EDPs via Fluxo de Gradiente em Espaços de Wasserstein

Autor: Davi Sales Barreira

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?

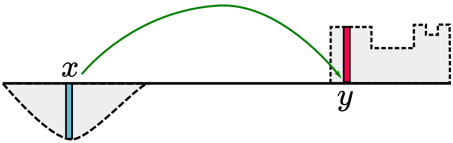


Figure 1: Massa não pode ser separada.

Kantorovich Problem - Relaxação do problema original de Monge.

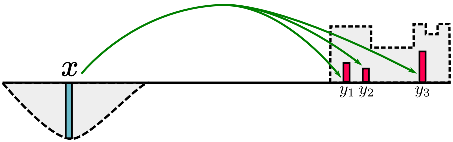


Figure 2: Massa pode ser separada.

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:

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

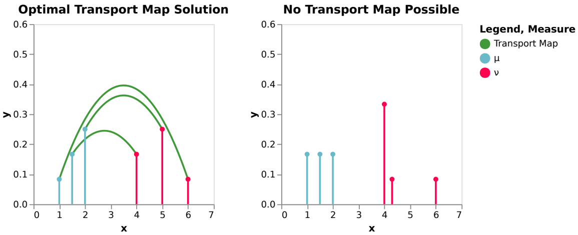


Figure 3: Exemplo de dois problemas de Transporte Ótimo.

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**.

When the cost function is a distance metric, the Dual Problem can be written in what is known as the Kantorovich-Rubinstein formulation.

Theorem (Kantorovich-Rubinstein)

Let (X, d) *be a Polish space with metric* d*, and cost function* $c(x, y) = d(x, y)$*. Then, for* $\mu, \nu \in \mathcal{P}(X)$*, the Kantorovich Problem is equivalent to*

$$\sup \left\{ \int_X \phi \, d\mu - \int_X \phi \, d\nu \ : \phi \in Lip_1(X) \right\} \tag{5}$$

Definition (Wasserstein Distance)

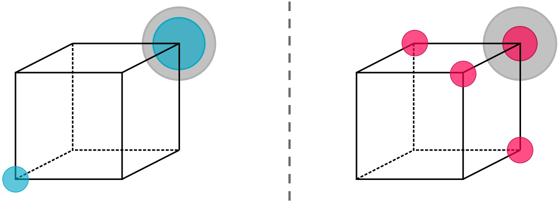
Let (X, d) be a Polish metric space, with $c : X \times X \to \mathbb{R}$ such that $c(x, y) = d(x, y)^p$, and $p \in [1, +\infty)$. For $\mu, \nu \in \mathcal{P}_p(X)$, the Wasserstein Distance is given by:

$$W_p(\mu, \nu) := \left(\inf_{\gamma \in \Pi(\mu, \nu)} \int_{X \times X} d(x, y)^p \; d\gamma \right)^{1/p} \tag{6}$$

$\mathcal{P}_p(X)$ is the space of probability measures with finite pth moment.

Optimal Transport Theory - Wasserstein Distance

The Wasserstein distance has many interesting properties which make it useful in Machine Learning applications. Two of them that are of utmost interest are the fact that it metrizes weak convergence and the incorporation of the ground geometry.



$$KL(\mu||\nu) = \text{low} \qquad KL(\mu||\nu) = \text{high}$$
$$W_p(\mu,\nu) = \text{high} \qquad W_p(\mu,\nu) = \text{low}$$

Figure 4: Comparison between Wasserstein distance and KL Divergence, based on Montavon et al. [55].

References I

FGV EMAp

[1] Martial Agueh and Guillaume Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2): 904–924, 2011.

[2] David Alvarez-Melis and Nicolò Fusi. Geometric dataset distances via optimal transport. *arXiv preprint arXiv:2002.02923*, 2020.

[3] David Alvarez-Melis and Tommi S Jaakkola. Gromov-wasserstein alignment of word embedding spaces. *arXiv preprint arXiv:1809.00013*, 2018.

[4] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

[5] Yuki Markus Asano, Christian Rupprecht, and Andrea Vedaldi. Self-labelling via simultaneous clustering and representation learning. *arXiv preprint arXiv:1911.05371*, 2019.

[6] Gary Bécigneul, Octavian-Eugen Ganea, Benson Chen, Regina Barzilay, and Tommi Jaakkola. Optimal transport graph neural networks. *arXiv preprint arXiv:2006.04804*, 2020.

[7] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[8] David Berthelot, Thomas Schumm, and Luke Metz. Began: Boundary equilibrium generative adversarial networks. *arXiv preprint arXiv:1703.10717*, 2017.

[9] Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and radon wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015.

[10] Nicolas Bonneel, Gabriel Peyré, and Marco Cuturi. Wasserstein barycentric coordinates: histogram regression using optimal transport. *ACM Trans. Graph.*, 35(4):71–1, 2016.

[11] Olivier Bousquet, Sylvain Gelly, Ilya Tolstikhin, Carl-Johann Simon-Gabriel, and Bernhard Schoelkopf. From optimal transport to generative modeling: the vegan cookbook. *arXiv preprint arXiv:1705.07642*, 2017.

[12] Charlotte Bunne, David Alvarez-Melis, Andreas Krause, and Stefanie Jegelka. Learning generative models across incomparable spaces. In *International Conference on Machine Learning*, pages 851–861. PMLR, 2019.

[13] Jiezhang Cao, Langyuan Mo, Yifan Zhang, Kui Jia, Chunhua Shen, and Mingkui Tan. Multi-marginal wasserstein gan. *arXiv preprint arXiv:1911.00888*, 2019.

[14] Liqun Chen, Shuyang Dai, Chenyang Tao, Dinghan Shen, Zhe Gan, Haichao Zhang, Yizhe Zhang, and Lawrence Carin. Adversarial text generation via feature-mover's distance. *arXiv preprint arXiv:1809.06297*, 2018.

[15] Lenaic Chizat and Francis Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. In *Advances in neural information processing systems*, pages 3036–3046, 2018.

[16] Nicolas Courty, Rémi Flamary, and Devis Tuia. Domain adaptation with regularized optimal transport. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 274–289. Springer, 2014.

[17] Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 39(9):1853–1865, 2016.

[18] Nicolas Courty, Rémi Flamary, Amaury Habrard, and Alain Rakotomamonjy. Joint distribution optimal transportation for domain adaptation. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL `https://proceedings.neurips.cc/paper/2017/file/0070d23b06b1486a538c0eaa45dd167a-Paper.pdf`.

[19] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26:2292–2300, 2013.

[20] Marco Cuturi and Arnaud Doucet. Fast computation of wasserstein barycenters. In *International conference on machine learning*, pages 685–693. PMLR, 2014.

[21] Marco Cuturi, Olivier Teboul, and Jean-Philippe Vert. Differentiable ranking and sorting using optimal transport. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL `https://proceedings.neurips.cc/paper/2019/file/d8c24ca8f23c562a5600876ca2a550ce-Paper.pdf`.

[22] Bharath Bhushan Damodaran, Benjamin Kellenberger, Rémi Flamary, Devis Tuia, and Nicolas Courty. Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 447–463, 2018.

[23] Ishan Deshpande, Ziyu Zhang, and Alexander G Schwing. Generative modeling using the sliced wasserstein distance. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3483–3491, 2018.

[24] Pierre Dognin, Igor Melnyk, Youssef Mroueh, Jerret Ross, Cicero Dos Santos, and Tom Sercu. Wasserstein barycenter model ensembling. *arXiv preprint arXiv:1902.04999*, 2019.

[25] Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and mmd using sinkhorn divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2681–2690. PMLR, 2019.

References III

[26] Rémi Flamary, Marco Cuturi, Nicolas Courty, and Alain Rakotomamonjy. Wasserstein discriminant analysis. *Machine Learning*, 107 (12):1923–1945, 2018.

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

[28] Charlie Frogner, Chiyuan Zhang, Hossein Mobahi, Mauricio Araya-Polo, and Tomaso Poggio. Learning with a wasserstein loss. *arXiv preprint arXiv:1506.05439*, 2015.

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

[30] Léo Gautheron, Ievgen Redko, and Carole Lartizien. Feature selection for unsupervised domain adaptation using optimal transport. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 759–776. Springer, 2018.

[31] Aude Genevay, Gabriel Peyré, and Marco Cuturi. Learning generative models with sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*, pages 1608–1617. PMLR, 2018.

[32] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014. URL http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

[33] Edouard Grave, Armand Joulin, and Quentin Berthet. Unsupervised alignment of embeddings with wasserstein procrustes. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1880–1890. PMLR, 2019.

[34] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028*, 2017.

[35] Xin Guo, Johnny Hong, Tianyi Lin, and Nan Yang. Relaxed wasserstein with applications to gans. *arXiv preprint arXiv:1705.07164*, 2017.

[36] GM Harshvardhan, Mahendra Kumar Gourisaria, Manjusha Pandey, and Siddharth Swarup Rautaray. A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review*, 38:100285, 2020.

[37] Gao Huang, Chuan Guo, Matt J Kusner, Yu Sun, Fei Sha, and Kilian Q Weinberger. Supervised word mover's distance. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper/2016/file/10c66082c124f8afe3df4886f5e516e0-Paper.pdf.

[38] Hicham Janati, Marco Cuturi, and Alexandre Gramfort. Wasserstein regularization for sparse multi-task regression. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 1407–1416. PMLR, 16–18 Apr 2019. URL `http://proceedings.mlr.press/v89/janati19a.html`.

[39] Ray Jiang, Aldo Pacchiano, Tom Stepleton, Heinrich Jiang, and Silvia Chiappa. Wasserstein fair classification. In *Uncertainty in Artificial Intelligence*, pages 862–872. PMLR, 2020.

[40] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[41] Soheil Kolouri, Phillip E Pope, Charles E Martin, and Gustavo K Rohde. Sliced-wasserstein autoencoder: An embarrassingly simple generative model. *arXiv preprint arXiv:1804.01947*, 2018.

[42] Wouter M Kouw and Marco Loog. An introduction to domain adaptation and transfer learning. *arXiv preprint arXiv:1812.11806*, 2018.

[43] Mathias Kraus and Stefan Feuerriegel. Personalized purchase prediction of market baskets with wasserstein-based sequence matching. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2643–2652, 2019.

[44] Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger. From word embeddings to document distances. In *International conference on machine learning*, pages 957–966. PMLR, 2015.

[45] Charlotte Laclau, Ievgen Redko, Basarab Matei, Younes Bennani, and Vincent Brault. Co-clustering through optimal transport. In *International Conference on Machine Learning*, pages 1955–1964. PMLR, 2017.

[46] Chen-Yu Lee, Tanmay Batra, Mohammad Haris Baig, and Daniel Ulbricht. Sliced wasserstein discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10285–10295, 2019.

[47] Na Lei, Kehua Su, Li Cui, Shing-Tung Yau, and Xianfeng David Gu. A geometric view of optimal transportation and generative model. *Computer Aided Geometric Design*, 68:1–21, 2019.

[48] Xuhong Li, Yves Grandvalet, Rémi Flamary, Nicolas Courty, and Dejing Dou. Representation transfer by optimal transport. *arXiv preprint arXiv:2007.06737*, 2020.

[49] Antoine Liutkus, Umut Simsekli, Szymon Majewski, Alain Durmus, and Fabian-Robert Stöter. Sliced-wasserstein flows: Nonparametric generative modeling via optimal transport and diffusions. In *International Conference on Machine Learning*, pages 4104–4113. PMLR, 2019.

[50] Giulia Luise, Alessandro Rudi, Massimiliano Pontil, and Carlo Ciliberto. Differential properties of sinkhorn approximation for learning with wasserstein distance. *arXiv preprint arXiv:1805.11897*, 2018.

[51] Chen Ma, Liheng Ma, Yingxue Zhang, Ruiming Tang, Xue Liu, and Mark Coates. Probabilistic metric learning with adaptive margin for top-k recommendation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1036–1044, 2020.

[52] Saurav Manchanda, Khoa Doan, Pranjul Yadav, and S Sathiya Keerthi. Regression via implicit models and optimal transport cost minimization. *arXiv preprint arXiv:2003.01296*, 2020.

[53] Facundo Mémoli. A spectral notion of gromov–wasserstein distance and related methods. *Applied and Computational Harmonic Analysis*, 30(3):363–401, 2011.

[54] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.

[55] 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`.

[56] Vaishnavh Nagarajan and J. Zico Kolter. Gradient descent gan optimization is locally stable. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30, pages 5585–5595. Curran Associates, Inc., 2017. URL `https://proceedings.neurips.cc/paper/2017/file/7e0a0209b929d097bd3e8ef30567a5c1-Paper.pdf`.

[57] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10): 1345–1359, 2009.

[58] Giorgio Patrini, Rianne van den Berg, Patrick Forré, Marcello Carioni, Samarth Bhargav, Max Welling, Tim Genewein, and Frank Nielsen. Sinkhorn autoencoders. In Ryan P. Adams and Vibhav Gogate, editors, *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pages 733–743, Tel Aviv, Israel, 22–25 Jul 2020. PMLR. URL `http://proceedings.mlr.press/v115/patrini20a.html`.

[59] Michaël Perrot, Nicolas Courty, Rémi Flamary, and Amaury Habrard. Mapping estimation for discrete optimal transport. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 4204–4212, 2016.

[60] 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.

[61] Ievgen Redko, Nicolas Courty, Rémi Flamary, and Devis Tuia. Optimal transport for multi-source domain adaptation under target shift. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 849–858. PMLR, 2019.

[62] Ievgen Redko, Emilie Morvant, Amaury Habrard, Marc Sebban, and Younès Bennani. A survey on domain adaptation theory. *arXiv preprint arXiv:2004.11829*, 2020.

[63] Laurent Risser, Quentin Vincenot, and Jean-Michel Loubes. Tackling algorithmic bias in neural-network classifiers using wasserstein-2 regularization. *arXiv e-prints*, pages arXiv–1908, 2019.

[64] Antoine Rolet, Marco Cuturi, and Gabriel Peyré. Fast dictionary learning with a smoothed wasserstein loss. In *Artificial Intelligence and Statistics*, pages 630–638. PMLR, 2016.

[65] Denis Rousselle and Stéphane Canu. Optimal transport for semi-supervised domain adaptation. In *Proceedings*, page 373. Presses universitaires de Louvain, 2015.

[66] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3723–3732, 2018.

[67] Masaki Saito, Eiichi Matsumoto, and Shunta Saito. Temporal generative adversarial nets with singular value clipping. In *Proceedings of the IEEE international conference on computer vision*, pages 2830–2839, 2017.

[68] Tim Salimans, Han Zhang, Alec Radford, and Dimitris Metaxas. Improving gans using optimal transport. *arXiv preprint arXiv:1803.05573*, 2018.

[69] Roman Sandler and Michael Lindenbaum. Nonnegative matrix factorization with earth mover's distance metric. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1873–1880. IEEE, 2009.

[70] Maziar Sanjabi, Jimmy Ba, Meisam Razaviyayn, and Jason D Lee. On the convergence and robustness of training gans with regularized optimal transport. *arXiv preprint arXiv:1802.08249*, 2018.

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

[72] Morgan A Schmitz, Matthieu Heitz, Nicolas Bonneel, Fred Ngole, David Coeurjolly, Marco Cuturi, Gabriel Peyré, and Jean-Luc Starck. Wasserstein dictionary learning: Optimal transport-based unsupervised nonlinear dictionary learning. *SIAM Journal on Imaging Sciences*, 11(1):643–678, 2018.

[73] Vivien Seguy and Marco Cuturi. Principal geodesic analysis for probability measures under the optimal transport metric. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL `https://proceedings.neurips.cc/paper/2015/file/f26dab9bf6a137c3b6782e562794c2f2-Paper.pdf`.

[74] Soroosh Shafieezadeh Abadeh, Peyman Mohajerin Mohajerin Esfahani, and Daniel Kuhn. Distributionally robust logistic regression. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL `https://proceedings.neurips.cc/paper/2015/file/cc1aa436277138f61cda703991069eaf-Paper.pdf`.

[75] Soroosh Shafieezadeh-Abadeh, Daniel Kuhn, and Peyman Mohajerin Esfahani. Regularization via mass transportation. *Journal of Machine Learning Research*, 20(103):1–68, 2019. URL `http://jmlr.org/papers/v20/17-633.html`.

[76] Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[77] Jian Shen, Yanru Qu, Weinan Zhang, and Yong Yu. Wasserstein distance guided representation learning for domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[78] Sidak Pal Singh and Martin Jaggi. Model fusion via optimal transport. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 22045–22055. Curran Associates, Inc., 2020. URL `https://proceedings.neurips.cc/paper/2020/file/fb2697869f56484404c8ceee2985b01d-Paper.pdf`.

[79] Aman Sinha, Hongseok Namkoong, Riccardo Volpi, and John Duchi. Certifying some distributional robustness with principled adversarial training. *arXiv preprint arXiv:1710.10571*, 2017.

[80] Hannah Snyder. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104: 333–339, 2019.

[81] Justin Solomon, Raif Rustamov, Leonidas Guibas, and Adrian Butscher. Wasserstein propagation for semi-supervised learning. In *International Conference on Machine Learning*, pages 306–314. PMLR, 2014.

[82] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*, pages 443–450. Springer, 2016.

[83] Vayer Titouan, Nicolas Courty, Romain Tavenard, and Rémi Flamary. Optimal transport for structured data with application on graphs. In *International Conference on Machine Learning*, pages 6275–6284. PMLR, 2019.

FGV EMAp

[84] Ilya Tolstikhin, Olivier Bousquet, Sylvain Gelly, and Bernhard Schoelkopf. Wasserstein auto-encoders. *arXiv preprint arXiv:1711.01558*, 2017.

[85] Nilesh Tripuraneni, Michael I Jordan, and Chi Jin. On the theory of transfer learning: The importance of task diversity. *arXiv preprint arXiv:2006.11650*, 2020.

[86] Rosanna Turrisi, Rémi Flamary, Alain Rakotomamonjy, and Massimiliano Pontil. Multi-source domain adaptation via weighted joint distributions optimal transport. *arXiv preprint arXiv:2006.12938*, 2020.

[87] 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).

[88] Titouan Vayer, Laetita Chapel, Rémi Flamary, Romain Tavenard, and Nicolas Courty. Fused gromov-wasserstein distance for structured objects: theoretical foundations and mathematical properties. *arXiv preprint arXiv:1811.02834*, 2018.

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

[90] Jake Williams, Abel Tadesse, Tyler Sam, Huey Sun, and George D Montanez. Limits of transfer learning. *arXiv preprint arXiv:2006.12694*, 2020.

[91] Eric Wong, Frank Schmidt, and Zico Kolter. Wasserstein adversarial examples via projected sinkhorn iterations. In *International Conference on Machine Learning*, pages 6808–6817. PMLR, 2019.

[92] Jiqing Wu, Zhiwu Huang, Dinesh Acharya, Wen Li, Janine Thoma, Danda Pani Paudel, and Luc Van Gool. Sliced wasserstein generative models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3713–3722, 2019.

[93] Kaiwen Wu, Allen Wang, and Yaoliang Yu. Stronger and faster wasserstein adversarial attacks. In *International Conference on Machine Learning*, pages 10377–10387. PMLR, 2020.

[94] Yujia Xie, Hanjun Dai, Minshuo Chen, Bo Dai, Tuo Zhao, Hongyuan Zha, Wei Wei, and Tomas Pfister. Differentiable top-k with optimal transport. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 20520–20531. Curran Associates, Inc., 2020. URL `https://proceedings.neurips.cc/paper/2020/file/ec24a54d62ce57ba93a531b460fa8d18-Paper.pdf`.

[95] Yujia Xie, Xiangfeng Wang, Ruijia Wang, and Hongyuan Zha. A fast proximal point method for computing exact wasserstein distance. In *Uncertainty in Artificial Intelligence*, pages 433–453. PMLR, 2020.

References IX

[96] Yuguang Yan, Wen Li, Hanrui Wu, Huaqing Min, Mingkui Tan, and Qingyao Wu. Semi-supervised optimal transport for heterogeneous domain adaptation. In *IJCAI*, volume 7, pages 2969–2975, 2018.

[97] Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Mannudeep K Kalra, Yi Zhang, Ling Sun, and Ge Wang. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, 37(6):1348–1357, 2018.

[98] Hai-Tao Yu, Adam Jatowt, Hideo Joho, Joemon M Jose, Xiao Yang, and Long Chen. Wassrank: Listwise document ranking using optimal transport theory. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 24–32, 2019.

[99] Xiaofeng Zhang, Jingbin Zhong, and Kai Liu. Wasserstein autoencoders for collaborative filtering. *Neural Computing and Applications*, pages 1–10, 2020.

[100] Junbo Zhao, Yoon Kim, Kelly Zhang, Alexander Rush, and Yann LeCun. Adversarially regularized autoencoders. In *International conference on machine learning*, pages 5902–5911. PMLR, 2018.

[101] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 2020.