Capítulo 3

Existência de Mapas de Transporte ótimo

Nesse capítulo, retornamos à formulação de Monge e estudaremos a existência e propriedades de mapas de transporte ótimo, ou seja a existência de minimisantes para o problema de Monge. Recordamos que a motivação principal para introduzir a formulação de Kantorovitch é a dificuldade em aplicar o método direto do cálculo das variações para a formulação de Monge. Isso é uma estratégia frequente no cálculo das variações; quando o método direto falha, procuramos uma *relaxação semi-contínua inferiormente* do problema para a qual o método direto é eficaz, e em seguida usamos as condições de otimalidade para identificar situações onde as soluções dessa relaxação correspondem na verdade à soluções do problema original. No contexto do problema de Kantorovitch, as condições de otimalidade que nos premitirão voltar ao problema de Monge é justamente a teoria de c-monotonicidade cíclica desenvolvida no último capítulo. Nesse capítulo vamos discutir como explorar essa teoria quando os espaços $\mathcal{X}$ e $\mathcal{Y}$ são (subconjuntos de) $\mathbb{R}^d$.

3.1 Consequências da c-monotonicidade cíclica

Recordemos o problema de Kantorovitch: dado um custo $c : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R} \cup \{+\infty\}$ e $\mu, \nu\mathscr{P}(\mathbb{R}^d)$ consideramos

$$\min_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} c\mathrm{d}\gamma = \sup_{\varphi\oplus\psi\leq c} \int_{\mathbb{R}^d} \varphi\mathrm{d}\mu + \int_{\mathbb{R}^d} \psi\mathrm{d}\nu, \tag{3.1}$$

onde a igualdade é a dualidade de Kantorovitch já provada. Nesse capítulo vamos supor sempre que o mínimo é atingido, para isso basta supor por exemplo que

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} c\mathrm{d}\mu\otimes\nu < +\infty,$$

logo existe um plano de transporte ótimo, assim como o supremo que define o problema dual, existem potenciais ótimos de Kantorovitch (φ, ψ).

Sabemos do Teorema 2.3.1 que para γ-quase todo ponto $(x, y) \in \operatorname{supp}\gamma$, vale que

$$\varphi(x) + \psi(y) = c(x, y). \tag{3.2}$$

Assumindo regularidade o suficiente dos potenciais de Kantorovitch, podemos diferenciar ambos os lados dessa equação com respeito à x enquanto mantemos y fixo. Dessa forma, temos que

$$\nabla\varphi(x) = \nabla_x c(x, y). \tag{3.3}$$

Assumindo que $y \mapsto \nabla_x c(x, y)$ é invertível, o ponto y será necessariamente dado por

$$y = (\nabla_x c(x, \cdot))^{-1}(\nabla\varphi(x)). \tag{3.4}$$

A inversibilidade de $\nabla_x c(x, \cdot)$ é uma consequência, por exemplo, da hipótese conhecida como *condição de torção*

$$\det D^2_{x,y} c(x, y) \neq 0.$$

Um caso mais simples, e importante é quando $c(x, y) = h(x - y)$, onde $h : \mathbb{R}^d \to \mathbb{R}$ é uma função estritamente convexa. Vamos provar à frente que se h é estritamente convexa, então seu gradiente ∇h está bem definido em $\mathscr{L}^d$-quase todo ponto de $\mathbb{R}^d$.[1]

O mesmo cálculo de antes nesse caso nos dá

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

O caso mais simples, e talvez o mais importante, é o custo quadrático quando $c(x, y) = \frac{1}{2}|x - y|^2$. Neste caso, obtemos o teorema de Brenier, que afirma que a aplicação de transporte ótimo é dada por

$$T = \mathrm{id} - \nabla\varphi = \nabla u, \tag{3.5}$$

e u é uma função convexa.

De modo geral, podemos definir uma aplicação

$$T \overset{\text{def.}}{=} \mathrm{id} - (\nabla h)^{-1}(\nabla\varphi), \tag{3.6}$$

com a qual obtemos heuristicamente a seguinte propriedade: assumindo que o gradiente $\nabla\varphi$ é bem definido, para quase todo ponto $(x, y) \in \operatorname{supp}\gamma$, segue que $y = T(x)$.

Portanto, para provar a existência de um mapa de transporte ótimo precisamos ou provar que os potenciais de Kantorovitch são muito regulares, por exemplo $\mathscr{C}^1$ para que os gradientes estejam sempre bem definidos, ou ao menos balancear alguma regularidade mais fraca dos potenciais de Kantorovitch com hipóteses sobre a medida de partida μ para que a aplicação T definida em (3.6) seja bem definida em μ-quase todo ponto.

No caso mais simples do custo quadrático, a primeira alternativa é diretamente ligada à questão de continuidade do transporte ótimo [2], que além de ser consideravelmente mais difícil, é falsa em situações muito razoáveis, como aprendemos desde o contra-exemplo proposto por Caffarelli.

A segunda alternativa requer explorar a regularidade que obtemos diretamente da teoria de c-monotonicidade cíclica, onde os potenciais herdam frequentemente a regularidade Lipschitz do custo c ou são funções convexas. Nesses casos podemos provar com os teoremas de Rademacher e/ou de Alexandrov que garante a existência do gradiente em $\mathscr{L}^d$-quase todo ponto.

[1] Na verdade, o conjunto onde ele não está definido é uma união enumerável de superfícies $\mathscr{C}^1$ de dimensão $d - 1$. Chamamos isso de um conjunto $(d - 1)$-rectifiável.

[2] Note que φ de ordem $\mathscr{C}^1$ implicaria um mapa de transporte contínuo.

3.2 Diferenciabilidade de funções convexas

No capítulo 2 nós vimos que funções convexas sempre admitem uma noção mais fraca de gradiente, os subgradientes, que se tornam o conjunto

$$\partial f(x) = \{p : f(y) \geq f(x) + \langle p, y - x \rangle \text{ para todo } y\},$$

sempre que f é uma função convexa. Esse conjunto é não vazio $\partial f(x) \neq \emptyset$ sempre que $x \in \operatorname{dom} f$.

Exercício 3.1. Prove que se $\partial f(x)$ é um conjunto unitário, se e somente se, o gradiente $\nabla f(x)$ existe e $\partial f(x) = \{\nabla f(x)\}$.

Agora vamos ver que $\nabla f(x)$ está bem definido para $\mathscr{L}^d$-quase todo ponto $x \in \mathbb{R}^d$. Para isso vamos usar o teorema de Rademacher para funções Lipschitz. Lembramos que $f : U \to \mathbb{R}$, onde U é um subconjunto aberto de $\mathbb{R}^d$, é Lipschitz quando existe uma constante L tal que

$$|f(x) - f(y)| \leq L|x - y|.$$

Por outro lado, $\operatorname{Lip}(f)$ denote a menor constante tal que essa desigualdade é verdadeira. Uma função $f : \mathbb{R}^d \to \mathbb{R}$ é localmente Lipschitz quando f é Lipschitz em todo conjunto aberto e limitado.

Teorema 3.2.1. *Se f é localmente Lipschitz, então o gradiente $\nabla f(x)$ existe para $\mathscr{L}^d$-quase todo ponto $x \in \mathbb{R}^d$.*

Nosso objetivo é mostrar que toda função convexa é localmente Lipschitz no interior de seu domínio, e portanto é diferenciável em quase todo ponto. Isso é uma consequência direta da convexidade.

Teorema 3.2.2. *Seja $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$. Então as seguintes afirmações valem*

1. *f é localmente limitada em seu domínio;*
2. *f é localmente Lipschitz em seu domínio;*
3. *f é diferenciável em $\mathscr{L}^d$-quase todo ponto.*

Demonstração. Seja um cubo $Q \subset \operatorname{dom} f$. Se os vértices do cubo são dados por $(v_i)_{i=1}^{2^d}$, então todo ponto $x \in Q$ pode ser escrito como uma combinação convexa dos vértices, ou seja

$$x = \sum_{i=1}^{2^d} \lambda_i v_i,$$

e da convexidade de f segue que

$$f(x) = f\left(\sum_{i=1}^{2^d} \lambda_i v_i\right) \leq \sum_{i=1}^{2^d} \lambda_i f(v_i).$$

Escolha $M \stackrel{\text{def.}}{=} \max_{i=1,\ldots 2^d} f(v_i)$. Segue que f é localmente limitada.

Agora vamos provar que para todo $x_0 \operatorname{int}\operatorname{dom} f$, f é limitada inferiormente em alguma bola $B_r(x_0)$. Suponha que esse não seja o caso e, para todo $n \in \mathbb{N}$, suponha que exista um ponto $x_n \in B_{1/n}(x_0)$ tal que $f(x_n) \leq -n$. Tome $y_n \overset{\text{def.}}{=} x_0 - (x_n - x_0)$, de modo que

$$x_n, y_n \in B_{1/n}(x_0) \text{ e } x_0 = \frac{1}{2}(x_n + y_n).$$

Para n suficientemente grande $B_{1/n}(x_0)$ está contido no cubo Q da parte anterior da prova. Desse modo, da convexidade de f teríamos que

$$f(x_0) \leq \frac{1}{2} f(x_n) + \frac{1}{2} f(y_n) \leq \frac{M-n}{2} \xrightarrow[n \to +\infty]{} -\infty,$$

o que contradiz o fato de que $f(x_0)$ é um valor finito. Dessa forma, deve existir algum $r > 0$ e $m \geq 0$ tal que $f \geq -m$ em $B_r(x_0)$.

Dado um cubo $Q \subset \operatorname{int}\operatorname{dom} f$ tal que $-m \leq f \leq M$ pelo item precedente, considere uma bola de raio $2R$ tal que $B_{2R} \subset Q$. Dados $x, y \in B_R$, escolha $z \overset{\text{def.}}{=} x + t(y - x)$, e tome t suficientemente grande para que $z \in \partial B_{2R}$. Dessa forma

$$t = \frac{|z - x|}{|y - x|} > 1,$$

pois como x, y, z são colineares, e podemos escrever $y = t^{-1} z + (1 - t^{-1}) x$, se $t \leq 1$, teríamos que $z \in B_R$. Nós obtemos então pela convexidade de f que

$$\begin{aligned} f(y) &\leq f(x) + \frac{1}{t}(f(z) - f(x)) \\ &\leq f(x) + \frac{M+m}{R}|y - x|. \end{aligned}$$

Logo f é localmente Lipschitz e segue do teorema de Rademacher que $\nabla f(x)$ está bem definido em $\mathscr{L}^d$ quase todo ponto. □

3.3 O teorema de Brenier

Vamos agora estudar o teorema de Brenier sobre a existência de aplicações de transporte ótimo para o custo quadrático

$$c(x, y) = \frac{1}{2}|x - y|^2.$$

Note que dadas duas medidas de probabilidade $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$ e um plano de transporte $\gamma \in \Pi(\mu, \nu)$, temos que

$$0 \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2}|x - y|^2 \mathrm{d}\gamma = \frac{1}{2}\int_{\mathbb{R}^d} |x|^2 \mathrm{d}\mu + \frac{1}{2}\int_{\mathbb{R}^d} |y|^2 \mathrm{d}\nu - \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle \, \mathrm{d}\gamma.$$

Portanto para que o valor do problema de Kantorovitch associado tenha valor finito, é necessário e suficiente que os momentos de ordem 2 de μ e ν sejam finitos, $M_2(\mu), M_2(\nu) < +\infty$. Dessa forma, convém definir o subconjunto

$$\mathscr{P}_2(\mathbb{R}^d) \overset{\text{def.}}{=} \left\{\mu \in \mathscr{P}(\mathbb{R}^d) : M_2(\mu) < +\infty\right\}, \text{ onde } M_2(\mu) \overset{\text{def.}}{=} \frac{1}{2}\int_{\mathbb{R}^d} |x|^2 \mathrm{d}\mu. \tag{3.7}$$

Como as quantidades $\frac{1}{2}\int_{\mathbb{R}^d} |x|^2 \mathrm{d}\gamma(x,y)$ são constantes e iguais à $M_2(\mu)$ para todo $\gamma \in \Pi(\mu,\nu)$, e de forma análoga para a segunda marginal, o problema de Kantorovitch para o custo quadrático é equivalente à formulação de *máxima correlação*

$$\mathcal{T}(\mu,\nu) \stackrel{\text{def.}}{=} \max_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle x,y\rangle \,\mathrm{d}\gamma(x,y) = \min_{u \text{ convexa}} \int_{\mathbb{R}^d} u(x)\mathrm{d}\mu + \int_{\mathbb{R}^d} u^*(y)\mathrm{d}\nu. \tag{3.8}$$

Exercício 3.2. Prove que para todo par de medidas $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ a formulação em máxima correlação é equivalente à formulação em termos de distância de Wasserstein

$$W_2^2(\mu,\nu) \stackrel{\text{def.}}{=} \min_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{1}{2}|x-y|^2\mathrm{d}\gamma = \max_{\varphi} \int_{\mathbb{R}^d} \varphi \mathrm{d}\mu + \int_{\mathbb{R}^d} \varphi^c \mathrm{d}\nu, \tag{3.9}$$

com $c = \frac{1}{2}|x-y|^2$.

Apesar de as formulações (3.8),(3.9) serem equivalentes, vamos ver que a formulação em máxima correlação é ligeiramente mais conveniente para provar o teorema de Brenier devido à sua relação com funções convexas. O primeiro resultado para isso é o seguinte, que pode ser formulado mesmo em espaços poloneses, sobre planos de transporte concentrados em gráficos.

Dados dois espaços poloneses $\mathcal{X},\mathcal{Y}$, dizemos que um conjunto $\Gamma \subset \mathcal{X}\times\mathcal{Y}$ é um gráfico se para todo par $x \in \mathcal{X}$, existe no máximo um $y \in \mathcal{Y}$ tal que $(x,y) \in \Gamma$. Note que é perfeitamente possível que para um dado x, não exista nenhum y com essa propriedade.

Proposição 3.3.1. *Dados dois espaços poloneses $\mathcal{X},\mathcal{Y}$, suponha que $\gamma \in \mathscr{P}(\mathcal{X}\times\mathcal{Y})$ seja tal que $(\pi_{\mathcal{X}})_\sharp\gamma = \mu$, e existe um gráfico $\Gamma \subset \mathcal{X}\times\mathcal{Y}$ tal que γ é concentrado em Γ. Então existe uma aplicação $T : \mathcal{X} \to \mathcal{Y}$ tal que $\gamma = (\mathrm{id}, T)_\sharp\mu$.*

Demonstração. Como γ é concentrado em Γ, e é uma medida de Radon, existe um conjunto boreliano $\Gamma_1 \subset \Gamma$ tal que $\gamma(\mathcal{X}\times\mathcal{Y}\setminus\Gamma_1) = 0$. Logo, como toda medida de Radon é regular interiormente, e Γ_1 é boreliano, existe uma sequência de conjuntos compactos $(K_k)_{k\in\mathbb{N}}$ encaixados e tais que $\gamma\left(\mathcal{X}\times\mathcal{Y}\setminus\bigcup_{k\in\mathbb{N}} K_k\right) = 0$.

Seja agora $C_k \stackrel{\text{def.}}{=} (\pi_{\mathcal{X}})(K_k)$ e defina $T_k : C_k \to \mathcal{Y}$, onde para cada $x \in C_k$, $T_k(x)$ é o único ponto de $\mathcal{Y}$ tal que $(x, T_k(x)) \in K_k$. Vamos observar algumas propriedades das aplicações T_k. Primeiro note que, como C_k são uma família crescente de conjuntos compactos, se $i < k$, então $C_i \subseteq C_k$ e segue que $T_k|_{C_i} = T_i$.

Além disso, note que T_k é contínuo em C_k para todo $k \in \mathbb{N}$. De fato, como K_k é um conjunto compacto, seja uma sequência $(x_n)_{n\in\mathbb{N}} \subset C_k$ tal que $x_n \to x$. Desse modo $(x_n, y_n = T_k(x_n))_{n\in\mathbb{N}} \subset K_k$, e por compacidade toda subsequência de (x_n,y_n) admite uma subsequência tal que $(x_{n_i}, y_{n_i}) \to (x,y) \in K_k$. Como K_k é um subconjunto do gráfico Γ, existe um único $y \in \mathcal{Y}$ tal que $(x,y) \in K_k$, e portanto $y = T_k(x)$. Como toda subsequência admite uma nova subsequência convergente, pela propriedade de Urysohn, concluímos que $T_k(x_n) \to T_k(x)$.

Podemos então definir uma aplicação global, observando que

$$\mu\left(\mathcal{X}\setminus\bigcup_{k\in\mathbb{N}} C_k\right) = \gamma\left(\mathcal{X}\times\mathcal{Y}\setminus\bigcup_{k\in\mathbb{N}} K_k\right) = 0.$$

Logo, fixemos um ponto $y_0 \in \mathcal{Y}$ qualquer, e podemos definir a aplicação de Borel $T : \mathcal{X} \to \mathcal{Y}$ como

$$T(x) \stackrel{\text{def.}}{=} \begin{cases} T_k(x), & \text{se } x \in C_k, \\ y_0, & \text{se } x \in \mathcal{X}\setminus\bigcup_{k\in\mathbb{N}} C_k. \end{cases}$$

Para concluir a prova, note que para toda função contínua e limitada $\varphi \in \mathscr{C}_b(\mathcal{X} \times \mathcal{Y})$, temos que $\varphi(x, y) = \varphi(x, T(x))$ para $(x, y) \in K_k$. Dessa forma, pelo teorema da convergência monótona

$$\int_{\mathcal{X}\times\mathcal{Y}} \varphi \mathrm{d}\gamma = \lim_{k\to\infty} \int_{K_k} \varphi \mathrm{d}\gamma = \lim_{k\to\infty} \int_{C_k} \varphi(x, T(x)) \mathrm{d}\mu = \int_{\mathcal{X}} \varphi(x, T(x)) \mathrm{d}\mu.$$

Disso concluímos que $\gamma = (\mathrm{id}, T)_\sharp \mu$ e o resultado segue. □

Esse resultado pode ser usado de forma abstrata, ou seja, não é necessário conhecer uma aplicação que defina o gráfico Γ, basta ser capaz de definir o gráfico Γ de forma abstrata como uma relação de teoria de conjuntos, e o seu ponto positivo é que ele fornece automaticamente uma aplicação de transporte mensurável, por isso ele é particularmente importante nas generalizações do Teorema de Brenier que vamos provar em seguida. No entanto, no caso do custo euclidiano quadrático, ou na formulação de máxima correlação, a teoria de c-monotonicidade cíclica implica que a aplicação de transporte construída é o gradiente de uma função convexa, que é automaticamente mensurável.

Teorema 3.3.1. *Sejam duas medidas de probabilidade $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$, tais que $\mu \ll \mathscr{L}^d$.*

1. *O problema de Kantorovitch na forma de máxima correlação* (3.8) *admite um único plano de transporte ótimo $\gamma \in \Pi(\mu, \nu)$, enquanto que a formulação dual admite maximisadores dado por um par de funções convexas (φ, φ^*), que é único módulo a soma de uma constante, como elemento de $L^1(\mu)$. Além disso, o plano de transporte ótimo assume a forma $\gamma = (\mathrm{id}, \nabla\varphi)_\sharp \mu$.*

2. *Conversamente, se φ é uma função convexa, s.c.i., e diferenciável em μ quase todo ponto tal que $|\nabla\varphi| \in L^2(\mu)$, então $T \overset{\text{def.}}{=} \nabla\varphi$ é sempre ótimo para o transporte de μ à $(\nabla\varphi)_\sharp\mu$.*

3. *Assuma também que $\nu \ll \mathscr{L}^d$, então o mapa de transporte ótimo de μ para ν é dado por $T_{\mu\to\nu} = \nabla\varphi$, enquanto o mapa de ν para μ é dado por $T_{\nu\to\mu} = \nabla\varphi^*$ e nós temos que*

$$T_{\nu\to\mu} \circ T_{\mu\to\nu} = \mathrm{id}\ \mu\text{-q.t.p. em } \mathbb{R}^d,\ T_{\mu\to\nu} \circ T_{\nu\to\mu} = \mathrm{id}\ \nu\text{-q.t.p. em } \mathbb{R}^d. \tag{3.10}$$

Demonstração. Começando pelo item (1), se γ é um plano de transporte ótimo entre μ e $\ni$, note pelo teorema 2.3.1 que existem potenciais ótimos da forma (φ, φ^*) tais que γ é concentrada no conjunto

$$\Gamma \overset{\text{def.}}{=} \{(x, y) : \varphi(x) + \varphi^*(y) = \langle x, y\rangle\} = \{(x, y) : y \in \partial\varphi(x)\}.$$

Segue então que γ é tal que

$$\gamma(\Gamma) = 1, \text{ com } \Gamma \overset{\text{def.}}{=} \left\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : y \in \partial\varphi(x)\right\}.$$

Por outro lado, definindo os conjuntos

$$\Gamma_0 \overset{\text{def.}}{=} \left\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : {}_{\text{não existe } \nabla\varphi(x)}^{y\in\partial\varphi(x)}\right\}, \quad \Gamma_1 \overset{\text{def.}}{=} \Gamma \setminus \Gamma_0.$$

Por outro lado, temos que

$$\gamma(\Gamma_0) \leq \gamma\left(\{x : \text{não existe } \nabla\varphi(x)\} \times \mathbb{R}^d\right) = \mu\left(\{x : \text{não existe } \nabla\varphi(x)\}\right) = 0,$$

pois $\mu \ll \mathscr{L}^d$ e toda função convexa é diferencial em $\mathscr{L}^d$-quase todo ponto.

Temos então que, para todo par de conjuntos borelianos A, B[3]

$$\begin{aligned}\gamma(A\times B) &= \gamma\left(\{(x,y)\in A\times B\}\cap\Gamma_1\right) = \gamma\left(\{(x,\nabla\varphi(x))\in A\times B\}\cap\Gamma_1\right)\\ &= \gamma\left((\mathrm{id},\nabla\varphi)^{-1}(A\times B)\times\mathbb{R}^d\right) = \mu\left((\mathrm{id},\nabla\varphi)^{-1}(A\times B)\right)\\ &= (\mathrm{id},\nabla\varphi)_\sharp\mu(A\times B).\end{aligned}$$

Segue que $\gamma = (\mathrm{id},\nabla\varphi)_\sharp\mu$.

Para obter a unicidade, note que se existirem dois planos de transporte ótimo γ_1 e γ_2, pelos argumentos à cima, temos que $\gamma_i = (\mathrm{id},\nabla\varphi_i)_\sharp\mu$, onde φ_1,φ_2 são funções convexas. Mas então podemos construir um novo plano de transporte ótimo

$$\bar\gamma \overset{\text{def.}}{=} \frac{\gamma_1+\gamma_2}{2},$$

que por sua vez será da forma $\bar\gamma = (\mathrm{id},\nabla\bar\varphi)_\sharp\mu$. Mas pela construção, este novo plano só pode ser induzido por uma aplicação de transporte se $\nabla\varphi_1 = \nabla\varphi_2$ em μ quase todo ponto. Segue que $\gamma_1=\gamma_2$.

Suponha agora que existem dois potenciais de Kantorovitch φ_1,φ_2, tais que $\|\varphi_1-\varphi_2\|_{L^1(\mu)} > 0$. Pelos argumentos anteriores, isso gera dois mapas de transporte ótimo $T_i = \nabla\varphi_i$ para $i=1,2$. Mas pelo argumento anterior temos que $\nabla(\varphi_1-\varphi_2)=0$ para μ-quase todo ponto. Portanto φ_1 e φ_2 divergem no máximo de uma constante no suporte de μ.

Consideremos agora a afirmação conversa no item 2. Assumindo que $\mu \ll \mathscr{L}^d$ e que $\nabla\varphi\in L^2(\mu)$, defina $\nu \overset{\text{def.}}{=} (\nabla\varphi)_\sharp\mu$, de forma que

$$M_2(\nu) = \frac{1}{2}\int_{\mathbb{R}^d}|\nabla\varphi|^2\mathrm{d}\mu < +\infty.$$

Dessa forma o problema de Kantorovitch admite um único plano de transporte ótimo. Além disso, como $\mu\ll\mathscr{L}^d$, definindo $\gamma \overset{\text{def.}}{=} (\mathrm{id},\nabla\varphi)_\sharp\mu$, segue da construção que para γ-quase todo par (x,y) temos que $y=\nabla\varphi(x)$ e segue do case de igualdade da identidade de Fenchel que

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

Pelo teorema 2.3.1, segue que φ é um potencial de Kantorovitch e que γ definido como à cima é ótimo para o transporte de μ à ν.

Para provar o item (3), como $\mu\ll\mathscr{L}^d$ existe um único plano de transporte ótimo $\gamma=\gamma_{\mu\to\nu} = (\mathrm{id},T_{\mu\to\nu})_\sharp\mu\in\Pi(\mu,\nu)$. Por outro lado, inverter as marginais de um plano de transporte preserva o custo de transporte por conta de simetria. Isso pode ser atingido através da aplicação $i:(x,y)\mapsto(y,x)$. Desse modo, se $\gamma\in\Pi(\mu,\nu)$, definindo o plano invertido $\gamma_{\mathrm{inv}} \overset{\text{def.}}{=} i_\sharp\gamma\in\Pi(\nu,\mu)$[4], temos que

$$\int\langle x,y\rangle\,\mathrm{d}\gamma(x,y) = \int\langle x',y'\rangle\,\mathrm{d}\gamma_{\mathrm{inv}}(x',y').$$

[3]Outra forma de ver isso é argumentando que para toda função $f\in\mathscr{C}_b(\mathbb{R}^d\times\mathbb{R}^d)$ temos que

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} f(x,y)\mathrm{d}\gamma(x,y) = \int_{\Gamma_1} f(x,y)\mathrm{d}\gamma(x,y) = \int_{\mathbb{R}^d\times\mathbb{R}^d} f(x,\nabla\varphi(x))\mathrm{d}\gamma(x,y) = \int_{\mathbb{R}^d\times\mathbb{R}^d} f(x,\nabla\varphi(x))\mathrm{d}\mu(x).$$

Não é claro qual argumento é o mais elementar.

[4]Verifique que $\gamma_{\mathrm{inv}}\in\Pi(\nu,\mu)$.

Dessa forma, $\gamma_{\nu\to\mu} \stackrel{\text{def.}}{=} i_\sharp\gamma_{\mu\to\nu}$ é o único plano ótimo para o transporte de ν à μ. Além disso, como $\nu \ll \mathscr{L}^d$, devemos ter que $\gamma_{\nu\to\mu} = (\mathrm{id}, T_{\nu\to\mu})_\sharp\nu$, e portanto, $\gamma_{\mu\to\nu} = i_\sharp\gamma_{\nu\to\mu} = (T_{\nu\to\mu}, \mathrm{id})_\sharp\nu$. Isso significa que o único plano de transporte ótimo $\gamma \in \Pi(\mu,\nu)$ é tal que

$$\gamma = (\mathrm{id}, T_{\mu\to\nu})_\sharp\mu = (T_{\nu\to\mu}, \mathrm{id})_\sharp\nu. \tag{3.11}$$

No que diz respeito aos potenciais de Kantorovitch, note também por simetria que se (φ,φ^*) são ótimos para o transporte de μ à ν, então (φ^*,φ) são ótimos para o transporte de ν à μ. Isso quer dizer que $T_{\mu\to\nu} = \nabla\varphi$ e $T_{\nu\to\mu} = \nabla\varphi^*$.

Finalmente, para concluir com a propriedade de inversão entre os mapas de transporte (3.10) note que para toda função mensurável $f = f(x,y)$ vale que

$$\int f(x,y)\mathrm{d}\gamma = \int f(x, T_{\mu\to\nu}(x))\mathrm{d}\mu(x) = \int f(T_{\nu\to\mu}(y), y)\mathrm{d}\nu(y).$$

Escolhendo $f(x,y) = |y - T_{\mu\to\nu}(x)|$, temos que

$$0 = \int f(x, T_{\mu\to\nu}(x))\mathrm{d}\mu(x) = \int |y - T_{\mu\to\nu}\circ T_{\nu\to\mu}(y)|\mathrm{d}\nu(y),$$

o que significa que $T_{\mu\to\nu}\circ T_{\nu\to\mu} = \mathrm{id}$ em ν-quase todo ponto de $\mathbb{R}^d$. De forma análoga, provamos a identidade recíproca para μ e o resultado segue. $\square$