Universidade Federal do Rio Grande do Sul

Instituto de Matemática e Estatística

Programa de Pós-Graduação em Matemática

Tópicos em Transporte Ótimo

Dissertação de Mestrado

Marcus Vinícius da Silva

Porto Alegre, Agosto de 2019.

Dissertação submetida por Marcus Vinícius da Silva [1] como requisito parcial para a obtenção do grau de Mestre em Matemática pelo Programa de Pós-Graduação em Matemática do Instituto de Matemática e Estatística da Universidade Federal do Rio Grande do Sul.

Professor Orientador:
Diego Marcon Farias (PPGMat-UFRGS)

Banca Examinadora:
Fagner Bernardini Rodrigues (PPGMat-UFRGS)
Rafael Rigão Souza (PPGMat-UFRGS)
Matheus Correia dos Santos (PPGMAp-UFRGS)

[1] Bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, e do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq

Conteúdo

Agradecimentos

Essa dissertação em especial, mas também minhas conquistas pessoais em geral, não teriam sido possíveis não fosse pelo apoio da minha família. Todos, mesmo quem não entende porque eu ainda estou estudando ("quer dizer que no doutorado vão ser mais quatro anos?? Minha nossa!") em todas oportunidades demonstraram estar torcendo pelo meu sucesso. Agradeço a todos vocês, mas duas pessoas foram especialmente importantes na minha trajetória.

Minha mãe, Maria Helena, que sempre me deu carinho, amor e condições para me desenvolver, é uma delas. Graças aos sacrifícios que minha mãe fez tive o privilégio de me dedicar em primeiro lugar aos estudos. Minha mãe é uma mulher sensacional e de energia infinita, suficiente para ser mãe e pai quando foi preciso; ela é meu modelo de resiliência (e também de empreendedorismo; já fez tanta coisa que é mais fácil listar o que ela não fez). Por isso e por todo amor envolvido, sou muito grato. A outra pessoa é a minha vó Celézia. Sempre tem meu bem estar em mente, muitas vezes até esquecendo do próprio. Existem muitas formas de dizer "eu te amo", entre as quais dar alimento, dar café e ralhar comigo todas as vezes que me viu carregando mochilas pesadas são as preferidas da minha vó. Por isso e por tanto mais, muito obrigado.

Ao longo da vida acadêmica encontrei diversos anjos da guarda de carne e osso. A professora Vanusa me proporcionou o primeiro contato com Matemática mais legal do que aquela vista em sala de aula. Foi ela quem me ensinou a passar um número na base decimal para base binária, o que absolutamente me fascinou na época, e também foi ela quem me inscreveu na OBMEP lá em 2005. Tanta coisa aconteceu depois disso e por causa disso, que é assustador pensar nas possibilidades. Por identificar em primeira mão o gosto pela Matemática na criaturinha que eu era e por me ajudar em tudo que foi possível para que eu pudesse desenvolver esse interesse, muito obri-

gado.

Através da OBMEP e da Vanusa acabei conhecendo o professor Rogério, outro anjo da guarda. Ele aprofundou meu interesse em Matemática nos cursos aos sábados do PIC-Jr., mas sobretudo foi como um mentor mais tarde nos estudos da graduação. Quando eu ainda cursava Ciência da Computação, era ele a primeira pessoa para quem eu pedia conselhos sobre quais disciplinas da Matemática eu deveria fazer como extracurriculares. E nunca deixei de me espantar com como ele *se importava* de fato com o que eu tinha a dizer. Mesmo sendo um cara ocupado, sempre tinha tempo para pensar de verdade sobre os problemas que um guri (eu) lhe apresentava. Por ser esse professor excepcional e uma influência muito positiva na minha vida, muito obrigado.

Um dos professores indicados pelo Rogério foi o professor Brietzke. Com ele fiz boa parte da graduação em Matemática, incluindo aí as Análises, e portanto é razoável supor que o meu jeito de ver a Matemática foi, pelo menos em parte, influenciado por ele. Mas o motivo de ele aparecer nesses agradecimentos não é tanto matemático, embora já fosse o suficiente. Desde que tive aula com ele, ele é meu modelo de humildade em sala de aula e fora dela. Apesar de deter um conhecimento muito grande, ele jamais me fez sentir como se eu não soubesse o suficiente e nem me deixou intimidado para fazer uma pergunta. Eu posso já não lembrar as respostas, mas permanece o sentimento de que ninguém é menos por ainda não saber algo, e por isso muito obrigado.

Mais diretamente relacionado a essa dissertação, tenho muito a agradecer ao meu amigo e orientador Diego. De portas abertas para receber os alunos (metaforicamente, mas também literalmente), o Diego nunca se furtou de tentar o máximo para me ajudar. Mesmo nos momentos em que eu já não acreditava no meu potencial, ele estava lá para sentar comigo e explicar os conceitos até eu estar razoavelmente tranquilo e sentindo que eu também entendia a Matemática acontecendo ali. Por ter paciência para me explicar tudo de novo pela n-ésima vez, e por fazer questão que eu estivesse curtindo o mestrado, muito obrigado.

O ambiente acadêmico é terreno fértil para o florescimento de ansiedade e frustração. As pessoas de fora não acreditam quando eu falo, mas o maior desafio para mim durante a graduação e o mestrado tem sido de ordem emocional, e não matemática. A Matemática que estudamos é difícil, mas nada que um pouco de dedicação (horas de estudo, discussão com os colegas, etc.) não resolva; lidar com a frustração de não aprender tudo na velocidade que se quer, isso sim é que é difícil. Para lidar com isso de forma saudável,

contei com a ajuda da Cláudia, minha psicóloga. Se hoje eu tenho uma postura muito mais paciente a respeito do meu próprio aprendizado, e se, mais geralmente, eu trato a mim mesmo com mais gentileza, é por causa dela. Por tudo isso, muito obrigado.

Eu sou um cara privilegiado em vários aspectos, e um deles é estar constantemente rodeado de pessoas sensacionais. Agradeço aos amigos do Coral da UFRGS, que em pouco tempo se tornaram essenciais para mim. Agradeço aos amigos da Sociedade Mateológica (Affonso, Alex, Bianca, Felipe, Gabriel, Jan, Lorenzo, Micheli, Nicole, Vinícius) e ao Jéferson e ao Madruga, por todas as longas conversas, por entreterem minhas suposições mais absurdas, por me manterem a uma distância segura de qualquer conversa trivial, pelos anos de amizade já idos e pelos que ainda estão por vir.

Feltes, Gleiciano, Izabella, Jader, Josué, Pessil e William (e mais uma galera) acompanharam mais de perto o processo todo, alguns desde antes do mestrado, na verdade. Muito obrigado por dividirem comigo esse caminho, por estarem comigo enquanto a gente aprende alguma Matemática (ou pelo menos se acostuma com ela), mas também obrigado por todo o resto. Quando tudo era confusão e eu não enxergava o que eu tinha que fazer, vocês traziam uma tocha e tudo ficava claro. A vida não é um problema de otimização, e com vocês eu sempre prefiro pegar o caminho mais longo se isso significar passarmos mais tempo juntos.

Um agradecimento especial à Taís, que foi a rocha em que me segurei em vários momentos chave. Pós-graduação é uma fase atribulada, mas fica bem mais fácil quando a gente pode dividir a carga com alguém especial. Muito obrigado pela parceria e pela paciência! Te amo!

Por último, agradeço às pessoas do Instituto de Matemática e Estatística da UFRGS, especialmente àquelas que fazem o Programa de Pós-Graduação em Matemática continuar crescendo. Muito obrigado aos colegas da pós, que criam um ambiente tão bom de se frequentar. Deveria ser o normal, mas nem toda pós é colaborativa e não competitiva como a nossa. Agradeço ainda às agências de fomento, CNPq e Capes, porque sem elas eu e muitos outros jamais poderíamos ter ocupado esse espaço. A situação brasileira atual é crítica, mas eu vejo muita gente boa que não se deixa abater e é essa gente que me dá esperança de que atravessaremos essa era medieval. Espero que consigamos manter a gratuidade e excelência do ensino superior público e que mais pessoas ainda tenham o privilégio que eu tive e continuo tendo de estudar e trabalhar com o que se ama.

Dedico essa dissertação para a minha mãe, que mesmo sabendo pouco do

formalismo de otimização, nunca foi nada menos do que a melhor possível.

Introdução

A área de Transporte Ótimo se originou em um problema formulado pelo matemático francês Gaspard Monge em 1781 [1]. O Problema de Monge (PM) consiste em mover toda a massa de um lugar para outro da forma mais eficiente possível. Formalmente, consiste em, dadas uma função custo $c : X \times Y \longrightarrow \mathbb{R}$ e probabilidades $\mu \in P(X)$ e $\nu \in P(Y)$, minimizar o funcional

$$I(T) = \int_X c(x, T(x))\, d\mu(x)$$

dentre todas aplicações $T : X \longrightarrow Y$ tais que $T_{\#}\mu = \nu$ (isto é, a medida ν é tal que $\nu(A) = \mu(T^{-1}(A))$ para todo $A \subseteq X$ boreliano; dizemos nesse caso que ν é o *push-forward* de μ por T). Se T satisfaz isso, T é chamada de aplicação de transporte. O valor $c(x, y)$ pode ser interpretado como o custo de levar uma unidade de massa do ponto x para o ponto y, e as probabilidades μ e ν podem ser interpretadas como a forma em que a massa está distribuída em X e Y, respectivamente.

O problema de minimizar o funcional acima pode ser muito difícil. De fato, dependendo das propriedades da função custo, dos espaços X e Y e das probabilidades μ e ν, pode ser que nem mesmo exista T aplicação de transporte. Por exemplo, se considerarmos $\mu = \delta_x$ (toda massa concentrada em um único ponto de X), a condição $T_{\#}\mu = \nu$ implica que $\nu = \delta_{T(x)}$. Logo, se ν não for concentrada em um ponto, estaremos procurando o minimizante do funcional custo em um conjunto vazio.

Mais tarde, em 1942, o matemático soviético Leonid Kantorovich fez contribuições importantes para a teoria e formulou uma versão relaxada do Problema de Monge [2]. O Problema de Kantorovich (PK) consiste em minimizar o funcional

$$J(\gamma) = \int_{X \times Y} c(x, y)\, d\gamma(x, y)$$

entre todas probabilidades $\gamma \in P(X \times Y)$ tais que μ e ν são as marginais de γ. Isto é, para todo $A \subseteq X$ e todo $B \subseteq Y$ conjuntos mensuráveis, γ é tal que $\gamma(A \times Y) = \mu(A)$ e $\gamma(X \times B) = \nu(B)$. Em outras palavras, procuramos minimizar o funcional entre todos $\gamma \in \textsc{Adm}(\mu, \nu) := \{\gamma \in P(X \times Y) \,|\, \pi^1_{\#}\gamma = \mu, \pi^2_{\#}\gamma = \nu\}$, onde π^1 e π^2 são as projeções na primeira e na segunda coordenada, respectivamente. Os elementos de $\textsc{Adm}(\mu, \nu)$ são chamados planos de transporte. O problema de estarmos procurando minimizante em um conjunto vazio não acontece aqui, uma vez que, por exemplo, a probabilidade $\gamma = \mu \times \nu \in \textsc{Adm}(\mu, \nu)$.

Note que se T é uma aplicação de transporte (ou seja, $T_{\#}\mu = \nu$), então ela dá origem a uma probabilidade $\gamma = (\mathrm{Id}, T)_{\#}\mu \in \textsc{Adm}(\mu, \nu)$. Isso mostra que a formulação de Kantorovich é mais fraca e que o conjunto onde estamos procurando um representante ótimo é, de certa forma, maior do que aquele do PM. Mostramos adiante que, sob certas condições, ao resolvermos o PK estamos (como um bônus) também resolvendo o PM. Fazemos isso mostrando que uma solução ótima para o PK é uma probabilidade cujo suporte está contido no gráfico de uma função. A partir daí podemos construir uma aplicação T que é solução do PM.

Exemplo 0.1 (Problema de Kantorovich discreto)**.** Consideremos o caso de uma empresa que tem lojas precisando de produtos e depósitos contendo esses produtos. A empresa precisa levar os produtos dos depósitos para as lojas, e quer minimizar o gasto com as viagens. Sabe-se que

- o custo de levar uma unidade de produto do depósito i para a loja j é de $c(i, j)$;
- a quantidade de produtos no depósito i é dado por $\mu(\{i\})$;
- a quantidade de produtos necessários na loja j é $\nu(\{j\})$.

Isto é, dados os conjuntos $X = \{1, \cdots, n\}$ (depósitos), $Y = \{1, \cdots, m\}$ (lojas), uma função custo $c : X \times Y \longrightarrow \mathbb{R}$ e probabilidades $\mu \in P(X)$ e $\nu \in P(Y)$, o PK consiste em determinar uma probabilidade $\pi \in P(X \times Y)$ que minimize

$$I[\pi] = \sum_{i,j}^{n,m} c(i, j)\pi(i, j),$$

e tal que μ e ν sejam suas marginais. Ou seja, representando π como uma matriz

$$\pi = \begin{bmatrix} \pi_{11} & \cdots & \pi_{1m} \\ \vdots & \ddots & \vdots \\ \pi_{n1} & \cdots & \pi_{nm} \end{bmatrix}$$

em que a i-ésima linha representa a quantidade de produto que o depósito i enviou para cada loja, e a j-ésima coluna representa a quantidade de produto que a loja j recebeu de cada depósito, queremos que

$$\sum_{j=1}^{m} \pi_{ij} = \mu(\{i\}) \text{ e } \sum_{i=1}^{n} \pi_{ij} = \nu(\{j\}).$$

Em outras palavras, queremos que cada depósito envie tudo o que tem disponível, e que cada loja receba tudo o que precisa. A maneira mais comum de se resolver esse problema e problemas semelhantes é através de programação linear. As notas de aula em [3] dão uma boa introdução ao assunto e a outros tópicos de otimização.

Nessa dissertação estamos interessados em problemas de otimização deste tipo, porém em contextos mais abstratos. No capítulo 3 demonstramos a existência e unicidade de uma solução para o PM no espaço Euclidiano usual, quando o custo é o quadrado da distância. Este resultado é conhecido como Teorema de Brenier [4]. Embora o capítulo seja dedicado principalmente à demonstração desse teorema, os primeiros resultados são bastante gerais e são úteis para os outros capítulos. No capítulo 4 provamos o Teorema de McCann [5], que nos permite enunciar resultado semelhante ao de Brenier, mas agora em uma variedade compacta com o custo dado pelo quadrado da distância Riemanniana. Por fim, no capítulo 5, mostramos um resultado mais geral, onde substituimos o custo dado pela distância Riemanniana por um custo dado pela ação mínima de um Lagrangiano. Nos capítulos 1 e 2 provamos resultados auxiliares que são usados nos capítulos 4 e 5.

A seguir são feitos alguns comentários a respeito de notação e definições usadas nessa dissertação.

- Se M é uma variedade diferenciável e $x \in M$, o espaço tangente a x é denotado por T_xM.

- O espaço cotangente a x é denotado por T_x^*M. Trata-se do dual de T_xM, isto é, $T_x^*M = \{f : T_xM \longrightarrow \mathbb{R} \mid f \text{ é linear}\}$.

- O fibrado tangente de uma variedade M é o conjunto $TM = \cup_{x \in M}\{x\} \times T_xM$. Da mesma forma, o fibrado cotangente é definido por $T^*M = \cup_{x \in M}\{x\} \times T_x^*M$.

- Se a dimensão do espaço ambiente X é n, dizemos que um conjunto é um *conjunto pequeno* se tiver medida finita pela medida de Hausdorff de dimensão $n - 1$.

- Dizemos que uma sequência (x_n) em um espaço normado X converge fracamente para $x \in X$ quando, para todo funcional linear limitado $f : X \longrightarrow \mathbb{R}$, tivermos que $\lim f(x_n) = f(x)$. Denota-se $x_n \rightharpoonup x$.

Capítulo 1

Funções Semicôncavas

O conceito de função côncava é muito importante em vários contextos de otimização, mas pode ser muito restritivo. Funções semicôncavas abrangem um conjunto muito maior de funções do que funções côncavas, enquanto mantem várias de suas propriedades. Em particular, a função custo definida a partir de um Lagrangiano é localmente semicôncava, como mostramos no Teorema 5.2. Isso nos permite falar sobre diferenciabilidade em quase toda parte dessa função.

A apresentação das definições e resultados desse capítulo é inspirada, principalmente, em [6].

Definição 1.1 (Módulo de continuidade)**.** Um módulo de continuidade, ou simplesmente um módulo, é uma função contínua não-decrescente $\omega : [0, +\infty) \to [0, +\infty)$ tal que $\omega(0) = 0$. Dizemos que ω é um módulo linear se for da forma $\omega(t) = kt$, para algum $k \geq 0$.

Definição 1.2 (Superdiferencial)**.** $p \in T_x^*M$ é um superdiferencial da função $f : M \longrightarrow \mathbb{R}$ em $x \in M$ se for a derivada no ponto x de uma função $g : U \subseteq M \longrightarrow \mathbb{R}$ que tangencia f por cima no ponto x. Isto é, g é tal que $g(y) \geq f(y)$ para todo $y \in U$, $g(x) = f(x)$ e $p = dg(x)$. O conjunto que contem os superdiferenciais de f no ponto x é denotado por $\partial^+ f(x)$.

Definição 1.3 (Função semicôncava)**.** Uma função $f : U \longrightarrow \mathbb{R}$ definida em um aberto $U \subseteq \mathbb{R}^n$ é dita ω-semicôncava em U (ou, equivalentemente, semicôncava em U com módulo ω) se, para cada $x \in U$, existir uma função linear $l_x : \mathbb{R}^n \longrightarrow \mathbb{R}$ tal que, para todo $y \in U$,

$$f(y) - f(x) \leq l_x(y - x) + |y - x|\omega(|y - x|).$$

Podemos dizer simplesmente que f é semicôncava, omitindo o módulo, quando isso não gerar confusão. Dizemos que f é localmente semicôncava em U se, para cada $x \in U$, existir uma vizinhança aberta V_x tal que f é semicôncava em V_x para algum módulo (que não precisa ser necessariamente o mesmo em cada V_x). A função linear l_x pode depender do ponto x.

Proposição 1.4. 1) *Dadas $f_i : U \longrightarrow \mathbb{R}$, $U \subseteq \mathbb{R}^n$, tais que f_i é semicôncava com módulo ω_i em U para $i \in \{1, \ldots, k\}$, temos:*

(*i*) *para quaisquer $\alpha_1, \ldots, \alpha_k \geq 0$, a função $\sum_{i=1}^k \alpha_i f_i$ é $\sum_{i=1}^k \alpha_i \omega_i$-semicôncava;*

(*ii*) *a função $\min_{1 \leq i \leq k} f_i$ é $\max_{1 \leq i \leq k} \omega_i$-semicôncava.*

2) *Toda função de classe C^1 é localmente semicôncava.*

Demonstração. 1) *i*) Como f_i é ω_i-semicôncava em U, para cada $x \in U$ existe $l_x^i : \mathbb{R}^n \longrightarrow \mathbb{R}$ tal que

$$f_i(y) - f_i(x) \leq l_x^i(y - x) + |y - x|\omega_i(|y - x|).$$

Multiplicando por α_i e somando em i ficamos com

$$\sum_{i=1}^k \alpha_i f_i(y) - \sum_{i=1}^k \alpha_i f_i(x) \leq \sum_{i=1}^k \alpha_i l_x^i(y - x) + |y - x| \sum_{i=1}^k \alpha_i \omega_i(|y - x|).$$

Como $\sum_{i=1}^k \alpha_i l_x^i$ é linear e $\sum_{i=1}^k \alpha_i \omega_i$ é módulo, segue que a função $\sum_{i=1}^k \alpha_i f_i$ é semicôncava com módulo $\sum_{i=1}^k \alpha_i \omega_i$.

ii) Sejam $x \in U$ e i_0 tal que $f_{i_0}(x) = \min_i f_i(x)$. Como f_{i_0} é ω_{i_0}-semicôncava, existe l_x que satisfaz

$$\begin{aligned} f_{i_0}(y) - f_{i_0}(x) &\leq l_x(y - x) + |y - x|\omega_{i_0}(|y - x|) \\ &\leq l_x(y - x) + |y - x| \max_i \omega_i(|y - x|). \end{aligned}$$

Mas $f_{i_0}(y) - f_{i_0}(x) \geq \min_i f_i(y) - \min_i f_i(x)$ pois i_0 é tal que $f_{i_0}(x) = \min_i f_i(x)$. Logo

$$\min_i f_i(y) - \min_i f_i(x) \leq l_x(y - x) + |y - x| \max_i \omega_i(|y - x|)$$

de onde segue que $\min_i f_i$ é $\max_i \omega_i$-semicôncava em U.

2) Seja C um subconjunto de U aberto convexo com fecho compacto (uma bola, por exemplo). A função $x \mapsto df(x)$ é contínua porque f é de classe C^1, portanto possui módulo de continuidade ω em C (pois usando a compacidade de $\bar{C}$ podemos definir uma função que funcione como módulo de continuidade). Ou seja, existe um módulo ω tal que $\|df(y) - df(x)\| \leq \omega(|y-x|)$. Definindo $\gamma(t) = ty + (1-t)x$, temos que

$$\begin{aligned}\frac{d}{dt}(f \circ \gamma)(t) &= df(\gamma(t)) \cdot \gamma'(t) \\ &= df(ty + (1-t)x) \cdot (y-x)\end{aligned}$$

de onde segue pelo Teorema Fundamental do Cálculo que

$$\int_0^1 df(ty + (1-t)x) \cdot (y-x)\, dt = \int_0^1 \frac{d}{dt}(f \circ \gamma)(t)\, dt = f(y) - f(x). \quad (1.1)$$

Como $\|df(y) - df(x)\| = \sup_{|v|\leq 1} |df(y) \cdot (v) - df(x) \cdot (v)|$, então

$$df(y) \cdot \left(\frac{y-x}{|y-x|}\right) - df(x) \cdot \left(\frac{y-x}{|y-x|}\right) \leq \|df(y) - df(x)\|.$$

Portanto

$$\begin{aligned}df(ty + (1-t)x) \cdot \left(\frac{y-x}{|y-x|}\right) - df(x) \cdot \left(\frac{y-x}{|y-x|}\right)& \\ &\leq \|df(ty + (1-t)x) - df(x)\| \\ &\leq \omega(|ty + (1-t)x - x|) \\ &= \omega(|t(y-x)|) \\ &\leq \omega(|y-x|)\end{aligned}$$

o que implica que

$$df(ty + (1-t)x) \cdot (y-x) \leq df(x) \cdot (y-x) + |y-x|\omega(|y-x|).$$

Integrando os dois lados dessa última desigualdade em t, e usando a equação (1.1), segue que

$$\begin{aligned}f(y) - f(x) &= \int_0^1 df(ty + (1-t)x) \cdot (y-x)\, dt \\ &\leq df(x) \cdot (y-x) + |y-x|\omega(|y-x|).\end{aligned}$$

Logo, para $x \in C$ existe $l_x = df(x)$ função linear tal que

$$f(y) - f(x) \leq df(x) \cdot (y - x) + |y - x|\omega(|y - x|).$$

Ou seja, f é ω-semicôncava em C. Como isso vale para qualquer $x \in U$, f é localmente semicôncava. $\square$

Lema 1.5. *Sejam U um subconjunto aberto de $\mathbb{R}^n$ e $f : U \longrightarrow \mathbb{R}$ uma função ω-semicôncava. Valem:*

(i) *Para todo K subconjunto compacto de U existe uma constante A tal que, se $x \in K$, então para toda função linear l_x satisfazendo*

$$f(y) - f(x) \leq l_x(y - x) + |y - x|\omega(|y - x|)$$

tem-se que $\|l_x\| \leq A$.

(ii) *f é localmente Lipschitz.*

Demonstração. Primeiramente vamos mostrar que f é localmente limitada.

Note que se K é compacto, então é limitado e portanto $K \subseteq B_R$ para algum raio R. Como f é ω-semicôncava, temos que

$$\begin{aligned} f(y) - f(x) &\leq l_x(y - x) + |y - x|\omega(|y - x|) \\ \Longrightarrow \ f(y) &\leq l_x(y - x) + |y - x|\omega(|y - x|) + f(x) \\ &\leq \|l_x\||y - x| + |y - x|\omega(|y - x|) + f(x) \\ &\leq \|l_x\|R + R\omega(R) + f(x). \end{aligned}$$

Portanto $f(y) \leq \|l_x\|R + R\omega(R) + f(x)$ para todo $y \in K$. Isto é, f é limitada superiormente em K.

Para mostrar a limitação por baixo, seja $x \in U$ e C um cubo compacto contido em U com $x \in C$. Se $\{y_1, \ldots, y_{2^n}\}$ são os vértices de C, podemos escrever x como uma combinação convexa, isto é, $x = \sum \alpha_i y_i$, onde $\alpha_i \geq 0$ e $\sum \alpha_i = 1$. Para cada i temos

$$f(y_i) - f(x) \leq l_x(y_i - x) + |y_i - x|\omega(|y_i - x|).$$

Multiplicando por α_i e somando em i, obtemos

$$\sum \alpha_i f(y_i) - \sum \alpha_i f(x) \leq l_x\left(\sum \alpha_i y_i - \sum \alpha_i x\right) + \sum \alpha_i |y_i - x|\omega(|y_i - x|).$$

Seja D_C o diâmetro de C. Como $\sum \alpha_i = 1$ e $x = \sum \alpha_i y_i$, segue das desigualdades acima que

$$\begin{aligned}\sum \alpha_i f(y_i) - f(x) \leq \sum \alpha_i |y_i - x| \omega(|y_i - x|) \leq D_C \omega(D_C) \\ \implies \sum \alpha_i f(y_i) - D_C \omega(D_C) \leq f(x).\end{aligned}$$

Note que $\sum \alpha_i f(y_i) \geq \min_i f(y_i)$, e que portanto

$$\min_i f(y_i) - D_C \omega(D_C) \leq f(x).$$

Logo f é localmente limitada inferiormente também, ou seja, f é localmente limitada.

$i)$ Suponha que $\bar{B}(x_0, 2r) \subseteq U$. Se $x \in \bar{B}(x_0, r)$ então $x - rv \in \bar{B}(x_0, 2r)$ para todo $v \in \mathbb{R}^n$ com $|v| = 1$. Portanto

$$\begin{aligned}f(x - rv) - f(x) &\leq l_x(x - rv - x) + |x - rv - x| \omega(|x - rv - x|) \\ &= -r l_x(v) + r\omega(r).\end{aligned}$$

Da compacidade de $\bar{B}(x_0, 2r)$ e como vimos que f é localmente limitada, podemos definir $\tilde{B} = \sup_{z \in \bar{B}(x_0, 2r)} |f(z)|$. Daí segue que

$$\begin{aligned}f(x - rv) - f(x) &\leq -r l_x(v) + r\omega(r) \\ \implies f(x) - f(x - rv) &\geq r l_x(v) - r\omega(r) \\ \implies l_x(v) &\leq \frac{f(x) - f(x - rv)}{r} + \omega(r) \\ &\leq \frac{2\tilde{B}}{r} + \omega(r).\end{aligned}$$

Podemos aplicar essa desigualdade para $-v$ no lugar de v, do que podemos concluir que, se $x \in \bar{B}(x_0, r)$, então $\|l_x\| \leq \frac{2\tilde{B}}{r} + \omega(r)$.

Como K é compacto, podemos cobri-lo com uma quantidade finita de bolas. Aplicando o raciocínio acima para cada bola e chamando a maior constante de A, obtemos que $\|l_x\| \leq A$.

$ii)$ Seja K um subconjunto compacto de U. Pela parte (i), existe A tal que, para quaisquer $x, y \in K$,

$$\begin{aligned}f(y) - f(x) &\leq l_x(y - x) + |y - x| \omega(|y - x|) \\ &\leq \|l_x\| |y - x| + |y - x| \omega(|y - x|) \\ &\leq (A + \omega(D_K)) |y - x|\end{aligned}$$

onde D_K é o diâmetro de K. Vale a mesma desigualdade com x e y trocados. Logo $|f(y) - f(x)| \leq L|y - x|$, onde $L = (A + \omega(D_K))$. Isto é, f é localmente Lipschitz. $\square$

Enunciaremos sem provar o seguinte resultado bastante conhecido:

Teorema 1.6 (Rademacher)**.** *Se $U \subseteq \mathbb{R}^n$ é um aberto e $f : U \longrightarrow \mathbb{R}$ é uma função de Lipschitz, então f é diferenciável em quase toda parte em relação à medida de Lebesgue.*

Temos imediatamente o seguinte corolário:

Corolário 1.7. *Se $f : U \longrightarrow \mathbb{R}$ é localmente ω-semicôncava então f é diferenciável em quase toda parte em relação à medida de Lebesgue.*

Para o caso de funções semicôncavas, existe um resultado ainda mais forte. O conjunto dos pontos em que a função f semicôncava não é diferenciável em um conjunto pequeno. Uma prova deste fato pode ser encontrada em [7, Section 4.1]. Para a definição de um conjunto pequeno, ver o final da Introdução.

Teorema 1.8. *Se $\varphi : U \longrightarrow \mathbb{R}$ é uma função semicôncava definida no aberto $U \subseteq \mathbb{R}^n$ então φ é diferenciável em cada ponto do complemento de um conjunto pequeno.*

Até aqui as funções consideradas estavam definidas no $\mathbb{R}^n$. O seguinte lema nos permite estender a definição de semiconcavidade para funções reais definidas em uma variedade.

Lema 1.9. *Sejam $U, V \subseteq \mathbb{R}^n$ abertos. Suponha que $F : V \longrightarrow U$ é de classe C^1. Se $f : U \longrightarrow \mathbb{R}$ é localmente semicôncava então $f \circ F : V \longrightarrow \mathbb{R}$ também é localmente semicôncava.*

Além disso, se F é de classe C^2 e f é localmente semicôncava com módulo linear, então o mesmo pode ser dito de $f \circ F$.

Demonstração. Pela natureza local do lema, podemos assumir, sem perda de generalidade, que f é semicôncava em U. Seja V' uma bola aberta tal que $\bar{V}' \subseteq V$. Pela continuidade da função $z \mapsto \|dF(z)\|$, sabemos que existe $C_{\bar{V}'} = \max_{z \in \bar{V}'} \|dF(z)\| < +\infty$. Como $dF(x)$ é a derivada de F, pela definição de derivada sabemos que existe um módulo $\hat{\omega}_{\bar{V}'}$ tal que

$$F(y) - F(x) = dF(x) \cdot (y - x) + (y - x)\hat{\omega}_{\bar{V}'}(|y - x|).$$

Pelo Teorema do Valor Médio temos que, para algum $z \in \bar{V}'$, vale que

$$\frac{|F(y) - F(x)|}{|y - x|} \leq \|dF(z)\|.$$

Logo

$$|F(y) - F(x)| \leq \|dF(z)\||y - x| \leq C_{\bar{V}'}|y - x|.$$

Como f é semicôncava em U, segue que

$$\begin{aligned} f(F(y)) - f(F(x)) &\leq l_{F(x)}(F(y) - F(x)) \\ &\quad + |F(y) - F(x)|\omega(|F(y) - F(x)|) \\ &= l_{F(x)}(dF(x) \cdot (y - x) + (y - x)\hat{\omega}_{\bar{V}'}(|y - x|)) \\ &\quad + |F(y) - F(x)|\omega(|F(y) - F(x)|) \\ &= l_{F(x)}(dF(x) \cdot (y - x)) + l_{F(x)}((y - x)\hat{\omega}_{\bar{V}'}(|y - x|)) \\ &\quad + |F(y) - F(x)|\omega(|F(y) - F(x)|) \\ &\leq l_{F(x)}(dF(x) \cdot (y - x)) + \|l_{F(x)}\||y - x|\hat{\omega}_{\bar{V}'}(|y - x|) \\ &\quad + C_{\bar{V}'}|y - x|\omega(C_{\bar{V}'}|y - x|). \end{aligned}$$

Como $F(\bar{V}')$ é compacto, pelo Lema 1.5 existe uma constante $\hat{C}_{\bar{V}'}$ que limita $\|l_{F(x)}\|$. Logo

$$\begin{aligned} f(F(y)) - f(F(x)) &\leq l_{F(x)}(dF(x) \cdot (y - x)) \\ &\quad + \hat{C}_{\bar{V}'}|y - x|\hat{\omega}_{\bar{V}'}(|y - x|) + C_{\bar{V}'}|y - x|\omega(C_{\bar{V}'}|y - x|) \end{aligned}$$

o que significa que $f \circ F$ é semicôncava em $\bar{V}'$ com módulo $r \mapsto \hat{C}_{\bar{V}'}\hat{\omega}_{\bar{V}'}(r) + C_{\bar{V}'}\omega(C_{\bar{V}'}r)$. Logo $f \circ F$ é localmente semicôncava em V.
Se F é de classe C^2, dF é localmente Lipschitz em U e podemos assumir que $\hat{\omega}_{\bar{V}'}$ é linear (pois é o módulo de continuidade de dF). □

Agora estamos autorizados a falar de semiconcavidade em variedades. Por definição, para que uma função $f : M \longrightarrow \mathbb{R}$ seja localmente semicôncava, basta que sua composição com alguma carta coordenada F seja localmente semicôncava. O lema anterior garante que esse conceito fica bem definido, e é coerente com o caso em que $M = \mathbb{R}^n$.

Definição 1.10. Uma família de funções $\{f_i : U \longrightarrow \mathbb{R}\}$ é dita uniformemente ω-semicôncava se cada f_i é ω-semicôncava, ou seja, se existir um módulo ω tal que f_i é ω-semicôncava, para todo i.

Note que a uniformidade se dá porque o mesmo módulo vale para todas as funções. A função linear l_x que aparece na definição de função semicôncava não precisa ser a mesma.

Teorema 1.11. *Seja $(f_i)_{i\in I}$ uma família uniformemente ω-semicôncava. Se a função $f(x) := \inf_{i\in I} f_i(x)$ é finita em todo U, então f é ω-semicôncava.*

Demonstração. Fixa $x_0 \in U$. Pela definição de ínfimo, podemos construir uma sequência (f_{i_n}) tal que $f_{i_n}(x_0)$ converge para $f(x_0)$ por cima. Seja $C \subseteq U$ um cubo com centro em x_0, e sejam $y_1, \ldots, y_{2^n}$ os vértices de C. Para todo $x \in C$ existem $\alpha_1, \ldots, \alpha_{2^n}$ tais que $\sum \alpha_j = 1$ e $\sum \alpha_j y_j = x$. Como cada f_i é ω-semicôncava, temos que

$$f_i(y_j) - f_i(x) \leq l_x(y_j - x) + |y_j - x|\omega(|y_j - x|).$$

Multiplicando por α_j e somando em j, segue que

$$\begin{aligned}
\sum \alpha_j f_i(y_j) - \sum \alpha_j f_i(x) \leq l_x &\left(\sum \alpha_j y_j - \sum \alpha_j x\right) \\
&+ \sum \alpha_j |y_j - x|\omega(|y_j - x|) \\
\implies \sum \alpha_j f_i(y_j) &\leq f_i(x) + \sum \alpha_j |y_j - x|\omega(|y_j - x|) \\
\implies \min_{1\leq j\leq 2^n} f_i(y_j) &\leq f_i(x) + \sum \alpha_j |y_j - x|\omega(|y_j - x|) \\
\implies \min_{1\leq j\leq 2^n} f_i(y_j) &\leq f_i(x) + D_C\omega(D_C) \\
\implies \min_{1\leq j\leq 2^n} f_i(y_j) - D_C\omega(D_C) &\leq f_i(x).
\end{aligned}$$

Como $f(y_j) = \inf_i f_i(y_j)$ é finito, existe $A \in \mathbb{R}$ tal que $A \leq f_i(x)$ para todo $i \in I$.

Seja $\varepsilon > 0$ tal que $\bar{B}(x_0, \varepsilon) \subseteq C$. Como f_i é ω-semicôncava e $x_0 \in U$, existe l_i tal que

$$f_i(y) - f_i(x_0) \leq l_i(y - x_0) + |y - x_0|\omega(|y - x_0|)$$

para todo $y \in U$. Para todo $v \in \mathbb{R}^n$ com $|v| = 1$ vale que $x_0 + \varepsilon v \in \bar{B}(x_0, \varepsilon)$, e portanto

$$\begin{aligned}
f_i(x_0 + \varepsilon v) - f_i(x_0) &\leq l_i(x_0 + \varepsilon v - x_0) \\
&\quad + |x_0 + \varepsilon v - x_0|\omega(|x_0 + \varepsilon v - x_0|) \\
\implies f_i(x_0 + \varepsilon v) - f_i(x_0) &\leq \varepsilon l_i(v) + \varepsilon\omega(\varepsilon).
\end{aligned}$$

Como visto acima $A \leq f_i(x)$ para todo $x \in C$. Além disso, já que $f_{i_n}(x_0) \searrow f(x_0)$, podemos assumir que f_{i_n} é limitada, isto é, existe $M \in \mathbb{R}$ tal que $f_{i_n}(x_0) \leq M$ para todo n. Logo

$$\begin{aligned}
&f_i(x_0 + \varepsilon v) \leq \varepsilon l_i(v) + \varepsilon\omega(\varepsilon) + f_i(x_0) \\
&\implies A \leq \varepsilon l_i(v) + \varepsilon\omega(\varepsilon) + f_i(x_0) \\
&\implies A \leq \varepsilon l_i(v) + \varepsilon\omega(\varepsilon) + M \\
&\implies \frac{A - M}{\varepsilon} - \omega(\varepsilon) \leq l_i(v).
\end{aligned}$$

Aplicando isso para $-v$ obtemos que $\|l_i\| \leq \frac{M-A}{\varepsilon} + \omega(\varepsilon)$. Extraindo uma subsequência, se necessário, podemos supor que $l_{i_n} \longrightarrow l$. Como para todo $y \in U$ temos $f(y) \leq f_{i_n}(y)$, passando ao limite em n obtemos que

$$\begin{aligned}
&f(y) \leq f_{i_n}(x_0) + l_{i_n}(y - x_0) + |y - x_0|\omega(|y - x_0|) \\
\implies &f(y) - f(x_0) \leq l(y - x_0) + |y - x_0|\omega(|y - x_0|).
\end{aligned}$$

Como $x_0 \in U$ foi escolhido arbitrariamente, isso conclui a demonstração. $\square$

O exemplo a seguir mostra que a semiconcavidade uniforme de uma família de funções não é automaticamente transportada por difeomorfismos, como a semiconcavidade local de uma função é.

Exemplo 1.12. Para $k \in \mathbb{R}$ definimos $f_k : \mathbb{R} \longrightarrow \mathbb{R}$ por $f_k(x) = kx$. É claro que a família $(f_k)_{k\in\mathbb{R}}$ é uniformemente ω-semicôncava para qualquer módulo ω. De fato,

$$\begin{aligned}
f_k(y) - f_k(x) &= ky - kx \\
&= k(y - x) \\
&\leq k(y - x) + |y - x|\omega(|y - x|)
\end{aligned}$$

pois $\omega \geq 0$.

Considere o difeomorfismo $\varphi : \mathbb{R}_+^* \longrightarrow \mathbb{R}_+^*$ dado por $\varphi(x) = x^2$. Mostraremos que não existe um aberto não-vazio $U \subseteq \mathbb{R}_+^*$ nem um módulo ω tais que a família $(f_k \circ \varphi\big|_U)$ seja uniformemente ω-semicôncava.

Suponha por absurdo que tenhamos, para algum aberto não-vazio U e algum módulo ω,

$$f_k \circ \varphi(y) - f_k \circ \varphi(x) \leq l_x(y - x) + |y - x|\omega(|y - x|)$$

onde l_x depende de k mas não de ω. Como $f_k \circ \varphi$ é diferenciável, devemos ter $l_x(y-x) = (f_k \circ \varphi)'(x)(y-x) = 2kx(y-x)$. Portanto, deveríamos ter

$$\begin{aligned}
&k(y^2 - x^2) \leq 2kx(y-x) + |y-x|\omega(|y-x|) \\
&\implies k(y^2 - x^2) - k(2xy - 2x^2) \leq |y-x|\omega(|y-x|) \\
&\implies k(y-x)^2 \leq |y-x|\omega(|y-x|).
\end{aligned}$$

Fixando $x, y \in U$ e definindo $h = y - x$, temos

$$kh^2 \leq |h|\omega(|h|) \implies k \leq \frac{\omega(|h|)}{|h|}$$

para todo k, absurdo.

Em virtude do exemplo acima, a seguinte definição é a mais razoável em variedades.

Definição 1.13. A família de funções $f_i : M \longrightarrow \mathbb{R}$, $i \in I$, definidas em uma variedade M de dimensão n, é dita uniformemente localmente semicôncava se existir uma cobertura aberta $(U_j)_{j \in J}$ de M, onde cada U_j é o domínio de uma carta coordenada $\varphi_j : U_j \longrightarrow V_j \subseteq \mathbb{R}^n$, tais que para todo $j \in J$ a família $(f_i \circ \varphi_j^{-1})_{i \in I}$ é uniformemente semicôncava em $V_j \subseteq \mathbb{R}^n$.

Corolário 1.14. *Se a família de funções $f_i : M \longrightarrow \mathbb{R}$, $i \in I$, é uniformemente localmente semicôncava e a função $f(x) = \inf_{i \in I} f_i(x)$ é finita em toda parte, então f é localmente semicôncava.*

Demonstração. A família de funções $(f_i)_{i \in I}$ ser uniformemente localmente semicôncava significa que $M \subseteq \bigcup U_j$, $\varphi_j : U_j \longrightarrow V_j \subseteq \mathbb{R}^n$ são tais que $(f_i \circ \varphi_j^{-1})_{i \in I}$ é uniformemente semicôncava para todo j. Isto é, cada $g_i = f_i \circ \varphi_j^{-1}$ é ω_j-semicôncava, portanto a família $(g_i)_{i \in I}$ é uniformemente semicôncava. Logo, pelo Teorema 1.11 a função $g(x) := \inf_{i \in I} g_i(x)$ é ω_j-semicôncava também. Daí segue que $g \circ \varphi_j$ é semicôncava, pois semiconcavidade permanece por difeomorfismos, pelo Lema 1.9. Isto é, $g \circ \varphi_j = f$ é semicôncava em U_j. Logo f é localmente semicôncava em M. □

A definição a seguir pode parecer estranha à primeira vista pelo acúmulo de advérbios, mas não entremos em pânico. Ela vai ser necessária nas Proposições 1.16 e 1.17 para mostrar que uma função definida a partir do ínfimo de funções semicôncavas também é semicôncava. Isto é necessário para provar a semiconcavidade local de uma função c-côncava, no Capítulo 5.

Definição 1.15. Seja $c : M \times N \longrightarrow \mathbb{R}$, com M uma variedade e N um espaço topológico. A família de funções $(c(\cdot, y))_{y \in N}$ é localmente uniformemente localmente semicôncava se para todo $y_0 \in N$ existe uma vizinhança V_0 de y_0 tal que $(c(\cdot, y))_{y \in V_0}$ é uniformemente localmente semicôncava em M.

Note que o primeiro "localmente"diz respeito ao espaço N, enquanto que o segundo diz respeito à variedade M.

Proposição 1.16. *Seja $c : M \times N \longrightarrow \mathbb{R}$, em que M é uma variedade e N é um espaço topológico, tal que a família $(c(\cdot, y))_{y \in N}$ é localmente uniformemente localmente semicôncava. Se $K \subseteq N$ é compacto e a função $f_K : U \longrightarrow \mathbb{R}$ dada por $f_K(x) = \inf_{y \in K} c(x, y)$ é finita em toda parte, então f_K é localmente semicôncava em M.*

Demonstração. A família $(c(\cdot, y))_{y \in N}$ é localmente uniformemente localmente semicôncava. Logo, para todo $y \in N$ existe V_y tal que a família $(c(\cdot, y))_{y \in V_y}$ é uniformemente localmente semicôncava.

Note que $K \subseteq \bigcup_{y \in K} V_y$. Como K é compacto, existem V_i, $i = 1, \ldots, l$ tais que $K \subseteq \bigcup_{i=1}^{k} V_i$. Em cada V_i a família $(c(\cdot, y))_{y \in V_i}$ é uniformemente localmente semicôncava. A função $f_i(x) = \inf_{y \in K \cap V_i} c(x, y)$ é finita em toda parte pois $f_i \geq f_K$ (ambas são definidas por ínfimos, mas f_i em um conjunto menor). Logo, pelo Corolário 1.14, f_i é localmente semicôncava para todo i. Como $f_K = \min_i f_i$, segue da Proposição 1.4 que f_K também é localmente semicôncava. □

Proposição 1.17. *Sejam M, N variedades. Se $c : M \times N \longrightarrow \mathbb{R}$ é localmente semicôncava em $M \times N$, então a família de funções $(c(\cdot, y))_{y \in N}$ é localmente uniformemente localmente semicôncava.*

Demonstração. Por hipótese, podemos cobrir $M \times N$ por uma família $(U_i \times W_j)_{i \in I, j \in J}$, onde U_i é aberto em M e W_j é aberto em N, com U_i domínio do mapa $\varphi_i : U_i \longrightarrow \tilde{U}_i \subseteq \mathbb{R}^n$ e W_j domínio do mapa $\psi_j : W_j \longrightarrow \tilde{W}_j \subseteq \mathbb{R}^m$, tais que $(\tilde{x}, \tilde{y}) \mapsto c(\varphi_i^{-1}(\tilde{x}), \psi_j^{-1}(\tilde{y}))$ é $\omega_{i,j}$-semicôncava em $\tilde{U}_i \times \tilde{W}_j$. Logo, a família $(c(\varphi_i^{-1}(\cdot), \psi_j^{-1}(\tilde{y})))_{\tilde{y} \in \tilde{W}_j}$ é uniformemente localmente semicôncava em $\tilde{U}_i$. Então, por definição, $(c(\cdot, y))_{y \in N}$ é localmente uniformemente localmente semicôncava. □

O seguinte corolário é consequência das Proposições 1.16 e 1.17.

Corolário 1.18. *Sejam M, N variedades, $c : M \times N \longrightarrow \mathbb{R}$ localmente semicôncava e $K \subseteq N$ compacto. Se $f_K(x) = \inf_{y \in K} c(x, y)$ for finita em toda parte, então $f_K : U \longrightarrow \mathbb{R}$ é localmente semicôncava em M.*

Capítulo 2

Mecânica Lagrangiana

A Mecânica Lagrangiana é uma formulação da Mecânica Clássica que combina a conservação do momento linear e a conservação da energia. Vários problemas físicos acabam se resumindo a encontrar uma trajetória minimizante para a ação de um Lagrangiano $\mathbb{A}_L$. Como é equivalente à Mecânica Newtoniana, pode ser mostrado que uma trajetória γ que minimiza $\mathbb{A}_L$ para o Lagrangiano $L(x,v) = \frac{m}{2}|v|^2 - U(x)$ é uma trajetória que descreve o movimento de uma partícula de massa m sujeita ao potencial U.

No Capítulo 4, mostramos a existência da aplicação de transporte ótimo para o custo $\frac{d^2}{2}$. Esse custo está relacionado a um Lagrangiano, o que motiva a generalização da teoria para custos que sejam provenientes de outros Lagrangianos. Considerar esses custos abre várias possibilidades interessantes de interpretação para o problema de transporte ótimo. Mudando o Lagrangiano podemos pensar no custo não só em termos econômicos, ou de distância, mas também, por exemplo, como a energia relacionada a algum sistema físico.

Neste capítulo, relembraremos alguns resultados importantes que são usados para mostrar a existência de aplicação de transporte ótimo quando o custo é definido a partir de um Lagrangiano. O que está aqui é baseado principalmente nos apêndices de [6].

Começamos pela definição do principal elemento da teoria que queremos construir.

Definição 2.1 (Lagrangiano)**.** Seja M uma variedade. Um Lagrangiano em M é uma função $L : TM \longrightarrow \mathbb{R}$ contínua e limitada inferiormente.

Note que se uma curva γ é definida sobre a variedade M, sua velocidade $\dot{\gamma}(t)$ pertence ao espaço tangente ao ponto $\gamma(t)$. Ou seja, o par $(\gamma(t), \dot{\gamma}(t))$

pertence a TM.

Definição 2.2 (Ação de um Lagrangiano)**.** Sejam $L : TM \longrightarrow \mathbb{R}$ um Lagrangiano em M e $\gamma : [a,b] \longrightarrow M$ uma curva absolutamente contínua. A ação do Lagrangiano L sobre a curva γ é dada por $\mathbb{A}_L(\gamma) = \int_a^b L(\gamma(s), \dot{\gamma}(s))\, ds$. Por conveniência, definimos $\mathbb{A}_L(\gamma) = +\infty$ se γ não for absolutamente contínua.

Agora que sabemos como associar a cada curva um número real, podemos nos perguntar qual curva, entre todas que tem as mesmas extremidades, que atinge o menor valor. Note que isso é muito semelhante ao que fazemos ao buscar curvas que ligam dois pontos e minimizam o funcional de comprimento em Geometria Riemanniana.

Definição 2.3 (Minimizante)**.** Seja L um Lagrangiano em M. Uma curva absolutamente contínua $\gamma : [a,b] \longrightarrow M$ é um L-minimizante se $\mathbb{A}_L(\gamma) \leq \mathbb{A}_L(\alpha)$ para toda curva absolutamente contínua $\alpha : [a,b] \longrightarrow M$ com $\gamma(a) = \alpha(a)$ e $\gamma(b) = \alpha(b)$.

Definição 2.4 (Lagrangiano de Tonelli)**.** Dizemos que $L : TM \longrightarrow \mathbb{R}$ é um Lagrangiano de Tonelli fraco se possuir as seguintes propriedades:

(a) L é de classe C^1;

(b) a função $L(x,\cdot) : T_xM \longrightarrow \mathbb{R}$ é estritamente convexa para qualquer $x \in M$;

(c) existe uma métrica Riemanniana completa g em M e uma constante $C \in \mathbb{R}$ tal que
$$L(x,v) \geq |v|_x + C$$
para todo $(x,v) \in TM$, onde $|\cdot|_x$ é a norma em T_xM obtida da métrica g;

(d) para todo $K \subseteq M$ compacto, a restrição de L para $T_KM = \bigcup_{x\in K} T_xM$ é superlinear nas fibras de TM, ou seja, para toda constante $A \geq 0$, existe uma constante $C(A,K) \in \mathbb{R}$ tal que
$$L(x,v) \geq A|v|_x + C(A,K)$$
para todo $(x,v) \in T_KM$.

Dizemos ainda que L é um Lagrangiano de Tonelli se satisfizer também as seguintes condições mais fortes:

(a') L é de classe C^2;

(b') para todo $(x,v) \in TM$, $\frac{\partial^2 L}{\partial v^2}(x,v)$ é positivo definida em T_xM.

Note que quando M é uma variedade compacta, se a condição (c) valer para alguma alguma métrica g, então valerá para todas, porque sobre um compacto todas são equivalentes.

Além disso, se L é um Lagrangiano de Tonelli fraco e $U : TM \longrightarrow \mathbb{R}$ é de classe C^1, então $L + U$ também é Lagrangiano de Tonelli fraco. Similarmente, se L é um Lagrangiano de Tonelli e U é de classe C^2, então $L + U$ também é Lagrangiano de Tonelli. Dessa forma, conhecendo um exemplo de Lagrangiano de Tonelli, podemos gerar vários.

Os próximos teoremas são resultados conhecidos de Cálculo de Variações em uma variável, que enunciaremos sem provar.

Teorema 2.5. *Seja L um Lagrangiano de Tonelli fraco definido numa variedade conexa M. Então para todo $a, b \in \mathbb{R}$ com $a < b$, e para todo $x, y \in M$, existe uma curva $\gamma : [a,b] \longrightarrow M$ absolutamente contínua que é L-minimizante.*

O Teorema 2.5 vale sob condições mais fracas do que L ser de classe C^1, mas essa hipótese é necessária no teorema a seguir.

Teorema 2.6. *Se L é um Lagrangiano de Tonelli fraco e γ é L-minimizante, então γ é de classe C^1.*

Além disso, para todo intervalo $[t_0, t_1]$ tal que $\gamma(t)$ está contido no domínio de uma carta coordenada, tem-se que

$$\frac{\partial L}{\partial v}(\gamma(t_1), \dot{\gamma}(t_1)) - \frac{\partial L}{\partial v}(\gamma(t_0), \dot{\gamma}(t_0)) = \int_{t_0}^{t_1} \frac{\partial L}{\partial x}(\gamma(s), \dot{\gamma}(s))\, ds.$$

Esta é uma versão integral das equações de Euler-Lagrange. Isso implica que $t \mapsto \frac{\partial L}{\partial v}(\gamma(t), \dot{\gamma}(t))$ é uma função de classe C^1 e que

$$\frac{d}{dt}\left(\frac{\partial L}{\partial v}(\gamma(t), \dot{\gamma}(t))\right) = \frac{\partial L}{\partial x}(\gamma(t), \dot{\gamma}(t)).$$

Mais geralmente, se L é de classe C^r, com $r \geq 2$, então γ é de classe C^r.

Definição 2.7 (Transformada Global de Legendre)**.** Seja L um Lagrangiano de classe C^1 em M. Sua Transformada Global de Legendre $\mathscr{L} : TM \longrightarrow T^*M$ é definida por $\mathscr{L}(x,v) = (x, \frac{\partial L}{\partial v}(x,v))$.

Proposição 2.8. *Se L é um Lagrangiano de Tonelli fraco em M, então sua Transformada Global de Legendre $\mathscr{L} : TM \longrightarrow T^*M$ é um homeomorfismo. Além disso, se L é um Lagrangiano de Tonelli de classe C^r, com $r \geq 2$, então $\mathscr{L}$ é de classe C^{r-1}.*

Demonstração. Primeiro mostramos que $\mathscr{L}$ é sobrejetiva.

Seja $p \in T_x^*M$. A condição (d) da definição de Lagrangiano de Tonelli fraco garante a existência de uma constante $C(A, K)$ para qualquer escolha de constante A e $K \subseteq M$ conjunto compacto. As escolhas espertas nesse caso são $A = |p|_x + 1$ e $K = \{x\}$. Segue que

$$L(x, v) \geq A|v|_x + C(A, K)$$

e logo

$$\begin{aligned} p(v) - L(x, v) &\leq p(v) - A|v|_x - C(A, K) \\ &\leq p(v) - |p|_x|v|_x - |v|_x - C(A, K). \end{aligned}$$

Mas temos que $p(v) \leq |p|_x|v|_x$. Logo

$$p(v) - L(x, v) \leq -|v|_x - C(A, K).$$

Como o lado direito vai para $-\infty$ conforme $|v|_x$ vai para $+\infty$, temos que a função $v \mapsto p(v) - L(x, v)$ deve atingir um máximo em um ponto v_p. Como é de classe C^1, no ponto v_p sua derivada deve ser nula. Ou seja:

$$\frac{\partial}{\partial v}(p(v) - L(x, v))|_{v=v_p} = p - \frac{\partial L}{\partial v}(x, v_p) = 0$$

o que implica que $p = \frac{\partial L}{\partial v}(x, v_p)$. Logo $(x, p) = \mathscr{L}(x, v_p)$.

Agora mostramos que $\mathscr{L}$ é injetiva.

Como para $x \neq x'$ temos $\mathscr{L}(x, v) \neq \mathscr{L}(x', v')$ diretamente, basta mostrar que para $v \neq v'$ teremos $\mathscr{L}(x, v) \neq \mathscr{L}(x, v')$. Para isso, considere a função $\varphi : [0, 1] \longrightarrow \mathbb{R}$ dada por $\varphi(t) = L(x, tv + (1-t)v')$. Pela condição (b) da definição de Lagrangiano de Tonelli fraco, φ é estritamente convexa. Como φ é de classe C^1, ser estritamente convexa implica que $\varphi'(0) \neq \varphi'(1)$. Como $\frac{dx}{dt} = 0$, temos que

$$\begin{aligned} \varphi'(t) &= \frac{d}{dt}L(x, tv + (1-t)v') \\ &= \frac{\partial L}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial L}{\partial v}(x, tv + (1-t)v')\frac{d}{dt}(tv + (1-t)v') \\ &= \frac{\partial L}{\partial v}(x, tv + (1-t)v')(v - v'). \end{aligned}$$

Logo $\frac{\partial L}{\partial v}(x, v) = \varphi'(1) \neq \varphi'(0) = \frac{\partial L}{\partial v}(x, v')$, e portanto $\mathscr{L}$ é injetiva.

Para que $\mathscr{L}$ seja homeomorfismo é necessário que tanto $\mathscr{L}$ quanto $\mathscr{L}^{-1}$ sejam contínuas. $\mathscr{L}$ é contínua porque L é C^1. Para mostrar que $\mathscr{L}^{-1}$ é contínua, mostramos que $\mathscr{L}$ é própria (isto é, pré-imagens de conjuntos compactos são relativamente compactas). Note que, se $P \subseteq T^*M$ é compacto, então existe uma constante C tal que $\|\frac{\partial L}{\partial v}(x, v)\|_x \leq C$ para todo $\frac{\partial L}{\partial v}(x, v) \in P$. Portanto, para mostrar que $\mathscr{L}$ é própria, basta mostrar que o conjunto $\{(x, v) \in TM \mid x \in K, \|\frac{\partial L}{\partial v}(x, v)\|_x \leq C\}$ é compacto (onde $K \subseteq M$ é compacto, como é costume para um conjunto denotado por essa letra). Mostramos que ele é compacto mostrando que está contido em um conjunto compacto. Logo, como é fechado, deve ser compacto também.

A convexidade estrita de $L(x, \cdot)$ implica que $\frac{\partial L}{\partial v}(x, v)(v) \geq L(x, v) - L(x, 0)$. De fato, como $L(x, \cdot)$ é estritamente convexa, sabemos que

$$L(x, tv + (1-t)v') > tL(x, v) + (1-t)L(x, v')$$

para todo $t \in (0, 1)$, e para todos $v, v' \in T_xM$. Fazendo $v' = 0$, obtemos

$$\begin{aligned} L(x, tv) &> tL(x, v) + L(x, 0) - tL(x, 0) \\ &= t(L(x, v) - L(x, 0)) + L(x, 0). \end{aligned}$$

Logo

$$\frac{L(x, tv) - L(x, 0)}{t} > L(x, v) - L(x, 0).$$

Isso vale para todo t. Logo, fazendo $t \longrightarrow 0^+$, obtemos que

$$\frac{\partial L}{\partial v}(x, v)(v) \geq L(x, v) - L(x, 0).$$

Ainda, L é contínua, e portanto existe uma constante C_1 tal que $L(x, 0) \leq C_1$ para todo $x \in K$. Então $\frac{\partial L}{\partial v}(x, v)(v) \geq L(x, v) - C_1$. Pela condição (d) da definição de Lagrangiano de Tonelli fraco, existe uma constante $C(A, K)$ tal que $L(x, v) \geq A|v|_x + C(A, K)$ para todo $(x, v) \in T_KM$. Logo

$$\frac{\partial L}{\partial v}(x, v)(v) \geq L(x, v) - C_1 \geq A|v|_x + C(A, K).$$

Onde juntamos as constantes em uma só porque fica mais bonito. Daí segue que

$$\frac{\partial L}{\partial v}(x, v)\left(\frac{v}{|v|_x}\right) \geq A + \frac{C(A, K)}{|v|_x} \implies \left\|\frac{\partial L}{\partial v}(x, v)\right\|_x \geq A + \frac{C(A, K)}{|v|_x}.$$

Podemos tomar $A = C + 1$, e daí temos

$$\left\| \frac{\partial L}{\partial v}(x,v) \right\|_x \geq C + 1 + \frac{C(C+1,K)}{|v|_x}.$$

Seja $(x,v) \in \{(x,v) \in TM \mid x \in K, \|\frac{\partial L}{\partial v}(x,v)\|_x \leq C\}$. Para esse (x,v), vale que

$$C \geq \left\| \frac{\partial L}{\partial v}(x,v) \right\|_x \geq C + 1 + \frac{C(C+1,K)}{|v|_x} \implies -C(C+1,K) \geq |v|_x.$$

Portanto

$$\{(x,v) \in TM \mid x \in K, \left\| \frac{\partial L}{\partial v}(x,v) \right\|_x \leq C\} \subseteq$$
$$\{(x,v) \in TM \mid x \in K, |v|_x \leq -C(C+1,K)\}$$

que é compacto, logo o conjunto anterior é compacto também.

Por fim, suponha que L seja um Lagrangiano de Tonelli de classe C^r. Pelo Teorema da Função Inversa, para mostrar que $\mathscr{L}$ é um difeomorfismo C^{r-1}, basta mostrar que sua derivada é invertível. Calculando as derivadas de $\mathscr{L}$ obtemos a matriz

$$\begin{bmatrix} Id & 0 \\ \frac{\partial^2 L}{\partial x \partial v}(x,v) & \frac{\partial^2 L}{\partial v^2}(x,v) \end{bmatrix}$$

que é invertível pois, pela definição de Lagrangiano de Tonelli, $\frac{\partial^2 L}{\partial v^2}(x,v) > 0$. □

Definição 2.9. Seja L um Lagrangiano em M. Definimos o seu Hamiltoniano $H : T^*M \longrightarrow \mathbb{R} \cup \{+\infty\}$ por

$$H(x,p) = \sup_{v \in T_x M} \Big(p(v) - L(x,v) \Big).$$

Dessa relação entre o Lagrangiano e seu Hamiltoniano podemos deduzir várias propriedades que o Hamiltoniano deve ter. A proposição seguinte lista essas propriedades.

Proposição 2.10. *Seja L um Lagrangiano de Tonelli fraco definido na variedade M. Seu Hamiltoniano H é finito em toda parte e goza das seguintes propriedades:*

(a^*) H é de classe C^1 e em coordenadas vale que

$$\begin{cases} \frac{\partial H}{\partial p}(\mathscr{L}(x,v)) = v \\ \frac{\partial H}{\partial x}(\mathscr{L}(x,v)) = -\frac{\partial L}{\partial x}(x,v) \end{cases}$$

(b^*) para cada $x \in M$, o mapa $H(x,\cdot) : T_x^*M \longrightarrow \mathbb{R}$ é estritamente convexo;

(d^*) para todo compacto $K \subseteq M$, a restrição de H a $T_K^*M = \bigcup_{x\in K} T_x^*M$ é superlinear nas fibras de T^*M. Isto é, para para toda constante $A \geq 0$ existe uma constante $C^*(A,K)$ tal que $H(x,p) \geq A|p|_x + C^*(A,K)$ para todo $(x,p) \in T_K^*M$.

Se L é de classe C^r, com $r \geq 2$, então valem também:

(a'^*) H é de classe C^r;

(b'^*) para todo $(x,p) \in T^*M$, a segunda derivada parcial $\frac{\partial^2 H}{\partial p^2}(x,p)$ é positiva definida em T_x^*M.

Demonstração. A menos de composição com alguma carta coordenada, podemos sem perda de generalidade assumir que $M = U \subseteq \mathbb{R}^n$. Como U é aberto, podemos ainda considerar uma bola V compactamente contida em U. Isso é útil porque normas advindas de métricas Riemannianas são equivalentes em variedades compactas. No caso, trabalhamos tendo em mente a norma Euclidiana usual. A propriedade (d) da Definição 2.4 passa a ser a seguinte: para qualquer escolha de constante $A \geq 0$ existe uma constante C_A tal que $L(x,v) \geq A|v| + C_A$ para todo $x \in V$ e para todo $v \in \mathbb{R}^n = T_{\bar{V}}M$.

Temos que $T^*V = V \times \mathbb{R}^{n*}$, em que $\mathbb{R}^{n*}$ é o dual de $\mathbb{R}^n$. Fixemos $R > 0$, e suponhamos que $|p| \leq R$. Para $A = R+1$, existe C_{R+1} tal que

$$L(x,v) \geq (R+1)|v| + C_{R+1}$$

$$\Longrightarrow -L(x,v) \leq -(R+1)|v| - C_{R+1}$$
$$\Longrightarrow p(v) - L(x,v) \leq p(v) - (R+1)|v| - C_{R+1}.$$

Mas $p(v) \leq |p||v| \leq R|v|$, logo

$$p(v) - L(x,v) \leq R|v| - (R+1)|v| - C_{R+1}$$

$$\Longrightarrow p(v) - L(x,v) \leq -|v| - C_{R+1}.$$

Como L é C^1, $L(\cdot, 0)$ é limitada em um compacto. Ou seja, existe uma constante $C \in \mathbb{R}$ tal que $L(x, 0) \leq C$ para todo $x \in V \subseteq \bar{V}$. Daí segue que, para v com $|v| > C - C_{R+1}$, temos

$$\begin{aligned} & -|v| < -C + C_{R+1} \\ \implies & p(v) - L(x,v) \leq -|v| - C_{R+1} \\ & < -C + C_{R+1} - C_{R+1} \\ \implies & p(v) - L(x,v) < -C \leq -L(x,0) \end{aligned}$$

para todo $x \in V$ e para todo v tal que $|v| > C - C_{R+1}$. Isso implica que

$$\begin{aligned} H(x,p) &= \sup_{v \in \mathbb{R}^m} p(v) - L(x,v) \\ &= \sup_{|v| \leq C - C_{R+1}} p(v) - L(x,v). \end{aligned}$$

Isto é, o supremo que define $H(x, p)$ é atingido em um ponto $v_{(x,p)}$ com $|v_{(x,p)}| \leq C - C_{R+1}$. Como a função $v \mapsto p(v) - L(x, v)$ é de classe C^1, a sua derivada no ponto $v_{(x,p)}$ deve ser nula.

$$0 = \frac{d}{dv}(p(v) - L(x,v)) = p - \frac{\partial L}{\partial v}(x, v_{(x,p)}) \implies p = \frac{\partial L}{\partial v}(x, v_{(x,p)}).$$

Logo $(x, p) = \mathscr{L}(x, v_{(x,p)})$ e $v_{(x,p)}$ é único porque $\mathscr{L}$ é injetiva. Como $p = \frac{\partial L}{\partial v}(x, v_{(x,p)})$ e $L(x, \cdot)$ é estritamente convexa (sua derivada é crescente, portanto invertível), obtemos que $v_{(x,p)} = \left(\frac{\partial L}{\partial v}\right)^{-1}(p)$. Para que a notação não fique muito carregada, definimos $g := \left(\frac{\partial L}{\partial v}\right)^{-1}$.

$$H(x,p) = p(v_{(x,p)}) - L(x, v_{(x,p)}) = p(g(p)) - L(x, g(p)).$$

Logo

$$\begin{aligned} H(x,p+h) - H(x,p) &= (p+h)(g(p+h)) - L(x, g(p+h)) \\ &\quad - p(g(p)) + L(x, g(p)) \\ &= p(g(p+h)) + h(g(p+h)) - L(x, g(p+h)) \\ &\quad - p(g(p)) + L(x, g(p)) \\ &= p(g(p+h) - g(p)) + h(g(p+h)) \\ &\quad - (L(x, g(p+h)) - L(x, g(p))). \end{aligned} \tag{2.1}$$

Por expansão de Taylor, temos que

$$\begin{aligned}L(x,g(p+h))-L(x,g(p)) &= \frac{\partial L}{\partial v}(g(p))\cdot(g(p+h)-g(p))+o(h)\\ &= p(g(p+h)-g(p))+o(h),\end{aligned}$$

pois $\frac{\partial L}{\partial v}(g(p))=p$. Substituindo em (2.1) ficamos com

$$\begin{aligned}H(x,p+h)-H(x,p) &= p(g(p+h)-g(p))+h(g(p+h))\\ &\quad -p(g(p+h)-g(p))+o(h)\\ &= h(g(p+h))+o(h).\end{aligned}$$

Fazendo $h\longrightarrow 0$ obtemos que

$$\frac{\partial H}{\partial p}(x,p)=g(p)=\left(\frac{\partial L}{\partial v}\right)^{-1}(p)=v_{(x,p)}.$$

Além disso $H(x,p)=p(v_{(x,p)})-L(x,v_{(x,p)})$ implica que $\frac{\partial H}{\partial x}(x,p)=\frac{\partial}{\partial x}(p(v)-L(x,v))=-\frac{\partial L}{\partial x}(x,v_{(x,p)})$. Como $(x,p)=\mathscr{L}(x,v_{(x,p)})$ podemos reescrever

$$\begin{aligned}\frac{\partial H}{\partial p}\circ\mathscr{L}(x,v)&=v\\ \frac{\partial H}{\partial x}\circ\mathscr{L}(x,v)&=-\frac{\partial L}{\partial x}(x,v),\end{aligned}$$

o que prova (a^*).

Quando L é um Lagrangiano de Tonelli de classe C^r, sabemos que $\mathscr{L}$ é de classe C^{r-1}. Disso segue que tanto $\frac{\partial H}{\partial p}$ quanto $\frac{\partial H}{\partial x}$ são C^{r-1}, e portanto H é C^r. Isso prova (a'^*).

Provaremos agora (b'^*). Já sabemos que $\frac{\partial H}{\partial p}(x,\frac{\partial L}{\partial v}(x,v))=v$. Diferenciando essa expressão em relação a v, obtemos

$$\frac{\partial^2 H}{\partial p^2}(\mathscr{L}(x,v))\cdot\frac{\partial^2 L}{\partial v^2}(x,v)=Id_{\mathbb{R}^m}.$$

Logo, $\frac{\partial^2 H}{\partial p^2}$ é a inversa de uma matriz positiva definida, de forma que deve ser positiva definida também.

Agora provamos a convexidade estrita do item (b^*). Sejam $p_1,p_2\in T_x^*M$ com $p_1\neq p_2$, e defina $p_3=tp_1+(1-t)p_2$ para algum $t\in(0,1)$. Sejam, ainda,

$v_1, v_2, v_3 \in T_xM$ tais que $p_i = \frac{\partial L}{\partial v}(x, v_i)$ para $i = 1, 2, 3$. Tais v_i existem e são diferentes porque $\mathscr{L}$ é bijetiva.

Para $i = 1, 2$ temos $H(x, p_i) = p_i(v_i) - L(x, v_i)$ pois o supremo da definição de H é atingido precisamente em $p_i = \frac{\partial L}{\partial v}(x, v_i)$. Daí segue que

$$\begin{aligned} H(x, p_3) &= p_3(v_3) - L(x, v_3) \\ &= t(p_1(v_3) - L(x, v_3)) + (1-t)(p_2(v_3) - L(x, v_3)). \end{aligned} \tag{2.2}$$

Como os supremos são atingidos em um único ponto, temos que

$$H(x, p_1) > p_1(v_3) - L(x, v_3)$$

e

$$H(x, p_2) > p_2(v_3) - L(x, v_3).$$

Juntando isso com (2.2), ficamos com

$$H(x, tp_1 + (1-t)p_2) < tH(x, p_1) + (1-t)H(x, p_2),$$

ou seja, $H(x, \cdot)$ é estritamente convexa em T_x^*M e (b^*) está provado.

Para provar (d^*), seja $K \subseteq M$ compacto. Como $H(x, p) \geq p(v) - L(x, v)$ para todo $(x, v) \in TM$ (pois é um supremo), temos que

$$H(x, p) \geq \sup_{|v|_x \leq A} p(v) + \inf_{x \in K, |v|_x \leq A} -L(x, v)$$

onde o primeiro termo é igual a $A|p|_x$ e o segundo é finito por compacidade, e portanto podemos definir $C^*(A, K) = \inf_{x \in K, |v|_x \leq A} -L(x, v)$. $\square$

Definição 2.11. O campo vetorial Hamiltoniano X_H associado ao Hamiltoniano H é definido por $X_H(x, p) = (\frac{\partial H}{\partial p}(x, p), -\frac{\partial H}{\partial x}(x, p))$ e a EDO associada é

$$\begin{cases} \dot{x} = \frac{\partial H}{\partial p}(x, p) \\ \dot{p} = -\frac{\partial H}{\partial x}(x, p)) \end{cases}$$

Observe que se $(x(t), p(t))$ é uma solução da EDO associada a X_H, então

$$\frac{d}{dt} H(x, p) = \frac{\partial H}{\partial x} \cdot \dot{x} + \frac{\partial H}{\partial p} \cdot \dot{p} = \frac{\partial H}{\partial x} \cdot \frac{\partial H}{\partial p} + \frac{\partial H}{\partial p} \cdot \left(-\frac{\partial H}{\partial x} \right) = 0.$$

Portanto H é constante ao longo de soluções da EDO associada a X_H.

Teorema 2.12. *Seja L um Lagrangiano de Tonelli fraco definido na variedade M. Se $\gamma : [a,b] \longrightarrow M$ é um L-minimizante, então $t \mapsto \mathscr{L}(\gamma(t), \dot\gamma(t))$ é uma solução de classe C^1 do campo vetorial X_H.*

Além disso, se L é um Lagrangiano de Tonelli, então existe um fluxo (parcial) ϕ_t^L de classe C^1 em TM tal que toda curva velocidade de um L-minimizante é parte de uma órbita de ϕ_t^L. Esse fluxo é chamado de Fluxo de Euler-Lagrange e é definido por $\phi_t^L = \mathscr{L}^{-1} \circ \phi_t^H \circ \mathscr{L}$ onde ϕ_t^H é o fluxo parcial de X_H.

Demonstração. Escrevendo $(x(t), p(t)) = \mathscr{L}(\gamma(t), \dot\gamma(t))$ ficamos com $x(t) = \gamma(t)$ e $p(t) = \frac{\partial L}{\partial v}(\gamma(t), \dot\gamma(t))$. Pelo Teorema 2.6, γ é de classe C^1 porque é L-minimizante. Portanto $x(t)$ também é de classe C^1 com $\dot x(t) = \dot\gamma(t)$. Da mesma forma, $p(t)$ é de classe C^1, e como $\frac{d}{dt}\frac{\partial L}{\partial v}(\gamma(t), \dot\gamma(t)) = \frac{\partial L}{\partial x}(\gamma(t), \dot\gamma(t))$ segue que $\dot p(t) = \frac{\partial L}{\partial x}(\gamma(t), \dot\gamma(t))$.

Como $(x(t), p(t)) = \mathscr{L}(\gamma(t), \dot\gamma(t))$ concluimos da Proposição 2.10 que $t \mapsto (x(t), p(t))$ satisfaz

$$\begin{cases} \dot x = \frac{\partial H}{\partial p}(x,p) \\ \dot p = -\frac{\partial H}{\partial x}(x,p)), \end{cases}$$

de forma que a Transformada de Legendre de uma curva velocidade de um L-minimizante é solução do campo vetorial Hamiltoniano X_H.

Se L é um Lagrangiano de Tonelli (ou seja, de classe C^2), então, pela Proposição 2.10, H é de classe C^2 também, e X_H é C^1, o que define um fluxo parcial ϕ_t^H. $\square$

Definição 2.13 (Energia)**.** Se L é um Lagrangiano C^1 definido na variedade M, então sua energia $E : TM \longrightarrow \mathbb{R}$ é definida por

$$E(x,v) = H \circ \mathscr{L}(x,v) = \frac{\partial L}{\partial v}(x,v)(v) - L(x,v).$$

Corolário 2.14. *Se L é um Lagrangiano de classe C^1 na variedade M e $\gamma : [a,b] \longrightarrow M$ é um L-minimizante, então a energia E é constante na curva velocidade $s \mapsto (\gamma(s), \dot\gamma(s))$.*

Demonstração. Note que $E(\gamma(s), \dot\gamma(s)) = H \circ \mathscr{L}(\gamma(s), \dot\gamma(s))$. A energia é constante nessa curva velocidade porque a função $s \mapsto \mathscr{L}(\gamma(s), \dot\gamma(s))$ é solução de X_H e H é constante em órbitas de X_H. $\square$

Proposição 2.15. *Se L é um Lagrangiano de Tonelli fraco definido na variedade M, então para todo subconjunto compacto $K \subseteq M$ e toda constante $C < +\infty$, o conjunto $\{(x,v) \in TM \,|\, x \in K, E(x,v) \leq C\}$ é compacto. Ou seja, a energia E é um mapa próprio (leva conjuntos compactos em conjuntos compactos) em todo subconjunto da forma $\pi^{-1}(K)$, onde π é a projeção natural.*

Demonstração. Como $E = H \circ \mathscr{L}$, a proposição segue do fato de H ser próprio e $\mathscr{L}$ ser um homeomorfismo. □

Enunciaremos agora, sem provar, o Teorema de Hopf-Rinow, importante resultado de Geometria Riemanniana que é usado na proposição que vem a seguir, bem como em outros pontos da dissertação.

Teorema 2.16 (Hopf-Rinow)**.** *Seja (M, g) uma variedade Riemanniana. São equivalentes:*

(i) *para $x \in M$ fixado, a aplicação exponencial $\exp_x$ está definida em todo o espaço tangente T_xM;*

(ii) *em M, conjuntos fechados e limitados (com respeito à distância Riemanniana) são compactos;*

(iii) *como espaço métrico, M é completo;*

(iv) *M é geodesicamente completa;*

(v) *existe exaustão de M por compactos: existe uma sequência (encaixada) de compactos $K_n \subseteq M$ com $K_n \subseteq \operatorname{int}(K_{n+1})$ e $\cup_n K_n = M$ tais que, para $x \in M$ e sequência $y_n \notin K_n$, vale*

$$\lim_{n \longrightarrow +\infty} d(x, y_n) = +\infty.$$

Além disso, os itens anteriores implicam no seguinte (mas não é equivalente):

(vi) *para quaisquer $x, y \in M$, existe geodésica minimizante $\gamma : [0, l] \longrightarrow M$ com $\gamma(0) = x$ e $\gamma(l) = y$.*

Proposição 2.17. *Seja L um Lagrangiano de Tonelli fraco definido na variedade M. Suponha que K seja um subconjunto compacto de M e $t > 0$. Então podemos achar um subconjunto compacto $\tilde{K} \subseteq M$ e uma constante $A \in \mathbb{R}$ tais que para qualquer $\gamma : [0,t] \longrightarrow M$ L-minimizante com $\gamma(0), \gamma(t) \in K$, temos que $\gamma([0,t]) \subseteq \tilde{K}$ e $|\dot{\gamma}(s)|_{\gamma(s)} \leq A$ para todo $s \in [0,t]$.*

Demonstração. A distância d é a que vem da métrica Riemanniana completa. Pelo Teorema de Hopf-Rinow, todas bolas limitadas fechadas nessa distância são compactas.

Escolhemos $x_0 \in K$ e $R > 0$ tais que $K \subseteq B(x_0, R)$. Sejam $x, y \in K$. Se $\alpha : [0,t] \longrightarrow M$ é uma geodésica com $\alpha(0) = x$ e $\alpha(t) = y$ cujo comprimento é $d(x,y)$ (que existe, pois a métrica é completa), a desigualdade triangular

$$d(x,y) \leq d(x,x_0) + d(x_0,y) < 2R$$

implica que, se $z \in \alpha([0,t])$, então

$$d(x_0,z) \leq d(x_0,x) + d(x,z) \leq d(x_0,x) + d(x,y) \leq 3R.$$

Ou seja, $\alpha([0,t]) \subseteq \bar{B}(x_0, 3R)$.

A curva α tem velocidade constante porque é uma geodésica. Isto é, $|\dot{\alpha}(s)|_{\alpha(s)} = \frac{d(x,y)}{t} \leq \frac{2R}{t}$, para todo $s \in [0,t]$. Por compacidade, o Lagrangiano L é limitado no conjunto $\mathscr{K} = \{(z,v) \in TM \mid z \in \bar{B}(x_0, 3R), |v|_z \leq \frac{2R}{t}\}$. Seja θ o limite superior de L em $\mathscr{K}$. A ação de α em $[0,t]$ é menor do que θt, pois

$$\mathbb{A}(\alpha) = \int_0^t L(\alpha(s), \dot{\alpha}(s))\, ds \leq \int_0^t \theta\, ds = \theta t.$$

Se $\gamma : [0,t] \longrightarrow M$ é um L-minimizante, então $\mathbb{A}(\gamma) \leq \mathbb{A}(\alpha)$, ou seja, $\int_0^t L(\gamma(s), \dot{\gamma}(s))\, ds \leq t\theta$. Pela condição (c) na Definição 2.4, existe C constante tal que para todo $(x,v) \in TM$ tem-se que $L(x,v) \geq |v|_x + C$. Logo

$$\begin{aligned} t\theta &\geq \int_0^t L(\gamma(s), \dot{\gamma}(s))\, ds \\ &\geq \int_0^t (|\dot{\gamma}(s)|_{\gamma(s)} + C)\, ds \\ &= Ct + \int_0^t |\dot{\gamma}(s)|_{\gamma(s)}\, ds \\ \Longrightarrow\ \theta - C &\geq \frac{1}{t} \int_0^t |\dot{\gamma}(s)|_{\gamma(s)}\, ds. \end{aligned}$$

Pelo Teorema do Valor Médio para Integrais, sabemos que existe $s_0 \in [0,t]$ tal que $|\dot{\gamma}(s_0)|_{\gamma(s_0)} = \frac{1}{t}\int_0^t |\dot{\gamma}(s)|_{\gamma(s)}\,ds$. Logo $\theta - C \geq |\dot{\gamma}(s_0)|_{\gamma(s_0)}$. Note que isso implica que $\gamma([0,t]) \subseteq \bar{B}(\gamma(0), t(\theta - C)) \subseteq \bar{B}(x_0, R + t(\theta - C))$. Essa segunda inclusão segue de que, se $\gamma(0) \in K$, então para todo $x \in \bar{B}(\gamma(0), t(\theta - C))$ temos

$$d(x, x_0) \leq d(x, \gamma(0)) + d(\gamma(0), x_0) \leq t(\theta - C) + R.$$

Fazendo $\tilde{K} = \bar{B}(x_0, R + t(\theta - C))$, definimos $\theta_1 = \sup\{E(z,v) \mid (z,v) \in TM, z \in \tilde{K}, |v|_z \leq \theta - C\}$, que é finito pela compacidade de $\tilde{K}$.

Finalmente, como $\gamma([0,t]) \subseteq \tilde{K}$ e $|\dot{\gamma}(s_0)|_{\gamma(s_0)} \leq \theta - C$, podemos concluir que $E(\gamma(s_0), \dot{\gamma}(s_0)) \leq \theta_1$. Mas, pelo Corolário 2.14 a energia E é constante para γ L-minimizante. Logo, a curva velocidade $s \mapsto (\gamma(s), \dot{\gamma}(s))$ está contida em $\tilde{\mathscr{K}} = \{(z,v) \in TM \mid z \in \tilde{K}, E(z,v) \leq \theta_1\}$, que é o compacto que queríamos. $\square$

Capítulo 3

Teorema de Brenier

Este capítulo é devotado para o Problema de Monge com $X = Y = \mathbb{R}^d$, em que a função custo é a distância quadrática advinda da norma de $\mathbb{R}^d$. Neste contexto, o Teorema de Brenier mostra que se a probabilidade μ não dá massa para conjuntos pequenos e ν é uma probabilidade qualquer então existe uma aplicação de transporte que é solução do Problema de Monge. Ainda, essa aplicação de transporte é o gradiente de uma função convexa e está relacionada a uma solução do Problema de Kantorovich.

Antes de chegarmos ao Teorema de Brenier, no entanto, precisamos de alguns resultados preliminares. Começamos por mostrar algumas propriedades do funcional custo e do conjunto de medidas em que buscamos o seu minimizante que asseguram que podemos usar o Método Direto do Cálculo de Variações para demonstrar a existência de plano de transporte ótimo no Problema de Kantorovich. Esses resultados valem em contextos bem gerais, e por isso não servirão apenas para o Teorema de Brenier, mas também para o de McCann no Capítulo 4 e para quando o custo vier de um Lagrangiano, no Capítulo 5. Dessa forma, vários dos resultados a seguir poderiam ter sido simplificados caso mudássemos um pouco as hipóteses. Em [8] é possível encontrar resultados semelhantes demonstrados para situações ligeiramente diferentes.

O Método Direto consiste basicamente em três passos. O primeiro passo é conseguir uma sequência minimizante de medidas γ_n. O segundo passo é usar alguma propriedade de compacidade para garantir que essa sequência tenha uma subsequência convergente. Por fim, mostrar que o limite dessa subsequência é o elemento que procuramos.

Esse capítulo é baseado principalmente em [9].

Um espaço polonês é um espaço topológico completo, metrizável e separável. Os exemplos que nos interessam nesta dissertação são o $\mathbb{R}^n$ e variedades Riemannianas em geral. Convergência fraca de medidas está definida no final da Introdução.

Lema 3.1. *Sejam X e Y espaços poloneses e $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ uma função custo que é semicontínua inferiormente e limitada por baixo. Então o funcional custo J, dado por*

$$J(\gamma) = \int_{X\times Y} c(x, y)\, d\gamma(x, y)$$

é semicontínuo inferiormente em $P(X \times Y)$ com a topologia da convergência fraca. Em outras palavras

$$\gamma_k \rightharpoonup \gamma \text{ em } P(X \times Y) \implies J(\gamma) \leq \liminf J(\gamma_k).$$

Demonstração. A semicontinuidade inferior e a limitação por baixo de c garantem que existe uma sequência crescente de funções contínuas c_n tais que pontualmente temos $c(x, y) = \lim_n c_n(x, y)$, ver [8, pág. 26]. Pelo Teorema da Convergência Monótona, segue que

$$\begin{aligned} J(\gamma) &= \int_{X\times Y} c\, d\gamma \\ &= \lim_n \int_{X\times Y} c_n(x, y)\, d\gamma \\ &= \lim_n \lim_k \int_{X\times Y} c_n(x, y)\, d\gamma_k \\ &\leq \liminf_k \int_{X\times Y} c\, d\gamma_k \\ &= \liminf_k J(\gamma_k). \end{aligned}$$

A terceira igualdade vem de $\gamma_k \rightharpoonup \gamma$ e a desigualdade vem de $c_n \leq c$. Logo o funcional custo é semicontínuo inferiormente. □

Definição 3.2. Seja X um espaço polonês. O conjunto $\mathcal{K} \subseteq P(X)$ é dito rígido se, para todo $\varepsilon > 0$, existir um compacto $K_\varepsilon \subseteq X$ tal que $\mu(X \backslash K_\varepsilon) < \varepsilon$ para qualquer $\mu \in \mathcal{K}$.

Dadas uma função mensurável $f : X \longrightarrow Y$ e uma probabilidade $\mu \in P(X)$, o *push-forward* de μ por f é a probabilidade $f_\# \mu \in P(Y)$ definida por $f_\# \mu(B) = \mu(f^{-1}(B))$, para $B \subseteq Y$ mensurável. No que segue, $\pi^1 : X \times Y \longrightarrow X$ é a projeção na primeira coordenada e $\pi^2 : X \times Y \longrightarrow Y$ a projeção na segunda coordenada.

Lema 3.3. *Sejam X e Y espaços poloneses. Se $\mathcal{K}_1 \subseteq P(X)$ e $\mathcal{K}_2 \subseteq P(Y)$ são ambos rígidos, então o conjunto*

$$\mathcal{K} = \{\gamma \in P(X \times Y) \,|\, \pi^1_\# \gamma \in \mathcal{K}_1 \text{ e } \pi^2_\# \gamma \in \mathcal{K}_2\}$$

também é rígido.

Demonstração. Primeiro, note que

$$(X \times Y) \backslash (K_1 \times K_2) \subseteq [(X \backslash K_1) \times Y] \cup [X \times (Y \backslash K_2)] \,. \tag{3.1}$$

Sejam $\varepsilon > 0$ e $\gamma \in \mathcal{K}$. Definimos $\mu := \pi^1_\# \gamma \in \mathcal{K}_1$ e $\nu := \pi^2_\# \gamma \in \mathcal{K}_2$. Como $\mathcal{K}_1$ e $\mathcal{K}_2$ são rígidos, existem $K_1 \subseteq X$ e $K_2 \subseteq Y$ compactos tais que $\mu(X \backslash K_1) < \frac{\varepsilon}{2}$ e $\nu(Y \backslash K_2) < \frac{\varepsilon}{2}$. Então, pelo que temos na equação (3.1), segue que

$$\begin{aligned} \gamma((X \times Y) \backslash (K_1 \times K_2)) &\leq \gamma((X \backslash K_1) \times Y) + \gamma(X \times (Y \backslash K_2)) \\ &= \mu(X \backslash K_1) + \nu(Y \backslash K_2) < \varepsilon. \end{aligned}$$

Portanto o conjunto $\mathcal{K}$ é rígido. □

O seguinte teorema, que enunciaremos sem provar, é um resultado clássico e será importante nos resultados que o seguem. Uma prova pode ser encontrada em [9].

Teorema 3.4 (Prokhorov)**.** *Se X é um espaço polonês, então $\mathcal{K} \subseteq P(X)$ é rígido se, e somente se, é relativamente compacto (isto é, possui a propriedade de que toda sequência tem subsequência convergente) com relação à convergência fraca.*

Agora estamos aptos a aplicar o Método Direto para garantir a existência de um plano de transporte ótimo. A definição de ínfimo nos permite construir uma sequência minimizante, que por estar em um conjunto relativamente compacto admite subsequência convergente. Mostramos que o limite dessa sequência é um elemento de $\textsc{Adm}(\mu, \nu) := \{\gamma \in P(X \times Y) \,|\, \pi^1_\# \gamma = \mu \text{ e } \pi^2_\# \gamma = \nu\}$, e obtemos, da semicontinuidade inferior do funcional J, que esse elemento é o que procurávamos.

Teorema 3.5 (Existência de Plano Ótimo)**.** *Suponha que a função custo é semicontínua inferiormente e limitada por baixo. Então existe um plano de transporte ótimo, isto é, um minimizante para o Problema de Kantorovich.*

Demonstração. Dadas $\mu \in P(X)$ e $\nu \in P(Y)$, os conjuntos $\{\mu\}$ e $\{\nu\}$ são rígidos. Logo, pelo Lema 3.3, o conjunto $\textsc{Adm}(\mu,\nu)$ é rígido também. Pelo Teorema de Prokhorov, $\textsc{Adm}(\mu,\nu)$ é relativamente compacto com relação à convergência fraca.

Pela definição de ínfimo, podemos construir uma sequência minimizante $\gamma_n \in \textsc{Adm}(\mu,\nu)$, tal que $J(\gamma_n) \leq \frac{1}{n} + \inf_\gamma J(\gamma)$. Como $\textsc{Adm}(\mu,\nu)$ é relativamente compacto, existe uma subsequência convergente $\gamma_{n_k} \rightharpoonup \gamma \in \overline{\textsc{Adm}(\mu,\nu)}$. Para mostrar que na verdade $\gamma \in \textsc{Adm}(\mu,\nu)$ (e não apenas ao seu fecho), considere $\varphi \in C_b(X)$. Temos

$$\begin{aligned}
\int_X \varphi(x)\, d(\pi^1_\#(\gamma))(x) &= \int_{X\times Y} \varphi(\pi^1(x,y))\, d\gamma(x,y) \\
&= \lim_{n_k} \int_{X\times Y} \varphi(\pi^1(x,y))\, d\gamma_{n_k}(x,y) \\
&= \lim_{n_k} \int_X \varphi(x)\, d(\pi^1_\#(\gamma_{n_k}))(x) \\
&= \lim_{n_k} \int_X \varphi(x)\, d\mu(x) \\
&= \int_X \varphi(x)\, d\mu(x),
\end{aligned}$$

ou seja, $\pi^1_\#(\gamma) = \mu$. Analogamente se mostra que $\pi^2_\#(\gamma) = \nu$. Portanto $\gamma \in \textsc{Adm}(\mu,\nu)$. Pelo Lema 3.1 o funcional custo J é semicontínuo inferiormente. Portanto $J(\gamma) \leq \liminf_{n_k} J(\gamma_{n_k}) \leq \liminf \frac{1}{n} + \inf_\alpha J(\alpha) = \inf_\alpha J(\alpha)$, e daí segue que γ é minimizante para o funcional custo J. Ou seja, γ é um plano de transporte ótimo. $\square$

Até aqui provamos a existência de uma solução para o Problema de Kantorovich. O nosso objetivo agora é caracterizar essa solução e mostrar que ela dá origem a uma solução do Problema de Monge. Isto é, para cada ponto x, ela leva a apenas um ponto y. Em outras palavras, mostraremos que o suporte da solução que obtemos está contido no gráfico de uma função (rigorosamente falando, isso será verdade exceto em um conjunto de medida desprezível). Para chegar lá, precisamos definir alguns conceitos antes.

Definição 3.6 (conjunto c-ciclicamente monótono)**.** Seja $c : X \times Y \longrightarrow \mathbb{R}$ uma funcão contínua. Um conjunto $\Gamma \subseteq X \times Y$ é c-ciclicamente monótono quando satisfaz, para todo subconjunto finito $\{(x_i, y_i)\}_{i=1,2,\cdots,N} \subseteq \Gamma$, que

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

para qualquer permutação $\sigma : \{1, 2, \cdots, N\} \longrightarrow \{1, 2, \cdots, N\}$.

Definição 3.7 (c-transformada)**.** Dada uma função $f : X \longrightarrow \mathbb{R} \cup \{-\infty\}$ que não é identicamente igual a $-\infty$, definimos sua c-transformada (ou função c-conjugada) como $f^c : Y \longrightarrow \mathbb{R} \cup \{-\infty\}$ dada por

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

Analogamente, se $g : Y \longrightarrow \mathbb{R} \cup \{-\infty\}$ então sua c-transformada $g^c : X \longrightarrow \mathbb{R} \cup \{-\infty\}$ é dada por

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

Definição 3.8 (função c-côncava)**.** Uma função $f : X \longrightarrow \mathbb{R} \cup \{-\infty\}$ é dita c-côncava quando existe $g : Y \longrightarrow \mathbb{R} \cup \{-\infty\}$ tal que $f = g^c$.

Definição 3.9 (c-superdiferencial)**.** Seja $f : X \longrightarrow \mathbb{R} \cup \{-\infty\}$ uma função c-côncava. O c-superdiferencial de f é o conjunto

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

Note que se considerarmos $c(x, y) = -\langle x, y \rangle$, recuperamos conceitos já conhecidos de Análise Convexa, a saber, conjunto ciclicamente monótono, Transformada de Legendre, função côncava e superdiferencial.

O teorema seguinte caracteriza o plano de transporte ótimo e é fundamental em nossa abordagem. A prova apresentada aqui é baseada em [11] e [12].

Teorema 3.10. *Seja $c : X \times Y \longrightarrow \mathbb{R}$ uma função custo que é contínua e limitada inferiormente. Sejam $\mu \in P(X)$ e $\nu \in P(Y)$ tais que, para alguma função $a \in L^1(\mu)$ e alguma função $b \in L^1(\nu)$, tenhamos que*

$$c(x, y) \leq a(x) + b(y). \tag{3.2}$$

Seja $\gamma \in \textsc{Adm}(\mu, \nu)$. 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* $\varphi : X \longrightarrow \mathbb{R} \cup \{-\infty\}$ *tal que* $\operatorname{supp} \gamma \subseteq \partial^c \varphi$.

Demonstração. Pela hipótese $c(x, y) \leq a(x) + b(y)$ e como c é limitada por baixo, temos que $c \in L^1(\tilde{\gamma})$ para qualquer $\tilde{\gamma} \in \textsc{Adm}(\mu, \nu)$. De fato:

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

(*i*) $\Longrightarrow$ (*ii*). Provaremos por contraposição. Suponha que o suporte de γ não seja c-ciclicamente monótono. Isto é, existem pontos $(x_i, y_i) \in \operatorname{supp} \gamma, i = 1, \cdots, n$ tais que

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

para alguma permutação σ. Construimos um plano de transporte $\tilde{\gamma}$ melhor (em relação ao custo) do que γ e, portanto, γ não é ótimo.

Seja $\varepsilon > 0$. Da continuidade de c, para cada $i \in \{1, \ldots, n\}$ podemos obter uma vizinhança $V_i = B_r(x_i) \times B_r(y_i)$ do ponto (x_i, y_i) tal que $|c(x, y) - c(x_i, y_i)| < \varepsilon$ para todo $(x, y) \in V_i$. Podemos tomar r pequeno o suficiente para que também tenhamos $|c(x, y) - c(x_i, y_{\sigma(i)})| < \varepsilon$ para todo $(x, y) \in B_r(x_i) \times B_r(y_{\sigma(i)})$.

Construimos $\tilde{\gamma}$ como segue:

$$\tilde{\gamma} := \gamma - \varepsilon_0 \sum_{i=1}^{n} \gamma_i + \varepsilon_0 \sum_{i=1}^{n} \tilde{\gamma}_i$$

onde $\gamma_i = \frac{1}{\gamma(V_i)} \gamma|_{V_i}$ e $\tilde{\gamma}_i = \mu_i \times \nu_{\sigma(i)} \in \textsc{Adm}(\mu_i, \nu_{\sigma(i)})$, onde $\mu_i = \pi^1_{\#} \gamma_i$ e $\nu_i = \pi^2_{\#} \gamma_i$. As medidas γ_i estão bem definidas porque $(x_i, y_i) \in \operatorname{supp} \gamma$ implica que $\gamma(V_i) > 0$. O valor de ε_0 será definido adiante para assegurar a positividade de $\tilde{\gamma}$.

Além da positividade de $\tilde{\gamma}$, temos que mostrar que suas marginais são μ e ν, e também mostrar que $\tilde{\gamma}$ se sai melhor do que γ em relação ao custo c.

Para ter positividade, seja $A \subseteq X \times Y$ um conjunto mensurável:

$$\varepsilon_0 \gamma_i(A) = \frac{\varepsilon_0}{\gamma(V_i)} \gamma(A \cap V_i) < \frac{\gamma(A)}{n}$$

onde a desigualdade acontece desde que $\varepsilon_0 < \frac{\min_i \gamma(V_i)}{n}$. Logo, para esta escolha de ε_0, garante-se que $\varepsilon_0 \sum_{i=1}^n \gamma_i(A) < \gamma(A)$, e que portanto $\tilde{\gamma}$ é positiva.

Para ver que $\tilde{\gamma} \in \textsc{Adm}(\mu, \nu)$, note que

$$\pi^1_\# \tilde{\gamma} = \pi^1_\# \left(\gamma - \varepsilon_0 \sum_{i=1}^n \gamma_i + \varepsilon_0 \sum_{i=1}^n \tilde{\gamma}_i \right) = \mu - \varepsilon_0 \sum_{i=1}^n \mu_i + \varepsilon_0 \sum_{i=1}^n \mu_i = \mu$$

e que

$$\pi^2_\# \tilde{\gamma} = \pi^2_\# \left(\gamma - \varepsilon_0 \sum_{i=1}^n \gamma_i + \varepsilon_0 \sum_{i=1}^n \tilde{\gamma}_i \right) = \nu - \varepsilon_0 \sum_{i=1}^n \nu_i + \varepsilon_0 \sum_{i=1}^n \nu_i = \nu.$$

Por fim, verificamos que $\tilde{\gamma}$ se sai melhor do que γ em relação ao custo c. Note que na primeira igualdade usamos que as integrais só diferem nas vizinhanças V_i.

$$\begin{aligned}
&\int_{X \times Y} c(x,y)\, d\gamma(x,y) - \int_{X \times Y} c(x,y)\, d\tilde{\gamma}(x,y) \\
&\qquad = \varepsilon_0 \sum_{i=1}^n \int_{X \times Y} c(x,y)\, d\gamma_i(x,y) - \varepsilon_0 \sum_{i=1}^n \int_{X \times Y} c(x,y)\, d\tilde{\gamma}_i(x,y) \\
&\qquad = \varepsilon_0 \sum_{i=1}^n \left(\int_{X \times Y} c(x,y)\, d\gamma_i(x,y) - \int_{X \times Y} c(x,y)\, d\tilde{\gamma}_i(x,y) \right) \\
&\qquad = \varepsilon_0 \sum_{i=1}^n \Bigg(\frac{1}{\gamma(V_i)} \int_{V_i} c(x,y)\, d\gamma(x,y) \\
&\qquad\qquad - \int_{B_r(x_i) \times B_r(y_{\sigma(i)})} c(x,y)\, d(\mu_i \times \nu_{\sigma(i)})(x,y) \Bigg) \\
&\qquad > \varepsilon_0 \sum_{i=1}^n \Big(c(x_i, y_i) - \varepsilon - c(x_i, y_{\sigma(i)}) - \varepsilon \Big) \\
&\qquad = \varepsilon_0 \Bigg(\sum_{i=1}^n (c(x_i, y_i) - c(x_i, y_{\sigma(i)})) - 2\varepsilon n \Bigg).
\end{aligned}$$

Como por hipótese $\sum_{i=1}^n (c(x_i, y_i) - c(x_i, y_{\sigma(i)})) > 0$, podemos escolher $\varepsilon > 0$ de tal forma que $\sum_{i=1}^n (c(x_i, y_i) - c(x_i, y_{\sigma(i)})) > 2\varepsilon n$. Assim obtemos que

$$\int_{X\times Y} c(x,y)\, d\gamma(x,y) > \int_{X\times Y} c(x,y)\, d\tilde{\gamma}(x,y)$$

e portanto γ não é plano de transporte ótimo.

$(ii) \implies (iii)$. Fixamos $(x_0, y_0) \in \operatorname{supp}\gamma$ e definimos

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

onde o ínfimo é avaliado entre todas as coleções de pontos $\{(x_i, y_i)\}_{i=1,2,\cdots,n} \subseteq \operatorname{supp}\gamma$, $n \geq 1$. Observe que as hipóteses de c assumir valores reais e que $\operatorname{supp}\gamma \neq \emptyset$ garantem que $\varphi < +\infty$. Pela definição de monotonicidade c-cíclica obtemos que $\varphi(x_0) \geq 0$. Em particular $\varphi \not\equiv 0$.

Para ver que φ é c-côncava, note que fazendo $y_n = y$ podemos escrever

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

onde

$$\begin{aligned}\psi(y) := -\inf\{-c(x_n,y) + [c(x_n,y_{n-1}) - c(x_{n-1},y_{n-1})] + \cdots + \\ + [c(x_1,y_0) - c(x_0,y_0)]\}.\end{aligned}$$

Passamos para a demonstração de que $\operatorname{supp}\gamma \subseteq \partial^c\varphi$. Isto é, se $(x,y) \in \operatorname{supp}\gamma$, então $\varphi(x) + \varphi^c(y) = c(x,y)$.

Como

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

então $\varphi(x) + \varphi^c(y) \leq c(x,y)$. Para verificar a desigualdade inversa, note que $\varphi^c = \psi^{cc} \geq \psi$ e portanto basta provar que $\varphi(x) + \psi(y) \geq c(x,y)$. Pela definição de ínfimo, dado $\varepsilon > 0$ existe um $\tilde{y} \in Y$ tal que $\varphi(x) + \varepsilon > c(x,\tilde{y}) - \psi(\tilde{y})$. Além disso, se considerarmos $x_n = x$ e $y_{n-1} = \tilde{y}$, temos

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

Tomando o ínfimo entre todas coleções finitas de pontos (x_i, y_i), ficamos com $-\psi(y) \leq -c(x,y) + c(x,\tilde{y}) - \psi(\tilde{y})$. Mas $\tilde{y}$ foi tomado tal que $-\psi(\tilde{y}) < \varphi(x) - c(x,\tilde{y}) + \varepsilon$. Logo

$$\begin{aligned}
-\psi(y) &\leq -c(x,y) + c(x,\tilde{y}) - \psi(\tilde{y}) \\
&< -c(x,y) + c(x,\tilde{y}) + \varphi(x) - c(x,\tilde{y}) + \varepsilon \\
&= -c(x,y) + \varphi(x) + \varepsilon.
\end{aligned}$$

Como isso vale para um $\varepsilon > 0$ qualquer, segue que $c(x,y) \leq \varphi(x) + \psi(y)$.

(iii) $\implies$ *(i)*. Temos que $\varphi(x) + \varphi^c(y) = c(x,y)$ para todo $(x,y) \in \operatorname{supp}\gamma \subseteq \partial^c\varphi$ e que $\varphi(x) + \varphi^c(y) \leq c(x,y)$ para todo $(x,y) \in X \times Y$. Daí temos que

$$\begin{aligned}
\int c\,d\gamma &= \int (\varphi(x) + \varphi^c(y))\,d\gamma \\
&= \int \varphi(x)\,d\mu + \int \varphi^c(y)\,d\nu \\
&= \int (\varphi(x) + \varphi^c(y))\,d\tilde{\gamma} \leq \int c\,d\tilde{\gamma}.
\end{aligned}$$

Portanto γ é plano de transporte ótimo. □

O Teorema 3.10 nos diz que dado um plano de transporte ótimo γ existe uma função c-côncava φ tal que $\operatorname{supp}\gamma \subseteq \partial^c\varphi$. No entanto, vale algo mais forte do que isso. Se $\tilde{\gamma}$ é outro plano de transporte ótimo, então $\operatorname{supp}\tilde{\gamma} \subseteq \partial^c\varphi$ para a mesma φ. De fato

$$\begin{aligned}
\int \varphi\,d\mu + \int \varphi^c\,d\nu = \int (\varphi(x) + \varphi^c(y))\,d\tilde{\gamma} \leq \int c(x,y)\,d\tilde{\gamma} \\
= \int c(x,y)\,d\gamma = \int (\varphi(x) + \varphi^c(y))\,d\gamma = \int \varphi\,d\mu + \int \varphi^c\,d\nu,
\end{aligned}$$

ou seja, para os pontos de $\operatorname{supp}\tilde{\gamma}$ também vale que $\varphi(x) + \varphi^c(y) = c(x,y)$, portanto $\operatorname{supp}\tilde{\gamma} \subseteq \partial^c\varphi$.

Os resultados mostrados até aqui nesse capítulo são bastante gerais e, de fato, são usados em diversos contextos. Pelo resto do capítulo, no entanto, estamos falando especificamente do caso em que $X = Y = \mathbb{R}^d$.

Lema 3.11. *Seja $c : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$ a função $c(x,y) = \frac{|x-y|^2}{2}$. Se φ é c-côncava, então $\phi(x) := \frac{|x|^2}{2} - \varphi(x)$ é convexa e semicontínua inferiormente. Neste caso $\partial^c\varphi(x) = \partial\phi(x)$, isto é, em x o c-superdiferencial de φ é o subdiferencial da função convexa ϕ.*

Demonstração. Pela definição de c-concavidade, temos que existe ψ tal que

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

Logo $\phi(x) = \frac{|x|^2}{2} - \varphi(x) = \sup_{y\in\mathbb{R}^d}\{\langle x, y\rangle - \left(\frac{|y|^2}{2} - \psi(y)\right)\}$ que é convexa e semicontínua inferiormente por ser supremo de funções afins. $\square$

Como toda função convexa e semicontínua inferiormente é localmente de Lipschitz, é uma consequência do Lema 3.11 que φ c-côncava com $\operatorname{supp}\gamma \subseteq \partial^c\varphi$ dada no Teorema 3.10 é localmente Lipschitz. Sendo localmente Lipschitz, sabemos que é diferenciável em quase todo ponto com relação à medida de Lebesgue. Logo, para quase todo x, $\partial^c\varphi(x)$ é um conjunto unitário. Isto é, se φ é diferenciável em um ponto $\bar{x} \in \mathbb{R}^d$, escrevemos $\bar{y} = \nabla\phi(\bar{x})$ e então, pelo Lema 3.11

$$\begin{aligned}
\phi(\bar{x}) &= \frac{|\bar{x}|^2}{2} - \varphi(\bar{x}) \\
\Longrightarrow\ \nabla\phi(\bar{x}) &= \bar{x} - \nabla\varphi(\bar{x}) \\
\Longrightarrow\ \bar{y} &= \bar{x} - \nabla\varphi(\bar{x}) \\
\Longrightarrow\ \partial^c\varphi(\bar{x}) &= \{\nabla\phi(\bar{x})\} = \{\bar{x} - \nabla\varphi(\bar{x})\}.
\end{aligned}$$

Note que a cadeia de implicações acima vale quase sempre, para a medida de Lebesgue. Se μ é absolutamente contínua em relação à medida de Lebesgue, então $\partial^c\varphi(\bar{x})$ se reduz a um único ponto μ-qtp também.

Teorema 3.12 (Brenier)**.** *Sejam $X = Y = \mathbb{R}^d$ e $c : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$ a função custo quadrática dada por $c(x,y) = \frac{|x-y|^2}{2}$. Se $\mu \in P(\mathbb{R}^d)$ não dá massa para conjuntos pequenos, e $\nu \in P(\mathbb{R}^d)$, então existe uma única aplicação de transporte ótimo $T : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ para o Problema de Monge. Além disso, $T = \nabla\phi$ é o gradiente de uma função convexa semicontínua inferiormente.*

Demonstração. Pelo Teorema 3.5, existe um plano de transporte ótimo $\gamma \in \textsc{Adm}(\mu, \nu)$ para o Problema de Kantorovich associado. Pelo Teorema 3.10 temos que $\operatorname{supp}\gamma$ é c-ciclicamente monótono e que existe $\varphi : \mathbb{R}^d \longrightarrow \mathbb{R} \cup \{-\infty\}$ c-côncava tal que $\operatorname{supp}\gamma \subseteq \partial^c\varphi$. Pelo que foi observado após o Lema 3.11, sabemos que $\partial^c\varphi(x) = \{x - \nabla\varphi(x)\} = \{\nabla\phi(x)\}$ para quase todo ponto em relação à medida de Lebesgue (e portanto μ-qtp também), onde $\phi : \mathbb{R}^d \longrightarrow \mathbb{R}$ é convexa e semicontínua inferiormente. Logo $\operatorname{supp}\gamma \subseteq \operatorname{graf} \nabla\phi$, para $\phi(x) = \frac{|x|^2}{2} - \varphi(x)$ convexa.

Falta mostrar que $T = \nabla\phi$ é uma aplicação de transporte ótimo e que é a única. Note que T é uma aplicação de transporte ótimo porque

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

A primeira igualdade vem de $(\mathrm{Id}, T)_{\#}\mu = \gamma$. A segunda vem da otimalidade de γ para o Problema de Kantorovich. A desigualdade final aparece porque entre os planos de transporte em $\textsc{Adm}(\mu, \nu)$, só um subconjunto deles pode ser visto como uma aplicação de transporte, de forma que o mínimo está sendo avaliado em um conjunto maior do que o conjunto onde está sendo avaliado o ínfimo.

Por fim, T é única. De fato, suponha que T_1 e T_2 são aplicações de transporte ótimas. Então $\gamma_1 = (\mathrm{Id}, T_1)_{\#}\mu$ e $\gamma_2 = (\mathrm{Id}, T_2)_{\#}\mu$ são planos ótimos para o Problema de Kantorovich. Logo, pelo Teorema 3.10 e pela observação seguinte, tanto γ_1 quanto γ_2 tem seus suportes contidos em $\partial\phi$, e a aplicação de transporte obtida a partir daí é $T = \nabla\phi$, portanto $T = T_1 = T_2$. $\square$

Capítulo 4

Teorema de McCann

Seja (M, g) uma variedade Riemanniana compacta sem bordo. Para cada $x \in M$, seja $g_x : T_xM \times T_xM \longrightarrow \mathbb{R}$ um produto interno. Deste produto interno, podemos definir a norma $|v|_x = \sqrt{g_x(v,v)}$ para $v \in T_xM$. Definimos ainda a distância entre dois pontos por $d(x,y) = \inf \int_0^1 |\dot{\gamma}(t)|\, dt$ onde o ínfimo é avaliado entre todas as curvas absolutamente contínuas $\gamma(t)$ com $\gamma(0) = x$ e $\gamma(1) = y$. Em outras palavras, em uma variedade Riemanniana a distância entre dois pontos é o comprimento da menor curva que liga os dois pontos, quando essa curva existe. Esta distância faz de M um espaço métrico, e a topologia de espaço métrico é a mesma que já existia a princípio em M. Como M é compacta, pelo Teorema de Hopf-Rinow temos que entre quaisquer dois pontos x, y existe geodésica que liga os dois pontos, e essa geodésica está definida para todo $t \in \mathbb{R}$.

Nos resultados que seguem, consideraremos a função custo dada por $c(x,y) = \frac{d(x,y)^2}{2}$, onde d é a distância Riemanniana definida acima. $p \in \partial^+\phi(x)$ é a notação para p ser um superdiferencial da função ϕ no ponto x, conforme a Definição 1.2.

Lema 4.1. *Sejam $\phi : M \longrightarrow \mathbb{R}$ e $h : \mathbb{R} \longrightarrow \mathbb{R}$, com $p \in \partial^+\phi(x)$ e $\tau \in \partial^+h(\phi(x))$. Se h é não-decrescente, então $\tau p \in \partial^+(h \circ \phi)(x)$.*

Demonstração. Como $p \in \partial^+\phi(x)$, temos que

$$\phi(\exp_x v) \leq \phi(x) + \langle p, v \rangle_x + o(|v|_x)$$

e, da mesma forma, $\tau \in \partial^+h(\phi(x))$ significa que

$$h(\phi(x) + \varepsilon) \leq h(\phi(x)) + \tau\varepsilon + o(\varepsilon).$$

Fazendo $\varepsilon = \langle p, v\rangle_x + o(|v|_x)$, obtemos que

$$\begin{aligned} h(\phi(\exp_x v)) &\leq h(\phi(x) + \langle p, v\rangle_x + o(|v|_x)) \\ &\leq h(\phi(x)) + \tau\langle p, v\rangle_x + o(|v|_x), \end{aligned}$$

onde, na primeira desigualdade, usamos que h é não-decrescente. Logo, $\tau p \in \partial^+(h \circ \phi)(x)$. □

Proposição 4.2. *Se M é uma variedade compacta, de classe C^∞ e sem bordo, então a função $\psi(x) = \frac{d^2(x,y)}{2}$ é semicôncava para cada y fixado.*

Demonstração. Sejam $\sigma : [0,1] \longrightarrow M$ a geodésica minimizante com $\sigma(0) = y$ e $\sigma(1) = x$, U uma vizinhança completamente normal de x e $w \in T_xM$ tal que $\exp_x w \in U$. Primeiro tratamos o caso em que $y \in U$.

$$\begin{aligned} \psi(\exp_x w) &= \frac{d^2(y, \exp_y(\exp_y^{-1}\exp_x w))}{2} \\ &= \frac{|\exp_y^{-1}\exp_x w|_y^2}{2}, \end{aligned} \tag{4.1}$$

pois $\exp_y^{-1}\exp_x w$ é a velocidade inicial da geodésica ligando os pontos y e $\exp_y(\exp_y^{-1}\exp_x w))$, e geodésicas tem velocidade de módulo constante. No que segue, vamos definir a função $f(w) = \exp_y^{-1}\exp_x w$ para simplificar. A notação é carregada mas a ideia é simples. Expandimos f em torno de 0,

$$f(w) = f(0) + f'(0) \cdot w + o(|w|_x)$$

ou seja

$$\begin{aligned} \exp_y^{-1}\exp_x w &= \exp_y^{-1}\exp_x 0 + d(\exp_y^{-1})(\exp_x 0) \cdot (d(\exp_x)(0)) \cdot w + o(|w|_x) \\ &= \dot\sigma(0) + d(\exp_y^{-1})(x) \cdot w + o(|w|_x), \end{aligned}$$

pois $d(\exp_x)(0) = \mathrm{Id}$. Logo

$$\begin{aligned} \frac{|\exp_y^{-1}\exp_x w|_y^2}{2} &= \frac{|\dot\sigma(0) + d(\exp_y^{-1})(x) \cdot w + o(|w|_x)|_y^2}{2} \\ &= \frac{|\dot\sigma(0)|_y^2}{2} + \langle\dot\sigma(0), d(\exp_y^{-1})(x) \cdot w\rangle_y + o(|w|_x). \end{aligned} \tag{4.2}$$

Aplicando a regra da cadeia para $f \circ f^{-1}(x) = x$, obtemos que

$$df(f^{-1}(x)) \cdot d(f^{-1})(x) \cdot w = w$$

e que portanto

$$d(f^{-1})(x) \cdot w = [df(f^{-1}(x))]^{-1} \cdot w,$$

ou seja,

$$\begin{aligned} d(\exp_y^{-1})(x) \cdot w &= [d(\exp_y)(\exp_y^{-1} x)]^{-1} \cdot w \\ &= [d(\exp_y)(\dot\sigma(0))]^{-1} \cdot w. \end{aligned}$$

Por fim, note que $\dot\sigma(1) = d(\exp_y)(\dot\sigma(0)) \cdot \dot\sigma(0)$. Logo, pelo Lema de Gauss temos que

$$\begin{aligned} \langle \dot\sigma(0), d(\exp_y^{-1})(x) \cdot w \rangle_y &= \langle \dot\sigma(0), [d(\exp_y)(\dot\sigma(0))]^{-1} \cdot w \rangle_y \\ &= \langle d(\exp_y)(\dot\sigma(0)) \cdot \dot\sigma(0), w \rangle_x \qquad (4.3) \\ &= \langle \dot\sigma(1), w \rangle_x. \end{aligned}$$

Juntando as Equações (4.1), (4.2) e (4.3), e vendo que $\frac{|\dot\sigma(0)|_y^2}{2} = \frac{d^2(y,x)}{2}$, concluimos que

$$\psi(\exp_x w) = \psi(x) + \langle \dot\sigma(1), w \rangle_y + o(|w|_x).$$

Isto é, se y está numa vizinhança totalmente normal de x, então a função ψ é diferenciável. Tratamos agora do caso em que $y \notin U$.

Como $d(x,y) = \sqrt{2\psi(x)}$, pela regra da cadeia obtemos que $\nabla_x d(x,y) = \frac{\dot\sigma(1)}{|\dot\sigma(1)|}$. Seja $z \in U$ um ponto que pertença à geodésica σ. Temos que $\nabla_x d(x,z) = \frac{\dot\sigma(1)}{|\dot\sigma(1)|}$, isto é,

$$d(z, \exp_x v) = d(x,z) + \left\langle \frac{\dot\sigma(1)}{|\dot\sigma(1)|_x}, v \right\rangle_x + o(|v|_x).$$

Usando a desigualdade triangular e a igualdade acima, obtemos

$$\begin{aligned} d(y, \exp_x v) &\leq d(y,z) + d(z, \exp_x v) \\ &\leq d(y,z) + d(x,z) + \left\langle \frac{\dot\sigma(1)}{|\dot\sigma(1)|_x}, v \right\rangle_x + o(|v|_x) \\ &= d(y,x) + \left\langle \frac{\dot\sigma(1)}{|\dot\sigma(1)|_x}, v \right\rangle_x + o(|v|_x) \end{aligned}$$

pois z pertence à geodésica ligando y a x e portanto $d(y,x) = d(y,z)+d(z,x)$. Segue que $\frac{\dot{\sigma}(1)}{|\dot{\sigma}(1)|_x}$ é um superdiferencial da função $d(y,\cdot)$ em x. Aplicando o Lema 4.1 com $h(d) = \frac{d^2}{2}$ e $\phi(x) = d(y,x)$, concluimos que

$$\begin{aligned}\frac{d^2(y,\exp_x v)}{2} &\leq \frac{d^2(y,x)}{2} + d(x,y)\left\langle \frac{\dot{\sigma}(1)}{|\dot{\sigma}(1)|_x}, v\right\rangle_x + o(|v|_x)\\ &= \frac{d^2(y,x)}{2} + \langle \dot{\sigma}(1), v\rangle_x + o(|v|_x)\end{aligned}$$

e daí segue que a função $\frac{d^2(\cdot,y)}{2}$ é semicôncava, com um superdiferencial em x dado pela velocidade no ponto x da geodésica que liga y a x. □

O Lema 3.11 nos dá uma boa equivalência para a existencia de uma função c-côncava quando estamos em $\mathbb{R}^d$. Como agora estamos em uma variedade, o que garante a diferenciabilidade em quase toda parte da função c-côncava é a proposição a seguir.

Proposição 4.3. *Sejam M uma variedade Riemanniana compacta, de classe C^∞ e sem bordo, e $\varphi : M \longrightarrow \mathbb{R}\cup\{-\infty\}$ uma função c-côncava que não é identicamente igual a $-\infty$. Então φ é Lipschitz, semicôncava e só assume valores reais. Além disso, para $y \in \partial^c\varphi(x)$, $\exp_x^{-1}(y) \subseteq -\partial^+\varphi(x)$. De modo inverso, se φ é diferenciável em x, então $\exp_x(-\nabla\varphi(x)) \in \partial^c\varphi(x)$.*

Demonstração. Provamos primeiro que φ assume valores reais. Como φ é c-côncava, existe $\phi : M \longrightarrow \mathbb{R}\cup\{-\infty\}$ tal que $\varphi = \phi^c$. Isto é, $\varphi(x) = \inf_{y\in M}\{\frac{d(x,y)^2}{2} - \phi(y)\}$. A função ϕ não pode ser igual a $-\infty$ em todos pontos, porque nesse caso φ seria igual a $+\infty$ em todos os pontos, e esse valor nem está em seu contradomínio. Logo $\phi < +\infty$ em algum ponto. Por esse fato e também porque $\frac{d^2}{2}$ é uniformemente limitada em M (pois é uma função contínua definida em um compacto), segue que φ assume valores reais.

Como M é compacta e de classe C^∞, as funções $\frac{d(\cdot,y)^2}{2}$ são uniformemente Lipschitz. De fato,

$$\frac{d(x,y)^2}{2} - \frac{d(z,y)^2}{2} = \frac{1}{2}(d(x,y)+d(z,y))(d(x,y)-d(z,y)) \leq Cd(x,z)$$

onde $C = \max_{x,y\in M} d(x,y)$. O mesmo pode ser feito trocando os papéis de x e z, logo vale que

$$\left|\frac{d(x,y)^2}{2} - \frac{d(z,y)^2}{2}\right| \leq Cd(x,z)$$

e como a mesma constante C vale para todo $y \in M$, as funções são uniformemente Lipschitz. Além disso, pela Proposição 4.2, essas funções também são uniformemente semicôncavas.

Definindo $h_y(x) = \frac{d(x,y)^2}{2} - \phi(y)$, temos que $\varphi(x) = \inf_{y \in M} h_y(x)$. Como as funções $\frac{d(\cdot,y)^2}{2}$ são semicôncavas, as funções h_y também são. De fato,

$$\begin{aligned} h_y(z) - h_y(x) &= \frac{d(z,y)^2}{2} - \phi(y) - \frac{d(x,y)^2}{2} + \phi(y) \\ &= \frac{d(z,y)^2}{2} - \frac{d(x,y)^2}{2}. \end{aligned}$$

Da compacidade de M segue, pelo Corolário 1.18, que φ é localmente semicôncava.

Sejam $y \in \partial^c \varphi(x)$ e $v \in \exp_x^{-1}(y)$. Da Proposição 4.2 sabemos que $-v$ pertence ao superdiferencial de $\frac{d(\cdot,y)^2}{2}$ em x, ou seja,

$$\frac{d(z,y)^2}{2} - \frac{d(x,y)^2}{2} \leq \langle -v, \exp_x^{-1}(z) \rangle + o(d(x,z)).$$

Como $y \in \partial^c \varphi(x)$, para esse y vale que $\varphi(x) = \frac{d(x,y)^2}{2} - \varphi^c(y)$ e para todo $z \in M$ vale que $\varphi(z) \leq \frac{d(z,y)^2}{2} - \varphi^c(y)$. Logo

$$\begin{aligned} \varphi(z) - \varphi(x) &\leq \frac{d(z,y)^2}{2} - \varphi^c(y) - \frac{d(x,y)^2}{2} + \varphi^c(y) \\ = \frac{d(z,y)^2}{2} - \frac{d(x,y)^2}{2} &\leq \langle -v, \exp_x^{-1}(z) \rangle + o(d(x,z)). \end{aligned}$$

Disso segue que $-v \in \partial^+ \varphi(x)$.

Ainda falta mostrar que o c-superdiferencial de φ no ponto x não é vazio. Como φ é c-côncava, existe uma função ϕ tal que $\varphi(x) = \inf_{y \in M} \frac{d(x,y)^2}{2} - \phi(y)$, e portanto podemos construir uma sequência $(y_n) \in M$ tal que $\varphi(x) = \lim_{n \to \infty} \frac{d(x,y_n)^2}{2} - \phi(y_n)$. Pela compacidade de M, podemos extrair uma subsequência convergente. Seja y o limite dessa subsequência. Pela continuidade de $\frac{d(z,\cdot)^2}{2}$ e de ϕ seque que $\varphi(x) = \frac{d(x,y)^2}{2} - \phi(y)$. Logo $y \in \partial^c \varphi(x)$. □

Podemos então mostrar o principal resultado desse capítulo, análogo ao Teorema de Brenier nesse contexto mais geral.

Teorema 4.4 (McCann). *Sejam M uma variedade Riemanniana de classe C^∞, compacta, sem bordo, e $\mu \in P(M)$ tal que μ dá medida nula para* *conjuntos pequenos. Então para toda $\nu \in P(M)$ existe (a menos de um conjunto de medida nula) apenas um plano de transporte ótimo de μ para ν e esse plano é induzido por um mapa T. Além disso, esse mapa T pode ser escrito como $x \mapsto \exp_x(-\nabla\varphi(x))$ para alguma função c-côncava $\varphi : M \longrightarrow \mathbb{R}$.*

Demonstração. A função custo é contínua, então ela é em particular semicontínua inferiormente. Como além disso ela é limitada por baixo, segue do Teorema 3.5 que existe plano de transporte ótimo.

Observe que, numa variedade compacta, $d^2(\cdot,\cdot)/2$ é limitada. Logo, a condição (3.2) é imediata, e pelo Teorema 3.10, se γ é um plano de transporte ótimo, então $\operatorname{supp}\gamma \subseteq \partial^c\varphi$ para alguma função c-côncava φ. Da Proposição 4.3 segue que φ é localmente semicôncava, portanto é diferenciável μ-qtp x (pois μ dá medida nula para conjuntos pequenos e sabemos que o conjunto em que φ não é diferenciável é pequeno).

Disso segue que o mapa $T(x) := \exp_x(-\nabla\varphi(x))$ está bem definido e seu gráfico deve ser de medida total para qualquer γ plano de transporte ótimo. $\square$

Capítulo 5

Custo advindo de um Lagrangiano

O Teorema de Brenier assegura a existência e a unicidade da aplicação de transporte ótimo quando o custo é dado pelo quadrado da distância Euclidiana. Essa aplicação é, basicamente, mover a massa contida no ponto x na direção do gradiente de uma função c-côncava. O Teorema de McCann, por sua vez, nos dá a existência e unicidade de aplicação de transporte ótimo quando o custo é dado pelo quadrado da distância Riemanniana em uma variedade compacta. Nesse caso, a aplicação se resume a mover a massa do ponto x na direção do gradiente de uma função, mas dessa vez ao longo de uma geodésica da variedade. Isto está de acordo com o Teorema de Brenier porque no espaço Euclidiano as geodésicas são retas.

Neste capítulo vamos considerar o Problema do Transporte Ótimo em um contexto ainda mais geral. Aqui o custo associado aos pontos x, y é o valor mínimo, dentre todas as curvas com extermidades x e y, da ação de um Lagrangiano. Note que isso generaliza o caso do Teorema de McCann porque o custo considerado lá pode ser encarado como o valor mínimo atingido pela ação do Lagrangiano $L(x,v) = \frac{|v|_x^2}{2}$.

Começamos definindo o novo custo, e mostramos que esse custo tem as propriedades necessárias para garantir a existência e unicidade de uma aplicação de transporte ótimo.

Definição 5.1. Seja $L : TM \longrightarrow \mathbb{R}$ um Lagrangiano limitado por baixo em uma variedade conexa M. Para $t > 0$, definimos o custo $c_{t,L} : M \times M \longrightarrow \mathbb{R}$ por $c_{t,L}(x,y) = \inf \mathbb{A}_L(\gamma)$ onde o ínfimo é avaliado sobre todas curvas absolutamente contínuas $\gamma : [0,t] \longrightarrow M$ com $\gamma(0) = x$ e $\gamma(t) = y$.

Observe que, através de uma mudança de variável na integral que define a ação $\mathbb{A}_L$, temos $c_{t,L} = c_{1,L^t}$, onde $L^t(x,v) = tL(x, \frac{v}{t})$.

Como a função custo definida acima é contínua (por consequência do Teorema 5.2) e limitada por baixo, segue do Teorema 3.5 que existe minimizante para o Problema de Kantorovich. E do Teorema 3.10 sabemos que se existe um plano de transporte ótimo, existe uma função c-côncava $\varphi : M \longrightarrow \mathbb{R} \cup \{-\infty\}$ tal que seu c-superdiferencial contem o suporte de qualquer plano de transporte ótimo.

O c-superdiferencial da função φ é o conjunto $\partial^c\varphi = \{(x,y) \in M \times M : \varphi(x) + \varphi^c(y) = c(x,y)\}$. Vale, em geral (isto é, em pontos arbitrários da variedade), que $\varphi(x) + \varphi^c(y) \leq c(x,y)$. Dessa forma, se γ é um plano de transporte ótimo, os pontos de $\operatorname{supp}\gamma \subseteq \partial^c\varphi$ são pontos em que a função $h(x,y) = \varphi(x) + \varphi^c(y) - c(x,y)$ atinge máximo, e portanto tem derivada nula. Isto é, se tiver derivada. Mostramos a seguir que podemos, de fato, diferenciar esta função em quase todo ponto.

Teorema 5.2. *Seja $L : TM \longrightarrow \mathbb{R}$ um Lagrangiano de Tonelli fraco definido em uma variedade compacta M. Então, para todo $t > 0$, o custo $c_{t,L}$ é uma função localmente semicôncava em $M \times M$. Além disso, um superdiferencial de $c_{t,L}$ em (x_0, y_0) é dado por*

$$(w_0, w_1) \mapsto \frac{\partial L}{\partial v}(\gamma(t), \dot\gamma(t))(w_1) - \frac{\partial L}{\partial v}(\gamma(0), \dot\gamma(0))(w_0)$$

onde $\gamma : [0,t] \longrightarrow M$ é um L-minimizante com $\gamma(0) = x_0$, $\gamma(t) = y_0$, e $(w_0, w_1) \in T_xM \times T_yM = T_{(x,y)}(M \times M)$.

Demonstração. Mostramos para o caso em que $M = \mathbb{R}^n$. A prova do caso geral é uma adaptação desta, e pode ser encontrada em [6, Theorem B.19]. Pelo que foi observado, basta mostrar para o custo $c := c_{1,L}$.

Sejam $x, y \in \mathbb{R}^n$. Mostramos a semiconcavidade de c em uma vizinhança de (x,y). Seja γ a curva que atinge o ínfimo na definição de $c(x,y)$, que existe pelo Teorema 2.5. Isto é, γ é tal que

$$c(x,y) = \int_0^1 L(\gamma(s), \dot\gamma(s))\, ds.$$

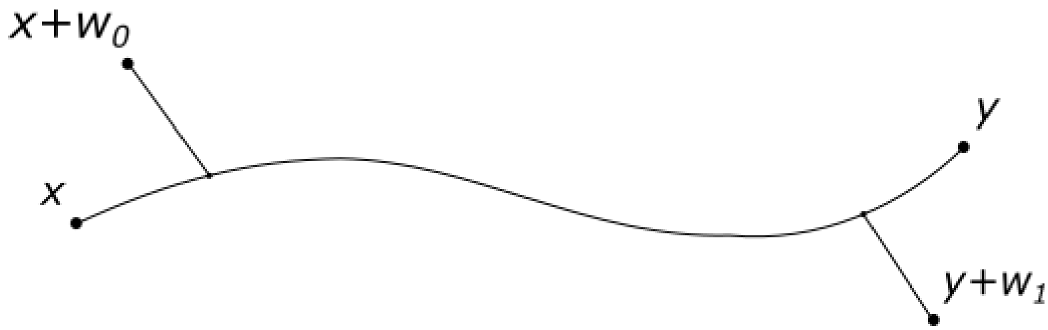

Figura 5.1: As curvas γ e $\tilde{\gamma}$.

Definimos uma segunda curva $\tilde{\gamma} : [0,1] \longrightarrow \mathbb{R}^n$, a partir de γ:

$$\tilde{\gamma}(s) = \begin{cases} \frac{\varepsilon - s}{\varepsilon} w_0 + \gamma(s), & s \in [0, \varepsilon) \\ \gamma(s), & s \in [\varepsilon, 1-\varepsilon) \\ \frac{s-(1-\varepsilon)}{\varepsilon} w_1 + \gamma(s), & s \in [1-\varepsilon, 1] \end{cases}$$

onde $w_0, w_1 \in \mathbb{R}^n = T_xM = T_yM$ e $0 < \varepsilon < \frac{1}{2}$.

Note que $\tilde{\gamma}(0) = w_0 + x$ e $\tilde{\gamma}(1) = w_1 + y$. Portanto, $\tilde{\gamma}$ entra na avaliação do ínfimo que define $c(w_0 + x, w_1 + y)$, de forma que

$$\begin{aligned} c(w_0 + x, w_1 + y) &\leq \int_0^1 L(\tilde{\gamma}(s), \dot{\tilde{\gamma}}(s))\, ds \\ &= \int_0^\varepsilon L\left(\frac{\varepsilon - s}{\varepsilon} w_0 + \gamma(s), -\frac{w_0}{\varepsilon} + \dot{\gamma}(s)\right) ds \\ &+ \int_\varepsilon^{1-\varepsilon} L\left(\gamma(s), \dot{\gamma}(s)\right) ds \\ &+ \int_{1-\varepsilon}^1 L\left(\frac{s-(1-\varepsilon)}{\varepsilon} w_1 + \gamma(s), \frac{w_1}{\varepsilon} + \dot{\gamma}(s)\right) ds \end{aligned}$$

e podemos concluir que, como γ é L-minimizante,

$$
\begin{aligned}
c(w_0 + x, &w_1 + y) - c(x,y) \leq \\
&\int_0^\varepsilon \left[L\left(\frac{\varepsilon - s}{\varepsilon} w_0 + \gamma(s), -\frac{w_0}{\varepsilon} + \dot\gamma(s) \right) - L(\gamma(s), \dot\gamma(s)) \right] ds \\
&+ \int_{1-\varepsilon}^1 \left[L\left(\frac{s - (1-\varepsilon)}{\varepsilon} w_1 + \gamma(s), \frac{w_1}{\varepsilon} + \dot\gamma(s) \right) - L(\gamma(s), \dot\gamma(s)) \right] ds.
\end{aligned} \tag{5.1}
$$

O termo do meio sumiu porque foi cancelado com o que veio de $c(x, y)$.

Da definição de derivada de L obtemos que para cada $s \in [0, 1]$ existe um módulo de continuidade ω_s tal que

$$
\begin{aligned}
L(z + \gamma(s), v + \dot\gamma(s)) &- L(\gamma(s), \dot\gamma(s)) \\
&= DL(\gamma(s), \dot\gamma(s))(z, v) + |(z, v)|\omega_s(|(z, v)|).
\end{aligned}
$$

Seja $K \subseteq \mathbb{R}^n$ um compacto tal que $x, y \in K$. Pela Proposição 2.17, a curva $s \mapsto (\gamma(s), \dot\gamma(s))$ fica contida em um compacto $\tilde K$. Por isso, podemos substituir ω_s por um módulo de continuidade ω para DL que funciona para todo s. Assim, a desigualdade (5.1) implica

$$
\begin{aligned}
c(w_0 + x, w_1 + y) - c(x,y) \leq &\int_0^\varepsilon DL(\gamma(s), \dot\gamma(s)) \left(\frac{\varepsilon - s}{\varepsilon} w_0, -\frac{w_0}{\varepsilon} \right) ds \\
&+ \int_0^\varepsilon \left| \left(\frac{\varepsilon - s}{\varepsilon} w_0, -\frac{w_0}{\varepsilon} \right) \right| \omega \left(\left| \left(\frac{\varepsilon - s}{\varepsilon} w_0, -\frac{w_0}{\varepsilon} \right) \right| \right) ds \\
&+ \int_{1-\varepsilon}^1 DL(\gamma(s), \dot\gamma(s)) \left(\frac{s - (1-\varepsilon)}{\varepsilon} w_1, \frac{w_1}{\varepsilon} \right) ds \\
&+ \int_{1-\varepsilon}^1 \left| \left(\frac{s - (1-\varepsilon)}{\varepsilon} w_1, \frac{w_1}{\varepsilon} \right) \right| \omega \left(\left| \left(\frac{s - (1-\varepsilon)}{\varepsilon} w_1, \frac{w_1}{\varepsilon} \right) \right| \right) ds.
\end{aligned}
$$

Note que $\left|\left(\frac{\varepsilon-s}{\varepsilon} w_0, -\frac{w_0}{\varepsilon}\right)\right| \leq \left|\left(\frac{\varepsilon}{\varepsilon} w_0, -\frac{w_0}{\varepsilon}\right)\right| \leq \left|\left(\frac{1}{\varepsilon} w_0, -\frac{w_0}{\varepsilon}\right)\right| = \frac{\sqrt{2}}{\varepsilon}|w_0|$, e similarmente $|\frac{s-(1-\varepsilon)}{\varepsilon} w_1, \frac{w_1}{\varepsilon}| \leq \frac{\sqrt{2}}{\varepsilon}|w_1|$. Além disso, temos que $|w_0| \leq$

$|(w_0, w_1)|$, o mesmo acontecendo com w_1. Logo, ficamos com

$$\begin{aligned} c(w_0 + x, w_1 + y) - c(x, y) \leq & \int_0^\varepsilon DL(\gamma(s), \dot\gamma(s)) \left(\frac{\varepsilon - s}{\varepsilon} w_0, -\frac{w_0}{\varepsilon} \right) ds \\ & + \int_{1-\varepsilon}^1 DL(\gamma(s), \dot\gamma(s)) \left(\frac{s - (1-\varepsilon)}{\varepsilon} w_1, \frac{w_1}{\varepsilon} \right) ds \\ & + 2\sqrt{2}\, |(w_0, w_1)|\, \omega \left(\frac{\sqrt{2}}{\varepsilon} |(w_0, w_1)| \right). \end{aligned} \tag{5.2}$$

Observe que a soma dos dois primeiros termos no lado direito da desigualdade (5.2) é linear em (w_0, w_1), e que o último termo é um módulo de semiconcavidade. Como esse módulo foi escolhido de modo a funcionar em um compacto $\tilde{K}$, e como $x, y \in \mathbb{R}^n$ foram escolhidos arbitrariamente, concluimos que a função c é localmente semicôncava em $\mathbb{R}^n \times \mathbb{R}^n$.

Para calcular o superdiferencial, basta calcular explicitamente as integrais da equação (5.2). Temos que

$$\begin{aligned} DL(\gamma(s), \dot\gamma(s)) \left(\frac{\varepsilon - s}{\varepsilon} w_0, -\frac{w_0}{\varepsilon} \right) = & \frac{\partial L}{\partial x}(\gamma(s), \dot\gamma(s)) \cdot \left(\frac{\varepsilon - s}{\varepsilon} w_0 \right) \\ & + \frac{\partial L}{\partial v}(\gamma(s), \dot\gamma(s)) \cdot \left(-\frac{w_0}{\varepsilon} \right). \end{aligned}$$

Usando a linearidade para passar $\frac{\varepsilon - s}{\varepsilon}$ e $\frac{1}{\varepsilon}$ para fora, e aplicando a equação de Euler-Lagrange (Teorema 2.6), ficamos com

$$\begin{aligned} DL(\gamma(s), \dot\gamma(s)) \left(\frac{\varepsilon - s}{\varepsilon} w_0, -\frac{w_0}{\varepsilon} \right) = & \frac{\varepsilon - s}{\varepsilon} \frac{d}{ds} \frac{\partial L}{\partial v}(\gamma(s), \dot\gamma(s)) \cdot w_0 \\ & - \frac{1}{\varepsilon} \frac{\partial L}{\partial v}(\gamma(s), \dot\gamma(s)) \cdot w_0. \end{aligned}$$

Então podemos resolver a primeira integral em (5.2) por partes, como segue:

$$\begin{aligned}
&\int_0^{\varepsilon} \left[\frac{\varepsilon - s}{\varepsilon}\frac{d}{ds}\frac{\partial L}{\partial v}(\gamma(s),\dot\gamma(s))\cdot w_0 - \frac{1}{\varepsilon}\frac{\partial L}{\partial v}(\gamma(s),\dot\gamma(s))\cdot w_0\right] ds\\
&= \left(\frac{\varepsilon - s}{\varepsilon}\right)\frac{\partial L}{\partial v}(\gamma(s),\dot\gamma(s))\cdot w_0\Big|_0^{\varepsilon} + \frac{1}{\varepsilon}\int_0^{\varepsilon}\frac{\partial L}{\partial v}(\gamma(s),\dot\gamma(s))\cdot w_0\, ds\\
&\quad - \frac{1}{\varepsilon}\int_0^{\varepsilon}\frac{\partial L}{\partial v}(\gamma(s),\dot\gamma(s))\cdot w_0\, ds\\
&= -\frac{\partial L}{\partial v}(\gamma(0),\dot\gamma(0))\cdot w_0.
\end{aligned}$$

Agora fazemos o mesmo para a segunda integral. Pelos mesmos argumentos de linearidade e Euler-Lagrange, temos que

$$\begin{aligned}
DL(\gamma(s),\dot\gamma(s))\left(\frac{s-(1-\varepsilon)}{\varepsilon}w_1, \frac{w_1}{\varepsilon}\right) &= \frac{s-(1-\varepsilon)}{\varepsilon}\frac{d}{ds}\frac{\partial L}{\partial v}(\gamma(s),\dot\gamma(s))\cdot w_1\\
&\quad + \frac{1}{\varepsilon}\frac{\partial L}{\partial v}(\gamma(s),\dot\gamma(s))\cdot w_1.
\end{aligned}$$

E resolvemos a segunda integral em (5.2) por partes, como segue:

$$\begin{aligned}
&\int_{1-\varepsilon}^{1}\left[\frac{s-(1-\varepsilon)}{\varepsilon}\frac{d}{ds}\frac{\partial L}{\partial v}(\gamma(s),\dot\gamma(s))\cdot w_1 + \frac{1}{\varepsilon}\frac{\partial L}{\partial v}(\gamma(s),\dot\gamma(s))\cdot w_1\right] ds\\
&= \frac{s-(1-\varepsilon)}{\varepsilon}\frac{\partial L}{\partial v}(\gamma(s),\dot\gamma(s))\cdot w_1\Big|_{1-\varepsilon}^{1} - \frac{1}{\varepsilon}\int_{1-\varepsilon}^{1}\frac{\partial L}{\partial v}(\gamma(s),\dot\gamma(s))\cdot w_1\, ds\\
&\quad + \frac{1}{\varepsilon}\int_{1-\varepsilon}^{1}\frac{\partial L}{\partial v}(\gamma(s),\dot\gamma(s))\cdot w_1\, ds\\
&= \frac{\partial L}{\partial v}(\gamma(1),\dot\gamma(1))\cdot w_1.
\end{aligned}$$

Logo, um superdiferencial de $c_{1,L}$ em (x,y) é dado por

$$(w_0,w_1)\mapsto \frac{\partial L}{\partial v}(\gamma(1),\dot\gamma(1))(w_1) - \frac{\partial L}{\partial v}(\gamma(0),\dot\gamma(0))(w_0). \qquad \square$$

Como φ é c-côncava, existe ϕ tal que $\varphi(x) = \inf_{y\in M} h_y(x)$ onde $h_y(x) = c(x,y) - \phi(y)$. Como c é localmente semicôncava, h_y também é, para cada y fixado. Como M é compacta, segue do Corolário 1.18 que φ é localmente semicôncava. Pelo Corolário 1.7, já que φ e c são localmente semicôncavas, então são diferenciáveis em quase toda parte. Além disso, nos pontos em que existe, o diferencial é igual ao superdiferencial. Então, em particular, a Proposição 5.2 mostra que nesses pontos $\frac{\partial}{\partial x}c(x,y) = -\frac{\partial L}{\partial v}(\gamma(0),\dot{\gamma}(0))$, onde γ é um L-minimizante entre x e y.

Dessa forma, podemos fazer a conta a seguir sem ter nada a temer. Para $(x,y) \in \operatorname{supp}\gamma$ a função h atinge máximo e portanto

$$h(x,y) = \varphi(x) + \varphi^c(y) - c(x,y)$$

$$\implies 0 = \frac{\partial}{\partial x}h(x,y) = \nabla\varphi(x) - \frac{\partial}{\partial x}c(x,y)$$

$$\implies \nabla\varphi(x) = \frac{\partial}{\partial x}c(x,y).$$

Se $\frac{\partial}{\partial x}c(x,y)$ também fosse diferenciável poderíamos aplicar o Teorema da Função Implícita aqui para determinar unicamente y a partir de uma função de x, resolvendo o Problema de Monge. Mas não queremos ter que exigir demais de c. Injetividade em y é o suficiente.

De posse de todas essas informações, podemos caracterizar a aplicação de transporte ótima nesse contexto.

Teorema 5.3. *Se L é um Lagrangiano de Tonelli definido em uma variedade compacta M, então para qualquer $t > 0$, o custo $c_{t,L} : M \times M \longrightarrow \mathbb{R}$ é tal que $\frac{\partial}{\partial x}c(x,y)$ é injetiva em y nos pontos em que essa derivada existe. Portanto, existe uma única (a menos de conjunto de medida nula) aplicação de transporte ótimo dada por $y = T(x) = \pi\phi_t(x,v)$, onde $\pi : TM \longrightarrow M$ é a projeção canônica, ϕ_t é o fluxo de Euler Lagrange de L, e $v \in T_xM$ é unicamente determinado pela equação*

$$\frac{\partial c}{\partial x}(x,y) = -\frac{\partial L}{\partial v}(x,v).$$

Além disso, a curva $\gamma : [0,t] \longrightarrow M$ dada por $\gamma(s) = \pi\phi_s(x,v)$ é o único L-minimizante com $\gamma(0) = x$ e $\gamma(t) = y$.

Demonstração. Se L é um Lagrangiano de Tonelli, pelo Teorema 2.12 o fluxo de Euler Lagrange ϕ_t é tal que a curva $\gamma(s) = \pi\phi_s(x,v)$ é um L-minimizante. Como toda curva velocidade de um L-minimizante está contida em uma órbita do fluxo de Euler-Lagrange, temos que a seguinte condição é satisfeita:

se $\gamma_i : [a_i, b_i] \longrightarrow M$, $i = 1, 2$ são dois L-minimizantes tais que $\gamma_1(t_0) = \gamma_2(t_0)$ e $\dot{\gamma_1}(t_0) = \dot{\gamma_2}(t_0)$ para algum $t_0 \in [a_1, b_1] \cap [a_2, b_2]$, então $\gamma_1 = \gamma_2$ em $[a_1, b_1] \cap [a_2, b_2]$.

Agora provamos que a velocidade $\dot{\gamma}(0)$ é unicamente determinada pela igualdade

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, v).$$

Seja $\gamma : [0, t] \longrightarrow M$ com $x = \gamma(0)$ e $y = \gamma(t)$ um L-minimizante. Do Teorema 5.2 temos

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, \dot{\gamma}(0)).$$

Como o mapa de classe C^1 $v \mapsto L(x, v)$ é estritamente convexo (porque L é um Lagrangiano de Tonelli fraco), a transformada de Legendre $v \in T_xM \mapsto \frac{\partial L}{\partial v}(x, v)$ é injetiva e portanto $\dot{\gamma}(0)$ é unicamente determinada pela equação.

Além disso, existe uma única curva $\gamma : [0, t] \longrightarrow M$ L-minimizante com $x = \gamma(0)$ e $y = \gamma(t)$. De fato, seja $\gamma_1 : [0, t] \longrightarrow M$ outra curva L-minimizante com $\gamma_1(0) = x$. Pelo que provamos acima,

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, \dot{\gamma_1}(0)),$$

e da unicidade temos $\dot{\gamma}(0) = \dot{\gamma_1}(0)$. Segue da condição dada no início da prova que $\gamma = \gamma_1$ no intervalo todo.

Por fim, provamos a injetividade em y de $\frac{\partial c_{t,L}}{\partial x}(x, y)$. Sejam (x, y) e (x, y_1) pontos onde $\frac{\partial c_{t,L}}{\partial x}$ existe, tais que

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = \frac{\partial c_{t,L}}{\partial x}(x, y_1).$$

Pelo visto acima, existe uma única γ L-minimizante tal que $x = \gamma(0)$, $y = \gamma(t)$ e

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, \dot{\gamma}(0)).$$

Respectivamente, existe uma única γ_1 L-minimizante tal que $x = \gamma_1(0)$ e $y_1 = \gamma_1(t)$ e

$$\frac{\partial c_{t,L}}{\partial x}(x, y_1) = -\frac{\partial L}{\partial v}(x, \dot{\gamma_1}(0)).$$

Mas então

$$-\frac{\partial L}{\partial v}(x, \dot{\gamma}(0)) = -\frac{\partial L}{\partial v}(x, \dot{\gamma_1}(0)),$$

e da injetividade da transformada de Legendre temos que $\dot{\gamma}(0) = \dot{\gamma}_1(0)$. Da condição dada no início da prova concluimos que $\gamma = \dot{\gamma}$. Em particular, $y = \gamma(t) = \gamma_1(t) = y_1$. $\square$

Note que enquanto no Teorema de McCann a aplicação consistia em seguir geodésicas (seguindo o fluxo geodésico), aqui a solução é seguir o fluxo Lagrangiano.

Considerações Finais

O caso em que a variedade M não é necessariamente compacta foi tratado por Fathi e Figalli [6]. Sem a compacidade, não se sabe se vale a semiconcavidade de uma função c-côncava. Para contornar isso, eles utilizam diferenciabilidade aproximada. Em [13] Figalli e Gigli mostram que vale semiconvexidade para os potenciais de Kantorovich associados ao custo dado pela distância Riemanniana quadrática, sem assumir nenhuma hipótese de compacidade ou sobre a curvatura de M.

Também se mostra em [13] esse resultado para um custo definido a partir de um Lagrangiano, porém com mais condições no Lagrangiano. Seria interessante verificar se a semiconcavidade local do custo definido por um Lagrangiano é válida também no caso não compacto, porque isto faria com que a prova usual que apresentamos no Teorema 5.3 fosse válida mesmo para o caso em que M não é compacto.

Bibliografia

[1] G. Monge. *Mémoire sur la théorie des déblais et des remblais*. Histoire de l'Académie Royale des Sciences de Paris, páginas 666-704, 1781.

[2] L. V. Kantorovich. *On the translocation of masses*. C. R. (Dokl.) Acad. Sci. URSS, páginas 225-226. 1942.

[3] L. Buriol, M. Ritt. *Otimização Combinatória*. Notas de aula. Disponível em: https://www.inf.ufrgs.br/ MRPRITT/lib/exe/fetch.php?media= inf05010:notas-9348.pdf (acesso em 18/07/2019).

[4] Y. Brenier. *Decomposition polaire et réarrangement monotone des champs de vecteurs*. C. R. Acad. Sci. Paris Ser. I Math., 305(19):805-808, 1987.

[5] R. J. McCann. *Polar factorization of maps on Riemannian manifolds*. Geometric and Functional Analysis, 11, páginas 589-608, 2001.

[6] A. Fathi, A. Figalli. *Optimal Transportation on Non-Compact Manifolds*. Israel Journal of Mathematics, páginas 1-59, 2010.

[7] P. Cannarsa, C. Sinestrari. *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control*. Progress in Nonlinear Differential Equations and Their Applications 58, Birkhäuser, Boston, 2004.

[8] C. Villani. *Topics in Optimal Transportation*, volume 58 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2003.

[9] D. Marcon. *Transporte Ótimo*. Notas de aula, 2018.

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

[11] L. Ambrosio, N. Gigli. *A user's guide to optimal transport*. Notas de aula, 2010.

[12] F. Santambrogio. *Optimal transport for applied mathematicians; Calculus of Variations, PDEs and modeling*, volume 87 of Progress in Nonlinear Differential Equations and their Applications. Birkhauser/Springer, Cham, 2015.

[13] A. Figalli, N. Gigli. *Local semiconvexity of Kantorovich potentials on non-compact manifolds*. ESAIM Control Optim. Calc. Var. 17, no. 3. 2011.