

## Capítulo 2

# Dualidade de Kantorovitch Rubinstein

Nesse capítulo vamos discutir um dos resultados mais fundamentais da teoria de transporte ótimo, a *dualidade de Kantorovitch-Rubinstein*. Como anteriormente, ao longo desse capítulo vamos assumir que  $(\mathcal{X}, d_{\mathcal{X}})$  e  $(\mathcal{Y}, d_{\mathcal{Y}})$  são espaços Poloneses. Então dados  $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  s.c.i., tal que o problema de Kantorovitch tem valor finito para duas distribuições de probabilidade  $\mu \in \mathcal{P}(\mathcal{X})$ ,  $\nu \in \mathcal{P}(\mathcal{Y})$ , a dualidade de Kantorovitch afirma que

$$\min_{\gamma \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c d\gamma = \sup_{\substack{\varphi \oplus \psi \leq c \\ \varphi \in \mathcal{C}_b(\mathcal{X}), \psi \in \mathcal{C}_b(\mathcal{Y})}} \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu, \quad (2.1)$$

onde a operação  $\oplus : \mathcal{C}_b(\mathcal{X}) \times \mathcal{C}_b(\mathcal{Y}) \rightarrow \mathcal{C}_b(\mathcal{X} \times \mathcal{Y})$  é definida como  $\varphi \oplus \psi(x, y) \stackrel{\text{def.}}{=} \varphi(x) + \psi(y)$ .

Além de ter um significado matemático profundo através da teoria mais geral de dualidade em análise e otimização convexa, esse resultado também tem uma interpretação econômica muito rica, um dos motivos pelos quais Kantorovitch ganhou o prêmio Nobel da economia.

Considere a seguinte situação:  $\mu$  representa a distribuição de padarias numa cidade que produzem pão, enquanto que  $\nu$  representa a distribuição de cafés da mesma que geram a demanda pelo pão produzido pelas padarias. Dessa forma  $c(x, y)$  representa o custo de levar uma unidade de pão da padaria  $x$  ao café  $y$ . O problema de Kantorovitch, também chamado o problema primal, representa portanto o custo total de transporte mínimo que um consórcio entre padarias e cafés pagaria para atender toda a demanda de pão, enquanto exaurindo a produção, assumindo que estas estão em equilíbrio e não há excedente de produção.

No supremo do lado direito, o problema dual, a quantidade  $\varphi(x)$  representa portanto o quanto uma transportadora deve cobrar de uma padaria  $x$  para retirar uma unidade de pão, enquanto que  $\psi(y)$  representa o preço de entrega de uma unidade de pão no café  $y$ . A restrição  $\varphi(x) + \psi(y) \leq c(x, y)$  significa que a transportadora não pode cobrar mais que o custo de transporte individual  $c(x, y)$  da padaria  $x$  e do café  $y$  juntos, pois se esse fosse o caso, o consórcio de cafés e padarias poderia efetuar o transporte por um custo melhor que a transportadora.

O leitor atento percebeu que o problema dual foi escrito como um sup, enquanto que o problema primal costumamos escrever como um min. Isso é por que nem sempre podemos demonstrar a existência de potenciais ótimos  $\varphi, \psi$ , pelo menos não em  $\mathcal{C}_b(\mathcal{X}), \mathcal{C}_b(\mathcal{Y})$ . Um objetivo importante

desse capítulo será obter condições que garantam existência de um par  $(\varphi, \psi)$  que maximize também o problema dual, os chamados *potenciais de Kantorovitch*. Dado um par de potenciais de Kantorovitch  $(\varphi, \psi)$ , se  $\gamma$  é um plano de transporte ótimo, vamos demonstrar que a condição

$$\varphi(x) + \psi(y) = c(x, y) \text{ para } \gamma\text{-quase todo par } (x, y) \quad (2.2)$$

é necessária e suficiente para a otimalidade de  $\gamma$ . Essa relação significa que o problema dual fornece uma precificação do serviço de transporte que atinge o equilíbrio do mercado.

Agora suponha que o potencial  $\psi$  está fixo, uma forma muito natural de escolher um  $\varphi$  de forma que este par seja admissível, ou seja,  $\varphi \oplus \psi \leq c$ , é definir o que chamamos de a *transformada*  $c$

$$\psi^c(x) \stackrel{\text{def.}}{=} \inf_{y \in \mathcal{Y}} c(x, y) - \psi(y), \text{ de formas que } \psi^c \oplus \psi \leq c. \quad (2.3)$$

Assim como a *transformada*  $\bar{c}$

$$\varphi^{\bar{c}}(y) \stackrel{\text{def.}}{=} \inf_{x \in \mathcal{X}} c(x, y) - \varphi(x), \text{ de formas que } \varphi \oplus \varphi^{\bar{c}} \leq c. \quad (2.4)$$

Toda essa discussão é válida para uma classe de custos  $c$  bem ampla, mas o berço dessa teoria começou com o caso  $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$  e

$$c(x, y) = -x \cdot y,$$

que aparece sem perda de generalidade da escolha  $c(x, y) = \frac{1}{2}|x - y|^2$ . Nesse caso o problema de Kantorovitch e seu dual são reescritos como

$$\max_{\gamma \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} x \cdot y d\gamma = \inf_{\varphi \oplus \psi \geq x \cdot y} \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu. \quad (2.5)$$

Nessa variante a noção de transformada- $c$  se torna a *transformada de Legendre* definida como

$$f^*(p) = \sup_{x \in \mathbb{R}^d} x \cdot p - f(x), \quad (2.6)$$

um objeto central da análise e otimização convexa.

Por conta disso, nessa capítulo vamos fazer um detour do nosso objetivo inicial de estudar o problema dual de Kantorovitch para entender algumas ferramentas de análise convexa, onde a transformada de Legendre será o objeto central que também nos servirá mais adiante. Um dos jeitos mais simples e clássicos de demonstrar a dualidade é um exercício de análise convexa, que se tornará um exercício para o leitor. Infelizmente essa prova só funciona em espaços métricos compactos, logo uma vez construídas essas ferramentas, vamos voltar à transformada- $c$ , que pode ser vista como uma generalização da transformada de Legendre.

## 2.1 Um pouco de análise convexa

Nessa sessão vamos estudar as propriedades de funções convexas num espaço Banach geral  $X$ , ou seja um espaço vetorial normado e completo. Seja um conjunto  $C \subset X$ , dizemos que  $C$  é convexo sempre que, dados quaisquer dois pontos de  $C$ , a reta conectando-os está contida em  $C$ , ou seja para todo  $t \in (0, 1)$

$$tx + (1 - t)y \in C \text{ para todos } x, y \in C.$$

Os principais exemplos de espaços de Banach que podemos citar são

$$\mathcal{C}_0(\mathcal{X}) \text{ e } \mathcal{M}_b(\mathcal{X}).$$

O primeiro, o conjunto das funções contínuas que convergem para zero no infinito, definidas em um espaço Polonês  $\mathcal{X}$  é completo pelo fato de que o limite uniforme de funções contínuas é também uma função contínua. Caso o leitor não esteja familiarizado com este fato, ele é fortemente encorajado à prová-lo.

**Exercício 2.1.** Seja uma sequência  $(f_n)_{n \in \mathbb{N}} \subset \mathcal{C}(\mathcal{X})$  que converge uniformemente para  $f$ , ou seja

$$\|f_n - f\|_\infty = \sup_{x \in \mathcal{X}} d_{\mathcal{X}}(f_n(x), f(x)) \xrightarrow{n \rightarrow \infty} 0.$$

Então  $f$  é contínua.

Como consequência, o espaço das medidas de Radon finitas sobre  $\mathcal{X}$ ,  $\mathcal{M}_b(\mathcal{X})$ , também é um espaço de Banach, pois é o espaço dual de um espaço vetorial normado.<sup>1</sup> O fato de que  $\mathcal{M}_b(\mathcal{X}) = \mathcal{C}_0(\mathcal{X})^*$  é um espaço de Banach portanto é completamente independente de  $\mathcal{C}_0(\mathcal{X})^*$  o ser, pois todo espaço dual é completo. Falando de conjuntos convexos, já cruzamos um muito relevante para o nosso estudo:

**Exercício 2.2.** Prove que  $\mathcal{P}(\mathcal{X})$  é um subconjunto convexo de  $\mathcal{M}_b(\mathcal{X})$ . Prove que o conjunto  $\Pi(\mu, \nu)$  de planos de transporte com marginais fixas é convexo.

Uma função  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  é dita convexa se o seu domínio

$$\text{dom } f \stackrel{\text{def.}}{=} \{x \in X : f(x) < +\infty\}$$

é um conjunto convexo, e para todo  $t \in (0, 1)$  vale que

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \text{ para todos } x, y \in X.$$

Por outro lado, uma função  $f$  é dita côncava se  $-f$  é convexa. Nesse caso o domínio de uma função côncava é dado por  $\text{dom } f \stackrel{\text{def.}}{=} \{x \in X : f(x) > -\infty\}$  ainda deve ser convexo e a desigualdade oposta é satisfeita: para todo  $t \in (0, 1)$  vale que

$$f(tx + (1-t)y) \geq tf(x) + (1-t)f(y) \text{ para todos } x, y \in X.$$

Por outro lado, uma função é dita estritamente convexa (resp. côncava) se as desigualdades à cima são estritas. No seguinte exercício o leitor é convidado à verificar diversas propriedades de funções convexas.

**Exercício 2.3.** Sejam  $X$  um espaço de Banach e  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  uma função.

- Prove que se  $f$  é estritamente convexa, semi-contínua inferiormente e não é  $f \equiv +\infty$ , então  $f$  admite um único minimizante.
- Uma função  $f$  é convexa se, e somente se, seu epígrafo é convexo.

---

<sup>1</sup>Com esses dois exemplos os leitor já percebe que, para os nossos objetivos, não basta desenvolver a teoria em dimensão finita.

- Uma função  $f$  é convexa se, e somente se, para todos  $x, y \in X$  a função sobre  $[0, 1]$  dada por  $t \mapsto f(tx + (1 - t)y)$  é convexa.
- Considere uma família arbitrária de funções convexas  $(f_\alpha)_{\alpha \in \Lambda}$ , então

$$f(x) \stackrel{\text{def.}}{=} \sup_{\alpha \in \Lambda} f_\alpha(x)$$

é uma função convexa.

- Seja  $C \subset X$  um conjunto fechado e convexo não vazio. Defina as funções indicatrizes (no sentido da análise convexa) e suporte de  $C$ ,  $\chi_C$  e  $\sigma_C$  respectivamente. Enquanto  $\chi_C : X \rightarrow \mathbb{R} \cup \{+\infty\}$  é definida sobre  $X$ , a função suporte é definida sobre o dual  $\sigma_C : X^* \rightarrow \mathbb{R} \cup \{+\infty\}$ :

$$\chi_C(x) \stackrel{\text{def.}}{=} \begin{cases} 0, & x \in C, \\ +\infty, & \text{se não,} \end{cases} \quad \sigma_C(p) \stackrel{\text{def.}}{=} \sup_{x \in C} \langle p, x \rangle = \sup_{x \in X} \langle p, x \rangle - \chi_C(x).$$

Prove que  $\sigma_C$  e  $\chi_C$  são convexas.

O fato fundamental que permite o desenvolvimento da análise convexa é a *versão geométrica do teorema de Hahn-Banach*, que garante a existência de um hiperplano que separe um conjunto convexo de um ponto exterior a ele.

**Teorema 2.1.1** (Hahn-Banach Geométrico). *Seja  $X$  um espaço de Banach,  $C \subset X$  um conjunto fechado, convexo, não vazio e  $x_0 \in X \setminus C$ . Então, existe  $p \in X^*$  e  $\alpha \in \mathbb{R}$  tais que*

$$\langle p, x \rangle \leq \alpha < \langle p, x_0 \rangle \text{ para todo } x \in C.$$

*Demonstração.* A prova segue da versão analítica do Teorema de Hahn-Banach, aplicada ao funcional de Minkowski. Primeiro note que à menos de uma translação do espaço podemos assumir que  $0 \in \text{int } K$ , o que não muda a conclusão pois esta será baseada em um funcional linear. Definimos então o *funcional de Minkowski*

$$p_K(x) \stackrel{\text{def.}}{=} \inf\{\alpha : \alpha^{-1}x \in K, \alpha > 0\}.$$

Temos as seguintes propriedades:

1.  $p_K(x) \geq 0$  para todo  $x \in X$ ;
2.  $p_K(\lambda x) = \lambda p_K(x)$  para todo  $\lambda > 0$  e  $x \in X$ ;
3.  $p_K(x + y) \leq p_K(x) + p_K(y)$  para todo  $x, y \in X$ ;
4. existe  $M$  tal que  $p_K(x) \leq M \|x\|$  para todo  $x \in X$ ;
5.  $K = \{x \in X : p_K(x) \leq 1\}$ .

O leitor é convidado a provar essas propriedades.

Com o objetivo de aplicar a forma analítica do teorema de Hahn-Banach seja  $Y = \mathbb{R}x_0$  e  $g(tx_0) = t$ . Segue então que  $g \leq p$  em  $Y$ . De fato,

- $t \leq 0$  então  $g(tx_0) = t \leq 0 \leq p_K(tx_0)$

- $t > 0$  então  $\frac{t}{\alpha}x_0 \notin K$  para todo  $\alpha \geq 1$  pois  $x_0 \notin K$ . Na verdade, como  $\text{dist}(x_0, K) > 0$ , temos ainda que  $p_K(tx_0) > t = g(tx_0)$ .

Logo o teorema de Hahn-Banach garante a existência de um funcional linear contínuo  $p \in X^*$  tal que  $p|_Y = g$  e  $p \leq p_K$  em  $X$ . Segue então que para todo  $x \in K$

$$\langle p, x \rangle \leq p_K(x) \leq 1 < g(x_0) = \langle p, x_0 \rangle.$$

Concluímos que o hiperplano  $\{\langle p, x \rangle = 1\}$  separa estritamente  $K$  e  $x_0$ .  $\square$

Desse resultado central serve de fundamento para diversos resultados da análise convexa, inclusive outras versões mais robustas de teoremas de separação. Nós estamos particularmente interessados na teoria de dualidade para problemas de otimização convexa, cujo objeto central é a transformada de Legendre. No exercício 2.3, nos deparamos com o primeiro exemplo de transformada de Legendre com as funções indicatriz e suporte de um conjunto convexo. De modo geral, podemos definir a transformada de Legendre para uma classe bem mais ampla de funções

$$\Gamma_0(X) \stackrel{\text{def.}}{=} \left\{ f : X \rightarrow \mathbb{R} \cup \{+\infty\} : \begin{array}{l} f \text{ é convexa, s.c.i.} \\ \text{e própria, } f \not\equiv +\infty \end{array} \right\}.$$

Dada  $f \in \Gamma_0(X)$ , a sua transformada de Legendre também é uma função  $f^* \in \Gamma_0(X^*)$  definida como

$$f^*(p) = \sup_{x \in X} \langle p, x \rangle - f(x).$$

Um resultado imediato da definição é a chamada *desigualdade de Fenchel*: para todo  $x \in X$  e  $p \in X^*$ , vale

$$f(x) + f^*(p) \geq \langle p, x \rangle.$$

De fato, como  $f^*(p) = \sup_{y \in X} \langle p, y \rangle - f(y)$ , basta tomar  $y = x$  no supremo para obter a desigualdade.

Em alguns casos, essa desigualdade se torna uma igualdade, isto é, existe  $(x, p)$  tal que

$$f(x) + f^*(p) = \langle p, x \rangle.$$

No caso em que  $X = \mathbb{R}^d$  e  $f \in \mathcal{C}^1$ , a condição de otimalidade do supremo definindo  $f^*(p)$  é dada por  $p - \nabla f(x) = 0$ . Escolhendo esses pares de valores conjugados  $(x, \nabla f(x))$ , temos a identidade

$$f(x) + f^*(\nabla f(x)) = \langle \nabla f(x), x \rangle.$$

Assim, a desigualdade de Fenchel se torna uma igualdade nos pontos de contato entre  $f$  e o plano tangente definido pelo gradiente. Isso sugere que, no caso não diferenciável, o papel do gradiente deve ser desempenhado por um conjunto de vetores que realizam esse contato. Isso motiva o estudo das propriedades de diferenciabilidade de funções convexas. Em todo ponto de seu domínio podemos dar uma noção de derivada direcional.

**Lema 2.1.1.** *Seja  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  uma função convexa e s.c.i. sobre um espaço de Banach  $X$ . Se  $x_0 \in \text{int}(\text{dom } f)$  então a derivada direcional está bem definida para todo  $v \in X$ . Isto é, o limite*

$$f'(x_0; v) = \lim_{t \downarrow 0} \frac{f(x_0 + tv) - f(x_0)}{t} = \inf_{t > 0} \frac{f(x_0 + tv) - f(x_0)}{t}$$

*existe está bem definido.*

*Demonstração.* Como  $x_0 \in \text{int}(\text{dom } f)$  existe  $\rho > 0$  tal que  $B_\rho(x_0) \subset \text{dom } f$  (bola aberta de raio  $\rho$ ). Logo, fixando  $v \in X$  para  $t$  suficientemente pequeno  $x_0 + tv \in \text{dom } f$  e portanto a expressão

$$\frac{f(x_0 + tv) - f(x_0)}{t}$$

está bem definida, vamos provar que ela é monótona. Isso implica que o limite existe e o resultado segue.

Pela convexidade de  $f$  e o exercício 2.3,  $t \mapsto \varphi(t) \stackrel{\text{def.}}{=} f(x + tv)$  é uma função convexa. Logo tome  $0 < s < t$  e seja  $\alpha = s/t < 1$ , de formas que  $s = \alpha t + (1 - \alpha)0$ . Pela convexidade de  $\varphi$  temos

$$\begin{aligned}\varphi(s) &\leq \alpha\varphi(t) + (1 - \alpha)\varphi(0) \\ \varphi(s) - \varphi(0) &\leq \alpha(\varphi(t) - \varphi(0)) \\ \frac{\varphi(s) - \varphi(0)}{s} &\leq \frac{\varphi(t) - \varphi(0)}{t}.\end{aligned}$$

Portanto o limite unilateral existe e é igual ao ínfimo pela monotonicidade.  $\square$

Sabemos bem que a existência de derivadas direcionais não garante a diferenciabilidade da função. Vamos ver no próximo capítulo que isso é verdade em quase todo ponto para funções convexas, mas isso não é o bastante para estudar problemas de otimização convexa pois muitas vezes é justamente no ponto de não diferenciabilidade que precisamos obter informação, por isso introduzimos a noção de subdiferencial, que por ser uma noção mais fraca de diferenciabilidade, promete resultados mais gerais para a existência de sub-gradientes.

**Definição 2.1.1** (Subdiferencial). Seja  $f \in \Gamma_0(X)$ . O *subdiferencial* de  $f$  no ponto  $x \in X$  é o conjunto

$$\partial f(x) \stackrel{\text{def.}}{=} \{p \in X^* : f(y) \geq f(x) + \langle p, y - x \rangle \quad \forall y \in X\}.$$

Os elementos de  $\partial f(x)$  são chamados de *subgradientes* de  $f$  em  $x$ .

Note que:

- Se  $f$  é diferenciável em  $x$ , então  $\partial f(x) = \{\nabla f(x)\}$ .
- Em geral, os elementos de  $\partial f(x)$  são exatamente os planos de suporte ao gráfico de  $f$  no ponto  $(x, f(x))$ .

**Proposição 2.1.1** (Subdiferencial não vazio em pontos interiores). *Seja  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  uma função convexa e s.c.i. sobre um espaço de Banach  $X$ . Se  $x_0 \in \text{int}(\text{dom } f)$  então  $\partial f(x_0) \neq \emptyset$ .*

*Demonstração.* Aplicando o Lemma 2.1.1 sobre a existência de derivadas direcionais em  $x_0 \in \text{int dom } f$ , podemos definir o funcional linear e contínuo

$$\varphi : X \ni v \mapsto \varphi(v) = f'(x_0; v),$$

que é sublinear pela convexidade de  $f$  (i.e.  $\varphi(\lambda v) = \lambda\varphi(v)$  para  $\lambda \geq 0$  e  $\varphi(v + w) \leq \varphi(v) + \varphi(w)$ ).

Pelo teorema de Hahn–Banach para funcionais sublineares, existe  $p \in X^*$  tal que

$$\langle p, v \rangle \leq \varphi(v) \quad \text{para todo } v \in X.$$

Da definição de  $\varphi$  como infimo para  $t > 0$  segue que, para todo  $v$  e todo  $t > 0$ ,

$$\langle p, v \rangle \leq \frac{f(x_0 + tv) - f(x_0)}{t},$$

isto é  $f(x_0 + tv) \geq f(x_0) + t\langle p, v \rangle$ . Tomando  $y = x_0 + tv$  e variando  $t > 0, v \in X$  obtemos

$$f(y) \geq f(x_0) + \langle p, y - x_0 \rangle \quad \text{para todo } y \in X,$$

o que mostra  $p \in \partial f(x_0)$ . Portanto  $\partial f(x_0) \neq \emptyset$ .  $\square$

**Observação 2.1.1.** A hipótese  $x_0 \in \text{int}(\text{dom } f)$  é essencial: em pontos de fronteira do domínio o subdiferencial pode ser vazio.

Para as funções convexas e s.c.i., a transformada de Legendre é um operador involutivo, em outras palavras, definindo o biconjugado  $f^{**}$  como

$$f^{**}(x) \stackrel{\text{def.}}{=} \sup_{p \in X^*} \langle p, x \rangle - f^*(p),$$

o teorema de Moreau nos diz que  $f^{**} = f$  para tais funções.

**Teorema 2.1.2** (Teorema de Moreau). *Seja  $f \in \Gamma_0(X)$ . Então  $f^{**} = f$ .*

*Demonstração.* A desigualdade  $f^{**} \leq f$  é imediata da definição de  $f^*$ . Fixemos  $x \in \text{dom } f$ , tomado o supremo dentre todos os  $p \in X^*$  na desigualdade de Fenchel, temos que

$$f(x) \geq \sup_{p \in X^*} \langle p, x \rangle - f^*(p) = f^{**}(x).$$

Vamos agora provar a desigualdade inversa, que  $f \leq f^{**}$ . Como  $f$  é uma função convexa, o seu epígrafo

$$\text{epi}(f) \stackrel{\text{def.}}{=} \{(x, t) \in \text{dom } f \times \mathbb{R} : t \geq f(x)\}$$

é também um conjunto convexo. Fixemos então um ponto  $x_0$  tal que  $f(x_0) < +\infty$ .

Dessa forma, para todo  $\varepsilon > 0$  o ponto  $(x_0, f(x_0) - \varepsilon) \notin \text{epi}(f)$ . Pela forma geométrica do Teorema de Hahn-Banach, existe um hiperplano que separa estritamente  $\text{epi}(f)$  e  $(x_0, f(x_0) - \varepsilon)$ . Ou seja, existem  $p \in X^*$ ,  $\alpha, \beta$  tais que

$$\langle p, x \rangle + \alpha f(x) \geq \beta > \langle p, x_0 \rangle + \alpha(f(x_0) - \varepsilon),$$

para todo  $x \in \text{dom}(f)$ . Tomando  $x = x_0$  do lado esquerdo, concluímos que  $\alpha$  deve ser estritamente positivo. Logo, dividindo ambos os lados por  $\alpha$  podemos assumir sem perda de generalidade que  $\alpha = 1$ .

Tomando  $q = -p$ , temos que para todo  $x \in \text{epi}(f)$

$$\langle q, x \rangle - f(x) \leq -\beta,$$

o que implica, tomado o supremo em  $x \in X$ , que

$$f^*(q) = \sup_{x \in X} \langle q, x \rangle - f(x) \leq -\beta < \langle q, x_0 \rangle - f(x_0) + \varepsilon.$$

Reorganizando os termos e tomado o supremo em  $q$ , obtemos que

$$f(x_0) - \varepsilon \leq \sup_{q \in X^*} \langle q, x_0 \rangle - f^*(q) = f^{**}(x_0).$$

Como  $\varepsilon > 0$  era arbitrário, o resultado segue.  $\square$

Estes resultados fornecem as ferramentas necessárias para abordarmos a dualidade de Kantorovich, primeiro no caso compacto como aplicação direta, e depois no caso geral através da teoria de  $c$ -convexidade.

## 2.2 Dualidade via Convexidade

Podemos agora propor uma prova elementar da dualidade de Kantorovicht utilizando as ferramentas de análise convexa desenvolvidas. Nossa objetivo é de introduzir as ideias principais nesse caso mais simples, para depois desenvolvermos a teoria completa usando a noção de monotonicidade  $c$ -cíclica.

Vamos chamar o problema de minimização em termos de planos de transporte de *problema primal*,

$$\min_{\gamma \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\gamma(x, y), \quad (\text{P})$$

enquanto o problema de maximização em termos de funções potenciais de *problema dual*

$$\sup_{\substack{\varphi \oplus \psi \leq c \\ \varphi \in \mathcal{C}(\mathcal{X}), \psi \in \mathcal{C}(\mathcal{Y})}} \int_{\mathcal{X}} \varphi(x) d\mu(x) + \int_{\mathcal{Y}} \psi(y) d\nu(y). \quad (\text{D})$$

O nosso objetivo então é provar que  $\min(\text{P}) = \max(\text{D})$ , e que ambos os problemas admitem soluções. Chamamos de *dualidade fraca* a desigualdade mais fácil de se obter pelo procedimento chamado de *dualização das restrições* que consiste em reescrever a função indicatriz  $\chi_{\Pi(\mu, \nu)}$  na sua forma dual.

**Lema 2.2.1.** *Dados dois espaços Poloneses  $\mathcal{X}$  e  $\mathcal{Y}$  e uma função custo  $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$ , para toda medida  $\gamma \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$  vale que*

$$\sup_{\varphi \in \mathcal{C}_b(\mathcal{X}), \psi \in \mathcal{C}_b(\mathcal{Y})} \left( \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu - \int_{\mathcal{X} \times \mathcal{Y}} \varphi \oplus \psi d\gamma \right) = \chi_{\Pi(\mu, \nu)}(\gamma) = \begin{cases} 0, & \text{se } \gamma \in \Pi(\mu, \nu), \\ +\infty, & \text{se não.} \end{cases}$$

Como consequência disso, vale a dualidade fraca:

$$\min(\text{P}) \geq \sup(\text{D}).$$

*Demonstração.* Primeiramente note que se  $\gamma \in \Pi(\mu, \nu)$ , então para todo par de funções  $\varphi \in \mathcal{C}_b(\mathcal{X})$  e  $\psi \in \mathcal{C}_b(\mathcal{Y})$  vale

$$\int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu - \int_{\mathcal{X} \times \mathcal{Y}} \varphi \oplus \psi d\gamma = 0$$

pois as marginais de  $\gamma$  são  $\mu$  e  $\nu$ . Portanto o supremo é igual a zero.

Se não for o caso, podemos definir a medida de Radon com sinal

$$\theta \stackrel{\text{def.}}{=} \mu \otimes \nu - (\pi_{0,\#}\gamma) \otimes (\pi_{1,\#}\gamma), \text{ que admite a decomposição de Hahn-Jordan } \theta = \theta_+ - \theta_-,$$

onde  $\theta_+, \theta_-$  são medidas de Radon positivas com suportes disjuntos. Podemos assumir sem perda de generalidade que  $\theta_+ \neq 0$ , à menos de mudança de sinal de  $\theta$ .

Logo existe um conjunto fechado  $\Sigma$  tal que  $\theta_+(\Sigma) > 0 = \theta_-(\Sigma)$ , assim como funções  $\varphi \in \mathcal{C}_b(\mathcal{X})$  e  $\psi \in \mathcal{C}_b(\mathcal{Y})$  tais que  $\varphi \oplus \psi$  é concentrada em  $\Sigma$ .<sup>2</sup> Dessa forma temos que

$$\int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu - \int_{\mathcal{X} \times \mathcal{Y}} \varphi \oplus \psi d\gamma = \int_{\mathcal{X} \times \mathcal{Y}} \varphi \oplus \psi d\theta > 0.$$

Substituindo  $\varphi$  e  $\psi$  por  $\lambda\varphi$  e  $\lambda\psi$  para  $\lambda > 0$ , e fazendo  $\lambda \rightarrow +\infty$  vemos que o supremo é infinito.

Passamos agora à prova da dualidade fraca. Com essa caracterização da função indicatriz  $\chi_{\Pi(\mu, \nu)}$ , temos que

$$\begin{aligned} \inf(P) &= \inf_{\gamma \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})} \int_{\mathcal{X} \times \mathcal{Y}} c d\gamma + \chi_{\Pi(\mu, \nu)}(\gamma) \\ &= \inf_{\gamma \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})} \sup_{\varphi \in \mathcal{C}_b(\mathcal{X}), \psi \in \mathcal{C}_b(\mathcal{Y})} \int_{\mathcal{X} \times \mathcal{Y}} (c - \varphi \oplus \psi) d\gamma + \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu \end{aligned}$$

O supremo então domina essa mesma quantidade avaliada em qualquer par de funções  $(\varphi, \psi)$  fixadas, e portanto para qualquer tal par temos

$$\inf(P) \geq \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu + \inf_{\gamma \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})} \int_{\mathcal{X} \times \mathcal{Y}} (c - \varphi \oplus \psi) d\gamma. \quad (2.7)$$

Temos então que estudar esse ínfimo em  $\gamma$  da mesma forma que fizemos na primeira parte para  $\chi_{\Pi(\mu, \nu)}$ . Se  $\varphi \oplus \psi \leq c$ , o integrando é não negativo e portanto o ínfimo é limitado inferiormente por 0. Mas essa cota inferior pode ser facilmente atingida por  $\gamma \equiv 0$ , de forma que o ínfimo é 0. Se não for o caso, podemos encontrar um ponto  $(x_0, y_0)$  tal que  $\varphi(x_0) + \psi(y_0) > c(x_0, y_0)$ , e então tomando  $\gamma = \lambda\delta_{(x_0, y_0)}$  e fazendo  $\lambda \rightarrow +\infty$  vemos que o ínfimo é  $-\infty$ . Portanto

$$\inf_{\gamma \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})} \int_{\mathcal{X} \times \mathcal{Y}} (c - \varphi \oplus \psi) d\gamma = \begin{cases} 0, & \text{se } \varphi \oplus \psi \leq c, \\ -\infty, & \text{se não.} \end{cases}$$

Voltando para (2.7), como o lado direito não depende de  $\gamma$ , temos que o inf do problema primal majora o supremo do lado direito entre todas as funções  $\varphi, \psi$ . Mas veja que pelo argumento que acabamos de realizar, esse supremo não muda se considerarmos apenas funções  $\varphi \oplus \psi \leq c$ , pois se não o lado direito é  $-\infty$  e a desigualdade é trivialmente satisfeita. Portanto

$$\inf(P) \geq \sup_{\substack{\varphi \oplus \psi \leq c \\ \varphi \in \mathcal{C}_b(\mathcal{X}), \psi \in \mathcal{C}_b(\mathcal{Y})}} \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu = \sup(D),$$

e a dualidade fraca segue. □

---

<sup>2</sup>Por exemplo, podemos considerar uma função sobre  $\mathcal{X} \times \mathcal{Y}$  que é igual a 1 em um conjunto compacto contido em  $\Sigma$  e nula fora de um conjunto aberto que contém esse conjunto compacto, e então aplicar o teorema de Tietze para obter uma função contínua sobre o produto. Em seguida usar Stone-Weierstrass para aproximar essa função por somas de funções do tipo  $\varphi \oplus \psi$ .

Ou mais simples ainda, tome  $\varphi$  e  $\psi$  como funções contínuas que são iguais a 1 em um conjunto compacto contido em  $\pi_0(\Sigma)$  e  $\pi_1(\Sigma)$  respectivamente, e que são nulas fora de um conjunto aberto que contém esses conjuntos compactos.

Vamos agora propor uma prova elementar da dualidade forte<sup>3</sup>, baseada no teorema de Moreau da análise convexa. Para isso, considere o funcional  $H : \mathcal{C}_0(\mathcal{X} \times \mathcal{Y}) \mapsto \mathbb{R} \cup \{+\infty\}$  definido como

$$H(p) \stackrel{\text{def.}}{=} - \sup_{\varphi \in \mathcal{C}_b(\mathcal{X}), \psi \in \mathcal{C}_b(\mathcal{Y})} \left\{ \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu : \varphi \oplus \psi \leq c - p \right\}.$$

Dessa forma temos que  $H(0) = -\sup(D)$ , e a provando que  $H$  é convexa e s.c.i., podemos aplicar o teorema de Moreau, na esperança que a relação  $H(0) = H^{**}(0)$  corresponda à dualidade forte.

**Teorema 2.2.1.** *Sejam  $\mathcal{X}$  e  $\mathcal{Y}$  espaços Poloneses e seja  $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$  uma função custo semi-contínua inferiormente tal que  $\inf(P) < +\infty$ . Então as seguintes afirmações são verdadeiras:*

1.  *$H$  é convexa e s.c.i. sobre  $\mathcal{C}_b(\mathcal{X} \times \mathcal{Y})$ , com respeito à convergência uniforme;*
2. *A transformada de Legendre de  $H$  é dada por*

$$H^*(\gamma) = \sup_{p \in \mathcal{C}_b(\mathcal{X} \times \mathcal{Y})} \langle p, \gamma \rangle - H(p) = \begin{cases} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\gamma(x, y), & \text{se } \gamma \in \Pi(\mu, \nu), \\ +\infty, & \text{se não;} \end{cases}$$

3. *Temos dualidade forte.*

*Demonstração.* (1) Para provar a convexidade de  $H$ , sejam  $p_0, p_1 \in \mathcal{C}_b(\mathcal{X} \times \mathcal{Y})$  e  $t \in [0, 1]$ . Tome pares  $(\varphi_i, \psi_i)$  admissíveis para  $H(p_i)$ , com  $i = 0, 1$  e defina

$$(\varphi_t, \psi_t) \stackrel{\text{def.}}{=} (1-t)(\varphi_0, \psi_0) + t(\varphi_1, \psi_1), \quad p_t \stackrel{\text{def.}}{=} (1-t)p_0 + tp_1.$$

Logo temos que  $\varphi_t \oplus \psi_t \leq c - p_t$  e, portanto,

$$-H(p_t) \geq (1-t) \left[ \int_{\mathcal{X}} \varphi_0 d\mu + \int_{\mathcal{Y}} \psi_0 d\nu \right] + t \left[ \int_{\mathcal{X}} \varphi_1 d\mu + \int_{\mathcal{Y}} \psi_1 d\nu \right].$$

Tomando o supremo sobre todos os pares admissíveis  $(\varphi_i, \psi_i)$  para  $i = 0, 1$ , primeiro para  $i = 0$  depois para  $i = 1$ , segue a convexidade de  $H$ .

Para provar que  $H$  é s.c.i., seja  $p_n \xrightarrow[n \rightarrow \infty]{\|\cdot\|_\infty} p$ . Fixemos  $\varepsilon > 0$  arbitrário, e tome usando a definição de supremo, um par  $\varphi_n \oplus \psi_n \leq c - p_n$  tal que

$$\int_{\mathcal{X}} \varphi_n d\mu + \int_{\mathcal{Y}} \psi_n d\nu \geq -H(p_n) - \varepsilon.$$

Defina o novo par admissível  $(\bar{\varphi}_n, \psi_n)$  com  $\bar{\varphi}_n = \varphi_n - \delta_n$  com  $\delta_n = \|p_n - p\|_\infty$ , de formas que  $\bar{\varphi}_n \oplus \psi_n \leq c - p$ . Logo

$$\int_{\mathcal{X}} \bar{\varphi}_n d\mu + \int_{\mathcal{Y}} \psi_n d\nu \geq -H(p_n) - \varepsilon - \delta_n.$$

---

<sup>3</sup>Veja por exemplo “Optimal Transport for applied mathematicians”, sessão 1.6.3; note que na prova dessa referência, usamos a existência de potenciais de Kantorovitch em domínios compactos. O que não é feito aqui, mas a ideia continua muito parecida.

Dessa forma,  $(\bar{\varphi}_n, \psi_n)$  se torna admissível para  $H(p)$  e nós temos que

$$-H(p) \geq \int_{\mathcal{X}} \bar{\varphi}_n d\mu + \int_{\mathcal{Y}} \psi_n d\nu \geq -H(p_n) - \varepsilon - \delta_n.$$

Tomando o  $\liminf$  quando  $n \rightarrow \infty$  obtemos

$$H(p) \leq \liminf_{n \rightarrow \infty} H(p_n) + \varepsilon.$$

Como  $\varepsilon > 0$  era arbitrário, a semicontinuidade inferior de  $H$  segue.

(2) Vamos agora calcular a transformada de Legendre de  $H$ . Seja  $\gamma \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$ , então

$$\begin{aligned} H^*(\gamma) &= \sup_p \langle p, \gamma \rangle - H(p) \\ &= \sup_p \langle p, \gamma \rangle + \sup_{\varphi \oplus \psi \leq c-p} \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu \\ &= \sup_p \langle c, \gamma \rangle + \sup_{\varphi \oplus \psi \leq c-p} \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu - \int_{\mathcal{X} \times \mathcal{Y}} (c-p) d\gamma \\ &= \begin{cases} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\gamma(x, y), & \text{se } \gamma \in \Pi(\mu, \nu), \\ +\infty, & \text{se não,} \end{cases} \end{aligned}$$

onde a última igualdade segue do Lema 2.2.1 aplicado à medida  $\gamma$ .

(3) Finalmente, aplicando o teorema de Moreau, temos que

$$\sup_{\varphi \oplus \psi} \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu = -H(0) = -H^{**}(0) = \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c d\gamma,$$

e a dualidade forte segue.  $\square$

Esse argumento, apesar de simples e geral, não nos fornece existência de potenciais de Kantorovich e que por sua vez, não podem ser empregados para extrair mais informação do plano de transporte ótimo. Por outro lado, dado um par  $(\varphi, \psi)$  admissível, podemos reescrever as restrições como: dado  $y \in \mathcal{Y}$  devemos ter que

$$\psi(y) \leq c(x, y) - \varphi(x) \text{ para todo } x \in \mathcal{X},$$

o que nos indica que o maior valor possível para  $\psi(y)$  deve ser dado pelo ínfimo das quantidades a direita, e não menos do que isso. Isso motiva a definição de  $c$ -transformada, que é o análogo da transformada de Legendre na teoria de transporte ótimo.

**Definição 2.2.1** ( $c$ -Transformada). Sejam  $\mathcal{X}, \mathcal{Y}$  conjuntos e  $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  uma função custo. Dada  $\varphi : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ , a transformada- $c$  de  $\varphi$  é a função  $\varphi^c : \mathcal{Y} \rightarrow \mathbb{R} \cup \{\pm\infty\}$  dada por

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

Por outro lado, a transformada  $c$  de uma função  $\psi : \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$ , é a função  $\psi^c : \mathcal{X} \rightarrow \mathbb{R} \cup \{\pm\infty\}$  dada por

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

Uma função  $\psi$  é dita  $c$ -cônica se existe  $\varphi$  tal que  $\psi = \varphi^c$ .

A motivação para a definição de funções  $c$ -côncavas vem do teorema de Moreau da análise convexa. De fato, uma consequência imediata desse teorema é que uma função  $\varphi$  é convexa se, e somente se, ela é a transformada de Legendre de uma outra função.

As seguintes propriedades da transformada  $c$  seguem diretamente da definição.

**Proposição 2.2.1.** *Sejam  $\mathcal{X}$  e  $\mathcal{Y}$  espaços poloneses. As seguintes propriedades são satisfeitas*

1. *Seja um par  $(\varphi, \psi)$  tal que  $\varphi \oplus \psi \leq c$ , então  $\psi \leq \varphi^c$ ;*
2. *Se  $c \in \mathcal{C}_b(\mathcal{X} \times \mathcal{Y})$  possui um módulo de continuidade, i.e.*

$$|c(x, y) - c(x', y')| \leq \omega_{\mathcal{X}}(d_{\mathcal{X}}(x, x')) + \omega_{\mathcal{Y}}(d_{\mathcal{Y}}(y, y')),$$

*então quaisquer funções  $c$ -côncavas  $\psi^c, \varphi^c$  tem módulos de continuidade  $\omega_{\mathcal{X}}$  e  $\omega_{\mathcal{Y}}$ .*

*Demonstração.* Propriedade (1) segue diretamente da definição. Para provar (2), considere  $y, y' \in \mathcal{Y}$ , dado  $\varepsilon > 0$ , pela definição da transformada  $c$ , deve existir  $x'$  tal que

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

Logo usando  $x'$  no ínfimo que define  $\varphi^c(y)$  temos que

$$\begin{aligned} \varphi^c(y) - \varphi^c(y') &\leq c(x', y) - \varphi(x') - (c(x', y') - \varphi(x') + \varepsilon) = c(x', y) - c(x', y') + \varepsilon \\ &\leq \omega_{\mathcal{Y}}(d_{\mathcal{Y}}(y, y')) + \varepsilon. \end{aligned}$$

Fazendo  $\varepsilon \rightarrow 0$  e trocando os papéis de  $y, y'$ , obtemos que

$$|\varphi^c(y) - \varphi^c(y')| \leq \omega_{\mathcal{Y}}(d_{\mathcal{Y}}(y, y')).$$

O mesmo argumento mostra que o módulo de continuidade de  $\psi^c$  é dado por  $\omega_{\mathcal{X}}$ . □

Usando esses dois resultados, junto de teoremas clássicos de compacidade de funções contínuas, ou seja, o teorema de Ascoli-Arzelà.

**Teorema 2.2.2** (Ascoli-Arzelà). *Seja  $\mathcal{X}$  um espaço polônés compacto. Um conjunto  $\mathcal{F} \subset \mathcal{C}(\mathcal{X})$  fechado na topologia induzida pela convergência uniforme é sequencialmente compacto se, e somente se*

- $\mathcal{F}$  é equilimitada, i.e. existe  $C > 0$  tal que  $\|f\|_{\infty} \leq C$  para toda  $f \in \mathcal{F}$ ;
- $\mathcal{F}$  é equicontínua, i.e. para todo  $\varepsilon > 0$  existe  $\delta$  tal que

$$\text{se } d_{\mathcal{X}}(x, y) < \delta, \text{ então } |f(x) - f(y)| \leq \varepsilon.$$

*Ou equivalentemente, existe um módulo de continuidade  $\omega$  comum à todas as funções  $f \in \mathcal{F}$ .*

Com isso, podemos provar que existe um par de potenciais de Kantorovich  $(\varphi, \psi)$ .

**Exercício 2.4.** Dados  $\mathcal{X}$  e  $\mathcal{Y}$  espaços Poloneses compactos e  $c$  uma função custo contínua. Prove que existe um par de potenciais de Kantorovich  $(\varphi, \psi)$  tais que  $\psi = \varphi^c$  e  $\varphi$  é uma função  $c$ -côncava. *[Dica: Para aplicar o método direto, construa uma sequência maximizante equicontínua usando a transformada  $c$ .]*

## 2.3 O caso geral: conjuntos c-ciclicamente monótonos

Na nossa intuição econômica anterior imagine o seguinte: dados um conjunto de fornecedores  $\mathcal{X}$  e um conjunto de consumidores  $\mathcal{Y}$ , suponha que temos um plano de transporte  $\gamma$  que associa cada fornecedor a um consumidor, e que esse plano é ótimo. Agora, considere uma troca cíclica entre vários pares  $(x_i, y_i)$  no suporte de  $\gamma$ , onde cada fornecedor  $x_i$  passa a fornecer para o consumidor  $y_{i+1}$  (com  $y_{n+1} = y_1$ ). A variação do custo total de transporte nessas situações é dada por

$$\sum_{i=1}^n c(x_i, y_{i+1}) - \sum_{i=1}^n c(x_i, y_i)$$

Se essa variação for estritamente positiva, então o plano original não seria ótimo. Portanto, para qualquer ciclo de trocas, o custo total não pode diminuir. Isso motiva a seguinte definição:

**Definição 2.3.1** (c-Monotonicidade Cíclica). Um conjunto  $\Gamma \subset \mathcal{X} \times \mathcal{Y}$  é *c-ciclicamente monótono* se para todo  $n \in \mathbb{N}$  e toda sequência finita  $\{(x_i, y_i)\}_{i=1}^n \subset \Gamma$ , vale

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

onde convencionamos  $y_{n+1} = y_1$ .

Para formalizar a intuição anterior nos diz que se  $\gamma$  é ótimo, então  $\gamma$ -quase todo par  $(x, y)$  deve estar contido em um conjunto c-CM. No entanto, veja que essa intuição não é completamente rigorosa, podemos incluir a esse conjunto uma quantidade não enumerável de pares  $(x, y)$  que  $\gamma$  não vê e quebrar a propriedade de ser c-CM. É mais seguro introduzir a noção de suporte de uma medida.

**Definição 2.3.2.** Seja  $\mathcal{X}$  um espaço polonês. O suporte de uma medida  $\mu \in \mathcal{P}(\mathcal{X})$  é o menor conjunto fechado  $S \subset \mathcal{X}$  tal que  $\mu(S) = 1$ . Ou ainda, podemos definir

$$\text{supp } \mu \stackrel{\text{def.}}{=} \{x \in \mathcal{X} : \mu(B_r(x)) > 0 \text{ para todo } r > 0\}.$$

Se um conjunto  $A$  é tal que  $\mu(\mathcal{X} \setminus A) = 0$ , então  $\text{supp } \mu \subset \overline{A}$ . Além disso, podemos construir novas medidas de Radon fazendo restrições ao seu suporte. Introduzimos a notação  $\mu \llcorner A$  como a medida definida por

$$\mu \llcorner A(B) = \mu(A \cap B) \text{ para todo boreliano } B.$$

**Proposição 2.3.1.** Sejam  $\mathcal{X}$  e  $\mathcal{Y}$  espaços poloneses,  $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  uma função custo contínua e  $\mu \in \mathcal{P}(\mathcal{X})$ ,  $\nu \in \mathcal{P}(\mathcal{Y})$ . Se  $\gamma$  é um plano de transporte ótimo para o problema de Kantorovich com custo  $c$  e marginais  $\mu$  e  $\nu$ , então o suporte de  $\gamma$  é c-ciclicamente monótono.

*Demonstração.* Seja  $\gamma$  um plano de transporte ótimo e assuma por contradição que  $\text{supp } \gamma$  não é c-CM. Então existe uma sequência finita de pontos  $\{(x_i, y_i)\}_{i=1}^n \subset \text{supp } \gamma$  tais que

$$\sum_{i=1}^n c(x_i, y_i) - \sum_{i=1}^n c(x_i, y_{i+1}) > \delta. \tag{2.8}$$

Nosso objetivo é de contradizer a otimalidade de  $\gamma$  escolhendo uma pequena região ao redor dos pontos do suporte  $\{(x_i, y_i)\}_{i=1}^n$  onde podemos alterar a distribuição de massa de  $\gamma$  e concentrá-la no

novo ciclo. Como  $c$  é um custo contínuo, existem vizinhanças abertas  $U_i \times V_i$  para cada sequência  $(x'_i, y'_i) \in U_i \times V_i$  vale que

$$\sum_{i=1}^n c(x'_i, y'_i) - \sum_{i=1}^n c(x'_i, y'_{i+1}) > \delta/2.$$

Logo podemos definir para todo  $i$  as medidas

$$\gamma_i \stackrel{\text{def.}}{=} \frac{\gamma \llcorner (U_i \times V_i)}{m_i}, \quad m_i \stackrel{\text{def.}}{=} \gamma(U_i \times V_i) > 0.$$

Estas estão todas bem definidas e não triviais pois como  $(x_i, y_i) \in \text{supp } \gamma$ , temos  $\gamma(U_i \times V_i) > 0$ . Como falamos no capítulo anterior, para qualquer espaço de medida  $(\Omega, \mathcal{F}, \mathbb{P})$  tal que  $\mathbb{P}$  não possui átomos, existem mapas  $(X_i, Y_i)$  tais que  $(X_i, Y_i)_\sharp \mathbb{P} = \gamma_i$ . Em outras palavras, sejam  $(X_i, Y_i)$  variáveis aleatórias com distribuição conjunta  $\gamma_i$ . Definimos então a nova medida

$$\bar{\gamma} \stackrel{\text{def.}}{=} \gamma + \varepsilon \sum_{i=1}^n [(X_i, Y_{i+1})_\sharp \mathbb{P} - (X_i, Y_i)_\sharp \mathbb{P}].$$

Logo  $\bar{\gamma} \in \Pi(\mu, \nu)$  pois além de ter as boas marginais,  $\gamma - \varepsilon \sum_{i=1}^n \gamma_i \geq 0$  para  $\varepsilon$  suficientemente pequeno, por construção. Por outro lado

$$\int_{\mathcal{X} \times \mathcal{Y}} c d\bar{\gamma} = \int_{\mathcal{X} \times \mathcal{Y}} c d\gamma + \varepsilon \int_{\Omega} \left( \sum_{i=1}^n c(X_i, Y_{i+1}) - c(X_i, Y_i) \right) d\mathbb{P} \leq \int_{\mathcal{X} \times \mathcal{Y}} c d\gamma - \varepsilon \frac{\delta}{2},$$

o que contradiz a otimalidade de  $\gamma$ .  $\square$

Vamos provar que os conjuntos  $c$ -ciclicamente monotônicos são exatamente os conjuntos de pontos de contato entre uma função  $c$ -convexa e o custo  $c$ . Em analogia ao subdiferencial de funções convexas, definimos o  $c$ -subdiferencial como

$$\partial^c \varphi(x) \stackrel{\text{def.}}{=} \left\{ y \in \mathcal{Y} : x \in \operatorname{argmin}_{x'} c(x', y) - \varphi(x') \right\}, \quad (2.9)$$

assim como o gráfico do operador subdiferencial que se torna

$$\operatorname{Graph} \partial^c \varphi \stackrel{\text{def.}}{=} \{(x, y) : y \in \partial^c \varphi(x)\} = \{(x, y) : \varphi(x) + \varphi^c(y) = c(x, y)\}. \quad (2.10)$$

Com essas definições, provamos o seguinte resultado.

**Proposição 2.3.2.** *Sejam  $\mathcal{X}, \mathcal{Y}$  espaços Poloneses  $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  uma função custo semi-contínua inferiormente. Então se  $\Gamma \subset \mathcal{X} \times \mathcal{Y}$  é um conjunto  $c$ -CM, existe uma função  $c$ -côncava  $\varphi : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$  tal que*

$$\Gamma \subset \operatorname{Graph} \partial^c \varphi.$$

*Demonstração.* Note que a condição  $\Gamma \subset \operatorname{Graph} \partial^c \varphi$  é equivalente à

$$\varphi(x) \leq c(x, \bar{y}) - c(\bar{x}, \bar{y}) + \varphi(\bar{x}), \quad \text{para todo } (\bar{x}, \bar{y}) \in \Gamma, \quad x \in \mathcal{X}. \quad (2.11)$$

De fato, temos que

$$\begin{aligned}\Gamma \subset \text{Graph } \partial^c \varphi &\iff \bar{y} \in \partial^c \varphi(\bar{x}), \text{ para todos } (\bar{x}, \bar{y}) \in \Gamma \\ &\iff c(\bar{x}, \bar{y}) - \varphi(\bar{x}) \leq c(x, \bar{y}) - \varphi(x), \text{ para todos } (\bar{x}, \bar{y}) \in \Gamma, x \in \mathcal{X}.\end{aligned}$$

Fixando um ponto  $x_0$  para o qual queremos construir  $\varphi$  tal que  $\varphi(x_0)$  satisfazendo (2.11), temos por um argumento induutivo que

$$\varphi(x) \leq c(x, y_N) - c(x_N, y_N) + \sum_{i=0}^{N-1} c(x_{i+1}, y_i) - c(x_i, y_i),$$

para toda sequência de pontos  $(x_i, y_i)_{i=0}^N \subset \Gamma$ .

Em particular, podemos definir  $\varphi$  como o ínfimo de todas essas quantidades

$$\varphi(x) \stackrel{\text{def.}}{=} \inf_{(x_i, y_i)_{i=0}^N \subset \Gamma} \left\{ c(x, y_N) - c(x_N, y_N) + \sum_{i=0}^{N-1} c(x_{i+1}, y_i) - c(x_i, y_i) \right\}.$$

Com essa definição, é fácil ver que

$$\varphi(x) = \inf_{(x_N, y_N) \in \Gamma} c(x, y_N) - c(x_N, y_N) + \phi(x_N), \quad (2.12)$$

basta inicialmente considerar o ínfimo na definição de  $\varphi$  com  $(x_N, y_N)$  fixos e em seguida minimizar em  $(x_N, y_N)$ . Além disso, pela monotonicidade  $c$ -cíclica, temos que  $\varphi(x_0) \geq 0$ ; enquanto que para obter  $\varphi(x_0) \leq 0$ , basta tomar  $N = 1$  e  $(x_0, y_0) = (x_1, y_1)$ .

Para concluir a prova, temos apenas que provar que existe uma função  $\psi(y)$  tal que  $\varphi = \psi^{\bar{c}}$ . Para, basta definir

$$-\psi(y) \stackrel{\text{def.}}{=} \inf_{(x_i, y_i)_{i=0}^N \subset \Gamma} \left\{ -c(x_N, y) + \sum_{i=0}^{N-1} c(x_{i+1}, y_i) - c(x_i, y_i) \right\}.$$

De (2.12), temos que  $\varphi = \psi^{\bar{c}}$ . □

Os resultados de base para provar a dualidade forte e a existência de potenciais de Kantorovitch são então as proposições 2.3.1 e 3.3.1. Mesmo que estas proposições sejam formuladas para custos contínuos, podemos obter o caso geral por aproximação.

**Teorema 2.3.1** (Dualidade de Kantorovich - Caso Geral). *Sejam  $\mathcal{X}, \mathcal{Y}$  espaços poloneses, fixemos  $\mu \in \mathcal{P}(\mathcal{X})$ ,  $\nu \in \mathcal{P}(\mathcal{Y})$  e uma função custo  $c : \mathcal{X} \times \mathcal{Y} \rightarrow [0, +\infty]$  e semi-contínua inferiormente tal que existam funções  $a \in L^1(\mu)$ ,  $b \in L^1(\nu)$  tais que*

$$c(x, y) \leq a(x) + b(y) \text{ o que implica } \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} cd\gamma < +\infty. \quad (2.13)$$

*Então as seguintes afirmações são verdadeiras*

1. Há dualidade forte

$$\begin{aligned} \min_{\gamma \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c d\gamma &= \sup_{\substack{\varphi \oplus \psi \leq c \\ \varphi \in \mathcal{C}_b(\mathcal{X}), \psi \in \mathcal{C}_b(\mathcal{Y})}} \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu \\ &= \max_{\substack{\varphi \oplus \psi \leq c \\ \varphi \in L^1(\mu), \psi \in L^1(\nu)}} \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu. \end{aligned} \quad (2.14)$$

Ou seja, há dualidade forte e o supremo é atingido por funções integráveis.

2. Um plano de transporte  $\gamma \in \Pi(\mu, \nu)$  é ótimo se, e somente se,  $\gamma$  está concentrada em um conjunto  $c$ -ciclicamente monótono.

Em outras palavras,  $\gamma$  é ótimo se, e somente se, existe uma função  $c$ -côncava  $\varphi$  tal que

$$c(x, y) = \varphi(x) + \varphi^c(y) \quad \text{para } \gamma\text{-quase todo } (x, y) \in \mathcal{X} \times \mathcal{Y}. \quad (2.15)$$

Nesse caso, o par de potenciais  $(\varphi, \varphi^c)$  é ótimo para o problema dual de Kantorovich.

*Demonação.* Vamos começar provando o item (1), no caso em que  $c \in \text{Lip}_b$ . Tome um plano de transporte ótimo  $\bar{\gamma}$  e aplique as proposições 2.3.1 e 2.3.2 em conjunto para obter uma função  $c$ -côncava  $\varphi$  tal que

$$\text{supp } \gamma \subset \{(x, y) \in \mathcal{X} \times \mathcal{Y} : \varphi(x) + \varphi^c(y) = c(x, y)\}.$$

Pela construção de  $\varphi$  como um  $c$ -transformada,  $\varphi$  herda a regularidade de  $c$ , ou seja,  $\varphi \in \text{Lip}_b(\mathcal{X})$ . Logo vale que

$$\min_{\gamma \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c d\gamma = \int_{\mathcal{X} \times \mathcal{Y}} c d\bar{\gamma} = \int_{\mathcal{X} \times \mathcal{Y}} \varphi(x) + \varphi^c(y) d\bar{\gamma} = \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \varphi^c d\nu$$

Como a dualidade fraca sempre é válida, temos que

$$\int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \varphi^c d\nu = \min(P) \geq \sup(D)$$

e logo  $\varphi$  e  $\varphi^c$  são potenciais de Kantorovich ótimos.

*Dualidade para  $c$  s.c.i.:* Pelo Lema 1.5.1 existe uma sequência  $(c_k)_{k \in \mathbb{N}}$  monótona crescente de funções  $k$ -Lipschitz e limitadas que convergem pontualmente para  $c$ . Para simplificar a notação, definimos

$$\mathcal{C}_k(\gamma) \stackrel{\text{def.}}{=} \begin{cases} \int_{\mathcal{X} \times \mathcal{Y}} c_k d\gamma, & \gamma \in \Pi(\mu, \nu), \\ +\infty, & \text{se não,} \end{cases}$$

e analogamente  $\mathcal{C}(\cdot)$ .

Usando a dualidade forte no caso Lipschitz, para cada  $k \in \mathbb{N}$  existe um plano de transporte ótimo  $\gamma_k$  e um par de potenciais de Kantorovich  $(\varphi_k, \psi_k)_{k \in \mathbb{N}}$  tais que  $\varphi_k \oplus \psi_k \leq c_k \leq c$ . Portanto, como a dualidade fraca é sempre válida, temos que

$$\begin{aligned} \min \mathcal{C} &\geq \sup_{\varphi \oplus \psi \leq c} \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu \geq \sup_{\varphi \oplus \psi \leq c_k} \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu \\ &= \int_{\mathcal{X}} \varphi_k d\mu + \int_{\mathcal{Y}} \psi_k d\nu = \min \mathcal{C}_k \end{aligned}$$

Dessa forma, para provar a dualidade forte, bastar provar que  $\liminf_{k \rightarrow \infty} \min \mathcal{C}_k \geq \min \mathcal{C}$ .

Para isso, note que  $(\gamma_k)_{k \in \mathbb{N}} \subset \Pi(\mu, \nu)$ , pois o conjunto de restrições é o mesmo para toda a sequência de problemas primais. Além disso, ele é sequencialmente compacto pelo teorema de Prokhorov, logo existe uma subsequência de  $\gamma_k$  (não renomeada) que converge na topologia estreita para  $\bar{\gamma}$ . Por outro lado, como  $c_k$  é uma sequência monótona crescente, fixemos um valor  $n \in \mathbb{N}$  de forma que para todo  $k > n$  temos  $c_k \geq c_n$  e, portanto,

$$\liminf_{k \rightarrow \infty} \min \mathcal{C}_k \geq \liminf_{k \rightarrow \infty} \int_{\mathcal{X} \times \mathcal{Y}} c_n d\gamma_k \geq \int_{\mathcal{X} \times \mathcal{Y}} c_n d\bar{\gamma}.$$

Tomando o limite quando  $n \rightarrow \infty$  no lado direito da desigualdade à cima, pelo teorema da convergência monótona de Beppo-Levi, obtemos

$$\liminf_{k \rightarrow \infty} \min \mathcal{C}_k \geq \lim_{n \rightarrow \infty} \int_{\mathcal{X} \times \mathcal{Y}} c_n d\bar{\gamma} = \int_{\mathcal{X} \times \mathcal{Y}} c d\bar{\gamma} \geq \min \mathcal{C}.$$

Pela discussão anterior a dualidade forte segue no caso  $c$  s.c.i. e  $\mathcal{X}, \mathcal{Y}$  espaços poloneses.

Seguindo para o item (2), começamos provando que todo plano de transporte ótimo  $\gamma$  se concentra num conjunto c-CM. Isso é verdade para o caso  $c \in \text{Lip}_b$  da Proposição 2.3.1. Para provar o caso geral  $c$  s.c.i., note pelo argumento anterior que

$$0 \leq \int_{\mathcal{X} \times \mathcal{Y}} (c - \varphi_k \oplus \psi_k) d\gamma \leq \min \mathcal{C} - \min \mathcal{C}_k \xrightarrow{k \rightarrow \infty} 0.$$

Logo  $c - \varphi_k \oplus \psi_k \xrightarrow[k \rightarrow \infty]{L^1(\gamma)} 0$ , e portanto existe uma subsequência (não renomeada) que converge  $\gamma$ -quase certamente. Definimos  $\Gamma$  como o conjunto onde essa convergência pontual ocorre, de formas que  $\gamma(\Gamma) = 1$ . Para provar que esse conjunto é c-CM, tome uma sequência finita de pontos  $(x_i, y_i)_{i=1}^n \subset \Gamma$ , então

$$\sum_{i=1}^n c(x_i, y_{i+1}) \geq \sum_{i=1}^n \varphi_k(x_i) + \psi_k(y_{i+1}) = \sum_{i=1}^n \varphi_k(x_i) + \psi_k(y_i) \xrightarrow{k \rightarrow \infty} \sum_{i=1}^n c(x_i, y_i),$$

e a convergência ocorre diretamente pela definição do conjunto  $\Gamma$ , e daí segue que este é c-CM.

Por outro lado, note pela hipótese (2.13), todo plano de transporte  $\gamma \in \Pi(\mu, \nu)$  tem custo finito. Logo no suporte de  $\gamma$ , podemos assumir que o custo  $c$  assume valores reais. Dessa forma, se existe um conjunto c-CM onde se concentra  $\gamma$ , pela Proposição 2.3.2, existe uma função  $c$ -cônica  $\varphi$  tal que (2.15) é verdade.

Sendo assim, temos que

$$\sup(D) \geq \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \varphi^c d\nu = \int_{\mathcal{X} \times \mathcal{Y}} c d\gamma \geq \inf(P).$$

Logo  $(\varphi, \varphi^c)$  são potenciais de Kantorovitch ótimos e  $\gamma$  é um plano de transporte ótimo.  $\square$