Efficient and Exact Multi-Marginal Optimal Transport with Pairwise Costs

Bohan Zhou, [1] Matthew Parno. [1]

[1]Dartmouth College

DARTMOUTH

Abstract

We provide an exact and efficient method to solve Multimarginal Optimal Transport (MMOT) under a family of cost functions:

$$\inf_{P\in\Gamma(\mu_1,\cdots,\mu_m)} \int c(x_1,\cdots,x_m)\mathrm{d}P(x_1,\cdots,x_m), \tag{1}$$

for the space $\boldsymbol{X} = X_1 \times \cdots \times X_m$ and prescribed marginal probability measures $(\mu_i)_{i=1}^m$. The set of transport plans $\Gamma(\mu_1,\cdots,\mu_m)$ is defined by

$$\Gamma(\mu_1,\cdots,\mu_m) \overset{\text{def}}{=} \{P \in \mathbb{P}(\boldsymbol{X}) \mid (\pi_i)_\# P = \mu_i, 1 \leqslant i \leqslant m\}.$$

We assume the cost function $c(x_1,\cdots,x_m)$ satisfies:

- The cost function is a summed pairwise cost functions

$$c(x_1,\cdots,x_m) = \sum_{1\leqslant i<j\leqslant m} c_{ij}(x_i,x_j);$$

- $c_{ij}(x_i,x_j) = h_{ij}(x_i - x_j)$ for some strictly convex function h_{ij}.

Preliminary

Background

1. c-transform and Duality theory:
   The c-transform of a function $f : X_1 \mapsto \mathbb{R}$ is given by

$$f^c(x_2) = \inf_{x_1} c(x_1,x_2) - f(x_1).$$

It it natural to have $f(x_1) + f^c(x_2) \leqslant c(x_1,x_2)$. The c-transform is a generalization of the Legendre transform $f^*(y) = \sup_x x\cdot y - f(x)$.
We say (f_1,f_2) are c-conjugate if $f_1 = f_2^c$ and $f_2 = f_1^c$.

The dual problem corresponding to (1) is given by

$$\sup_{(f_1,\cdots,f_m)} \sum_{i=1}^m \int_{X_i} f_i(x_i)\mathrm{d}\mu_i, \tag{2}$$

where $f_i \in L^1(\mu_i)$ and $\sum_{i=1}^m f_i(x_i) \leqslant c(x_1,\cdots,x_m)$.
[Kel84] provided a general duality theorem: there exists a c-conjugate solution to (2). We have the relationship between the primal solution and dual solution:

$$\sum_{i=1}^m f_i(x_i) = c(x_1,\cdots,x_m) \qquad P\text{−a.e.}. \tag{3}$$

2. Gradient in Hilbert space and back-and-forth method by [JL20] to solve 2-marginal OT under cost $c(x_1,x_2) = h(x_1 - x_2)$ for some strictly convex function h:
   - For functional $I(f) = \int f\mathrm{d}\mu_1 + \int f^c\mathrm{d}\mu_2$, first find the first variance δI by a perturbation lemma [GM96];
   - Define the gradient in a Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$:

$$\langle \nabla_{\mathcal{H}} I(f), g\rangle = \delta I(g;f).$$

   - [JL20] picked the space $\dot{H}^1$ with the inner product $\langle f_1,f_2\rangle_{\dot{H}^1} = \int \nabla f_1 \cdot \nabla f_2 \mathrm{d}x$ for the dual variables. The gradient is given by

$$I(f) = (-\Delta)^{-1}(\mu_1 - (S_{f^c})_\# \mu_2), \tag{4}$$

   where the Brenier map $S_f(x_1) \overset{\text{def}}{=} x_1 - \nabla h^*(\nabla f(x_1))$.
   - To solve (2) for $m = 2$, [JL20] used a gradient-ascent scheme to update two functionals of type $I(f)$, depending on each dual variables $(f_i)_{i=1}^2$, in a back-and-forth fashion.

Current Computational Methods

Here, we list several MMOT solvers to our best knowledge. In general, entropy-regularized based methods may suffer from numerical instability and blurring issues. LP based methods may not be practical in solving large-scale problems.

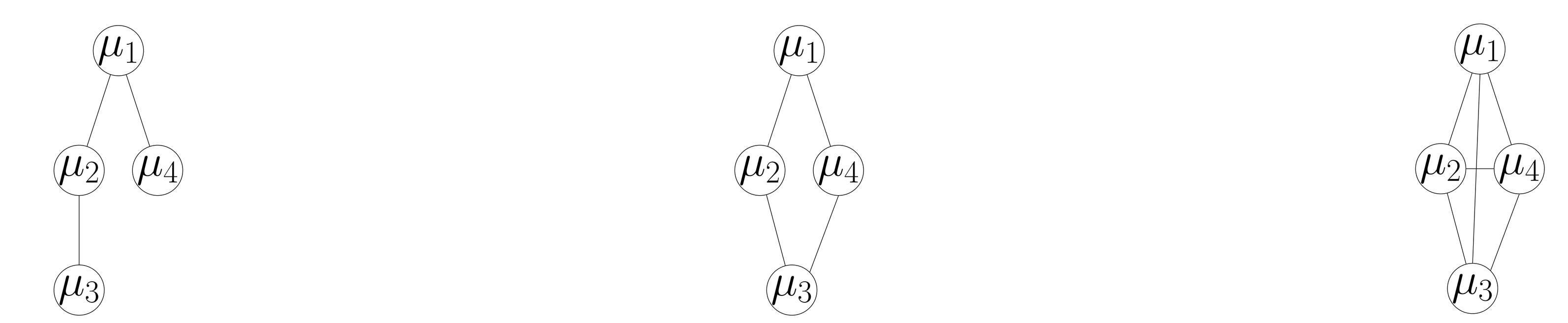
- [BCC+15]: Entropy-regularized MMOT on primal variables.
- [HRCK21]: Entropy-regularized MMOT with structure on dual variables.
- [ABA22]: Solving exact MMOT with structure via ellipsoid algorithm with oracle.
- [NX22]: LP-based method to approximate MMOT with controllable level of sub-optimality.

Main results

Our Strategy

1. **Graphical Representation of MMOT** Given a summed pairwise cost function, we can represent the relationship between their marginals on a graph, each node stores a marginal μ_i with its dual variable f_i, each edge stores pairwise cost c_{ij}.
2. **Unrolling MMOT into a tree representation** We prove an equivalent theorem that any MMOT that has a graphical representation with possible cycle is equivalent to another MMOT of a tree representation. The proof is via duplicating nodes and generalized gluing lemma. We also show that the cost of duplicating is limited by the number of edges in original graph.
3. **Solving MMOT on the rooted tree via gradient-ascent** By leveraging c-transform to get rid of the constraint, we will use gradient ascent on the remaining $(m-1)$ dual variables in the space $\dot{H}^1$. The key in this step is that we introduce a "net potential", which help us to define the gradient and to compute the c-transform.

Step 1: Graphical Representation



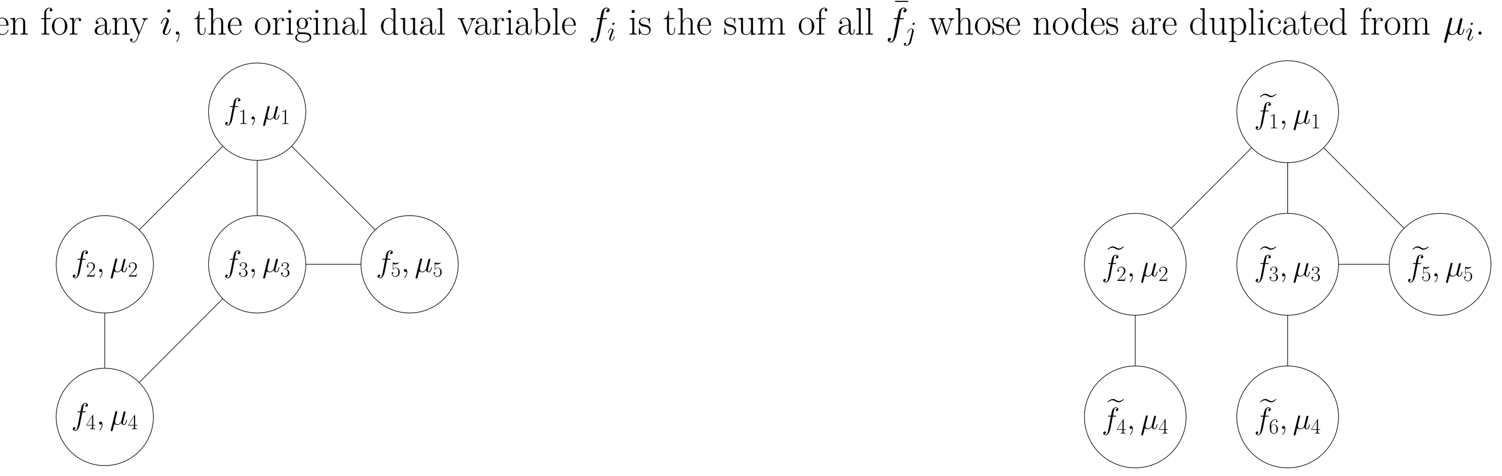
(a) $c = c_{12} + c_{23} + c_{14}$. (b) $c = c_{12} + c_{14} + c_{23} + c_{34}$. (c) $c = \sum_{i<j} c_{ij}$. Barycenter costs.

Step 2: Tree Representation [Equivalence theorem]

Given a cost function $c(x_1,\cdots,x_m)$ that corresponds to $G = (V,E)$ with possible cycles, there exists a cost function $\bar{c}(x_1,\cdots,x_n)$ that corresponds to tree $\bar{G} = (\bar{V},\bar{E})$ with $n = |\bar{V}| = |E| + 1$ nodes, such that

$$\inf_{P^{(m)}\in\Gamma(\mu_1,\cdots,\mu_m)} \int c(x_1,\cdots,x_m)\mathrm{d}P^{(m)} = \inf_{P^{(n)}\in\Gamma(\mu_1,\cdots,\mu_n)} \int \bar{c}(x_1,\cdots,x_n)\mathrm{d}P^{(n)},$$

where $(\mu_k)_{k=m+1}^n$ are duplicated from $(\mu_i)_{i=1}^m$ in the "unrolling" process.
Furthermore, let $P^{(m)}$ and $(f_i)_{i=1}^m$ be the optimal primal and dual solutions to the original MMOT. And $P^{(n)}$ and $(\bar{f}_i)_{i=1}^n$ be the optimal primal and dual solutions to the new MMOT.
Then for any i, the original dual variable f_i is the sum of all $\bar{f}_j$ whose nodes are duplicated from μ_i.



Step 3: Gradient-ascent on Rooted Tree

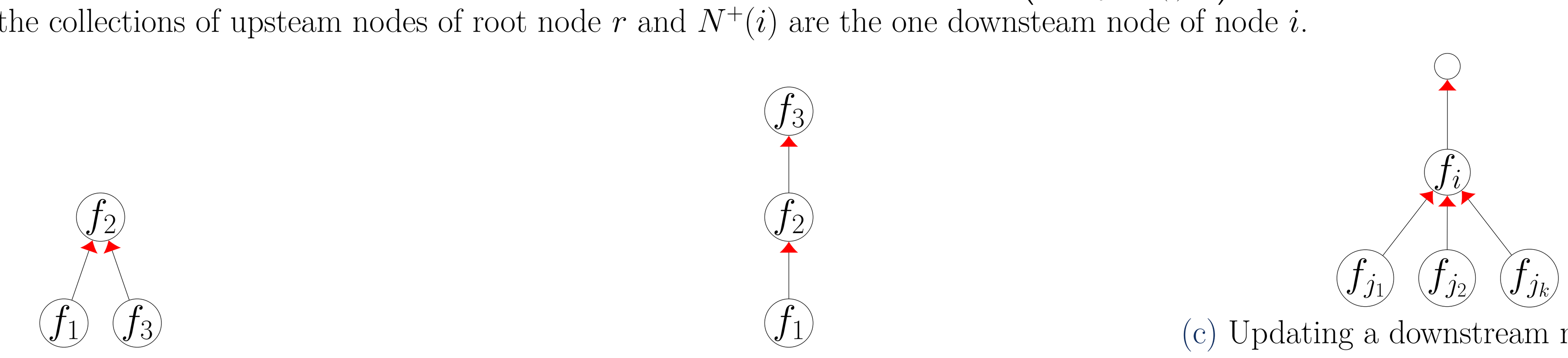
Define $I_r(f_1,\ldots,f_{r-1},f_{r+1},\ldots,f_m) \overset{\text{def}}{=} I(f_1,\ldots,f_{r-1},(\sum_{i\neq r} f_i)^c, f_{r+1},\ldots,f_m)$. The updates are:

$$\begin{cases} f_i \leftarrow f_i - \sigma\nabla_{\dot{H}^1} I_r(f_i); \\ f_r \leftarrow (\sum_{i\neq r} f_i)^c. \end{cases} \tag{5}$$

$$\begin{cases} \nabla_{\dot{H}^1} I_r(f_i) = (-\Delta)^{-1}\left(\mu_i - (S_{f_i'})_\# \mu_{N^+(i)}\right) & \text{(6a)} \\ f_r(x_r) = \sum_{i\in N^-(r)} f_i'(x_r). & \text{(6b)} \end{cases}$$

where the net potential f_i' at edge $(i, N^+(i))$ we introduced, are recursively defined by $f_i' = \left(f_i - \sum_{j\in N^-(i)} f_j'\right)^{c_{iN^+(i)}}$.

Here $N^-(r)$ are the collections of upsteam nodes of root node r and $N^+(i)$ are the one downsteam node of node i.



(a) A directed tree $c = c_{12} + c_{23}$ (b) Another directed tree for $c = c_{12} + c_{23}$ (c) Updating a downstream node of leaf nodes

Numerical Results

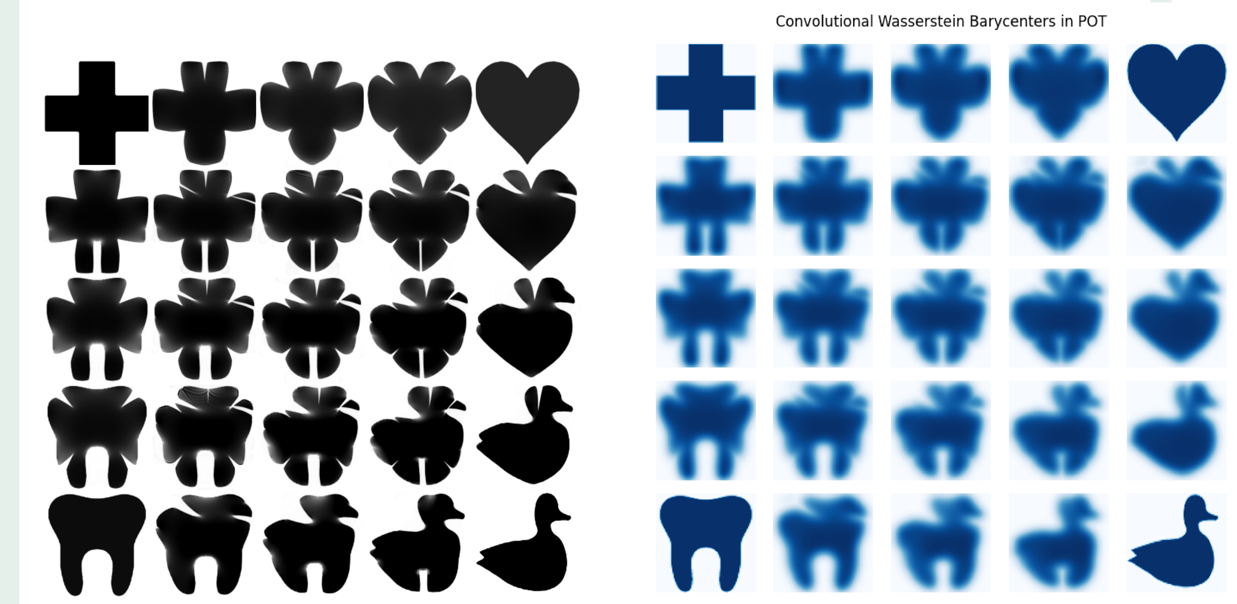
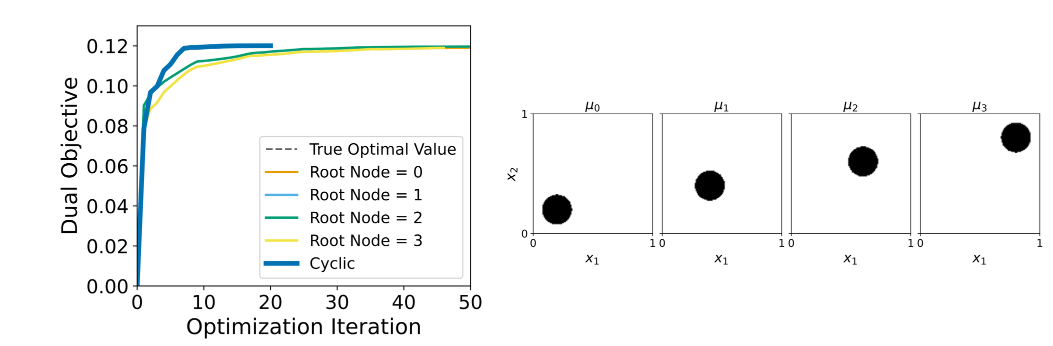
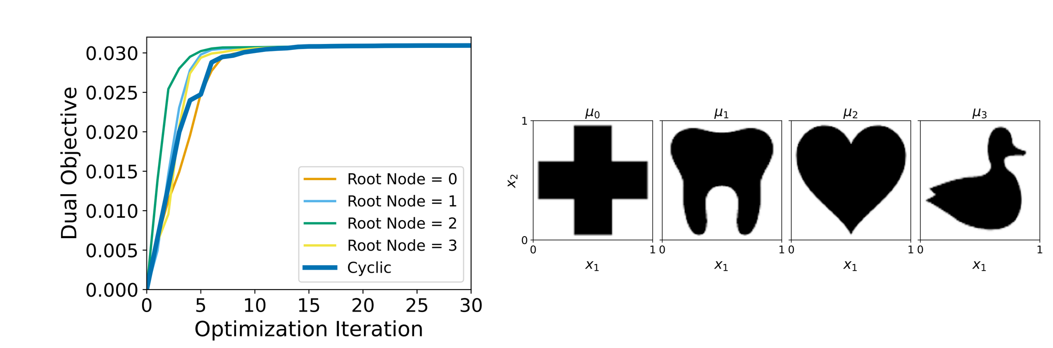


Figure: Left plot: sharp Wasserstein barycenter via our method. Right plot: blurred Wasserstein barycenter via entropy-regularized based method in POT package, regularization parameter is 0.004. Both 4-marginals are given at four corners.



(a) Impact of cycling the root node with pure translation.



(b) Impact of cycling the root node with shape deformation.

Figure: Performance between fixed root node and cyclic root node. The cyclic root node is a variant of the main algorithm which forces c-conjugate condition.

References

[ABA21] J.M. Altschuler and E. Boix-Adserà. Polynomial-time algorithms for multimarginal optimal transport problems with structure, *Math. Program.*, 2022.
[BCC+15] JD. Benamou, G. Carlier, M. Cuturi, L. Nenna and G. Peyré. Iterative Bregman projections for regularized transportation problems, *SIAM J. Sci. Comput.*, 2015.
[HRCK21] I. Haasler, R. Singh, Q. Zhang, J. Karlsson and Y. Chen. Multi-marginal optimal transport and probabilistic graphical models, *IEEE Trans. Inform. Theory*, 2021. [GM96] W. Gangbo and R.J. McCann. The geometry of optimal transportation, *Acta Math.*, 1996.
[JL20] M. Jacobs and F. Léger. A fast approach to optimal transport: the back-and-forth method, *Numerische Mathematik*, 2020.
[Kel84] H. G. Kellerer. Duality theorems for marginal problems, *Z. Wahrsch. Verw. Gebiete*, 1984.
[NX22] A. Neufeld and Q. Xiang. Numerical method for feasible and approximately optimal solutions of multi-marginal optimal transport beyond discrete measures, *ArXiv*, 2022.
[PZgithub] M. Parno and —-. MMOT2d,`https://github.com/simda-muri/mmot`
[ZP22] —- and M. Parno. Exact and Efficient Multi-marginal Optimal Transport with Pairwise Cost. *ArXiv*, 2022.