FEASIBLE APPROXIMATION OF MATCHING EQUILIBRIA FOR LARGE-SCALE MATCHING FOR TEAMS PROBLEMS

ARIEL NEUFELD AND QIKUN XIANG

ABSTRACT. We propose a numerical algorithm for computing approximately optimal solutions of the matching for teams problem. Our algorithm is efficient for problems involving large number of agent categories and allows for non-discrete agent type measures. Specifically, we parametrize the so-called transfer functions and develop a parametric version of the dual formulation, which we tackle to produce feasible and approximately optimal solutions for the primal and dual formulations. These solutions yield upper and lower bounds for the optimal value, and the difference between these bounds provides a direct sub-optimality estimate of the computed solutions. Moreover, we are able to control the sub-optimality to be arbitrarily close to 0. We subsequently prove that the approximate primal and dual solutions converge when the sub-optimality goes to 0 and their limits constitute a true matching equilibrium. Thus, the outputs of our algorithm are regarded as an approximate matching equilibrium. We also analyze the theoretical computational complexity of our parametric formulation as well as the sparsity of the resulting approximate matching equilibrium. In the numerical experiments, we study three matching for teams problems: a problem of business location distribution, the well-known 2-Wasserstein barycenter problem, and a high-dimensional problem involving 100 agent categories. Through the numerical results, we showcase that the proposed algorithm can produce high-quality approximate matching equilibria in these settings, provide quantitative insights about the optimal city structure in the business location distribution problem, and that the sub-optimality estimates computed by our algorithm are much less conservative than theoretical estimates.

1. INTRODUCTION

The goal of this paper is to provide an algorithm which constructs approximately optimal solutions of the matching for teams problem in theoretical economics involving a large number of agent categories. The matching for teams problem was introduced by Carlier and Ekeland [21], and the setting is described as follows.

Assumption 1.1 (Matching for teams [21, Section 2.4])**.** *We make the following assumptions:*

(A1) *for $i = 1, \dots, N$ (where $N \in \mathbb{N}$), $(\mathcal{X}_i, d_{\mathcal{X}_i})$ is a compact metric space (with metric $d_{\mathcal{X}_i}$) and $\boldsymbol{\mathcal{X}} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$;*
(A2) *$(\mathcal{Z}, d_{\mathcal{Z}})$ is a compact metric space;*
(A3) *for $i = 1, \dots, N$, $\mu_i \in \mathcal{P}(\mathcal{X}_i)$ is a probability measure on $\mathcal{X}_i$;*
(A4) *for $i = 1, \dots, N$, $c_i : \mathcal{X}_i \times \mathcal{Z} \to \mathbb{R}$ is continuous.*

Using the terminologies of Carlier and Ekeland [21], the matching for teams problem describes an economic game involving N categories of agents (e.g., one category of consumer and $N-1$ categories of different producers), where the number of agents in each category is infinite. The spaces $\mathcal{X}_1, \dots, \mathcal{X}_N$, referred to as type spaces, each represents the types of agents from a category. The space $\mathcal{Z}$, referred to as the quality space, represents a type of indivisible good with various qualities. In category i, the distribution of agent types is characterized by the probability measure $\mu_i \in \mathcal{P}(\mathcal{X}_i)$. Moreover, the function $c_i(x_i, z)$ represents the cost for an agent with type $x_i \in \mathcal{X}_i$ to be matched to a unit of good with quality $z \in \mathcal{Z}$. In order for a unit of good with quality $z \in \mathcal{Z}$ to be traded, one agent from each category must come together to form a team and exchange money within the team. The goal is to find a *matching equilibrium* defined as follows.

Key words and phrases. matching for teams; optimal transport; linear semi-infinite optimization.

AN gratefully acknowledges the financial support by his Nanyang Assistant Professorship Grant (NAP Grant) *Machine Learning based Algorithms in Finance and Insurance*.

Definition 1.2 (Matching equilibrium [21, Definition 1]). *Let Assumption 1.1 hold. A matching equilibrium consists of continuous functions $\varphi_1, \dots, \varphi_N : \mathcal{Z} \to \mathbb{R}$, probability measures $\gamma_1, \dots, \gamma_N \in \mathcal{P}(\mathcal{X}_i \times \mathcal{Z})$, and a probability measure $\nu \in \mathcal{P}(\mathcal{Z})$ such that:*

(ME1) *for $i = 1, \dots, N$, $\gamma_i \in \Gamma(\mu_i, \nu)$, where $\Gamma(\mu_i, \nu)$ denotes the set of couplings of μ_i and ν, i.e., $\Gamma(\mu_i, \nu)$ contains all probability measures on $\mathcal{X}_i \times \mathcal{Z}$ whose marginals on $\mathcal{X}_i$ and $\mathcal{Z}$ are μ_i and ν;*

(ME2) *$\sum_{i=1}^N \varphi_i(z) = 0$ for all $z \in \mathcal{Z}$;*

(ME3) *for $i = 1, \dots, N$, $\varphi_i^{c_i}(x_i) + \varphi_i(z) = c_i(x_i, z)$ for γ_i-almost all $(x_i, z) \in \mathcal{X}_i \times \mathcal{Z}$, where $\varphi_i^{c_i} : \mathcal{X}_i \to \mathbb{R}$ is called the c_i-transform of φ_i and is defined by*

$$\varphi_i^{c_i}(x_i) := \inf_{z \in \mathcal{Z}} \left\{ c_i(x_i, z) - \varphi_i(z) \right\} \qquad \forall x_i \in \mathcal{X}_i.$$

In Definition 1.2, $\varphi_i(z)$ represents the amount of money received by an agent of category i when trading a unit of good with quality $z \in \mathcal{Z}$, $\nu \in \mathcal{P}(\mathcal{Z})$ represents the distribution of the qualities of traded goods, and for $i = 1, \dots, N$, $\gamma_i \in \Gamma(\mu_i, \nu)$ describes the matching between agents of category i and qualities of traded goods. The condition (ME1) ensures that every agent is matched to some good. The condition (ME2) is called the balance condition as it requires each team to be self-financed, e.g., all money paid by the consumers will be transferred to the producers. The condition (ME3) requires that an agent of type $x_i \in \mathcal{X}_i$ is matched to a unit of good with quality z only if z minimizes the net cost, i.e., $z \in \arg\min_{z' \in \mathcal{Z}} \left\{ c_i(x_i, z') - \varphi_i(z') \right\}$. In the following, we present two concrete applications of the matching for teams problem that we will analyze in this paper.

Application 1.3 (Equilibrium of business location distribution). *Let us consider the study of the geographic distribution of a type of business in a city by modeling the locations of business outlets as well as the employees' workplace choices as a game involving N agent populations. In this game, the first $N-1$ agent populations represent $N-1$ categories of employees and the N-th agent population represents the business owners. Specifically, the set $\mathcal{Z} \subset \mathbb{R}^2$ represents the locations (in longitude and latitude) in the city where business outlets could possibly be located at. For $i = 1, \dots, N-1$, the set $\mathcal{X}_i \subset \mathbb{R}^2$ represents the locations in the city where the i-th category of employees can reside and $\mu_i \in \mathcal{P}(\mathcal{X}_i)$ represents the distribution of the dwellings of the i-th category of employees. The set $\mathcal{X}_N$ represents the possible locations of the suppliers of the business in the city and $\mu_N \in \mathcal{P}(\mathcal{X}_N)$ denotes their geographic distribution. Moreover, for $i = 1, \dots, N-1$, the cost function $c_i(\cdot, \cdot)$ represents the commuting cost of the i-th category of employees, i.e., $c_i(\boldsymbol{x}_i, \boldsymbol{z})$ denotes the cost of commuting from an employee's home located at $\boldsymbol{x}_i \in \mathcal{X}_i$ to a business outlet located at $\boldsymbol{z} \in \mathcal{Z}$. Furthermore, the cost function $c_N : \mathcal{X}_N \times \mathcal{Z} \to \mathbb{R}$ represents the restocking cost of a business outlet, i.e., $c_N(\boldsymbol{x}_N, \boldsymbol{z})$ denotes the cost of transporting goods from a supplier located at $\boldsymbol{x}_N \in \mathcal{X}_N$ to a business outlet located at $\boldsymbol{z} \in \mathcal{Z}$.*

Under this setting, we are interested in finding a matching equilibrium $(\varphi_i : \mathcal{Z} \to \mathbb{R})_{i=1:N}$, $\big(\gamma_i \in \mathcal{P}(\mathcal{X}_i \times \mathcal{Z})\big)_{i=1:N}$, and $\nu \in \mathcal{P}(\mathcal{Z})$. For $i = 1, \dots, N-1$, $\varphi_i(\boldsymbol{z})$ denotes the amount of salary earned by an employee working at a business outlet located at $\boldsymbol{z} \in \mathcal{Z}$. $\varphi_N(\boldsymbol{z})$ denotes the negative of the total amount of salary paid out by a business outlet located at $\boldsymbol{z} \in \mathcal{Z}$ to the $N-1$ categories of employees. $\nu \in \mathcal{P}(\mathcal{Z})$ describes the geographic distribution of the business outlets in the city and $\gamma_N \in \mathcal{P}(\mathcal{X}_N \times \mathcal{Z})$ describes how the business outlets choose the suppliers to restock from. Moreover, for $i = 1, \dots, N-1$, $\gamma_i \in \mathcal{P}(\mathcal{X}_i \times \mathcal{Z})$ describes where the employees in the i-th category choose to work at depending on where they reside. At equilibrium, the condition $\gamma_i \in \Gamma(\mu_i, \nu)$ for $i = 1, \dots, N-1$ requires each employee to work at some business outlet, and the condition $\gamma_N \in \Gamma(\mu_N, \nu)$ requires each supplier to be supplying some business outlet. The balance condition $\sum_{i=1}^N \varphi_i(\boldsymbol{z}) = 0$ for all $\boldsymbol{z} \in \mathcal{Z}$ ensures that the total amount of salary paid out by each business owner, i.e., $-\varphi_N(\boldsymbol{z})$, is equal to the total amount of salary received by the $N-1$ categories of employees, i.e., $\sum_{i=1}^{N-1} \varphi_i(\boldsymbol{z})$. Finally, for $i = 1, \dots, N$, the condition $\varphi_i^{c_i}(\boldsymbol{x}_i) + \varphi_i(\boldsymbol{z}) = c_i(\boldsymbol{x}_i, \boldsymbol{z})$ for γ_i-almost all $(\boldsymbol{x}_i, \boldsymbol{z}) \in \mathcal{X}_i \times \mathcal{Z}$ states that each employee acts rationally when choosing the workplace, i.e., an employee that resides at $\boldsymbol{x}_i$ minimizes the commuting cost $c_i(\boldsymbol{x}_i, \boldsymbol{z})$ minus the salary $\varphi_i(\boldsymbol{z})$ when deciding the workplace, and each business owner acts rationally when choosing the location of the business outlet, i.e., a business owner that restocks from a supplier at $\boldsymbol{x}_N$ minimizes

the cost $c_N(\boldsymbol{x}_N, \boldsymbol{z})$ of transporting goods plus the total salary $-\varphi_N(\boldsymbol{z})$ paid out to the employees. In this application, the computation of matching equilibria can not only aid the business owners when choosing the locations of business outlets, but also help city planners to improve transportation efficiency in the city.

Application 1.4 (p-Wasserstein barycenter [1])**.** *When $\mathcal{X}_1 = \ldots = \mathcal{X}_N = \mathcal{Z}$ and $c_i(x, z) := \lambda_i d_{\mathcal{Z}}(x, z)^p$ for $i = 1, \ldots, N$ where $p \in [1, \infty)$, $\lambda_i > 0$, and $\sum_{i=1}^N \lambda_i = 1$, an optimizer of* (MT) *is known as a barycenter of $\mu_1, \ldots, \mu_N \in \mathcal{P}(\mathcal{Z})$ in the Wasserstein space of order p with weights $\lambda_1, \ldots, \lambda_N$. The most widely studied setting is the 2-Wasserstein barycenter problem, where $\mathcal{X}_1 = \ldots = \mathcal{X}_N = \mathcal{Z} \subset \mathbb{R}^d$ and $c_i(\boldsymbol{x}, \boldsymbol{z}) := \lambda_i \|\boldsymbol{x} - \boldsymbol{z}\|_2^2$. In recent years, Wasserstein barycenters have found widespread applications in a variety of fields, including statistical inference [13, 50, 62, 63], unsupervised clustering [57, 72, 73], pattern recognition [65], texture mixing [58], color transfer [45, 48], shape interpolation [61, 69], etc.*

In the following, for $i = 1, \ldots, N$, and for $\mu_i \in \mathcal{P}(\mathcal{X}_i)$, $\nu \in \mathcal{P}(\mathcal{Z})$, let $W_{c_i}(\mu_i, \nu)$ denote the optimal transportation cost between μ_i and ν under the cost function $c_i(\cdot, \cdot)$, i.e.,

$$W_{c_i}(\mu_i, \nu) := \inf_{\gamma_i \in \Gamma(\mu_i, \nu)} \left\{ \int_{\mathcal{X}_i \times \mathcal{Z}} c_i(x, z)\, \gamma_i(\mathrm{d}x, \mathrm{d}z) \right\}.$$

Carlier and Ekeland [21] have proved the existence of matching equilibria and characterized them via three optimization problems, as detailed below.

Theorem 1.5 (Existence and characterization of matching equilibria [21, Section 4.2 & Proposition 1 & Theorem 3])**.** *Let Assumption 1.1 hold. Then, the following statements hold.*

(i) There exist continuous functions $\tilde{\varphi}_1, \ldots, \tilde{\varphi}_N : \mathcal{Z} \to \mathbb{R}$, probability measures $\tilde{\gamma}_1, \ldots, \tilde{\gamma}_N \in \mathcal{P}(\mathcal{X}_i \times \mathcal{Z})$, and a probability measure $\tilde{\nu} \in \mathcal{P}(\mathcal{Z})$ that constitute a matching equilibrium.

(ii) $(\tilde{\varphi}_i)_{i=1:N}$, $(\tilde{\gamma}_i)_{i=1:N}$, and $\tilde{\nu}$ are a matching equilibrium if and only if (ME1')–(ME3') *below hold:*

(ME1') *$\tilde{\nu}$ is an optimizer of the following problem:*

$$\inf_{\nu \in \mathcal{P}(\mathcal{Z})} \left\{ \sum_{i=1}^N W_{c_i}(\mu_i, \nu) \right\}; \tag{MT}$$

(ME2') *$(\tilde{\varphi}_i)_{i=1:N}$ is an optimizer of the following problem:*

$$\sup \left\{ \sum_{i=1}^N \int_{\mathcal{X}_i} \varphi_i^{c_i}\, \mathrm{d}\mu_i : (\varphi_i)_{i=1:N} \text{ are continuous}, \sum_{i=1}^N \varphi_i = 0 \right\}; \tag{MT*}$$

(ME3') *for $i = 1, \ldots, N$, $\tilde{\gamma}_i$ is an optimizer of the following problem:*

$$\inf_{\gamma_i \in \Gamma(\mu_i, \tilde{\nu})} \left\{ \int_{\mathcal{X}_i \times \mathcal{Z}} c_i(x, z)\, \gamma_i(\mathrm{d}x, \mathrm{d}z) \right\}. \tag{MT$_{\mathsf{cp}}$}$$

(iii) (MT) *and* (MT*) *have identical optimal values.*

Our objective is to develop a numerical algorithm for efficiently computing feasible and approximately optimal solutions of the problems (MT), (MT*), and (MT$_{\mathsf{cp}}$) when the number N of agent categories is large, and to apply it to the concrete applications discussed above. Moreover, we will show that the computed approximate optimizers, which are referred to as an *approximate matching equilibrium*, converge to a true matching equilibrium when their sub-optimality goes to 0.

Related work. It is well-known that the problem (MT) admits an equivalent multi-marginal optimal transport (MMOT) reformulation, which has been discussed by Carlier and Ekeland [21, Section 6]. There are numerous existing studies about the computation of MMOT and related problems. Many of these studies either only consider discrete measures (see, e.g., [5, 8, 10, 38, 51, 66]) or approximate the problems via discretization of non-discrete measures (see, e.g., [36, 41]). Some studies develop regularization-based methods for approximating MMOT and related problems involving non-discrete measures. These methods typically involve solving an infinite-dimensional optimization problem

parametrized by deep neural networks; see, e.g., [30, 31, 33, 35, 44]. See also [29, 34, 55] for the theoretical properties of entropic regularization and the Sinkhorn algorithm. One downside of neural network based methods is the challenge posed by the non-convexity of the objective function when training neural networks, and there is hence no theoretical guarantee on the quality of the approximate solutions represented by trained neural networks. Recently, Alfonsi, Coyaud, Ehrlacher, and Lombardi [2] and Neufeld and Xiang [54] developed approximation schemes for MMOT problems via relaxation of the marginal constraints into finitely many linear constraints. In particular, Neufeld and Xiang [54] developed a numerical algorithm which is capable of constructing a feasible and approximately optimal solution of the MMOT problem and computing a sub-optimality estimate of the constructed solution. Our numerical approach, however, is tailored to the structure of the problems (MT), (MT*), and (MT$_{\mathsf{cp}}$) without relying on the MMOT formulation.

We would like to point out some existing studies about equilibrium/optimal spatial structure described by measures that are similar to Application 1.3. Lucas and Rossi-Hansberg [52] and Carlier and Ekeland [20] studied the equilibrium structure of a city by analyzing the equilibrium distribution of business and residential districts while considering the positive externality of labor. Buttazzo and Santambrogio [19] and Carlier and Santambrogio [22] considered the optimal structure of a city rather than the equilibrium structure, when taking the congestion effect into account. Blanchet and Carlier [14] analyzed the spatial equilibrium of agents' preferences for land in an economy. Blanchet, Mossay, and Santambrogio [15] used the notion of Cournot–Nash equilibrium to model agents' choices of holiday destinations. Besbes, Castro, and Lobel [12] modeled the equilibrium in the interaction between drivers and customers in a ride-hailing platform.

The 2-Wasserstein barycenter problem has recently become a highly active research area due to its widespread applications discussed in Application 1.4. Most studies about the computation of Wasserstein barycenter focus on the case where $\mu_1, \dots, \mu_N$ are discrete measures with finite support; see, e.g., [3, 4, 7, 16–18, 39, 43, 57, 69–71]. Some notable theoretical results about the computation of discrete 2-Wasserstein barycenter are listed below.

- There exists a sparsely supported 2-Wasserstein barycenter $\hat{\nu}$ of $\mu_1, \dots, \mu_N$ with $|\mathrm{supp}(\hat{\nu})| \leq \sum_{i=1}^N |\mathrm{supp}(\mu_i)| - N + 1$ (where $\mathrm{supp}(\cdot)$ denotes the support of a probability measure); see, e.g., [16, Proposition 1].
- Altschuler and Boix-Adserà [3] showed that there exists a polynomial-time algorithm for the exact computation of discrete 2-Wasserstein barycenter in any fixed dimensions.
- Altschuler and Boix-Adserà [4] showed that the exact computation of discrete 2-Wasserstein barycenter is NP-hard in the dimension of the underlying space $\mathcal{Z}$.

Chizat [25], Luise, Salzo, Pontil, and Ciliberto [53], and Xie, Wang, Wang, and Zha [70] have developed regularization-based methods for approximating discrete 2-Wasserstein barycenter. Moreover, there are also numerical methods for computing 2-Wasserstein barycenter when $\mu_1, \dots, \mu_N$ are continuous. Some of these methods are only applicable to specific families of probability measures, such as Gaussian or the *location-scatter family*; see, e.g., [6, 24]. Some studies consider the case where the measures $\mu_1, \dots, \mu_N$ are unknown with sample access, and develop stochastic optimization algorithms for approximating a 2-Wasserstein barycenter with fixed support; see, e.g., [26, 47, 64, 74]. Recently, numerical methods for continuous 2-Wasserstein barycenter based on neural network parametrization or generative neural networks have been developed; see, e.g., [27, 37, 45, 46, 50]. These methods also suffer from the aforementioned downside of neural network based methods due to the non-convexity of the training objective, posing challenges to the subsequent theoretical analyses.

Carlier, Oberman, and Oudet [23] proposed a numerical method for (MT) with general cost functions $c_1, \dots, c_N$. After discretizing the underlying spaces $\mathcal{X}_1, \dots, \mathcal{X}_N, \mathcal{Z}$, they developed a linear programming approximation of (MT) where the number of decision variables scales linearly with the number N of agent categories. They subsequently proved the convergence of their computed approximate optimizers to an optimizer of (MT). By discretizing $\mathcal{X}_1, \dots, \mathcal{X}_N, \mathcal{Z}$, Carlier et al. [23] also developed another numerical method for approximating (MT*) in the 2-Wasserstein barycenter

case by a non-smooth concave maximization problem. From an optimizer of this non-smooth concave maximization problem, they were able to reconstruct an approximate optimizer of (MT) whose support lies in a pre-specified finite set based on the discretization of $\mathcal{Z}$.

Compared to existing methods for the computation of matching for teams and p-Wasserstein barycenter, our numerical approach is applicable to general cost functions $c_1, \ldots, c_N$ as well as general probability measures $\mu_1, \ldots, \mu_N$ that are not necessarily discrete and not restricted to any family of measures. The approximate optimizers of (MT) constructed through our approach do not have pre-specified support. Thus, our approach belongs to the so-called *free support* approaches. When the measures $\mu_1, \ldots, \mu_N$ are non-discrete, our approach produces two approximate optimizers of (MT): one is a discrete measure with sparse support, and the other is a (possibly) non-discrete measure. Moreover, our approach simultaneously constructs *feasible and approximately optimal solutions* of (MT), (MT*), and (MT$_{\mathsf{cp}}$). The feasibility of these solutions provides us with upper and lower bounds for the optimal value of (MT) and (MT*) that can be computed by our numerical algorithm. Most importantly, the difference between the computed upper and lower bounds corresponds to a sub-optimality bound of the computed solutions that is often much less conservative than sub-optimality bounds obtained through purely theoretical analyses. Furthermore, we also perform analysis about the theoretical computational complexity of our approach as well as the sparsity of the constructed discrete measure.

Contributions and outline of the paper. Specifically, this paper makes the following contributions.

(1) We introduce a parametric formulation of matching for teams that is a linear semi-infinite programming (LSIP) problem. We show that one can construct *feasible approximate optimizers* of the problems (MT), (MT*), and (MT$_{\mathsf{cp}}$) (which are referred to as *approximate matching equilibria*) from an approximate optimizer of the parametric formulation (see Theorem 2.13).

(2) We establish important theoretical results about the aforementioned LSIP problem and the constructed *approximate matching equilibria*, including:
- theoretical computational complexity of the LSIP problem and its dual (see Theorem 2.5),
- the existence of an approximate optimizer of (MT) with sparse support (see Corollary 2.14),
- the convergence of the constructed *approximate matching equilibria* to true matching equilibria (see Theorem 2.16),
- an explicit estimate for the "size" of the parametric formulation in order to control the sub-optimality of the constructed *approximate matching equilibria* when the spaces $\mathcal{X}_1, \ldots, \mathcal{X}_N, \mathcal{Z}$ are compact subsets of Euclidean spaces (see Theorem 2.20).

(3) We develop a numerical algorithm that is able to compute ϵ-*approximate matching equilibria* for any given $\epsilon > 0$ and we analyze its convergence (see Theorem 3.7).

(4) We perform three numerical experiments on problems involving one- and two-dimensional type spaces $\mathcal{X}_1, \ldots, \mathcal{X}_N$ to demonstrate the performance of the proposed algorithm. In all three experiments, we showcase that the sub-optimality bounds computed by our algorithm are much less conservative compared to their purely theoretical bounds, which highlights a practical advantage of the proposed algorithm compared to existing methods for similar problems. Specifically, in the first experiment, we examine an instance of the business location distribution problem in Application 1.3 and we draw concrete insights as well as recommendations from the computed approximate matching equilibria that can aid the city planners in improving the ecomonic efficiency of the city. In the second experiment, we showcase the computation of approximate 2-Wasserstein barycenters (Application 1.4) via our algorithm. Moreover, in the third experiment, we analyze the empirical computational cost of our algorithm and demonstrate that it is capable of solving large problem instances with $N = 100$ agent categories.

The rest of this paper is organized as follows. Section 2 introduces the parametric formulation of the matching for teams problem and the construction of *approximate matching equilibria*. In Section 3, we present the details of the numerical algorithm that we develop as well as its properties. In Section 4, we apply the developed algorithm to the three aforementioned matching for teams problems to demonstrate its performance in practice. The proof of the theoretical results can be found in the appendix of the arXiv version of the paper at https://arxiv.org/abs/2308.03550.

Notions and notations. Throughout this paper, all vectors are assumed to be column vectors. We denote vectors and vector-valued functions by boldface symbols. In particular, for $n \in \mathbb{N}$, we denote by $\mathbf{0}_n$ the vector in $\mathbb{R}^n$ with all entries equal to 0, i.e., $\mathbf{0}_n := (\underbrace{0, \dots, 0}_{n \text{ times}})^\mathsf{T}$. We also use $\mathbf{0}$ when the dimension is unambiguous. We denote by $\langle \cdot, \cdot \rangle$ the Euclidean dot product, i.e., $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \boldsymbol{x}^\mathsf{T} \boldsymbol{y}$, and we denote by $\| \cdot \|_p$ the p-norm of a vector for $p \in [1, \infty]$. A subset of a Euclidean space is called a polyhedron or a polyhedral convex set if it is the intersection of finitely many closed half-spaces. In particular, a subset of a Euclidean space is called a polytope if it is a bounded polyhedron. For a subset A of a Euclidean space, let $\mathrm{aff}(A)$, $\mathrm{conv}(A)$, $\mathrm{cone}(A)$ denote the affine hull, convex hull, and conic hull of A, respectively. Moreover, let $\mathrm{cl}(A)$, $\mathrm{int}(A)$, $\mathrm{relint}(A)$, $\mathrm{relbd}(A)$ denote the closure, interior, relative interior, and relative boundary of A, respectively.

For a Polish space $(\mathcal{Y}, d_\mathcal{Y})$ with its corresponding metric $d_\mathcal{Y}(\cdot, \cdot)$, let $\mathcal{B}(\mathcal{Y})$ denote the Borel subsets of $\mathcal{Y}$ and let $\mathcal{P}(\mathcal{Y})$ denote the set of Borel probability measures on $\mathcal{Y}$. We denote by δ_y the Dirac measure at any $y \in \mathcal{Y}$ and we denote by $\mathrm{supp}(\mu)$ the support of any probability measure $\mu \in \mathcal{P}(\mathcal{Y})$. Moreover, we denote by $\mathcal{C}(\mathcal{Y})$ the set of all continuous functions on $\mathcal{Y}$ and we denote by $\mathcal{L}^1(\mathcal{Y}, \mu)$ the set of μ-integrable functions on $\mathcal{Y}$ with respect to a probability measure $\mu \in \mathcal{P}(\mathcal{Y})$. Furthermore, we use $\Gamma(\cdot, \dots, \cdot)$ to denote the set of couplings of measures, i.e., the set of measures with fixed marginals, as detailed in the following definition.

Definition 1.6 (Coupling)**.** *For $m \in \mathbb{N}$ Polish spaces $(\mathcal{Y}_1, d_{\mathcal{Y}_1}), \dots, (\mathcal{Y}_m, d_{\mathcal{Y}_m})$ and probability measures $\nu_1 \in \mathcal{P}(\mathcal{Y}_1)$, …, $\nu_m \in \mathcal{P}(\mathcal{Y}_m)$, let $\Gamma(\nu_1, \dots, \nu_m)$ denote the set of couplings of $\nu_1, \dots, \nu_m$, defined as*

$$\Gamma(\nu_1, \dots, \nu_m) := \Big\{ \gamma \in \mathcal{P}(\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m) : \textit{the marginal of } \gamma \textit{ on } \mathcal{Y}_j \textit{ is } \nu_j \textit{ for } j = 1, \dots, m \Big\}.$$

For any $\mu, \nu \in \mathcal{P}(\mathcal{Y})$, let $W_1(\mu, \nu)$ denote the Wasserstein metric of order 1 between μ and ν, which is given by

$$W_1(\mu, \nu) := \inf_{\gamma \in \Gamma(\mu, \nu)} \left\{ \int_{\mathcal{Y} \times \mathcal{Y}} d_\mathcal{Y}(x, y) \, \gamma(\mathrm{d}x, \mathrm{d}y) \right\}.$$

2. Approximation of matching for teams

In this section, we develop a parametric version of the problem (MT*) by parametrizing the transfer functions $(\varphi_i)_{i=1:N}$. Specifically, for $i = 1, \dots, N$, we consider a finite set $\mathcal{G}_i = \{g_{i,1}, \dots, g_{i,m_i}\}$ of $m_i \in \mathbb{N}$ continuous functions on $\mathcal{X}_i$, and we consider a finite set $\mathcal{H} = \{h_1, \dots, h_k\}$ of $k \in \mathbb{N}$ continuous functions on $\mathcal{Z}$. Subsequently, we parametrize (MT*) by requiring the transfer functions $(\varphi_i)_{i=1:N}$ to be linear combinations of $\mathcal{H}$ as well as requiring $\sum_{i=1}^N \varphi_i = 0$, and replacing the integrand $\varphi_i^{c_i}$ in the objective of (MT*) with a function $\psi_i : \mathcal{X}_i \to \mathbb{R}$ which is a linear combination of $\mathcal{G}_i$ plus a constant, for $i = 1, \dots, N$. By requiring that $\psi_i(x_i) + \varphi_i(z) \le c_i(x_i, z)$ for all $(x_i, z) \in \mathcal{X}_i \times \mathcal{Z}$ and $i = 1, \dots, N$, we guarantee that $\psi_i \le \varphi_i^{c_i}$ for $i = 1, \dots, N$ and thus this parametric version of (MT*) provides a lower bound for (MT*). Through this parametric formulation, we reduce the decision space of (MT*) from infinite dimensional to finite dimensional, which results in a linear semi-infinite programming (LSIP) problem.

In the following, Section 2.1 introduces the parametric formulation of (MT*). In Section 2.2, we establish the strong duality between the parametric formulation and its dual, which is a minimization problem over probability measures subject to finitely many moment-based constraints. We analyze the theoretical computational complexity of the parametric formulation and its dual in Section 2.3. In Section 2.4, we construct approximate matching equilibria via approximate optimizers of the parametric formulation and its dual and show their convergence towards a matching equilibrium when the approximation error goes to 0. Moreover, we discuss the existence of an approximate optimizer of the dual of the parametric formulation which has sparse support. In Section 2.5, we consider the case where the underlying spaces $\mathcal{X}_1, \dots, \mathcal{X}_N$, and $\mathcal{Z}$ are all Euclidean, and we show that the approximation error of our approach can be controlled to be arbitrarily close to 0 through explicit choices of the functions $\mathcal{G}_1, \dots, \mathcal{G}_N, \mathcal{H}$.

2.1. **The parametric formulation of matching for teams.** We let $\mathcal{G}_i := \{g_{i,1}, \ldots, g_{i,m_i}\}$ be a set of $m_i \in \mathbb{N}$ continuous $\mathbb{R}$-valued functions on $\mathcal{X}_i$, for $i = 1, \ldots, N$, and we let $\mathcal{H} := \{h_1, \ldots, h_k\}$ be a set of $k \in \mathbb{N}$ continuous $\mathbb{R}$-valued functions on $\mathcal{Z}$. The precise choices of the functions $\mathcal{G}_1, \ldots, \mathcal{G}_N, \mathcal{H}$ will be specified later in Section 2.5. For notational simplicity, let the vector-valued functions $\boldsymbol{g}_1 : \mathcal{X}_1 \to \mathbb{R}^{m_1}, \ldots, \boldsymbol{g}_N : \mathcal{X}_N \to \mathbb{R}^{m_N}$, and $\boldsymbol{h} : \mathcal{Z} \to \mathbb{R}^k$ be defined as

$$\boldsymbol{g}_i(x_i) := (g_{i,1}(x_i), \ldots, g_{i,m_i}(x_i))^\mathsf{T} \qquad \forall x_i \in \mathcal{X}_i,\ \forall 1 \le i \le N, \tag{2.1}$$

$$\boldsymbol{h}(z) := (h_1(z), \ldots, h_k(z))^\mathsf{T} \qquad \forall z \in \mathcal{Z}. \tag{2.2}$$

Moreover, let the vectors $\bar{\boldsymbol{g}}_1 \in \mathbb{R}^{m_1}, \ldots, \bar{\boldsymbol{g}}_N \in \mathbb{R}^{m_N}$ be defined as

$$\bar{\boldsymbol{g}}_i := \left(\textstyle\int_{\mathcal{X}_i} g_{i,1} \,\mathrm{d}\mu_i, \ldots, \int_{\mathcal{X}_i} g_{i,m_i} \,\mathrm{d}\mu_i \right)^\mathsf{T} \qquad \forall 1 \le i \le N. \tag{2.3}$$

With these notations, the parametric formulation of (MT*) is given by the following linear semi-infinite programming (LSIP) problem:

$$\begin{array}{lll} \underset{(y_{i,0}, \boldsymbol{y}_i, \boldsymbol{w}_i)}{\text{maximize}} & \displaystyle\sum_{i=1}^N y_{i,0} + \langle \bar{\boldsymbol{g}}_i, \boldsymbol{y}_i \rangle & \\ \text{subject to} & y_{i,0} + \langle \boldsymbol{g}_i(x_i), \boldsymbol{y}_i \rangle + \langle \boldsymbol{h}(z_i), \boldsymbol{w}_i \rangle \le c_i(x_i, z_i) & \\ & \qquad \forall (x_i, z_i) \in \mathcal{X}_i \times \mathcal{Z},\ \forall 1 \le i \le N, & (\mathsf{MT}^*_{\mathsf{par}}) \\ & \displaystyle\sum_{i=1}^N \boldsymbol{w}_i = \boldsymbol{0}_k, & \\ & y_{i,0} \in \mathbb{R},\ \boldsymbol{y}_i \in \mathbb{R}^{m_i},\ \boldsymbol{w}_i \in \mathbb{R}^k \quad \forall 1 \le i \le N. & \end{array}$$

In (MT$^*_{\mathsf{par}}$), we have that $\psi_i(\cdot) := y_{i,0} + \langle \boldsymbol{g}_i(\cdot), \boldsymbol{y}_i \rangle$ and $\varphi_i(\cdot) := \langle \boldsymbol{h}(\cdot), \boldsymbol{w}_i \rangle$ are both continuous for $i = 1, \ldots, N$. The semi-infinite inequality constraint in (MT$^*_{\mathsf{par}}$) requires that $\psi_i(x_i) + \varphi_i(z_i) \le c_i(x_i, z_i)$ for all $(x_i, z_i) \in \mathcal{X}_i \times \mathcal{Z}$, for $i = 1, \ldots, N$. The equality constraint $\sum_{i=1}^N \boldsymbol{w}_i = \boldsymbol{0}_k$ guarantees that $\sum_{i=1}^N \varphi_i = 0$. Thus, one can observe that (MT$^*_{\mathsf{par}}$) provides a lower bound for (MT*).

2.2. **Duality results.** In this subsection, we derive the dual optimization problem of (MT$^*_{\mathsf{par}}$). To begin, let us first present the notion of *moment sets*.

Definition 2.1 (Moment set [54, Definition 2.7])**.** *Let $(\mathcal{Y}, d_\mathcal{Y})$ be a compact metric space. For a collection $\mathcal{G}$ of $\mathbb{R}$-valued Borel measurable functions on $\mathcal{Y}$, let $\mathcal{P}(\mathcal{Y}; \mathcal{G}) := \{\mu \in \mathcal{P}(\mathcal{Y}) : \mathcal{G} \subseteq \mathcal{L}^1(\mathcal{Y}, \mu)\}$. Let $\overset{\mathcal{G}}{\sim}$ be defined as the following equivalence relation on $\mathcal{P}(\mathcal{Y}; \mathcal{G})$: for all $\mu, \nu \in \mathcal{P}(\mathcal{Y}; \mathcal{G})$,*

$$\mu \overset{\mathcal{G}}{\sim} \nu \quad \Leftrightarrow \quad \forall g \in \mathcal{G},\ \int_\mathcal{Y} g \,\mathrm{d}\mu = \int_\mathcal{Y} g \,\mathrm{d}\nu. \tag{2.4}$$

For every $\mu \in \mathcal{P}(\mathcal{Y}; \mathcal{G})$, let $[\mu]_\mathcal{G} := \{\nu \in \mathcal{P}(\mathcal{Y}; \mathcal{G}) : \nu \overset{\mathcal{G}}{\sim} \mu\}$ be the equivalence class of μ under $\overset{\mathcal{G}}{\sim}$. We call $[\mu]_\mathcal{G}$ the moment set centered at μ characterized by test functions $\mathcal{G}$. In addition, let $\overline{W}_{\mu,\mathcal{G}}$ denote the supremum W_1-metric between μ and members of $[\mu]_\mathcal{G}$, and let $\overline{W}_\mathcal{G}$ denote the supremum W_1-metric between any two probability measures that are $\overset{\mathcal{G}}{\sim}$-equivalent, i.e.,

$$\overline{W}_{\mu,\mathcal{G}} := \sup_{\nu \in [\mu]_\mathcal{G}} \{W_1(\mu, \nu)\}, \quad \overline{W}_\mathcal{G} := \sup \{W_1(\nu, \nu') : \nu, \nu' \in \mathcal{P}(\mathcal{Y}; \mathcal{G}),\ \nu \overset{\mathcal{G}}{\sim} \nu'\}.$$

Note that $\overline{W}_{\mu,\mathcal{G}} \le \overline{W}_\mathcal{G} < \infty$ due to the compactness of $\mathcal{Y}$.

The following theorem reveals that the dual optimization problem of (MT$^*_{\mathsf{par}}$) is a relaxation of (MT) through the moment-based equivalence relation (2.4). It also shows that the strong duality holds.

Theorem 2.2 (Strong duality). *Let Assumption 1.1 hold. For $i = 1, \dots, N$, let $m_i \in \mathbb{N}$ and let $\mathcal{G}_i := \{g_{i,1}, \dots, g_{i,m_i}\} \subset \mathcal{C}(\mathcal{X}_i)$. Let $k \in \mathbb{N}$ and let $\mathcal{H} := \{h_1, \dots, h_k\} \subset \mathcal{C}(\mathcal{Z})$. Moreover, let $\boldsymbol{g}_i(\cdot)$, $\boldsymbol{h}(\cdot)$, and $\bar{\boldsymbol{g}}_i$ be given by (2.1), (2.2), and (2.3), respectively. Then, the optimal value of (MT$^*_{\text{par}}$) is equal to the optimal value of the following optimization problem:*

$$\inf \left\{ \sum_{i=1}^{N} \int_{\mathcal{X}_i \times \mathcal{Z}} c_i \, \mathrm{d}\theta_i : \theta_i \in \Gamma(\bar{\mu}_i, \bar{\nu}_i),\ \bar{\mu}_i \overset{\mathcal{G}_i}{\sim} \mu_i,\ \bar{\nu}_i \overset{\mathcal{H}}{\sim} \bar{\nu}_1 \ \forall 1 \le i \le N \right\}. \tag{MT$_{\text{par}}$}$$

The following proposition provides sufficient conditions for the set of optimizers of (MT$^*_{\text{par}}$) to be non-empty and bounded.

Proposition 2.3 (Optimizers of (MT$^*_{\text{par}}$)). *Let Assumption 1.1 hold. For $i = 1, \dots, N$, let $m_i \in \mathbb{N}$ and let $\mathcal{G}_i := \{g_{i,1}, \dots, g_{i,m_i}\} \subset \mathcal{C}(\mathcal{X}_i)$. Let $k \in \mathbb{N}$ and let $\mathcal{H} := \{h_1, \dots, h_k\} \subset \mathcal{C}(\mathcal{Z})$. Moreover, let $\boldsymbol{g}_i(\cdot)$, $\boldsymbol{h}(\cdot)$, and $\bar{\boldsymbol{g}}_i$ be given by (2.1), (2.2), and (2.3), respectively. Then, the following statements hold.*

(i) If $\operatorname{supp}(\mu_i) = \mathcal{X}_i$ for $i = 1, \dots, N$, then the set of optimizers of (MT$^*_{\text{par}}$) is non-empty.

(ii) Suppose that, for $i = 1, \dots, N$, $\operatorname{supp}(\mu_i) = \mathcal{X}_i$ and that there exist $m_i + 1$ points $x_{i,1}, \dots, x_{i,m_i+1} \in \mathcal{X}_i$ such that the $m_i + 1$ vectors $\boldsymbol{g}_i(x_{i,1}), \dots, \boldsymbol{g}_i(x_{i,m_i+1}) \in \mathbb{R}^{m_i}$ are affinely independent. Moreover, suppose that there exist $k + 1$ points $z_1, \dots, z_{k+1} \in \mathcal{Z}$ such that the $k + 1$ vectors $\boldsymbol{h}(z_1), \dots, \boldsymbol{h}(z_{k+1}) \in \mathbb{R}^k$ are affinely independent. Then, the set of optimizers of (MT$^*_{\text{par}}$) is non-empty and bounded.

2.3. **Theoretical computational complexity of (MT$^*_{\text{par}}$) and (MT$_{\text{par}}$).** In this subsection, we analyze the theoretical computational complexity of the LSIP problem (MT$^*_{\text{par}}$) by viewing it as a so-called *convex feasibility* problem. Subsequently, the theoretical computational complexity of (MT$_{\text{par}}$) can also be analyzed. In our analysis, the theoretical computational complexity of (MT$^*_{\text{par}}$) and (MT$_{\text{par}}$) is quantified in terms of the number of calls to an associated global minimization oracle defined as follows.

Definition 2.4 (Global minimization oracle for (MT$^*_{\text{par}}$)). *Let Assumption 1.1 hold. For $i = 1, \dots, N$, let $m_i \in \mathbb{N}$ and $\mathcal{G}_i := \{g_{i,1}, \dots, g_{i,m_i}\} \subset \mathcal{C}(\mathcal{X}_i)$. Let $k \in \mathbb{N}$ and let $\mathcal{H} := \{h_1, \dots, h_k\} \subset \mathcal{C}(\mathcal{Z})$, where $h_l : \mathcal{Z} \to \mathbb{R}$ is non-negative for $l = 1, \dots, k$. Moreover, let $(\boldsymbol{g}_i : \mathcal{X}_i \to \mathbb{R}^{m_i})_{i=1:N}$, $\boldsymbol{h} : \mathcal{Z} \to \mathbb{R}^k$, and $(\bar{\boldsymbol{g}}_i \in \mathbb{R}^{m_i})_{i=1:N}$ be defined in (2.1), (2.2), and (2.3). A procedure $\texttt{Oracle}(\cdot, \cdot, \cdot)$ is called a global minimization oracle for (MT$^*_{\text{par}}$) if, for every $i \in \{1, \dots, N\}$, every $\boldsymbol{y}_i \in \mathbb{R}^{m_i}$, and every $\boldsymbol{w}_i \in \mathbb{R}^k$, a call to $\texttt{Oracle}(i, \boldsymbol{y}_i, \boldsymbol{w}_i)$ returns a minimizer $(x_i^\star, z_i^\star) \in \mathcal{X}_i \times \mathcal{Z}$ of the global minimization problem $\inf_{x_i \in \mathcal{X}_i,\, z_i \in \mathcal{Z}} \{c_i(x_i, z_i) - \langle \boldsymbol{g}_i(x_i), \boldsymbol{y}_i \rangle - \langle \boldsymbol{h}(z_i), \boldsymbol{w}_i \rangle\}$ as well as its corresponding objective value $\beta_i^\star := c_i(x_i^\star, z_i^\star) - \langle \boldsymbol{g}_i(x_i^\star), \boldsymbol{y}_i \rangle - \langle \boldsymbol{h}(z_i^\star), \boldsymbol{w}_i \rangle$.*

The following theorem states that, under some mild conditions, there exists an algorithm for solving both (MT$^*_{\text{par}}$) and (MT$_{\text{par}}$) whose computational complexity is polynomial in N, m, k, and the computational cost of each call to $\texttt{Oracle}(\cdot, \cdot, \cdot)$. In the theorem, we denote the computational complexity of the multiplication of two $m \times m$ matrices by $O(m^\omega)$. For example, with the standard matrix multiplication procedure, the computational complexity of this operation is $O(m^3)$. However, it is known that $\omega < 2.376$; see, e.g., [28].

Theorem 2.5 (Theoretical computational complexity of (MT$^*_{\text{par}}$) and (MT$_{\text{par}}$)). *Let Assumption 1.1 hold. For $i = 1, \dots, N$, let $m_i \in \mathbb{N}$ and $\mathcal{G}_i := \{g_{i,1}, \dots, g_{i,m_i}\} \subset \mathcal{C}(\mathcal{X}_i)$. Let $k \in \mathbb{N}$ and let $\mathcal{H} := \{h_1, \dots, h_k\} \subset \mathcal{C}(\mathcal{Z})$, where $h_l : \mathcal{Z} \to \mathbb{R}$ is non-negative for $l = 1, \dots, k$. Let $m := \sum_{i=1}^N m_i$ and let $(\boldsymbol{g}_i : \mathcal{X}_i \to \mathbb{R}^{m_i})_{i=1:N}$, $\boldsymbol{h} : \mathcal{Z} \to \mathbb{R}^k$, and $(\bar{\boldsymbol{g}}_i \in \mathbb{R}^{m_i})_{i=1:N}$ be defined in (2.1), (2.2), and (2.3). Let $\texttt{Oracle}(\cdot, \cdot, \cdot)$ be the global minimization oracle in Definition 2.4. Suppose that[1], for $i = 1, \dots, N$, $\|\boldsymbol{g}_i(x_i)\|_2 \le 1$ for all $x_i \in \mathcal{X}_i$, and that $\|\boldsymbol{h}(z)\|_2 \le 1$ for all*

[1]Since $\mathcal{X}_i$ is compact and $g_{i,1}, \dots, g_{i,m_i}$ are continuous, one may replace $g_{i,j}$ by $\max_{x_i \in \mathcal{X}_i} \{\|\boldsymbol{g}_i(x_i)\|_2\}^{-1} g_{i,j}$ for $j = 1, \dots, m_i$ to guarantee that $\|\boldsymbol{g}_i(x_i)\|_2 \le 1$ for all $x_i \in \mathcal{X}_i$. The same can be done to $h_1, \dots, h_k$ to guarantee that $\|\boldsymbol{h}(z)\|_2 \le 1$ for all $z \in \mathcal{Z}$. Due to the linearity of the objective and the constraints of (MT$^*_{\text{par}}$), this rescaled version of (MT$^*_{\text{par}}$) is equivalent to the original problem without rescaling.

$z \in \mathcal{Z}$. Moreover, suppose that ($\mathsf{MT}^*_{\mathsf{par}}$) has an optimizer $(y^\star_{1,0}, \boldsymbol{y}_1^{\star\mathsf{T}}, \boldsymbol{w}_1^{\star\mathsf{T}}, \dots, y^\star_{N,0}, \boldsymbol{y}_N^{\star\mathsf{T}}, \boldsymbol{w}_N^{\star\mathsf{T}})^\mathsf{T}$ and let $M_{\mathsf{opt}} := \|(y^\star_{1,0}, \boldsymbol{y}_1^{\star\mathsf{T}}, \boldsymbol{w}_1^{\star\mathsf{T}}, \dots, y^\star_{N,0}, \boldsymbol{y}_N^{\star\mathsf{T}}, \boldsymbol{w}_N^{\star\mathsf{T}})^\mathsf{T}\|_2$. Furthermore, let T denote the computational cost of each call to $\texttt{Oracle}(\cdot,\cdot,\cdot)$ and let $\epsilon_{\mathsf{LSIP}} > 0$ be an arbitrary positive tolerance value. Then, the following statements hold.

- *(i) There exists an algorithm which computes an ϵ_{LSIP}-optimizer of ($\mathsf{MT}^*_{\mathsf{par}}$) with computational complexity $O\big((m+Nk)\log((m+Nk)M_{\mathsf{opt}}/\epsilon_{\mathsf{LSIP}})(NT+(m+Nk)^\omega)\big)$.*
- *(ii) Suppose that there exist $M_{\mathsf{max}} > 0$ and finite sets $\mathcal{C}_1 \subseteq \mathcal{X}_1 \times \mathcal{Z}, \dots, \mathcal{C}_N \subseteq \mathcal{X}_N \times \mathcal{Z}$, such that for every $(y_{1,0}, \boldsymbol{y}_1^\mathsf{T}, \boldsymbol{w}_1^\mathsf{T}, \dots, y_{N,0}, \boldsymbol{y}_N^\mathsf{T}, \boldsymbol{w}_N^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{m+N(k+1)}$ satisfying $y_{i,0} + \langle \boldsymbol{g}_i(x_i), \boldsymbol{y}_i\rangle + \langle \boldsymbol{h}(z_i), \boldsymbol{w}_i\rangle \le c_i(x_i,z_i)\ \forall (x_i,z_i) \in \mathcal{C}_i$ for $i = 1,\dots,N$, $\sum_{i=1}^N \boldsymbol{w}_i = \mathbf{0}_k$, and $\sum_{i=1}^N y_{i,0} + \langle \bar{\boldsymbol{g}}_i, \boldsymbol{y}_i\rangle \ge \sum_{i=1}^N y^\star_{i,0} + \langle \bar{\boldsymbol{g}}_i, \boldsymbol{y}_i^\star\rangle$, it holds that $\big\|(y_{1,0}, \boldsymbol{y}_1^\mathsf{T}, \boldsymbol{w}_1^\mathsf{T}, \dots, y_{N,0}, \boldsymbol{y}_N^\mathsf{T}, \boldsymbol{w}_N^\mathsf{T})^\mathsf{T}\big\|_2 \le M_{\mathsf{max}}$. Let $\mathrm{poly}\big(N, m, k, \log\big(\frac{M_{\mathsf{max}}}{\epsilon_{\mathsf{LSIP}}}\big)\big)$ denote the collection of functions that are asymptotically polynomial in N, m, k, and $\log\big(\frac{M_{\mathsf{max}}}{\epsilon_{\mathsf{LSIP}}}\big)$. Moreover, suppose that $\sum_{i=1}^N |\mathcal{C}_i| = \mathrm{poly}\big(N, m, k, \log\big(\frac{M_{\mathsf{max}}}{\epsilon_{\mathsf{LSIP}}}\big)\big)$. Then, there exists an algorithm which computes a pair of ϵ_{LSIP}-optimizers of ($\mathsf{MT}^*_{\mathsf{par}}$) and ($\mathsf{MT}_{\mathsf{par}}$) with $\mathrm{poly}\big(N, m, k, \log\big(\frac{M_{\mathsf{max}}}{\epsilon_{\mathsf{LSIP}}}\big)\big)$ many calls to $\texttt{Oracle}(\cdot,\cdot,\cdot)$ and $\mathrm{poly}\big(N, m, k, \log\big(\frac{M_{\mathsf{max}}}{\epsilon_{\mathsf{LSIP}}}\big)\big)$ additional computational time.*

Remark 2.6. *Recall that Proposition 2.3(i) has provided a sufficient condition to guarantee the existence of an optimizer $(y^\star_{1,0}, \boldsymbol{y}_1^{\star\mathsf{T}}, \boldsymbol{w}_1^{\star\mathsf{T}}, \dots, y^\star_{N,0}, \boldsymbol{y}_N^{\star\mathsf{T}}, \boldsymbol{w}_N^{\star\mathsf{T}})^\mathsf{T}$ of ($\mathsf{MT}^*_{\mathsf{par}}$). Moreover, recall that Proposition 2.3(ii) has provided sufficient conditions to guarantee the non-emptiness and boundedness of the set of optimizers of ($\mathsf{MT}^*_{\mathsf{par}}$). Under these conditions, the existence of the constant $M_{\mathsf{max}} > 0$ and the finite sets $\mathcal{C}_1 \subseteq \mathcal{X}_1 \times \mathcal{Z}, \dots, \mathcal{C}_N \subseteq \mathcal{X}_N \times \mathcal{Z}$ in Theorem 2.5(ii) follows from the equivalence between (i) and (iii) in [40, Corollary 9.3.1].*

Remark 2.7. *Notice that $\texttt{Oracle}(\cdot,\cdot,\cdot)$ performs minimization over $\mathcal{X}_i \times \mathcal{Z}$, and thus when k and $m_1, \dots, m_N$ do not depend on the number N of agent categories, the computational complexity of $\texttt{Oracle}(\cdot,\cdot,\cdot)$ is independent of N and the theoretical computational complexity of ($\mathsf{MT}^*_{\mathsf{par}}$) is polynomial in N. This will be numerically tested and verified in Section 4.3 in an experiment.*

Remark 2.8. *In Theorem 2.5, the dependence of the constants M_{opt}, M_{max} on N, m, and k is not studied. The analysis of this dependence as well as a specific choice of the finite sets $\mathcal{C}_1, \dots, \mathcal{C}_N$ in Theorem 2.5(ii) will be presented later in Proposition 2.23 under more specific assumptions on the spaces $\mathcal{X}_1, \dots, \mathcal{X}_N, \mathcal{Z}$, the cost functions $c_1, \dots, c_N$, and the test functions $\mathcal{G}_1, \dots, \mathcal{G}_N, \mathcal{H}$.*

2.4. **Construction and convergence of approximate matching equilibria.** In this subsection, we show how approximate matching equilibria can be constructed from approximate optimizers of ($\mathsf{MT}^*_{\mathsf{par}}$) and ($\mathsf{MT}_{\mathsf{par}}$) and we show their convergence to a true matching equilibrum. The construction requires an operation on $\mathcal{P}(\mathcal{X}_i \times \mathcal{Z})$ called *reassembly* [54, Definition 2.4], which is a direct consequence of the gluing lemma of probability measures (see, e.g., [68, Lemma 7.6]). Moreover, we also need an operation on a collection of probability measures that is called *binding*. These two operations are presented in the following definitions.

Definition 2.9 (Reassembly (see [54, Definition 2.4]))**.** *Let Assumption 1.1 hold and let $\nu \in \mathcal{P}(\mathcal{Z})$. For any $i \in \{1,\dots,N\}$ and any $\hat{\theta}_i \in \mathcal{P}(\mathcal{X}_i \times \mathcal{Z})$, let its marginal on $\mathcal{X}_i$ and $\mathcal{Z}$ be denoted by $\hat{\mu}_i$ and $\hat{\nu}_i$, respectively. Let $\bar{\mathcal{X}}_i := \mathcal{X}_i$ and let $\bar{\mathcal{Z}} := \mathcal{Z}$ in order to differentiate different copies of the same space. $\tilde{\theta}_i \in \mathcal{P}(\mathcal{X}_i \times \mathcal{Z})$ is called a reassembly of $\hat{\theta}_i$ with marginals μ_i and ν if there exists $\gamma \in \mathcal{P}(\mathcal{X}_i \times \mathcal{Z} \times \bar{\mathcal{X}}_i \times \bar{\mathcal{Z}})$ which satisfies the following conditions:*

- *(i) the marginal of γ on $\mathcal{X}_i \times \mathcal{Z}$ is $\hat{\theta}_i$;*
- *(ii) the marginal of γ on $\mathcal{X}_i \times \bar{\mathcal{X}}_i$, denoted by η_i, is an optimal coupling of $\hat{\mu}_i$ and μ_i under the cost function $d_{\mathcal{X}_i}$, i.e., $\eta_i \in \Gamma(\hat{\mu}_i, \mu_i)$ satisfies $\int_{\mathcal{X}_i \times \bar{\mathcal{X}}_i} d_{\mathcal{X}_i}(x_i, \bar{x}_i)\, \eta_i(\mathrm{d}x_i, \mathrm{d}\bar{x}_i) = W_1(\hat{\mu}_i, \mu_i)$;*
- *(iii) the marginal of γ on $\mathcal{Z} \times \bar{\mathcal{Z}}$, denoted by ζ_i, is an optimal coupling of $\hat{\nu}_i$ and ν under the cost function $d_{\mathcal{Z}}$, i.e., $\zeta_i \in \Gamma(\hat{\nu}_i, \nu)$ satisfies $\int_{\mathcal{Z} \times \bar{\mathcal{Z}}} d_{\mathcal{Z}}(z, \bar{z})\, \zeta_i(\mathrm{d}z, \mathrm{d}\bar{z}) = W_1(\hat{\nu}_i, \nu)$;*
- *(iv) the marginal of γ on $\bar{\mathcal{X}}_i \times \bar{\mathcal{Z}}$ is $\tilde{\theta}_i$.*

Let $R(\hat{\theta}_i;\mu_i,\nu)\subset\Gamma(\mu_i,\nu)$ denote the set of reassemblies of $\hat{\theta}_i$ with marginals μ_i and ν. In particular, $R(\hat{\theta}_i;\mu_i,\nu)$ is non-empty, as shown by [54, Lemma 2.5].

Definition 2.10 (Binding)**.** *Let Assumption 1.1 hold and let $\nu\in\mathcal{P}(\mathcal{Z})$. For $i=1,\dots,N$, let $\gamma_i\in\mathcal{P}(\mathcal{X}_i\times\mathcal{Z})$ be such that the marginal of γ_i on $\mathcal{Z}$ is ν. Then, $\tilde{\mu}\in\mathcal{P}(\mathcal{X}_1\times\cdots\times\mathcal{X}_N)$ is called a binding of $\gamma_1,\dots,\gamma_N$ if there exists $\gamma\in\mathcal{P}(\mathcal{X}_1\times\cdots\times\mathcal{X}_N\times\mathcal{Z})$ which satisfies the following conditions:*

(i) for $i=1,\dots,N$, the marginal of γ on $\mathcal{X}_i\times\mathcal{Z}$ is γ_i;

(ii) the marginal of γ on $\mathcal{X}_1\times\cdots\times\mathcal{X}_N$ is $\tilde{\mu}$.

Let $B(\gamma_1,\dots,\gamma_N)$ denote the set of bindings of $\gamma_1,\dots,\gamma_N$.

The following lemma shows that the set of bindings is non-empty.

Lemma 2.11. *Let Assumption 1.1 hold and let $\nu\in\mathcal{P}(\mathcal{Z})$. For $i=1,\dots,N$, let $\gamma_i\in\mathcal{P}(\mathcal{X}_i\times\mathcal{Z})$ be such that the marginal of γ_i on $\mathcal{Z}$ is ν. Then, there exists a binding $\tilde{\mu}\in B(\gamma_1,\dots,\gamma_N)$ of $\gamma_1,\dots,\gamma_N$.*

Next, let us define the function $\bar{c}:\mathcal{X}_1\times\cdots\times\mathcal{X}_N\to\mathbb{R}$ by:

$$\bar{c}(x_1,\dots,x_N):=\inf_{z\in\mathcal{Z}}\left\{\sum_{i=1}^N c_i(x_i,z)\right\}\qquad\forall x_1\in\mathcal{X}_1,\ \dots,\ \forall x_N\in\mathcal{X}_N. \tag{2.5}$$

Due to the continuity of $c_1,\dots,c_N$ and the compactness of $\mathcal{Z}$, it follows from [11, Proposition 7.32 & Proposition 7.33] that $\bar{c}$ is continuous and there exists a Borel measurable function $\tilde{z}:\boldsymbol{\mathcal{X}}\to\mathcal{Z}$ such that

$$\sum_{i=1}^N c_i\big(x_i,\tilde{z}(x_1,\dots,x_N)\big)=\bar{c}(x_1,\dots,x_N)\qquad\forall x_1\in\mathcal{X}_1,\ \dots,\ \forall x_N\in\mathcal{X}_N. \tag{2.6}$$

In the following, in order to control the approximation error of ($\mathsf{MT}^*_{\mathsf{par}}$) and ($\mathsf{MT}_{\mathsf{par}}$), we impose the assumption that the cost functions $c_1,\dots,c_N$ are Lipschitz continuous. Thus, we extend Assumption 1.1 as follows.

Assumption 2.12. *In addition to (A1)–(A4) in Assumption 1.1, we assume that:*

(A4+) *For $i=1,\dots,N$, there exist constants $L^{(1)}_{c_i}>0$ and $L^{(2)}_{c_i}>0$ such that $c_i:\mathcal{X}_i\times\mathcal{Z}\to\mathbb{R}$ satisfies $|c_i(x,z)-c_i(x',z')|\le L^{(1)}_{c_i}d_{\mathcal{X}_i}(x,x')+L^{(2)}_{c_i}d_{\mathcal{Z}}(z,z')$ for all $x,x'\in\mathcal{X}_i$, $z,z'\in\mathcal{Z}$.*

The construction of approximate matching equilibria through parametrizing transfer functions is detailed in the following theorem.

Theorem 2.13 (Approximation of matching for teams)**.** *Let Assumption 2.12 hold. For $i=1,\dots,N$, let $m_i\in\mathbb{N}$ and let $\mathcal{G}_i:=\{g_{i,1},\dots,g_{i,m_i}\}\subset\mathcal{C}(\mathcal{X}_i)$. Let $k\in\mathbb{N}$ and let $\mathcal{H}:=\{h_1,\dots,h_k\}\subset\mathcal{C}(\mathcal{Z})$. Let $\boldsymbol{g}_i(\cdot)$, $\boldsymbol{h}(\cdot)$, and $\bar{\boldsymbol{g}}_i$ be given by (2.1), (2.2), and (2.3), respectively. Moreover, let $\epsilon_{\mathsf{LSIP}}>0$ be arbitrary, let $\epsilon_{\mathsf{approx}}:=\epsilon_{\mathsf{LSIP}}+(N-1)\max_{1\le i\le N}\{L^{(2)}_{c_i}\}\overline{W}_{\mathcal{H}}+\sum_{i=1}^N L^{(1)}_{c_i}\overline{W}_{\mu_i,\mathcal{G}_i}$, and let $(\hat{y}_{i,0},\hat{\boldsymbol{y}}_i,\hat{\boldsymbol{w}}_i)_{i=1:N}$ and $(\hat{\theta}_i)_{i=1:N}$ be feasible solutions of ($\mathsf{MT}^*_{\mathsf{par}}$) and ($\mathsf{MT}_{\mathsf{par}}$) that satisfy*[2]

$$\sum_{i=1}^N\int_{\mathcal{X}_i\times\mathcal{Z}}c_i\,\mathrm{d}\hat{\theta}_i\le\left(\sum_{i=1}^N\hat{y}_{i,0}+\langle\bar{\boldsymbol{g}}_i,\hat{\boldsymbol{y}}_i\rangle\right)+\epsilon_{\mathsf{LSIP}}. \tag{2.7}$$

Furthermore, for $i=1,\dots,N$, let $\hat{\mu}_i$ and $\hat{\nu}_i$ denote the marginals of $\hat{\theta}_i$ on $\mathcal{X}_i$ and $\mathcal{Z}$, respectively, and let $\hat{\nu}\in\mathcal{P}(\mathcal{Z})$ satisfy[3]

$$\sum_{i=1}^N W_1(\hat{\nu},\hat{\nu}_i)\le(N-1)\overline{W}_{\mathcal{H}}. \tag{2.8}$$

Let $\tilde{z}:\boldsymbol{\mathcal{X}}\to\mathcal{Z}$ be a Borel measurable function satisfying (2.6). Then, the following statements hold.

(i) $\sum_{i=1}^N\hat{y}_{i,0}+\langle\bar{\boldsymbol{g}}_i,\hat{\boldsymbol{y}}_i\rangle$ is a lower bound for the optimal value of (MT).

[2] In particular, $(\hat{y}_{i,0},\hat{\boldsymbol{y}}_i,\hat{\boldsymbol{w}}_i)_{i=1:N}$ is an ϵ_{LSIP}-optimizer of ($\mathsf{MT}^*_{\mathsf{par}}$) and $(\hat{\theta}_i)_{i=1:N}$ is an ϵ_{LSIP}-optimizer of ($\mathsf{MT}_{\mathsf{par}}$).

[3] A sufficient condition for (2.8) to hold is when $\hat{\nu}=\hat{\nu}_{\hat{i}}$ for some $\hat{i}\in\{1,\dots,N\}$.

(ii) *Let $z_0 \in \mathcal{Z}$ be arbitrary and let*

$$\begin{aligned}
\tilde{\varphi}_{i,0} &:= \inf_{x_i \in \mathcal{X}_i} \left\{ c_i(x_i, z_0) - \hat{y}_{i,0} - \langle \boldsymbol{g}_i(x_i), \hat{\boldsymbol{y}}_i \rangle \right\} && \forall 1 \le i \le N-1, \\
\tilde{\varphi}_i(z) &:= \inf_{x_i \in \mathcal{X}_i} \left\{ c_i(x_i, z) - \hat{y}_{i,0} - \langle \boldsymbol{g}_i(x_i), \hat{\boldsymbol{y}}_i \rangle \right\} - \tilde{\varphi}_{i,0} && \forall z \in \mathcal{Z},\ \forall 1 \le i \le N-1, \\
\tilde{\varphi}_N(z) &:= -\sum_{i=1}^{N-1} \tilde{\varphi}_i(z) && \forall z \in \mathcal{Z}.
\end{aligned} \tag{2.9}$$

Then, $(\tilde{\varphi}_i)_{i=1:N}$ is an $\epsilon_{\mathsf{approx}}$-optimizer of (MT). Moreover, for $i = 1, \dots, N-1$, $\tilde{\varphi}_i$ is $L^{(2)}_{c_i}$-Lipschitz continuous.*

(iii) *$\hat{\nu}$ is an $\epsilon_{\mathsf{approx}}$-optimizer of (MT).*

(iv) *For $i = 1, \dots, N$, let $\hat{\gamma}_i \in R(\hat{\theta}_i; \mu_i, \hat{\nu})$. Then, $\hat{\gamma}_i \in \Gamma(\mu_i, \hat{\nu})$ and $\int_{\mathcal{X}_i \times \mathcal{Z}} c_i \, \mathrm{d}\hat{\gamma}_i \le W_{c_i}(\mu_i, \hat{\nu}) + \epsilon_{\mathsf{approx}}$.*

(v) *Let $(\hat{\gamma}_i)_{i=1:N}$ be defined as in statement (iv), let $\tilde{\mu} \in B(\hat{\gamma}_1, \dots, \hat{\gamma}_N)$, and let $\tilde{\nu} := \tilde{\mu} \circ \tilde{z}^{-1}$. Then, $\tilde{\nu}$ is an $\epsilon_{\mathsf{approx}}$-optimizer of (MT).*

(vi) *Let $\tilde{\mu}$, $\tilde{\nu}$ be defined as in statement (v) and let $\tilde{\gamma}_i := \tilde{\mu} \circ (\pi_i, \tilde{z})^{-1}$ for $i = 1, \dots, N$, where $\pi_i : \boldsymbol{\mathcal{X}} \to \mathcal{X}_i$ denotes the projection function onto $\mathcal{X}_i$. Then, $\tilde{\gamma}_i \in \Gamma(\mu_i, \tilde{\nu})$ and $\int_{\mathcal{X}_i \times \mathcal{Z}} c_i \, \mathrm{d}\tilde{\gamma}_i \le W_{c_i}(\mu_i, \tilde{\nu}) + \epsilon_{\mathsf{approx}}$.*

By observing the connection between Theorem 2.13(ii)–(vi) and the characterization of matching equilibria in (ME1')–(ME3'), we refer to $\left((\tilde{\varphi}_i)_{i=1:N}, (\hat{\gamma}_i)_{i=1:N}, \hat{\nu}\right)$ and $\left((\tilde{\varphi}_i)_{i=1:N}, (\tilde{\gamma}_i)_{i=1:N}, \tilde{\nu}\right)$ as $\epsilon_{\mathsf{approx}}$-approximate matching equilibria.

As a consequence of Theorem 2.13, one can construct an approximate optimizer of (MT) which is supported on at most $\min_{1 \le i \le N}\{m_i\} + k + 2$ points via the parametric formulation. This is summarized in the following corollary.

Corollary 2.14 (Approximate optimizer of (MT) with sparse support)**.** *Let the assumptions of Theorem 2.13 hold. Then, there exist $q \in \mathbb{N}$ with $1 \le q \le \min_{1 \le i \le N}\{m_i\} + k + 2$, $\alpha_1 > 0, \dots, \alpha_q > 0$ satisfying $\sum_{l=1}^{q} \alpha_l = 1$, $z_1 \in \mathcal{Z}, \dots, z_q \in \mathcal{Z}$, such that $\hat{\nu} := \sum_{l=1}^{q} \alpha_l \delta_{z_l} \in \mathcal{P}(\mathcal{Z})$ is an $\epsilon_{\mathsf{approx}}$-optimizer of (MT).*

Remark 2.15. *As shown by Corollary 2.14, $\hat{\nu} \in \mathcal{P}(\mathcal{Z})$ in Theorem 2.13(iii) can be chosen to be a discrete probability measure with sparse support. In contrast, $\tilde{\nu} \in \mathcal{P}(\mathcal{Z})$ in Theorem 2.13(v) can be a non-discrete probability measure even when $\hat{\nu}$ is discrete, due to the presence of the reassembly and binding steps. A discrete probability measure $\hat{\nu} \in \mathcal{P}(\mathcal{Z})$ can be interpreted as an approximate matching equilibrium in which agents only trade finitely many distinct types of goods. On the other hand, a non-discrete probability measure $\tilde{\nu} \in \mathcal{P}(\mathcal{Z})$ can be interpreted as an approximate matching equilibrium in which agents trade uncountably many types of goods.*

The notion of $\epsilon_{\mathsf{approx}}$-approximate matching equilibrium is justified since when given a sequence of $\epsilon^{(l)}_{\mathsf{approx}}$-approximate matching equilibria constructed as above where $\lim_{l \to \infty} \epsilon^{(l)}_{\mathsf{approx}} = 0$, one can extract a subsequence that converges to a true matching equilibrium. This is detailed in the next theorem.

Theorem 2.16 (Construction of matching equilibria)**.** *Let Assumption 2.12 hold. Let $\left(\epsilon^{(l)}_{\mathsf{LSIP}} \in (0, \infty)\right)_{l \in \mathbb{N}}$, let $\left(\mathcal{G}^{(l)}_i\right)_{i=1:N,\, l \in \mathbb{N}}$ and $\left(\mathcal{H}^{(l)}\right)_{l \in \mathbb{N}}$ be collections of continuous functions, and let*

$$\epsilon^{(l)}_{\mathsf{approx}} := \epsilon^{(l)}_{\mathsf{LSIP}} + (N-1) \max_{1 \le i \le N} \left\{ L^{(2)}_{c_i} \right\} \overline{W}_{\mathcal{H}^{(l)}} + \sum_{i=1}^{N} L^{(1)}_{c_i} \overline{W}_{\mu_i, \mathcal{G}^{(l)}_i}$$

satisfy $\lim_{l \to \infty} \epsilon^{(l)}_{\mathsf{approx}} = 0$. Moreover, for each $l \in \mathbb{N}$, let $\left(\tilde{\varphi}^{(l)}_i\right)_{i=1:N}$, $\hat{\nu}^{(l)}$, $\left(\hat{\gamma}^{(l)}_i\right)_{i=1:N}$, $\tilde{\nu}^{(l)}$, and $\left(\tilde{\gamma}^{(l)}_i\right)_{i=1:N}$ be constructed in Theorem 2.13 (with $(\mathcal{G}_i)_{i=1:N} \leftarrow \left(\mathcal{G}^{(l)}_i\right)_{i=1:N}$, $\mathcal{H} \leftarrow \mathcal{H}^{(l)}$, and $\epsilon_{\mathsf{LSIP}} \leftarrow \epsilon^{(l)}_{\mathsf{LSIP}}$). Then, the following statements hold.

(i) *For* $i = 1, \dots, N$, $\left(\tilde{\varphi}_i^{(l)}\right)_{l\in\mathbb{N}}$ *has at least one accumulation point in* $\mathcal{C}(\mathcal{Z})$ *with respect to the metric of uniform convergence.*

(ii) $\left(\hat{\nu}^{(l)}\right)_{l\in\mathbb{N}}$ *has at least one accumulation point in* $\left(\mathcal{P}(\mathcal{Z}), W_1\right)$ *and for* $i = 1, \dots, N$, $\left(\hat{\gamma}_i^{(l)}\right)_{l\in\mathbb{N}}$ *has at least one accumulation point in* $\left(\mathcal{P}(\mathcal{X}_i \times \mathcal{Z}), W_1\right)$.

(iii) $\left(\tilde{\nu}^{(l)}\right)_{l\in\mathbb{N}}$ *has at least one accumulation point in* $\left(\mathcal{P}(\mathcal{Z}), W_1\right)$ *and for* $i = 1, \dots, N$, $\left(\tilde{\gamma}_i^{(l)}\right)_{l\in\mathbb{N}}$ *has at least one accumulation point in* $\left(\mathcal{P}(\mathcal{X}_i \times \mathcal{Z}), W_1\right)$.

Furthermore, let $(l_k)_{k\in\mathbb{N}} \subseteq \mathbb{N}$ *be an increasing subsequence such that* $\left(\hat{\nu}^{(l_k)}\right)_{k\in\mathbb{N}}$ *converges in* $\left(\mathcal{P}(\mathcal{Z}), W_1\right)$ *to* $\hat{\nu}^{(\infty)}$, $\left(\tilde{\nu}^{(l_k)}\right)_{k\in\mathbb{N}}$ *converges in* $\left(\mathcal{P}(\mathcal{Z}), W_1\right)$ *to* $\tilde{\nu}^{(\infty)}$, *and for* $i = 1, \dots, N$, $\left(\tilde{\varphi}_i^{(l_k)}\right)_{k\in\mathbb{N}}$ *converges uniformly to* $\tilde{\varphi}_i^{(\infty)} \in \mathcal{C}(\mathcal{Z})$, $\left(\hat{\gamma}_i^{(l_k)}\right)_{k\in\mathbb{N}}$ *converges in* $\left(\mathcal{P}(\mathcal{X}_i \times \mathcal{Z}), W_1\right)$ *to* $\hat{\gamma}_i^{(\infty)}$, *and* $\left(\tilde{\gamma}_i^{(l_k)}\right)_{k\in\mathbb{N}}$ *converges in* $\left(\mathcal{P}(\mathcal{X}_i \times \mathcal{Z}), W_1\right)$ *to* $\tilde{\gamma}_i^{(\infty)}$. *Then,*

(iv) $\left(\tilde{\varphi}_i^{(\infty)}\right)_{i=1:N}$, $\left(\hat{\gamma}_i^{(\infty)}\right)_{i=1:N}$, $\hat{\nu}^{(\infty)}$ *constitute a matching equilibrium,*

(v) $\left(\tilde{\varphi}_i^{(\infty)}\right)_{i=1:N}$, $\left(\tilde{\gamma}_i^{(\infty)}\right)_{i=1:N}$, $\tilde{\nu}^{(\infty)}$ *constitute a matching equilibrium.*

In order to control the approximation error $\epsilon_{\mathsf{approx}}$ in Theorem 2.13 to be arbitrarily close to 0, we need to explicitly construct the test functions $\mathcal{G}_1, \dots, \mathcal{G}_N, \mathcal{H}$. This is the aim of the next subsection.

2.5. **Explicit construction of moment sets on a Euclidean space.** In this subsection, we consider the case where the underlying spaces $\mathcal{X}_1, \dots, \mathcal{X}_N, \mathcal{Z}$ are all compact subsets of Euclidean spaces. We adopt explicit constructions of continuous test functions $\mathcal{G}_1, \dots, \mathcal{G}_N, \mathcal{H}$ by Neufeld and Xiang [54] such that $(\overline{W}_{\mu_i,\mathcal{G}_i})_{i=1:N}$ and $\overline{W}_{\mathcal{H}}$ can be controlled to be arbitrarily close to 0. These constructions ensure that the error term $\epsilon_{\mathsf{approx}}$ in Theorem 2.13 can be controlled to be arbitrarily close to 0. In the following, we work under Assumption 2.17 and consider test functions $\mathcal{G}_1, \dots, \mathcal{G}_N, \mathcal{H}$ that are constructed via either Setting 2.18 or Setting 2.19 presented below.

Assumption 2.17. *In addition to Assumption 2.12, we make the following assumptions.*

(i) *For* $i = 1, \dots, N$, $\mathcal{X}_i \subset \mathbb{R}^{d_i}$ *for* $d_i \in \mathbb{N}$, *and* $d_{\mathcal{X}_i}$ *is induced by a norm* $\|\cdot\|$ *on* $\mathbb{R}^{d_i}$.

(ii) $\mathcal{Z} \subset \mathbb{R}^{d_0}$ *for* $d_0 \in \mathbb{N}$, *and* $d_{\mathcal{Z}}$ *is induced by a norm* $\|\cdot\|$ *on* $\mathbb{R}^{d_0}$.

Setting 2.18. *Let Assumption 2.17 hold. In this setting,* $(\mathcal{G}_i)_{i=1:N}$ *and* $\mathcal{H}$ *are constructed as follows.*

- *For* $i = 1, \dots, N$:
  - *let* $-\infty < \underline{M}_{i,j} < \overline{M}_{i,j} < \infty$ *for* $j = 1, \dots, d_i$ *satisfy* $\mathcal{X}_i \subseteq \times_{j=1}^{d_i}[\underline{M}_{i,j}, \overline{M}_{i,j}]$;
  - *for* $j = 1, \dots, d_i$, *let* $n_{i,j} \in \mathbb{N}$, $\kappa_{i,j,l} := \underline{M}_{i,j} + \frac{l}{n_{i,j}}(\overline{M}_{i,j} - \underline{M}_{i,j})$ *for* $l = 0, \dots, n_{i,j}$;
  - *define* $\mathcal{G}_i$ *by*

$$\mathcal{G}_i := \left\{ \mathcal{X}_i \ni (x_1, \dots, x_{d_i})^{\mathsf{T}} \mapsto \max_{1\le j\le d_i} \left\{ \tfrac{n_{i,j}}{\overline{M}_{i,j} - \underline{M}_{i,j}} (x_j - \kappa_{i,j,l_j})^+ \right\} \in \mathbb{R} : \right.$$
$$\left. 0 \le l_j \le n_{i,j} \ \forall 1 \le j \le d_i \right\};$$

  - *define* $\epsilon_i := 2\left\| \left(\frac{\overline{M}_{i,1} - \underline{M}_{i,1}}{n_{i,1}}, \dots, \frac{\overline{M}_{i,d_i} - \underline{M}_{i,d_i}}{n_{i,d_i}} \right)^{\mathsf{T}} \right\|$.
- *Moreover, let* $-\infty < \underline{M}_{0,j} < \overline{M}_{0,j} < \infty$ *for* $j = 1, \dots, d_0$ *satisfy* $\mathcal{Z} \subseteq \times_{j=1}^{d_0}[\underline{M}_{0,j}, \overline{M}_{0,j}]$;
- *for* $j = 1, \dots, d_0$, *let* $n_{0,j} \in \mathbb{N}$, $\kappa_{0,j,l} := \underline{M}_{0,j} + \frac{l}{n_{0,j}}(\overline{M}_{0,j} - \underline{M}_{0,j})$ *for* $l = 0, \dots, n_{0,j}$;
- *define* $\mathcal{H}$ *by*

$$\mathcal{H} := \left\{ \mathcal{Z} \ni (z_1, \dots, z_{d_0})^{\mathsf{T}} \mapsto \max_{1\le j\le d_0} \left\{ \tfrac{n_{0,j}}{\overline{M}_{0,j} - \underline{M}_{0,j}} (z_j - \kappa_{0,j,l_j})^+ \right\} \in \mathbb{R} : \right.$$
$$\left. 0 \le l_j \le n_{0,j} \ \forall 1 \le j \le d_0 \right\};$$

- *define* $\epsilon_0 := 2\left\| \left(\frac{\overline{M}_{0,1} - \underline{M}_{0,1}}{n_{0,1}}, \dots, \frac{\overline{M}_{0,d_0} - \underline{M}_{0,d_0}}{n_{0,d_0}} \right)^{\mathsf{T}} \right\|$.

Setting 2.19. *Let Assumption 2.17 hold. Let $V(C)$ denote the set of extreme points of a polytope C. In this setting, $(\mathcal{G}_i)_{i=1:N}$ and $\mathcal{H}$ are constructed as follows.*

- *For $i = 1, \dots, N$:*
  - *let $\mathfrak{C}_i$ be a finite collection of d_i-simplices[4] in $\mathbb{R}^{d_i}$ which satisfies $\bigcup_{C\in\mathfrak{C}_i} C \supseteq \mathcal{X}_i$, and if $C_1, C_2 \in \mathfrak{C}_i$ and $C_1 \cap C_2 \neq \emptyset$ then $C_1 \cap C_2$ is a face[4] of both C_1 and C_2;*
  - *for every extreme point $\boldsymbol{v}$ of some $C \in \mathfrak{C}_i$, let $g_{i,\boldsymbol{v}} : \mathcal{X}_i \to \mathbb{R}$ be defined[5] as follows:*

$$g_{i,\boldsymbol{v}}(\boldsymbol{x}) := \sum_{\boldsymbol{u}\in V(F)} \lambda^F_{\boldsymbol{u}}(\boldsymbol{x}) \mathbb{1}_{\{\boldsymbol{u}=\boldsymbol{v}\}} \qquad \forall \boldsymbol{x} \in \operatorname{relint}(F) \cap \mathcal{X}_i,$$

  where $\boldsymbol{x} = \sum_{\boldsymbol{u}\in V(F)} \lambda^F_{\boldsymbol{u}}(\boldsymbol{x})\boldsymbol{u}$ and F is a face of some $C \in \mathfrak{C}_i$;
  - *let $\boldsymbol{v}_{i,0}$ be an extreme point of some $C \in \mathfrak{C}_i$, define $\mathcal{G}_i := \{g_{i,\boldsymbol{v}} : \boldsymbol{v}$ is an extreme point of some $C \in \mathfrak{C}_i$ and $\boldsymbol{v} \neq \boldsymbol{v}_{i,0}\}$, and define $\epsilon_i := 2 \max_{C\in\mathfrak{C}_i} \max_{\boldsymbol{v},\boldsymbol{v}'\in V(C)} \{\|\boldsymbol{v} - \boldsymbol{v}'\|\}$.*
- *Moreover, let $\mathfrak{C}_0$ be a finite collection of d_0-simplices in $\mathbb{R}^{d_0}$ which satisfies $\bigcup_{C\in\mathfrak{C}_0} C \supseteq \mathcal{Z}$, and if $C_1, C_2 \in \mathfrak{C}_0$ and $C_1 \cap C_2 \neq \emptyset$ then $C_1 \cap C_2$ is a face of both C_1 and C_2;*
- *for every extreme point $\boldsymbol{v}$ of some $C \in \mathfrak{C}_0$, let $h_{\boldsymbol{v}} : \mathcal{Z} \to \mathbb{R}$ be defined[5] as follows:*

$$h_{\boldsymbol{v}}(\boldsymbol{z}) := \sum_{\boldsymbol{u}\in V(F)} \lambda^F_{\boldsymbol{u}}(\boldsymbol{z}) \mathbb{1}_{\{\boldsymbol{u}=\boldsymbol{v}\}} \qquad \forall \boldsymbol{z} \in \operatorname{relint}(F) \cap \mathcal{Z},$$

  where $\boldsymbol{z} = \sum_{\boldsymbol{u}\in V(F)} \lambda^F_{\boldsymbol{u}}(\boldsymbol{z})\boldsymbol{u}$ and F is a face of some $C \in \mathfrak{C}_0$;
- *let $\boldsymbol{v}_{0,0}$ be an extreme point of some $C \in \mathfrak{C}_0$, define $\mathcal{H} := \{h_{\boldsymbol{v}} : \boldsymbol{v}$ is an extreme point of some $C \in \mathfrak{C}_0$ and $\boldsymbol{v} \neq \boldsymbol{v}_{0,0}\}$, and define $\epsilon_0 := 2 \max_{C\in\mathfrak{C}_0} \max_{\boldsymbol{v},\boldsymbol{v}'\in V(C)} \{\|\boldsymbol{v} - \boldsymbol{v}'\|\}$.*

Under these two settings, we are able to control the approximation error in Theorem 2.13 to be arbitrarily close to 0 as well as construct ϵ-approximate matching equilibria for any $\epsilon > 0$. This is stated in Theorem 2.20(ii) below. Moreover, statement (iii) of Theorem 2.20 provides a scalability result which bounds the number of test functions in $(\mathcal{G}_i)_{i=1:N}$ and $\mathcal{H}$ needed to control the approximation error. Furthermore, statement (iv) of Theorem 2.20 presents a sufficient condition for the affine independence condition in Proposition 2.3(ii) to hold.

Theorem 2.20 (Controlling the approximation error in Theorem 2.13)**.** *Let Assumption 2.17 hold, let $L^{(2)}_{\bar{c}} := \max_{1\le i\le N} \{L^{(2)}_{c_i}\}$, and suppose that $(\mathcal{G}_i)_{i=1:N}$ and $\mathcal{H}$ are constructed according to Setting 2.18 or Setting 2.19. Let $\epsilon_{\mathsf{LSIP}} > 0$ be arbitrary, and define $\bar{\epsilon}_{\mathsf{approx}} := \epsilon_{\mathsf{LSIP}} + (N-1)L^{(2)}_{\bar{c}}\epsilon_0 + \sum_{i=1}^N L^{(1)}_{c_i}\epsilon_i$. Moreover, let $(\tilde{\varphi}_i)_{i=1:N}$, $\hat{\nu}$, $(\hat{\gamma}_i)_{i=1:N}$, $\tilde{\nu}$, $(\tilde{\gamma}_i)_{i=1:N}$ be constructed via the procedure in Theorem 2.13. Then, the following statements hold.*

(i) Theorem 2.13 holds with $\epsilon_{\mathsf{approx}} \leftarrow \bar{\epsilon}_{\mathsf{approx}}$, i.e., $\big((\tilde{\varphi}_i)_{i=1:N}, (\hat{\gamma}_i)_{i=1:N}, \hat{\nu}\big)$ and $\big((\tilde{\varphi}_i)_{i=1:N}, (\tilde{\gamma}_i)_{i=1:N}, \tilde{\nu}\big)$ constitute $\bar{\epsilon}_{\mathsf{approx}}$-approximate matching equilibria.

(ii) For any $\epsilon > 0$ and any $\epsilon_{\mathsf{LSIP}} \in (0, \epsilon)$, the test functions $(\mathcal{G}_i)_{i=1:N}$ and $\mathcal{H}$ can be constructed via either Setting 2.18 or Setting 2.19 such that $\bar{\epsilon}_{\mathsf{approx}} \le \epsilon$.

(iii) For $i = 1, \dots, N$, let $C_{i,\|\cdot\|} \ge 1$ be a constant that satisfies $\|\boldsymbol{x}_i\| \le C_{i,\|\cdot\|}\|\boldsymbol{x}_i\|_2$ for all $\boldsymbol{x}_i \in \mathcal{X}_i$. Similarly, let $C_{0,\|\cdot\|} \ge 1$ be a constant that satisfies $\|\boldsymbol{z}\| \le C_{0,\|\cdot\|}\|\boldsymbol{z}\|_2$ for all $\boldsymbol{z} \in \mathcal{Z}$. Then, in statement (ii), $(\mathcal{G}_i)_{i=1:N}$ and $\mathcal{H}$ can be constructed via Setting 2.18 such that $|\mathcal{G}_i| = \prod_{j=1}^{d_i} \left(1 + \left\lceil \frac{4NL^{(1)}_{c_i}(\overline{M}_{i,j} - \underline{M}_{i,j})C_{i,\|\cdot\|}\sqrt{d_i}}{\epsilon - \epsilon_{\mathsf{LSIP}}} \right\rceil\right)$ for $i = 1, \dots, N$ and $|\mathcal{H}| = \prod_{j=1}^{d_0} \left(1 + \left\lceil \frac{4(N-1)L^{(2)}_{\bar{c}}(\overline{M}_{0,j} - \underline{M}_{0,j})C_{0,\|\cdot\|}\sqrt{d_0}}{\epsilon - \epsilon_{\mathsf{LSIP}}} \right\rceil\right)$.

(iv) Suppose that $(\mathcal{G}_i)_{i=1:N}$ and $\mathcal{H}$ are constructed via Setting 2.19. Let $m_i := |\mathcal{G}_i|$ for $i = 1, \dots, N$ and let $k := |\mathcal{H}|$. Then,

- *if $\operatorname{int}(\mathcal{X}_i) \cap \operatorname{int}(C) \neq \emptyset$ for all $C \in \mathfrak{C}_i$, then there exist $m_i + 1$ points $\boldsymbol{x}_{i,1}, \dots, \boldsymbol{x}_{i,m_i+1} \in \mathcal{X}_i$ such that the $m_i + 1$ vectors $\boldsymbol{g}_i(\boldsymbol{x}_{i,1}), \dots, \boldsymbol{g}_i(\boldsymbol{x}_{i,m_i+1}) \in \mathbb{R}^{m_i}$ are affinely independent;*

[4] The definitions of d_i-simplices and faces of a convex set can be found in, for example, [60, Section 1 & Section 18].

[5] Note that $g_{i,\boldsymbol{v}}(\boldsymbol{x})$ is well-defined for every $\boldsymbol{x} \in \mathcal{X}_i$ and $h_{\boldsymbol{v}}(\boldsymbol{z})$ is well-defined for every $\boldsymbol{z} \in \mathcal{Z}$ due to statements (i) and (ii) of [54, Proposition 3.10].

- *if* $\operatorname{int}(\mathcal{Z}) \cap \operatorname{int}(C) \neq \emptyset$ *for all* $C \in \mathfrak{C}_0$*, then there exist* $k+1$ *points* $\boldsymbol{z}_1, \dots, \boldsymbol{z}_{k+1} \in \mathcal{Z}$ *such that the* $k+1$ *vectors* $\boldsymbol{h}(\boldsymbol{z}_1), \dots, \boldsymbol{h}(\boldsymbol{z}_{k+1}) \in \mathbb{R}^k$ *are affinely independent.*

Remark 2.21. *Theorem 2.20(iii) provides insights about the scalability of the approximation scheme developed in this section. For simplicity, let* $\mathcal{X}_i \subseteq [\underline{M}, \overline{M}]^d$ *for* $d \in \mathbb{N}$*,* $-\infty < \underline{M} < \overline{M} < \infty$*, let* $d_{\mathcal{X}_i}$ *be induced by the Euclidean norm* $\|\cdot\|_2$ *on* $\mathbb{R}^d$ *for* $i = 1, \dots, N$*, let* $\mathcal{Z} \subseteq [\underline{M}, \overline{M}]^d$*,* $d_{\mathcal{Z}}$ *be induced by the Euclidean norm* $\|\cdot\|_2$ *on* $\mathbb{R}^d$*, and let* c_i *be* L_c*-Lipschitz continuous (with respect to the 1-product metric on* $\mathcal{X}_i \times \mathcal{Z}$*) for some* $L_c > 0$ *for* $i = 1, \dots, N$*. Then, by Theorem 2.20, the total number of test functions in* $\mathcal{G}_1, \dots, \mathcal{G}_N$ *and* $\mathcal{H}$ *to control the approximation error in Theorem 2.13 to below* ϵ *is of the order* $O\Big(N\big(\frac{4NL_c(\overline{M}-\underline{M})\sqrt{d}}{\epsilon-\epsilon_{\mathsf{LSIP}}}\big)^d\Big)$*, which is exponential in the dimension* d *of the underlying spaces. On the other hand, when* $(\overline{M} - \underline{M})$*,* d*, and* L_c *are fixed, the total number of test functions grows polynomially in the number* N *of agent categories.*

Under Setting 2.19, we can derive an explicit construction of a linear programming (LP) relaxation of the LSIP problem ($\mathsf{MT}^*_{\mathsf{par}}$) with bounded superlevel sets under some mild additional assumptions. This is detailed in the proposition below.

Proposition 2.22 (Explicit construction of LP relaxation of ($\mathsf{MT}^*_{\mathsf{par}}$) with bounded superlevel sets)**.** *Let Assumption 2.17 hold and suppose that* $(\mathcal{G}_i)_{i=1:N}$*,* $\mathcal{H}$ *are constructed according to Setting 2.19. For* $i = 1, \dots, N$*, let* $\widehat{\mathcal{X}}_i := \{\boldsymbol{v}_{i,0}, \boldsymbol{v}_{i,1}, \dots, \boldsymbol{v}_{i,m_i}\}$ *be an enumeration of the finite set* $\big\{\boldsymbol{v} \in \mathbb{R}^{d_i} : \boldsymbol{v}$ *is an extreme point of some* $C \in \mathfrak{C}_i\big\}$ *(i.e., the cardinality of this set is* $m_i + 1 \in \mathbb{N}$*) and denote let* $g_{i,j} := g_{i,\boldsymbol{v}_{i,j}}$ *for* $j = 0, 1, \dots, m_i$*. Let* $\widehat{\mathcal{Z}} := \{\boldsymbol{v}_{0,0}, \boldsymbol{v}_{0,1}, \dots, \boldsymbol{v}_{0,k}\}$ *be an enumeration of the finite set* $\big\{\boldsymbol{v} \in \mathbb{R}^{d_0} : \boldsymbol{v}$ *is an extreme point of some* $C \in \mathfrak{C}_0\big\}$ *(i.e., the cardinality of this set is* $k+1 \in \mathbb{N}$*) and denote* $h_l := h_{\boldsymbol{v}_{0,l}}$ *for* $l = 0, 1, \dots, k$*. Moreover, let* $\boldsymbol{g}_i(\cdot)$*,* $\boldsymbol{h}(\cdot)$*, and* $\bar{\boldsymbol{g}}_i$ *be given by (2.1), (2.2), and (2.3), respectively. Furthermore, let us assume in addition that for* $i = 1, \dots, N$*,* $\widehat{\mathcal{X}}_i \subseteq \mathcal{X}_i$*,* $\int_{\mathcal{X}_i} g_{i,\boldsymbol{v}_{i,j}} \, \mathrm{d}\mu_i > 0$ *for* $j = 0, 1, \dots, m_i$*, and that* $\widehat{\mathcal{Z}} \subseteq \mathcal{Z}$*. Then, the following LP relaxation of* ($\mathsf{MT}^*_{\mathsf{par}}$) *has bounded superlevel sets:*

$$
\begin{aligned}
\underset{(y_{i,0}, \boldsymbol{y}_i, \boldsymbol{w}_i)}{\text{maximize}} \quad & \sum_{i=1}^N y_{i,0} + \langle \bar{\boldsymbol{g}}_i, \boldsymbol{y}_i \rangle \\
\textit{subject to} \quad & y_{i,0} + \langle \boldsymbol{g}_i(\boldsymbol{x}_i), \boldsymbol{y}_i \rangle + \langle \boldsymbol{h}(\boldsymbol{z}_i), \boldsymbol{w}_i \rangle \leq c_i(\boldsymbol{x}_i, \boldsymbol{z}_i) \\
& \qquad\qquad \forall (\boldsymbol{x}_i, \boldsymbol{z}_i) \in \widehat{\mathcal{X}}_i \times \widehat{\mathcal{Z}}, \ \forall 1 \leq i \leq N, \\
& \sum_{i=1}^N \boldsymbol{w}_i = \boldsymbol{0}_k, \\
& y_{i,0} \in \mathbb{R}, \ \boldsymbol{y}_i \in \mathbb{R}^{m_i}, \ \boldsymbol{w}_i \in \mathbb{R}^k \quad \forall 1 \leq i \leq N.
\end{aligned}
\tag{2.10}
$$

Setting 2.19 and the choice of the finite sets $\widehat{\mathcal{X}}_1, \dots, \widehat{\mathcal{X}}_N, \widehat{\mathcal{Z}}$ in Proposition 2.22 also provide explicit upper bounds for the constants M_{opt} and M_{max} in the computational complexity results in Theorem 2.5 that depend on N, m, and k. This allows us to express the computational complexities in Theorem 2.5 purely in terms of N, m, k, and ϵ_{LSIP}.

Proposition 2.23 (Upper bounds for the constants in Theorem 2.5)**.** *Let all assumptions of Proposition 2.22 hold. Let us assume that, for* $i = 1, \dots, N$*,* $\mathcal{X}_i = \bigcup_{C \in \mathfrak{C}_i} C$*, and that* $\mathcal{Z} = \bigcup_{C \in \mathfrak{C}_0} C$*. Moreover, let us assume*[6] *that* $\max_{\boldsymbol{x}_i \in \mathcal{X}_i, \, \boldsymbol{z} \in \mathcal{Z}} \big\{c_i(\boldsymbol{x}_i, \boldsymbol{z})\big\} = 0$ *for* $i = 1, \dots, N$*, and let* $D_1 := \max_{1 \leq i \leq N} \big\{ \max_{\boldsymbol{x}_i, \boldsymbol{x}_i' \in \mathcal{X}_i} \big\{\|\boldsymbol{x}_i - \boldsymbol{x}_i'\|\big\}\big\}$*,* $D_2 := \max_{\boldsymbol{z}, \boldsymbol{z}' \in \mathcal{Z}} \big\{\|\boldsymbol{z} - \boldsymbol{z}'\|\big\}$*,* $L_{\bar{c}}^{(1)} := \max_{1 \leq i \leq N} \big\{L_{c_i}^{(1)}\big\}$*,* $L_{\bar{c}}^{(2)} := \max_{1 \leq i \leq N} \big\{L_{c_i}^{(2)}\big\}$*. Furthermore, let* $\texttt{Oracle}(\cdot, \cdot, \cdot)$ *be the global minimization oracle in Definition 2.4 and let* T *denote the computational cost of each call to* $\texttt{Oracle}(\cdot, \cdot, \cdot)$*. Then, the following statements hold.*

[6] This assumption can be satisfied by subtracting a constant from c_i, i.e., $c_i \leftarrow c_i - \max_{\boldsymbol{x}_i \in \mathcal{X}_i, \, \boldsymbol{z} \in \mathcal{Z}} \big\{c_i(\boldsymbol{x}_i, \boldsymbol{z})\big\}$.

(i) It holds that $h_l : \mathcal{Z} \to \mathbb{R}$ *is non-negative for* $l = 1, \dots, k$, $\max_{\boldsymbol{x}_i \in \mathcal{X}_i} \{\|\boldsymbol{g}_i(\boldsymbol{x}_i)\|_2\} \le 1$ *for* $i = 1, \dots, N$, *and* $\max_{\boldsymbol{z} \in \mathcal{Z}} \{\|\boldsymbol{h}(\boldsymbol{z})\|_2\} \le 1$.

(ii) There exists an optimizer $(y^\star_{1,0}, \boldsymbol{y}_1^{\star\mathsf{T}}, \boldsymbol{w}_1^{\star\mathsf{T}}, \dots, y^\star_{N,0}, \boldsymbol{y}_N^{\star\mathsf{T}}, \boldsymbol{w}_N^{\star\mathsf{T}})^\mathsf{T}$ *of* ($\mathsf{MT}^*_{\mathsf{par}}$) *that satisfies* $\|(y^\star_{1,0}, \boldsymbol{y}_1^{\star\mathsf{T}}, \boldsymbol{w}_1^{\star\mathsf{T}}, \dots, y^\star_{N,0}, \boldsymbol{y}_N^{\star\mathsf{T}}, \boldsymbol{w}_N^{\star\mathsf{T}})^\mathsf{T}\|_2 \le \big(L_{\bar{c}}^{(1)} D_1 + L_{\bar{c}}^{(2)} D_2\big)(2N+1)(m+N(k+1))^{\frac{1}{2}}$.

(iii) If the values of D_1, D_2, $L_{\bar{c}}^{(1)}$, $L_{\bar{c}}^{(2)}$ *do not depend on* N, m, k, ϵ_{LSIP}, *then there exists an algorithm which computes an* ϵ_{LSIP}*-optimizer of* ($\mathsf{MT}^*_{\mathsf{par}}$) *with computational complexity*[7] $O\big((m+Nk)\log((m+Nk)/\epsilon_{\mathsf{LSIP}})(NT + (m+Nk)^\omega)\big)$.

If we assume in addition that $\rho_{\mathsf{min}} := \min_{1 \le i \le N, 0 \le j \le m_i} \{\int_{\mathcal{X}_i} g_{i,j} \, \mathrm{d}\mu_i\} > 0$, *then the following statements hold.*

(iv) For every $(y_{1,0}, \boldsymbol{y}_1^\mathsf{T}, \boldsymbol{w}_1^\mathsf{T}, \dots, y_{N,0}, \boldsymbol{y}_N^\mathsf{T}, \boldsymbol{w}_N^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{m+N(k+1)}$ *satisfying* $y_{i,0} + \langle \boldsymbol{g}_i(\boldsymbol{x}_i), \boldsymbol{y}_i\rangle + \langle \boldsymbol{h}(\boldsymbol{z}_i), \boldsymbol{w}_i\rangle \le c_i(\boldsymbol{x}_i, \boldsymbol{z}_i)\ \forall (\boldsymbol{x}_i, \boldsymbol{z}_i) \in \widehat{\mathcal{X}}_i \times \widehat{\mathcal{Z}}$ *for* $i = 1, \dots, N$, $\sum_{i=1}^N \boldsymbol{w}_i = \boldsymbol{0}_k$, *and* $\sum_{i=1}^N y_{i,0} + \langle \bar{\boldsymbol{g}}_i, \boldsymbol{y}_i\rangle \ge \alpha^\star$ *where* $\alpha^\star$ *denotes the optimal value of* ($\mathsf{MT}^*_{\mathsf{par}}$), *it holds that* $\big\|(y_{1,0}, \boldsymbol{y}_1^\mathsf{T}, \boldsymbol{w}_1^\mathsf{T}, \dots, y_{N,0}, \boldsymbol{y}_N^\mathsf{T}, \boldsymbol{w}_N^\mathsf{T})^\mathsf{T}\big\|_2 \le \big(L_{\bar{c}}^{(1)} D_1 + L_{\bar{c}}^{(2)} D_2\big) N^2 (2 + \rho_{\mathsf{min}}^{-1})(m + N(k+1))^{\frac{1}{2}}$.

(v) If the values of D_1, D_2, $L_{\bar{c}}^{(1)}$, $L_{\bar{c}}^{(2)}$ *do not depend on* N, m, k, ϵ_{LSIP}, *then there exists an algorithm which computes a pair of* ϵ_{LSIP}*-optimizers of* ($\mathsf{MT}^*_{\mathsf{par}}$) *and* ($\mathsf{MT}_{\mathsf{par}}$) *with* $\mathrm{poly}\big(N, m, k, -\log(\rho_{\mathsf{min}}\epsilon_{\mathsf{LSIP}})\big)$ *number of calls to* `Oracle`$(\cdot,\cdot,\cdot)$ *plus* $\mathrm{poly}\big(N, m, k, -\log(\rho_{\mathsf{min}}\epsilon_{\mathsf{LSIP}})\big)$ *additional computational time.*

Remark 2.24. *When the conditions* $\widehat{\mathcal{X}}_i \subseteq \mathcal{X}_i$, $\widehat{\mathcal{Z}} \subseteq \mathcal{Z}$ *in Proposition 2.22, or the conditions* $\mathcal{X}_i = \bigcup_{C \in \mathfrak{C}_i} C$, $\mathcal{Z} = \bigcup_{C \in \mathfrak{C}_0} C$ *in Proposition 2.23 fail to hold, one may extend* $\mathcal{X}_i$ *to* $\widetilde{\mathcal{X}}_i := \bigcup_{C \in \mathfrak{C}_i} C$, *extend* $\mathcal{Z}$ *to* $\widetilde{\mathcal{Z}} := \bigcup_{C \in \mathfrak{C}_0} C$, *and extend the definition of* $c_i : \mathcal{X}_i \times \mathcal{Z} \to \mathbb{R}$ *to* $\tilde{c}_i : \widetilde{\mathcal{X}}_i \times \widetilde{\mathcal{Z}} \to \mathbb{R}$ *as follows:*

$$\tilde{c}_i(\boldsymbol{x}_i, \boldsymbol{z}) := \min_{(\boldsymbol{x}_i', \boldsymbol{z}') \in \mathcal{X}_i \times \mathcal{Z}} \Big\{ c_i(\boldsymbol{x}_i', \boldsymbol{z}') + L_{c_i}^{(1)} \|\boldsymbol{x}_i - \boldsymbol{x}_i'\| + L_{c_i}^{(2)} \|\boldsymbol{z} - \boldsymbol{z}'\| \Big\} \qquad \forall (\boldsymbol{x}_i, \boldsymbol{z}) \in \widetilde{\mathcal{X}}_i \times \widetilde{\mathcal{Z}}.$$

Such an extension satisfies $\tilde{c}_i(\boldsymbol{x}_i, \boldsymbol{z}) = c_i(\boldsymbol{x}_i, \boldsymbol{z})$ *for all* $(\boldsymbol{x}_i, \boldsymbol{z}) \in \mathcal{X}_i \times \mathcal{Z}$ *as well as* $\big|\tilde{c}_i(\boldsymbol{x}_i, \boldsymbol{z}) - \tilde{c}_i(\boldsymbol{x}_i', \boldsymbol{z}')\big| \le L_{c_i}^{(1)} \|\boldsymbol{x}_i - \boldsymbol{x}_i'\| + L_{c_i}^{(2)} \|\boldsymbol{z} - \boldsymbol{z}'\|$ *for all* $(\boldsymbol{x}_i, \boldsymbol{z}), (\boldsymbol{x}_i', \boldsymbol{z}') \in \widetilde{\mathcal{X}}_i \times \widetilde{\mathcal{Z}}$. *Thus, the analyses in this subsection can be carried out with* $\mathcal{X}_i$, $\mathcal{Z}$, c_i *replaced by* $\widetilde{\mathcal{X}}_i$, $\widetilde{\mathcal{Z}}$, $\tilde{c}_i$.

3. Numerical method

In this section, we develop a numerical method for approximately solving the matching for teams problem. To begin, we solve the LSIP problem ($\mathsf{MT}^*_{\mathsf{par}}$) by developing a so-called cutting-plane discretization algorithm inspired by Conceptual Algorithm 11.4.1 in [40]. The core idea of the algorithm is to replace the semi-infinite constraint in ($\mathsf{MT}^*_{\mathsf{par}}$) with a finite subcollection of constraints, which relaxes ($\mathsf{MT}^*_{\mathsf{par}}$) into a linear programming (LP) problem. Subsequently, one iteratively adds more constraints to the existing subcollection until the approximation error falls below a pre-specified tolerance threshold. The addition of constraints can be thought of as introducing "cuts" to restrict the feasible set of an LP relaxation of ($\mathsf{MT}^*_{\mathsf{par}}$).

Before presenting the algorithm, let us first introduce the following lemma, which deals with the construction of an optimal coupling in the discrete-to-discrete case, the discrete-to-continuous case (also known as the semi-discrete case), and the one-dimensional case. This lemma combines classical results about discrete optimal transport (see, e.g., [56, Section 2.3] and [9, Section 1.3]), semi-discrete optimal transport (see, e.g., [49] and [56, Section 5.2]), and one-dimensional optimal transport (see, e.g., [59, Section 3.1]).

Lemma 3.1 (Construction of optimal coupling)**.** *Let* $(\mathcal{Y}, d_\mathcal{Y})$ *be a compact metric space and let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. For* $n \in \mathbb{N}$, $\big(\alpha_i \in (0,1]\big)_{i=1:n}$, *distinct points* $(x_i \in \mathcal{Y})_{i=1:n}$ *with*

[7] Recall that we denote the computational complexity of the multiplication of two $m \times m$ matrices by $O(m^\omega)$.

$\sum_{i=1}^n \alpha_i = 1$, let $Y : \Omega \to \mathcal{Y}$ be a random variable such that $\mathbb{P}[Y = x_i] = \alpha_i$ for $i = 1, \dots, n$. Let $\nu_1 \in \mathcal{P}(\mathcal{Y})$ denote the law of Y and let $\nu_2 \in \mathcal{P}(\mathcal{Y})$. Suppose that any one of the following assumptions hold:

- (C1) *The discrete-to-discrete case. $\nu_2 = \sum_{i=1}^{n_2} \beta_i \delta_{y_i}$ for $n_2 \in \mathbb{N}$, $\big(\beta_i \in (0,1]\big)_{i=1:n_2}$, distinct points $(y_i \in \mathcal{Y})_{i=1:n_2}$ with $\sum_{i=1}^{n_2} \beta_i = 1$.*
- (C2) *The discrete-to-continuous case. $\mathcal{Y} \subset \mathbb{R}^d$ for $d \in \mathbb{N}$, $d_{\mathcal{Y}}$ is induced by a norm $\|\cdot\|$ on $\mathbb{R}^d$ under which the closed unit ball $\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| \le 1\}$ is a strictly convex set[8], ν_2 is absolutely continuous with respect to the Lebesgue measure on $\mathcal{Y}$.*
- (C3) *The one-dimensional case. $\mathcal{Y} \subset \mathbb{R}$ and $d_{\mathcal{Y}}$ is the Euclidean distance on $\mathbb{R}$.*

Let the random variable $\bar{Y} : \Omega \to \mathcal{Y}$ be defined according to the procedures below in the three cases (C1)–(C3).

- *The discrete-to-discrete case. Suppose that (C1) holds and let $(\gamma^\star_{i,j})_{i=1:n,\, j=1:n_2}$ be an optimizer of the following linear programming (LP) problem:*

$$\begin{aligned} \underset{(\gamma_{i,j})}{\text{minimize}} \quad & \sum_{i=1}^{n}\sum_{j=1}^{n_2} d_{\mathcal{Y}}(x_i, y_j)\gamma_{i,j} \\ \text{subject to} \quad & \sum_{j=1}^{n_2} \gamma_{i,j} = \alpha_i \ \forall 1 \le i \le n, \qquad \sum_{i=1}^{n} \gamma_{i,j} = \beta_j \ \forall 1 \le j \le n_2, \\ & \gamma_{i,j} \ge 0 \qquad \forall 1 \le i \le n,\ \forall 1 \le j \le n_2. \end{aligned}$$

  Let $\bar{Y} : \Omega \to \mathcal{Y}$ be such that $\mathbb{P}[\bar{Y} = y_j | Y = x_i] = \gamma^\star_{i,j}$ for $i = 1, \dots, n$, $j = 1, \dots, n_2$.
- *The discrete-to-continuous case. Suppose that (C2) holds and let $\big(\phi_i^\star\big)_{i=1:n} \subset \mathbb{R}$ be an optimizer of the following concave maximization problem (which always exists; see, e.g., [54, Proposition 3.2]): $\sup_{\phi_1,\dots,\phi_n \in \mathbb{R}} \Big\{ \sum_{i=1}^n \phi_i \alpha_i - \int_{\mathcal{Y}} \max_{1\le i\le n} \big\{\phi_i - d_{\mathcal{Y}}(x_i, y)\big\}\, \nu_2(\mathrm{d}y) \Big\}$. For $i = 1, \dots, n$, let $V_i := \big\{\boldsymbol{z} \in \mathcal{Y} : \phi_i^\star - d_{\mathcal{Y}}(x_i, \boldsymbol{z}) = \max_{1\le k\le n} \big\{\phi_k^\star - d_{\mathcal{Y}}(x_k, \boldsymbol{z})\big\}\big\}$. Let $\bar{Y} : \Omega \to \mathcal{Y}$ be such that $\mathbb{P}[\bar{Y} \in E | Y = x_i] = \frac{\nu_2\big(E \cap V_i\big)}{\nu_2(V_i)}$ for all $E \in \mathcal{B}(\mathcal{Y})$, for $i = 1, \dots, n$.*
- *The one-dimensional case. Suppose that (C3) holds, let $F^{-1}_{\nu_2}(t) := \inf\big\{y \in \mathcal{Y} : \nu_2\big(\mathcal{Y} \cap (-\infty, y]\big) \ge t\big\}$ for $t \in [0,1]$, and let $\bar{Y} : \Omega \to \mathcal{Y}$ be constructed via the following procedure.*
  - *Step 1: sort the sequence $(x_1, \dots, x_n)$ into ascending order $x^{(1)} < x^{(2)} < \cdots < x^{(n)}$ and let $\sigma(x_i)$ denote the order of x_i in the sorted sequence, i.e., $\big\{\sigma(x_i) : 1 \le i \le n\big\} = \{1, \dots, n\}$ and $x^{(\sigma(x_i))} \equiv x_i$ for $i = 1, \dots, n$.*
  - *Step 2: for $j = 0, 1, \dots, n$, let $F(j) := \sum_{1\le i\le n,\, \sigma(x_i)\le j} \alpha_i$.*
  - *Step 3: let $U : \Omega \to [0,1]$ be a uniform random variable on $[0,1]$ that is independent of Y.*
  - *Step 4: let $\bar{Y} := F^{-1}_{\nu_2}\Big(UF(\sigma(Y)) + (1-U)F(\sigma(Y) - 1)\Big)$.*

Then, in all three cases, the law $\gamma^\star \in \mathcal{P}(\mathcal{Y} \times \mathcal{Y})$ of the random variable $(Y, \bar{Y}) : \Omega \to \mathcal{Y} \times \mathcal{Y}$ satisfies $\gamma^\star \in \Gamma(\nu_1, \nu_2)$ and $\int_{\mathcal{Y}\times\mathcal{Y}} d_{\mathcal{Y}}(x,y)\, \gamma^\star(\mathrm{d}x, \mathrm{d}y) = W_1(\nu_1, \nu_2)$.

We will work with the following assumptions throughout this section.

Assumption 3.2. *In addition to Assumption 2.12, we make the following assumptions:*

(i) For $i = 1, \dots, N$, $\mathrm{supp}(\mu_i) = \mathcal{X}_i$, and at least one of the assumptions (C1)–(C3) in Lemma 3.1 is satisfied with respect to $\mathcal{Y} \leftarrow \mathcal{X}_i$, $d_{\mathcal{Y}} \leftarrow d_{\mathcal{X}_i}$, $\nu_2 \leftarrow \mu_i$.

(ii) For $i = 1, \dots, N$, $\mathcal{G}_i := \big\{g_{i,1}, \dots, g_{i,m_i}\big\} \subset \mathcal{C}(\mathcal{X}_i)$ for $m_i \in \mathbb{N}$. Moreover, for $\boldsymbol{g}_i : \mathcal{X}_i \to \mathbb{R}^{m_i}$ defined in (2.1), there exist $m_i + 1$ points $x_{i,1}, \dots, x_{i,m_i+1} \in \mathcal{X}_i$ such that the $m_i + 1$ vectors $\boldsymbol{g}_i(x_{i,1}), \dots, \boldsymbol{g}_i(x_{i,m_i+1}) \in \mathbb{R}^{m_i}$ are affinely independent.

[8] For example, under the p-norm, this condition is satisfied for all $1 < p < \infty$ (by the Minkowski inequality), but fails when $p = 1$ or $p = \infty$ (assuming $d > 1$).

(iii) $\mathcal{H} := \{h_1, \ldots, h_k\} \subset \mathcal{C}(\mathcal{Z})$ *for* $k \in \mathbb{N}$. *Moreover, for* $\boldsymbol{h} : \mathcal{Z} \to \mathbb{R}^k$ *defined in (2.2), there exist* $k+1$ *points* $z_1, \ldots, z_{k+1} \in \mathcal{Z}$ *such that the* $k+1$ *vectors* $\boldsymbol{h}(z_1), \ldots, \boldsymbol{h}(z_{k+1}) \in \mathbb{R}^k$ *are affinely independent.*

Note that by Proposition 2.3, Assumption 3.2 is sufficient to guarantee that the set of optimizers of ($\mathsf{MT}^*_{\mathsf{par}}$) is non-empty and bounded, which is crucial for proving the convergence of the cutting-plane discretization algorithm in this section.

Remark 3.3. *Recall that Setting 2.19 is a concrete setting in which Assumption 3.2(ii) and Assumption 3.2(iii) hold. This has been shown in Theorem 2.20(iv).*

Algorithm 1 provides a concrete implementation of the cutting-plane discretization method for solving ($\mathsf{MT}^*_{\mathsf{par}}$). A key step in Algorithm 1 is to solve LP relaxations of ($\mathsf{MT}^*_{\mathsf{par}}$) in which the semi-infinite constraint $y_{i,0} + \langle \boldsymbol{g}_i(x_i), \boldsymbol{y}_i \rangle + \langle \boldsymbol{h}(z_i), \boldsymbol{w}_i \rangle \leq c_i(x_i, z_i)\ \forall (x_i, z_i) \in \mathcal{X}_i \times \mathcal{Z}$ is replaced by a finite subcollection of constraints for $i = 1, \ldots, N$. Specifically, for any $r \in \mathbb{Z}_+$ and any finite set $\mathcal{C}_i^{(r)} \subseteq \mathcal{X}_i \times \mathcal{Z}$, let ($\mathsf{MT}^{*(r)}_{\mathsf{par}}$) denote the following LP problem:

$$
\begin{aligned}
\underset{(y_{i,0}, \boldsymbol{y}_i, \boldsymbol{w}_i)}{\text{maximize}} \quad & \sum_{i=1}^N y_{i,0} + \langle \bar{\boldsymbol{g}}_i, \boldsymbol{y}_i \rangle \\
\text{subject to} \quad & y_{i,0} + \langle \boldsymbol{g}_i(x_i), \boldsymbol{y}_i \rangle + \langle \boldsymbol{h}(z_i), \boldsymbol{w}_i \rangle \leq c_i(x_i, z_i) \\
& \qquad \forall (x_i, z_i) \in \mathcal{C}_i^{(r)},\ \forall 1 \leq i \leq N, \qquad (\mathsf{MT}^{*(r)}_{\mathsf{par}}) \\
& \sum_{i=1}^N \boldsymbol{w}_i = \boldsymbol{0}_k, \\
& y_{i,0} \in \mathbb{R},\ \boldsymbol{y}_i \in \mathbb{R}^{m_i},\ \boldsymbol{w}_i \in \mathbb{R}^k \quad \forall 1 \leq i \leq N.
\end{aligned}
$$

The LP problem ($\mathsf{MT}^{*(r)}_{\mathsf{par}}$) admits the following dual LP problem:

$$
\begin{aligned}
\underset{(\theta_{i,x,z}), \boldsymbol{\xi}}{\text{minimize}} \quad & \sum_{i=1}^N \sum_{(x,z) \in \mathcal{C}_i^{(r)}} \theta_{i,x,z} c_i(x,z) \\
\text{subject to} \quad & \sum_{(x,z) \in \mathcal{C}_i^{(r)}} \theta_{i,x,z} = 1 && \forall 1 \leq i \leq N, \\
& \sum_{(x,z) \in \mathcal{C}_i^{(r)}} \theta_{i,x,z} \boldsymbol{g}_i(x) = \bar{\boldsymbol{g}}_i && \forall 1 \leq i \leq N, \qquad (\mathsf{MT}^{(r)}_{\mathsf{par}}) \\
& \sum_{(x,z) \in \mathcal{C}_i^{(r)}} \theta_{i,x,z} \boldsymbol{h}(z) = \boldsymbol{\xi} && \forall 1 \leq i \leq N, \\
& \theta_{i,x,z} \in \mathbb{R}_+ \qquad \forall (x,z) \in \mathcal{C}_i^{(r)},\ \forall 1 \leq i \leq N, \\
& \boldsymbol{\xi} \in \mathbb{R}^k.
\end{aligned}
$$

Remark 3.4 explains the assumptions of Algorithm 1. The properties of Algorithm 1 are presented in Proposition 3.5.

Remark 3.4 (Details of Algorithm 1)**.** *Let Assumption 3.2 hold. Algorithm 1 is inspired by the Conceptual Algorithm 11.4.1 in [40]. Below is a list explaining the inputs to Algorithm 1.*

- $(\mathcal{X}_i)_{i=1:N}$ *and* $(c_i)_{i=1:N}$ *are given by Assumption 1.1.*
- $(\boldsymbol{g}_i : \mathcal{X}_i \to \mathbb{R}^{m_i})_{i=1:N}$, $\boldsymbol{h} : \mathcal{Z} \to \mathbb{R}^k$, *and* $(\bar{\boldsymbol{g}}_i \in \mathbb{R}^{m_i})_{i=1:N}$ *are defined in (2.1), (2.2), and (2.3), respectively.*
- $\mathcal{C}_i^{(0)} \subseteq \mathcal{X}_i \times \mathcal{Z}$ *is a finite set for* $i = 1, \ldots, N$. *The sets* $\left(\mathcal{C}_i^{(0)}\right)_{i=1:N}$ *are chosen such that the LP problem* ($\mathsf{MT}^{*(r)}_{\mathsf{par}}$) *with* $r = 0$ *has bounded superlevel sets. The existence of such* $\left(\mathcal{C}_i^{(0)}\right)_{i=1:N}$ *is shown in Proposition 3.5(i).*

Algorithm 1: Cutting-plane discretization algorithm for solving ($\mathsf{MT}^*_{\mathsf{par}}$) and ($\mathsf{MT}_{\mathsf{par}}$)

Input: $(\mathcal{X}_i)_{i=1:N}$, $(c_i)_{i=1:N}$, $(\boldsymbol{g}_i : \mathcal{X}_i \to \mathbb{R}^{m_i})_{i=1:N}$, $(\bar{\boldsymbol{g}}_i \in \mathbb{R}^{m_i})_{i=1:N}$, $\boldsymbol{h} : \mathcal{Z} \to \mathbb{R}^k$, $\big(\mathcal{C}_i^{(0)} \subseteq \mathcal{X}_i \times \mathcal{Z}\big)_{i=1:N}$, $\texttt{Oracle}(\cdot,\cdot,\cdot)$, $\epsilon_{\mathsf{LSIP}} > 0$

Output: $\alpha^{\mathsf{UB}}_{\mathsf{MT}_{\mathsf{par}}}$, $\alpha^{\mathsf{LB}}_{\mathsf{MT}_{\mathsf{par}}}$, $(\hat{y}_{i,0}, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{w}}_i)_{i=1:N}$, $(\hat{\theta}_i)_{i=1:N}$

1 $r \leftarrow 0$.
2 **while** *true* **do**
3 Solve the LP problem ($\mathsf{MT}^{*(r)}_{\mathsf{par}}$), denote the computed optimal value by $\alpha^{(r)}$, denote the computed optimizer by $\big(y^{(r)}_{i,0}, \boldsymbol{y}^{(r)}_i, \boldsymbol{w}^{(r)}_i\big)_{i=1:N}$, and denote the corresponding optimizer of the dual LP problem ($\mathsf{MT}^{(r)}_{\mathsf{par}}$) by $\big(\theta^{(r)}_{i,x,z}\big)_{(x,z)\in\mathcal{C}_i^{(r)}, i=1:N}$, $\boldsymbol{\xi}^{(r)}$.
4 **for** $i = 1, \ldots, N$ **do**
5 Call $\texttt{Oracle}\big(i, \boldsymbol{y}^{(r)}_i, \boldsymbol{w}^{(r)}_i\big)$ and let $\big(x^{\star(r)}_i, z^{\star(r)}_i, \beta^{(r)}_i\big)$ be its output.
6 Let $\mathcal{C}^{\star(r)}_i \subseteq \mathcal{X}_i \times \mathcal{Z}$ be a finite set such that $\big(x^{\star(r)}_i, z^{\star(r)}_i\big) \in \mathcal{C}^{\star(r)}_i$.
7 $\mathcal{C}^{(r+1)}_i \leftarrow \mathcal{C}^{(r)}_i \cup \mathcal{C}^{\star(r)}_i$.
8 **if** $\sum_{i=1}^N y^{(r)}_{i,0} - \beta^{(r)}_i \le \epsilon_{\mathsf{LSIP}}$ **then**
9 Skip to Line 11.
10 $r \leftarrow r + 1$.
11 $\alpha^{\mathsf{UB}}_{\mathsf{MT}_{\mathsf{par}}} \leftarrow \alpha^{(r)}$, $\alpha^{\mathsf{LB}}_{\mathsf{MT}_{\mathsf{par}}} \leftarrow \alpha^{(r)} - \Big(\sum_{i=1}^N y^{(r)}_{i,0} - \beta^{(r)}_i\Big)$.
12 **for** $i = 1, \ldots, N$ **do**
13 $\hat{y}_{i,0} \leftarrow \beta^{(r)}_i$, $\hat{\boldsymbol{y}}_i \leftarrow \boldsymbol{y}^{(r)}_i$, $\hat{\boldsymbol{w}}_i \leftarrow \boldsymbol{w}^{(r)}_i$.
14 $\hat{\theta}_i \leftarrow \sum_{(x,z)\in\mathcal{C}_i^{(r)}} \theta^{(r)}_{i,x,z}\delta_{(x,z)}$.
15 **return** $\alpha^{\mathsf{UB}}_{\mathsf{MT}_{\mathsf{par}}}$, $\alpha^{\mathsf{LB}}_{\mathsf{MT}_{\mathsf{par}}}$, $(\hat{y}_{i,0}, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{w}}_i)_{i=1:N}$, $(\hat{\theta}_i)_{i=1:N}$.

- $\texttt{Oracle}(\cdot,\cdot,\cdot)$ *is the global minimization oracle in Definition 2.4. We assume that a numerical procedure can be implemented to solve this global minimization problem.*
- $\epsilon_{\mathsf{LSIP}} > 0$ *is a pre-specified numerical tolerance value (see Proposition 3.5).*

We would like to remark that when solving the LP problem ($\mathsf{MT}^{*(r)}_{\mathsf{par}}$) in Line 3 by the dual simplex algorithm (see, e.g., [67, Chapter 6.4]) or the interior point algorithm (see, e.g., [67, Chapter 18]), one can obtain a corresponding optimizer of the dual LP problem ($\mathsf{MT}^{(r)}_{\mathsf{par}}$) from the outputs of these algorithms.

Proposition 3.5 (Properties of Algorithm 1)**.** *Let Assumption 3.2 hold. Then,*

(i) there exist finite sets $\big(\mathcal{C}_i^{(0)} \subseteq \mathcal{X}_i \times \mathcal{Z}\big)_{i=1:N}$ such that the LP problem ($\mathsf{MT}^{*(r)}_{\mathsf{par}}$) with $r = 0$ has bounded superlevel sets.

Moreover, assume that all inputs of Algorithm 1 are set according to Remark 3.4. Then, the following statements hold.

(ii) Algorithm 1 terminates after finitely many iterations.
(iii) $\alpha^{\mathsf{LB}}_{\mathsf{MT}_{\mathsf{par}}} \le (\mathsf{MT}^*_{\mathsf{par}}) \le \alpha^{\mathsf{UB}}_{\mathsf{MT}_{\mathsf{par}}}$ where $\alpha^{\mathsf{UB}}_{\mathsf{MT}_{\mathsf{par}}} - \alpha^{\mathsf{LB}}_{\mathsf{MT}_{\mathsf{par}}} \le \epsilon_{\mathsf{LSIP}}$.
(iv) $(\hat{y}_{i,0}, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{w}}_i)_{i=1:N}$ is an ϵ_{LSIP}-optimal solution of ($\mathsf{MT}^*_{\mathsf{par}}$) and $\sum_{i=1}^N \hat{y}_{i,0} + \langle \bar{\boldsymbol{g}}_i, \hat{\boldsymbol{y}}_i\rangle = \alpha^{\mathsf{LB}}_{\mathsf{MT}_{\mathsf{par}}}$.
(v) $(\hat{\theta}_i)_{i=1:N}$ is an ϵ_{LSIP}-optimal solution of ($\mathsf{MT}_{\mathsf{par}}$) where $\hat{\theta}_i$ has finite support for $i = 1, \ldots, N$ and $\sum_{i=1}^N \int_{\mathcal{X}_i\times\mathcal{Z}} c_i \,\mathrm{d}\hat{\theta}_i = \alpha^{\mathsf{UB}}_{\mathsf{MT}_{\mathsf{par}}}$.

The concrete procedure for computing approximate matching equilibria based on the outputs of Algorithm 1 is shown in Algorithm 2. Remark 3.6 explains the assumptions and details of Algorithm 2. The properties of Algorithm 2 are presented in Theorem 3.7.

Algorithm 2: Construction of approximate matching equilibria

Input: $(\mathcal{X}_i)_{i=1:N}$, $\mathcal{Z}$, $(\mu_i)_{i=1:N}$, $(c_i)_{i=1:N}$, $\epsilon_{\mathsf{LSIP}} > 0$, $(\mathcal{G}_i)_{i=1:N}$, $\mathcal{H}$, $\tilde{z}(\cdot)$, $\texttt{Oracle}(\cdot,\cdot,\cdot)$

Output: $(\tilde{\varphi}_i)_{i=1:N}$, $\hat{\nu}$, $(\hat{\gamma}_i)_{i=1:N}$, $\tilde{\nu}$, $(\tilde{\gamma}_i)_{i=1:N}$, $\alpha^{\mathsf{LB}}_{\mathsf{MT}}$, $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$, $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$, $\hat{\epsilon}_{\mathsf{sub}}$, $\tilde{\epsilon}_{\mathsf{sub}}$

1 Let $(\boldsymbol{g}_i(\cdot))_{i=1:N}$, $\boldsymbol{h}(\cdot)$, and $(\bar{\boldsymbol{g}}_i)_{i=1:N}$ be defined by (2.1), (2.2), and (2.3).

2 Construct finite sets $\left(\mathcal{C}_i^{(0)} \subseteq \mathcal{X}_i \times \mathcal{Z}\right)_{i=1:N}$ such that the LP problem ($\mathsf{MT}^{*(r)}_{\mathsf{par}}$) with $r = 0$ has bounded superlevel sets.

3 $\left(\alpha^{\mathsf{UB}}_{\mathsf{MT}_{\mathsf{par}}}, \alpha^{\mathsf{LB}}_{\mathsf{MT}_{\mathsf{par}}}, (\hat{y}_{i,0}, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{w}}_i)_{i=1:N}, (\hat{\theta}_i)_{i=1:N}\right) \leftarrow$ the outputs of Algorithm 1 with inputs $\left((\mathcal{X}_i)_{i=1:N}, (c_i)_{i=1:N}, (\boldsymbol{g}_i(\cdot))_{i=1:N}, (\bar{\boldsymbol{g}}_i)_{i=1:N}, \boldsymbol{h}(\cdot), \left(\mathcal{C}_i^{(0)}\right)_{i=1:N}, \texttt{Oracle}(\cdot,\cdot,\cdot), \epsilon_{\mathsf{LSIP}}\right)$.

4 Choose an arbitrary $z_0 \in \mathcal{Z}$ and let $(\tilde{\varphi}_i)_{i=1:N}$ be defined as in (2.9).

5 Let $\hat{\nu}_i$ denote the marginal of $\hat{\theta}_i$ on $\mathcal{Z}$ for $i = 1, \ldots, N$. Choose an arbitrary $\hat{i} \in \{1, \ldots, N\}$.

6 $K \leftarrow |\mathrm{supp}(\hat{\nu}_{\hat{i}})|$. Express $\hat{\nu}_{\hat{i}} = \sum_{l=1}^{K} \alpha_l \delta_{z_l}$ for $(\alpha_l \in (0,1])_{l=1:K}$ satisfying $\sum_{l=1}^{K} \alpha_l = 1$ and $(z_l \in \mathcal{Z})_{l=1:K}$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $Z : \Omega \to \mathcal{Z}$ be a random variable such that $\mathbb{P}[Z = z_l] = \alpha_l$ for $l = 1, \ldots, K$. $\hat{\nu} \leftarrow$ the law of Z.

7 **for** $i = 1, \ldots, N$ **do**

8 Let a random variable $Z_i : \Omega \to \mathcal{Z}$ be constructed via Lemma 3.1 with $\mathcal{Y} \leftarrow \mathcal{Z}$, $d_{\mathcal{Y}} \leftarrow d_{\mathcal{Z}}$, $Y \leftarrow Z$, $\nu_2 \leftarrow \hat{\nu}_i$, $\bar{Y} \leftarrow Z_i$.

9 Let a random variable $X_i : \Omega \to \mathcal{X}_i$ be such that $\mathbb{P}[X_i \in E | Z_i = z] = \frac{\hat{\theta}_i(E \times \{z\})}{\hat{\nu}_i(\{z\})}$ for every $z \in \mathrm{supp}(\hat{\nu}_i)$ and every $E \in \mathcal{B}(\mathcal{X}_i)$.

10 Let a random variable $\bar{X}_i : \Omega \to \mathcal{X}_i$ be constructed via Lemma 3.1 with $\mathcal{Y} \leftarrow \mathcal{X}_i$, $d_{\mathcal{Y}} \leftarrow d_{\mathcal{X}_i}$, $Y \leftarrow X_i$, $\nu_2 \leftarrow \mu_i$, $\bar{Y} \leftarrow \bar{X}_i$.

11 Define a random variable $\bar{Z} : \Omega \to \mathcal{Z}$ as $\bar{Z} := \tilde{z}(\bar{X}_1, \ldots, \bar{X}_N)$. $\tilde{\nu} \leftarrow$ the law of $\bar{Z}$.

12 **for** $i = 1, \ldots, N$ **do**

13 $\hat{\gamma}_i \leftarrow$ the law of $(\bar{X}_i, Z)$, $\hat{\alpha}_i \leftarrow \mathbb{E}[c_i(\bar{X}_i, Z)]$. $\tilde{\gamma}_i \leftarrow$ the law of $(\bar{X}_i, \bar{Z})$, $\tilde{\alpha}_i \leftarrow \mathbb{E}[c_i(\bar{X}_i, \bar{Z})]$.

14 $\alpha^{\mathsf{LB}}_{\mathsf{MT}} \leftarrow \alpha^{\mathsf{LB}}_{\mathsf{MT}_{\mathsf{par}}}$, $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}} \leftarrow \sum_{i=1}^{N} \hat{\alpha}_i$, $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}} \leftarrow \sum_{i=1}^{N} \tilde{\alpha}_i$.

15 $\hat{\epsilon}_{\mathsf{sub}} \leftarrow \hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}} - \alpha^{\mathsf{LB}}_{\mathsf{MT}}$, $\tilde{\epsilon}_{\mathsf{sub}} \leftarrow \tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}} - \alpha^{\mathsf{LB}}_{\mathsf{MT}}$.

16 **return** $(\tilde{\varphi}_i)_{i=1:N}$, $\hat{\nu}$, $(\hat{\gamma}_i)_{i=1:N}$, $\tilde{\nu}$, $(\tilde{\gamma}_i)_{i=1:N}$, $\alpha^{\mathsf{LB}}_{\mathsf{MT}}$, $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$, $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$, $\hat{\epsilon}_{\mathsf{sub}}$, $\tilde{\epsilon}_{\mathsf{sub}}$.

Remark 3.6 (Details of Algorithm 2)**.** *Let Assumption 3.2 hold. Below is a list explaining the inputs to Algorithm 2.*

- $(\mathcal{X}_i)_{i=1:N}$, $\mathcal{Z}$, $(\mu_i)_{i=1:N}$, *and* $(c_i)_{i=1:N}$ *are given by Assumption 1.1.*
- $\epsilon_{\mathsf{LSIP}} > 0$ *is a pre-specified error tolerance value when calling Algorithm 1 in Line 3 (see Proposition 3.5 and Theorem 3.7).*
- $(\mathcal{G}_i)_{i=1:N}$ *contain test functions for* $\mu_1, \ldots, \mu_N$ *that satisfy Assumption 3.2(ii).* $\mathcal{H}$ *contains test functions on* $\mathcal{Z}$ *that satisfy Assumption 3.2(iii). The choice of* $(\mathcal{G}_i)_{i=1:N}$ *and* $\mathcal{H}$ *controls the sub-optimality of the outputs of Algorithm 2 (see Theorem 3.7).*
- $\tilde{z} : \boldsymbol{\mathcal{X}} \to \mathcal{Z}$ *is a Borel measurable function which satisfies (2.6).*
- $\texttt{Oracle}(\cdot,\cdot,\cdot)$ *is the global minimization oracle in Definition 2.4.*

The list below provides further explanations of some lines in Algorithm 2.

- *Line 2 constructs finite sets* $\left(\mathcal{C}_i^{(0)} \subseteq \mathcal{X}_i \times \mathcal{Z}\right)_{i=1:N}$ *which are then used as input to Algorithm 1. This is possible due to Proposition 3.5(i), Assumption 3.2(ii), and Assumption 3.2(iii).*
- *Line 6 requires the support of* $\hat{\nu}_{\hat{i}}$ *to be finite, which is guaranteed by Proposition 3.5(v).*
- *Line 8 constructs an optimal coupling between* $\hat{\nu}_{\hat{i}}$ *and* $\hat{\nu}_i$ *via Lemma 3.1. This is possible since the assumption (C1) is satisfied due to the finite support of* $\hat{\nu}_i$*. Moreover, the random variable* $X_i : \Omega \to \mathcal{X}_i$ *in Line 9 is well-defined due to the finite support of* $\hat{\nu}_i$*.*
- *Line 10 constructs an optimal coupling between the law of* X_i *(which is equal to the marginal of* $\hat{\theta}_i$ *on* $\mathcal{X}_i$*) and* μ_i *via Lemma 3.1. This is possible due to Proposition 3.5(v) and Assumption 3.2(i).*

Theorem 3.7 (Properties of Algorithm 2)**.** *Let Assumption 3.2 hold. Assume in addition that all inputs of Algorithm 2 are set according to Remark 3.6. Moreover, let* $\left(\overline{\overline{W}}_{\mu_i,\mathcal{G}_i}\right)_{i=1:N}$ *satisfy* $\overline{\overline{W}}_{\mu_i,\mathcal{G}_i} \geq \overline{W}_{\mu_i,\mathcal{G}_i}$ *for* $i = 1,\dots,N$*, let* $\overline{\overline{W}}_{\mathcal{H}}$ *satisfy* $\overline{\overline{W}}_{\mathcal{H}} \geq \overline{W}_{\mathcal{H}}$*, and let* ϵ_{theo} *be defined as*

$$\epsilon_{\mathsf{theo}} := \epsilon_{\mathsf{LSIP}} + \left(\sum_{i=1}^{N} L_{c_i}^{(1)} \overline{\overline{W}}_{\mu_i,\mathcal{G}_i}\right) + \left(\sum_{i\neq\hat{\imath}} L_{c_i}^{(2)}\right) \overline{\overline{W}}_{\mathcal{H}},$$

where $\hat{\imath} \in \{1,\dots,N\}$ *is chosen in Line 5. Then, the following statements hold.*

(i) $(\tilde{\varphi}_i)_{i=1:N}$ *is an* $\tilde{\epsilon}_{\mathsf{sub}}$*-optimizer of* (MT*) *and for* $i = 1,\dots,N-1$*,* $\tilde{\varphi}_i$ *is* $L_{c_i}^{(2)}$*-Lipschitz continuous.*

(ii) $\hat{\nu}$ *is an* $\hat{\epsilon}_{\mathsf{sub}}$*-optimizer of* (MT)*.*

(iii) For $i = 1,\dots,N$*,* $\hat{\gamma}_i \in \Gamma(\mu_i,\hat{\nu})$ *and* $\int_{\mathcal{X}_i\times\mathcal{Z}} c_i \,\mathrm{d}\hat{\gamma}_i \leq W_{c_i}(\mu_i,\hat{\nu}) + \hat{\epsilon}_{\mathsf{sub}}$*.*

(iv) $\tilde{\nu}$ *is an* $\tilde{\epsilon}_{\mathsf{sub}}$*-optimizer of* (MT)*.*

(v) For $i = 1,\dots,N$*,* $\tilde{\gamma}_i \in \Gamma(\mu_i,\tilde{\nu})$ *and* $\int_{\mathcal{X}_i\times\mathcal{Z}} c_i \,\mathrm{d}\tilde{\gamma}_i \leq W_{c_i}(\mu_i,\tilde{\nu}) + \tilde{\epsilon}_{\mathsf{sub}}$*.*

(vi) $\alpha_{\mathsf{MT}}^{\mathsf{LB}} < (\mathsf{MT}) < \tilde{\alpha}_{\mathsf{MT}}^{\mathsf{UB}} < \hat{\alpha}_{\mathsf{MT}}^{\mathsf{UB}}$ *and* $\tilde{\epsilon}_{\mathsf{sub}} \leq \hat{\epsilon}_{\mathsf{sub}} \leq \epsilon_{\mathsf{theo}}$*.*

In particular, $\left((\tilde{\varphi}_i)_{i=1:N},(\hat{\gamma}_i)_{i=1:N},\hat{\nu}\right)$ *constitute an* $\hat{\epsilon}_{\mathsf{sub}}$*-approximate matching equilibrium and* $\left((\tilde{\varphi}_i)_{i=1:N},(\tilde{\gamma}_i)_{i=1:N},\tilde{\nu}\right)$ *constitute an* $\tilde{\epsilon}_{\mathsf{sub}}$*-approximate matching equilibrium.*

Next, assume in addition to the above assumptions that Assumption 2.17 also holds. Moreover, suppose that for arbitrary $\bar{\epsilon}_0 > 0, \bar{\epsilon}_1 > 0, \dots, \bar{\epsilon}_N > 0$*, one can construct* $(\mathfrak{C}_i)_{i=1:N}$*,* $(\mathcal{G}_i)_{i=1:N}$*,* $\mathfrak{C}_0$*,* $\mathcal{H}$ *via Setting 2.19 such that* $\epsilon_0 \leq \bar{\epsilon}_0$*,* $\mathrm{int}(\mathcal{Z}) \cap \mathrm{int}(C)$ *for all* $C \in \mathfrak{C}_0$*, and for* $i = 1,\dots,N$*,* $\epsilon_i \leq \bar{\epsilon}_i$*,* $\mathrm{int}(\mathcal{X}_i) \cap \mathrm{int}(C) \neq \emptyset$ *for all* $C \in \mathfrak{C}_i$*. Then,*

(vii) for any $\epsilon > 0$ *and* $\epsilon_{\mathsf{LSIP}} \in (0,\epsilon)$*, one can construct* $(\mathcal{G}_i)_{i=1:N}$*,* $\mathcal{H}$ *such that* $\tilde{\epsilon}_{\mathsf{sub}} \leq \hat{\epsilon}_{\mathsf{sub}} \leq \epsilon$*.*

Remark 3.8. *A concrete case in which the assumptions of Theorem 3.7(vii) are satisfied is when each of* $\mathcal{X}_1,\dots,\mathcal{X}_N,\mathcal{Z}$ *can be expressed as a union of simplices that are interior-disjoint. This follows from [54, Proposition 3.9].*

Remark 3.9 (Sub-optimality estimates in Algorithm 2 and a priori upper bound)**.** *From a theoretical perspective, Theorem 3.7(vii) states that, for any given* $\epsilon > 0$*, there exist explicit choices of the inputs* ϵ_{LSIP}*,* $(\mathcal{G}_i)_{i=1:N}$*, and* $\mathcal{H}$ *of Algorithm 2 such that* $\left((\tilde{\varphi}_i)_{i=1:N},(\hat{\gamma}_i)_{i=1:N},\hat{\nu}\right)$ *and* $\left((\tilde{\varphi}_i)_{i=1:N},(\tilde{\gamma}_i)_{i=1:N},\tilde{\nu}\right)$ *computed by Algorithm 2 are* ϵ*-approximate matching equilibria. One such choice is to let* $\epsilon_{\mathsf{LSIP}} \in (0,\epsilon)$ *be arbitrary and then let* $\mathcal{G}_1,\dots,\mathcal{G}_N,\mathcal{H}$ *be defined as in Setting 2.19 such that* $\epsilon_i \leq \frac{\epsilon-\epsilon_{\mathsf{LSIP}}}{2NL_{c_i}^{(1)}}$ *for* $i = 1,\dots,N$ *and* $\epsilon_0 \leq \frac{\epsilon-\epsilon_{\mathsf{LSIP}}}{2(N-1)L_{\bar{c}}^{(2)}}$ *(where* $L_{\bar{c}}^{(2)} := \max_{1\leq i\leq N}\left\{L_{c_i}^{(2)}\right\}$*). However, in practice, one often specifies* $\epsilon_{\mathsf{LSIP}} > 0$*,* $(\mathcal{G}_i)_{i=1:N}$*, and* $\mathcal{H}$ *in the inputs of Algorithm 2 and subsequently uses the values of* $\hat{\epsilon}_{\mathsf{sub}}$ *and* $\tilde{\epsilon}_{\mathsf{sub}}$ *in the output to estimate the sub-optimality of the computed solutions. The term* ϵ_{theo} *in Theorem 3.7(vi) provides a theoretical upper bound for the computed sub-optimality estimates* $\hat{\epsilon}_{\mathsf{sub}}$ *and* $\tilde{\epsilon}_{\mathsf{sub}}$*.* ϵ_{theo} *is referred to as an a priori upper bound for* $\hat{\epsilon}_{\mathsf{sub}}$ *and* $\tilde{\epsilon}_{\mathsf{sub}}$ *since it is based on the upper estimates* $\left(\overline{\overline{W}}_{\mu_i,\mathcal{G}_i}\right)_{i=1:N}$ *of* $\left(\overline{W}_{\mu_i,\mathcal{G}_i}\right)_{i=1:N}$ *and the upper estimate* $\overline{\overline{W}}_{\mathcal{H}}$ *of* $\overline{W}_{\mathcal{H}}$ *that can be computed independent of Algorithm 2. The computed sub-optimality estimates* $\hat{\epsilon}_{\mathsf{sub}}$ *and* $\tilde{\epsilon}_{\mathsf{sub}}$ *are typically much less conservative than their a priori upper bound* ϵ_{theo}*, as we will demonstrate in the numerical experiments in Section 4.*

4. Numerical experiments

In this section, we perform three numerical experiments to demonstrate the numerical algorithm (i.e., Algorithm 2) that we have developed. The first experiment in Section 4.1 studies the business location distribution problem introduced in Application 1.3 where $\mathcal{X}_1,\dots,\mathcal{X}_N,\mathcal{Z} \subset \mathbb{R}^2$ to demonstrate the convergence of the approximate matching equilibria constructed by Algorithm 2 to true matching equilibria. The second experiment in Section 4.2 examines the performance of our algorithm for the 2-Wasserstein barycenter problem in Application 1.4 where $\mathcal{X}_1,\dots,\mathcal{X}_N,\mathcal{Z} \subset \mathbb{R}^2$. The third experiment in Section 4.3 investigates the empirical scalability of our algorithm in a problem

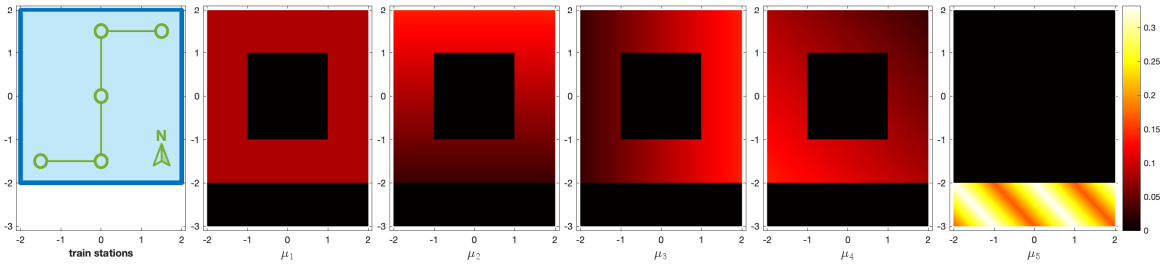

FIGURE 4.1. Experiment 1 – The leftmost panel shows the railway line and the locations of the train stations in the city. The rest of the panels show the probability density functions of $\mu_1, \ldots, \mu_N$.

where $\mathcal{X}_1, \ldots, \mathcal{X}_N \subset \mathbb{R}$ and $\mathcal{Z} \subset \mathbb{R}^2$. The code used in the numerical experiments is available on GitHub[9].

4.1. **Experiment 1: business location distribution.** In our first numerical experiment, we study the business location distribution problem in Application 1.3. Let us consider a square-shaped city and a type of business which hires 4 categories of employees, that is, this matching for teams problem involves 5 categories of agents and $N = 5$. The costs for the employees to travel on foot in the city are given by the city block distance, i.e., the travel cost between two locations $\boldsymbol{x} = (x_1, x_2)^\mathsf{T}$ and $\boldsymbol{x}' = (x_1', x_2')^\mathsf{T}$ is $\|\boldsymbol{x} - \boldsymbol{x}'\|_1 = |x_1 - x_1'| + |x_2 - x_2'|$. Moreover, there is a railway line that runs through the city with 5 train stations at locations $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_5 \in \mathbb{R}^2$ (see the leftmost panel of Figure 4.1) where the travel cost between consecutive train stations is 0.1. Therefore, for $i = 1, \ldots, 4$, we define the commuting cost of the i-th category of employees from $\boldsymbol{x}_i$ to $\boldsymbol{z}$ by

$$c_i(\boldsymbol{x}_i, \boldsymbol{z}) := 0.15 \left(\min_{1 \le j \le 5,\, 1 \le k \le 5} \left\{ \|\boldsymbol{x}_i - \boldsymbol{u}_j\|_1 + \|\boldsymbol{z} - \boldsymbol{u}_k\|_1 + 0.1|j - k| \right\} \wedge \|\boldsymbol{x}_i - \boldsymbol{z}\|_1 \right),$$

where 0.15 is the weight factor associated to the commuting costs. On the other hand, we define the restocking cost of business outlets to be a scaled city block distance $c_5(\boldsymbol{x}_5, \boldsymbol{z}) := 0.4\|\boldsymbol{x}_5 - \boldsymbol{z}\|_1$.

The specific settings of this experiment are detailed below.

Assumption 4.1. *We assume that the following statements hold.*

- $N = 5$. *For* $i = 1, \ldots, 4$, $\mathcal{X}_i = [-2, 2]^2 \setminus (-1, 1)^2 \subset \mathbb{R}^2$, *and* $\mathcal{X}_5 = [-2, 2] \times [-3, -2] \subset \mathbb{R}^2$. *Moreover,* $d_{\mathcal{X}_i}(\boldsymbol{x}_i, \boldsymbol{x}_i') := \|\boldsymbol{x}_i - \boldsymbol{x}_i'\|_2$ *for* $i = 1, \ldots, 5$.
- $\mathcal{Z} = [-2, 2]^2 \subset \mathbb{R}^2$ *and* $d_{\mathcal{Z}}(\boldsymbol{z}, \boldsymbol{z}') := \|\boldsymbol{z} - \boldsymbol{z}'\|_2$.
- *For* $i = 1, \ldots, 5$, $\mu_i \in \mathcal{P}(\mathcal{X}_i)$ *is absolutely continuous with respect to the Lebesgue measure on* $\mathcal{X}_i$ *and* $\mathrm{supp}(\mu_i) = \mathcal{X}_i$. *The density of* μ_i *is a continuous piece-wise affine and positive function on* $\mathcal{X}_i$.
- *For* $i = 1, \ldots, 4$, $c_i : \mathcal{X}_i \times \mathcal{Z} \to \mathbb{R}$ *is given by* $c_i(\boldsymbol{x}_i, \boldsymbol{z}) := 0.15\Big(\min_{1 \le j \le 5,\, 1 \le k \le 5} \left\{ \|\boldsymbol{x}_i - \boldsymbol{u}_j\|_1 + \|\boldsymbol{z} - \boldsymbol{u}_k\|_1 + 0.1|j - k| \right\} \wedge \|\boldsymbol{x}_i - \boldsymbol{z}\|_1 \Big)$. $c_5 : \mathcal{X}_5 \times \mathcal{Z} \to \mathbb{R}$ *is given by* $c_5(\boldsymbol{x}_5, \boldsymbol{z}) := 0.4\|\boldsymbol{x}_5 - \boldsymbol{z}\|_1$.
- *The test functions* $\mathcal{G}_1, \ldots, \mathcal{G}_N, \mathcal{H}$ *are constructed via Setting 2.19 such that Assumption 3.2(ii) and Assumption 3.2(iii) are satisfied.*

Figure 4.1 shows the shape of the sets $\mathcal{X}_1, \ldots, \mathcal{X}_N$, as well as the probability density functions of $\mu_1, \ldots, \mu_N$ as color plots (see the color bar on the right for the scale). In order to compute approximate matching equilibria for this business location distribution problem, we fix $\epsilon_{\mathsf{LSIP}} = 2 \times 10^{-4}$ and test 5 different combinations of test functions $\mathcal{G}_1, \ldots, \mathcal{G}_N, \mathcal{H}$. Specifically, the test functions in $\mathcal{G}_1, \ldots, \mathcal{G}_N$ are defined based on increasingly finer partitions of $\mathcal{X}_1, \ldots, \mathcal{X}_N$ into triangles, and the value of $|\mathcal{G}_1| + \cdots + |\mathcal{G}_N|$ ranges between 109 and 14284. Moreover, the test functions in $\mathcal{H}$ are

[9] https://github.com/qikunxiang/MatchingForTeams

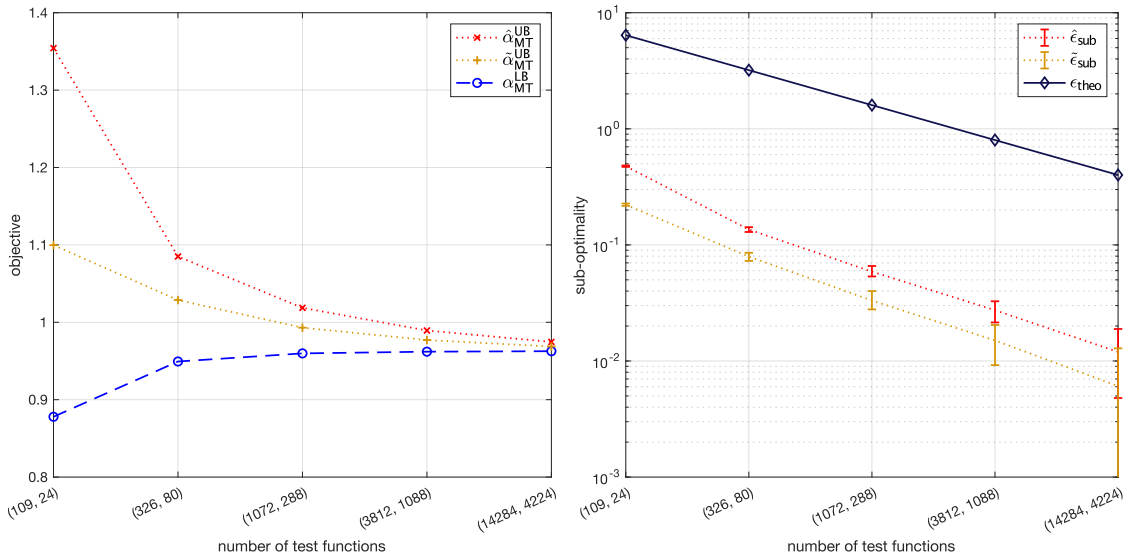

FIGURE 4.2. Experiment 1 – Left: the lower bound $\alpha^{\mathsf{LB}}_{\mathsf{MT}}$ and the upper bounds $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$, $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ for the optimal value of (MT) computed by Algorithm 2. Right: the computed sub-optimality estimates $\hat{\epsilon}_{\mathsf{sub}}$, $\tilde{\epsilon}_{\mathsf{sub}}$ and their a priori upper bound ϵ_{theo} on the log-scale. The tuples on the bottom indicate the number of test functions $(|\mathcal{G}_1| + \cdots + |\mathcal{G}_N|, |\mathcal{H}|)$.

defined based on increasingly finer partitions of $\mathcal{Z}$ into triangles, and the value of $|\mathcal{H}|$ ranges between 24 and 4224. The values of $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$, $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ in Line 14 of Algorithm 2 are computed via Monte Carlo integration using 10^4 independent random samples. Moreover, each Monte Carlo integration is repeated 100 times in order to examine the Monte Carlo error.

Figure 4.2 shows the computed upper and lower bounds for (MT), the sub-optimality estimates $\hat{\epsilon}_{\mathsf{sub}}$, $\tilde{\epsilon}_{\mathsf{sub}}$, and their a priori theoretical upper bound ϵ_{theo} derived from Theorem 3.7(vi). The left panel of Figure 4.2 shows the upper bounds $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$, $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ and the lower bound $\alpha^{\mathsf{LB}}_{\mathsf{MT}}$ computed by Algorithm 2. The horizontal axis shows the number of test functions used, i.e., each tuple corresponds to $\big(|\mathcal{G}_1| + \cdots + |\mathcal{G}_N|, |\mathcal{H}|\big)$. It can be seen that $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ is considerably better than $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$. The differences between the upper bounds $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ and $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ and the lower bound $\alpha^{\mathsf{LB}}_{\mathsf{MT}}$ are large when $|\mathcal{G}_1|+\cdots+|\mathcal{G}_N| = 109$ and $|\mathcal{H}| = 24$. When $|\mathcal{G}_1| + \cdots + |\mathcal{G}_N|$ and $|\mathcal{H}|$ increase, the difference between the bounds decreases. The right panel of Figure 4.2 compares the computed sub-optimality estimates $\hat{\epsilon}_{\mathsf{sub}}$ and $\tilde{\epsilon}_{\mathsf{sub}}$ with their a priori theoretical upper bound ϵ_{theo} on the log-scale. The error bars indicate the Monte Carlo errors in the computation of the upper bounds $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ and $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$. The results show that the value of ϵ_{theo} is around 10 to 60 times larger than $\hat{\epsilon}_{\mathsf{sub}}$ and $\tilde{\epsilon}_{\mathsf{sub}}$. This demonstrates that not only does Algorithm 2 produce feasible solutions of (MT), (MT*), and (MT$_{\mathsf{cp}}$), it also produces sub-optimality estimates $\hat{\epsilon}_{\mathsf{sub}}$, $\tilde{\epsilon}_{\mathsf{sub}}$ of these feasible solutions that are much less conservative than suggested by an a priori theoretical analysis, as discussed in Remark 3.9. Specifically, when $|\mathcal{G}_1|+\cdots+|\mathcal{G}_N| = 14284$ and $|\mathcal{H}| = 4224$, the value of ϵ_{theo} is equal to 0.4002, which indicates that the solutions computed by Algorithm 2 could be poor. Despite that, the values of the computed sub-optimality estimates $\hat{\epsilon}_{\mathsf{sub}}$ and $\tilde{\epsilon}_{\mathsf{sub}}$ are 0.0120 and 0.0061, respectively. By Theorem 2.16, this shows that $\big((\tilde{\varphi}_i)_{i=1:N}, (\hat{\gamma}_i)_{i=1:N}, \hat{\nu}\big)$ and $\big((\tilde{\varphi}_i)_{i=1:N}, (\tilde{\gamma}_i)_{i=1:N}, \tilde{\nu}\big)$ computed by Algorithm 2 are approximate matching equilibria that are close to true matching equilibria.

Finally, Figure 4.3, Figure 4.4, and Figure 4.5 illustrate $\hat{\nu}$, $\tilde{\nu}$, $(\tilde{\gamma}_i)_{i=1:N}$, and $(\tilde{\varphi}_i)_{i=1:N}$ from the outputs of Algorithm 2. The top row of Figure 4.3 shows $\hat{\nu}$ as bubble plots, where the locations of the red circles represent the atoms in $\hat{\nu}$, and the size and opacity of each circle represents the probability of each atom. The bottom row of Figure 4.3 shows $\tilde{\nu}$ as grayscale color plots superimposed with bubble plots. The reason for this choice is that for every combination of test functions, $\tilde{\nu}$ is a mixed probability measure containing a discrete component with two atoms and a non-discrete component.

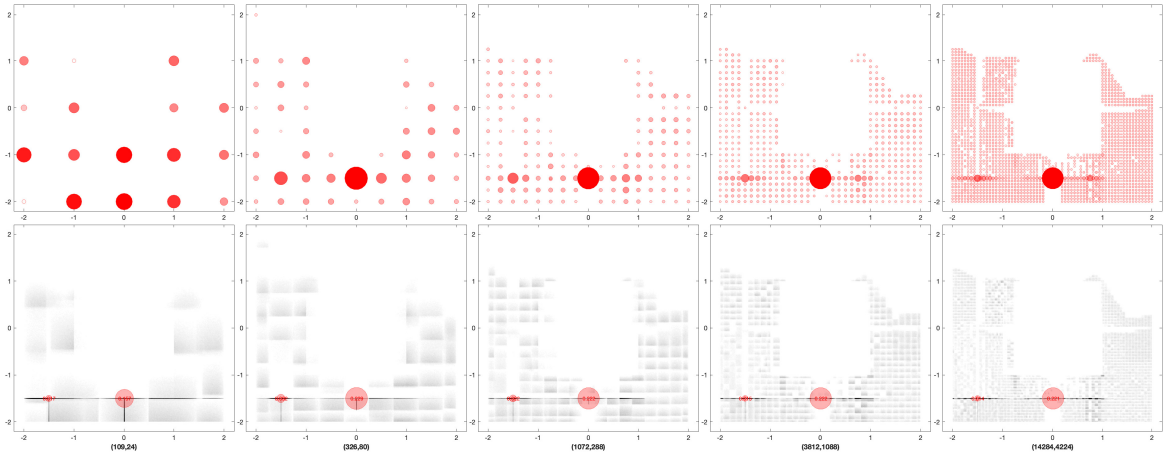

FIGURE 4.3. Experiment 1 – The probability measures $\hat{\nu}$ and $\tilde{\nu}$ computed by Algorithm 2. The top row shows $\hat{\nu}$ as bubble plots. The bottom row shows $\tilde{\nu}$ as color plots (in grayscale) superimposed with bubble plots. The tuple on the bottom of each column indicates the number of test functions $(|\mathcal{G}_1| + \cdots + |\mathcal{G}_N|, |\mathcal{H}|)$.

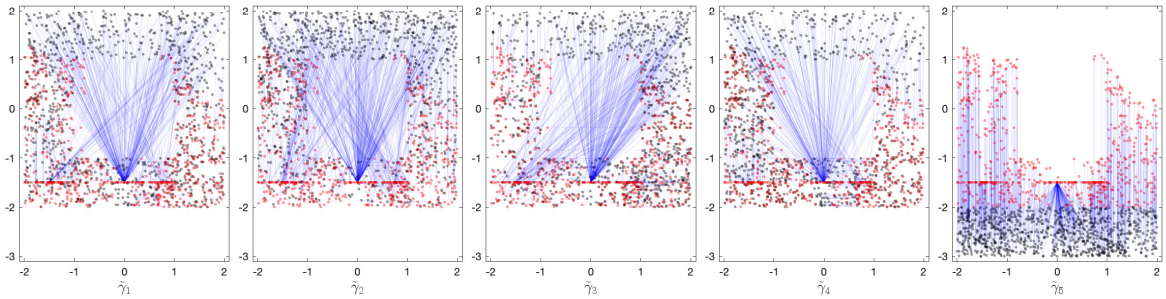

FIGURE 4.4. Experiment 1 – 1000 Coupled samples from the probability measures $(\tilde{\gamma}_i)_{i=1:N}$ computed by Algorithm 2. The black dots and red dots represent the locations of employees/suppliers and business outlets, respectively. The blue lines connecting the black and red dots indicate the coupling of the samples.

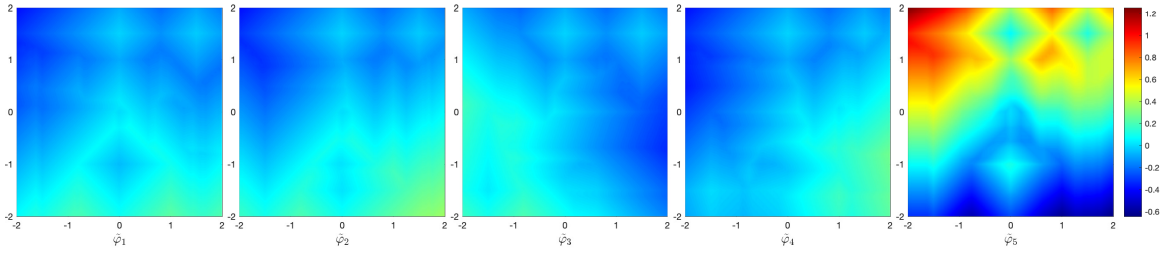

FIGURE 4.5. Experiment 1 – Color plots showing the approximately optimal transfer functions $(\tilde{\varphi}_i)_{i=1:N}$ computed by Algorithm 2.

The probabilities of the two atoms in $\tilde{\nu}$ are shown as text in the bubble plot, while the histograms of the non-discrete part of $\tilde{\nu}$ are shown as grayscale color plots. Moreover, it can be observed from the bottom row of Figure 4.3 that there is some probability in $\tilde{\nu}$ concentrated on a horizontal line. Figure 4.4 shows 1000 coupled samples from the approximately optimal couplings $(\tilde{\gamma}_i)_{i=1:N}$ computed by Algorithm 2 when $|\mathcal{G}_1| + \cdots + |\mathcal{G}_N| = 14284$ and $|\mathcal{H}| = 4224$, where the black dots and red dots represent the locations of $\boldsymbol{x}_i$ and $\boldsymbol{z}$ in the samples, and the blue lines connecting the dots indicate the coupling between the locations in the samples. Samples from $(\hat{\gamma}_i)_{i=1:N}$ look very similar to those

from $(\tilde{\gamma}_i)_{i=1:N}$ and are thus not shown here. These coupled samples illustrate how the employees in each category choose the business outlets to work at, as well as how the business outlets choose the suppliers to restock from. Figure 4.5 shows the transfer functions $(\tilde{\varphi}_i)_{i=1:N}$ computed by Algorithm 2 in color plots (see the color bar on the right for the scale). Recall that $\tilde{\varphi}_5 = -(\tilde{\varphi}_1 + \tilde{\varphi}_2 + \tilde{\varphi}_3 + \tilde{\varphi}_4)$ by our construction in (2.9). The following insights can be drawn from Figure 4.3, Figure 4.4, and Figure 4.5.

- As the number of test functions increases, the structures of both the discrete probability measure $\hat{\nu}$ and the non-discrete probability measure $\tilde{\nu}$ increase in sophistication.
- As discussed in Remark 2.15, $\hat{\nu}$ can be interpreted as an approximate matching equilibrium where the business outlets are only located at finitely many locations. In particular, it can be observed that $\hat{\nu}$ is sparser than indicated by the theoretical result in Corollary 2.14. When $|\mathcal{G}_1| + \cdots + |\mathcal{G}_N| = 14284$ and $|\mathcal{H}| = 4224$, the two atoms in $\hat{\nu}$ with the largest probabilities are $(0, -1.5)^\mathsf{T}$ and $(-1.5, -1.5)^\mathsf{T}$, which correspond to the two train stations in the south.
- $\tilde{\nu}$ represents an approximate matching equilibrium where business outlets are distributed over uncountably many locations. It shows some highly non-trivial features including the presence of two atoms at $(0, -1.5)^\mathsf{T}$ and $(-1.5, -1.5)^\mathsf{T}$, non-zero probability on a one-dimensional subspace, and an absolutely continuous component. Moreover, it can be observed from Figure 4.2 that $\tilde{\nu}$ is considerably more optimal than $\hat{\nu}$. This demonstrates that, when the business is close to an equilibrium state, around 22.1% of its outlets will be located at the train station at $(0, -1.5)^\mathsf{T}$, around 1.4% of its outlets will be location at the train station at $(-1.5, -1.5)^\mathsf{T}$, and the rest of its outlets will not be concentrated at specific locations but will instead be dispersed into a continuum of locations, where a considerable portion of outlets will be dispersed along a line segment passing through the two train stations.
- The approximately optimal couplings $\tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}_3, \tilde{\gamma}_4$ demonstrate how the presence of the train in the city impacts the workplace choices of the employees. Since the train connects the northern part of the city with the southern part, many residents of in the northern part of the city will commute to the southern part by train. On the other hand, some residents in the western and eastern parts of the city that live far from train stations will work at business outlets that are nearby.
- The approximately optimal transfer functions $\tilde{\varphi}_1, \tilde{\varphi}_2, \tilde{\varphi}_3, \tilde{\varphi}_4$ can be interpreted as "salary maps" up to the addition of constants, which indicate the amount of salary paid out to each category of employees at each business outlet location. Using the "salary map" $\tilde{\varphi}_i$, the i-th category of employees can determine an approximately optimal workplace location based on where they reside, through solving the minimization problem $\min_{\boldsymbol{z} \in \mathcal{Z}} \{c_i(\boldsymbol{x}_i, \boldsymbol{z}) - \tilde{\varphi}_i(\boldsymbol{z})\}$. On the other hand, $\tilde{\varphi}_5 = -(\tilde{\varphi}_1 + \tilde{\varphi}_2 + \tilde{\varphi}_3 + \tilde{\varphi}_4)$ corresponds to the negative of the total salary paid out to the employees at each business outlet location. Due to all the suppliers being located in the southern part of the city, business outlets that are located in the northern part of the city need to spend more on restocking and thus will pay lower salaries. Besides this observation, it can be seen from Figure 4.5 that the approximately optimal transfer functions are continuous piece-wise affine and exhibit complex patterns due to the presence of the train stations.

Using the computed approximate matching equilibria, business owners can get insights for determining how to optimally set up the business outlets. Moreover, they can aid city planners in analyzing the effects of transportation infrastructures, e.g., highways, railways, on shaping the geographic structure of business outlets at equilibrium. For example, in the particular problem instance that we have analyzed in this subsection, the experimental results above offer the following recommendations.

- The railway operators can use the computed approximately optimal couplings $(\tilde{\gamma}_i)_{i=1:N}$ to gauge the demand for the train services at each train station. Therefore, the railway operations can be planned accordingly to cater to the commuters and avoid congestion.
- A second railway line could be added to the city to connect residents in the western and eastern parts of the city to the southern part due to the proximity of the southern part to the suppliers.

The effects of a second railway line may be analyzed by incorporating it into the cost functions $(c_i)_{i=1:N}$ and subsequently analyzing the resulting new equilibrium. This can potentially increase the overall economic efficiency of the city.

- Since a computed approximate matching equilibrium only reflects an approximately optimal structure of the city at equilibrium, the actual structure of the city may differ from a computed equilibrium. Therefore, the relevant decision makers can utilize the computed approximate matching equilibrium to implement additional policies to incentivize and facilitate the shift towards the optimal structure in order to improve the overall economic efficiency of the city.

4.2. **Experiment 2: 2-Wasserstein barycenter.** In the second numerical experiment, we study the well-known 2-Wasserstein barycenter problem for two-dimensional probability measures $\mu_1, \dots, \mu_N$, which has been introduced in Application 1.4. Specifically, we study two instances of the problem. The first instance considers the case where $\mu_1, \dots, \mu_N$ are from the same *location-scatter family* as defined by Álvarez-Esteban et al. [6, Definition 2.1]. In this case, an approximate optimizer of (MT) can be computed to high precision with a fixed-point scheme. We treat the approximate optimizer and its corresponding objective value computed this way as the ground truths, and we compare the values of $\hat{\alpha}_{\mathsf{MT}}^{\mathsf{UB}}$, $\tilde{\alpha}_{\mathsf{MT}}^{\mathsf{UB}}$, $\alpha_{\mathsf{MT}}^{\mathsf{LB}}$, and $\tilde{\nu}$ computed by Algorithm 2 with them to show the correctness of our numerical approach. In the second instance, we consider the general case where $\mu_1, \dots, \mu_N$ are not from the same *location-scatter family*, and we showcase the performance of Algorithm 2 for approximately computing a 2-Wasserstein barycenter of $\mu_1, \dots, \mu_N$.

4.2.1. *Location-scatter family.* In the first problem instance, we take $\mu_1, \dots, \mu_N$ from the same *location-scatter family* as defined by Álvarez-Esteban et al. [6] below.

Definition 4.2 (Location-scatter family of probability measures [6, Definition 2.1])**.** *Let $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$ be a probability measure with zero mean and identity covariance matrix. The location-scatter family generated by μ_0 is defined as the following family of probability measures:*

$$\mathcal{F}(\mu_0) := \left\{\mu_0 \circ L^{-1} : L : \mathbb{R}^d \ni \boldsymbol{x} \mapsto \mathbf{A}\boldsymbol{x} + \boldsymbol{b} \in \mathbb{R}^d,\ \mathbf{A} \in \mathbb{S}_{++}^d,\ \boldsymbol{b} \in \mathbb{R}^d\right\},$$

where $\mathbb{S}_{++}^d$ denotes the set of $d \times d$ symmetric positive definite matrices.

Below is a list presenting the settings of this problem instance.

Assumption 4.3. *We assume that the following statements hold.*

- *$\mu_0 \in \mathcal{P}(\mathbb{R}^2)$ is supported on a square and is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$. It is the mixture of 3 bivariate Gaussian measures truncated to the square (see the leftmost panel of Figure 4.6 for its probability density function).*
- *$N = 5$. For $i = 1, \dots, 5$, $\mu_i \in \mathcal{F}(\mu_0)$. μ_i is restricted to its support $\operatorname{supp}(\mu_i) =: \mathcal{X}_i$ and treated as an element of $\mathcal{P}(\mathcal{X}_i)$. Moreover, $d_{\mathcal{X}_i}(\boldsymbol{x}_i, \boldsymbol{x}_i') := \|\boldsymbol{x}_i - \boldsymbol{x}_i'\|_2$.*
- *$\mathcal{Z} = \operatorname{conv}(\mathcal{X}_1 \cup \cdots \cup \mathcal{X}_5)$ and $d_{\mathcal{Z}}(\boldsymbol{z}, \boldsymbol{z}') := \|\boldsymbol{z} - \boldsymbol{z}'\|_2$.*
- *For $i = 1, \dots, 5$, $c_i : \mathcal{X}_i \times \mathcal{Z} \to \mathbb{R}$ is given by $c_i(\boldsymbol{x}_i, \boldsymbol{z}) := \frac{1}{5}\big(\|\boldsymbol{z}\|_2^2 - 2\langle \boldsymbol{x}_i, \boldsymbol{z}\rangle\big) = \frac{1}{5}\|\boldsymbol{x}_i - \boldsymbol{z}\|_2^2 - \frac{1}{5}\|\boldsymbol{x}_i\|_2^2$.*
- *The test functions $\mathcal{G}_1, \dots, \mathcal{G}_N, \mathcal{H}$ are constructed via Setting 2.19 such that Assumption 3.2(ii) and Assumption 3.2(iii) are satisfied.*

Figure 4.6 shows the probability density functions of $\mu_0, \mu_1, \dots, \mu_N$ as color plots (see the color bar on the right for the scale). The above problem instance considers the 2-Wasserstein barycenter of $\mu_1, \dots, \mu_N$ with equal weights $\lambda_1 = \cdots = \lambda_N = \frac{1}{N}$. Note that for $i = 1, \dots, N$, the quadratic term $\frac{1}{N}\|\boldsymbol{x}_i\|_2^2$ has been subtracted from the cost function c_i since it does not affect the optimal solution of (MT). Since $\mu_1, \dots, \mu_N \in \mathcal{F}(\mu_0)$, (MT) has a unique optimizer which is the 2-Wasserstein barycenter of $\mu_1, \dots, \mu_N$, and this barycenter also belongs to $\mathcal{F}(\mu_0)$ [6, Corollary 4.5]. Moreover, an approximate optimizer of (MT) can be computed to high precision with a fixed-point scheme; see [6, Corollary 4.5]. Therefore, we denote the optimal value of (MT) computed by this fixed-point scheme by α_{fp} and treat it as the ground truth, and we treat the approximate optimizer of (MT) computed this way as the true 2-Wasserstein barycenter of $\mu_1, \dots, \mu_N$.

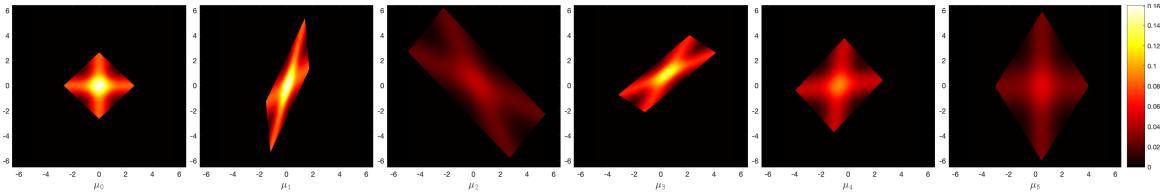

FIGURE 4.6. Experiment 2 (location-scatter family) – The leftmost panel shows the probability density function of the initial probability measure μ_0 that defines a location-scatter family $\mathcal{F}(\mu_0)$. The rest of the panels show the probability density functions of the probability measures $\mu_1, \ldots, \mu_N$ which all belong to the location-scatter family generated by μ_0.

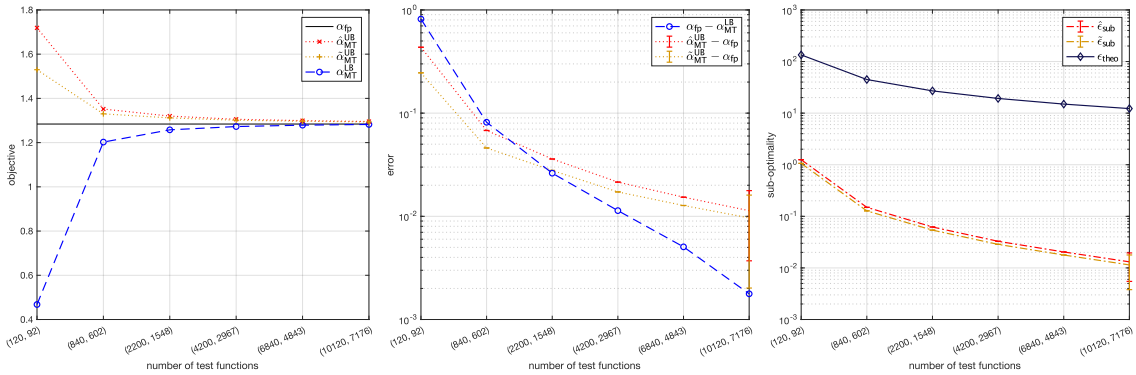

FIGURE 4.7. Experiment 2 (location-scatter family) – Left: the lower bound $\alpha^{\mathsf{LB}}_{\mathsf{MT}}$ and the upper bounds $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$, $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ for the optimal value of (MT) computed by Algorithm 2 as well as the true optimal value α_{fp} computed via the fixed-point scheme of Álvarez-Esteban et al. [6]. Center: the differences between the computed bounds $\alpha^{\mathsf{LB}}_{\mathsf{MT}}$, $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$, $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ and the true optimal value α_{fp} on the log-scale. Right: the computed sub-optimality estimates $\hat{\epsilon}_{\mathsf{sub}}$, $\tilde{\epsilon}_{\mathsf{sub}}$ and their a priori upper bound ϵ_{theo} on the log-scale. The tuples on the bottom indicate the number of test functions $(|\mathcal{G}_1| + \cdots + |\mathcal{G}_N|, |\mathcal{H}|)$.

To approximate the 2-Wasserstein barycenter, we set $\epsilon_{\mathsf{LSIP}} = 10^{-3}$ and test 6 different combinations of test functions $\mathcal{G}_1, \ldots, \mathcal{G}_N, \mathcal{H}$. Similar to Experiment 1, the test functions in $\mathcal{G}_1, \ldots, \mathcal{G}_N, \mathcal{H}$ are defined based on increasingly finer triangular partitions of $\mathcal{X}_1, \ldots, \mathcal{X}_N, \mathcal{Z}$ with $|\mathcal{G}_1| + \cdots + |\mathcal{G}_N|$ ranging between 120 and 10120 and $|\mathcal{H}|$ ranging between 92 and 7176. The values of $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$, $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ are computed via Monte Carlo integration using 10^6 independent random samples, and each Monte Carlo integration is repeated 100 times to examine the Monte Carlo error.

Figure 4.7 shows the computed upper and lower bounds for (MT), the sub-optimality estimates $\hat{\epsilon}_{\mathsf{sub}}$, $\tilde{\epsilon}_{\mathsf{sub}}$, and their a priori theoretical upper bound ϵ_{theo} derived from Theorem 3.7(vi). Note that we have added a constant term $\sum_{i=1}^N \int_{\mathcal{X}_i} \frac{1}{N} \|\boldsymbol{x}_i\|_2^2 \, \mu_i(\mathrm{d}\boldsymbol{x}_i)$ to the objective of the matching for teams problem. This is to add back the subtracted term $\frac{1}{N}\|\boldsymbol{x}_i\|_2^2$ from the cost function c_i for better interpretability of the objective values. The left panel of Figure 4.7 shows the upper bounds $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$, $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ and the lower bound $\alpha^{\mathsf{LB}}_{\mathsf{MT}}$ computed by Algorithm 2. It also shows the value of α_{fp} computed via the fixed-point scheme for reference. The horizontal axis shows the number of test functions used, i.e., each tuple corresponds to $\big(|\mathcal{G}_1| + \cdots + |\mathcal{G}_N|, |\mathcal{H}|\big)$. The computed bounds look largely similar to the left panel of Figure 4.2 where all bounds improve as the number of test function increases. However, it can be seen that the two upper bounds $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ and $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ are comparable in this instance, with $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ being only slightly smaller than $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$. Since the ground truth α_{fp} is known for this problem, we show the differences between the computed bounds and α_{fp} on the log-scale in the center

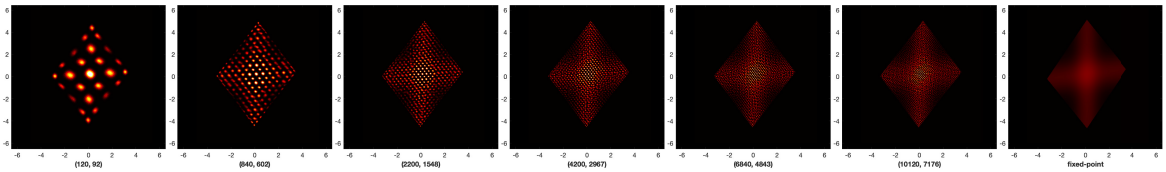

FIGURE 4.8. Experiment 2 (location-scatter family) – Histograms of the approximate 2-Wasserstein barycenter $\tilde{\nu}$ computed by Algorithm 2. The tuples on the bottom indicate the number of test functions $(|\mathcal{G}_1| + \cdots + |\mathcal{G}_N|, |\mathcal{H}|)$. The rightmost panel shows the probability density function of the true 2-Wasserstein barycenter computed via the fixed-point scheme of Álvarez-Esteban et al. [6].

panel of Figure 4.7, where the error bars show the Monte Carlo errors in the computation of the upper bounds. When $|\mathcal{G}_1| + \cdots + |\mathcal{G}_N| = 10120$ and $|\mathcal{H}| = 7176$, the value of $\alpha_{\mathsf{fp}} - \alpha_{\mathsf{MT}}^{\mathsf{LB}}$ is 0.0018, and the values of $\hat{\alpha}_{\mathsf{MT}}^{\mathsf{UB}} - \alpha_{\mathsf{fp}}$ and $\tilde{\alpha}_{\mathsf{MT}}^{\mathsf{UB}} - \alpha_{\mathsf{fp}}$ are 0.0113 and 0.0097, respectively. This shows that the lower bound $\alpha_{\mathsf{MT}}^{\mathsf{LB}}$ is close to being optimal, and the computed sub-optimality estimates $\hat{\epsilon}_{\mathsf{sub}}$, $\tilde{\epsilon}_{\mathsf{sub}}$ are dominated by the sub-optimality of the computed upper bounds. This also shows that the computed approximate 2-Wasserstein barycenter $\tilde{\nu}$ is not far from the true 2-Wasserstein barycenter. The right panel of Figure 4.7 compares the computed sub-optimality estimates $\hat{\epsilon}_{\mathsf{sub}}$ and $\tilde{\epsilon}_{\mathsf{sub}}$ with their a priori theoretical upper bound ϵ_{theo} on the log-scale, with error bars indicating the Monte Carlo error in the upper bounds. Here, the value of ϵ_{theo} is around 100 to 1000 times larger than $\hat{\epsilon}_{\mathsf{sub}}$ and $\tilde{\epsilon}_{\mathsf{sub}}$. This demonstrates that the sub-optimality estimates $\hat{\epsilon}_{\mathsf{sub}}$, $\tilde{\epsilon}_{\mathsf{sub}}$ of the computed approximate 2-Wasserstein barycenters are much less conservative than suggested by an a priori theoretical analysis, as discussed in Remark 3.9.

Figure 4.8 shows the histograms of the approximate 2-Wasserstein barycenter $\tilde{\nu}$ as well as the true 2-Wasserstein barycenter computed via the fixed-point scheme. One can observe that $\tilde{\nu}$ is a continuous probability measure that approximates the true 2-Wasserstein barycenter by a finite number of "blobs". This is due to the reassembly and binding steps that were carried out in Lines 7–11 of Algorithm 2. The continuous probability measure $\tilde{\nu}$ is formed via a sequence of "gluing" operations that "glue together" the discrete probability measure $\hat{\nu}$ with the continuous probability measures $\mu_1, \ldots, \mu_N$ via the gluing lemma; see, e.g., [68, Lemma 7.6]. When $|\mathcal{G}_1| + \cdots + |\mathcal{G}_N| = 10120$ and $|\mathcal{H}| = 7176$, the histogram of $\tilde{\nu}$ looks very similar to the probability density function of the true 2-Wasserstein barycenter. This is in agreement with the computed sub-optimality estimate $\tilde{\epsilon}_{\mathsf{sub}}$, whose value is 0.0114. Overall, this problem instance serves as a sanity check and illustrates that our algorithm has correctly computed the lower and upper bounds for the optimal value of (MT) as our theoretical results have demonstrated. In Section 4.2.2, we will study a general problem instance in which the fixed-point scheme of Álvarez-Esteban et al. [6] is not applicable and the ground truth optimal value of (MT) is unavailable.

4.2.2. *The general case.* Next, we examine a general instance of the 2-Wasserstein barycenter problem, where $\mu_1, \ldots, \mu_N$ do not belong to the same *location-scatter family*. The detailed settings of this problem instance are shown below.

Assumption 4.4. *We assume that the following statements hold.*

- $N = 20$. *For* $i = 1, \ldots, 20$, $\mathcal{X}_i$ *is a parallelogram in* $\mathbb{R}^2$, $\mu_i \in \mathcal{P}(\mathcal{X}_i)$ *is absolutely continuous with respect to the Lebesgue measure on* $\mathcal{X}_i$. *It is the mixture of 2 or 3 bivariate Gaussian measures truncated to* $\mathcal{X}_i$. *Moreover,* $d_{\mathcal{X}_i}(\boldsymbol{x}_i, \boldsymbol{x}_i') := \|\boldsymbol{x}_i - \boldsymbol{x}_i'\|_2$.
- $\mathcal{Z} = \mathrm{conv}(\mathcal{X}_1 \cup \cdots \cup \mathcal{X}_{20})$ *and* $d_{\mathcal{Z}}(\boldsymbol{z}, \boldsymbol{z}') := \|\boldsymbol{z} - \boldsymbol{z}'\|_2$.
- *For* $i = 1, \ldots, 20$, $c_i : \mathcal{X}_i \times \mathcal{Z} \to \mathbb{R}$ *is given by* $c_i(\boldsymbol{x}_i, \boldsymbol{z}) := \frac{1}{20}\big(\|\boldsymbol{z}\|_2^2 - 2\langle \boldsymbol{x}_i, \boldsymbol{z}\rangle\big) = \frac{1}{20}\|\boldsymbol{x}_i - \boldsymbol{z}\|_2^2 - \frac{1}{5}\|\boldsymbol{x}_i\|_2^2$.
- *The test functions* $\mathcal{G}_1, \ldots, \mathcal{G}_N, \mathcal{H}$ *are chosen via Setting 2.19 such that Assumption 3.2(ii) and Assumption 3.2(iii) are satisfied.*

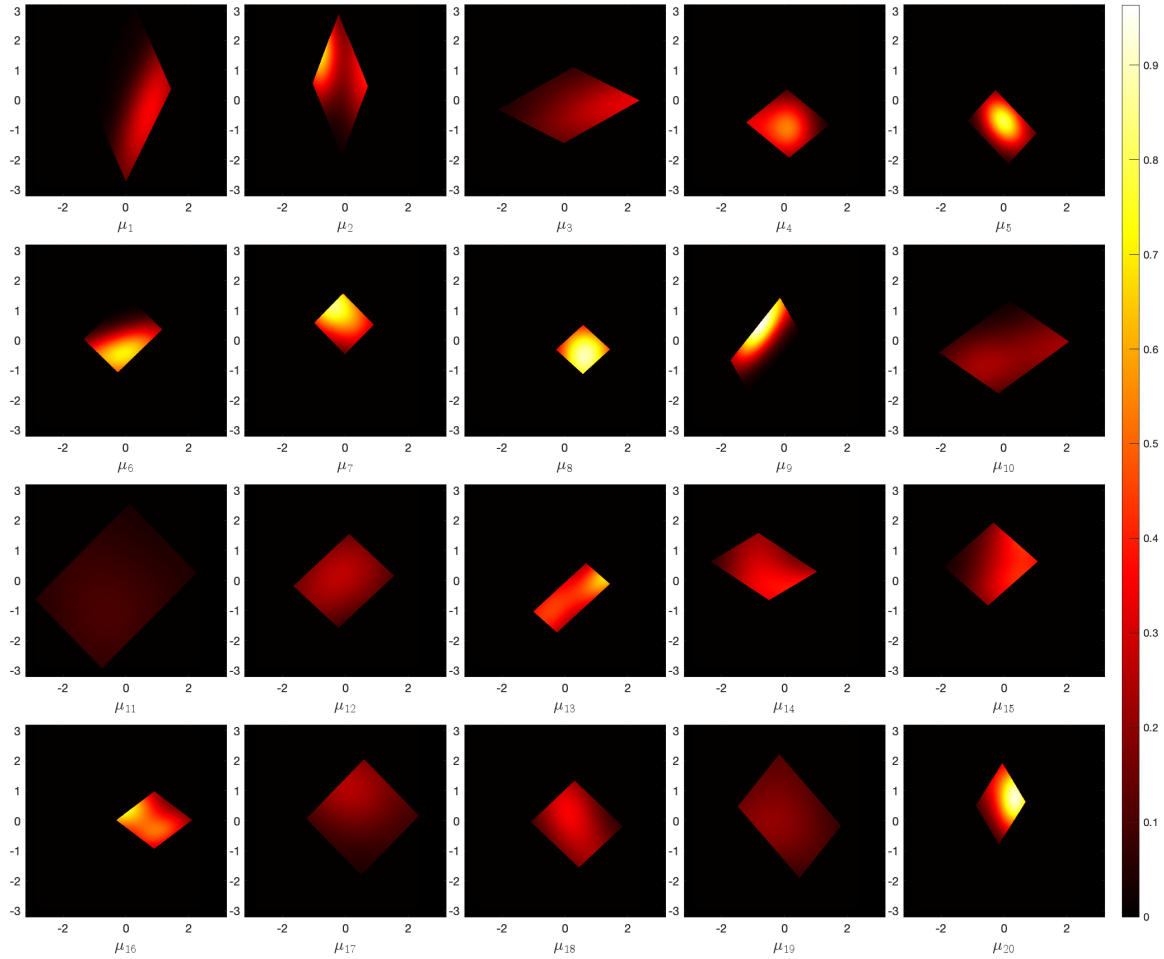

FIGURE 4.9. Experiment 2 (general case) – Probability density functions of $\mu_1, \dots, \mu_N$.

Figure 4.9 shows the probability density functions of $\mu_1, \dots, \mu_N$ as color plots (see the color bar on the right for the scale). Same as the previous problem instance, this problem also considers the 2-Wasserstein barycenter of $\mu_1, \dots, \mu_N$ with equal weights $\lambda_1 = \cdots = \lambda_N = \frac{1}{N}$. In this case, *the fixed-point scheme of Álvarez-Esteban et al. [6] does not apply and there is no known method for computing the true 2-Wasserstein barycenter of* $\mu_1, \dots, \mu_N$.

To approximate the 2-Wasserstein barycenter, we fix $\epsilon_{\mathsf{LSIP}} = 2 \times 10^{-4}$ and test 6 different combinations of test functions $\mathcal{G}_1, \dots, \mathcal{G}_N, \mathcal{H}$. Same as before, the test functions in $\mathcal{G}_1, \dots, \mathcal{G}_N, \mathcal{H}$ are defined based on increasingly finer triangular partitions of $\mathcal{X}_1, \dots, \mathcal{X}_N, \mathcal{Z}$ with $|\mathcal{G}_1| + \cdots + |\mathcal{G}_N|$ ranging between 700 and 19200 and $|\mathcal{H}|$ ranging between 76 and 1869. The values of $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$, $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ are computed via Monte Carlo integration using 10^6 independent random samples, and each Monte Carlo integration is repeated 100 times to examine the Monte Carlo error.

Figure 4.10 shows the computed upper and lower bounds for (MT), the sub-optimality estimates $\hat{\epsilon}_{\mathsf{sub}}$, $\tilde{\epsilon}_{\mathsf{sub}}$, and their a priori theoretical upper bound ϵ_{theo} derived from Theorem 3.7(vi). The left panel of Figure 4.10 shows the upper bounds $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$, $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ and the lower bound $\alpha^{\mathsf{LB}}_{\mathsf{MT}}$ computed by Algorithm 2. The horizontal axis shows the number of test functions used, i.e., each tuple corresponds to $\big(|\mathcal{G}_1| + \cdots + |\mathcal{G}_N|, |\mathcal{H}|\big)$. Observe that the results look similar to Figure 4.7. The right panel of Figure 4.10 compares the computed sub-optimality estimates $\hat{\epsilon}_{\mathsf{sub}}$ and $\tilde{\epsilon}_{\mathsf{sub}}$ with their a priori theoretical upper bound ϵ_{theo} on the log-scale, with error bars indicating the Monte Carlo error in the upper bounds. Similar to Figure 4.7, the value of ϵ_{theo} is around 100 to 1000 times larger than $\hat{\epsilon}_{\mathsf{sub}}$ and $\tilde{\epsilon}_{\mathsf{sub}}$. Again, the sub-optimality estimates $\hat{\epsilon}_{\mathsf{sub}}$, $\tilde{\epsilon}_{\mathsf{sub}}$ of the computed approximate 2-Wasserstein

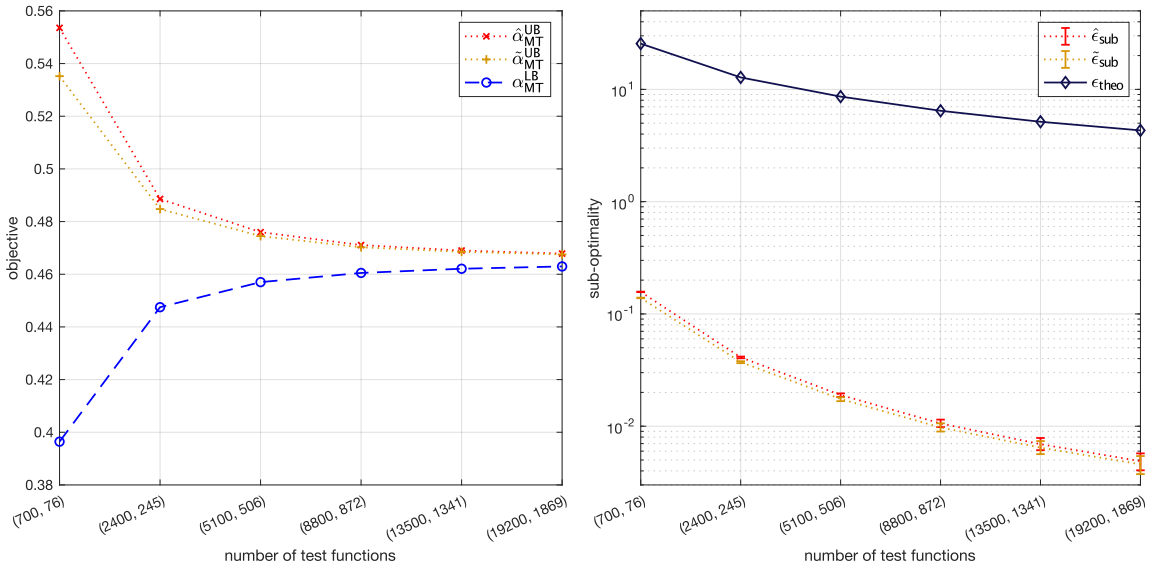

FIGURE 4.10. Experiment 2 (general case) – Left: the lower bound $\alpha^{\mathsf{LB}}_{\mathsf{MT}}$ and the upper bounds $\hat{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$, $\tilde{\alpha}^{\mathsf{UB}}_{\mathsf{MT}}$ for the optimal value of (MT) computed by Algorithm 2. Right: the computed sub-optimality estimates $\hat{\epsilon}_{\mathsf{sub}}$, $\tilde{\epsilon}_{\mathsf{sub}}$ and their a priori upper bound ϵ_{theo} on the log-scale. The tuples on the bottom indicate the number of test functions $(|\mathcal{G}_1| + \cdots + |\mathcal{G}_N|, |\mathcal{H}|)$.

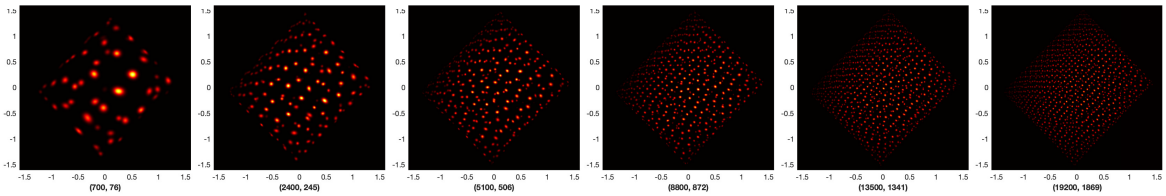

FIGURE 4.11. Experiment 2 (general case) – Histograms of the approximate 2-Wasserstein barycenter $\tilde{\nu}$ computed by Algorithm 2. The tuples on the bottom indicate the number of test functions $(|\mathcal{G}_1| + \cdots + |\mathcal{G}_N|, |\mathcal{H}|)$.

barycenters are much less conservative than suggested by an a priori theoretical analysis, as discussed in Remark 3.9.

Figure 4.11 shows the histograms of the approximate 2-Wasserstein barycenter $\tilde{\nu}$. Similar to Figure 4.8, one can observe that $\tilde{\nu}$ is a continuous probability measure that approximates the true 2-Wasserstein barycenter by a finite number of "blobs". Since the value of the computed sub-optimality estimate $\tilde{\epsilon}_{\mathsf{sub}}$ is 0.0046, the approximate 2-Wasserstein barycenter $\tilde{\nu}$ when $|\mathcal{G}_1| + \cdots + |\mathcal{G}_N| = 19200$ and $|\mathcal{H}| = 1869$ is close to the true 2-Wasserstein barycenter.

4.3. **Experiment 3: one-dimensional type spaces.** In the third numerical experiment, we examine the scalability of Algorithm 1 in terms of how its empirical running time changes with the number N of agent categories in the matching for teams problem. To that end, let us study the following matching for teams problem with one-dimensional type spaces, i.e., $\mathcal{X}_1, \ldots, \mathcal{X}_N \subset \mathbb{R}$ and a two-dimensional quality space, i.e., $\mathcal{Z} \subset \mathbb{R}^2$.

Example 4.5. *The matching for teams problem is specified as follows.*

- *For $i = 1, \ldots, N$, the type space of the i-th category of agents is $\mathcal{X}_i = [0, 1] \subset \mathbb{R}$. The type space $\mathcal{X}_i$ represents a scalar-valued preference variable of agents in the i-th category that is between 0 and 1. Moreover, $d_{\mathcal{X}_i}(x_i, x_i') := |x_i - x_i'|$.*

- *The quality space $\mathcal{Z}$ is given by $\mathcal{Z} = \{(z_1, z_2)^\mathsf{T} \in \mathbb{R}^2 : z_1 \geq 0,\ z_2 \geq 0,\ z_1 + z_2 \leq 1\} \subset \mathbb{R}^2$. Thus, each good is characterized by two non-negative quality variables whose sum is less than or equal to 1. 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 Lebesgue measure on $\mathcal{X}_i$, and its density is a continuous piece-wise affine and positive function on $\mathcal{X}_i$. μ_i represents the distribution of the preference variable within the i-th agent category.*
- *For $i = 1, \ldots, N$, the cost function $c_i : \mathcal{X}_i \times \mathcal{Z} \to \mathbb{R}$ is given by $c_i(x_i, \boldsymbol{z}) := \frac{1}{N} l_i\big(|x_i - \langle \boldsymbol{s}_i, \boldsymbol{z}\rangle|\big)$, where $\boldsymbol{s}_i \in \mathbb{R}^2$, $\|\boldsymbol{s}_i\|_2 = 1$, and*

$$l_i(t) := \begin{cases} 0 & \text{if } 0 \leq t \leq \theta_{i,1}, \\ t - \theta_{i,1} & \text{if } \theta_{i,1} < t \leq \theta_{i,2}, \\ \theta_{i,2} - \theta_{i,1} & \text{if } t > \theta_{i,2}. \end{cases}$$

  Here, the vector $\boldsymbol{s}_i$ represents the weights these agents use when assessing the goods based on the two quality variables in $\mathcal{Z}$. An agent evaluates a good with quality $\boldsymbol{z} \in \mathcal{Z}$ by comparing her assessment $\langle \boldsymbol{s}_i, \boldsymbol{z}\rangle$ of the good and her preference variable $x_i \in \mathcal{X}_i$. The cost function $c_i(x_i, \boldsymbol{z})$ is equal to 0 if the absolute difference $\big|x_i - \langle \boldsymbol{s}_i, \boldsymbol{z}\rangle\big|$ is below a threshold $\theta_{i,1}$. Moreover, the cost grows linearly when the absolute difference $\big|x_i - \langle \boldsymbol{s}_i, \boldsymbol{z}\rangle\big|$ is above the threshold $\theta_{i,1}$ but below a second threshold $\theta_{i,2}$, and the cost remains constant when $\big|x_i - \langle \boldsymbol{s}_i, \boldsymbol{z}\rangle\big|$ exceeds the second threshold $\theta_{i,2}$. The factor $\frac{1}{N}$ in c_i guarantees that the magnitude of the matching for teams problem (e.g., the magnitude of the optimal value of (MT)) does not increase with the number N of agent categories. In particular, this allows us to set an a priori theoretical sub-optimality upper bound ϵ_{theo} based on Theorem 3.7 that remains constant for all values of N.
- *The test functions $\mathcal{G}_1, \ldots, \mathcal{G}_N, \mathcal{H}$ are constructed via Setting 2.19 such that Assumption 3.2(ii) and Assumption 3.2(iii) are satisfied. It holds that $m_i = 49$ for all i, and it holds that $k = 560$.*

In order to investigate the performance of our method, we generate 10 problem instances (or scenarios) of Example 4.5, where for $i = 1, \ldots, N$ the probability density function of μ_i, the vector $\boldsymbol{s}_i$, and the two thresholds $\theta_{i,1}, \theta_{i,2}$ are independently randomly generated.

In this experiment, we fix $\epsilon_{\mathsf{LSIP}} = 5 \times 10^{-5}$. The global minimization problem $\texttt{Oracle}(\cdot, \cdot, \cdot)$ in Algorithm 1 is formulated into a mixed-integer programming problem[10] and are subsequently solved by the mixed-integer solver of Gurobi [42]. We would like to remark that the global minimization problems $\texttt{Oracle}\big(1, \boldsymbol{y}_1^{(r)}, \boldsymbol{w}_1^{(r)}\big), \ldots, \texttt{Oracle}\big(N, \boldsymbol{y}_N^{(r)}, \boldsymbol{w}_N^{(r)}\big)$ in Line 5 of Algorithm 1 can be solved in parallel. However, we choose to solve them sequentially in our implementation of Algorithm 1 in order not to over-complicate the running time analysis.

We apply Algorithm 2 to the 10 randomly generated problem instances and record the computed values of the sub-optimality estimate $\hat{\epsilon}_{\mathsf{sub}}$ as well as the running time of Algorithm 1 for $N = 4, 6, 8, 10, 12, 14, 16, 18, 20, 50, 80, 100$ agent categories. $\hat{\epsilon}_{\mathsf{sub}}$ is computed via Monte Carlo integration using 10^7 independent samples. We only examine the values of $\hat{\epsilon}_{\mathsf{sub}}$ here because $\tilde{\epsilon}_{\mathsf{sub}} \leq \hat{\epsilon}_{\mathsf{sub}}$ and the computation of $\tilde{\epsilon}_{\mathsf{sub}}$ requires us to solve a global minimization problem (2.5), which is more computationally costly than the computation of $\hat{\epsilon}_{\mathsf{sub}}$. In addition, we only examine the running time of Line 3 in Algorithm 2, i.e., the running time of Algorithm 1. The running time of the rest of Algorithm 2 consists mostly of time spent computing $\hat{\alpha}_{\mathsf{MT}}^{\mathsf{UB}}$ via Monte Carlo integration in Line 13, which can be parallelized and is negligible compared to the running time of Line 3 when N is large.

The left part of Table 4.1 shows the average and maximum values of the sub-optimality estimate $\hat{\epsilon}_{\mathsf{sub}}$ and the average and maximum running time of Algorithm 1 when $N = 4, 6, 8, 10, 12, 14, 16, 18, 20, 50, 80, 100$. From the results, it can be seen that the values of $\hat{\epsilon}_{\mathsf{sub}}$ computed by our algorithm are about two orders of magnitude smaller than the a priori theoretical upper bound $\epsilon_{\mathsf{theo}} = 0.1293$.

[10] Additional details about the mixed-integer programming formulation can be found in the online appendix on GitHub: https://github.com/qikunxiang/MatchingForTeams.

TABLE 4.1. Experiment 3 – Computed sub-optimality estimate, running time of Algorithm 1, and support sparsity (note that $\epsilon_{\mathsf{theo}} = 0.1293$).

N	Avg. $\hat{\epsilon}_{\mathsf{sub}}$ $[\times 10^{-4}]$	Max. $\hat{\epsilon}_{\mathsf{sub}}$ $[\times 10^{-4}]$	Avg. time [s]	Max. time [s]	Avg. $\lvert\mathrm{supp}(\hat{\nu})\rvert$	Max. $\lvert\mathrm{supp}(\hat{\nu})\rvert$	$\min_{1\le i\le N}\{m_i\}$ $+k+2$
4	9.771	26.625	537	1123	87.1	104	611
6	10.636	23.440	1283	2438	104.8	137	611
8	7.319	11.492	1958	2781	99.9	132	611
10	8.419	16.667	2820	3542	113.0	161	611
12	8.557	20.180	3834	4377	114.8	168	611
14	8.291	17.612	4695	5513	118.3	176	611
16	7.461	15.005	5826	7629	116.1	170	611
18	6.999	12.227	6462	7398	109.3	164	611
20	6.869	12.815	6884	8009	120.6	203	611
50	5.464	6.968	17462	19717	140.1	175	611
80	4.927	6.355	33949	38496	151.1	187	611
100	4.822	6.412	42348	49769	156.8	205	611

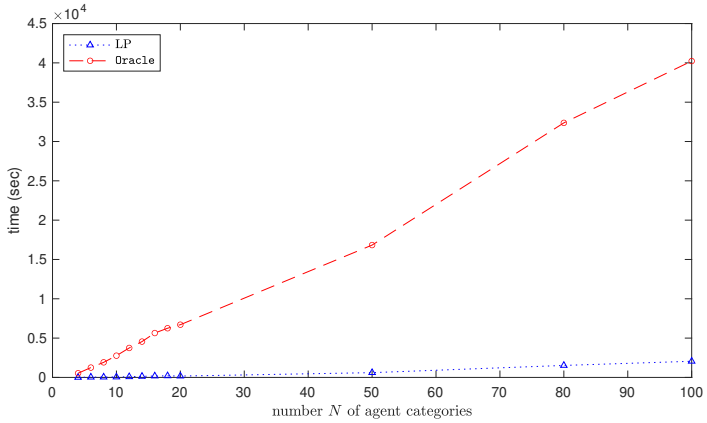

FIGURE 4.12. Experiment 3 – Running time of the LP solver (Line 3) and the global minimization oracle (Line 5) in Algorithm 1.

Figure 4.12 shows the total running time of the LP solver in Line 3 of Algorithm 1, and the total running time of $\texttt{Oracle}(\cdot,\cdot,\cdot)$ in Line 5 of Algorithm 1. It can be seen that the total running time of $\texttt{Oracle}(\cdot,\cdot,\cdot)$ in Algorithm 1 is much longer relative to the LP solver for all values of N, showing that almost all of the running time is spent on computing $\texttt{Oracle}(\cdot,\cdot,\cdot)$. An interesting observation is that the total running time of $\texttt{Oracle}(\cdot,\cdot,\cdot)$ seems to grow approximately linearly in N. This is in line with our discussion in Remark 2.7. Under the assumptions in Example 4.5, for $i = 1,\dots,N$ and for $\boldsymbol{y}_i \in \mathbb{R}^{m_i}$, $\boldsymbol{w}_i \in \mathbb{R}^k$, the global minimization problem $\texttt{Oracle}(\cdot,\cdot,\cdot)$ is given by

$$\begin{aligned} \underset{x_i,\,\boldsymbol{z}_i}{\text{minimize}} \quad & l_i\big(x_i - \langle \boldsymbol{s}_i, \boldsymbol{z}_i\rangle\big) - \langle \boldsymbol{g}_i(x_i), \boldsymbol{y}_i\rangle - \langle \boldsymbol{h}(\boldsymbol{z}_i), \boldsymbol{w}_i\rangle \\ \text{subject to} \quad & \underline{\kappa}_i \le x_i \le \overline{\kappa}_i, \\ & \boldsymbol{z}_i \in \mathcal{Z}, \end{aligned}$$

which corresponds to minimizing a continuous piece-wise affine function with 3 variables, since $\mathcal{X}_i \times \mathcal{Z} \subset \mathbb{R}^3$. Thus, the computational cost of this minimization indeed does not depend on N. Since each iteration of Algorithm 1 solves N such minimization problems, the total running time of $\texttt{Oracle}(\cdot,\cdot,\cdot)$ in each iteration of Algorithm 1 increases linearly in N. In addition, in Algorithm 1, the number of iterations till convergence grows slowly relative to the growth of N, which makes the total running time of $\texttt{Oracle}(\cdot,\cdot,\cdot)$ in Algorithm 1 grow approximately linearly in N. Moreover, this

shows that in a computing environment with sufficient parallelization capabilities, a suitable parallel implementation of the for-loop in Line 4 can drastically reduce the running time of Algorithm 1.

Remark 4.6. *The LP problem* ($\mathsf{MT}^{*(r)}_{\mathsf{par}}$) *that is solved in Line 3 of Algorithm 1 has a structure which allows it to be efficiently solved by parallel algorithms based on operator splitting methods; see, e.g., [32]. In this experiment, we use the standard LP solver provided by Gurobi [42] since it is reasonably efficient. The investigation of the parallelization aspects of Algorithm 1 is left as future work.*

Furthermore, recall that we have shown in Corollary 2.14 the existence of an approximate optimizer $\hat{\nu}$ of (MT) with $|\mathrm{supp}(\hat{\nu})| \leq \min_{1\leq i\leq N}\{m_i\} + k + 2$. The right part of Table 4.1 shows the average and maximum values of $|\mathrm{supp}(\hat{\nu})|$, where $\hat{\nu}$ is the discrete approximate optimizer of (MT) computed by Algorithm 2. It shows that $\hat{\nu}$ computed by Algorithm 2 is even more sparse than what Corollary 2.14 suggests, even though $|\mathrm{supp}(\hat{\nu})|$ still increases with N. A possible explanation of this phenomenon is as follows. As discussed by Carlier et al. [23, Section 2.2], one can restrict the quality space $\mathcal{Z}$ to any subset $\mathcal{Z}' \subseteq \mathcal{Z}$ satisfying $\mathcal{Z}' \supseteq \tilde{z}(\boldsymbol{\mathcal{X}})$ without affecting the optimal value of (MT). This suggests that there could be many test functions in $\mathcal{H}$ that are redundant since they are identical when their domains are restricted to a suitable choice of $\mathcal{Z}'$.

5. Proof of theoretical results

All proof of the theoretical results can be found in the appendix of the arXiv version of the paper at https://arxiv.org/abs/2308.03550.

References

[1] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM J. Math. Anal.*, 43(2): 904–924, 2011.

[2] A. Alfonsi, R. Coyaud, V. Ehrlacher, and D. Lombardi. Approximation of optimal transport problems with marginal moments constraints. *Math. Comp.*, 90(328):689–737, 2021.

[3] J. M. Altschuler and E. Boix-Adserà. Wasserstein barycenters can be computed in polynomial time in fixed dimension. *J. Mach. Learn. Res.*, 22(1):1532–4435, 2021.

[4] J. M. Altschuler and E. Boix-Adserà. Wasserstein barycenters are NP-hard to compute. *SIAM J. Math. Data Sci.*, 4(1):179–203, 2022.

[5] J. M. Altschuler and E. Boix-Adserà. Polynomial-time algorithms for multimarginal optimal transport problems with structure. *Math. Program.*, 199(1-2):1107–1178, 2023.

[6] P. C. Álvarez-Esteban, E. del Barrio, J. A. Cuesta-Albertos, and C. Matrán. A fixed-point approach to barycenters in Wasserstein space. *J. Math. Anal. Appl.*, 441(2):744–762, 2016.

[7] E. Anderes, S. Borgwardt, and J. Miller. Discrete Wasserstein barycenters: Optimal transport for discrete data. *Math. Methods Oper. Res.*, 84(2):389–409, 2016.

[8] F. A. Ba and M. Quellmalz. Accelerating the Sinkhorn algorithm for sparse multi-marginal optimal transport via fast Fourier transforms. *Algorithms*, 15(9):311, 2022.

[9] J.-D. Benamou. Optimal transportation, modelling and numerical simulation. *Acta Numer.*, 30: 249–325, 2021.

[10] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):A1111–A1138, 2015.

[11] D. P. Bertsekas and S. E. Shreve. *Stochastic optimal control: the discrete time case*, volume 139 of *Mathematics in Science and Engineering*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1978.

[12] O. Besbes, F. Castro, and I. Lobel. Surge pricing and its spatial supply response. *Management Science*, 67(3):1350–1367, 2021.

[13] J. Bigot, E. Cazelles, and N. Papadakis. Penalization of barycenters in the Wasserstein space. *SIAM J. Math. Anal.*, 51(3):2261–2285, 2019.

[14] A. Blanchet and G. Carlier. Optimal transport and Cournot-Nash equilibria. *Math. Oper. Res.*, 41(1):125–145, 2016.

[15] A. Blanchet, P. Mossay, and F. Santambrogio. Existence and uniqueness of equilibrium for a spatial model of social interactions. *Internat. Econom. Rev.*, 57(1):31–59, 2016.
[16] S. Borgwardt. An LP-based, strongly-polynomial 2-approximation algorithm for sparse Wasserstein barycenters. *Int. J. Oper. Res.*, 22(2):1511–1551, 2022.
[17] S. Borgwardt and S. Patterson. A column generation approach to the discrete barycenter problem. *Discrete Optim.*, 43:100674, 2022.
[18] S. Borgwardt and S. Patterson. An integer program for pricing support points of exact barycenters. *Preprint, arXiv:2210.14135*, 2022.
[19] G. Buttazzo and F. Santambrogio. A model for the optimal planning of an urban area. *SIAM J. Math. Anal.*, 37(2):514–530, 2005.
[20] G. Carlier and I. Ekeland. The structure of cities. *Journal of Global Optimization*, 29(4): 371–376, 2004.
[21] G. Carlier and I. Ekeland. Matching for teams. *Econom. Theory*, 42(2):397–418, 2010.
[22] G. Carlier and F. Santambrogio. A variational model for urban planning with traffic congestion. *ESAIM Control Optim. Calc. Var.*, 11(4):595–613, 2005.
[23] G. Carlier, A. Oberman, and E. Oudet. Numerical methods for matching for teams and Wasserstein barycenters. *ESAIM Math. Model. Numer. Anal.*, 49(6):1621–1642, 2015.
[24] S. Chewi, T. Maunu, P. Rigollet, and A. J. Stromme. Gradient descent algorithms for Bures-Wasserstein barycenters. In *Conference on Learning Theory*, pages 1276–1304. PMLR, 2020.
[25] L. Chizat. Doubly regularized entropic Wasserstein barycenters. *Preprint, arXiv:2303.11844*, 2023.
[26] S. Claici, E. Chien, and J. Solomon. Stochastic Wasserstein barycenters. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 999–1008. PMLR, 2018.
[27] S. Cohen, M. Arbel, and M. P. Deisenroth. Estimating barycenters of measures in high dimensions. *Preprint, arXiv:2007.07105*, 2020.
[28] D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progressions. *J. Symbolic Comput.*, 9(3):251–280, 1990.
[29] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, page 2292–2300, Red Hook, NY, USA, 2013.
[30] L. De Gennaro Aquino and C. Bernard. Bounds on multi-asset derivatives via neural networks. *Int. J. Theor. Appl. Finance*, 23(8):2050050, 31, 2020.
[31] L. De Gennaro Aquino and S. Eckstein. MinMax methods for optimal transport and beyond: Regularization, approximation and numerics. In *Advances in Neural Information Processing Systems*, volume 33, pages 13818–13830, 2020.
[32] J. Eckstein. A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multi-block alternating direction method of multipliers. *J. Optim. Theory Appl.*, 173(1):155–182, 2017.
[33] S. Eckstein and M. Kupper. Computation of optimal transport and related hedging problems via penalization and neural networks. *Appl. Math. Optim.*, 83(2):639–667, 2021.
[34] S. Eckstein and M. Nutz. Quantitative stability of regularized optimal transport and convergence of Sinkhorn's algorithm. *SIAM J. Math. Anal.*, 54(6):5922–5948, 2022.
[35] S. Eckstein, M. Kupper, and M. Pohl. Robust risk aggregation with neural networks. *Mathematical Finance*, 30(4):1229–1272, 2020.
[36] S. Eckstein, G. Guo, T. Lim, and J. Obłój. Robust pricing and hedging of options on multiple assets and its numerics. *SIAM J. Financial Math.*, 12(1):158–188, 2021.
[37] J. Fan, A. Taghvaei, and Y. Chen. Scalable computations of Wasserstein barycenter via input convex neural networks. *Preprint, arXiv:2007.04462*, 2020.
[38] G. Friesecke, A. S. Schulz, and D. Vögler. Genetic column generation: Fast computation of high-dimensional multimarginal optimal transport problems. *SIAM J. Sci. Comput.*, 44(3): A1632–A1654, 2022.

[39] D. Ge, H. Wang, Z. Xiong, and Y. Ye. Interior-point methods strike back: Solving the Wasserstein barycenter problem. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.
[40] M. A. Goberna and M. A. López. *Linear semi-infinite optimization*. John Wiley & Sons, 1998.
[41] G. Guo and J. Obłój. Computational methods for martingale optimal transport problems. *Ann. Appl. Probab.*, 29(6):3311–3347, 2019.
[42] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2024. URL http://www.gurobi.com.
[43] F. Heinemann, A. Munk, and Y. Zemel. Randomized Wasserstein barycenter computation: Resampling with statistical guarantees. *SIAM J. Math. Data Sci.*, 4(1):229–259, 2022.
[44] P. Henry-Labordère. (Martingale) optimal transport and anomaly detection with neural networks: A primal-dual algorithm. *Available at SSRN 3370910*, 2019.
[45] A. Korotin, L. Li, J. Solomon, and E. Burnaev. Continuous Wasserstein-2 barycenter estimation without minimax optimization. *Preprint, arXiv:2102.01752*, 2021.
[46] A. Korotin, V. Egiazarian, L. Li, and E. Burnaev. Wasserstein iterative networks for barycenter estimation. *Preprint, arXiv:2201.12245*, 2022.
[47] R. Krawtschenko, C. A. Uribe, A. Gasnikov, and P. Dvurechensky. Distributed optimization with quantization for computing Wasserstein barycenters. *Preprint, arXiv:2010.14325*, 2020.
[48] M. Kuang and E. G. Tabak. Sample-based optimal transport and barycenter problems. *Comm. Pure Appl. Math.*, 72(8):1581–1630, 2019.
[49] B. Lévy. A numerical algorithm for L_2 semi-discrete optimal transport in 3D. *ESAIM Math. Model. Numer. Anal.*, 49(6):1693–1715, 2015.
[50] L. Li, A. Genevay, M. Yurochkin, and J. M. Solomon. Continuous regularized Wasserstein barycenters. In *Advances in Neural Information Processing Systems*, volume 33, pages 17755–17765. Curran Associates, Inc., 2020.
[51] T. Lin, N. Ho, M. Cuturi, and M. I. Jordan. On the complexity of approximating multimarginal optimal transport. *J. Mach. Learn. Res.*, 23(65):1–43, 2022.
[52] R. E. Lucas, Jr. and E. Rossi-Hansberg. On the internal structure of cities. *Econometrica*, 70 (4):1445–1476, 2002.
[53] G. Luise, S. Salzo, M. Pontil, and C. Ciliberto. Sinkhorn barycenters with free support via Frank-Wolfe algorithm. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.
[54] 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.
[55] M. Nutz and J. Wiesel. Entropic optimal transport: convergence of potentials. *Probability Theory and Related Fields*, 2021.
[56] G. Peyré and M. Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.
[57] G. Puccetti, L. Rüschendorf, and S. Vanduffel. On the computation of Wasserstein barycenters. *J. Multivariate Anal.*, 176:104581, 16, 2020.
[58] J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its application to texture mixing. In *Scale Space and Variational Methods in Computer Vision*, pages 435–446, 2012.
[59] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems: Volume I: Theory*. Springer Science & Business Media, 1998.
[60] R. T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.
[61] J. Solomon, F. de Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas. Convolutional Wasserstein distances: efficient optimal transportation on geometric domains. *ACM Trans. Graph.*, 34(4), 2015.
[62] S. Srivastava, V. Cevher, Q. Dinh, and D. Dunson. WASP: Scalable bayes via barycenters of subset posteriors. In *Artificial Intelligence and Statistics*, pages 912–920. PMLR, 2015.

[63] S. Srivastava, C. Li, and D. B. Dunson. Scalable Bayes via barycenter in Wasserstein space. *J. Mach. Learn. Res.*, 19(1):312–346, 2018.

[64] M. Staib, S. Claici, J. M. Solomon, and S. Jegelka. Parallel streaming Wasserstein barycenters. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[65] E. G. Tabak, G. Trigila, and W. Zhao. Distributional barycenter problem through data-driven flows. *Pattern Recognition*, 130:108795, 2022.

[66] N. Tupitsa, P. Dvurechensky, A. Gasnikov, and C. A. Uribe. Multimarginal optimal transport by accelerated alternating minimization. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 6132–6137. IEEE, 2020.

[67] R. J. Vanderbei. *Linear programming—foundations and extensions*, volume 285 of *International Series in Operations Research & Management Science*. Springer, Cham, 2020. Fifth edition.

[68] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[69] J. von Lindheim. Simple approximative algorithms for free-support Wasserstein barycenters. *Comput. Optim. Appl.*, 85(1):213–246, 2023.

[70] Y. Xie, X. Wang, R. Wang, and H. Zha. A fast proximal point method for computing exact Wasserstein distance. In *Uncertainty in Artificial Intelligence*, pages 433–453. PMLR, 2020.

[71] L. Yang, J. Li, D. Sun, and K.-C. Toh. A fast globally linearly convergent algorithm for the computation of Wasserstein barycenters. *J. Mach. Learn. Res.*, 22:21–37, 2021.

[72] J. Ye and J. Li. Scaling up discrete distribution clustering using ADMM. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 5267–5271. IEEE, 2014.

[73] J. Ye, P. Wu, J. Z. Wang, and J. Li. Fast discrete distribution clustering using Wasserstein barycenter with sparse support. *IEEE Trans. Signal Process.*, 65(9):2317–2332, 2017.

[74] C. Zhang, H. Qian, and J. Xie. An asynchronous decentralized algorithm for Wasserstein barycenter problem. *Preprint, arXiv:2304.11653*, 2023.

DIVISION OF MATHEMATICAL SCIENCES, NANYANG TECHNOLOGICAL UNIVERSITY, 21 NANYANG LINK, 637371 SINGAPORE

Email address: ariel.neufeld@ntu.edu.sg

DIVISION OF MATHEMATICAL SCIENCES, NANYANG TECHNOLOGICAL UNIVERSITY, 21 NANYANG LINK, 637371 SINGAPORE

Email address: qikun.xiang@ntu.edu.sg