Numerical method for feasible and approximately optimal solutions of multi-marginal optimal transport beyond discrete measures

Qikun Xiang

Nanyang Technological University, Singapore

April 15, 2022

Joint work with:

Ariel Neufeld,
Nanyang Technological University, Singapore

Financial support from the SIAM Student Travel Award and from
the Division of Mathematical Sciences, NTU are gratefully acknowledged.

Multi-marginal optimal transport (MMOT)

- **Given:**
  - N probability measures $\mu_1, \ldots, \mu_N$ on Polish spaces $(\mathcal{X}_1, d_{\mathcal{X}_1}), \ldots, (\mathcal{X}_N, d_{\mathcal{X}_N})$;
  - $\boldsymbol{\mathcal{X}} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$, $d_{\boldsymbol{\mathcal{X}}} := \bigoplus_{i=1}^N d_{\mathcal{X}_i}$;
  - cost function $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \to \mathbb{R}$.
- **Goal:** minimization over the set of couplings of $\mu_1, \ldots, \mu_N$:

$$\inf_{\mu \in \Gamma(\mu_1, \ldots, \mu_N)} \int_{\boldsymbol{\mathcal{X}}} f(x_1, \ldots, x_N)\, \mu(\mathrm{d}x_1, \ldots, \mathrm{d}x_N), \tag{MMOT}$$

where

$$\Gamma(\mu_1, \ldots, \mu_N) := \big\{\mu \in \mathcal{P}(\boldsymbol{\mathcal{X}}) : \text{the marginal of } \mu \text{ on } \mathcal{X}_i \text{ is } \mu_i \quad \forall 1 \leq i \leq N\big\}.$$

- **Dual problem** (MMOT*):

$$\sup \left\{ \sum_{i=1}^N \int_{\mathcal{X}_i} g_i \, \mathrm{d}\mu_i : \sum_{i=1}^N g_i(x_i) \leq f(x_1, \ldots, x_N) \quad \forall (x_1, \ldots, x_N) \in \boldsymbol{\mathcal{X}} \right\}.$$

Applications of MMOT

- **Multi-population matching** (a.k.a. matching for teams) in theoretical economics
- Wasserstein barycenter (special case of multi-population matching)
- Density functional theory in quantum mechanics
- Mathematical finance
- **Distributionally robust optimization** (DRO) in operations research

Existing approaches: discretization & regularization

- **Assuming discrete marginals**: Benamou, Carlier, Cuturi, Nenna, and Peyré [2015], Puccetti, Rüschendorf, and Vanduffel [2020], Tupitsa, Dvurechensky, Gasnikov, and Uribe [2020], . . .
- **Replacing non-discrete marginals by discrete approximations**: Carlier, Oberman, and Oudet [2015], Eckstein, Guo, Lim, and Obłój [2021], . . .
- **Entropic regularization and the Sinkhorn algorithm**: Cuturi [2013], Benamou, Carlier, and Nenna [2019], Peyré and Cuturi [2019], . . .
- **Regularization + neural networks**: Eckstein and Kupper [2019], Henry-Labordère [2019], Eckstein, Kupper, and Pohl [2020], Cohen, Arbel, and Deisenroth [2020], De Gennaro Aquino and Eckstein [2020], . . .
- Do not produce feasible solutions of the primal nor the dual of MMOT.

Contributions

- We develop a numerical algorithm for approximately solving the MMOT problem.
- The algorithm produces a feasible solution of the dual MMOT through relaxation.

  $\Rightarrow$ **lower bound**
- The algorithm produces a feasible solution of the primal MMOT.

  $\Rightarrow$ **upper bound**
- For any $\tilde{\epsilon} > 0$, we are able to control upper bound $-$ lower bound $\leq \tilde{\epsilon}$.

  $\Rightarrow$ **$\tilde{\epsilon}$-optimal primal and dual solutions**

Relaxation

- Relax the marginal constraints into finitely many linear constraints.
  - Fixing a marginal can be seen as having infinitely many linear constraints.
- Couplings are replaced by a *moment set*:

Definition (Moment set)

For $i = 1, \ldots, N$, $[\mu_i]_{\mathcal{G}_i}$ is called a moment set centered at μ_i characterized by functions $\mathcal{G}_i \subset \mathcal{L}^1(\mathcal{X}_i, \mu_i)$: $\quad \nu_i \in [\mu_i]_{\mathcal{G}_i} \quad \Leftrightarrow \quad \int_{\mathcal{X}_i} g_i \,\mathrm{d}\mu_i = \int_{\mathcal{X}_i} g_i \,\mathrm{d}\nu_i \; \forall g_i \in \mathcal{G}_i$.
Moreover,

$$\Gamma([\mu_1]_{\mathcal{G}_1}, \ldots, [\mu_N]_{\mathcal{G}_N}) := \Big\{\mu \in \Gamma(\nu_1, \ldots, \nu_N) : \nu_i \in [\mu_i]_{\mathcal{G}_i} \; \forall 1 \leq i \leq N\Big\}.$$

- **Relaxed MMOT problem**:

$$\inf_{\mu \in \Gamma([\mu_1]_{\mathcal{G}_1}, \ldots, [\mu_N]_{\mathcal{G}_N})} \int_{\boldsymbol{\mathcal{X}}} f(x_1, \ldots, x_N) \, \mu(\mathrm{d}x_1, \ldots, \mathrm{d}x_N). \qquad (\mathrm{MMOT}_{\mathrm{relax}})$$

Constructing a feasible solution of the primal MMOT

Definition (Reassembly)

Let $\bar{\mathcal{X}}_i := \mathcal{X}_i$ for $i = 1, \ldots, N$. $\tilde{\mu}$ is called a reassembly of $\hat{\mu} \in \mathcal{P}(\mathcal{X}_1 \times \cdots \times \mathcal{X}_N)$ with the marginals $\mu_1, \ldots, \mu_N$ if there exists a probability measure $\gamma \in \mathcal{P}(\mathcal{X}_1 \times \cdots \times \mathcal{X}_N \times \bar{\mathcal{X}}_1 \times \cdots \times \bar{\mathcal{X}}_N)$ such that:

1. the marginal of γ on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_N$ is $\hat{\mu}$;
2. for $i = 1, \ldots, N$, the marginal $\gamma_i \in \Gamma(\hat{\mu}_i, \mu_i)$ of γ on $\mathcal{X}_i \times \bar{\mathcal{X}}_i$ satisfies $\int_{\mathcal{X}_i \times \bar{\mathcal{X}}_i} d_{\mathcal{X}_i}(x, y)\, \gamma_i(\mathrm{d}x, \mathrm{d}y) = W_1(\hat{\mu}_i, \mu_i)$;
3. the marginal of γ on $\bar{\mathcal{X}}_1 \times \cdots \times \bar{\mathcal{X}}_N$ is $\tilde{\mu}$.

The set of *reassemblies* is denoted by $R(\hat{\mu}; \mu_1, \ldots, \mu_N) \subset \Gamma(\mu_1, \ldots, \mu_N)$.

- Idea: morphing $\hat{\mu}$ in an "optimal" way to turn its marginals into $\mu_1, \ldots, \mu_N$.
- Allows us to construct a feasible solution of (MMOT) via an infeasible one.

Controlling the approximation error

Theorem (Approximation of MMOT)

Suppose that:

1. $f : \mathcal{X} \to \mathbb{R}$ *is* L_f*-Lipschitz continuous for* $L_f > 0$*;*
2. *for* $i = 1, \ldots, N$*,* $\mathcal{G}_i \subset \mathcal{L}^1(\mathcal{X}_i, \mu_i)$*;*
3. $\hat{\mu}$ *is an* ϵ*-optimizer of* $(\mathrm{MMOT}_{\mathrm{relax}})$ *for* $\epsilon > 0$*, i.e.,*

$$\int_{\mathcal{X}} f \, \mathrm{d}\hat{\mu} \leq \inf_{\mu \in \Gamma([\mu_1]_{\mathcal{G}_1}, \ldots, [\mu_N]_{\mathcal{G}_N})} \left\{ \int_{\mathcal{X}} f \, \mathrm{d}\mu \right\} + \epsilon;$$

4. $\tilde{\epsilon} := \epsilon + L_f \left(\sum_{i=1}^{N} \sup_{\nu_i \in [\mu_i]_{\mathcal{G}_i}} \left\{ W_1(\mu_i, \nu_i) \right\} \right)$.

Then, every $\tilde{\mu} \in R(\hat{\mu}; \mu_1, \ldots, \mu_N)$ *is an* $\tilde{\epsilon}$*-optimizer of* (MMOT)*, i.e.,*

$$\int_{\mathcal{X}} f \, \mathrm{d}\tilde{\mu} \leq \inf_{\mu \in \Gamma(\mu_1, \ldots, \mu_N)} \left\{ \int_{\mathcal{X}} f \, \mathrm{d}\mu \right\} + \tilde{\epsilon}.$$

Important questions

- **Question 1**: for any $\epsilon > 0$, can we obtain an ϵ-optimizer $\hat{\mu}$ of $(\text{MMOT}_{\text{relax}})$?
  - **Answer 1**: yes, through proving the strong duality between $(\text{MMOT}_{\text{relax}})$ and its dual problem, which is a linear semi-infinite programming (LSIP) problem. Subsequently, a suitable cutting-plane algorithm (e.g., Conceptual Algorithm 11.4.1 of Goberna and López [1998]) can be used to obtain an ϵ-optimizer $\hat{\mu}$ of $(\text{MMOT}_{\text{relax}})$.
- **Question 2**: can we explicitly construct a $\tilde{\mu} \in R(\hat{\mu}; \mu_1, \ldots, \mu_N)$?
  - **Answer 2**: yes, we can explicitly construct a reassembly using the copula theory and Laguerre diagrams.
- **Question 3**: can we control $\sup_{\nu_i \in [\mu_i]_{\mathcal{G}_i}} \left\{ W_1(\mu_i, \nu_i) \right\}$ to be arbitrarily close to 0 for $i = 1, \ldots, N$?
  - **Answer 3**: yes, for any $\epsilon_i > 0$ we can explicitly construct a finite collection $\mathcal{G}_i \subset \mathcal{L}^1(\mathcal{X}_i, \mu_i)$ such that $\sup_{\nu_i \in [\mu_i]_{\mathcal{G}_i}} \left\{ W_1(\mu_i, \nu_i) \right\} \le \epsilon_i$.

Numerical algorithm

- Putting these pieces together, we develop a numerical algorithm, whose properties are summarized as follows.

Theorem (Properties of the proposed algorithm)

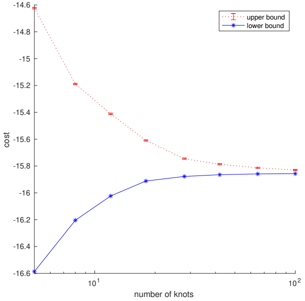
Under suitable conditions, for any $\tilde{\epsilon} > 0$, the proposed algorithm produces the outputs: $\tilde{\mu} \in \mathcal{P}(\boldsymbol{\mathcal{X}})$, $\hat{y}_0 \in \mathbb{R}$, $\hat{\boldsymbol{y}} \in \mathbb{R}^m$, $\mathrm{MMOT}^{\mathrm{LB}}$, $\mathrm{MMOT}^{\mathrm{UB}}$, and the following statements hold.

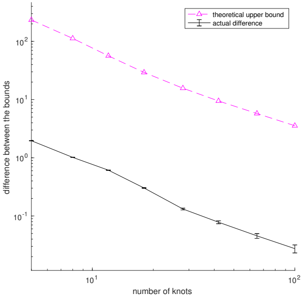
1. $\mathrm{MMOT}^{\mathrm{LB}} \leq (\mathrm{MMOT}) \leq \mathrm{MMOT}^{\mathrm{UB}}$;
2. $\hat{y}_0 + \langle \hat{\boldsymbol{y}}, \boldsymbol{g}(\cdot) \rangle$ *corresponds to an* $\hat{\epsilon}$*-optimizer of* (MMOT^*), *where* $\hat{\epsilon} := \mathrm{MMOT}^{\mathrm{UB}} - \mathrm{MMOT}^{\mathrm{LB}} \leq \tilde{\epsilon}$;
3. $\tilde{\mu}$ *is an* $\hat{\epsilon}$*-optimizer of* (MMOT), *where* $\hat{\epsilon} := \mathrm{MMOT}^{\mathrm{UB}} - \mathrm{MMOT}^{\mathrm{LB}} \leq \tilde{\epsilon}$.

A numerical example

- **Settings:**
  - $N = 50$;
  - $f(\boldsymbol{x}) := \left(\sum_{k=1}^{2} \left|\langle \boldsymbol{s}_k^+, \boldsymbol{x}\rangle - t_k^+\right|\right) - \left(\sum_{k=1}^{2} \left|\langle \boldsymbol{s}_k^-, \boldsymbol{x}\rangle - t_k^-\right|\right)$, where $\boldsymbol{s}_1^+$, $\boldsymbol{s}_2^+$, $\boldsymbol{s}_1^-$, $\boldsymbol{s}_2^-$ are uniformly randomly generated from the unit sphere in $\mathbb{R}^N$, and t_1^+, t_2^+, t_1^-, t_2^- are randomly generated real constants;
  - for $i = 1, \dots, N$, μ_i is a mixture of normal distributions with randomly generated parameters truncated to $[-10, 10]$.

- **Result:**





Further applications of the approximation scheme

- Multi-population matching (a.k.a. matching for teams)
  - Introduced by Carlier and Ekeland [2010] and includes the Wasserstein barycenter problem as a special case.
  - Carlier and Ekeland [2010] showed that the multi-population matching problem can be reformulated into an MMOT problem.
  - Therefore, we can apply the machinery we developed for solving MMOT problems to obtain an approximately optimal matching equilibrium.
  - Moreover, we adopt an alternative relaxation scheme to parametrize the transfer functions in the problem, which also results in an approximately optimal matching equilibrium.

Further application: two-stage DRO

- Two-stage distributionally robust optimization (DRO) with marginal constraints:

$$\inf_{\boldsymbol{a}\in S_1}\left\{\langle \boldsymbol{c}_1, \boldsymbol{a}\rangle + \sup_{\mu\in\Gamma(\mu_1,\ldots,\mu_N)}\int_{\mathcal{X}} Q(\boldsymbol{a},\boldsymbol{x})\,\mu(\mathrm{d}\boldsymbol{x})\right\},$$

where $Q(\boldsymbol{a},\boldsymbol{x})$ is the cost incurred by the second-stage problem:

$$\begin{aligned} Q(\boldsymbol{a},\boldsymbol{x}) := \quad & \underset{\boldsymbol{z}}{\text{minimize}} && \langle \boldsymbol{c}_2, \boldsymbol{z}\rangle \\ & \text{subject to} && \mathbf{A}_{\mathrm{in}}\boldsymbol{z} \le \mathbf{V}_{\mathrm{in}}\boldsymbol{a} + \mathbf{W}_{\mathrm{in}}\boldsymbol{x} + \boldsymbol{b}_{\mathrm{in}}, \\ & && \mathbf{A}_{\mathrm{eq}}\boldsymbol{z} = \mathbf{V}_{\mathrm{eq}}\boldsymbol{a} + \mathbf{W}_{\mathrm{eq}}\boldsymbol{x} + \boldsymbol{b}_{\mathrm{eq}}, \\ & && \boldsymbol{z} \in \mathbb{R}^{K_2}. \end{aligned}$$

- We can similarly replace $\Gamma(\mu_1,\ldots,\mu_N)$ by $\Gamma([\mu_1]_{\mathcal{G}_1},\ldots,[\mu_N]_{\mathcal{G}_N})$ for suitable $\mathcal{G}_i \subset \mathcal{L}^1(\mathcal{X}_i,\mu_i)$ for $i=1,\ldots,N$ to relax the two-stage DRO problem.

- Subsequently, we are able to compute an approximately optimal solution of the two-stage DRO problem such that the approximation error can be controlled to be arbitrarily close to 0.

References

1. A. Neufeld and Q. Xiang. Numerical method for feasible and approximately optimal solutions of multi-marginal optimal transport beyond discrete measures. Preprint, arXiv:2203.01633, 2022.
URL: https://arxiv.org/abs/2203.01633

2. A. Neufeld and Q. Xiang. Numerical methods for approximately optimal equilibrium in matching for teams beyond discrete measures. Working paper.

3. A. Neufeld and Q. Xiang. Numerical method for approximately optimal solutions of two-stage distributionally robust optimization with marginal constraints. Working paper.

4. M. A. Goberna and M. A. López. *Linear semi-infinite optimization.* John Wiley & Sons, 1998.

5. G. Carlier and I. Ekeland. Matching for teams. *Economic Theory*, 42(2):397–418, 2010.