Universidade Federal do Rio Grande do Sul

Instituto de Matemática

Programa de Pós-Graduação em Matemática

O Problema de Monge-Kantorovich para o Custo Quadrático

Dissertação de Mestrado

Guilherme Ost de Aguiar

Porto Alegre, 15 de julho de 2011

Dissertação submetida por Guilherme Ost de Aguiar[1] como requisito parcial para a obtenção do grau de Mestre em Matemática pelo programa de Pós-Graduação em Matemática do Instituto de Matemática da Universidade Federal do Rio Grande do Sul.

Professor Orientador:

Prof. Dr. Rafael Rigão Souza

Banca Examinadora:

Prof. Dr. Artur Oscar Lopes

Profa. Dra. Joana Mohr

Prof. Dr. Rogério Ricardo Steffenon

Data da Apresentação: 15 de julho de 2011

[1]Bolsista do Conselho Nacional do Desenvolvimento Científico e Tecnológico - CNPq

A Aline,

e a minha família.

Agradecimentos

Em primeiro lugar, agradeço aos meus pais, Danilo e Ledir, por todo apoio, carinho e incentivo dado a mim durante este período.

A Aline pelo afeto e por todos os momentos de alegrias que passamos juntos, assim como pelo companheirismo em todas horas de estudo.

Agradeço ao meu orientador, Rafael Rigão Souza, pela orientação que recebi, pelo comprometimento que tiveste com a minha formação e pelos diversos conselhos.

Aos meus amigos pelos momentos de descontração e pela compreensão nos momentos em que não estive presente.

Resumo

Abordamos o problema do transporte ótimo de Monge-Kantorovich no caso em que o custo é dado pelo quadrado da distância. Tal custo tem uma estrutura que permite a obtenção de resultados mais ricos do que o caso geral. Nosso objetivo é determinar se há soluções para tal problema e caracterizá-las.

Além disso, tratamos informalmente do problema de transporte ótimo para um custo geral.

Palavras-chave: Monge-Kantorovich. Planos de Transporte. A Dualidade de Kantorovich. Subdiferencial.

Abstract

We analyze the Monge-Kantorovich optimal transportation problem in the case where the cost function is given by the square of the Euclidean norm. Such cost has a structure which allow us to get more interesting results than the general case. Our main purpose is to determine if there are solutions to such problem and characterize them.

We also give an informal treatment to the optimal transportation problem in the general case.

Keywords: Monge-Kantorovich. Transference plans. The Kantorovich duality. Subdifferential.

Sumário

Capítulo 1

Introdução

O problema de transporte de massa ótimo foi introduzido em 1781, em um famoso paper publicado por Gaspard Monge: *Mémoire sur la théorie déblais et des remblais.* Desde então, tal problema tornou-se um tema clássico na teoria das probabilidades, na economia e na otimização.

Informalmente, este problema consiste em transportar um volume de areia fixado para uma determinada região de mesmo volume, considerando-se que há um custo atribuído a cada unidade de areia transportada, levando-se em conta a sua localização atual x e sua nova localização y. Assim, a questão central é determinar uma maneira de se realizar o transporte de forma que o custo total seja mínimo.

Embora Monge tenha sido o precursor deste tema, vamos definir o problema de transporte ótimo segundo a abordagem de Leonid V. Kantorovich, e a partir daí obter o problema de minimização de Monge. Kantorovich propôs o problema de transporte ótimo como segue. Inicialmente, consideram-se dois espaços de probabilidade $(X, \mathcal{A}^X, \mu)$ e $(Y, \mathcal{A}^Y, \nu)$, que modelam, respectivamente, a quantidade de areia envolvida e a região de destino, e uma função custo $c : X \times Y \longrightarrow \mathbb{R}^+ \cup \{+\infty\}$ mensurável em relação à σ-álgebra produto $\mathcal{A}^X \times \mathcal{A}^Y$ que nos fornece o custo do transporte de uma unidade de massa de uma localização x para uma localização y. Aqui não nos preocupamos em dizer quem são os espaços X e Y. A fim de situar o leitor, adiantamos que vamos considerar o caso em que X e Y são espaços *poloneses.* Em seguida, define-se o conjunto $\Pi(\mu, \nu)$ de todas as medidas de probabilidade admissíveis em $\mathcal{A}^X \times \mathcal{A}^Y$, cujo papel é modelar os possíveis planos de transporte. Diz-se que uma medida de probabilidade π em $\mathcal{A}^X \times \mathcal{A}^Y$

é admissível se

$$\pi[A \times Y] = \mu[A], \qquad \pi[X \times B] = \nu[B], \tag{1.1}$$

para todos os conjuntos $A \in \mathcal{A}^X$ e $B \in \mathcal{A}^Y$. Desta forma, o problema de transporte ótimo de Kantorovich consiste em minimizar a função *custo total de transporte*

$$I[\pi] = \int_{X\times Y} c(x,y)d\pi(x,y) \tag{1.2}$$

definido em $\Pi(\mu,\nu)$. Qualquer probabilidade em $\Pi(\mu,\nu)$ que satisfaça

$$I[\pi] = \inf_{\pi\in\Pi(\mu,\nu)} I[\pi],$$

será chamada de *plano de transporte ótimo.*

No problema de minimização proposto por Kantorovich, a priori, não se excluía a possibilidade de que a massa localizada em uma determinada posição pudesse ser repartida em algumas porções e que estas pudessem ser enviadas para diferentes destinos. Monge, por outro lado, sob as mesmas condições de Kantorovich, exigia adicionalmente que para cada localização x estivesse associado um único destino y. Em termos de planos de transporte, isto significa que π em (1.2) seja da forma

$$d\pi(x,y) = d\pi_T(x,y) \equiv d\mu(x)\delta_{[y=T(x)]} \tag{1.3}$$

onde T é uma função mensurável de X em Y.

Consequentemente, obtém-se associado ao problema de Monge a seguinte função custo total de transporte

$$I[T] = \int_X c(x,T(x))d\mu(x) \tag{1.4}$$

definida no conjunto das funções mensuráveis T de X em Y tais que π_T pertença ao conjunto $\Pi(\mu,\nu)$. Neste caso , equivalentemente a (1.1), $\Pi(\mu,\nu)$ passa a ser caracterizada pela condição

$$\forall B \in \mathcal{A}^Y \quad \nu[B] = \mu[T^{-1}(B)]. \tag{1.5}$$

Quando uma função mensurável T de X em Y satisfizer (1.5), diremos que ν é a *medida imagem* de μ por T e escreveremos $\nu = T\#\mu$. Assim o problema de Monge consiste em minimizar a função custo em (1.4) sobre o conjunto das funções mensuráveis T tais que $\nu = T\#\mu$. Qualquer função mensurável T de X em Y que satisfaça

$$I[T] = \inf_{\nu=T\#\mu} \int_X c(x,T(x))d\mu(x),$$

será chamado de *transporte ótimo.*

1.1 Objetivos

Diante destas considerações preliminares, surgem naturalmente duas questões:

- Existem planos de transporte minimizantes para o problema de Monge-Kantorovich?
- Se tais existirem (e de fato existem), podemos caracterizá-los?

O objetivo desta dissertação é estudar o problema de Monge-Kantorovich, buscando responder a estas duas questões quando o custo é dado pelo quadrado da distância.

1.2 Estrutura

Esta dissertação de mestrado é baseada na obra *Topics in Optimal Transportation* cuja autoria é de Cédric Villani. Iremos estruturá-la da seguinte forma: no capítulo 2, introduziremos os conceitos básicos, certos exemplos, carateterizaremos o conjunto dos planos de transportes admissíveis e apresentaremos o teorema de dualidade para um custo geral. O estudo feito neste capítulo é superficial. Em vista de um estudo detalhado indicamos a dissertação de mestrado de **Aline**, [9].

O capítulo 3 é dedicado ao estudo minucioso do problema de Monge-Kantorovich no caso específico que o custo é dado pelo quadrado da distância.

Capítulo 2

A Dualidade de Kantorovich

Neste capítulo, o principal objeto é o teorema de Dualidade de Kantorovich. A sua demonstração rigorosa é delicada, extensa e requer algumas ferramentas de análise convexa cuja utilidade para o restante da dissertação é ínfima. Sendo assim, preferimos omitir a demonstração rigorosa e apresentar uma versão informal.

Inicialmente, fixamos a notação utilizada durante a dissertação. Em seguida, definimos os conceitos básicos e enunciamos duas importantes caracterizações para o conjunto dos planos de transporte admissíveis $\Pi(\mu, \nu)$. A parte final deste capítulo é reservado ao teorema da dualidade.

2.1 Notação

Em toda a dissertação, qualquer que seja o espaço métrico (X, d), que denotaremos por simplicidade por X, vamos equipá-lo com a topologia induzida pela distância, e denotaremos por $B(x, r)$ a bola de centro x e raio r. Além disso, dados dois espaços métricos (X, d) e $(Y, \tilde{d})$, consideraremos o espaço métrico produto $(X \times Y, [d + \tilde{d}])$ onde $[d + \tilde{d}]((x_1, y_1), (x_2, y_2)) = d(x_1, x_2) + \tilde{d}(y_1, y_2)$, $\forall x_1, x_2 \in X$ e $\forall y_1, y_2 \in Y$.

Dado $A \subset X$, diremos que $x \in A$ é um ponto interior se existe $r > 0$ tal que a bola aberta $B(x, r) \subset A$. Denotaremos por A^c o complementar do conjunto A, por $\text{Int}(A)$ o conjunto dos pontos interiores de A e escreveremos $\bar{A}$ para significar o fecho de A.

Considere o espaço métrico mensurável $(X, \mathcal{B}(X))$, onde $\mathcal{B}(X)$ corresponde a σ-álgebra de Borel de X, isto é, a σ-álgebra gerada pelo abertos de X. Denotaremos por $M_+(X)$ o conjunto das medidas não-negativas σ-finitas em $\mathcal{B}(X)$ e por $P(X)$ o conjunto de todas as probabilidades definidas em $\mathcal{B}(X)$.

Dada uma medida μ em $(X, \mathcal{B}(X))$, para $p \in [1, +\infty)$ o conjunto

$$\mathcal{L}^p(d\mu) = \{f : X \to \mathbb{R} \cup \{+\infty\}; \int |f|^p d\mu < \infty\}$$

representará o espaço de Lebesgue de ordem p em relação a μ, com a usual identificação das funções que coincidem em quase toda parte. O espaço das funções contínuas e limitadas definidas em X será denotado por $C_b(X)$.

2.2 Definições e resultados básicos

Esta seção é o nosso ponto de partida.

Definição 2.2.1. Um espaço métrico X completo e separável é dito um espaço *polonês.*

Daremos a seguir dois exemplos de espaços poloneses.

Exemplo 2.2.2. Um exemplo trivial de espaço polonês é o conjunto dos números reais $\mathbb{R}$. Sabemos que $\mathbb{R}$ é completo e que o conjunto dos números racionais $\mathbb{Q}$ é um conjunto enumerável denso em $\mathbb{R}$. Isto mostra que $\mathbb{R}$ é de fato um espaço polonês.

Exemplo 2.2.3. O espaço polonês de maior interesse nesta dissertação ocorre quando $X = \mathbb{R}^n$. Argumentando de maneira similar ao do exemplo anterior, mostra-se que $\mathbb{R}^n$ é um espaço polonês.

Antes de enunciarmos as duas caracterizações para o conjunto dos planos de transporte admissíveis $\Pi(\mu, \nu)$, cujas demonstrações encontram-se em **Villani**,[11], vamos relembrar sua definição.

Definição 2.2.4. Sejam $(X, \mathcal{A}^X, \mu)$ e $(Y, \mathcal{A}^Y, \nu)$ dois espaços de probabilidade quaisquer. Definimos $\Pi(\mu, \nu)$ como o conjunto de todas as medidas não-negativas π em $\mathcal{A}^X \times \mathcal{A}^Y$ que satisfazem $\forall A \in \mathcal{A}^X$, $B \in \mathcal{A}^Y$

$$\pi[A \times Y] = \mu[A] \quad \text{e} \quad \pi[X \times B] = \nu[B]. \tag{2.1}$$

Observe que imediatamente de (2.1), $\pi[X \times Y] = \mu[X] = 1$ e, portanto, π é uma probabilidade em $\mathcal{A}^X \times \mathcal{A}^Y$.

Ambas caracterizações se farão presentes durante a dissertação, sendo a segunda mais frequente.

Lema 2.2.5. *Sejam* $(X, \mathcal{A}^X, \mu)$ *e* $(Y, \mathcal{A}^Y, \nu)$ *dois espaços de probabilidade.* $\pi \in \Pi(\mu, \nu)$ *se, e somente se, é uma medida de probabilidade em* $\mathcal{A}^X \times \mathcal{A}^Y$ *tal que* $\forall(\varphi, \psi) \in \mathcal{L}^1(d\mu) \times \mathcal{L}^1(d\nu)$*,* π *satisfaz*

$$\int\limits_{X \times Y} [\varphi(x) + \psi(y)] d\pi = \int\limits_{X} \varphi d\mu + \int\limits_{Y} \psi d\nu.$$

Lema 2.2.6. *Sejam* (X, d) *e* $(Y, \tilde{d})$ *espaços métricos,* $\mu \in P(X)$ *e* $\nu \in P(Y)$*. Então* $\pi \in \Pi(\mu, \nu)$ *se, e somente se,* $\pi \in P(X \times Y)$ *tal que* $\forall(\varphi, \psi) \in C_b(X) \times C_b(Y)$*,* π *satisfaz*

$$\int\limits_{X \times Y} [\varphi(x) + \psi(y)] d\pi = \int\limits_{X} \varphi d\mu + \int\limits_{Y} \psi d\nu. \tag{2.2}$$

2.3 Teorema da Dualidade

A seguir enunciamos a poderosa ferramenta de dualidade devida à **Kantorovich** cujo papel é fundamental na teoria do transporte.

Teorema 2.3.1. ***(Dualidade de Kantorovich)*** *Sejam* X *e* Y *espaços poloneses,* $\mu \in P(X)$ *e* $\nu \in P(Y)$ *e seja* $c : X \times Y \to \mathbb{R}^+ \cup \{+\infty\}$ *uma função custo semicontínua inferiormente.*

Sempre que $\pi \in P(X \times Y)$ *e* $(\varphi, \psi) \in \mathcal{L}^1(d\mu) \times \mathcal{L}^1(d\nu)$*, defina*

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

Defina $\Pi(\mu, \nu)$ *como o conjunto de todas as probabilidades de Borel* π *em* $X \times Y$ *que satisfazem* $\forall A \subset X$*,* $B \subset Y$ *mensuráveis*

$$\pi[A \times Y] = \mu[A] \quad e \quad \pi[X \times B] = \nu[B]$$

e defina Φ_c *como o conjunto de todas as funções mensuráveis* $(\varphi, \psi) \in \mathcal{L}^1(d\mu) \times \mathcal{L}^1(d\nu)$ *satisfazendo*

$$\varphi(x) + \psi(y) \leq c(x, y), \tag{2.3}$$

para $d\mu$*-quase todo ponto em* X *e* $d\nu$*-quase todo ponto em* Y*.*

Então,

$$\inf_{\pi \in \Pi(\mu, \nu)} I[\pi] = \sup_{(\varphi, \psi) \in \Phi_c} J(\varphi, \psi). \tag{2.4}$$

Ainda, o ínfimo do lado esquerdo de (2.4) é atingido, e o valor do supremo não muda ao restringir a definição de Φ_c *às funções contínuas e limitadas.*

Uma prova rigorosa para esta importante ferramenta pode ser encontrada em **Villani**, [11]. Faremos uma prova **informal**. A ideia da prova é transformar nosso problema em outro do tipo inf sup, e trocar as operações ao aplicar o **princípio do minimax**, isto é, substituir "inf sup" por "sup inf".

Demonstração. Note que

$$\inf_{\pi \in \Pi(\mu,\nu)} I[\pi] = \inf_{\pi \in M_+(X\times Y)} \left(I[\pi] + \begin{cases} 0, & \text{se } \pi \in \Pi(\mu,\nu) \\ +\infty, & \text{caso contrário} \end{cases} \right). \tag{2.5}$$

Afirmação 1.

$$\sup_{(\varphi,\psi)} \left[\int \varphi d\mu + \int \psi d\nu - \int [\varphi(x) + \psi(y)] d\pi(x,y) \right] = \begin{cases} 0, & \text{se } \pi \in \Pi(\mu,\nu) \\ +\infty, & \text{caso contrário,} \end{cases}$$

onde o supremo percorre todos os pares $(\varphi,\psi) \in \mathcal{L}^1(d\mu) \times \mathcal{L}^1(d\nu)$.

Prova da afirmação 1. Se $\pi \in \Pi(\mu,\nu)$, então pelo lema 2.2.5

$$\int \varphi d\mu + \int \psi d\nu - \int [\varphi(x) + \psi(y)] d\pi(x,y) = 0, \quad \forall\ (\varphi,\psi) \in \mathcal{L}^1(d\mu) \times \mathcal{L}^1(d\nu),$$

de forma que o supremo é zero.

Caso contrário, existiria um par $(\varphi_0,\psi_0) \in \mathcal{L}^1(d\mu) \times \mathcal{L}^1(d\nu)$ tal que, sem perda de generalidade, $\int [\varphi_0(x) + \psi_0(y)] d\pi(x,y) > \int \varphi_0 d\mu + \int \psi_0 d\nu$. Fazendo a escolha $(\varphi,\psi) = (\lambda\varphi_0, \lambda\psi_0)$ com $\lambda \to +\infty$, vemos que o supremo é $+\infty$, o que prova a afirmação.

Logo, o lado direito de (2.5) é dado por

$$\inf_{\pi \in M_+(X\times Y)} \sup_{(\varphi,\psi)} \left\{ \int c d\pi + \int_X \varphi d\mu + \int_Y \psi d\nu - \int [\varphi(x) + \psi(y)] d\pi(x,y) \right\}.$$

Assumindo que podemos invocar o princípio do minimax (é neste ponto que a nossa prova deixa de ser rigorosa), podemos reescrever esta expressão como

$$\sup_{(\varphi,\psi)} \inf_{\pi \in M_+(X\times Y)} \left\{ \int c d\pi + \int_X \varphi d\mu + \int_Y \psi d\nu - \int [\varphi(x) + \psi(y)] d\pi(x,y) \right\}$$

$$= \sup_{(\varphi,\psi)} \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu - \sup_{\pi \in M_+(X\times Y)} \int [\varphi(x) + \psi(y) - c(x,y)] d\pi(x,y) \right\}. \tag{2.6}$$

Afirmação 2.

$$\sup_{\pi \in M_+(X\times Y)} \int [\varphi(x) + \psi(y) - c(x,y)] d\pi(x,y) = \begin{cases} 0, & \text{se } (\varphi,\psi) \in \Phi_c \\ +\infty, & \text{caso contrário.} \end{cases} \tag{2.7}$$

Prova da afirmação 2. Neste momento, vamos assumir que os pares $(\varphi, \psi) \in \Phi_c$ satisfazem a desigualdade (2.3) para todos os pontos. O importante aqui é que tal restrição não afeta o valor do supremo em (2.4)! (A prova deste fato é feita na seção 3.3 para uma outra classe de funções, porém a demonstração é análoga.)

Definamos $\xi(x, y) \equiv \varphi(x) + \psi(y) - c(x, y)$. Se existe algum ponto (x_0, y_0) tal que $\xi(x_0, y_0) > 0$, então ao escolhermos $\pi = \lambda\delta_{(x_0,y_0)}$ com $\lambda \to +\infty$, vemos que o supremo é infinito. Por outro lado, se $\xi \leq 0$, então é imediato que o supremo é obtido para $\pi = 0$. Assim, usando nossa observação inicial, terminamos a prova da afirmação.

Substituindo a igualdade de (2.7) em (2.6), obtemos

$$\begin{aligned} \inf_{\pi \in \Pi(\mu,\nu)} I[\pi] &= \sup_{(\varphi,\psi)} \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu - \left\{ \begin{array}{ll} 0, & \text{se } (\varphi, \psi) \in \Phi_c \\ +\infty, & \text{caso contrário} \end{array} \right\} \right\} \\ &= \sup_{(\varphi,\psi) \in \Phi_c} J(\varphi, \psi). \end{aligned}$$

$\square$

Salientamos que a semicontinuidade da função custo não foi utilizada durante a nossa prova informal. Contudo, esta hipótese é indispensável para a prova rigorosa. Além disso, note que não mostramos que o ínfimo em (2.4) é atingido. Isto será o conteúdo da seção 3.2. Além disso, perceba que a dualidade de Kantorovich não se preocupa em dizer se o supremo é ou não atingido. Esta questão será esclarecida, no próximo capítulo, seção 3.5, no caso em que custo é o quadrado da distância.

Capítulo 3

Geometria do Transporte Ótimo

Daqui em diante, vamos nos concentrar em analisar o problema de transporte quando o custo é o quadrado da distância em $\mathbb{R}^n$. Embora estejamos diminuindo consideravelmente o grau de generalidade da função custo e dos espaços envolvidos, é sob estas condições que ocorrem a maioria dos casos importantes para aplicação da teoria de transporte.

O objetivo principal desta dissertação encontra-se neste capítulo e consiste em demonstrar a caracterização de otimalidade (teorema 3.6.3): quando medidas de probabilidade μ e ν possuem momentos de segunda ordem finitos, um plano de transferência é ótimo se, e somente se, está concentrado em um subdiferencial de uma função convexa.

A fim de alcançar nosso objetivo, precisamos de uma série de resultados. Começamos na seção 3.1, onde algumas considerações são feitas sobre o novo custo c e as medidas de probabilidade μ e ν. Na seção 3.2, mostramos a existência de uma plano de transporte minimizante e, em seguida, na seção 3.3, reescrevemos o problema dual aproveitando-se da particular forma do custo. A partir daí, na seção 3.4, fazemos uma breve exposição sobre algumas ferramentas de análise convexa que se fazem presentes nos capítulos seguintes, e, na seção 3.5, obtemos alguns resultados importantes como, por exemplo, o teorema 3.5.7. Na seção 3.6, terminamos esta dissertação ao demonstrarmos o teorema 3.6.3 e exibirmos dois exemplos importantes.

3.1 Dualidade para o custo quadrático

Nosso objetivo final é demonstrar o teorema 3.6.3. Antes de chegarmos lá, passaremos por uma série de resultados preliminares. Denotaremos por $X = Y = \mathbb{R}^n$, e a função custo será $c(x,y) = \|x-y\|^2$, onde $\|x\| = \langle x,x\rangle^{1/2}$ e $\langle x,y\rangle$ é um produto interno em $\mathbb{R}^n$. Para elegância das provas definiremos $c(x,y) = \|x-y\|^2/2$, levando-se em conta que o fator $1/2$ não tem influência no problema de Kantorovich. Assim,

$$I[\pi] = \int_{\mathbb{R}^n\times\mathbb{R}^n} \frac{\|x-y\|^2}{2} d\pi(x,y). \tag{3.1}$$

Sejam μ,ν duas medidas de probabilidade de Borel em $\mathbb{R}^n$ com *momentos de segunda ordem finitos*, isto é,

$$M := \int_{\mathbb{R}^n} \frac{\|x\|^2}{2} d\mu(x) + \int_{\mathbb{R}^n} \frac{\|y\|^2}{2} d\nu(y) < +\infty. \tag{3.2}$$

Esta condição assegura que o funcional I é sempre finito em $\Pi(\mu,\nu)$. De fato,

$$\frac{\|x-y\|^2}{2} \leq \frac{\|x-y\|^2 + \|x+y\|^2}{2} = \|x\|^2 + \|y\|^2,\ \forall (x,y) \in \mathbb{R}^n\times\mathbb{R}^n.$$

Logo, para qualquer $\pi \in \Pi(\mu,\nu)$

$$\begin{aligned} I[\pi] &= \int_{\mathbb{R}^n\times\mathbb{R}^n} \frac{\|x-y\|^2}{2} d\pi(x,y) \\ &\leq \int_{\mathbb{R}^n\times\mathbb{R}^n} \left(\|x\|^2 + \|y\|^2\right) d\pi(x,y) \\ &= \int_{\mathbb{R}^n} \|x\|^2 d\mu(x) + \int_{\mathbb{R}^n} \|y\|^2 d\nu(y) \\ &= 2M, \end{aligned}$$

onde usamos o lema 2.2.6 para garantir a segunda igualdade.

Relembramos que o teorema da Dualidade de Kantorovich afirma que

$$\inf_{\pi\in\Pi(\mu,\nu)} I[\pi] = \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi). \tag{3.3}$$

Isto será o ponto inicial da nossa investigação.

3.2 O problema primal

Como primeiro passo, mostramos a existência de um plano minimizante para problema de minimização de Kantorovich. Mais precisamente,

Proposição 3.2.1. ***(Existência de uma medida de probabilidade ótima)*** *O problema de minimização* $\inf\{I[\pi]; \pi \in \Pi(\mu,\nu)\}$ *admite um minimizante.*

Antes de provarmos esta proposição precisaremos de algumas definições, e de um teorema cuja prova será omitida. Começamos com a

Definição 3.2.2. Seja X um espaço métrico e μ uma medida de probabilidade definida em $\mathcal{B}(X)$.

(i) μ é *tight* se para qualquer $\varepsilon > 0$ existe um compacto K_ε tal que $\mu[K_\varepsilon^c] \leq \varepsilon$.

(ii) Uma família $\mathcal{P}$ de medidas de probabilidades definidas em $\mathcal{B}(X)$ é *tight* se para qualquer $\varepsilon > 0$ existe um compacto K_ε tal que

$$\sup_{\mu \in \mathcal{P}} \mu[K_\varepsilon^c] \leq \varepsilon.$$

Lema 3.2.3. *Qualquer medida de probabilidade definida na σ-álgebra de Borel de um espaço polonês é tight.*

Faremos a prova para o caso particular em que $X = \mathbb{R}^n$, levando-se em conta que este é o caso que utilizamos nesta dissertação.

Demonstração. Com efeito, $X = \bigcup_{n \geq 1} B_n$, onde $B_n = B[0, n]$ é compacto para todo n natural. Logo, $B_n \uparrow X$,e, portanto, $\mu[B_n] \uparrow 1$. Assim, para qualquer $\varepsilon > 0$, existe um $K_\varepsilon = B_{n(\varepsilon)}$ tal que $\mu[K_\varepsilon^c] = 1 - \mu[K_\varepsilon] \leq \varepsilon$. □

Definição 3.2.4. Seja X um espaço métrico e um subconjunto $\mathcal{P} \subset P(X)$. Dizemos que $\mathcal{P}$ é sequencialmente relativamente compacto se para qualquer sequência $(\mu_k)_{k \in \mathbb{N}}$ em $\mathcal{P}$ pode-se extrair uma subsequência, denotada por $(\mu_{n_k})_{k \in \mathbb{N}}$, e uma medida de probabilidade μ_* definida em $\mathcal{B}(X)$, tal que para qualquer $\varphi \in C_b(X)$,

$$\lim_{k \to \infty} \int_X \varphi d\mu_k = \int_X \varphi d\mu_*.$$

A prova do teorema a abaixo, devida à **Prokhorov**, é omitida. Para detalhes consultar **Billingsley**, [4].

Teorema 3.2.5. *Sejam X um espaço polonês e $\mathcal{P}$ uma família tight de medidas de probabilidade em $P(X)$. Então $\mathcal{P}$ é sequencialmente relativamente compacto em relação à topologia fraca.*

Demonstração da proposição 3.2.1. Vamos dividi-la em 5 etapas.

Afirmação 1. $\Pi(\mu, \nu)$ é não vazio.

Prova da afirmação 1. De fato, pela teoria da medida existe uma medida π definida em $\mathcal{B}(\mathbb{R}^n)\times\mathcal{B}(\mathbb{R}^n)=\mathcal{B}(\mathbb{R}^n\times\mathbb{R}^n)$ tal que $\pi(A\times B)=\mu(A)\nu(B)$ para todo $A,B\in\mathcal{B}(R^n)$. Escolhendo $A=B=\mathbb{R}^n$, segue que π é uma medida de probabilidade.

Além disso, como ν é uma medida de probabilidade

$$\begin{aligned}\pi(A\times\mathbb{R}^n) &= \mu(A)\nu(\mathbb{R}^n)\\ &= \mu(A),\ \forall A\in B(R^n).\end{aligned}$$

Analogamente, obtemos que $\pi(\mathbb{R}^n\times B)=\nu(B),\ \forall B\in B(R^n)$. Portanto, $\pi\in\Pi(\mu,\nu)$, o que completa a afirmação.

Afirmação 2. $\Pi(\mu,\nu)$ é sequencialmente relativamente compacto em relação à topologia fraca.

Prova da afirmação 2. Do lema 3.2.3, temos que μ e ν são medidas de probabilidade tights. Logo, dado $\varepsilon>0$ existem conjuntos compactos $K,L\subset\mathbb{R}^n$ tais que

$$\mu[\mathbb{R}^n\setminus K]\leq\varepsilon/2 \ \text{ e } \ \nu[\mathbb{R}^n\setminus L]\leq\varepsilon/2.$$

Assim qualquer que seja $\pi\in\Pi(\mu,\nu)$,

$$\pi[(\mathbb{R}^n\times\mathbb{R}^n)\setminus(K\times L)]\leq\pi[\mathbb{R}^n\times(\mathbb{R}^n\setminus L)]+\pi[(\mathbb{R}^n\setminus K)\times\mathbb{R}^n]\leq\varepsilon.$$

A desigualdade acima implica que

$$\sup_{\pi\in\Pi(\mu,\nu)}\pi[(\mathbb{R}^n\times\mathbb{R}^n)\setminus(K\times L)]\leq\varepsilon.$$

Portanto, o conjunto $\Pi(\mu,\nu)$ é tight e pelo teorema 3.2.5 é também sequencialmente relativamente compacto em relação à topologia fraca. Isto completa a afirmação feita.

Afirmação 3. $\Pi(\mu,\nu)$ é fechado em relação à topologia fraca.

Prova da afirmação 3. Com efeito, seja $(\pi_k)_{k\in\mathbb{N}}$ uma sequência em $\Pi(\mu,\nu)$ que converge fracamente para medida π_*. Assim, pela definição de convergência fraca

$$\forall f\in C_b(\mathbb{R}^n\times\mathbb{R}^n),\ \lim_{k\to\infty}\int_{\mathbb{R}^n\times\mathbb{R}^n}f(x,y)d\pi_k(x,y)=\int_{\mathbb{R}^n\times\mathbb{R}^n}f(x,y)d\pi_*(x,y).$$

Tomando $f\equiv 1$, temos imediatamente que π_* é uma probabilidade. Além disso, para qualquer par de funções $(\varphi,\psi)\in C_b(\mathbb{R}^n)\times C_b(\mathbb{R}^n)$,

$$\begin{aligned}\int_{\mathbb{R}^n\times\mathbb{R}^n}[\varphi(x)+\psi(y)]d\pi_*(x,y) &= \lim_{k\to\infty}\int_{\mathbb{R}^n\times\mathbb{R}^n}[\varphi(x)+\psi(y)]d\pi_k(x,y)\\ &= \int_{\mathbb{R}^n}\varphi(x)d\mu(x)+\int_{\mathbb{R}^n}\psi(y)d\nu(y),\end{aligned}$$

onde na primeira igualdade usamos a definição de convergência fraca e na segunda o fato de que $\pi_k \in \Pi(\mu,\nu)$. Pelo lema 2.2.6, segue que $\pi_* \in \Pi(\mu,\nu)$, concluindo a afirmacão.

Afirmação 4. $\Pi(\mu,\nu)$ é sequencialmente compacto em relação à topologia fraca.

Prova da afirmação 4. Seja $(\pi_k)_{k\in\mathbb{N}}$ uma sequência em $\Pi(\mu,\nu)$. Pelo passo 2, existe uma subsequência $(\pi_{k_m})_{m\in\mathbb{N}}$ que converge fracamente para alguma medida de probabilidade π_*. Mas pelo passo 3, $\pi_* \in \Pi(\mu,\nu)$. Portanto, $\Pi(\mu,\nu)$ é sequencialmente compacto em relação à topologia fraca. Assim, provamos a afirmação.

Agora concluiremos o resultado. Pelo passo 1, $\Pi(\mu,\nu)$ é não vazio. Agora como o custo é contínuo, deduzimos que I é contínuo (em relação à topologia fraca) e, portanto, do passo 4 e do fato que o funcional I em $\Pi(\mu,\nu)$ é finito (ver seção 3.1), concluímos a existência de um plano de transporte $\pi_* \in \Pi(\mu,\nu)$ tal que

$$\inf_{\pi\in\Pi(\mu,\nu)} I[\pi] = I[\pi_*].$$

Portanto, a medida de probabilidade π_* é um minimizante de I, isto é, π_* é um plano de transporte ótimo para o problema de Kantorovich. $\square$

Desta forma, respondemos a primeira das nossas questões. Uma importante consequência da proposição acima é o fato de que o espaço $P_2(\mathbb{R}^n)$ das medidas de probabilidade de Borel com segundo momento finito é metrizável. A métrica que definimos é dada por

$$W_2(\mu,\nu) = \left[\inf_{\pi\in\Pi(\mu,\nu)} \left\{\int_{\mathbb{R}^n\times\mathbb{R}^n} \|x-y\|^2 d\pi(x,y)\right\}\right]^{1/2}.$$

Esta métrica é chamada de *métrica quadrática de Wasserstein.*

Provar que tal métrica é de fato uma métrica nos desviaria dos propósitos iniciais desta dissertação. Sendo assim, optamos por omiti-la. Como referência para este assunto indicamos o capítulo 7 de **Villani**, [11]. Salientamos, ainda, que não utilizaremos este resultado no restante da dissertação.

3.3 O problema Dual

Como segundo passo, estudaremos o problema dual, cuja estrutura particular do custo nos conduzirá a um resultado muito mais preciso.

A condição para (φ, ψ) pertencer a Φ_c é

$$\varphi(x) + \psi(y) \leq \frac{\|x-y\|^2}{2}, \tag{3.4}$$

para $d\mu-$quase todo x e $d\nu-$quase todo y em $\mathbb{R}^n$. Fazendo uso da forma particular do custo quadrático, podemos expandir o lado direito de (3.4) e rearranjar os termos envolvidos para mostrar que (3.4) equivale a

$$\langle x, y\rangle \leq \left[\frac{\|x\|^2}{2} - \varphi(x)\right] + \left[\frac{\|y\|^2}{2} - \psi(y)\right]. \tag{3.5}$$

Definindo

$$\tilde{\varphi}(x) = \frac{\|x\|^2}{2} - \varphi(x), \qquad \tilde{\psi}(y) = \frac{\|y\|^2}{2} - \psi(y),$$

temos que (3.5) equivale a

$$\langle x, y\rangle \leq \tilde{\varphi}(x) + \tilde{\psi}(y). \tag{3.6}$$

Seja então $\tilde{\Phi}_c$ o conjunto de todos os pares $(\tilde{\varphi}, \tilde{\psi}) \in \mathcal{L}^1(d\mu) \times \mathcal{L}^1(d\nu)$ (tomando valores em $\mathbb{R} \cup \{+\infty\}$) tal que para quase todo x, y em $\mathbb{R}^n$,

$$\langle x, y\rangle \leq \tilde{\varphi}(x) + \tilde{\psi}(y).$$

A partir de (3.4) e (3.6), vemos que existe uma bijeção entre os conjuntos Φ_c e $\tilde{\Phi}_c$. Este fato será relevante para provarmos a proposição abaixo.

Proposição 3.3.1. *Sob as mesmas condições do teorema 2.3.1, temos que*

$$\inf_{(\tilde{\varphi}, \tilde{\psi}) \in \tilde{\Phi}_c} \left\{ J(\tilde{\varphi}, \tilde{\psi}) \right\} = \sup_{\pi \in \Pi(\mu,\nu)} \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x, y\rangle \, d\pi(x,y) \right\}. \tag{3.7}$$

Demonstração. Usando (3.2) e a bijeção entre Φ_c e $\tilde{\Phi}_c$, vemos que

$$\begin{aligned}
\sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi) &= \sup_{(\varphi,\psi)\in\Phi_c} \left\{ \int_{\mathbb{R}^n} \varphi(x) d\mu(x) + \int_{\mathbb{R}^n} \psi(y) d\nu(y) \right\} \\
&= \sup_{(\tilde{\varphi},\tilde{\psi})\in\tilde{\Phi}_c} \left\{ \int_{\mathbb{R}^n} \left(\frac{\|x\|^2}{2} - \tilde{\varphi}(x) \right) d\mu(x) + \int_{\mathbb{R}^n} \left(\frac{\|y\|^2}{2} - \tilde{\psi}(y) \right) d\nu(y) \right\} \\
&= M + \sup_{(\tilde{\varphi},\tilde{\psi})\in\tilde{\Phi}_c} \left\{ -\int_{\mathbb{R}^n} \tilde{\varphi}(x) d\mu(x) - \int_{\mathbb{R}^n} \tilde{\psi}(y) d\nu(y) \right\} \\
&= M - \inf_{(\tilde{\varphi},\tilde{\psi})\in\tilde{\Phi}_c} \left\{ \int_{\mathbb{R}^n} \tilde{\varphi}(x) d\mu(x) + \int_{\mathbb{R}^n} \tilde{\psi}(y) d\nu(y) \right\} \\
&= M - \inf_{(\tilde{\varphi},\tilde{\psi})\in\tilde{\Phi}_c} \left\{ J(\tilde{\varphi}, \tilde{\psi}) \right\}. \qquad (3.8)
\end{aligned}$$

Agora, usando (3.2) e a definição de $\Pi(\mu,\nu)$, temos que

$$\begin{aligned}
\inf_{\pi\in\Pi(\mu,\nu)} I[\pi] &= \inf_{\pi\in\Pi(\mu,\nu)}\left\{\int_{\mathbb{R}^n\times\mathbb{R}^n}\frac{\|x-y\|^2}{2}d\pi(x,y)\right\}\\
&= \inf_{\pi\in\Pi(\mu,\nu)}\left\{\int_{\mathbb{R}^n\times\mathbb{R}^n}\left[\frac{\|x\|^2}{2}+\frac{\|y\|^2}{2}-\langle x,y\rangle\right]d\pi(x,y)\right\}\\
&= \inf_{\pi\in\Pi(\mu,\nu)}\left\{\int_{\mathbb{R}^n}\frac{\|x\|^2}{2}d\mu(x)+\int_{\mathbb{R}^n}\frac{\|y\|^2}{2}d\nu(y)-\int_{\mathbb{R}^n\times\mathbb{R}^n}\langle x,y\rangle\, d\pi(x,y)\right\}\\
&= M-\sup_{\pi\in\Pi(\mu,\nu)}\left\{\int_{\mathbb{R}^n\times\mathbb{R}^n}\langle x,y\rangle\, d\pi(x,y)\right\}. \qquad (3.9)
\end{aligned}$$

Assim, ao aplicar o teorema 2.3.1, obtemos (3.7), a partir das identidades (3.8) e (3.9). □

No que segue, por simplicidade de notação, omitiremos o símbolo ˜ dos pares de funções pertencentes a $\tilde{\Phi}_c$.

Nosso próximo passo é mostrar que o problema dual, nesta nova formulação, possui um minimizante. Para tanto, precisaremos trabalhar com funções convexas. Assim, neste momento, fazemos uma breve digressão para revisar alguns fatos de análise convexa.

3.4 Tópicos de análise convexa

Nesta seção, damos algumas definições e relembramos alguns fatos sobre análise convexa que são utilizados nas seções seguintes. Enunciar precisamente cada um destes fatos nos exigiria bastante cuidado e esforço que não proporcionaria uma melhor compreensão do que eles significam. Portanto, preferimos apresentá-los de forma simplificada. Alguns dos resultados aqui enunciados são demonstrados no Apêndice. Indicamos como referência o livro de análise convexa do **Rockafellar**, [10].

O principal resultado desta seção é a caracterização fundamental do subdiferencial de uma função convexa semicontínua inferiormente.

Definição 3.4.1. Seja φ uma função definida em $\mathbb{R}^n$ tomando valores em $\mathbb{R}\cup\{+\infty\}$:

(i) φ é *própria* se não é identicamente $+\infty$;

(ii) O *domínio* de φ, $\mathrm{Dom}(\varphi)$, é dado por

$$\mathrm{Dom}(\varphi)=\{x\in\mathbb{R}^n;\ \varphi(x)<+\infty\};$$

(iii) φ é *convexa* se $\forall x,y\in\mathbb{R}^n, \forall t\in[0,1]$

$$\varphi(tx+(1-t)y)\leq t\varphi(x)+(1-t)\varphi(y); \qquad (3.10)$$

(iv) φ é *semicontínua inferiormente* se para cada $\lambda \in \mathbb{R}$ o conjunto

$$[\varphi \leq \lambda] = \{x \in \mathbb{R}^n; \varphi(x) \leq \lambda\}$$

é fechado, ou equivalentemente, se para qualquer sequência $(x_n)_{n\in\mathbb{N}}$ que converge para um certo ponto $x \in \mathbb{R}^n$, tem-se $\liminf_{n\to\infty} \varphi(x_n) \geq \varphi(x)$.

Na teoria de regularidade de funções convexas definidas em $\mathbb{R}^n$, surge naturalmente a noção de conjunto com dimensão de Hausdorff no máximo $n - 1$. Chamamos, nesta dissertação, estes conjuntos de *Small Sets*. Segundo **Villani**, [11], podemos pensar que os small sets são subconjuntos de $\mathbb{R}^n$ com medida nula em relação à Lebesgue, sem prejudicar o entendimento.

Exemplo 3.4.2. A fronteira de um conjunto convexo é um small set. Em particular, $\partial(\mathrm{Dom}(\varphi))$ é um small set, se φ for uma função convexa.

1) Diferenciabilidade. Durante a demonstração do teorema 3.6.3, precisaremos de alguns resultados clássicos de diferenciabilidade. São os seguintes:

- Uma função convexa própria é automaticamente contínua e localmente Lipschitz no $\mathrm{Int}(\mathrm{Dom}(\varphi))$;
- Teorema de **Rademacher**: o gradiente $\bigtriangledown\varphi$, de uma função convexa φ, é localmente limitado e está definido em quase todo ponto. Além disso, o conjunto dos pontos onde $\bigtriangledown\varphi$ não existe é um small set.

Definição 3.4.3. Seja $\varphi : \mathbb{R}^n \longmapsto \mathbb{R}^n$ uma função. Diremos que φ é *monótona* se

$$\forall x, y \in \mathbb{R}^n, \qquad \langle \varphi(x) - \varphi(y), x - y\rangle \geq 0.$$

2) O gráfico de uma função convexa permanece acima de sua tangente. Para todos os pontos x onde φ é diferenciável, temos a relação crucial

$$\forall z \in \mathbb{R}^n, \qquad \varphi(z) \geq \varphi(x) + \langle \bigtriangledown\varphi(x), z - x\rangle, \tag{3.11}$$

a qual expressa o fato geométrico de que o gráfico de φ se encontra acima de seu hiperplano tangente em um ponto x. Em particular, $\bigtriangledown\varphi$ é monótono: dados quaisquer pontos de diferenciabilidade para φ, digamos x, y, temos

$$\begin{aligned} \langle \bigtriangledown\varphi(x), x - y\rangle &= -\langle \bigtriangledown\varphi(x), y - x\rangle \\ &\geq \varphi(x) - \varphi(y) \geq \langle \bigtriangledown\varphi(y), x - y\rangle, \end{aligned}$$

onde na primeira desigualdade fizemos a escolha $z = y$ em (3.11), e na segunda $z = x$.

Além disso, temos a *"continuidade do gradiente".* Sejam φ uma função convexa própria, x um ponto de diferenciabilidade de φ, e $y = \bigtriangledown\varphi(x)$. Então dado $\varepsilon > 0$, existe $\delta > 0$ tal que

$$\bigtriangledown\varphi\left(B_\delta(x)\right) \subset B_\epsilon(y). \tag{3.12}$$

3) Subdiferenciabilidade. O *subdiferencial* $\partial\varphi$ de uma função convexa φ é uma aplicação definida por

$$\partial\varphi(x) = \{y \in \mathbb{R}^n : \forall z \in \mathbb{R}^n, \quad \varphi(z) \geq \varphi(x) + \langle y, z - x\rangle\}, \quad \forall x \in \mathbb{R}^n. \tag{3.13}$$

Nesta dissertação, vamos identificar a aplicação subdiferencial $\partial\varphi$ com o seu gráfico, denotado por $\mathrm{Graph}(\partial\varphi)$, isto é, consideraremos o subdiferencial como um subconjunto de $\mathbb{R}^n \times \mathbb{R}^n$. Os resultados que utilizaremos sobre o subdiferencial são:

- Para todo $x \in \mathrm{Int}(\mathrm{Dom}(\varphi))$, o subdiferencial de uma função convexa $\partial\varphi(x)$ é não vazio;

- Uma função convexa φ é diferenciável em um ponto x se, e somente se, $\partial\varphi(x)$ contém um único ponto, o qual é então $\bigtriangledown\varphi(x)$;

- Se φ é uma função semicontínua inferiormente, então $\mathrm{Graph}(\partial\varphi)$ é um conjunto fechado.

As demonstrações destes fatos se encontram no apêndice.

4) Funções conjugadas. Seja $\varphi : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ uma função própria qualquer. Então definimos a sua função *convexa conjugada* ou *transformada de Legende*, por

$$\varphi^*(y) = \sup_{x\in\mathbb{R}^n} \left[\langle x, y\rangle - \varphi(x)\right]. \tag{3.14}$$

Note que φ^* é uma função definida em $\mathbb{R}^n$ tomando valores em $\mathbb{R} \cup \{+\infty\}$.

Afirmamos que φ^* é uma função convexa própria semicontínua inferiormente. Para vermos isso, procederemos da seguinte forma. Podemos estender a noção de transformada de Legende para funções impróprias (isto é, $\varphi \equiv +\infty$). Nestes casos, $\varphi \equiv +\infty \Leftrightarrow \varphi^* \equiv -\infty$. Portanto, se φ é própria, obrigatoriamente φ^* é uma função própria.

Note que para quaisquer $(y, \bar{y}, t) \in \mathbb{R}^n \times \mathbb{R}^n \times [0, 1]$,

$$\begin{aligned} t\varphi^*(y) + (1-t)\varphi^*(\bar{y}) &\geq \sup_{x\in\mathbb{R}^n}\left[\langle x, ty\rangle - t\varphi(x) + \langle x, (1-t)\bar{y}\rangle - (1-t)\varphi(x)\right] \\ &= \sup_{x\in\mathbb{R}^n}\left[\langle x, tay + (1-t)\bar{y}\rangle - \varphi(x)\right] \\ &= \varphi^*(tay + (1-t)\bar{y}). \end{aligned}$$

Assim, vemos que φ^* é convexa.

Dado $\lambda \in \mathbb{R}$, considere uma sequência $(y_n)_{n\in\mathbb{N}}$ em $[\varphi^* \leq \lambda]$, convergindo para y. Por definição, temos que

$$\langle x, y_n\rangle - \varphi(x) \leq \lambda \quad , \forall x \in \mathbb{R}^n, \forall n \in \mathbb{N}.$$

Como $y_n \to y$, deduzimos que

$$\langle x, y\rangle - \varphi(x) \leq \lambda \quad , \forall x \in \mathbb{R}^n,$$

ou seja, $y \in [\varphi^* \leq \lambda]$. Portanto, φ^* é semicontínua inferiormente. Isto completa nossa afirmação.

Imediatamente desta afirmação concluímos que se φ é uma função própria definida em $\mathbb{R}^n$ tomando valores em $\mathbb{R} \cup \{+\infty\}$, então a função

$$\varphi^{**}(x) = \sup_{y\in\mathbb{R}^n} \left[\langle y, x\rangle - \varphi^*(y)\right], \tag{3.15}$$

é convexa própria e semicontínua inferiormente.

A partir da definição,

$$\forall x, y \in \mathbb{R}^n, \quad \langle x, y\rangle \leq \varphi(x) + \varphi^*(y). \tag{3.16}$$

A caracterização dos casos de igualdade em (3.16) será crucial para nós.

Proposição 3.4.4. ***(Caracterização do subdiferencial)*** *Seja φ uma função convexa própria definida em $\mathbb{R}^n$. Então, para todo $x, y \in \mathbb{R}^n$,*

$$\langle x, y\rangle = \varphi(x) + \varphi^*(y) \Longleftrightarrow y \in \partial\varphi(x).$$

Demonstração. Dados $x, y \in \mathbb{R}^n$ quaisquer,

$$\begin{aligned} \langle x, y\rangle = \varphi(x) + \varphi^*(y) &\Longleftrightarrow \forall z \in \mathbb{R}^n, \quad \langle x, y\rangle \geq \varphi(x) + \langle y, z\rangle - \varphi(z) \\ &\Longleftrightarrow \forall z \in \mathbb{R}^n, \quad \varphi(z) \geq \varphi(x) + \langle y, z - x\rangle \\ &\Longleftrightarrow y \in \partial\varphi(x), \end{aligned}$$

onde na primeira equivalência usamos (3.14) e na terceira (3.13). □

Terminamos esta seção com uma lema que será usado no final do teorema 3.5.7.

Lema 3.4.5. *Seja $\varphi : \mathbb{R}^n \longmapsto \mathbb{R} \cup \{+\infty\}$ uma função própria. Então $\varphi^* = \varphi^{***}$.*

Demonstração. Por definição, para todo $x, y \in \mathbb{R}^n$

$$\varphi(x) \geq \langle x, y\rangle - \varphi^*(y) \Longrightarrow \varphi(x) \geq \sup_{y\in\mathbb{R}^n} [\langle x, y\rangle - \varphi^*(y)] = \varphi^{**}(x),\ \forall x \in \mathbb{R}^n.$$

Portanto, $\varphi \geq \varphi^{**}$. Em particular, $\varphi^* \geq \varphi^{***}$.

Por outro lado, usando que $\varphi \geq \varphi^{**}$,

$$\varphi^{***}(y) = \sup_{x\in\mathbb{R}^n} [\langle y, x\rangle - \varphi^{**}(x)] \geq \sup_{x\in\mathbb{R}^n} [\langle y, x\rangle - \varphi(x)] = \varphi^*(y),\ \forall y \in \mathbb{R}^n.$$

□

Definição 3.4.6. Seja (φ, ψ) um par de funções próprias. Diremos que o par (φ, ψ) é *conjugado* se $\varphi^* = \psi$ e $\psi^* = \varphi$. Em particular $\varphi^{**} = \varphi$ e ambas são funções convexas próprias semicontínuas inferiormente.

3.5 De volta ao problema Dual

Nosso principal objetivo nesta seção é provar o teorema 3.5.7. Para isso introduziremos uma importante ferramenta, o método da **convexificação dupla**, contido no item (ii) do lema abaixo.

Lema 3.5.1. *(i) O ínfimo em (3.7) não se altera quando restrito aos pares de funções $(\varphi, \psi) \in \mathcal{L}^1(d\mu) \times \mathcal{L}^1(d\nu)$ que satisfazem a desigualdade (3.6) para todo x, y em $\mathbb{R}^n$.*

(ii) Se o par $(\varphi, \psi) \in \tilde{\Phi}_c$, onde $\tilde{\Phi}_c$ é dado em (3.6)*, então as funções φ^{**} e φ^* são mensuráveis, $(\varphi^{**}, \varphi^*) \in \tilde{\Phi}_c$ e $J(\varphi^{**}, \varphi^*) \leq J(\varphi, \psi) < +\infty$.*

Demonstração. **(i)** Seja $(\varphi, \psi) \in \tilde{\Phi}_c$. Pela definição de $\tilde{\Phi}_c$, existem conjuntos mensuráveis $N_X,\ N_Y$, com $\mu[N_X] = \nu[N_Y] = 0$ tal que a desigualdade (3.6) se verifica para todo $(x, y) \in N_X^c \times N_Y^c$. Definamos

$$\bar{\varphi}(x) = \begin{cases} \varphi(x), & \text{se } x \in N_X^c \\ +\infty, & \text{se } x \in N_X \end{cases} \quad \text{e} \quad \bar{\psi}(y) = \begin{cases} \psi(y), & \text{se } y \in N_Y^c \\ +\infty, & \text{se } y \in N_Y \end{cases}.$$

Como $\bar{\varphi} = \varphi\chi_{[N_X^c]} + \infty\chi_{[N_X]}$, vemos que $\bar{\varphi}$ se escreve como uma soma de funções mensuráveis e, portanto, é mensurável. Analogamente, $\bar{\psi}$ é mensurável. Ainda, por construção, o par $(\bar{\varphi}, \bar{\psi})$ satisfaz (3.6) para todo x, y em $\mathbb{R}^n$ e $J(\varphi, \psi) = J(\bar{\varphi}, \bar{\psi})$.

(ii) Seja $(\varphi, \psi) \in \tilde{\Phi}_c$. A partir do item (i), podemos supor que a desigualdade (3.6) é satisfeita para todo par $x, y \in \mathbb{R}^n$. Logo, temos que

$$\varphi^*(y) = \sup_{x \in \mathbb{R}^n} [\langle x, y \rangle - \varphi(x)] \leq \psi(y), \quad \forall\, y \in \mathbb{R}^n. \tag{3.17}$$

Como o supremo de funções mensuráveis é mensurável (ver **Bartle**, [2]), temos que φ^* é uma função mensurável. Ainda, a desigualdade (3.17) implica que

$$J(\varphi, \psi) = \int_{\mathbb{R}^n} \varphi(x) d\mu(x) + \int_{\mathbb{R}^n} \psi(y) d\nu(y) \geq \int_{\mathbb{R}^n} \varphi(x) d\mu(x) + \int_{\mathbb{R}^n} \varphi^*(y) d\nu(y) = J(\varphi, \varphi^*). \tag{3.18}$$

Também da desigualdade (3.17), temos que para todo ponto $x \in \mathbb{R}^n$, vale que

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

é uma função mensurável tal que $J(\varphi, \varphi^*) \geq J(\varphi^{**}, \varphi^*)$. A última desigualdade combinada a (3.18) implica em $J(\varphi, \psi) \geq J(\varphi^{**}, \varphi^*)$.

Além disso, pela definição de φ^{**},

$$\langle x, y \rangle \leq \varphi^{**}(x) + \varphi^*(y). \tag{3.19}$$

□

Chamamos de *método da convexificação dupla* a construção do par $(\varphi^{**}, \varphi^*)$ a partir de um par $(\varphi, \psi) \in \tilde{\Phi}_c$.

O próximo resultado é um corolário do lema anterior e nos mostra que o ínfimo de J em $\tilde{\Phi}_c$ é inalterado se restringirmos J a um subconjunto de $\tilde{\Phi}_c$, o qual é feito dos pares de funções convexas conjugadas semicontínuas inferiormente $(\varphi^{**}, \varphi^*)$, onde $\varphi \in \mathcal{L}^1(d\mu)$.

Proposição 3.5.2.

$$\inf_{(\varphi,\psi) \in \tilde{\Phi}_c} J(\varphi, \psi) = \inf_{\varphi \in \mathcal{L}^1(d\mu)} J(\varphi^{**}, \varphi^*). \tag{3.20}$$

Demonstração. Note que

$$\begin{aligned} \inf_{\varphi \in \mathcal{L}^1(d\mu)} J(\varphi^{**}, \varphi^*) &\leq \inf\{J(\varphi^{**}, \varphi^*) : (\varphi, \psi) \in \tilde{\Phi}_c\} \\ &\leq \inf_{(\varphi,\psi) \in \tilde{\Phi}_c} J(\varphi, \psi). \end{aligned} \tag{3.21}$$

onde a segunda desigualdade é consequência do item (ii) do lema 3.5.1.

Agora, note que

$$\inf_{(\varphi,\psi) \in \tilde{\Phi}_c} J(\varphi, \psi) \leq \inf\{J(\varphi^{**}, \varphi^*) : (\varphi^{**}, \varphi^*) \in \tilde{\Phi}_c, \varphi \in \mathcal{L}^1(d\mu)\}. \tag{3.22}$$

Por outro lado,

$$\inf_{\varphi\in\mathcal{L}^1(d\mu)} J(\varphi^{**},\varphi^*) = \inf\left\{J(\varphi^{**},\varphi^*) : \int \varphi^{**}d\mu < +\infty, \int \varphi^* d\nu < +\infty, \varphi\in\mathcal{L}^1(d\mu)\right\}. \tag{3.23}$$

Afirmação. Para toda função $\varphi \in \mathcal{L}^1(d\mu)$, tem-se $\int \varphi^* d\nu > -\infty$. Da mesma forma, para toda $\varphi \in \mathcal{L}^1(d\nu)$, tem-se $\int \varphi^* d\mu > -\infty$. Em particular, se $\varphi^* \in \mathcal{L}^1(d\nu)$, então $\int \varphi^{**} d\mu > -\infty$.

Prova da afirmação. Pela definição de φ^*, $\forall x, y \in \mathbb{R}^n$, temos que

$$\varphi(x) + \varphi^*(y) \geq \langle x, y\rangle \geq -\left[\frac{\|x\|^2}{2} + \frac{\|y\|^2}{2}\right].$$

Seja $\pi \in \Pi(\mu,\nu)$ qualquer. A última desigualdade nos permite deduzir que

$$\begin{aligned}\int \varphi^*(y)d\nu(y) &= \int \varphi^*(y)d\pi(x,y)\\ &\geq -\int \left[\frac{\|x\|^2}{2} + \frac{\|y\|^2}{2}\right] d\pi(x,y) - \int \varphi(x)d\pi(x,y)\\ &= -\int \frac{\|x\|^2}{2}d\mu(x) - \int \frac{\|y\|^2}{2}d\nu(y) - \int \varphi(x)d\mu(x) > -\infty,\end{aligned}$$

onde na primeira igualdade usamos o lema 2.2.6, e a desigualdade estrita é justificada pelo fato de que φ é integrável e μ, ν possuirem momentos de segunda ordem finitos. Raciocinando de forma análoga no outro caso concluímos a prova da afirmação.

A partir da nossa afirmação, concluímos de (3.23) que $\inf_{\varphi\in\mathcal{L}^1(d\mu)} J(\varphi^{**},\varphi^*)$ é igual a

$$\inf\left\{J(\varphi^{**},\varphi^*) : (\varphi^{**},\varphi^*) \in \mathcal{L}^1(d\mu)\times\mathcal{L}^1(d\nu), \varphi\in\mathcal{L}^1(d\mu)\right\},$$

que por sua vez é igual a

$$\inf\left\{J(\varphi^{**},\varphi^*) : (\varphi^{**},\varphi^*) \in \tilde{\Phi}_c, \varphi\in\mathcal{L}^1(d\mu)\right\},$$

pois (φ^{**},φ^*) satisfaz (3.19). Assim, de (3.21), obtemos que

$$\inf_{(\varphi,\psi)\in\tilde{\Phi}_c} J(\varphi,\psi) \leq \inf_{\varphi\in\mathcal{L}^1(d\mu)} J(\varphi^{**},\varphi^*). \tag{3.24}$$

que juntamente com (3.21) termina a prova desta proposição. □

Lema 3.5.3. ***(Lema da convexificação dupla)*** *Sejam μ,ν medidas de probabilidade em $\mathbb{R}^n$ com suportes em subconjuntos X e Y de $\mathbb{R}^n$, satisfazendo*

$$\tilde{M} \equiv \int_X \frac{\|x\|^2}{2}d\mu(x) + \int_Y \frac{\|y\|^2}{2}d\nu(y) < +\infty.$$

Quando φ, ψ são funções mensuráveis com valores em $\mathbb{R} \cup \{+\infty\}$, introduzimos

$$\varphi^*(y) = \sup_{x \in X}[\langle x, y\rangle - \varphi(x)], \tag{3.25}$$

$$\psi^*(x) = \sup_{y \in Y}[\langle x, y\rangle - \psi(y)]. \tag{3.26}$$

Seja $\tilde{\Phi}_c$ definido por (3.6), e seja $(\varphi_k, \psi_k)_{k\in\mathbb{N}}$ uma sequência minimizante para J em $\tilde{\Phi}_c$, isto é, $J(\varphi_k, \psi_k) \to \inf_{(\varphi,\psi)\in\tilde{\Phi}_c} J(\varphi, \psi)$. Então,

(i) Pode-se modificar (φ_k, ψ_k) em conjuntos de medida zero (em relação a μ, ν) de tal forma que a desigualdade (3.6) valha para todo $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$, sem alterar os valores de $J(\varphi_k, \psi_k)$.

(ii) Existe uma sequência de números reais $(a_k)_{k\in\mathbb{N}}$ tal que

$$(\bar{\varphi}_k, \bar{\psi}_k) = (\varphi_k^{**} - a_k, \varphi_k^* + a_k) \tag{3.27}$$

ainda é uma sequência minimizante para J em $\tilde{\Phi}_c$, satisfazendo as limitações inferiores

$$\forall x \in X,\ \forall y \in Y, \qquad \bar{\varphi}_k(x) \geq -\frac{\|x\|^2}{2}, \quad \bar{\psi}_k(y) \geq -\frac{\|y\|^2}{2}, \tag{3.28}$$

e as limitações superiores

$$\liminf_{k\to\infty} \inf_{x\in X} \left(\bar{\varphi}_k(x) + \frac{\|x\|^2}{2}\right) \leq \inf_{\tilde{\Phi}} J + \tilde{M}, \tag{3.29}$$

$$\liminf_{k\to\infty} \inf_{y\in Y} \left(\bar{\psi}_k(y) + \frac{\|y\|^2}{2}\right) \leq \inf_{\tilde{\Phi}} J + \tilde{M}. \tag{3.30}$$

Demonstração. **(i)**. Decorre do primeiro item do lema 3.5.1.

(ii). **Afirmação 1**. Mostraremos a existência da sequência de números reais $(a_k)_{k\in\mathbb{N}}$ de forma que a sequência de pares em (3.27) cumpra (3.28).

Prova da afirmação 1. Seja $(\varphi_k, \psi_k)_{k\in\mathbb{N}}$ qualquer sequência minimizante para J em $\tilde{\Phi}_c$. Ao invocar a parte (i) acima, podemos assumir que para todo x e y, $\varphi_k(x) \geq \langle x, y\rangle - \psi_k(y)$. Como ψ_k não é identicamente $+\infty$, então existe $y_0 \in Y$ tal que $\psi_k(y_0) = -b_0 < +\infty$. Sendo assim, deduzimos que φ_k é limitada inferiormente pela função afim $\langle x, y_0\rangle + b_0$. Além disso, segue que

$$\varphi_k^*(y_0) = \sup_{x\in X}[\langle x, y_0\rangle - \varphi_k(x)] \leq -b_0. \tag{3.31}$$

Usando a definição de φ_k^*, temos que para todo $x \in X,\ y \in Y$,

$$\varphi_k(x) + \varphi_k^*(y) \geq \langle x, y \rangle .$$

Como φ_k não é identicamente $+\infty$, então existe $x_0 \in X$ tal que $\varphi_k(x_0) = -c_0 < +\infty$. Logo, da desigualdade acima, concluímos que para todo $y \in Y$

$$\varphi_k^*(y) \geq \langle x_0, y \rangle + c_0. \tag{3.32}$$

Portanto, de (3.31) e (3.32) obtemos

$$-\frac{\|x_0\|^2}{2} + c_0 \leq \inf_{y\in Y} \left[\frac{\|y\|^2}{2} + \langle x_0, y\rangle + c_0\right] \leq \inf_{y\in Y} \left[\varphi_k^*(y) + \frac{\|y\|^2}{2}\right] \leq -b_0 + \frac{\|y_0\|^2}{2}, \tag{3.33}$$

onde a primeira desigualdade é válida, pois $\forall y \in Y$, $\frac{\|y+x_0\|^2}{2} \geq 0$.

A primeira e última desigualdades em (3.33) implicam que

$$a_k \equiv \inf_{y\in Y} \left[\varphi_k^*(y) + \frac{\|y\|^2}{2}\right]$$

é finito. Definamos

$$(\bar{\varphi}_k, \bar{\psi}_k) = (\varphi_k^{**} - a_k, \varphi_k^* + a_k).$$

Note que $\bar{\varphi}_k = (\bar{\psi}_k)^*$. Com efeito, para todo $x \in X$

$$\begin{aligned}
(\bar{\psi}_k)^*(x) &= (\varphi_k^* - a_k)^*(x) \\
&= \sup_{y\in Y}[\langle x, y\rangle - (\varphi_k^*(x) - a_k)] \\
&= \sup_{y\in Y}[\langle x, y\rangle - \varphi_k^*(x)] + a_k \\
&= \varphi_k^{**}(x) + a_k = \bar{\varphi}_k(x),
\end{aligned}$$

onde na segunda e na quarta igualdades usamos a definição da transformada de Legendre.

Além disso, pela definição de $\bar{\psi}_k$ e a_k

$$\inf_{y\in Y} \left[\bar{\psi}_k(y) + \frac{\|y\|^2}{2}\right] = \inf_{y\in Y} \left[\varphi_k^*(y) - a_k + \frac{\|y\|^2}{2}\right] = \inf_{y\in Y} \left[\varphi_k^*(y) + \frac{\|y\|^2}{2}\right] - a_k = 0. \tag{3.34}$$

Em particular, para todo $y \in Y$, $\bar{\psi}_k(y) \geq -\frac{\|y\|^2}{2}$.

Agora, provemos que para todo $x \in X$, $\bar{\varphi}_k(x) \geq -\frac{\|x\|^2}{2}$. De fato,

$$\begin{aligned}
\bar{\varphi}_k(x) + \frac{\|x\|^2}{2} &= (\bar{\psi}_k)^*(x) + \frac{\|x\|^2}{2} \\
&= \sup_{y\in Y} \left[\langle x, y\rangle - \bar{\psi}_k(y) + \frac{\|x\|^2}{2}\right] \\
&\geq \sup_{y\in Y} \left[-\frac{\|y\|^2}{2} - \bar{\psi}_k(y)\right] \\
&= -\inf_{y\in Y} \left[\frac{\|y\|^2}{2} + \bar{\psi}_k(y)\right] = 0,
\end{aligned}$$

onde na primeira igualdade usamos que $\bar{\varphi}_k = (\bar{\psi}_k)^*$, na segunda a definição da transformada de Legendre e na última a identidade (3.34), e na desigualdade usamos novamente $\frac{\|x+y\|^2}{2} \geq 0$. Com isso, provamos a afirmação 1.

Afirmação 2. Provaremos que para todo $k \in \mathbb{N}$ o par $(\bar{\varphi}_k, \bar{\psi}_k) \in \tilde{\Phi}_c$ e que a sequência formada por estes pares constitui uma sequência minimizante para J em $\tilde{\Phi}_c$.

Prova da afirmação 2. Observando que para todo $k \in \mathbb{N}$, $(\varphi_k, \psi_k) \in \tilde{\Phi}_c$ temos que as funções $\bar{\varphi}_k, \bar{\psi}_k$ são mensuráveis tais que

$$J(\bar{\varphi}_k, \bar{\psi}_k) = J(\varphi_k^{**}, \varphi_k^*) \leq J(\varphi_k, \psi_k) < +\infty, \tag{3.35}$$

onde na primeira igualdade usamos o fato de que $\forall a \in \mathbb{R}, J(\varphi, \psi) = J(\varphi + a, \psi - a)$, e na primeira desigualdade o lema 3.5.1, item (ii). Em particular, para qualquer $k \in \mathbb{N}$,

$$\int_X \left(\frac{\|x\|^2}{2} + \bar{\varphi}_k(x) \right) d\mu(x) + \int_Y \left(\frac{\|y\|^2}{2} + \bar{\psi}_k(x) \right) d\nu(x) = \tilde{M} + J(\bar{\varphi}_k, \bar{\psi}_k) < +\infty.$$

De (3.28) vemos que ambos os integrandos são não-negativos e, portanto, deduzimos que eles são integráveis. Logo $\forall k \in \mathbb{N},\ \bar{\varphi}_k \in \mathcal{L}^1(d\mu)$ e $\bar{\psi}_k \in \mathcal{L}^1(d\nu)$. Também do lema 3.5.1 item (ii) segue que $\forall k \in \mathbb{N}\ (\bar{\varphi}_k, \bar{\psi}_k) \in \tilde{\Phi}_c$. Fazendo k tender para o infinito em (3.35), e levando-se em conta que (φ_k, ψ_k) é uma sequência minimizante para J em $\tilde{\Phi}_c$, concluímos que $(\bar{\varphi}_k, \bar{\psi}_k)$ constitui uma sequência minimizante, o que prova a afirmação.

Afirmação 3. As desigualdades (3.29) e (3.30) se verificam.

Prova da afirmação 3. Note que para qualquer $k \in \mathbb{N}$,

$$\begin{aligned}
\tilde{M} + J(\bar{\varphi}_k, \bar{\psi}_k) &= \int_X \left(\frac{\|x\|^2}{2} + \bar{\varphi}_k(x) \right) d\mu(x) + \int_Y \left(\frac{\|y\|^2}{2} + \bar{\psi}_k(x) \right) d\nu(x) \\
&\geq \inf_{x \in X} \left(\frac{\|x\|^2}{2} + \bar{\varphi}_k(x) \right) \mu(X) + \inf_{y \in Y} \left(\frac{\|y\|^2}{2} + \bar{\psi}_k(x) \right) \nu(Y) \\
&= \inf_{x \in X} \left(\frac{\|x\|^2}{2} + \bar{\varphi}_k(x) \right) + \inf_{y \in Y} \left(\frac{\|y\|^2}{2} + \bar{\psi}_k(x) \right).
\end{aligned}$$

Como ambas expressões dentro dos parênteses são não-negativas, segue, por exemplo que

$$\liminf_{k \to \infty} \inf_{x \in X} \left(\bar{\varphi}_k(x) + \frac{\|x\|^2}{2} \right) \leq \liminf_{k \to \infty} J(\bar{\varphi}_k, \bar{\psi}_k) + \tilde{M} = \inf_{(\varphi, \psi) \in \tilde{\Phi}_c} J(\varphi, \psi) + \tilde{M},$$

onde na última igualdade usamos que $(\bar{\varphi}_k, \bar{\psi}_k)$ constitui uma sequência minimizante para J em $\tilde{\Phi}_c$. Isto mostra que (3.29) se verifica. De maneira análoga, mostramos que a desigualdade (3.30) é válida. Isto conclui a afirmação 3 e a prova do item (ii).

□

Durante a demonstração do teorema 3.5.7 necessitaremos da noção de conjunto uniformemente integrável e de alguma caracterização para estes conjuntos. Começamos com a

Definição 3.5.4. Seja μ uma medida finita. Um subconjunto $K \subset \mathcal{L}^1(d\mu)$ é *uniformemente integrável* se

$$\lim_{c\to\infty} \sup_{f\in K} \int_{|f|\geq c} |f| d\mu = 0.$$

Definição 3.5.5. Sejam μ uma medida de probabilidade e $(\varphi_n)_{n\in\mathbb{N}}$ uma sequência de funções em $\mathcal{L}^1(d\mu)$. Dizemos que $(\varphi_n)_{n\in\mathbb{N}}$ converge fracamente para φ se para qualquer ψ contínua,

$$\lim_{n\to\infty} \int \varphi_n \psi d\mu = \int \varphi\psi d\mu.$$

A caracterização que utilizaremos é devida à **Dunford-Pettis** e sua demonstração será omitida, podendo ser encontrada em **Balder**, [1].

Proposição 3.5.6. ***(Dunford-Pettis)*** *Um subconjunto $K \subset \mathcal{L}^1(d\mu)$ é uniformemente integrável se, e somente se, é relativamente compacto em relação à topologia fraca.*

Finalmente estamos em condições de provar o

Teorema 3.5.7. ***(Existência de um par ótimo de funções convexas conjugadas)*** *Sejam μ e ν duas medidas de probabilidade em $\mathbb{R}^n$, com momentos de segunda ordem finitos. Seja $\tilde{\Phi}_c$ definido por (3.6). Então, existe um par (φ, φ^*) de funções convexas conjugadas próprias semicontínuas inferiormente definidas em $\mathbb{R}^n$, tal que*

$$\inf_{(\varphi,\psi)\in\tilde{\Phi}_c} J(\varphi, \psi) = J(\varphi, \varphi^*).$$

Demonstração. Vamos dividir a demonstração em 6 passos.

Passo 1. Vamos substituir a função $\langle x, y\rangle$ por uma função não-negativa; para tal adicionaremos uma adequada função integrável em ambos os lados da desigualdade que define $\tilde{\Phi}_c$. Por exemplo,

$$(\varphi, \psi) \in \tilde{\Phi}_c \Longleftrightarrow \left[\varphi(x) + \frac{\|x\|^2}{2}\right] + \left[\psi(y) + \frac{\|y\|^2}{2}\right] \geq \frac{\|x\|^2}{2} + \frac{\|y\|^2}{2} + \langle x, y\rangle = \frac{\|x + y\|^2}{2}.$$

Seja $(\varphi_k, \psi_k)_{k\in\mathbb{N}}$ uma sequência minimizante para J em $\tilde{\Phi}_c$. Pelo item (ii) do lema 3.5.3, podemos assumir que

$$0 \leq \varphi_k(x) + \frac{\|x\|^2}{2}, \quad 0 \leq \psi_k(y) + \frac{\|y\|^2}{2}, \ \forall x, y \in \mathbb{R}^n,$$

de forma que

$$\left[\varphi_k(x) + \frac{\|x\|^2}{2}\right] + \left[\psi_k(y) + \frac{\|y\|^2}{2}\right] \geq \frac{\|x+y\|^2}{2} \geq 0, \ \forall x, y \in \mathbb{R}^n.$$

O ponto importante aqui é que tudo é não-negativo!

Passo 2. Vamos obter algumas estimativas que nos garantirão integrabilidade uniforme. Para cada $l \in \mathbb{N}$, definamos $\varphi_k^{(l)}, \psi_k^{(l)}$ por truncamento:

$$\varphi_k^{(l)}(x) + \frac{\|x\|^2}{2} = \min\left(\varphi_k(x) + \frac{\|x\|^2}{2}, l\right), \ \forall x \in \mathbb{R}^n$$

e

$$\psi_k^{(l)}(y) + \frac{\|y\|^2}{2} = \min\left(\varphi_k(y) + \frac{\|y\|^2}{2}, l\right), \ \forall y \in \mathbb{R}^n.$$

A fim de simplificar a escrita, vamos definir

$$O_k^{(l)}(x) = \min\left(\varphi_k(x) + \frac{\|x\|^2}{2}, l\right), \ \forall x \in \mathbb{R}^n$$

e

$$\tilde{O}_k^{(l)}(y) = \min\left(\psi_k(y) + \frac{\|y\|^2}{2}, l\right), \ \forall y \in \mathbb{R}^n.$$

A partir de suas definições e do passo 1, obtemos

$$\begin{cases} 0 \leq O_k^{(l)}(x) \leq l, \ \forall x \in \mathbb{R}^n, \\ \\ 0 \leq \tilde{O}_k^{(l)}(y) \leq l, \ \forall y \in \mathbb{R}^n, \end{cases} \tag{3.36}$$

$$\begin{cases} \varphi_k^{(l)}(x) \leq \varphi_k(x), \forall x \in \mathbb{R}^n, \\ \\ \psi_k^{(l)}(x) \leq \psi_k(y), \forall y \in \mathbb{R}^n, \end{cases} \tag{3.37}$$

e usando, adicionalmente, que $O_k^{(l)}(x) + \tilde{O}_k^{(l)}(y) \geq \min\left(\frac{\|x+y\|^2}{2}, l\right)$ (ver apêndice, lema 3.6.10), também deduzimos que

$$\varphi_k^{(l)}(x) + \psi_k^{(l)}(y) \ \ \geq \ \ \min\left[\frac{\|x+y\|^2}{2}, l\right] - \left(\frac{\|x\|^2}{2} + \frac{\|y\|^2}{2}\right), \tag{3.38}$$

De (3.37) concluímos que

$$J(\varphi_k^{(l)}, \psi_k^{(l)}) \leq J(\varphi_k, \psi_k). \tag{3.39}$$

Agora, note que

$$\begin{cases} O_k^{(l)}(x) \leq O_k^{(l+1)}(x), \ \forall x \in \mathbb{R}^n \\ \\ \tilde{O}_k^{(l)}(y) \leq \tilde{O}_k^{(l+1)}(y), \forall y \in \mathbb{R}^n. \end{cases} \tag{3.40}$$

As duas desigualdades de (3.40) nos permitem concluir

$$\begin{cases} \varphi_k^{(1)} \leq \varphi_k^{(2)} \leq \cdots \leq \varphi_k^{(l)} \leq \cdots, \\ \\ \psi_k^{(1)} \leq \psi_k^{(2)} \leq \cdots \leq \psi_k^{(l)} \leq \cdots, \end{cases} \tag{3.41}$$

Salientamos que as relações importantes são: (3.36), (3.38), (3.39) e (3.41).

Passo 3. Provemos que a sequência $(\varphi_k^{(l)})_{k\in\mathbb{N}^*}$ é fracamente relativamente compacta. Sabemos, de (3.36), que para cada $l \in \mathbb{N}$ e para todo $x \in \mathbb{R}^n$,

$$\varphi_k^{(l)}(x) = -\frac{\|x\|^2}{2} + O_k^{(l)}(x)$$

e $|O_k^{(l)}(x)| \leq l$.

Seja $K_l = \{O_k^{(l)}, k \in \mathbb{N}\}$. Observe que para todo k, $|O_k^{(l)}| \leq l$, de forma que $K_l \subset \mathcal{L}^1(d\mu)$ e

$$\sup_{O_k^{(l)} \in K_l} \int_{|O_k^{(l)}(x)| \geq c} |O_k^{(l)}| d\mu \leq l \sup_{O_k^{(l)} \in K_l} \mu(x \in \mathbb{R}^n : |O_k^{(l)}(x)| \geq c) = 0, \ \forall c > l + 1.$$

Logo, K_l é um conjunto uniformemente integrável e, por Dunford-Pettis, é relativamente compacto em relação à topologia fraca. Como $-\frac{\|x\|^2}{2}$ é uma função fixada em $\mathcal{L}^1(d\mu)$, deduzimos que o conjunto $\tilde{K}_l = \{\varphi_k^{(l)}, k \in \mathbb{N}\}$ é relativamente compacto em relação à topologia fraca.

Passo 4. Vamos obter algumas estimativas para o limite fraco de alguma subsequência de $(\varphi_k^{(l)})_{k\in\mathbb{N}}$. Pelo passo 3, para $l = 1$, existe um subconjunto infinito $\mathbb{N}_1 \subset \mathbb{N}$ e $\varphi^{(1)} \in \mathcal{L}^1(d\mu)$ tal que $\varphi_k^{(1)} \longrightarrow \varphi^{(1)}$ fracamente, quando $k \to \infty$ com $k \in \mathbb{N}_1$. É claro que qualquer subconjunto de um conjunto relativamente compacto em relação à topologia fraca é relativamente compacto em relação à topologia fraca. Portanto, existe um subconjunto infinito $\mathbb{N}_2 \subset \mathbb{N}_1$ e $\varphi^{(2)} \in \mathcal{L}^1(d\mu)$ tal que $\varphi_k^{(2)} \longrightarrow \varphi^{(2)}$ fracamente, quando $k \to \infty$ com $k \in \mathbb{N}_2$. Prosseguindo desta maneira, obtemos uma sequência de conjuntos infinitos $\mathbb{N} \supset \mathbb{N}_1 \supset \cdots \supset \mathbb{N}_l \supset \cdots$ tal que $\varphi_k^{(l)} \longrightarrow \varphi^{(l)} \in \mathcal{L}^1(d\mu)$ fracamente ,quando $k \to \infty$ com $k \in \mathbb{N}_l$. Definamos o conjunto $\mathbb{N}^* \subset \mathbb{N}$ estipulando que seu l-ésimo elemento (na ordem crescente dos números naturais) seja o l-ésimo elemento de $\mathbb{N}_l$. Logo, para cada $l \in \mathbb{N}$ a sequência $(\varphi_k^{(l)})_{k\in\mathbb{N}^*}$ é, a partir do seu l-ésimo elemento, uma subsequência de $(\varphi_k^{(l)})_{k\in\mathbb{N}_l}$ o que implica que $\varphi_k^{(l)} \longrightarrow \varphi^{(l)}$, quando $k \to \infty$ com $k \in \mathbb{N}^*$. Então, como a convergência fraca preserva desigualdades, deduzimos, respectivamente, de (3.38), (3.39) ,(3.41) que

$$\varphi^{(l)}(x) + \psi^{(l)}(y) \geq \min\left[\frac{\|x+y\|^2}{2}, l\right] - \left(\frac{\|x\|^2}{2} + \frac{\|y\|^2}{2}\right), \forall x, y \in \mathbb{R}^n, \tag{3.42}$$

$$J(\varphi^{(l)}, \psi^{(l)}) = \lim_{k\to\infty, k\in\mathbb{N}^*} J(\varphi_k^{(l)}, \psi_k^{(l)}) \leq \lim_{k\to\infty, k\in\mathbb{N}^*} J(\varphi_k, \psi_k) = \inf_{\tilde{\Phi}_c} J, \tag{3.43}$$

$$\begin{cases} \varphi^{(1)} \leq \varphi^{(2)} \leq \cdots \leq \varphi^{(l)} \leq \cdots, \\ \\ \psi^{(1)} \leq \psi^{(2)} \leq \cdots \leq \psi^{(l)} \leq \cdots. \end{cases} \tag{3.44}$$

Passo 5. Vamos mostrar que existe um par (φ, ψ) ótimo para o problema dual. Argumentando de modo similar a do passo 4, de (3.36), concluímos que para cada $l \in \mathbb{N}$

$$\begin{cases} 0 \leq \varphi^{(l)}(x) + \frac{\|x\|^2}{2}, \ \forall x \in \mathbb{R}^n, \\ \\ 0 \leq \psi^{(l)}(y) + \frac{\|y\|^2}{2}, \ \forall y \in \mathbb{R}^n, \end{cases}$$

de forma que

$$\begin{aligned} \int_{\mathbb{R}^n} |\varphi^{(l)}(x)| d\mu(x) &= \int_{\mathbb{R}^n} \left|\varphi^{(l)}(x) + \frac{\|x\|^2}{2} - \frac{\|x\|^2}{2}\right| d\mu(x) \\ &\leq \int_{\mathbb{R}^n} \left(\varphi^{(l)}(x) + \frac{\|x\|^2}{2}\right) d\mu(x) + \int_{\mathbb{R}^n} \frac{\|x\|^2}{2} d\mu(x), \end{aligned}$$

e

$$\int_{\mathbb{R}^n} |\psi^{(l)}(y)| d\nu(y) \leq \int_{\mathbb{R}^n} \left(\psi^{(l)}(y) + \frac{\|y\|^2}{2}\right) d\nu(y) + \int_{\mathbb{R}^n} \frac{\|y\|^2}{2} d\nu(y).$$

Somando as desigualdades acima, vemos que

$$\int_{\mathbb{R}^n} |\varphi^{(l)}(x)| d\mu(x) + \int_{\mathbb{R}^n} |\psi^{(l)}(y)| d\nu(y) \leq J(\varphi^{(l)}, \psi^{(l)}) + 2M \leq \inf_{\tilde{\Phi}_c} J + 2M < +\infty. \tag{3.45}$$

Além disso, temos que ambas são limitadas inferiormente por uma função em $\mathcal{L}^1$ fixada:

$$\begin{cases} -\frac{\|x\|^2}{2} \leq \varphi^{(l)}(x), \ \forall x \in \mathbb{R}^n, \\ \\ -\frac{\|y\|^2}{2} \leq \psi^{(l)}(y), \ \forall y \in \mathbb{R}^n. \end{cases}$$

De (3.44), vemos que estas sequências são crescentes. Um dos fatos importantes aqui é que sob estas hipóteses o teorema da convergência monótona ainda é válido!

Definamos

$$\varphi(x) = \lim_{l\to\infty} \varphi^{(l)}(x), \qquad \psi(y) = \lim_{l\to\infty} \psi^{(l)}(y),$$

quando os limites existirem. Vamos mostrar que estas funções estão bem definidas para quase todos os pontos e que o par (φ, ψ) é ótimo.

Note φ e ψ são mensuráveis, já que o limite de funções mensuráveis é mensurável. Pelo teorema da convergência monótona,

$$\int_{\mathbb{R}^n} \varphi d\mu = \lim_{l\to\infty} \int_{\mathbb{R}^n} \varphi^{(l)} d\mu, \qquad \int_{\mathbb{R}^n} \psi d\nu = \lim_{l\to\infty} \int_{\mathbb{R}^n} \varphi^{(l)} d\nu, \tag{3.46}$$

o que implica,

$$\int_{\mathbb{R}^n} \varphi d\mu \leq \lim_{l\to\infty} \int_{\mathbb{R}^n} |\varphi^{(l)}| d\mu < +\infty, \quad \int_{\mathbb{R}^n} \psi d\mu \leq \lim_{l\to\infty} \int_{\mathbb{R}^n} |\psi^{(l)}| d\nu < +\infty$$

onde usamos (3.45) para garantir que os limites são finitos. Assim, temos que o par $(\varphi, \psi) \in \mathcal{L}^1(d\mu) \times \mathcal{L}^1(d\nu)$, o que implica φ finito para $d\mu$-quase todo ponto e ψ finito para $d\nu$-quase todo ponto. A partir disso, já que $(\varphi^{(l)})_{l\in\mathbb{N}}$ e $(\psi^{(l)})_{l\in\mathbb{N}}$ são sequências crescentes, deduzimos que as funções estão φ, ψ estão bem definidas para quase todos os pontos. Nos pontos em estas funções não estão definidas, podemos defini-las como $+\infty$.

Ao passar o limite em (3.42), vemos que $\varphi(x) + \psi(y) \geq \langle x, y\rangle$, $\forall x, y \in \mathbb{R}^n$, de forma que o par $(\varphi, \psi) \in \tilde{\Phi}_c$. Por (3.43) e (3.46), concluímos que

$$\inf_{\tilde{\Phi}_c} J \leq J(\varphi, \psi) = \lim_{l\to\infty} J(\varphi^{(l)}, \psi^{(l)}) \leq \inf_{\tilde{\Phi}_c} J.$$

Portanto, deduzimos que (φ, ψ) é um par ótimo.

Passo 6. Finalizaremos a prova. Ao aplicar o item (ii) do lema 3.5.1 no par ótimo (φ, ψ), ou seja, o método da convexificação dupla, obtemos um par de funções próprias convexas conjugadas semicontínuas inferiormente $(\varphi^{**}, \varphi^*) \in \tilde{\Phi}_c$ tal que

$$\inf_{\tilde{\Phi}_c} J \leq J(\varphi^{**}, \varphi^*) \leq J(\varphi, \psi) = \inf_{\tilde{\Phi}_c} J.$$

Agora, seja $\xi \equiv \varphi^{**}$. Pelo lema 3.4.5, temos $\xi^* = \varphi^*$ de forma que (ξ, ξ^*) é o par de funções convexas próprias conjugadas semicontínuas inferiormente tal que

$$\inf_{\tilde{\Phi}_c} J(\varphi, \psi) = J(\xi, \xi^*).$$

□

3.6 Demonstração do critério de otimalidade

Lembramos que escrevemos $\nu = T\#\mu$ quando

$$\nu[B] = \mu[T^{-1}(B)], \quad \forall B \in \mathcal{B}(\mathbb{R}^n)\,,$$

onde $T : \mathbb{R}^n \to \mathbb{R}^n$ uma função mensurável. É fácil provar (ver **Fernandez**, [6]) que $\nu = T\#\mu$ se, e somente se,

$$\int_{\mathbb{R}^n} (\psi \circ T(x))d\mu(x) = \int_{\mathbb{R}^n} \psi(y)d\nu(y) \tag{3.47}$$

para toda $\psi \in \mathcal{L}^1(d\nu)$.

Definição 3.6.1. Seja $T : \mathbb{R}^n \to \mathbb{R}^n$ uma função mensurável. Definimos a aplicação $(Id \times T) : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$, por

$$(Id \times T)(x) = (x, T(x)).$$

Durante a demonstração do critério de otimalidade precisaremos também da

Definição 3.6.2. Seja X um espaço métrico e μ uma medida de probabilidade definida em $\mathcal{B}(X)$. O *suporte* de μ, $\mathrm{Supp}(\mu)$, é o menor subconjunto fechado de X para o qual cada vizinhança aberta de um ponto $x \in X$ tem medida estritamente positiva.

Agora estamos prontos para provar o resultado mais relevante desta dissertação. Aqui usaremos diversas vezes os resultados da seção 3.4.

Teorema 3.6.3 (**Teorema do Transporte Ótimo para Custo Quadrático).** *Sejam μ, ν medidas de probabilidade em $\mathbb{R}^n$, com momentos de segunda ordem finitos no sentido de (3.2). Consideramos o problema de transporte de Monge-Kantorovich associado a função custo quadrática $c(x,y) = \|x-y\|^2$.*

(i) ***(Critério de otimização de Knott-Smith)*** *$\pi \in \Pi(\mu,\nu)$ é ótimo se, e somente se, existe uma função convexa φ tal que*

$$\mathit{Supp}(\pi) \subset \mathit{Graph}(\partial\varphi), \tag{3.48}$$

ou equivalentemente,

$$y \in \partial\varphi(x) \;\; \textit{para } d\pi\textit{-quase todo ponto } (x,y) \in \mathbb{R}^n \times \mathbb{R}^n, \tag{3.49}$$

onde $\partial\varphi$ é o subdiferencial de φ (ver a definição de subdiferencial na equação 3.13). Além disso, neste caso, o par (φ, φ^*) é um minimizante para o problema dual

$$\inf\left\{\int_{\mathbb{R}^n}\varphi d\mu + \int_{\mathbb{R}^n}\psi d\nu;\ \forall(x,y),\ \langle x,y\rangle \le \varphi(x)+\psi(y)\right\}.$$

(ii) ***(Teorema de Brenier)*** *Se μ não atribui massa para small sets, então existe um único π ótimo, dado por*

$$d\pi(x,y) = d\mu(x)\delta[y = \nabla\varphi(x)], \tag{3.50}$$

ou equivalentemente,

$$\pi = (Id \times \nabla\varphi)\#\mu, \tag{3.51}$$

onde $\nabla\varphi$ é o único (isto é, unicidade determinada em $d\mu$-quase toda parte) gradiente de uma função convexa tal que $\nabla\varphi\#\mu = \nu$. Além disso,

$$Supp(\nu) = \overline{\nabla\varphi(Supp(\mu))}.$$

(iii) Como um corolário, sob as mesmas hipóteses de (ii), $\nabla\varphi$ é a única solução para o problema do transporte de Monge:

$$\int_{\mathbb{R}^n}\|x-\nabla\varphi(x)\|^2 d\mu(x) = \inf_{T\#\mu=\nu}\int_{\mathbb{R}^n}\|x-T(x)\|^2 d\mu(x),$$

ou equivalentemente,

$$\int_{\mathbb{R}^n}\langle x, \nabla\varphi(x)\rangle\, d\mu(x) = \sup_{T\#\mu=\nu}\int_{\mathbb{R}^n}\langle x, T(x)\rangle\, d\mu(x).$$

Demonstração. Vamos primeiro mostrar a equivalência entre (3.50) e (3.51).

Afirmação. Seja $T : \mathbb{R}^n \to \mathbb{R}^n$ uma função mensurável.

a) se $\nu = T\#\mu$, então $(Id \times T)\#\mu \in \Pi(\mu,\nu)$. Além disso, $[d(Id\times T)\#\mu$-quase toda parte $y = T(x)]$.

b) se $\pi \in \Pi(\mu,\nu)$ é tal que $[d\pi -$ quase toda parte $y = T(x)]$, então $\pi = (Id \times T)\#\mu$.

Prova da afirmação. **(a)** De fato, para todos $A, B \in \mathcal{B}(\mathbb{R}^n)$,

$$(Id\times T)\#\mu[A\times\mathbb{R}^n] = \mu[(Id\times T)^{-1}(A\times\mathbb{R}^n)] = \mu[A],$$

onde usamos na primeira igualdade a definição de $(Id \times T)\#\mu$.

Como $\nu = T\#\mu$, obtemos que

$$(Id \times T)\#\mu[\mathbb{R}^n \times B] = \mu[(Id \times T)^{-1}(\mathbb{R}^n \times B)] = \mu[T^{-1}(B)] = \nu[B].$$

Para segunda parte, defina $D = \text{Graph}(T)$ e note que

$$(Id \times T)\#\mu[D] = \mu(\mathbb{R}^n) = 1.$$

Isto prova o item (a) da afirmação.

(b) Suponha que $[d\pi - \text{quase toda parte } y = T(x)]$. Então para qualquer $C \in \mathcal{B}(\mathbb{R}^{2n})$,

$$\pi[C] = \pi[C \cap \text{Graph}(T)].$$

Por outro lado, defina $A = \{x \in \mathbb{R}^n : (x, T(x)) \in C\} = (Id \times T)^{-1}(C)$. Logo, como $\pi \in \Pi(\mu, \nu)$,

$$\mu[(Id \times T)^{-1}(C)] = \mu[A] = \pi[A \times \mathbb{R}^n] = \pi[C \cap \text{Graph}(T)] = \pi[C].$$

Isto prova o item (b) da nossa afirmação.

Imediatamente da afirmação acima segue que se $\pi \in \Pi(\mu, \nu)$ e $\nu = T\#\mu$, então

$$[d\pi - \text{quase toda parte } y = T(x)] \Leftrightarrow \pi = (Id \times T)\#\mu.$$

Em outras palavras, acabamos de provar a equivalência entre (3.50) e (3.51).

Procedendo como na seção 3.3, reduzimos o problema de minimização de Monge-Kantorovich e sua formulação dual para o problema em (3.7).

(i) ($\Longrightarrow$) Seja $\pi \in \Pi(\mu, \nu)$ uma plano de transferência ótimo. A existência de tal plano segue da proposição 3.2.1. Pelo teorema 3.5.7, existe uma par de funções convexas (φ, φ^*), ótimo para o problema dual. Então pela identidade (3.7),

$$\begin{aligned}\int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x, y \rangle \, d\pi(x, y) &= \int_{\mathbb{R}^n} \varphi(x) d\mu(x) + \int_{\mathbb{R}^n} \varphi^*(y) d\nu(y) \\ &= \int_{\mathbb{R}^n \times \mathbb{R}^n} [\varphi(x) + \varphi^*(y)] \, d\pi(x, y),\end{aligned}$$

onde na segunda igualdade usamos o lema 2.2.5. Equivalentemente,

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} [\varphi(x) + \varphi^*(y) - \langle x, y \rangle] \, d\pi(x, y) = 0.$$

Pela desigualdade (3.16), vemos que a função dentro do colchete é não-negativa, de forma que ela deve se anular para $d\pi$ quase todo (x, y). Pela proposição 3.4.4, obtemos (3.49).

($\Longleftarrow$) Suponha que exista uma função convexa φ tal que $\pi \in \Pi(\mu, \nu)$ satisfaz (3.49). Então pelos mesmos argumentos,

$$\begin{aligned}\sup_{\pi \in \Pi(\mu,\nu)} \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x, y \rangle \, d\pi(x,y) \right\} &\geq \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x, y \rangle \, d\pi(x,y) \\ &= \int_{\mathbb{R}^n \times \mathbb{R}^n} [\varphi(x) + \varphi^*(y)] \, d\pi(x,y) \\ &\geq \inf_{(\tilde{\varphi}, \tilde{\psi}) \in \tilde{\Phi}_c} \left\{ J(\tilde{\varphi}, \tilde{\psi}) \right\}.\end{aligned}$$

A partir da identidade (3.7), deduzimos que o plano de transferência π e o par (φ, φ^*) são ótimos. Isto conclui a prova de (i).

(ii) A prova será dividida em 3 passos.

Passo 1. Assuma que μ atribua medida nula aos small sets. Seja $\pi \in \Pi(\mu, \nu)$ um plano ótimo e seja φ a função convexa dada pelo item (i). Como $\varphi \in \mathcal{L}^1(\mu)$, ela é $d\mu$-quase toda parte finita: $\mu[\text{Dom}(\varphi)] = 1$. Por outro lado, pelo exemplo 3.4.2, $\partial\text{Dom}(\varphi)$ é um small set, de forma que

$$1 = \mu[\text{Dom}\varphi] \leq \mu[\text{Int}(\text{Dom}(\varphi))] + \mu[\partial\text{Dom}(\varphi)] = \mu[\text{Int}(\text{Dom}(\varphi))].$$

Agora, no $\text{Int}(\text{Dom}(\varphi))$, o conjunto de não diferenciabilidade de φ é um small set. Logo, $\mu[\{x \in \text{Int}(\text{Dom}(\varphi)) : x \text{ é um ponto de diferenciabilidade de } \varphi\}] = 1$. Assim, para $d\mu$-quase todo ponto x, o subdiferencial de φ em um ponto x é $\{\bigtriangledown\varphi(x)\}$ (aqui usamos fortemente a condição de que no $\text{Int}(\text{Dom}(\varphi))$ o subdiferencial é não vazio). Observando que uma afirmação válida $d\mu$-quase todo x é também verdadeira para $d\pi$-quase todo par (x, y), obtemos, pelo item (i), que $y = \bigtriangledown\varphi(x)$ para $d\pi$- quase todo par (x, y).

Passo 2. Até aqui provamos que qualquer plano de transferência é da forma $(Id \times \bigtriangledown\varphi)\#\mu$ para alguma função convexa φ tal que $\bigtriangledown\varphi\#\mu = \nu$, e que existe pelo menos um tal plano de transferência. Agora mostraremos a unicidade do plano ótimo π e também a unicidade do gradiente $\bigtriangledown\varphi$ de uma função convexa tal que $\bigtriangledown\varphi\#\mu = \nu$.

Sejam $\bar{\pi}$ outro plano de transferência ótimo e $\bar{\varphi}$ a sua função convexa associada tal que $\bigtriangledown\bar{\varphi}\#\mu = \nu$. Considere o par ótimo $(\bar{\varphi}, \bar{\varphi}^*)$ para o problema dual, assim como (φ, φ^*). Logo,

$$\int_{\mathbb{R}^n} \bar{\varphi} d\mu + \int_{\mathbb{R}^n} \bar{\varphi}^* d\nu = \inf_{\tilde{\Phi}_c} J = \int_{\mathbb{R}^n} \varphi d\mu + \int_{\mathbb{R}^n} \varphi^* d\nu. \tag{3.52}$$

Pelo lema 2.2.6, o lado esquerdo de (3.52) é igual a

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} [\bar{\varphi}(x) + \bar{\varphi}^*(y)] \, d\pi(x,y),$$

enquanto o lado direito de (3.52), do fato de π ser um plano ótimo associado a φ e de (3.7), é igual a

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} \langle x, y\rangle \, d\pi(x,y).$$

Portanto

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} [\bar{\varphi}(x) + \bar{\varphi}^*(y)] \, d\pi(x,y) \quad = \quad \int_{\mathbb{R}^n\times\mathbb{R}^n} \langle x, y\rangle \, d\pi(x,y).$$

Como $\pi = (Id \times \bigtriangledown\varphi)\#\mu$, a equação acima pode ser reescrita como

$$\int_{\mathbb{R}^n} [\bar{\varphi}(x) + \bar{\varphi}^*(\bigtriangledown\varphi(x))] \, d\mu(x) = \int_{\mathbb{R}^n} \langle x, \bigtriangledown\varphi(x)\rangle \, d\mu(x).$$

Assim,

$$\int_{\mathbb{R}^n} [\bar{\varphi}(x) + \bar{\varphi}^*(\bigtriangledown\varphi(x)) - \langle x, \bigtriangledown\varphi(x)\rangle] \, d\mu(x) = 0.$$

Como a definição de transformada de Legendre nos diz que o integrando é não-negativo, vemos que ele deve se anular $d\mu$-quase toda parte. Usando a proposição 3.4.4, concluímos que

$$\bigtriangledown\varphi(x) \in \partial\bar{\varphi}, \quad \text{para } d\mu - \text{quase todo } x.$$

Como $\bar{\varphi}$ é diferenciável $d\mu$-quase toda parte, concluímos que

$$\bigtriangledown\varphi(x) = \bigtriangledown\bar{\varphi}(x), \quad \text{para } d\mu - \text{quase todo } x.$$

Isto prova tanto a unicidade do gradiente de uma função convexa $\bigtriangledown\varphi$ tal que $\bigtriangledown\varphi\#\mu = \nu$, como a unicidade do plano de transferência ótimo π.

Passo 3. Agora vamos provar que $\mathrm{Supp}(\nu) = \overline{\bigtriangledown\varphi(\mathrm{Supp}(\mu))}$. Seja $x \in \mathrm{Supp}(\mu)$ um ponto de diferenciabilidade para φ, e $y = \bigtriangledown\varphi(x)$. Pela continuidade do gradiente (ver (3.12)), nós sabemos que para qualquer $\varepsilon > 0$ existe um $\delta > 0$ com $\bigtriangledown\varphi(B_\delta(x)) \subset B_\varepsilon(y)$. Em particular,

$$\nu[B_\varepsilon(y)] \geq \nu[\bigtriangledown\varphi(B_\delta(x))] = \mu[\bigtriangledown\varphi^{-1}(\bigtriangledown\varphi(B_\delta(x)))] \geq \mu[(B_\delta(x)],$$

onde na igualdade usamos que $\nu = \bigtriangledown\varphi\#\mu$. Como x pertence ao suporte de μ, $\mu[(B_\delta(x)] > 0$ e, portanto, a partir da desigualdade acima, concluímos que $\nu[B_\varepsilon(y)] > 0$. Como ε foi arbitrário, deduzimos que $y \in \mathrm{Supp}(\nu)$. Isto mostra que

$$\bigtriangledown\varphi(\mathrm{Supp}(\mu)) \subset \mathrm{Supp}(\nu). \tag{3.53}$$

Por outro lado,

$$\nu[\triangledown\varphi(\mathrm{Supp}(\mu))] = \mu[\triangledown\varphi^{-1}(\triangledown\varphi(\mathrm{Supp}(\mu)))] \geq \mu[\mathrm{Supp}(\mu)] = 1.$$

Logo, temos que ν está concentrado em $\triangledown\varphi(\mathrm{Supp}(\mu))$, e portanto (pela definição de suporte de uma medida)

$$\mathrm{Supp}(\nu) \subset \overline{\triangledown\varphi(\mathrm{Supp}(\mu)}. \tag{3.54}$$

Combinando (3.53) e (3.54), deduzimos que $\mathrm{Supp}(\nu) = \overline{\triangledown\varphi(\mathrm{Supp}(\mu))}$.

(iii) Relembremos que o problema de transporte de Monge é o problema de transporte de Kantorovich quando os planos de transporte tem a formal especial

$$d\pi_T(x,y) = d\mu(x)\delta_{[y=T(x)]}, \tag{3.55}$$

onde $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ é uma aplicação mensurável.

Assim, para qualquer medida de probabilidade π_T, temos

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} \xi(x,y)d\pi_T(x,y) = \int_{\mathbb{R}^n} \xi(x,T(x))d\mu(x), \tag{3.56}$$

para qualquer função mensurável definida em $\mathbb{R}^n \times \mathbb{R}^n$. Em particular, o custo associado a estes planos é

$$I[T] = I[\pi_T] = \int_{\mathbb{R}^n} c(x,T(x))d\mu(x).$$

Afirmamos que se π_T pertence a $\Pi(\mu,\nu)$, então T em (3.55) é tal que $\nu = T\#\mu$. De fato, por (3.56) e pelo lema 2.2.6,

$$\begin{aligned}\int_{\mathbb{R}^n}[\varphi(x)+\psi\circ T(x)]d\mu(x) &= \int_{\mathbb{R}^n}[\varphi(x)+\psi(y)]d\pi_T(x,y)\\ &= \int_{\mathbb{R}^n}\varphi(x)d\mu(x) + \int_{\mathbb{R}^n}\psi(y)d\nu(y),\end{aligned}$$

para todo par $(\varphi,\psi) \in C_b(\mathbb{R}^n) \times C_b(\mathbb{R}^n)$. Cancelando $\int_{\mathbb{R}^n}\varphi(x)d\mu(x)$ em ambos os lados, obtemos a condição

$$\int_{\mathbb{R}^n}(\psi\circ T(x))d\mu(x) = \int_{\mathbb{R}^n}\psi(y)d\nu(y), \tag{3.57}$$

que, conforme lembrado no início desta seção, é equivalente a $\nu = T\#\mu$.

Assim, aplicando o item (ii) a este caso particular, vemos que existe único $\triangledown\varphi$ tal que

$$\int_{\mathbb{R}^n}\langle x, \nabla\varphi(x)\rangle\, d\mu(x) = \sup_{\pi_T\in\Pi(\mu,\nu)}\int_{\mathbb{R}^n}\langle x, T(x)\rangle\, d\mu(x) = \sup_{T\#\mu=\nu}\int_{\mathbb{R}^n}\langle x, T(x)\rangle\, d\mu(x).$$

$\square$

Como aplicação direta do teorema de Brenier, obtemos o

Exemplo 3.6.4. Sejam μ e ν duas medidas de probabilidade, em $P(\mathbb{R})$, quaisquer. Se μ não atribui massa a conjuntos de dimensão de Hausdorff zero, então existe um único $\nabla\varphi$ (unicidade determinada em quase todos os pontos) de uma função convexa φ, tal que $\nu = \nabla\varphi\#\mu$. Além disso, o par (φ, φ^*) satisfaz

$$J(\varphi, \varphi^*) = \inf\{J(\varphi, \psi) : \forall(x, y),\ \langle x, y\rangle \leq \varphi(x) + \psi(y)\}.$$

Agora vamos analisar um exemplo interessante de transporte entre duas medidas.

Exemplo 3.6.5. Consideremos o problema de transportar duas medidas sem átomos definidas em $\mathbb{R}$ mas com suportes compactos. Para simplificar pensemos que ambas são suportadas no intervalo $[0, 1]$ (a generalização a intervalos fechados quaisquer é muito simples). É fácil ver que se μ é a medida de Lebesgue em $[0, 1]$ e ν é dada por uma densidade f então uma função $\theta : [0, 1] \to [0, 1]$ que transporta μ em ν (i.e. $\nu = \theta\#\mu$) é a função que satisfaz, para $x \in [0, 1]$,

$$\int_0^{\theta(x)} f(z)dz = x\,.$$

No caso mais geral em que μ e ν são apenas equivalentes a Lebesgue (i.e. tem densidades estritamente positivas em $[0, 1]$), então θ deve satisfazer a equação

$$\nu([0, \theta(x)]) = \mu([0, x])\,.$$

θ assim definida será um homeomorfismo de $[0, 1]$ em $[0, 1]$ (de fato pode-se mostrar que θ e sua inversa são Holder continuas). Ainda, e mais importante, tal transporte é a solução de Monge para o problema de transporte entre μ e ν, minimizando o custo médio quadrático. Uma demonstração este fato, em um contexto mais geral, pode ser encontrada no teorema 2.18 de **Villani**, [11].

Apêndice - Algumas Demonstrações

Lema 3.6.6. *Seja $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função convexa. Então para todo $x \in$ Int(Dom(φ)), o subdiferencial $\partial\varphi(x)$ é não-vazio.*

Durante a demonstração deste lema faremos uso dos seguintes resultados:

- Uma função convexa própria é automaticamente contínua no interior do seu domínio;
- **A primeira forma geométrica de Hahn-Banach:** Sejam dois conjuntos convexos e não vazios A e B contidos em um espaço vetorial E, tal que $A \cap B = \emptyset$. Se A é aberto, então existe um hiperplano fechado que separa A e B;
- Se $C \subset E$ é um conjunto convexo, então $\mathrm{Int}(C)$ é convexo. Se além disso $\mathrm{Int}(C) \neq \emptyset$, então $\overline{C} = \overline{\mathrm{Int}(C)}$.

Demonstração do Lema 3.6.6. Seja $x_0 \in \mathrm{Int}(\mathrm{Dom}(\varphi))$. Sendo φ contínua em x_0, existe $\delta > 0$ tal que

$$x \in B(x_0, \delta) \Rightarrow |\varphi(x) - \varphi(x_0)| < 1. \tag{3.58}$$

Assim, o conjunto $D = B(x_0, \delta) \times (\varphi(x_0) + 1, +\infty)$ é um aberto contido no epigráfico de φ. Isto mostra que $A = \mathrm{Int}(\mathrm{epi}(\varphi))$ é não vazio. Como φ é convexa, então A também é convexo (aqui usamos o terceiro resultado acima citado).

Por definição de epigráfico, temos que o ponto $(x_0, \varphi(x_0)) \in \partial\mathrm{epi}(\varphi)$, de forma que o conjunto convexo $B = \{(x_0, \varphi(x_0))\}$ é disjunto de A. Pela primeira forma geométrica de Hahn-Banach, existe um funcional linear não-nulo $\Psi \in (\mathbb{R}^{n+1})^*$ e um escalar $\alpha \in \mathbb{R}$ tais que

$$\Psi(x, \lambda) \geq \alpha \geq \Psi(x_0, \varphi(x_0)), \quad \forall (x, \lambda) \in A, \tag{3.59}$$

Por continuidade, podemos substituir em (3.59), A por $\overline{A} = \overline{\mathrm{epi}(\varphi)}$ (novamente fizemos uso do terceiro resultado acima citado). Em particular, temos

$$\Psi(x, \lambda) \geq \alpha \geq \Psi(x_0, \varphi(x_0)), \quad \forall (x, \lambda) \in \mathrm{epi}(\varphi). \tag{3.60}$$

Pelo teorema de Reisz, existe um vetor v em $\mathbb{R}^n$ e $k \in \mathbb{R}$ tais que $\Psi(x, \lambda) = \langle v, x \rangle + k\lambda$ para todo $(x, \lambda) \in \mathbb{R}^{n+1}$. Logo, (3.60) pode ser reescrita como

$$\langle v, x \rangle + k\lambda \geq \alpha \geq \langle v, x_0 \rangle + k\varphi(x_0), \quad \forall (x, \lambda) \in \text{epi}(\varphi). \tag{3.61}$$

Escolhendo $x = x_0$ em (3.61) e fazendo $\lambda \to +\infty$, deduzimos que $k \geq 0$.

Afirmação. $k > 0$.

Prova da afirmação. Suponhamos, por absurdo, que $k = 0$. Imediatamente de (3.61), vemos que

$$\langle v, x \rangle \geq \alpha \geq \langle v, x_0 \rangle, \quad \forall x \in \text{Dom}(\varphi). \tag{3.62}$$

Como $\Psi \not\equiv 0$, temos que v é não nulo. A partir de (3.58), temos que $B(x_0, \delta) \subset \text{Dom}(\varphi)$. Logo, de (3.62), obtemos que

$$\langle v, x_0 + \delta z \rangle \geq \alpha \geq \langle v, x_0 \rangle, \quad \forall z \in B(0, 1).$$

Estas desigualdades são equivalentes a

$$\alpha \geq \langle v, x_0 \rangle \geq \alpha - \delta \langle v, z \rangle, \quad \forall z \in B(0, 1).$$

Portanto, subtraindo α, trocando o sinal e usando o fato de que $\delta > 0$, deduzimos que

$$\langle v, z \rangle \geq 0, \forall z \in B(0, 1).$$

Trocando z por $-z$, concluímos que $\langle v, z \rangle = 0, \forall z \in B(0, 1)$, e consequentemente que $\langle v, z \rangle = 0$, $\forall z \in \mathbb{R}^n$, o que força $v = 0$, uma contradição. Isto completa nossa afirmação.

Fazendo $\lambda = \varphi(x)$ em (3.61) obtemos,

$$\varphi(x) \geq \varphi(x_0) + \left\langle -\frac{v}{k}, x - x_0 \right\rangle, \quad \forall x \in \text{Dom}(\varphi). \tag{3.63}$$

Como para $x \in (\text{Dom}(\varphi))^c$, $\varphi(x) = +\infty$, temos que desigualdade vale trivialmente, e portanto vemos que o vetor $-\frac{v}{k} \in \partial\varphi(x_0)$. $\square$

Lema 3.6.7. *Seja $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função convexa. Se x é um ponto de diferenciabilidade, então vale (3.11). Além disso, $\partial\varphi(x) = \{\nabla\varphi(x)\}$.*

Demonstração. Para quaisquer $z \in \mathbb{R}^n$ e $\lambda \in (0, 1)$, a convexidade de φ implica

$$\begin{aligned} \varphi(x + \lambda(z - x)) &= \varphi(\lambda z + (1 - \lambda)x) \\ &\leq \lambda\varphi(z) + (1 - \lambda)\varphi(x) = \varphi(x) + \lambda(\varphi(z) - \varphi(x)). \end{aligned} \tag{3.64}$$

Por definição de diferenciabilidade, existe uma vizinhança U_x de x que para todo h tal que $x + h \in U_x$,

$$\varphi(x+h) - \varphi(x) = \langle \bigtriangledown\varphi(x), h\rangle + r(h),$$

onde $\lim_{h\to 0} \frac{r(h)}{\|h\|} = 0$.

Fazendo na relação acima $h = \lambda(z - x)$, para λ suficientemente pequeno e positivo, obtemos de (3.64) que

$$\begin{aligned} \varphi(z) - \varphi(x) &\geq \frac{\varphi(x + \lambda(z-x)) - \varphi(x)}{\lambda} \\ &= \langle \bigtriangledown\varphi(x), z - x\rangle + r(\lambda(z-x))/\lambda, \quad \forall z \in \mathbb{R}^n. \end{aligned} \tag{3.65}$$

Afirmação 1. Para qualquer $z \in \mathbb{R}^n$, $\lim_{\lambda\to 0} r(\lambda(z-x))/\lambda = 0$.

Prova da afirmação 1. De fato, $\lim_{\lambda\to 0} r(\lambda(z-x))/\lambda = \lim_{h\to 0} \frac{r(h)}{\|h\|}\|h\| = 0$, onde na primeira igualdade fizemos a mudança $h = \lambda(z - x)$ e na segunda usamos a definição de $r(h)$.

Como consequência da afirmação 1 e de (3.65), temos que $\forall z \in \mathbb{R}^n$,

$$\varphi(z) \geq \varphi(x) + \langle \bigtriangledown\varphi(x), z - x\rangle,$$

o que nos mostra a validade de (3.11) que, por sua vez, implica em $\bigtriangledown\varphi(x) \in \partial\varphi(x)$.

Afirmação 2. $\partial\varphi(x)$ contém apenas um ponto.

Prova da afirmação 2. Seja $y \in \partial\varphi(x)$. Então por definição de subdiferencial,

$$\varphi(x + \lambda z) - \varphi(x) \geq \langle y, x + \lambda z - x\rangle = \lambda \langle y, z\rangle, \quad \forall z \in \mathbb{R}^n,\ \forall \lambda \in (0,1). \tag{3.66}$$

Dividindo por λ a desigualdade acima e fazendo $\lambda \to 0$ vemos que

$$\langle \bigtriangledown\varphi(x), z\rangle \geq \langle y, z\rangle, \quad \forall z \in \mathbb{R}^n,$$

uma vez que x é um ponto de diferenciabilidade. Equivalentemente,

$$\langle \bigtriangledown\varphi(x) - y, z\rangle \geq 0, \quad \forall z \in \mathbb{R}^n. \tag{3.67}$$

Trocando z por $-z$ na relação acima, deduzimos que $\langle \bigtriangledown\varphi(x) - y, z\rangle = 0$, $\forall z \in \mathbb{R}^n$, o que força $y = \bigtriangledown\varphi(x)$. Isto completa afirmação 2 e o lema. □

Lema 3.6.8. *Seja* $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função convexa semicontínua inferiormente. Então para cada sequência* $(x_k)_{k\in\mathbb{R}}$ *com* $x_k \to x$*, tem-se*

$$\liminf_{k\to\infty} \varphi(x_k) \geq \varphi(x).$$

Demonstração. Como φ é uma função semicontínua inferiormente, então para todo $\lambda \in \mathbb{R}$, $[\varphi \leq \lambda]$ é um conjunto fechado.

Afirmação. Para todo x e para todo $\varepsilon > 0$, existe uma vizinhança $U_x(\varepsilon)$ de x tal que

$$\varphi(y) \geq \varphi(x) - \varepsilon, \quad \forall y \in \ U_x(\varepsilon).$$

Prova da afirmação. Vamos separar a prova em 2 casos.

Caso 1: Se $x \in \mathrm{Dom}(\varphi)$, então defina $\lambda = \varphi(x) - \varepsilon$. Logo, $x \in [\varphi > \lambda]$ aberto, de forma que existe vizinhança $U_x(\varepsilon)$ de x contida em $[\varphi > \lambda]$. Assim, para todo $y \in U_x(\varepsilon)$, $\varphi(y) > \varphi(x) - \varepsilon$.

Caso 2: Se $x \in (\mathrm{Dom}(\varphi))^c$, então para todo $\lambda \in \mathbb{R}$, $x \notin [\varphi \leq \lambda]$. Logo, temos que $x \notin F = \bigcap\limits_{\lambda \in \mathbb{R}} [\varphi \leq \lambda]$ fechado. Assim, $x \in F^c$ aberto, de forma que existe vizinhança $U_x(\varepsilon)$ de x tal que $\varphi(y) = +\infty$, para todo $y \in U_x(\varepsilon)$. Isto conclue a afirmação.

Considere $x_k \to x$. Então pela afirmação, para todo k natural existe uma vizinhança $U_x(k)$ de x tal que $\varphi(y) \geq \varphi(x) - 1/k$, para todo $y \in U_x(k)$. Sem perda de generalidade, suponhamos que $x_k \in U_n(k)$. Assim, para todo k,

$$\varphi(x_k) \geq \varphi(x) - 1/k \Longrightarrow \liminf_{k \to \infty} \varphi(x_k) \geq \varphi(x).$$

□

Lema 3.6.9. *Seja $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função convexa semicontínua inferiormente. Então $G(\partial\varphi)$ é um conjunto fechado no seguinte sentido:*

$$\left. \begin{array}{r} x_k \longrightarrow x \\ \partial\varphi(x_k) \ni y_k \longrightarrow y \end{array} \right\} \Longrightarrow y \in \partial\varphi(x). \tag{3.68}$$

Demonstração. Suponha que $x_k \to x$ e $\partial\varphi(x_k) \ni y_k \to y$. Pela definição de subdiferencial,

$$\varphi(z) \geq \varphi(x_k) + \langle y_k, z - x_k \rangle, \quad \forall z \in \mathbb{R}^n,\ \forall k \in \mathbb{N}.$$

Pelo lema 3.6.8, temos que $\forall z \in \mathbb{R}^n$,

$$\varphi(z) \geq \liminf_{k \to \infty} [\varphi(x_k) + \langle y_k, z - x_k \rangle] \geq \varphi(x) + \langle y, z - x \rangle,$$

onde usamos que $x_k \to x$ e $y_k \to y$. □

Lema 3.6.10. *Suponha que para todo $(x, y, k) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{N}$ as seguintes relações se verificam:*

$$0 \leq \varphi_k(x) + \frac{\|x\|^2}{2}, \quad 0 \leq \psi_k(y) + \frac{\|y\|^2}{2},$$

e

$$\left[\varphi_k(x) + \frac{\|x\|^2}{2}\right] + \left[\psi_k(y) + \frac{\|y\|^2}{2}\right] \geq \frac{\|x+y\|^2}{2} \geq 0.$$

Então todo $(x, y, k) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{N}$,

$$O_k^{(l)}(x) + \tilde{O}_k^{(l)}(y) \geq \min\left(\varphi_k(x) + \frac{\|x\|^2}{2} + \psi_k(y) + \frac{\|y\|^2}{2}, l\right) \geq \min\left(\frac{\|x+y\|^2}{2}, l\right). \quad (3.69)$$

Demonstração. A segunda desigualdade em (3.69) é imediata a partir da terceira relação.

A prova da primeira desigualdade será divida em 2 casos:

Caso 1: suponhamos que

$$\min\left(\varphi_k(x) + \frac{\|x\|^2}{2} + \psi_k(y) + \frac{\|y\|^2}{2}, l\right) = \varphi_k(x) + \frac{\|x\|^2}{2} + \psi_k(y) + \frac{\|y\|^2}{2}.$$

Então segue, por exemplo, que

$$l \geq \varphi_k(x) + \frac{\|x\|^2}{2} + \psi_k(y) + \frac{\|y\|^2}{2} \geq \varphi_k(x) + \frac{\|x\|^2}{2},$$

onda na segunda desigualdade usamos a primeira relação.

Desta forma, temos que $O_k^{(l)}(x) = \varphi_k(x) + \frac{\|x\|^2}{2}$. Analogamente, deduzimos que, neste caso, $\tilde{O}_k^{(l)}(y) = \psi_k(y) + \frac{\|y\|^2}{2}$. Isto mostra que o lema é válido para o caso 1.

Caso 2: suponhamos que $\min\left(\varphi_k(x) + \frac{\|x\|^2}{2} + \psi_k(y) + \frac{\|y\|^2}{2}, l\right) = l$. Isto significa que

$$l \leq \varphi_k(x) + \frac{\|x\|^2}{2} + \psi_k(y) + \frac{\|y\|^2}{2}.$$

Imediatamente das suas definições, segue que $l \leq O_k^{(l)}(x) + \tilde{O}_k^{(l)}(y)$, o que completa a demonstração do segundo caso e do lema. $\square$

Referências Bibliográficas

[1] Balder, E. J., *An extension of the Dunford-Pettis theorem with applications to a bocce-type oscillation restriction creterion*, University of Utrecht, 1997.

[2] Bartle, R., *The Elements of Integration and Lebesgue Measure*, Wiley, New York, 1995.

[3] Brezis, H., *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer, New York, 2010.

[4] Billingsley, P., *Convergence of Probability Measure*, Wiley, Chicago, 1999.

[5] Billingsley, P., *Probability and Measure*, Wiley, Chicago, 1995.

[6] Fernandez, P., *Medida e Integração*, IMPA, Rio de Janeiro, 1996.

[7] Fitzpatrick, P., *Real Analysis*, China Machine Press, China, 2010.

[8] Lima, E. L., *Espaços métricos*, IMPA, Rio de Janeiro, 2007.

[9] Oliveira, A. D., *O Teorema da Dualidade de Kantorovich para o Transporte Ótimo*, dissertação de mestrado, UFRGS, 2011.

[10] Rockafellar, R. T., *Convex Analysis*, Princeton University Press, New Jersey, 1972.

[11] Villani, C., *Topics in Optimal Transportation*, American Mathematical Society, 2003.

[12] Villani, C., *Optimal transport, old and new*, Springer, 2008.