The Monge Gap: A Regularizer to Learn All Transport Maps

https://github.com/Sisha3342/MongeGap

Team: A. Palanevich
Project suggested by: M. Gazdieva

Yandex School of Data Analysis, December 2023

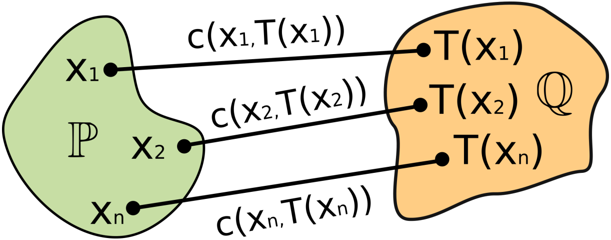
Table of Contents

Table of Contents

Optimal transport

The Monge problem

$$W_c(\mathbb{P}, \mathbb{Q}) = \inf_{T\sharp\mathbb{P}=\mathbb{Q}} \int_\Omega c(\mathbf{x}, T(\mathbf{x}))d\mathbb{P}(\mathbf{x}).$$



Existing approaches

Recent papers consider maps $T = \nabla f_\theta$, where f_θ is an input convex neural network (ICNN), and fit θ with SGD using samples. Despite their mathematical elegance, fitting OT maps with ICNNs raises many challenges[1]:

[1] Alexander Korotin et al. "Do neural optimal transport solvers work? a continuous wasserstein-2 benchmark". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 14593–14605.

Existing approaches

Recent papers consider maps $T = \nabla f_\theta$, where f_θ is an input convex neural network (ICNN), and fit θ with SGD using samples. Despite their mathematical elegance, fitting OT maps with ICNNs raises many challenges[1]:

- Some of their parameters must be non-negative.

[1] Korotin et al., "Do neural optimal transport solvers work? a continuous wasserstein-2 benchmark".

Existing approaches

Recent papers consider maps $T = \nabla f_\theta$, where f_θ is an input convex neural network (ICNN), and fit θ with SGD using samples. Despite their mathematical elegance, fitting OT maps with ICNNs raises many challenges[1]:

- Some of their parameters must be non-negative.
- Parameters initialization.

[1] Korotin et al., "Do neural optimal transport solvers work? a continuous wasserstein-2 benchmark".

Existing approaches

Recent papers consider maps $T = \nabla f_\theta$, where f_θ is an input convex neural network (ICNN), and fit θ with SGD using samples. Despite their mathematical elegance, fitting OT maps with ICNNs raises many challenges[1]:

- Some of their parameters must be non-negative.
- Parameters initialization.
- Min-max minimization problem.

[1] Korotin et al., "Do neural optimal transport solvers work? a continuous wasserstein-2 benchmark".

Existing approaches

Recent papers consider maps $T = \nabla f_\theta$, where f_θ is an input convex neural network (ICNN), and fit θ with SGD using samples. Despite their mathematical elegance, fitting OT maps with ICNNs raises many challenges[1]:

- Some of their parameters must be non-negative.
- Parameters initialization.
- Min-max minimization problem.
- Relevance of the imposed constraints.

[1] Korotin et al., "Do neural optimal transport solvers work? a continuous wasserstein-2 benchmark".

Table of Contents

Key contributions

A new paper proposes a radically different approach: the Monge gap[2], which introduces a regularizer to solve the OT problem. It has the following advantages:

[2] Théo Uscidda and Marco Cuturi. "The Monge Gap: A Regularizer to Learn All Transport Maps". In: *arXiv preprint arXiv:2302.04953* (2023).

Key contributions

A new paper proposes a radically different approach: the Monge gap[2], which introduces a regularizer to solve the OT problem. It has the following advantages:

- Generalizes to arbitrary costs.

[2] Uscidda and Cuturi, "The Monge Gap: A Regularizer to Learn All Transport Maps".

Key contributions

A new paper proposes a radically different approach: the Monge gap[2], which introduces a regularizer to solve the OT problem. It has the following advantages:

- Generalizes to arbitrary costs.
- No architecture constraints.

[2] Uscidda and Cuturi, "The Monge Gap: A Regularizer to Learn All Transport Maps".

Key contributions

A new paper proposes a radically different approach: the Monge gap[2], which introduces a regularizer to solve the OT problem. It has the following advantages:

- Generalizes to arbitrary costs.
- No architecture constraints.
- Simple minimization algorithm.

[2] Uscidda and Cuturi, "The Monge Gap: A Regularizer to Learn All Transport Maps".

Definition and properties

The Monge Gap

$$\mathcal{M}_{\rho}^{c}(T) = \int_{\Omega} c(\mathbf{x}, T(\mathbf{x}))\mathrm{d}\rho(\mathbf{x}) - W_c(\rho, T\sharp\rho).$$

Definition and properties

The Monge Gap

$$\mathcal{M}_\rho^c(T) = \int_\Omega c(\mathbf{x}, T(\mathbf{x}))\mathrm{d}\rho(\mathbf{x}) - W_c(\rho, T\sharp\rho).$$

Properties:

- For any vector field $T, \mathcal{M}_\rho^c(T) \geq 0$.

Definition and properties

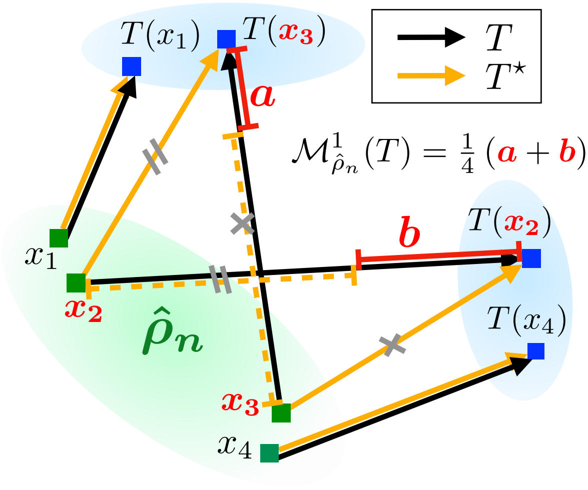
The Monge Gap

$$\mathcal{M}_{\rho}^{c}(T) = \int_{\Omega} c(\mathbf{x}, T(\mathbf{x}))\mathrm{d}\rho(\mathbf{x}) - W_c(\rho, T\sharp\rho).$$

Properties:

- For any vector field $T, \mathcal{M}_{\rho}^{c}(T) \geq 0$.
- T is a c-OT map between ρ and $T\sharp\rho \leftrightarrow \mathcal{M}_{\rho}^{c}(T) = 0$.

Geometrical intuition



Minimization problem: generic costs

We directly parametrize T as a neural network T_θ and solve the optimization problem:

Loss for generic costs

$$\min_{\theta} \mathcal{L}(\theta) = \Delta(T_\theta \sharp \mu, \nu) + \lambda_{\mathsf{MG}} \mathcal{M}_{\mu}^{c}(T).$$

Minimization problem: structured costs

For structured costs, i.e., $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$ with h beeing strictly convex, the proposed method can be modified.

OT map for structured costs

$$T_\theta : \mathbf{x} \mapsto \mathbf{x} - \nabla h^* \circ F_\theta(\mathbf{x}).$$

[3] Chen-Hao Chao et al. "Quasi-Conservative Score-based Generative Models". In: *arXiv preprint arXiv:2209.12753* (2022).

Minimization problem: structured costs

For structured costs, i.e., $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$ with h beeing strictly convex, the proposed method can be modified.

OT map for structured costs

$$T_\theta : \mathbf{x} \mapsto \mathbf{x} - \nabla h^* \circ F_\theta(\mathbf{x}).$$

Because F_θ is a conservative vector field, we use an additional regularizer[3] to point it and optimize the following loss:

Loss for structured costs

$$\min_\theta \mathcal{L}(\theta) = \Delta(T_\theta \sharp \mu, \nu) + \lambda_{\mathrm{MG}} \mathcal{M}_\rho^c(T_\theta) + \lambda_{\mathrm{cons}} \mathcal{C}_\rho(F_\theta).$$

[3] Chao et al., "Quasi-Conservative Score-based Generative Models".

Table of Contents

Wasserstein-2 benchmark

We use the Korotin eq al. 2021 benchmark for pairs of gaussians mixtures μ, ν in dimension $d \in \{2, 4, \ldots, 26\}$ for which the optimal map for the squared Euclidean cost is known. We report the next metrics:

Wasserstein-2 benchmark

We use the Korotin eq al. 2021 benchmark for pairs of gaussians mixtures μ, ν in dimension $d \in \{2, 4, \ldots, 26\}$ for which the optimal map for the squared Euclidean cost is known. We report the next metrics:

Unexplained variance percentage

$$\mathcal{L}_2^{\mathrm{UV}}(\hat{T}) = 100 \cdot \frac{\mathbb{E}_\mu \left\| \hat{T}(X) - T^*(X) \right\|_2^2}{\mathrm{Var}_\nu(X)}.$$

Wasserstein-2 benchmark

We use the Korotin eq al. 2021 benchmark for pairs of gaussians mixtures μ, ν in dimension $d \in \{2, 4, \ldots, 26\}$ for which the optimal map for the squared Euclidean cost is known. We report the next metrics:

Unexplained variance percentage

$$\mathcal{L}_2^{\mathrm{UV}}(\hat{T}) = 100 \cdot \frac{\mathbb{E}_\mu \left\| \hat{T}(X) - T^*(X) \right\|_2^2}{\mathrm{Var}_\nu(X)}.$$

The Sinkhorn divergence

$$S_{\ell_2^2,\epsilon}(\nu, \hat{T}\sharp\mu) = W_{\ell_2^2,\epsilon}(\mu,\nu) - \frac{1}{2}(W_{\ell_2^2,\epsilon}(\mu,\mu) + W_{\ell_2^2,\epsilon}(\nu,\nu)).$$

Results

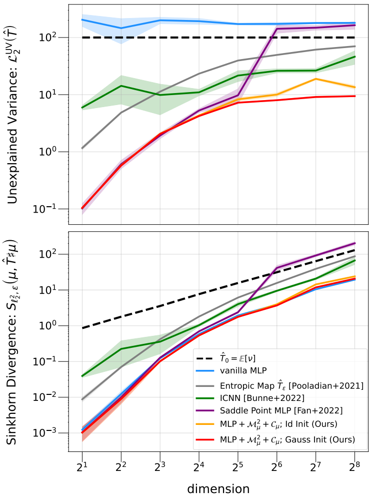


Figure: Reference

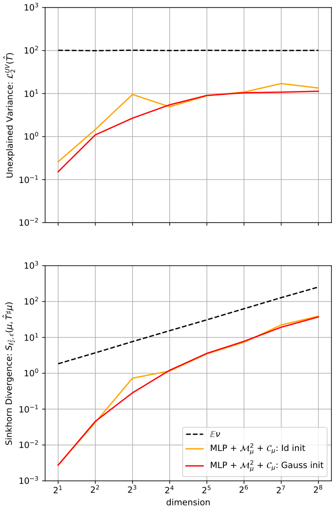
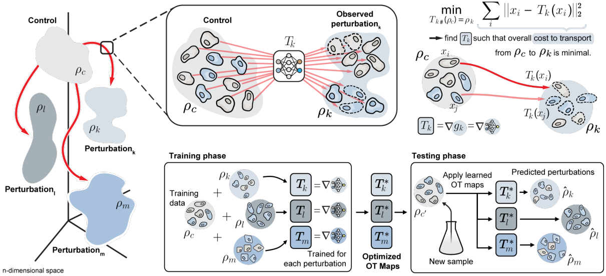


Figure: Replication

Single-cell genomics

Predicting the response of cells to a perturbation is a central question in biology. We predict responses of cells populations to cancer treatments using the proteomic dataset[4]. The dataset used can be accessed here.



[4] Charlotte Bunne et al. “Learning Single-Cell Perturbation Responses using Neural Optimal Transport”. In: *bioRxiv* (2021). DOI: 10.1101/2021.12.15.472775. eprint: https://www.biorxiv.org/content/early/2021/12/15/2021.12.15.472775.full.pdf. URL: https://www.biorxiv.org/content/early/2021/12/15/2021.12.15.472775.

Results

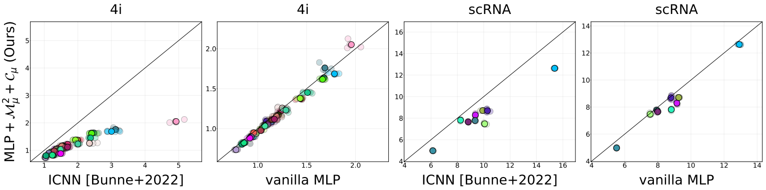


Figure: Reference

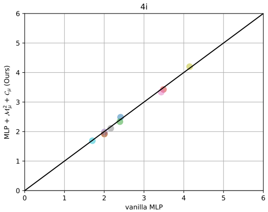


Figure: Replication

Table of Contents

Conclusion

- Studied the proposed regularization approach.
- Implemented regularizers/models and assessed their performance on selected tasks.
- Got quite close results for the $\mathcal{L}_2^{\mathrm{UV}}$, while the replication's $S_{\ell_2^2,\epsilon}$ is worse than the original one.

References

- Bunne, Charlotte et al. "Learning Single-Cell Perturbation Responses using Neural Optimal Transport". In: *bioRxiv* (2021). DOI: 10.1101/2021.12.15.472775. eprint: https://www.biorxiv.org/content/early/2021/12/15/2021.12.15.472775.full.pdf. URL: https://www.biorxiv.org/content/early/2021/12/15/2021.12.15.472775.
- Chao, Chen-Hao et al. "Quasi-Conservative Score-based Generative Models". In: *arXiv preprint arXiv:2209.12753* (2022).
- Korotin, Alexander et al. "Do neural optimal transport solvers work? a continuous wasserstein-2 benchmark". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 14593–14605.
- Uscidda, Théo and Marco Cuturi. "The Monge Gap: A Regularizer to Learn All Transport Maps". In: *arXiv preprint arXiv:2302.04953* (2023).