Primal-dual hybrid algorithms for chi-squared regularized Optimal Transport: statistical-computational trade-offs and applications to Wasserstein Barycenters

Denys Ruban DRUBA015@UOTTAWA.CA
University of Ottawa, Canada

Augusto Gerolin AGEROLIN@UOTTAWA.CA,GEROLIN@IMPA.BR
University of Ottawa, Canada
Instituto Nacional de Matemática Pura e Aplicada, Brazil

Abstract

We investigate the interplay between optimization and statistics in regularized Optimal Transport problems [9, 11]—the standard approach for computational Optimal Transport [5, 15]. While regularization parameter improves optimization and tractability, it inevitably introduces bias. In this paper, we design a computational algorithm based on chi-square regularized Primal-Dual Hybrid Gradient (PDHG) that are not only efficient and robust, but also capable of accurately recovering key statistical properties, including Monge maps, Kantorovich potentials and Wasserstein barycenters. Our implementation is JAX-based, GPU-paralellizable, and incorporates adaptive step sizes and restart strategies, the latter ensuring a linear convergence rate [2].

1. Introduction

Let $\varepsilon > 0$, Ω be a subset of $\mathbb{R}^d$, and $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ be a lower semicontinuous convex function, superlinear at infinity. For simplicity, assume that $\Phi \in C^1((0,\infty))$ and $\Phi(1) = \Phi'(1) = 0$.

In this paper, we investigate a novel primal-dual hybrid gradient (PDHG) algorithm [4] solving a convex-regularized $2-$Wasserstein distance between probability measures $\mu, \nu \in \mathcal{P}(\Omega)$ [9, 11]

$$W^2_{2,\Phi,\varepsilon}(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} C_{2,\Phi,\varepsilon}(\pi) := \min_{\pi\in\Pi(\mu,\nu)} \int_{\Omega\times\Omega} \|x-y\|_2^2 d\pi(x,y) + \varepsilon G(\pi|\xi), \tag{1}$$

where $\Pi(\mu,\nu)$ denotes the set of *transport plans* between μ and ν, and $G : \mathcal{P}(\Omega\times\Omega) \to \mathbb{R}\cup\{+\infty\}$ is the Φ-entropy, that is, $G(\pi|\xi) = \int_{\Omega\times\Omega} \Phi(\frac{d\pi}{d\xi})d\xi$ if $\pi \ll \xi$ is a absolutely continuous measure with respect to a reference measure ξ and $+\infty$ otherwise.

Particular choices of regularization includes $\Phi(t) = t(\log\, t-1)+1$ for $t \geq 0$ (Shannon entropy) [5, 15], $\Phi(t) = \frac{1}{2}(t_+)^2$ (Chi-square) [6, 8] and $\Phi(t) = \frac{1}{p(p-1)}(t^p - p(t-1))$, $p > 1$ (Tsallis) [11, 13]. In the limit when $\varepsilon \to 0^+$, the problem (1) converges to the $2-$Wasserstein distance [16, 17]

$$W_2^2(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \int_{\Omega\times\Omega} \|x-y\|_2^2 d\pi(x,y) = \max_{\substack{u\in L^1(\mu), v\in L^1(\nu)\\ u(x)+v(y)\leq\|x-y\|^2}} \Big\{ \int_\Omega u d\mu + \int_\Omega v d\nu \Big\}, \tag{2}$$

where the potentials u, v are the so-called Kantorovich potentials. Due to Brenier's Theorem [3], the solution $\pi = (\mathrm{Id}, T)_{\#}\mu$ in (2) is concentrated on the (Brenier) map $T_\sharp\mu = \nu$, $T(x) = x - \nabla u(x)$.

The main objective of this work is to leverage regularization in combination with the PDHG optimizer to design computational algorithms that are both robust and efficient, while faithfully preserving statistical properties. We focus on accurately estimating transport maps, Kantorovich potentials, and the gradient of the 2-Wasserstein distance as well as 2−Wasserstein Barycenters.

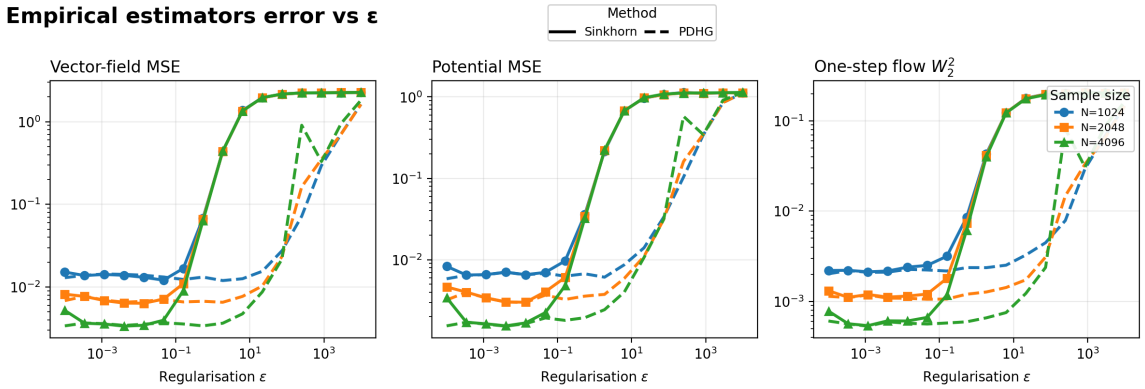


Figure 1: We evaluate the bias of empirical estimators of the Kantorovich potential (middle), the associated vector field $v(x) = T(x) - x$ (left), and the displacement interpolation $T_t(x) = x + tv(x), t \in [0, 1]$ (right) for one–dimensional Gaussian distributions as a function of the regularization parameter ε, see details in sections 2.2, A.3 and A.5. For each ε, 100 independent samples are drawn, and the mean squared error is computed according to (27). The results show that PDHG with chi-square regularization maintains lower bias even for relatively larger ε compared to the (Shannon-entropy) Sinkhorn algorithm.

Main contributions and organization of the paper:

- In section 2, we introduce a novel algorithm for computing the chi-square regularized 2-Wasserstein distance. Our method is GPU-parallelizable [1], and stable as the regularization parameter $\varepsilon \to 0^+$, reaching machine precision. This enables principled control over accuracy versus computational cost, allowing ε to be selected according to the desired level of precision/efficiency in applications.
- Through experiments in section 3, we focus on the interplay between optimization and statistics: we illustrate, in academic instances, that PDHG optimizer achieves a favorable trade-off between optimization and statistical fidelity, enabling the estimation of key properties of such as Brenier map T, Kantorovich potentials (see eq. (2)), and displacement interpolation, see Figure 1.
- Finally, in section 4, we apply our methodology to compute 2-Wasserstein barycenters.

2. Primal-dual hybrid gradient algorithm for convex-regularized optimal transport

2.1. Convex-regularized 2-Wasserstein distance: strong-duality

The 2-Wasserstein distance with general convex regularization in (1) admits a dual formulation [9]

$$W^2_{2,\Phi,\varepsilon}(\mu,\nu) = \sup\{\, D_\varepsilon(u,v) \;:\; u \in L^\Phi(\Omega,\mu), v \in L^\Phi(\Omega,\nu)\,\} \tag{3}$$

$$:= \sup_{u,v \in L^\Phi} \left\{ \int_\Omega u\, d\mu + \int_\Omega v\, d\nu - \varepsilon \int_{\Omega\times\Omega} \Phi\left(\frac{u(x) + v(y) - \|x-y\|_2^2}{\varepsilon} \right) d\xi \right\}, \tag{4}$$

where the class of potentials (u, v) in equation 3 is given by [9, 11]

$$L^{\Phi}(\Omega,\nu) = \{f : \Omega \to \mathbb{R} : \|f\|_{L^{\Phi}} < \infty\}\,, \text{ where } \|f\|_{L^{\Phi}} = \inf\left\{\varepsilon \geq 0 : \int_{\Omega} \Phi\left(\frac{|f|}{\varepsilon}\right) d\nu < \infty\right\}.$$

2.2. Chi-squared regularization with adaptive step-size and restart

In what follows, we specialize the PDHG algorithm to the chi-square regularization of the $2-$Wasserstein distance, obtained by choosing $\Phi(t) = \frac{1}{2}t_+^2$ and $\xi = \mu \otimes \nu$ in equation (1). A key feature of the chi-square regularization in PDHG algorithms is that it admits closed-form expressions for the proximal map, which greatly simplifies the computations.

Indeed, let $\mu, \nu \in \mathcal{P}(\Omega)$ be absolute continuous measures with densities ρ_μ, ρ_ν. Assume $\pi \ll \xi$ and denote by $\tilde{\pi}$ the density of π with respect to ξ, then the proximal map for $\Phi = \frac{1}{2}t_+^2$ is given by

$$\mathrm{prox}_{\tau C_{2,\Phi,\varepsilon}}(\tilde{\pi}) = \left(\mathrm{Id} + \tau \partial C_{2,\Phi,\varepsilon}\right)^{-1}(\tilde{\pi}) = \left(\mathrm{Id} + \tau\varepsilon\Phi'\right)^{-1}(\tilde{\pi} - \tau c) = \left(\frac{\tilde{\pi}(x,y) - \tau c(x,y)}{1+\tau\varepsilon}\right)_+.$$

The PDHG iterations $(\pi^k, u^k, v^k)_{k\in\mathbb{N}}$ for the chi-square regularized optimal transport primal (1) and dual (3) problems are – see details in the suplementary material in equation (10)

$$\begin{aligned}
\pi^{k+1}(x,y) &= \left(\frac{\tilde{\pi}^k(x,y) - \tau\big(c(x,y) - u^k(x) - v^k(y)\big)}{1+\tau\varepsilon}\right)_+ \mu\otimes\nu,\\
u^{k+1}(x) &= u^k(x) + \sigma\left(\rho_\mu(x) - \int_\Omega \left(2\tilde{\pi}^{k+1}(x,y) - \tilde{\pi}^k(x,y)\right) d\xi(y)\right),\\
v^{k+1}(y) &= v^k(y) + \sigma\left(\rho_\nu(y) - \int_\Omega \left(2\tilde{\pi}^{k+1}(x,y) - \tilde{\pi}^k(x,y)\right) d\xi(x)\right).
\end{aligned} \tag{5}$$

Our implementation described in details Algorithms 2-4 in section A.3 in the supplementary material. It incorporates adaptive step sizes and restart strategies, the latter ensuring a linear convergence rate for linear programming [2] – rather than sublinear convergence on Chambolle-Pock PDHG [4]. Adaptive restart strategies for PDHG were originally developed in [1] under the name PDLP, and later adapted to Optimal Transport [10].

3. Experiments: interplay between optimization and statistics

In the following experiments, we apply the PDHG algorithm (Algorithm 2 in section A.3) to two cases: (i) μ and ν are Gaussian distributions, and (ii) μ and ν are Cauchy distributions, both with randomly chosen parameters. Our focus is on estimating the Monge maps, Kantorovich potentials, and displacement interpolations. Further details regarding the optimization aspects of the problem are provided in Section A.3 of the supplementary material, while Section A.5 discusses the statistical properties and the methodology used to estimate these quantities.

All experiments are conducted on an NVIDIA H100 GPU using implementations written entirely in JAX. We evaluate one-dimensional source–target pairs with $n = 8192$ support points, varying the regularization parameter over $\varepsilon \in \{10^{-3}, 10^{-2}, \ldots, 10^{5}\}$ with stopping criteria 10^{-6} tolerance on marginal error. Figure 2 reports both the wall-clock runtime and the attained objective precision for each ε. For comparison, we benchmark against Sinkhorn-based algorithms for

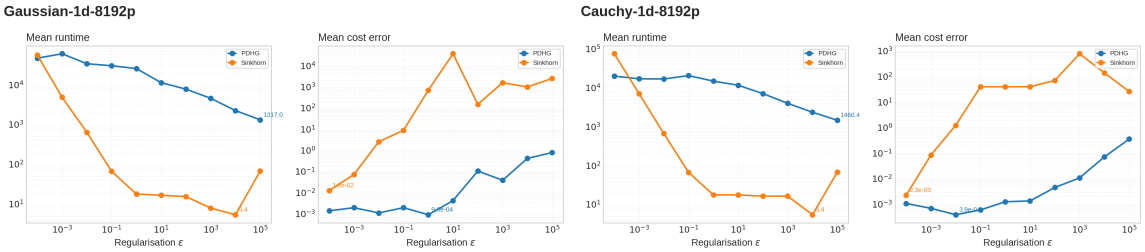


Figure 2: Mean runtime and objective precision (mean cost error) of Sinkhorn (Shannon-entropy) and PDHG (chi-square) in function of the regularization parameter ε. Marginals μ and ν are 20 different 1D Gaussian (left) and Cauchy (right) distributions with random parameters and $n = 8192$ points. These plots show the interplay between precision and runtime. Plots show that PDHG has more flat runtime curve and higher precision even for larger values of ε, while Sinkhorn blows up in runtime on smaller (around 10^{-3}) regularization values attaining the same preicison as PDHG.

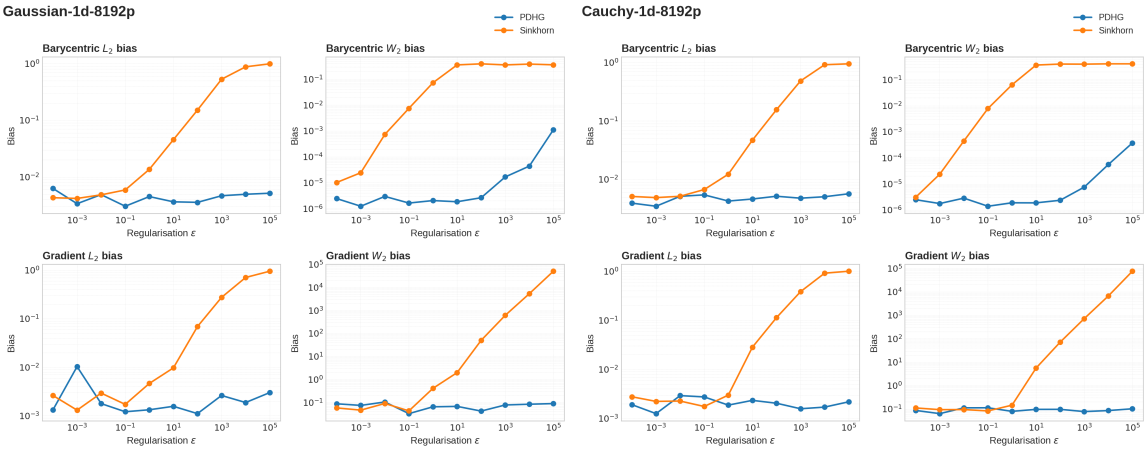


Figure 3: ℓ^2 and W_2 push-forward errors (25) of the optimal transport map in (2) in function of the regularization $\varepsilon \in \{10^{-3}, \ldots, 10^5\}$ for one-dimensional Gaussian (left) and Cauchy (right) datasets. The results indicate that, even for comparatively large values of ε, quadratic PDHG attains higher accuracy in estimating the exact transport map in (1).

the Shannon-entropy regularized 2-Wasserstein distance implemented using JAX and executed on GPU.

Table 1 in the supplementary material section (A.4) summarizes runtime and error for computing Shannon-entropy regularized (Sinkhorn) and chi-square regularised (PDHG) 2-Wasserstein distance in our experimental setup. Figure 3 displays the resulting error curves. As expected, decreasing the regularization parameter ε brings both methods closer to the exact 2-Wasserstein distance.

We also performed the experiments on estimating the Kantorovich potentials, as well as estimating the displacement interpolation for Gaussian 1D distributions. Figure 1 in page 2 shows the

displacement interpolation step as a function of ε. The plots illustrate the theoretical behavior that as $\varepsilon \to 0$, the bias approaches the discretization bias, showing that quadratic regularization exhibits smaller bias for finite values of ε.

However, the computational trade-offs differ substantially: Sinkhorn is extremely efficient when a relatively large regularization ($\varepsilon \gtrsim 10^{-2}$) is suitable, but its runtime increases by nearly an order of magnitude once ε approaches 10^{-4} and numerical instabilities occur for smaller ε. In contrast, PDHG exhibits almost linear computational time scaling with respect to ε, remaining practical even in the low-bias regime where Sinkhorn becomes prohibitively slow. Quadratic regularization is more robust in providing a statistically more accurate estimator of the Monge map, even for relatively large regularization parameters. It also enjoys favorable properties such as sparsity of the transport plan and nice convergence rates toward the unregularized solution as $\varepsilon \to 0$ [8, 14].

4. Wassestein Barycenters

Let $k \in \mathbb{N}$ and $(\lambda_j)_{j=1}^k$ be a sequence of non-negative numbers such that $\sum_{j=1}^k \lambda_k = 1$. Consider $\nu_1, \dots, \nu_k \in \mathcal{P}(\Omega)$ be i.i.d. input measures. The $2-$Wasserstein Barycenter and its corresponding (regularized) empirical estimator are defined as a solution of the following variational problems,

$$\mu^\star \in \arg\min_{\mu} \ \mathbb{E}_{\nu\sim\mathcal{P}}\, W_2^2(\mu,\nu) \quad \text{and} \quad \hat{\mu}_\varepsilon \ \in \ \arg\min_{\mu} \ \sum_{i=1}^k \lambda_i W_{2,\Phi,\varepsilon}^2(\mu,\nu_i). \tag{6}$$

The measure $\hat{\mu}_\varepsilon$ is an estimator of $\mu^\star$ and the effect of regularization on the right-hand-size of (6) renders the objective strongly convex on grids, ensuring uniqueness and numerical stability.

We compute 2-Wasserstein Barycenters on three benchmark settings: (i) one-dimensional Gaussian distributions (Fig. 5); (ii) geometric two-dimensional shapes (disc-square-diamond shapes in Fig. 4); and (iii) the MNIST dataset (Fig. 4), using both Shannon-entropy (Sinkhorn) and chi-square regularized (PDHG) approaches.

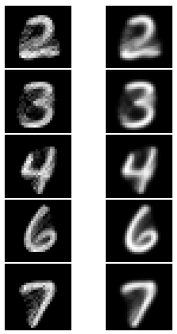
Implementation of PDHG is written in JAX and we use POT library for Sinkhorn $2-$Wasserstein Barycenter. Experiments are executed on a single NVIDIA H100 GPU. Computation time is dominated by the iterative solvers, with stopping criteria set to a marginal error tolerance of 10^{-6} for both PDHG and Sinkhorn methods.

5. Conclusions and future work

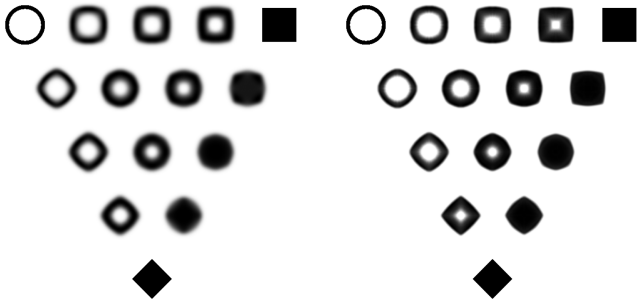
In this work, we introduced a primal–dual hybrid gradient (PDHG) algorithm for the chi-square regularized 2-Wasserstein distance and benchmarked it on high-resolution 1-D and 2-D transport tasks. Our evaluation considered runtime, transport cost accuracy, and the quality of downstream statistical quantities such as Monge maps, displacement interpolation and Wasserstein barycenters.

The experiments reveal a clear trade-off between existing methods: Sinkhorn remains the algorithm of choice when fast approximations with moderate precision are sufficient, whereas PDHG excels in regimes where low bias and accurate statistical estimates are essential - even for relatively large regularization ε (see Fig. 3). Importantly, PDHG achieves this reliability at a computational cost that is often comparable to, or lower than, that of Sinkhorn in the small-bias regime.

These findings highlight the complementary roles of Sinkhorn and PDHG, and position chi-square regularization as a practical alternative to entropy regularization for applications in machine learning and data science that demand both computational efficiency and statistical fidelity.



MNIST barycenters: PDHG left ($\varepsilon = 0.1$) vs. Sinkhorn right ($\varepsilon = 10^{-4}$)



Shape barycenters: Sinkhorn left ($\varepsilon = 10^{-3}$), PDHG right ($\varepsilon = 20$)

Figure 4: 2-Wasserstein barycenters (6) of a disc, square, and diamond on a 128×128 grid and uniform barycenter of samples size 20 of MNIST digits. Experiments on both geometric shapes and MNIST digits demonstrate that chi-square regularization preserves sharp corners and produces barycenters closer to the unregularized case by gently spreading the mass, whereas entropic regularization tends to introduce blurriness. Higher resolution images are given at Figure 6 in supplementary material.

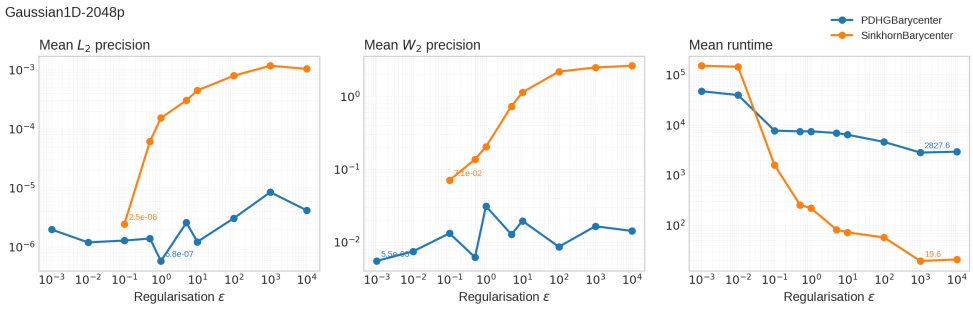


Figure 5: Mean ℓ_2 and W_2 errors and mean runtime for the $2-$Wasserstein barycenter when ν_1 and ν_2 are 20 different one-dimensional Gaussians ($n = 2048$) with random parameters as a function of regularization ε. Our results show that the PDHG optimizer with chi-square regularization achieves higher-precision barycenters at the cost of slower computational time compared to POT Sinkhorn Barycenter. Also, PDHG is stable for small entropic regularization (from $\varepsilon \leq 10^{-2}$) while Sinkhorn encounters numerical issues on this grid.

In future work, we aim to establish convergence guarantees for our algorithms and extend the analysis to high-dimensional settings. In particular, we plan to investigate minibatch OT [7] with chi-square regularization, and empirically characterize the sample complexity regimes for both chi-square and entropic regularizations, highlighting the differences in their bias–variance trade-offs. We also intend to explore applications to sampling and gradient flows, examining how the choice of regularizer influences the geometry, stability, and convergence behavior of these flows.

6. Acknowledgments

D.R. and A.G. thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support under the Multi-Marginal Optimal Transport grant (Grant No. RGPIN-2022-05207), the Mitacs Accelerate Program, and the Canada Research Chairs Program (Grant No. CRC-2021-00234). Part of this research was conducted while the authors were visiting the Institute for Pure and Applied Mathematics (IPAM), which is supported by the National Science Foundation (Grant No. DMS-1925919).

References

[1] David Applegate, Mateo Díaz, Oliver Hinder, Haihao Lu, Miles Lubin, Brendan O'Donoghue, and Warren Schudy. Practical large-scale linear programming using primal-dual hybrid gradient, 2022. URL https://arxiv.org/abs/2106.04756.

[2] David Applegate, Oliver Hinder, Haihao Lu, and Miles Lubin. Faster first-order primal-dual methods for linear programming using restarts and sharpness. *Mathematical Programming*, 201(1–2):133–184, October 2022. ISSN 1436-4646. doi: 10.1007/s10107-022-01901-9. URL http://dx.doi.org/10.1007/s10107-022-01901-9.

[3] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991. doi: 10.1002/cpa.3160440402.

[4] A Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 2011.

[5] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transportation distances, 2013. URL https://arxiv.org/abs/1306.0895.

[6] Montacer Essid and Justin Solomon. Quadratically-regularized optimal transport on graphs, 2018. URL https://arxiv.org/abs/1704.08200.

[7] Kilian Fatras, Younes Zine, Szymon Majewski, Rémi Flamary, Rémi Gribonval, and Nicolas Courty. Minibatch optimal transport distances; analysis and applications, 2021. URL https://arxiv.org/abs/2101.01792.

[8] Dirk A. Lorenz, Paul Manns, and Christian Meyer. Quadratically regularized optimal transport, 2019. URL https://arxiv.org/abs/1903.01112.

[9] Mahler Lorenz, D. H. Orlicz space regularization of continuous optimal transport problems. *Appl Math Optim 85*, 2022.

[10] Haihao Lu and Jinwen Yang. PDOT: a practical primal-dual algorithm and a gpu-based solver for optimal transport, 2024. URL https://arxiv.org/abs/2407.19689.

[11] Simone Di Marino and Augusto Gerolin. Optimal transport losses and sinkhorn algorithm with general convex regularization. 2020.

[12] Simone Di Marino and Augusto Gerolin. An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm. *Journal of Scientific Computing*, 85 (2):27, 2020.

[13] Boris Muzellec, Richard Nock, Giorgio Patrini, and Frank Nielsen. Tsallis regularized optimal transport and ecological inference, 2016. URL https://arxiv.org/abs/1609.04495.

[14] Marcel Nutz. Quadratically regularized optimal transport: Existence and multiplicity of potentials, 2025. URL https://arxiv.org/abs/2404.06847.

[15] Gabriel Peyré and Marco Cuturi. Computational optimal transport, 2020. URL https://arxiv.org/abs/1803.00567.

[16] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*. Birkhäuser, 2015.

[17] Cédric Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.

Appendix A. Suplementary material

A.1. Primal-Dual Hybrid Gradient (PDHG)

Let $\mathcal{X}$ and $\mathcal{Y}$ be Hilbert spaces with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. Consider the following optimization problem

$$\min_{x \in \mathcal{X}} f(x) + g(Kx) \tag{7}$$

where

- $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ are proper, lower-semicontinuous, convex functions,
- $K : \mathcal{X} \to \mathcal{Y}$ is a bounded linear operator.

(7) equals to the corresponding saddle-point formulation:

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x) + \langle Kx, y \rangle - g^*(y) \tag{8}$$

where $g^*(y) = \sup_{z \in \mathcal{Y}} \langle y, z \rangle - g(z)$ is conjugate of g [4]. Further we assume that there exists $(\hat{x}, \hat{y}) \in \mathcal{X} \times \mathcal{Y}$ a solution for (8) which satisfies

$$\begin{aligned} K\hat{x} &\in \partial g^*(\hat{y}), \\ -(K^*\hat{y}) &\in \partial f(\hat{x}) \end{aligned}$$

PDHG iterations [4] Choose step sizes $\tau, \sigma > 0$ such that

$$\tau\sigma\|K\|^2 < 1, \tag{9}$$

where $||K||$ is the operator norm of K. Given (x^k, y^k), one PDHG update reads

$$\begin{aligned} y^{k+1} &:= \operatorname{prox}_{\sigma g^*}\big(y^k + \sigma K \bar{x}^k\big), \\ x^{k+1} &:= \operatorname{prox}_{\tau f}\big(x^k - \tau K^* y^{k+1}\big), \\ \bar{x}^{k+1} &:= x^{k+1} + \theta\big(x^{k+1} - x^k\big), \qquad 0 \le \theta \le 1, \end{aligned} \tag{10}$$

where $\operatorname{prox}_{\lambda h}(z) := \arg\min_u \big\{h(u) + \frac{1}{2\lambda}\|u - z\|^2\big\}$ and K^* denotes the adjoint of K. The following theorem gives the result about convergence and covergence rates for vanilla PDHG Algorithm 1.

Theorem 1 (Pock-Chambolle [4]) *Let $\mathcal{X}, \mathcal{Y}$ be finite Hilbert spaces, assume $B_1 \subset \mathcal{X}, B_2 \subset \mathcal{Y}$, define the partial primal-dual gap*

$$G_{B_1 \times B_2}(x, y) = \max_{y' \in B_2} \langle y', Kx \rangle - g^*(y') + f(x) - \min_{x' \in B_1} \langle y, Kx' \rangle - g^*(y) + f(x'). \tag{11}$$

Choose $\theta = 1$, τ, σ such that $\tau\sigma\|K\|^2 < 1$, take $(x^k, \bar{x}^k, y^k)$ be defined by (10). Then $x_N = \frac{1}{N}\sum_{k=1}^N x_k$ and $y_N = \frac{1}{N}\sum_{k=1}^N y_k$, for any bounded $B_1 \times B_2 \subset \mathcal{X} \times \mathcal{Y}$ such that the saddle point of (8) $(\hat{x}, \hat{y}) \in B_1 \times B_2$, the parial gap has the following bound:

$$G_{B_1 \times B_2}(x_N, y_N) \le \frac{C(B_1, B_2)}{N} \tag{12}$$

where

$$C(B_1, B_2) = \sup_{(x,y) \in B_1 \times B_2} \frac{\|x - x^0\|}{2\tau} + \frac{\|y - y^0\|}{2\sigma}.$$

Moreover $x^k \to \hat{x}$ and $y^k \to \hat{y}$.

A.2. Caractherization of solution of convex-regularized Optimal Transport

The following theorem characterizes the primal-dual solutions for convex-regularized optimal transport in a compact subset, see also [8, 11, 12].

Theorem 2 *[11] Let $\varepsilon > 0$, $\Omega \subset \mathbb{R}^d$ be a compact subset and $\Phi \in C^1((0,\infty))$ is an entropy i.e. nonnegative lower semicontinuous convex and superlinear at infinity function, $\Psi = \Phi^*$ is convex conjugate. Let $\mu \in \mathcal{P}(\Omega), \nu \in \mathcal{P}(\Omega)$, $\xi \in \mathcal{P}(\Omega \times \Omega)$ a reference measure; then given $u^* \in L^\Phi(\Omega, \mu)$ and $v^* \in L^\Phi(\Omega, \nu)$, the following are equivalent*

- *(Existance of maximizers) u^* and v^* are maximizers for (3)*
- *(Complementary slackness) let $\pi^* = \Psi'\left(\frac{u(x)+v(y)-\|x-y\|_2^2}{\varepsilon}\right) \cdot \xi$, then $\pi^* \in \Pi(\mu,\nu)$*
- *(Duality) $W^2_{2,\Phi,\varepsilon}(\mu,\nu) = D_\varepsilon(u^*, v^*)$.*

Finally, the maximizers exists and $\pi^* = \Psi'\left(\frac{u(x)+v(y)-\|x-y\|_2^2}{\varepsilon}\right)$ is the unique minimizer of the problem (1) [11, Thoerem 3.4].

Remark. In Theorem 2 we use quadratic cost, but since the domain Ω is compact in $\mathbb{R}^d$, this cost is bounded, so [11, Thoerem 3.4] is applicable for our case.

A.3. Details on the implementation of PDHG for convex-regularized Optimal Transport

Given the grid points $x_i, y_j \in \Omega$, discrete measures $\mu = \sum_{i=1}^n \mu_i \delta_{x_i}$ and $\nu = \sum_{i=1}^n \nu_j \delta_{y_j}$ and cost matrix $c \in \mathbb{R}^{n\times n}$ defined by $c_{ij} = \|x_i - y_j\|_2^2$, we define the reference measure $\xi_{i,j}$ to be the product $\mu_i\nu_j$ either the unfirom measure over discrete set of points. Then the discrete version of the problem (1) is defined as

$$\min_{\pi \in \mathbb{R}_+^{n\times n}} \left\{ \sum_{ij} c_{ij}\pi_{ij} + \varepsilon \sum_{ij} \Phi\left(\frac{\pi_{ij}}{\mu_i\nu_j}\right)\mu_i\nu_j : \pi 1_n = \mu, \pi^\top 1_n = \nu \right\} \tag{13}$$

which equals to the corresponding Dual Problem

$$\max_{u,v\in\mathbb{R}^n} \left\{ \sum_i u_i\mu_i + \sum_j v_j\nu_j - \varepsilon \sum_{ij} \Psi\left(\frac{u_i + v_j - c_{ij}}{\varepsilon}\right)\mu_i\nu_j \right\}. \tag{14}$$

Let us recast the discrete OT problem (13) as the saddle-point problem (8):

$$\mathcal{X} := \mathbb{R}^{n\times n}, \quad \mathcal{Y} := \mathbb{R}^{2n}, \quad x := \pi \in \mathcal{X}, \quad y := (u,v) \in \mathcal{Y}.$$

- *Primal objective $f : \mathbb{R}^{n\times n} \to \mathbb{R} \cup \{+\infty\}$:*

$$f(\pi) = \langle c, \pi\rangle \; + \; \varepsilon \sum_{i,j} \Phi\Big(\tfrac{\pi_{ij}}{\mu_i\nu_j}\Big)\mu_i\nu_j \; + \; \iota_{\mathbb{R}_+^{n\times n}}(\pi), \tag{15}$$

where $\iota_{\mathbb{R}_+^{n\times n}}(z) = 0$ if $z \geq 0$ entrywise and $+\infty$ otherwise.

- *Linear operator* $K : \mathbb{R}^{n\times n} \to \mathbb{R}^{2n}$:

$$K\pi := \begin{pmatrix} \pi \mathbf{1}_n \\ \pi^\top \mathbf{1}_n \end{pmatrix} = \begin{pmatrix} \mathbf{1}_n^\top \otimes \mathbf{1}_n \\ \mathbf{1}_n \otimes \mathbf{1}_n^\top \end{pmatrix} \mathrm{vec}(\pi) \tag{16}$$

- *Marginal constraints function* $g : \mathbb{R}^{2n} \to \mathbb{R} \cup \{+\infty\}$:

$$\begin{aligned} g(a,b) = \iota_{\{\mu\}}(u) + \iota_{\{\nu\}}(v) &= \begin{cases} 0, & a = \mu,\ b = \nu, \\ +\infty, & \text{otherwise} \end{cases} \\ g(K\pi) &= \begin{cases} 0, & \pi 1_n = \mu, \pi^\top 1_n = \nu, \\ +\infty, & \text{otherwise.} \end{cases} \end{aligned} \tag{17}$$

with convex conjugate

$$g^*(u,v) = \langle \mu, u\rangle + \langle \nu, v\rangle \quad ((u,v)^\top \in \mathbb{R}^{2n}). \tag{18}$$

Computing proximals

$$\begin{aligned} \mathrm{prox}_{\sigma g^*}(u,v) = \mathrm{argmin}_{u',v'} \left\{ \frac{1}{2}\|u-u'\|^2 + \frac{1}{2}\|v-v'\|^2 + \sigma\langle u', \mu\rangle + \sigma\langle v', \nu\rangle \right\} = \\ = \begin{pmatrix} u - \sigma\mu \\ v - \sigma\nu \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \mathrm{prox}_{\tau f}(\pi) = \mathrm{argmin}_{\gamma \in \mathbb{R}_+^{n\times n}} \left\{ \sum_{ij} \frac{1}{2}(\pi_{ij} - \gamma_{ij})^2 + \tau \sum_{ij} c_{ij}\gamma_{ij} + \tau \sum_{ij} \varepsilon \Phi\left(\frac{\gamma_{ij}}{\mu_i \nu_j}\right) \mu_i \nu_j \right\} \\ \left[\mathrm{prox}_{\tau f}(\pi)\right]_{ij} = \mu_i \nu_i \,\mathrm{prox}_{\lambda_{ij}\Phi}\left(\frac{\pi_{ij} - \tau c_{ij}}{\mu_i \nu_j}\right), \quad \lambda_{ij} = \frac{\tau\varepsilon}{\mu_i \nu_j}. \end{aligned} \tag{19}$$

In the case of chi-square regularization, i.e. $\Phi(t) = \frac{1}{2}t_+^2$, the proximal map (19) can be computed explicitly

$$\mathrm{prox}_{\lambda_{ij}\Phi}\left(\frac{\pi_{ij} - \tau c_{ij}}{\mu_i \nu_j}\right) = \left(\frac{\pi_{ij} - \tau c_{ij}}{\mu_i \nu_j + \tau\varepsilon}\right)_+. \tag{20}$$

And the proximal update formula is

$$\left[\mathrm{prox}_{\tau f}(\pi)\right]_{ij} = \frac{\mu_i \nu_j}{\mu_i \nu_j + \tau\varepsilon}(\pi_{ij} - \tau c_{ij})_+ = \left(\frac{\pi_{ij} - \tau c_{ij}}{1 + \lambda_{ij}}\right)_+ \tag{21}$$

Combining this updates, we can define the Chambolle-Pock's PDHG algorithm for discrete Optimal Transport (13) as Algorithm 1.

A.4. PDHG Algorithm 2

To improve convergence speed, stability, and robustness to step-size selection, we adopt a restarted, adaptive variant of PDHG. The approach originates with Applegate et al. [1] for general-purpose linear programming and was recently adapted to Optimal Transport by Haihao Lu, Jinwen Yang [10]. We further develop a Restarted Adaptive PDHG for chi-square regularized OT, Algorithm 2 presents a pseudocode of this algorithm along with adpative step function (Algorithm 3) and primal weight update (Algorithm 4).

Algorithm 1: Vanilla PDHG for regularised OT

Data: $\mu, \nu \in \mathbb{R}_+^d$;
cost matrix $C \in \mathbb{R}^{d\times d}$;
step sizes $\tau, \sigma > 0$;
regularisation parameter $\varepsilon > 0$;
π^0, u^0, v^0 initial values;
while $\|\pi^k \mathbf{1} - \mu\| + \|(\pi^k)^\top \mathbf{1} - \nu\| > \mathrm{tol}$ **do**

$\pi_{ij}^{k+1} \leftarrow \mu_i \nu_j \ \mathrm{prox}_{\lambda_{ij}\Phi}\Big(\frac{\pi_{ij}^k - \tau\big(C_{ij} - (u_i^k + v_j^k)\big)}{\mu_i \nu_j}\Big)$;

$\bar{\pi}^{k+1} \leftarrow \pi^{k+1} + \theta\big(\pi^{k+1} - \pi^k\big)$;

$u_i^{k+1} \leftarrow u_i^k + \sigma\Big(\mu_i - \sum_j \bar{\pi}_{ij}^{k+1}\Big)$;

$v_j^{k+1} \leftarrow v_j^k + \sigma\Big(\nu_j - \sum_i \bar{\pi}_{ij}^{k+1}\Big)$;

end

Algorithm 2: Adaptive Restarted PDHG for chi-square regularized OT

Data: marginals $\mu \in \mathbb{R}_+^m$, $\nu \in \mathbb{R}_+^n$;
cost $C \in \mathbb{R}^{m\times n}$;
reg. $\varepsilon > 0$, tolerance $\texttt{tol} \geq 0$
initial $z^{0,0} = (\pi^{0,0}, u^{0,0}, v^{0,0})$;
step size $\hat{\eta}^{0,0} \leftarrow 1/\|K\|_\infty$, primal weight $\omega^0 > 0$.
initialize outer counter $n \leftarrow 0$, total inner steps $k \leftarrow 0$;
repeat

$t \leftarrow 0$;

repeat

$(z^{n,t+1}, \eta^{n,t+1}, \hat{\eta}^{n,t+1}) \leftarrow \texttt{AdaptiveStepOfPDHG\_OT}(z^{n,t}, \omega^n, \hat{\eta}^{n,t}, k, \varepsilon)$;

$\bar{z}^{n,t+1} \leftarrow \frac{1}{\sum_{i=1}^{t+1}\eta^{n,i}} \sum_{i=1}^{t+1} \eta^{n,i} z^{n,i}$;

$z_c^{n,t+1} \leftarrow \texttt{GetRestartCandidate}\big(z^{n,t+1}, \bar{z}^{n,t+1}\big)$;

$t \leftarrow t+1$, $k \leftarrow k+1$;

until *restart or termination criteria hold*;

$z^{n+1,0} \leftarrow z_c^{n,t}$, $n \leftarrow n+1$; `// restart the outer loop`

$\omega^n \leftarrow \texttt{PrimalWeightUpdate}\big(z^{n,0}, z^{n-1,0}, \omega^{n-1}\big)$;

until *termination criteria hold*;
Output: $z^{n,0}$.

Algorithm 3: `AdaptiveStepOfPDHG_OT`$(z^{n,t}, \omega^n, \hat{\eta}^{n,t}, k, \varepsilon)$

Input: $z^{n,t} = (\pi, u, v)$;
primal weight $\omega^n > 0$;
current step $\hat{\eta}^{n,t} > 0$;
counter k;
reg. $\varepsilon > 0$
Output: $(z' = (\pi', u', v'),\ \eta,\ \hat{\eta}')$
$(\pi, u, v) \leftarrow z^{n,t}; \quad \eta \leftarrow \hat{\eta}^{n,t};$
for $i = 1, 2, \ldots 10$ **do**

$$\pi'_{ij} \leftarrow \left(\frac{\pi^k_{ij} - \tau\left(C_{ij} - (u^k_i + v^k_j)\right)}{1 + \varepsilon\frac{\eta}{\omega^n}}\right);$$

$$u^{k+1}_i \leftarrow u^k_i + \eta\omega^n\Big(\mu_i - \textstyle\sum_j \left(2\pi'_{ij} - \pi_{ij}\right)\Big);$$

$$v^{k+1}_j \leftarrow v^k_j + \eta\omega^n\Big(\nu_j - \textstyle\sum_i \bar{\pi}^{k+1}_{ij}\Big);$$

$$\bar{\eta} \leftarrow \frac{\|(\pi' - \pi,\ u' - u,\ v' - v)\|^2}{2\left\langle (u' - u,\ v' - v),\ K(\pi' - \pi)\right\rangle};$$

$$\hat{\eta}' \leftarrow \min\Big((1 - (k+1)^{-0.3})\,\bar{\eta},\ (1 + (k+1)^{-0.6})\,\eta\Big);$$

  if $\eta \le \bar{\eta}$ **then**
    return $(\,(\pi', u', v'),\ \eta,\ \hat{\eta}'\,)$
  end
  $\eta \leftarrow \hat{\eta}';$
end

Algorithm 4: `PrimalWeightUpdate`$\left(z^{n,0}, z^{n-1,0}, \omega^{n-1}\right)$

$\Delta^n_\pi = \|\pi^{n,0} - \pi^{n-1,0}\|_2, \quad \Delta^n_{(u,v)} = \|u^{n,0} - u^{n-1,0}\|_2 + \|v^{n,0} - v^{n-1,0}\|;$
if $\Delta^n_\pi > \varepsilon_{\text{zero}}$ **and** $\Delta^n_{(u,v)} > \varepsilon_{\text{zero}}$ **then**

$$\textbf{return } \exp\Big(\theta\log\big(\tfrac{\Delta^n_{(u,v)}}{\Delta^n_\pi}\big) + (1-\theta)\log(\omega^{n-1})\Big)$$

else
  return ω^{n-1}
end

Adaptive restarts strategy. We adaptively restart the PDHG algorithm each outer iteration. Algorithm 2 initially selects a restart candidate at each outer iteration, choosing between the current iterate and the average iterate based on a greedy principle. A restart criteria is assessed to determine if there is a constant factor decay in the progress metric. If such decay is observed, a restart is triggered. As a metric of measuring the progress we use the marginal error

$$\texttt{MarginalError}(\pi) := \|\pi\mathbf{1} - \mu\|_2^2 + \|\pi^\top\mathbf{1} - \nu\|_2^2 = \|K\,\pi - (\mu,\nu)^\top\|_2^2. \tag{22}$$

More specifically, the restart scheme is described as follows

Restart criteria Given the parameters $\beta_{\text{necessary}} \in (0,1)$ and $\beta_{\text{sufficient}} \in (0, \beta_{\text{necessary}})$. Default parameters we use is $\beta_{\text{necessary}} = 0.8$, $\beta_{\text{sufficient}} = 0.2$. The algorithm restarts if the following holds:

(1) (*Sufficient decay in Marginal Error*)

$$\texttt{MarginalError}\big(\pi_c^{n,t+1}\big) \le \beta_{\text{sufficient}}\texttt{MarginalError}(\pi^{n,0})$$

(2) (*Necessary decay + no local progress in Marginal error*)

$$\begin{cases}\texttt{MarginalError}\big(\pi_c^{n,t+1}\big) \le \beta_{\text{necessary}}\texttt{MarginalError}(\pi^{n,0})\\ \texttt{MarginalError}\big(\pi_c^{n,t+1}\big) \ge \texttt{MarginalError}(\pi^{n,t})\end{cases}$$

Choosing restart candidate

$$\texttt{GetRestartCandidate}\big(z^{n,t+1},\,\bar{z}^{n,t+1}\big) := \begin{cases} z^{n,t+1}, & \texttt{MarginalErr}\big(\pi^{n,t+1}\big) < \texttt{MarginalErr}\big(\bar{\pi}^{n,t+1}\big),\\ \bar{z}^{n,t+1}, & \text{otherwise.}\end{cases}$$

We stop the algorithm when primal feasibility is satisfied up to given tolerance, i.e. the plan has right marginals up to certain error.

$$\texttt{MarginalError}\big(\pi\big) \le \texttt{tol}$$

This error is computational much cheaper then, for instance, KKT error introduced in [10].

Experiments on 1D distributions and 2D Gaussian are given in Table 1. This table shows the precision and computational time for the Algorithm 2 (PDHG) and Sinkhorn. Both implemented in JAX for and executed on H100 GPU.

Test	Reg.	M	Time	Err.	Err. Bar
G-4096	$1e^{-1}$	sink	0.03	8.89	$7.3e^{-3}$
	$1e^{-2}$	sink	0.24	2.6	$7.2e^{-4}$
	$1e^{-3}$	sink	1.73	0.072	$2.3e^{-5}$
	$1e^{-4}$	sink	20.20	0.013	$1.1e^{-5}$
	$1e^{5}$	pdhg	0.46	1.04	$1.2e^{-3}$
	$1e^{3}$	pdhg	0.82	0.10	$3.1e^{-5}$
	$1e^{2}$	pdhg	2.70	0.009	$3.5e^{-6}$
	$1e^{1}$	pdhg	6.13	0.002	$1.3e^{-6}$
	$1e^{-1}$	pdhg	8.36	$8.8e^{-4}$	$1.7e^{-6}$
	$1e^{-3}$	pdhg	9.22	0.002	$2.3e^{-6}$
G-8192	$1e^{-1}$	sink	0.07	8.9	$7.3e^{-3}$
	$1e^{-2}$	sink	0.64	2.55	$7.2e^{-4}$
	$1e^{-3}$	sink	4.86	0.07	$2.3e^{-5}$
	$1e^{-4}$	sink	56.19	0.012	$1.1e^{-5}$
	$1e^{5}$	pdhg	1.32	0.79	$1.2e^{-3}$
	$1e^{3}$	pdhg	4.58	0.03	$3.1e^{-5}$
	$1e^{2}$	pdhg	7.78	0.10	$3.5e^{-6}$
	$1e^{1}$	pdhg	11.32	0.004	$1.3e^{-6}$
	$1e^{-1}$	pdhg	30.13	0.002	$8.1e^{-5}$
	$1e^{-3}$	pdhg	60.24	0.001	$8.1e^{-5}$

Test	Reg.	M	Time	Err.	Err. Bar
C-4096	$1e^{-1}$	sink	0.026	39.91	0.0008
	$1e^{-2}$	sink	0.25	1.23	0.0004
	$1e^{-3}$	sink	2.58	0.085	$2.3e^{-5}$
	$1e^{-4}$	sink	27.20	0.002	$2.0e^{-6}$
	$1e^{5}$	pdhg	0.43	0.89	0.001
	$1e^{3}$	pdhg	0.78	0.031	$2.1e^{-5}$
	$1e^{2}$	pdhg	1.48	0.011	$4.8e^{-6}$
	$1e^{1}$	pdhg	2.63	0.003	$2.0e^{-6}$
	$1e^{-1}$	pdhg	5.60	$5.0e^{-4}$	$1.7e^{-6}$
	$1e^{-3}$	pdhg	8.14	$9.8e^{-4}$	$6.7e^{-7}$
G2D-64	$1e^{-1}$	sink	0.03	4.03	
	$1e^{-2}$	sink	0.26	0.72	
	$1e^{-3}$	sink	2.893	0.019	
	$1e^{-4}$	sink	29.59	$7.6e^{-5}$	
	$1e^{5}$	pdhg	0.46	2.09	
	$1e^{3}$	pdhg	0.56	0.21	
	$1e^{2}$	pdhg	0.97	0.14	
	$1e^{1}$	pdhg	1.80	0.012	
	$1e^{-1}$	pdhg	3.066	$2.3e^{-4}$	
	$1e^{-3}$	pdhg	3.99	$2.0e^{-4}$	

Table 1: Comparison of Sinkhorn and PDHG for the cases: 1D Gaussians 4096 and 8192 points (**G-8192** and **G-4096**), Cauchy 4096 points (**C-4096**) and Gaussian 2D 64×64 (**G2D-64**). Columns: **Test** – instance name; **Reg.** – regularisation ε; **M** – method; **Time** – runtime (s); **Err.** – cost error; **Err. Bar** – push-forward error (W_2) via the barycentric map (23).

A.5. Statistical properties

Approximation of the Monge Map Given the discrete measures $\mu = \sum_{i=1}^{n} a_i \, \delta_{x_i}$, $\nu = \sum_{j=1}^{m} b_j \, \delta_{y_j}$ were $a_i, b_j > 0$, $x_i, y_j \in \mathbb{R}^d$, and the regularised optimal plans $\pi_\varepsilon^{\text{ent}}$ and $\pi_\varepsilon^{\text{chi}^2}$ (PDHG and Sinkhorn respectively) together with their associated dual potentials $(u^{\text{ent}}, v^{\text{ent}})$ and $(u^{\text{chi}^2}, v^{\text{chi}^2})$. We estimate the Monge map $T : \text{supp}(\mu) \to \text{supp}(\nu)$ via two standard constructions:

(1) **Barycentric projection.**

$$T_1(x_i) \;=\; \frac{1}{a_i} \sum_{j=1}^{m} \pi_{ij}^{\text{reg}} \, y_j, \qquad i = 1, \ldots, n, \tag{23}$$

where "reg" stands for either "ent" or "chi^2".

(2) **Gradient of the discrete potential.** Define the convex Brenier potential $\psi^{\text{reg}}(x_i) := \frac{1}{2}\|x_i\|_2^2 - u_i^{\text{reg}}$. We interpolate ψ^{reg} to obtain a continuously differentiable function $\tilde{\psi}^{\text{reg}}$ and set

$$T_2(x_i) \;=\; \nabla \tilde{\psi}^{\text{reg}}(x_i). \tag{24}$$

Error Metrics For measuring the error of an estimator T_M with $M \in \{1, 2\}$, we form the push-forward measure $\hat{\nu}^M = (T_M)_{\#}\mu = \sum_{i=1}^{n} a_i \, \delta_{T_M(x_i)}$. Approximation quality is assessed by

$$\text{Err}_{\ell^2}^{\text{reg},M} = \| \hat{\nu}^M - \nu \|_{\ell^2}^2, \tag{25}$$

$$\text{Err}_{W_2}^{\text{reg},M} = W_2\big(\hat{\nu}^M, \, \nu\big). \tag{26}$$

Estimating the potential, vector field, and forward gradient–flow step (1D Gaussians). Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and let $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, $\widehat{\nu}_n = \frac{1}{n}\sum_{j=1}^M \delta_{y_j}$ be empirical measures with i.i.d. samples $x_i \sim \mu$, $y_j \sim \nu$ and uniform weights. We consider the quadratic cost $c(x,y) = \frac{1}{2}(x-y)^2$.

For $\mu = \mathcal{N}(m_1, \sigma_1^2)$ and $\nu = \mathcal{N}(m_2, \sigma_2^2)$ define $a := \sigma_2/\sigma_1$ and $b := m_2 - a\, m_1$. Then the optimal transport map and the associated vector field are

$$T^\star(x) = a\,x + b, \qquad g^\star(x) = T^\star(x) - x = (a-1)x + b.$$

A Kantorovich potential (up to an additive constant) is

$$\phi^\star(x) = \tfrac{1}{2}(1-a)\,x^2 - b\,x.$$

These will serve as references for statistical evaluation.

We write a regularized Wasserstein distance between $\widehat{\mu}_n$ and $\widehat{\nu}_n$ as

$$W^2_{2,\Phi,\varepsilon}(\widehat{\mu}_n, \widehat{\nu}_n) = \min_{\pi \in \mathbb{R}_+^{n\times n}} \Big\{ \langle C, \pi\rangle + \varepsilon \Phi(\pi) \;:\; \pi \mathbf{1} = \tfrac{1}{n}\mathbf{1},\ \pi^\top \mathbf{1} = \tfrac{1}{n}\mathbf{1} \Big\},$$

where $C_{ij} = \frac{1}{2}(x_i - y_j)^2$, Φ is either *Shannon entropy* or *chi-sqare*. Let $(\widehat{\pi}_{\Phi,\varepsilon}, \widehat{u}_{\Phi,\varepsilon}, \widehat{v}_{\Phi,\varepsilon})$ denote an optimal primal–dual triplet. PDHG or Sinkhorn solver returns a coupling $\widehat{\pi}_{\Phi,\varepsilon}$ and dual vectors $(\widehat{u}_{\Phi,\varepsilon}, \widehat{v}_{\Phi,\varepsilon})$.

(i) Potential estimator. Potentials are defined up to an additive constant. When comparing to $\phi^\star$, we fit the best constant $c^\star$ in least squares: $c^\star = \arg\min_c \frac{1}{N}\sum_i (\widehat{u}_{\Phi,\eta,i} - \phi^\star(x_i) - c)^2$.

(ii) Monge map / vector field estimator (barycentric projection). Given a coupling $\widehat{\pi}_{\Phi,\varepsilon}$, the barycentric projection yields an estimator of the map at sample locations:

$$\widehat{T}_{\Phi,\varepsilon}(x_i) \;=\; \frac{\sum_{j=1}^M \widehat{\pi}_{\Phi,\varepsilon,ij}\, y_j}{\sum_{j=1}^M \widehat{\pi}_{\Phi,\varepsilon,ij}}, \qquad \widehat{g}_{\Phi,\varepsilon}(x_i) \;=\; \widehat{T}_{\Phi,\varepsilon}(x_i) - x_i.$$

(iii) Displacement interpolation. For a step size $\tau \in (0,1]$, the explicit Euler update of particles is

$$x_i^{(\tau)} \;=\; x_i + \tau\, \widehat{g}_{\Phi,\varepsilon}(x_i) \;=\; (1-\tau)\, x_i + \tau\, \widehat{T}_{\Phi,\varepsilon}(x_i),$$

and the pushed empirical measure is

$$\widehat{\mu}_\tau^{\Phi,\varepsilon} \;=\; \frac{1}{N}\sum_{i=1}^N \delta_{x_i^{(\tau)}}.$$

In the Gaussian reference case, $\mu_\tau = \mathcal{N}\big((1-\tau)m_s + \tau m_t,\ (1-\tau)\sigma_s + \tau\sigma_t\big)$, which provides a closed–form target for evaluating the one–step discrepancy.

Error Metrics We measure the error of the estimators in the following way

$$\begin{aligned} \mathrm{MSE}_\phi &= \mathbb{E}_{x\sim\mu}\left[\hat{u}_{\Phi,\varepsilon}(x) - \phi^*(x) - c^*\right]^2 \\ \mathrm{MSE}_g &= \mathbb{E}_{x\sim\mu}\left[\hat{g}_{\Phi,\varepsilon}(x) - g^*(x)\right]^2 \\ \mathrm{MSE}_{\text{displacement}} &= \mathbb{E}_{x\sim\mu} W_2^2(\hat{\mu}_\tau^{\Phi,\varepsilon}(x), \mu_\tau) \end{aligned} \tag{27}$$

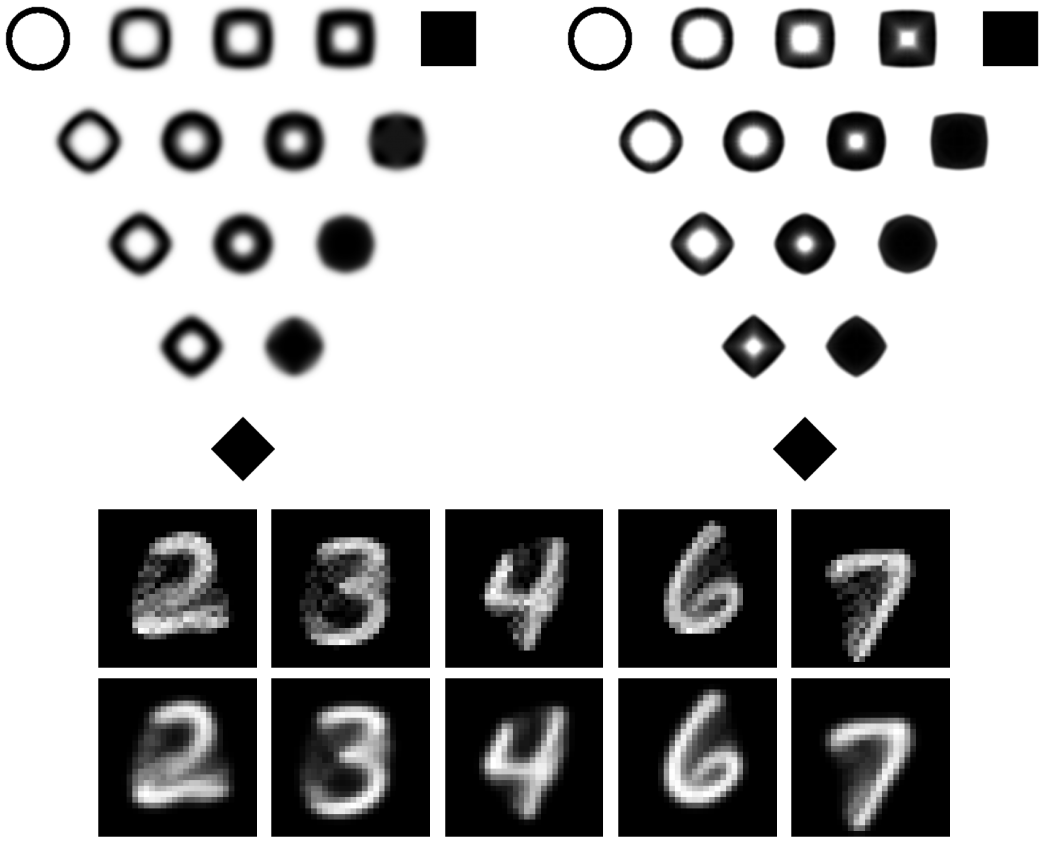


Figure 6: Highigher resoltuion version of Figure 4