Optimal Transport with Proximal Splitting

Nicolas Papadakis, Gabriel Peyré, Edouard Oudet

Hamza Golubovic, Jonathan Friesen, Alex Albors Juez

University of Washington

May 30th, 2024

Problem formulation

- Benamou-Brenier's formulation [1] is given by :

$$\begin{aligned} \underset{(m,f)\in\mathcal{C}}{\text{minimize}} \quad & \int_{[0,1]^d}\int_0^1 J(m,f)\,dtdx \\ \text{subject to} \quad & J(m,f) = \begin{cases} \frac{\|m\|^2}{2f} & \text{if } f>0, \\ 0 & \text{if } (m,f)=(0,0)\ , \\ \infty & \text{otherwise} \end{cases} \\ & (m,f)\in\mathcal{C}_0 \end{aligned}$$

$$\mathcal{C}_0 = \{\,(m,f);\ \partial_t f + \mathrm{div}_x(m) = 0,\ m(0,\cdot) = m(1,\cdot) = 0,\ f(\cdot,0) = f^0,\ f(\cdot,1) = f^1\,\}$$

- Changed variables $(v,f)\mapsto(m,f)$ to obtain convexity.

Problem discretization: Staggered grids $\mathcal{G}_s^x$, $\mathcal{G}_s^t$

- We describe the case $(x,t) \in [0,1] \times [0,1]$ for simplicity.

$$\mathcal{G}_c = \{(x_i = i/N,\ t_j = j/P) \in [0,1]^2\ ;\ 0 \le i \le N, 0 \le j \le P\}$$

$$\mathcal{G}_s^x = \Big\{\Big(x_i = (i+1/2)/N, t_j = j/P\Big)\ ;\ -1 \le i \le N, 0 \le j \le P\Big\},$$

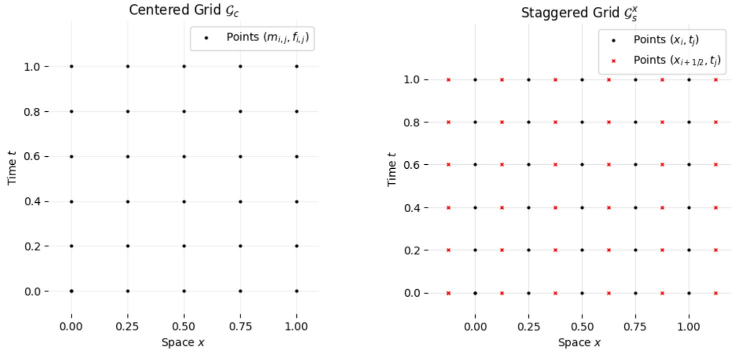


Figure 1: Centered and Staggered Grids

Discretized operators

1. Interpolator $(\mathcal{G}_s^t, \mathcal{G}_s^x) \to \mathcal{G}_c$:

$$\mathcal{I}(U)_{ij} = \begin{cases} m_{i,j} & = (\bar{m}_{i-1,j} + \bar{m}_{i,j})/2, \\ f_{i,j} & = (\bar{f}_{i,j-1} + \bar{f}_{i,j})/2. \end{cases}$$

2. Boundary operator (acting on staggered grid):

$$b(U) = \left(\left(\bar{m}_{-1,j}, \bar{m}_{N,j} \right)_{j=0}^{P}, \left(\bar{f}_{i,-1}, \bar{f}_{i,P} \right)_{i=0}^{N} \right)$$

3. Divergence operator:

$$\operatorname{div}(U)_{i,j} = N(\bar{m}_{i,j} - \bar{m}_{i-1,j}) + P(\bar{f}_{i,j} - \bar{f}_{i,j-1}).$$

4. Abbreviate conditions with $AU = (\operatorname{div}(U), b(U))$ and $y = (0, 0, 0, f^0, f^1)$ so that $AU = y$.

Discretized objective

$$\underset{U \in \mathcal{E}_s}{\text{minimize}} \quad \sum_{k \in \mathcal{G}_c} J(\mathcal{I}(U)_k) + \iota_{\mathcal{C}}(U)$$

with

$$\mathcal{C} = \{U \in \mathcal{G}_s \,;\, \operatorname{div}(U) = 0 \text{ and } b(U) = b_0\}$$

now becomes

$$\underset{(U,V) \in (\mathcal{E}_s, \mathcal{E}_c)}{\text{minimize}} \quad \sum_{k \in \mathcal{G}_c} J(V) + \iota_{\mathcal{C}}(U) + \iota_{\mathcal{C}_{s,c}}((U,V))$$

with the new set constraint

$$\mathcal{C}_{s,c} = \{z = (U,V) \in \mathcal{E}_s \times \mathcal{E}_c \,:\, V = \mathcal{I}(U)\}$$

Intro to Proximal algorithms

- Given a closed, proper convex, l.s.c. function $f : \mathcal{H} \to \mathbb{R}$.

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad f(x)$$

- (Sub)-Gradient descent does $x_{t+1} = x_t - \gamma \partial f(x_t)$.
- Can we do better if we know structure of f?

$$\text{i.e.} \qquad f(x) = \sum_j f_j(x_j) + g(x), \;\; f(x) = g(x) + h(x)$$

- **Proximal operator**: $\mathrm{Prox}_{f,\gamma}(x) = \arg\min_{z \in \mathcal{H}} f(z) + \frac{1}{2\gamma} \|x - z\|^2$.

Understanding the Proximal operator

- Intuition: Minimize f while staying close to a given point.
- Proximal operator can be thought of as implicit gradient descent:

$$\mathrm{Prox}_{\gamma,f}(x) = (I + \gamma\partial f)^{-1}(x)$$

- The update $x_{t+1} = \mathrm{Prox}_{\gamma,f}(x_t) \implies x_{t+1} = x_t - \gamma\partial f(x_{t+1})$.
- $x = \mathrm{Prox}_{\gamma,f}(x) \iff x$ is a minimizer.
- $(I + \gamma\partial f)^{-1}$ is known as the **Resolvent** $R_{\gamma,\partial f}$ in the broader context of monotone operators. For our purposes, $R_{\gamma,\partial f} = \mathrm{Prox}_{\gamma f}$.
- $C_{\gamma,\partial f} = 2R_{\gamma,\partial f} - I = 2\,\mathrm{Prox}_{\gamma f} - I$ is called the **Cayley** operator [4].

Proximal mappings

$$\underset{z \in \mathcal{H}}{\text{minimize}} \quad G_1(z) + G_2(z)$$

- $C_{\gamma,\partial G_2}(x) = 2\,\mathrm{Prox}_{\gamma,G_2}(x) - x.$

Theorem

$0 \in \partial(G_1 + G_2)(x) = \partial G_1(x) + \partial G_2(x)$ *if and only if* $C_{\gamma,\partial G_1} \circ C_{\gamma,\partial G_2}(z) = z$ *where* $x = \mathrm{Prox}_{\gamma,G_2}(z)$. *[3]*

- For f closed, proper convex, $C_{\gamma,\partial G_1} \circ C_{\gamma,\partial G_2}$ is non-expansive.
- Banach fixed point convergence is not ensured. (Consider rotations or $x \mapsto -x$).
- What to do?

Douglas-Rachford Iteration

$$\underset{z \in \mathcal{H}}{\text{minimize}} \quad G_1(z) + G_2(z)$$

- Consider the average operator $\frac{1}{2}(I + C_{\partial G_1} \circ C_{\partial G_2})$ instead of $C_{\partial G_1} \circ C_{\partial G_2}$. Same fixed points, but obtain convergence at rate $1/k$ [3]
- In the language of proximal operators, this amounts to

$$\begin{cases} z_{k+1} &= z_k + \mathrm{Prox}_{\gamma G_1}(2x_{k+1} - z_k) - x_k \\ x_{k+1} &= \mathrm{Prox}_{\gamma G_2}(z_{k+1}) \end{cases}$$

Yields $(z_k, x_k) \in \mathcal{H}^2$ with $x_k \to x^\star$.

Proximal operators for Benamou-Brenier

- Asymmetric Douglas-Rachford (A-DR)

$$\underset{(U,V)\in(\mathcal{E}_s,\mathcal{E}_c)}{\text{minimize}} \quad \underbrace{\sum_{k\in\mathcal{G}_c} J(V) + \iota_{\mathcal{C}}(U)}_{G_1} + \underbrace{\iota_{\mathcal{C}_{s,c}}((U,V))}_{G_2}$$

- Can obtain another version swapping the roles of G_1 and G_2.
- Symmetric Douglas-Rachford (S-DR)

$$\underset{(U,V,\tilde{U},\tilde{V})\in(\mathcal{E}_s,\mathcal{E}_c)^2}{\text{minimize}} \quad \underbrace{\sum_{k\in\mathcal{G}_c} J(V) + \iota_{\mathcal{C}}(U) + \iota_{\mathcal{C}_{s,c}}((\tilde{U},\tilde{V}))}_{G_1} + \underbrace{\iota_{\{U=\tilde{U},V=\tilde{V}\}}}_{G_2}$$

ADMM and DR

- We can cast the problem to have the form

$$\min \sum_{k\in\mathcal{G}_c} J(V) + \iota_{\mathcal{C}}(V) = \underbrace{\sum_{k\in\mathcal{G}_c} J(V)}_{F} + (\underbrace{\iota_{\{y\}}}_{G} \circ A)(V)$$

With dual problem given by

$$\max_{z\in\mathcal{G}_c^*} -F^*(-A^*z) - G^*(z) = \min_{z\in\mathcal{G}_c^*} \underbrace{F^*(-A^*z)}_{G_2} + \underbrace{G^*(z)}_{G_1}$$

- Applying Douglas-Rachford to this

$$\begin{cases} s_{\ell+1} & = \mathrm{Prox}_{F^*\circ(-A^*)/\gamma}(-A^*q_{(\ell)} - u_{(\ell)}), \\ q_{\ell+1} & = \mathrm{Prox}_{G^*/\gamma}(-A^*s_{\ell+1} + u_\ell) \\ u_{\ell+1} & = u_\ell + (-A^*s_{\ell+1} - q_{\ell+1}) \end{cases} \qquad \text{which is ADMM}$$

Primal Dual for Benamou-Brenier

$$\underset{z \in \mathcal{H}}{\text{minimize}} \quad G_1(z) + G_2 \circ A(z)$$

- In the context of Benamou-Brenier:
  $G_2 = \sum J(m_k, f_k),\ A = \mathcal{I},\ G_1 = \iota_C$.
- We iteratively compute a sequence $(U^{(l)}, \Upsilon^{(l)}, V^{(l)})$ via

$$\begin{cases} V^{(l+1)} & = \text{Prox}_{\mathcal{J}^*}(V^{(l)} + \mathcal{I}(\Upsilon^{(l)})) \\ U^{(l+1)} & = \text{Prox}_{\iota_C^*}(V^{(l)} - \mathcal{I}^*(V^{(l+1)})) \\ \Upsilon^{(l+1)} & = U^{(l+1)} + \theta(U^{(l+1)} - U^{(l)}) \end{cases}$$

- Can be seen as a preconditioned ADMM [2] with $U^{(l+1)} \to U^\star$.
- $\mathcal{J}^*$ represents Fenchel conjugate function.

Proximal operators for Benamou-Brenier

- Proximal of indicator is projection onto set: $\mathrm{Prox}_{\iota_C} = \mathrm{Proj}_C$.

$$\mathrm{Proj}_C = (I - A(AA^*)^{-1})A^*)^{-1} + A^*(AA^*)^{-1}y$$
$$\mathrm{Proj}_{C_{s,c}} = (I + \mathcal{I}^*\mathcal{I})^{-1}(U + \mathcal{I}^*(V))$$

- Prox_{G_1}? $G_1 = \sum J(m_k, f_k) + \iota_C$ The function is separable!

$$\mathrm{Prox}_{G_1}(U, V) = (\mathrm{Proj}_C(U), \mathrm{Prox}_J(V_k)_{k\in\mathcal{G}_c})$$

- Finally, need to compute $\mathrm{Prox}_{\gamma J}(V_k)$:

$$\mathrm{Prox}_{\gamma J}(\tilde{m}, \tilde{f}) = \begin{cases} \left(\frac{f^\star \tilde{m}}{f^\star + \gamma}, f^\star\right) & \text{if } f^\star > 0 \\ (0, 0) & \text{otherwise} \end{cases}$$

and $f^\star$ largest real root of $P(x) = (x - \tilde{f})(x + \gamma)^2 - \frac{\gamma}{2}\|\tilde{m}\|^2 = 0$.

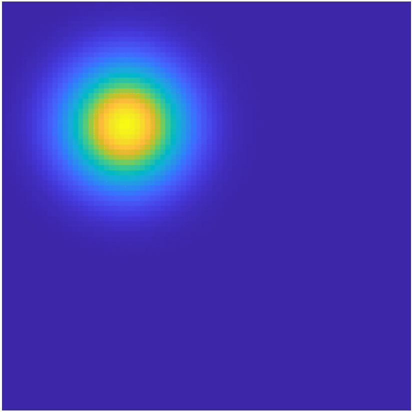
Generalized Objective

$$\mathcal{J}_\beta^w(V) := \sum_{k \in \mathcal{G}_c} w_k J_\beta(m_k, f_k)$$

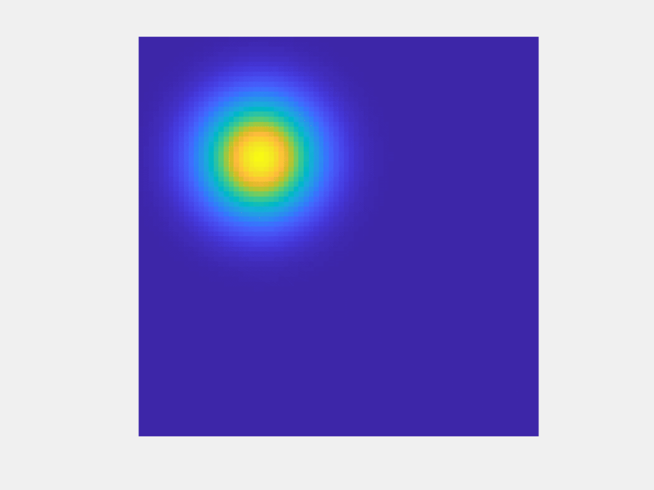
$$\text{where } J_\beta(m,f) = \begin{cases} \frac{\|m\|^2}{2f^\beta} & \text{if } f > 0 \\ 0 & \text{if } (m,f) = (0,0) \\ \infty & \text{otherwise} \end{cases}$$

- $w = (w_k)_{k \in \mathcal{G}_c}$ weights in $\mathbb{R}_+ \cup \{\infty\}$, which are both time and space dependent
- Objective remains convex for $\beta \in [0,1]$
  - For $\beta = 1$, corresponds to L^2-Wasserstein Metric
  - For $\beta = 0$, corresponds to H^{-1} Sobolev Metric
  - Linear Interpolation of metrics for $\beta \in (0,1)$

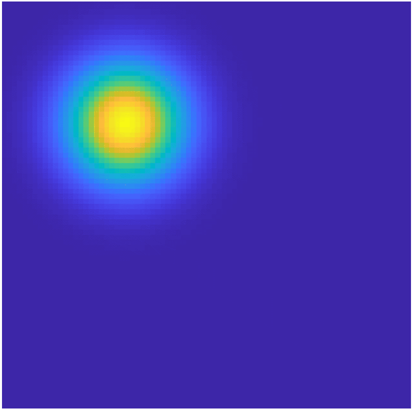
Gaussian Mixture Model, $\beta = 1$



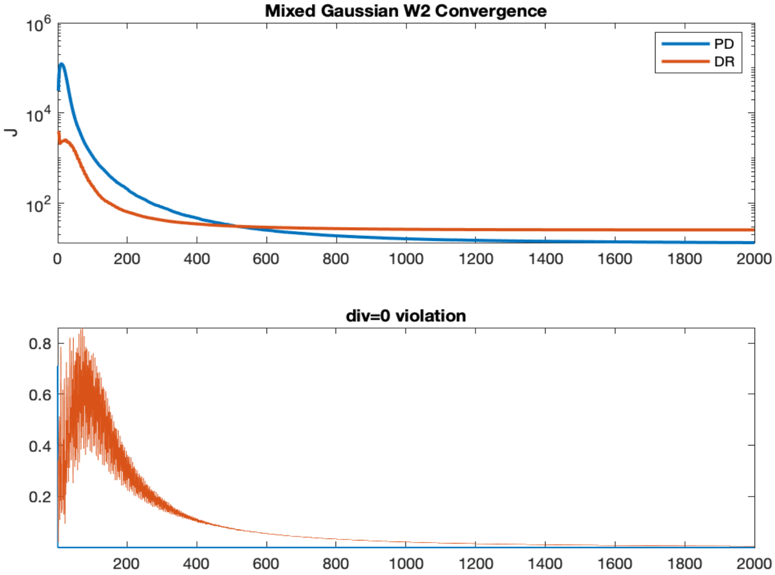
Gaussian Mixture Model, $\beta = 0$



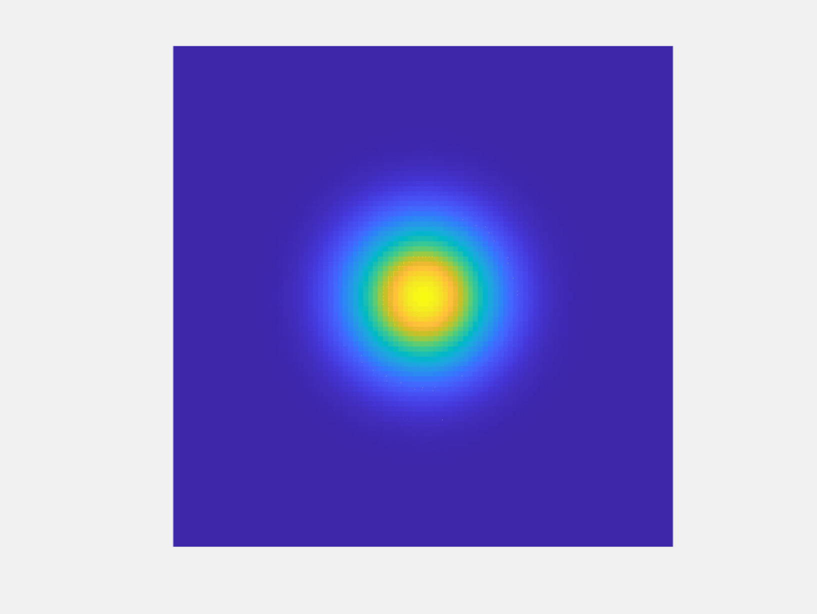
Gaussian Mixture Model, $\beta = 1/2$



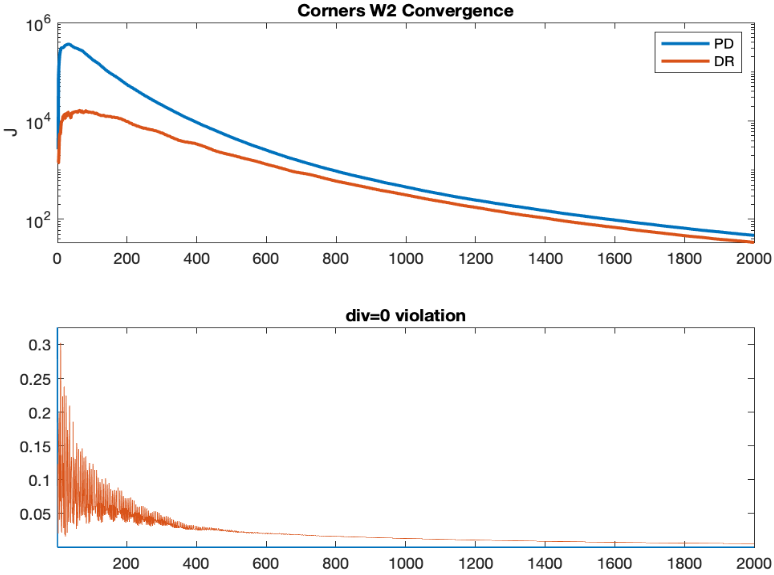
Gaussian Mixture Model



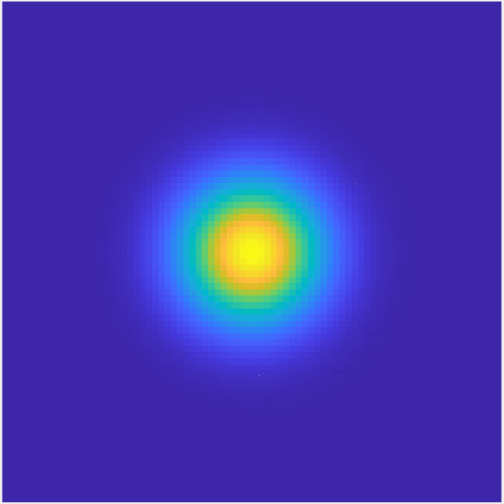
Corners Dataset



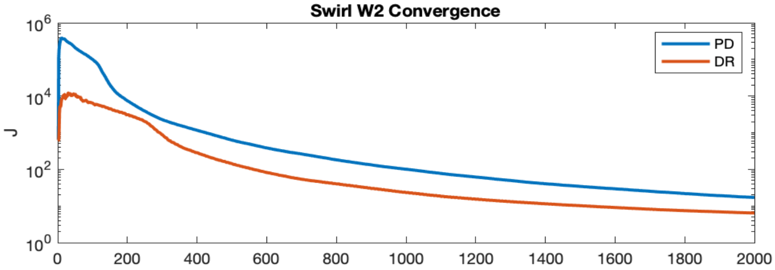
Corners Dataset

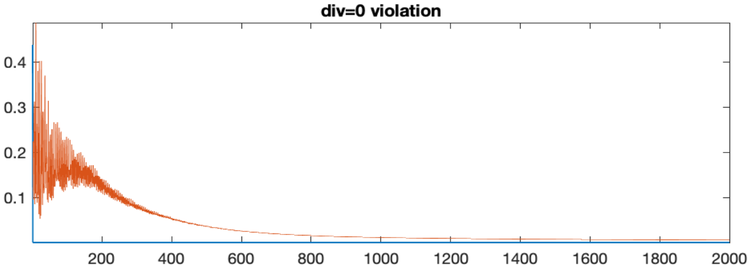


Swirl Dataset

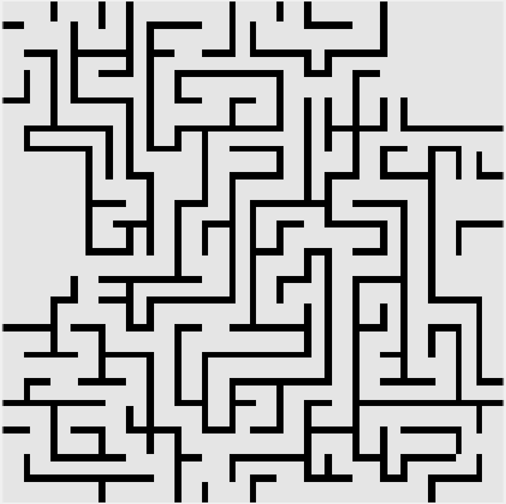


Swirl Dataset

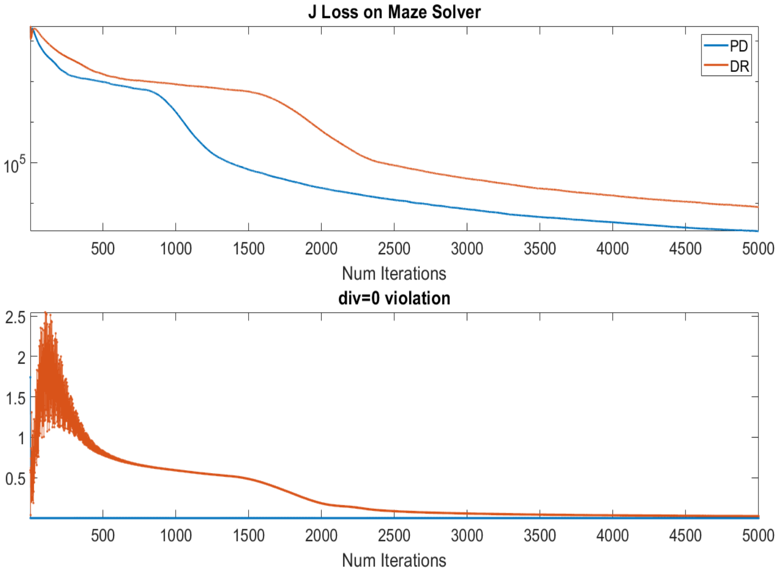




Maze



Maze


J Loss on Maze Solver
PD
DR
10^5
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
Num Iterations
div=0 violation
2.5
2
1.5
1
0.5
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
Num Iterations

References

[1] Jean-David Benamou and Yann Brenier. “A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem”. In: *Numerische Mathematik* 84 (Jan. 2000), pp. 375–393.

[2] Antonin Chambolle and Thomas Pock. “A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging”. In: *Journal of Mathematical Imaging and Vision* 40.1 (2011), pp. 120–145.

[3] Jonathan Eckstein. *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*. PhD Thesis, Massachusetts Institute of Technology. Cambridge, MA: MIT Press, 1990.

[4] Neal Parikh and Stephen Boyd. *Proximal Algorithms*. Foundations and Trends in Optimization, Vol. 1, No. 3, pp. 127-239. 2013.