Learning Costs for Structured Monge Displacements

and related numerical experiments

Oussama Zekri[1]
Project for the course *Computational Optimal Transport*, Master MVA

[1] Research student, oussama.zekri@ens-paris-saclay.fr

17 January 2024

école normale supérieure paris—saclay

Presentation of the problem

OT

In OT, we seek among all mappings T that can transform one probability measure into another, the one who minimise a cost $c(\mathbf{x}, T(\mathbf{x}))$.
[Monge, 1781] formulation :

$$T^\star := \arg\inf_{T_\sharp \mu = \nu} \int_{\mathbb{R}^d} c(\mathbf{x}, T(\mathbf{x}))\,\mathrm{d}\mu \tag{1}$$

In general : ℓ_2^2 distance is used as the cost function.

- [Brenier, 1991]'s theorem, fairly well documented...
- Is this really the best way to make distribution alignments that make sense?

We consider in this work, costs of the form $c(\mathbf{x}, T(\mathbf{x})) = h(\mathbf{x} - T(\mathbf{x}))$, where $h(\cdot) = \frac{1}{2}\|\cdot\|_2^2 + \tau(\cdot)$, whith τ, a *regularizer* term.

- [Brenier, 1991]'s theorem not available. Monges maps: do they really exist?
- How does this term of regularization allow for meaningful alignments?

Proposed methods

The Monge problem

Solving (1) directly is not so easy... Idea : Kantorovich Relaxation [Santambrogio, 2015]

$$
\begin{aligned}
f^\star, g^\star &:= \operatorname{arg\,sup}_{\substack{f,g:\mathbb{R}^d\to\mathbb{R}\\ f\oplus g\leq h}} \int_{\mathbb{R}^d} f \,\mathrm{d}\mu + \int_{\mathbb{R}^d} g\,\mathrm{d}\nu \\
&= \operatorname{arg\,sup}_{f:\mathbb{R}^d\to\mathbb{R},\, h\text{-concave}} \int_{\mathbb{R}^d} f\,\mathrm{d}\mu + \int_{\mathbb{R}^d} \bar{f}^h \mathrm{d}\nu
\end{aligned} \tag{2}
$$

Theorem (Generalization of Brenier's theorem)

If the solution $f^\star$ of (2) is h-concave and differentiable, and if h is stricly convex, we have that

$$T^\star(\mathbf{x}) = \mathbf{x} - (\nabla h)^{-1}(\nabla f^\star(\mathbf{x})) = \mathbf{x} - \nabla h^* \circ \nabla f^\star(\mathbf{x})\,, \tag{3}$$

where the convex conjugate of h reads: $h^*(\mathbf{x}) := \max_{\mathbf{y}} \langle \mathbf{y}, \mathbf{x}\rangle - h(\mathbf{y})\,.$

Proposed methods

Existence of Monge maps

Proposition, Existence of Monge map under convex regularizer

Suppose that the regularizer term τ is convex. There exists an optimal Monge map $T^\star : \mathbb{R}^d \to \mathbb{R}^d$ that solves (1). Furthermore, $T^\star$ can be written as $T^\star(\mathbf{x}) = \mathbf{x} - prox_\tau(\mathbf{x}) \circ \nabla f^\star(\mathbf{x})$, where $f^\star$ is defined as in (2).

Remark, Sufficient condition

Note that the convex condition on τ **is only a sufficient condition** to get the existence of a Monge map. Consider, for example, the regularizer $\tau_{STVS}(\mathbf{x}) := \gamma^2 \mathbf{1}_d^T \left(asinh(\frac{\mathbf{x}}{2\gamma}) + \frac{1}{2} - \frac{1}{2} e^{-2\sigma(\mathbf{x})} \right)$ introduced in [Schreck et al., 2015]. It is highlighted in [Cuturi et al., 2023] that τ_{STVS} is non-convex but $h_{STVS} = \frac{1}{2}\| \ \|_2^2 + \tau_{STVS}$ is $\frac{1}{2}$-strongly convex, so stricly convex : the Theorem 4 applies here.

Proposed methods

Regularizers

- $\tau_{\ell_1}(\mathbf{x}) = \gamma\|\mathbf{x}\|_1$ with $\mathrm{prox}_{\tau_{\ell_1}}(\mathbf{x}) = \mathrm{ST}_\gamma(\mathbf{x}) = \left(1 - \frac{\gamma}{|\mathbf{x}|}\right)_+ \odot \mathbf{x}$

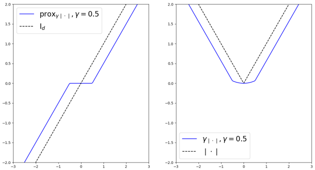

Figure: Soft-tresholding operator in dimension 1.

- $\tau_{\mathbf{b}}(\mathbf{x}) = \frac{\gamma}{2}(\mathbf{b}^\top\mathbf{x})^2$ with $\mathrm{prox}_{\tau_{\mathbf{b}}}(\mathbf{x}) = \gamma(\mathrm{Id} + \mathbf{b}\mathbf{b}^\top)^{-1}\mathbf{x}$

Let $\mathbf{M} = \mathbf{U}^\top\mathbf{\Sigma}\mathbf{V}$ be an SVD of $\mathbf{M}$, where $\mathbf{\Sigma}$ contains the vector of singular vectors in decreasing order $\sigma = (\sigma_i)_{i\in\{1,\ldots,d\}}$ of $\mathbf{M}$.

- $\tau_*(M) := \gamma\|\mathbf{M}\|_* = \gamma\|\sigma\|_1$ with $\mathrm{prox}_{\tau_*}(\mathbf{M}) = \mathbf{U}^\top\mathrm{prox}_{\gamma\|\cdot\|_1}(\sigma)\mathbf{V}$
- $\tau_{\mathrm{rk}}(M) := \gamma\|\mathbf{M}\|_{\mathrm{rk}} = \gamma\|\sigma\|_0$ with $\mathrm{prox}_{\tau_{\mathrm{rk}}}(\mathbf{M}) = \mathbf{U}^\top\mathrm{prox}_{\gamma\|\cdot\|_0}(\sigma)\mathbf{V}$

Numerical findings

Setup and parameters of the experiments

We use the Optimal Transport Tools library, OTT-Jax [Cuturi et al., 2022], in order to run our experiments.

We generate a source point cloud $\mathbf{x}$ made up of 30 points, with mean $[0, 0]$, and a target point clouds $\mathbf{y}$, divided in two target point clouds, made up of 25 points each, with mean $[5, 0]$ and $[0, 8]$ respectively. See Figure 2.

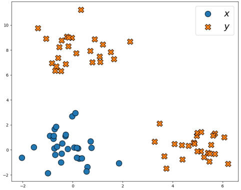

Figure: In blue, the source point cloud made up with 30 points, with mean $[0, 0]$. In orange, the target point cloud made up with 25 points each, with mean $[5, 0]$ and $[0, 8]$ respectively.

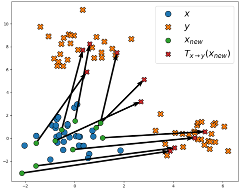

Figure: We consider the classical cost $h = \ell_2^2$, i.e. without regularizer ($\tau = 0$). 10 new points displayed in the source distribution in green. Entropics maps displayed as arrows.

Numerical findings

The ℓ_1 regularizer

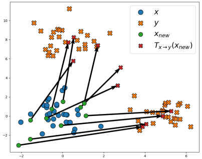

Figure: Entropic maps ℓ_1-norm, $\gamma = 0.1$

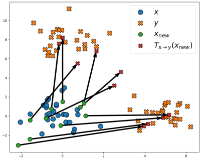

Figure: Entropic maps ℓ_1-norm, $\gamma = 1$

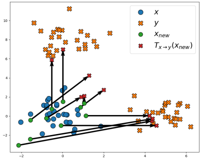

Figure: Entropic maps ℓ_1-norm, $\gamma = 10$

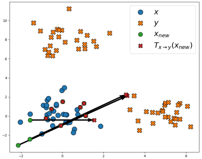

Figure: Entropic maps ℓ_1-norm, $\gamma = 100$

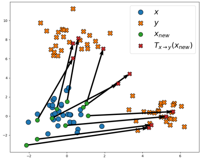

Figure: Entropic maps $\mathbf{b}$-directional, $\gamma = 0.1$

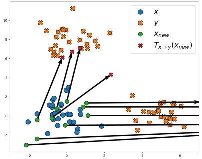

Figure: Entropic maps $\mathbf{b}$-directional, $\gamma = 1$

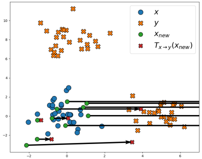

Figure: Entropic maps $\mathbf{b}$-directional, $\gamma = 10$

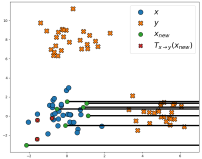

Figure: Entropic maps $\mathbf{b}$-directional, $\gamma = 100$

Numerical findings

Datasets and explanations for matrix OT

Dataset [Meaney, 2022], which is a collection of X-ray projection images of a pine cone imaged in a cone-beam computed tomography (CBCT) scanner, see Figure 12.

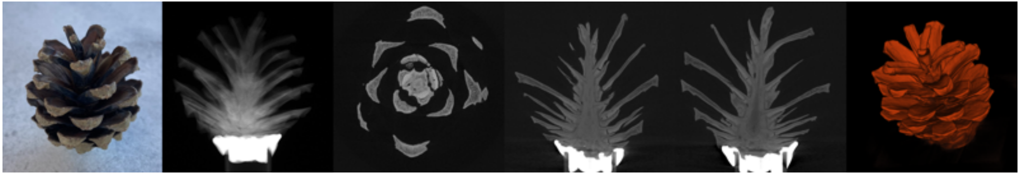

Figure: Overview of the dataset [Meaney, 2022]

We will consider the indices (i, j) of the matrix as a point cloud, and the value of the matrix in (i, j), noted $M_{i,j}$ as the mass.
This brings us back to a Discrete Optimal Transport problem. ($\mu = \sum_{i=1}^{n} p_i^s \delta_{x_i^s}$, the source distribution, and $\nu = \sum_{i=1}^{m} p_i^t \delta_{x_i^t}$, the target distribution.)

Numerical findings

Nuclear norm regularizer

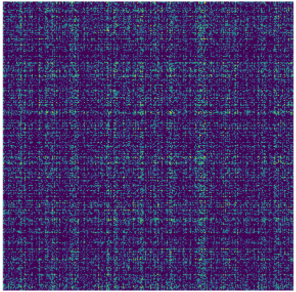

Figure: Approximated OT matrix between two distributions of the dataset, with $h = \ell_2^2$. Datas' reference used : 20201118.0701 and 20201118.0702.

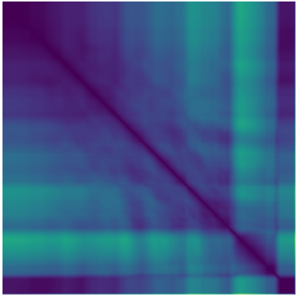

Figure: Approximated OT matrix between two distributions of the dataset, with $h = \ell_2^2 + \tau_*$. Datas' reference used : 20201118.0701 and 20201118.0702.

Experiments on these matrices are not really interpretable as predicted... It tells us nothing about the "low-rank" nature of the nuclear norm...

Numerical experiments

- Existence of established Monge maps in a rather satisfactory case.
- Some of the results confirm our theoretical guarantees: the ℓ_1 regularizer does induce sparsity, and the $\mathbf{b}$-directional regularizer induces penalization in the direction of $\mathbf{b}$.
- OT between matrices : we have not successfully showcased the intended properties.
- Need to rethink the experiments to show that low-rank matrices are favored by our two matrix regularizers.

End

Thank you for your attention

Bibliography

Brenier, Y. (1991).
Polar factorization and monotone rearrangement of vector-valued functions.
Communications on Pure and Applied Mathematics, 44(4).

Cuturi, M., Klein, M., and Ablin, P. (2023).
Monge, bregman and occam: Interpretable optimal transport in high-dimensions with feature-sparse maps.

Cuturi, M., Meng-Papaxanthos, L., Tian, Y., Bunne, C., Davis, G., and Teboul, O. (2022).
Optimal transport tools (ott): A jax toolbox for all things wasserstein.

Meaney, A. (2022).
Cone-Beam Computed Tomography Dataset of a Pine Cone.

Monge, G. (1781).
Mémoire sur la théorie des déblais et des remblais.
De l'Imprimerie Royale.

Definitions

For all $\mathbf{x}, \mathbf{y}$ we write $(f \oplus g)(\mathbf{x}, \mathbf{y}) := f(\mathbf{x}) + g(\mathbf{y})$ and for any function $f : \mathbb{R}^d \to \mathbb{R}$, we define its h-transform as

$$\bar{f}^h(\mathbf{y}) := \min_{\mathbf{x}} h(\mathbf{x} - \mathbf{y}) - f(\mathbf{x}). \tag{4}$$

Definition (h-concave functions)

A function f is said to be h-concave if there exists a function g such that it is itself the h-transform of g, i.e., $f = \bar{g}^h$.