EDPs via Fluxo de Gradiente em Espaços de Wasserstein

Autor: Davi Sales Barreira

Sumário

Sumário

Ideia Geral

O espaço de Wasserstein se trata de um espaço métrico de medidas de probabilidade embutido com a métrica de Wasserstein.

Um Fluxo de Gradiente é um sistema de equações onde a evolução do sistema se dá através da descida de gradiente.

A ideia geral dessa apresentação é mostrar como algumas EDPs podem ser reformuladas em termos de um Fluxo de Gradiente em um espaço de Wasserstein. Apresentaremos como reformular a equação de calor, porém, esse método é mais geral, sendo aplicável para muitas outras EDPs.

Motivação

Por que interpretar EDPs como Fluxo de Gradiente em Wasserstein?

1. Estética. Veremos que é uma bela interpretação que permite entender as EDPs de outro ponto de vista;
2. Reformulação permite utilizar outros ferramentais para demonstrar, por exemplo, taxas de convergência, existência e unicidade;
3. Esquema de discretização de fluxos de gradiente como algoritmo para aproximar soluções fracas para as EDPs.

Problema de Monge - Qual a maneira ótima de transporta massa de uma configuração para outra?

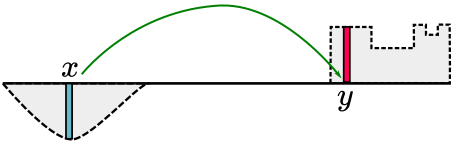

Figure 1: Massa não pode ser separada.

Kantorovich Problem - Relaxação do problema original de Monge.

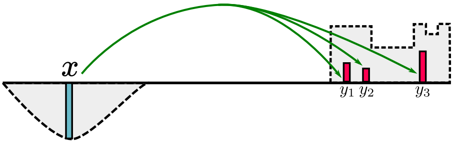

Figure 2: Massa pode ser separada.

Definition (Problema de Monge)

Dadas duas medidas de probabilidade $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ e uma função de custo $c : X \times Y \to [0, +\infty]$, resolva:

$$(MP) \qquad \inf \left\{ \int_X c(x, T(x)) d\mu \quad : \quad T_{\#}\mu = \nu \right\} \tag{1}$$

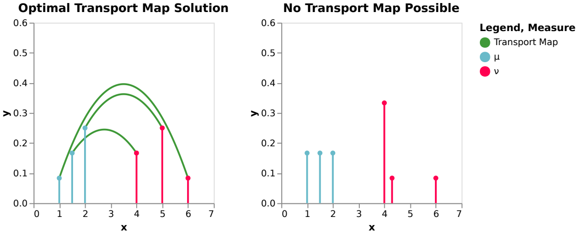

Figure 3: Exemplo de dois problemas de Transporte Ótimo.

Definition (Acoplamento)

Sejam (X, μ) e (Y, ν) espaços de probabilidade. Para $\gamma \in \mathcal{P}(X \times Y)$, dizemos que γ é um acoplamento de (μ, ν) se $(\pi_X)_\# \gamma = \mu$ e $(\pi_Y)_\# \gamma = \nu$. Chamamos $\Pi(\mu, \nu)$ do conjunto de **Planos de Transporte**:

$$\Pi(\mu, \nu) := \{\gamma \in \mathcal{P}(X \times Y) \ : \ (\pi_X)_\# \gamma = \mu \quad \text{and} \quad (\pi_Y)_\# \gamma = \nu\} \tag{2}$$

Definition (Problema de Kantorovich)

Dadas duas medidas de probabilidade $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ e a função de custo $c : X \times Y \to [0, +\infty]$, resolva:

$$(KP) \qquad \inf \left\{ \int_{X \times Y} c(x, y) d\gamma \ : \ \gamma \in \Pi(\mu, \nu) \right\} \tag{3}$$

O Problema de Kantorovich tem uma formulação dual, que para certas condições de regularidade possui a mesma solução ótima que o problema primal (dualidade forte).

Definition (Problema Dual)

Dadas $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ e custo $c : X \times Y \to \mathbb{R}_+$. O Problema Dual é

$$\text{(DP)} \quad \sup \left\{ \int_X \phi \, d\mu + \int_Y \psi \, d\nu \ : \phi \in C_b(X) \ , \psi \in C_b(Y) \ , \ \phi \oplus \psi \leq c \right\} \tag{4}$$

Funções ϕ, ψ são chamdas de **Potenciais de Kantorovich**.

Definition (Distância de Wasserstein)

Seja (X, d) um espaço métrico polonês, com $c : X \times X \to \mathbb{R}$ tal que $c(x, y) = d(x, y)^p$, e $p \in [1, +\infty)$. Para $\mu, \nu \in \mathcal{P}_p(X)$, a distância de Wasserstein é dada por:

$$W_p(\mu, \nu) := \left(\inf_{\gamma \in \Pi(\mu,\nu)} \int_{X \times X} d(x, y)^p \ d\gamma \right)^{1/p} \tag{5}$$

$\mathcal{P}_p(X)$ é o espaço de medidas de probabilidade com p-ésimo momento.

References I

FGV EMAp

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

[2] Luigi Ambrosio, Elia Brué, and Daniele Semola. Lectures on optimal transport, 2021.

[3] H.H. Bauschke and P.L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.

[4] R.I. Bot. *Conjugate duality in convex optimization*, volume 637. Springer Science & Business Media, 2009.

[5] G. Carlier and C. Poon. On the total variation wasserstein gradient flow and the tv-jko scheme. *ESAIM: Control, Optimisation and Calculus of Variations*, 25:42, 2019.

[6] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An introduction to total variation for image analysis. *Theoretical foundations and numerical methods for sparse recovery*, 9(263-340):227, 2010.

[7] L.C. Evans and R.F. Gariepy. *Measure theory and fine properties of functions*. CRC press, 2015.

[8] Rémi FLAMARY. *Transport optimal pour l'apprentissage statistique*. PhD thesis, Télécom Paristech, 2019.

[9] Rémi Flamary. Optimal transport for machine learning. page 97, November 2019.

[10] David JH Garling. *Analysis on Polish spaces and an introduction to optimal transportation*, volume 89. Cambridge University Press, 2018.

[11] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the fokker–planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17, 1998.

[12] Grégoire Montavon, Klaus-Robert Müller, and Marco Cuturi. Wasserstein training of restricted boltzmann machines. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29, pages 3718–3726. Curran Associates, Inc., 2016. URL `https://proceedings.neurips.cc/paper/2016/file/728f206c2a01bf572b5940d7d9a8fa4c-Paper.pdf`.

[13] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[14] R.T. Rockafellar. *Conjugate duality and optimization*. SIAM, 1974.

[15] R. Rossi and G. Savaré. Tightness, integral equicontinuity and compactness for evolution problems in banach spaces. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 2(2):395–431, 2003.

[16] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94, 2015.

References II

[17] Filippo Santambrogio. {Euclidean, metric, and Wasserstein} gradient flows: an overview. *Bulletin of Mathematical Sciences*, 7(1): 87–154, 2017.

[18] user125646 (https://math.stackexchange.com/users/125646/user125646). How to show that the set of all lipschitz functions on a compact set x is dense in c(x)? Mathematics Stack Exchange. URL `https://math.stackexchange.com/q/665686`. URL:https://math.stackexchange.com/q/665686 (version: 2014-02-07).

[19] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[20] Larry Wasserman. Statistical methods for machine learning - lecture notes, 2018.