

Transporte Ótimo – Notas de Aula

Diego Marcon

1 de Junho de 2018

As principais referências são [1, 2, 20, 22, 23].

1 Problema de Monge

O Problema de Monge foi publicado ainda no século XVIII por Gaspard Monge¹ e ainda chama a atenção de muitos matemáticos, devido a sua beleza e ampla gama de aplicações. O problema consiste em transportar *de maneira ótima* uma quantidade de massa de X para Y , sem dividir massa. Em termos matemáticos mais precisos,

Problema de Monge. [16] Dadas duas medidas de probabilidade $\mu \in \mathcal{P}(X)$ e $\nu \in \mathcal{P}(Y)$, queremos minimizar um funcional da forma

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

dentre todas as aplicações mensuráveis $T : X \rightarrow Y$ que satisfazem $T_{\#}\mu = \nu$.



¹Matemático francês que viveu de 9 de Maio de 1746 até 28 de Julho de 1818. Nascido em Beaune, Côte-d'Or, França, começou a ensinar física no Collège de la Trinité, onde estudava, em Lyon quando tinha apenas 17 anos. É conhecido por seu trabalho em Geometria Descritiva e um dos pioneiros de Geometria Diferencial. Seu nome aparece em transporte ótimo, equações diferenciais parciais (na equação de Monge-Ampère), Teorema de Monge em geometria Euclideana plana, etc. Monge faleceu aos 72 anos em Paris, França. Algumas destas informações retiradas de https://en.wikipedia.org/wiki/Gaspard_Monge em 03 de Janeiro de 2018. Segue um retrato de Monge (que também pode ser encontrado no link acima) e uma foto minha com uma rua de Paris dedicada a ele.

Comentários sobre os objetos que mencionamos acima:

- Em muitos dos resultados que veremos, de forma bastante geral, X e Y são **espaços polonenses**, isto é, espaços métricos completos e separáveis. No entanto, nosso maior interesse vai ser o espaço Euclidiano usual \mathbb{R}^d e/ou uma variedade Riemanniana completa M .
- Vamos denotar por $\mathcal{P}(X)$ o espaço das medidas Borelianas² em X cuja massa total é unitária:

$$\mu(X) = \int_X d\mu(x) = 1 \quad \text{e} \quad \nu(Y) = \int_Y d\nu(y) = 1.$$

- A função $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ é chamada de **função custo**. Intuitivamente, $c(x, y)$ nos indica o preço a ser pago para transportar a massa do ponto x para o ponto y .
- A condição $T_{\#}\mu = \nu$ quer dizer que a medida ν seja o **push-forward** (ou medida imagem) de μ pela aplicação T . Aparece naturalmente, pois é justamente esta a condição que indica que a massa em X é transportada para Y ³. Mais precisamente, temos

$$\nu(B) := \mu(T^{-1}(B)) \quad \text{para } B \subseteq Y \text{ mensurável.} \quad (2)$$

ou ainda,

$$\int_X (h \circ T) d\mu = \int_Y h d\nu \quad \text{para toda função } h : Y \rightarrow \mathbb{R} \cup \{+\infty\} \text{ mensurável.} \quad (3)$$

Se $T_{\#}\mu = \nu$, nós dizemos que T é uma **aplicação de transporte** (ou um mapa de transporte).

- Uma aplicação de transporte que minimiza (1) é dita uma **aplicação de transporte ótimo** (ou um mapa de transporte ótimo).

Observamos que, independentemente da escolha da função custo c , o Problema de Monge é não trivial, pelas seguintes razões:

- (i) Pode ser que não existam aplicações de transporte ou, em outras palavras, é possível que

$$\{T; T_{\#}\mu = \nu\} = \emptyset.$$

De fato, se $\mu = \delta_{x_0}$, tem-se que $T_{\#}\mu = \delta_{T(x_0)}$. Logo, não existem aplicações de transporte se $\nu \neq \delta_{T(x_0)}$.

- (ii) A restrição $T_{\#}\mu = \nu$. Inicialmente, vamos pensar no caso em que μ e ν são absolutamente contínuas com respeito a medida de Lebesgue de \mathbb{R}^d . Isto significa que

$$\mu(A) = \int_A f(x) dx \quad \text{e} \quad \nu(B) = \int_B g(y) dy,$$

ou ainda, que $d\mu(x) = f(x) dx$ e $d\nu = g(y) dy$. Deste modo, condição (3) se escreve como

$$\int_{\mathbb{R}^d} h(T(x))f(x) dx = \int_{\mathbb{R}^d} h(y)g(y) dy \quad \text{para } h : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\} \text{ mensurável.}$$

²Lembra que uma medida μ é dita de Borel quando todo conjunto Boreliano é μ -mensurável.

³Por definição, dada uma medida μ em X , a medida $T_{\#}\mu$ em Y é definida por

$$(T_{\#}\mu)(B) := \mu(T^{-1}(B)) \quad \text{para } B \subseteq Y \text{ mensurável.}$$

É um exercício padrão de teoria da medida (faça! – basta aproximar funções mensuráveis por funções simples) que esta condição é equivalente a

$$\int_X (h \circ T) d\mu = \int_Y h d(T_{\#}\mu) \quad \text{para toda função } h : Y \rightarrow \mathbb{R} \cup \{+\infty\} \text{ mensurável.}$$

A operação de push-forward é *covariante* no sentido que, a partir de $T : X \rightarrow Y$, define uma aplicação $T_{\#} : \mathcal{M}(X) \rightarrow \mathcal{M}(Y)$.

No caso (excepcional) que $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ é um difeomorfismo, podemos usar a fórmula de mudança de variáveis para integrais múltiplas (fazendo $y = T(x)$):

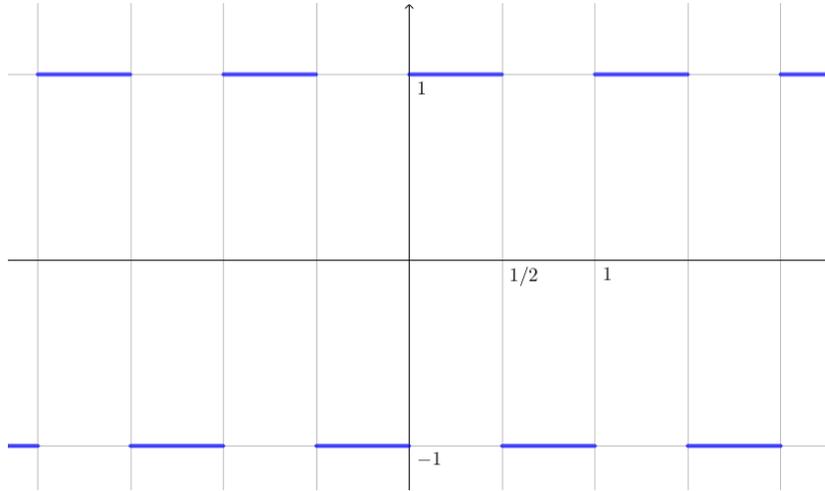
$$\int_{\mathbb{R}^d} h(T(x))f(x) dx = \int_{\mathbb{R}^d} h(T(x))g(T(x))|\det \nabla T(x)| dx \quad \text{para } h : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\} \text{ mensurável.}$$

Observando que $g \circ T$ pode ser qualquer função mensurável de $X = \mathbb{R}^d$, concluímos que

$$f(x) = g(T(x))|\det \nabla T(x)|, \quad \text{para quase todo ponto.} \quad (4)$$

A equação (4) representa a restrição $T_{\#}\mu = \nu$ nos termos mais familiares de medidas absolutamente contínuas (quando T é suficientemente suave). Note que (4) é uma equação *altamente não linear* que T deve satisfazer!

- (iii) Além do já exposto anteriormente, a restrição $T_{\#}\mu = \nu$ não é sequencialmente fechada com respeito à convergência fraca. Ou seja, **não é verdade que** $\mu_n \rightharpoonup \mu, T_{\#}\mu_n = \nu \implies T_{\#}\mu = \nu$. Um contra-exemplo é obtido a partir da “onda quadrada”:



que pode ser escrita como

$$f(x) = \begin{cases} 1, & \text{se } 0 \leq x < 1/2 \\ -1, & \text{se } 1/2 \leq x < 1. \\ 1 - \text{periódica} \end{cases}$$

Definindo $f_n(x) = f(nx)$, mostre (como exercício) que

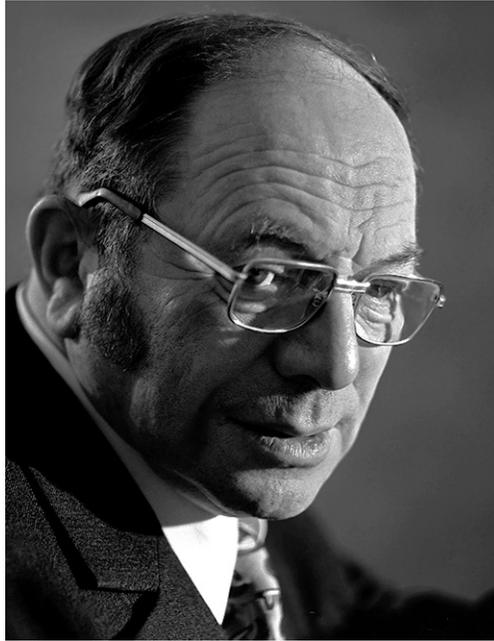
$$(f_n)_{\#}\mathcal{L}|_{[0,1]} = \frac{\delta_{-1} + \delta_1}{2}, \quad \text{mas } f_n \rightharpoonup 0 \quad \text{e} \quad 0_{\#}\mathcal{L}|_{[0,1]} = \delta_0.$$

2 Problema de Kantorovich

Como já comentamos, o Problema de Monge data de 1781 [16]. Surpreendentemente, uma primeira contribuição importante para este problema apareceu apenas em 1942 [13, 12] com trabalhos de Leonid Kantorovich⁴. Como já de costume atualmente, Kantorovich que considerou

⁴Matemático e economista soviético que viveu entre 19 de Janeiro de 1912 e 7 de Abril de 1986. Nascido em São Petersburgo, Império Russo (atual Rússia), começou seus estudos na Universidade Estatal de Leningrado (atual Universidade Estatal de São Petersburgo) aos 14 anos e, aos 22 anos, já ocupava uma vaga de professor titular. É conhecido pelo problema de alocação de recursos que estudamos (e que lhe rendeu o Prêmio Nobel em Ciências Econômicas em 1975) e vários problemas em análise funcional e teoria de operadores. Seu nome aparece no Teorema de Kantorovich (sobre convergência do Método de Newton), Problemas de Monge-Kantorovich, K -espaços. Kantorovich faleceu aos 74 anos em Moscou, URSS (atual Rússia). Informações retiradas de https://en.wikipedia.org/wiki/Leonid_Kantorovich em 04 de Janeiro de 2018. Segue um retrato de Kantorovich (que também pode ser encontrado no link acima)

uma versão mais relaxada (ou fraca) do problema, de maneira bastante parecida com o que se faz na busca por soluções fracas em equações diferenciais parciais ou minimizantes em cálculo de variações.



A ideia aqui é permitir que a massa em cada ponto $x \in X$ seja dividida e transportada para mais de um ponto em Y . Isto equivale a exigir que uma probabilidade $\gamma \in \mathcal{P}(X \times Y)$ descreva o transporte. Intuitivamente, o infinitesimal $d\gamma(x, y)$ indica qual a porção da massa que estava em $x \in X$ que foi transportada para $y \in Y$. Neste problema, estamos pensando que toda a massa que estava em $x \in X$ deve ser transportada para o espaço Y e que toda massa que chega em y é igual à quantidade de massa que o ponto $y \in Y$ suporta. Estas condições podem ser escritas (informalmente) como

$$d\mu(x) = \int_Y d\gamma(x, y) \quad \text{e} \quad d\nu(y) = \int_X d\gamma(x, y).$$

De forma mais precisa, para todo $A \subseteq X$ e $B \subseteq Y$ mensuráveis:

$$\gamma(A \times Y) = \mu(A) \quad \text{e} \quad \gamma(X \times B) = \nu(B). \quad (5)$$

Uma outra forma de escrever (5): sejam $\pi^1 : X \times Y \rightarrow X$ e $\pi^2 : X \times Y \rightarrow Y$ as projeções usuais nas coordenadas; então, (5) é equivalente a

$$\pi^1_{\#}\gamma = \mu \quad \text{e} \quad \pi^2_{\#}\gamma = \nu. \quad (6)$$

Nós dizemos que probabilidades $\gamma \in \mathcal{P}(X \times Y)$ que satisfazem (5) ou (6) têm **marginais** μ e ν . No caso de γ ter marginais μ e ν , dizemos que γ é um **plano de transporte**. O conjunto dos planos de transporte é denotado por $\text{ADM}(\mu, \nu)$ ou $\Pi(\mu, \nu)$ (segundo Ambrosio&Gigli [1] ou Villani [22], respectivamente).

Estamos finalmente prontos para enunciar o

Problema de Kantorovich. Dadas duas medidas de probabilidade $\mu \in \mathcal{P}(X)$ e $\nu \in \mathcal{P}(Y)$, queremos minimizar

$$\mathcal{C}[\gamma] := \int_{X \times Y} c(x, y) d\gamma(x, y) \quad (7)$$

dentre todas as probabilidades $\gamma \in \mathcal{P}(X \times Y)$ com marginais μ e ν .

Um plano de transporte que minimiza (7) é dito um **plano de transporte ótimo**. Seguem algumas vantagens desta formulação mais fraca:

- (i) O conjunto dos planos de transporte $\text{ADM}(\mu, \nu)$ é sempre não vazio: de fato, a medida produto $\mu \times \nu \in \text{ADM}(\mu, \nu)$. No entanto, esta medida representa uma escolha muito ruim: **clarify**.
- (ii) Além disso, como veremos, $\text{ADM}(\mu, \nu)$ é convexo e compacto na topologia da convergência fraca⁵ de $\mathcal{P}(X \times Y)$.
- (iii) A restrição $\gamma \in \text{ADM}(\mu, \nu)$ é linear.
- (iv) A aplicação $\gamma \mapsto \int c d\gamma$ é linear.
- (v) É possível ver uma aplicação de transporte T como um plano de transporte especial, “suportado” no gráfico de T : como exercício, verificar que

$$T_{\#}\mu = \nu \implies \gamma_T := (Id, T)_{\#}\mu \in \text{ADM}(\mu, \nu).$$

A formulação de Kantorovich do Problema de Monge atinge justamente o seu objetivo: fornecer compacidade e semicontinuidade o suficiente para que possamos utilizar técnicas do Cálculo de Variações e construir um minimizante para (7) como limite sequências minimizantes.

3 Alguns exemplos

Iniciamos com uma possível aplicação e depois mostramos alguns exemplos para dar uma ideia das hipóteses necessárias.

Exemplo 1 (Caso discreto). Podemos pensar na seguinte situação: um número n de fábricas deseja transportar os seus produtos para um número m de distribuidores. Digamos:

- (i) a fábrica x_i tem uma quantidade s_i de produtos;
- (ii) o distribuidor y_j tem uma demanda de d_j ;
- (iii) o custo de transportar da fábrica i para o distribuidor j é denotado por $c(x_i, y_j)$;
- (iv) γ_{ij} representa a quantidade de produto que sai do ponto x_i e vai para o ponto y_j .

Nestas condições, gostaríamos de minimizar

$$\sum_{i,j} c(x_i, y_j) \gamma_{ij}.$$

Com a hipótese adicional que a produção e a demanda coincidem (e são iguais a um), chegamos às seguintes condições de compatibilidade:

1. Tudo que está na fábrica x_i é transportado: $\sum_{j=1}^m \gamma_{ij} = s_i$;
2. A demanda do ponto de distribuição y_j é atendida (no valor exato): $\sum_{i=1}^n \gamma_{ij} = d_j$;
3. Quantidade de produtos e demanda: $\sum s_i = \sum d_j = 1$.

Para ver que este é um problema de transporte, como estávamos discutindo, basta definir as medidas discretas:

$$\mu = \sum_i s_i \delta_{x_i}, \quad \nu = \sum_j d_j \delta_{y_j} \quad \text{e} \quad \gamma = \sum_{i,j} \gamma_{ij} \delta_{(x_i, y_j)}.$$

⁵Na verdade, fraca*.

Exemplo 2. Como já vimos, quando $\mu = \delta_{x_0}$ é uma medida de Dirac e ν não é, não existe aplicação de transporte ótima, porque sequer existe *alguma* aplicação de transporte.

Quando, por outro lado, $\nu = \delta_{y_0}$, obtemos um caso trivial. Neste caso, toda a massa de X deve ser transportada para y_0 :

$$\min \int_{X \times Y} c \, d\gamma = \int_X c(x, y_0) \, d\mu(x).$$

Para o próximo exemplo, vamos considerar $X = \mathbb{R}$ e $c : X \times X \rightarrow \mathbb{R}$ o custo $c(x, y) = |x - y|$. Observe que, dada $f : X \rightarrow \mathbb{R}$ função 1-Lipschitz, vale que

$$\int_X f(x) \, d(\nu - \mu)(x) = \int_X [f(T(x)) - f(x)] \, d\mu(x) \leq \int_X |T(x) - x| \, d\mu(x),$$

de modo que

$$\sup \left\{ \int_X f(x) \, d(\nu - \mu)(x); f \in \text{Lip}_1(X) \right\} \leq \inf_{T \# \mu = \nu} \int_X |T(x) - x| \, d\mu(x). \quad (8)$$

Voltaremos a esta propriedade quando falarmos de dualidade.

Exemplo 3 (Book shifting). **Escrever interpretação de mover livros** Sejam $\mu = \chi_{[0,n]} \mathcal{L}^1$ e $\nu = \chi_{[1,n+1]} \mathcal{L}^1$. Observamos que o custo da aplicação de transporte

$$T(x) = x + 1$$

é n . Por outro lado, a função $f(x) = x$ é pertencente a $\text{Lip}_1(X)$ e satisfaz

$$\int_X f(x) \, d(\nu - \mu)(x) = \int_1^{n+1} x \, dx - \int_0^n x \, dx = n.$$

Portanto, levando em conta (8), T é uma aplicação de transporte ótimo.

Por outro lado, S definida por

$$S(x) = \begin{cases} t + n, & \text{se } x \in [0, 1] \\ t, & \text{se } x \in [1, n] \end{cases}$$

também é uma aplicação de transporte ótimo. **Caso custo quadrático melhor vários movimentos pequenos.**

Exemplo 4. Exemplo Villani grosso com segmento no meio

4 Algumas preliminares

Devido ao relaxamento das restrições do problema de Kantorovich, para mostrar existência de planos de transporte ótimo, nós podemos aplicar o método direto do cálculo de variações. Para o leitor não familiarizado, vamos dar a ideia (bastante breve) do método a seguir, até por ser importante em várias situações.

Suponhamos que queremos minimizar uma função ou um funcional $J[x]$ dentre todos os x pertencentes a um espaço topológico E . Em outras palavras, vamos considerar

$$\inf_{x \in E} J[x].$$

O método direto do cálculo de variações é aplicável quando conseguirmos verificar os três (na verdade dois) passos fundamentais:

1. Considerar uma *sequência minimizante*: $x_n \in E$ tal que $J[x_n] \rightarrow \inf_E J$. Pela definição de ínfimo, sempre existem sequências minimizantes. Logo, este passo é sempre trivial.

2. *Compacidade*: por alguma propriedade de espaço E , concluir que existe uma subsequência convergente $x_{n_k} \rightarrow x_0$. No caso de E ser compacto, esta propriedade é válida, mas em muitos dos casos interessantes, E não é compacto.

3. *Semicontinuidade inferior*: se J é semicontínua inferiormente⁶, então

$$J[x_0] \leq \liminf_{k \rightarrow +\infty} J[x_{n_k}] = \lim_{k \rightarrow +\infty} J[x_{n_k}] = \inf_E J.$$

Isto prova que, na realidade, $J[x_0] = \inf_E J$ e, portanto, $x_0 \in E$ é um mínimo de J

Este é o **método direto do cálculo de variações**.

Nós vamos considerar, no caso do Problema de Kantorovich, convergência de elementos de medidas de probabilidade. Para isto, definimos a seguir alguns dos conceitos que serão necessários.

Dizemos que $\rho_n \in \mathcal{P}(X)$ **converge fracamente**⁷ para $\rho \in \mathcal{P}(X)$ quando

$$\int_X f(x) d\rho_n(x) \rightarrow \int_X f(x) d\rho(x), \text{ para toda } f \in C_b^0(X). \quad (9)$$

Lembramos que $C_b^0(X)$ é o espaço das funções contínuas e limitadas de X em \mathbb{R} . É possível verificar (mas não muito relevante para nossos propósitos) que esta convergência é induzida por uma distância ou, em outras palavras, que a topologia gerada por esta convergência é metrizável.

Um subconjunto $\mathcal{K} \subset \mathcal{P}(X)$ é dito **rigido**⁸ quando, para todo $\varepsilon > 0$, existe um conjunto compacto $K_\varepsilon \subset X$ tal que

$$\rho(X \setminus K_\varepsilon) < \varepsilon, \text{ para todo } \rho \in \mathcal{K}.$$

Intuitivamente, um conjunto de probabilidades é rígido quando não há massa “escapando” para infinito em nenhuma das medidas de \mathcal{K} . Esta condição nos dá uma importante caracterização para compacidade relativa em $\mathcal{P}(X)$ no teorema devido a Yuri Prokhorov⁹:

⁶Lembre que, sendo E um espaço metrizável, $J : E \rightarrow \mathbb{R}$ é semicontínua inferiormente se, e somente se, para todo $x \in E$, tem-se $J[x] \leq \liminf_{x_n \rightarrow x} J[x_n]$.

⁷Para aqueles familiarizados com Análise Funcional, esta convergência caracteriza, na realidade, a topologia fraca-* de $(C_b^0(X))^*$. O conjunto $\mathcal{P}(X)$ pode ser identificado como um subconjunto convexo da esfera unitária de $(C_b^0(X))^*$. Em alguns textos em inglês, também se diz que ρ_n “narrowly converges to” ρ .

⁸No inglês, “tight”.

⁹Matemático soviético que viveu de 15 de Dezembro de 1929 até 16 de Julho de 2013. Nascido em Moscou, URSS (atual Rússia), obteve seu doutorado na Universidade Estatal de Moscou em 1949, sob orientação do renomado Andrei Kolmogorov. Sua maior área de interesse é a probabilidade, onde ficou conhecido principalmente pela métrica de Lévy-Prokhorov e pelo Teorema de Prokhorov que tratamos nestas notas. Faleceu há aproximadamente 5 anos (notas de 2018) em Moscou, URSS (atual Rússia), aos 83 anos. Informações retiradas de https://en.wikipedia.org/wiki/Yuri_Prokhorov em 05 de Janeiro de 2018. Segue uma foto retirada de fineartamerica.com/featured/yuri-prokhorov.



Teorema de Prokhorov. *Se X é um espaço polonês, então $\mathcal{K} \subset \mathcal{P}(X)$ é rígido se, e somente se, é relativamente compacto com relação à convergência fraca.*

Vamos apresentar uma prova deste teorema que aparece em [3] e indicar os passos da prova (mais funcional analítica) presente em <https://terrytao.wordpress.com> (o mais conhecido “blog do Tao”).

Demonstração. \Leftarrow Esta é a parte do teorema que é menos interessante aos nossos propósitos. Considere uma família de conjuntos abertos “crescendo” para X :

$$A_n \subseteq A_{n+1} \subseteq X \quad \text{e} \quad \bigcup_n A_n = X. \quad (10)$$

Em particular, para qualquer $\mu \in \mathcal{P}(X)$,

$$\lim_n \mu(A_n) = \mu(X). \quad (11)$$

Afirmamos que, para cada $\varepsilon > 0$, existe $n \in \mathbb{N}$ tal que $\rho(A_n) > 1 - \varepsilon$ para todo $\rho \in \mathcal{K}$. De fato, caso contrário teríamos que, para algum $\varepsilon > 0$ e para todo $n \in \mathbb{N}$, é possível encontrar $\rho_n \in \mathcal{K}$ de maneira que $\rho_n(A_n) \leq 1 - \varepsilon$. Por hipótese, \mathcal{K} é relativamente compacto. Então, existe uma subsequência convergente: $\rho_n \rightarrow \mu \in \mathcal{K}$. Logo¹⁰, lembrando (10),

$$\mu(A_n) \leq \liminf_k \rho_{n_k}(A_n) \leq \liminf_k \rho_{n_k}(A_{n_k}) \leq 1 - \varepsilon.$$

Mas isto seria uma contradição com (11) e nossa afirmação está demonstrada.

Agora, X é um espaço métrico separável, de modo que, para cada $k \in \mathbb{N}$, existe uma sequência de bolas abertas de raio $1/k$ que cobrem X . Digamos,

$$X = \bigcup_{i=1}^{+\infty} B_{1/k}(x_i).$$

¹⁰A primeira desigualdade é um resultado padrão de Teoria da Medida: Se μ_n é uma sequência de medidas de probabilidade (ou, mais geralmente, localmente finitas) de Borel, então $\mu_n \rightarrow \mu$ se, e somente se, para todo $A \subseteq X$ aberto,

$$\mu(A) \leq \liminf_n \mu_n(A).$$

Ver e.g. [15, Seções 2.4 e 4.3] ou [10, Capítulo 1].

Pela nossa afirmação, existe $n_k \in \mathbb{N}$ tal que

$$\rho \left(\bigcup_{i=1}^{n_k} B_{1/k}(x_i) \right) > 1 - \varepsilon \quad \text{para todo } \rho \in \mathcal{K}.$$

Definimos

$$K_\varepsilon := \overline{\bigcap_{k=1}^{+\infty} \bigcup_{i=1}^{n_k} B_{1/k}(x_i)}.$$

Assim,

(i) K_ε é totalmente limitado, pois é o fecho de um conjunto totalmente limitado: para cada $\delta > 0$, consideramos $1/k < 2\delta$ e $K_\varepsilon \subseteq \bigcup_{i=1}^{n_k} B_{1/k}(x_i)$, uma reunião de bolas centradas em pontos de K_ε e de diâmetro δ .

(ii) K_ε é completo, pois é um subconjunto fechado de um espaço completo.

Portanto, K_ε é compacto¹¹ e obtemos o que queríamos:

$$\rho(K_\varepsilon) = \lim_k \rho \left(\bigcup_{i=1}^{n_k} B_{1/k}(x_i) \right) \geq 1 - \varepsilon.$$

$\boxed{\implies}$ Esta é a implicação mais importante. No entanto, observe da prova abaixo que sequer é preciso X ser polonês. Consideramos $\rho_n \in \mathcal{K}$ e queremos mostrar que existe $\rho \in \mathcal{K}$ tal que $\rho_{n_k} \rightarrow \rho$, para alguma subsequência ρ_{n_k} . Para isto, devemos construir uma medida de probabilidade ρ em X que satisfaça

$$\rho(A) \leq \liminf_n \rho_n(A) \quad \text{para todo aberto } A. \quad (12)$$

Pela hipótese de rigidez, é possível, para cada $\ell \in \mathbb{N}$, encontrar um compacto K_ℓ tal que

$$\rho_n(K_\ell) > 1 - \frac{1}{\ell}.$$

Podemos supor, sem perda de generalidade, que $K_1 \subseteq K_2 \subseteq K_3 \subseteq \dots$ (poderíamos substituir K_2 por $K_1 \cup K_2$ sem problemas).

Como todo conjunto métrico compacto é separável, temos que $Y := \bigcup_\ell K_\ell$ é também separável; logo, possui base enumerável¹². Lembre que uma base enumerável é uma coleção enumerável $\mathcal{A} := \{A_i\}_{i=1}^{+\infty}$ de abertos em Y que tem a propriedade de “gerar” qualquer aberto de Y , isto é, todo aberto $O \subseteq Y$ pode ser escrito como $O = \bigcup A_{i_k}$.

Agora, consideramos a família \mathcal{H} de todas as uniões finitas de conjuntos da forma $\overline{A_i} \cap K_j$, que também é uma família enumerável.

A vantagem das construções acima é que, sendo \mathcal{H} enumerável, conseguimos extrair uma subsequência ρ_{n_k} de modo que $\rho_{n_k}(H)$ é convergente para todo $H \in \mathcal{H}$. Isto é feito com um “argumento diagonal”, como segue. Note que $\rho_n(H_1)$ possui uma subsequência convergente $\rho_{n_k}^1(H_1)$ pois é uma sequência limitada de números reais. Em seguida, extraímos uma subsequência convergente de $\rho_{n_k}^1(H_2)$, que pode ser denotada por $\rho_{n_k}^2(H_2)$. Pela construção, temos também que $\rho_{n_k}^2(H_1)$ é convergente. E assim por diante.

Vamos denotar o limite obtido no parágrafo anterior por

$$\alpha(H) := \lim_k \rho_{n_k}(H) \quad \text{para } H \in \mathcal{H}.$$

Observe que α satisfaz

- $\alpha(\emptyset) = 0$ e $H_1 \subset H_2 \implies \alpha(H_1) \leq \alpha(H_2)$;

¹¹Ver, por exemplo, [14, Seção 7.4] para referência em topologia de espaços métricos compactos.

¹²Uma possível referência para estes conceitos e resultados de topologia geral é Lima [14, Capítulo 8].

- $\alpha(H_1 \cup H_2) \leq \alpha(H_1) + \alpha(H_2)$ e, se $H_1 \cap H_2 = \emptyset$, temos $\alpha(H_1 \cup H_2) = \alpha(H_1) + \alpha(H_2)$.

Gostaríamos de, a partir de α , construir uma medida de probabilidade satisfazendo (12). Vamos definir uma probabilidade que, para A aberto, satisfaz

$$\rho(A) = \sup_{H \in \mathcal{H}, H \subseteq A} \alpha(H).$$

Vamos denotar o supremo acima por ρ_1 e deixar os vários passos abaixo como exercício (que estão resolvidos em [3, Seção 1.5]):

- (i) Suponhamos que $A \subset F$ com A aberto e F fechado, e existe $H \in \mathcal{H}$ tal que $F \subset H$. Então existe H_0 tal que $A \subset H_0 \subset F$.
- (ii) Usar passo anterior para verificar que ρ_1 é subaditiva, para uniões finitas de conjuntos abertos.
- (iii) Verificar que ρ_1 é σ -subaditiva (ou seja, subaditiva para uniões enumeráveis) em conjuntos abertos.
- (iv) Verificar que $\gamma(M) = \inf_{M \subset G} \rho(G)$, definido para qualquer subconjunto $M \subset X$, é uma medida exterior (basta mostrar σ -subaditividade).
- (v) Todo Boreliano é γ -mensurável (basta mostrar que todo fechado é γ -mensurável).

Segue dos itens acima que a restrição de γ aos Borelianos é uma medida de Borel. Esta é, por definição, a medida ρ que queremos. Temos que, para A aberto, $\rho(A) = \gamma(A) = \sup_{H \in \mathcal{H}, H \subseteq A} \alpha(H)$.

Além disso, $\rho \in \mathcal{P}(X)$, pois $K_\ell \in \mathcal{H}$ (compactos possuem subcoberturas finitas) e logo

$$1 \geq \rho(X) \geq \sup_\ell \alpha(K_\ell) \geq \sup_\ell \left(1 - \frac{1}{\ell}\right) = 1 \implies \rho(X) = 1.$$

Resta ainda verificar que $\rho_{n_k} \rightarrow \rho$, ou seja, (12). Para qualquer $H \in \mathcal{H}$ com $H \subseteq A$, temos

$$\alpha(H) = \lim_k \rho_{n_k}(H) \leq \liminf_k \rho_{n_k}(A).$$

Fazendo o supremo em H , concluímos (12). □

Observação 5. No caso de ser X compacto, todo subconjunto de $\mathcal{P}(X)$ é trivialmente rígido e, portanto, toda sequência de probabilidades admite subsequência convergente. Isto pode ser provado, mais diretamente, com técnicas de Análise Funcional. O Teorema da Representação de Riesz [21, Teorema 7.3] diz que

$$(C_b(X))^* = M(X) = \text{conjunto das medidas com sinal em } X. \quad (13)$$

Segue do Teorema de Banach-Alaoglu [8, Seção 15.2] que a bola fechada de $M(X)$ é compacta na topologia da convergência fraca-* (que é a nossa convergência fraca).

Sendo X localmente compacto e σ -compacto, temos que existe uma subsequência convergente (em um processo diagonal parecido com o que fizemos acima). Para garantir que o limite ρ tem massa total unitária (ou seja, para provar que a medida limite é de probabilidade), é necessário supor rigidez. \triangleleft

5 Existência de plano de transporte ótimo

Vamos aplicar o método direto do cálculo de variações, como já havíamos comentado, de maneira bastante *straightforward*.

Inicialmente, mostramos que o funcional custo é semicontínuo inferiormente.

Lema 6. *Sejam X, Y espaços poloneses e $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função custo que satisfaz*

(1) c semicontinua inferiormente;

(2) c limitada por baixo.

Então, o funcional custo é semicontínuo inferiormente em $\mathcal{P}(X \times Y)$, com a topologia da convergência fraca. Em outras palavras,

$$\gamma_k \rightarrow \gamma \text{ em } \mathcal{P}(X \times Y) \implies \int_{X \times Y} c \, d\gamma \leq \liminf_k \int_{X \times Y} c \, d\gamma_k.$$

Demonstração. As condições (1) e (2) garantem que existe uma sequência crescente de funções contínuas c_n tal que, pontualmente, $c(x, y) = \lim_n c_n(x, y)$. Pelo Teorema da Convergência Monótona:

$$\int_{X \times Y} c \, d\gamma = \lim_n \int_{X \times Y} c_n \, d\gamma = \lim_n \lim_k \int_{X \times Y} c_n \, d\gamma_k \leq \liminf_k \int_{X \times Y} c \, d\gamma_k.$$

A segunda igualdade é a definição da convergência $\gamma_k \rightarrow \gamma$ e a desigualdade final segue de $c_n \leq c$, para todo n . \square

Em seguida, observamos resultado que, junto com Prokhorov, vai fornecer compacidade.

Lema 7. *Sejam X, Y espaços poloneses. Se $\mathcal{K}_1 \subset \mathcal{P}(X)$ e $\mathcal{K}_2 \subset \mathcal{P}(Y)$ são ambos rígidos, então também é rígido o conjunto*

$$\left\{ \gamma \in \mathcal{P}(X \times Y); \pi_{\#}^1 \gamma \in \mathcal{K}_1 \text{ e } \pi_{\#}^2 \gamma \in \mathcal{K}_2 \right\}.$$

Demonstração. É direta da desigualdade:

$$\gamma((X \times Y) \setminus (K_1 \times K_2)) \leq \mu(X \setminus K_1) + \nu(Y \setminus K_2),$$

que é válida para qualquer $\gamma \in \text{ADM}(\mu, \nu)$. \square

Teorema 8. *Suponhamos que a função custo c é semicontínua inferiormente e limitada por baixo. Então, existe um plano de transporte ótimo, ou seja, um minimizantes para o Problema de Kantorovich.*

Demonstração. Dada uma medida de Borel μ , o conjunto $\mathcal{K}_1 = \{\mu\} \subset \mathcal{P}(X)$ é rígido¹³. Também é rígido o conjunto o conjunto $\mathcal{K}_2 = \{\nu\} \subset \mathcal{P}(Y)$. O conjunto dos planos de transporte $\text{ADM}(\mu, \nu)$ é rígido em $\mathcal{P}(X \times Y)$. Logo, pelo Teorema de Prokhorov, é relativamente compacto. Assim, dada uma sequência minimizante $\gamma_n \in \text{ADM}(\mu, \nu)$, existe uma subsequência convergente: $\gamma_{n_k} \rightarrow \gamma \in \overline{\text{ADM}(\mu, \nu)}$. Vamos mostrar que, de fato, $\gamma \in \text{ADM}(\mu, \nu)$: dada qualquer $\varphi \in C_b(X)$, temos

$$\int_X \varphi(x) \, d(\pi_{\#}^1 \gamma)(x) = \int_{X \times Y} \varphi(\pi^1(x, y)) \, d\gamma(x, y) = \lim_n \int_{X \times Y} \varphi \, d\gamma_n = \lim_n \int_X \varphi \, d(\pi_{\#}^1 \gamma_n) = \int_X \varphi(x) \, d\mu(x),$$

o que mostra que $\pi_{\#}^1 \gamma = \mu$. De forma parecida, $\pi_{\#}^2 \gamma = \nu$.

Utilizando o lema anterior, nosso funcional é semicontínuo inferiormente e o teorema está demonstrado. \square

Observação 9. Se fosse c contínua, não haveria necessidade do Lema 6, seguiria direto da convergência fraca. Se fossem X e Y compactos, não haveria necessidade do Lema 7 (nem do Teorema de Prokhorov), pois poderíamos extrair uma subsequência convergente pelo Teorema de Banach-Alaoglu.

¹³Mais geralmente, se m é uma medida de Borel e $E \subseteq X$ é um Boreliano com $m(E) < +\infty$, então, para cada $\varepsilon > 0$, existe um compacto $K \subset E$ tal que $m(E \setminus K) < \varepsilon$. Ver, por exemplo, [15, Seção 2.3].

6 Caracterização de otimalidade de planos de transporte

Iniciamos considerando um exemplo, como em [1].

Exemplo 10. Sejam $X = Y = \mathbb{R}^d$ o custo quadrático

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

e medidas $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ suportadas em conjuntos finitos (com N elementos). Então um plano $\gamma \in \text{ADM}(\mu, \nu)$ é ótimo se, e somente se,

$$\sum_{i=1}^N \frac{|x_i - y_i|^2}{2} \leq \sum_{i=1}^N \frac{|x_i - y_{\sigma(i)}|^2}{2},$$

onde $(x_i, y_i) \in \text{supp } \gamma$ e σ é uma permutação do conjunto $\{1, 2, \dots, N\}$. Cancelando os termos ao quadrado, temos que $\gamma \in \text{ADM}(\mu, \nu)$ é ótimo se, e somente se,

$$\sum_{i=1}^N \langle x_i, y_i \rangle \geq \sum_{i=1}^N \langle x_i, y_{\sigma(i)} \rangle. \quad (14)$$

Acontece que esta é uma definição já conhecida de Análise Convexa. Um conjunto $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$ é dito **ciclicamente monótono** quando satisfaz (14) para quaisquer escolhas de pontos $(x, y) \in \Gamma$ e qualquer permutação $\sigma : \{1, 2, \dots, N\} \rightarrow \{1, 2, \dots, N\}$.

Um teorema devido a Rockafellar [18, Theorem 1] (que também pode ser encontrado no livro [19, Theorem 24.8]), fornece uma caracterização de subdiferenciais de funções convexas como conjuntos com esta propriedade de monotonicidade cíclica. Eis o enunciado:

Teorema 11 (Rockafellar). *Um conjunto $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$ é ciclicamente monótono se, e somente se, existe uma função convexa própria $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d \times \{+\infty\}$, semicontínua inferiormente, tal que*

$$\Gamma \subseteq \partial\varphi := \bigcup_{x \in \mathbb{R}^d} (x, \partial\varphi(x)).$$

Acima, estamos denotando por $\partial\varphi(x)$ o subdiferencial de φ em x , como definido na Subseção 13.3.2.

Não provamos este teorema, porque nós vamos provar um resultado mais geral, que inclui este, e que pode ser encontrado em [1, Theorem 1.13] e/ou [20, Theorems 1.37 & 1.38]. Antes de enunciar a generalização, precisamos estender todas as definições envolvidas, que é o que passamos a fazer.

Seja $c : X \times Y \rightarrow \mathbb{R}$ uma função contínua de valores reais. Um conjunto $\Gamma \subseteq X \times Y$ é dito **c -ciclicamente monótono** quando satisfaz a seguinte propriedade: para todo subconjunto finito $\{(x_i, y_i)\}_{i=1,2,\dots,N} \subseteq \Gamma$, vale

$$\sum_{i=1}^N c(x_i, y_i) \geq \sum_{i=1}^N c(x_i, y_{\sigma(i)}),$$

para qualquer permutação σ do conjunto $\{1, 2, \dots, N\}$.

Dada uma função $f : X \rightarrow \mathbb{R} \cup \{-\infty\}$, não identicamente $-\infty$, definimos a sua **c -transformada**, também conhecida como **função c -conjugada** como a função $f^c : Y \rightarrow \mathbb{R} \cup \{-\infty\}$ dada por

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

Analogamente, definimos a c -transformada de $g : Y \rightarrow \mathbb{R} \cup \{-\infty\}$ como a função $g^c : X \rightarrow \mathbb{R} \cup \{-\infty\}$ dada por

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

Uma função $f : X \rightarrow \mathbb{R} \cup \{-\infty\}$ é dita **c -côncava** quando existe $g : Y \rightarrow \mathbb{R} \cup \{-\infty\}$ tal que $f = g^c$. As seguintes propriedades são de verificação deixadas como exercício.

Exercício 1. Suponhamos que c é uma função contínua de valores reais e considere uma função $f : X \rightarrow \mathbb{R} \cup \{-\infty\}$. Então

(i) $f^{cc} \geq f$.

(ii) $f^{cc} = f$ se, e somente se, f é c -côncava.

(iii) Se f é c -côncava, então f é semicontínua inferiormente.

Finalmente, dada $f : X \rightarrow \mathbb{R} \cup \{-\infty\}$ uma função c -côncava, definimos o c -**superdiferencial** de f como o conjunto

$$\partial^c f = \{(x, y) \in X \times Y; f(x) + f^c(y) = c(x, y)\}. \quad (15)$$

Exercício 2. Considere o caso $X = Y = \mathbb{R}^d$ e seja $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ a função $c(x, y) = -\langle x, y \rangle$. Mostre que:

(i) um conjunto $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$ é c -ciclicamente monótono se, e somente se, é ciclicamente monótono.

(ii) uma função $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{-\infty\}$ é c -côncava se, e somente se, f é côncava e semicontínua superiormente.

(iii) o c -superdiferencial de uma função c -côncava é o superdiferencial usual, de modo que

$$\partial f(x) := \{y \in \mathbb{R}^d; f(z) \leq f(x) + \langle y, z - x \rangle \quad \forall z\} = \{y \in \mathbb{R}^d; f(x) + f^c(y) = \langle x, y \rangle\}$$

(iv) a c -transformada é a Transformada de Legendre-Fenchel usual de uma função côncava.

Agora, estamos prontos para enunciar (e provar) uma generalização do Teorema de Rockafellar.

Teorema 12. Suponhamos que $c : X \times Y \rightarrow \mathbb{R}$ é contínua e limitada inferiormente. Sejam $\mu \in \mathcal{P}(X)$ $\nu \in \mathcal{P}(Y)$ tais que, para alguma função $a \in L^1(\mu)$ e alguma $b \in L^1(\nu)$, vale que

$$c(x, y) \leq a(x) + b(y). \quad (16)$$

Além disso, seja $\gamma \in \text{AdM}(\mu, \nu)$. São equivalentes:

(i) o plano de transporte γ é ótimo;

(ii) o suporte de γ é c -ciclicamente monótono;

(iii) existe uma função c -côncava $\varphi : X \rightarrow \mathbb{R} \cup \{-\infty\}$ tal que $\text{supp } \gamma \subseteq \partial^c \varphi$.

Demonstração. A hipótese (16) e a limitação por baixo indicam que a função custo c está em $L^1(\tilde{\gamma})$, para qualquer que seja $\tilde{\gamma} \in \text{AdM}(\mu, \nu)$, pois

$$\int_{X \times Y} c(x, y) d\tilde{\gamma}(x, y) \leq \int_{X \times Y} (a(x) + b(y)) d\tilde{\gamma}(x, y) = \int_X a(x) d\mu(x) + \int_Y b(y) d\nu(y) < +\infty.$$

Passamos a mostrar as equivalências do teorema.

$(i) \implies (ii)$ Supomos, por contradição, que o suporte de γ não é c -ciclicamente monótono: existe $n \geq 1$ e pontos $(x_i, y_i) \in \text{supp } \gamma$, $i = 1, 2, \dots, n$, tais que

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

A ideia é conseguir construir um competidor $\tilde{\gamma}$ cujo custo total é menor do que o custo total de γ , contradizendo (i). Para isto, utilizamos a continuidade de c para obter vizinhanças $V_i := B_r(x_i) \times B_r(y_i)$ dos pontos (x_i, y_i) tais que

$$|c(x, y) - c(x_i, y_i)| < \varepsilon. \quad (17)$$

O número $\varepsilon > 0$ será escolhido convenientemente mais tarde.

Agora, para outra constante a ser escolhida $\varepsilon_0 > 0$, definimos

$$\tilde{\gamma} := \gamma - \varepsilon_0 \sum_{i=1}^n \gamma_i + \varepsilon_0 \sum_{i=1}^n \tilde{\gamma}_i, \quad \text{onde } \gamma_i := \frac{1}{\gamma(V_i)} \gamma \llcorner V_i \quad \text{e} \quad \tilde{\gamma}_i := \mu_i \times \nu_{\sigma(i)} \in \text{ADM}(\mu_i, \nu_{\sigma(i)}),$$

onde $\mu_i := \pi_{\#}^1 \gamma_i$ e $\nu_i := \pi_{\#}^2 \gamma_i$. As medidas γ_i foram escolhidas suportadas na vizinhança V_i e estão bem definidas porque

$$(x_i, y_i) \in \text{supp } \gamma \implies \gamma(V_i) > 0.$$

As medidas $\tilde{\gamma}_i$ foram escolhidas para, como veremos, “corrigir” as marginais (poderíamos ter escolhido qualquer medida positiva com marginais μ_i e ν_i) e se obter $\tilde{\gamma} \in \text{ADM}(\mu, \nu)$. Verifiquemos então que $\tilde{\gamma}$ satisfaz o que afirmamos.

1. Positividade: $\tilde{\gamma} > 0$ pois, escolhendo ε_0 apropriadamente, conseguimos $\gamma - \varepsilon_0 \sum \gamma_i > 0$:

$$\varepsilon_0 \gamma_i(A) = \frac{\varepsilon_0}{\gamma(V_i)} \gamma(A \cap V_i) < \frac{\gamma(A)}{n} \quad \text{se escolhermos } \varepsilon_0 < \frac{\min_i \gamma(V_i)}{n}.$$

Logo, para esta escolha de ε_0 ,

$$\varepsilon_0 \sum_{i=1}^n \gamma_i(A) < \gamma(A) \quad \forall A \subseteq X \times Y.$$

2. As marginais estão corretas, isto é, $\tilde{\gamma} \in \text{ADM}(\mu, \nu)$:

$$\pi_{\#}^1 \tilde{\gamma} = \mu - \varepsilon_0 \sum_{i=1}^n \pi_{\#}^1 \gamma_i + \varepsilon_0 \sum_{i=1}^n \pi_{\#}^1 \tilde{\gamma}_i = \mu \quad \text{e} \quad \pi_{\#}^2 \tilde{\gamma} = \nu - \varepsilon_0 \sum_{i=1}^n \pi_{\#}^2 \gamma_i + \varepsilon_0 \sum_{i=1}^n \pi_{\#}^2 \tilde{\gamma}_{\sigma(i)} = \nu$$

pois as marginais foram escolhidas convenientemente.

3. O competidor $\tilde{\gamma}$ de fato seria melhor do γ : utilizando a continuidade (17),

$$\begin{aligned} \int_{X \times Y} c(x, y) d\gamma(x, y) - \int_{X \times Y} c(x, y) d\tilde{\gamma}(x, y) &= \varepsilon_0 \sum_{i=1}^n \int_{X \times Y} c(x, y) d\gamma_i(x, y) - \varepsilon_0 \sum_{i=1}^n \int_{X \times Y} c(x, y) d\tilde{\gamma}_i(x, y) \\ &\geq \varepsilon_0 \sum_{i=1}^n (c(x_i, y_i) - \varepsilon) - \varepsilon_0 \sum_{i=1}^n (c(x_i, y_{\sigma(i)}) + \varepsilon) \\ &= \varepsilon_0 \left(\sum_{i=1}^n c(x_i, y_i) - \sum_{i=1}^n c(x_i, y_{\sigma(i)}) - 2n\varepsilon \right) > 0 \end{aligned}$$

com a escolha conveniente, que apenas é permitida no caso de (ii) ser falso,

$$0 < \varepsilon < \frac{1}{2n} \sum_{i=1}^n [c(x_i, y_i) - c(x_i, y_{\sigma(i)})].$$

(ii) \implies (iii) Fixamos $(x_0, y_0) \in \text{supp } \gamma$ e, explicitamente, definimos $\varphi : X \rightarrow \mathbb{R} \cup \{-\infty\}$ por

$$\varphi(x) := \inf \left\{ [c(x, y_n) - c(x_n, y_n)] + [c(x_n, y_{n-1}) - c(x_{n-1}, y_{n-1})] + \cdots + [c(x_1, y_0) - c(x_0, y_0)] \right\},$$

onde o ínfimo é em $n \geq 1$ e quaisquer escolhas de pontos $\{(x_i, y_i)\}_{i=1,2,\dots,n} \subseteq \text{supp } \gamma$. Observamos que as hipóteses de c assumir valores reais (finitos) e $\text{supp } \gamma \neq \emptyset$ garantem que $\varphi < +\infty$.

A definição de monotonicidade c -cíclica implica $\varphi(x_0) \geq 0$. Em particular, $\varphi \not\equiv -\infty$. Este é o único lugar onde monotonicidade é utilizada. Apesar de não dispensável para os nossos

propósitos, note que na verdade $\varphi(x_0) = 0$, pois escolhendo $n = 1$ e $(x_1, y_1) = (x_0, y_0)$, temos $\varphi(x_0) \leq c(x_0, y_1) - c(x_1, y_1) + c(x_1, y_0) - c(x_0, y_0) = 0$.

Para provar que φ é c -côncava, notamos que podemos escrever

$$\varphi(x) = \inf_y \{c(x, y) - \psi(y)\},$$

onde (a ideia é tomar $y = y_n$ na definição de φ)

$$\psi(y) := - \inf \left\{ -c(x_n, y) + [c(x_n, y_{n-1}) - c(x_{n-1}, y_{n-1})] + \cdots + [c(x_1, y_0) - c(x_0, y_0)] \right\}, \quad (18)$$

com ínfimo relativo a $n \geq 1$, $\{(x_i, y_i)\}_{i=1,2,\dots,N} \subseteq \text{supp } \gamma$ e $y_n = y$. Isto mostra que φ é c -côncava.

Finalmente, devemos mostrar que $\varphi(x) + \varphi^c(y) = c(x, y)$ sempre que $(x, y) \in \text{supp } \gamma$. A desigualdade

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

é imediata da definição de φ^c . Logo, devemos verificar a desigualdade oposta. Já que $\varphi^c = \psi^{cc} \geq \psi$, basta verificarmos que vale

$$\varphi(x) + \psi(y) \geq c(x, y).$$

Por (18) e pela definição de ínfimo, dado $\varepsilon > 0$, existe \tilde{y} que satisfaz

$$c(x, \tilde{y}) - \psi(\tilde{y}) < \varphi(x) + \varepsilon.$$

Além disso, pela definição de ψ :¹⁴

$$-\psi(y) \leq -c(x, y) + c(x, \tilde{y}) - \psi(\tilde{y}) \leq -c(x, y) + \varphi(x) + \varepsilon.$$

Sendo $\varepsilon > 0$ arbitrário, esta parte do teorema está demonstrada.

(iii) \implies (i) Por hipótese,

$$\begin{aligned} \varphi(x) + \varphi^c(y) &= c(x, y) \quad \forall (x, y) \in \text{supp } \gamma \text{ e} \\ \varphi(x) + \varphi^c(y) &\leq c(x, y) \quad \forall (x, y) \in X \times Y. \end{aligned}$$

Logo, γ é um plano de transporte ótimo, pois

$$\int c \, d\gamma = \int (\varphi(x) + \varphi^c(y)) \, d\gamma = \int \varphi(x) \, d\mu(x) + \int_Y \varphi^c(y) \, d\nu(y) = \int (\varphi(x) + \varphi^c(y)) \, d\tilde{\gamma} \leq \int c \, d\tilde{\gamma}. \quad \square$$

Observe que, portanto, a otimalidade de γ no Problema de Kantorovich é uma propriedade que depende apenas do suporte da medida γ . Uma consequência interessante (da prova) do teorema acima é o seguinte:

Exercício 3. Se um plano de transporte ótimo tem o suporte contido no c -superdiferencial da função c -côncava φ , então todo plano de transporte ótimo tem suporte contido no c -superdiferencial da mesma função φ .

¹⁴Considerando $x_n = x, y_{n-1} = \tilde{y}$, tem-se

$$\begin{aligned} -\psi(y) &\leq -c(x, y) + [c(x, \tilde{y}) - c(x_{n-1}, \tilde{y})] + [c(x_{n-1}, y_{n-2}) - c(x_{n-2}, y_{n-2})] + \cdots + [c(x_1, y_0) - c(x_0, y_0)] \\ &\leq [-c(x, y) + c(x, \tilde{y})] - c(x_{n-1}, \tilde{y}) + [c(x_{n-1}, y_{n-2}) - c(x_{n-2}, y_{n-2})] + \cdots + [c(x_1, y_0) - c(x_0, y_0)] \end{aligned}$$

Agora basta considerar o ínfimo em $n - 1 \geq 1$ e $\{(x_i, y_i)\}_{i=1}^{n-1} \subseteq \text{supp } \gamma$.

7 Existência de aplicações ótimas – custo quadrático em \mathbb{R}^d

Nesta seção analisamos o Problema de Monge com $X = Y = \mathbb{R}^d$ e $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ a função custo quadrática dada por

$$c(x, y) = \frac{|x - y|^2}{2}. \quad (19)$$

De acordo com as seções anteriores, dadas $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, sempre existe um plano de transporte ótimo $\gamma \in \text{ADM}(\mu, \nu)$. Além disso, existe uma função c -côncava φ tal que

$$\text{supp } \gamma \subseteq \partial^c \varphi. \quad (20)$$

Nosso objetivo passa a ser verificar que, para μ -quase todo x , o c -superdiferencial $\partial^c \varphi(x)$ é um conjunto unitário, isto é, com um único ponto. Observamos uma importante conexão de funções c -côncavas com funções convexas no presente contexto.

Lema 13. *Consideramos $X = Y = \mathbb{R}^d$ e $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ o custo quadrático dado por (19). Então,*

$$\varphi \text{ é } c\text{-côncava} \iff \phi(x) := \frac{|x|^2}{2} - \varphi(x) \text{ é convexa e semicontínua inferiormente.}$$

Neste caso, $\partial^c \varphi(x) = \partial \phi(x)$.

Demonstração. Pela definição de c -côncavidade:

$$\varphi(x) = \inf_{y \in \mathbb{R}^d} \left\{ \frac{|x - y|^2}{2} - \psi(y) \right\} = \frac{|x|^2}{2} - \sup_{y \in \mathbb{R}^d} \left\{ \langle x, y \rangle - \left(\frac{|y|^2}{2} - \psi(y) \right) \right\}$$

Logo,

$$\phi(x) = \frac{|x|^2}{2} - \varphi(x) = \sup_{y \in \mathbb{R}^d} \left\{ \langle x, y \rangle - \left(\frac{|y|^2}{2} - \psi(y) \right) \right\}$$

que é convexa e semicontínua inferiormente, como supremo de funções afins. As demais afirmações ficam como exercício. \square

Uma consequência do lema acima é que a função φ em (20) é localmente Lipschitz e, portanto, diferenciável quase sempre (com respeito à medida de Lebesgue). Assim, $\partial^c \varphi(x)$ é um conjunto unitário para Lebesgue quase todo ponto. Mais explicitamente, se φ é diferenciável em um ponto $\bar{x} \in \mathbb{R}^d$, escrevemos $\bar{y} = \nabla \varphi(\bar{x})$ e então, pelo Lema 13,

$$\bar{y} = \bar{x} - \nabla \phi(\bar{x}), \text{ isto é, } \partial^c \varphi(\bar{x}) = \{ \nabla \varphi(\bar{x}) \} = \{ \bar{x} - \nabla \phi(\bar{x}) \} \mathcal{L}^d - \text{qtp.}$$

Para fechar a nossa linha de raciocínio, precisamos garantir o c -superdiferencial de φ é unitário para μ -quase todo ponto. A condição que mais obviamente implica nesta propriedade é

$$\mu \ll \mathcal{L}^d, \text{ isto é, } \mu \text{ é absolutamente contínua com a medida de Lebesgue.} \quad (21)$$

Mais sutil é a observação de que o conjunto dos pontos de não diferenciabilidade de uma função convexa tem dimensão (de Hausdorff) no máximo $d - 1$ ¹⁵, de modo que basta assumir que

$$\mu(A) = 0 \text{ para qualquer conjunto } A \text{ tal que } \mathcal{H}^{d-1}(A) < +\infty. \quad (22)$$

O Teorema de Brenier¹⁶ é o que resume toda nossa discussão para encontrar um minimizante para o Problema de Monge. Foi publicado em 1987 [4] (ver também [5]).

¹⁵Mais precisamente, é um conjunto \mathcal{H}^{d-1} -retificável, isto é, a menos de um conjunto \mathcal{H}^{d-1} -negligível, está contido em uma união enumerável de gráficos de funções Lipschitz da forma $g : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$

¹⁶Yann Brenier [foi um matem.....](#)



Teorema 14 (Brenier, 1987). *Sejam $X = Y = \mathbb{R}^d$ e $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ a função custo quadrática dada por (19). Se $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfaz (21) ou (22) e $\nu \in \mathcal{P}(\mathbb{R}^d)$, então existe uma única aplicação de transporte ótimo $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ para o Problema de Monge. Além disso, $T = \nabla \phi$ é o gradiente de uma função convexa semicontínua inferiormente $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$.*

Demonstração. Considere um plano de transporte ótimo $\gamma \in \text{ADM}(\mu, \nu)$ para o Problema de Kantorovich associado. Temos que $\text{supp } \gamma$ é c -ciclicamente monótono, de modo que existe uma função c -côncava $\varphi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{-\infty\}$ tal que

$$\text{supp } \gamma \subseteq \partial^c \varphi.$$

Pela discussão acima, e pelas hipóteses (21) e (22), podemos escrever

$$\partial^c \varphi(x) = \{\nabla \phi(x)\} \text{ para } \mu - \text{qtp},$$

onde $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ é convexa e semicontínua inferiormente. Logo,

$$\text{supp } \gamma \subseteq \text{graph } \nabla \phi, \text{ para } \phi(x) := \frac{|x|^2}{2} - \varphi(x) \text{ é convexa.}$$

Provamos abaixo que $T = \nabla \phi$ é uma aplicação de transporte ótimo:

$$\begin{aligned} \int_X \frac{|x - T(x)|^2}{2} d\mu(x) &= \int_{X \times Y} \frac{|x - y|^2}{2} d\gamma(x, y) \\ &= \min_{\tilde{\gamma} \in \text{ADM}(\mu, \nu)} \int_{X \times Y} \frac{|x - y|^2}{2} d\tilde{\gamma}(x, y) \\ &\leq \inf_{S \# \mu = \nu} \int_X \frac{|x - S(x)|^2}{2} d\mu(x) \end{aligned}$$

A unicidade segue de: se T_1 e T_2 são aplicações ótimas, então

$$\gamma_1 := (Id, T_1) \# \mu \text{ e } \gamma_2 := (Id, T_2) \# \mu$$

são planos ótimos para o Problema de Kantorovich¹⁷. Logo, o plano

$$\gamma := \frac{\gamma_1 + \gamma_2}{2}$$

¹⁷Isto segue de um resultado que ainda não provamos, que vale com bastante generalidade: se X e Y são espaços poloneses, $\mu \in \mathcal{P}(X)$ é uma medida não-atômica e $\nu \in \mathcal{P}(Y)$, então $\min(\text{Pr.Kantorovich}) = \inf(\text{Pr.Monge})$. Apesar de resultados parciais anteriores existirem, estamos enunciando um resultado devido a Aldo Pratelli [17]. A estratégia da demonstração é provar que o conjunto dos planos de transporte que são induzidos por aplicações de transporte é denso em $\text{ADM}(\mu, \nu)$, com respeito a convergência fraca.

também é um plano de transporte ótimo, de modo que $\text{supp } \gamma \subseteq \nabla \bar{\phi}$ para alguma função $\bar{\phi}$. Isto só pode acontecer se $T_1 = T_2$ para μ -quase todo ponto. Mas daí também $\text{supp } \gamma_1 \subseteq \nabla \bar{\phi}$ e $\text{supp } \gamma_2 \subseteq \nabla \bar{\phi}$
fazer desenho □

7.1 A equação de Monge-Ampère

Como já havíamos discutido, a condição $\nabla \phi \# \mu = \nu$ pode ser encarada como uma mudança de variáveis. No caso em que ambas as densidades são absolutamente contínuas com respeito a medida de Lebesgue:

$$d\mu(x) = f(x) dx \quad \text{e} \quad d\nu = g(y) dy,$$

temos

$$\int_{\mathbb{R}^d} h(\nabla \phi(x)) f(x) dx = \int_{\mathbb{R}^d} h(y) g(y) dy \quad \text{para } h : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\} \text{ mensurável.}$$

No caso de ser $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ uniformente convexa, tem-se $T = \nabla \phi$ injetiva. Se for também diferenciável, podemos usar a fórmula de mudança de variáveis para integrais múltiplas para obter

$$\int_{\mathbb{R}^d} h(\nabla \phi(x)) f(x) dx = \int_{\mathbb{R}^d} h(\nabla \phi(x)) g(T(x)) (\det \nabla^2 \phi(x)) dx \quad \text{para } h : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\} \text{ mensurável.}$$

Observando que $g \circ T$ pode ser qualquer função mensurável de $X = \mathbb{R}^d$, concluímos que

$$f(x) = g(\nabla \phi(x)) (\det \nabla^2 \phi(x)), \quad \text{para } \mathcal{L}^d\text{-quase todo ponto.} \quad (23)$$

Se $g > 0$, temos uma **equação de Monge-Ampère**:

$$\det \nabla^2 \phi(x) = \frac{f(x)}{g(\nabla \phi(x))}, \quad (24)$$

que é uma EDP (não-uniformente) elíptica no caso em que ϕ é convexa.

Teorema 15 (Teorema de Regularidade (Caffarelli; Caffarelli & Urbas)). *Suponhamos que $X, Y \subseteq \mathbb{R}^d$ são abertos conexos e limitados, com Y convexo. Sejam*

$$\mu = f \mathcal{L}^d|_X \quad \text{e} \quad \nu = g \mathcal{L}^d|_Y$$

tais que $0 < c_1 \leq f, g \leq c_2$. Seja $T = \nabla \phi$ a única solução de Brenier do problema de Monge. Então

- (i) $T \in C^\alpha$;
- (ii) $f \in C^\alpha(X), g \in C^\alpha(Y) \implies T \in C^{1,\alpha}$, isto é, $\phi \in C^{2,\alpha}$.

Adicionalmente, um problema em aberto de Ambrosio, Gigli [1] é o agora

Teorema 16 (De Philippis, Figalli (Inventiones 2013)). *Suponhamos que $X, Y \subseteq \mathbb{R}^d$ são abertos conexos e limitados, com Y convexo. Sejam*

$$\mu = f \mathcal{L}^d|_X \quad \text{e} \quad \nu = g \mathcal{L}^d|_Y$$

tais que $0 < c_1 \leq f, g \leq c_2$. Seja $T = \nabla \phi$ a única solução de Brenier do problema de Monge. Então $T \in W_{\text{loc}}^{1,1}(X)$.

7.2 Uma aplicação para o problema isoperimétrico

Dado $E \subset \mathbb{R}^n$ um conjunto mensurável (na realidade, de perímetro finito), a **desigualdade isoperimétrica** diz que

$$P(E) \geq P(B) \quad \text{sempre que} \quad |E| = |B|,$$

onde B é uma bola qualquer com mesmo volume de E . Uma outra forma de enunciar este resultado é o seguinte:

$$|E| = |B| = \omega_n r^n \implies r = |E|^{1/n} \omega_n^{-1/n} \implies P(B) = n \omega_n r^{n-1} = n \omega_n^{1/n} |E|^{1-1/n}$$

Teorema 17. *Seja $E \subset \mathbb{R}^n$ de perímetro finito (ou suave, limitado). Então*

$$P(E) \geq n |B_1|^{\frac{1}{n}} |E|^{\frac{n-1}{n}}.$$

Demonstração (adaptação de prova devida à Gromov). Suponhamos $|E| = |B_1|$ e consideremos

$$d\mu(x) = \chi_E(x) dx \quad \text{e} \quad d\nu(y) = \chi_{B_1}(y) dy.$$

Seja $T(x) = \nabla\phi(x)$ a transformação de transporte ótimo dada pelo Teorema de Brenier. Notamos que $T(E) = B_1$. Logo, temos

(i) $|T(x)| \leq 1$, para todo $x \in E$;

(ii) $T_{\#}\mu = \nu \implies \det \nabla^2\phi(x) = \frac{f(x)}{g(T(x))} = \frac{|E|}{|B_1|} = 1.$

(iii) Desigualdade aritmética geométrica: já que $\lambda_i \geq 0$,

$$\det \nabla^2\phi(x) = \prod_{i=1}^n \lambda_i \leq \left(\frac{1}{n} \sum_{i=1}^n \lambda_i \right)^n = \frac{\Delta\phi(x)^n}{n^n}.$$

Assim, pelo Teorema de Gauss-Green, temos

$$\begin{aligned} P(E) &= \int_{\partial E} d\sigma \geq \int_{\partial E} |T(x)| d\sigma \geq \int_{\partial E} T(x) \cdot \eta d\sigma = \int_E \operatorname{div} T(x) dx \\ &= \int_E \Delta\phi(x) dx = n \int_E (\det \nabla^2\phi(x))^{1/n} dx = n \int_E dx = n |E| = n |B_1|^{\frac{1}{n}} |E|^{\frac{n-1}{n}}. \quad \square \end{aligned}$$

8 Dualidade a partir de monotonicidade cíclica

O Problema de Monge-Kantorovich admite uma versão dual, assim como vários problemas deste tipo. A ideia por trás é um princípio minimax. Podemos reescrever o Problema de Monge-Kantorovich como

$$\inf_{\gamma \in \operatorname{ADM}(\mu, \nu)} \int c d\gamma = \inf_{\gamma \in \mathcal{M}_+} \left(\int c d\gamma + \begin{cases} 0, & \text{se } \gamma \in \operatorname{ADM}(\mu, \nu) \\ +\infty, & \text{se } \gamma \notin \operatorname{ADM}(\mu, \nu) \end{cases} \right),$$

onde \mathcal{M}_+ é o conjunto das medidas de Borel positivas. A ideia é transformar o Problema de Kantorovich em um problema sem restrições. Note, no entanto, que não estaremos “minimizand” a função da direita caso $\gamma \notin \operatorname{ADM}(\mu, \nu)$.

Agora, notamos que

$$\sup_{(\varphi, \psi) \in C_b(X) \times C_b(Y)} \left\{ \left[\int \varphi d\mu + \int \psi d\nu \right] - \int [\varphi(x) + \psi(y)] d\gamma(x, y) \right\} = \begin{cases} 0, & \text{se } \gamma \in \operatorname{ADM}(\mu, \nu) \\ +\infty, & \text{se } \gamma \notin \operatorname{ADM}(\mu, \nu) \end{cases}.$$

Logo, podemos escrever

$$\inf_{\gamma \in \text{ADM}(\mu, \nu)} \int c \, d\gamma = \inf_{(\varphi, \psi) \in \mathcal{M}_+} \sup_{\gamma \in \mathcal{M}_+} \left\{ \int [c(x, y) - \varphi(x) - \psi(y)] \, d\gamma(x, y) + \int \varphi \, d\mu + \int \psi \, d\nu \right\}.$$

O princípio minimax consiste em trocar a ordem do inf-sup acima. Em alguns casos, isto pode ser justificado rigorosamente com técnicas de Análise Funcional; é assim que está feito em [22, Seção 1.1]. Aqui, nós vamos apenas esboçar o argumento por dualidade e provar a dualidade por monotonicidade cíclica, como em [1, Theorem 1.17].

Então, em um passo *formal*, trocamos de ordem o ínfimo e o supremo:

$$\begin{aligned} \inf_{\gamma \in \text{ADM}(\mu, \nu)} \int c \, d\gamma &= \sup_{(\varphi, \psi)} \inf_{\gamma \in \mathcal{M}_+} \left\{ \int [c(x, y) - \varphi(x) - \psi(y)] \, d\gamma(x, y) + \int \varphi \, d\mu + \int \psi \, d\nu \right\} \\ &= \sup_{(\varphi, \psi)} \left\{ \int \varphi \, d\mu + \int \psi \, d\nu - \sup_{\gamma \in \mathcal{M}_+} \int [c(x, y) - \varphi(x) - \psi(y)] \, d\gamma(x, y) \right\} \end{aligned}$$

Olhamos para o supremo dentro das chaves acima.

1. Se $c(x, y) - \varphi(x) - \psi(y) \leq 0$ para todo (x, y) , então a integral é negativa (não-positiva) e o supremo é atingido quando $\gamma \equiv 0$;
2. Se, por outro lado, existe (x_0, y_0) tal que $c(x_0, y_0) - \varphi(x_0) - \psi(y_0) > 0$, então a escolha $\delta_n := n \cdot \delta_{(x_0, y_0)}$ mostra que o supremo vai para $+\infty$.

Desta discussão concluímos a **Dualidade de Kantorovich**:

$$\inf_{\gamma \in \text{ADM}(\mu, \nu)} \int c \, d\gamma = \sup_{\varphi(x) + \psi(y) \leq c(x, y)} \int \varphi \, d\mu + \int \psi \, d\nu. \quad (25)$$

Uma interpretação interessante da dualidade acima, que é devida a Caffarelli e pode ser encontrada em [22, Subseção 1.1.3], é o *problema do remetente* (ou transportador – no inglês, *shipper's problem*). Eis a interpretação. Você quer fazer o transporte de modo a minimizar o custo – Problema de Kantorovich – e eu sou um transportador que faz a seguinte proposta: vou toda a massa que está em $x \in X$ cobrando uma taxa $\varphi(x)$ e vou transportar todo o necessário para $y \in Y$ cobrando uma taxa $\psi(y)$. Você deverá me pagar, na totalidade, o valor

$$\int \varphi(x) \, d\mu(x) + \int \psi(y) \, d\nu(y).$$

Minha proposta é muito interessante porque, além disso, garanto que $\varphi(x) + \psi(y) \leq c(x, y)$; em outras palavras, você pagará, no máximo, o que já deveria pagar de qualquer maneira. A dualidade de Kantorovich afirma que, se eu fizer o meu trabalho de forma ótima, vou receber o valor mínimo que você deveria pagar. Todo mundo sai feliz! Você paga o mínimo que conseguiria e eu recebo este dinheiro para fazer o transporte.

Finalmente, passamos ao teorema. Vamos escrever que $(\varphi, \psi) \in \text{ADM}$ quando $\varphi \in L^1(\mu)$, $\psi \in L^1(\nu)$ e $\varphi + \psi \leq c$ em todo ponto.

Teorema 18 (Dualidade de Kantorovich). *Sejam $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ e $c : X \times Y \rightarrow \mathbb{R}$ contínua, limitada por baixo e tal que*

$$c(x, y) \leq a(x) + b(y), \quad \text{com } a \in L^1(\mu) \text{ e } b \in L^1(Y).$$

Então

$$\inf_{\gamma \in \text{ADM}(\mu, \nu)} \int c \, d\gamma = \sup_{(\varphi, \psi) \in \text{ADM}} \int \varphi \, d\mu + \int \psi \, d\nu$$

e o máximo é atingido por um par (φ, ψ^c) , para alguma função c -côncava φ .

Demonstração. Para qualquer $\gamma \in \text{ADM}(\mu, \nu)$ e para $(\varphi, \psi) \in \text{ADM}$, temos

$$\int c \, d\gamma \geq \int (\varphi + \psi) \, d\gamma = \int \varphi \, d\mu + \int \psi \, d\nu,$$

de modo que $\inf \geq \sup$.

Por outro lado, considerando $\gamma \in \text{ADM}(\mu, \nu)$ um plano de transporte ótimo, sabemos que existe (Teorema 12) uma função φ c -côncava com $\text{supp } \gamma \subseteq \partial^c \varphi$. Logo, (ver 15)

$$\int c \, d\gamma \geq \int (\varphi + \varphi^c) \, d\gamma = \int \varphi \, d\mu + \int \varphi^c \, d\nu. \quad \square$$

8.1 Fórmula de Kantorovich–Rubinstein

Consideramos (X, d) um espaço polonês, $Y = X$ e $c(x, y) = d(x, y)$. Por dualidade,

$$\inf_{\gamma \in \text{ADM}(\mu, \nu)} \int_{X \times X} d(x, y) \, d\gamma(x, y) = \sup_{(\phi, \phi^c)} \int_X \phi(x) \, d\mu(x) + \int_X \phi^c(y) \, d\nu(y).$$

com ϕ função c -côncava.

Observamos que, neste caso $c = d$, temos

$$\phi \text{ é } d\text{-côncava} \iff \phi \text{ é 1-Lipschitz. Além disso, } \phi^d = -\phi.$$

De fato, ϕ é c -côncava se, e somente se, $\phi = \phi^{cc}$; logo, escolhendo $y = z$ abaixo,

$$\phi(x) = \phi^{cc}(x) = \inf_{y \in X} \sup_{z \in X} \left\{ d(x, y) - d(y, z) + \phi(z) \right\} \leq d(x, z) + \phi(z).$$

Trocando papéis de x e z , obtemos

$$|\phi(x) - \phi(z)| \leq d(x, z).$$

Reciprocamente, se ϕ é 1-Lipschitz, tem-se

$$\phi(x) \leq d(x, z) + \phi(z), \quad \forall z \implies \phi(x) \leq \inf_{z \in X} \{d(x, z) + \phi(z)\} = (-\phi)^c(x).$$

Igualdade vale para $z = x$, de modo que $\phi = (-\phi)^c$. Em particular, ϕ é c -côncava.

A discussão acima implica em

Teorema 19 (Fórmula de Kantorovich–Rubinstein). Para X um espaço polonês e $c = d : X \times X \rightarrow \mathbb{R}$, temos

$$\inf_{\gamma \in \text{ADM}(\mu, \nu)} \int_{X \times X} d(x, y) \, d\gamma(x, y) = \sup_{\|\phi\|_{\text{Lip}} \leq 1} \int_X \phi(x) \, d(\mu - \nu)(x).$$

9 Custos estritamente convexos

Em $X = Y = \mathbb{R}^d$, vamos considerar a função custo $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ dada por

$$c(x, y) := h(x - y),$$

onde $h : \mathbb{R}^d \rightarrow \mathbb{R}$ é uma função (que em seguida vamos supor estritamente convexa).

Seja $\gamma \in \text{ADM}(\mu, \nu)$ um plano de transporte ótimo. Logo, pelo Teorema 12, existe uma função c -côncava φ tal que $\text{supp } \gamma \subseteq \partial^c \varphi$. Logo,

$$\begin{aligned} \varphi(x) + \varphi^c(y) &\leq c(x, y) \quad \forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d, \\ \varphi(x) + \varphi^c(y) &= c(x, y) \quad \forall (x, y) \in \text{supp } \gamma. \end{aligned}$$

Logo, para $(x, y) \in \text{supp } \gamma$, a função

$$x \mapsto \varphi(x) - c(x, y_0)$$

assume um valor mínimo em x_0 . Se for φ diferenciável em x_0 e, por exemplo, $c \in C^1$, podemos concluir que

$$\nabla\varphi(x_0) = \nabla_x c(x_0, y_0).$$

Uma condição adicional natural para escrevermos $y_0 = T(x_0)$ é conseguir inverter a identidade acima.

Dizemos que $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ satisfaz a **condição twist** quando, para todo $x_0 \in \mathbb{R}^d$,

$$y \mapsto \nabla_x c(x_0, y)$$

é injetiva (logo, invertível na sua imagem).

Teorema 20. *Seja $\Omega \subset \mathbb{R}^d$ e $c(x, y) = h(x - y)$ com $h \in C^1(\Omega)$ uma função estritamente convexa. Suponhamos $\mu \ll dx$ e $\partial\Omega$ de medida nula. Então, existe (única μ -qtp) aplicação de transporte ótimo. Além disso, existe φ função c -côncava tal que*

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

Demonstração. As hipóteses sobre h garantem que toda função c -côncava (tem mesmo módulo de continuidade que c e logo) é localmente de Lipschitz em Ω . Desta forma, φ é diferenciável em (Lebesgue) quase todo ponto do interior de Ω . Além disso, $|\partial\Omega| = 0$ por hipótese. Segue que para μ -quase todo ponto, φ é diferenciável e

$$\nabla\varphi(x_0) = \nabla h(x_0 - y_0)$$

Sendo $h \in C^1$ e estritamente convexa em um compacto, h é uniformemente convexa e portanto injetiva (c satisfaz a condição twist). Segue que podemos escrever, como queríamos,

$$x_0 - y_0 = (\nabla h)^{-1}(\nabla\varphi(x_0)) \quad \text{para } \mu\text{-qtp.} \quad \square$$

10 Variedades Riemannianas

Nesta seção, vamos revisar conceitos fundamentais de variedades diferenciáveis e de geometria Riemanniana. Várias referências foram consultadas: Carmo, Conlon, Lee, Gallot Hulin, Villani, Rodney (apostila)...

10.1 Variedades diferenciáveis e diferenciabilidade de funções

Uma variedade diferenciável é um espaço topológico que localmente se parece muito com o espaço euclidiano. De forma mais precisa, uma **variedade diferenciável** (ou de classe C^∞) de dimensão n é um espaço topológico¹⁸ M junto com um conjunto de **cartas** $\phi_\alpha : U_\alpha \rightarrow \mathbb{R}^n$, onde os conjuntos U_α abertos são abertos em M e cobrem M , e as funções de transição $\phi_\alpha \circ \phi_\beta^{-1}$ são de classe C^∞ onde estão definidas. Ao conjunto das cartas $\{\phi_\alpha\}$ chamamos **atlas**.

Uma função f é dita **diferenciável** em M quando $f \circ \phi_\alpha^{-1} : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ é diferenciável, para todas as parametrizações ϕ_α . Isto equivale a dizer que, quando vista “em coordenadas”, f é uma função diferenciável. A condição de as funções de transição serem suaves é fundamental para que este conceito esteja bem definido. No caso em que duas parametrizações em um sistema

De forma análoga, f é de classe C^k ou de classe C^∞ quando $f \circ \phi_\alpha^{-1}$ é de classe C^k ou de classe C^∞ , respectivamente.

Exemplo 21. Superfícies diferenciáveis, espaço projetivo, etc, podem ser equipados com estruturas de variedades diferenciáveis.

¹⁸Em livros de análise matemática e/ou geometria diferencial, se pede que M seja de Hausdorff e com base enumerável. Isto é o que garante existência de partições da unidade e é o que fazemos aqui.

10.2 Espaço tangente

Quando a variedade está “dentro” de um espaço euclidiano, poderíamos definir vetores tangentes assim como fazemos para superfícies diferenciáveis, a saber, como elementos da imagem da diferencial de uma parametrização qualquer. No caso mais abstrato que nos encontramos, é necessário obter alguma caracterização diferente.

Intuitivamente, nos espaços do tipo \mathbb{R}^n , campos de vetores aparecem como um conjunto de setas, uma sobre cada ponto do espaço. Usualmente, campos vetoriais podem representar campos de força ou a velocidade de um fluido. Cada uma dessas setas representa a direção e o sentido que uma partícula andaria sob o efeito do tal campo vetorial. Ou então, cada uma das setas poderia indicar a derivada direcional a ser calculada de funções escalares. Estas duas maneiras de pensar indicam os caminhos mais comuns de se definir o espaço tangente a uma variedade diferenciável.

Vamos inicialmente pensar em $f : \mathbb{R}^n \rightarrow \mathbb{R}$. A derivada direcional de f no ponto x e na direção de um vetor v pode ser calculada ao considerar uma curva α com $\alpha(0) = x$ e $\alpha'(0) = v$; tem-se

$$\frac{\partial f}{\partial v}(x) = \frac{d}{dt}f(\alpha(t))\Big|_{t=0} = \sum_{i=1}^n \alpha'_i(0) \frac{\partial f}{\partial x_i}(x) = \left(\sum_{i=1}^n \alpha'_i(0) \frac{\partial}{\partial x_i} \Big|_{t=0} \right) f, \quad (26)$$

onde $\alpha(t) = (\alpha_1(t), \alpha_2(t), \dots, \alpha_n(t))$. Em particular, $\alpha'_i(0)$ são as coordenadas de $v \in \mathbb{R}^n$. É comum escrever pensar em v como um funcional $v : C^\infty(\mathbb{R}^n) \rightarrow \mathbb{R}$ e pensar em $v(f)$ como a derivada de f em x e na direção v .

Um dos pontos de vista é pensar nos vetores tangentes à variedades como velocidades de curvas que passam por x com o “mesmo” vetor tangente. Mais precisamente, dadas duas curvas diferenciáveis α e β com $\alpha(0) = x = \beta(0)$, definimos uma relação de equivalência por¹⁹

$$\alpha \sim \beta \iff \frac{d}{dt}\phi(\alpha(t))\Big|_{t=0} = \frac{d}{dt}\phi(\beta(t))\Big|_{t=0} \quad \text{para alguma carta } \phi.$$

Um **vetor tangente** é uma classe de equivalência $v = [\gamma]$ para alguma curva diferenciável γ com $\gamma(0) = x$ e o conjunto dos vetores tangentes a M no ponto $x \in M$ é chamado de **espaço tangente** e denotado por $T_x M$.

Proposição 22. *Se M é uma variedade diferenciável de dimensão n , então $T_x M$ é um espaço vetorial de dimensão n .*

Demonstração. Dado um ponto $x \in M$ e uma carta ϕ que parametriza uma vizinhança de x em M , definimos $d\phi_x : T_x M \rightarrow \mathbb{R}^n$ por

$$d\phi_x(v) := \frac{d}{dt}\phi(\gamma(t))\Big|_{t=0},$$

onde γ é qualquer representante da classe de equivalência v . Esta aplicação está bem definida, como pode ser visto diretamente da definição da relação de equivalência. Também pela definição da relação de equivalência, vemos que $d\phi_x$ é injetiva. Para mostrar que é também sobrejetiva, dado $w \in \mathbb{R}^n$, definimos

$$\alpha(t) = \phi^{-1}(\phi(x) + tw) \quad \text{de modo que} \quad \frac{d}{dt}\phi(\alpha(t))\Big|_{t=0} = w.$$

Sendo $d\phi_x$ uma bijeção, podemos equipar $T_x M$ com a estrutura de espaço vetorial induzida pela aplicação inversa $d\phi_x^{-1}$, tornando $T_x M$ um espaço vetorial de dimensão n . De maneira mais explícita, as operações em $T_x M$ são definidas como

$$v + w := d\phi_x^{-1}(d\phi_x(v) + d\phi_x(w)) \quad \text{e} \quad cv := d\phi_x^{-1}(c d\phi_x(v)).$$

¹⁹Pela diferenciabilidade das funções de transição, a igualdade vale, na realidade, para toda carta que contém as curvas α e β . Logo, o espaço tangente é independente da parametrização escolhida.

Sendo as funções de transição difeomorfismos, temos que

$$(\psi \circ \phi^{-1})'(\phi(x)) : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

é um isomorfismo. Portanto, as duas parametrizações quaisquer de vizinhanças de x geram em $T_x M$ a mesma estrutura de espaço vetorial. \square

É também comum utilizar a notação $\gamma'(0) = [\gamma]$, onde γ é qualquer representante da classe de equivalência $[\gamma]$. Assim, continuamos com a fórmula usual

$$d\phi_{\alpha(0)}(\gamma'(0)) = \left. \frac{d}{dt} \phi(\gamma(t)) \right|_{t=0},$$

embora esta seja agora apenas a definição da aplicação $d\phi_x$.

Um outro ponto de vista, menos geométrico mas bastante conveniente, é olhar para vetores tangentes como objetos que agem sobre funções diferenciáveis da mesma maneira que “tomar derivadas direcionais”. Em outras palavras, um vetor tangente é visto como o que nos permite fazer a derivada direcional de funções suaves. Nós dizemos que $v_x : C^\infty(M) \rightarrow \mathbb{R}$ é um **vetor tangente** a M no ponto $x \in M$ quando é uma **derivação** em x , isto é, um funcional linear que satisfaz a regra de Leibniz no ponto x :

$$v_x(fg) = v_x(f) \cdot g(x) + f(x) \cdot v_x(g).$$

Esta propriedade é claramente motivada pela regra da diferenciação de um produto de funções. Ao espaço das derivações como acima chamamos de **espaço tangente** a M em x e denotamos por $\tilde{T}_x M$ (o til será deixado de lado quando mostrarmos que as duas definições são equivalentes). Naturalmente, $\tilde{T}_x M$ tem uma estrutura de espaço vetorial (dada pela soma de funcionais lineares e a multiplicação de um funcional linear por um escalar).

Observamos que, dado $[\gamma] \in T_x M$, podemos associar uma derivação em $\tilde{T}_x M$ pela identidade

$$v_\gamma(f) := \left. \frac{d}{dt} f(\gamma(t)) \right|_{t=0} \quad \text{para toda } f \in C^\infty(M). \quad (27)$$

Para verificar que de fato (27) define um elemento de $\tilde{T}_x M$, consideramos uma carta $\phi : U \rightarrow \mathbb{R}^n$ em uma vizinhança U de x . Em coordenadas, temos

$$v_\gamma(f) = \left. \frac{d}{dt} (f \circ \phi^{-1} \circ \phi \circ \gamma)(t) \right|_{t=0} = d(f \circ \phi^{-1})(\phi(x)) \cdot \left. \frac{d}{dt} (\phi \circ \gamma)(t) \right|_{t=0}.$$

Escrevendo as variáveis de \mathbb{R}^n da inversa da parametrização como $\phi^{-1}(x^1, x^2, \dots, x^n)$ e a curva γ na carta ϕ por

$$\phi(\gamma(t)) = (x^1(t), x^2(t), \dots, x^n(t)),$$

temos

$$v_\gamma(f) = \sum_{i=1}^n \frac{dx^i}{dt}(0) \frac{\partial (f \circ \phi^{-1})}{\partial x^i}(\phi(x)).$$

Comparar esta fórmula com (26). A linearidade de v_γ segue daí:

$$\begin{aligned} v_\gamma(af + bg) &= \sum_{i=1}^n \frac{dx^i}{dt}(0) \frac{\partial ((af + bg) \circ \phi^{-1})}{\partial x^i}(\phi(x)) \\ &= a \sum_{i=1}^n \frac{dx^i}{dt}(0) \frac{\partial (f \circ \phi^{-1})}{\partial x^i}(\phi(x)) + b \sum_{i=1}^n \frac{dx^i}{dt}(0) \frac{\partial (g \circ \phi^{-1})}{\partial x^i}(\phi(x)) \\ &= av_\gamma(f) + bv_\gamma(g). \end{aligned}$$

A regra de Leibniz é ainda mais direta:

$$v_\gamma(fg) = \left. \frac{d}{dt} [f(\gamma(t))g(\gamma(t))] \right|_{t=0} = \left. \frac{d}{dt} f(\gamma(t)) \right|_{t=0} g(\gamma(0)) + f(\gamma(0)) \left. \frac{d}{dt} g(\gamma(t)) \right|_{t=0}.$$

Logo, $[\gamma] \in T_x M \implies v_\gamma \in \tilde{T}_x M$.

Observamos ainda que, definindo

$$\frac{\partial}{\partial x^i} \Big|_x : C^\infty(M) \rightarrow \mathbb{R} \quad \text{por} \quad \frac{\partial}{\partial x^i} \Big|_x (f) = \frac{\partial(f \circ \phi^{-1})}{\partial x^i}(\phi(x)),$$

obtemos uma derivação linear, ou seja, um elemento de $\tilde{T}_x M$. Pelas contas acima, temos

$$v_\gamma = \sum_{i=1}^n (x^i)'(0) \frac{\partial}{\partial x^i} \Big|_x,$$

que é exatamente a fórmula (26), mas em sua versão para a variedade M . Em seguida, vamos mostrar que toda derivação linear é dessa forma, o que mostra que as duas noções de espaço tangente que definimos são equivalentes. Além disso, se obtem uma base:

$$T_x M = \text{Span} \left\{ \frac{\partial}{\partial x^1} \Big|_x, \frac{\partial}{\partial x^2} \Big|_x, \dots, \frac{\partial}{\partial x^n} \Big|_x \right\}.$$

Tudo isto é consequência da proposição seguinte.

Proposição 23. *A aplicação*

$$\begin{aligned} \Phi : T_x M &\rightarrow \tilde{T}_x M \\ [\gamma] &\mapsto v_\gamma \end{aligned}$$

é um isomorfismo de espaços vetoriais.

Demonstração. Inicialmente, vamos mostrar que Φ é injetiva: dados $v, w \in T_x M$ satisfazendo $\Phi(v) = \Phi(w)$, escolhamos respectivos representantes α e β . Assim, devemos ter

$$\frac{d}{dt} f(\alpha(t)) \Big|_{t=0} = \frac{d}{dt} f(\beta(t)) \Big|_{t=0} \quad \text{para toda } f \in C^\infty(M).$$

Dada uma parametrização $\phi : U \subset M \rightarrow \mathbb{R}^n$, consideramos $f = \phi : U \rightarrow \mathbb{R}^n$ e temos²⁰

$$\frac{d}{dt} (\phi \circ \alpha)(t) \Big|_{t=0} = \frac{d}{dt} (\phi \circ \beta)(t) \Big|_{t=0}.$$

Logo, $\alpha \sim \beta$ e $v = w$.

Finalmente, mostramos a sobrejetividade de Φ . Devemos mostrar que, dada uma derivação qualquer $v \in \tilde{T}_x M$, podemos encontrar uma curva γ satisfazendo (27). Seja $\phi : U \rightarrow \mathbb{R}^n$ uma carta em uma vizinhança de $x = \phi^{-1}(x_0)$ e seja $\phi^i := \pi^i \circ \phi : U \rightarrow \mathbb{R}$ a i -ésima coordenada de ϕ . Definimos

$$a^i := v(\phi^i) \in \mathbb{R} \quad \text{e} \quad a = (a^1, a^2, \dots, a^n) \in \mathbb{R}^n.$$

Para t suficientemente pequeno, está bem definida a curva

$$\gamma(t) := \phi^{-1}(ta) \in U \subset M.$$

Desta forma, pela regra da cadeia usual do espaço euclidiano, segue que

$$v_\gamma(f) = \frac{d}{dt} f(\gamma(t)) \Big|_{t=0} = d(f \circ \phi^{-1})(\phi(x)) \cdot (\phi \circ \gamma)'(0) = \sum_{i=1}^n \frac{\partial(f \circ \phi^{-1})}{\partial x^i}(\phi(x)) v(\phi^i)$$

²⁰Estritamente falando, deveríamos ter uma função ϕ definida em M e não apenas em um subconjunto $U \subset M$. Este detalhe técnico está relacionado com a existência de funções de corte e, conseqüentemente, com as restrições topológicas que impomos sobre M . Não entraremos nestes detalhes.

Por outro lado, o Teorema Fundamental do Cálculo aplicado para a função $f \circ \phi^{-1} : \phi(U) \subset \mathbb{R}^n \rightarrow \mathbb{R}$ implica que

$$\begin{aligned} (f \circ \phi^{-1})(y) - (f \circ \phi^{-1})(x_0) &= \int_0^1 \frac{d}{dt} (f \circ \phi^{-1})(x_0 + t(y - x_0)) dt \\ &= \sum_{i=1}^n (y^i - x_0^i) \int_0^1 \frac{\partial (f \circ \phi^{-1})}{\partial x^i} (x_0 + t(y - x_0)) dt \\ &=: \sum_{i=1}^n (y^i - x_0^i) g_i(y), \end{aligned}$$

onde, conforme indicado, definimos

$$g_i(y) = \int_0^1 \frac{\partial (f \circ \phi^{-1})}{\partial x^i} (x_0 + t(y - x_0)) dt.$$

Escrevendo $w = \phi^{-1}(y) \in M$, obtemos

$$f(w) - f(x) = \sum_{i=1}^n (\phi^i(w) - \phi^i(x)) (g_i \circ \phi)(w).$$

Aplicando v em ambos os lados, utilizando que v é linear e a regra de Leibniz, obtemos²¹

$$\begin{aligned} v(f) &= \sum_{i=1}^n v(\phi^i - \phi^i(x)) (g_i \circ \phi)(x) + (\phi^i(x) - \phi^i(x)) v(g_i \circ \phi) \\ &= \sum_{i=1}^n a^i g_i(x_0) = \sum_{i=1}^n \frac{\partial (f \circ \phi^{-1})}{\partial x^i} (\phi(x)) v(\phi^i) \end{aligned} \quad \square$$

10.3 Geometria Riemanniana

Para fazer geometria, em geral, é interessante a existência de alguma estrutura adicional. Algo que permita, por exemplo, medir comprimentos ou ângulos. A definição seguinte vai neste sentido.

Uma **variedade Riemanniana** é uma variedade diferenciável M tal que, para todo $x \in M$, existe uma forma quadrática $g_x : T_x M \times T_x M \rightarrow \mathbb{R}$, simétrica, positiva definida que satisfaz

$$x \mapsto g_x \text{ de classe } C^\infty.$$

A função g é chamada de **métrica Riemanniana** e define um produto interno em cada espaço tangente $T_x M$. Para $v, w \in T_x M$, são comuns as notações

$$g_x(v, w) = \langle v, w \rangle_x.$$

A métrica Riemanniana fornece uma maneira de transformar M em um espaço métrico, com distância

$$d(x, y) = \inf \int_0^1 \sqrt{g_{\gamma(t)}(\gamma'(t), \gamma'(t))} dt,$$

onde o ínfimo é tomado sobre todas as curvas $\gamma : [0, 1] \rightarrow M$ que satisfazem $\gamma(0) = x$ e $\gamma(1) = y$.

Note que, denotando $|v|_x = \sqrt{g_x(v, v)}$, temos a fórmula usual de comprimento de curvas e a distância acima mede a menor curva que une dois pontos de M :

$$d(x, y) = \inf \int_0^1 |\gamma'(t)|_{\gamma(t)} dt.$$

²¹Deveríamos, mais precisamente, observar que v depende apenas dos valores de f em uma vizinhança do ponto x . Isto, novamente, depende da existência de funções de corte. Mais elementar é a observação que, se $f \equiv c$ é uma função constante, então $v(c) = 0$: de fato, a regra de Leibniz implica que

$$v(1) = v(1 \cdot 1) = 1 \cdot v(1) + v(1) \cdot 1 = 2 \cdot v(1) \implies v(1) = 0 \implies v(c) = c \cdot v(1) = 0.$$

Exercício 4. Verificar que

$$d(x, y) = \inf \sqrt{\int_0^1 g_{\gamma(t)}(\gamma'(t), \gamma'(t)) dt}.$$

A vantagem desta maneira de escrever a distância é que minimizantes automaticamente são geodésicas de velocidade constante.

10.4 Forma de volume

Se A está em uma vizinhança parametrizada por uma carta ϕ , temos

10.5 Aplicação exponencial

11 Teorema de McCann

Supor M variedade Riemanniana compacta e $c : M \times M \rightarrow \mathbb{R}$ a função custo dada por

$$c(x, y) = \frac{d(x, y)^2}{2}.$$

Temos.

Proposição 24. A função custo $c = d^2/2$ é semicôncava me M .

Demonstração. Villani [23, Apeêndice do Capítulo 10]. □

12 Espaços de Wasserstein

Nesta seção, vamos ver como a teoria de transporte ótimo permite definir uma distância no espaço das medidas de probabilidade em X .

Consideramos um espaço Polonês (X, d) e duas medidas de probabilidade $\mu, \nu \in \mathcal{P}(X)$. Definimos a **distância de Wasserstein** entre μ e ν como

$$W_p(\mu, \nu) := \left(\inf_{\gamma \in \text{ADM}(\mu, \nu)} \int_X d(x, y)^p d\gamma(x, y) \right)^{1/p}. \quad (28)$$

Vamos mostrar que W_p satisfaz as propriedades de distância. No entanto, dadas duas medidas quaisquer, é possível ter $W_p(\mu, \nu) = +\infty$. Para um $x_0 \in X$ qualquer fixado, definimos o **Espaço de Wasserstein** de ordem $p \in [1, +\infty)$ por²²

$$\mathcal{P}_p(X) = \left\{ \mu \in \mathcal{P}(X); \int_X d(x, x_0)^p d\mu(x) < +\infty \right\}.$$

Em outras palavras, o espaço de Wasserstein é o espaço das medidas de probabilidade com momento de ordem p finito. Vamos, neste curso, sempre pensar em $\mathcal{P}_p(X)$ equipado com a distância de Wasserstein.

Proposição 25. A distância de Wasserstein W_p , definida em (28), de fato define uma distância no conjunto $\mathcal{P}_p(X)$.

Demonstração. Vamos verificar cada um dos axiomas de métrica.

²²É um exercício simples mostrar que a definição é independente do $x_0 \in X$ escolhido.

- 1) *Identidade de indiscerníveis.* Temos $W_p(\mu, \mu) \geq 0$. Além disso, $\gamma = (Id \times Id)_\# \mu$ é claramente admissível e tem custo nulo. Segue que $W_p(\mu, \mu) = 0$.

Por outro lado, se $W_p(\mu, \nu) = 0$ e γ é um plano de transporte ótimo, devemos ter

$$\text{supp } \gamma \subseteq \{(x, y) \in X \times X; x = y\}.$$

Logo, para γ quase todo ponto, $\pi^X = \pi^Y$; portanto, $\mu = \pi_\#^X \gamma = \pi_\#^Y \gamma = \nu$

- 2) *Simetria.* Dado um plano de transporte ótimo γ entre μ e ν , temos que $(y, x)_\# \gamma$ é um plano de transporte ótimo entre ν e μ , mostrando que $W_p(\mu, \nu) = W_p(\nu, \mu)$.
- 3) *Desigualdade triangular.* Para a desigualdade triangular, vamos utilizar resultados de desintegração de medidas e um lema de colagem (Lema 32), ambos tópicos que estão indicados no apêndice, na Subseção 13.1.1. Utilizamos o Lema de Colagem para construir uma medida que “cole” dois planos ótimos. Sejam $\mu_1, \mu_2, \mu_3 \in \mathcal{P}(X)$ e considere γ_{12} plano ótimo entre μ_1 e μ_2 , γ_{23} plano ótimo entre μ_2 e μ_3 . Temos “mesma projeção” na variável x_2 , isto é,

$$\pi_\#^2 \gamma_{12} = \mu_2 = \pi_\#^1 \gamma_{23}.$$

Pelo Lema de Colagem, existe $\gamma \in \mathcal{P}(X \times X \times X)$ tal que

$$\pi_\#^{12} \gamma = \gamma_{12} \quad \text{e} \quad \pi_\#^{23} \gamma = \gamma_{23}.$$

Observamos que $\pi_\#^1 \gamma = \mu_1$ e $\pi_\#^3 \gamma = \mu_3$; logo $\pi_\#^{13} \gamma \in \text{ADM}(\mu_1, \mu_3)$. A partir destas considerações, escrevemos:

$$\begin{aligned} W_p(\mu_1, \mu_3)^p &\leq \int_X d(x_1, x_3)^p d(\pi_\#^{13} \gamma)(x_1, x_3) \\ &= \int_X d(\pi^{13}(x_1, x_2, x_3))^p d\gamma(x_1, x_2, x_3) \\ &= \int_X d(x_1, x_3)^p d\gamma(x_1, x_2, x_3) \\ &\leq \int_X [d(x_1, x_2) + d(x_2, x_3)]^p d\gamma(x_1, x_2, x_3). \end{aligned}$$

Agora, pela desigualdade triangular para a distância d de X , obtemos

$$\begin{aligned} W_p(\mu_1, \mu_3) &\leq \left(\int_X d(x_1, x_2)^p d\gamma(x_1, x_2, x_3) \right)^{1/p} + \left(\int_X d(x_2, x_3)^p d\gamma(x_1, x_2, x_3) \right)^{1/p} \\ &= \left(\int_X d(x_1, x_2)^p d(\pi_\#^{12} \gamma)(x_1, x_2) \right)^{1/p} + \left(\int_X d(x_2, x_3)^p d(\pi_\#^{23} \gamma)(x_2, x_3) \right)^{1/p} \\ &= W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3), \end{aligned}$$

como queríamos mostrar. □

Observação 26. Note que a distância de Wasserstein pode ser encarada como um tipo de norma L^p . Em espaços de Banach, podemos escrever

$$W_p(\mu, \nu) = \left(\inf_{\gamma \in \text{ADM}(\mu, \nu)} \int_X \|x - y\|^p d\gamma(x, y) \right)^{1/p} = \inf_{\gamma \in \text{ADM}(\mu, \nu)} \|x - y\|_{L^p(\gamma)}.$$

Observação 27. A desigualdade de Hölder implica que

$$p \leq q \implies W_p(\mu, \nu) \leq W_q(\mu, \nu);$$

de fato, para qualquer $\gamma \in \text{Adm}(\mu, \nu)$, aplicamos Hölder com $q/p \geq 1$:

$$\int_X d(x, y)^p d\gamma(x, y) \leq \left(\int_X d(x, y)^q d\gamma(x, y) \right)^{p/q} \cdot \gamma(X \times X)^{q/(1-p)} = \left(\int_X d(x, y)^q d\gamma(x, y) \right)^{p/q}.$$

Por outro lado, se d é uma distância limitada, temos

$$\int_X d(x, y)^q d\gamma(x, y) = \int_X d(x, y)^{p+(q-p)} d\gamma(x, y) \leq \text{diam}(X)^{q-p} \int_X d(x, y)^p d\gamma(x, y),$$

o que implica

$$W_q(\mu, \nu) \leq \text{diam}(X)^{1-p/q} W_p(\mu, \nu)^{p/q}.$$

Em particular, as distâncias são equivalentes quando a distância é limitada.

Vamos analisar, a partir de agora, propriedades topológicas do espaço métrico $(\mathcal{P}_p(X), W_p)$. Inicialmente, estudamos a convergência de sequências com a distância de Wasserstein. O resultado que segue pode ser encontrado em [22, Theorem 7.12] ou [1, Proposition 2.4]. Nossa apresentação é como a de [22].

Proposição 28. *Consideramos uma sequência $\mu_k \in \mathcal{P}_p(X)$ e $\mu_k \rightarrow \mu \in \mathcal{P}_p(X)$. Então, as seguintes afirmações são equivalentes:*

(i) *Convergência em distância: $W_p(\mu_k, \mu) \rightarrow 0$.*

(ii) *Convergência fraca $\mu_k \rightarrow \mu$ e convergência dos momentos de ordem p :*

$$\lim_{k \rightarrow +\infty} \int_X d(x, x_0)^p d\mu_k(x) = \int_X d(x, x_0)^p d\mu(x).$$

(iii) *Convergência fraca $\mu_k \rightarrow \mu$ e*

$$\limsup_{k \rightarrow +\infty} \int_X d(x, x_0)^p d\mu_k(x) \leq \int_X d(x, x_0)^p d\mu(x).$$

(iv) *Convergência fraca $\mu_k \rightarrow \mu$ e*

$$\lim_{R \rightarrow +\infty} \limsup_{k \rightarrow +\infty} \int_{X \setminus B_R(x_0)} d(x, x_0)^p d\mu_k(x) = 0.$$

(v) *Para toda função teste φ contínua e com crescimento até ordem p :*

$$|\varphi(x)| \leq C(1 + d(x, x_0)^p),$$

vale que

$$\lim_{k \rightarrow +\infty} \int_X \varphi(x) d\mu_k(x) = \int_X \varphi(x) d\mu(x).$$

Demonstração. Vamos iniciar provando que as quatro últimas afirmações são equivalentes.

(ii) \iff (iii). Que (ii) implica (iii) é trivial. A recíproca segue de semicontinuidade inferior com respeito à convergência fraca: $\mu_k \rightarrow \mu$ implica que

$$\int_X d(x, x_0)^p d\mu(x) \leq \liminf_{k \rightarrow +\infty} \int_X d(x, x_0)^p d\mu_k(x).$$

(ii) \implies (iv). Definimos funções contínuas limitadas:

$$f_R(x) := \inf \{d(x, x_0)^p, R^p\}.$$

Logo, por convergência fraca, temos

$$\lim_{k \rightarrow +\infty} \int_X f_R(x) d\mu_k(x) = \int_X f_R(x) d\mu(x).$$

Além disso, $f_R \nearrow d(\cdot, x_0)^p$, de modo que o Teorema da Convergência Monótona implica

$$\lim_{R \rightarrow +\infty} \int_X f_R(x) d\mu(x) = \int_X d(x, x_0)^p d\mu(x).$$

Além disso, por hipótese, temos

$$\lim_{k \rightarrow +\infty} \int_X d(x, x_0)^p d\mu_k(x) = \int_X d(x, x_0)^p d\mu(x),$$

de modo que

$$\lim_{R \rightarrow +\infty} \lim_{k \rightarrow +\infty} \int_X [d(x, x_0)^p - f_R(x)] d\mu_k(x) = \lim_{k \rightarrow +\infty} \int_X d(x, x_0)^p d\mu_k(x) - \lim_{R \rightarrow +\infty} \int_X f_R(x) d\mu(x) = 0.$$

Agora,

$$d(x, x_0) \geq 2R \iff -R^p \geq -\frac{d(x_0, x)^p}{2^p} \iff d(x, x_0)^p - R^p \geq (1 - 2^{-p})d(x, x_0)^p.$$

Observando que

$$d(x, x_0)^p - f_R(x) = \begin{cases} 0, & \text{se } d(x, x_0) < R \\ d(x, x_0)^p - R^p, & \text{se } d(x, x_0) \geq R \end{cases}$$

e que

$$(1 - 2^{-p}) \int_{X \setminus B_{2R}(x_0)} d(\cdot, x_0)^p d\mu_k \leq \int_{X \setminus B_{2R}(x_0)} [d(\cdot, x_0)^p - R^p] d\mu_k \leq \int_X [d(\cdot, x_0)^p - f_R] d\mu_k,$$

temos (iv):

$$\lim_{R \rightarrow +\infty} \limsup_{k \rightarrow +\infty} \int_{X \setminus B_{2R}(x_0)} d(x, x_0)^p d\mu_k(x) = 0.$$

(iv) \implies (v). Seja φ como em (v). Dado $\varepsilon > 0$, nossa hipótese garante que existe $R > 1$ tal que

$$\limsup_{k \rightarrow +\infty} \int_{X \setminus B_R(x_0)} d(x, x_0)^p d\mu_k(x) < \varepsilon.$$

Como μ tem momento de ordem p finito, aumentando $R > 1$, se necessário, podemos assumir também que

$$\int_{X \setminus B_R(x_0)} d(x, x_0)^p d\mu < \varepsilon.$$

Seja $\xi \in C^0$ uma função de corte: satisfaz $0 \leq \xi \leq 1$, tem suporte compacto e é tal que $\xi \equiv 1$ em $B_R(x_0)$. Temos (note que, em $X \setminus B_R(x_0)$, $1 < R \leq d(x, x_0)$)

$$\begin{aligned} \int_X \varphi d\mu_k - \int_X \varphi d\mu &= \int_X \varphi \xi d(\mu_k - \mu) + \int_X \varphi(1 - \xi) d\mu_k - \int_X \varphi(1 - \xi) d\mu \\ &\leq \int_X \varphi \xi d(\mu_k - \mu) + C \int_{X \setminus B_R(x_0)} (1 + d(x, x_0)^p) d\mu_k - \int_X \varphi d\mu \\ &\leq \int_X \varphi \xi d(\mu_k - \mu) + 2C \int_{X \setminus B_R(x_0)} d(x, x_0)^p d\mu_k - \int_X \varphi d\mu. \end{aligned}$$

Em valor absoluto (e aplicando raciocínio análogo ao anterior na última integral):

$$\left| \int_X \varphi d(\mu_k - \mu) \right| \leq \left| \int_X \varphi \xi d(\mu_k - \mu) \right| + 2C \int_{X \setminus B_R(x_0)} d(x, x_0)^p d\mu_k + 2C \int_{X \setminus B_R(x_0)} d(x, x_0)^p d\mu.$$

Observando que $\varphi \xi$ é contínua e limitada, convergência fraca implica que o primeiro termo do lado direito da estimativa acima vai para zero quando $k \rightarrow +\infty$. Assim, tomando limite superior quando $k \rightarrow +\infty$, obtemos

$$\limsup_{k \rightarrow +\infty} \left| \int_X \varphi d(\mu_k - \mu) \right| \leq 2C \limsup_{k \rightarrow +\infty} \int_{X \setminus B_R(x_0)} d(x, x_0)^p d\mu_k + 2C \int_{X \setminus B_R(x_0)} d(x, x_0)^p d\mu \leq 4C\varepsilon.$$

Sendo $\varepsilon > 0$ arbitrário, segue a afirmação (v).

(v) \implies (ii). É imediato, pois tomando φ limitada qualquer temos a convergência fraca e tomando $\varphi \equiv d(\cdot, x_0)^p$ temos a convergência dos momentos de ordem p .

Resta mostrar que a convergência em W_p é equivalente a qualquer dos itens (ii) – (v). Suponhamos que $\mu_k \rightarrow \mu$ em métrica. Vamos mostrar que vale (iii). Sendo convergente com a métrica W_p , segue que μ_k é uma sequência de Cauchy; logo, pelo Lema 29 abaixo, $\{\mu_k\}$ forma um conjunto rígido. Pelo Teorema 4 (de Prokhorov), existe uma subsequência que converge fracamente para $\tilde{\mu} \in \mathcal{P}_p(X)$. Por semicontinuidade inferior,

$$W_p(\mu, \tilde{\mu}) \leq \liminf_{k \rightarrow +\infty} W_p(\mu, \mu_k) = 0,$$

de modo que $\mu = \tilde{\mu}$. Agora, dado $\varepsilon > 0$, usamos a desigualdade triangular e obtemos²³

$$d(x, x_0)^p \leq (d(y, x_0) + d(x, y))^p \leq (1 + \varepsilon) d(y, x_0)^p + C_\varepsilon d(x, y)^p.$$

Integrando com respeito a γ_k , onde γ_k é plano de transporte ótimo entre μ_k e μ ,

$$\int_X d(x, x_0)^p d\mu_k \leq (1 + \varepsilon) \int_X d(y, x_0)^p d\mu + C_\varepsilon \int_X d(x, y)^p d\gamma_k.$$

O último termo vai para zero porque $W_p(\mu_k, \mu) \rightarrow 0$. Logo,

$$\limsup_{k \rightarrow +\infty} \int_X d(x, x_0)^p d\mu_k \leq (1 + \varepsilon) \int_X d(y, x_0)^p d\mu.$$

Mandando $\varepsilon \rightarrow 0$, obtemos (iii).

Por outro lado, suponhamos (ii). Temos que $\mu_k \rightarrow \mu$ e

$$\lim_{k \rightarrow +\infty} \int_X d(x, x_0)^p d\mu_k(x) = \int_X d(x, x_0)^p d\mu(x).$$

Agora, considere um plano de transporte ótimo γ_k entre μ_k e μ . Pelo Teorema de Prokhorov, $\{\mu_k\} \subset \mathcal{P}(X)$ é rígido. Logo, $\{\gamma_k\} \subset \mathcal{P}(X \times X)$ é rígido (marginais em conjuntos rígidos – ver Lema 7). Novamente pelo Teorema de Prokhorov, (a menos de uma subsequência) tem-se

$$\gamma_k \rightarrow \gamma \in \mathcal{P}(X \times X).$$

Por estabilidade (**provar estabilidade!**), temos que $\gamma \in \text{ADM}(\mu, \mu)$ é também ótima. Segue que γ é o plano trivial $(\text{Id}, \text{Id})_{\#}\mu$. Além disso, qualquer subsequência convergente deve convergir para este limite, de modo que toda a sequência converge para γ . Temos

$$\begin{aligned} W_p(\mu_k, \mu)^p &= \int d(x, y)^p d\gamma_k(x, y) \\ &= \int \inf \{d(x, y), R\}^p d\gamma_k + \int (d(x, y)^p - R^p)_+ d\gamma_k \end{aligned} \tag{29}$$

²³Aqui estamos utilizando que existe $C_\varepsilon > 0$ tal que

$$(a + b)^p \leq (1 + \varepsilon)a^p + C_\varepsilon b^p.$$

Para provar isto, divida tudo por b^p e note que a função $g(x) = (x+1)^p - (1+\varepsilon)x^p$ é contínua e satisfaz $g(0) = 1$, $g(x) \rightarrow -\infty$ quando $x \rightarrow +\infty$; logo, é limitada por cima em $(0, +\infty)$.

Note que, dado $\varepsilon > 0$ e x, y tais que $d(x, y) > R$, temos que

$$\max \{d(x, x_0), d(y, x_0)\}$$

é maior do que ambos $R/2$ e $d(x, y)/2$. Isto pode ser reescrito como

$$\chi_{\{d(x, y) > R\}} \leq \chi_{\{d(x, x_0) > R/2 \text{ e } d(x, x_0) > d(x, y)/2\}} + \chi_{\{d(y, x_0) > R/2 \text{ e } d(y, x_0) > d(x, y)/2\}}$$

Assim,

$$\begin{aligned} [d(x, y)^p - R^p]_+ &= [d(x, y)^p - R^p] \chi_{\{d(x, y) > R\}} \\ &\leq d(x, y)^p \chi_{\{d(x, y) > R\}} \\ &\leq d(x, y)^p \chi_{\{d(x, x_0) > R/2 \text{ e } d(x, x_0) > d(x, y)/2\}} + d(x, y)^p \chi_{\{d(y, x_0) > R/2 \text{ e } d(y, x_0) > d(x, y)/2\}} \\ &\leq 2^p d(x, x_0)^p \chi_{\{d(x, x_0) > R/2\}} + 2^p d(y, x_0)^p \chi_{\{d(y, x_0) > R/2\}}. \end{aligned}$$

Voltando à integral em (29):

$$W_p(\mu_k, \mu)^p \leq \int \inf \{d(x, y), R\}^p d\gamma_k + 2^p \int_{X \setminus B_{R/2}(x_0)} d(x, x_0) d\mu_k(x) + 2^p \int_{X \setminus B_{R/2}(x_0)} d(y, x_0) d\mu(y).$$

Fazendo $k \rightarrow +\infty$, obtemos

$$\limsup_{k \rightarrow +\infty} W_p(\mu_k, \mu)^p \leq 2^p \limsup_{k \rightarrow +\infty} \left(\int_{X \setminus B_{R/2}(x_0)} d(x, x_0) d\mu_k(x) + \int_{X \setminus B_{R/2}(x_0)} d(y, x_0) d\mu(y) \right).$$

Fazendo $R \rightarrow +\infty$ e utilizando (iii), obtemos

$$\limsup_{k \rightarrow +\infty} W_p(\mu_k, \mu)^p \leq 0,$$

o que implica, como queríamos mostrar, que

$$\lim_{k \rightarrow +\infty} W_p(\mu_k, \mu)^p = 0. \quad \square$$

Em seguida, provamos que toda sequência de Cauchy em \mathcal{P}_p é rígida, exatamente como em [23, Lemma 6.14].

Lema 29. *Seja X um espaço Polonês e $\{\mu_n\}$ uma sequência de Cauchy em $\mathcal{P}_p(X)$ com a distância de Wasserstein W_p . Então, $\{\mu_n\}$ é rígida.*

Demonstração. De acordo com a Observação 27, $\{\mu_n\}$ também é de Cauchy com respeito à métrica W_1 . Temos que, dado $\varepsilon > 0$, existe $N \in \mathbb{N}$ tal que

$$n \geq N \implies W_1(\mu_n, \mu_N) < \varepsilon^2.$$

Afirmção: dado $n \in \mathbb{N}$, existe $j \in \{1, 2, \dots, N\}$ tal que

$$W_1(\mu_n, \mu_j) < \varepsilon^2.$$

De fato, se $n \geq N$ escolhemos $j = N$; se $n < N$ escolhemos $j = n$.

O conjunto $\{\mu_1, \mu_2, \dots, \mu_N\}$ é um conjunto finito de medidas de radon; logo, é rígido. Isto significa que existe um conjunto compacto K tal que

$$\mu_j(X \setminus K) < \varepsilon \quad \text{para todo } j \in \{1, 2, \dots, N\}.$$

Sendo K compacto, podemos cobrir K por um número finito de bolas de raio ε :

$$K \subseteq \bigcup_{i=1}^m B_\varepsilon(x_i) =: U$$

Aumentando U pelo fator ε :

$$U_\varepsilon := \{x \in U; d(x, U) < \varepsilon\},$$

podemos considerar uma função contínua ϕ que satisfaça $\chi_U \leq \phi \leq \chi_{U_\varepsilon}$. Um exemplo de tal função pode ser

$$\phi(x) = \left(1 - \frac{d(x, U)}{\varepsilon}\right)_+,$$

que é $1/\varepsilon$ -Lipschitz. Temos, pela fórmula de dualidade de Kantorovich-Rubinstein (Teorema 19),

$$\mu_n(U_\varepsilon) = \int \chi_{U_\varepsilon} d\mu_n \geq \int \phi d\mu_n = \int \phi d\mu_j + \int \phi d\mu_n - \int \phi d\mu_j \geq \mu_j(U) - \frac{W_1(\mu_n, \mu_j)}{\varepsilon}.$$

Para $j \leq N$ temos que $\mu_j(U) \geq \mu_j(K) \geq 1 - \varepsilon$ e que, para cada n , podemos usar a afirmação acima para limitar W_1 . Logo, para todo $n \in \mathbb{N}$,

$$\mu_n(U_\varepsilon) \geq 1 - 2\varepsilon.$$

Fosse U_ε compacto, teríamos terminado a prova. Mas pode não ser. Um truque para contornar esta dificuldade é o seguinte²⁴: trocando ε por $\varepsilon/2^{\ell+1}$ acima, obtemos

$$\mu_n\left(\bigcup_{i=1}^{m(\ell)} B_{2^\ell}(x_i)\right) \geq 1 - \frac{\varepsilon}{2^\ell}$$

e definimos

$$\mathcal{K} := \bigcap_{\ell=1}^{+\infty} \overline{\bigcup_{i=1}^{m(\ell)} B_{2^\ell}(x_i)}.$$

Assim, por construção, \mathcal{K} é totalmente limitado: dado qualquer $\delta > 0$, existe uma cobertura finita por bolas abertas de raio δ . Sendo totalmente limitado e fechado em um espaço métrico completo, \mathcal{K} é compacto. \square

Teorema 30. *Seja X um espaço polonês. Então o espaço de Wasserstein $(\mathcal{P}_p(X), W_p)$ também é um espaço polonês. Além disso, tem-se*

- (i) *Um conjunto $\mathcal{K} \subseteq \mathcal{P}_p(X)$ é relativamente compacto se, e somente se, \mathcal{K} é rígido (justo) e p uniformemente integrável.*
- (ii) *Dada uma sequência $\mu_n, \mu \in \mathcal{P}_p(X)$, vale que*

$$W_p(\mu_n, \mu) \rightarrow 0 \iff \begin{cases} \mu_n \rightharpoonup \mu \text{ (convergência fraca que vínhamos considerando)} \\ \mu_n \text{ tem } p\text{-momentos uniformemente integráveis.} \end{cases}$$

Este resultado aparece em [2, Proposition 7.1.5]. Apresentamos uma prova baseada nos argumentos encontrados em [2, 23].

Demonstração. Vimos na Proposição 25 que $(\mathcal{P}_p(X), W_p)$ é um espaço métrico. Passamos a provar que é um espaço separável: consideramos $\{x_i\}_{i \in \mathbb{N}} \subset X$ um subconjunto denso enumerável de X e definimos

$$\mathcal{D} := \left\{ \sum_{i=1}^N r_i \delta_{x_i}; r_i \in \mathbb{Q} \text{ e } N \in \mathbb{N} \right\}.$$

Em outras palavras, \mathcal{D} é o conjunto de combinações lineares finitas e com coeficientes racionais de massas unitárias nos pontos de $\{x_i\}$. Já que $\mu \in \mathcal{P}_p(X)$, temos que, dado $\varepsilon > 0$, existe um compacto $K \subset X$ tal que

$$\int_{X \setminus K} d(x, x_0)^p d\mu(x) < \varepsilon^p.$$

²⁴Ver prova do Teorema de Prokhorov – Teorema 4 – onde o mesmo truque é utilizado.

Utilizamos este $\varepsilon > 0$ para obter uma cobertura finita (com centro em alguns x_i) do compacto K :

$$K \subseteq \bigcup_{i=1}^N B_{\varepsilon/2}(x_i).$$

Além disso, obtemos uma cobertura disjunta ao considerar:

$$K \subseteq \bigcup_{i=1}^N \tilde{B}_k \quad \text{onde } \tilde{B}_k := B_\varepsilon(x_k) \setminus \bigcup_{j < k} B_\varepsilon(x_j).$$

Definimos

$$f(x) := \begin{cases} x_k, & \text{se } x \in \tilde{B}_k \cap K, \quad k = 1, 2, \dots, N, \\ x_0, & \text{se } x \in X \setminus K \end{cases}.$$

Notamos que $f_{\#}\mu \in \mathcal{D}$, pois

$$(f_{\#}\mu)(A) = \mu(f^{-1}(A)) = \mu(X \setminus K)\delta_{x_0} + \sum_{i=1}^N \mu(\tilde{B}_i \cap K)\delta_{x_i} =: \sum_{i=0}^N a_i \delta_{x_i}.$$

Além disso, temos $d(x, f(x)) < \varepsilon$ para todo $x \in K$, de modo que

$$\begin{aligned} \int_X d(x, f(x))^p d\mu(x) &= \int_K d(x, f(x))^p d\mu(x) + \int_{X \setminus K} d(x, x_0)^p d\mu(x) \\ &\leq \varepsilon^p \mu(K) + \varepsilon^p = 2\varepsilon^p. \end{aligned}$$

Utilizando que $(I, f)_{\#}\mu \in \text{ADM}(\mu, f_{\#}\mu)$, temos que

$$W_p(\mu, f_{\#}\mu) \leq \sqrt[p]{2} \cdot \varepsilon,$$

o que mostra que \mathcal{D} é denso em $\mathcal{P}_p(X)$.

Para mostrar que $\mathcal{P}_p(X)$ é completo, basta supor que X é um espaço métrico completo. De fato, seja μ_k uma sequência de Cauchy em $\mathcal{P}_p(X)$. Então, pelo Lema 29, $\{\mu_k\}$ é um conjunto rígido. Segue do Teorema de Prokhorov (ver Teorema 4) que existe uma subsequência fracamente convergente $\mu_{k_j} \rightarrow \mu$. Por semicontinuidade inferior do custo total, temos que

$$\int_X d(x, x_0)^p d\mu(x) \leq \liminf_{j \rightarrow +\infty} \int_X d(x, x_0)^p d\mu_{k_j}(x) < +\infty$$

e, logo, $\mu \in \mathcal{P}_p(X)$. Por semicontinuidade da distância W_p , temos

$$W_p(\mu, \mu_{k_i}) \leq \liminf_{i \rightarrow +\infty} W_p(\mu_{k_j}, \mu_{k_i}),$$

de modo que

$$\limsup_{i \rightarrow +\infty} W_p(\mu, \mu_{k_i}) \leq \limsup_{i, j \rightarrow +\infty} W_p(\mu_{k_j}, \mu_{k_i}) = 0.$$

Segue que $\mu_{k_j} \rightarrow \mu \in \mathcal{P}_p(X)$ em distância. Sendo μ_k uma sequência de Cauchy com subsequência convergente, μ_k deve ser toda convergente. \square

13 Apêndices

13.1 Conceitos de teoria da medida

Referências básicas de Teoria da Medida e de Teoria Geométrica da Medida incluem: Evans, Gariepy [10], Folland [11], Maggi [15].

13.1.1 Desintegração de medidas e lema de colagem

Nesta seção enunciamos o Teorema de Desintegração de Medidas, que informalmente pode ser pensado como o processo inverso de construir uma medida produto.

Teorema 31 (Teorema de Desintegração). *Sejam X e Y espaços Poloneses (mais geralmente, poderiam ser espaços de Radon), $\phi : X \rightarrow Y$ mensurável e $\mu \in \mathcal{P}(X)$ e $\nu \in \mathcal{P}(Y)$ probabilidades tais que*

$$\nu = \phi_{\#}\mu.$$

Então, existe uma família de medidas de probabilidade $\{\mu_y\} \subseteq \mathcal{P}(X)$ tais que

(i) a função $y \mapsto \mu_y$ é Borel-mensurável, no seguinte sentido: para todo conjunto de Borel $B \subseteq X$, vale que a função $y \mapsto \mu_y(B)$ é Borel-mensurável;

(ii) μ_y está concentrada na “fibra” $\phi^{-1}(y)$, isto é, $\mu_y(X \setminus \phi^{-1}(y)) = 0$. Em particular,

$$\mu_y(E) = \mu_y(E \cap \phi^{-1}(y)).$$

(iii) Para toda função positiva $f : X \rightarrow \mathbb{R}^+$, tem-se

$$\int_X f(x) d\mu(x) := \int_Y \left[\int_{\phi^{-1}(y)} f(x) d\mu_y(x) \right] d\nu(y). \quad (30)$$

Em particular, para $E \subseteq X$ mensurável, temos

$$\mu(E) = \int_Y \mu_y(E) d\nu(y) = \int_Y \mu_y(E \cap \phi^{-1}(y)) d\nu(y).$$

Escrevendo a equação “em forma diferencial”, temos, mais informalmente, a desintegração

$$d\mu(x) = d\mu_y(x) d\nu(y).$$

No Lema de Colagem abaixo, utilizamos a seguinte notação (informal, mas esperamos que intuitiva) para projeções

$$\pi^Y(x, y) = y, \quad \pi^Y(y, z) = y, \quad \pi^Y(x, y, z) = y, \quad \pi^{X,Y}(x, y, z) = (x, y), \quad \text{etc.}$$

Observe que denotamos funções com domínios diferentes com o mesmo símbolo π^Y , mas o contexto deve deixar claro qual estamos considerando. Além disso, para $\pi^Y : X \times Y \rightarrow Y$, cada fibra $(\pi^Y)^{-1}(y)$ pode ser identificada com X de modo que podemos considerar a desintegração em medidas $\{\mu_y\} \subset \mathcal{P}(X)$ (ao invés de $\mathcal{P}(X \times Y)$). É isto que faremos na demonstração abaixo.

Lema 32 (Lema de Colagem). *Sejam X, Y e Z espaços Poloneses e duas medidas*

$$\gamma^1 \in \mathcal{P}(X \times Y) \quad \text{e} \quad \gamma^2 \in \mathcal{P}(Y \times Z)$$

com as mesmas projeções em Y , isto é,

$$\pi_{\#}^Y \gamma^1 = \pi_{\#}^Y \gamma^2.$$

Então existe uma medida $\gamma \in \mathcal{P}(X \times Y \times Z)$ que é “de colagem”, isto é, tal que

$$\pi_{\#}^{X,Y} \gamma = \gamma^1 \quad \text{e} \quad \pi_{\#}^{Y,Z} \gamma = \gamma^2.$$

Demonstração. A prova é uma aplicação do Teorema de Desintegração. Denotamos $\nu = \pi_{\#}^Y \gamma^1 = \pi_{\#}^Y \gamma^2$. Por desintegração para cada uma das medidas, escrevemos

$$d\gamma^1(x, y) = d\gamma_y^1(x) d\nu(y) \quad \text{e} \quad d\gamma^2(y, z) = d\gamma_y^2(z) d\nu(y).$$

Definimos, então

$$d\gamma(x, y, z) := d(\gamma_y^1(x) \times \gamma_y^2(z)) d\nu(y). \quad \square$$

13.2 Semicontinuidade

Seja X um espaço topológico. Uma função $f : X \rightarrow [-\infty, +\infty)$ é dita **semicontínua inferiormente em** $x_0 \in X$ quando, dado $\varepsilon > 0$, existe uma vizinhança U de x_0 tal que

$$f(x) \geq f(x_0) - \varepsilon \text{ para todo } x \in U.$$

Dizemos simplesmente **semicontínua inferiormente** quando f é semicontínua inferiormente em todo ponto de X .

Proposição 33. *Seja X um espaço topológico e $f : X \rightarrow [-\infty, +\infty)$. São equivalentes:*

- (a) f é semicontínua inferiormente.
- (b) Para todo $a \in \mathbb{R}$, $\{x \in X; f(x) > a\}$ é aberto.
- (c) Para todo $a \in \mathbb{R}$, $\{x \in X; f(x) \leq a\}$ é fechado.

Além disso, se X é um espaço métrico, então as anteriores também são equivalentes a:

- (d) Para todo $x \in \mathbb{R}$ e $x_n \rightarrow x$, tem-se $f(x) \leq \liminf_{n \rightarrow +\infty} f(x_n)$.

Demonstração. Vejamos que (a) \implies (b). Sejam dados $a \in \mathbb{R}$ e $x_0 \in \{x \in X; f(x) > a\}$. Escolha $\varepsilon > 0$ tal que $f(x_0) - \varepsilon > a$. Sendo f semicontínua inferiormente em x_0 , existe uma vizinhança U de x_0 tal que

$$f(x) \geq f(x_0) - \varepsilon > a, \text{ para todo } x \in U.$$

Logo, $U \subset \{x \in X; f(x) > a\}$, donde todo ponto de $\{x \in X; f(x) > a\}$ é interior.

Agora, mostraremos a recíproca (b) \implies (a). Dado $\varepsilon > 0$ e $x_0 \in X$. Consideramos $a = f(x_0) - \varepsilon$. Por hipótese, existe uma vizinhança de x_0 tal que, para x nesta vizinhança, $f(x) > a = f(x_0) - \varepsilon$.

A equivalência (b) \iff (c) é imediata.

Finalmente, mostramos que em um espaço métrico, (a) \iff (d). Se f é semicontínua em $x \in X$, então dado $\varepsilon > 0$, existe $\delta > 0$ tal que $d(x, y) < \delta$ implica $f(y) \geq f(x) - \varepsilon$. Utilizando que $x_n \rightarrow x$, temos que

$$n \geq n_0 \implies d(x, x_n) < \delta \implies f(x_n) \geq f(x) - \varepsilon.$$

Logo, $f(x) \leq \liminf_{n \rightarrow +\infty} f(x_n)$. Para a recíproca, vamos mostrar a contrapositiva. Suponhamos que não vale (a). Desta forma, existe $\varepsilon > 0$ tal que, para todo $\delta_n = 1/n > 0$, existe x_n tal que $d(x_n, x) < 1/n$ mas $f(x_n) < f(x) - \varepsilon$. Isto implica que $\liminf f(x_n) \leq f(x) - \varepsilon < f(x)$, de modo que não vale (e). \square

Seguem algumas outras propriedades de funções semicontínuas.

Proposição 34. *Uma função $f : X \rightarrow \mathbb{R}$ é contínua se, e somente se, ambas f e $-f$ são semicontínuas inferiormente.*

Demonstração. Imediato. \square

Mais importante é a seguinte.

Proposição 35. *Seja X um espaço topológico. Então*

- (i) *Se X é um espaço de Hausdorff²⁵ normal²⁶ e $f : X \rightarrow [0, 1]$ é semicontínua inferiormente, então*

$$f(x) = \sup \{h(x); h \text{ é contínua e } 0 \leq h \leq f\}.$$

Em outras palavras, f é pontualmente o supremo das funções contínuas menores do que f .

²⁵Um espaço topológico é dito de Hausdorff ou T_2 quando “separa” pontos: se $x_1, x_2 \in X$ e $x_1 \neq x_2$, então existem abertos disjuntos U, V tais que $x_1 \in U$ e $x_2 \in V$. Em particular, num espaço de Hausdorff, todo conjunto com um ponto só é fechado.

²⁶Um espaço topológico é dito normal quando “separa” conjuntos fechados: se F_1, F_2 são fechados disjuntos, então existem abertos disjuntos U, V tais que $F_1 \subset U$ e $F_2 \subset V$.

(ii) Se X é um espaço metrizável e $f : X \rightarrow [0, 1]$ é semicontinua inferiormente, então podemos escrever

$$f(x) = \lim_{k \rightarrow +\infty} f_k(x) \text{ pontualmente,}$$

onde, para todo $k \in \mathbb{N}$, f_k é Lipschitz-contínua e $f_k \leq f_{k+1}$.

Demonstração. (i) Seja $x_0 \in X$. Se $f(x_0) = 0$, não há o que demonstrar. Suponhamos que $f(x_0) > 0$ e seja $a \in (0, f(x_0))$ arbitrário. Como f é semicontínua inferiormente, o conjunto

$$U = \{x \in X; f(x) > a\}$$

é aberto. Sendo X espaço de Hausdorff, o conjunto com apenas um ponto $\{x_0\} \subset X$ é fechado. Como X é também um espaço normal, o Lema de Urysohn implica que existe uma função contínua $h : X \rightarrow [0, 1]$ tal que $h(x_0) = a$ e $h(x) = 0$ em $X \setminus U$. Esta propriedade de h pode ser reescrita como

$$0 \leq h \leq a1_U \leq f.$$

Esta última desigualdade segue de: $f(x) > a$ em U e $f(x) \geq 0$ em $X \setminus U$. Sendo a arbitrário, podemos encontrar $h \in C_c(X)$ com $h(x_0) = a$ tão próximo de $f(x_0)$ quanto se queira.

(ii) Sendo X um espaço metrizável, podemos definir $f_k : X \rightarrow [0, 1]$ explicitamente como

$$f_k(x) = \inf_{y \in X} \{f(y) + kd(x, y)\}.$$

Agora vamos verificar que esta sequência satisfaz o que queríamos.

- Claramente, $f_k \leq f_{k+1}$.
- Em seguida, vamos mostrar que f_k é Lipschitz com constante de Lipschitz k . Sejam $x_1, x_2 \in X$. Para qualquer $z \in X$, temos

$$f_k(x_1) \leq f(z) + kd(x_1, z) \leq \{f(z) + kd(x_2, z)\} + kd(x_1, x_2).$$

Fazendo o supremo em $z \in X$, obtemos

$$f_k(x_1) \leq f_k(x_2) + kd(x_1, x_2), \text{ para quaisquer } x_1, x_2 \in X,$$

o que implica que (invertendo os papéis de x_1 e x_2)

$$|f_k(x_1) - f_k(x_2)| \leq kd(x_1, x_2), \text{ para quaisquer } x_1, x_2 \in X.$$

- Finalmente, vamos mostrar convergência pontual, isto é, que para $x \in X$ fixado, tem-se $\lim_k f_k(x) = f(x)$. Aqui que entra a semicontinuidade inferior: dado $\varepsilon > 0$, existe $\delta > 0$ tal que

$$d(x, z) < \delta \implies f(z) \geq f(x) - \varepsilon.$$

Em particular,

$$d(x, y) < \delta \implies f(y) + kd(x, y) \geq f(x) - \varepsilon + kd(x, y) \geq f(x) - \varepsilon. \quad (31)$$

Por outro lado, tomando k suficientemente grande,

$$d(x, y) \geq \delta \implies f(y) + kd(x, y) \geq f(y) + k\delta \stackrel{k \gg 1}{\geq} f(x) - \varepsilon. \quad (32)$$

Juntando (31) e (32), temos

$$k \gg 1 \implies f(y) + kd(x, y) \geq f(x) - \varepsilon, \forall y \in X \implies f_k(x) \geq f(x) - \varepsilon.$$

A desigualdade $f_k(x) \leq f(x) \leq f(x) + \varepsilon$ é imediata. \square

Observação 36. A prova do item (i) acima funciona para X espaço de Hausdorff localmente compacto. Espaços de Hausdorff compactos são normais (exercício!). Daí, no caso localmente compacto, é possível de verificar que o Lema de Urysohn funciona nos seguintes termos: dado K compacto e U vizinhança aberta de K , então existe $f \in C_c(X)$ tal que $1_K \leq f \leq 1_U$.

Observação 37. Espaços métricos são automaticamente de Hausdorff e normais, de modo que o item (i) vale para qualquer espaço métrico. De qualquer maneira, o item (ii) é mais forte, além de ter uma construção mais explícita.

Observação 38. Como pode ser visto em [9, Seção 1.7.15(c)], o item (ii) acima pode ser generalizado para espaços topológicos mais gerais. Segundo [9], é um resultado de Tong que funções semicontínuas inferiormente podem ser aproximadas pontualmente por uma sequência crescente de funções contínuas se, e somente se, X é um espaço topológico T_1 e perfeitamente normal.²⁷

13.3 Análise convexa

A referência mais conhecida para este assunto é Rockafellar [19]. O material desta seção é uma ampliação de Villani [22, Subseção 2.1.3], que já é bastante inspirado em [19].

Uma função $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é uma **função convexa própria**²⁸ quando f não é identicamente $+\infty$ e, para todo $x, y \in \mathbb{R}^n$ e para todo $0 < \lambda < 1$, tem-se

$$f[(1 - \lambda)x + \lambda y] \leq (1 - \lambda)f(x) + \lambda f(y). \quad (33)$$

Quando a desigualdade é estrita em (33), dizemos que f é **estritamente convexa**.

O **domínio efetivo** de uma função convexa f é o conjunto

$$\text{dom } f := \{x \in \mathbb{R}^n; f(x) < +\infty\}.$$

É fácil de ver que $\text{dom } f$ é um conjunto convexo. É também fácil ver que:

Proposição 39 (Desigualdade de Jensen - versão discreta). A função $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é convexa se, e somente se, para todo $\lambda_1 \geq 0, \lambda_2 \geq 0, \dots, \lambda_m \geq 0$ com $\lambda_1 + \lambda_2 + \dots + \lambda_m = 1$, temos

$$f[\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_m x_m] \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \dots + \lambda_m f(x_m).$$

Uma propriedade geométrica que segue da definição é que o gráfico de uma função convexa está acima do plano tangente:

Proposição 40. Se $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é uma função convexa própria que é diferenciável em $x \in \text{dom } f$, então

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle,$$

para todo $y \in \mathbb{R}^n$. Em particular, ∇f é monótono em pontos de diferenciabilidade:

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0.$$

Demonstração. Reescrevendo (33) como

$$\frac{f[x + \lambda(y - x)] - f(x)}{\lambda} \leq f(y) - f(x)$$

e mandando $\lambda \rightarrow 0$, obtemos o resultado. □

Na verdade, esta propriedade de o gráfico estar acima dos planos tangentes caracteriza funções convexas:

²⁷ Ser um espaço T_1 é menos do que ser Hausdorff: dados dois pontos distintos, existe um conjunto aberto que contém apenas um deles. Além disso, um espaço topológico X é dito *perfeitamente normal* quando, além de ser normal, todo conjunto fechado puder ser escrito como interseção enumerável de conjuntos abertos (isto é, todo fechado é um G_δ).

²⁸ No contexto de Análise Convexa, a palavra *própria* indica que $f > -\infty$ em todo ponto e que $f(x) < +\infty$ para pelo menos um ponto [19, Seção 4].

Proposição 41. Suponhamos que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa própria de classe C^1 que satisfaz

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle,$$

para todo $x, y \in \mathbb{R}^n$. Então f é uma função convexa.

Demonstração. É sempre verdade que (como se vê tomando $x = y$)

$$f(y) \leq \sup_{x \in \mathbb{R}^n} [f(x) + \langle \nabla f(x), y - x \rangle].$$

Logo, nossa hipótese garante que, na verdade,

$$f(y) = \sup_{x \in \mathbb{R}^n} [f(x) + \langle \nabla f(x), y - x \rangle].$$

O resultado segue da observação que o supremo pontual de uma família qualquer de funções convexas é também uma função convexa. \square

Exercício 5. Mostre que o supremo pontual de uma família qualquer de funções convexas é também uma função convexa, isto é, dadas f_λ convexas, temos que

$$f(y) := \sup_{\lambda \in \Lambda} f_\lambda(y)$$

é convexa.

Em seguida, mostramos outra caracterização de funções convexas, mas com derivadas de ordem dois.

Proposição 42. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função de classe C^2 . Então f é convexa se, e somente se, $\nabla^2 f(x)$ é positiva semi-definida, para todo $x \in \text{dom } f$, isto é,

$$\langle \nabla^2 f(x)y, y \rangle \geq 0,$$

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

Demonstração. \implies Como $x = \frac{x+ty}{2} + \frac{x-ty}{2}$, a definição de função convexa implica que

$$2f(x) \leq f(x+ty) + f(x-ty).$$

Logo,

$$\frac{f(x+ty) + f(x-ty) - 2f(x)}{t^2} \geq 0.$$

Fazendo $t \rightarrow 0$, obtemos o que queríamos.

\impliedby A recíproca segue de duas aplicações do Teorema Fundamental do Cálculo e da proposição anterior:

$$\begin{aligned} f(y) - f(x) &= \int_0^1 \frac{d}{dt} f[x + t(y-x)] dt \\ &= \int_0^1 \langle \nabla f[x + t(y-x)], y-x \rangle dt \\ &= \langle \nabla f(x), y-x \rangle + \left\langle \int_0^1 [\nabla f[x + t(y-x)] - \nabla f(x)] dt, y-x \right\rangle \\ &= \langle \nabla f(x), y-x \rangle + \left\langle \int_0^1 \int_0^1 \frac{d}{ds} \nabla f[x + st(y-x)] ds dt, y-x \right\rangle \\ &= \langle \nabla f(x), y-x \rangle + \left\langle \int_0^1 \int_0^1 \nabla^2 f[x + st(y-x)] (t(y-x)) ds dt, y-x \right\rangle \\ &= \langle \nabla f(x), y-x \rangle + \int_0^1 \int_0^1 t \langle \nabla^2 f[x + st(y-x)] (y-x), y-x \rangle ds dt \\ &\geq \langle \nabla f(x), y-x \rangle. \end{aligned}$$

\square

13.3.1 Regularidade

A questão seguinte de nosso interesse é a análise da regularidade de funções convexas. Por exemplo, dada uma função convexa própria,

- será f contínua?
- será f diferenciável?

Passamos a estudar estas questões.

Teorema 43. *Suponhamos que $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é uma função convexa própria. Então, f é localmente limitada e localmente Lipschitz em $\text{int}(\text{dom } f)$. Em particular²⁹, f é diferenciável em quase todo ponto de $\text{dom } f$. Além disso, para $x \in \text{int}(\text{dom } f)$,*

$$\sup_{z \in B_{r/2}(x)} |f(z)| \leq C \int_{B_r(x)} |f(y)| \, dy \quad (34)$$

e

$$\text{ess sup}_{z \in B_{r/2}(x)} |\nabla f(z)| \leq \frac{C}{r} \int_{B_r(x)} |f(y)| \, dy. \quad (35)$$

Vamos apresentar três provas de que funções convexas são localmente de Lipschitz. A primeira vale apenas em dimensão um, mas é bastante instrutiva. A segunda, que pode ser encontrada em Evans, Gariepy [10, Seção 6.3], utiliza técnicas de aproximações e funções corte e automaticamente implica em (34) e (35), a partir de molificadores³⁰. Finalmente, apresentamos uma prova em dimensão $n \geq 1$ que é mais elementar.

Demonstração em dimensão 1. Se $n = 1$, $\text{int dom } f$ é um intervalo (a, b) . Em dimensão um, é possível caracterizar funções convexas como: dados $a < x < b$, temos

$$\frac{f(x) - f(a)}{x - a} \leq \frac{f(b) - f(a)}{b - a} \leq \frac{f(b) - f(x)}{b - x}. \quad (36)$$

Para ver isso, tomamos $t = \frac{x - a}{b - a} \in (0, 1)$ de modo que $x = (1 - t)a + tb$. Logo,

$$f(x) \leq (1 - t)f(a) + tf(b) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a).$$

Então,

$$\frac{f(x) - f(a)}{x - a} \leq \frac{f(b) - f(a)}{b - a}.$$

Considerando $s \in (0, 1)$ tal que $x = sa + (1 - s)b$, conseguimos a segunda desigualdade em (36).

Considere $(c, d) \subset \subset (a, b)$. Iterando (36), temos que, para $a < c < x < y < d < b$,

$$\frac{f(c) - f(a)}{c - a} \leq \frac{f(y) - f(x)}{y - x} \leq \frac{f(b) - f(d)}{b - d}.$$

Escolhendo

$$M := \max \left\{ \left| \frac{f(c) - f(a)}{c - a} \right|, \left| \frac{f(b) - f(d)}{b - d} \right| \right\},$$

temos $|f(y) - f(x)| \leq M|x - y|$, para quaisquer $x, y \in (c, d)$. □

Em seguida apresentamos a demonstração encontrada em Evans & Gariepy.

²⁹O Teorema de Rademacher afirma que funções Lipschitz são diferenciáveis em quase todo ponto.

³⁰Molificadores

Demonstração encontrada em [10]. Suponhamos inicialmente que f é convexa e suave. Assim, automaticamente f é localmente de Lipschitz e apenas precisamos mostrar (34) e (35).

Seja $y \in B_r(x)$ e $z \in B_{r/2}(x)$. Pela Proposição 40, temos

$$f(y) \geq f(z) + \langle \nabla f(z), y - z \rangle,$$

de modo que, integrando com respeito a y em $B_{r/2}(z)$, obtemos³¹

$$\int_{B_{r/2}(z)} f(y) \, dy \geq f(z) |B_{r/2}(z)| + \int_{B_{r/2}(z)} \langle \nabla f(z), y - z \rangle \, dy = f(z) |B_{r/2}(z)|.$$

Logo, já que $B_{r/2}(z) \subset B_r(x)$, temos

$$f(z) \leq \frac{1}{|B_{r/2}(z)|} \int_{B_{r/2}(z)} f(y) \, dy \leq \frac{|B_r(x)|}{|B_{r/2}(z)|} \int_{B_r(x)} |f(y)| \, dy = C_n \int_{B_r(x)} |f(y)| \, dy. \quad (37)$$

Em seguida, consideramos uma função de corte $\xi \in C_c^\infty(\mathbb{R}^n)$ tal que

$$0 \leq \xi \leq 1, \quad \xi \equiv 1 \text{ em } B_{r/2}(x), \quad \xi \equiv 0 \text{ fora de } B_r(x) \text{ e } |\nabla \xi| \leq C/r.$$

Temos

$$f(z) \geq f(y) + \langle \nabla f(y), z - y \rangle \implies f(z)\xi(y) \geq f(y)\xi(y) + \xi(y)\langle \nabla f(y), z - y \rangle.$$

Integrando com respeito a y em $B_r(x)$ e utilizando o Teorema da Divergência, obtemos

$$\begin{aligned} f(z) \int_{B_r(x)} \xi(y) \, dy &\geq \int_{B_r(x)} f(y)\xi(y) \, dy + \int_{B_r(x)} \xi(y)\langle \nabla f(y), z - y \rangle \, dy \\ &= \int_{B_r(x)} f(y)\xi(y) \, dy - \int_{B_r(x)} f(y) \operatorname{div} [\xi(y)(z - y)] \, dy \geq -C \int_{B_r(x)} |f(y)| \, dy. \end{aligned}$$

Esta última desigualdade vem de ser a função $y \mapsto \xi(y) - \operatorname{div} [\xi(y)(z - y)]$ suave e de suporte compacto em \mathbb{R}^n (logo limitada). Logo, já que $\xi \geq 1$, temos

$$f(z) \geq -C \int_{B_r(x)} |f(y)| \, dy. \quad (38)$$

Assim, (37) e (38) implicam (34) para f suave.

Passamos à prova de (35). Para $z \in B_{r/2}(x)$, definimos

$$S_z = \left\{ y \in \mathbb{R}^n; \frac{r}{4} \leq |y - z| \leq \frac{r}{2} \text{ e } \langle \nabla f(z), y - z \rangle \geq \frac{1}{2} |\nabla f(z)| |y - z| \right\}.$$

Geometricamente, S_z é o conjunto dos pontos $y \in \mathbb{R}^n$ que estão no anel $B_{r/2}(z) \setminus B_{r/4}(z)$ e cujo ângulo θ entre o gradiente $\nabla f(z)$ e o vetor $(y - z)$ satisfaz $\cos \theta \geq 1/2$ (ou $-\pi/3 \leq \theta \leq \pi/3$). Desta forma, S_z contem todo um setor de ângulo $2\pi/3$ do anel. Logo,

$$|S_z| \geq \frac{2\pi}{3} \left(\frac{1}{2^n} - \frac{1}{4^n} \right) r^n \equiv C_n r^n \equiv C |B_{r/2}(x)|.$$

Agora, observe que, para todo $y \in S_z$,

$$f(y) \geq f(z) + \langle \nabla f(z), y - z \rangle \geq f(z) + \frac{1}{2} |\nabla f(z)| |y - z| \geq f(z) + \frac{r}{8} |\nabla f(z)|.$$

³¹Utilizando coordenadas polares e o Teorema da Divergência:

$$\int_{B_{r/2}(z)} \langle \nabla f(z), y - z \rangle \, dy = \int_0^{r/2} \rho^n \, d\rho \int_{\mathbb{S}^1} \omega \cdot \nabla f(z) \, d\sigma(\omega) = C(r) \int_{B_1(0)} \operatorname{div}_w (\nabla f(z)) \, dw = 0.$$

Logo,

$$|\nabla f(z)| \leq \frac{8}{r} [f(y) - f(z)].$$

Assim, integrando com respeito a y em S_z , obtemos

$$|\nabla f(z)| \leq \frac{C}{r} \int_{B_{r/2}(x)} |f(y) - f(z)| dy \stackrel{(34)}{\leq} \frac{C}{r} \int_{B_r(x)} |f(y)| dy,$$

que é equivalente a (35).

Para funções convexas (não necessariamente suaves), consideramos, para $\varepsilon > 0$, as aproximações $f^\varepsilon := \eta_\varepsilon * f \in C^\infty(\mathbb{R}^n)$, onde η_ε são os molificadores usuais; ver e.g. [10, Seção 4.2]. Em particular, $f^\varepsilon \rightarrow f$ em $L^1_{\text{loc}}(\mathbb{R}^n)$ e $f^\varepsilon(x) \rightarrow f(x)$ para quase todo ponto $x \in \mathbb{R}^n$.

Observe que f^ε é também convexa, para todo $\varepsilon > 0$: dados $x, y \in \mathbb{R}^n$ e $\lambda \in (0, 1)$, temos

$$\begin{aligned} f^\varepsilon((1-\lambda)x + \lambda y) &= \int_{\mathbb{R}^n} f(z - (1-\lambda)x - \lambda y) \eta(z) dz \\ &= \int_{\mathbb{R}^n} f((1-\lambda)(z-x) + \lambda(z-y)) \eta(z) dz \\ &\leq (1-\lambda) \int_{\mathbb{R}^n} f(z-x) \eta(z) dz + \lambda \int_{\mathbb{R}^n} f(z-y) \eta(z) dz \\ &= (1-\lambda)f^\varepsilon(x) + \lambda f^\varepsilon(y). \end{aligned}$$

Logo, podemos aplicar (34) e (35) para f^ε :

$$\begin{aligned} \sup_{B_{r/2}(x)} |f^\varepsilon| &\leq C \int_{B_r(x)} |f^\varepsilon| \implies \text{ess sup}_{B_{r/2}(x)} |f| \leq C \int_{B_r(x)} |f|, \\ \text{ess sup}_{z \in B_{r/2}(x)} |\nabla f^\varepsilon(z)| &\leq \frac{C}{r} \int_{B_r(x)} |f^\varepsilon(y)| dy \leq M \text{ uniformemente em } 0 < \varepsilon \leq \varepsilon_0. \end{aligned} \quad (39)$$

Sendo f^ε suave, para quase todo $y, z \in B_{r/2}(x)$, temos

$$|f^\varepsilon(z) - f^\varepsilon(y)| = \left| \int_0^1 \nabla f^\varepsilon(z + t(y-z)) dt \cdot (y-z) \right| \leq M|y-z|.$$

Assim, utilizando a convergência pontual (já que M é independente de ε):

$$|f(z) - f(y)| \leq M|y-z|.$$

Segue que f é localmente Lipschitz. Em particular, f é contínua e o Teorema de Rademacher implica que é diferenciável em quase todo ponto, de modo que é possível fazer $\varepsilon \rightarrow 0$ em (39) para obter (35). \square

Finalmente, apresentamos uma terceira prova de que f é localmente Lipschitz, mais elementar que a anterior (mas que não implica nas estimativas (34) e (35)).

Demonstração. Considere $x_0 \in \text{int dom } f$ e $Q_r(x_0) \subset \text{int dom } f$ um cubo fechado centrado em x_0 cujas arestas tem comprimento $2r$ e são paralelas com os eixos coordenados. Observe que $B_r(x_0) \subset Q_r(x_0)$. Sendo f convexa, o máximo de f em $Q_r(x_0)$ é atingido em um dos vértices x_k do cubo (exercício). Logo,

$$f(y) \leq \sup_k f(x_k) =: M < +\infty, \text{ para todo } y \in B_r(x_0).$$

Em seguida, vamos mostrar que f é também limitada por baixo em $B_r(x_0)$. De fato, dado $x \in B_r(x_0)$, temos que $y := 2x_0 - x \in B_r(x_0)$. Logo,

$$x_0 = \frac{x+y}{2} \implies f(x_0) \leq \frac{f(x) + f(y)}{2} \leq \frac{f(x) + M}{2} \implies f(x) \geq 2f(x_0) - M.$$

Logo, $|f| \leq C$ em $B_r(x_0)$, o que mostra que f é localmente limitada em $\text{int dom } f$.

Agora, fixamos $\delta \in (0, r)$. Dados $x, y \in B_{r-\delta}(x_0)$, temos que

$$y = \frac{\delta}{\delta + \|y - x\|} x + \frac{\|y - x\|}{\delta + \|y - x\|} z, \quad \text{onde } z = y + \delta \frac{y - x}{\|y - x\|} \in B_r(x_0).$$

Usando convexidade,

$$f(y) \leq \frac{\delta}{\delta + \|y - x\|} f(x) + \frac{\|y - x\|}{\delta + \|y - x\|} f(z).$$

Logo,

$$f(y) - f(x) \leq \frac{f(z) - f(x)}{\delta + \|y - x\|} \|y - x\| \leq \frac{2C}{\delta} \|y - x\|.$$

Trocando os papéis de $x, y \in B_{r-\delta}(x_0)$, concluímos que f é de Lipschitz em $B_{r-\delta}(x_0)$. \square

Em seguida, estudamos diferenciabilidade de segunda ordem para funções convexas. O Teorema de Alexandrov diz que funções convexas são duas vezes diferenciáveis para quase todo ponto, que é um resultado notável, já que diferenciabilidade de primeira ordem é apenas quase sempre.

Iniciamos mostrando que derivadas de segunda ordem de funções convexas são “medidas de Radon com sinal”. Integração por partes nos diz que μ^{ij} da proposição abaixo é, no sentido das distribuições, igual a $\partial_{ij}^2 f$.

Proposição 44. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Então existem medidas de Radon com sinal μ^{ij} tais que*

$$\int_{\mathbb{R}^n} f(x) \frac{\partial^2 \phi}{\partial x_i \partial x_j}(x) dx = \int_{\mathbb{R}^n} \phi(x) d\mu^{ij}(x),$$

para qualquer $\phi \in C_c^\infty(\mathbb{R}^n)$ e quaisquer $i, j \in \{1, 2, \dots, n\}$. Além disso, $\mu^{ij} = \mu^{ji}$ e μ^{ii} são positivas.

Deomnstração [10]. Assim como na segunda prova do Teorema 43, a molificação $f^\varepsilon = \eta_\varepsilon * f$ é suave e convexa. Logo, pela Proposição 42, temos que, para qualquer $x \in \mathbb{R}^n$ e qualquer $\xi \in \mathbb{R}^n$,

$$\langle \nabla f^\varepsilon(x) \xi, \xi \rangle = \sum_{i,j} \frac{\partial^2 f^\varepsilon}{\partial x_i \partial x_j} \xi_i \xi_j \geq 0.$$

Integração por partes implica que

$$\int_{\mathbb{R}^n} f^\varepsilon(x) \frac{\partial^2 \phi}{\partial x_i \partial x_j}(x) \xi_i \xi_j dx = \int_{\mathbb{R}^n} \phi(x) \frac{\partial^2 f^\varepsilon}{\partial x_i \partial x_j}(x) \xi_i \xi_j dx \geq 0, \quad \forall \phi \in C_c^\infty(\mathbb{R}^n), \phi \geq 0.$$

Fazendo $\varepsilon \rightarrow 0$, obtemos

$$L(\phi) := \int_{\mathbb{R}^n} f(x) \frac{\partial^2 \phi}{\partial x_i \partial x_j}(x) \xi_i \xi_j dx \geq 0, \quad \forall \phi \in C_c^\infty(\mathbb{R}^n), \phi \geq 0. \quad (40)$$

O lado esquerdo de (40) define um funcional linear $L : C_c^\infty(\mathbb{R}^n) \rightarrow \mathbb{R}$ que é positivo (isto é, vale (40)). Logo, pelo Teorema da Representação de Riesz (ver [10, Seção 1.8]), existe uma medida de Radon μ^ξ tal que

$$L(\phi) = \int_{\mathbb{R}^n} f(x) \frac{\partial^2 \phi}{\partial x_i \partial x_j}(x) \xi_i \xi_j dx \geq 0, \quad \forall \phi \in C_c^\infty(\mathbb{R}^n).$$

Definimos $\mu^{ii} = \mu^{e_i}$, isto é, considerando os vetores da base canônica de \mathbb{R}^n . Para $i \neq j$, consideramos

$$\xi = \frac{e_i + e_j}{\sqrt{2}},$$

que tem entradas $1/\sqrt{2}$ nas posições i e j e demais entradas nulas. Observe que

$$\sum_{k,l} \frac{\partial^2 \phi}{\partial x_k \partial x_l} \xi_k \xi_l = \frac{1}{2} \frac{\partial^2 \phi}{\partial x_i \partial x_i} + \frac{\partial^2 \phi}{\partial x_i \partial x_j} + \frac{1}{2} \frac{\partial^2 \phi}{\partial x_j \partial x_j} \implies \frac{\partial^2 \phi}{\partial x_i \partial x_j} = \sum_{k,l} \frac{\partial^2 \phi}{\partial x_k \partial x_l} \xi_k \xi_l - \frac{1}{2} \frac{\partial^2 \phi}{\partial x_i \partial x_i} - \frac{1}{2} \frac{\partial^2 \phi}{\partial x_j \partial x_j}.$$

Portanto,

$$\begin{aligned} \int_{\mathbb{R}^n} f \frac{\partial^2 \phi}{\partial x_i \partial x_j} dx &= \int_{\mathbb{R}^n} f \sum_{k,l} \frac{\partial^2 \phi}{\partial x_k \partial x_l} \xi_k \xi_l dx - \frac{1}{2} \int_{\mathbb{R}^n} f \frac{\partial^2 \phi}{\partial x_i \partial x_i} dx - \frac{1}{2} \int_{\mathbb{R}^n} f \frac{\partial^2 \phi}{\partial x_j \partial x_j} dx \\ &= \int_{\mathbb{R}^n} \phi d\mu^\xi - \frac{1}{2} \int_{\mathbb{R}^n} \phi d\mu^{ii} - \frac{1}{2} \int_{\mathbb{R}^n} \phi d\mu^{jj} \\ &= \int_{\mathbb{R}^n} \phi d\left(\mu^\xi - \frac{1}{2}\mu^{ii} - \frac{1}{2}\mu^{jj}\right) =: \int_{\mathbb{R}^n} \phi d\mu^{ij}. \quad \square \end{aligned}$$

Pela decomposição de Lebesgue, podemos escrever μ^{ij} como uma soma de uma medida absolutamente contínua e uma medida singular:

$$\mu^{ij} = \mu_{ac}^{ij} + \mu_s^{ij},$$

onde

$$\mu_{ac}^{ij} \ll \mathcal{L}^n \quad \text{e} \quad \mu_s^{ij} \perp \mathcal{L}^n.$$

Pelo Teorema de Radon-Nikodym,

$$d\mu_{ac}^{ij}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x) dx,$$

onde $\frac{\partial^2 f}{\partial x_i \partial x_j}$ não é uma derivada “de verdade”, mas apenas uma notação para a densidade da medida absolutamente contínua μ_{ac}^{ij} . Denotamos por $\nabla^2 f(x)$ a matriz

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right].$$

O Teorema de Alexandrov afirma que, para quase todo ponto, esta “Hessiana” cumpre seu papel.

Teorema de Alexandrov. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Então, para quase todo $x \in \mathbb{R}^n$,*

$$f(x+v) = f(x) + \langle \nabla f(x), v \rangle + \frac{\langle \nabla^2 f(x)v, v \rangle}{2} + o(|v|^2) \quad \text{quando } v \rightarrow 0.$$

Demonstração. **Evans Gariepy ou Villani** □

13.3.2 Subdiferenciabilidade

Em análise não suave, o que substitui o conceito de diferenciabilidade é o conceito de subdiferenciabilidade (ou superdiferenciabilidade). Dada uma função convexa $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $x \in \mathbb{R}^n$, o **subdiferencial** $\partial f(x)$ no ponto x é o conjunto dos pontos $p \in \mathbb{R}^n$ tais que

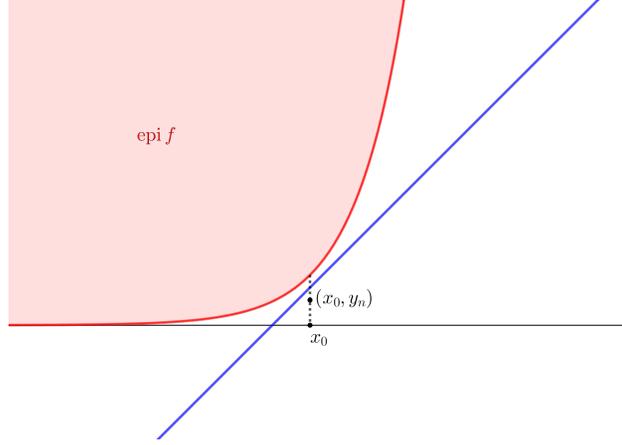
$$f(y) \geq f(x) + \langle p, y - x \rangle \quad \forall y \in \mathbb{R}^n.$$

Esta definição deve ser comparada com as Proposições 40 e 41. Um ponto $p \in \partial f(x)$ representa geometricamente um hiperplano suporte em $f(x)$: o gráfico de f está todo acima deste hiperplano.

Desenho em dimensão um

Proposição 45. *Se $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa, então $\partial f(x_0) \neq \emptyset$ para todo $x_0 \in \text{int}(\text{dom } f)$. Além disso, f é diferenciável em x_0 se, e somente se, $\partial f(x_0)$ consiste de um ponto só. Neste último caso, deve ser $\partial f(x_0) = \{\nabla f(x_0)\}$.*

Demonstração. Considere $A = \text{epi } f = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}; y \geq f(x)\}$, o epigráfico de f , que é um conjunto convexo. Pelo Teorema 43, f é contínua e portanto $\text{epi } f$ é também fechado. Considere $y_k = f(x_0) - 1/k$ e $B = \{(x_0, y_k)\}$, que é convexo e compacto. Por construção, $A \cap B = \emptyset$. Logo, o Teorema de Hahn-Banach³² (ver, por exemplo, [6, Theorem 1.7]) implica que existe um hiperplano que separa estritamente estes dois conjuntos e é possível obter (41) abaixo.



Mais analiticamente, existe um funcional linear $\Phi_k : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}$ e $\alpha_k \in \mathbb{R}$ tais que $\Phi_k > \alpha_k$ em A e $\Phi_k < \alpha_k$ em B . Observe que a aplicação $\varphi_k : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por

$$\varphi_k(x) = \Phi_k(x, 0)$$

é linear. Logo, pelo Teorema de Representação de Riesz (ver, por exemplo, [7, Teorema 8.22]), existe $p_k \in \mathbb{R}^n$ tal que

$$\Phi_k(x, 0) = \langle p_k, x \rangle \quad \forall x \in \mathbb{R}^n.$$

Além disso, por linearidade, podemos decompor

$$\Phi_k(x, y) = \Phi_k(x, 0) + \Phi_k(0, y) = \varphi_k(x) + y \Phi_k(0, 1) = \langle p_k, x \rangle + \beta_k y,$$

onde $\beta_k := \Phi_k(0, 1)$.

Em particular, escolhendo pontos $(x, f(x)) \in A$, temos

$$\langle p_k, x \rangle + \beta_k f(x) > \alpha_k > \langle p_k, x_0 \rangle + \beta_k y_k.$$

Substituindo $x = x_0$ acima, concluímos que $\beta_k > 0$. Logo, definindo $q_k = p_k/\beta_k$ e lembrando a definição de y_k , temos que,

$$f(x) > f(x_0) + \langle q_k, x_0 - x \rangle - \frac{1}{k} \quad \forall x \in \mathbb{R}^n. \quad (41)$$

Escolhendo convenientemente $x = x_0 - q_k/\|q_k\|$ (de modo que $x_0 - x = q_k/\|q_k\|$), obtemos que

$$\|q_k\| = \langle q_k, x_0 - x \rangle < f\left(x_0 - \frac{q_k}{\|q_k\|}\right) - f(x) + \frac{1}{k} \leq M,$$

já que f é localmente limitada. Segue que, a menos de tomar uma subsequência, $q_k \rightarrow q \in \mathbb{R}^n$. Passando ao limite em (41), obtemos

$$f(x) \geq f(x_0) + \langle q, x_0 - x \rangle \quad \forall x \in \mathbb{R}^n,$$

³²Em \mathbb{R}^n , não é necessário utilizar o Teorema de Hahn-Banach em toda sua generalidade. É um exercício mostrar diretamente que, em dimensão finita, dados A e B fechados, sendo um deles compacto, existe $w \in \mathbb{R}^n$ e $a > 0$ tais que

$$\langle w, x - y \rangle \geq a \quad \forall x \in A, y \in B.$$

Dica: Definir $W := A - B = \{a - b; a \in A, b \in B\}$ e minimizar $\|x\|$ em W (que é fechado, convexo e $0 \notin W$).

de modo que $q \in \partial f(x_0)$.

Suponhamos que f é diferenciável em x_0 . Então, pela Proposição 40, $\nabla f(x_0) \in \partial f(x_0)$. Por outro lado, dado $p \in \partial f(x_0)$, temos

$$f(x_0) + p \cdot (x - x_0) \leq f(x) = f(x_0) + \nabla f(x_0) \cdot (x - x_0) + o(\|x - x_0\|) \quad \forall x \in \mathbb{R}^n.$$

Ou ainda,

$$\langle p - \nabla f(x_0), x - x_0 \rangle \leq o(\|x - x_0\|) \quad \forall x \in \mathbb{R}^n.$$

Logo, só pode ser $p = \nabla f(x_0)$.

falta recíproca. □

13.3.3 Transformada de Legendre

Dada uma função convexa própria $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, definimos a sua **função convexa conjugada** $f^* : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ como

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

Mostramos que f^* é de fato convexa³³:

$$\begin{aligned} f^*((1-\lambda)y + \lambda z) &\leq \sup_{x \in \mathbb{R}^n} (\langle x, (1-\lambda)y + \lambda z \rangle - f(x)) \\ &= \sup_{x \in \mathbb{R}^n} ((1-\lambda)\langle x, y \rangle + \lambda\langle x, z \rangle - f(x)) \\ &\leq (1-\lambda) \sup_{x \in \mathbb{R}^n} (\langle x, y \rangle - f(x)) + \lambda \sup_{x \in \mathbb{R}^n} (\langle x, z \rangle - f(x)) \\ &= (1-\lambda)f^*(y) + \lambda f^*(z). \end{aligned}$$

Além disso, f^* é semicontínua inferiormente, caracterizando possíveis descontinuidades em $\partial(\text{dom } f^*)$. Temos também a **desigualdade de Young**³⁴:

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

Proposição 46. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa própria e semicontínua inferiormente. Então, para quaisquer $x, y \in \mathbb{R}^n$, as seguintes afirmações são equivalentes:*

- (i) $\langle x, y \rangle = f(x) + f^*(y)$.
- (ii) $y \in \partial f(x)$.
- (iii) $x \in \partial f^*(y)$.

Demonstração. Vamos mostrar que (i) \iff (ii). A equivalência (i) \iff (iii) segue deste caso, pois vamos mostrar em seguida que, nas hipóteses da nossa proposição, $f^{**} = f$.

Uma das desigualdades é sempre satisfeita. Logo, temos

$$\begin{aligned} \langle x, y \rangle = f(x) + f^*(y) &\iff \langle x, y \rangle \geq f(x) + f^*(y) \iff \forall z, \langle x, y \rangle \geq f(x) + \langle z, y \rangle - f(z) \\ &\iff \forall z, f(z) \geq f(x) + \langle z - x, y \rangle \iff y \in \partial f(x). \end{aligned} \quad \square$$

Dadas duas funções convexas próprias $f, g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, definimos a **inf-convolução** (convolução infimal) de f e g como

$$(f \square g)(x) = \inf_{y \in \mathbb{R}^n} (f(x-y) + g(y)) = \inf_{y \in \mathbb{R}^n} (f(y) + g(x-y)).$$

É fácil de ver que

$$(f \square g)^* = f^* + g^*.$$

³³De fato, nosso argumento mostra que o supremo de uma família qualquer de funções convexas é ainda convexa.

³⁴Nome da desigualdade motivado pelo caso particular, que é mais famoso da desigualdade:

$$f(x) = \frac{|x|^p}{p} \quad \text{e} \quad \frac{1}{p} + \frac{1}{q} = 1 \implies f^*(y) = \frac{|y|^q}{q}.$$

Proposição 47. Para $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função própria, são equivalentes:

(i) f é convexa e semicontínua inferiormente.

(ii) $f = g^*$ para alguma função própria $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$

(iii) $f^{**} = f$.

Demonstração. Vamos mostrar que (i) \implies (iii) \implies (ii) \implies (i). A única que não é clara é a primeira dessas implicações. Suponhamos então que f é convexa e semicontínua inferiormente. Pela desigualdade de Fenchel (ou Young), tem-se que, fixado qualquer $x \in \mathbb{R}^n$,

$$\langle x, y \rangle \leq f(x) + f^*(y), \quad \forall y.$$

Segue que

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

Sendo f convexa e semicontínua inferiormente, o subdiferencial $\partial f(x)$ é não vazio. Tomando $p \in \partial f(x)$, temos, por definição,

$$f(z) \geq f(x) + \langle p, z - x \rangle \quad \text{ou, equivalentemente,} \quad f(x) \leq f(z) + \langle p, x - z \rangle.$$

Portanto,

$$\begin{aligned} f^{**}(x) &= \sup_{y \in \mathbb{R}^n} \{\langle x, y \rangle - f^*(y)\} \\ &= \sup_{y \in \mathbb{R}^n} \{\langle x, y \rangle - \sup_{z \in \mathbb{R}^n} \{\langle y, z \rangle - f(z)\}\} \\ &= \sup_{y \in \mathbb{R}^n} \inf_{z \in \mathbb{R}^n} \{\langle y, x - z \rangle + f(z)\} \\ &\geq \inf_{z \in \mathbb{R}^n} \{\langle p, x - z \rangle + f(z)\} \geq f(x). \end{aligned} \quad \square$$

Referências

- [1] L. Ambrosio and N. Gigli. *A user's guide to optimal transport*. Lecture Notes, 2010.
- [2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
- [3] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [4] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.*, 305(19):805–808, 1987.
- [5] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.
- [6] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.
- [7] Hamilton Prado Bueno. *Álgebra Linear - um segundo curso*, volume 06 of *Textos Universitários*. SBM, Rio de Janeiro, 2006.
- [8] C. R. de Oliveira. *Introdução à análise funcional*. Projeto Euclides. IMPA, Rio de Janeiro, 2012.
- [9] Ryszard Engelking. *General topology*, volume 6 of *Sigma Series in Pure Mathematics*. Heldermann Verlag, Berlin, second edition, 1989. Translated from the Polish by the author.
- [10] L. Evans and R. Gariepy. *Measure theory and fine properties of functions*. CRC Press, 1992.
- [11] Gerald B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, 1984. Modern techniques and their applications, A Wiley-Interscience Publication.
- [12] L. V. Kantorovich. On a problem of Monge. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 312(Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 11):15–16, 2004.
- [13] L. V. Kantorovich. On mass transportation. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 312(Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 11):11–14, 2004.
- [14] Elon Lages Lima. *Elementos de Topologia Geral*. Coleção Textos Universitários. Editora SBM, 2009.
- [15] Francesco Maggi. *Sets of finite perimeter and geometric variational problems*, volume 135 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2012. An introduction to geometric measure theory.
- [16] G. Monge. Mémoire sur la théorie des déblais et des remblais. In *Histoire de l'Académie Royale des Sciences de Paris*, pages 666–704, 1781.
- [17] Aldo Pratelli. On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation. *Ann. Inst. H. Poincaré Probab. Statist.*, 43(1):1–13, 2007.
- [18] R. T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific J. Math.*, 17:497–510, 1966.
- [19] R. Tyrrell Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.

- [20] Filippo Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.
- [21] Elias M. Stein and Rami Shakarchi. *Functional analysis*, volume 4 of *Princeton Lectures in Analysis*. Princeton University Press, Princeton, NJ, 2011. Introduction to further topics in analysis.
- [22] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
- [23] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

Índice

Desintegração de uma medida, 35

Distância de Wasserstein, 27

Dualidade de Kantorovich, 20

espaço de Wasserstein, 27

Fórmula de Kantorovich–Rubinstein, 21

Lema de colagem, 35

Problema de Kantorovich, 3

Problema de Monge, 1

Teorema de Desintegração de Medidas, 35