \delta_{x_0}$ y $\nu = \frac{1}{2}\delta_{y_0} + \frac{1}{2}\delta_{y_1}$. Si $T : X \to Y$ es cualquier función, entonces $T_{\#}\mu(\{T(x_0)\}) = 1$ y por lo tanto $T_{\#}\mu = \delta_{T(x_0)} \neq \frac{1}{2}\delta_{y_0} + \frac{1}{2}\delta_{y_1} = \nu$. ▲

No existen estas incompatibilidades esenciales cuando ν tiene átomos y μ es no atómica, pues nada impide considerar una T que mande un subconjunto de X de medida positiva a un punto de Y. En este sentido, un resultado de Pratelli prueba que cuando X e Y son polacos y μ es no atómica siempre existe una $T : X \to Y$ que realiza el transporte [26].

Otra dificultad del problema de Monge es que aun cuando μ y ν son tales que éste resulta factible, la igualdad $T_{\#}\mu = \nu$ no se mantiene por límites débiles y por lo tanto el conjunto de las T candidatas a realizar el transporte óptimo no es cerrado en sentido débil.

Ejemplo 1.1.4. Sea $f : \mathbb{R} \to \mathbb{R}$ la extensión 1-periódica de la función dada por

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

y sea $(f_n)_n$ la sucesión de funciones de $\mathbb{R}$ a $\mathbb{R}$ definida por $f_n(x) = f(nx)$ para n en $\mathbb{N}$ y x en $\mathbb{R}$. Sea μ la restricción de la medida de Lebesgue al $[0, 1]$ y $\nu = (\delta_{-1} + \delta_1)/2$. Es inmediato que

$$(f_n)_{\#}\mu = \nu \tag{1.1}$$

para todo n en $\mathbb{N}$. Por el teorema de Riemann–Lebesgue $(f_n)_n$ converge débilmente a $f \equiv 0$. Luego la igualdad (1.1) no se preserva por límites débiles, pues $f_{\#}\mu = \delta_0 \neq \nu$. ▲

A las dificultades mencionadas anteriormente se agrega también la no linealidad en T de la asignación $T \mapsto \int_X c(x, T(x))d\mu(x)$. Para entender el tipo de no linealidad en cuestión consideremos el caso particular $X = Y = \mathbb{R}^n$ y supongamos que μ y ν están dadas por densidades f y g respectivamente. Tenemos

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

para todo boreliano $E \subseteq \mathbb{R}^n$. Adicionalmente, supongamos que el transporte de μ en ν pueda ser realizado por un difeomorfismo T de clase C^1. Dado un boreliano $E \subseteq \mathbb{R}^n$, tenemos

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

por el teorema de cambio de variables. Como T transporta μ en ν tenemos también

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

Por lo tanto T está sujeta a la condición no lineal

$$f(x) = g(T(x))|\det DT(x)| \;, \tag{1.2}$$

conocida como **ecuación de Monge–Ampère**.

Otra característica del problema de Monge es que no permite la posibilidad de dividir masa para transportar μ en ν: para cada $x \in X$ toda la masa en x va a parar a $T(x)$. Con respecto a este punto, una posibilidad sería relajar el problema para permitir que, al menos en principio, un transporte de μ en ν pudiera mandar la masa de un punto de X a varios puntos de Y y esperar que así aumenten las chances de que el transporte de μ en ν sea un problema factible. Esta relajación del problema de Monge se conoce como el problema de Kantorovich y quizás su característica más notable sea que elimina de una vez todas las dificultades del problema de Monge mencionadas anteriormente.

Consideremos dos medidas de probabilidad μ y ν como antes y modelemos el problema de transportar μ en ν permitiendo didivir la masa. Necesitamos decir cómo se repartirá la masa $\mu(x)$ inicialmente en x y decir también que todas las contribuciones a y sumarán exactamente $\nu(y)$. Para esto consideramos las medidas de probabilidad $\gamma \in \mathcal{P}(X \times Y)$ que satisfacen

$$\begin{aligned} d\mu(x) &= \int_Y d\gamma(x,y) \quad , \quad \text{para cada } x \in X \\ d\nu(y) &= \int_X d\gamma(x,y) \quad , \quad \text{para cada } y \in Y. \end{aligned} \tag{1.3}$$

O bien, en términos de medidas de conjuntos,

$$\begin{aligned} \mu(E) &= \int_{E\times Y} d\gamma(x,y) \quad , \quad \text{para cada boreliano } E \subseteq X \\ \nu(F) &= \int_{X\times F} d\gamma(x,y) \quad , \quad \text{para cada boreliano } F \subseteq Y. \end{aligned} \tag{1.4}$$

Una condición equivalente a (1.4) es

$$\begin{aligned}
\int_X \chi_E(x)\,d\mu(x) &= \int_{X\times Y} \chi_E(x)\,d\gamma(x,y) \quad , \quad \text{para cada boreliano } E \subseteq X \\
\int_Y \chi_F(y)\,d\nu(y) &= \int_{X\times Y} \chi_F(y)\,d\gamma(x,y) \quad , \quad \text{para cada boreliano } F \subseteq X\ .
\end{aligned} \tag{1.5}$$

Si π^1 y π^2 son las proyecciones en la primera y la segunda coordenada, cualquiera de estas condiciones es equivalente a pedir $\pi^1_{\#}\gamma = \mu$ y $\pi^2_{\#}\gamma = \nu$.

Definición 1.1.5. Cuando γ cumple las condiciones equivalentes (1.3), (1.4) y (1.5), decimos que μ y ν son las **medidas marginales** de γ. Notamos $\mathcal{A}(\mu,\nu)$ al conjunto de las medidas de probabilidad $\gamma \in \mathcal{P}(X\times Y)$ que tienen marginales μ y ν en ese orden.

Dado que $\gamma \in \mathcal{P}(X\times Y)$ si y sólo si $\pi^1_{\#}\gamma = \mu$ y $\pi^2_{\#}\gamma = \nu$, el siguiente resultado es una consecuencia inmediata del lema 1.1.1.

Lema 1.1.6. *Sean X e Y espacios métricos y μ, ν medidas borelianas de probabilidad en X e Y respectivamente. Una medida $\gamma \in \mathcal{P}(X\times Y)$ está en $\mathcal{A}(\mu,\nu)$ si y sólo si verifica*

$$\begin{aligned}
\int_X \varphi(x)\,d\mu(x) &= \int_{X\times Y} \varphi(x)\,d\gamma(x,y) \\
\int_Y \psi(y)\,d\nu(y) &= \int_{X\times Y} \psi(y)\,d\gamma(x,y)\ .
\end{aligned}$$

para todo par de funciones $\varphi : X \to \overline{\mathbb{R}}$, $\psi : Y \to \overline{\mathbb{R}}$ medibles.

Observación 1.1.7. Notar que dadas μ en $\mathcal{P}(X)$ y ν en $\mathcal{P}(Y)$, el conjunto $\mathcal{A}(\mu,\nu)$ es no vacío. Como X e Y son de medida finita, se puede probar que existe una única medida $\mu\otimes\nu$ en $\mathcal{P}(X\times Y)$ que cumple

$$(\mu\otimes\nu)(B_1\times B_2) = \mu(B_1)\nu(B_2)$$

para todo B_1 boreliano de X y todo B_2 boreliano de Y. La medida $\mu\otimes\nu$ se llama la **medida producto** de μ y ν y es un elemento de $\mathcal{A}(\mu,\nu)$. En el contexto de los problemas de transporte de masa se la interpreta como el plan de transporte que para cada x en X distribuye la masa $\mu(x)$ equitativamente en todo el soporte de ν. ▲

Problema del transporte óptimo de Kantorovich

Dados dos espacios métricos X e Y junto con dos medidas $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y una función medible $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$, minimizar la aplicación

$$\gamma \mapsto \int_{X\times Y} c(x,y)d\gamma(x,y) \tag{1.6}$$

entre todas las γ de $\mathcal{A}(\mu,\nu)$.

Definición 1.1.8. En el contexto de problemas de transporte decimos que cada γ en $\mathcal{A}(\mu,\nu)$ es un **plan de transporte** de μ a ν. Siguiendo la terminología utilizada en programación lineal podemos llamar **solución factible** del problema de Kantorovich a cada plan de transporte, pues cada uno de ellos satisface las restricciones que se le imponen — i.e. ser un elemento de $\mathcal{A}(\mu,\nu)$ — aunque no necesariamente sea una solución óptima del problema. Cada $T : X \to Y$ que cumple $T_{\#}\mu = \nu$ se dice **función de transporte**. En la bibliografía suelen llamarse **transport plans** y **transport maps** respectivamente.

Observación 1.1.9. La observación 1.1.7 muestra que para medidas μ, ν cualesquiera siempre existe al menos un plan de transporte, mientras que 1.1.2 y 1.1.3 muestran que no siempre existen funciones de transporte. Dicho de otro modo, el problema de Kantorovich siempre es factible mientras que el de Monge puede no serlo. ▲

Notar que si $T : X \to Y$ realiza el transporte de μ a ν en el sentido de Monge, entonces ese mismo transporte se puede pensar en el sentido de Kantorovich. Al menos formalmente podemos considerar $d\gamma(x,y) = d\mu(x)d\delta_{T(x)}(y)$ y entonces

$$\begin{aligned}\int_{X\times Y} c(x,y)\,d\gamma(x,y) &= \int_{X\times Y} c(x,y)\,d\mu(x)d\delta_{T(x)}(y) = \int_X \Big(\int_Y c(x,y)\,d\delta_{T(x)}(y)\Big)\,d\mu(x) = \\ &= \int_X \Big(\int_{\{T(x)\}} c(x,y)\,d\delta_{T(x)}(y)\Big)\,d\mu(x) = \int_X c(x,T(x))\,d\mu(x)\,.\end{aligned} \tag{1.7}$$

En términos precisos, lo que hacemos es considerar una medida $\gamma_T \in \mathcal{P}(X \times Y)$ dada por el pushforward de μ por $I \times T : X \to X \times Y$, donde $I \times T(x) = (x, T(x))$ para todo x en X. Así, tenemos

$$\begin{aligned}\gamma_T &\coloneqq (I\times T)_{\#}\mu \\ \gamma_T(E) &= \mu((I\times T)^{-1}(E))\end{aligned} \tag{1.8}$$

para todo E boreliano de $X \times Y$. Es inmediato a partir de (1.8) que vale

$$\pi^1_{\#}\gamma_T = \mu \quad \text{y} \quad \pi^2_{\#}\gamma_T = \nu\,.$$

Cualquier medida de la forma γ_T está concentrada en el gráfico de T. Ese es el contenido del siguiente lema, que permite dar una versión rigurosa de (1.7) y será útil más adelante.

Lema 1.1.10. *Sean X e Y espacios métricos y μ una medida de probabilidad boreliana en X. Si $T : X \to Y$ es boreliana, entonces la medida γ_T dada por $(I \times T)_{\#}\mu$ está concentrada en*

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

y para toda $f : X \times Y \to \overline{\mathbb{R}}$ medible vale

$$\int_X f(x, T(x))\, d\mu(x) = \int_{X\times Y} f(x, y)\, d\gamma_T(x, y)\ . \tag{1.9}$$

Demostración. Notar que $(I \times T)^{-1}(\Gamma(T)) = X$ y entonces si E es un medible de $X \times Y$ resulta

$$\gamma_T(E \cap \Gamma(T)) = \mu\big(\,(I \times T)^{-1}(E \cap \Gamma(T))\,\big) = \mu\big(\,(I \times T)^{-1}(E) \cap (I \times T)^{-1}(\Gamma(T))\,\big) =$$
$$= \mu\big(\,(I \times T)^{-1}(E) \,\cap\, X\,\big) = \mu\big(\,(I \times T)^{-1}(E)\,\big) = \gamma_T(E)\ .$$

Luego γ_T está concentrada en $\Gamma(T)$. Se puede escribir $X \times Y$ como una unión disjunta

$$X \times Y = \Gamma(T) \sqcup Z$$

donde Z es un medible de medida 0. Observar que para (x, y) en $\Gamma(T)$ vale

$$f(x, y) = f(x, T(x))\ . \tag{1.10}$$

Notar que las integrales sobre Z son nulas porque Z tiene medida nula. Teniendo en cuenta el lema 1.1.6 y (1.10), tenemos

$$\int_X f(x, T(x))\, d\mu(x) = \int_{X\times Y} f(x, T(x))\, d\gamma_T(x, y) = \int_{\Gamma(T)} f(x, T(x))\, d\gamma_T(x, y) +$$
$$+ \int_Z f(x, T(x))\, d\gamma_T(x, y) = \int_{\Gamma(T)} f(x, T(x))\, d\gamma_T(x, y) = \int_{\Gamma(T)} f(x, y)\, d\gamma_T(x, y) =$$
$$= \int_{\Gamma(T)} f(x, y)\, d\gamma_T(x, y) + \int_Z f(x, y)\, d\gamma_T(x, y) = \int_{X\times Y} f(x, y)\, d\gamma_T(x, y)$$

y esto prueba el lema. ■

Una consecuencia de 1.1.10 es que funciones de transporte difiriendo en un conjunto de medida positiva inducen planes de transporte distintos.

Lema 1.1.11. *Sean $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y $c : X \times Y \to \mathbb{R}$ una función de costo. Si $T_1 : X \to Y$ y $T_2 : X \to Y$ transportan μ en ν y difieren en un conjunto de medida μ-positiva, entonces $\gamma_{T_1} \neq \gamma_{T_2}$.*

Demostración. Sean $\Gamma_1 = \Gamma(T_1)$, $\Gamma_2 = \Gamma(T_2)$ y $E = \{x \in X : T_1(x) = T_2(x)\}$. Notar que $\Gamma_1 \cap \Gamma_2 = \Gamma_1 \cap (E \times Y) = \Gamma_2 \cap (E \times Y)$ y que por hipótesis $\mu(E) < 1$. Por 1.1.10 vale que γ_{T_1} y γ_{T_2} están concentradas en Γ_1 y Γ_2 respectivamente. Luego

$$\gamma_{T_2}(\Gamma_1) = \gamma_{T_2}(\Gamma_1 \cap \Gamma_2) = \gamma_{T_2}(\Gamma_1 \cap (E \times Y)) \leq \gamma_{T_2}(E \times Y) = \mu(E) < 1 = \gamma_{T_1}(\Gamma_1)$$

y por lo tanto $\gamma_{T_1} \neq \gamma_{T_2}$. ■

1.2. Existencia de minimizantes

Formulamos el problema de Kantorovich como una relajación del problema de Monge y vimos que con esta relajación siempre hay soluciones factibles. Probaremos ahora que bajo condiciones bastante generales siempre existen soluciones factibles óptimas — es decir, minimizantes — para el problema de Kantorovich. Para enunciar y probar los resultados de esta sección y las siguientes se requerirán algunos preliminares sobre medidas en espacios métricos. Estas definiciones, notaciones y resultados pueden consultarse en B.1.

Para demostrar el resultado se necesita un lema previo sobre rigidez de familias en un espacio producto.

Lema 1.2.1. *Sean X e Y espacios métricos y μ, ν medidas borealianas de probabilidad en X e Y respectivamente. Si $\mathcal{F}_1 \subset \mathcal{P}(X)$ y $\mathcal{F}_2 \subset \mathcal{P}(Y)$ son rígidas, entonces también es rígida*

$$\mathcal{A}(\mathcal{F}_1, \mathcal{F}_2) := \left\{\gamma \in \mathcal{P}(X \times Y) : \pi^1_{\#}\gamma \in \mathcal{F}_1 \ \ y \ \ \pi^2_{\#}\gamma \in \mathcal{F}_2\right\}.$$

Demostración. Sean $\varepsilon > 0$ y $K_1 \subset X$, $K_2 \subset Y$ compactos tales que $\mu(X \backslash K_1) < \frac{\varepsilon}{2}$ y $\nu(Y \backslash K_2) < \frac{\varepsilon}{2}$ para toda μ en $\mathcal{F}_1$ y toda ν en $\mathcal{F}_2$. De la inclusión

$$(X \times Y) \backslash (K_1 \times K_2) \subset (X \backslash K_1 \times Y) \cup (X \times Y \backslash K_2)$$

resulta

$$\gamma(X \times Y \backslash K_1 \times K_2) \leq \gamma(X \backslash K_1 \times Y) + \gamma(X \times Y \backslash K_2) = \mu(X \backslash K_1) + \nu(Y \backslash K_2) < 2 \cdot \frac{\varepsilon}{2} = \varepsilon$$

para toda $\gamma \in \mathcal{A}(\mathcal{F}_1, \mathcal{F}_2)$. Luego $\mathcal{A}(\mathcal{F}_1, \mathcal{F}_2)$ es rígida. ■

Proposición 1.2.2. *Sean X e Y espacios métricos polacos y $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ medidas borelianas de probabilidad. Si $c : X \times Y \to \mathbb{R}$ es acotada por abajo y semicontinua inferiormente, entonces existe un minimizante para el problema de Kantorovich.*

Demostración. Basta ver que $\mathcal{A}(\mu, \nu)$ es compacto en $\mathcal{P}(X \times Y)$ y que la aplicación definida por (1.6) es semicontinua inferiormente.

Como X e Y son polacos, entonces μ y ν son rígidas por B.1.14. Por el lema anterior, $\mathcal{A}(\mu, \nu)$ es rígido porque $\mathcal{F}_1 = \{\mu\}$ y $\mathcal{F}_2 = \{\nu\}$ lo son. Luego $\mathcal{A}(\mu, \nu)$ tiene clausura compacta por B.1.17. Sea $(\gamma_n)_n \subset \mathcal{A}(\mu, \nu)$ tal que $\gamma_n \Rightarrow \gamma$. Veamos que $\gamma \in \mathcal{A}(\mu, \nu)$. Si E es un boreliano de X tenemos

$$\begin{aligned} \pi^1_{\#}\gamma(E) = \gamma(E \times Y) = \int_{X \times Y} \chi_{E \times Y}\, d\gamma = \lim_n \int_{X \times Y} \chi_{E \times Y}\, d\gamma_n = \\ = \lim_n \gamma_n(E \times Y) = \lim_n \mu(E) = \mu(E)\ . \end{aligned}$$

Esto dice que $\pi^1_{\#}\gamma = \mu$ y del mismo modo se ve que $\pi^2_{\#}\gamma = \nu$. Luego $\gamma \in \mathcal{A}(\mu, \nu)$ y entonces $\mathcal{A}(\mu, \nu)$ es cerrado en $\mathcal{P}(X \times Y)$. Por lo tanto $\mathcal{A}(\mu, \nu) = \overline{\mathcal{A}(\mu, \nu)}$ es compacto.

Probemos ahora que la aplicación definida por (1.6) es semicontinua inferiormente. Sea $(\gamma_n)_n \subset \mathcal{A}(\mu, \nu)$ tal que $\gamma_n \Rightarrow \gamma$. Como c es a la vez semicontinua y acotada inferiormente, existe sucesión $(c_k)_k$ de funciones continuas y acotadas tales que $c_k \nearrow c$. Para cada n tenemos

$$\int_{X \times Y} c_k\, d\gamma_n \leq \int_{X \times Y} c\, d\gamma_n.$$

Como $\gamma_n \Rightarrow \gamma$, entonces tomando $\liminf_n$ a ambos lados resulta

$$\int_{X \times Y} c_k\, d\gamma \leq \liminf_n \int_{X \times Y} c\, d\gamma_n\ .$$

No podemos asegurar que la expresión del lado derecho de la desigualdad coincida con $\int_{X \times Y} c\, d\gamma$ porque c no necesariamente está acotada. Como $X \times Y$ tiene medida finita, cada c_k está en $L^1(\gamma)$ y por el teorema de convergencia monótona resulta de la anterior desigualdad

$$\int_{X \times Y} c\, d\gamma \leq \liminf_n \int_{X \times Y} c\, d\gamma_n.$$

Luego (1.6) define una aplicación semicontinua inferiormente. ■

El conjunto $\mathcal{P}(X) \times \mathcal{P}(Y)$ es convexo cuando se lo piensa con la estructura lineal heredada del espacio normado $\mathcal{M}(X) \times \mathcal{M}(Y)$. Así, cuando c es una función de costo semicontinua inferiormente y acotada por abajo, está bien definida la aplicación

$$\begin{aligned} \mathcal{P}(X) \times \mathcal{P}(Y) &\longrightarrow (-\infty; +\infty] \\ (\mu, \nu) &\longmapsto \min_{\gamma \in \mathcal{A}(\mu,\nu)} I(\gamma) \end{aligned} \tag{1.11}$$

y resulta convexa.

Proposición 1.2.3. *Con las mismas hipótesis que en 1.2.2, la función definida por (1.11) es convexa.*

Demostración. Sean (μ_0, ν_0) y (μ_1, ν_1) en $\mathcal{P}(X) \times \mathcal{P}(Y)$. Por hipótesis existen planes de transporte $\gamma_0 \in \mathcal{A}(\mu_0, \nu_0)$ y $\gamma_1 \in \mathcal{A}(\mu_1, \nu_1)$ que realizan los transportes óptimos de μ_0 a ν_0 y de μ_1 a ν_1 respectivamente.

Para cada λ en $[0; 1]$ consideremos el par $(\mu_\lambda, \nu_\lambda) \in \mathcal{P}(X) \times \mathcal{P}(Y)$ dado por

$$\mu_\lambda = \lambda\mu_1 + (1-\lambda)\mu_0$$
$$\nu_\lambda = \lambda\nu_1 + (1-\lambda)\nu_0$$

y consideremos también el plan $\gamma_\lambda \in \mathcal{A}(\mu_\lambda, \nu_\lambda)$ dado por

$$\gamma_\lambda = \lambda\gamma_1 + (1-\lambda)\gamma_0 \ .$$

Llamemos $\mathcal{I}$ a la aplicación definida por (1.11) y notemos que

$$(\mu_\lambda, \mu_\lambda) = \lambda(\mu_1, \nu_1) + (1-\lambda)(\mu_0, \nu_0) \ .$$

Por lo tanto tenemos

$$\mathcal{I}(\lambda(\mu_1, \nu_1) + (1-\lambda)(\mu_0, \nu_0)) = \mathcal{I}(\mu_\lambda, \nu_\lambda) = \min_{\gamma \in \mathcal{A}(\mu_\lambda, \nu_\lambda)} \int_{X\times Y} c(x, y)\, d\gamma(x, y) \leq$$

$$\leq \int_{X\times Y} c(x, y)\, d\gamma_\lambda(x, y) = \lambda \int_{X\times Y} c(x, y)\, d\gamma_1(x, y) + (1-\lambda) \int_{X\times Y} c(x, y)\, d\gamma_0(x, y) =$$
$$= \lambda\mathcal{I}(\mu_1, \nu_1) + (1-\lambda)\mathcal{I}(\mu_0, \nu_0) \ ,$$

lo que prueba la convexidad de la aplicación $\mathcal{I}$ definida por (1.11). ■

Una vez probada la existencia de planes minimizantes para el problema de Kantorovich aparecen naturalmente algunas preguntas. Quizás la más obvia sea sobre la unicidad, si la hay y bajo qué condiciones. No estableceremos resultados de unicidad hasta el capítulo

4, cuando probemos unicidad para problemas de Monge–Kantorovich con costos dados por una potencia de la distancia euclídea. Mientras tanto, el siguiente ejemplo muestra que puede haber unicidad en un problema de Monge y no haberla en el problema de Kantorovich asociado.

Ejemplo 1.2.4. Sean los subconjuntos de $\mathbb{R}^2$ dados por $X = \{x_1 = (0,0),\, x_2 = (1,1)\}$ e $Y = \{y_1 = (0,1),\, y_2 = (1,0)\}$. Definamos las medidas

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

y tomemos como función de costo $c(x, y) = |x - y|$ la distancia euclídea. La única función $T : X \to Y$ que transporta μ en ν está dada por $T(x_1) = y_1$ y $T(x_2) = y_2$. Luego hay unicidad en el problema de Monge y se puede ver que γ_T es óptima para el problema de Kantorovich. También es óptima la medida $\gamma \in \mathcal{A}(\mu, \nu)$ dada por

$$\gamma = \frac{1}{3}\left[\delta_{x_1,y_2} + \delta_{x_2,y_1} + \delta_{x_2,y_2}\right]$$

y por lo tanto no hay unicidad para el problema de Kantorovich, pues $\gamma \neq \gamma_T$. ▲

Pensando en el problema de Kantorovich como una relajación del problema original de Monge tiene sentido preguntar bajo qué condiciones se obtienen soluciones igualmente eficientes al permitir dividir masa; es decir, estudiar en qué casos sucede

$$\inf_{T : T_{\#}\mu = \nu} I(T) = \min_{\gamma \in \mathcal{A}(\mu,\nu)} I(\gamma)\ . \tag{1.12}$$

La pregunta es difícil de responder. La dificultad radica en el hecho de que para probar la igualdad (1.12) es necesario construir funciones $T : X \to Y$ que realicen el transporte a un costo tan cercano a mín (1.6) como se desee, pero es complicado garantizar la existencia de funciones de transporte para el problema de Monge.

Antes de enunciar los resultados que se conocen en este sentido fijemos las notaciones

$$I(\gamma) = \int_{X\times Y} c(x, y)\, d\gamma(x, y) \quad \text{e} \quad I(T) = \int_X c(x, T(x))\, d\mu(x)$$

para γ en $\mathcal{A}(\mu, \nu)$ y $T : X \to Y$ tal que $T_{\#}\mu = \nu$.

Como corolario del teorema de dualidad de Kantorovich que demostraremos en 2.4.1 probaremos algunos resultados de existencia de funciones de transporte para medidas soportadas en dominios de $\mathbb{R}^d$. En particular, en 4.2.2 probaremos la existencia de minimizantes para el problema de Monge suponiendo la continuidad absoluta de μ respecto de la medida de Lebesgue d-dimensional y suponiendo que los costos sean inducidos por una función estrictamente convexa de la distancia euclídea. Como consecuencia de ese resultado, habremos probado el siguiente teorema.

Teorema. *Sea B la bola de $\mathbb{R}^d$ y sean $\mu, \nu \in \mathcal{P}(B)$ con μ absolutamente continua respecto de la medida de Lebesgue d-dimensional. Pongamos $X = Y = B$ y consideremos una función de costo $c : X \times Y \to \mathbb{R}$ dada por $c(x, y) = g(x - y)$ con $g : \mathbb{R}^d \to \mathbb{R}$ de clase C^1 y estrictamente convexa. Entonces existe $T : X \to Y$ que realiza el transporte de μ a ν e induce una solución óptima al problema de Kantorovich. Es decir, T verifica*

$$I(T) = \text{mín}(1.6)$$

y por lo tanto vale la igualdad (1.12).

En [19] Gangbo prueba el siguiente resultado con respecto a la existencia de soluciones factibles para el problema de Monge.

Teorema (Gangbo). *Si X e Y son espacios polacos y $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ son no atómicas, entonces existe $T : X \to Y$ que transporta μ en ν.*

Por otra parte, Ambrosio establece en [1] la igualdad (1.12) para medidas no atómicas soportadas en compactos convexos de $\mathbb{R}^n$ cuando la función de costo es continua y acotada.

Teorema (Ambrosio). *Si $X = Y$ es un compacto convexo en $\mathbb{R}^n$, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y μ es no atómica, entonces existe $T : X \to Y$ que transporta que μ en ν. Para el caso particular en el que $c : X \times Y \to \mathbb{R}_{\geq 0}$ es continua y acotada vale la igualdad (1.12).*

Los resultados más fuertes son de 2007 y se deben a Pratelli. En [26], él probó los siguientes resultados.

Teorema (Pratelli). *Sean X e Y espacios polacos, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y μ no atómica. Entonces existe $T : X \to Y$ que transporta μ en ν satisfaciendo*

$$I(T) \leq I(\mu \otimes \nu) ,$$

donde $\mu \otimes \nu$ denota la medida producto de μ y ν en $X \times Y$.

Teorema (Pratelli). *Sean X e Y espacios polacos, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y μ no atómica. Sea $c : X \times Y \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ una función continua y posiblemente no acotada. Entonces, dados $\varepsilon > 0$ y una solución factible γ para el problema de Kantorovich, existe $T_\varepsilon : X \to Y$*

que transporta μ en ν satisfaciendo

$$I(T_\varepsilon) \leq I(\gamma) + \varepsilon \ .$$

Por lo tanto, para una tal c vale la igualdad (1.12).

Este último teorema es en algún sentido el mejor posible dado que deja de valer sin continuidad o sin no–atomicidad de las medidas. En [26] puede verse un ejemplo en el que se muestra la necesidad de la hipótesis de continuidad de c para la validez del resultado.

Por otra parte, cuando las medidas son atómicas no se puede asegurar la existencia de una $T : X \to Y$ que realice el transporte y la no validez de la igualdad (1.12) resulta una cuestión trivial en esos casos. Más interesante es la siguiente pregunta: si existe al menos una $T : X \to Y$ que transporte μ en ν, ¿es necesaria la hipótesis de no–atomicidad para probar la igualdad (1.12)? La respuesta es que sí, como muestra el siguiente ejemplo dado por Pratelli en [26].

Ejemplo 1.2.5. Sean $X = Y = \mathbb{R}^2$. Consideremos dos bolas disjuntas en $\mathbb{R}^2$ de área $\frac{1}{2}$, digamos B_1 y B_2. Sean $x_1, x_2 \in \mathbb{R}^2$ puntos distintos y que no estén en ninguna de las bolas.

Notemos $\mathcal{L}$ a la medida 2-dimensional de Lebesgue y para $E \subset \mathbb{R}^2$ medible notemos $\mathcal{L}\big|_E$ a la restricción de $\mathcal{L}$ a E.

Consideremos las medidas de probabilidad

$$\mu = \mu_1 = \frac{1}{2}\delta_{x_1} + \mathcal{L}\big|_{B_1}$$
$$\nu = \mu_2 = \frac{1}{2}\delta_{x_2} + \mathcal{L}\big|_{B_2}$$

y la distancia euclídea $c(x,y) = |x - y|$ como costo. Es fácil ver que existen funciones $T : X \to Y$ que transportan μ en ν. Por ejemplo, la dada por

$$T(x) = \begin{cases} x_2 & , \text{ si } \ x = x_1 \\ x + (r_2 - r_1) & , \text{ si } \ x \in B_1 \end{cases} ,$$

donde r_1 y r_2 son los centros de las bolas B_1 y B_2 respectivamente. Necesariamente se tiene $T(x_1) = x_2$ para cualquier T que realice el transporte de μ a ν y esta observación resulta en la siguiente cota inferior para $I(T)$:

$$I(T) = \int_X |x - T(x)| \, d\mu(x) \geq \int_{\{x_1\}} |x - T(x)| \, d\mu(x) = \frac{1}{2}|x_1 - x_2| \ .$$

Como la cota vale para toda T que transporte μ en ν, resulta

$$\inf_{T\,:\,T_{\#}\mu=\nu} I(T) \;\geq\; \frac{1}{2}|x_1 - x_2|\;. \tag{1.13}$$

La idea del ejemplo es tomar B_1 lejos de B_2, x_1 cerca de B_2 y x_2 cerca de B_1. Si la distancia entre B_1 y B_2 es mucho mayor que las distancias entre x_1 y B_2 y entre x_2 y B_1, es intuitivamente claro que la solución óptima al problema de Kantorovich consiste en concentrar la masa de B_1 en x_2 y deshacer el átomo de μ en x_1 distribuyéndolo en B_2.

Figura 1.1: soportes de μ y ν

Es claro que $I(\gamma_*) < +\infty$ para cualquier minimizante γ_* del problema de Kantorovich. Si en (Fig. 1.1) movemos B_2 y x_1 rígidamente hacia la derecha — i.e. sin generar movimiento relativo entre ellos — , el valor óptimo del problema de Kantorovich no va a modificarse. Moviendo B_2 y x_1 lo suficientemente hacia la derecha como para que $|x_1 - x_2| > 2 \cdot I(\gamma_*)$, habremos conseguido

$$\inf_{T\,:\,T_{\#}\mu=\nu} I(T) \;\geq\; \frac{1}{2}|x_1 - x_2| \;>\; I(\gamma_*) \;=\; \min_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma)\;.$$

Esto prueba que el ínfimo del problema de Monge es estrictamente mayor que el mínimo del problema de Kantorovich. Notar que en este ejemplo el ínfimo del problema de Monge es también un mínimo. ▲

1.3. El caso finito

Si X e Y son conjuntos finitos y μ, ν son combinaciones convexas de deltas, el problema de Kantorovich se reduce a un problema de programación lineal finita. Consideremos $X = \{x_1, x_2, \ldots, x_m\}$, $Y = \{y_1, y_2, \ldots, y_n\}$ y supongamos que

$$\mu = \sum_{i=1}^{m} \mu_i \delta_{x_i} \quad \text{y} \quad \nu = \sum_{j=1}^{n} \nu_j \delta_{y_j}\;,$$

donde μ_i y ν_j son pesos no negativos para cada i, j. Para transportar masa desde x_i hasta y_j tenemos costo c_{ij} y por lo tanto $c : X \times Y \to \mathbb{R}$ está dada por $c(x_i, y_j) = c_{ij}$ para cada i, j. Toda medida $\gamma \in \mathcal{A}(\mu, \nu)$ es de la forma

$$\gamma = \sum_{i=1}^{m} \sum_{j=1}^{n} \gamma_{ij} \delta_{x_i, y_j} \ ,$$

donde los pesos $(\gamma_{ij})_{ij}$ satisfacen

$$\sum_{j=1}^{n} \gamma_{ij} = \mu_i \ , \ 1 \leq i \leq m \qquad \text{y} \qquad \sum_{i=1}^{m} \gamma_{ij} = \nu_j \ , \ 1 \leq j \leq n \ .$$

Luego, el problema de transporte óptimo de Kantorovich es un problema de programación lineal finita:

$$\begin{aligned} \text{minimizar} \quad & \sum_{i=1}^{m} \sum_{j=1}^{n} \gamma_{ij} c_{ij} \\ \text{sujeto a} \quad & \sum_{j=1}^{n} \gamma_{ij} = \mu_i \qquad 1 \leq i \leq m \\ & \sum_{i=1}^{m} \gamma_{ij} = \nu_j \qquad 1 \leq j \leq n \ . \end{aligned} \tag{1.14}$$

La medida producto $\mu \otimes \nu = \sum_{ij} \mu_i \nu_j \delta_{x_i, y_j}$ muestra que el problema es factible. Por otra parte, para cualquier solución factible γ resulta $0 \leq \gamma_{ij} \leq 1$ y entonces la existencia de un minimizante es inmediata por tratarse de un problema de minimización de un funcional continuo en un compacto no vacío de $\mathbb{R}^{m \times n}$. La dificultad en este caso, relevante para las aplicaciones, es la obtención de algoritmos eficientes para encontrar minimizantes. La bibliografía es extensa y se conocen algoritmos eficientes para resolver estos problemas desde mediados del siglo pasado. Estos temas pueden leerse en [24].

Capítulo 2

Dualidad de Kantorovich

El objetivo principal de este capítulo es probar la equivalencia entre dos formulaciones distintas del problema de Kantorovich: la mencionada anteriormente, llamada **primal**, y una alternativa llamada **dual** que estará dada por un problema de maximización. Los valores óptimos de estos problemas coincidirán y es en este sentido que ambos serán equivalentes. Como aplicación de este resultado, en el capítulo 4 obtendremos soluciones explícitas para el problema de Monge a partir de maximizantes del problema dual.

Veremos primero los teoremas de dualidad en el caso finito y las dificultades que aparecen al intentar generalizar estos resultados al caso no finito. Luego estudiremos los teoremas de dualidad en el contexto de espacios polacos. La última sección estará destinada a probar una forma especial de dualidad cuando el costo proviene de una métrica.

2.1. Dualidad en el caso finito

Consideremos el problema de estudiar la eficiencia de una solución factible para un problema de transporte finito. Siguiendo con la notación de 1.3, en el caso finito el problema de Kantorovich se reduce al programa lineal dado por

$$\begin{aligned} &\text{minimizar} && \sum_{i=1}^{m}\sum_{j=1}^{n}\gamma_{ij}c_{ij} && \\ &\text{sujeto a} && \sum_{j=1}^{n}\gamma_{ij}=\mu_i && 1\leq i\leq m \\ & && \sum_{i=1}^{m}\gamma_{ij}=\nu_j && 1\leq j\leq n\,. \end{aligned} \tag{2.1}$$

Cada solución factible $\gamma = (\gamma_{ij})_{ij}$ de (2.1) tiene costo $p = \sum_{i=1}^m \sum_{j=1}^n \gamma_{ij} c_{ij}$. Llamando p^* al costo mínimo, una medida de la eficiencia de γ como solución de (2.1) está dada por el error relativo de p como aproximación de p^*. Como p^* es desconocido, conocer una cota inferior $\overline{p}$ para p^* es útil para acotar el error de aproximación:

$$0 \leq \frac{p - p^*}{p^*} \leq \frac{p - \overline{p}}{p^*} \leq \frac{p - \overline{p}}{\overline{p}} \ .$$

Una forma de sistematizar la construcción de cotas inferiores para p^* consiste en tomar combinaciones lineales de las restricciones de (2.1) que acoten por abajo la función objetivo del problema (2.1). Para esto consideremos números reales $(\varphi_i)_i$ y $(\psi_j)_j$ a determinar y multipliquemos las restricciones de (2.1) para obtener

$$\begin{aligned} \sum_{j=1}^n \varphi_i \gamma_{ij} &= \varphi_i \mu_i \qquad 1 \leq i \leq m \\ \sum_{i=1}^m \psi_j \gamma_{ij} &= \psi_j \nu_j \qquad 1 \leq j \leq n \ . \end{aligned}$$

Sumando estas expresiones sobre i y j tenemos

$$\sum_{i=1}^m \varphi_i \mu_i + \sum_{j=1}^n \psi_j \nu_j = \sum_{i=1}^m \sum_{j=1}^n \varphi_i \gamma_{ij} + \sum_{j=1}^n \sum_{i=1}^m \psi_j \gamma_{ij} = \sum_{i=1}^m \sum_{j=1}^n (\varphi_i + \psi_j) \gamma_{ij}$$

y esto dice que si $(\varphi_i)_i$ y $(\psi_j)_j$ son elegidos de modo que

$$\varphi_i + \psi_j \leq c_{ij} \tag{2.2}$$

para todo par i, j de índices, entonces tendremos

$$\sum_{i=1}^m \varphi_i \mu_i + \sum_{j=1}^n \psi_j \nu_j \leq \sum_{i=1}^m \sum_{j=1}^n \gamma_{ij} c_{ij} \ . \tag{2.3}$$

Notar que en el lado izquierdo de esta última expresión tendremos una cota inferior para el costo de cualquier solución factible de (2.1) siempre que $(\varphi_i)_i$ y $(\psi_j)_j$ y verifiquen (2.2). Buscar la mejor de las cotas inferiores para el lado derecho de (2.3) induce el siguiente programa lineal:

$$\begin{aligned} \text{maximizar} \quad & \sum_{i=1}^m \varphi_i \mu_i + \sum_{j=1}^n \psi_j \nu_j \\ \text{sujeto a} \quad & \varphi_i + \psi_j \leq c_{ij} \qquad 1 \leq i \leq m \, , \ 1 \leq j \leq n \ . \end{aligned} \tag{2.4}$$

Llamamos **primal** al problema (2.1) y decimos que (2.4) es su problema **dual**. Notando p^* y d^* a los valores óptimos de los problemas primal y dual respectivamente, es claro a partir de la construcción del problema dual que $d^* \leq p^*$. Si nuestro objetivo es conocer el valor óptimo del problema primal, saber que $d^* = p^*$ nos permitiría estudiar el problema dual en lugar del primal y calcular su valor óptimo. En principio sólo podríamos afirmar $d^* \leq p^*$ y, de hecho, existen problemas de optimización en los que tiene sentido una dualidad como la anterior pero $d^* < p^*$.

Llamamos **gap de dualidad** a la diferencia $p^* - d^*$ y **teorema de dualidad** a cualquier teorema que dé condiciones bajo las cuales el gap de dualidad es cero. La utilidad de un tal teorema es clara: permite pasar de un problema a su dual, el cual es equivalente al primero en el sentido de que comparte con éste su valor óptimo. En [24] o [7] pueden verse los dos teoremas fundamentales de dualidad en dimensión finita que enunciamos a continuación.

Dados $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^{m\times 1}$ y $c \in \mathbb{R}^{n\times 1}$, consideremos los problemas

$$(P) \quad \begin{array}{rl} \text{minimizar} & c^t x \\ \text{sujeto a} & Ax = b\ ,\ x \geq 0 \end{array} \qquad \text{y} \qquad (D) \quad \begin{array}{rl} \text{maximizar} & y^t b \\ \text{sujeto a} & A^t y \leq c\ ,\ y \in \mathbb{R}^{m\times 1} \end{array}$$

donde las desigualdades matriciales deben leerse componente a componente. Notar que (D) es el problema dual del primal (P). Con estas notaciones tenemos los siguientes resultados.

Teorema 2.1.1 (Teorema débil de dualidad en dimensión finita). *Dadas soluciones factibles x e y de los problemas primal y dual respectivamente, se tiene $y^t b \leq c^t x$. Es decir, el costo de cualquier solución factible del problema dual es menor o igual que el costo de cualquier solución factible del problema primal.*

Teorema 2.1.2 (Teorema fuerte de dualidad en dimensión finita). *Si el problema primal admite un minimizante, entonces el problema dual también y los valores óptimos de los problemas coinciden. Es decir, si el problema primal admite un minimizante el gap de dualidad es cero.*

Una consecuencia inmediata del teorema fuerte de dualidad es que el problema (2.1) tiene gap de dualidad nulo: como se vio en 1.3, la existencia de un minimizante para el primal es una consecuencia de la compacidad y por lo tanto aplica el teorema 2.1.2.

En el caso general no es tan fácil obtener teoremas de dualidad como lo es en caso finito. En primer lugar, no es tan simple enunciar un problema dual que tenga chances de satisfacer un teorema de dualidad. En segundo lugar, una vez enunciado el problema dual es mucho más complicado probar teoremas en el contexto de espacios polacos abstractos y medidas no atómicas que en $\mathbb{R}^n$ con medidas atómicas.

Cambiando sumas por integrales podemos dar una expresión para el problema primal totalmente análoga a la dada en el caso finito:

$$\begin{aligned} &\text{minimizar} \quad \int_{X\times Y} c(x,y)\, d\gamma(x,y) \\ &\text{sujeto a} \quad \gamma \in \mathcal{A}(\mu,\nu)\ . \end{aligned} \tag{2.5}$$

Las manipulaciones necesarias para deducir el problema dual en el caso finito dependen esencialmente de la linealidad de las sumas. Hay un único detalle adicional que es la permutación de dos sumatorias, pero en todo caso sería fácil de resolver en el caso general trabajando con funciones de costo no negativas. La diferencia importante con el caso finito es que ahora $(\varphi_i)_i$ y $(\psi_j)_j$ deben ser funciones y entonces importan cuestiones de medibilidad e integrabilidad.

Debemos especificar un espacio para las funciones φ y ψ de modo que las integrales cuya suma se desea maximizar no se indefinan. Restringiéndonos a pares (φ, ψ) dados por funciones borelianas

$$\begin{aligned} \varphi : X \to [-\infty; +\infty), &\quad \varphi \in L^1(\mu) \\ \psi : Y \to [-\infty; +\infty), &\quad \varphi \in L^1(\nu)\ , \end{aligned}$$

el problema dual se puede formular como

$$\begin{aligned} &\text{maximizar} \quad \int_X \varphi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y) \\ &\text{sujeto a} \quad \varphi(x) + \psi(y) \le c(x,y) \qquad \text{para todo } (x,y) \in X \times Y\ . \end{aligned} \tag{2.6}$$

Esta formulación es esencialmente mínima en el sentido de que con menos hipótesis ni siquiera tiene sentido el enunciado. El siguiente ejemplo tomado de [6] muestra que sin hipótesis adicionales no es posible garantizar que el gap de dualidad sea cero.

Ejemplo 2.1.3. Sean $X = Y = [0,1]$, $\mu = \nu$ la medida de Lebesgue en $[0,1]$ y c la función de costo dada por

$$c(x,y) = \begin{cases} 0 & , \text{ si } 0 \le y < x \le 1\ , \\ 1 & , \text{ si } 0 \le y = x \le 1\ , \\ +\infty & , \text{ si } 0 \le x < y \le 1\ . \end{cases}$$

Notar que $\gamma = (\mathrm{Id} \times \mathrm{Id})_{\#}\mu \in \mathcal{A}(\mu,\nu)$ tiene costo 1. Para probar que el valor óptimo p^* del problema primal es 1 basta ver que $(\mathrm{Id} \times \mathrm{Id})_{\#}\mu$ es el único plan de transporte con costo finito.

Tomemos $\gamma \in \mathcal{A}(\mu,\nu)$ con costo finito y definamos $A = \{(x,y) \in X \times Y\ :\ y \le x\}$.

Por la finitud del costo de γ resulta $\gamma(A^c) = 0$ y por lo tanto $\gamma(E) = \gamma(E \cap A)$ para todo E en $X \times Y$. Para $t \in (0,1)$ consideremos

$$\begin{aligned} T_t =& \big\{(x,y) \in X \times Y \ : \ 0 \le y \le x \, , \ 0 \le x \le t \ \big\} \, , \\ R_t =& \big\{(x,y) \in X \times Y \ : \ 0 \le y < t \, , \ t < x \le 1 \ \big\} \, . \end{aligned}$$

Tenemos

$$\begin{aligned} &\mu([0,t]) = \gamma([0,t] \times [0,1]) = \gamma(([0,t] \times [0,1]) \cap A) = \gamma(T_t) \ , \\ &\nu([0,t]) = \gamma([0,1] \times [0,t]) = \gamma(([0,1] \times [0,t]) \cap A) = \gamma(T_t \cup R_t) = \gamma(T_t) + \gamma(R_t) \ . \end{aligned}$$

Como $\mu = \nu$ resulta $\gamma(R_t) = 0$ para todo $t \in (0,1)$. Si $(x,y) \in X \times Y$ es tal que $y < x$, tomando $t \in (y,x)$ resulta $(x,y) \in R_t$. Tomando una enumeración de los racionales $\{t_n\}_{n \in \mathbb{N}} = \mathbb{Q} \cap (0,1)$ tenemos

$$\{(x,y) \in [0,1] \times [0,1] \ : \ y < x\} = \bigcup_{n \in \mathbb{N}} R_{t_n} \ .$$

Por lo tanto $\gamma\big(\{(x,y) \in [0,1] \times [0,1] \ : \ x \neq y\}\big) = 0$ y entonces γ está concentrada en la diagonal. Pero la única medida $\gamma \in \mathcal{A}(\mu,\nu)$ concentrada en la diagonal es $(\mathrm{Id} \times \mathrm{Id})_{\#}\mu$. Esto termina de probar que $p^* = 1$.

Veamos ahora que el valor óptimo del problema dual es $d^* = 0$ y que, por lo tanto, el gap de dualidad es $p^* - d^* = 1 > 0$.

Sea (φ, ψ) un par admisible para el problema dual. Dado $\varepsilon > 0$ fijo, para cada $x \in [\varepsilon, 1]$ se tiene $\varphi(x) + \psi(x - \varepsilon) \le c(x, x-\varepsilon) = 0$ y entonces

$$\begin{aligned} 0 \ge \int_\varepsilon^1 \varphi(x) + \psi(x-\varepsilon)\, dx = \int_\varepsilon^1 \varphi(x)\, dx + \int_\varepsilon^1 \psi(x-\varepsilon)\, dx = \\ = \int_\varepsilon^1 \varphi(x)\, dx + \int_0^{1-\varepsilon} \psi(x)\, dx \ . \end{aligned}$$

Tomando límite cuando $\varepsilon \to 0^+$ resulta

$$0 \ge \int_0^1 \varphi(x)\, dx + \int_0^1 \psi(x)\, dx = \int_X \varphi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y) \ .$$

Tomando supremo sobre todos los pares admisibles (φ, ψ) resulta $d^* \le 0$ y, como el par $(0,0)$ es admisible, entonces $d^* = 0$. ▲

2.2. Teorema de dualidad de Kantorovich I

Antes de enunciar y probar los teoremas de dualidad de Kantorovich fijemos la notación que usaremos en esta sección. Sean X e Y espacios polacos y consideremos μ y ν medidas borelianas de probabilidad en X e Y respectivamente.

Definimos los funcionales lineales I y J por

$$I(\gamma) = \int_{X\times Y} c(x,y)\, d\gamma(x,y) \quad , \quad J(\varphi,\psi) = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu$$

para $\gamma \in \mathcal{P}(X\times Y)$ y (φ,ψ) en $L^1(\mu)\times L^1(\nu)$.

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

$$\varphi(x) + \psi(y) \leq c(x,y)$$

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. Cuando sea necesario enfatizar la diferencia entre $\mathcal{F}_c$ y $\mathcal{F}_c \cap C_b$, usaremos la notación $\mathcal{F}_c \cap L^1$ para $\mathcal{F}_c$. Recordar que notamos $\mathcal{A}(\mu,\nu)$ al conjunto de planes de transporte de μ a ν.

Llamamos **problema primal de Kantorovich** al problema

$$\begin{array}{cc} \text{minimizar} & I(\gamma) \\ \text{sujeto a} & \gamma \in \mathcal{A}(\mu,\nu) \end{array}$$

y **problema dual de Kantorovich** al problema

$$\begin{array}{cc} \text{maximizar} & J(\varphi,\psi) \\ \text{sujeto a} & (\varphi,\psi) \in \mathcal{F}_c\ . \end{array}$$

Notar que, en principio, el mínimo y el máximo de los problemas primal y dual podrían ser sólo ínfimo y supremo respectivamente.

Proposición 2.2.1 (Teorema débil de dualidad). *Sean X e Y espacios polacos y sean $\mu \in \mathcal{P}(X)$ y $\nu \in \mathcal{P}(Y)$ medidas borelianas de probabilidad. Para cualquier función de costo $c : X\times Y \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ boreliana valen las desigualdades*

$$\sup_{\mathcal{F}_c \cap C_b} J(\varphi,\psi) \leq \sup_{\mathcal{F}_c \cap L^1} J(\varphi,\psi) \leq \inf_{\gamma \in \mathcal{A}(\mu,\nu)} I(\gamma)\ . \tag{2.7}$$

Más aun, el ínfimo en (2.7) es un mínimo.

Demostración. La desigualdad a la izquierda de (2.7) es trivial a partir de la inclusión $\mathcal{F}_c \cap C_b \subset \mathcal{F}_c \cap L^1$. Luego basta probar la desigualdad de la derecha. Sean (φ, ψ) en $\mathcal{F}_c \cap L^1$ y γ en $\mathcal{A}(\mu, \nu)$. Por definición de $\mathcal{A}(\mu, \nu)$ vale

$$J(\varphi, \psi) = \int_X \varphi \, d\mu + \int_Y \psi \, d\nu = \int_{X \times Y} \varphi(x) + \psi(y) \; d\gamma(x, y) \; .$$

Además vale

$$\varphi(x) + \psi(y) \leq c(x, y)$$

para μ-c.t. $x \in X$ y ν-c.t. $y \in Y$ y, entonces, es posible considerar borelianos N_x, N_y satisfaciendo $\mu(N_x) = \nu(N_y) = 0$ y tales que esta desigualdad valga en $N_x^c \times N_y^c$. Tenemos

$$(N_x^c \times N_y^c)^c \subset (N_x \times Y) \cup (X \times N_y)$$

y por lo tanto

$$\gamma((N_x^c \times N_y^c)^c) \leq \gamma(N_x \times Y) + \gamma(X \times N_y) = \mu(N_x) + \nu(N_y) = 0,$$

donde la anteúltima igualdad es consecuencia de que γ esté en $\mathcal{A}(\mu, \nu)$. Así, la desigualdad

$$\varphi(x) + \psi(y) \leq c(x, y)$$

vale en γ-c.t.p. dado que vale en $N_x^c \times N_y^c$. Luego vale

$$J(\varphi, \psi) = \int_{X \times Y} \varphi(x) + \psi(y) \; d\gamma(x, y) \;\; \leq \int_{X \times Y} c(x, y) \; d\gamma(x, y) = I(\gamma).$$

Tomando supremo en (φ, ψ) e ínfimo en γ resulta la desigualdad a la derecha de (2.7).

Que el ínfimo en (2.7) es un mínimo resulta de la existencia de minimizantes para el problema de Kantorovich probada en 1.2.2. ■

El siguiente teorema es la generalización al caso no finito del teorema fuerte de dualidad enunciado en 2.1.2 para el caso finito. Notar que se necesitan hipótesis adicionales sobre la función de costo c para poder probar el resultado.

Teorema 2.2.2 (Dualidad de Kantorovich - primera versión). *Sean X e Y espacios polacos dotados de sendas medidas borelianas de probabilidad $\mu \in \mathcal{P}(X)$ y $\nu \in \mathcal{P}(Y)$. Sea $c : X \times Y \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ una función de costo semicontinua inferiormente. Entonces*

$$\sup_{(\varphi, \psi) \in \mathcal{F}_c} J(\varphi, \psi) \;=\; \inf_{\gamma \in \mathcal{A}(\mu, \nu)} I(\gamma) \; . \tag{2.8}$$

Adicionalmente, el ínfimo de (2.8) es un mínimo y el supremo permanece inalterado si el supremo se toma sobre $\mathcal{F}_c \cap C_b$.

Para probar esta versión del teorema de dualidad de Kantorovich vamos a escribir el problema $\inf_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma)$ como un problema de tipo "ínf sup" y probar que éste puede ser reescrito como uno de tipo "sup ínf".

Antes de dar una demostración rigurosa dejemos en claro su estructura formal. Para esto definamos primero la **función característica** de $\mathcal{A}(\mu,\nu)$:

$$\chi_{\mathcal{A}}(\gamma) := \begin{cases} 0 & , \text{ si } \gamma \in \mathcal{A}(\mu,\nu) \\ +\infty & , \text{ si no} \end{cases} .$$

Un truco común en optimización restringida es permitir agrandar el conjunto sobre el que se está optimizando a la vez que se agrega una penalización en la función objetivo a optimizar. La utilidad de esta penalización es hacer que el punto donde se alcanza el óptimo permanezca en el dominio original del problema, idealmente en el mismo lugar o controladamente cerca del punto en el que se alcanzaba el óptimo en el problema original.

En nuestro caso la idea funciona particularmente bien: si pensamos $\mathcal{A}(\mu,\nu)$ como un subconjunto de $M_+(X\times Y)$ podemos afirmar

$$\inf_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma) \quad = \inf_{\gamma\in M_+(X\times Y)} I(\gamma) + \chi_{\mathcal{A}}(\gamma). \tag{2.9}$$

Es fácil probar que

$$\chi_{\mathcal{A}}(\gamma) = \sup_{(\varphi,\psi)} \left\{ \int_X \varphi\, d\mu + \int_Y \psi\, d\nu - \int_{X\times Y} \varphi(x) + \psi(y)\, d\gamma(x,y) \right\}, \tag{2.10}$$

donde el supremo se toma sobre los pares de funciones $(\varphi,\psi) \in C_b(X)\times C_b(Y)$. Si ahora reemplazamos (2.10) en (2.9) obtenemos

$$\inf_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma) = \inf_{\gamma\in M_+(X\times Y)} \sup_{(\varphi,\psi)} \left\{ \int_X \varphi\, d\mu + \int_Y \psi\, d\nu - \int_{X\times Y} (\varphi(x)+\psi(y)-c(x,y))\, d\gamma(x,y) \right\}.$$

La parte esencial de la demostración consiste probar que es posible invertir el orden del ínfimo y el supremo. Asumiendo que es posible hacer esto, tenemos

$$\inf_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma) = \sup_{(\varphi,\psi)} \left\{ \int_X \varphi\, d\mu + \int_Y \psi\, d\nu - \sup_{\gamma\in M_+(X\times Y)} \int_{X\times Y} (\varphi(x)+\psi(y)-c(x,y))\, d\gamma(x,y) \right\}.$$

Similarmente, podemos definir la función característica de $\mathcal{F}_c$ y probar

$$\chi_{\mathcal{F}_c}(\varphi,\psi) := \begin{cases} 0 & , \text{ si } (\varphi,\psi) \in \mathcal{F}_c \\ +\infty & , \text{ si no} \end{cases} = \sup_{\gamma \in M_+(X\times Y)} \int_{X\times Y} (\varphi(x) + \psi(y) - c(x,y))\, d\gamma(x,y) \,.$$

Así, obtenemos

$$\inf_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma) \quad = \sup_{(\varphi,\psi)\in\mathcal{F}_c} \left\{ \int_X \varphi\, d\mu + \int_Y \psi\, d\nu \right\} ,$$

que es la igualdad buscada.

La parte delicada de la demostración es el paso en el que se invierten el supremo y el ínfimo. Una forma de hacer esto rigurosamente es utilizando un teorema de dualidad de análisis convexo y separando la demostración en casos, pues cuando X e Y no son compactos aparecen algunas dificultades técnicas adicionales.

Vamos a estructurar la demostración del teorema de dualidad de Kantorovich 2.2.2 en tres pasos, probando resultados de generalidad creciente:

1. Probar el teorema de dualidad de Kantorovich para el caso en el que X e Y son compactos y c es continua. En este paso vamos a necesitar el teorema de dualidad de Fenchel–Rockafellar A.2.7 para invertir el orden en el que se toman un supremo y un ínfimo. Este es el contenido de la proposición 2.2.3.
2. Probar el teorema de dualidad de Kantorovich para el caso en el que c es acotada y uniformemente continua. Esto se prueba en 2.2.4.
3. Probar el teorema de dualidad de Kantorovich 2.2.2. Esto se prueba en la página 36.

Probemos primero la dualidad de Kantorovich asumiendo que X e Y son compactos y c continua. Para la demostración de la dualidad en este caso particular vamos a invocar el teorema de dualidad de Fenchel–Rockafellar A.2.7.

Proposición 2.2.3. *Sean X e Y espacios métricos compactos y μ, ν medidas borelianas de probabilidad en X e Y respectivamente. Si $c : X \times Y \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ es una función de costo continua, entonces*

$$\sup_{(\varphi,\psi)\in\mathcal{F}_c} J(\varphi,\psi) \;=\; \inf_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma) \,.$$

Adicionalmente, en esta igualdad el ínfimo es un mínimo y el supremo no cambia si se toma sobre $\mathcal{F}_c \cap C_b$.

Demostración. El espacio producto $X \times Y$ es compacto porque X e Y los son y, por lo tanto, c es acotada por ser continua en un compacto. Consideremos $C_b(X \times Y)$, el espacio de las funciones continuas y acotadas en $X \times Y$ equipado con la norma del supremo. En virtud del teorema de Riesz–Markov su espacio dual puede ser identificado con $M(X \times Y)$, el espacio de las medidas de Radon regulares en $X \times Y$ normado por la variación total. Con esta correspondencia, cada funcional lineal no negativa en $C_b(X \times Y)$ está definida por una medida boreliana regular y no negativa.

Consideremos ahora dos funciones Θ y Ξ definidas en $C_b(X \times Y)$ así:

$$\begin{aligned} \Theta : C_b(X \times Y) &\longrightarrow \mathbb{R} \cup \{+\infty\} \\ u &\longmapsto \begin{cases} 0 & , \text{ si } u(x,y) \geq -c(x,y) \\ +\infty & , \text{ si no} \end{cases} \end{aligned}$$

$$\begin{aligned} \Xi : C_b(X \times Y) &\longrightarrow \mathbb{R} \cup \{+\infty\} \\ u &\longmapsto \begin{cases} \int_X \varphi \, d\mu + \int_Y \psi \, d\nu & , \text{ si } u(x,y) = \varphi(x) + \psi(y) \\ +\infty & , \text{ si no} \end{cases} \end{aligned}.$$

Notar que Ξ está bien definida. Para verlo, notar que si $\varphi(x) + \psi(y) = \widetilde{\varphi}(x) + \widetilde{\psi}(y)$, entonces existe s en $\mathbb{R}$ tal que $\varphi = \widetilde{\varphi} + s$ y $\psi = \widetilde{\psi} - s$. Siendo μ y ν medidas de probabilidad resulta

$$\int_X \varphi \, d\mu + \int_Y \psi \, d\nu = \int_X \widetilde{\varphi} \, d\mu + \int_X s \, d\mu + \int_Y \widetilde{\psi} \, d\nu - \int_Y s \, d\nu = \int_X \widetilde{\varphi} \, d\mu + \int_Y \widetilde{\psi} \, d\nu.$$

Es claro que Θ es convexa dado que $D(\Theta)$ es convexo y Θ es convexa cuando se la restringe a $D(\Theta)$. La convexidad de Ξ se prueba análogamente. Observemos además que $\Theta(1) = 0 < +\infty$ y que $\Xi(1) = 1 < +\infty$. Para probar que Θ es continua en 1 basta observar que si $u \in C_b(X \times Y)$ cumple $\|u - 1\|_\infty < \frac{1}{2}$, entonces $-c \leq 0 < \frac{1}{2} < u$ y por lo tanto $|\Theta(u) - \Theta(1)| = |0 - 0| = 0$.

A partir de las observaciones anteriores podemos aplicar el teorema de dualidad de Fenchel–Rockafellar A.2.7 a las funciones Θ y Ξ para concluir

$$\operatorname*{ínf}_{u \in E} \big\{\Theta(u) + \Xi(u)\big\} = \operatorname*{máx}_{\gamma \in E^*} \big\{ - \Theta^*(-\gamma) - \Xi^*(\gamma)\big\}. \tag{2.11}$$

Veamos que la igualdad (2.11) implica la tesis del teorema de dualidad de Kantorovich.

En primer lugar, el lado izquierdo de (2.11) está dado por

$$\inf\Big\{\int_X \varphi\,d\mu + \int_Y \psi\,d\nu \;:\; \varphi(x)+\psi(y)\geq -c(x,y)\Big\}$$

$$= \inf\Big\{-\Big[\int_X (-\varphi)\,d\mu + \int_Y (-\psi)\,d\nu\Big] \;:\; (-\varphi)(x)+(-\psi)(y)\leq c(x,y)\Big\}$$

$$= \inf\Big\{-\Big[\int_X \varphi\,d\mu + \int_Y \psi\,d\nu\Big] \;:\; \varphi(x)+\psi(y)\leq c(x,y)\Big\}$$

$$= -\sup\Big\{J(\varphi,\psi) \;:\; (\varphi,\psi)\in \mathcal{F}_c\cap C_b\Big\}. \tag{2.12}$$

Es decir, el lado izquierdo de (2.11) concide con $-\sup_{\mathcal{F}_c\cap C_b} J(\varphi,\psi)$.

Para computar el lado derecho de (2.11) necesitamos conocer primero las transformadas de Legendre–Fenchel de Θ y Ξ. Para cada $\gamma\in M(X\times Y)$ tenemos

$$\begin{aligned}\Theta^*(-\gamma) &= \sup_{u\in C_b(X\times Y)}\Big\{-\int_{X\times Y} u(x,y)\,d\gamma(x,y) \;:\; u(x,y)\geq -c(x,y)\Big\}\\ &= \sup_{u\in C_b(X\times Y)}\Big\{\int_{X\times Y} u(x,y)\,d\gamma(x,y) \;:\; u(x,y)\leq c(x,y)\Big\}\,.\end{aligned} \tag{2.13}$$

Si γ no es una medida no negativa, puede afirmarse que existe una función no positiva $v\in C_b(X\times Y)$ — y por lo tanto menor o igual que c — tal que $\int v\,d\gamma > 0$. Tomando $u=nv$ con $n\to+\infty$ se ve que el supremo (2.13) es $+\infty$. Si γ es no negativa el supremo (2.13) es claramente $\int c\,d\gamma$. Tenemos

$$\Theta^*(-\gamma) = \begin{cases} \int_{X\times Y} c(x,y)\,d\gamma(x,y) & ,\ \text{si } \gamma\in M_+(X\times Y)\\ +\infty & ,\ \text{si no}\end{cases}\,. \tag{2.14}$$

Un razonamiento similar muestra que $\Xi^*(\gamma)$ es la función característica de $\mathcal{A}(\mu,\nu)$:

$$\Xi^*(\gamma) = \begin{cases} 0 & ,\ \text{si } \gamma\in \mathcal{A}(\mu,\nu)\\ +\infty & ,\ \text{si no}\end{cases}\,. \tag{2.15}$$

Ahora sí computamos el lado derecho de (2.11) usando (2.14) y (2.15)

$$\max_{\gamma\in E^*}\Big\{-\Theta^*(-\gamma)-\Xi^*(\gamma)\Big\} = -\min_{\gamma\in E^*}\Big\{\Theta^*(-\gamma)+\Xi^*(\gamma)\Big\} = -\min_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma)\,. \tag{2.16}$$

Juntando (2.16) con (2.12) tenemos la igualdad

$$\min_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma) = \sup_{\mathcal{F}_c\cap C_b} J(\varphi,\psi)$$

y ésta, junto con 2.2.1, prueba la dualidad de Kantorovich bajo las hipótesis asumidas. ■

Notar que cuando X e Y son compactos cualquier c continua es uniformemente continua y acotada en $X\times Y$. Para la siguiente versión de la dualidad de Kantorovich vamos a prescindir de la hipótesis de compacidad de X e Y pero manteniendo la continuidad uniforme y la acotación de c.

Proposición 2.2.4. *Sean X e Y espacios polacos y μ, ν medidas borelianas de probabilidad en X e Y respectivamente. Si $c : X\times Y \to \mathbb{R}_{\geq 0}\cup\{+\infty\}$ es una función de costo acotada y uniformemente continua, entonces*

$$\sup_{(\varphi,\psi)\in\mathcal{F}_c} J(\varphi,\psi) \;=\; \inf_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma)\;.$$

Adicionalmente, en esta igualdad el ínfimo es un mínimo y el supremo no cambia si se toma sobre $\mathcal{F}_c\cap C_b$.

Demostración. Por 1.2.2 existe un plan óptimo γ_* para el problema de Kantorovich

$$I(\gamma_*) = \inf_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma).$$

La idea de la demostración es restringir el problema a un producto de compactos $X_0\times Y_0$ en el que se pueda usar 2.2.3 para después obtener un par de funciones $(\overline{\varphi}_0,\overline{\psi}_0)$ que dé un valor de $J(\overline{\varphi}_0,\overline{\psi}_0)$ cercano a $I(\gamma_*)$. Esta cercanía dependerá de qué tan bien aproxime $X_0\times Y_0$ a $X\times Y$ en un sentido que precisaremos más adelante. Lo importante para la prueba es que la aproximación pueda hacerse arbitrariamente buena.

Sea $\delta > 0$. El espacio producto $X\times Y$ es polaco porque X e Y lo son y, por lo tanto, $\gamma_*\in M(X\times Y)$ es rígida. Luego existen $X_0\subset X$ e $Y_0\subset Y$ compactos tales que $\mu(X\backslash X_0)<\delta$, $\nu(Y\backslash Y_0)<\delta$ y entonces

$$\gamma_*((X_0\times Y_0)^c) = \gamma_*((X\times Y)\backslash(X_0\times Y_0)) < 2\delta. \tag{2.17}$$

Vamos a considerar el problema de Kantorovich y su formulación dual restringidas a X_0 e Y_0. Para distinguir los problemas notaremos con subíndices 0 los objetos del problema restringido como detallamos a continuación.

Definamos primero

$$\gamma_{*0} = \frac{1_{X_0 \times Y_0}}{\gamma_*(X_0 \times Y_0)} \cdot \gamma_* \;\in\; \mathcal{P}(X_0 \times Y_0) \tag{2.18}$$

y también $\mu_0 = \pi_1 \gamma_{*0}$ y $\nu_0 = \pi_2 \gamma_{*0}$. Consideremos las medidas de probabilidad μ_0 en X_0, ν_0 en Y_0 y notemos $\mathcal{A}_0(\mu_0, \nu_0)$ al conjunto de las medidas borelianas de probabilidad en $X_0 \times Y_0$ con marginales μ_0 y ν_0.

Para $\gamma_0 \in \mathcal{A}_0(\mu_0, \nu_0)$ y (φ_0, ψ_0) en $L^1(\mu_0) \times L^1(\nu_0)$ definamos

$$I_0(\gamma_0) = \int_{X_0 \times Y_0} c(x,y)\, d\gamma_0(x,y) \quad \text{y} \quad J_0(\varphi_0, \psi_0) = \int_{X_0} \varphi_0\, d\mu_0 + \int_{Y_0} \psi_0\, d\nu_0 \;.$$

Notemos $\mathcal{F}_{c0}$ al conjunto de pares (φ_0, ψ_0) en $L^1(\mu_0) \times L^1(\nu_0)$ tales que

$$\varphi_0(x) + \psi_0(y) \leq c(x,y)$$

para μ_0-c.t. $x \in X_0$ y ν_0-c.t. $y \in Y_0$.

Por 2.2.3 la dualidad de Kantorovich es válida en $X_0 \times Y_0$ y por lo tanto existen $\widetilde{\gamma}_0 \in \mathcal{A}_0(\mu_0, \nu_0)$ y $(\widetilde{\varphi}_0, \widetilde{\psi}_0)$ en $\mathcal{F}_{c0}$ tales que

$$\begin{gathered} I_0(\widetilde{\gamma}_0) = \inf_{\gamma_0 \in \mathcal{A}_0(\mu_0,\nu_0)} I_0(\gamma_0) \;=\; \sup_{(\varphi_0,\psi_0) \in \mathcal{F}_{c0}} J_0(\varphi_0, \psi_0) \\ J_0(\widetilde{\varphi}_0, \widetilde{\psi}_0) \geq \sup_{(\varphi_0,\psi_0) \in \mathcal{F}_{c0}} J_0(\varphi_0, \psi_0) \;-\delta \;. \end{gathered} \tag{2.19}$$

Definamos $\widetilde{\gamma} \in \mathcal{A}(\mu, \nu)$ como

$$\widetilde{\gamma} = \gamma_*(X_0 \times Y_0) \;\cdot\; \widetilde{\gamma}_0 \;+\; 1_{(X_0 \times Y_0)^c} \;\cdot\; \gamma_* \;. \tag{2.20}$$

Para verificar que efectivamente $\widetilde{\gamma} \in \mathcal{A}(\mu, \nu)$ se debe usar (2.18) y tener en cuenta que $\pi_1 \gamma_{*0} = \mu_0 = \pi_1 \widetilde{\gamma}_0$ y $\pi_2 \gamma_{*0} = \nu_0 = \pi_2 \widetilde{\gamma}_0$. Tenemos

$$\begin{gathered} I(\widetilde{\gamma}) = \int_{X \times Y} c(x,y)\, d\widetilde{\gamma}(x,y) = \gamma_*(X_0 \times Y_0) I_0(\widetilde{\gamma}_0) + \int_{(X_0 \times Y_0)^c} c(x,y)\, d\gamma_*(x,y) \leq \\ \leq I_0(\widetilde{\gamma}_0) + 2\delta \|c\|_\infty = \inf I_0 + 2\delta \|c\|_\infty \end{gathered}$$

y entonces

$$\inf_{\gamma \in \mathcal{A}(\mu,\nu)} I(\gamma) \;=\; \inf_{\gamma_0 \in \mathcal{A}_0(\mu_0,\nu_0)} I_0(\gamma_0) \;+\; 2\delta \|c\|_\infty \;.$$

Ahora vamos a modificar $(\widetilde{\varphi}_0, \widetilde{\psi}_0)$ para obtener un par $(\overline{\varphi}_0, \overline{\psi}_0)$ que dé un valor de $J(\overline{\varphi}_0, \overline{\psi}_0)$ cercano a $I(\gamma_*)$. Más precisamente, un valor lo suficientemente cercano de modo tal que la separación entre $J(\overline{\varphi}_0, \overline{\psi}_0)$ e $I(\gamma_*)$ tienda a cero cuando $\gamma \to 0^+$.

Podemos asumir que $\widetilde{\varphi}_0(x) + \widetilde{\psi}_0(y) \leq c(x,y)$ para todo $(x,y) \in X_0 \times Y_0$ si tenemos el cuidado de redefinir $\widetilde{\varphi}$ y $\widetilde{\psi}$ como $-\infty$ en conjuntos nulos adecuados.

Supongamos adicionalmente que $\delta < 1$, lo cual no invalida ningún argumento anterior. Como $J_0(0,0) = 0$ entonces $\sup J_0 \geq 0$ y por lo tanto

$$\int_{X_0 \times Y_0} \widetilde{\varphi}_0(x) + \widetilde{\psi}_0(y)\, d\gamma_0(x,y) = J_0(\widetilde{\varphi}_0, \widetilde{\psi}_0) \geq \sup J_0 - \delta \geq -\delta \geq -1$$

para todo $\gamma_0 \in \mathcal{A}_0(\mu_0, \nu_0)$. Luego existe $(x_0, y_0) \in X_0 \times Y_0$ tal que $\widetilde{\varphi}_0(x_0) + \widetilde{\psi}_0(y_0) \geq -1$. Para todo $s \in \mathbb{R}$ el par $(\widetilde{\varphi}_0 - s, \widetilde{\psi}_0 + s)$ sigue estando en $\mathcal{F}_{c0}$ y deja invariante el valor de J_0. Es posible elegir s de modo que $\widetilde{\varphi}_0(x_0) + s \geq -\frac{1}{2}$ y $\widetilde{\psi}_0(y_0) - s \geq -\frac{1}{2}$. Luego, cambiando $(\widetilde{\varphi}_0, \widetilde{\psi}_0)$ por un par de la forma $(\widetilde{\varphi}_0 - s, \widetilde{\psi}_0 + s)$ si es necesario, podemos suponer que $\widetilde{\varphi}_0(x_0) \geq -\frac{1}{2}$ y $\widetilde{\psi}_0(y_0) \geq -\frac{1}{2}$.

Para todo $(x,y) \in X_0 \times Y_0$ tenemos

$$\widetilde{\varphi}_0(x) \leq c(x, y_0) - \widetilde{\psi}_0(y_0) \leq c(x, y_0) + \frac{1}{2} \tag{2.21}$$

$$\widetilde{\psi}_0(y) \leq c(x_0, y) - \widetilde{\varphi}_0(x_0) \leq c(x_0, y) + \frac{1}{2}\,. \tag{2.22}$$

Para cada $x \in X$ definimos

$$\overline{\varphi}_0(x) = \inf_{y \in Y_0} c(x,y) - \widetilde{\psi}_0(y). \tag{2.23}$$

Notar que $\overline{\varphi}_0$ es continua en X. Como $(\widetilde{\varphi}_0, \widetilde{\psi}_0) \in \mathcal{F}_{c0}$ tenemos $\widetilde{\varphi}_0 \leq \overline{\varphi}_0$ en X_0 y entonces $J_0(\overline{\varphi}_0, \widetilde{\psi}_0) \geq J_0(\widetilde{\varphi}_0, \widetilde{\psi}_0)$. Podemos controlar $\overline{\varphi}_0$ por arriba y por abajo. Por (2.21) resulta

$$\overline{\varphi}_0(x) = \inf_{y \in Y_0} c(x,y) - \widetilde{\psi}_0(y) \leq c(x, y_0) - \widetilde{\psi}_0(y_0) \leq c(x, y_0) + \frac{1}{2}$$

y esto da la cota

$$\overline{\varphi}_0(x) \leq c(x, y_0) - \widetilde{\psi}_0(y_0) \leq c(x, y_0) + \frac{1}{2}$$

para todo $x \in X$.

Por otra parte, la desigualdad

$$c(x,y) - \widetilde{\psi}_0(y) = c(x,y) - c(x_0,y) + c(x_0,y) - \widetilde{\psi}_0(y) \geq (c(x,y) - c(x_0,y)) - \frac{1}{2}$$

resulta de (2.22). Tomando ínfimo a ambos lados de ésta conseguimos

$$\overline{\varphi}_0(x) = \inf_{y\in Y_0} c(x,y) - \widetilde{\psi}_0(y) \geq \inf_{y\in Y_0} c(x,y) - c(x_0,y) - \frac{1}{2} \tag{2.24}$$

para todo $x \in X$.

Para cada $y \in Y$ definimos

$$\overline{\psi}_0(y) = \inf_{x\in X} c(x,y) - \overline{\varphi}_0(x). \tag{2.25}$$

Por definición de $\overline{\psi}_0$ tenemos que $(\overline{\varphi}_0, \overline{\psi}_0)$ está en $\mathcal{F}_c$ y por lo tanto su restricción a $X_0 \times Y_0$ está en $\mathcal{F}_{c0}$. Además, por (2.23) y (2.25) vale que $\widetilde{\psi}_0 \leq \overline{\psi}_0$ en Y_0 y entonces $J_0(\overline{\varphi}_0, \overline{\psi}_0) \geq J_0(\overline{\varphi}_0, \widetilde{\psi}_0) \geq J_0(\widetilde{\varphi}_0, \widetilde{\psi}_0)$.

Se pueden obtener cotas para $\overline{\psi}_0$ similares a las obtenidas para $\overline{\varphi}_0$:

$$\overline{\psi}_0(y) \geq \inf_{x\in X} c(x,y) - c(x,y_0) - \frac{1}{2} \tag{2.26}$$

$$\overline{\psi}_0(y) \leq c(x_0,y) - \overline{\varphi}_0(x_0) \leq c(x_0,y) - \widetilde{\varphi}_0(x_0) \leq c(x_0,y) + \frac{1}{2}\ .$$

Como consecuencia de (2.24) y (2.26) tenemos las desigualdades

$$\begin{aligned} \overline{\varphi}_0(x) &\geq -2\|c\|_\infty - \frac{1}{2} \\ \overline{\psi}_0(y) &\geq -2\|c\|_\infty - \frac{1}{2} \end{aligned} \tag{2.27}$$

para todo $x \in X$ e $y \in Y$.

Ahora acotamos $J(\overline{\varphi}_0, \overline{\psi}_0)$ por abajo

$$\begin{aligned} J(\overline{\varphi}_0, \overline{\psi}_0) &= \int_X \overline{\varphi}_0\, d\mu + \int_Y \overline{\psi}_0\, d\nu = \int_{X\times Y} \overline{\varphi}_0(x) + \overline{\psi}_0(y)\ \ d\gamma_*(x,y) = \\ &= \int_{X_0\times Y_0} \overline{\varphi}_0(x) + \overline{\psi}_0(y)\ \ d\gamma_*(x,y) + \int_{(X_0\times Y_0)^c} \overline{\varphi}_0(x) + \overline{\psi}_0(y)\ \ d\gamma_*(x,y). \end{aligned} \tag{2.28}$$

Reescribiendo (2.28) según (2.18) tenemos

$$J(\overline{\varphi}_0, \overline{\psi}_0) = \gamma_*(X_0\times Y_0) \int_{X_0\times Y_0} \overline{\varphi}_0(x) + \overline{\psi}_0(y)\ \ d\gamma_{*0}(x,y) + \int_{(X_0\times Y_0)^c} \overline{\varphi}_0(x) + \overline{\psi}_0(y)\ \ d\gamma_*(x,y). \tag{2.29}$$

Recordando la definición de J_0 y aplicando las cotas (2.17) y (2.27) en (2.29) resulta

$$\begin{aligned} J(\overline{\varphi}_0, \overline{\psi}_0) \geq (1-2\delta) J_0(\overline{\varphi}_0, \overline{\psi}_0) - (4\|c\|_\infty + 1)\gamma_*((X_0 \times Y_0)^c) \geq \\ \geq (1-2\delta) J_0(\overline{\varphi}_0, \overline{\psi}_0) - 2(4\|c\|_\infty + 1)\delta. \end{aligned} \tag{2.30}$$

Como $J_0(\overline{\varphi}_0, \overline{\psi}_0) \geq J_0(\widetilde{\varphi}_0, \widetilde{\psi}_0)$ y tenemos (2.19), a partir de (2.30) resulta

$$\begin{aligned} J(\overline{\varphi}_0, \overline{\psi}_0) &\geq (1-2\delta) J_0(\widetilde{\varphi}_0, \widetilde{\psi}_0) - 2(4\|c\|_\infty + 1)\delta \geq \\ &\geq (1-2\delta)(\text{ínf}\, I_0 - \delta) - 2(4\|c\|_\infty + 1)\delta \geq \\ &\geq (1-2\delta)(\text{ínf}\, I - \delta(2\|c\|_\infty + 1)) - 2(4\|c\|_\infty + 1)\delta. \end{aligned} \tag{2.31}$$

Finalmente, tenemos

$$\sup J \geq (1-2\delta)(\text{ínf}\, I - \delta(2\|c\|_\infty + 1)) - 2(4\|c\|_\infty + 1)\delta.$$

Cuando $\delta \to 0^+$ resulta

$$\sup_{\mathcal{F}_c} J \geq \inf_{\mathcal{A}(\mu,\nu)} I$$

y por lo tanto la proposición queda demostrada. ■

Ahora sí estamos en condiciones de dar la demostración de 2.2.2 en toda su generalidad.

Demostración de 2.2.2. Recordar que por 2.2.1 basta ver que

$$\inf_{\gamma \in \mathcal{A}(\mu,\nu)} I(\gamma) \leq \sup_{(\varphi,\psi) \in \mathcal{F}_c \cap C_b} J(\varphi, \psi)\,. \tag{2.32}$$

Por A.1.4 podemos escribir $c = \sup_n c_n$ para una sucesión no decreciente $(c_n)_n$ de funciones uniformemente continuas y no negativas en $X \times Y$. Cambiando c_n por ínf $\{c_n, n\}$ si es necesario, podemos asumir que las c_n están acotadas sin perder generalidad.

Consideramos la familia de problemas de Kantorovich en $\mathcal{A}(\mu, \nu)$ inducidos por los funcionales

$$I_n(\gamma) = \int_{X \times Y} c_n \, d\gamma \tag{2.33}$$

para cada n en $\mathbb{N}$.

Por el caso particular de la dualidad de Kantorovich probado en 2.2.4 sabemos que para cada n en $\mathbb{N}$ vale la igualdad

$$\inf_{\gamma\in\mathcal{A}(\mu,\nu)} I_n(\gamma) = \sup_{(\varphi,\psi)\in\mathcal{F}_{c_n}} J(\varphi,\psi)\,. \tag{2.34}$$

Dado que $c_n \leq c$ para todo n, es claro que si (φ,ψ) está en $\mathcal{F}_{c_n}$ entonces también está en $\mathcal{F}_c$. Esta observación prueba las desigualdades

$$\sup_{(\varphi,\psi)\in\mathcal{F}_{c_n}} J(\varphi,\psi) \leq \sup_{(\varphi,\psi)\in\mathcal{F}_c} J(\varphi,\psi) \tag{2.35}$$

para todo n en $\mathbb{N}$. Notar que este mismo argumento prueba también las desigualdades en $C_b(X)\times C_b(Y)$ y en $L^1(\mu)\times L^1(\nu)$.

Dado que $(c_n)_n$ es no decreciente y $c_n \leq c$ para todo n, entonces dados $m \leq n$ tenemos

$$I_m(\gamma) \leq I_n(\gamma) \leq I(\gamma) \tag{2.36}$$

para toda γ en $\mathcal{A}(\mu,\nu)$. De la desigualdad a la izquierda de (2.36) resulta que

$$\inf_{\gamma\in\mathcal{A}(\mu,\nu)} I_n(\gamma)$$

define una sucesión monótona en n. Por la desigualdad a la derecha de (2.36) esta sucesión es acotada y por lo tanto

$$\lim_n \inf_{\gamma\in\mathcal{A}(\mu,\nu)} I_n(\gamma) = \sup_n \inf_{\gamma\in\mathcal{A}(\mu,\nu)} I_n(\gamma) \leq \inf_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma)\,. \tag{2.37}$$

Veamos que

$$\sup_n \inf_{\gamma\in\mathcal{A}(\mu,\nu)} I_n(\gamma) = \inf_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma)\,. \tag{2.38}$$

Dado que ya probamos (2.37), basta probar

$$\lim_n \inf_{\gamma\in\mathcal{A}(\mu,\nu)} I_n(\gamma) \geq \inf_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma) \tag{2.39}$$

para obtener (2.38).

Por 1.2.2, para cada n existe una medida γ_n en $\mathcal{A}(\mu,\nu)$ y, con el mismo argumento que en 1.2.2, podemos suponer que $(\gamma_n)_n$ converge débilmente a $\gamma_* \in \mathcal{A}(\mu,\nu)$. Dados $m \leq n$ y cambiando γ por γ_n en la desigualdad a la izquierda de (2.36), tenemos

$$\lim_n \inf_\gamma I_n(\gamma) = \lim_n I_n(\gamma_n) = \limsup_n I_n(\gamma_n) \geq \limsup_n I_m(\gamma_n) \geq I_m(\gamma_*)\,.$$

Por el teorema de convergencia monótona tenemos $I_m(\gamma_*) \xrightarrow[m]{} I(\gamma_*)$ y, junto con la desigualdad anterior, esto prueba (2.38). Combinado (2.34) con (2.35) tenemos

$$\inf_{\gamma \in \mathcal{A}(\mu,\nu)} I_n(\gamma) \leq \sup_{(\varphi,\psi)\in \mathcal{F}_c \cap C_b} J(\varphi, \psi) .$$

Tomando supremo del lado izquierdo y usando (2.38) se prueba la desigualdad (2.32). ■

2.3. Convexidad generalizada I

En esta sección vamos a definir la c-transformada y el concepto de c-concavidad. Estas nociones permiten generalizar a un contexto abstracto los conceptos de concavidad y de transformada de Legendre–Fenchel — ver A.2.3 —. En conjunto, resultarán útiles para obtener una versión refinada del teorema de dualidad 2.2.2 que, a su vez, permitirá probar la existencia de minimizantes para el problema de Monge. Continuaremos desarrollando la teoría de convexidad generalizada en el capítulo siguiente y daremos una demostración alternativa del teorema de dualidad vía una caracterización de las soluciones óptimas del problema primal de Kantorovich.

Definición 2.3.1. Dada una función $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$, para cada $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ no constantemente $-\infty$ definimos su ***c*-transformada** como

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

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

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

Otro nombre para la c-transformada es ***c*-conjugada**.

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

$$\varphi(x) + \varphi^c(y) \leq c(x, y) , \tag{2.40}$$

de manera análoga a lo que sucede en (A.2). Si inicialmente se tiene

$$\varphi(x) + \psi(y) \leq c(x, y)$$

para todo (x, y), es fácil ver que $\psi \leq \varphi^c$ y $\varphi \leq \psi^c$.

Salvo que haya indefiniciones del tipo $\infty - \infty$, tiene sentido c-transformar funciones más de una vez. Por ejemplo, para φ como antes se tiene

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

Probaremos en 2.3.6 que para toda φ vale $\varphi^{ccc} = \varphi^c$ y es por este motivo que no tiene sentido c-transformar más de dos veces.

Observación 2.3.2. Al menos formalmente, la transformada de Legendre–Fenchel de φ — cuya definición puede verse en A.2.3 — se correspondería con la c-transformada para el caso particular

$$c(x, y) = -\langle y, x \rangle = -\operatorname{ev}_x(y) = -y(x)$$

del siguiente modo:

$$\varphi^c(y) = \inf_{x \in X} -\langle y, x \rangle - \varphi(x) = -\sup_{x \in X} \langle y, x \rangle + \varphi(x) = -\sup_{x \in X} \langle y, x \rangle - (-\varphi)(x) = -(-\varphi)^*(y)\,.$$

Luego $\varphi^c = -(-\varphi)^*$. Como en el contexto de espacios normados tenemos las nociones usuales de concavidad y convexidad, se tiene $\varphi^c = -(-\varphi)^*$ cóncava por ser la función $(-\varphi)^*$ convexa. ▲

Definición 2.3.3. Una función $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ se dice ***c*-cóncava** si existe alguna función $\psi : Y \to \mathbb{R} \cup \{-\infty\}$, $\psi \not\equiv -\infty$, tal que $\varphi = \psi^c$. Análogamente, una función $\psi : Y \to \mathbb{R} \cup \{-\infty\}$ se dice ***c*-cóncava** si existe $\varphi : X \to \mathbb{R} \cup \{-\infty\}$, $\varphi \not\equiv -\infty$, 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 $\psi = \varphi^c$.

Observación 2.3.4. Continuando con la observación 2.3.2, recordemos que gracias a A.2.6 es posible afirmar que toda función convexa y semicontinua inferiormente es supremo de funciones afines. Análogamente, toda función cóncava y semicontinua superiormente es ínfimo de funciones afines. Notar que si φ es c-cóncava entonces existe ψ tal que

$$\varphi(x) = \inf_{y \in Y} c(x, y) - \psi(y).$$

Podemos pensar que cada aplicación

$$x \mapsto c(x, y) - \psi(y) \tag{2.41}$$

define una función afín generalizada. La analogía está justificada pues cuando

$$c(x,y) = -\langle y, x\rangle$$

las aplicaciones definidas por (2.41) son afines en el sentido usual. Con esta analogía, la definición de c-concavidad diría que una función c-cóncava se puede escribir como el ínfimo de funciones afines generalizadas. El siguiente lema muestra que este hecho caracteriza a las funciones c-cóncavas. ▲

Lema 2.3.5. *Una función $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ es c-cóncava si y sólo si existe una familia $\{(y_i, t_i)\}_{i\in I} \subset Y \times \mathbb{R} \cup \{+\infty\}$ tal que*

$$\varphi(x) = \inf_{i\in I} c(x, y_i) + t_i$$

para todo x en X.

Proposición 2.3.6. *Sean X e Y conjuntos no vacíos y $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$. Sean las funciones $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ y $\psi : Y \to \mathbb{R} \cup \{-\infty\}$. Entonces valen las siguientes afirmaciones:*

1. *Si φ^{cc} está bien definida, resulta $\varphi \leq \varphi^{cc}$ y se tiene la equivalencia:*

$$\varphi = \varphi^{cc} \iff \varphi \text{ es } c\text{-cóncava}.$$

2. *Si están bien definidas φ^{cc} y φ^{ccc}, entonces $\varphi^{c} = \varphi^{ccc}$.*

3. *Valen propiedades análogas para ψ y sus c-transformadas.*

Observación 2.3.7. Notar que el primer ítem de la proposición es una versión abstracta del teorema de biconjugación de Fenchel A.2.6. ▲

Demostración de 2.3.6. Por (2.40) vale

$$\varphi(x) \leq c(x,y) - \varphi^{c}(y)$$

para todo par (x, y). Tomando ínfimo en y resulta

$$\varphi \leq \varphi^{cc}. \tag{2.42}$$

Es claro que para probar una desigualdad análoga para φ^c, es decir $\varphi^c \leq (\varphi^c)^{cc}$, basta un razonamiento análogo al anterior. Combinando esta observación con (2.42) resulta

$$\varphi^{ccc}(y) = \inf_{x\in X} c(x,y) - \varphi^{cc}(x) \leq \inf_{x\in X} c(x,y) - \varphi(x) = \varphi^c(x) \leq (\varphi^c)^{cc}(y) = \varphi^{ccc}(y)$$

para todo y en Y. Luego $\varphi^c = \varphi^{ccc}$, lo que demuestra el segundo ítem.

Es claro que si $\varphi = \varphi^{cc}$ entonces φ es c-cóncava, pues $\varphi = \varphi^{cc} = (\varphi^c)^c$ y entonces $\varphi = \psi^c$ tomando $\psi = \varphi^c$. La afirmación recíproca resulta del segundo ítem, pues si $\varphi = \psi^c$ para alguna ψ, entonces $\varphi^{cc} = (\psi^c)^{cc} = \psi^{ccc} = \psi^c = \varphi$.

Para probar el tercer ítem basta observar que las definiciones de c-transformadas son simétricas para X e Y. ■

2.4. Teorema de dualidad de Kantorovich II

Usando el lenguaje y los resultados de la sección anterior es posible refinar el teorema de dualidad de Kantorovich probando que el supremo del problema dual es en realidad un máximo. Esta segunda formulación del teorema de dualidad es más fuerte que la primera y apenas menos general.

Teorema 2.4.1 (Dualidad de Kantorovich - segunda versión). *Sean X e Y espacios polacos dotados de sendas medidas borelianas de probabilidad $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y sea $c : X \times Y \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ una función de costo semicontinua inferiormente. Sea $\mathcal{A}(\mu,\nu)$ el conjunto de las $\gamma \in \mathcal{P}(X \times Y)$ tales que $\pi^1_{\#}\gamma = \mu$ y $\pi^2_{\#}\gamma = \nu$.*

Supongamos, además, que existen funciones no negativas c_X en $L^1(\mu)$ y c_Y en $L^1(\nu)$ tales que

$$c(x,y) \leq c_X(x) + c_Y(y) \tag{2.43}$$

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

$$\min_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma) = \max\left\{ \int_X \varphi\, d\mu + \int_Y \varphi^c\, d\nu \;:\; (\varphi,\varphi^c) \in L^1(\mu)\times L^1(\nu) \text{ y } \varphi \text{ } c\text{-cóncava} \right\}. \tag{2.44}$$

Luego, existen un par de funciones c-conjugadas (φ,φ^c) en $L^1(\mu)\times L^1(\nu)$ y una medida boreliana de probabilidad $\gamma_* \in \mathcal{A}(\mu,\nu)$ tales que

$$I(\gamma_*) = J(\varphi,\varphi^c).$$

Demostración. En [33] puede verse una demostración para el caso en el que X e Y son compactos y c continua. La extensión por aproximación al caso general puede verse en [32]. En [37] hay indicaciones para una demostración directa del caso general. En 3.3 daremos una demostración distinta basada en la exposición de [2]. ■

Definición 2.4.2. Se llama **potencial c-cóncavo** para el problema de Kantorovich en cuestión a cualquier función c-cóncava φ que induzca un maximizante (φ, φ^c) para el problema dual definido en 2.4.1. Otros nombres posibles son **potencial de Kantorovich** o simplemente **potencial**, cuando no hay ambigüedad sobre el costo c.

Observación 2.4.3. Con la hipótesis adicional de que c sea continua se puede probar que si φ está en $L^1(\mu)$ y es c-cóncava entonces φ^c es medible y valen

$$\int_Y \varphi^c \, d\nu < +\infty$$

y

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

para ν casi todo $y \in Y$ y para todo $x \in X$. Si además $\varphi = \psi^c$ para ψ en $L^1(\nu)$, entonces φ^c está en $L^1(\nu)$ y por lo tanto (φ, φ^c) está en $\mathcal{F}_c$.

Para verlo consideremos φ en $L^1(\mu)$ y c-cóncava. Supongamos que φ^c es $+\infty$ en un subconjunto F de Y de medida positiva. Luego

$$\varphi^c(y) = \inf_{x \in X} c(x, y) - \varphi(x) = +\infty$$

para todo y en F y por lo tanto

$$c(x, y) - \varphi(x) = +\infty \tag{2.45}$$

para todo (x, y) en $X \times F$. Por (3.17) vale

$$c(x, y) - \varphi(x) \leq c_X(x) - \varphi(x) + c_Y(y) \tag{2.46}$$

para todo (x, y) en $X \times Y$. Notar que no tenemos indefiniciones en las restas porque φ es c-cóncava y c es acotada inferiormente. Para γ en $\mathcal{A}(\mu, \nu)$ vale

$$+\infty = \int_{X \times F} c(x, y) - \varphi(x) \, d\gamma(x, y) \leq \nu(F) \int_X c_X(x) - \varphi(x) \, d\mu(x) + \int_Y c_Y(y) \, d\nu(y) < +\infty \tag{2.47}$$

por (2.45), (2.46), (3.17) y porque φ está en $L^1(\mu)$. Pero esto es un absurdo y por lo tanto

φ^c es distinta de $+\infty$ en casi todo punto. Luego la suma $\varphi(x) + \varphi^c(y)$ está definida para ν casi todo $y \in Y$, para todo $x \in X$ y vale

$$\varphi(x) + \varphi^c(y) \leq c(x, y) \, . \tag{2.48}$$

La medibilidad de φ^c resulta de que ésta es un supremo puntual de funciones continuas, pues estamos suponiendo que c es continua.

Notemos que si $\varphi = \psi^c$ entonces $\varphi^c = \psi^{cc} \geq \psi$. Combinando esto con (2.46) y (2.48) tenemos

$$\psi(y) \leq \varphi^c(y) \leq c_X(x) - \varphi(x) + c_Y(y)$$

para todo x en X y ν casi todo y en Y. Luego $\int_Y \varphi^c \, d\nu < +\infty$ y si ψ está en $L^1(\nu)$ entonces φ^c está en $L^1(\nu)$. ▲

La siguiente proposición es un corolario de 2.4.1 y nos dice que los minimizantes del problema primal de Kantorovich están caracterizados por su soporte.

Proposición 2.4.4 (Caracterización de soluciones óptimas del problema dual). *Con las mismas notaciones que en 2.2.2 y 2.4.1 y las hipótesis de 2.4.1, consideremos un par (φ, ψ) en $\mathcal{F}_c$ que maximice el problema dual de Kantorovich. Entonces $\gamma \in \mathcal{P}(X \times Y)$ es un minimizante del problema primal de Kantorovich si y sólo si γ está concentrada en*

$$\Big\{ \ (x, y) \in X \times Y \ : \ \varphi(x) + \psi(y) = c(x, y) \ \Big\} \, . \tag{2.49}$$

Demostración. Sea (φ, ψ) como en el enunciado de la proposición. Si $\gamma \in \mathcal{A}(\mu, \nu)$ está concentrada en (2.49) entonces

$$\gamma\big(\operatorname{sop}(\gamma) \, \Delta \, \{ \ (x, y) \in X \times Y \ : \ \varphi(x) + \psi(y) = c(x, y) \ \} \big) = 0.$$

Luego

$$I(\gamma) = \int_{X \times Y} c(x, y) \, d\gamma(x, y) = \int_{\operatorname{sop}(\gamma)} c(x, y) \, d\gamma(x, y) = \int_{\operatorname{sop}(\gamma)} \varphi(x) + \psi(y) \ \ d\gamma(x, y) =$$

$$= \int_{X \times Y} \varphi(x) + \psi(y) \ \ d\gamma(x, y) = \int_X \varphi(x) \, d\mu(x) + \int_Y \psi(y) \, d\nu(y) = J(\varphi, \psi)$$

y entonces γ es óptima.

Para ver la afirmación recíproca, consideremos $\gamma \in \mathcal{A}(\mu, \nu)$ minimizante del problema primal de Kantorovich. Razonando como antes podemos concluir

$$0 = \int_{X\times Y} c(x,y) - \varphi(x) - \psi(y) \; d\gamma(x,y) \; . \tag{2.50}$$

Como el par (φ, ψ) está en $\mathcal{F}_c$, el integrando de (2.50) es no negativo en casi todo punto y por lo tanto la igualdad (2.50) implica que $c(x,y) - \varphi(x) - \psi(y) = 0$ para γ-c.t.p. Luego (2.49) tiene medida total y entonces γ está concentrada allí. ■

2.5. Dualidad para costos métricos

En esta sección vamos a estudiar problemas de transporte con **costos métricos** — i.e. costos dados por una métrica —. Dado un espacio métrico X, nos van a interesar problemas de transporte de X en X en los que el costo está dado por una métrica d *que no necesariamente define la topología de* X. En particular, nos interesará refinar el teorema de dualidad de Kantorovich para costos métricos.

Empecemos por caracterizar las funciones c-cóncavas en el caso métrico.

Definición 2.5.1. Sea (X,d) un espacio métrico. Notamos $\mathrm{Lip}(X)$ al conjunto de las funciones Lipschitz de X en $\mathbb{R}$. Para cada φ en $\mathrm{Lip}(X)$ notamos $\|\varphi\|_{\mathrm{Lip}}$ a la constante de Lipschitz óptima, es decir

$$\|\varphi\|_{\mathrm{Lip}} = \sup_{x\neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x,y)} \; .$$

En particular, nos va a interesar considerar el subconjunto de $\mathrm{Lip}(X)$ formado por las funciones con constante de Lipschitz óptima menor o igual que 1. Para ello usamos la notación

$$\mathrm{Lip}_1(X) = \left\{\varphi \in \mathrm{Lip}(X) : \|\varphi\|_{\mathrm{Lip}} \leq 1\right\} \; .$$

Proposición 2.5.2. *Sea (X,d) un espacio métrico. Entonces, una función $\varphi : X \to \mathbb{R}$ es d-cóncava si y sólo si pertenece a $\mathrm{Lip}_1(X)$ y, en tal caso, resulta $\varphi^d = -\varphi$.*

Demostración. Sea $\varphi : X \to \mathbb{R}$ función d-cóncava. Luego existe $\psi : X \to \mathbb{R} \cup \{-\infty\}$ tal que $\varphi = \psi^d$. Dado que

$$\varphi(x) = \inf_{y\in Y} d(x,y) - \psi(y) \; ,$$

podemos suponer que ψ nunca toma el valor $-\infty$ porque φ es finita. Para cada y en X definimos φ_y según

$$\varphi_y(x) = d(x,y) - \psi(y)$$

y por lo tanto

$$\varphi(x) = \inf_{y \in Y} \varphi_y(x) \; . \tag{2.51}$$

Se verifica que

$$|\varphi_y(x) - \varphi_y(x')| = |d(x,y) - d(x',y)| \leq d(x,x') \tag{2.52}$$

y entonces es una consecuencia de (2.51) y (2.52) que

$$|\varphi(x) - \varphi(x')| \leq d(x,x') \; .$$

Luego $\varphi \in \mathrm{Lip}_1(X)$. Recíprocamente, dada $\varphi \in \mathrm{Lip}_1(X)$ tenemos

$$\varphi(x) \leq \varphi(y) + d(x,y)$$

para $x, y \in X$ cualesquiera y por lo tanto vale

$$\varphi(x) \leq \inf_{y \in X} \varphi(y) + d(x,y) \leq \varphi(x) + d(x,x) = \varphi(x)$$

para todo x en X. Luego

$$\varphi(x) = \inf_{y \in X} d(x,y) + \varphi(y) = \inf_{y \in X} d(x,y) - (-\varphi)(y)$$

para todo x y entonces $\varphi = (-\varphi)^d$, lo que prueba la d-concavidad de φ. Notar que $-\varphi$ también es un elemento de $\mathrm{Lip}_1(X)$ y por lo tanto $-\varphi = (-(-\varphi))^d = \varphi^d$. ■

Definición 2.5.3. Sea X un espacio polaco y sea d una métrica en X semicontinua inferiormente. Dadas μ y ν en $\mathcal{P}(X)$, definimos $\mathcal{T}_d(\mu,\nu)$ como el valor óptimo para el problema de Kantorovich con costo d y medidas μ y ν. Más precisamente,

$$\mathcal{T}_d(\mu,\nu) = \inf_{\gamma \in \mathcal{A}(\mu,\nu)} \int_{X \times X} d(x,y)\, d\gamma(x,y)$$

y su buena definición resulta de la proposición 1.2.2.

A continuación enunciamos una versión del teorema de dualidad para el caso en que el costo c está dado por una métrica. Notar que para la dualidad basta establecer la igualdad entre $\mathcal{T}_d(\mu,\nu)$ y un supremo o máximo como los que aparecen en las dualidades 2.2.2 y 2.4.1.

Teorema 2.5.4 (Kantorovich–Rubinstein). *Sea X un espacio polaco y sea d una métrica en X semicontinua inferiormente. Dadas μ y ν en $\mathcal{P}(X)$, se tiene la dualidad*

$$\mathcal{T}_d(\mu,\nu) = \max\left\{ \int_X \varphi\, d(\mu-\nu) \ :\ \varphi \in \mathrm{Lip}_1(X) \cap L^1(|\mu-\nu|)\right\} . \tag{2.53}$$

Demostración. Es claro que con las hipótesis exigidas estamos en condiciones de aplicar el teorema de dualidad 2.4.1. Por dicho teorema y por 2.5.2 resulta

$$\begin{aligned}
\mathcal{T}_d(\mu,\nu) &= \max\left\{ \int_X \varphi\, d\mu + \int_X \varphi^d\, d\nu \ :\ (\varphi,\varphi^d) \in L^1(\mu)\times L^1(\nu) \ \text{ y } \ \varphi \ \ d\text{-cóncava} \right\} = \\
&= \max\left\{ \int_X \varphi\, d\mu + \int_X (-\varphi)\, d\nu \ :\ \varphi \in L^1(\mu)\cap L^1(\nu) \ \text{ y } \ \varphi \in \mathrm{Lip}_1(X) \right\} = \\
&= \max\left\{ \int_X \varphi\, d(\mu-\nu) \ :\ \varphi \in L^1(|\mu-\nu|) \ \text{ y } \ \varphi \in \mathrm{Lip}_1(X) \right\} .
\end{aligned}$$

Luego vale (2.53). ■

Observación 2.5.5. Para la discusión que sigue, es importante recordar que todos los problemas y resultados discutidos hasta el momento siguen siendo válidos siempre que μ y ν sean medidas positivas con $0 < \mu(X) = \nu(Y) < +\infty$. Lo importante es que la masa se conserve y no que valga exactamente 1. ▲

Una consecuencia de 2.5.4 es que en un problema de transporte con costos dados por una métrica, es posible conseguir un plan de transporte óptimo en el que la masa compartida por las medidas permanece inmóvil.

Tomemos tres medidas positivas $\mu,\nu,\sigma \in X$ y consideremos el problema de transportar $\mu+\sigma$ en $\nu+\sigma$ con costo dado por una métrica d en X. Si los soportes de μ y ν fueran disjuntos, la masa compartida por $\mu+\sigma$ y $\nu+\sigma$ estaría dada por σ. En cualquier caso, aun si los soportes no son disjuntos, una porción de la masa compartida por $\mu+\sigma$ y $\nu+\sigma$ está dada por σ.

Una forma de hacer el transporte de $\mu+\sigma$ a $\nu+\sigma$ es dejando la masa de σ inmóvil. Podríamos conseguir un plan auxiliar $\widetilde{\gamma}$ que transporte μ en ν de manera óptima y tomar un plan de transporte γ dado por

$$\gamma = \gamma_\sigma + \widetilde{\gamma} , \tag{2.54}$$

donde γ_σ es el transporte de σ en σ dado por la identidad. Como el costo d es una métrica en X, el costo de γ coincide con el de $\widetilde{\gamma}$ — porque el de γ_σ es cero —. Esto prueba la desigualdad

$$\mathcal{T}_d(\mu+\sigma,\nu+\sigma) \le \mathcal{T}_d(\mu,\nu) . \tag{2.55}$$

Es un corolario del teorema de Kantorovich–Rubinstein que valga la igualdad en (2.55).

Corolario 2.5.6. *Sea X un espacio métrico polaco y sea d una métrica en X semicontinua inferiormente. Sean μ, ν y σ medidas borelianas no negativas en X tales que*

$$\mu(X) = \nu(X) < +\infty \quad y \quad \sigma(X) < +\infty \ .$$

Entonces vale que

$$\mathcal{T}_d(\mu + \sigma, \nu + \sigma) = \mathcal{T}_d(\mu, \nu) \ .$$

Demostración. Basta observar que

$$\int_X \varphi \, d((\mu + \sigma) - (\nu + \sigma)) = \int_X \varphi \, d(\mu - \nu)$$

y usar la dualidad (2.53). ∎

Capítulo 3

Optimalidad de Soluciones Factibles

En este capítulo seguimos estudiando la convexidad generalizada y, en particular, la noción de monotonía c-cíclica. Como aplicación de este concepto probamos un teorema de caracterización de soluciones óptimas para el problema primal de Kantorovich y lo usamos para dar una demostración alternativa del teorema de dualidad.

3.1. Convexidad generalizada II

En esta sección X e Y son conjuntos no vacíos y $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ una función de costo. En las siguientes definiciones y proposiciones no asumiremos ninguna regularidad sobre estos objetos a menos que la enunciemos explícitamente.

Definición 3.1.1. Un conjunto $\Gamma \subset X \times Y$ se dice ***c*-cíclicamente monótono** si para cada $m \geq 1$ y para cada familia de pares $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$ en Γ vale que

$$\sum_{i=1}^{m} c(x_i, y_i) \leq \sum_{i=1}^{m} c(x_i, y_{i-1}) \tag{3.1}$$

donde $y_0 \coloneqq y_m$.

Dado que toda permutación puede escribirse como producto de ciclos disjuntos y dado que en la definición de conjunto c-cíclicamente monótono el m es variable, es fácil probar que pedir (3.1) es equivalente a pedir

$$\sum_{i=1}^{m} c(x_i, y_i) \leq \sum_{i=1}^{m} c(x_i, y_{\sigma(i)}) \tag{3.2}$$

para toda σ permutación de $\{1, 2, \dots, m\}$.

Definición 3.1.2. Sea $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ una función c-cóncava. La ***c*-superdiferencial** de φ se define como la relación $\partial^c\varphi$ en $X \times Y$ dada por

$$(x, y) \in \partial^c\varphi \quad \iff \quad \forall z \in X \ , \ \ \varphi(z) \leq \varphi(x) + [\, c(z, y) - c(x, y)\,] \ . \tag{3.3}$$

Dado x en X, notamos $\partial^c\varphi(x)$ al conjunto de los y en Y tales que (x, y) está en $\partial^c\varphi$ y lo llamamos la c-superdiferencial de φ en x. Se definen análogamente la c-superdiferencial y la c-superdiferencial en un punto para funciones c-cóncavas $\psi : Y \to \mathbb{R} \cup \{-\infty\}$.

Lema 3.1.3. *Si $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ es una función c-cóncava entonces*

$$\partial^c\varphi = \big\{(x, y) \in X \times Y \ : \ \varphi(x) + \varphi^c(y) = c(x, y)\big\} \tag{3.4}$$

y por lo tanto todo subconjunto $\Gamma \subset \partial^c\varphi$ es c-cíclicamente monótono. En particular, $\partial^c\varphi$ es c-cíclicamente monótono.

Demostración. Veamos primero que vale la igualdad (3.4). Si $(x, y) \in \partial^c\varphi$, por definición de c-transformada y por (3.3), vale

$$\varphi^c(y) \leq c(x, y) - \varphi(x) \leq c(z, y) - \varphi(z) \ \ \forall z \in X \ . \tag{3.5}$$

Tomando ínfimo en z resulta

$$\varphi^c(y) \leq c(x, y) - \varphi(x) \leq \varphi^c(y) \ . \tag{3.6}$$

Recíprocamente, tomemos (x, y) verificando la igualdad de (3.4). Por la definición de c-transformada tenemos

$$c(x, y) - \varphi(x) = \varphi^c(y) \leq c(z, y) - \varphi(z) \ \ \forall z \in X \tag{3.7}$$

y entonces (x, y) está en $\partial^c\varphi$ por (3.3).

Si $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$ son elementos de $\Gamma \subset \partial^c\varphi$, entonces

$$\sum_{i=1}^{m} c(x_i, y_i) = \sum_{i=1}^{m} [\varphi(x_i) + \varphi^c(y_i)] = \sum_{i=1}^{m} \varphi(x_i) + \sum_{i=1}^{m} \varphi^c(y_i) = \sum_{i=1}^{m} \varphi(x_i) + \sum_{i=1}^{m} \varphi^c(y_\sigma(i)) =$$
$$= \sum_{i=1}^{m} [\varphi(x_i) + \varphi^c(y_\sigma(i))] \leq \sum_{i=1}^{m} c(x_i, y_\sigma(i)) \tag{3.8}$$

para cualquier permutación σ y por lo tanto Γ es c-cíclicamente monótono. Notar que la desigualdad en (3.8) es consecuencia de la definición de c-transformada. ■

El siguiente resultado es una afirmación recíproca a la del lema 3.1.3 conocida como teorema de Rüschendorf. Se trata de una versión abstracta del teorema de Rockafellar que caracteriza las subdiferenciales de funciones convexas. La referencia original es [31].

Teorema 3.1.4 (Rüschendorf). *Supongamos que c es finita. Un subconjunto $\Gamma \subset X \times Y$ es c-cíclicamente monótono si y sólo si está contenido en la c-superdiferencial de una función c-cóncava φ. Más aun, dado $(\overline{x}, \overline{y})$ en Γ, es posible tomar φ cumpliendo $\varphi(\overline{x}) = 0$ y tal que para todo x en X valga*

$$\varphi(x) \leq c(x, \overline{y}) - c(\overline{x}, \overline{y}) \ . \tag{3.9}$$

Demostración. Si Γ está contenido en $\partial^c\varphi$ para una función c-cóncava φ entonces Γ es c-cíclicamente monótono por 3.1.3.

Supongamos que Γ es c-cíclicamente monótono y tomemos $(\overline{x}, \overline{y}) \in \Gamma$ fijo. Definamos una función $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ por

$$\varphi(x) = \inf_{\substack{N \in \mathbb{N} \\ (x_i, y_i) \in \Gamma \\ 1 \leq i \leq N}} [c(x, y_1) - c(x_1, y_1)] + [c(x_1, y_2) - c(x_2, y_2)] + \cdots + [c(x_N, \overline{y}) - c(\overline{x}, \overline{y})] \tag{3.10}$$

para cada x en X. Tomando $N = 1$ y $(x_1, y_1) = (\overline{x}, \overline{y})$ tenemos $\varphi(\overline{x}) \leq 0$. Por ser Γ un conjunto c-cíclicamente monótono, cada elemento de la familia sobre la que tomamos ínfimo en (3.10) es no negativo y entonces $\varphi(\overline{x}) \geq 0$. Luego $\varphi(\overline{x}) = 0$ y además $\varphi(x) < +\infty$ para todo x en X por ser c finita. Esto prueba que $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ está bien definida y que $\varphi \not\equiv -\infty$. Tomando $N = 1$ y $(x_1, y_1) = (\overline{x}, \overline{y})$ obtenemos (3.9).

Por 2.3.5 φ es c-cóncava y entonces sólo falta ver que $\Gamma \subset \partial^c\varphi$. Tomemos $(\widetilde{x}, \widetilde{y}) \in \Gamma$. Poniendo $(x_1, y_1) = (\widetilde{x}, \widetilde{y})$ y separando el primer corchete en (3.10) se tiene

$$\varphi(x) \leq c(x, \widetilde{y}) - c(\widetilde{x}, \widetilde{y}) + \varphi(\widetilde{x})$$

para todo x en X. Luego $(\widetilde{x}, \widetilde{y})$ verifica (3.3) y por lo tanto está en $\partial^c\varphi$. ■

3.2. Caracterización de soluciones óptimas

El objetivo de esta sección es probar un criterio de optimalidad para el problema primal de Kantorovich dando condiciones equivalentes a la optimalidad para los planes de transporte. Este criterio complementa el criterio de optimalidad obtenido en 2.4.4 para el problema dual de Kantorovich.

Proposición 3.2.1 (Condición necesaria de optimalidad para el problema de Kantorovich). *Sean X e Y espacios polacos, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y una función $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ continua y acotada inferiormente. Si γ es un minimizante para el problema de Kantorovich verificando $I(\gamma) < +\infty$, entonces el soporte de γ es c-cíclicamente monótono.*

Observación 3.2.2. La demostración de esta proposición consistirá en suponer que el soporte de γ no es c-cíclicamente monótono y mostrar que, entonces, es posible mejorar el costo. Como se verá, el mejoramiento en el costo es finito y por lo tanto hay dos casos en los que no se llega a una contradicción: cuando el costo de γ es $-\infty$ o cuando es $+\infty$. Las hipótesis que pedimos son tales que eliminan estos casos, donde los planes óptimos no necesariamente estarían soportados en conjuntos c-cíclicamente monótonos.

Es posible construir ejemplos admitiendo planes óptimos que no tengan soportes c-cíclicamente monótonos si se permite que el costo óptimo en el problema de Kantorovich sea infinito. Es fácil construir un ejemplo que admita alguna $\gamma \in \mathcal{A}(\mu, \nu)$ con soporte no c-cíclicamente monótono y en el que toda $\gamma \in \mathcal{A}(\mu, \nu)$ tenga costo $+\infty$. Veamos un ejemplo en el que el costo óptimo sea $-\infty$ y se pueda conseguir con una $\gamma \in \mathcal{A}(\mu, \nu)$ que no tenga soporte c-cíclicamente monótono. ▲

Ejemplo 3.2.3. Consideremos $X = Y = [0; +\infty)$, $\mu = \nu = \mathcal{L}\big|_{[0;+\infty)}$ y

$$c(x, y) = -x - |x - y| \,.$$

Es inmediato que si $T : X \to Y$ es la identidad entonces $T_{\#}\mu = \nu$ y por lo tanto γ_T es una solución factible para el problema de Kantorovich con

$$\operatorname{sop}(\gamma_T) = \big\{(x, x) : x \in [0; +\infty)\big\} \,.$$

Además, el plan γ_T es óptimo porque $I(\gamma_T) = -\infty$. Sin embargo el soporte de γ_T no es

c-cíclicamente monótono. Si (x_1, y_1) y (x_2, y_2) son elementos distintos de $\mathrm{sop}(\gamma_T)$ entonces

$$c(x_1, y_1) + c(x_2, y_2) = -x_1 - x_2 > -x_1 - x_2 - 2|x_1 - x_2| = c(x_1, y_2) + c(x_2, y_1) .$$

Luego c no es c-cíclicamente monótono. Notar que la única hipótesis de 3.2.1 que no se verifica es que c esté acotada inferiormente. ▲

Demostración de 3.2.1. Sea γ una solución óptima para el problema de Kantorovich con $I(\gamma) < +\infty$. Si escribimos $c_0 = \inf c$ tenemos

$$-\infty < c_0 = \int_{X\times Y} c_0 \, d\gamma(x, y) \leq \int_{X\times Y} c(x, y) \, d\gamma(x, y)$$

y entonces $I(\gamma)$ es finito. Supongamos que $\mathrm{sop}(\gamma)$ no es c-cíclicamente monótono. Luego existen $m \geq 1$ y $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$ en $\mathrm{sop}(\gamma)$ tales que

$$\sum_{i=1}^{m} c(x_i, y_i) > \sum_{i=1}^{m} c(x_i, y_{\sigma(i)}) .$$

Sea $\varepsilon > 0$ tal que

$$\varepsilon < \frac{1}{2m}\Big(\sum_{i=1}^{m} c(x_i, y_i) - \sum_{i=1}^{m} c(x_i, y_{\sigma(i)})\Big)$$

y tomemos

$$0 < \delta := \frac{1}{2m}\Big(\sum_{i=1}^{m} c(x_i, y_i) - \sum_{i=1}^{m} c(x_i, y_{\sigma(i)})\Big) - \varepsilon . \tag{3.11}$$

Por la continuidad de c es posible tomar $r > 0$ tal que

$$\begin{aligned} c(x, y) &> c(x_i, y_i) - \varepsilon && \text{para todo } (x, y) \text{ en } B_r(x_i) \times B_r(y_i) \\ c(x, y) &< c(x_i, y_{\sigma(i)}) + \varepsilon && \text{para todo } (x, y) \text{ en } B_r(x_i) \times B_r(y_{\sigma(i)}) . \end{aligned} \tag{3.12}$$

Tomemos $V_i = B_r(x_i) \times B_r(y_i)$ para cada $1 \leq i \leq m$. Notar que, como cada par (x_i, y_i) está en

$$\mathrm{sop}(\gamma) = \Big\{(x, y) \in X \times Y : \gamma(U) > 0 \text{ para todo } U \subset X \times Y \text{ entorno abierto de } (x, y)\Big\} ,$$

resulta $\gamma(V_i) > 0$ para todo $1 \leq i \leq m$. Consideremos

$$\gamma_i := \frac{1}{\gamma(V_i)}\gamma\big|_{V_i} \quad , \qquad \mu_i := \pi_1\gamma_i \quad , \qquad \nu_i := \pi_2\gamma_i \qquad y \qquad \widetilde{\gamma}_i := \mu_i \otimes \nu_{\sigma(i)}$$

para cada $1 \leq i \leq m$. Tomemos $\alpha > 0$ tal que

$$\alpha < \frac{1}{m} \min_{1 \leq i \leq m} \gamma(V_i) \tag{3.13}$$

y consideremos la medida

$$\widetilde{\gamma} = \gamma - \alpha \sum_{i=1}^{m} \gamma_i + \alpha \sum_{i=1}^{m} \widetilde{\gamma}_i \ \in M(X \times Y)\ .$$

Para cada $1 \leq i \leq m$ tenemos $\alpha\gamma_i \leq \frac{1}{m}\gamma$, pues

$$\alpha\gamma_i \ \leq\ \frac{\gamma(V_i)}{m}\gamma_i \ =\ \frac{\gamma(V_i)}{m}\frac{1}{\gamma(V_i)}\gamma\big|_{V_i} \ =\ \frac{1}{m}\gamma\big|_{V_i} \ \leq\ \frac{1}{m}\gamma$$

por (3.13). Luego $\gamma - \alpha\sum_{i=1}^{m}\gamma_i$ es positiva y entonces $\widetilde{\gamma}$ también lo es. Además, $\widetilde{\gamma}$ tiene marginales μ y ν:

$$\pi_1\widetilde{\gamma} = \pi_1\gamma - \alpha\sum_{i=1}^{m}\pi_1\gamma_i + \alpha\sum_{i=1}^{m}\pi_1\widetilde{\gamma}_i = \mu - \alpha\sum_{i=1}^{m}\mu_i + \alpha\sum_{i=1}^{m}\mu_i = \mu$$
$$\pi_2\widetilde{\gamma} = \pi_2\gamma - \alpha\sum_{i=1}^{m}\pi_2\gamma_i + \alpha\sum_{i=1}^{m}\pi_2\widetilde{\gamma}_i = \nu - \alpha\sum_{i=1}^{m}\nu_i + \alpha\sum_{i=1}^{m}\nu_{\sigma(i)} = \nu$$

Estas observaciones prueban que $\widetilde{\gamma}$ está en $\mathcal{A}(\mu,\nu)$. Se deduce a partir de (3.11) y (3.12) que

$$\begin{aligned} I(\widetilde{\gamma}) = \int_{X\times Y} c\,d\widetilde{\gamma} = I(\gamma) - \alpha\Big(\sum_{i=1}^{m}\int_{X\times Y} c\,d\gamma_i - \sum_{i=1}^{m}\int_{X\times Y} c\,d\widetilde{\gamma}_i\Big) \leq \\ \leq I(\gamma) - \alpha\Big(\Big(\sum_{i=1}^{m} c(x_i,y_i) - \sum_{i=1}^{m} c(x_i,y_{\sigma(i)})\Big) - 2m\varepsilon\Big) \leq \\ \leq I(\gamma) - 2m\alpha\Big(\frac{1}{2m}\Big(\sum_{i=1}^{m} c(x_i,y_i) - \sum_{i=1}^{m} c(x_i,y_{\sigma(i)})\Big) - \varepsilon\Big) = \leq I(\gamma) - 2m\alpha\delta < I(\gamma)\ , \end{aligned}$$

donde la última desigualdad resulta de la finitud de $I(\gamma)$. Pero esto contradice la optimalidad de γ y por lo tanto $\operatorname{sop}(\gamma)$ debe ser c-cíclicamente monótono. ∎

Luego de la demostración de 3.2.1 y del ejemplo 3.2.3, quizás las preguntas más obvias sean sobre la necesidad de la hipótesis de continuidad de c y sobre la validez de una afirmación recíproca para 3.2.1. En el primer caso, la respuesta es que sí es posible relajar la continuidad de c a semicontinuidad inferior. La demostración del siguiente resultado puede verse en [20] y [3].

Proposición 3.2.4. *Sean X e Y espacios polacos, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y una función $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ semicontinua y acotada inferiormente. Si γ es un minimizante para el problema de Kantorovich verificando $I(\gamma) < +\infty$, entonces el soporte de γ es c-cíclicamente monótono.*

Con respecto a la segunda pregunta, la respuesta también es afirmativa y es posible probar la validez de la afirmación recíproca bajo condiciones igual de generales. En [3] se probó una afirmación recíproca bajo cierta hipótesis de finitud de momentos de c, para luego ser extendida al caso general en [27] y [34]. En particular, en [34] se prueba el resultado que enunciamos a continuación.

Proposición 3.2.5. *Sean X e Y espacios polacos, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y una función $c : X \times Y \to \mathbb{R}_{\geq 0}$ finita y semicontinua inferiormente. Si $\gamma \in \mathcal{A}(\mu, \nu)$ es c-cíclicamente monótono, entonces es un minimizante para el problema de Kantorovich.*

No probaremos estos resultados pero sí probaremos a continuación un teorema de caracterización de soluciones óptimas para el problema de Kantorovich asumiendo una hipótesis de finitud sobre c. En particular, una afirmación recíproca a la de 3.2.1 está contenida en dicho teorema.

Teorema 3.2.6 (Caracterización de soluciones óptimas del problema primal). *Sean X e Y espacios polacos, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ y $c : X \times Y \to \mathbb{R}$ una función continua y acotada inferiormente.*

Supongamos que existen funciones no negativas c_X en $L^1(\mu)$ y c_Y en $L^1(\nu)$ tales que

$$c(x, y) \leq c_X(x) + c_Y(y) \tag{3.14}$$

para todo (x, y) en $X \times Y$. Sea $\gamma \in \mathcal{A}(\mu, \nu)$ una solución factible para el problema de Kantorovich. Entonces son equivalentes:

1. *El plan de transporte γ es óptimo.*

2. *El soporte de γ es c-cíclicamente monótono.*

3. *Existe una función c-cóncava φ tal que $\max\{\varphi, 0\} \in L^1(\mu)$ y $\mathrm{sop}(\gamma) \subset \partial^c \varphi$.*

Observación 3.2.7. Asumiendo las hipótesis del teorema y escribiendo $c_0 = \inf c$, por (3.14) tenemos

$$c_0 \leq c(x, y) \leq c_X(x) + c_Y(y)$$

para todo (x, y) en $X \times Y$. Luego, dada γ en $\mathcal{A}(\mu, \nu)$ resulta

$$-\infty < c_0 \leq \int_{X \times Y} c(x, y)\, d\gamma(x, y) \leq \int_X c_X(x)\, d\mu(x) + \int_Y c_Y(y)\, d\nu(y) < +\infty \,.$$

Esto dice que c está en $L^1(\gamma)$ para toda γ en $\mathcal{A}(\mu, \nu)$, impidiendo que el costo óptimo en el problema de Kantorovich sea $-\infty$ o $+\infty$. También tenemos

$$\begin{aligned} \int_{\operatorname{sop}(\gamma)} c_X(x) + c_Y(y)\, d\gamma(x, y) = \int_{X \times Y} c_X(x) + c_Y(y)\, d\gamma(x, y) = \\ = \int_X c_X(x)\, d\mu(x) + \int_Y c_Y(y)\, d\nu(y) < +\infty \end{aligned}$$

y entonces $c_X(x) + c_Y(y)$ es finita en γ casi todo punto. Esto dice que para cada γ en $\mathcal{A}(\mu, \nu)$ las funciones c_X y c_Y son ambas finitas para casi todo (x, y) en $\operatorname{sop}(\gamma)$. ▲

Demostración de 3.2.6. Para ver 1) $\Rightarrow$ 2) asumamos que vale 1) y notemos que por la observación anterior tenemos hipótesis suficientes par aplicar 3.2.1. Luego vale 2).

Para ver 2) $\Rightarrow$ 3) supongamos que el soporte de γ es c-cíclicamente monótono. Por la observación anterior podemos tomar $(\overline{x}, \overline{y})$ en el soporte de γ de modo que c_X sea finita en $\overline{x}$ y c_Y en $\overline{y}$. Dado que c es finita por hipótesis, el teorema 3.1.4 nos permite considerar una función c-cóncava φ con $\operatorname{sop}(\gamma) \subset \partial^c \varphi$ y verificando (3.9). Luego

$$\varphi(x) \leq c(x, \overline{y}) - c(\overline{x}, \overline{y}) \leq c_X(x) + c_Y(\overline{y}) - c(\overline{x}, \overline{y})$$

y entonces

$$0 \leq \text{máx}\,\{\varphi(x), 0\} \leq c_X(x) + |c_Y(\overline{y}) - c(\overline{x}, \overline{y})| \,.$$

Así, $\text{máx}\,\{\varphi, 0\}$ está en $L^1(\mu)$ porque c_X lo está y $|c_Y(\overline{y}) - c(\overline{x}, \overline{y})|$ es finito.

Veamos 3) $\Rightarrow$ 1). Para esto supongamos que existe una función c-cóncava φ tal que $\text{máx}\,\{\varphi, 0\} \in L^1(\mu)$ y $\operatorname{sop}(\gamma) \subset \partial^c \varphi$. Por 3.1.3 vale

$$\varphi(x) + \varphi^c(y) = c(x, y) \quad \forall\, (x, y) \in \operatorname{sop}(\gamma) \subset \partial^c \varphi \tag{3.15}$$

y por definición de c-transformada

$$\varphi(x) + \varphi^c(y) \leq c(x, y) \quad \forall\, (x, y) \in X \times Y \,. \tag{3.16}$$

Sea $\widetilde{\gamma} \in \mathcal{A}(\mu, \nu)$. Para probar la optimalidad de γ basta ver que $\int c\, d\gamma \leq \int c\, d\widetilde{\gamma}$ sin ninguna hipótesis adicional sobre $\widetilde{\gamma}$. Por (3.15) y (3.16) tenemos

$$\int_{X\times Y} c(x,y)\,d\gamma(x,y) = \int_{\mathrm{sop}(\gamma)} c(x,y)\,d\gamma(x,y) = \int_{\mathrm{sop}(\gamma)} \varphi(x)+\varphi^c(y)\,d\gamma(x,y) =$$
$$= \int_{X\times Y} \varphi(x)+\varphi^c(y)\,d\gamma(x,y) = \int_X \varphi(x)\,d\mu(x) + \int_Y \varphi^c(y)\,d\mu(y) =$$
$$= \int_{X\times Y} \varphi(x)+\varphi^c(y)\,d\widetilde{\gamma}(x,y) \leq \int_{X\times Y} c(x,y)\,d\widetilde{\gamma}(x,y)$$

y entonces vale 1). ■

3.3. Teorema de dualidad de Kantorovich III

Notar que la demostración del teorema 3.2.6 no depende de los teoremas de dualidad 2.2.2 y 2.4.1. De hecho, es posible dar una demostración del teorema de dualidad de Kantorovich a partir de 3.2.6 y es lo que haremos a continuación.

Teorema 3.3.1 (Dualidad de Kantorovich - tercera versión). *Sean X e Y espacios polacos dotados de sendas medidas borelianas de probabilidad $\mu \in \mathcal{P}(X)$ y $\nu \in \mathcal{P}(Y)$. Sea $\mathcal{A}(\mu,\nu)$ el conjunto de las $\gamma \in \mathcal{P}(X\times Y)$ tales que $\pi^1_\#\gamma = \mu$ y $\pi^2_\#\gamma = \nu$. Sea $c : X\times Y \to \mathbb{R}$ una función de costo continua y acotada inferiormente.*

Para $\gamma \in \mathcal{P}(X\times Y)$ y (φ,ψ) en $L^1(\mu)\times L^1(\nu)$ definimos

$$I(\gamma) = \int_{X\times Y} c(x,y)\,d\gamma(x,y) \quad , \quad J(\varphi,\psi) = \int_X \varphi\,d\mu + \int_Y \psi\,d\nu\ .$$

Supongamos, además, que existen funciones no negativas c_X en $L^1(\mu)$ y c_Y en $L^1(\nu)$ tales que

$$c(x,y) \leq c_X(x) + c_Y(y) \tag{3.17}$$

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

$$\min_{\gamma\in\mathcal{A}(\mu,\nu)} I(\gamma) = \max\left\{ \int_X \varphi\,d\mu + \int_Y \varphi^c\,d\nu \ : \ (\varphi,\varphi^c) \in L^1(\mu)\times L^1(\nu) \ \text{ y } \ \varphi \ \text{ c-cóncava} \right\}. \tag{3.18}$$

Luego, existen un par de funciones c-conjugadas (φ,φ^c) en $L^1(\mu)\times L^1(\nu)$ y una medida boreliana de probabilidad $\gamma_* \in \mathcal{A}(\mu,\nu)$ tales que

$$I(\gamma_*) = J(\varphi,\varphi^c).$$

Demostración. La demostración de que el máximo — que en principio es un supremo — a la derecha de (3.18) es menor o igual que el ínfimo, se hace igual que en 2.2.1. Por lo tanto, para probar el teorema basta ver que existe φ c-concava tal que $(\varphi, \varphi^c) \in L^1(\mu) \times L^1(\nu)$ y

$$\min_{\gamma \in \mathcal{A}(\mu,\nu)} I(\gamma) \;= J(\varphi, \varphi^c)\ . \tag{3.19}$$

Tomemos γ en $\mathcal{A}(\mu, \nu)$ minimizante de I. Por el teorema 3.2.6, existe $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ función c-cóncava tal que $\mathrm{sop}(\gamma) \subset \partial^c\varphi$ y $\max\{\varphi, 0\} \in L^1(\mu)$. Al igual que en la demostración de 3.2.6, es posible tomar $(\overline{x}, \overline{y})$ en el soporte de γ y φ de modo que $c_X(\overline{x})$ y $c_Y(\overline{y})$ sean finitos, $\varphi(\overline{x}) = 0$ y φ verifique (3.9).

Por definición de c-transformada y por (3.17), tenemos

$$\varphi^c(y) \leq c(\overline{x}, y) - \varphi(\overline{x}) = c(\overline{x}, y) \leq c_X(\overline{x}) + c_Y(y)$$

para todo y en Y. Luego

$$0 \leq \max\{\varphi^c(y), 0\} \leq c_X(\overline{x}) + c_Y(y)$$

y entonces $\max\{\varphi^c, 0\} \in L^1(\nu)$. Como tenemos $\max\{\varphi, 0\} \in L^1(\mu)$ y $\max\{\varphi^c, 0\} \in L^1(\nu)$, para ver que φ y φ^c están en L^1 basta probar que no integran $-\infty$. Razonando como en 3) $\Rightarrow$ 1) de la demostración de 3.2.6, y recordando que $c \in L^1(\gamma)$, tenemos

$$-\infty < \int_{X\times Y} c(x, y)\, d\gamma(x, y) = \int_X \varphi(x)\, d\mu(x) + \int_Y \varphi^c(y)\, d\nu(y)\ . \tag{3.20}$$

Luego $(\varphi, \varphi^c) \in L^1(\mu) \times L^1(\nu)$ y por (3.20) se cumple la igualdad (3.19). ∎

Capítulo 4

Minimizantes para el problema de Monge

.

Siguiendo con la discusión inciada en 1.2 , vamos a estudiar ahora los problemas de transporte que tienen lugar en dominios de $\mathbb{R}^d$ con costos dados por una potencia de la distancia euclídea. Tenemos que considerar tres casos distintos. Por un lado están los casos en los que $c(x,y) = |x-y|^p$ con $p > 1$, que serían los casos convexos. Por otro, los casos cóncavos con costos $c(x,y) = |x-y|^p$ para $0 < p < 1$. Por último, el caso límite $c(x,y) = |x-y|$. Este último caso quizás sea el más natural y fue el costo considerado por Monge en la formulación original del problema de transporte.

Los primeros resultados de existencia de minimizantes para el problema de Monge en este contexto fueron probados por Brenier en [11], por McCann y Gangbo en [20] y por Sudakov en [35]. En esta sección vamos a estudiar los tres tipos de costos y a dar una prueba completa para el caso convexo. Para los otros casos se mencionarán resultados conocidos y se harán algunas observaciones, dando las referencias correspondientes donde pueden encontrarse pruebas completas.

Antes de avanzar probemos el siguiente lema que permitirá demostrar la existencia de minimizantes para el problema de Monge.

Lema 4.0.2. *Con las mismas hipótesis y notaciones que en el segundo teorema de dualidad 2.4.1, sea $\gamma \in \mathcal{A}(\mu,\nu)$ un plan de transporte no necesariamente óptimo para el problema de Kantorovich. Supongamos que γ está concentrada en $C \subset X \times Y$. Entonces:*

1. *El conjunto $\big\{\, x \in X \,:\, \exists y \in Y \text{ tal que } (x,y) \in C \big\}$ tiene medida μ total.*

2. *Si para cada $x \in X$ existe a lo sumo un $y \in Y$ con $(x, y) \in C$, entonces γ es de la forma γ_T con $T : X \to Y$ definida en μ casi todo punto según la relación*

$$T(x) = y \quad \Longleftrightarrow \quad (x, y) \in C \ . \tag{4.1}$$

Luego, T es un minimizante para el problema de Monge si γ es un plan óptimo.

Demostración. Llamemos X_0 a la proyección de C sobre X :

$$X_0 = \Big\{ \ x \in X \ : \ \exists y \in Y \text{ tal que } (x, y) \in C \Big\} \ .$$

Notar que $C \subset X_0 \times Y$. Si no valiera *1*, entonces $\mu(X_0) < 1$ y resultaría

$$1 = \int_{X \times Y} d\gamma = \int_C d\gamma \leq \int_{X_0 \times Y} d\gamma = \gamma(X_0 \times Y) = \mu(X_0) < 1 \ ,$$

lo cual es un absurdo. Esto prueba *1*.

Supongamos ahora que para cada $x \in X$ existe a lo sumo un $y \in Y$ con $(x, y) \in C$. Tomemos $y_0 \in Y$ y definamos $T : X \to Y$ igual a y_0 en el complemento de X_0 y según (4.1) en X_0. Probemos que $\gamma = \gamma_T$.

Sea $E \subset X \times Y$ boreliano. Notar que γ_T está concentrada en $\Gamma(T)$ y que $\gamma(\Gamma(T)^c) = 0$. Luego

$$\begin{aligned} \gamma(E) = \int_{X \times Y} \chi_E(x, y) \, d\gamma(x, y) = \int_{\Gamma(T)} \chi_E(x, y) \, d\gamma(x, y) = \\ = \int_{\Gamma(T)} \chi_E(x, T(x)) \, d\gamma(x, y) = \int_{X \times Y} \chi_E(x, T(x)) \, d\gamma(x, y) = \int_X \chi_E(x, T(x)) \, d\mu(x) \end{aligned} \tag{4.2}$$

donde la última igualdad resulta del lema 1.1.6. Por otra parte,

$$\gamma_T(E) = \int_{X \times Y} \chi_E(x, y) \, d\gamma_T(x, y) = \int_X \chi_E(x, T(x)) \, d\mu(x) \ , \tag{4.3}$$

siendo la última igualdad una consecuencia del lema 1.1.10. Combinando (4.2) con (4.3) resulta $\gamma = \gamma_T$ por ser E un boreliano arbitrario. Esto prueba que $T_{\#}\mu = \nu$ porque $\gamma_T = \gamma \in \mathcal{A}(\mu, \nu)$.

Por el lema 1.1.10, la igualdad (1.9) para $f = c$ prueba que el costo de T coincide con el de γ y entonces T es un minimizante del problema de Monge si γ es óptima. ∎

4.1. Caso cuadrático : $p = 2$

Este caso particular del problema de Monge–Kantorovich se conoce como el problema de Monge–Kantorovich con **costo cuadrático**, dado por $c(x, y) = |x - y|^2$. Si bien en la sección siguiente demostraremos la existencia de minimizantes para el problema de Monge en el caso de costos estrictamente convexos generales, vamos a demostrar primero un resultado clásico de existencia y caracterización de minimizantes para el caso cuadrático probado por Brenier.

Notar que el costo cuadrático es equivalente al costo $c(x, y) = \frac{1}{2}|x - y|^2$, en el sentido de que los minimizantes son los mismos para ambos costos. Para obtener expresiones más elegantes estudiaremos este costo equivalente en lugar del costo $c(x, y) = |x - y|^2$.

En esta sección la función de costo c estará dada por $c(x, y) = \frac{1}{2}|x - y|^2$ y, para cada función $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$, notaremos $\overline{\varphi}$ a la función dada por $\overline{\varphi}(x) = \frac{1}{2}|x|^2 - \varphi(x)$.

Lema 4.1.1. *Sea $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$. Entonces φ es c-cóncava si y sólo si $\overline{\varphi}$ es convexa y semicontinua inferiormente. Además, en ese caso vale*

$$y \in \partial^c\varphi(x) \iff y \in \partial\overline{\varphi}(x)$$

para todo x en $\mathbb{R}^d$ y por lo tanto $\partial^c\varphi = \partial\overline{\varphi}$.

Demostración. Dada φ y x en $\mathbb{R}^d$ tenemos:

$$\begin{aligned}\varphi(x) = \inf_y \frac{1}{2}|x - y|^2 - \psi(y) &\iff \varphi(x) = \inf_y \frac{1}{2}|x|^2 + \langle x, -y\rangle + \frac{1}{2}|y|^2 - \psi(y)\\ &\iff \varphi(x) - \frac{1}{2}|x|^2 = \inf_y \langle x, -y\rangle + \left(\frac{1}{2}|y|^2 - \psi(y)\right)\\ &\iff \frac{1}{2}|x|^2 - \varphi(x) = \sup_y \langle x, y\rangle - \left(\frac{1}{2}|y|^2 - \psi(y)\right) \iff \overline{\varphi}(x) = \inf_y \langle y, x\rangle - \overline{\psi}(y)\ .\end{aligned}$$

Luego $\overline{\varphi}$ es c-cóncava si y sólo si $\overline{\varphi} = \overline{\psi}^*$ y por lo tanto $\overline{\varphi}$ es c-cóncava si y sólo si $\overline{\varphi}$ es convexa y semicontinua inferiormente — ver teorema A.2.6.

Para x e y en $\mathbb{R}^d$, por (3.3) tenemos:

$$\begin{aligned}y \in \partial^c\varphi(x) &\iff \varphi(z) \leq \varphi(x) + \frac{1}{2}|z - y|^2 - \frac{1}{2}|x - y|^2 \ \ \forall z\\ &\iff \varphi(z) \leq \varphi(x) + \frac{1}{2}|z|^2 - \langle z, y\rangle - \frac{1}{2}|x|^2 + \langle x, y\rangle \ \ \forall z\end{aligned}$$

$$
\begin{aligned}
&\iff \varphi(z) \leq \varphi(x) + \frac{1}{2}|z|^2 - \langle z - x, y\rangle - \frac{1}{2}|x|^2 \;\; \forall z \\
&\iff \frac{1}{2}|x|^2 - \varphi(x) + \langle y, z - x\rangle \leq \frac{1}{2}|z|^2 - \varphi(z) \;\; \forall z \\
&\iff \overline{\varphi}(x) + \langle y, z - x\rangle \leq \overline{\varphi}(z) \;\; \forall z \iff y \in \partial\overline{\varphi}(x)
\end{aligned}
$$

y esto completa la demostración. ■

Diremos que un conjunto $E \subset \mathbb{R}^d$ es un **conjunto pequeño** si su dimensión de Hausdorff es menor o igual que $d-1$. Puede probarse que cuando $\varphi : \mathbb{R}^d \to \mathbb{R}$ es una función convexa entonces su conjunto de puntos de no diferenciabilidad es un conjunto pequeño [37].

Teorema 4.1.2 (Brenier). *Sean $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ con segundo momento finito. Si μ no da masa a conjuntos pequeños de $\mathbb{R}^d$, entonces el problema de Monge–Kantorovich con costo cuadrático para las medidas μ y ν admite un único minimizante $\gamma \in \mathcal{A}(\mu, \nu)$ dado por el gradiente de una función convexa φ:*

$$\gamma = (\mathrm{Id} \times \nabla\varphi)_{\#}\mu \,.$$

Por lo tanto, $\nabla\varphi$ es el un único minimizante para el problema de Monge correspondiente, donde la unicidad debe entenderse como unicidad en μ casi todo punto.

Demostración. Sea γ un plan óptimo para el problema de Kantorovich definido por μ, ν y el costo $c(x,y) = \frac{1}{2}|x-y|^2$. Tenemos

$$
\begin{aligned}
0 \leq c(x,y) = \frac{1}{2}|x-y|^2 = \frac{1}{2}|x|^2 - \langle x, y\rangle + \frac{1}{2}|y|^2 \leq \\
\frac{1}{2}|x|^2 - |x||y| + \frac{1}{2}|y|^2 = \frac{1}{2}(|x| + |y|)^2 \leq |x|^2 + |y|^2
\end{aligned}
$$

para todo x, y en $\mathbb{R}^d$. Tomando $c_X(x) = \frac{1}{2}|x|^2$ y $c_Y(y) = \frac{1}{2}|y|^2$, se verifican las hipótesis de 3.2.6 y por lo tanto existe una función φ c-cóncava tal que $\mathrm{sop}(\gamma) \subset \partial^c\varphi$. Por el lema 4.1.1, $\overline{\varphi}$ es una función convexa y entonces el conjunto de puntos en los que $\overline{\varphi}$ es no diferenciable tiene dimensión de Hausdorff menor o igual que $d-1$. Luego $\overline{\varphi}$ es diferenciable salvo en un conjunto de medida μ nula y por lo tanto

$$\gamma\Big(\partial\overline{\varphi} \,\Delta \left\{ (x, \nabla\overline{\varphi}(x)) \;:\; \overline{\varphi} \text{ diferenciable en } x \right\} \Big) = 0 \,. \tag{4.4}$$

Por (4.4) y porque $\mathrm{sop}(\gamma) \subset \partial^c\varphi = \partial\overline{\varphi}$, se deduce que γ está concetrada en el gráfico de $\nabla\overline{\varphi}$. Por el lema 4.0.2, $\gamma = \gamma_{\nabla\overline{\varphi}}$ y $\nabla\overline{\varphi}$ es un minimizante para el problema de Monge.

Notar que $\text{sop}(\gamma) \subset \partial^c\varphi$, la optimalidad de γ y 3.1.3 implican que (φ, φ^c) es un maximizante para el problema dual. Así, si $\widetilde{\gamma}$ es otro minimizante para el problema de Kantorovich, podemos concluir por 2.4.4 y 3.1.3 que $\widetilde{\gamma}$ está concentrada en el gráfico de $\nabla\overline{\varphi}$. Razonando como antes se deduce $\widetilde{\gamma} = \gamma_{\nabla\overline{\varphi}} = \gamma$. Luego $\gamma_{\nabla\overline{\varphi}}$ es el único minimizante para el problema de Kantorovich y por lo tanto $\nabla\overline{\varphi}$ es el único minimizante para el problema de Monge. ■

Observación 4.1.3. Notar que si μ es absolutamente continua respecto de la medida de Lebesgue d-dimensional entonces μ no da masa a conjuntos pequeños. ▲

4.2. Caso estrictamente convexo : $1 < p$

Como se vió en la demostración del teorema de Brenier, los teoremas de dualidad dan la posibilidad de construir explícitamente soluciones para el problema de Monge a partir de un par (φ, ψ) de funciones c-conjugadas maximizantes del problema dual. En esta sección explotaremos ese hecho construyendo minimizantes para el problema de Monge con cualquier costo estrictamente convexo.

Definición 4.2.1. Vamos a llamar **dominio** a cualquier subconjunto no vacío y conexo de $\mathbb{R}^d$ que coincida con la clausura de su interior. Por ejemplo, son dominios las clausuras de abiertos conexos y no vacíos.

Teorema 4.2.2. *Sea Ω un dominio compacto de $\mathbb{R}^d$ y μ, ν medidas de probabilidad en Ω. Sea $g : \mathbb{R}^d \to \mathbb{R}$ una función estrictamente convexa de clase C^1 y, tomando $X = Y = \Omega$, consideremos $c : X \times Y \to \mathbb{R}$ dada por $c(x, y) = g(x - y)$. Si μ es absolutamente continua respecto de la medida de Lebesgue d-dimensional y $\mu(\partial\Omega) = 0$, entonces el problema de Monge es factible y admite un minimizante $T : X \to Y$. Más aun, existe un potencial φ para el problema de modo tal que*

$$T(x) = x - ((\nabla g)^{-1} \circ \nabla\varphi)(x) \tag{4.5}$$

para μ-c.t. x en X. Además, el problema de Kantorovich tiene solución única γ_T bajo estas mismas hipótesis.

Demostración. Por 2.4.1 podemos considerar un par de funciones c-conjugadas (φ, ψ) que maximice el problema dual. Tenemos

$$\varphi(x) = \inf_{y \in Y} g(x - y) - \psi(y)$$

para todo x en X. Veamos que φ satisface una condición de Lipschitz. Para cada y en Y notemos φ_y a la función en X dada por

$$\varphi_y(x) = g(x-y) - \psi(y) \ .$$

y observemos que

$$\varphi(x) = \inf_{y \in Y} \varphi_y(x) \ . \tag{4.6}$$

Con esta notación tenemos

$$\varphi_y(x) - \varphi_y(x') = g(x-y) - g(x'-y) = \langle\, \nabla g(\theta(x,x',y))\, ,\, x - x'\, \rangle \ .$$

Por la compacidad de Ω existe una bola B centrada en 0 que contiene a Ω y entonces para x,x' e y en Ω vale que $x-y$, $x'-y$ y $x-x'$ están en $2B$. Como $2B$ es convexa, entonces $\theta(x,x',y)$ también está en $2B$. Luego

$$|\varphi_y(x) - \varphi_y(x')| \leq \|\nabla g\|_{L^\infty(2B)} |x - x'| \tag{4.7}$$

para x, x' en X e y en Y.

A partir de (4.7) y (4.6) tenemos

$$\varphi(x) \leq \varphi_y(x) \leq \varphi_y(x') + \|\nabla g\|_{L^\infty(2B)} |x - x'|$$

y por lo tanto

$$\varphi(x) \leq \varphi_y(x') + \|\nabla g\|_{L^\infty(2B)} |x - x'|\, ,$$

de donde se concluye

$$\varphi(x) - \varphi(x') \leq \|\nabla g\|_{L^\infty(2B)} |x - x'|$$

tomando ínfimo en y.

Repitiendo la cuenta con x y x' permutados se concluye

$$|\varphi(x) - \varphi(x')| \leq \|\nabla g\|_{L^\infty(2B)} |x - x'|$$

y por lo tanto φ verifica una condición de Lipschitz con constante $\|\nabla g\|_{L^\infty(2B)}$.

Como φ es Lipschitz, podemos restringir φ al interior de Ω y el teorema de Rademacher nos dice que $\varphi : \Omega^\circ \to \mathbb{R}$ es diferenciable en casi todo punto respecto de la medida de Lebesgue d-dimensional. Como μ es absolutamente continua respecto de la medida de

Lebesgue, entonces φ es diferenciable en Ω° salvo por un conjunto S de medida μ nula.

Sean $x_0 \in \Omega^\circ \setminus S$ e $y_0 \in \Omega$ tales que

$$\varphi(x_0) + \psi(y_0) = c(x_0, y_0) = g(x_0 - y_0) \ . \tag{4.8}$$

Veamos que y_0 queda unívocamente determinado por x_0. Para esto, consideremos la aplicación

$$x \mapsto g(x - y_0) - \varphi(x) \ . \tag{4.9}$$

Dado que $\varphi = \psi^c$, tenemos

$$g(x - y_0) - \varphi(x) \geq \psi(y_0)$$

para todo x en Ω y por (4.8) la igualdad vale en el caso de que x sea x_0 . Luego, la aplicación definida por (4.9) tienen un mínimo local en x_0 y a partir de la anulación de su gradiente allí se concluye

$$\nabla g(x_0 - y_0) = \nabla \varphi(x_0) \ . \tag{4.10}$$

Por ser g estrictamente convexa y de clase C^1, su gradiente es una función inyectiva y por lo tanto es posible despejar y_0 en (4.10) para obtener

$$y_0 = x_0 - ((\nabla g)^{-1} \circ \nabla \varphi)(x_0) \ . \tag{4.11}$$

Hemos probado que si x está en $\in \Omega^\circ \setminus S$ entonces existe a lo sumo un y en Ω tal que (x, y) satisface la condición (4.8).

Tomemos $\gamma \in \mathcal{A}(\mu, \nu)$ minimizante del problema de Kantorovich. Por 2.4.4, sabemos que γ está concentrada en

$$\Big\{ \ (x, y) \in X \times Y \ : \ \varphi(x) + \psi(y) = c(x, y) = g(x - y) \ \Big\} \ .$$

Considerando

$$C = \Big\{ \ (x, y) \in (\Omega^\circ \setminus S) \times \Omega \ : \ \varphi(x) + \psi(y) = g(x - y) \ \Big\} \ ,$$

estamos en condiciones de aplicar el lema 4.0.2 y entonces (4.5) define un minimizante para el problema de Monge. Notar que definimos γ en términos de φ y de g, pero no en términos de γ. Esta observación, junto con el hecho de que $\gamma = \gamma_T$ como consecuencia del lema 4.0.2, prueba que γ es única. ■

Observación 4.2.3. Tomando $g(x) = |x|^p = \left(\sum_i x_i^2\right)^{\frac{p}{2}}$, se ve que el teorema anterior cubre el caso de costos de la forma $c(x, y) = |x - y|^p$ con $p > 1$. Notar que el caso $c(x, y) = |x - y|$ requiere que $p = 1$ y en ese caso g deja de ser estrictamente convexa. ▲

Una discusión de los resultados más importantes para el caso de costos convexos, y no sólo estrictamente convexos, puede leerse en [33]. El resultado más reciente que se menciona allí es de 2014 y prueba la existencia de minimizantes para el problema de Monge en el caso de costos de la forma $c(x, y) = g(x-y)$ con $g : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ convexa. La referencia es [8].

4.3. Caso estrictamente cóncavo : $0 < p < 1$

En el caso cóncavo sucede algo que en el caso convexo no: la masa compartida por μ y ν queda fija por cualquier plan de transporte óptimo. Antes de enunciar precisamente el resultado recordemos las siguientes definiciones.

Definición 4.3.1. Dadas dos medidas μ y ν en una misma σ-álgebra, decimos que μ y ν son **mutuamente singulares** si existen A y B disjuntos tales que μ está concentrada en A y ν está concentrada en B.

Dos medidas positivas que no son mutuamente singulares comparten masa. La siguiente definición dice cómo construir una medida que represente esa masa.

Definición 4.3.2. Dadas dos medidas positivas μ y ν en una misma σ-álgebra, definimos la **parte común** de μ y ν como la medida positiva $\mu \wedge \nu$ dada por $\mu - [\mu - \nu]_+$, donde $[\mu - \nu]_+$ es la parte positiva de $\mu - \nu$. Equivalentemente, $\mu \wedge \nu$ puede definirse como $\nu - [\nu - \mu]_+$.

Proposición 4.3.3. *Sean μ y ν medidas borealianas de probabilidad en $\Omega \subset \mathbb{R}^d$ y sea el costo $c(x, y) = l(|x - y|)$, donde $l : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ es una función estrictamente cóncava, de clase C^1, creciente y cumpliendo $l(0) = 0$. Sea $\gamma = \gamma_D + \gamma_O$ un minimizante para el problema de Kantorovich, donde γ_D es la restricción de γ a la diagonal*

$$D = \{(x, x) : x \in \Omega\}$$

y γ_O es la restricción de γ al complemento de D en $\Omega \times \Omega$. Entonces las medidas $\pi^1_\# \gamma_O$ y $\pi^2_\# \gamma_O$ son mutuamente singulares.

Notar que esta proposición dice cómo deben ser los planes de transporte óptimos de μ a ν: la masa que comparten se queda quieta y se transporta de manera óptima la masa de μ no compartida por ν. Cuando μ y ν son mutuamente disjuntas la observación es irrelevante. Sin embargo, la observación señala una restricción importante para el caso en que no es así. Si μ y ν son tales que para algún boreliano E vale

$$\mu(E) > (\mu \wedge \nu)(E) = \nu(E) > 0 \ ,$$

entonces cualquier plan de transporte óptimo γ deberá trasladar masa de E hacia fuera de E. Pero la masa compartida por μ y ν en E deberá permanecer quieta y esto dice que no puede existir un plan óptimo de la forma γ_T. Es decir, en este caso el problema de Monge no sería factible. Notar la diferencia con el caso convexo en el que no existían estos impedimentos.

El siguiente teorema afirma que sí existen minimizantes para el problema de Monge cuando las medidas son mutuamente singulares .

Teorema 4.3.4. *Sean μ y ν dos medidas borelianas de probabilidad mutuamente singulares en $\mathbb{R}^d$. Sea $l : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ una función estrictamente cóncava, de clase C^1, creciente y cumpliendo $l(0) = 0$. Consideremos $c : X \times Y \to \mathbb{R}$ dada por $c(x, y) = l(|x - y|)$. Si μ es absolutamente continua respecto de la medida de Lebesgue d-dimensional, entonces el problema de Kantorovich tiene un único minimizante y está inducido por una aplicación $T : X \to Y$. Luego, T es un minimizante para el problema de Monge.*

La demostración de 4.3.4 puede hacerse repitiendo la estructura de la demostración de 4.2.2, pero hay que corregir dos detalles: la inyectividad de ∇g y la diferenciabilidad en casi todo punto de φ.

En primer lugar, ∇g sigue siendo inyectiva pero por motivos distintos — antes usamos la convexidad estricta de g —. En este caso tenemos $g(x) = l(|x|)$ y

$$\nabla g(x) = l'(|x|) \cdot \frac{x}{|x|} \, . \tag{4.12}$$

Por ser l estrictamente cóncava resulta l' inyectiva. Si $\nabla g(x_1) = \nabla g(x_1)$ se concluye $x_1 = x_2$ tomando módulo en (4.12) y luego usando la inyectividad de l'. Esto prueba la inyectividad de ∇g.

Por otra parte, ya no es cierto que la función de costo sea Lipschitz porque cerca de cero l puede tener pendiente no acotada, por ejemplo si $l(x) = \sqrt{x}$. Esto impide concluir que φ sea diferenciable en casi todo punto. Para el caso estrictamente convexo lo hicimos probando que φ era Lipschitz y aplicando el teorema de Rademacher, pero el argumento dependía de que c fuera Lipschitz. Ahora c no es Lipschitz porque l no lo es. Para solucionar este problema vamos a definir una noción débil de diferenciabilidad y a probar que φ es diferenciable en ese sentido.

Observemos primero que $f : \Omega \subset \mathbb{R}^d \to \mathbb{R}$ es diferenciable en $x_0 \in \Omega$ y $v = \nabla f(x_0)$ es su gradiente en ese punto si y sólo si para todo $\varepsilon > 0$ el conjunto

$$A(x_0, v, \varepsilon) = \Big\{ x \in \Omega \,:\, |f(x) - f(x_0) - v \cdot (x - x_0)| > \varepsilon |x - x_0| \Big\} \tag{4.13}$$

está a distancia positiva de x_0 o, equivalentemente, cumple $B(x_0; \delta) \cap A(x_0, v, \varepsilon) = \emptyset$ para δ pequeño. Podemos cambiar esta condición por otra más débil y definir la noción de gradiente aproximado de f en x_0.

Definición 4.3.5. Sea $\Omega \subset \mathbb{R}^d$ un abierto, $f : \Omega \subset\to \mathbb{R}$ y $x_0 \in \Omega$. Decimos que $v \in \mathbb{R}^d$ es el **gradiente aproximado** de f en x_0, y lo notamos $\nabla_{\mathrm{ap}} f(x_0)$, si para todo $\varepsilon > 0$ el conjunto $A(x_0, v, \varepsilon)$ tiene densidad cero en x_0. Es decir, si para todo $\varepsilon > 0$ vale

$$\lim_{\delta \to 0} \frac{|A(x_0, v, \varepsilon) \cap B(x_0; \delta)|}{|B(x_0; \delta)|} = 0 \,.$$

Cuando existe $\nabla_{\mathrm{ap}} f(x_0)$ decimos que f es **aproximadamente diferenciable** en x_0.

El gradiente aproximado comparte algunas propiedades con el gradiente usual, como la unicidad y la linealidad. Es claro a partir de la definición que si f es diferenciable en x_0 entonces también es aproximadamente diferenciable en x_0 y allí coinciden su gradiente aproximado y su gradiente en sentido usual. En [15] se puede ver en detalle la teoría de diferenciabilidad aproximada.

Para poder demostrar lo que nos interesa vamos a necesitar el siguiente teorema tipo Rademacher.

Teorema 4.3.6. *Sea $\Omega \subset \mathbb{R}^d$ un abierto y $f, g : \Omega \subset\to \mathbb{R}$ funciones con g Lipschitz. Sea $E \subset \Omega$ un boreliano en el que f y g coinciden. Entonces f es aproximadamente diferenciable en casi todo E y vale $\nabla_{ap} f(x) = \nabla g(x)$ para casi todo x en E.*

Con este resultado es posible probar que φ es aproximadamente diferenciable en casi todo el soporte de μ y entonces la definición de T dada por (4.5) para el caso convexo se corresponde con la dada por

$$T(x) = x - ((\nabla g)^{-1} \circ \nabla_{\mathrm{ap}} \varphi)(x) \tag{4.14}$$

para este caso.

Para probar que φ es aproximadamente diferenciable en casi todo el soporte de μ hay que dar un argumento más o menos cuidadoso que puede leerse en [33]. La idea consiste en ver que φ coincide en ciertos abiertos U_i con ciertas funciones φ_i que sí son Lipschitz, y en cada uno de esos abiertos usar el teorema 4.3.6. En cada abierto U_i existe $\nabla_{\mathrm{ap}}\varphi$ en casi todo punto y como los abiertos U_i cubren casi todo el soporte de μ, entonces existe $\nabla_{\mathrm{ap}}\varphi$ en un conjunto en el que μ está concentrada. Este último paso es el que requiere la hipótesis de que μ y ν sean mutuamente singulares.

Una vez probado que φ es aproximadamente diferenciable en un conjunto que concentra la masa de μ, podemos definir T en μ casi todo punto según (4.14) y proceder como en el caso convexo.

4.4. Caso euclídeo : $p = 1$

En este caso no es posible arreglar la demostración del caso estrictamente convexo como se hizo para el caso estrictamente cóncavo. Veamos dónde falla el argumento.

Consideremos un par (φ, ψ) de funciones c-conjugadas que maximice el problema dual. Ahora $g(x) = |x|$ es también una función convexa aunque no estrictamente convexa. Es posible probar, al igual que en el caso estrictamente convexo, que φ satisface una condición de Lipschitz — en este caso, con constante 1 —. Usando el teorema de Rademacher podemos conseguir S de medida μ nula tal que φ es diferenciable en $\Omega^\circ \backslash S$. Si (x_0, y_0) es tal que

$$\varphi(x_0) + \psi(y_0) = c(x_0, y_0) = g(x_0 - y_0) \ ,$$

podemos razonar como en el caso convexo y probar que

$$\nabla g(x_0 - y_0) = \nabla\varphi(x_0) \ . \tag{4.15}$$

Hasta este punto el argumento se pudo repetir sin problemas porque dependió sólo de la convexidad de g y no fue necesaria la convexidad estricta. La diferencia entre este caso y el caso $p > 1$ aparece ahora: no podemos decir que ∇g sea una función inyectiva porque g no es estrictamente convexa. De hecho, tenemos

$$\nabla g(x) = \frac{x}{|x|}$$

y entonces (4.15) se transforma en

$$\frac{x_0 - y_0}{|x_0 - y_0|} = \nabla\varphi(x_0) \ . \tag{4.16}$$

Notar que $\nabla\varphi(x_0)$ tiene norma 1 y que y_0 no queda unívocamente determinado por x_0, pues y_0 podría ser cualquier elemento de Ω de la forma

$$y_0 = x_0 - \lambda \cdot \nabla\varphi(x_0)$$

con λ positivo. Esto dice que para el caso $p = 1$ sólo podemos conocer la dirección en la que movería la masa una aplicación T minimizante para el problema de Monge. De hecho, para $p = 1$ no hay un minimizante único en el problema de Monge.

Ejemplo 4.4.1. Supongamos que tenemos una distribución de masa dada por n masas puntuales idénticas en los enteros $0, 1, \ldots, n-1$ y deseamos transportarlas a los enteros $1, 2, \ldots, n$ minimizando el costo dado por la distancia euclídea. Más precisamente, tomemos $X = \{0, 1, \ldots, n-1\}$, $Y = \{1, 2, \ldots, n\}$,

$$\mu = \frac{1}{n}\sum_{i=0}^{n-1}\delta_i \qquad y \qquad \nu = \frac{1}{n}\sum_{i=1}^{n}\delta_i \ .$$

Consideremos el costo dado por la distancia euclídea $c(x, y) = |x - y|$. Tenemos al menos dos funciones $S, T : X \to Y$ que transportan μ en ν con costo mínimo:

$$S(i) = i + 1 \quad \text{para} \ \ 0 \leq i < n \ \ ,$$

$$T(i) = \begin{cases} i & , \text{ si } \ 0 < i < n \\ n & , \text{ si } \ i = 0 \end{cases} \ .$$

En ambos casos el costo es n, que es óptimo. ▲

Ejemplo 4.4.2. Una versión continua del ejemplo anterior está dada por $X = [0, 1]$, $Y = [1, 2]$ y μ, ν las restricciones de la medida de Lebesgue en $\mathbb{R}$ a X e Y respectivamente. Tomando el costo $c(x, y) = |x - y|$, dos transportes óptimos están dados por $S(x) = x + 1$ y $T(x) = 2 - x$, ambos con costo 1. ▲

Los ejemplos anteriores muestran que si se espera obtener un teorema de unicidad de funciones de transporte — como los probados para los casos estrictamente convexo y estrictamente cóncavo —, es necesario imponer algún tipo de restricción adicional. Vale el siguiente resultado.

Proposición 4.4.3 (Principio de selección para el caso $p = 1$). *Sean μ y ν en $\mathcal{P}(\mathbb{R}^d)$ con momentos de primer orden finitos y notemos por $|x-y|$ a la distancia euclídea en $\mathbb{R}^d$. Sea $\mathcal{O}(\mu,\nu)$ el cojunto de minimizantes para el problema de Kantorovich*

$$\operatorname*{ínf}_{\gamma\in\mathcal{A}(\mu,\nu)} \left\{ \int_{X\times Y} |x-y|\, d\gamma(x,y) \right\}. \tag{4.17}$$

Asumamos, además, que μ es absolutamente continua con respecto a la medida de Lebesgue en $\mathbb{R}^d$. Entonces existe un único minimizante de la forma $\gamma = (I\times T)_{\#}\mu$ para el problema (4.17), si se le pide a γ la condición adicional de ser un minimizante del problema

$$\operatorname*{ínf}_{\gamma\in\mathcal{O}(\mu,\nu)} \left\{ \int_{X\times Y} g(|x-y|)\, d\gamma(x,y) \right\}$$

para toda función $g : \mathbb{R}_{\geq 0} \to \mathbb{R}$ estrictamente convexa y acotada por abajo.

Se puede leer una idea de la demostración de 4.4.3 en [37]. En [33] y en [3] se pueden leer demostraciones completas de este y otros resultados relacionados.

Parte II

Aplicaciones

Capítulo 5

Desigualdades

En este capítulo damos demostraciones para dos desigualdades utilizando técnicas de transporte óptimo. La técnica consiste en explotar la existencia de la aplicación de Brenier y aplicar la desigualdad aritmético–geométrica para probar una desigualdad de traza–determinante. Las desigualdades que probaremos serán la desigualdad isoperimétrica y la de Sobolev–Gagliardo–Nirenberg.

5.1. Desigualdad isoperimétrica

La desigualdad isoperimétrica clásica establece que el círculo es la figura de mayor área de entre todas las que tengan su mismo perímetro. Más precisamente, dados un círculo de radio R y una región plana E de perímetro $2\pi R$, se tiene $\text{Área}(E) \leq \pi R^2$. Equivalentemente, dada una región plana E, se verifica

$$\text{Área}(E) \leq \frac{1}{4\pi} \, \mathrm{P}^2(E) \; ,$$

donde $\mathrm{P}(E)$ es el perímetro de E.

Para poder generalizar la noción de perímetro de un conjunto — i.e. medida del borde a una dimensión menor que la del conjunto — a dimesiones mayores es necesario contar con una forma de medir subconjuntos de $\mathbb{R}^d$ de dimensión menor que d. La forma estándar de hacerlo es con medidas de Hausdorff. Dado un espacio métrico X y un subconjunto E de X, para cada $d \in \mathbb{R}_{\geq 0}$ definimos la **medida de Hausdorff s-dimensional** de E como

$$\mathcal{H}^s(E) = \lim_{\varepsilon \to 0^+} \inf_{\{U_i\}} \sum_i \left(\text{diám}(U_i)\right)^s \; ,$$

donde los ínfimos se toman sobre todos los cubrimientos $\{U_i\}$ de E tales que diám$(U_i) < \varepsilon$ para todo i. Puede leerse la teoría de medidas de Hausdorff en [15], [17] o [21]. Para enunciar la versión general de la desigualdad isoperimétrica bastará la siguiente definición: un subconjunto E de X se dice ***m*-rectificable** si existe una función Lipschitz de $\mathbb{R}^m$ en X que contenga a E en su imagen.

Proposición 5.1.1 (Desigualdad isoperimétrica). *Sea $E \subset \mathbb{R}^d$ un abierto acotado con frontera $(d-1)$-rectificable y sea B la bola unitaria de $\mathbb{R}^d$. Entonces se tiene*

$$\mathrm{P}(E) \geq d\mathcal{L}(B)^{\frac{1}{d}}\mathcal{L}(E)^{1-\frac{1}{d}} , \tag{5.1}$$

donde $\mathrm{P}(E) = \mathcal{H}^{d-1}(\partial E)$ y $\mathcal{L}$ es la medida de Lebesgue d-dimensional.

En [17] puede verse una demostración de un enunciado más general, que funciona incluso en caso de que la frontera de E no sea $(d-1)$-rectificable.

Vamos a dar una demostración de 5.1.1 usando transporte óptimo. En la demostración quedarán pendientes algunos de los detalles relacionados con la regularidad de los objetos intervinientes: los cambios de variable, la buena definición de integrales y la validez del teorema de la divergencia. Estos detalles pueden leerse en el capítulo 4 de [37] o en el capítulo 12 de [38]. Una versión del teorema de la divergencia para campos Lipschitz — y por lo tanto con divergencia existiendo sólo en casi todo punto —puede leerse al final del capítulo 4 en [17].

Sean E y B como en el enunciado de 5.1.1. Consideremos

$$\mu = \frac{1}{\mathcal{L}(E)}\mathcal{L}\big|_E \quad y \quad \mu = \frac{1}{\mathcal{L}(B)}\mathcal{L}\big|_B \ , \tag{5.2}$$

donde $\mathcal{L}$ es la medida de lebesgue d-dimensional. Tomemos un transporte $T : E \to B$ de μ a ν, óptimo para el costo cuadrático. Se verifica que T es esencialmente única y está dada por el gradiente de una función convexa φ. Por (1.2) y (5.2) tenemos

$$\frac{1}{\mathcal{L}(E)} = \frac{1}{\mathcal{L}(B)}|\det DT(x)| \tag{5.3}$$

para todo x en E. Por ser φ una función convexa, su hessiano es semidefinido positivo y por lo tanto $DT(x)$ es una matriz semidefinida positiva para todo x. Tenemos el siguiente lema.

Lema 5.1.2. *Sea $A \in \mathbb{R}^{n\times n}$ una matriz semidefinida positiva. Entonces vale*

$$(det(A))^{\frac{1}{n}} \leq \frac{1}{n} \cdot \mathrm{tr}(A) \ . \tag{5.4}$$

Demostración. Cambiando a una base en la que A sea triangular se puede ver que el determinante de A es el producto de sus autovalores y que la traza es la suma de estos. Se obtiene (5.4) aplicando la desigualdad aritmético–geométrica a estas expresiones del determinante y la traza. ∎

Aplicando el lema a $DT(x)$ obtenemos

$$|\det DT(x)|^{\frac{1}{d}} \leq \frac{1}{d}\, \nabla \cdot T(x) \tag{5.5}$$

para todo x en E. Por (5.3) y (5.5) resulta

$$\frac{1}{\mathcal{L}(E)^{\frac{1}{d}}} \leq \frac{1}{d\, \mathcal{L}(B)^{\frac{1}{d}}}\, \nabla \cdot T(x) \tag{5.6}$$

para todo x en E. Luego

$$\mathcal{L}(E)^{1-\frac{1}{d}} = \int_E \frac{1}{\mathcal{L}(E)^{\frac{1}{d}}}\, dx \leq \frac{1}{d\, \mathcal{L}(B)^{\frac{1}{d}}} \int_E \nabla \cdot T(x)\, dx = \frac{1}{d\, \mathcal{L}(B)^{\frac{1}{d}}} \int_{\partial E} \langle T(x), n(x) \rangle\, d\mathcal{H}^{d-1}(x)\ , \tag{5.7}$$

donde $n : \partial E \to S^{d-1}$ es el campo de vectores normales salientes y la última igualdad es consecuencia de aplicar el teorema de la divergencia. Como $T(x) \in B$ para todo x en E, tenemos $\langle T(x), n(x) \rangle \leq 1$. Se concluye

$$\mathcal{L}(E)^{1-\frac{1}{d}} \leq \frac{1}{d\, \mathcal{L}(B)^{\frac{1}{d}}}\, \mathrm{P}(E)$$

a partir de (5.7) y por lo tanto vale (5.1).

5.2. Desigualdad de Sobolev–Gagliardo–Nirenberg

En esta sección vamos a usar transporte óptimo para demostrar de la desigualdad de Sobolev–Gagliardo–Nirenberg, también conocida como desigualdad de Sobolev. Al igual que en la anterior demostración, las de las cuestiones de regularidad quedarán pendientes. Para una demostración completa se puede consultar el capítulo 6 de [37].

A continuación damos las definiciones indispensables para enunciar la desigualdad de Sobolev. Para un estudio más profundo de los espacios de Sobolev y sus aplicaciones a las ecuaciones diferenciales se puede consultar [16] o [12].

Vamos a notar $C_c^\infty(\mathbb{R}^d)$ al conjunto de las funciones C^∞ de soporte compacto definidas en $\mathbb{R}^d$ y $L^1_{\text{loc}}(\mathbb{R}^d)$ al de las localmente integrables en $\mathbb{R}^d$ — es decir, al de las funciones medibles que tienen integral finita sobre cada compacto de $\mathbb{R}^d$ —. Para un vector F de d funciones en $L^p(\mathbb{R}^d)$ definimos

$$\|F\|_p := \|F\|_{L^p} := \left(\sum_{i=1}^{d} \|F_i\|_p^p\right)^{\frac{1}{p}}.$$

Definición 5.2.1. Sea $1 \leq p \leq \infty$. Dada $f \in L^1_{\text{loc}}(\mathbb{R}^d)$ e $1 \leq i \leq d$, decimos que $g \in L^1_{\text{loc}}(\mathbb{R}^d)$ es la **derivada débil de $\boldsymbol{f}$ respecto de $\boldsymbol{x_i}$** si

$$\int_{\mathbb{R}^d} f\, \frac{\partial \varphi}{\partial x_i} = -\int_{\mathbb{R}^d} g\, \varphi \tag{5.8}$$

para toda φ en $C_c^\infty(\mathbb{R}^d)$.

Se puede ver que cada derivada débil de f, si existe, es única. Más precisamente, dos derivadas débiles de f respecto de la misma variable difieren sólo en un conjunto de medida cero. Cuando existen la derivada en el sentido usual y también la derivada débil respecto de la misma variable, ambas coinciden en casi todo punto. Usamos las mismas notaciones que para las derivadas en sentido usual: $\frac{\partial f}{\partial x_i}$ y ∇f.

Observación 5.2.2. La intuición detrás de la definición de derivadas débiles es que la validez de (5.8) permite las mismas manipulaciones formales que la integración por partes, aun cuando no existan las derivadas en sentido clásico. De algún modo, definimos las derivadas débiles pensando en las mínimas condiciones que permitan el formalismo de la integración por partes. ▲

Definición 5.2.3. El **espacio de Sobolev** $W^{1,p}(\mathbb{R}^d)$ es el subespacio de $L^p(\mathbb{R}^d)$ formado por las funciones de dicho espacio que tienen derivada débil respecto de x_i en $L^p(\mathbb{R}^d)$ para cada $1 \leq i \leq d$. Se verifica que $W^{1,p}(\mathbb{R}^d)$ es un espacio de Banach con la norma

$$\|f\|_{W^{1,p}} = \|f\|_p + \sum_{i=1}^{d} \left\|\frac{\partial f}{\partial x_i}\right\|_p. \tag{5.9}$$

Observación 5.2.4. La norma $\|\cdot\|_{W^{1,p}}$ es equivalente a la norma dada por

$$\left(\|f\|_p^p + \sum_{i=1}^{d} \left\|\frac{\partial f}{\partial x_i}\right\|_p^p\right)^{\frac{1}{p}} = \left(\|f\|_p^p + \|\nabla f\|_p^p\right)^{\frac{1}{p}}$$

por la equivalencia de normas en $\mathbb{R}^{d+1}$. ▲

Teorema 5.2.5 (Sobolev–Gagliardo–Nirenberg). *Sea* $1 \leq p < d$. *Entonces*

$$W^{1,p}(\mathbb{R}^d) \subset L^{p^*}(\mathbb{R}^d) \quad \text{donde } p^* \text{ está dado por } \frac{1}{p^*} = \frac{1}{p} - \frac{1}{d}\,. \tag{5.10}$$

Más aun, existe una constante $C = C(p,d)$ *que depende sólo de* p *y* d *tal que*

$$\|f\|_{L^{p^*}} \leq C\|\nabla f\|_{L^p} \tag{5.11}$$

para toda f *en* $W^{1,p}(\mathbb{R}^d)$.

Observación 5.2.6. El exponente $p^* = \frac{dp}{d-p}$ determinado por (5.10) se conoce como el **conjugado de Sobolev de p**. Notar que $p < p^* \leq p(p+1)$.

Se puede ver que p^* es el único exponente para el que puede haber una inclusión como (5.10). Para verlo, supongamos que existen $1 \leq q \leq \infty$ y $C > 0$ tales que vale

$$\|\varphi\|_{L^q} \leq C\|\nabla\varphi\|_{L^p} \tag{5.12}$$

para toda φ en $C_c^\infty(\mathbb{R}^d)$. Fijemos una φ no nula y observemos que $\|\nabla\varphi\|_{L^p} \neq 0$ por ser φ un elemento no nulo de $C_c^\infty(\mathbb{R}^d)$. Para cada $\lambda > 0$ consideremos $\varphi_\lambda(x) = \varphi(\lambda x)$. Reemplazando φ_λ en (5.12), se obtiene

$$\|\varphi\|_{L^q} \leq C\lambda^{(1+\frac{d}{q}-\frac{d}{p})}\|\nabla\varphi\|_{L^p} \tag{5.13}$$

para todo $\lambda > 0$. Haciendo tender λ a 0 y a $+\infty$ se deduce que necesariamente debe ser $1 + \frac{d}{q} - \frac{d}{p} = 0$ y por lo tanto $q = \frac{dp}{d-p} = p^*$. ▲

Observación 5.2.7. La inclusión

$$W^{1,p}(\mathbb{R}^d) \hookrightarrow L^{p^*}(\mathbb{R}^d)$$

se conoce como el **embedding de Sobolev** y su continuidad es una consecuencia de 5.2.5. En efecto, tenemos

$$\|f-g\|_{L^{p^*}} \leq C\|\nabla(f-g)\|_{L^p} = C\left(\sum_{i=1}^{d}\left\|\frac{\partial(f-g)}{\partial x_i}\right\|_p^p\right)^{\frac{1}{p}} \leq C\cdot A\|f-g\|_{W^{1,p}}\,,$$

donde la última desigualdad es consecuencia de la observación 5.2.4. ▲

Demostración de 5.2.5. Se puede ver que basta probar el resultado para $f \geq 0$ — ver problema 18 del capítulo 5 en [16] —. Dada $f \geq 0$ en $W^{1,p}(\mathbb{R}^d)$ podemos conseguir una sucesión $(f_n)_n$ de funciones no negativas en $C_c^\infty(\mathbb{R}^d)$ que converja a f en $W^{1,p}(\mathbb{R}^d)$ — ver teorema 9.2 en [12] —. En particular, la sucesión $(f_n)_n$ convergerá a f en $L^{p^*}(\mathbb{R}^d)$ y $\|\nabla f_n\|_p$ convergerá a $\|\nabla f\|_p$. Probando primero el resultado para funciones no negativas en $C_c^\infty(\mathbb{R}^d)$, tendremos

$$\|f_n\|_{L^{p^*}} \leq C\|\nabla f_n\|_{L^p} \tag{5.14}$$

para todo n. Por lo tanto $(f_n)_n$ será de Cauchy en $L^{p^*}(\mathbb{R}^d)$ y entonces convergente allí. Por unicidad del límite en $L^p(\mathbb{R}^d) \cap L^{p^*}(\mathbb{R}^d)$ se tiene que $(f_n)_n$ converge a f en $L^{p^*}(\mathbb{R}^d)$. Luego se deduce la validez de (5.11) tomando límite en (5.14).

Por lo anterior podemos suponer que f es no negativa y está en $C_c^\infty(\mathbb{R}^d)$. El resultado es claro si $\|f\|_{L^{p^*}} = 0$. Dada f tal que $\|f\|_{L^{p^*}} \neq 1$, siempre es posible dividir (5.11) por $\|f\|_{L^{p^*}}$ y reducirse el caso $\|f\|_{L^{p^*}} = 1$. Luego, asumiendo que $\|f\|_{L^{p^*}} = 1$, basta probar que

$$\|\nabla f\|_{L^p} = \left(\int |\nabla f(x)|^p \right)^{\frac{1}{p}} \geq K \tag{5.15}$$

para alguna constante K independiente de f. Notar que $|\nabla f(x)|$ denota la norma p de $\nabla f(x)$ punto a punto — i.e. como elemento de $\mathbb{R}^d$ —. Fijemos $g : \mathbb{R}^d \to \mathbb{R}$ estrictamente positiva, de clase C^∞ e integrando 1 en $\mathbb{R}^d$. Tomando g de este modo, y porque asumimos $\|f\|_{L^{p^*}} = 1$, podemos definir dos medidas de probabilidad en $\mathbb{R}^d$ según

$$\mu = f^{p^*}\mathcal{L} \quad y \quad \mu = g\,\mathcal{L} \ ,$$

donde $\mathcal{L}$ es la medida de Lebesgue d-dimensional. Tomemos como T la única función de transporte de μ a ν óptima para el costo cuadrático, dada por el gradiente de una función convexa. Por (1.2) tenemos

$$g(T(x)) = \frac{f^{p^*}(x)}{\det(DT(x))} \tag{5.16}$$

para todo x en $\mathbb{R}^d$ — recordar que estamos ignorando los detalles de regularidad —.

Cambiando variables tenemos

$$\int g^{1-\frac{1}{d}}(y) = \int g^{-\frac{1}{d}}(y)g(y) = \int (g \circ T)^{-\frac{1}{d}}(x)g(T(x))\det(DT(x)) =$$
$$= \int (g \circ T)^{-\frac{1}{d}}(x) f^{p^*}(x) = \int (\det(DT(x)))^{-\frac{1}{d}}(x)(f^{p^*}(x))^{1-\frac{1}{d}} \ ,$$

donde las dos últimas igualdades son consecuencia de (5.16). Podemos probar la misma desigualdad determinante–traza que probamos en (5.5) y obtener

$$\int g^{1-\frac{1}{d}}(y) \leq \frac{1}{d}\int (\nabla \cdot T(x))(f^{p^*}(x))^{1-\frac{1}{d}} = \frac{1}{d}\int (\nabla \cdot T(x))(f(x))^{1+\frac{p^*}{q}} \ , \tag{5.17}$$

donde q verifica $\frac{1}{p}+\frac{1}{q}=1$ y por lo tanto $p^*\big(1-\frac{1}{d}\big)=1+\frac{p^*}{q}$. Como f tiene soporte compacto, podemos usar el teorema de la divergencia en (5.17) para concluir

$$\begin{aligned}\int g^{1-\frac{1}{d}}(y) \leq \frac{1}{d}\int (\nabla\cdot T(x))(f(x))^{1+\frac{p^*}{q}} = -\frac{1}{d}\int T\cdot\nabla(f^{1+\frac{p^*}{q}})(x) = \\ = -\frac{1}{d}\Big(1+\frac{p^*}{q}\Big)\int f^{\frac{p^*}{q}}(x)(T\cdot\nabla f)(x)\,. \end{aligned}\tag{5.18}$$

Tomando módulo en última integral y aplicando la desigualdad de Hölder resulta

$$\begin{aligned}\int g^{1-\frac{1}{d}}(y) \leq \frac{1}{d}\Big(1+\frac{p^*}{q}\Big)\int f^{\frac{p^*}{q}}(x)|T(x)||\nabla f(x)| \leq \\ \leq \frac{1}{d}\Big(1+\frac{p^*}{q}\Big)\Big(\int f^{p^*}(x)|T(x)|^q\Big)^{\frac{1}{q}}\Big(\int |\nabla f(x)|^p\Big)^{\frac{1}{p}}\,. \end{aligned}\tag{5.19}$$

Usando (5.16) y cambiando variables se obtiene

$$\begin{aligned}\int g^{1-\frac{1}{d}}(y) \leq \frac{1}{d}\Big(1+\frac{p^*}{q}\Big)\Big(\int g(T(x))\det(DT(x))|T(x)|^q\Big)^{\frac{1}{q}}\Big(\int |\nabla f(x)|^p\Big)^{\frac{1}{p}} = \\ = \frac{1}{d}\Big(1+\frac{p^*}{q}\Big)\Big(\int g(y)|y|^q\Big)^{\frac{1}{q}}\Big(\int |\nabla f(x)|^p\Big)^{\frac{1}{p}}\,. \end{aligned}$$

Como g es arbitraria y q depende sólo de p, esto prueba (5.15). ■

En [37] se puede ver qué funciones son las que hacen de la desigualdad una igualdad. La prueba del resultado es similar a la dada aquí pero teniendo el cuidado de obtener los casos igualdad en cada desigualdad que aprece. Una expresión explícita de las constantes óptimas en términos de la función Γ pueden verse en [4] o en [36]. En [18] se puede ver una prueba de la equivalencia entre la desigualdad isoperimétrica y el caso $p=1$ de la desigualdad de Sobolev. La equivalencia debe ser entendida en el sentido de que si una desigualdad vale con constante K entonces la otra también.

Proposición 5.2.8. *Sea $K\in\mathbb{R}$. Entonces son equivalentes:*

1. *para todos los conjuntos medibles de medida finita vale $P(E)\geq K|E|^{\frac{d-1}{d}}$;*

2. *para toda f en $W^{1,1}(\mathbb{R}^d)$ vale $\|\nabla f\|_{L^1}\geq K\|f\|_{L^{\frac{d}{d-1}}}$*

Notar que $1^*=\frac{d}{d-1}$ y que notamos $P(E)$ al perímetro de E. Así, es posible obtener la constante óptima para el caso $p=1$ tomando $K_0=\max_K K$, donde el máximo se toma sobre todos los K que verifican *1.* Luego, con la notación de 5.2.5 y notando que K_0 es la constante de la desigualdad isoperimétrica, tenemos

$$C = C(1,d) = \frac{1}{K_0} = \frac{1}{d|B|^{\frac{1}{d}}} = \frac{(\Gamma(\frac{n}{2}+1))^{\frac{1}{d}}}{d \cdot \sqrt{\pi}} \ .$$

Capítulo 6

Problemas de Transporte

En este capítulo estudiaremos dos modelos continuos de transporte. Por continuos debe entenderse no–discretos. Empezaremos por un modelo simple debido a Beckmann [5] que no incorpora efectos de congestión. Seguiremos con otro modelo más sofisticado debido a Carlier, Jimenez y Santambrogio [14] que sí incorpora efectos de congestión.

6.1. El problema de Beckmann

En [5], Beckmann propuso lo que él llamó un **modelo continuo de transporte**, que describimos a continuación siguiendo a [13] y [33]. Consideremos un *área urbana* dada por un abierto Ω, que supondremos acotado, conexo y con borde suave. Consideremos dos medidas borelianas μ y ν en $\overline{\Omega}$ no negativas y finitas representando, respectivamente, la distribución de residentes y la distribución de producción en el área urabana Ω. La medida $\mu - \nu$ representa el exceso de demanda. Para satisfacer sus necesidades de consumo, los residentes se moverán a través de la región Ω siguiendo trayectorias definidas por un campo $\mathbf{w} : \Omega \to \mathbb{R}^2$. Para cada subregión $K \subset \Omega$ definimos una noción de equilibrio estableciendo que el flujo neto saliente de residentes igual al exceso de demanda:

$$\int_{\partial K} \mathbf{w} \cdot \mathbf{n} \, dS = (\mu - \nu)(K) \ . \tag{6.1}$$

En particular, si $K = \Omega$, (6.1) dice que

$$\int_{\partial \Omega} \mathbf{w} \cdot \mathbf{n} \, dS = (\mu - \nu)(\Omega) \ . \tag{6.2}$$

La interpretación es clara: si el exceso de demanda neto en el área urbana Ω es positivo, entonces la demanda de los residentes sólo puede ser satisfecha si una cantidad positiva de ellos cruza la frontera de Ω; si el exceso de demanda neto en el área urbana Ω es negativo, el equilibrio sólo se puede lograr con un flujo entrante de consumidores provenientes desde fuera de la región Ω.

Observación 6.1.1. Dado que la condición de equilibrio (6.1) vale para toda subregión K, bajo hipótesis de regularidad razonables podemos dar una formulación diferencial de (6.1). Si μ y ν son absolutamente continuas con densidades f y g respectivamente, el teorema de la divergencia permite reformular la condición (6.1) como

$$\nabla \cdot \mathbf{w} = f - g \quad \text{en} \;\; \Omega\,. \tag{6.3}$$

Notar que (6.3) es una forma de ecuación de continuidad, común en problemas de física. En este contexto, $f - g$ es una medida del exceso local de demanda. ▲

En la formulación dada por Beckmann en [5], se define una función en el borde de Ω representando la densidad neta de exportación punto a punto y se le pide una condición de compatibilidad similar a (6.2). Para evitar estos cuidados nos vamos a restringir al caso en que $\mu(\Omega) = \nu(\Omega) > 0$. Así, podremos suponer que μ y ν son medidas de probabilidad, normalizando primero si fuera necesario. Con esta restricción tiene sentido pensar que el área urbana Ω se autoabastece. Por lo tanto, podremos pedir la condición

$$\mathbf{w} \cdot \mathbf{n} = 0 \tag{6.4}$$

en el borde de Ω para forzar al campo $\mathbf{w}$ a no atravesarlo, lo que se interpreta como forzar a los residentes a abastecerse en Ω.

El objetivo del problema será, logrando una situación de equilibrio, minimizar el costo de transporte de los residentes. Vamos a suponer que el costo de transporte es uniforme — la posibilidad de que no lo sea sí es tenida en cuenta en [5] —. Podemos formular el problema así:

$$\min_{\mathbf{w}} \|\mathbf{w}\|_1 \tag{6.5}$$

donde el mínimo se toma sobre sobre los campos $\mathbf{w}$ satisfaciendo (6.1) y (6.4).

El inconveniente que tiene esta formulación es que podría no existir un minimizante en $L^1(\mathbb{R}^2)$, dado que $L^1(\mathbb{R}^2)$ no es un espacio reflexivo. Para solucionar esta dificultad vamos a dar una formulación débil del problema considerando medidas vectoriales en lugar de campos, a la vez que lo generalizamos a dominios de $\mathbb{R}^d$.

Decimos que una medida vectorial w en $\mathcal{M}^d(\overline{\Omega})$ satisface la condición de equilibrio si vale

$$\int_{\overline{\Omega}} \nabla\varphi \cdot w = \int_{\overline{\Omega}} \varphi \, d(\nu - \mu) \tag{6.6}$$

para toda φ en $C^1(\overline{\Omega})$. Notar que en el caso de considerar campos vectoriales, la condición (6.6) se obtiene integrando por partes en (6.3). Reformulamos el problema de Beckmann como

$$\text{mín}\left\{\|w\| : w \in \mathcal{M}^d(\overline{\Omega}) \text{ tales que } \int_{\overline{\Omega}} \nabla\varphi \cdot w = \int_{\overline{\Omega}} \varphi \, d(\nu - \mu) \ \ \forall \varphi \in C^1(\overline{\Omega})\right\}. \tag{6.7}$$

Siguiendo las definiciones dadas en B.2, (6.7) dice que w debe ser una medida vectorial con divergencia $\mu - \nu$. Luego, podemos formular el problema de Beckmann como

$$\text{mín}\left\{\|w\| = |w|(\Omega) : w \in \mathcal{M}^d_{\text{div}}(\Omega) \text{ tales que } \nabla \cdot w = \mu - \nu\right\}.$$

Proposición 6.1.2. *Sea Ω un dominio convexo y acotado de $\mathbb{R}^d$ y μ, ν dos medidas de probabilidad en Ω. Entonces el problema*

$$\text{ínf}\left\{\|w\| = |w|(\Omega) : w \in \mathcal{M}^d_{\text{div}}(\Omega) \textit{ tales que } \nabla \cdot w = \mu - \nu\right\} \tag{6.8}$$

admite un minimizante. Más aun, su valor óptimo coincide con

$$\min_{\gamma \in \mathcal{A}(\mu,\nu)} \int_{X \times X} |x - y| \, d\gamma(x, y) , \tag{6.9}$$

que es el valor óptimo para un problema de Kantorovich asociado a (6.8) con costo dado por la distancia euclídea.

Demostración. Veamos primero que (6.9) $\leq$ (6.8). Sea $w \in \mathcal{M}^d_{\text{div}}(\Omega)$ tal que $\nabla \cdot w = \mu - \nu$. Para cada φ en $\text{Lip}_1(\Omega) \cap C^1(\Omega)$ tenemos

$$\|w\| = |w|(\Omega) = \int_\Omega 1 \, d|w| \geq \int_\Omega (-\nabla\varphi) \cdot dw = \int_\Omega \varphi \, d(\mu - \nu) . \tag{6.10}$$

Por el teorema 2.5.4 es posible tomar una sucesión $(\varphi_n)_n$ en $\text{Lip}_1(\Omega) \cap C^1(\Omega)$ convergiendo uniformemente a un potencial de Kantorovich φ para el problema (6.9). Aplicando (6.10) a cada φ_n y tomando límite en n resulta $\|w\| \geq$ (6.9). Tomando ínfimo en w resulta (6.9) $\leq$ (6.8).

Para probar la otra desigualdad, tomemos un plan óptimo γ para el problema de

Kantorovich (6.9) y consideremos $\psi_\gamma \in C_0(\Omega, \mathbb{R}^d)^*$ dada por

$$\psi_\gamma(\phi) = \int_{\Omega\times\Omega} \Big(\int_0^1 \phi(\omega_{x,y}(t)) \cdot \omega'_{x,y}(t)\, dt \Big)\, d\gamma(x,y) \tag{6.11}$$

para toda ϕ en $C_0(\Omega, \mathbb{R}^d)$, donde $\omega_{x,y}$ es la parametrización estándar del segmento $[x, y]$. Notar que es necesaria la convexidad de Ω para la buena definición de ψ_γ. Por el teorema de Riesz–Markov existe una única medida vectorial $w_\gamma \in \mathcal{M}^d(\Omega)$ tal que

$$\psi_\gamma(\phi) = \int_\Omega \phi \cdot dw_\gamma \tag{6.12}$$

para toda ϕ en $C_0(\Omega, \mathbb{R}^d)$. En particular, para cada ϕ en $C_0(\Omega, \mathbb{R})$ vale

$$\begin{aligned}\int_\Omega \nabla\phi \cdot dw_\gamma = \psi_\gamma(\nabla\phi) = \int_{\Omega\times\Omega} \Big(\int_0^1 \nabla\phi(\omega_{x,y}(t)) \cdot \omega'_{x,y}(t)\, dt \Big)\, d\gamma(x,y) = \\ = \int_{\Omega\times\Omega} \Big(\phi(\omega_{x,y}(t)) \Big)\Big|_{t=0}^{t=1} d\gamma(x,y) = \int_{\Omega\times\Omega} \phi(y) - \phi(x)\, d\gamma(x,y) = -\int_\Omega \phi\, d(\mu - \nu)\ .\end{aligned}$$

Esto dice que $w_\gamma \in \mathcal{M}^d_{\mathrm{div}}(\Omega)$ y $\nabla \cdot w_\gamma = \mu - \nu$. Basta ver que $\|w_\gamma\| \leq$ (6.9).

Definamos $\eta_\gamma \in C_0(\Omega, \mathbb{R})^*$ por

$$\eta_\gamma(\phi) = \int_{\Omega\times\Omega} \Big(\int_0^1 \phi(\omega_{x,y}(t)) |\omega'_{x,y}(t)|\, dt \Big)\, d\gamma(x,y) \tag{6.13}$$

para toda ϕ en $C_0(\Omega, \mathbb{R})$. Por el teorema de Riesz–Markov existe una única medida positiva σ_γ en $\mathcal{M}(\Omega)$ tal que

$$\eta_\gamma(\phi) = \int_\Omega \phi\, d\sigma_\gamma$$

para toda ϕ en $C_0(\Omega, \mathbb{R})$. Tomemos un potencial de Kantorovich $\varphi \in \mathrm{Lip}_1(\Omega)$ para (6.9) y veamos que $w_\gamma = -\nabla\varphi\, \sigma_\gamma$. Se puede probar que para $x \neq y$ en $\mathrm{sop}(\gamma)$ y t en $(0, 1)$ vale

$$\omega'_{x,y}(t) = -(x-y) = -|x-y| \frac{x-y}{|x-y|} = -|x-y| \nabla\varphi(\omega_{x,y}(t))\ . \tag{6.14}$$

Para una demostración de este hecho ver el lema 3.6 de [33]. Notar que cuando $x = y$ los extremos de (6.14) coinciden y entonces

$$\omega'_{x,y}(t) = -|\omega'_{x,y}(t)| \nabla\varphi(\omega_{x,y}(t)) \tag{6.15}$$

para todos los x, y en $\mathrm{sop}(\gamma)$ y todo t en $(0, 1)$.

Dada ϕ en $C_0(\Omega, \mathbb{R}^d)$, por (6.15) tenemos

$$\begin{aligned}\psi_\gamma(\phi) &= \int_{\Omega\times\Omega} \Big(\int_0^1 \phi(\omega_{x,y}(t)) \cdot \omega'_{x,y}(t)\, dt \Big)\, d\gamma(x,y) \\ &= \int_{\Omega\times\Omega} -\Big(\int_0^1 \phi(\omega_{x,y}(t)) \cdot |\omega'_{x,y}(t)| \nabla\varphi(\omega_{x,y}(t))\, dt \Big)\, d\gamma(x,y) \\ &= \int_{\Omega\times\Omega} \Big(\int_0^1 -\phi(\omega_{x,y}(t)) \cdot \nabla\varphi(\omega_{x,y}(t)) |\omega'_{x,y}(t)|\, dt \Big)\, d\gamma(x,y) \\ &= \eta(-\phi \cdot \nabla\varphi)\ .\end{aligned} \tag{6.16}$$

Luego

$$\int_\Omega \phi \cdot dw_\gamma = \psi_\gamma(\phi) = \eta(-\phi\cdot\nabla\varphi) = \int_\Omega -\phi\cdot\nabla\varphi\, d\sigma_\gamma = \int_\Omega \phi \cdot d(\nabla\varphi\, \sigma_\gamma) \tag{6.17}$$

para toda ϕ en $C_0(\Omega, \mathbb{R}^d)$ y por lo tanto $w_\gamma = -\nabla\varphi\, \sigma_\gamma$ como consecuencia de la unicidad en el teorema de Riesz–Markov.

Tenemos

$$\|w_\gamma\| = |w_\gamma|(\Omega) = \int_\Omega d|w_\gamma| = \int_\Omega d|\nabla\varphi\, \sigma_\gamma| \leq \int_\Omega |\nabla\varphi|\, d\sigma_\gamma\ ,$$

y como $\varphi \in \mathrm{Lip}_1(\Omega)$, entonces

$$\begin{aligned}\|w_\gamma\| \leq \int_\Omega |\nabla\varphi|\, d\sigma_\gamma \leq \int_\Omega d\sigma_\gamma = \eta_\gamma(1) \\ = \int_{\Omega\times\Omega} \Big(\int_0^1 |\omega'_{x,y}(t)|\, dt \Big)\, d\gamma(x,y) = \int_{\Omega\times\Omega} |x-y|\, d\gamma(x,y)\ .\end{aligned}$$

Esto prueba que $\|w_\gamma\| \leq$ (6.9). ■

Puede demostrarse que todo minimizante del problema de Beckmann es de la forma w_γ para algún plan de transporte óptimo γ del problema de Kantorovich asociado. También puede probarse que w_γ no depende del plan γ elegido cuando μ es absolutamente continua. Como consecuencia de estos resultados, el problema (6.8) admite un único minimizante cuando μ es absolutamente continua. Las demostraciones de estos resultados pueden verse en [33].

6.2. Un modelo de transporte con congestión

Una característica que poseen el modelo de Beckmann y el problema de Monge es que no tienen en cuenta efectos de congestión: las buenas soluciones son mejores que las malas porque minimizan un costo que depende sólo de la longitud o costo de los caminos. En un modelo de transporte, esta característica se vuelve un defecto dado que en general es poco realista suponer despreciables los efectos de congestión. Formularemos a continuación un modelo de transporte debido a Carlier, Jimenez y Santambrogio que sí incorpora efectos de congestión. Los detalles y las demostraciones omitidas pueden consultarse en [14].

Consideremos un dominio compacto $\Omega \subset \mathbb{R}^d$ que sea la clausura de un abierto, acotado, conexo y convexo. Supongamos dadas dos medidas borelianas de probabilidad μ_0 y μ_1 en Ω que cumplirán, respectivamente, los roles de las medidas μ y ν del problema de Beckmann. En lo que sigue, Ω, μ_0 y μ_1 permanecerán fijos.

Consideremos el conjunto $\mathcal{C}$ de las curvas absolutamente continuas de $C([0,1],\Omega)$. Podemos pensar que $\mathcal{C}$ es la realización de $W^{1,1}([0,1],\Omega)$ en $C([0,1],\Omega)$. De este modo, $\mathcal{C}$ es un espacio de Banach y por lo tanto tiene una topología inducida por la norma, lo que permite considerar medidas borelianas en $\mathcal{C}$. Para cada par de elementos x, y en Ω notamos

$$\mathcal{C}_{x,y} = \{ \ \sigma \in \mathcal{C} \ : \ \sigma(0) = x \ , \ \sigma(1) = y \ \} \ .$$

Para φ continua en Ω y σ en $\mathcal{C}$ definimos

$$L_\varphi(\sigma) = \int_0^1 \varphi(\sigma(t))|\dot{\sigma}(t)| \, dt \ .$$

Para cada $t \in [0,1]$ notamos e_t al morfismo de evaluación en t para elementos de $\mathcal{C}$.

Consideremos el conjunto de medidas de probabilidad en $\mathcal{C}$ dado por

$$\mathcal{Q}(\mu_0, \mu_1) = \{ \ Q \in \mathcal{P}(\mathcal{C}) \ : \ e_{0\#}Q = \mu_0 \ , \ e_{1\#}Q = \mu_1 \} \ .$$

A cada elemento Q de $\mathcal{Q}(\mu_0, \mu_1)$ lo llamamos un **patrón de transporte** de μ_0 a ν_0.

Por el teorema de Riesz–Markov cada medida $Q \in \mathcal{Q}(\mu_0, \mu_1)$ induce una única medida $i_Q \in \mathcal{M}(\Omega)$ que llamaremos **intensidad de tráfico** y está dada por

$$\int_\Omega \varphi \, di_Q = \int_{\mathcal{C}} L_\varphi(\sigma) \, dQ(\sigma) = \int_{\mathcal{C}} \Big(\int_0^1 \varphi(\sigma(t))|\dot{\sigma}(t)| \, dt \Big) \, dQ(\sigma) \tag{6.18}$$

para toda φ en $C(\Omega, \mathbb{R})$. Para cada $E \subset \Omega$ boreliano, $i_Q(E)$ es una medida de la cantidad de tráfico que pasa por E. Más precisamente, para $E \subset \Omega$ boreliano tenemos

$$i_Q(E) = \int_\Omega \chi_E \, di_Q = \int_{\mathcal{C}} \Big(\int_0^1 \chi_E(\sigma(t))|\dot\sigma(t)| \, dt \Big) \, dQ(\sigma) \tag{6.19}$$

$$= \int_{\mathcal{C}} \Big(\int_{\sigma^{-1}(E)} |\dot\sigma(t)| \, dt \Big) \, dQ(\sigma) = \int_{\mathcal{C}} \mathcal{H}^1(E \cap \mathrm{Im}(\sigma)) \, dQ(\sigma) \; . \tag{6.20}$$

Notar que i_Q es positiva pero podría no ser finita. Puede probarse que si la medida μ_0 es absolutamente continua respecto de la medida de Lebesgue d-dimensional $\mathcal{L}^d$ entonces i_Q también lo es. Cometeremos el abuso de notar i_Q tanto a la medida i_Q como a su densidad respecto de $\mathcal{L}^d$.

Una vez cuantificada la intensidad de tráfico podemos introducir el efecto de congestión. La idea es asignar a cada curva σ de $\mathcal{C}_{x,y}$ un costo que dependa crecientemente del tráfico existente en las zonas que atraviesa. Para esto fijemos una función $g : \Omega \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ independiente de Q y tal que $g(x, \cdot)$ sea no decreciente para cada x en Ω. Definamos $\xi_Q : \Omega \to \mathbb{R}_{\geq 0}$ por

$$\xi_Q(x) = \begin{cases} g(x, i_Q(x)) & , \text{ si } \; i_Q \ll \mathcal{L}^d \\ +\infty & , \text{ si no} \end{cases} \tag{6.21}$$

para todo x en Ω. La idea de ξ_Q es modelar el impacto que tiene en el punto x una cierta intensidad de tráfico $i_Q(x)$. Ahora podemos definir una métrica c_{ξ_Q} en Ω dada por

$$c_{\xi_Q}(x, y) = \inf_{\sigma \in \mathcal{C}_{x,y}} L_{\xi_Q}(\sigma) \tag{6.22}$$

para cada par de puntos x, y en Ω. Cualquier curva σ en $\mathcal{C}_{x,y}$ que verifique $L_{\xi_Q}(\sigma) = c_{\xi_Q}(x, y)$ será una **geodésica** para c_{ξ_Q}.

Por otra parte, a cada Q en $\mathcal{Q}(\mu_0, \mu_1)$ le asociamos la medida $\gamma_Q = (e_0 \times e_1)_\# Q \in \mathcal{P}(\Omega \times \Omega)$. Si E y F son borelianos de Ω entonces

$$\gamma_Q(E \times F) = Q(\{ \, \sigma \in \mathcal{C} \; : \; \sigma(0) \in E, \; \sigma(1) \in F \, \}) \; .$$

Notar que $Q \in \mathcal{Q}(\mu_0, \mu_1)$ implica que γ_Q es un elemento de $\mathcal{A}(\mu_0, \mu_1)$. Nos va a interesar considerar patrones de transporte Q tales que γ_Q sea un minimizante del problema

$$\inf_{\gamma \in \mathcal{A}(\mu_0, \mu_1)} \int_{\Omega \times \Omega} c_{\xi_Q}(x, y) \, d\gamma(x, y) \; . \tag{6.23}$$

Además de encontrar un patrón de transporte Q tal que γ_Q minimice (6.23), queremos que esto suceda en una situación de equilibrio. Pensemos en un problema de transporte en el que agentes distribuidos según μ_0 deben trasladarse para conseguir una distribución μ_1. Queremos que la solución óptima de nuestro modelo nos diga cuál sería una asignación de rutas a cada uno de los agentes de modo tal que se minimizara el costo total de transporte y de forma estable. La estabilidad la vamos a entender en el sentido de equilibrios de Nash: toda ruta de x a y en el soporte de Q debe ser la mejor posible dada la congestión generada por Q. Esto puede formalizarse exigiéndole a Q la siguiente condición: para todo par de elementos x, y en Ω, si $\sigma \in \mathcal{C}_{x,y}$ está en el soporte de Q entonces $L_{\xi_Q}(\sigma) = c_{\xi_Q}(x, y)$. En otras palabras, estamos pidiendo que Q esté concentrada sobre geodésicas de c_{ξ_Q}.

Con estas dos condiciones formulamos la siguiente definición.

Definición 6.2.1. Un **patrón de transporte de equilibrio** es una medida $Q \in \mathcal{Q}(\mu_0, \mu_1)$ tal que

$$Q(\{ \ \sigma \in \mathcal{C} \ : \ L_{\xi_Q}(\sigma) = c_{\xi_Q}(\sigma(0), \sigma(1)) \ \}) = 1$$

y tal que γ_Q es un minimizante del problema

$$\inf_{\gamma \in \mathcal{A}(\mu_0, \mu_1)} \int_{\Omega \times \Omega} c_{\xi_Q}(x, y) \, d\gamma(x, y) \ .$$

En [14] se prueba que bajo ciertas hipótesis siempre existe un patrón de transporte de equilibrio. Entre otras hipótesis, se necesita poder garantizar que existe al menos una Q en $\mathcal{Q}(\mu_0, \mu_1)$ con i_Q satisfaciendo cierta condición de integrabilidad. La existencia de una tal Q puede probarse si se pide que μ_0 y μ_1 estén en L^q para cierto q. Por ejemplo, dado que Ω es acotado, basta que μ_0 y μ_1 estén en L^∞.

Capítulo 7

Metrización

En este capítulo estudiamos las distancias de Wasserstein, que permiten metrizar el espacio de medidas borelianas en un espacio métrico. Estas distancias se construyen usando el costo de transporte entre las medidas en cuestión cuando se toma como función de costo una potencia de la métrica del espacio métrico ambiente. En la última parte estudiamos el caso finito y una versión modificada de las distancias de Wasserstein útil en las aplicaciones llamada earth mover's distance.

7.1. Distancias de Wasserstein

Una aplicación del transporte óptimo es dar una noción de distancia entre dos medidas de probabilidad definidas en un mismo espacio métrico (X, d). Recordar que $\mathcal{P}(X)$ denota el conjunto de medidas borelianas de probabilidad en X. Para μ, ν en $\mathcal{P}(X)$ y $p \geq 0$ tenemos por

$$\mathcal{T}_{d^p}(\mu, \nu) = \inf_{\gamma \in \mathcal{A}(\mu,\nu)} \int_{X \times X} d^p(x, y)\, d\gamma(x, y) \tag{7.1}$$

y definimos la notación abreviada $\mathcal{T}_p(\mu, \nu) = \mathcal{T}_{d^p}(\mu, \nu)$.

Notemos $\mathcal{P}_p(X)$ al conjunto de medidas borelianas de probabilidad en (X, d) que tienen momentos de orden p finitos. Se tiene el siguiente teorema cuya demostración puede verse en [37].

Teorema 7.1.1. *Sea (X, d) un espacio métrico polaco. Entonces:*

- *Para cada $p \in [1, +\infty)$ tenemos una métrica en $\mathcal{P}_p(X)$ dada por $W_p = \mathcal{T}_p^{1/p}$.*

- *Para cada $p \in [0, 1)$ tenemos una métrica en $\mathcal{P}_p(X)$ dada por $W_p = \mathcal{T}_p$.*

Llamamos **distancia de Wasserstein de exponente p** a W_p. En el caso $p = 2$ la llamamos **distancia de Wasserstein cuadrática** y en el caso $p = 1$ la llamamos **distancia de Kantorovich–Rubinstein**.

Recordar que cuando (X, d) es un espacio métrico se define en $\mathcal{P}(X)$ la noción de convergencia débil: $(\mu_n)_n$ converge débilmente a μ si para toda $f \in C_b(X)$ se tiene

$$\int_X f\, d\mu_n \xrightarrow[n]{} \int_X f\, d\mu. \tag{7.2}$$

La topología en $\mathcal{P}(X)$ resultante de esta noción de convergencia se conoce como topología débil y tiene la propiedad de ser Hausdorff y separable cuando (X, d) es polaco. Tiene sentido entonces preguntar si esta topología es metrizable. La respuesta es que sí y en el apéndice B.1 puede verse una forma de hacer esto vía la métrica de Prokhorov d_P. Veamos a continuación que esto también es posible con distancias de Wasserstein.

Definición 7.1.2. Dado un espacio métrico (X, d), una familia de medidas de probabilidad $\mathcal{F} \subset \mathcal{P}_p(X)$ se dice ***p*-uniformemente integrable** si para todo $x_0 \in X$ y todo $\varepsilon > 0$ existe $R > 0$ tal que

$$\sup_{\mu \in \mathcal{F}} \int_{X \setminus B_R(x_0)} d^p(x, x_0)\, d\mu(x) < \varepsilon$$

o, equivalentemente, si se verifica

$$\lim_{R \to +\infty} \sup_{\mu \in \mathcal{F}} \int_{X \setminus B_R(x_0)} d^p(x, x_0)\, d\mu(x) = 0$$

para todo $x_0 \in X$.

Observación 7.1.3. Para probar la p-integrabilidad uniforme de una familia de medidas $\mathcal{F} \subset \mathcal{P}_p(X)$ basta verificar cualquiera de las condiciones para un único x_0 en X. Notar también que cuando X es acotado la definición se verifica trivialmente para cualquier familia de medidas tomando R suficientemente grande. ▲

Definición 7.1.4. Dado un espacio métrico (X, d) y una función $\varphi : X \to \mathbb{R}$, decimos que φ satisface una **condición de crecimiento de orden p** si existen $x_0 \in X$ y $C \in \mathbb{R}$ tales que

$$|\varphi(x)| \leq C[d^p(x, x_0) + 1]$$

para todo $x \in X$.

Observación 7.1.5. Basta verificar la condición para un único x_0 en X. Si φ es acotada la definición se verifica trivialmente tomando $C = \|\varphi\|_\infty$. Notar que cuando X es acotado una condición de crecimiento de cualquier orden para φ es equivalente a que φ esté acotada. ▲

La demostración del siguiente teorema puede verse en [37].

Teorema 7.1.6. *Sea* $p \in (0, +\infty)$, $(\mu_k)_k$ *una sucesión de medidas de probabilidad en* $\mathcal{P}_p(X)$ *y* $\mu \in \mathcal{P}(X)$. *Entonces son equivalentes:*

1. $W_p(\mu_k, \mu) \xrightarrow[k \to +\infty]{} 0$

2. $(\mu_k)_k$ *es* p*-uniformemente integrable y converge débilmente a* μ

3. $(\mu_k)_k$ *converge débilmente a* μ *y convergen los momentos de orden* p

$$\int_X d^p(x, x_0)\, d\mu_k(x) \xrightarrow[k \to +\infty]{} \int_X d^p(x, x_0)\, d\mu(x)$$

para cada $x_0 \in X$

4. *para toda función* φ *en* X *que satisface una condición de crecimiento de orden* p *vale*

$$\int_X \varphi\, d\mu_n \xrightarrow[k \to +\infty]{} \int_X \varphi\, d\mu$$

Dos métricas equivalentes d_1 y d_2 en X definen los mismos abiertos y, por lo tanto, las mismas funciones continuas y las mismas medidas borelianas. Esto dice que tanto $\mathcal{P}(X)$ como $C_b(X)$ dependen sólo de la topología del espacio métrico (X, d) y no de la métrica d. Luego, el espacio topológico resultante de darle a $\mathcal{P}(X)$ la topología débil depende sólo de la topología de X y no de una métrica en particular.

Una consecuencia del párrafo anterior es que para estudiar la metrizabilidad de la topología débil en $\mathcal{P}(X)$ es posible cambiar la métrica de (X, d) por cualquier otra métrica equivalente a d que resulte conveniente. En particular, se puede cambiar d por la métrica equivalente y acotada $\frac{d}{d+1}$. Combinando esta observación con el hecho de que tanto las condiciones de crecimiento como la integrabilidad uniforme se verifican trivialmente para espacios métricos acotados, resulta el siguiente corolario a partir del teorema 7.1.6.

Corolario 7.1.7. *Sea* (X, d) *un espacio métrico y sea la métrica* $\widetilde{d} = \frac{d}{d+1}$ *en* X, *equivalente a la métrica* d. *Entonces para todo* $p \in (0, +\infty)$ *la distancia* W_p *calculada a partir de* $\widetilde{d}$ *metriza la topología débil en* $\mathcal{P}(X)$.

7.2. El caso finito

En muchas aplicaciones computacionales son útiles las distancias de Wasserstein como una forma de medir la distancia entre objetos discretos. Entre los ejemplos clásicos están medir distancias entre dos histogramas o entre imágenes. Los histogramas se piensan como medidas no negativas sobre conjuntos discretos unidimensionales $[\![1, n]\!]$. Para el caso de las imágenes, típicamente como medidas no negativas sobre conjuntos discretos bidimensionales $[\![1, m]\!] \times [\![1, n]\!]$ tomando valores en $[\![0, 255]\!]$.

Sea (X, d) un espacio métrico. Llamamos **signatura** a todo conjunto de pares de la forma

$$\mathcal{S} = \{(x_1, \mu_1), (x_2, \mu_2), \ldots, (x_m, \mu_m)\},$$

donde $x_i \in X$ y $\mu_i \in \mathbb{R}_{\geq 0}$ para todo $i = 1, \ldots, m$. El conjunto de **clusters** de $\mathcal{S}$ es el conjunto $\{x_1, x_2, \ldots, x_m\}$ y, en él, tenemos una medida dada por los pesos $\mu = (\mu_i)_i$.

Nos interesa definir una noción de distancia entre signaturas que guarde alguna relación con la métrica d. Pensemos, por ejemplo, que la signatura $\mathcal{S}$ puede representar una imagen en escala de grises donde cada punto x_i es un píxel de la imagen y μ_i es el nivel de blanco de ese píxel. Para definir esta noción de distancia entre signaturas vamos a considerar el costo de transportar las medidas que definen sobre sus respectivos conjuntos de clusters.

En este punto aparece un problema y es que las medidas pueden no ser de probabilidad. O peor aun, pueden tener masas totales distintas. Para solucionar esta dificultad vamos a considerar un problema de transporte modificado que tendrá la forma de un problema de flujo máximo.

Sean X_1, X_2 subconjuntos finitos de X y

$$\mathcal{S}_1 = \{(x_1^1, \mu_1^1), (x_2^1, \mu_2^1), \ldots, (x_m^1, \mu_m^1)\} \;\text{ y }\; \mathcal{S}_2 = \{(x_1^2, \mu_1^2), (x_2^2, \mu_2^2), \ldots, (x_n^2, \mu_n^2)\}$$

dos signaturas asociadas a ellos. Notemos $d_{ij} = d(x_i^1, x_j^2)$ para $i = 1, \ldots, m$ y $j = 1, \ldots, n$.

Para el problema de transportar $\mathcal{S}_1$ en $\mathcal{S}_2$ el programa lineal (1.14) se transforma en

$$\begin{aligned}
\text{minimizar} \quad & \sum_{i=1}^{m} \sum_{j=1}^{n} \gamma_{ij} d_{ij} \\
\text{sujeto a} \quad & \sum_{j=1}^{n} \gamma_{ij} \leq \mu_i^1 \qquad 1 \leq i \leq m & (7.3) \\
& \sum_{i=1}^{m} \gamma_{ij} \leq \mu_j^2 \qquad 1 \leq j \leq n & (7.4) \\
& \sum_{i=1}^{m} \sum_{j=1}^{n} \gamma_{ij} = \min \left\{ \sum_{i=1}^{m} \mu_i^1 \,, \; \sum_{j=1}^{n} \mu_j^2 \right\}. & (7.5)
\end{aligned}$$

Las desigualdades en (7.3) y (7.4) se necesitan porque, a priori, se desconoce cuál de las dos medidas es más masiva y por lo tanto no se sabe si fluirá toda la masa de X_1 dejando lugar vacío en X_2 o si no podrá fluir toda la masa de X_1. La restricción (7.5) fuerza el mayor flujo posible desde X_1 hasta X_2. En el caso de que $\mathcal{S}_1$ y $\mathcal{S}_2$ tengan la misma masa el programa lineal anterior será equivalente a (1.14).

La utilidad de las distancias de Wasserstein en este caso es medir la similaridad entre objetos discretos de algún tipo, que representamos como signaturas distintas en un mismo espacio métrico (X, d). Pero, como señalamos anteriormente, tenemos un problema porque las distancias de Wasserstein sólo están definidas para el caso en el que las medidas son de probabilidad. Para solucionar esta dificultad hay dos posibilidades.

La primera de ellas consiste en normalizar las medidas μ^1 y μ^2 para que ambas sean de probabilidad y calcular la distancia de Wasserstein entre las medidas normalizadas. Al hacer este cambio la noción de distancia resultante entre signaturas ya no es una métrica sino una pseudométrica: una medida μ y cualquier versión escalada $\lambda\mu$ de ésta coinciden después de una normalización. Así, con la notación anterior tenemos una pseudométrica d entre signaturas dada por

$$d(\mathcal{S}_1, \mathcal{S}_2) = W_1\left(\frac{\mu^1}{\mu^1(X_1)}, \frac{\mu^2}{\mu^2(X_2)}\right) ,$$

donde W_1 se calcula con la distancia d inicial. Que la métrica y la pseudométrica tengan el mismo nombre se justifica por el hecho de que la pseudométrica d extiende a la métrica d en el siguiente sentido: si consideramos dos signaturas de la forma $\mathcal{S}_1 = \{(x_1^1, \mu_1^1)\}$ y $\mathcal{S}_2 = \{(x_1^2, \mu_1^2)\}$, entonces vale que $d(\mathcal{S}_1, \mathcal{S}_2) = d(x_1^1, x_1^2)$.

La segunda posibilidad consiste en resolver el programa lineal y tomar como distancia entre $\mathcal{S}_1$ y $\mathcal{S}_2$ el cociente entre el costo total y el flujo total. La noción de distancia resultante entre signaturas se conoce como **earth mover's distance** y se nota EMD. Más precisamente, se tiene

$$\mathrm{EMD}(\mathcal{S}_1, \mathcal{S}_2) = \frac{\sum_{i=1}^m \sum_{j=1}^n \gamma_{ij}^* d_{ij}}{\sum_{i=1}^m \sum_{j=1}^n \gamma_{ij}^*} ,$$

donde $\gamma^* = (\gamma_{ij}^*)_{ij}$ es un minimizante para el problema de flujo asociado a $\mathcal{S}_1$ y $\mathcal{S}_2$.

Una aplicación de esta noción de distancia al procesamiento de imágenes puede verse en [29]. Se aplica allí la noción de distancia EMD para responder consultas basadas en contenido — color o textura — a una base de datos de imágenes.

Existen muchas otras aplicaciones de la teoría de transporte óptimo al procesamiento de imágenes. En [23] pueden verse técnicas recientes aplicadas a problemas de transferencia de color y a problemas de segmentación.

Apéndices

Apéndice A

Convexidad

En este capítulo repasamos las definiciones y teoremas de la teoría de convexidad necesarios para desarrollar la teoría de transporte óptimo. En la primera sección tratamos la semicontinuidad y en la segunda la convexidad propiamente dicha. Las referencias para el capítulo son [12] y [10].

A.1. Semicontinuidad

Definición A.1.1. Sea X un espacio topológico, $x_0 \in X$ y $f : X \to \mathbb{R} \cup \{-\infty, +\infty\}$. La función f se dice **semicontinua superiormente en x_0** :

- si $f(x_0) > -\infty$ y para todo $\varepsilon > 0$ existe un entorno U de x_0 en X tal que para todo x en U resulta $f(x) < f(x_0) + \varepsilon$;
- o bien, si $f(x_0) = -\infty$ y $f(x) \to -\infty$ cuando x tiende a x_0.

Es fácil ver que cuando X es un espacio métrico f resulta semicontinua superiormente en x_0 si y sólo si

$$\limsup_{x \to x_0} f(x) \leq f(x_0).$$

Decimos que f es **semicontinua superiormente** si f es semicontinua superiormente en x para todo x en X. Análogamente, la función f se dice **semicontinua inferiormente en x_0** :

- si $f(x_0) < +\infty$ y para todo $\varepsilon > 0$ existe un entorno U de x_0 en X tal que para todo x en U se tiene $f(x_0) - \varepsilon < f(x)$;
- o bien, si $f(x_0) = +\infty$ y $f(x) \to +\infty$ cuando x tiende a x_0.

Cuando X es un espacio métrico f resulta semicontinua inferiormente en x_0 si y sólo si

$$\liminf_{x \to x_0} f(x) \geq f(x_0).$$

Decimos que f es **semicontinua inferiormente** si f es semicontinua inferiormente en x para todo x en X.

A partir de las definiciones anteriores es claro que, tanto puntual como globalmente, f es continua si y sólo si es semicontinua superior e inferiormente.

En general no es cierto que ínfimos o supremos puntuales de familias arbitrarias de funciones continuas definan funciones continuas. La siguiente proposición muestra que sí se preserva la semicontinuidad.

Proposición A.1.2. *Sea X un espacio topológico y $(f_i)_i$ una familia de funciones definidas en X tomando valores en $\mathbb{R} \cup \{-\infty, +\infty\}$.*

Si cada f_i es semicontinua superiormente, entonces la función $f : X \to \mathbb{R} \cup \{-\infty, +\infty\}$ dada por

$$f(x) = \inf_i f_i(x)$$

para todo x en X es semicontinua superiormente.

Análogamente, cuando cada f_i es semicontinua inferiormente y la función f se define como

$$f(x) = \sup_i f_i(x)$$

para todo x en X, entonces f es semicontinua inferiormente.

En el contexto de espacios métricos vale una propiedad recíproca que permite expresar funciones semicontinuas como límites puntuales de funciones continuas.

Proposición A.1.3. *Sea X un espacio métrico y $f : X \to \mathbb{R} \cup \{-\infty, +\infty\}$.*

Si f es semicontinua superiormente, existe una sucesión $(f_n)_n$ de funciones continuas definidas en X tales que $f = \inf_n f_n$.

Análogamente, si f es semicontinua inferiormente, entonces existe una sucesión $(f_n)_n$ de funciones continuas definidas en X tales que $f = \sup_n f_n$.

El siguiente resultado es un refinamiento de la proposición anterior que resultará útil para probar el teorema de dualidad de Kantorovich.

Proposición A.1.4. *Sea (X, d) un espacio métrico y $F : X \to \mathbb{R} \cup \{+\infty\}$ una función no negativa y semicontinua inferiormente. Para cada n en $\mathbb{N}$ definamos una función F_n en X dada por*

$$F_n(x) = \inf_{y \in X} \left\{ F(y) + nd(x, y) \right\}.$$

Luego $(F_n)_n$ es una sucesión no decreciente de funciones no negativas y uniformemente continuas en X tales que $F = \sup F_n$.

Otra propiedad importante de las funciones semicontinuas es que alcanzan extremos sobre compactos.

Proposición A.1.5. *Sea X un espacio topológico y K un compacto de X. Entonces toda función $f : K \to \mathbb{R} \cup \{-\infty\}$ semicontinua superiormente alcanza máximo absoluto. Análogamente, si f es semicontinua inferiormente entonces f alcanza mínimo absoluto.*

El siguiente corolario es inmediato a partir de la proposición anterior.

Corolario A.1.6. *Sea X un espacio topológico y A un precompacto de X.*

Si $f : \overline{A} \to \mathbb{R} \cup \{-\infty\}$ es semicontinua superiormente, entonces existe x_0 en $\overline{A}$ tal que $f(x_0) \geq f(x)$ para todo x en A.

Si $f : \overline{A} \to \mathbb{R} \cup \{+\infty\}$ es semicontinua inferiormente, entonces existe x_0 en $\overline{A}$ tal que $f(x_0) \leq f(x)$ para todo x en A.

A.2. Convexidad

En esta sección E es un espacio normado y E^* su dual topológico — i.e. el de los funcionales lineales y continuos —. Dada una función $\varphi : E \to (-\infty; +\infty]$ definimos su **dominio** como

$$D(\varphi) = \left\{ x \in E : \ \varphi(x) < +\infty \right\}$$

y su **epigrafo** como

$$\mathrm{epi}(\varphi) = \left\{ (x, z) \in E \times \mathbb{R} : \ \varphi(x) \leq z \right\}.$$

Notar que $\mathrm{epi}(\varphi)$ es un subconjunto de $E \times \mathbb{R} = E \times (-\infty; +\infty)$ y por lo tanto no contiene ningún par de la forma (x, z) con $z = +\infty$, aunque claramente estos pares sí satisfacen $\varphi(x) \leq z$.

Decimos que φ es **propia** si existe al menos un punto de E en el cual es finita. Es decir, φ es propia si $D(\varphi)$ es no vacío.

Decimos que φ es **convexa** si dados x_1, x_2 en E cualesquiera y $\lambda \in [0;1]$ vale

$$\varphi(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\varphi(x_1) + (1-\lambda)\varphi(x_2)\,,$$

y que es **estrictamente convexa** si es convexa y para $x_1 \neq x_2$ y $\lambda \in (0;1)$ vale la desigualdad estricta

$$\varphi(\lambda x_1 + (1-\lambda)x_2) < \lambda\varphi(x_1) + (1-\lambda)\varphi(x_2).$$

Notar que dados $D \subset E$ y una función convexa φ definida en D, es posible extender φ a E preservando la convexidad si se define φ como $+\infty$ en $E \backslash D$.

A continuación se listan algunas propiedades elementales de las funciones convexas.

- Si φ es convexa entonces $D(\varphi)$ es un subconjunto convexo de E y $\mathrm{epi}(\varphi)$ es un subconjunto convexo de $E \times \mathbb{R}$.

- Si φ y ψ son convexas también lo es cualquier combinación cónica de ellas; es decir, si α y β son números reales no negativos, entonces $\alpha\varphi + \beta\psi$ es convexa.

- Si $(\varphi_i)_i$ es una familia de funciones convexas entonces $\varphi = \sup_i \varphi_i$ también lo es.

El siguiente lema es una consecuencia del teorema de Hahn–Banach.

Lema A.2.1. *Sea E un espacio normado y $\varphi : E \to (-\infty; +\infty]$ es una función convexa. Entonces existen x^* en E^* y β en $\mathbb{R}$ tales que*

$$\varphi(x) \geq \langle x^*, x\rangle + \beta \tag{A.1}$$

para todo x en E. Si x_0 en E es tal que $\varphi(x_0) < +\infty$, entonces es posible elegir x^* y β para que en (A.1) valga la igualdad cuando $x = x_0$.

Observación A.2.2. El lema anterior no es más que una generalización de la existencia de hiperplanos soporte para conjuntos convexos aplicada al epigrafo de φ. ▲

Definición A.2.3. Si $\varphi : E \to (-\infty; +\infty]$ es propia definimos la **transformada de Legendre–Fenchel de φ** como la función $\varphi^* : E^* \to (-\infty; +\infty]$ dada por

$$\varphi^*(x^*) = \sup_{x \in E} \left\{ \langle x^*, x\rangle - \varphi(x) \right\}\,.$$

También se conoce a φ^* como la **conjugación de Fenchel de φ**.

Observación A.2.4. Es necesario probar la buena defición de φ^* : por ser φ propia, existe x_0 en E tal que $\varphi(x_0) < +\infty$ y por lo tanto

$$\varphi^*(x^*) = \sup_{x \in E} \big\{ \langle x^*, x\rangle - \varphi(x) \big\} \geq \langle x^*, x_0\rangle - \varphi(x_0) > -\infty$$

para todo x^* en E^*. ▲

Proposición A.2.5 (Desigualdad de Young). *Si φ es una función convexa propia en E, entonces φ^* es una función convexa propia en E^*. Más aun, φ^* es semicontinua inferiormente y satisface*

$$\varphi(x) + \varphi^*(x^*) \geq \langle x^*, x\rangle \tag{A.2}$$

para todo x en E y todo x^* en E^*.

Demostración. El lema A.2.1 afirma que existen x^* y β satisfaciendo (A.1). Luego

$$\beta \geq \langle x^*, x\rangle - \varphi(x)$$

para todo x en E y entonces

$$\varphi^*(x^*) = \sup_{x \in E} \big\{ \langle x^*, x\rangle - \varphi(x) \big\} \leq \beta\,,$$

con lo cual φ^* no es idénticamente $+\infty$.

Dado x en E, la asignación

$$x^* \mapsto \langle x^*, x\rangle - \varphi(x)$$

define una función afín — y por lo tanto continua y convexa —. Luego φ^* es semicontinua inferiormente por A.1.2 y convexa por ser supremo de convexas. ■

Si al conjugar una función propia φ resulta φ^* propia, nada impide iterar el proceso de conjugación y definir $\varphi^{**} : E \to [-\infty; +\infty]$ por

$$\varphi^{**}(x) = \sup_{x^* \in E^*} \big\{ \langle x^*, x\rangle - \varphi^*(x^*) \big\}.$$

Aunque φ^{**} no resulte propia, sí está bien definida como función. Una consecuencia de (A.2) es que dado x en E resulta

$$\varphi(x) \geq \langle x^*, x\rangle - \varphi^*(x^*) \tag{A.3}$$

para todo x^* en E^*. Tomando supremo en x^* se concluye $\varphi^{**} \leq \varphi$. No se puede esperar que valga la igualdad $\varphi^{**} = \varphi$ si φ no es semicontinua inferiormente o no es convexa, pues la doble conjugada $\varphi^{**} = (\varphi^*)^*$ siempre tiene estas propiedades por la demostración de A.2.5. El siguiente teorema, cuya demostración puede encontrarse en [10], afirma que esas condiciones necesarias para la igualdad son también suficientes.

Teorema A.2.6 (Teorema de biconjugación de Fenchel). *Si $\varphi : E \to (-\infty; +\infty]$ es una función propia, entonces son equivalentes:*

1. *φ es semicontinua inferiormente y convexa*

2. *φ^* es también propia y vale $\varphi = \varphi^{**}$.*

A continuación se enuncia un teorema de dualidad. En particular, este resultado se usa para probar el teorema de dualidad de Kantorovich.

Teorema A.2.7 (Dualidad de Fenchel–Rockafellar). *Sea E un espacio normado, E^* su dual topológico y Θ, Ξ dos funciones convexas en E tomando valores en $\mathbb{R} \cup \{+\infty\}$. Sean Θ^* y Ξ^* las transformadas de Fenchel–Rockafellar de Θ y Ξ respectivamente. Supongamos que existe x_0 en E tal que se satisfacen simultáneamete $\Theta(x_0) < +\infty$, $\Xi(x_0) < +\infty$ y Θ continua en x_0. Entonces vale la igualdad*

$$\inf_{x \in E} \big\{\Theta(x) + \Xi(x)\big\} = \max_{x^* \in E^*} \big\{ -\Theta^*(-x^*) - \Xi^*(x^*)\big\}. \tag{A.4}$$

Observación A.2.8. Es parte del contenido del teorema que el máximo a la derecha de (A.4) es un máximo y no sólo un supremo. La demostración de este resultado es una aplicación del teorema de Hahn–Banach y puede leerse en [37] o en [12]. ▲

Apéndice B

Medida y Probabilidad

En este capítulo resumimos las definiciones y resultados fundamentales sobre medidas en espacios 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 capítulo requieren esa hipótesis para poder probarse. Al final del capítulo mencionamos algunos resultados sobre medidas vectoriales que se usan en el capítulo 6 para estudiar problemas de transporte.

Las referencias para este capítulo son [28], [30], [9] y [25]. Los resultados sobre medidas vectoriales pueden consultarse en [33].

B.1. Medida y probabilidad en espacios métricos

En lo que sigue X siempre es un espacio métrico o topológico. Notamos $\mathcal{B}_X$, o $\mathcal{B}$ si no hay ambigüedad, a la σ-álgebra de Borel de X. Llamamos **medida boreliana** en X a cualquier medida definida en el espacio medible $(X, \mathcal{B}_X)$. Notamos $\mathcal{M}(X)$ a la colección de las medidas con signo borelianas finitas en X, $\mathcal{M}^+(X)$ al subconjunto de $\mathcal{M}(X)$ formado por las medidas no negativas y $\mathcal{P}(X)$ al de las medidas borelianas de probabilidad en X.

El conjunto $\mathcal{M}(X)$ admite estructura de espacio vectorial y de hecho toda medida μ en $\mathcal{M}(X)$ 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 μ en $\mathcal{M}(X)$ 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

$$|\mu|(X) = \sup\Big\{ \sum_i |\mu(E_i)| \;:\; (E_i)_i \text{ partición boreliana y numerable de } X \Big\}\,.$$

Vale que $\|\mu\| = |\mu|(X)$ es una norma en $\mathcal{M}(X)$ y que el espacio $(\mathcal{M}(X), \|\mu\|)$ es de Banach. Más aun, si conservamos sólo las medidas "buenas" de $(\mathcal{M}(X), \|\mu\|)$, este espacio resulta ser el dual de las funciones continuas en X que tienden a cero en el infinito.

Definición B.1.1. Una medida μ en un espacio topológico X es **regular** si para cada $B \in \mathcal{B}$ vale

$$\mu(B) = \sup\big\{\mu(F) \;:\; F \subset B \text{ cerrado}\big\} = \text{ínf}\,\big\{\mu(U) \;:\; U \supset B \text{ abierto}\big\}.$$

Teorema B.1.2. *Toda medida boreliana y finita en un espacio métrico es regular. Luego, dado un espacio métrico X junto con dos medidas borelianas y finitas μ y ν, son equivalentes:*

1. $\mu = \nu$
2. *μ y ν coinciden sobre los cerrados*
3. *μ y ν coinciden sobre los abiertos*

Definición B.1.3. Dado un espacio topológico X, definimos los siguiente espacios de funciones continuas:

$$\begin{aligned}
C(X) &= \big\{f : X \to \mathbb{R} \;:\; f \text{ continua}\big\}\,,\\
C_b(X) &= \big\{\, f \in C(X) \;:\; f \text{ acotada}\,\big\}\,,\\
C_0(X) &= \big\{\, f \in C(X) \;:\; \lim_{x\to\infty} f(x) = 0\big\}\,,\\
C_c(X) &= \big\{\, f \in C(X) \;:\; \mathrm{sop}(f) \text{ compacto}\,\big\}\,.
\end{aligned}$$

Los espacios $C_b(X)$ y $C_0(X)$ son de Banach con la norma $\|\cdot\|_\infty$. Con esa misma norma, $C_c(X)$ es normado y es un subespacio denso de $C_0(X)$. En general, $C_c(X)$ no es un espacio de Banach con la norma $\|\cdot\|_\infty$.

Teorema B.1.4 (Riesz–Markov para medidas con signo). *Sea un espacio topológico X localmente compacto y Hausdorff. Dado un funcional continuo ψ en $C_0(X)^*$, 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.

B.1.1. Convergencia débil de medidas

Tenemos en $\mathcal{M}(X)$ una topología inducida por la norma variación total. La dualidad del teorema de Riesz–Markov sugiere considerar en $\mathcal{M}(X)$ una topología débil cuando pensamos a este espacio como el dual de las funciones continuas que tienden a cero en el infinito. Notar que en términos precisos esta topología sería la topología débil-$*$ en $\mathcal{M}(X)$.

Definición B.1.5. Dada una sucesión $(\mu_n)_n$ de medidas borelianas y finitas en un espacio métrico 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 $\mu_n \Rightarrow \mu$.

El siguiente teorema da algunas condiciones equivalentes a la convergencia débil de medidas en $\mathcal{P}(X) \subset \mathcal{M}(X)$.

Teorema B.1.6. *Sea X un espacio métrico, $(\mu_n)_n$ una sucesión de medidas de probabilidad en X y $\mu \in \mathcal{P}(X)$. Son equivalentes:*

1. $\mu_n \Rightarrow \mu$
2. $\int_X f\, d\mu_n \xrightarrow[n]{} \int_X f\, d\mu$ *para toda f uniformemente continua y acotada en X*
3. $\limsup_n \mu_n(F) \leq \mu(F)$ *para todo cerrado $F \subset X$*
4. $\liminf_n \mu_n(U) \geq \mu(U)$ *para todo abierto $U \subset X$*
5. $\mu_n(B) \xrightarrow[n]{} \mu(B)$ *para todo boreliano $B \in \mathcal{B}$ 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 para toda sucesión de medidas de probabilidad $(\mu_n)_n$ y para toda $\mu \in \mathcal{P}(X)$ suceda que $(\mu_n)_n$ converge débilmente 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 B.1.7. Sea X un espacio métrico. Dado un subconjunto no vacío A de X y $\alpha > 0$, definimos

$$A_\alpha = \left\{ x \in X \,:\, d(x, A) < \alpha \right\},$$

donde $d(x, A) = \inf \left\{ d(x, a) \,:\, a \in A \right\}$. Definimos también $\emptyset_\alpha = \emptyset$ para todo $\alpha > 0$. Para cada par de medidas $\mu, \nu \in \mathcal{P}(X)$ definimos

$$d_P(\mu, \nu) = \inf \left\{ \alpha > 0 \,:\, \mu(B) \leq \nu(B_\alpha) + \alpha \;\text{ y }\; \nu(B) \leq \mu(B_\alpha) + \alpha \text{ para todo } B \in \mathcal{B}_X \right\}.$$

El siguiente teorema muestra que tenemos un espacio métrico $(\mathcal{P}(X), d_P)$ y que éste metriza la convergencia débil en $\mathcal{P}(X)$. La métrica d_P se llama **métrica de Prokhorov** inducida por d.

Teorema B.1.8. *Sea (X, d) un espacio métrico. Entonces:*

1. *d_P es una métrica en $\mathcal{P}(X)$.*
2. *Si $(\mu_n)_n$ es una sucesión en $\mathcal{P}(X)$ y $\mu \in \mathcal{P}(X)$, entonces $d_P(\mu_n, \mu) \underset{n}{\longrightarrow} 0$ implica $\mu_n \Rightarrow \mu$.*
3. *Si (X, d) es un espacio métrico separable entonces $(\mathcal{P}(X), d_P)$ también lo es y, en ese caso, para toda sucesión $(\mu_n)_n$ y toda μ en $\mathcal{P}(X)$ resulta $d_P(\mu_n, \mu) \underset{n}{\longrightarrow} 0$ si y sólo si $\mu_n \Rightarrow \mu$.*

Una consecuencia del teorema anterior es la unicidad de los límites débiles en $\mathcal{P}(X)$, pues son límites para una noción de convergencia metrizable.

B.1.2. Rigidez y Teorema de Prokhorov

Cuando se construye una medida de probabilidad como límite de medidas, a veces es razonable hacerlo considerando primero una sucesión de medidas que satisfagan cada vez más fuertemente un cierta propiedad para después extraer de ésta una subsucesión que converja a una medida límite. Por este motivo, es útil contar con algún teorema que dé condiciones necesarias para poder extraer sucesiones convergentes de una familia de medidas. El teorema de Prokhorov 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.

Un 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 B.1.9. *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 que resuelva la problemática planteada. Exigiendo completitud y separabilidad a (X, d), podremos dar un tal teorema usando la siguiente noción.

Definición B.1.10. Decimos que una medida boreliana y finita μ en X es **rígida** si para todo $\varepsilon > 0$ existe $K \subset X$ compacto tal que $\mu(X \backslash K) < \varepsilon$ o, equivalentemente, tal que $\mu(K) \geq \mu(X) - \varepsilon$.

Notar que cuando X es compacto toda medida boreliana y finita en X es rígida.

Lema B.1.11. *Si X es un espacio métrico y μ una medida boreliana, finita y rígida en él, entonces μ está soportada en un subconjunto σ-compacto de X.*

Teorema B.1.12. *Si μ es una medida boreliana, finita y rígida en X, entonces para cada $B \in \mathcal{B}$ vale que*

$$\mu(B) = \sup\big\{\mu(K) \,:\, K \subset B \text{ compacto}\big\}.$$

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 B.1.13. Un **espacio polaco** es un espacio métrico completo y separable.

Teorema B.1.14. *Si X es un espacio polaco entonces toda medida boreliana y finita en X es rígida.*

Definición B.1.15. Si X es un espacio métrico, una familia de medidas de probabilidad $\mathcal{F} \subset \mathcal{P}(X)$ se dice **rígida** 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 B.1.16. Cuando $\mathcal{F}$ es un singleton $\{\mu\}$, $\mathcal{F}$ es rígida si y sólo si μ lo es. ▲

Teorema B.1.17 (Prokhorov). *Sea (X, d) un espacio métrico completo y separable y $\mathcal{F}$ un subconjunto de $\mathcal{P}(X)$. Consideremos en $\mathcal{P}(X)$ la métrica de Prokhorov d_P. Son equivalentes:*

1. *$\overline{\mathcal{F}}$ es compacto en $\mathcal{P}(X)$.*
2. *$\mathcal{F}$ es rígida.*

B.2. Medidas vectoriales

Definición B.2.1. Dado un espacio topológico X, una **medida vectorial** en X es una función σ-aditiva $\mu : \mathcal{B}(X) \to \mathbb{R}^d$ para algún d en $\mathbb{N}$. Para cada $d \in \mathbb{N}$, notamos $\mathcal{M}^d(X)$ al conjunto de medidas vectoriales en X tomando valores en $\mathbb{R}^d$. A cada medida $\mu \in \mathcal{M}^d(X)$ le asignamos una medida positiva $|\mu| \in \mathcal{M}^+(X)$ dada por

$$|\mu|(E) = \sup\Big\{\sum_i |\mu(E_i)| \,:\, (E_i)_i \text{ partición numerable de } E\Big\} \tag{B.1}$$

para cada boreliano E. Llamamos **medida variación total** de μ a $|\mu|$.

Observación B.2.2. En (B.1), la notación $|\mu(E_i)|$ denota la norma euclídea del vector $\mu(E_i)$. Notar que cuando $d = 1$ se recuperan las definiciones usuales de medida con signo y medida variación total para medidas con signo. ▲

Para cada medida μ en $\mathcal{M}^d(X)$ podemos considerar las medidas con signo $\pi_i \circ \mu$ en $\mathcal{M}(X)$, donde $\pi_i : \mathbb{R}^d \to \mathbb{R}$ es la proyección en la i-ésima coordenada. Recíprocamente, dadas medidas $\mu_1, \ldots, \mu_d$ en $\mathcal{M}(X)$ podemos costruir la medida $\mu \in \mathcal{M}^d(X)$ dada por $\mu = (\mu_1, \ldots, \mu_d)$. Esto da una correspondencia entre $\mathcal{M}^d(X)$ y $(\mathcal{M}(X))^d$.

Definición B.2.3. Dada una medida vectorial μ en $\mathcal{M}^d(X)$, definimos $L^1(X, \mu)$ como el conjunto de las $f = (f_1, f_2, \ldots, f_d) : X \to \mathbb{R}^d$ tales que $f_i \in L^1(X, \mu_i)$ para todo i, donde $\mu_i = \pi_i \circ \mu$. Dada f en $L^1(X, \mu)$, definimos

$$\int_X f \cdot d\mu \coloneqq \sum_{i=1}^{d} \int_X f_i \, d\mu_i \; .$$

Cuando X es un espacio topológico localmente compacto y Hausdorff y $\psi \in C_0(X, \mathbb{R}^d)$ un funcional continuo, es posible aplicar B.1.4 coordenada a coordenada para probar el siguiente resultado.

Teorema B.2.4 (Riesz–Markov para medidas vectoriales). *Dado X localmente compacto y Hausdorff y dado un funcional continuo ψ en $C_0(X, \mathbb{R}^d)^*$, existe una única medida vectorial μ en $\mathcal{M}^d(X)$ tal que*

$$\psi(f) = \int_X f \cdot d\mu$$

para toda f en $C_0(X, \mathbb{R}^d)$.

Definición B.2.5. Dado un dominio acotado $\Omega \subset \mathbb{R}^d$ y μ en $\mathcal{M}^d(\Omega)$, decimos que ν en $\mathcal{M}(\Omega)$ es la **divergencia** de μ si vale

$$\int_\Omega \nabla\varphi \cdot d\mu = -\int_\Omega \varphi d\nu \tag{B.2}$$

para toda φ en $C^1(\overline{\Omega})$ y en ese caso notamos $\nabla \cdot \mu = \nu$ o $\operatorname{div} \mu = \nu$. Notamos $\mathcal{M}^d_{\mathrm{div}}(\Omega)$ al conjunto de las medidas vectoriales μ en $\mathcal{M}^d(\Omega)$ para las que existe $\operatorname{div} \mu$ en $\mathcal{M}(\Omega)$.

Bibliografía

[1] L. Ambrosio. Lecture notes on optimal transport problems. In *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*, pages 1–52. Springer Berlin Heidelberg, 2003.

[2] L. Ambrosio and N. Gigli. A user's guide to optimal transport. In *Modelling and Optimisation of Flows on Networks: Cetraro, Italy 2009, Editors: Benedetto Piccoli, Michel Rascle*, pages 1–155. Springer Berlin Heidelberg, 2013.

[3] L. Ambrosio and A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation. In *Optimal Transportation and Applications: Lectures given at the C.I.M.E. Summer School, held in Martina Franca, Italy, September 2-8, 2001*, pages 123–160. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[4] T. Aubin and Y. Li. On the best Sobolev inequality. *Journal de Mathématiques Pures et Appliquées*, 78(4):353 – 387, 1999.

[5] M. Beckmann. A continuous model of transportation. *Econometrica: Journal of the Econometric Society*, pages 643–660, 1952.

[6] M. Beiglboeck, C. Leonard, and W. Schachermayer. A general duality theorem for the Monge–Kantorovich transport problem, 2009, arXiv:0911.4347.

[7] D. Bertsimas and J. Tsitsiklis. *Introduction to linear optimization*. Athena Scientific Belmont, MA, 1997.

[8] S. Bianchini and M. Bardelloni. The decomposition of optimal transportation problems with convex cost, 2014, arXiv:1409.0515.

[9] P. Billingsley. *Convergence of probability measures*. Wiley, New York, 1999.

[10] J. Borwein and A. Lewis. *Convex analysis and nonlinear optimization: theory and examples*. Springer Science & Business Media, 2010.

[11] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[12] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.

[13] G. Carlier. Optimal transportation and economic applications. https://www.ceremade.dauphine.fr/~carlier/ima-transport-lecture-notes.pdf, 2010.

[14] G. Carlier, C. Jimenez, and F. Santambrogio. Optimal transportation with traffic congestion and Wardrop equilibria. *SIAM Journal on Control and Optimization*, 47(3):1330–1350, 2008.

[15] L. Evans and R. Gariepy. *Measure theory and fine properties of functions*. CRC Press, 1992.

[16] L. C. Evans. *Partial differential equations*. Graduate studies in mathematics. American Mathematical Society, second edition, 2010.

[17] H. Federer. *Geometric measure theory*. Springer, 2014.

[18] N. Fusco. The stability of the isoperimetric inequality. http://math.cmu.edu/~iantice/cna_2013/lecturenotes/fusco.pdf, 2013.

[19] W. Gangbo. The Monge mass transfer problems and its applications. *NSF-CBMS Conference. Contemporary Mathematics*, 1999.

[20] W. Gangbo and R. McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, 1996.

[21] P. Mattila. *Geometry of sets and measures in Euclidean spaces: fractals and rectifiability*. Cambridge university press, 1999.

[22] G. Monge. *Mémoire sur la théorie des déblais et des remblais*. In Histoire de l'Academie Royale des Sciences de Paris, 1781.

[23] N. Papadakis. *Optimal Transport for Image Processing*. Accreditation to supervise research, Université de Bordeaux, December 2015.

[24] C. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Dover, 1998.

[25] K. R. Parthasarathy. *Probability measures on metric spaces*, volume 352. American Mathematical Soc., 1967.

[26] A. Pratelli. On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation. In *Annales de l'Institut Henri Poincare (B) Probability and Statistics*. Elsevier, 2007.

[27] A. Pratelli. On the sufficiency of c-cyclical monotonicity for optimality of transport plans. *Mathematische Zeitschrift*, 2008.

[28] H.L. Royden. *Real Analysis*. Mathematics and statistics. Macmillan, 1988.

[29] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40, 2000.

[30] W. Rudin. *Real and complex analysis (3rd)*. New York: McGraw-Hill Inc, 1986.

[31] L. Rüschendorf. On c-optimal random variables. *Statistics & Probability Letters*, 27(3):267–270, 1996.

[32] F. Santambrogio. c-cyclical monotonicity of the support of optimal transport plans. http://www.math.u-psud.fr/~santambr/SupportcCM.pdf, 2015.

[33] F. Santambrogio. *Optimal transport for applied mathematicians*. Progress in Nonlinear Differential Equations and their applications. Springer, 2015.

[34] W. Schachermayer and J. Teichmann. Characterization of optimal transport plans for the Monge-Kantorovich problem. *Proceedings of the American Mathematical Society*, 137(2):519–529, 2009.

[35] V. Sudakov. *Geometric problems in the theory of infinite-dimensional probability distributions*. American Mathematical Soc., 1979.

[36] G. Talenti. Best constant in Sobolev inequality. *Annali di Matematica Pura ed Applicata*, 110(1):353–372, 1976.

[37] C. Villani. *Topics in optimal transportation*. American Mathematical Society, 2003.

[38] C. Villani. *Optimal transport: old and new*. Springer Science & Business Media, 2008.