Feasible approximation of matching equilibria in the matching for teams problem beyond discrete measures

*Supplementary material*

Ariel Neufeld, Qikun Xiang

# 1 Experiment 1

Recall that the setting used in the first numerical experiment is as follows.

**Assumption 1.1** (Assumption 4.1 in the paper)**.** *We assume that the following statements hold.*

- *For $i = 1, \dots, N$, $\mathcal{X}_i = \bigcup_{C\in\mathfrak{C}_i} C \subset \mathbb{R}^2$, where $\mathfrak{C}_i$ is a finite collection of triangles such that whenever $C_1 \cap C_2 \neq \emptyset$ for distinct $C_1, C_2 \in \mathfrak{C}_i$ then $C_1 \cap C_2$ is a face (i.e., a vertex or an edge) of $C_1$ and $C_2$. Moreover, $d_{\mathcal{X}_i}(\boldsymbol{x}_i, \boldsymbol{x}'_i) := \|\boldsymbol{x}_i - \boldsymbol{x}'_i\|_2$.*
- *$\mathcal{Z} = \bigcup_{C\in\mathfrak{C}_0} C \subset \mathbb{R}^2$, where $\mathfrak{C}_0$ is a finite collection of triangles such that whenever $C_1 \cap C_2 \neq \emptyset$ for distinct $C_1, C_2 \in \mathfrak{C}_0$ then $C_1 \cap C_2$ is a face (i.e., a vertex or an edge) of $C_1$ and $C_2$. Moreover, $d_{\mathcal{Z}}(\boldsymbol{z}, \boldsymbol{z}') := \|\boldsymbol{z} - \boldsymbol{z}'\|_2$.*
- *For $i = 1, \dots, N$, $\mu_i \in \mathcal{P}(\mathcal{X}_i)$ is absolutely continuous with respect to the Lebsegue measure on $\mathcal{X}_i$ and $\mathrm{supp}(\mu_i) = \mathcal{X}_i$.*
- *For $i = 1, \dots, N$, $c_i : \mathcal{X}_i \times \mathcal{Z} \to \mathbb{R}$ is given by $c_i(\boldsymbol{x}_i, \boldsymbol{z}) := \frac{1}{N}\big(\|\boldsymbol{z}\|_2^2 - 2\langle \boldsymbol{x}_i, \boldsymbol{z}\rangle\big)$.*
- *For $i = 1, \dots, N$, let $\{\boldsymbol{v}_{i,0}, \boldsymbol{v}_{i,1}, \dots, \boldsymbol{v}_{i,m_i}\}$ be an arbitrary enumeration of the finite set $V(\mathfrak{C}_i) := \big\{\boldsymbol{v} \in \mathbb{R}^2 : \boldsymbol{v}$ is an extreme point of some $C \in \mathfrak{C}_i\big\}$ (i.e., the cardinality of this set is $m_i + 1 \in \mathbb{N}$), let $\mathcal{G}_i := \big\{g_{i,\boldsymbol{v}_{i,1}}, \dots, g_{i,\boldsymbol{v}_{i,m_i}}\big\}$, and let $\boldsymbol{g}_i(\boldsymbol{x}_i) := \big(g_{i,\boldsymbol{v}_{i,1}}(\boldsymbol{x}_i), \dots, g_{i,\boldsymbol{v}_{i,m_i}}(\boldsymbol{x}_i)\big)^\mathsf{T}$ for all $\boldsymbol{x}_i \in \mathcal{X}_i$, where the functions $(g_{i,\boldsymbol{v}_{i,j}} : \mathcal{X}_i \to \mathbb{R})_{j=1:m_i}$ are defined in Theorem 2.17(vii).*
- *Let $\{\boldsymbol{v}_{0,0}, \boldsymbol{v}_{0,1}, \dots, \boldsymbol{v}_{0,k}\}$ be an arbitrary enumeration of the finite set $V(\mathfrak{C}_0) := \big\{\boldsymbol{v} \in \mathbb{R}^2 : \boldsymbol{v}$ is an extreme point of some $C \in \mathfrak{C}_0\big\}$ (i.e., the cardinality of this set is $k + 1 \in \mathbb{N}$), let $\mathcal{H} := \big\{h_{\boldsymbol{v}_{0,1}}, \dots, h_{\boldsymbol{v}_{0,k}}\big\}$, and let $\boldsymbol{h}(\boldsymbol{z}) := \big(h_{\boldsymbol{v}_{0,1}}(\boldsymbol{z}), \dots, h_{\boldsymbol{v}_{0,k}}(\boldsymbol{z})\big)^\mathsf{T}$ for all $\boldsymbol{z} \in \mathcal{Z}$, where the functions $(h_{\boldsymbol{v}_{0,l}} : \mathcal{Z} \to \mathbb{R})_{l=1:k}$ are defined in Theorem 2.17(vii).*

Thus, under Assumption 1.1, for $i = 1, \dots, N$ and any $\boldsymbol{y}_i \in \mathbb{R}^{m_i}$, the function $\mathcal{X}_i \ni \boldsymbol{x}_i \mapsto \langle \boldsymbol{g}_i(\boldsymbol{x}_i), \boldsymbol{y}_i\rangle \in \mathbb{R}$ is a continuous function that is piece-wise affine on each $C \in \mathfrak{C}_i$. Similarly, for any $\boldsymbol{w} \in \mathbb{R}^k$, the function $\mathcal{Z} \ni \boldsymbol{z} \mapsto \langle \boldsymbol{h}(\boldsymbol{z}), \boldsymbol{w}\rangle \in \mathbb{R}$ is continuous and piece-wise affine on each $C \in \mathfrak{C}_{\mathcal{Z}}$.

In the following, we introduce the detailed implementation of the global minimization oracle $\texttt{Oracle}(\cdot,\cdot,\cdot)$ used in the approximation of the matching for teams problem. For $i=1,\ldots,N$ and for any $\boldsymbol{y}_i\in\mathbb{R}^{m_i}$, $\boldsymbol{w}_i\in\mathbb{R}^k$, $\texttt{Oracle}(i,\boldsymbol{y}_i,\boldsymbol{w}_i)$ solves the global minimization problem $\inf_{\boldsymbol{x}\in\mathcal{X}_i,\,\boldsymbol{z}\in\mathcal{Z}}\big\{c_i(\boldsymbol{x},\boldsymbol{z})-\langle\boldsymbol{g}_i(\boldsymbol{x}),\boldsymbol{y}_i\rangle-\langle\boldsymbol{h}(\boldsymbol{z}),\boldsymbol{w}_i\rangle\big\}$. Under Assumption 1.1, for any $\boldsymbol{y}_i\in\mathbb{R}^{m_i}$, $\boldsymbol{w}_i\in\mathbb{R}^k$, we have

$$
\begin{aligned}
&\inf_{\boldsymbol{x}\in\mathcal{X}_i,\,\boldsymbol{z}\in\mathcal{Z}}\big\{c_i(\boldsymbol{x},\boldsymbol{z})-\langle\boldsymbol{g}_i(\boldsymbol{x}),\boldsymbol{y}_i\rangle-\langle\boldsymbol{h}(\boldsymbol{z}),\boldsymbol{w}_i\rangle\big\}\\
&\quad=\inf_{\boldsymbol{x}\in\mathcal{X}_i,\,\boldsymbol{z}\in\mathcal{Z}}\big\{\lambda_i\|\boldsymbol{z}\|_2^2-2\lambda_i\langle\boldsymbol{x},\boldsymbol{z}\rangle-\langle\boldsymbol{g}_i(\boldsymbol{x}),\boldsymbol{y}_i\rangle-\langle\boldsymbol{h}(\boldsymbol{z}),\boldsymbol{w}_i\rangle\big\}\\
&\quad=\min_{\boldsymbol{z}\in\mathcal{Z}}\Big\{\lambda_i\|\boldsymbol{z}\|_2^2-\max_{\boldsymbol{x}\in\mathcal{X}_i}\Big\{2\lambda_i\langle\boldsymbol{x},\boldsymbol{z}\rangle+\langle\boldsymbol{g}_i(\boldsymbol{x}),\boldsymbol{y}_i\rangle\Big\}-\langle\boldsymbol{h}(\boldsymbol{z}),\boldsymbol{w}_i\rangle\Big\}\\
&\quad=\min_{\boldsymbol{z}\in\mathcal{Z}}\Big\{\lambda_i\|\boldsymbol{z}\|_2^2-\max_{C\in\mathfrak{C}_i}\Big\{\max_{\boldsymbol{x}\in C}\Big\{2\lambda_i\langle\boldsymbol{x},\boldsymbol{z}\rangle+\langle\boldsymbol{g}_i(\boldsymbol{x}),\boldsymbol{y}_i\rangle\Big\}\Big\}-\langle\boldsymbol{h}(\boldsymbol{z}),\boldsymbol{w}_i\rangle\Big\}\\
&\quad=\min_{\boldsymbol{z}\in\mathcal{Z}}\Big\{\lambda_i\|\boldsymbol{z}\|_2^2-\max_{C\in\mathfrak{C}_i}\Big\{\max_{\boldsymbol{x}\in V(C)}\Big\{2\lambda_i\langle\boldsymbol{x},\boldsymbol{z}\rangle+\langle\boldsymbol{g}_i(\boldsymbol{x}),\boldsymbol{y}_i\rangle\Big\}\Big\}-\langle\boldsymbol{h}(\boldsymbol{z}),\boldsymbol{w}_i\rangle\Big\}\\
&\quad=\min_{\boldsymbol{z}\in\mathcal{Z}}\Big\{\lambda_i\|\boldsymbol{z}\|_2^2-\max_{\boldsymbol{x}\in V(\mathfrak{C}_i)}\Big\{2\lambda_i\langle\boldsymbol{x},\boldsymbol{z}\rangle+\langle\boldsymbol{g}_i(\boldsymbol{x}),\boldsymbol{y}_i\rangle\Big\}-\langle\boldsymbol{h}(\boldsymbol{z}),\boldsymbol{w}_i\rangle\Big\}\\
&\quad=\min_{\boldsymbol{x}\in V(\mathfrak{C}_i)}\Big\{\min_{\boldsymbol{z}\in\mathcal{Z}}\Big\{\lambda_i\|\boldsymbol{z}\|_2^2-2\lambda_i\langle\boldsymbol{x},\boldsymbol{z}\rangle-\langle\boldsymbol{h}(\boldsymbol{z}),\boldsymbol{w}_i\rangle\Big\}-\langle\boldsymbol{g}_i(\boldsymbol{x}),\boldsymbol{y}_i\rangle\Big\}\\
&\quad=\min_{\boldsymbol{x}\in V(\mathfrak{C}_i)}\Big\{\min_{C\in\mathfrak{C}_{\mathcal{Z}}}\Big\{\min_{\boldsymbol{z}\in C}\Big\{\lambda_i\|\boldsymbol{z}\|_2^2-2\lambda_i\langle\boldsymbol{x},\boldsymbol{z}\rangle-\langle\boldsymbol{h}(\boldsymbol{z}),\boldsymbol{w}_i\rangle\Big\}\Big\}-\langle\boldsymbol{g}_i(\boldsymbol{x}),\boldsymbol{y}_i\rangle\Big\}.
\end{aligned}
\tag{1.1}
$$

Since the function $\mathcal{Z}\ni\boldsymbol{z}\mapsto\langle\boldsymbol{h}(\boldsymbol{z}),\boldsymbol{w}_i\rangle\in\mathbb{R}$ is continuous and piece-wise affine on every $C\in\mathfrak{C}_{\mathcal{Z}}$, for fixed $\boldsymbol{x}\in V(\mathfrak{C}_i)$ and $C\in\mathfrak{C}_{\mathcal{Z}}$, the innermost minimization problem in (1.1) corresponds to minimizing a quadratic function over a triangle. In the following, let us fix an arbitrary $\boldsymbol{x}\in V(\mathfrak{C}_i)$ and an arbitrary $C\in\mathfrak{C}_{\mathcal{Z}}$. Let us assume that $C=\mathrm{conv}\big(\{\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{z}_3\}\big)$ for $\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{z}_3\in V(\mathfrak{C}_{\mathcal{Z}})$. Therefore, we have

$$
\begin{aligned}
&\min_{\boldsymbol{z}\in C}\Big\{\lambda_i\|\boldsymbol{z}\|_2^2-2\lambda_i\langle\boldsymbol{x},\boldsymbol{z}\rangle-\langle\boldsymbol{h}(\boldsymbol{z}),\boldsymbol{w}_i\rangle\Big\}\\
&\quad=\min\Big\{\lambda_i\|\alpha_1\boldsymbol{z}_1+\alpha_2\boldsymbol{z}_2+\alpha_3\boldsymbol{z}_3\|_2^2-2\lambda_i\langle\boldsymbol{x},\alpha_1\boldsymbol{z}_1+\alpha_2\boldsymbol{z}_2+\alpha_3\boldsymbol{z}_3\rangle\\
&\qquad\qquad-\langle\alpha_1\boldsymbol{h}(\boldsymbol{z}_1)+\alpha_2\boldsymbol{h}(\boldsymbol{z}_2)+\alpha_3\boldsymbol{h}(\boldsymbol{z}_3),\boldsymbol{w}_i\rangle:\\
&\qquad\qquad\alpha_1\ge0,\ \alpha_2\ge0,\ \alpha_3\ge0,\ \alpha_1+\alpha_2+\alpha_3=1\Big\}.
\end{aligned}
\tag{1.2}
$$

After introducing auxiliary variables $\zeta_1\ge0$, $\zeta_2\ge0$, $\zeta_3\ge0$, and $\xi\in\mathbb{R}$, we derive the following Karush-Kuhn-Tucker (KKT) optimality conditions for (1.2):

$$
\begin{aligned}
&2\lambda_i\langle\alpha_1^\star\boldsymbol{z}_1+\alpha_2^\star\boldsymbol{z}_2+\alpha_3^\star\boldsymbol{z}_3,\boldsymbol{z}_1\rangle-2\lambda_i\langle\boldsymbol{x},\boldsymbol{z}_1\rangle-\langle\boldsymbol{h}(\boldsymbol{z}_1),\boldsymbol{w}_i\rangle-\zeta_1^\star+\xi^\star=0,\\
&2\lambda_i\langle\alpha_1^\star\boldsymbol{z}_1+\alpha_2^\star\boldsymbol{z}_2+\alpha_3^\star\boldsymbol{z}_3,\boldsymbol{z}_2\rangle-2\lambda_i\langle\boldsymbol{x},\boldsymbol{z}_2\rangle-\langle\boldsymbol{h}(\boldsymbol{z}_2),\boldsymbol{w}_i\rangle-\zeta_2^\star+\xi^\star=0,\\
&2\lambda_i\langle\alpha_1^\star\boldsymbol{z}_1+\alpha_2^\star\boldsymbol{z}_2+\alpha_3^\star\boldsymbol{z}_3,\boldsymbol{z}_3\rangle-2\lambda_i\langle\boldsymbol{x},\boldsymbol{z}_3\rangle-\langle\boldsymbol{h}(\boldsymbol{z}_3),\boldsymbol{w}_i\rangle-\zeta_3^\star+\xi^\star=0,\\
&\alpha_1^\star+\alpha_2^\star+\alpha_3^\star=1,\\
&\alpha_1^\star\ge0,\ \alpha_2^\star\ge0,\ \alpha_3^\star\ge0,\\
&\zeta_1^\star\ge0,\ \zeta_2^\star\ge0,\ \zeta_3^\star\ge0.
\end{aligned}
\tag{1.3}
$$

The vectorized version of (1.3) is given by

$$
\begin{pmatrix}
2\lambda_i\langle \boldsymbol{z}_1,\boldsymbol{z}_1\rangle & 2\lambda_i\langle \boldsymbol{z}_1,\boldsymbol{z}_2\rangle & 2\lambda_i\langle \boldsymbol{z}_1,\boldsymbol{z}_3\rangle & -1 & 0 & 0 & 1\\
2\lambda_i\langle \boldsymbol{z}_2,\boldsymbol{z}_1\rangle & 2\lambda_i\langle \boldsymbol{z}_2,\boldsymbol{z}_2\rangle & 2\lambda_i\langle \boldsymbol{z}_2,\boldsymbol{z}_3\rangle & 0 & -1 & 0 & 1\\
2\lambda_i\langle \boldsymbol{z}_3,\boldsymbol{z}_1\rangle & 2\lambda_i\langle \boldsymbol{z}_3,\boldsymbol{z}_2\rangle & 2\lambda_i\langle \boldsymbol{z}_3,\boldsymbol{z}_3\rangle & 0 & 0 & -1 & 1\\
1 & 1 & 1 & 0 & 0 & 0 & 0
\end{pmatrix}
\begin{pmatrix}\alpha_1^\star\\ \alpha_2^\star\\ \alpha_3^\star\\ \zeta_1^\star\\ \zeta_2^\star\\ \zeta_3^\star\\ \xi^\star\end{pmatrix}
=
\begin{pmatrix}
2\lambda_i\langle \boldsymbol{x},\boldsymbol{z}_1\rangle + \langle \boldsymbol{h}(\boldsymbol{z}_1),\boldsymbol{w}_i\rangle\\
2\lambda_i\langle \boldsymbol{x},\boldsymbol{z}_2\rangle + \langle \boldsymbol{h}(\boldsymbol{z}_2),\boldsymbol{w}_i\rangle\\
2\lambda_i\langle \boldsymbol{x},\boldsymbol{z}_3\rangle + \langle \boldsymbol{h}(\boldsymbol{z}_3),\boldsymbol{w}_i\rangle\\
1
\end{pmatrix},
$$

$$
\begin{pmatrix}\alpha_1^\star\\ \alpha_2^\star\\ \alpha_3^\star\\ \zeta_1^\star\\ \zeta_2^\star\\ \zeta_3^\star\end{pmatrix} \geq \mathbf{0}_6.
$$

Let $\mathbf{Z} := \begin{pmatrix}\boldsymbol{z}_1^\top\\ \boldsymbol{z}_2^\top\\ \boldsymbol{z}_3^\top\end{pmatrix}$, $\mathbf{Z}_{[1,2]} := \begin{pmatrix}\boldsymbol{z}_1^\top\\ \boldsymbol{z}_2^\top\end{pmatrix}$, $\mathbf{Z}_{[1,3]} := \begin{pmatrix}\boldsymbol{z}_1^\top\\ \boldsymbol{z}_3^\top\end{pmatrix}$, $\mathbf{Z}_{[2,3]} := \begin{pmatrix}\boldsymbol{z}_2^\top\\ \boldsymbol{z}_3^\top\end{pmatrix}$, let $\boldsymbol{h}(\mathbf{Z}) := \begin{pmatrix}\boldsymbol{h}(\boldsymbol{z}_1)^\top\\ \boldsymbol{h}(\boldsymbol{z}_2)^\top\\ \boldsymbol{h}(\boldsymbol{z}_3)^\top\end{pmatrix}$, and let $\mathbf{1}_2 := (1,1)^\top$, $\mathbf{1}_3 := (1,1,1)^\top$, $\boldsymbol{e}_1 := (1,0,0)^\top$, $\boldsymbol{e}_2 := (0,1,0)^\top$, $\boldsymbol{e}_3 := (0,0,1)^\top$. We have the following seven case where we can simplify the KKT optimality conditions.

Case 1: $\zeta_1^\star = \zeta_2^\star = \zeta_3^\star = 0$.

$$
\begin{pmatrix}\alpha_1^\star\\ \alpha_2^\star\\ \alpha_3^\star\\ \xi^\star\end{pmatrix}
= \begin{pmatrix} 2\lambda_i\mathbf{Z}\mathbf{Z}^\top & \mathbf{1}_3\\ \mathbf{1}_3^\top & 0\end{pmatrix}^{-1}
\begin{pmatrix} 2\lambda_i\mathbf{Z}\boldsymbol{x} + \boldsymbol{h}(\mathbf{Z})\boldsymbol{w}_i\\ 1\end{pmatrix},\quad
\begin{pmatrix}\alpha_1^\star\\ \alpha_2^\star\\ \alpha_3^\star\end{pmatrix} \geq \mathbf{0}_3.
$$

Case 2: $\zeta_1^\star = \zeta_2^\star = \alpha_3^\star = 0$.

$$
\begin{pmatrix}\alpha_1^\star\\ \alpha_2^\star\\ \zeta_3^\star\\ \xi^\star\end{pmatrix}
= \begin{pmatrix} 2\lambda_i\mathbf{Z}\mathbf{Z}_{[1,2]}^\top & -\boldsymbol{e}_3 & \mathbf{1}_3\\ \mathbf{1}_2^\top & 0 & 0\end{pmatrix}^{-1}
\begin{pmatrix} 2\lambda_i\mathbf{Z}\boldsymbol{x} + \boldsymbol{h}(\mathbf{Z})\boldsymbol{w}_i\\ 1\end{pmatrix},\quad
\begin{pmatrix}\alpha_1^\star\\ \alpha_2^\star\\ \zeta_3^\star\end{pmatrix} \geq \mathbf{0}_3.
$$

Case 3: $\zeta_1^\star = \alpha_2^\star = \zeta_3^\star = 0$.

$$
\begin{pmatrix}\alpha_1^\star\\ \alpha_3^\star\\ \zeta_2^\star\\ \xi^\star\end{pmatrix}
= \begin{pmatrix} 2\lambda_i\mathbf{Z}\mathbf{Z}_{[1,3]}^\top & -\boldsymbol{e}_2 & \mathbf{1}_3\\ \mathbf{1}_2^\top & 0 & 0\end{pmatrix}^{-1}
\begin{pmatrix} 2\lambda_i\mathbf{Z}\boldsymbol{x} + \boldsymbol{h}(\mathbf{Z})\boldsymbol{w}_i\\ 1\end{pmatrix},\quad
\begin{pmatrix}\alpha_1^\star\\ \alpha_3^\star\\ \zeta_2^\star\end{pmatrix} \geq \mathbf{0}_3.
$$

Case 4: $\alpha_1^\star = \zeta_2^\star = \zeta_3^\star = 0$.

$$\begin{pmatrix} \alpha_2^\star \\ \alpha_3^\star \\ \zeta_1^\star \\ \xi^\star \end{pmatrix} = \begin{pmatrix} 2\lambda_i \mathbf{Z}\mathbf{Z}_{[2,3]}^\mathsf{T} & -\boldsymbol{e}_1 & \mathbf{1}_3 \\ \mathbf{1}_2^\mathsf{T} & 0 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 2\lambda_i \mathbf{Z}\boldsymbol{x} + \boldsymbol{h}(\mathbf{Z})\boldsymbol{w}_i \\ 1 \end{pmatrix}, \quad \begin{pmatrix} \alpha_2^\star \\ \alpha_3^\star \\ \zeta_1^\star \end{pmatrix} \geq \mathbf{0}_3.$$

Case 5: $\zeta_1^\star = \alpha_2^\star = \alpha_3^\star = 0$.

$$\begin{pmatrix} \alpha_1^\star \\ \zeta_2^\star \\ \zeta_3^\star \\ \xi^\star \end{pmatrix} = \begin{pmatrix} 2\lambda_i \mathbf{Z}\boldsymbol{z}_1 & -\boldsymbol{e}_2 & -\boldsymbol{e}_3 & \mathbf{1}_3 \\ 1 & 0 & 0 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 2\lambda_i \mathbf{Z}\boldsymbol{x} + \boldsymbol{h}(\mathbf{Z})\boldsymbol{w}_i \\ 1 \end{pmatrix}, \quad \begin{pmatrix} \alpha_1^\star \\ \zeta_2^\star \\ \zeta_3^\star \end{pmatrix} \geq \mathbf{0}_3.$$

Case 6: $\alpha_1^\star = \zeta_2^\star = \alpha_3^\star = 0$.

$$\begin{pmatrix} \alpha_2^\star \\ \zeta_1^\star \\ \zeta_3^\star \\ \xi^\star \end{pmatrix} = \begin{pmatrix} 2\lambda_i \mathbf{Z}\boldsymbol{z}_2 & -\boldsymbol{e}_1 & -\boldsymbol{e}_3 & \mathbf{1}_3 \\ 1 & 0 & 0 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 2\lambda_i \mathbf{Z}\boldsymbol{x} + \boldsymbol{h}(\mathbf{Z})\boldsymbol{w}_i \\ 1 \end{pmatrix}, \quad \begin{pmatrix} \alpha_2^\star \\ \zeta_1^\star \\ \zeta_3^\star \end{pmatrix} \geq \mathbf{0}_3.$$

Case 7: $\alpha_1^\star = \alpha_2^\star = \zeta_3^\star = 0$.

$$\begin{pmatrix} \alpha_3^\star \\ \zeta_1^\star \\ \zeta_2^\star \\ \xi^\star \end{pmatrix} = \begin{pmatrix} 2\lambda_i \mathbf{Z}\boldsymbol{z}_3 & -\boldsymbol{e}_1 & -\boldsymbol{e}_2 & \mathbf{1}_3 \\ 1 & 0 & 0 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 2\lambda_i \mathbf{Z}\boldsymbol{x} + \boldsymbol{h}(\mathbf{Z})\boldsymbol{w}_i \\ 1 \end{pmatrix}, \quad \begin{pmatrix} \alpha_3^\star \\ \zeta_1^\star \\ \zeta_2^\star \end{pmatrix} \geq \mathbf{0}_3.$$

These seven cases can help us compute (1.2) efficiently, which will then allow us to compute (1.1).

## 2 Semi-discrete optimal transport in $\mathbb{R}^2$

In this section, we will present the numerical method that we use to compute the semi-discrete optimal transport under the setting of Experiment 1 (see Assumption 1.1 above as well as Assumption 4.1 in the paper). This corresponds to the following setting.

**Assumption 2.1** (Semi-discrete optimal transport in two dimensions)**.** *We assume that the following statements hold.*

- $\mathcal{X} \subset \mathbb{R}^2$ *is a convex polytope,* $d_\mathcal{X}(\boldsymbol{x}, \boldsymbol{x}') = \|\boldsymbol{x} - \boldsymbol{x}'\|_2$*;*
- $n \in \mathbb{N}$, $\alpha_i \in (0, 1]$, $\boldsymbol{x}_i \in \mathcal{X}$ *for* $i = 1, \ldots, n$, $\sum_{i=1}^n \alpha_i = 1$, $\nu_1 = \sum_{i=1}^n \alpha_i \delta_{\boldsymbol{x}_i}$*;*
- $\nu_2 \in \mathcal{P}(\mathcal{X})$ *has a bounded density function* $f : \mathcal{X} \to \mathbb{R}_+$ *with respect to the Lebesgue measure on* $\mathbb{R}^2$.

In the case where the support of $\nu_2$ is non-convex, we can take $\mathcal{X}$ to be the convex hull of the support of $\nu_2$. It follows from Lemma 3.1 in the paper that an optimal coupling between $\nu_1$ and $\nu_2$ is characterized as follows. Let $h : \mathbb{R}^n \to \mathbb{R}$ be defined (see also equation (3.1) in the paper) by

$$h(\phi_1, \ldots, \phi_n) := \left( \sum_{i=1}^{n} \alpha_i \phi_i \right) - \int_{\mathcal{X}} \max_{1 \le i \le n} \left\{ \phi_i - \|\boldsymbol{x}_i - \boldsymbol{y}\|_2 \right\} \nu_2(\mathrm{d}\boldsymbol{y}) \qquad \forall (\phi_1, \ldots, \phi_n) \in \mathbb{R}^n.$$

Moreover, for $i = 1, \ldots, n$ and for all $(\phi_1, \ldots, \phi_n) \in \mathbb{R}^n$, let $V_i(\phi_1, \ldots, \phi_n) \subset \mathcal{X}$ be defined (see also equation (3.2) in the paper) as follows

$$V_i(\phi_1, \ldots, \phi_n) := \Big\{ \boldsymbol{y} \in \mathcal{X} : \phi_i - \|\boldsymbol{x}_i - \boldsymbol{y}\|_2 = \max_{1 \le k \le n} \left\{ \phi_k - \|\boldsymbol{x}_k - \boldsymbol{y}\|_2 \right\} \Big\}.$$

It follows from the proof of [1, Proposition 3.1.2] that for any $(\phi_1, \ldots, \phi_n) \in \mathbb{R}^n$ and $i \neq j$, $V_i(\phi_1, \ldots, \phi_n) \cap V_j(\phi_1, \ldots, \phi_n)$ is $\nu_2$-negligible. Hence, $h(\phi_1, \ldots, \phi_n)$ can be rewritten as

$$h(\phi_1, \ldots, \phi_n) = \sum_{i=1}^{n} \left( \alpha_i \phi_i - \int_{V_i(\phi_1, \ldots, \phi_n)} \phi_i - \|\boldsymbol{x}_i - \boldsymbol{y}\|_2 \, \nu_2(\mathrm{d}\boldsymbol{y}) \right).$$

Furthermore, it follows from the proof of [1, Proposition 3.1.2] that $h$ is differentiable with gradient $\nabla h : \mathbb{R}^n \to \mathbb{R}^n$ given by

$$\nabla h(\phi_1, \ldots, \phi_n) = \big( \alpha_1 - \nu_2(V_1(\phi_1, \ldots, \phi_n)), \ldots, \alpha_n - \nu_2(V_n(\phi_1, \ldots, \phi_n)) \big)^{\mathsf{T}} \qquad \forall (\phi_1, \ldots, \phi_n) \in \mathbb{R}^n.$$

Lemma 3.1 in the paper states that an optimal coupling of $\nu_1$ and $\nu_2$ is characterized by the sets $\big(V_i(\phi_1^\star, \ldots, \phi_n^\star)\big)_{i=1:n}$ where $(\phi_1^\star, \ldots, \phi_n^\star) \in \arg\max_{(\phi_1, \ldots, \phi_n)} \big\{ h(\phi_1, \ldots, \phi_n) \big\}$. Therefore, if we are able to evaluate $\int_{V_i(\phi_1, \ldots, \phi_n)} \phi_i - \|\boldsymbol{x}_i - \boldsymbol{y}\|_2 \, \nu_2(\mathrm{d}\boldsymbol{y})$ and $\nu_2(V_i(\phi_1, \ldots, \phi_n))$ for $i = 1, \ldots, n$ and any $(\phi_1, \ldots, \phi_n) \in \mathbb{R}^n$, we are then able to compute $(\phi_1^\star, \ldots, \phi_n^\star) \in \arg\max_{(\phi_1, \ldots, \phi_n)} \big\{ h(\phi_1, \ldots, \phi_n) \big\}$ via any first-order convex optimization method.

An important observations is that for any $(\phi_1, \ldots, \phi_n) \in \mathbb{R}^n$ and any $i \in \{1, \ldots, n\}$, the set $V_i(\phi_1, \ldots, \phi_n)$ is characterized by straight lines and hyperbolas, i.e.,

$$V_i(\phi_1, \ldots, \phi_n) = \mathcal{X} \cap \bigcap_{k \neq i} \Big\{ \boldsymbol{y} \in \mathbb{R}^n : \|\boldsymbol{x}_i - \boldsymbol{y}\|_2 - \|\boldsymbol{x}_k - \boldsymbol{y}\|_2 \le \phi_i - \phi_k \Big\}.$$

For $i = 1, \ldots, n$, let $\overline{\rho}_{\mathcal{X},i} : [0, 2\pi] \to \mathbb{R}_+$ be defined by

$$\overline{\rho}_{\mathcal{X},i}(\theta) := \max \Big\{ \rho \ge 0 : \boldsymbol{x}_i + \rho \left( \begin{smallmatrix} \cos\theta \\ \sin\theta \end{smallmatrix} \right) \in \mathcal{X} \Big\} \qquad \forall \theta \in [0, 2\pi].$$

Since $\mathcal{X}$ is a convex polytope, which can be characterized by the intersection of finitely many closed half-spaces, the function $\overline{\rho}_{\mathcal{X},i}(\cdot)$ can be efficiently evaluated via the polar equations of straight lines. Similarly, for $i = 1, \ldots, n$, $k \neq i$, let $\overline{\rho}_{i,k} : [0, 2\pi] \to \mathbb{R}_+$ be defined by

$$\overline{\rho}_{i,k}(\theta) := \max \Big\{ \max \Big\{ \rho \ge 0 : \rho - \big\| \boldsymbol{x}_k - \boldsymbol{x}_i - \rho \left( \begin{smallmatrix} \cos\theta \\ \sin\theta \end{smallmatrix} \right) \big\|_2 \le \phi_i - \phi_k \Big\}, 0 \Big\} \qquad \forall \theta \in [0, 2\pi].$$

Since the set $\Big\{ \boldsymbol{y} \in \mathbb{R}^n : \|\boldsymbol{x}_i - \boldsymbol{y}\|_2 - \|\boldsymbol{x}_k - \boldsymbol{y}\|_2 \le \phi_i - \phi_k \Big\}$, unless empty, can be characterized by a hyperbola, the function $\overline{\rho}_{i,k}(\cdot)$ can be efficiently evaluated via the polar equation of hyperbola. Finally, letting $\overline{\rho}_i(\theta) := \min \big\{ \overline{\rho}_{\mathcal{X},i}(\theta), \min_{k \neq i} \big\{ \overline{\rho}_{i,k}(\theta) \big\} \big\}$ for all $\theta \in [0, 2\pi]$ for $i = 1, \ldots, n$, the terms

$\int_{V_i(\phi_1,\ldots,\phi_n)} \phi_i - \|\boldsymbol{x}_i - \boldsymbol{y}\|_2 \, \nu_2(\mathrm{d}\boldsymbol{y})$ and $\nu_2(V_i(\phi_1,\ldots,\phi_n))$ can be re-expressed as follows:

$$\int_{V_i(\phi_1,\ldots,\phi_n)} \phi_i - \|\boldsymbol{x}_i - \boldsymbol{y}\|_2 \, \nu_2(\mathrm{d}\boldsymbol{y}) = \int_0^{2\pi} \int_0^{\overline{\rho}_i(\theta)} \rho(\phi_i - \rho) f\big(\boldsymbol{x}_i + \rho\big(\begin{smallmatrix}\cos\theta\\ \sin\theta\end{smallmatrix}\big)\big) \, \mathrm{d}\rho \, \mathrm{d}\theta \qquad \text{for } i = 1,\ldots,n,$$
$$\nu_2(V_i(\phi_1,\ldots,\phi_n)) = \int_0^{2\pi} \int_0^{\overline{\rho}_i(\theta)} \rho f\big(\boldsymbol{x}_i + \rho\big(\begin{smallmatrix}\cos\theta\\ \sin\theta\end{smallmatrix}\big)\big) \, \mathrm{d}\rho \, \mathrm{d}\theta \qquad \text{for } i = 1,\ldots,n.$$

If the inner integrals $\int_0^{\overline{\rho}_i(\theta)} \rho(\phi_i - \rho) f\big(\boldsymbol{x}_i + \rho\big(\begin{smallmatrix}\cos\theta\\ \sin\theta\end{smallmatrix}\big)\big) \, \mathrm{d}\rho$ and $\int_0^{\overline{\rho}_i(\theta)} \rho f\big(\boldsymbol{x}_i + \rho\big(\begin{smallmatrix}\cos\theta\\ \sin\theta\end{smallmatrix}\big)\big) \, \mathrm{d}\rho$ are piece-wise continuous in $\theta$ and can be efficiently computed, e.g., when $f(\cdot)$ is piece-wise affine, then the terms $\int_{V_i(\phi_1,\ldots,\phi_n)} \phi_i - \|\boldsymbol{x}_i - \boldsymbol{y}\|_2 \, \nu_2(\mathrm{d}\boldsymbol{y})$ and $\nu_2(V_i(\phi_1,\ldots,\phi_n))$ can be accurately approximated via any numerical quadrature procedure.

# 3 Experiment 2

## 3.1 Setting

Recall that the setting used in the second numerical experiment is as follows.

**Assumption 3.1** (Assumption 4.3 in the paper)**.** *We assume that the following statements hold.*

- *For $i = 1,\ldots,N$, $\mathcal{X}_i = [\underline{\kappa}_i, \overline{\kappa}_i] \subset \mathbb{R}$, where $-\infty < \underline{\kappa}_i < \overline{\kappa}_i < \infty$. Moreover, $d_{\mathcal{X}_i}(x_i, x_i') := |x_i - x_i'|$.*
- *$\mathcal{Z} = \bigcup_{C \in \mathfrak{C}_0} C \subset \mathbb{R}^2$ is a polytope, where $\mathfrak{C}_0$ is a finite collection of triangles such that whenever $C_1 \cap C_2 \neq \emptyset$ for distinct $C_1, C_2 \in \mathfrak{C}_0$ then $C_1 \cap C_2$ is a face (i.e., a vertex or an edge) of $C_1$ and $C_2$. Moreover, $d_{\mathcal{Z}}(\boldsymbol{z}, \boldsymbol{z}') := \|\boldsymbol{z} - \boldsymbol{z}'\|_2$.*
- *For $i = 1,\ldots,N$, $\mu_i \in \mathcal{P}(\mathcal{X}_i)$ is absolutely continuous with respect to the Lebsegue measure on $\mathcal{X}_i$ and $\mathrm{supp}(\mu_i) = \mathcal{X}_i$.*
- *For $i = 1,\ldots,N$, $c_i : \mathcal{X}_i \times \mathcal{Z} \to \mathbb{R}$ is given by $c_i(x_i, \boldsymbol{z}) := l_i\big(x_i - \langle \boldsymbol{s}_i, \boldsymbol{z}\rangle\big)$ where $\boldsymbol{s}_i \in \mathbb{R}^2$ and $l_i : [\lambda_{i,0}, \lambda_{i,n_i}] \to \mathbb{R}$ is a continuous function that is piece-wise affine on $[\lambda_{i,0}, \lambda_{i,1}], \ldots, [\lambda_{i,n_i-1}, \lambda_{i,n_i}]$ for $n_i \in \mathbb{N}$, $-\infty < \lambda_{i,0} < \lambda_{i,1} < \cdots < \lambda_{i,n_i} < \infty$, satisfying $\lambda_{i,0} \leq \underline{\kappa}_i - \max_{\boldsymbol{z} \in \mathcal{Z}} \big\{\langle \boldsymbol{s}_i, \boldsymbol{z}\rangle\big\}$, $\lambda_{i,n_i} \geq \overline{\kappa}_i - \min_{\boldsymbol{z} \in \mathcal{Z}} \big\{\langle \boldsymbol{s}_i, \boldsymbol{z}\rangle\big\}$.*
- *For $i = 1,\ldots,N$, $\underline{\kappa}_i = \kappa_{i,0} < \kappa_{i,1} < \cdots < \kappa_{i,m_i} = \overline{\kappa}_i$, and $g_{i,0}, g_{i,1}, \ldots, g_{i,m_i} : \mathcal{X}_i \to \mathbb{R}$ are defined as follows:*

  $$\begin{aligned} g_{i,0}(x_i) &:= \frac{(\kappa_{i,1} - x_i)^+}{\kappa_{i,1} - \kappa_{i,0}}, \\ g_{i,j}(x_i) &:= \min\left\{\frac{(x_i - \kappa_{i,j-1})^+}{\kappa_{i,j} - \kappa_{i,j-1}}, \frac{(\kappa_{i,j+1} - x_i)^+}{\kappa_{i,j+1} - \kappa_{i,j}}\right\} \qquad \forall 1 \leq j \leq m_i - 1, \\ g_{i,m_i}(x_i) &:= \frac{(x_i - \kappa_{i,m_i-1})^+}{\kappa_{i,m_i} - \kappa_{i,m_i-1}}. \end{aligned}$$

  *Let $\mathcal{G}_i := \{g_{i,1}, \ldots, g_{i,m_i}\}$ and let $\boldsymbol{g}_i(x_i) := \big(g_{i,1}(x_i), \ldots, g_{i,m_i}(x_i)\big)^\mathsf{T}$ for all $x_i \in \mathcal{X}_i$.*
- *Let $\{\boldsymbol{v}_{0,0}, \boldsymbol{v}_{0,1}, \ldots, \boldsymbol{v}_{0,k}\}$ be an arbitrary enumeration of the finite set $V(\mathfrak{C}_0) := \big\{\boldsymbol{v} \in \mathbb{R}^2 : \boldsymbol{v}$ is an extreme point of some $C \in \mathfrak{C}_0\big\}$ (i.e., the cardinality of this set is $k + 1 \in \mathbb{N}$), let $\mathcal{H} :=$*

$\{h_{\boldsymbol{v}_{0,1}}, \dots, h_{\boldsymbol{v}_{0,k}}\}$, *and let* $\boldsymbol{h}(\boldsymbol{z}) := \big(h_{\boldsymbol{v}_{0,1}}(\boldsymbol{z}), \dots, h_{\boldsymbol{v}_{0,k}}(\boldsymbol{z})\big)^{\mathsf{T}}$ *for all* $\boldsymbol{z} \in \mathcal{Z}$*, where the functions* $(h_{\boldsymbol{v}_{0,l}} : \mathcal{Z} \to \mathbb{R})_{l=1:k}$ *are defined in Theorem 2.17(vii).*

In the following subsections, we introduce the detailed implementations of the function $\bar{c}(x_1, \dots, x_N)$ and the global minimization oracles used in the approximation of the matching for teams problem.

### 3.2 Implementation of the function $\bar{c}(x_1, \dots, x_N)$

Under Assumption 3.1, using the mixed-integer formulation of piece-wise affine functions by Vielma, Ahmed, and Nemhauser [2], the minimization problem $\bar{c}(x_1, \dots, x_N) := \min_{\boldsymbol{z}\in\mathcal{Z}} \big\{ \sum_{i=1}^{N} c_i(x_i, \boldsymbol{z}) \big\}$ can be formulated into the following mixed-integer linear programming problem:

$$
\begin{aligned}
\underset{\boldsymbol{z},\,(\zeta_{i,t}),\,(\iota_{i,t})}{\text{minimize}} \quad & \sum_{i=1}^{N} l_i(\lambda_{i,0}) + \sum_{t=1}^{n_i} \big(l_i(\lambda_{i,t}) - l_i(\lambda_{i,t-1})\big)\zeta_{i,t} \\
\text{subject to} \quad & \text{for } i = 1, \dots, N: \\
& \begin{cases}
\zeta_{i,t} \in \mathbb{R} & \forall 1 \le t \le n_i, \\
\iota_{i,t} \in \{0,1\} & \forall 1 \le t \le n_i - 1, \\
\zeta_{i,1} \le 1, \ \zeta_{i,n_i} \ge 0, & \\
\zeta_{i,t+1} \le \iota_{i,t} \le \zeta_{i,t} & \forall 1 \le t \le n_i - 1, \\
\lambda_{i,0} + \sum_{t=1}^{n_i} (\lambda_{i,t} - \lambda_{i,t-1})\zeta_{i,t} = x_i - \langle \boldsymbol{s}_i, \boldsymbol{z} \rangle, &
\end{cases} \\
& \boldsymbol{z} \in \mathcal{Z}.
\end{aligned}
$$

### 3.3 Implementation of $\texttt{Oracle}(\cdot,\,\cdot,\,\cdot)$

For $i = 1, \dots, N$ and any $\boldsymbol{y}_i \in \mathbb{R}^{m_i}$, $\boldsymbol{w}_i \in \mathbb{R}^k$, $\texttt{Oracle}(i, \boldsymbol{y}_i, \boldsymbol{w}_i)$ solves the global minimization problem $\inf_{x_i\in\mathcal{X}_i,\,\boldsymbol{z}\in\mathcal{Z}} \big\{ c_i(x_i, \boldsymbol{z}) - \langle \boldsymbol{g}_i(x_i), \boldsymbol{y}_i \rangle - \langle \boldsymbol{h}(\boldsymbol{z}), \boldsymbol{w}_i \rangle \big\}$. Under Assumption 3.1, for any $\boldsymbol{y}_i = (y_{i,1}, \dots, y_{i,m_i})^{\mathsf{T}} \in \mathbb{R}^{m_i}$, $\boldsymbol{w}_i = (w_{i,1}, \dots, w_{i,k})^{\mathsf{T}} \in \mathbb{R}^k$, we have

$$
\begin{aligned}
&\inf_{x_i\in\mathcal{X}_i,\,\boldsymbol{z}\in\mathcal{Z}} \big\{ c_i(x_i, \boldsymbol{z}) - \langle \boldsymbol{g}_i(x_i), \boldsymbol{y}_i \rangle - \langle \boldsymbol{h}(\boldsymbol{z}), \boldsymbol{w}_i \rangle \big\} \\
&\qquad = \min_{x_i\in\mathcal{X}_i,\,\boldsymbol{z}\in\mathcal{Z}} \big\{ l_i\big(x_i - \langle \boldsymbol{s}_i, \boldsymbol{z} \rangle\big) - \langle \boldsymbol{g}_i(x_i), \boldsymbol{y}_i \rangle - \langle \boldsymbol{h}(\boldsymbol{z}), \boldsymbol{w}_i \rangle \big\}.
\end{aligned} \tag{3.1}
$$

Notice that $\tilde{g}_i^{\boldsymbol{y}_i}(x_i) := \langle \boldsymbol{g}_i(x_i), \boldsymbol{y}_i \rangle$ is a continuous function on $\mathcal{X}_i$ that is piece-wise affine on $[\kappa_{i,0}, \kappa_{i,1}], \dots, [\kappa_{i,m_i-1}, \kappa_{i,m_i}]$, where $\tilde{g}_i^{\boldsymbol{y}_i}(\kappa_{i,0}) = 0$ and $\tilde{g}_i^{\boldsymbol{y}_i}(\kappa_{i,j}) = y_{i,j}$ for $j = 1, \dots, m_i$. Let $y_{i,0} \equiv 0$. Moreover, notice that $\tilde{h}^{\boldsymbol{w}_i}(\boldsymbol{z}) := \langle \boldsymbol{h}(\boldsymbol{z}), \boldsymbol{w}_i \rangle$ is a continuous function on $\mathcal{Z}$ that is piece-wise affine on each $C \in \mathfrak{C}_{\mathcal{Z}}$. We define $\mathcal{C}_{\boldsymbol{v}} := \big\{ C \in \mathfrak{C}_{\mathcal{Z}} : \boldsymbol{v} \in V(C) \big\}$ for each $\boldsymbol{v} \in V(\mathfrak{C}_{\mathcal{Z}})$. Subsequently, using the mixed-integer formulation of piece-wise affine functions in [2], (3.1) can be formulated into the following mixed-integer linear programming problem:

$$
\begin{aligned}
\underset{\substack{x_i,\boldsymbol{z},(\zeta_t),(\iota_t),\\(\xi_j),(\eta_j),(\beta_{\boldsymbol{v}}),(\chi_C)}}{\text{minimize}} \quad & l_i(\lambda_{i,0})+\sum_{t=1}^{n_i}\big(l_i(\lambda_{i,t})-l_i(\lambda_{i,t-1})\big)\zeta_t+\sum_{j=1}^{m_i}(y_{i,j-1}-y_{i,j})\xi_j-\Bigg(\sum_{\boldsymbol{v}\in V(\mathfrak{C}_{\mathcal{Z}})}\tilde{h}^{\boldsymbol{w}_i}(\boldsymbol{v})\beta_{\boldsymbol{v}}\Bigg)\\
\text{subject to} \quad & \begin{cases}\zeta_t\in\mathbb{R} & \forall 1\le t\le n_i,\\ \iota_t\in\{0,1\} & \forall 1\le t\le n_i-1,\\ \zeta_1\le 1,\ \zeta_{n_i}\ge 0, & \\ \zeta_{t+1}\le\iota_t\le\zeta_t & \forall 1\le t\le n_i-1,\\ \lambda_{i,0}+\sum_{t=1}^{n_i}(\lambda_{i,t}-\lambda_{i,t-1})\zeta_t=x_i-\langle\boldsymbol{s}_i,\boldsymbol{z}\rangle, &\end{cases}\\
& \begin{cases}\xi_j\in\mathbb{R} & \forall 1\le j\le m_i,\\ \eta_i\in\{0,1\} & \forall 1\le j\le m_i-1,\\ \xi_1\le 1,\ \xi_{m_i}\ge 0, & \\ \xi_{j+1}\le\eta_j\le\xi_j & \forall 1\le j\le m_i-1,\\ \kappa_{i,0}+\sum_{j=1}^{m_i}(\kappa_{i,j}-\kappa_{i,j-1})\xi_j=x_i, &\end{cases}\\
& \begin{cases}\beta_{\boldsymbol{v}}\ge 0 & \forall \boldsymbol{v}\in V(\mathfrak{C}_{\mathcal{Z}}),\\ \chi_C\in\{0,1\} & \forall C\in\mathfrak{C}_{\mathcal{Z}},\\ \sum_{\boldsymbol{v}\in V(C)}\beta_{\boldsymbol{v}}=1 & \forall C\in\mathfrak{C}_{\mathcal{Z}},\\ \beta_{\boldsymbol{v}}\le\sum_{C\in\mathcal{C}_{\boldsymbol{v}}}\chi_C & \forall \boldsymbol{v}\in V(\mathfrak{C}_{\mathcal{Z}}),\\ \sum_{\boldsymbol{v}\in V(\mathfrak{C}_{\mathcal{Z}})}\beta_{\boldsymbol{v}}\boldsymbol{v}=\boldsymbol{z}. &\end{cases}
\end{aligned}
\tag{3.2}
$$

## 3.4 Alternative implementation of the global minimization oracle for $(\mathsf{MT}^*_{\mathsf{par}})$

Let us now present the alternative mixed-integer formulations of piece-wise affine functions in [2] where the number of binary variables is logarithmic in the number of "pieces". Let us again work under Assumption 3.1 and fix an arbitrary $i\in\{1,\dots,N\}$ as well as $\boldsymbol{y}_i=(y_{i,1},\dots,y_{i,m_i})^\mathsf{T}\in\mathbb{R}^{m_i}$ and $\boldsymbol{w}_i=(w_{i,1},\dots,w_{i,k})^\mathsf{T}\in\mathbb{R}^k$. Again, let us define $\tilde{g}_i^{\boldsymbol{y}_i}(x_i):=\langle\boldsymbol{g}_i(x_i),\boldsymbol{y}_i\rangle$ and $\tilde{h}^{\boldsymbol{w}_i}(\boldsymbol{z}):=\langle\boldsymbol{h}(\boldsymbol{z}),\boldsymbol{w}_i\rangle$. Let $q_1:=\lceil\log_2(n_i)\rceil$, and let $(\mathsf{b}_1^r)_{r=1:n_i}\subseteq\{0,1\}^{q_1}$ be a sequence of distinct binary-valued vectors such that $\mathsf{b}_1^r$ and $\mathsf{b}_1^{r+1}$ differ by at most one component for $r=1,\dots,n_i-1$. For $s=1,\dots,q_1$, define

$$
\begin{aligned}
\mathcal{L}_1^s&:=\big\{r:r=0,\ [\mathsf{b}_1^1]_s=0\big\}\cup\big\{r:1\le r\le n_i-1,\ [\mathsf{b}_1^r]_s=0,\ [\mathsf{b}_1^{r+1}]_s=0\big\}\cup\big\{r:r=n_i,\ [\mathsf{b}_1^{n_i}]_s=0\big\},\\
\mathcal{R}_1^s&:=\big\{r:r=0,\ [\mathsf{b}_1^1]_s=1\big\}\cup\big\{r:1\le r\le n_i-1,\ [\mathsf{b}_1^r]_s=1,\ [\mathsf{b}_1^{r+1}]_s=1\big\}\cup\big\{r:r=n_i,\ [\mathsf{b}_1^{n_i}]_s=1\big\}.
\end{aligned}
$$

Similarly, let $q_2:=\lceil\log_2(m_i)\rceil$, and let $(\mathsf{b}_2^r)_{r=1:m_i}\subseteq\{0,1\}^{q_2}$ be a sequence of distinct binary-valued vectors such that $\mathsf{b}_2^r$ and $\mathsf{b}_2^{r+1}$ differ by at most one component for $r=1,\dots,m_i-1$. For $s=1,\dots,q_2$, define

$$
\begin{aligned}
\mathcal{L}_2^s&:=\big\{r:r=0,\ [\mathsf{b}_2^1]_s=0\big\}\cup\big\{r:1\le r\le m_i-1,\ [\mathsf{b}_2^r]_s=0,\ [\mathsf{b}_2^{r+1}]_s=0\big\}\cup\big\{r:r=m_i,\ [\mathsf{b}_2^{m_i}]_s=0\big\},\\
\mathcal{R}_2^s&:=\big\{r:r=0,\ [\mathsf{b}_2^1]_s=1\big\}\cup\big\{r:1\le r\le m_i-1,\ [\mathsf{b}_2^r]_s=1,\ [\mathsf{b}_2^{r+1}]_s=1\big\}\cup\big\{r:r=m_i,\ [\mathsf{b}_2^{m_i}]_s=1\big\}.
\end{aligned}
$$

Moreover, let $q_3 := \lceil \log_2(|\mathfrak{C}_{\mathcal{Z}}|) \rceil$ and let $B : \mathfrak{C}_{\mathcal{Z}} \to \{0,1\}^{q_3}$ be an arbitrary injective function. For $s = 1, \ldots, q_3$, let $\mathcal{C}_0^s := \{C \in \mathfrak{C}_{\mathcal{Z}} : [B(C)]_s = 0\}$ and let $\mathcal{C}_1^s := \{C \in \mathfrak{C}_{\mathcal{Z}} : [B(C)]_s = 1\}$. We can then formulate (3.1) into the following mixed-integer linear programming problem:

$$\begin{aligned}
\underset{\substack{x_i,\,\boldsymbol{z},\,(\zeta_t),\,(\iota_{s_1}),\\ (\xi_j),\,(\eta_{s_2}),\,(\beta_{C,\boldsymbol{v}}),\,(\chi_{s_3})}}{\text{minimize}} \quad & \sum_{t=0}^{n_i} l_i(\lambda_{i,t})\zeta_t - \left(\sum_{j=1}^{m_i} y_{i,j}\xi_j\right) - \left(\sum_{C\in\mathfrak{C}_{\mathcal{Z}}} \sum_{\boldsymbol{v}\in V(C)} \tilde{h}^{\boldsymbol{w}_i}(\boldsymbol{v})\beta_{C,\boldsymbol{v}}\right) \\
\text{subject to} \quad & \begin{cases}
\zeta_t \geq 0 & \forall 0 \leq t \leq n_i, \\
\iota_{s_1} \in \{0,1\} & \forall 1 \leq s_1 \leq q_1, \\
\sum_{t=0}^{n_i} \zeta_t = 1, & \\
\sum_{r\in\mathcal{L}_1^{s_1}} \zeta_r \leq \iota_{s_1},\ \sum_{r\in\mathcal{R}_1^{s_1}} \zeta_r \leq (1-\iota_{s_1}) & \forall 1 \leq s_1 \leq q_1, \\
\sum_{t=0}^{n_i} \lambda_{i,t}\zeta_t = x_i - \langle \boldsymbol{s}_i, \boldsymbol{z} \rangle, &
\end{cases} \\
& \begin{cases}
\xi_j \geq 0 & \forall 0 \leq j \leq m_i, \\
\eta_{s_2} \in \{0,1\} & \forall 1 \leq s_2 \leq q_2, \\
\sum_{j=0}^{m_i} \xi_j = 1, & \\
\sum_{r\in\mathcal{L}_2^{s_2}} \xi_r \leq \eta_{s_2},\ \sum_{r\in\mathcal{R}_2^{s_2}} \xi_r \leq (1-\eta_{s_2}) & \forall 1 \leq s_2 \leq q_2, \\
\sum_{j=1}^{m_i} \kappa_{i,j}\xi_j = x_i, &
\end{cases} \\
& \begin{cases}
\beta_{C,\boldsymbol{v}} \geq 0 & \forall \boldsymbol{v} \in V(C),\ \forall C \in \mathfrak{C}_{\mathcal{Z}}, \\
\chi_{s_3} \in \{0,1\} & \forall 1 \leq s_3 \leq q_3, \\
\sum_{C\in\mathfrak{C}_{\mathcal{Z}}} \sum_{\boldsymbol{v}\in V(C)} \beta_{C,\boldsymbol{v}} = 1, & \\
\sum_{C\in\mathcal{C}_1^{s_3}} \sum_{\boldsymbol{v}\in V(C)} \beta_{C,\boldsymbol{v}} \leq \chi_{s_3} & \forall 1 \leq s_3 \leq q_3, \\
\sum_{C\in\mathcal{C}_0^{s_3}} \sum_{\boldsymbol{v}\in V(C)} \beta_{C,\boldsymbol{v}} \leq (1-\chi_{s_3}) & \forall 1 \leq s_3 \leq q_3, \\
\sum_{C\in\mathfrak{C}_{\mathcal{Z}}} \sum_{\boldsymbol{v}\in V(C)} \beta_{C,\boldsymbol{v}}\boldsymbol{v} = \boldsymbol{z}. &
\end{cases}
\end{aligned} \tag{3.3}$$

In the numerical experiment, we have tested both formulations (3.2) and (3.3). The results showed that the formulation (3.3) is faster. Therefore, the experimental results in the paper are based on the formulation (3.3).

## 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:2203.01633v3*, 2022. 5

[2] J. P. Vielma, S. Ahmed, and G. Nemhauser. Mixed-integer models for nonseparable piecewise-linear optimization: unifying framework and extensions. *Oper. Res.*, 58(2):303–315, 2010. 7, 8