Learning Costs for Structured Monge Displacements

Oussama Zekri*
ENS Paris-Saclay
oussama.zekri@ens-paris-saclay.fr

January 17, 2024

Abstract

Optimal transport theory has brought ML with valuable tools for inferring a push-forward map between probability densities. The implementation of OT remains challenging because of the ever-growing need for efficient, low-capacity methods. [Cuturi et al., 2023] noticed that many uses of optimal transport are limited to taking the ℓ_2^2 norm as the cost.

With the aim of continuing the paper [Klein et al., 2023], we focus on the use of a regularizing term τ added to this classic cost, in order to have a cost in the form $c(x, y) = h(x - y)$ where $h = \ell_2^2 + \tau$. The idea is to evaluate the impact of different regularizing terms, and to highlight the characteristics specific to each of them, which thus give a new utility to optimal transport.

We will also be interested in establishing the existence of Monge maps, as numerous works have done in the past, and extending our tool to costs between matrix spaces.

1 Introduction

About Optimal Transport and computation Tackling Optimal Transport (OT) problems at a large scale presents huge challenges. The primary obstacle, notably on the computational front, is evident: addressing the [Kantorovich, 2006] problem with discrete measures of size n requires solving a linear program (LP) with a complexity of $O(n^3 \log(n))$.

Such complexity is unthinkable for today's problems, and we need to be able to develop methods to overcome this computational difficulty.

Previous work We consider in this work cost functions c of the form $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$, where $h : \mathbb{R}^d \to \mathbb{R}$ ($d \geq 1$) has the particular form $h(\cdot) = \frac{1}{2}\|\cdot\|_2^2 + \tau(\cdot)$, where τ is a *regularizer* term. It was introduced in [Cuturi et al., 2023]. In considering this new type of cost, we must also consider previous work on

*Department of Teaching and Research of Mathematics, ENS Paris-Saclay, France.

the existence of Monge maps. [Ambrosio and Pratelli, 2003, Ambrosio et al., 2004, Evans and Gangbo, 1999, Carlier et al., 2010].

We will focus in particular on [Santambrogio, 2015], which provides valuable results on costs where h is strictly convex.

Despite a few attempts, very few papers are interested in extending optimal transport to more exotic spaces such as matrix spaces. We will nevertheless quote this thesis [Vayer, 2020], which is very useful to read.

Software tools The Python softwares [Flamary et al., 2021, POT] and [Cuturi et al., 2022, OTT-Jax] have recently arrived, making it possible to apply optimal transport and its metrics.

While the first one is based on classical libraries such as [Harris et al., 2020, Numpy], the latter is based on [Bradbury et al., 2018, JAX], a powerful tool developed by Google. However, both libraries are sorely lacking in metrics. It may therefore be a good idea to add new metrics for optimal transport, when they seem interesting.

Contributions Our contributions are threefold.

- We introduced new regularizers, and therefore new cost functions, for which we show very precisely the existence of Monge maps.
- We add to OTT-Jax three new costs metrics : one for a $\mathbf{b}$-directional regularizer, and two for nuclear norm and rank regularizer, with their proximal operators. Note that our codes are available on our Github Repository for this project.
- Experiments have been introduced to highlight the specific characteristics of each of the regularizers studied.

2 Background

Notations We denote $\mathcal{M}(\mathcal{X})$ the set of Radon measures on the space $\mathcal{X}$.

Definition 2.1 (Push-forward [Peyré and Cuturi, 2020])**.** *Given mesurables spaces $\mathcal{X}$ and $\mathcal{Y}$, a mapping $T : \mathcal{X} \to \mathcal{Y}$ and a measure $\mu \in \mathcal{M}(\mathcal{X})$, the push-forward measure $\nu = T_\sharp \mu \in \mathcal{M}(\mathcal{Y})$ is defined by,*

$$\forall g \in \mathcal{C}(\mathcal{Y}), \quad \int_{\mathcal{Y}} g(y) \mathrm{d}\nu(y) = \int_{\mathcal{X}} g(T(x)) \mathrm{d}\mu(x)$$

Existence of Monge maps We consider in this work cost functions c of the form $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$, where $h : \mathbb{R}^d \to \mathbb{R}$ ($d \geq 1$) has the particular form $h(\cdot) = \frac{1}{2}\|\cdot\|_2^2 + \tau(\cdot)$, where τ is a *regularizer* term. The Monge problem [Monge, 1781], central in OT, seeks a map $T : \mathbb{R}^d \to \mathbb{R}^d$ that *push-forward* (Def. 2.1) a measure

$\mu \in \mathcal{P}(\mathbb{R}^d)$ onto $\nu \in \mathcal{P}(\mathbb{R}^d)$ which minimizing an average transport cost (as quantified by h in our case) of the form :

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

A natural first question to ask concerns the existence of such an optimal Monge map $T^\star$. In fact, many people have asked this question in different contexts. These include the following works [Ambrosio and Pratelli, 2003, Ambrosio et al., 2004, Evans and Gangbo, 1999, Carlier et al., 2010], which deals with the existence of the map $T^\star$.

To answer this question, we can first recall Brenier's theorem [Brenier, 1991], which establishes the existence (and uniqueness) of such a Monge map $T^\star$ when the regularization term τ is zero (so that $h(\cdot) = \frac{1}{2}\|\cdot\|_2^2$).

However, the whole point of our method lies in this regularization term τ. The existence (or not) of a Monge map, depending on the τ regularization term, will be discussed further in Section 3.

3 Method

3.1 The Monge problem

Solving the Monge problem (1) directly is not so easy. In fact, the constraints of this optimization problem are not even convex. The idea is to relax this problem, as in [Santambrogio, 2015, Proposition 1.11.], by introducing the semi-dual formulations :

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

where 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

$$\bar{f}^h(\mathbf{y}) := \min_{\mathbf{x}} h(\mathbf{x} - \mathbf{y}) - f(\mathbf{x}). \tag{3}$$

Definition 3.1 (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., $f = \bar{g}^h$.*

Now we have everything we need to recall [Santambrogio, 2015, Theorem 1.17.].

Theorem 3.2 (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*

$$T^\star(\mathbf{x}) = \mathbf{x} - (\nabla h)^{-1}(\nabla f^\star(\mathbf{x})) = \mathbf{x} - \nabla h^* \circ \nabla f^\star(\mathbf{x})\,, \tag{4}$$

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

Proof. The result follows from [Santambrogio, 2015, Theorem 1.17.]. □

As we mentioned in Section 2, the Brenier's theorem [Brenier, 1991] is a particular case of this : $h = \frac{1}{2}\| \cdot \|_2^2$ is indeed stricly convex, and $\nabla h = (\nabla h)^{-1} = \text{Id}$. It leads us to $T(\mathbf{x}) = \mathbf{x} - \nabla f^\star(\mathbf{x})$.

An important assumption of the Theorem 3.1 is the strict convexity of h. Let us introduce a useful sufficient condition on the regularizer term of h (i.e. τ) for this Theorem to be true.

Proposition 3.3 (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 $T^\star(\mathbf{x}) = \mathbf{x} - \text{prox}_\tau(\mathbf{x}) \circ \nabla f^\star(\mathbf{x})$, where $f^\star$ is defined as in (2).*

Proof. Let us recall that $\forall \mathbf{x} \in \mathbb{R}^d$, $h(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2 + \tau(\mathbf{x})$. If τ is convex, h is strictly convex, because it is the sum of a convex term and a strictly convex term. the Theorem 3.1 applies here. Futhermore, $\nabla h^*(\mathbf{x}) = -\text{prox}_\tau(\mathbf{x})$, so we have that $T^\star(\mathbf{x}) = \mathbf{x} - \text{prox}_\tau(\mathbf{x}) \circ \nabla f^\star(\mathbf{x})$. □

Remark 3.4 (Optimal Monge displacements)**.** *The previous formulation generates Optimal Monge Displacements [Klein et al., 2023], noted $\Delta\mathbf{x} = T^\star(\mathbf{x}) - \mathbf{x} = -\text{prox}_\tau(\mathbf{x}) \circ \nabla f^\star(\mathbf{x})$. They represent the displacement of a point x during transport.*

Remark 3.5 (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 $\tau_{\text{STVS}}(\mathbf{x}) := \gamma^2 \mathbf{1}_d^T \left(\text{asinh}(\frac{\mathbf{x}}{2\gamma}) + \frac{1}{2} - \frac{1}{2} e^{-2\sigma(\mathbf{x})} \right)$ introduced in [Schreck et al., 2015]. It is highlighted in [Cuturi et al., 2023] that τ_{STVS} is non-convex but $h_{\text{STVS}} = \frac{1}{2}\| \|_2^2 + \tau_{\text{STVS}}$ is $\frac{1}{2}$-strongly convex, so stricly convex : the Theorem 3.1 applies here.*

3.2 Regularizers considered and theoretical guarantees

Let us introduce the regularizers we will be considering in this work, developing the interest of each of them.

The first two are more classical regularizers, while the last two involve matrix spaces.

3.2.1 The ℓ_1 regularizer

We denote by τ_{ℓ_1}, the regularizer induced by ℓ_1 norm. It is defined by $\tau_{\ell_1}(\mathbf{x}) = \gamma\|\mathbf{x}\|_1$. The convexity of this regularizer implies the existence of Monge map thanks to the Proposition 3.3.

The main interest of this regularizer is that it induces sparsity, see [Bach, 2008].

Its proximal operator is the well-known tresholding operator, see Figure 1.

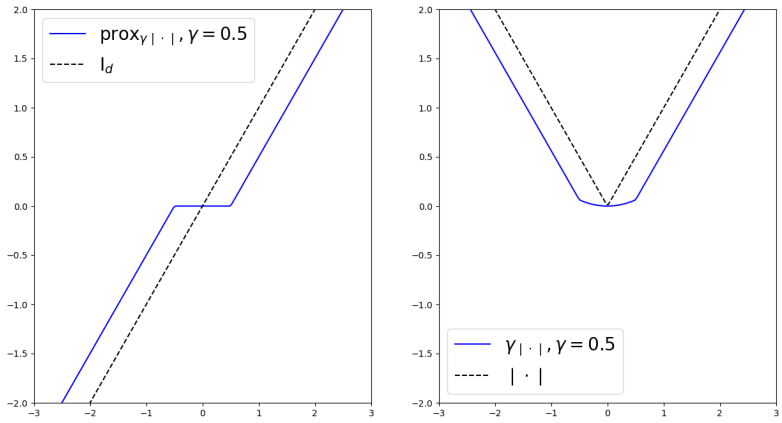

Figure 1: Soft-tresholding operator in dimension 1. On the left, the proximal operator of the absolute value in blue, and the identity operator in black. On the right, the Moreau enveloppe of the absolute value in blue, and the absolute value in black.

It is defined by

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

where $(a)_+ := \max\{0, a\}$ and $\odot$ denotes the element-wise product operator.

3.2.2 The $\mathbf{b}$-directional regularizer

We denote by $\tau_{\mathbf{b}}$, the $\mathbf{b}$-directional regularizer. It is defined by $\tau_{\mathbf{b}}(\mathbf{x}) = \frac{\gamma}{2}(\mathbf{b}^\top \mathbf{x})^2$. The convexity of this regularizer implies the existence of Monge map thanks to the Proposition 3.3.

The main interest of this regularizer is that it induces penalization in the direction of a vector $\mathbf{b}$, see [Klein et al., 2023].

Its proximal operator is defined by

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

In fact, if $h_{\mathbf{b}}(\mathbf{x}) := \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \frac{\gamma}{2}(\mathbf{b}^\top \mathbf{x})^2$, we can compute that $\nabla h_{\mathbf{b}}(\mathbf{x}) = \mathbf{b}\mathbf{b}^\top \mathbf{x} - \mathbf{y} + \mathbf{x}$

3.2.3 The nuclear norm regularizer

Inspired by the very recent paper [Sebbouh et al., 2023], we introduce τ_*, the nuclear norm regularizer.

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 $\sigma = (\sigma_i)_{i \in \{1,\dots,d\}}$ of $\mathbf{M}$. The nuclear norm of $\mathbf{M}$ is defined as $\|\mathbf{M}\|_* = \|\sigma\|_1$.

The nuclear norm regularizer is therefore defined by $\tau_*(M) := \gamma \|\mathbf{M}\|_* = \gamma \|\sigma\|_1$.

Its proximal operator is defined by

$$\text{prox}_{\tau_*}(\mathbf{M}) = \mathbf{U}^\top \text{prox}_{\gamma\|\cdot\|_1}(\sigma)\mathbf{V}$$

3.2.4 The rank regularizer

Still inspired by the very recent paper [Sebbouh et al., 2023], we also introduce τ_{rk}.

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 $\sigma = (\sigma_i)_{i \in \{1,\dots,d\}}$ of $\mathbf{M}$. The rank norm of $\mathbf{M}$ is defined as $\|\mathbf{M}\|_{\text{rk}} = \|\sigma\|_0$.

The rank norm regularizer is therefore defined by $\tau_{\text{rk}}(M) := \gamma \|\mathbf{M}\|_{\text{rk}} = \gamma \|\sigma\|_0$.

Its proximal operator is defined by

$$\text{prox}_{\tau_{\text{rk}}}(\mathbf{M}) = \mathbf{U}^\top \text{prox}_{\gamma\|\cdot\|_0}(\sigma)\mathbf{V}$$

These two matrix regularizers force us to re-specify the context of the optimal transport problem. We need to show why Monge maps also exist in this case.

3.2.5 On the existence of Monge maps for matrix regularizers

If $d_x > d_y$, we are no longer in the context of Theorem 3.1. That is why we will assume in our experiments that $d = d_x = d_y$. In fact, we can introduce an other matrix that first reduces the dimension d_x to equalize it with d_y, as is done in [Sebbouh et al., 2023, Figure 1]. But we'll assume this step has already been taken.

Then, we consider functions h of the form :

- $h_*(\sigma) = \frac{1}{2}\|\sigma\|_2^2 + \gamma \|\sigma\|_1$ for the nuclear norm.
- $h_{\text{rk}}(\sigma) = \frac{1}{2}\|\sigma\|_2^2 + \gamma \|\sigma\|_0$ for the rank regularizer.

Note that we considered σ, the vector of the singular values of $\mathbf{M}$, as a variable for our h function, and not $\mathbf{M}$.

In both cases, Proposition 3.3 applies, because τ_* and τ_{rk} are convex as functions of the variable σ. That is why, in this way, we recover Theorem 3.1.

3.3 On the computation of h-transforms

In order to carry out our experiments, we need to determine a practical way of computing h-transforms [Klein et al., 2023].

If f is concave and smooth, we can use proximal gradient descent with stricly positive step-size $\lambda > 0$, in order to optimize $\bar{f}^h$. When we have acces to $\mathrm{prox}_{\lambda h}$ this algorithm has the following iterations :

$$\mathbf{x} \leftarrow \mathbf{y} + \mathrm{prox}_{\lambda h}(\mathbf{x} - \mathbf{y} + \lambda \nabla f(\mathbf{x})). \tag{5}$$

Thanks to this procedure, we obtain the h-transform of f as well as its gradient.

Futhermore, because h is the sum of a regularizer τ and a quadratic term, thanks to [Parikh et al., 2014, Section 2.1.1], the proximal operator of λh is given by

$$\mathrm{prox}_{\lambda h}(\mathbf{z}) = \mathrm{prox}_{\frac{\lambda\gamma}{\lambda+1}\tau}\left(\frac{\mathbf{z}}{1+\lambda}\right).$$

The following proposition exploits this :

Proposition 3.6. *Assume f is L-smooth and concave and that $\lambda < 2/L$. Then, iterations (5) converge to a point $\mathbf{x}^\star(\mathbf{y}) = \arg\min_{\mathbf{x}} h(\mathbf{x} - \mathbf{y}) - f(\mathbf{x})$. Furthermore, we have*

$$\bar{f}^h(\mathbf{y}) = h(\mathbf{x}^\star(\mathbf{y}) - \mathbf{y}) - f(\mathbf{x}^\star(\mathbf{y})), \text{ and } \nabla \bar{f}^h(\mathbf{y}) = -\nabla h(\mathbf{x}^\star(\mathbf{y}) - \mathbf{y}). \tag{6}$$

Proof. The convergence of iterates (5) follows from [Beck and Teboulle, 2009, Theorem 1] or [Rockafellar, 1976, Theorem 1]. [Bauschke and Combettes, 2011, Proposition 18.7] then give the last identities. □

Equipped with h-concave potentials, we can now use them to produce OT maps for h.

Theorem 3.7. *Let μ be a measure and push it forward using $T_f^h := \mathrm{Id} - \mathrm{prox}_{\gamma\tau} \circ \nabla \bar{f}^h$ then T_f^h is optimal for cost h between μ and $(T_f^h)_\sharp \mu$.*

Proof. The result follows from [Santambrogio, 2015, Theorem 1.17]. □

With all this in mind, we are now ready to move on to the experiments.

4 Experiments

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

With Jax, the process of finding the solution of the dual problem (2) is already implemented. We compute entropic maps instead of the theoritical maps given in Section 3, which will make the computations easier.

Note that our codes are available on our Github Repository for this project.

This section is divided into two subsections. The first subsection, Section 4.1, deals with experiments on vector regularizers, with experiments on 2D Gaussian point clouds.

The second subsection, Section 4.2, deals with experiments on matrix regularizers.

4.1 Vector regularizers

4.1.1 Setup and parameters of the experiments

We therefore begin by considering 2D Gaussian point clouds.

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.

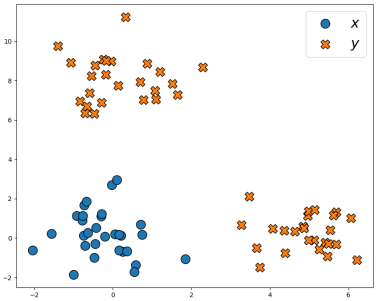

Figure 2: In blue, the source point cloud made up with 30 points, with mean $[0, 0]$. In orange, the target point cloud made up with 25 points each, with mean $[5, 0]$ and $[0, 8]$ respectively.

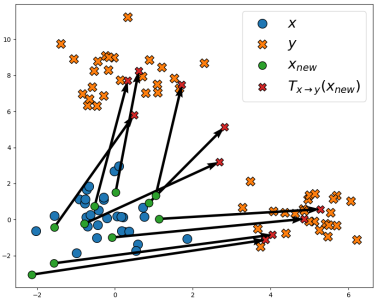

Figure 3: We consider the classical cost $h = \ell_2^2$, i.e. without regularizer ($\tau = 0$). 10 new points displayed in the source distribution in green. Entropics maps displayed as arrows.

We generate 10 new test points in the source point cloud, and display the entropic maps to the target distributions of these points, with the classical ℓ_2^2 costs as function h, i.e. with $\tau = 0$.

Figure 3 shows a classic optimal transport plan, which will serve as a reference for future experiments with regularizer terms.

4.1.2 The ℓ_1 regularizer

We consider the same point clouds and test points, this time with an ℓ_1-norm regularizer term in the cost function.

Entropy maps are displayed for different gamma values : $\gamma \in \{0.1, 1, 10, 100\}$

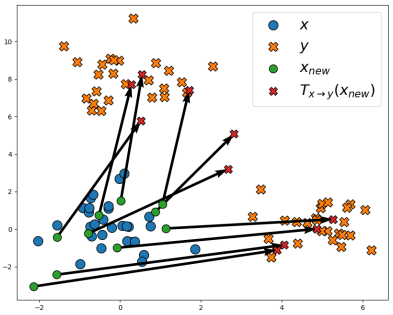

Figure 4: Entropic maps for ℓ_1-norm regularizer and $\gamma = 0.1$

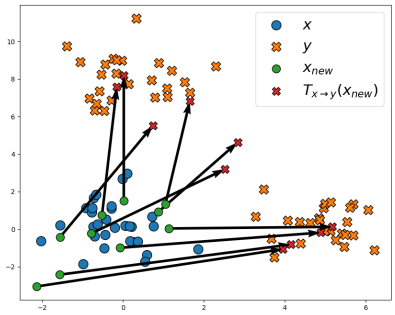

Figure 5: Entropic maps for ℓ_1-norm regularizer and $\gamma = 1$

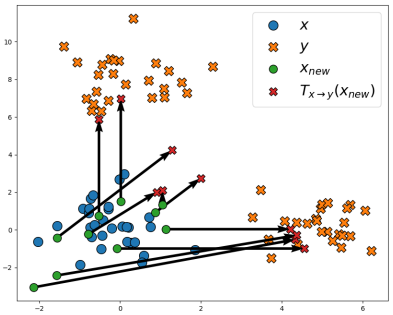

Figure 6: Entropic maps for ℓ_1-norm regularizer and $\gamma = 10$

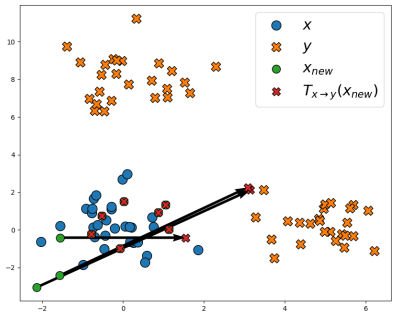

Figure 7: Entropic maps for ℓ_1-norm regularizer and $\gamma = 100$

As expected, we see that the ℓ_1 norm induces sparsity. That is also why we talk about *sparsity-inducing norm*, see [Bach, 2008]. A lot of entropic maps have a sparse displacement behavior : the higher the γ, the more the entropic maps move on a single coordinate, see Figure 6 for instance.

Futhermore, if γ is too high, some maps seem to clip due to the excessive penalty.

4.1.3 The $\mathbf{b}$-directional regularizer

We consider the same point clouds and test points, this time with the $\mathbf{b}$-directional regularizer term in the cost function. We choose to take $\mathbf{b} = [1, 0]$.

Entropy maps are displayed for different gamma values : $\gamma \in \{0.1, 1, 10, 100\}$

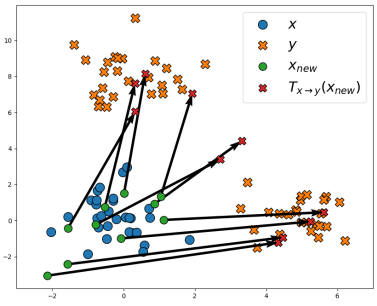

Figure 8: Entropic maps for $\mathbf{b}$-directional regularizer with $\mathbf{b} = [1, 0]$ and $\gamma = 0.1$

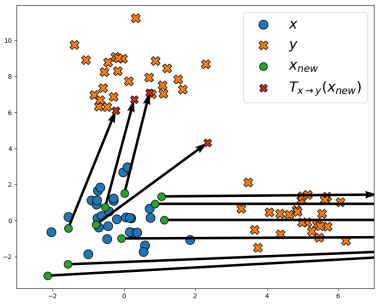

Figure 9: Entropic maps for $\mathbf{b}$-directional regularizer with $\mathbf{b} = [1, 0]$ and $\gamma = 1$

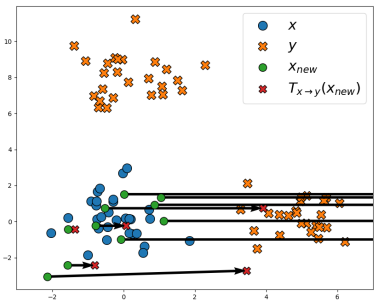

Figure 10: Entropic maps for $\mathbf{b}$-directional regularizer with $\mathbf{b} = [1, 0]$ and $\gamma = 10$

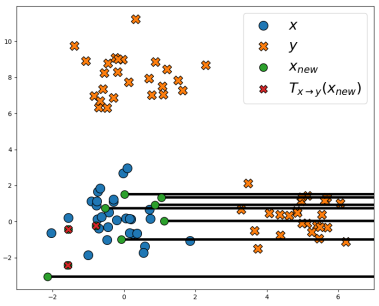

Figure 11: Entropic maps for $\mathbf{b}$-directional regularizer with $\mathbf{b} = [1, 0]$ and $\gamma = 100$

As expected, we see that the $\mathbf{b}$-directional regularizer induces penalisation in the direction of the vector $\mathbf{b}$. A lot of entropic maps seem to follow the direction imposed by $\mathbf{b}$: the higher the γ, the more the entropic maps are oriented like $\mathbf{b}$, see Figure 10 for instance.

Futhermore, as above, if γ is too high, some maps seem to clip due to the excessive penalty.

4.2 Matrix regularizers

4.2.1 Setup and parameters of the experiments

For these experiments, we will take inspiration from the practical session number 3, from another MVA course: Foundations of Distributed and Large Scale Computing Optimization.

This practical session deals with image reconstruction in X-ray tomography. It involves a *tomography matrix* $H \in \mathbb{R}^{M \times N}$, which is sparse and encodes the geometry of the measurements.

We will use the 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.

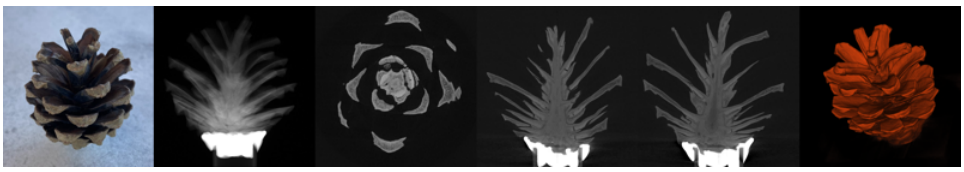

Figure 12: Overview of the dataset [Meaney, 2022]

To perform optimal transport between matrices, we will consider the indices (i, j) of the matrix as the 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. In fact, we can define $\mu = \sum_{i=1}^{n} p_i^s \delta_{x_i^s}$, the source distribution, and $\nu = \sum_{i=1}^{m} p_i^t \delta_{x_i^t}$, the target distribution.

4.2.2 Without regularizer

First, we calculated the approximated OT matrix between two distributions of the dataset, without any regularizer.

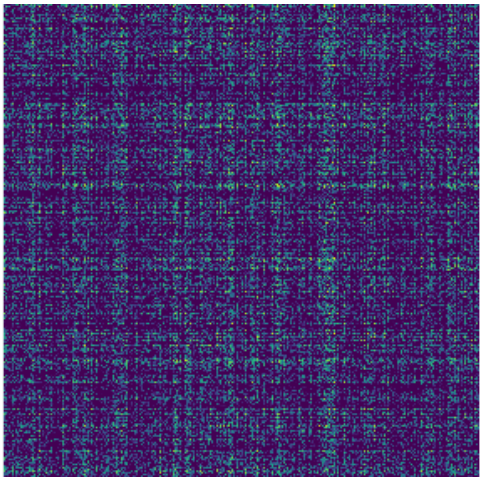

Figure 13: Approximated OT matrix between two distributions of the dataset, with $h = \ell_2^2$. Datas' reference used : 20201118.0701 and 20201118.0702.

The result is a matrix that doesn't tell the whole story...

4.2.3 With the nuclear norm regularizer

Then, we calculated the approximated OT matrix between the two same distributions of the dataset, without the nuclear norm regularizer introduced in Section 3.2.3.

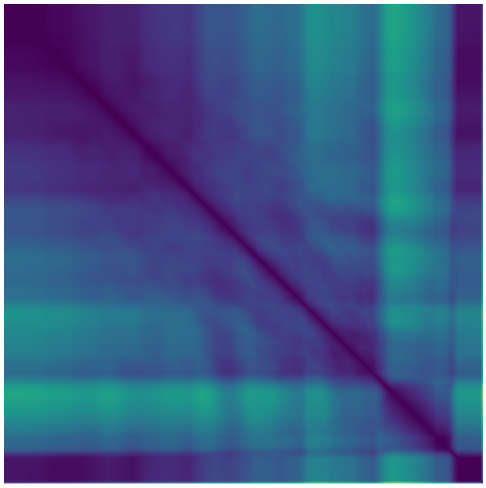

Figure 14: Approximated OT matrix between two distributions of the dataset, with $h = \ell_2^2 + \tau_*$. Datas' reference used : 20201118.0701 and 20201118.0702.

This alignment seems better than the previous one because it looks less scattered, but it is still really difficult to interpret. It tells us nothing about the "low-rank" nature of the nuclear norm...

Experiments on these matrices are not really interpretable as predicted...

5 Conclusion

We have obtained results that confirm our theoretical guarantees: the ℓ_1 regularizer does induce sparsity, and the $\mathbf{b}$-directional regularizer induces penalization in the direction of $\mathbf{b}$.

With regard to our extension to optimal transport between matrices, we haven't really been able to demonstrate the desired properties. We may have approached the problem in the wrong way, and our choice of aligning two matrices is certainly not what makes the low-rank properties we want to highlight interpretable.

To improve on this, we would have to rethink our experiments, perhaps along the lines of [Sebbouh et al., 2023], in order to bring about the desired results.

6 Connexion with the course

The work of [Klein et al., 2023] is deeply connected to optimal transport notions introduced in [Peyré and Cuturi, 2020]. In particular, it deals a lot with [Peyré and Cuturi, 2020,

Section 2.] of the course focuses on [Monge, 1781]'s formulation and [Brenier, 1991]'s theorem, since it is based on a generalization of the latter.

References

[Ambrosio et al., 2004] Ambrosio, L., Kirchheim, B., and Pratelli, A. (2004). Existence of optimal transport maps for crystalline norms. *Duke Mathematical Journal*, 125(2):207–241. 2, 3

[Ambrosio and Pratelli, 2003] Ambrosio, L. and Pratelli, A. (2003). Existence and stability results in the l1 theory of optimal transportation. *Optimal Transportation and Applications: Lectures given at the CIME Summer School, held in Martina Franca, Italy, September 2-8, 2001*, pages 123–160. 2, 3

[Bach, 2008] Bach, F. (2008). Learning with sparsity-inducing norms. 4, 9

[Bauschke and Combettes, 2011] Bauschke, H. H. and Combettes, P. L. (2011). *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer. 7

[Beck and Teboulle, 2009] Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202. 7

[Bradbury et al., 2018] Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q. (2018). JAX: composable transformations of Python+NumPy programs. 2

[Brenier, 1991] Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4). 3, 4, 13

[Carlier et al., 2010] Carlier, G., De Pascale, L., and Santambrogio, F. (2010). A strategy for non-strictly convex transport costs and the example of $\|x - y\|^p$ in r^2. *Communications in Mathematical Sciences*, 8(4):931–941. 2, 3

[Cuturi et al., 2023] Cuturi, M., Klein, M., and Ablin, P. (2023). Monge, bregman and occam: Interpretable optimal transport in high-dimensions with feature-sparse maps. 1, 4

[Cuturi et al., 2022] 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. 2, 7

[Evans and Gangbo, 1999] Evans, L. C. and Gangbo, W. (1999). *Differential equations methods for the Monge-Kantorovich mass transfer problem*. American Mathematical Soc. 2, 3

[Flamary et al., 2021] Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., Gautheron, L., Gayraud, N. T., Janati, H., Rakotomamonjy, A., Redko, I., Rolet, A., Schutz, A., Seguy, V., Sutherland, D. J., Tavenard, R., Tong, A., and Vayer, T. (2021). Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8. 2

[Harris et al., 2020] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362. 2

[Kantorovich, 2006] Kantorovich, L. (2006). On the translocation of masses. *Journal of Mathematical Sciences*, 133:1381–1382. 1

[Klein et al., 2023] Klein, M., Pooladian, A.-A., Ablin, P., Ndiaye, E., Niles-Weed, J., and Cuturi, M. (2023). Learning costs for structured monge displacements. 1, 4, 5, 6, 12

[Meaney, 2022] Meaney, A. (2022). Cone-Beam Computed Tomography Dataset of a Pine Cone. 11

[Monge, 1781] Monge, G. (1781). *Mémoire sur la théorie des déblais et des remblais*. De l'Imprimerie Royale. 2, 13

[Parikh et al., 2014] Parikh, N., Boyd, S., et al. (2014). Proximal algorithms. *Foundations and trends® in Optimization*, 1(3):127–239. 7

[Peyré and Cuturi, 2020] Peyré, G. and Cuturi, M. (2020). Computational optimal transport. 2, 12, 13

[Rockafellar, 1976] Rockafellar, R. T. (1976). Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898. 7

[Santambrogio, 2015] Santambrogio, F. (2015). Optimal transport for applied mathematicians. 2, 3, 4, 7

[Schreck et al., 2015] Schreck, A., Fort, G., Corff, S. L., and Moulines, E. (2015). A shrinkage-thresholding metropolis adjusted langevin algorithm for bayesian variable selection. 4

[Sebbouh et al., 2023] Sebbouh, O., Cuturi, M., and Peyré, G. (2023). Structured transforms across spaces with cost-regularized optimal transport. 5, 6, 12

[Vayer, 2020] Vayer, T. (2020). *A contribution to Optimal Transport on incomparable spaces*. PhD thesis. Thèse de doctorat dirigée par Courty, NicolasChapel, Laetitia et Tavenard, Romain Informatique Lorient 2020. 2