Fast Optimal Transport through Sliced Generalized Wasserstein Geodesics

Joint work with Guillaume Mahey, Gilles Gasso, Clément Bonet and Nicolas Courty
NeurIPS 2023 [5]

Laetitia Chapel
laetitia.chapel@irisa.fr

IRISA, Rennes, France
Institut Agro Rennes-Angers

Workshop on Optimal Transport: from theory to applications, Berlin 2024
Table of Contents

Background on Optimal Transport
  Optimal transport and Wasserstein distance
  Transport map and Wasserstein Geodesics
  Curvature of the Wasserstein space
  Wasserstein Generalized Geodesics
  Computational Optimal Transport

Sliced Wasserstein Generalized Geodesic
  SWGG with a PWD-like formulation
  SWGG with a Generalized Geodesic formulation

Experimental results
  Computational aspects
  Gradient flows
  Pan sharpening / image colorization
  Point cloud matchings
  Optimal transport dataset distances

Conclusion

Bibliography
Background on Optimal Transport
Optimal transport and Wasserstein distance

Optimal transport and Wasserstein distance

\[
\mathcal{O}T(\mu_1, \mu_2) \triangleq \inf_{\gamma \in \Gamma(\mu_1, \mu_2)} \int_{X \times Y} c(x, y) \, d\gamma(x, y)
\]

where \( \Gamma(\mu_1, \mu_2) \) \( \overset{\text{def}}{=} \{ \gamma \in \mathcal{M}_+(X \times Y) \, \text{s.t.} \, (\pi_x)_\# \gamma = \mu_1 \text{ and } (\pi_y)_\# \gamma = \mu_2 \} \) with \( \pi_x : X \times Y \to X \).
Background on Optimal Transport

Optimal transport and Wasserstein distance

- Optimal transport and Wasserstein distance

\[
\mathcal{OT}(\mu_1, \mu_2) \triangleq \inf_{\gamma \in \Gamma(\mu_1, \mu_2)} \int_{X \times Y} c(x, y) \, d\gamma(x, y)
\]

where \( \Gamma(\mu_1, \mu_2) \) def = \( \{ \gamma \in \mathcal{M}_+(X \times Y) \text{ s.t. } (\pi_x)_\# \gamma = \mu_1 \text{ and } (\pi_y)_\# \gamma = \mu_2 \} \) with \( \pi_x : X \times Y \to X \).

- Linear loss

- Marginal constraints

The transport plan \( \gamma(x, y) \) specifies for each pair \( (x, y) \) how many particles go from \( x \) to \( y \)

- Wasserstein distance when \( c(x, y) = |x - y|^p \)

\[
\mathcal{W}_p(\mu_1, \mu_2) \triangleq \left( \inf_{\gamma \in \Gamma(\mu_1, \mu_2)} \int_{X \times Y} c(x, y) \, d\gamma(x, y) \right)^{1/p}
\]
In some cases, the optimal plan $\gamma^*$ is a Monge map of the form $(\text{Id}, T)\#\mu_1$, e.g. for $p = 2$

$$\mathcal{W}_p^p(\mu_1, \mu_2) \triangleq \inf_T \int \|x - T(x)\|_2^2 \, d\mu_1(x)$$

where $T$ is a transport map and $T\#\mu_1 = \mu_2$
In some cases, the optimal plan $\gamma^*$ is a Monge map of the form $(Id, T)\#\mu_1$, e.g. for $p = 2$

$$\mathcal{W}_p^p(\mu_1, \mu_2) \triangleq \inf_T \int \|x - T(x)\|^2 d\mu_1(x)$$

where $T$ is a transport map and $T\#\mu_1 = \mu_2$

Defines for each particle located at $x$ what is its destination $T(x)$
Background on Optimal Transport
Transport map and Wasserstein Geodesics

- In some cases, the optimal plan $\gamma^*$ is a Monge map of the form $(\text{Id}, T)\#\mu_1$, e.g. for $p = 2$

$$\mathcal{W}_p^p(\mu_1, \mu_2) \triangleq \inf_T \int \|x - T(x)\|^2_2 \, d\mu_1(x)$$

where $T$ is a transport map and $T\#\mu_1 = \mu_2$

- Wasserstein geodesics $\mu^{1\rightarrow2}(t) \triangleq (tT^{1\rightarrow2} + (1 - t)\text{Id})\#\mu_1$ with $T^{1\rightarrow2}$ the optimal map

For short, we denote $\mu^{1\rightarrow2}$ for $t = 0.5$
The Wasserstein space is of positive curvature

\[ \mathcal{W}_2^2(\mu^{1\rightarrow 2}, \nu) \geq \frac{1}{2} \mathcal{W}_2^2(\mu_1, \nu) + \frac{1}{2} \mathcal{W}_2^2(\nu, \mu_2) - \frac{1}{4} \mathcal{W}_2^2(\mu_1, \mu_2) \]

or equivalently

\[ \mathcal{W}_2^2(\mu_1, \mu_2) \geq 2 \mathcal{W}_2^2(\mu_1, \nu) + 2 \mathcal{W}_2^2(\nu, \mu_2) - 4 \mathcal{W}_2^2(\mu^{1\rightarrow 2}, \nu) \]

for \( \nu \) a \textbf{pivot measure}. 

Parallelogram law in \( \mathbb{R}^d \)

Positive curvature of \( \mathcal{W} \) space
Background on Optimal Transport

Curvature of the Wasserstein space

- The Wasserstein space is of positive curvature

\[ \mathcal{W}^2_2(\mu_1 \to \mu_2, \nu) \geq \frac{1}{2} \mathcal{W}^2_2(\mu_1, \nu) + \frac{1}{2} \mathcal{W}^2_2(\nu, \mu_2) - \frac{1}{4} \mathcal{W}^2_2(\mu_1, \mu_2) \]

or equivalently

\[ \mathcal{W}^2_2(\mu_1, \mu_2) \geq 2 \mathcal{W}^2_2(\mu_1, \nu) + 2 \mathcal{W}^2_2(\nu, \mu_2) - 4 \mathcal{W}^2_2(\mu_1 \to \mu_2, \nu) \]

for \( \nu \) a pivot measure.

- The Wasserstein space is flat when \( \mu_1, \mu_2, \nu \) are 1d

\[ \mathcal{W}^2_2(\mu_1, \mu_2) = 2 \mathcal{W}^2_2(\mu_1, \nu) + 2 \mathcal{W}^2_2(\nu, \mu_2) - 4 \mathcal{W}^2_2(\mu_1 \to \mu_2, \nu) \]
Background on Optimal Transport
Wasserstein Generalized Geodesics

- Has been introduced by Ambrosio et al. [1]

- Wasserstein Geodesic: \( \mu^{1 \rightarrow 2}(t) \triangleq (t \ T^{1 \rightarrow 2} + (1 - t)Id) \# \mu_1 \)

- Wasserstein Generalized Geodesic: \( \mu^{1 \rightarrow 2}_{g}(t) \triangleq (t \ T^{\nu \rightarrow \mu_2} + (1 - t) T^{\nu \rightarrow \mu_1}) \# \nu \)

  for \( \nu \) a pivot measure.
Background on Optimal Transport

Wasserstein Generalized Geodesics

- Has been introduced by Ambrosio et al. [1]
- Wasserstein Geodesic: $\mu_1^{\rightarrow 2}(t) \triangleq (t \ T_1^{\rightarrow 2} + (1 - t) Id) \# \mu_1$
- **Wasserstein Generalized Geodesic:** $\mu_g^{\rightarrow 2}(t) \triangleq (t \ T_{\nu^{\rightarrow \mu_2}} + (1 - t) \ T_{\nu^{\rightarrow \mu_1}} ) \# \nu$
  for $\nu$ a **pivot measure**.
- Negative curvature:
  \[ \mathcal{W}_2^2(\mu_g^{\rightarrow 2}, \nu) \leq \frac{1}{2} \mathcal{W}_2^2(\mu_1, \nu) + \frac{1}{2} \mathcal{W}_2^2(\nu, \mu_2) - \frac{1}{4} \mathcal{W}_2^2(\mu_1, \mu_2) \]
Background on Optimal Transport
Wasserstein Generalized Geodesics

- Has been introduced by Ambrosio et al. [1]
- Wasserstein Geodesic:
  \[ \mu^{1 \rightarrow 2}(t) \triangleq (t \ T^{1 \rightarrow 2} + (1 - t)I) \# \mu_1 \]
- Wasserstein Generalized Geodesic:
  \[ \mu^{g \rightarrow 2}(t) \triangleq (t \ T^{\nu \rightarrow \mu_2} + (1 - t) \ T^{\nu \rightarrow \mu_1}) \# \nu \]
  for \( \nu \) a pivot measure.
- Negative curvature:
  \[ W_2^2(\mu^{g \rightarrow 2}, \nu) \leq \frac{1}{2} W_2^2(\mu_1, \nu) + \frac{1}{2} W_2^2(\nu, \mu_2) - \frac{1}{4} W_2^2(\mu_1 \rightarrow \mu_2) \]

- Wasserstein distance:
  \[ W_2^2(\mu_1, \mu_2) = 2 W_2^2(\mu_1, \nu) + 2 W_2^2(\nu, \mu_2) - 4 W_2^2(\mu^{g \rightarrow 2}, \nu) \]
  with \( W_2^2(\mu_1, \mu_2) \geq W_2^2(\mu_1, \mu_2) \)
For $\mu_1 = \sum_{i=1}^{n} h_i \delta_{x_i}$ and $\mu_2 = \sum_{j=1}^{m} g_j \delta_{y_j}$ and a quadratic cost, we solve

$$\mathcal{W}_2^2(\mu_1, \mu_2) \triangleq \min_{\gamma \in \Gamma(\mu_1, \mu_2)} \sum_{i,j} c(x_i, y_j) \gamma_{i,j}$$

$\rightarrow$ linear solvers with $O(n^3 \log(n))$ complexity
Computational Optimal Transport

Discrete formulation of OT

- For \( \mu_1 = \sum_{i=1}^{n} h_i \delta_{x_i} \) and \( \mu_2 = \sum_{j=1}^{m} g_j \delta_{y_j} \) and a quadratic cost, we solve

\[
\mathcal{W}_2^2(\mu_1, \mu_2) \triangleq \min_{\gamma \in \Gamma(\mu_1, \mu_2)} \sum_{i,j} c(x_i, y_j) \gamma_{i,j}
\]

\( \rightarrow \) linear solvers with \( O(n^3 \log(n)) \) complexity

- When \( \mu_1 \) and \( \mu_2 \) are 1D distributions and \( n = m \) with uniform masses, the solution is given by

\[
\mathcal{W}_2^2(\mu_1, \mu_2) \triangleq \frac{1}{n} \sum_{i=1}^{n} (x_{\sigma(i)} - y_{\tau(i)})^2
\]

\( \rightarrow \) the optimal transport plan respects the ordering of the elements \( x_{\sigma(i-1)} \leq x_{\sigma(i)} \) and \( y_{\tau(i-1)} \leq y_{\tau(i)} \), complexity \( O(n \log(n)) \) and \( O(n + n \log(n)) \) for computing the distance
Computational Optimal Transport
Geodesic in 1D

- In 1D, the middle of the geodesic can be easily computed
  \[(x_{\sigma(i)} + y_{\tau(i)})/2\]

- And when we take the pivot measure \(\nu\) to be the middle of the geodesic \(\mu^{1 \rightarrow 2}\), we have
  \[\mathcal{W}_2^2(\mu_1, \mu_2) = \mathcal{W}_\nu^2(\mu_1, \mu_2) = 2\mathcal{W}_2^2(\mu_1, \nu) + 2\mathcal{W}_2^2(\nu, \mu_2)\]
Computational Optimal Transport
Sliced Wasserstein on $\mathbb{R}^d$

1. Slice the distribution along lines $\theta \in S^{d-1}$
2. Project $\mu_1$ and $\mu_2$ onto $\theta$: $P_\theta \# \mu$, with $P_\theta : \mathbb{R}^d \rightarrow \mathbb{R}, x \mapsto \langle x, \theta \rangle$
3. Compute 1d Wasserstein onto the projected samples in 1d
4. Average all the distances

$$SW_2^2(\mu_1, \mu_2) \triangleq \int_{S^{d-1}} W_2^2(P_\theta \# \mu_1, P_\theta \# \mu_2) d\omega(\theta),$$

with $\omega$ uniform distribution on $S^{d-1}$.

→ provides a lower bound of $W_2^2(\mu_1, \mu_2)$ with complexity $O(Ln + Ln \log(n))$, $L$ number of lines
Computational Optimal Transport

Projected Wasserstein Distance on $\mathbb{R}^d$

1. Slice the distribution along lines $\theta \in S^{d-1}$
2. Project $\mu_1$ and $\mu_2$ onto $\theta$: $P_{\#}\mu$, with $P^\theta : \mathbb{R}^d \to \mathbb{R}, x \mapsto \langle x, \theta \rangle$
3. Compute $\mathbb{R}^d$ Wasserstein onto the permutations obtained by sorting the projections
4. Average all the distances (mettre un theta en indice dans les sigma)

$$P\text{WD}^2_2(\mu_1, \mu_2) \triangleq \int_{S^{d-1}} \frac{1}{n} \sum_{i=1}^{n} \| x_{\sigma_{\theta}(i)} - y_{\tau_{\theta}(i)} \|_2^2 d\omega(\theta),$$

with $\omega$ uniform distribution on $S^{d-1}$.

$\rightarrow$ provides an upper bound of $W^2_2(\mu_1, \mu_2)$ with complexity $O(Ln d + Ln \log(n))$, $L$ number of lines.
Sliced Wasserstein Generalized Geodesic

SWGG with a PWD-like formulation

1. Slice the distribution along lines $\theta \in S^{d-1}$
2. Project $\mu_1$ and $\mu_2$ onto $\theta$: $P^\theta_#\mu$, with $P^\theta : \mathbb{R}^d \to \mathbb{R}, x \mapsto \langle x, \theta \rangle$
3. Compute $\mathbb{R}^d$ Wasserstein onto the permutations obtained by sorting the projections
4. Take the minimum over all the distances

$$\text{SWGG}_2^2(\mu_1, \mu_2, \theta) \triangleq \frac{1}{n} \sum_{i=1}^{n} \left\| x_{\sigma \theta}(i) - y_{\tau \theta}(i) \right\|^2_2,$$

$$\text{min-SWGG}_2^2(\mu_1, \mu_2) \triangleq \min_{\theta \in S^{d-1}} \text{SWGG}_2^2(\mu_1, \mu_2, \theta)$$
Sliced Wasserstein Generalized Geodesic

SWGG with a PWD-like formulation

Properties of min-SWGG

- It comes with a transport map: let $\theta^*$ be the optimal projection direction
  
  $T(x_i) = y_{\tau_0^{-1}(\sigma_{\theta^*}(i))}, \quad \forall 1 \leq i \leq n.$

- It is an upper bound of $\mathcal{W}$ and a lower bound of $\mathcal{PWD}$
  
  $\mathcal{W}_2^2 \leq \text{min-SWGG}_2^2 \leq \mathcal{PWD}_2^2$

  and $\mathcal{W}_2^2 = \text{min-SWGG}_2^2$ when $d > 2n$ [2]

- Complexity $O(Lnd + Ln \log(n))$ with $L$ number of lines

- The Monte-Carlo search over the $L$ lines is effective in low dimension only

→ how to design gradient descent techniques for finding $\theta^*$?

→ further properties, such as sample complexity?
**Sliced Wasserstein Generalized Geodesic**

**SWGG with a Generalized Geodesic formulation**

1. Slice the distribution along lines $\theta \in S^{d-1}$

2. Project $\mu_1$ and $\mu_2$ onto $\theta$: $Q^\theta_\#\mu$, with $Q^\theta : \mathbb{R}^d \rightarrow \mathbb{R}^d, x \mapsto \theta \langle x, \theta \rangle$

3. Define the pivot measure $\nu$ to be the Wasserstein mean of the measure $Q^\theta_\#\mu_1$ and $Q^\theta_\#\mu_2$

$$\nu = \mu_{\theta \rightarrow 2}^{1 \rightarrow 2} \triangleq \arg \min_{\mu} \mathcal{W}_2^2 (Q^\theta_\#\mu_1, \mu) + \mathcal{W}_2^2 (\mu, Q^\theta_\#\mu_2)$$

4. Take the minimum over all the following distances

$$\text{SWGG}_2^2(\mu_1, \mu_2, \theta) = 2\mathcal{W}_2^2 (\mu_1, \mu_{\theta \rightarrow 2}^{1 \rightarrow 2}) + 2\mathcal{W}_2^2 (\mu_{\theta \rightarrow 2}^{1 \rightarrow 2}, \mu_2) - 4\mathcal{W}_2^2 (\mu_{g, \theta}^{1 \rightarrow 2}, \mu_{\theta \rightarrow 2}^{1 \rightarrow 2})$$

→ the two formulations are equivalent (for continuous or discrete distributions)
**Sliced Wasserstein Generalized Geodesic**

**SWGG with a Generalized Geodesic formulation**

**Why this reformulation?**
- Define a gradient descent algorithm for optimizing over $\theta$.
- Rewrite the problem as an OT formulation with a restricted constraint set.
- Define new properties for SWGG.

**Properties of min-SWGG**
- Weak convergence.
- Translation invariance.
- SWGG is equal to $\mathcal{W}$ when one of the distributions ($\mu_2$) is supported on a line of direction $\theta$:

  $W_2^2(\mu_1, \mu_2) = W_2^2(\mu_1, Q_\theta # \mu_1) + W_2^2(Q_\theta # \mu_1, \mu_2)$

  that can be computed with a closed form.
Sliced Wasserstein Generalized Geodesic
SWGG with a Generalized Geodesic formulation

Gradient descent for optimizing over $\theta$:

- $\text{min-SWGG}_2^2(\mu_1, \mu_2) = \min_{\theta \in S^{d-1}} \frac{1}{n} \sum_{i=1}^{n} \| x_{\sigma(i)} - y_{\tau(i)} \|_2^2$ is not amenable to optimization.

- $\text{min-SWGG}_2^2(\mu_1, \mu_2) = \min_{\theta \in S^{d-1}} \mathcal{W}_2^2(\mu_1, \mu_1^{\rightarrow 2}) + 2\mathcal{W}_2^2(\mu_1^{\rightarrow 2}, \mu_2) - 4\mathcal{W}_2^2(\mu_{g,\theta}^{1 \rightarrow 2}, \mu_{\theta}^{1 \rightarrow 2})$ can be computed with a $O(dn + n \log(n))$ complexity, but $\mathcal{W}_2^2(\mu_{g,\theta}^{1 \rightarrow 2}, \mu_{\theta}^{1 \rightarrow 2})$ is still piecewise linear with $\theta \rightarrow$ rely on the blurred Wasserstein distance [3].
Sliced Wasserstein Generalized Geodesic

SWGG with a Generalized Geodesic formulation

OT with a restricted constraint set

- Discrete optimal transport, with $n = m$ and uniform masses

$$\mathcal{W}_2^2(\mu_1, \mu_2) = \min_{\gamma \in \Gamma(\mu_1, \mu_2)} \sum_{i,j} c(x_i, y_j) \gamma_{i,j}$$

where $\Gamma(\mu_1, \mu_2) = \{\gamma \in \mathbb{R}^{n \times n} \text{ s.t. } \gamma 1_n = 1_n / n, \gamma^\top 1_n = 1_n / n\}$ (Birkhoff polytope).

- min-SWGG

$$\text{min-SWGG}_2^2(\mu_1, \mu_2) = \min_{\gamma_\theta \in \Pi(\mu_1, \mu_2)} \sum_{i,j} c(x_i, y_j) \gamma_{\theta i,j}$$

where $\Pi(\mu_1, \mu_2) = \{\gamma_\theta \in \mathbb{R}^{n \times n} \text{ s.t. it is constructed from the permutahedron of the proj. distributions}\}$
Sliced Wasserstein Generalized Geodesic

SWGG with a Generalized Geodesic formulation

OT with a restricted constraint set

- Discrete optimal transport, with \( n = m \) and uniform masses

\[
\mathcal{W}_2^2(\mu_1, \mu_2) = \min_{\gamma \in \Gamma(\mu_1, \mu_2)} \sum_{i,j} c(x_i, y_j) \gamma_{i,j}
\]

where \( \Gamma(\mu_1, \mu_2) = \{\gamma \in \mathbb{R}^{n \times n} \text{ s.t. } \gamma 1_n = 1_n/n, \gamma^\top 1_n = 1_n/n\} \) (Birkhoff polytope).

- min-SWGG

\[
\text{min-SWGG}_2^2(\mu_1, \mu_2) = \min_{\gamma_\theta \in \Pi(\mu_1, \mu_2)} \sum_{i,j} c(x_i, y_j) \gamma_{\theta,i,j}
\]

where \( \Pi(\mu_1, \mu_2) = \{\gamma_\theta \in \mathbb{R}^{n \times n} \text{ s.t. it is constructed from the permutahedron of the proj. distributions}\} \)

- \( \Pi(\mu_1, \mu_2) \subset \Gamma(\mu_1, \mu_2) \)

- Gives a sample complexity similar to Sinkhorn \( n^{-1/2} \) measures lying on smaller dimensional subspaces has a better sample complexity than between the original measures
Experimental results

Computational aspects

- Two Gaussian distributions $\mu_1$ and $\mu_2$

\[ W_2^2 = 32.4 \]
\[ W_2^2 = 346.1 \]
\[ W_2^2 = 3836.0 \]

\[ d = 2 \quad d = 20 \quad d = 200 \]
Experimental results

Gradient flows

- Initial $\mu_1$: uniform distribution, different target distributions
Experimental results
Pan sharpening / image colorization, using the map

- One distribution is supported on a line

- Construct a super-resolution multi-chromatic satellite image from a high-resolution mono-chromatic image (source) and low-resolution multi-chromatic image (target)
Experimental results

Point cloud matchings, using the map

- Iterative Closest Point iterative algorithm for aligning point clouds
- Based on several one-to-one correspondences between points

<table>
<thead>
<tr>
<th>$n$</th>
<th>500</th>
<th>3000</th>
<th>150 000</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>3.54 (0.02)</td>
<td>96.9 (0.30)</td>
<td>23.3 (59.37)</td>
</tr>
<tr>
<td>OT</td>
<td>0.32 (0.18)</td>
<td>48.4 (58.46)</td>
<td>.</td>
</tr>
<tr>
<td>min-SWGG</td>
<td>0.05 (0.04)</td>
<td>37.6 (0.90)</td>
<td>6.7 (105.75)</td>
</tr>
</tbody>
</table>

(the lower the better, timings into parenthesis)
Experimental results

Optimal transport dataset distances

- For computing distances between datasets
- Cumbersome to compute in practice since it lays down on solving multiple OT problems

![Distance Matrix](image)

**Figure:** OTDD results ($\times 10^2$) distances for min-SWGG (left) and Sinkhorn divergence (right) for various datasets.
Conclusion

- Sliced Wasserstein Generalized Geodesic
  - provides an upper bound for Wasserstein
  - comes with an associated transport map
  - has a $O(Lnd + n \log(n))$ complexity
  - has good statistical properties

- Not the only approximation method based on a pivot measure
  - Factored coupling [4], where $\nu = \arg\min_{\mu \in \mathcal{P}(\mathbb{R}^k)} \{ \mathcal{W}_2^2(\mu, \mu_1) + \mathcal{W}_2^2(\mu, \mu_1) \}$

- Subspace detours [6], where $\nu = \arg\min_{\nu \in \mathcal{P}(\mathbb{R}^d)} \{ \mathcal{W}_2^2(P^E \# \mu_1, \nu) + \mathcal{W}_2^2(\nu, P^E \# \mu_2) \}$

- Some open questions
  - how do the Birkhoff polytope and the considered permutahedron relate?
  - concentration results?
  - extension to incomparable spaces through a pivot measure?
Fast Optimal Transport through Sliced Generalized Wasserstein Geodesics

Joint work with Guillaume Mahey, Gilles Gasso, Clément Bonet and Nicolas Courty
NeurIPS 2023 [5]

Laetitia Chapel
laetitia.chapel@irisa.fr

IRISA, Rennes, France
Institut Agro Rennes-Angers

Workshop on Optimal Transport: from theory to applications, Berlin 2024
Bibliography


