Efficient and Exact Multimarginal Optimal Transport with Pairwise Costs

Bohan Zhou, Dartmouth College
Joint work with
Matthew Parno, Dartmouth College.

PIMS-IFDS-NSF Summer School

June 29, 2022

ONR MURI #N00014-20-1-2595.

Multimarginal OT

$$\inf_{P} \int_{X} c(x_1, \cdots, x_m) \mathrm{d}P(x_1, \cdots, x_m), \tag{1}$$

satisfying all given marginals $(\mu_i)_{i=1}^{m}$.

Figure: SAR images on the Beaufort Sea from the Sentinel-1 satellite.

(a) 03-01-21 (b) 03-03-21 (c) 03-05-21 (d) 03-07-21 (e) 03-09-21

Wasserstein Barycenter

Find the barycenter μ^* of $(\mu_i)_{i=1}^m$:

$$\mu^* = \arg\min_{\mu} \sum_{i=1}^{m} \lambda_i W_2^2(\mu_i, \mu) \tag{2}$$

- double-minimization.
- It is equivalent to

$$\inf_{P \in \Gamma(\mu_1, \cdots, \mu_m)} \int \left(\sum_{i<j} \frac{\lambda_i \lambda_j}{2} |x_i - x_j|^2 \right) \mathrm{d}P, \tag{3}$$

and then $u^* = (T)_\# P^*$ where $T(x_1, \cdots, x_m) = \sum \lambda_i x_i$ is the Euclidean barycenter map.

Compute MMOT

Original MMOT:

$$\inf_P \int_X c(x_1, \cdots, x_m) \mathrm{d}P(x_1, \cdots, x_m), \tag{4}$$

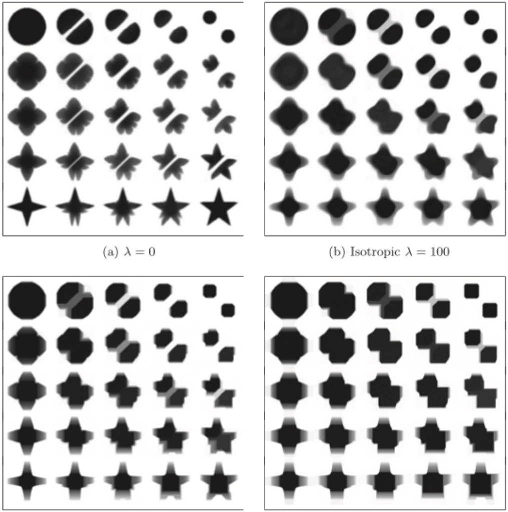
Regularized MMOT:

$$\inf_P \int_X c \mathrm{d}P + \varepsilon \int \log(P) \mathrm{d}P = \inf_P \mathrm{KL}(P \mid K), \tag{5}$$

Blurred Wasserstein barycenter

Convolutional Wasserstein Barycenters in POT

Naive deblurring by total variation



(a) $\lambda = 0$ (b) Isotropic $\lambda = 100$

(c) Anisotropic $\lambda = 500$ (d) Anisotropic $\lambda = 2000$

From [CP18]

Sharp Wasserstein interpolation

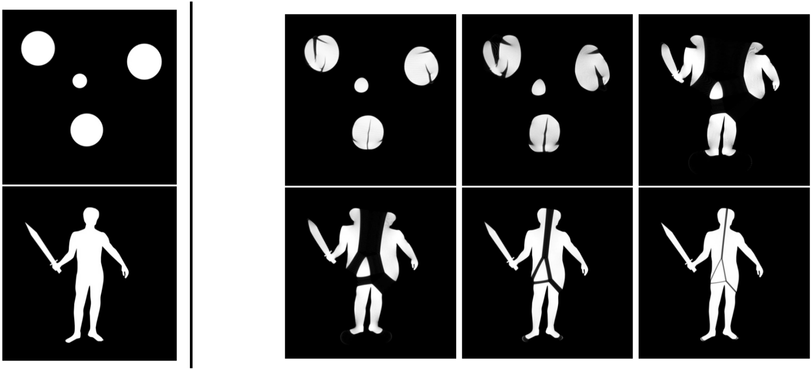


Figure: Jacobs and Léger [JL20]

Dual formulations

$$\sup_{f_1(x_1)+f_2(x_2)\leqslant c(x_1,x_2)} \int_{X_1} f_1(x_1)\mathrm{d}\mu_1 + \int_{X_2} f_2(x_2)\mathrm{d}\mu_2;$$

$$\sup_{f_1 \text{ is } c\text{-concave}} \int_{X_1} f_1(x_1)\mathrm{d}\mu_1 + \int_{X_2} f_1^c(x_2)\mathrm{d}\mu_2;$$

$$\sup_{f_2 \text{ is } c\text{-concave}} \int_{X_1} f_2^c(x_1)\mathrm{d}\mu_1 + \int_{X_2} f_2(x_2)\mathrm{d}\mu_2,$$

Recall for $c(x_1, x_2) = \frac{1}{2}|x_1 - x_2|^2$:

$$(x_1 - \nabla f_1(x_1))_{\#}\mu_1 = \mu_2 \qquad \text{and} \qquad (x_2 - \nabla f_1^c(x_2))_{\#}\mu_2 = \mu_1.$$

Two ingredients

For the functional:

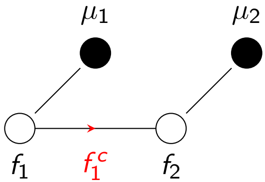
$$I(f) = \int f \mathrm{d}\mu_1 + \int f^c \mathrm{d}\mu_2,$$

We want to do gradient ascent in the class $\{f \mid f \text{ is } c\text{-concave}\}$:

$$f^{(k+1)} = f^{(k)} + \sigma \nabla I(f^{(k)}).$$

- $\dot{H}^1$ gradient: $\nabla_{\dot{H}^1} I(f) = (-\Delta)^{-1}(\mu_1 - (S_{f^c})_{\#}\mu_2)$ for $S_f(x) = x - \nabla f(x)$.
- Back-and-forth between two spaces for f_1 and f_2.

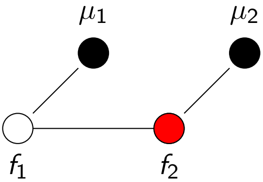
BF pipeline



1. Initial guess on f_1, get $f_2 \leftarrow f_1^c$.

$$
\begin{aligned}
f_2^{(n)} &= (f_1^{(n)})^c \\
f_2^{(n+\frac{1}{2})} &= f_2^{(n)} + \sigma \nabla_{\dot{H}^1} I_2(f_2^{(n)}) \\
f_1^{(n+\frac{1}{2})} &= (f_2^{(n+\frac{1}{2})})^c \\
f_1^{(n+1)} &= f_1^{(n+\frac{1}{2})} + \sigma \nabla_{\dot{H}^1} I_1(f_1^{(n+\frac{1}{2})})
\end{aligned}
$$

BF pipeline

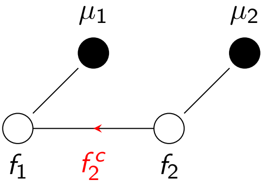


$$
\begin{aligned}
f_2^{(n)} &= (f_1^{(n)})^c \\
f_2^{(n+\frac{1}{2})} &= f_2^{(n)} + \sigma \nabla_{\dot{H}^1} I_2(f_2^{(n)}) \\
f_1^{(n+\frac{1}{2})} &= (f_2^{(n+\frac{1}{2})})^c \\
f_1^{(n+1)} &= f_1^{(n+\frac{1}{2})} + \sigma \nabla_{\dot{H}^1} I_1(f_1^{(n+\frac{1}{2})})
\end{aligned}
$$

check if $\mu_2 = (S_{f_2^c})_{\#}\mu_1$

1. Initial guess on f_1, get $f_2 \leftarrow f_1^c$.
2. Gradient ascent on f_2.

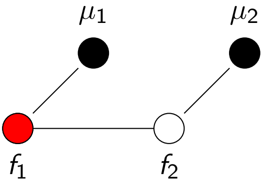
BF pipeline



$$f_2^{(n)} = (f_1^{(n)})^c$$
$$f_2^{(n+\frac{1}{2})} = f_2^{(n)} + \sigma \nabla_{\dot{H}^1} I_2(f_2^{(n)})$$
$$f_1^{(n+\frac{1}{2})} = (f_2^{(n+\frac{1}{2})})^c$$
$$f_1^{(n+1)} = f_1^{(n+\frac{1}{2})} + \sigma \nabla_{\dot{H}^1} I_1(f_1^{(n+\frac{1}{2})})$$

1. Initial guess on f_1, get $f_2 \leftarrow f_1^c$.
2. Gradient ascent on f_2.
3. Bounce back to $f_1 \leftarrow f_2^c$.

BF pipeline



$$
\begin{aligned}
f_2^{(n)} &= (f_1^{(n)})^c \\
f_2^{(n+\frac{1}{2})} &= f_2^{(n)} + \sigma \nabla_{\dot{H}^1} I_2(f_2^{(n)}) \\
f_1^{(n+\frac{1}{2})} &= (f_2^{(n+\frac{1}{2})})^c \\
f_1^{(n+1)} &= f_1^{(n+\frac{1}{2})} + \sigma \nabla_{\dot{H}^1} I_1(f_1^{(n+\frac{1}{2})})
\end{aligned}
$$

check if $\mu_1 = (S_{f_1^c})_{\#}\mu_2$

1. Initial guess on f_1, get $f_2 \leftarrow f_1^c$.
2. Gradient ascent on f_2.
3. Bounce back to $f_1 \leftarrow f_2^c$.
4. Gradient ascent on f_1.

MMOT duality

Dual form:

$$\inf_{P\in\Gamma(\mu_1,\cdots,\mu_m)}\int_X c(x_1,\cdots,x_m)\mathrm{d}P(x_1,\cdots,x_m); \tag{6}$$

$$= \sup_{f_1+\cdots+f_m\leqslant c}\sum_{i=1}^m\int f_i\mathrm{d}\mu_i \tag{7}$$

Conjugate tuples:

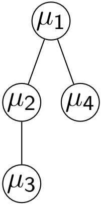
$$\widetilde{f}_i(x) = \inf\left\{c(y_1,\cdots,x,y_{i+1},\cdots,y_N) - \sum_{j\neq i}\widetilde{f}_j(y_j)\right\}, \tag{8}$$

Problems of BFM

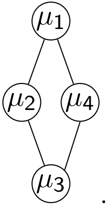
- General c-transform in high dimension is hard;
- No longer bounce between two spaces.
- The dual variables (f_i) are not the right one to define pushforward maps between marginals [ZP22].
- The close form of $\partial_i I(f_i; f_1, \cdots, f_m)$ is hard to find.

Pairwise cost

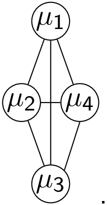
For $c = \sum_{i<j} c_{ij}(x_i, x_j)$ and $c_{ij} = h_{ij}(x_i - x_j)$ for strictly convex function h_{ij}.



(a) Four-marginal problem with $c = c_{12} + c_{23} + c_{14}$.



(b) Four-marginal problem with $c = c_{12} + c_{14} + c_{23} + c_{34}$.



(c) Four-marginal problem with $c = \sum_{i<j} c_{ij}$.

Why pairwise cost?

- High dimension c-transform into 1D c-transform:

$$
\begin{aligned}
&(f_2(x_2)+f_3(x_3))^c(x_1)\\
=&\inf_{x_2,x_3} c(x_1,x_2,x_3)-f_2(x_2)-f_3(x_3)\\
=&\inf_{x_2}\left(c_{12}(x_1,x_2)-f_2(x_2)+\inf_{x_3} c_{23}(x_2,x_3)-f_3(x_3)\right)
\end{aligned}
$$

- A natural graph structure to update nodes.

Applications

- Wasserstein barycenter:

$$\sum_{1\leqslant i<j\leqslant m} \lambda_i \lambda_j \left|x_i - x_j\right|^2 \qquad \text{(Gangbo-Święch)}$$

- Brenier's discrete least action:

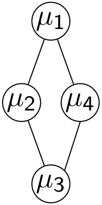
$$\sum_{i=1}^{m-1} \lambda_i \left|x_i - x_{i+1}\right|^2 \qquad \text{(Brenier)}$$

- Density Functional Theory

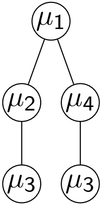
$$\sum_{1\leqslant i<j\leqslant m} \frac{1}{\left|x_i - x_j\right|} \qquad \text{(Coulomb interaction)}$$

Decycling

By duplicating marginals:

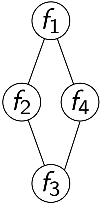


becomes

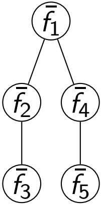


.

It turns into a new MMOT in terms of dual variables:



becomes



.

Equivalence Theorem

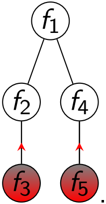
Theorem (Z. and Parno, Preprint)

Given a cost $c(x_1, \cdots, x_m)$ that corresponds to an undirected graph $G = (V, E)$ with possible cycles ($m = |V|$), there exists a cost $\bar{c}(x_1, \cdots, x_n)$ that corresponds to an undirected tree $\bar{G} = (\bar{V}, \bar{E})$ with $n = \left|\bar{V}\right| = |E| + 1$.
The MMOT of $\bar{c}$ is equal to MMOT of c, and the optimal dual solutions $(\bar{f}_i)_{i=1}^n$ can recover the optimal dual solution $(f_i)_{i=1}^m$.

MMOT on a rooted tree: Pipeline

1. Generate (out)-fluxes:

$$f_3' = (f_3)^{c_{23}};$$
$$f_3 \leftarrow f_3 - \sigma(\Delta)^{-1}(\mu_3 - (S_{f_3'})_{\#}\mu_2).$$



MMOT on a rooted tree: Pipeline

1. Generate (out)-fluxes:

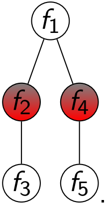
$$f_3' = (f_3)^{c_{23}};$$

$$f_3 \leftarrow f_3 - \sigma(\Delta)^{-1}(\mu_3 - (S_{f_3'})_{\#}\mu_2).$$

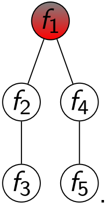
2. Generate net (out)-fluxes:

$$f_2' = (f_2 - \sum_{\text{all in-fluxes}} f_j')^{c_{12}};$$

$$f_2 \leftarrow f_2 - \sigma(\Delta)^{-1}(\mu_2 - (S_{f_2'})_{\#}\mu_1).$$



MMOT on a rooted tree: Pipeline



1. Generate (out)-fluxes:

$$f_3' = (f_3)^{c_{23}};$$
$$f_3 \leftarrow f_3 - \sigma(\Delta)^{-1}(\mu_3 - (S_{f_3'})_{\#}\mu_2).$$

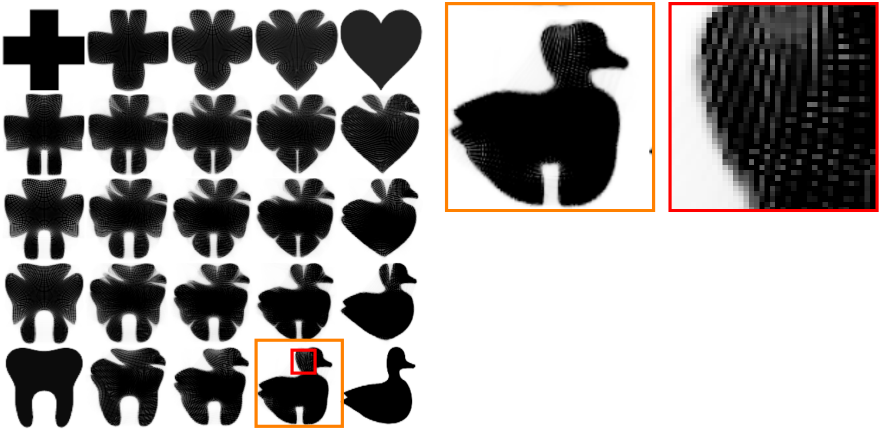
2. Generate net (out)-fluxes:

$$f_2' = (f_2 - \sum_{\text{all in-fluxes}} f_j')^{c_{12}};$$
$$f_2 \leftarrow f_2 - \sigma(\Delta)^{-1}(\mu_2 - (S_{f_2'})_{\#}\mu_1).$$

3. Get the root node by c-transform:

$$f_1 \leftarrow \sum_{\text{all in-flues}} f_j'.$$

Sharp interpolation



(a) (b)

Pushforward of measures

$$f(x) = x \qquad g(x) = x + 10^{-3}\sin(100x)$$

$$(f)_{\#}\mathscr{L} = p_u\mathscr{L} \qquad (g)_{\#}\mathscr{L} = p_v\mathscr{L}$$

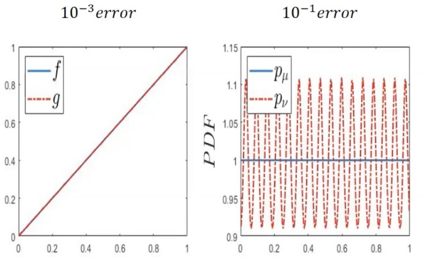
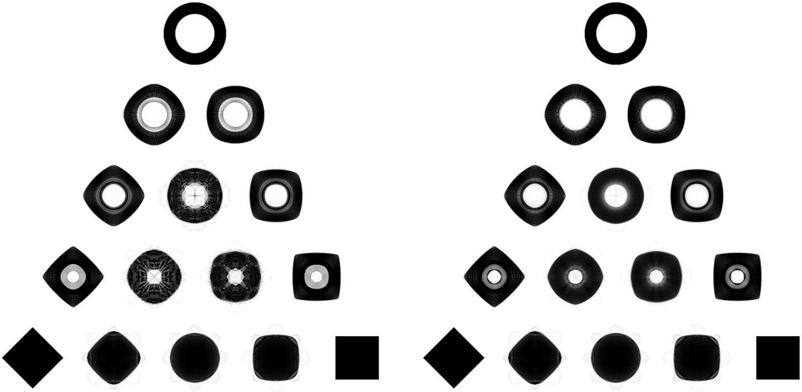


Figure: Example by Ditkowski-Fibich-Sagiv [DFS20]

Healing by the inverse of Laplacian

$$\nabla_{\dot{H}^1} I_1(f_1) = (-\Delta)^{-1}\left(\mu_1 - (S_{f_1^c})_{\#}\mu_2\right)$$



(a) 10-step gradient ascent

(b) 500-step gradient ascent

Take-aways

- An efficient and exact method to MMOT with pairwise costs.
- Python package will be available after submission.
- Possible improvement from computing pushforward of measures and picking stronger norm.
- Lack of convergence analysis except under strong assumptions.

[CP18] Marco Cuturi and Gabriel Peyré. Semidual regularized optimal transport. *SIAM Rev.*, 60(4):941–965, 2018. Revised reprint of "A smoothed dual approach for variational Wasserstein problems" [MR3466197].

[DFS20] Adi Ditkowski, Gadi Fibich, and Amir Sagiv. Density estimation in uncertainty propagation problems using a surrogate model. *SIAM/ASA Journal on Uncertainty Quantification*, 8(1):261–300, 2020.

[JL20] Matt Jacobs and Flavien Léger. A fast approach to optimal transport: the back-and-forth method. *Numer. Math.*, 146(3):513–544, 2020.

[ZP22] Bohan Zhou and Matthew Parno. Efficient and exact multimarginal optimal transport with pairwise costs. *Preprint on request*, 2022.