Optimal Transport for Signed Measures

Tobías Liaudat

Under the supervision of Marco Cuturi

CosmosClub Seminar
CosmoStat, CEA-Saclay

11/02/2019

Outline

Optimal Transport metric

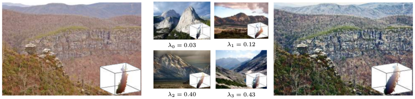
Wasserstein distance $\longrightarrow$ Useful to compare histograms and point clouds. (Typical scenario in machine learning tasks)

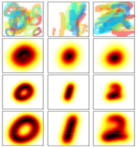
Entropic regularization $\longrightarrow$ Allows fast calculation of an approximate solution.

Examples:

- Bag of Features.
- Color histograms.
- Barycenter calculation.
- Measures with non overlapping support.
- Generative models.

And for signal processing tasks?





CREST

Optimal Transport metric

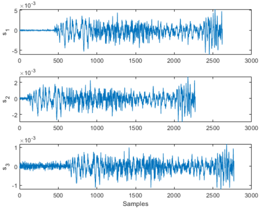
First obstacle ⟶ Signals are in general signed.

Objective ⟶ Extend the Wasserstein distance to signed measures.

Direct applications:

- Blind Source Separation
- Dictionary Learning

More to explore..



Outline

Measures

Positive Radon Measures μ on a set X

Continuous	Discrete
$d\mu(x) = m_\mu(x)dx$	$\mu = \sum_i \mu_i \delta_{x_i}$

Measure of sets $A \subset X$:

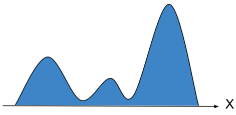
$$\mu(A) = \int_A d\mu(x) \in \mathbb{R} \qquad \mu(A) = \sum_i \mu_i \delta_{x_i}(A) \in \mathbb{R}$$

Integration against continuous functions:

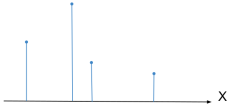
$$\int_X g\ d\mu = \int_X g(x) m_\mu(x) dx \in \mathbb{R} \qquad \int_X g\ d\mu = \sum_i \mu_i g(x_i) \in \mathbb{R}$$

Probability measures: $\mu(X) = \int_X d\mu(x) = 1$

Continuous density: $m_\mu(x)$



Discrete density



Norms and Strong Topologies

Some norms induce the strong topology on the space X.

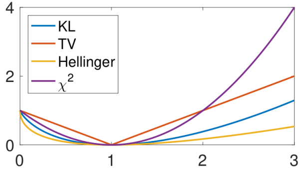
Some examples on the densities

The L^p norms:

$$\|m_\alpha - m_\beta\|_{L^p} := \left(\int_X (m_\alpha(x) - m_\beta(x))^p \, \mathrm{d}x\right)^{1/p}.$$

Csiszar divergences:

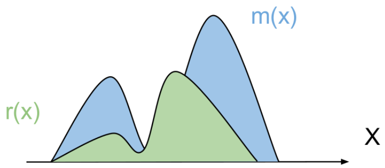
$$\mathcal{D}_\varphi(\alpha|\beta) := \int_X \varphi\left(\frac{\mathrm{d}\alpha}{\mathrm{d}\beta}\right)\mathrm{d}\beta + \varphi'_\infty \alpha^\perp(X), \quad \left(\frac{\mathrm{d}\alpha}{\mathrm{d}x} \leftrightarrow \frac{\mathrm{d}\beta}{\mathrm{d}x}\right) \longrightarrow \left(\frac{\mathrm{d}\alpha}{\mathrm{d}\beta} \leftrightarrow 1\right)$$



$$\begin{aligned}
\mathcal{X}^2 : &\quad \varphi(s) = |s-1|^2 \\
\text{TV norm} : &\quad \varphi(s) = |s-1| \\
\text{Hellinger} : &\quad \varphi(s) = |s-1|^2 \\
\text{KL} : &\quad \varphi(s) = s\log(s) \\
\text{Generalized KL} : &\quad \varphi(s) = s\log(s) - s + 1
\end{aligned}$$

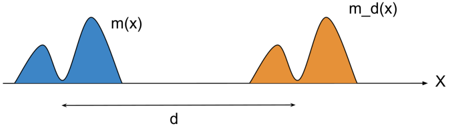
Norms and Strong Topologies

Idea: Norms inducing strong topologies compare vertical values (same support).



Not useful to compare measures with disjoint support.

$$\|m - m_d\|_{L^p} = \text{cst} \neq f(d)$$



Norms and Weak Topologies

Instead, norms inducing weak topologies can overcome disjoint supports. They metrize the weak convergence.

Weak convergence of Radon measures

Radon measure: $\int_A f \mathrm{d}\mu$

On a compact domain X, $\forall$ continuous function f:

$$\int_A f \mathrm{d}\mu_n \xrightarrow{n\to\infty} \int_A f \mathrm{d}\mu$$

Convergence in law of random vectors

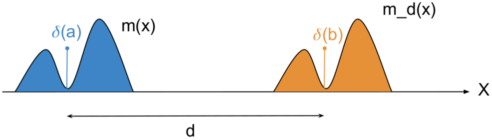
Random vector: $\mathbb{P}(X \in A)$

$\forall$ set A:

$$\mathbb{P}(X_n \in A) \xrightarrow{n\to\infty} \mathbb{P}(X \in A)$$

The Wasserstein distance metrizes the weak convergence.

$$W_p\left(\delta_a, \delta_b\right) = d(a, b) = \mathrm{d}$$



Outline

Monge's Formulation

Monge's transport (1784)

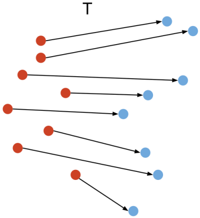
Let $c(x, y)$ be a cost function defined for points $(x, y) \in X \times Y$ and $T : X \to Y$ a map so that:

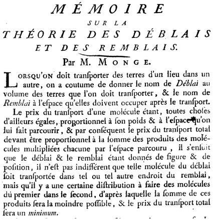
$$\min_{\nu = T_{\#}\mu} \int_X c\,(x, T(x))\,\mathrm{d}\nu(x)$$

The condition $\nu = T_{\#}\mu$ ensures that all the mass from μ is transported to ν by the map T and $T_{\#}$ is the push-forward operator.

Problems:

- Non-uniqueness
- Non-existence




MÉMOIRE
SUR LA
THÉORIE DES DÉBLAIS
ET DES REMBLAIS.
Par M. MONGE.

LORSQU'ON doit tranſporter des terres d'un lieu dans un autre, on a coutume de donner le nom de *Déblai* au volume des terres que l'on doit tranſporter, & le nom de *Remblai* à l'eſpace qu'elles doivent occuper après le tranſport.

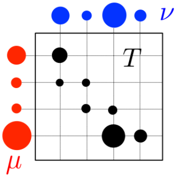
Le prix du tranſport d'une molécule étant, toutes choſes d'ailleurs égales, proportionnel à ſon poids & à l'eſpace qu'on lui fait parcourir, & par conſéquent le prix du tranſport total devant être proportionnel à la ſomme des produits des molécules multipliées chacune par l'eſpace parcouru, il s'enſuit que le déblai & le remblai étant donnés de figure & de poſition, il n'eſt pas indifférent que telle molécule du déblai ſoit tranſportée dans tel ou tel autre endroit du remblai, mais qu'il y a une certaine diſtribution à faire des molécules du premier dans le ſecond, d'après laquelle la ſomme de ces produits ſera la moindre poſſible, & le prix du tranſport total ſera un *minimum*.

Kantorovitch's Formulation - Relaxation

Discrete problem:

$$\mu = \sum_i \mu_i \delta_{x_i}, \quad \nu = \sum_j \nu_j \delta_{y_j}, \quad C_{i,j} = c(x_i, y_j) \geq 0.$$

Broaden the feasible maps.



Couplings - Kantorovitch's relaxation

$$\mathcal{U}(\mu,\nu) := \left\{ T \in \mathbb{R}_+^{n\times m} \ : \ T\mathbb{1}_m = \mu \ , \ T^T\mathbb{1}_n = \nu \right\}$$

Kantorovitch's formulation

$$L_C(\mu,\nu) = \min_{T\in\mathcal{U}(\mu,\nu)} \sum_{i,\,j} T_{i,\,j} C_{i,\,j}$$

If the cost is chosen as a distance then $L_{D^p}(\mu,\nu) = W_p^p(\mu,\nu)$, the **Wasserstein distance**.

Image credits: [Peyré&Cuturi18]

Kantorovitch's Formulation

The 3 settings:

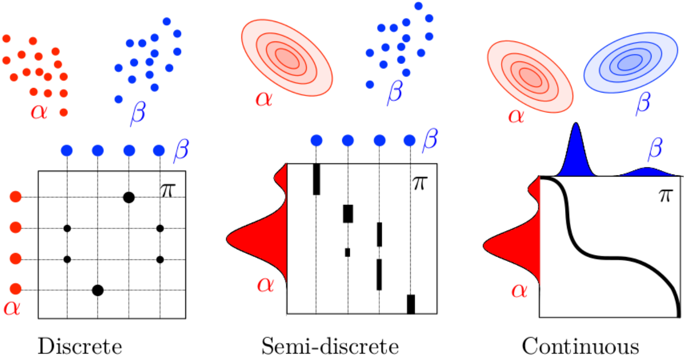


Image credits: [Peyré&Cuturi18]

Outline

Entropic Regularization

Regularized OT [Cuturi13]

$$L_C^{\varepsilon}(\mu, \nu) = \min_{T \in \mathcal{U}(\mu,\nu)} \sum_{i,j} T_{i,j} C_{i,j} - \varepsilon H(T),$$

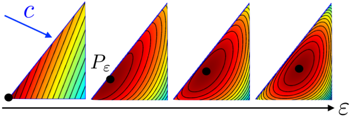
where H is the entropy:

$$H(T) := - \sum_{i,j} T_{i,j} \log(T_{i,j}) - T_{i,j}.$$

Main implications:

- **Approximate** solution $\longrightarrow$ Regulated by ε
- **Fast** solver $\longrightarrow$ Sinkhorn's algorithm
- Easier to **differentiate** $\longrightarrow$ Use as a loss

Entropic Regularization



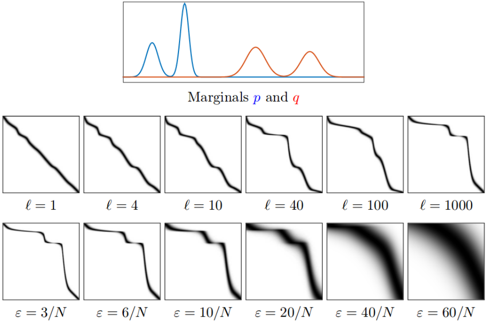


Image credits: [Peyré&Cuturi18]

Sinkhorn's Algorithm

Reformulation of regularized OT

The problem can be rewritten as:

$$\min_{T \in \mathcal{U}(\mu,\nu)} \mathrm{KL}(T|K) \quad , \quad \text{where } K = e^{-D^p/\varepsilon}$$

Property [Cuturi13][Sinkhorn&Knopp67]

Given the regularized OT problem, one has unique vectors a, b so that:

$$T = \mathrm{diag}(a) K \mathrm{diag}(b) \quad , \quad \text{with } T \in \mathcal{U}(\mu,\nu)$$

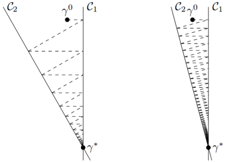
Row constraint: $T\mathbb{1}_m = \mu \iff a \odot (Kb) = \mu$

Column constraint: $T^T\mathbb{1}_n = \nu \iff b \odot (K^T a) = \nu$

Sinkhorn's iterations:

- $a \leftarrow \frac{\mu}{Kb}$
- $b \leftarrow \frac{\nu}{K^T a}$

Interpretation: Iterative projections.



Outline

Unbalanced OT Formulation

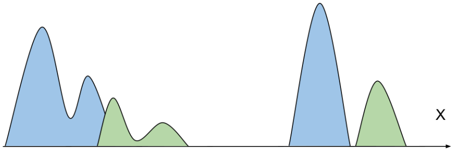
Generalized Sinkhorn formulation [Peyré&Cuturi18][Chizat...17]

Generalize the regularized formulation by relaxing the constraints:

$$W_c^{\varepsilon}(\mu,\nu) := \min_{T\in\mathbb{R}_+^{n\times m}} \sum_{i,j} T_{i,j}C_{i,j} - \varepsilon H(T) + F_1(T\mathbb{1}_m|\mu) + F_2(T^T\mathbb{1}_n|\nu).$$

- Idea: Use a "strong" norm for the constraints.
- Not necessary that $\|\mu\| = \|\nu\|$.
- The functions $F(\cdot)$ penalizes the not transported mass.
- Fast calculation with a Sinkhorn's algorithm adaptation.
- Not a distance.

Examples: $F(\cdot|p) = i_{\{=\}}(\cdot|p)$, $F(\cdot|p) = \lambda\mathrm{KL}(\cdot|p)$, $F(\cdot|p) = \lambda\mathrm{TV}(\cdot|p)$.



Generalized Wasserstein Distance

Generalized Wasserstein distance [Piccoli&Rossi14]

The formulation is a distance for unbalanced measures:

$$W_p^{(a,b)}(\mu,\nu) := \left(\inf_{\substack{\tilde{\mu},\tilde{\nu}\in\mathcal{M}(\mathbb{R}^d) \\ \|\tilde{\mu}\|=\|\tilde{\nu}\|}} a^p \left(\|\mu-\tilde{\mu}\|_1 + \|\nu-\tilde{\nu}\|_1\right)^p + b^p W_p^p(\tilde{\mu},\tilde{\nu}) \right)^{1/p}$$

Verify the properties of a distance [Piccoli&Rossi14]:

- Triangle inequality: $W_p^{(a,b)}(\mu,\eta) \leq W_p^{(a,b)}(\mu,\nu) + W_p^{(a,b)}(\nu,\eta)$.
- Symmetric: $W_p^{(a,b)}(\mu,\nu) = W_p^{(a,b)}(\nu,\mu)$.
- $W_p^{(a,b)}(\mu,\nu) = 0 \iff \mu = \nu$.
- The infimum is always attained.
- b (resp. a) parametrizes the ease of transport (resp. creation/cancellation) of mass.

$\longrightarrow$ Hard to calculate.

Outline

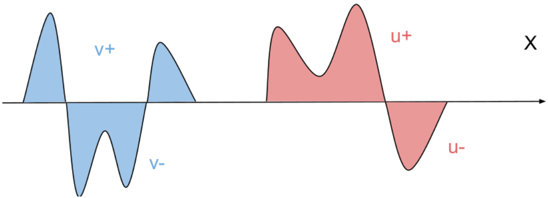
Signed Formulation

Idea: Re-use the unbalanced formulation and adapt is to signed values.

Define a decomposition:

$$\mu = \mu_+ - \mu_- \in \mathbb{R}^n, \quad \text{where } \mu_+, \mu_- \in \mathbb{R}_+^n$$
$$\nu = \nu_+ - \nu_- \in \mathbb{R}^m, \quad \text{where } \nu_+, \nu_- \in \mathbb{R}_+^m$$

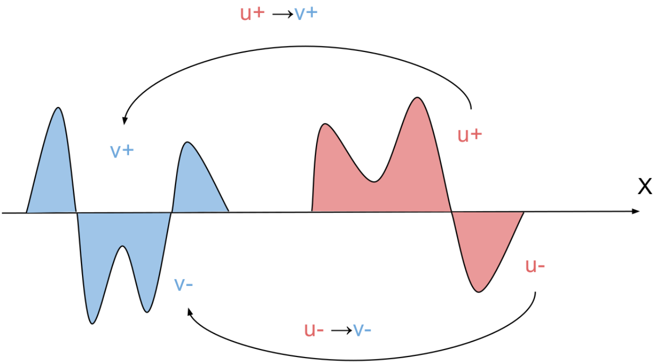
where the Jordan's decomposition is the one such that for μ (resp. ν) $\text{supp}(\mu_+) \cap \text{supp}(\mu_-) = \{\emptyset\}$.



Signed Formulation

Interactions between the signed parts?

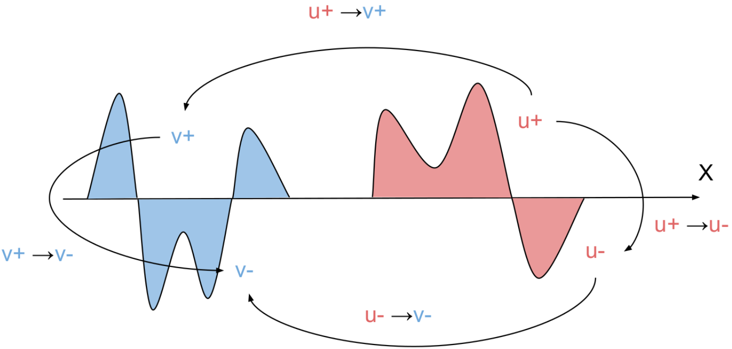
$$\mu_+ \longrightarrow \nu_+ \quad , \quad \mu_- \longrightarrow \nu_- \quad \textbf{Transportation}$$



Signed Formulation

Interactions between the signed parts?

$$\mu_+ \longrightarrow \nu_+ \quad , \quad \mu_- \longrightarrow \nu_- \qquad \textbf{Transportation}$$
$$\mu_+ \longrightarrow \mu_- \quad , \quad \nu_+ \longrightarrow \nu_- \qquad \textbf{Cancellation}$$



Signed Formulation

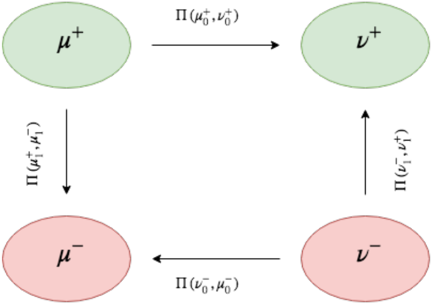
Define the signed transport in terms of classic optimal transport [Mainini12]

$$\mathbb{W}_p(\mu, \nu) = W_p(\mu_+ + \nu_-, \mu_- + \nu_+)$$

Actions taking place:

- **Transport** between same sign measures. Ex: $\mu_{0,+} \to \nu_{0,+}$
- **Cancellation** between different sign measures. Ex: $\mu_{1,+} \to \mu_{1,-}$

$\to$ Only works if the decomposition is balanced.
$\to$ Which decomposition to use?
$\to$ Properties?



Signed Formulation

Use the unbalanced formulation with TV regularization.

Signed optimal transport

$$\mathbb{W}_p^{(a,b)}(\mu,\nu) = W_p^{(a,b)}(\mu_+ + \nu_-, \mu_- + \nu_+),$$

where:

$$\left(W_p^{(a,b)}\right)^p(\mu,\nu) := \inf_{\substack{\tilde{\mu},\tilde{\nu}\in\mathcal{M}(\mathbb{R}^d)\\ \|\tilde{\mu}\|=\|\tilde{\nu}\|}} a^p\left(\|\mu-\tilde{\mu}\|_1 + \|\nu-\tilde{\nu}\|_1\right)^p + b^p W_p^p(\tilde{\mu},\tilde{\nu}).$$

- $\rightarrow$ ~~Only works if the decomposition is balanced.~~
- $\rightarrow$ Which decomposition to use?
- $\rightarrow$ Properties?

Signed Formulation

Signed optimal transport

$$\mu^{+} = \mu_0^{+} + \mu_1^{+} + \tilde{\mu}^{+}\,, \quad \nu^{+} = \nu_0^{+} + \nu_1^{+} + \tilde{\nu}^{+}\,,$$
$$\mu^{-} = \mu_0^{-} + \mu_1^{-} + \tilde{\mu}^{-}\,, \quad \nu^{-} = \nu_0^{-} + \nu_1^{-} + \tilde{\nu}^{-}\,,$$

where $\tilde{\mu}^{+}, \tilde{\mu}^{-} \in \mathbb{R}^n$ and $\tilde{\nu}^{+}, \tilde{\nu}^{-} \in \mathbb{R}^m$.

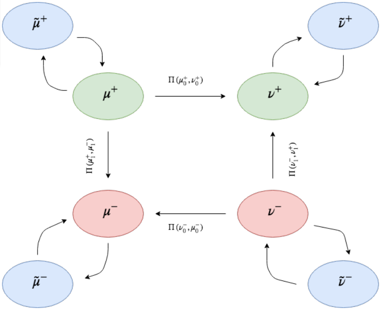
Actions taking place:

- **Transport** between same sign measures. Ex: $\mu_{0,+} \rightarrow \nu_{0,+}$
- **Cancellation** between different sign measures. Ex: $\mu_{1,+} \rightarrow \mu_{1,-}$
- **Creation / Destruction** To manage the unbalance scenario.. Ex: $\tilde{\mu}_{+} \leftrightarrow \mu_{+}$

→ ~~Only works if the decomposition is balanced.~~

→ Which decomposition to use?

→ Properties?



Signed Formulation

→ ~~Only works if the decomposition is balanced.~~
→ Which decomposition to use?

Proof of Prop.1 [Piccoli&Rossi18]

$\mathbb{W}_1(\mu, \nu)$ does not depends on the decomposition.
Based on the Lemma 4.

Lemma 4 [Piccoli&Rossi18]

Property of the Generalized Wasserstein Distance:
$W_1^{(a,b)}(\mu + \eta, \nu + \eta) = W_1^{(a,b)}(\mu, \nu)$

Strategy → Use the Jordan decomposition.

Signed Formulation

→ ~~Only works if the decomposition is balanced.~~
→ ~~Which decomposition to use?~~
→ Properties?

Prop.1 [Piccoli&Rossi18]

$\mathbb{W}_1(\mu, \nu)$ is a **distance** on the space $\mathcal{M}(\mathbb{R}^d)$ of signed measures with finite mass.

Lemma 5 [Piccoli&Rossi18]

- $\mathbb{W}_1^{(a,b)}(\mu, \nu) = 0 \iff \mu = \nu$,
- $\mathbb{W}_1^{(a,b)}(\mu + \eta, \nu + \eta) = \mathbb{W}_1^{(a,b)}(\mu, \nu)$,
- $\mathbb{W}_1^{(a,b)}(\mu_1 + \mu_2, \nu_1 + \nu_2) \leq \mathbb{W}_1^{(a,b)}(\mu_1, \nu_1) + \mathbb{W}_1^{(a,b)}(\mu_2, \nu_2)$.

Signed Formulation

→ ~~Only works if the decomposition is balanced.~~
→ ~~Which decomposition to use?~~
→ ~~Properties?~~

→ Hard to calculate.

Use the entropic regularization → Easier to compute.

Signed regularized optimal transport

$$\mathbb{W}_1^{(\lambda,\varepsilon)}(\mu,\nu) = W_1^{(\lambda,\varepsilon)}(\mu_+ + \nu_-, \mu_- + \nu_+),$$

where:

$$W_1^{(\lambda,\varepsilon)}(\mu,\nu) := \min_{\substack{\tilde{\mu},\tilde{\nu}\in\mathcal{M}(\mathbb{R}^d)\\ \|\tilde{\mu}\|=\|\tilde{\nu}\|}} \lambda\left(\|\mu-\tilde{\mu}\|_1 + \|\nu-\tilde{\nu}\|_1\right) + W_1^{\varepsilon}(\tilde{\mu},\tilde{\nu}).$$

CREST

Outline

1. Introduction
2. Notions on Optimal Transport
3. From Unbalanced to Signed OT
   - Signed Optimal Transport
   - Calculation
   - To Conclude

Primal formulation

$$W_1^{(\lambda,\varepsilon)}(\mu,\nu) := \min_{\substack{\tilde{\mu},\tilde{\nu}\in\mathcal{M}(\mathbb{R}^d) \\ \|\tilde{\mu}\|=\|\tilde{\nu}\|}} \lambda\left(\|\mu-\tilde{\mu}\|_1 + \|\nu-\tilde{\nu}\|_1\right) + W_1^{\varepsilon}(\tilde{\mu},\tilde{\nu}).$$

$$W_1^{(\lambda,\varepsilon)}(\mu,\nu) := \min_{T\in\mathbb{R}_+^{n\times m}} \lambda\left(\|\mu - T\mathbb{1}_m\|_1 + \|\nu - T^T\mathbb{1}_m\|_1\right) + \langle T, D_1\rangle - \varepsilon H(T).$$

Primal formulation

$$W_1^{(\lambda,\varepsilon)}(\mu,\nu) = \min_{T\in\mathbb{R}_+^{n\times m}} F_1\left(T\mathbb{1}_m\right) + F_2\left(T^T\mathbb{1}_m\right) + \varepsilon \mathrm{KL}(T|K)$$

where $(K)_{i,j} = e^{(D_1)_{i,j}/\varepsilon}$.

Remember:

$$T(x,y) = a(x)K(x,y)b(y) \quad , \quad (a,b) := \left(e^{-u/\varepsilon}, e^{-v/\varepsilon}\right).$$

Dual formulation

Fenchel-Rockafellar Theorem

f and g being lower semi-continuous and proper convex functions defined in E and F resp. Let A be a linear operator and A^* its adjoint. It holds:

$$\sup_{x \in E} -f(-x) - g(Ax) = \min_{y^* \in F^*} f^*(A^* y^*) + g^*(y^*).$$

Dual formulation

$$\max_{u,v} -F_1^*(u) - F_2^*(v) - \varepsilon \left\langle e^{-u/\varepsilon}, K e^{-v/\varepsilon} \right\rangle$$

Block coordinate relaxation

$$u^{(l+1)} = \arg\max_{u} -F_1^*(u) - \varepsilon \left\langle e^{-u/\varepsilon}, K e^{-v^{(l)}/\varepsilon} \right\rangle \qquad (P_u)$$

$$v^{(l+1)} = \arg\max_{v} -F_2^*(v) - \varepsilon \left\langle e^{-u^{(l+1)}/\varepsilon}, K e^{-v/\varepsilon} \right\rangle \qquad (P_v)$$

Solving the dual formulation

Observation: Use the Fenchel-Rockafellar theorem again.

$$\sup_u -F_1^*(u) - \varepsilon \left\langle e^{-u/\varepsilon}, Ke^{-v^{(l)}/\varepsilon} \right\rangle = \min_s F_1(s) + \varepsilon \mathrm{KL}(s|Ke^{v^{(l)}/\varepsilon})$$

The minimizer of the right part $s^\star$ belongs to the subdiferential of $u \mapsto \left\langle e^{-u/\varepsilon}, Ke^{-v^{(l)}/\varepsilon} \right\rangle$ at the point $u^\star$, the maximizer of the left part.

$$s^\star = e^{u^\star/\epsilon}(Ke^{v^{(l)}/\varepsilon})$$

The right part looks like a proximal operator!

$$\mathrm{Prox}^{\mathrm{KL}}_{F_1/\varepsilon}(t) := \arg\min_r F_1(r) + \varepsilon \mathrm{KL}(r|t)$$

We can rewrite:

$$(s^\star =)\ \mathrm{Prox}^{\mathrm{KL}}_{F_1/\varepsilon}(Ke^{v^{(l)}/\varepsilon}) = e^{u^\star/\epsilon}(Ke^{v^{(l)}/\varepsilon})$$

$$e^{u^\star/\epsilon} = \frac{\mathrm{Prox}^{\mathrm{KL}}_{F_1/\varepsilon}(Ke^{v^{(l)}/\varepsilon})}{(Ke^{v^{(l)}/\varepsilon})}$$

Solving the dual formulation

Rewrite the iterations using the Prox operation in terms of $(a, b) = \left(e^{-u/\varepsilon}, e^{-v/\varepsilon}\right)$:

Algorithm iterations [Chizat..16]

$$a^{(l+1)} = \frac{\text{prox}^{\text{KL}}_{F_1/\epsilon}(K b^{(l)})}{K b^{(l)}} \qquad (P_u)$$

$$b^{(l+1)} = \frac{\text{prox}^{\text{KL}}_{F_2/\epsilon}(K^T a^{(l+1)})}{K^T a^{(l+1)}} \qquad (P_v)$$

Explicit formulas for the prox operators!

Reconstruction after convergence

$$T^\star(x, y) = a^\star(x) K(x, y) b^\star(y)$$

Direct calculation of the distance with $T^\star$.

Outline

Conclusion & To-Do List

Some conclusions:

- We formulate an approximation to an optimal transport distance to signed measures.
- A fast algorithm for its calculation.

To-Do list:

- Use as a loss function.
- Analyze its differentiability.
- Extend applications to work with signed measures.
- Analyze theoretical properties of the signed regularized formulation.