Capítulo 1

O problema de Monge-Kantorovitch

1.1 Introdução

Em 1781, o matemático Francês Gaspar Monge propos uma formulação matemática para a seguinte pergunta:

> *Dada uma distribuição de recursos e uma distribuição de demandas, como transportar tais recursos à demanda de forma a realizar o mínimo trabalho possível?*

A sua intuição original era que tais recursos seriam de fato materiais de construção distribuídos num terreno, enquanto que a demanda seria uma distribuição desejada dos recursos à disposição no mesmo terreno, por exemplo para a construção de uma casa.

De forma geral, podemos tomar um espaço ambiente bem mais abstrato onde se distribuem tais recursos e demandas. Sendo assim, vamos considerar que os espaços ambiente onde se distribuem tais recursos e demandas são dados por um espaços métricos $(\mathcal{X}, d_X)$ e $(\mathcal{Y}, d_{\mathcal{Y}})$. Faremos a hipótese mais à frente que estes são espaços Poloneses, ou seja metrizáveis, completos e separáveis (que admitem um subconjunto denso e enumerável). Sendo assim, a intuição de Monge é recuperada tomando $\mathcal{X}$ e $\mathcal{Y}$ como o espaço euclidiano $\mathbb{R}^d$, com a topologia usual. No que diz respeito às distribuições de recursos e demandas, na linguagem moderna da teoria da medida, estas são representadas por medidas de probabilidade $\mu \in \mathscr{P}(\mathcal{X})$ e $\nu \in \mathscr{P}(\mathcal{Y})$.

Segundo a intuição original de Monge, o trabalho de levar uma unidade de massa de um ponto x para um ponto y seria dado pela distância euclidiana $|x-y|$, mas ainda na filosofia de considerar um problema de transporte de massa abstrato, podemos imaginar que o custo unitário de transportar uma unidade de massa de $x \in \mathcal{X}$ para $y \in \mathcal{Y}$ seria dado por uma função

$$c : (x, y) \in \mathcal{X} \times \mathcal{Y} \mapsto c(x, y) \in \mathbb{R}.$$

Por outro lado, o transporte em si é realizado por um mapa mensurável cujo domínio é o espaço de recursos e imagem dada pelo espaço de demandas $T : \mathcal{X} \to \mathcal{Y}$, satisfazendo a restrição de realizar o transporte de μ à ν. Na linguagem moderna da teoria da medida isso quer dizer que a *medida de transporte de T sobre μ*, $T_\sharp\mu$, também conhecida na literatura em inglês como medida de *push-forward*, deve coincidir com ν. Em outras palavras, para todo conjunto Borel mensurável $A \subset \mathcal{Y}$ deve-se ter que

$$T_\sharp\mu(A) \stackrel{\text{def.}}{=} \mu(T^{-1}(A)) = \nu(A).$$

Juntando todos esses elementos, o problema proposto é portanto escrito numa linguagem moderna como

$$\inf_{T_\sharp \mu = \nu} \int_{\mathcal{X}} c\left(x, T(x)\right) \mathrm{d}\mu(x). \tag{M}$$

No caso $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ e $c(x,y) = |x-y|$, Monge provou que se um mapa T é ótimo, então não há cruzamentos entre os trajetos de transporte, o que em si já é uma propriedade interessante, mas ele não provou a existência de um mapa ótimo, do mesmo jeito que a comunidade de cálculo das variações da época não havia começado a se perguntar quando, e em quais classes de funções, problemas variacionais admitem um minimizante.

Por outro lado, mesmo a existência de um mapa T que realize o transporte entre duas medidas quaisquer μ e ν é um problema não trivial. De fato, imagine que μ admita um átomo num ponto x, ou seja $\mu(\{x\}) > 0$, para todo mapa mensurável T a medida de transporte $T_\sharp \mu$ conterá um átomo no ponto $T(x)$. Logo se ν não contiver átomos, não existe nenhum mapa que faça o transporte de μ à ν! Veremos que esse problema é facilmente contornável se escolhermos uma boa classe de medidas μ. Na verdade é suficiente que μ não contenha átomos para que exista um mapa fazendo o transporte de μ para qualquer medida ν, mas infelizmente esse não é a única dificuldade da formulação de Monge. A restrição de fixar $T_\sharp \mu = \nu$ é extremamente não linear o que a torna muito mais difícil de passar ao limite, sendo então o maior impedimento em aplicar o *método direto do cálculo das variações* que veremos mais à frente nesse capítulo.

1.2 Uma alternativa probabilística: a formulação de Kantorovitch

Como vimos anteriormente, um dos problemas centrais na formulação de Monge é a existência de um mapa que realize o transporte de uma medida para outra. Também discutimos que basta que a medida inicial não contenha átomos para que um tal mapa exista, então uma alternativa é fixar um espaço de probabilidade do começo $(\Omega, \mathcal{F}, \mathbb{P})$ tal que a medida de probabilidade $\mathbb{P}$ não contenha átomos, de forma que para qualquer par $(\mu, \nu) \in \mathscr{P}(\mathcal{X}) \times \mathscr{P}(\mathcal{Y})$ existam mapas $X : \Omega \to \mathcal{X}$ e $Y : \Omega \to \mathcal{Y}$ tais que

$$X_\sharp \mathbb{P} = \mu, \quad Y_\sharp \mathbb{P} = \nu.$$

Em outras palavras, para cada par de leis de probabilidade μ, ν, existe um par de variáveis aleatórias X, Y com estas respectivas leis, também escrevemos $X \sim \mu$ e $Y \sim \nu$. Na notação de probabilidade escrevemos: para todo conjunto boreliano $A \subseteq \mathcal{X}$

$$\mathbb{P}(\{X \in A\}) = \mu(A) = \mathbb{P}(X^{-1}(A)).$$

Dessa forma, podemos escrever uma formulação probabilística do problema de Monge (M), também conhecida como o problema de Kantorovitch

$$\min_{X \sim \mu,\ Y \sim \nu} \mathbb{E}[c(X,Y)].$$

Essa formulação é equivalente à considerarmos um problema de minimização sobre todas as medidas de probabilidade no espaço produto $\mathcal{X} \times \mathcal{Y}$ cujas marginais são μ e ν, os chamados *planos de transporte*. Em outras palavras consideramos o espaço

$$\Pi(\mu, \nu) \overset{\text{def.}}{=} \left\{ \gamma \in \mathscr{P}(\mathcal{X} \times \mathcal{Y}) : {(\pi_0)}_\sharp \gamma = \mu,\ {(\pi_1)}_\sharp \gamma = \nu \right\},$$

onde $\pi_0 = \pi_{\mathcal{X}} : (x, y) \mapsto x$ e $\pi_1 = \pi_{\mathcal{Y}} : (x, y) \mapsto y$. A condição $(\pi_0)_\sharp \gamma = \mu$ portanto é equivalente à

$$\mu(B) = \gamma(B \times \mathcal{Y}) \text{ para todo boreliano } B \subset \mathcal{X},$$

e dizemos que μ é uma marginal de γ. De forma análoga, a condição $(\pi_1)_\sharp \gamma = \nu$ diz que ν é a outra marginal de γ. Dessa forma, o problema de Kantorovitch pode ser escrito em termos de planos de transporte como

$$\min_{\gamma \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \mathrm{d}\gamma(x, y) \equiv \min_{X \sim \mu,\ Y \sim \nu} \mathbb{E}[c(X, Y)]. \tag{K}$$

Dessa formulação, fazemos as seguintes observações:

- A formulação (K) não tem mais o problema de inexistência de um competidor admissivel, pois a medida produto $\mu \otimes \nu$ pertence sempre à $\Pi(\mu, \nu)$, logo o conjunto de planos de transporte admissíveis é sempre não vazio.

- Qualquer mapa de transporte T que realize o transporte de μ à ν induz um plano de transporte $\gamma_T \overset{\text{def.}}{=} (\mathrm{id}, T)_\sharp \mu \in \Pi(\mu, \nu)$. Dessa forma

  $$\min\,(\mathrm{K}) \leq \inf\,(\mathrm{M}).$$

  Portanto, se um plano de transporte ótimo é da forma γ_T para algum mapa T, então T é um mapa ótimo para o problema de Monge. De fato, esse é o caso, temos que

  $$\inf\,(\mathrm{M}) \leq \int_{\mathcal{X}} c(x, T(x)) \mathrm{d}\mu(x) = \min\,(\mathrm{K}) \leq \inf\,(\mathrm{M}).$$

  Veremos mais à frente que sob certas hipóteses sobre a medida μ e a função custo c, um plano de transporte ótimo é sempre da forma γ_T.

- Note que por linearidade da integral, o funcional $\gamma \mapsto \displaystyle\int_{\mathcal{X} \times \mathcal{Y}} c \mathrm{d}\gamma$ é linear; enquanto que o conjunto de planos de tranporte $\Pi(\mu, \nu)$ é convexo. A formulação (K) se torna então um problema de minimização linear sobre um conjunto convexo, o que facilita muito a análise de existência.

  Portanto, a estratégia natural para obter existência de um mapa de transporte ótimo para o problema de Monge é inicialmente se apoiar sobre os resultados mais facilmente obtidos para o problema de Kantorovitch e em seguida utilizar as condições de otimalidade para esse problema para mostrar que o plano de transporte ótimo é da forma γ_T para algum mapa T.

Por isso nosso foco será de nos concentrar inicialmente nos resultados fundamentais da topologia de espaços Poloneses e das medidas de probabilidade sobre tais espaços, em particular no Teorema de Weierstrass que garante existência de minimizantes de funções semi contínuas inferiormente em conjuntos compactos, e cuja prova é essencialmente o argumento conhecido como o método direto do calculo das variações. Isso culminará em um teorema de existência com hipóteses mínimas para o problema de Kantorovitch. Em seguida, vamos explorar a estrutura do problema de Kantorovitch como a minimização de um funcional linear num conjunto convexo, para isso vamos desenvolver ferramentas de análise convexa que nos dará muita informação quanto à estrutura de planos de transporte ótimo. Tudo isso, nos permitirá atacar o problema de Monge.

1.3 O método direto do cálculo das variações em espaços métricos

O objetivo dessa sessão não é de construir toda a teoria de espaços métricos, uma vez que já existem excelentes referências para o assunto, em particular em literatura matemática em português.

O método direto do cálculo das variações:

Geralmente, em cursos de análise na reta, é apresentado o Teorema de Weierstrass que afirma:

> *Seja $f : [a, b] \to \mathbb{R}$ uma função contínua, então existem pontos $x_\star, x^\star \in [a, b]$ atingindo o ínfimo e o supremo de f no intervalo $[a, b]$, ou seja,*
>
> $$f(x_\star) = \min_{x\in[a,b]} f(x), \quad f(x^\star) = \max_{x\in[a,b]} f(x).$$

Demonstração. De fato, como f é contínua num intervalo compacto, então f é limitada, logo o conjunto $f([a, b]) \subset \mathbb{R}$ é limitado. Portanto, o supremo e o ínfimo de $f([a, b])$ existem. Seja $m = \inf f([a, b])$, então existe uma sequência $(x_n)_{n\in\mathbb{N}} \subset [a, b]$ tal que $f(x_n) \to m$. Como $[a, b]$ é compacto, então existe uma subsequência $(x_{n_k})_{k\in\mathbb{N}}$ convergente para algum ponto $x_\star \in [a, b]$. Pela continuidade de f, temos que

$$f(x_\star) = f\left(\lim_{k\to+\infty} x_{n_k}\right) = \lim_{k\to+\infty} f(x_{n_k}) = m.$$

De forma análoga, existe um ponto $x^\star$ tal que $f(x^\star) = \max f([a, b])$. □

A prova é simples e elegante, no entanto, as propriedades fundamentais que garantem a mesma conclusão são a noção de *semi-continuidade* e *compacidade*, que se estendem sem grandes dificuldades para espaços métricos gerais.

Um espaço métrico é um conjunto $\mathcal{X}$ munido de uma função $d : \mathcal{X} \times \mathcal{X} \to [0, +\infty)$, chamada de métrica ou distância, que satisfaz as seguintes propriedades:

- $d(x, y) = 0$ se e somente se $x = y$;
- $d(x, y) = d(y, x)$ para todo $x, y \in \mathcal{X}$;
- $d(x, z) \leq d(x, y) + d(y, z)$ para todo $x, y, z \in \mathcal{X}$ (desigualdade triangular).

Estamos particularmente interessados em *espaços Poloneses*, isto é métricos completos e separáveis. Um espaço métrico é dito completo se toda sequência de Cauchy converge para um ponto do espaço.[1] Por outro lado, um espaço métrico é dito separável se existe um subconjunto denso enumerável.[2]

A semi-continuidade e compacidade são propriedades topológicas que estão normalmente em competição entre si; propriedades de continuidade são favorecidas por noções de convergência (topologias) mais fracas, mais fáceis de serem verificadas, enquanto que a compacidade em geral necessita mais estrutura para essa noção de convergência. Em seguida vamos relembrar os conceitos principais da topologia de espaços métricos.

[1] Lembramos que uma sequência $(x_n)_{n\in\mathbb{N}} \subset \mathcal{X}$ é dita de Cauchy se para todo $\varepsilon > 0$ existe $N \in \mathbb{N}$ tal que para todo $m, n \geq N$ temos $d(x_n, x_m) < \varepsilon$.

[2] Ou seja, existe um conjunto $D = (x_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}$ tal que para todo $x \in \mathcal{X}$ e todo $\varepsilon > 0$ existe $n \in \mathbb{N}$ tal que $d(x, x_n) < \varepsilon$.

Definição 1.3.1. Um função $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ é semi-contínua inferiormente (s.c.i.) se seus sub-níveis são fechados, ou seja para todo $\ell \in \mathbb{R}$ o conjunto

$$\{f \leq \ell\} = \{x \in \mathcal{X} : f(x) \leq \ell\}$$

é fechado.

Uma função $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ é sequencialmente semi-contínua inferiormente se para toda sequência $(x_n)_{n\in\mathbb{N}} \subset \mathcal{X}$ convergente para $x_\star \in \mathcal{X}$ temos

$$\begin{aligned} f(x_\star) \leq \liminf_{n\to+\infty} f(x_n) &\overset{\text{def.}}{=} \inf \left\{ L : f(x_{n_k}) \xrightarrow[k\to\infty]{} L, \right\} \\ &= \lim_{n\to+\infty} \left(\inf_{m\geq n} f(x_m) \right). \end{aligned}$$

A equivalência das noções de semi-continuidade é uma propridade geral da topologia de espaços métricos e vem da caracterização de conjunto fechado via sequências[3] O leitor é convidado à verificar essa equivalência e demais propriedades de funçoes semi-contínuas inferiormente.

Exercício 1.1. Seja $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ uma função s.c.i., o domínio de f é definido como $\operatorname{dom} f \overset{\text{def.}}{=} \{x \in \mathcal{X} : f(x) < +\infty\}$.

(a) Prove que, se $x \in \operatorname{dom} f$ então para todo $\varepsilon > 0$, existe uma vizinhança aberta V de x tal que

$$f(y) > f(x) - \varepsilon, \text{ para todo } y \in V.$$

(b) Se $f(x) = +\infty$, então para todo $M > 0$, existe uma vizinhança aberta V de x tal que

$$f(y) > M, \text{ para todo } y \in V.$$

(c) As propriedades dos itens anteriores não necessitam da topologia de espaços métricos. Mostre que semi-continuidade inferior e semi-continuidade inferior por sequências são equivalentes em um espaço métrico $\mathcal{X}$.

(d) O epígrafo de uma função f é definido como

$$\operatorname{epi}(f) \overset{\text{def.}}{=} \{(x, t) \in \mathcal{X} \times \mathbb{R} : f(x) \leq t\}.$$

Mostre que f é s.c.i. se, e somente se, $\operatorname{epi}(f)$ é subconjunto fechado de $\mathcal{X} \times \mathbb{R}$.

(e) Seja $(f_\alpha)_{\alpha\in\Lambda}$ uma família de funçoes s.c.i.; mostre que a função definida como

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

é semi-contínua inferiormente. *[Dica: Como escrever o epígrafo de f em função dos epígrafos de f_α?]*

---

[3] Lembramos que F subconjunto de um espaço métrico $\mathcal{X}$ é dito fechado se, e somente se, toda sequência $(x_n)_{n\in\mathbb{N}} \subset F$ que converge para um ponto x, tem-se que $x \in F$.

Definição 1.3.2. Um subconjunto $K \subset \mathcal{X}$ é compacto se para toda cobertura aberta de K, ou seja uma família $(A_i)_{i \in I}$ de abertos tais que $K \subset \bigcup_{i \in I} A_i$, existe uma subfamília finita $(A_{i_k})_{k=1}^N$ tal que $K \subset \bigcup_{k=1}^N A_{i_k}$.

Um conjunto $K \subset \mathcal{X}$ é sequencialmente compacto se toda sequência $(x_n)_{n \in \mathbb{N}} \subset K$ admite uma subsequência convergente para um ponto $x_\star \in K$.

Ums das razões pelas quais escolhemos trabalhar com espaços Poloneses é que nesses espaços as noções de compacidade e compacidade sequencial coincidem, assim como as noções de semi-continuidade e semi-continuidade sequencial.

Exercício 1.2. Mostre que um subconjunto K de um espaço métrico completo $\mathcal{X}$ é compacto se e somente se é sequencialmente compacto.

Outra caracterização importante de compacidade em espaços métricos é via conjuntos totalmente limitados.

Definição 1.3.3. Dizemos que um conjunto $K \subset \mathcal{X}$ é totalmente limitado se para todo $\varepsilon > 0$ existe um número finito de bolas de raio ε que cobrem K, ou seja existem pontos $(x_i)_{i=1}^N \subset \mathcal{X}$ tais que

$$K \subset \bigcup_{i=1}^N B(x_i, \varepsilon).$$

Exercício 1.3. Mostre que um subconjunto K de um espaço métrico completo $\mathcal{X}$ é compacto se e somente se é fechado e totalmente limitado.

Com essas definições em mãos, podemos enunciar e provar o Teorema de Weierstrass num espaço métrico geral, cuja prova é o argumento conhecido como o método direto do cálculo das variações.

Teorema 1.3.1 (Teorema de Weierstrass/Método direto do cálculo das variações)**.** *Seja uma função $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ semi-contínua inferiormente num espaço Polonês $\mathcal{X}$, e seja $K \subset \mathcal{X}$ um subconjunto compacto. Se $\inf_K f < +\infty$, então existe $x_\star \in K$ tal que*

$$f(x_\star) = \inf_K f = \min_K f.$$

Demonstração. Como $\inf_K f < +\infty$, pela definição de ínfimo, existe uma sequência minimizante, isto é $(x_n)_{n \in \mathbb{N}} \subset K$ tal que

$$f(x_n) \xrightarrow[n \to \infty]{} \inf_K f.$$

Podemos assumir sem perda de generalidade que $x_n \to x \in K$. De fato, como K é um subconjunto compacto de um espaço métrico completo, também é sequencialmente compacto, e portanto existe uma subsequência convergente, não reindexada. Esta subsequência também é uma sequência minimizante, logo obtemos da semi-continuidade inferior de f que

$$f(x_\star) \leq \liminf_{n \to \infty} f(x_n) = \inf_K f.$$

Como $x \in K$, o ínfimo de f em K é atingido em $x_\star$. $\square$

De modo geral, o problema de minimizar uma função f em todo o espaço $\mathcal{X}$ também é importante. Sem a hipótese de compacidade do espaço, necessitamos de hipóteses suplementares para a função f que comumente corresponde à uma hipótese de coercividade da função, o que será discutido no exercício seguinte.

Exercício 1.4. Dizemos que uma função $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ é coerciva se seus sub-conjuntos de nível são compactos, ou seja para todo $\ell \in \mathbb{R}$ o conjunto $\{f \leq \ell\}$ é compacto.

(a) Se $\mathcal{X} = \mathbb{R}^d$, então f é coerciva se, e somente se, $f(x) \to 0$ quando $x \to +\infty$.

(b) Dê um contra-exemplo de que a tese do item (a) não é verdade em geral.

(c) Seja uma função $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ coerciva e semi-contínua inferiormente num espaço Polonês $\mathcal{X}$. Se $\inf_{\mathcal{X}} f < +\infty$, então existe $x_\star \in \mathcal{X}$ tal que

$$f(x_\star) = \inf_{\mathcal{X}} f = \min_{\mathcal{X}} f.$$

Voltando ao problema que nos interessa, a formulação de Kantorovitch para o problema de transporte ótimo, para aplicar as ideais discutidas até agora, é precisamos estudar o espaço das medidas de Radon sob um espaço Polonês $\mathcal{X}$, em particular os critérios de semi-continuidade inferior e de compacidade desses objetos.

1.4 O espaço das medidas de Radon

Geralmente, num curso de teoria da medida ou probabilidade, fixamos um espaço de medida $(\Omega, \mathcal{F}, \mu)$ e o foco do estudo está nas propriedades de funçōes mensuráveis com respeito à σ-algebra $\mathcal{F}$. No entanto, diversas aplicações demandam uma maior flexibilidade, por exemplo no nosso problema de interesse, as variáveis em si são dadas por uma medida de probabilidade; no caso de processos estocáticos a lei das variáveis aleatórias em questão pode ser vista como uma curva numa classe de medidas de probabilidade.

Neste contexto, a alternativa é fixar uma σ-álgebra comum como "domínio" dessa classe de medidas. Para isso, dado um espaço Polonês $\mathcal{X}$, fixamos a σ-álgebra de Borel $\mathscr{B}$ gerada pelos abertos de $\mathcal{X}$. Chamamos tais medidas de Borelianas ou de Borel, no entanto na definição de medida de Radon que daremos em seguida, impomos mais condições de regularidade que serão importantes para obter boas propriedades de compacidade.

Definição 1.4.1 (Medida de Radon)**.** Seja $\mathcal{X}$ um espaço Polonês. Uma medida de Radon μ em $\mathcal{X}$ é uma medida de Borel finita em compactos, regular externa e interna, isto é:

- Para todo conjunto Borel $A \subset \mathcal{X}$, $\mu(A) = \inf\{\mu(U) : A \subset U, U \text{ aberto}\}$ (regularidade externa);
- Para todo conjunto Borel $A \subset \mathcal{X}$, $\mu(A) = \sup\{\mu(K) : K \subset A, K \text{ compacto}\}$ (regularidade interna);
- Para todo compacto $K \subset \mathcal{X}$, $\mu(K) < +\infty$.

Note que essa definição faz sentido para um espaço topológico geral[4], mas no caso de espaços Poloneses, toda medida Borel μ finita, ou seja $\mu(\mathcal{X}) < +\infty$ é regular, e portanto de Radon. Isso é uma consequência do *Teorema de Ulam*, cuja prova será feita passo à passo no seguinte exercício.

Exercício 1.5. O Teorema de Ulam afirma que se μ é uma medida de Borel, não negativa, num espaço Polonês $\mathcal{X}$, então para todo $\varepsilon > 0$, existe um compacto $K \subset \mathcal{X}$ tal que $\mu(\mathcal{X} \setminus K) < \varepsilon$.

Exercício 1.6. Use o Teorema de Ulam para provar que toda medida de Borel num espaço Polonês é regular interna e externamente.

Dada qualquer medida de Radon μ, podemos definir a integral de uma função Borel mensurável com respeito à μ da mesma forma que fazemos num curso tradicional de teoria da medida. O único problema é que o conjunto de medida nula onde f não está bem definida varia com respeito à cada medida μ. Para contornar isso, nos restringimos à funções contínuas ou semi-contínuas, já que essas propriedades de continuidade são finalmente propriedades intrísecas da topologia métrica de $\mathcal{X}$. Dessa forma, para toda função f semi-contínua a quantidade

$$\int_{\mathcal{X}} f(x)\mathrm{d}\mu(x)$$

está bem definida para toda medida de Radon μ.

Isso significa que cada medida positiva e finita μ determina um funcional linear contínuo sobre o espaço das funções contínuas e limitadas $\mathscr{C}_b(\mathcal{X})$

$$\mathscr{C}_b(\mathcal{X}) \ni f \mapsto \int_{\mathcal{X}} f(x)\mathrm{d}\mu(x) \leq \mu(\mathcal{X}) \left\|f\right\|_\infty,$$

ou sobre um espaço de funções apropriado. De modo geral, nos interessamos pelos seguintes espaços de funções

$$\begin{aligned}
\mathscr{C}_b(\mathcal{X}) &\overset{\text{def.}}{=} \{f \in \mathscr{C}(\mathcal{X}) : \left\|f\right\|_\infty < +\infty\} \text{ contínuas e limitadas} \\
\mathscr{C}_c(\mathcal{X}) &\overset{\text{def.}}{=} \{f \in \mathscr{C}(\mathcal{X}) : \operatorname{supp} f \text{ é compacto }\} \text{ contínuas à supporte compacto} \\
\mathscr{C}_0(\mathcal{X}) &\overset{\text{def.}}{=} \left\{f \in \mathscr{C}(\mathcal{X}) : \begin{matrix}\text{para todo } \varepsilon > 0 \text{ existe} \\ K \text{ compato t.q.} \\ |f(x)| < \varepsilon \text{ em } \mathcal{X} \setminus K\end{matrix}\right\} \text{ que convergem para 0 no finito.}
\end{aligned}$$

Exercício 1.7. Verifique que $\mathscr{C}_b(\mathcal{X}) \subseteq \mathscr{C}_c(\mathcal{X}) \subseteq \mathscr{C}_0(\mathcal{X})$, e temos igualdade se $\mathcal{X}$ é compacto.

Por outro lado, o teorema de representação de Riesz afirma que todo funcional linear, contínuo sobre $\mathscr{C}_0(\mathcal{X})$ pode ser representado por uma medida de Radon usando a operação de integral como produto de dualidade.

Teorema 1.4.1 (Teorema de Riesz)**.** *Seja $\mathcal{X}$ um espaço Polonês. Todo funcional linear contínuo $L : \mathscr{C}_0(\mathcal{X}) \to \mathbb{R}$ é da forma*

$$L(f) = \int_{\mathcal{X}} f(x)\mathrm{d}\mu(x)$$

para uma única medida de Radon, com sinal, finita μ em $\mathcal{X}$.

[4] As hipóteses mínimas para estudar tais objetos são de considerar $\mathcal{X}$ um espaço topológico Hausdorff.

Dessa forma, podemos definir o conjunto das medidas de Radon finitas como o dual do espaço de Banach $\mathscr{C}_0(\mathcal{X})$, *i.e.*

$$\mathscr{M}_b(\mathcal{X}) \stackrel{\text{def.}}{=} (\mathscr{C}_0(\mathcal{X}))^\star,$$

o que também induz a norma de variação total

$$\|\mu\|_{\mathscr{M}(\mathcal{X})} \stackrel{\text{def.}}{=} \sup\left\{\left|\int_{\mathcal{X}} f(x)\mathrm{d}\mu(x)\right| : f \in \mathscr{C}_0(\mathcal{X}),\ \|f\|_\infty \leq 1\right\},$$

assim como a convergência fraca estrela, ou fraca-$\star$, obtida através dessa dualidade. Dizemos que uma sequência $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{M}_b(\mathcal{X})$ converge na topologia fraca estrela para $\mu \in \mathscr{M}_b(\mathcal{X})$, denotado por $\mu_n \xrightarrow[n\to\infty]{\star} \mu$, se para todo $f \in \mathscr{C}_0(\mathcal{X})$,

$$\int_{\mathcal{X}} f\mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int_{\mathcal{X}} f\mathrm{d}\mu.$$

Ou seja, a topologia fraca estrela é a menor topologia para a qual a aplicação $\mu \mapsto \displaystyle\int_{\mathcal{X}} f\,d\mu$ é contínua para toda $f \in \mathscr{C}_0(\mathcal{X})$.

Como o dual de um espaço de Banach, temos diretamente acesso ao *teorema de Banach-Alaoglu-Bourbaki* que dá um princípio de compacidade para a bola unitária de $\mathscr{M}_b(\mathcal{X})$:

Teorema 1.4.2. *A bola unitário fechada*

$$B_{\mathscr{M}} \stackrel{\text{def.}}{=} \left\{\mu \in \mathscr{M}_b(\mathcal{X}) : \|\mu\|_{\mathscr{M}(\mathcal{X})} \leq 1\right\}$$

é compacta com respeito à convergência fraca-$\star$.

Estamos particularmente interessados no conjunto de medidas positivas e no conjunto de medidas de probabilidade

$$\mathscr{M}_+(\mathcal{X}) \stackrel{\text{def.}}{=} \{\mu \in \mathscr{M}_b(\mathcal{X}) : 0 \leq \mu\}, \quad \mathscr{P}(\mathcal{X}) \stackrel{\text{def.}}{=} \left\{\mu \in \mathscr{M}_+(\mathcal{X}) : \|\mu\|_{\mathscr{M}(\mathcal{X})} = 1\right\}.$$

Exercício 1.8. Seja $\mu \in \mathscr{M}_+(\mathcal{X})$, prove que $\|\mu\|_{\mathscr{M}(\mathcal{X})} = \mu(\mathcal{X})$.

Exercício 1.9. Mostre que a aplicação $\mu \mapsto \|\mu\|_{\mathscr{M}(\mathcal{X})}$ é semi-contínua inferiormente para a topologia fraca estrela, isto é, se $\mu_n \xrightarrow[n\to\infty]{\star} \mu$ então

$$\|\mu\|_{\mathscr{M}(\mathcal{X})} \leq \liminf_{n\to\infty} \|\mu_n\|_{\mathscr{M}(\mathcal{X})}.$$

[Dica: Use o exercício 1.1].

O exercício 1.9 mostra um fenômeno muito importante com respeito à convergência fraca estrela: uma sequência de medidas convergindo nessa topologia pode perder massa no infinito. O que significa que o espaço de medidas de probabilidade $\mathscr{P}(\mathcal{X})$ não é um subconjunto fechado de $\mathscr{M}_b(\mathcal{X})$ nessa topologia. Isso reduz muito a aplicabilidade direta do teorema 1.4.2 pois qualquer subsequência convergente de $\mathscr{P}(\mathcal{X})$ obtida com esse princípio de compacidade pode perder massa no infinito de forma que seu limite não seja mais uma medida de probabilidade! No entanto, à esse estágio do texto, ressaltamos que o leitor já tem todas as ferramentas necessárias para demonstrar a existência de um plano de transporte ótimo em um caso particular (mais ainda assim muito usado) do problema de Kantorovitch:

Exercício 1.10. Suponha que $\mathcal{X}$ e $\mathcal{Y}$ são espaços métricos compactos, por exemplo $\mathcal{X} = \mathcal{Y} = \Omega$ um fechado e limitado de $\mathbb{R}^d$, e c é um custo contínuo. Prove que o problema de Kantorovitch admite um plano de transporte ótimo.

Para contornar isso, introduzimos a noção de *convergência fraca* de medidas de probabilidade, também chamada de *convergência estreita* pois não é de fato uma convergência fraca da análise funcional, mas isso não importa muito.

Definição 1.4.2 (Convergência fraca de medidas)**.** Uma sequência $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}(\mathcal{X})$ converge fracamente para $\mu \in \mathscr{P}(\mathcal{X})$, denotado por $\mu_n \xrightarrow[n\to\infty]{} \mu$, se para toda função contínua e limitada $f \in \mathscr{C}_b(\mathcal{X})$,

$$\int_{\mathcal{X}} f \mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int_{\mathcal{X}} f \mathrm{d}\mu.$$

Para essa noção de convergência temos o direito de usar a função constante igual à 1 como função teste, logo se $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}(\mathcal{X})$ converge para μ na topologia estreita, então

$$1 = \mu_n(\mathcal{X}) = \int_{\mathcal{X}} 1 \mathrm{d}\mu_x \xrightarrow[n\to\infty]{} \int_{\mathcal{X}} 1 \mathrm{d}\mu = \mu(\mathcal{X}),$$

e portanto $\mu \in \mathscr{P}(\mathcal{X})$. Essa será portanto a topologia que em geral vamos tabalhar com o espaço de medidas de probabilidade. Note que pelo exercício 1.7, quando $\mathcal{X}$ é um espaço compacto as topologias estreita e fraca estrela coincidem e portanto temos compacidade diretamente do teorema de Banach-Alaoglu, valido para a tologia frace estrela em espaçosn de Banach em geral. A conveniência disso é que $\mathscr{P}(\mathcal{X})$ será automaticamente compacto quando $\mathcal{X}$ é compacto, se não for o caso teremos de desenvolver ferramentas para trabalhar melhor com o topologia estreita.

1.5 Semi-continuidade inferior e compacidade em $\mathscr{P}(\mathcal{X})$

Passamos agora ao estudo de critérios de semi-continuidade e compacidade com respeito respeito à topologia estreita. Note que como já discutimos que $\mathscr{C}_0(\mathcal{X}) \subseteq \mathscr{C}_b(\mathcal{X})$, obter tais resultados para a topologia estreita nos dará automaticamente os mesmos resultados para a topologia fraca estrela. Por definição, temos que se $f \in \mathscr{C}_b(\mathcal{X})$, então automaticamente temos que o funcional

$$\mathscr{P}(\mathcal{X}) \ni \mu \mapsto \int_{\mathcal{X}} f(x) \mathrm{d}\mu(x)$$

é contínuo para a topologia estreita.

Podemos generalizar esse resultado para o caso em que f é s.c.i. através do resultado seguinte

Lema 1.5.1. *Seja $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ uma função s.c.i. e limitada inferiormente, então existe uma sequência de funções $(f_k)_{k\in\mathbb{N}} \subset \mathscr{C}_b(\mathcal{X})$, definidas como*

$$f_k(x) \stackrel{\text{def.}}{=} \inf_{y\in\mathcal{X}} f(y) + k d_{\mathcal{X}}(x, y)$$

satisfazendo as seguintes propriedades

- *f_k é Lipschitz contínua com constante k;*

- *f_k forma uma sequência monótona, $f_k(x) \leq f_{k+1}(x) \xrightarrow[k\to\infty]{} f(x)$ para todo $x \in \mathcal{X}$.*

Demonstração. Defina a função

$$\bar{f}_k(x) \overset{\text{def.}}{=} \inf_{y\in\mathcal{X}} f(y) + kd_{\mathcal{X}}(x,y).$$

Vamos provar que $\bar{f}_k$ é Lipschitz contínua com constante k. Primeiramente, com uma aplicação simples do método direto, temos que para todo $x \in \mathcal{X}$, existe $y_x \in \mathcal{X}$ tal que $\bar{f}_k(x) = f(y_x) + kd_{\mathcal{X}}(x, y_x)$. Logo, para quaisquer $x, x' \in \mathcal{X}$, temos que

$$\bar{f}_k(x') \leq f(y_x) + kd_{\mathcal{X}}(x, y_x) \leq \bar{f}_k(x) + kd_{\mathcal{X}}(x, x').$$

Trocando o papel de x e x', obtemos a desigualdade reversa, o que implica que $\bar{f}_k$ é Lipschitz contínua com constante k.

Além disso, temos que $\bar{f}_k \leq \bar{f}_{k+1}$, pois para todo $x \in \mathcal{X}$

$$\bar{f}_k(x) = \inf_{y\in\mathcal{X}} f(y) + kd_{\mathcal{X}}(x,y) \leq \inf_{y\in\mathcal{X}} f(y) + (k+1)d_{\mathcal{X}}(x,y) = \bar{f}_{k+1}(x).$$

Desse modo, a sequência $\big(\bar{f}_k(x)\big)$ é monotoniamente crescente para todo $x \in \mathcal{X}$ e portanto converge. Pela definição de ínfimo, seja uma sequência $(x_k)_{k\in\mathbb{N}} \subset \mathcal{X}$ tal que

$$f(x_k) + kd_{\mathcal{X}}(x, x_k) \leq \bar{f}_k(x) + \frac{1}{k} \leq f(x) + \frac{1}{k}.$$

Logo necessariamente temos que $d_{\mathcal{X}}(x, x_k) \leq \frac{1}{k^2}$, o que implica que $\bar{f}_k(x) \xrightarrow[k\to\infty]{} f(x)$.

Para concluir a prova, podemos definir $f_k \overset{\text{def.}}{=} \min\{\bar{f}_k, k\}$, o que preserva as propriedades anteriores e garante que f_k é limitada. □

Usando esse resultado, nós recuperamos o resultado de semi-continuidade desejado. Com um pouco mais de trabalho podemos provar o teorema de Portmanteau, que dá várias caracterizações equivalentes da convergência estreita.

Teorema 1.5.1 (Portmanteau)**.** *Seja $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}(\mathcal{X})$. A seguintes afirmações são equivalentes:*

1. *$\mu_n \xrightarrow[n\to\infty]{} \mu$;*
2. *para toda função s.c.i. limitada inferiormente $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ temos que*

$$\int_{\mathcal{X}} f \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_{\mathcal{X}} f \mathrm{d}\mu_n.$$

3. *para toda função s.c.s. limitada superiormente $g : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ temos que*

$$\limsup_{n\to\infty} \int_{\mathcal{X}} g \mathrm{d}\mu_n \leq \int_{\mathcal{X}} g \mathrm{d}\mu.$$

4. *para todo conjunto aberto $A \subseteq \mathcal{X}$ vale que*

$$\limsup_{n\to\infty} \mu_n(A) \leq \mu(A).$$

5. *para todo conjunto fechado $F \subseteq \mathcal{X}$ vale que*

$$\mu(F) \leq \liminf_{n\to\infty} \mu_n(F).$$

6. *para todo conjunto mensurável $E \subseteq \mathcal{X}$ tal que $\mu(\partial E) = 0$ vale que*

$$\mu_n(E) \xrightarrow[n\to\infty]{} \mu_n(E).$$

Demonstração. Os items (2) e (3), e (4) e (5) são equivalentes entre si por meio da função complementar.

A implicação (1) $\implies$ (2) segue diretamente do Lema 1.5.1 e da definição de convergência estreita. Para ir de (2) $\implies$ (1), basta notar que se $f \in \mathscr{C}_b(\mathcal{X})$, então f é s.c.i. e s.c.s., logo aplicando (2) e (3) obtemos a igualdade necessária para a convergência estreita.

A implicação (2) $\implies$ (4) é obtida aplicando o item (2) à função $f = \chi_A$, que é s.c.i. para todo aberto A. Para a implicação inversa, podemos assumir que $f \in \mathscr{C}_b(\mathcal{X})$, novamente usando o Lema 1.5.1.

Para isso, seja $f \in \mathscr{C}_b(\mathcal{X})$, de modo que $f(\mathcal{X}) \subset [a, b]$. Seja uma partição do intervalo $[a, b]$ dada por

$$a = t_0 < t_1 < \cdots < t_m = b.$$

Como μ é uma medida de Radon finita, existe apenas um número finito de pontos t_i tais que $\mu(f^{-1}(\{t_i\})) > 0$. Portanto, podemos escolher a partição de modo que $\mu(f^{-1}(\{t_i\})) = 0$ para todo $i = 1, \ldots, m-1$ e $t_{i+1} - t_i < 1/N$ para todo i.

Defina os conjuntos

$$A_i \overset{\text{def.}}{=} \{x \in \mathcal{X} : t_i < f(x) \leq t_{i+1}\}, \quad i = 0, \ldots, m-1.$$

Podemos então definir

$$f_N(x) \overset{\text{def.}}{=} \sum_{i=0}^{m-1} a_i 1_{A_i}(x), \text{ com } a_i \overset{\text{def.}}{=} \inf_{A_i} f,$$

de modo que

$$f_N(x) \leq f(x),$$

e $f_N(x) \xrightarrow[N\to\infty]{} f(x)$ para todo $x \in \mathcal{X}$.

Note que cada conjunto A_i é mensurável e satisfaz $\mu(\partial A_i) = 0$ por construção. Portanto, aplicando (4) a cada conjunto A_i, temos que

$$\mu(A_i) = \mu(\operatorname{int} A_i) \leq \liminf_{n\to\infty} \mu_n(\operatorname{int} A_i) \leq \liminf_{n\to\infty} \mu_n(A_i).$$

Além disso, pelo lema de Fatou temos que

$$\begin{aligned}\int_{\mathcal{X}} f_N \mathrm{d}\mu = \sum_{i=0}^{m-1} a_i \mu(A_i) &\leq \sum_{i=0}^{m-1} a_i \liminf_{n\to\infty} \mu_n(A_i) \\ &\leq \liminf_{n\to\infty} \sum_{i=0}^{m-1} a_i \mu_n(A_i) = \liminf_{n\to\infty} \int_{\mathcal{X}} f_N \mathrm{d}\mu_n \leq \liminf_{n\to\infty} \int_{\mathcal{X}} f \mathrm{d}\mu_n.\end{aligned}$$

Finalmente, fazendo $N \to \infty$ e usando o Teorema da Convergência Monótona, concluímos que

$$\int_{\mathcal{X}} f \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_{\mathcal{X}} f \mathrm{d}\mu_n,$$

o que mostra (1). □

O segundo ingrediente que precisamos para estudar a existência de problemas variacionais em $\mathscr{P}(\mathcal{X})$ é a compacidade. Como discutimos anteriormente, o que impede que o teorema de Banach-Alaoglu-Bourbaki seja um critério de compacidade é o fenômeno de perde de massa no infinito quando passamos ao limite.

Teorema 1.5.2 (Prokhorov)**.** *Se $\mathcal{X}$ é um espaço métrico separável e completo, então um subconjunto $\mathcal{K} \subset \mathscr{P}(\mathcal{X})$ é relativamente (sequencialmente) compacto para a topologia fraca se, e somente se, é tight, isto é: para todo $\varepsilon > 0$, existe um compacto $K_\varepsilon \subset \mathcal{X}$ tal que*

$$\mu(K_\varepsilon) \geq 1 - \varepsilon, \quad \forall \mu \in \mathcal{K}.$$

Demonstração. Primeiro mostramos que se o conjunto $\mathcal{K}$ é tight, então ele é compacto. Nesse caso, existe uma sequência de compactos K_k de $\mathcal{X}$ tal que

$$\omega_k \overset{\text{def.}}{=} \sup_{\mu\in\mathcal{K}} \mu(\mathcal{X} \setminus K_k) \leq \frac{1}{k} \xrightarrow[k\to\infty]{} 0. \tag{1.1}$$

Em particular, podemos escolher $(K_k)_{k\in\mathbb{N}}$ como uma sequência monótona, ou seja $K_k \subset K_{k+1}$ para todo k. De fato, dada uma tal sequência, podemos definir uma segunda por indução como

$$K_1' = K_1 \text{ e } K_{k+1}' = K_k' \cup K_{k+1}.$$

Como a união finita de compactos é compacta, a nova sequência é formada por conjuntos compactos encaixados ainda satisfazendo a condição (1.1).

Considere agora uma sequência $(\mu_n)_{n\in\mathbb{N}}$, vamos construir um ponto de acumulação para essa sequência. Dado $k \in \mathbb{N}$, defina

$$\mu_{n,k} \overset{\text{def.}}{=} \mu_n \llcorner K_k.$$

Logo fixado $k \in \mathbb{N}$, temos que $(\mu_{n,k})_{n\in\mathbb{N}}$ está contido na bola unitária de $\mathscr{M}(K_k)$. Pelo teorema de Banach-Alaoglu-Bourbaki, existe uma subsequência $\sigma : \mathbb{N} \to \mathbb{N}$ tal que

$$\mu_{\sigma(n),k} \xrightarrow[n\to\infty]{\star} \mu^k \text{ para todo } k \in \mathbb{N},$$

onde, como cada K_k é compacto não há perda de massa no infinito, $1 - \omega_k \leq \mu^k(\mathcal{X}) \leq 1$. A nova sequência $\left(\mu^k\right)_{k\in\mathbb{N}}$ é monótona, ou seja $\mu^k \leq \mu^{k+1}$ para todo k, no sentido de medidas. Em outras palavras, para todo $0 \leq \phi \in \mathscr{C}_0(\mathcal{X})$ seque que

$$\int_{\mathcal{X}} \phi \mathrm{d}\mu^k \leq \int_{\mathcal{X}} \phi \mathrm{d}\mu^{k+1}.$$

Isso é uma consequência direta da inclusão $K_k \subset K_{k+1}$ e da monotonia da integral, o que implica que $\mu_{n,k} \leq \mu_{n,k+1}$ para n fixo.

Dessa forma, para todo $0 \leq \phi \in C_b(\mathcal{X})$, temos que a sequência de números reais

$$\int_{\mathcal{X}} \phi \mathrm{d}\mu^k$$

é monótona crescente e limitada superiormente por $\|\phi\|_\infty$. Podemos então definir um funcional linear limitado sobre $\mathscr{C}_b(\mathcal{X})$, e em particular sobre $\mathscr{C}_0(\mathcal{X})$, dado por

$$L\phi \overset{\text{def.}}{=} \lim_{k\to\infty} \int_{\mathcal{X}} \phi \mathrm{d}\mu^k.$$

Pelo teorema de representação de Riesz, existe uma medida de Radon $0 \leq \mu$ tal que

$$Lf = \int_{\mathcal{X}} f \mathrm{d}\mu, \text{ para toda } f \in \mathscr{C}_0(\mathcal{X}).$$

Por outro lado, pela definição de L, segue que μ^k converge para μ também na topologia estreita e portanto podemos usar a constante 1 como função teste, de formas que

$$1 = \lim_{k\to+\infty} 1 - \omega_k \leq \lim_{k\to+\infty} \mu^k(\mathcal{X}) = \mu(\mathcal{X}) \leq 1.$$

Logo μ é uma medida de probabilidade como queríamos.

Para provar a volta, seja $\mathcal{K}$ um conjunto (sequencialmente) compacto na topologia estreita. Como $\mathcal{X}$ é separável, existe um subconjunto denso e enumerável $(x_i)_{i\in\mathbb{N}}$. Vamos provar que para todo $j \in \mathbb{N}$, existe k_j tal que

$$\mu\left(\mathcal{X} \setminus \bigcup_{i=1}^{k_j} B(x_i, 1/j)\right) < 2^{-j}\varepsilon \text{ para toda } \mu \in \mathcal{K}. \tag{1.2}$$

Uma vez que isso esteja provado o resultado segue, pois definindo

$$K \overset{\text{def.}}{=} \bigcap_{j=1}^{\infty} \bigcup_{i=1}^{k_j} \overline{B}(x_i, 1/j),$$

note que K é compacto, pois é completo e totalmente limitado. Além disso

$$\mu(\mathcal{X} \setminus K) \leq \sum_{j=1}^{\infty} \mu\left(\mathcal{X} \setminus \bigcup_{j=1}^{k_j} B(x_i, 1/j)\right) < \sum_{j=1}^{\infty} 2^{-j}\varepsilon = \varepsilon, \text{ para todo } \mu \in \mathcal{K}.$$

Vamos então provar (1.2) por contradição. Suponha que exista j_0 e uma sequência $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{K}$ tal que

$$\mu_n\left(\mathcal{X} \setminus \bigcup_{i=1}^{n} B(x_i, 1/j_0)\right) \geq 2^{-j_0}\varepsilon.$$

Em particular, fixado um $k \in \mathbb{N}$, se $n > k$, então

$$\mu_n\left(\mathcal{X} \setminus \bigcup_{i=1}^{k} B(x_i, 1/j_0)\right) \geq \mu_n\left(\mathcal{X} \setminus \bigcup_{i=1}^{n} B(x_i, 1/j_0)\right) \geq 2^{-j_0}\varepsilon.$$

Por compacidade, existe uma subsequência de $(\mu_n)_{n\in\mathbb{N}}$ que converge na topologia estreita para μ. Segue então do teorema de Portmanteau 1.5.1 que

$$\mu\left(\mathcal{X}\setminus\bigcup_{i=1}^{k}B(x_i,1/j_0)\right)\geq\limsup_{n\to\infty}\mu_n\left(\mathcal{X}\setminus\bigcup_{i=1}^{k}B(x_i,1/j_0)\right)\geq 2^{-j_0}\varepsilon,$$

uma vez que para todo $k\in\mathbb{N}$ o conjunto $\mathcal{X}\setminus\bigcup_{i=1}^{k}B(x_i,1/j_0)$ é fechado. Como esta sequência de conjuntos é encaixados, obtemos da densidade de $(x_i)_{i\in\mathbb{N}}$ que

$$\bigcap_{k\in\mathbb{N}}\mathcal{X}\setminus\bigcup_{i=1}^{k}B(x_i,1/j_0)=\emptyset.$$

Finalmente, obtemos uma contradição fazendo $k\to\infty$ já que

$$\mu\left(\mathcal{X}\setminus\bigcup_{i=1}^{k_i}B(x_i,1/j_0)\right)\xrightarrow[k_i\to\infty]{}0$$

para toda medida de probabilidade. □

Observamos que no teorema à cima, o conceito de compacidade utilizado é o de compacidade sequencial. Isso não implica em perda alguma de generalidade pois a convergência estreita é metrizável. Uma das possíveis métricas sendo inclusive a distância de Wasserstein, definida através da teoria de transporte ótimo como veremos mais à frente.

Note também que quando o conjunto $\mathcal{K}$ é unitário, e portanto obviamente compacto pelo critério de compacidade por sequências, *i.e.* $\mathcal{K}=\{\mu\}$, a condição de ser tight é justamente a conclusão do teorema de Ulam dado no exercício 1.5.

Terminamos essa seção com uma condições suficiente para que uma família de medidas em $\mathscr{P}(\mathbb{R}^d)$ seja tight através de uma majoração uniforme de seus momentos. O momento de ordem p, ou p-momento de uma medida $\mu\in\mathscr{P}(\mathbb{R}^d)$ é definido como

$$M_p(\mu)\overset{\text{def.}}{=}\int_{\mathbb{R}^d}|x|^d\mathrm{d}\mu(x). \tag{1.3}$$

Note que, se X é uma variável aleatória com lei μ, temos que $M_p(\mu)=\mathbb{E}\left[|X|^p\right]$.

Proposição 1.5.1. *Seja $\mathcal{K}\subset\mathscr{P}(\mathbb{R}^d)$ tal que $C\overset{\text{def.}}{=}\sup_{\mu\in\mathcal{K}}M_p(\mu)<+\infty$. Então $\mathcal{K}$ é tight.*

Demonstração. Pelo teorema de Heine-Borel, a bola fechada de raio R e centrada em 0, $\overline{B}_R$, é compacta em $\mathbb{R}^d$. Além disso, para todo $\mu\in\mathcal{K}$ temos que

$$R^p\mu(\mathbb{R}^d\setminus\overline{B}_R)\leq\int_{\mathbb{R}^d\setminus\overline{B}_R}|x|^p\mathrm{d}\mu(x)\leq M_p(\mu)\leq C.$$

Logo, dado um $\varepsilon>0$, basta tomar $K_\varepsilon=\overline{B}_R$ tal que $C/R^p<\varepsilon$. □

1.6 Existência de planos de transporte ótimo

Usando os resultados anteriores o teorema de existência de um plano de transporte ótimo se torna uma aplicação direta do método direto do cálculo das variações. Primeiramente demonstramos que conjunto to planos de transporte à marginais fixas é tight, e portanto compacto pelo teorema de Prokhorov.

Lema 1.6.1. *Dados $\mu \in \mathscr{P}(\mathcal{X})$ e $\nu \in \mathscr{P}(\mathcal{Y})$, o conjunto dos planos de transporte $\Pi(\mu,\nu)$ com marginais μ e ν é compacto na topologia estreita de $\mathscr{P}(\mathcal{X}\times\mathcal{Y})$.*

Demonstração. Pela compacidade dos conjuntos unitários $\{\mu\}$ e $\{\nu\}$ em $\mathscr{P}(\mathcal{X}), \mathscr{P}(\mathcal{Y})$, temos pela recíproca do teorema de Prokhorov que para todo $\varepsilon > 0$ existem compactos $K \subseteq \mathcal{X}$ e $L \subseteq$ tais que

$$\mu(\mathcal{X}\setminus K) < \varepsilon/2 \text{ e } \nu(\mathcal{X}\setminus L) < \varepsilon/2.$$

Dessa forma, considere o compacto $K \times L$ de $\mathcal{X}\times\mathcal{Y}$, de forma que

$$(\mathcal{X}\times\mathcal{Y}) \setminus (K \times L) \subseteq (\mathcal{X}\times(\mathcal{Y}\setminus L)) \cup ((\mathcal{X}\setminus K)\times\mathcal{Y})$$

e portanto para $\gamma \in \Pi(\mu,\nu)$ temos que

$$\gamma\left((\mathcal{X}\times\mathcal{Y}) \setminus (K\times L)\right) \leq \gamma\left(\mathcal{X}\times(\mathcal{Y}\setminus L)\right) + \gamma\left((\mathcal{X}\setminus K)\times\mathcal{Y}\right) = \mu(\mathcal{X}\setminus K) + \nu(\mathcal{Y}\setminus L) < \varepsilon.$$

Como o compacto $K \times L$ confere a estimação desejada para todos os planos de transporte γ, segue que $\Pi(\mu,\nu)$ é compacto. □

Dessa forma, o resultado de existência sai diretamente aplicando-se o método direto do cálculo das variações, Teorema 1.3.1.

Teorema 1.6.1. *Dados $\mu \in \mathscr{P}(\mathcal{X})$, $\nu \in \mathscr{P}(\mathcal{Y})$ e $c : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}_+$ semi-contínua inferiormente, o problema de Kantorovitch admite um plano de transporte ótimo.*

Demonstração. O resultado é uma aplicação direta do teorema 1.3.1 com o critério de semi-continuidade de 1.5.1 e a compacidade de $\Pi(\mu,\nu)$ do Lemma 1.6.1. □