Decentralized and Equitable Optimal Transport

Ivan Lau, Shiqian Ma, and César A. Uribe

***Abstract*— This paper considers the decentralized (discrete) optimal transport (D-OT) problem. In this setting, a network of agents seeks to design a transportation plan jointly, where the cost function is the sum of privately held costs for each agent. We reformulate the D-OT problem as a constraint-coupled optimization problem and propose a *single-loop* decentralized algorithm with an iteration complexity of $O(1/\epsilon)$ that matches existing centralized first-order approaches. Moreover, we propose the decentralized *equitable* optimal transport (DE-OT) problem. In DE-OT, in addition to cooperatively designing a transportation plan that minimizes transportation costs, agents seek to ensure equity in their individual costs. The iteration complexity of the proposed method to solve DE-OT is also $O(1/\epsilon)$. This rate improves existing centralized algorithms, where the best iteration complexity obtained is $O(1/\epsilon^2)$.**

I. Introduction

Optimal Transport (OT) problem is a well-studied problem tracing back to the early work of Monge [1] and Kantorovich [2]. It has recently gained interest in the machine learning community due to its wide-ranging applications (see [3] and references therein). In the standard OT setting, there is one cost function, and the goal is to design a minimal-cost plan that transports "mass" from one probability distribution to another. The key challenge of OT in modern applications is the computational aspect since the probability distributions are typically high-dimensional.

Recently, Scetbon et al. [4] and Huang et al. [5] studied a variant of the OT problem, known as Equitable Optimal transport (EOT), and they showed that EOT is related to problems in economics such as fair cake-cutting problem and resource allocation. In EOT problem, there are N agents, each with a cost function, and the goal is to design a plan that minimizes the sum of the agents' transportation costs under the constraint that the cost is shared equally.

Existing works on OT and EOT consider a centralized approach in which a single agent/server designs the transportation plan. Motivated using a decentralized optimization framework to solve large-scale optimization, we study a decentralized variant of OT and EOT problems in this paper.

IL (ivanphlau@gmail.com) is with the Department of Computer Science, National University of Singapore, Singapore. Part of the research was done when the author was at the Department of Electrical and Computer Engineering, Rice University, Houston, TX, USA.

SM (sqma@rice.edu) is with the Department of Computational Applied Mathematics and Operations Research, Rice University, Houston, TX, USA. Research of SM is supported in part by NSF grants DMS-2243650, CCF-2308597, CCF-2311275 and ECCS-2326591, and a startup fund from Rice University.

CAU (cauribe@rice.edu) is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX, USA. The work of CAU and IL was partially supported by the National Science Foundation under Grants No. 2211815 and No. 2213568.

Decentralized OT (D-OT). Given two discrete probability distributions $p = (p_i)_{i=1}^n, q = (q_j)_{j=1}^n \in \Delta^n$ and a cost matrix $C \in \mathbb{R}_+^{n\times n}$ with $C_{ij} \geq 0$ corresponds to the unit cost of moving from p_i to q_j, the Kantorovich formulation of (discrete) OT is equivalent to solving the following linear programming (LP) problem

$$\min_{X\in\mathbb{R}_+^{n\times n}} \sum_{i,j} C_{ij}X_{ij} \text{ s.t. } X\mathbf{1}_n = p \text{ and } X^\intercal\mathbf{1}_n = q. \quad (1)$$

Here the optimization variable X and the objective function $\langle C, X\rangle := \sum_{i,j} C_{ij}X_{ij}$ are referred to as the transportation plan and the transportation cost, respectively, and the constraint $X\mathbf{1}_n = p$ and $X^\intercal\mathbf{1}_n = q$ is referred to as the marginal constraint. Since the OT problem is an LP with $2n$ equality constraints and n^2 variables, finding an exact solution is infeasible in practice for large n. In general, we aim to find an ϵ-approximate solution $\widehat{X}$ such that and

$$\langle C, \widehat{X}\rangle - \langle C, X^*\rangle + \left\|\widehat{X}\mathbf{1}_n - p\right\| + \left\|\widehat{X}^\intercal\mathbf{1}_n - q\right\| \leq \epsilon, \quad (2)$$

where X^* is a minimizer. Given an ϵ-approximate solution, one can find an $O(\epsilon)$-approximate solution satisfying the marginal constraint using [6, Lemma 7]. In this paper, we consider a decentralized variant of OT (D-OT), in which Problem (2) is solved by n agents collboratively over a network. In particular, agent k has access to the k^{th} column c_k of C and they work on optimizing the corresponding column x_k of X. Furthermore, each agent can exchange information only with its immediate neighbors. We formally formulate D-OT as a decentralized finite sum problem:

$$\min_{x_i\in\mathbb{R}_+^n} \sum_{i=1}^n c_i^\intercal x_i \text{ s.t. } \sum_{i=1}^n x_i = p \text{ and } x_i^\intercal\mathbf{1}_n = q_i, \quad (3)$$

where q_i is entry i of q.

Decentralized Equitable Optimal Transport (DE-OT). We now describe the EOT formulation by [4] using the above notations. In this setting, we have N agents with each agent k being given a cost matrix $C^k \in \mathbb{R}_+^{n\times n}$ where C_{ij}^k corresponds to the unit cost of moving from p_i to q_j. The EOT problem aims to find a transportation plan $(X^k)_{k=1}^N$ such that the transportation cost $\langle C^k, X^k\rangle$ among all the agents are equal to each other, and the sum of the cost is minimized:

$$\begin{aligned} &\min_{X^k\in\mathbb{R}_+^{n\times n}} \sum_{k=1}^N \langle C^k, X^k\rangle \\ &\text{s.t. } \langle C^k, X^k\rangle = \langle C^l, X^l\rangle \text{ for all } k, l \in [N], \\ &\left(\sum_{k=1}^N X^k\right)\mathbf{1}_n = p \text{ and } \left(\sum_{k=1}^N X^k\right)^\intercal \mathbf{1}_n = q. \end{aligned} \quad (4)$$

The algorithms proposed in [4], [5] consider a centralized setting, where $X^1, \ldots, X^N$ are designed by a central authority who has perfect information of $C^1, \ldots, C^N$. In this paper, we consider a decentralized variant of EOT (DE-OT), in which agent $k \in [N]$ knows only C^k, and works only on local optimization variable X^k such that collectively (X^k) solves (4). Like D-OT, each agent can communicate only with their immediate neighbors in some network. Our goal is to find an ϵ-approximate solution $\left(\widehat{X}^k\right)$ such that the sum of optimality gap (5), marginal constraint violation (6), and equitable constraint violation (7) is at most ϵ, i.e.,

$$\sum_{k=1}^N \langle C^k, \widehat{X}^k \rangle - \sum_{k=1}^N \langle C^k, (X^*)^k \rangle \tag{5}$$

$$+ \left\| \left(\sum_{k=1}^N \widehat{X}^k \right) \mathbf{1}_n - p \right\| + \left\| \left(\sum_{k=1}^N \widehat{X}^k \right)^\intercal \mathbf{1}_n - q \right\| \tag{6}$$

$$+ \frac{1}{N} \sum_{k=1}^N \left\| \langle C^k, \widehat{X}^k \rangle - \frac{1}{N} \sum_{k=1}^N \langle C^k, \widehat{X}^k \rangle \right\| \leq \epsilon. \tag{7}$$

The main contributions of this paper are as follows.

- We provide the first study of D-OT problem (3) and the first study of DE-OT problem (4), and show their equivalence with instances of distributed constraint-coupled optimization (DCCO) problems.
- We propose a single-loop decentralized algorithm that finds ϵ-approximate solutions to Problem (3) in $O(1/\epsilon)$ iterations. This iteration complexity matches existing centralized first-order approaches.
- We propose a single-loop decentralized algorithm that finds ϵ-approximate solution to Problem (4) in $O(1/\epsilon)$ iterations, This iteration complexity improves over existing centralized algorithms. Furthermore, our algorithm guarantees that the equitable constraint is satisfied at a rate of $O(1/k)$, while the existing centralized algorithms do not have such guarantees.

This paper is organized as follows. In Section II, we describe existing approaches for decentralized and equitable OT problems. Section III shows the reformulation of Problems (3) and (4) as distributed constrained-coupled optimization problems. Section IV introduces our single-loop decentralized algorithm that finds ϵ-approximate solutions in $O(1/\epsilon)$ iterations. We conclude with numerical results in Section V, and discussions and future work in Section VI.

II. Existing approaches for decentralized and equitable OT problems

The related work on optimal transport is extensive (e.g., see [7] and the references therein); we only provide a brief outline here, emphasizing the most closely related works.

Algorithms for (centralized) OT. Traditional LP algorithms are not scalable due to their arithmetic complexity of $\tilde{O}(n^3)$. In comparison, the approach of solving the *entropic-regularized* OT [8] have initiated a productive line of research. Existing algorithms to solve this approximation problem include Sinkhorn [8] and Greenkhorn [6] which have a complexity of $\tilde{O}(n^2/\epsilon^2)$ [9], as well as first-order methods such as primal-dual methods [9]–[11] with a complexity of $\tilde{O}(n^{2.5}/\epsilon)$, alternating minimization with a complexity of $\tilde{O}(n^{2.5}/\epsilon)$, dual extrapolation [12] with a complexity of $\tilde{O}(n^2/\epsilon)$, and extragradient [13] with a complexity of $\tilde{O}(n^2/\epsilon)$. However, the entropy term causes the plan to be fully dense, which can be undesirable in certain applications. Consequently, there has been a growing interest [14], [15] in Euclidean-regularized OT since it results in sparse transportation plans. Additionally, algorithms for Euclidean-regularized OT tend to be more robust than entropic-regularized OT when a small regularization parameter is used.

Remark II.1. While this paper focuses on theoretical complexities, we remark that an algorithm with a better complexity may not necessarily be faster in practice than another algorithm with a worse complexity. For instance, the methods with iteration complexity of $O(1/\epsilon)$ above are generally slower than Sinkhorn [8] and Greenkhorn [6] in practice. Likewise, Sinkhorn and Greenkhorn are slower than off-the-shelf LP solver for small regularization constant ϵ.

Algorithms for (centralized) EOT. The authors in [4], [5] did not solve the EOT Problem (4) directly, and instead solved the following formulation. which are equivalent [4, Proposition 1] when the entries of the cost matrices are the same (e.g., all non-negative):

$$\begin{aligned} &\min_{X^k \in \mathbb{R}_+^{n\times n}} \max_{\mathbf{p}=(p_k)\in\Delta_+^N} \sum_{k=1}^N p_k \langle C^k, X^k \rangle \\ &\text{s.t.} \; \left(\sum_{k=1}^N X^k\right) \mathbf{1}_n = p \text{ and } \left(\sum_{k=1}^N X^k\right)^\intercal \mathbf{1}_n = q. \end{aligned} \tag{8}$$

In particular, [4] proposed to solve an entropy-regularized approximation to (8) and designed a projected alternating maximization (PAM) algorithm to solve its dual. The iteration complexity of this approach is shown by [5] to be $O(1/\epsilon^2)$. The authors in [5] also gave an accelerated version of PAM, but they did not manage to show an improved iteration complexity. We note that these algorithms require projection onto simplex at each iteration, making it hard to be implemented in a decentralized manner.

Decentralization for OT. The work in [16]–[18] considered a decentralized optimal transport problem with an agent for each row and an agent for each column, with the row and column agents connected through a bipartite graph. Different from their framework, each agent in D-OT works on one column of the plan, and the connected undirected network can be arbitrary.

III. Reformulation to DCCO

In this section, we reformulate Problems (3) and (4) into distributed constraint-coupled optimization (DCCO) problems [19], [20], where a set of agents cooperatively minimize the sum of objective functions subject to a coupling affine equality constraint

$$\min_{x_i} \sum_{i=1}^{N} f_i(x_i) + g_i(x_i) \qquad \text{s.t.} \qquad \sum_{i=1}^{N} A_i x_i = b, \quad (9)$$

where f_i is smooth and convex, while g_i is convex but possibly non-smooth.

A. Reformulating D-OT as a DCCO problem

We use the following notations to be consistent with the optimization literature. Agents are indexed using i (instead of k), while the decision variable of agent i and the i^{th} column of the cost matrix C are denoted by x_i and c_i. With these notations, we reformulate D-OT to a DCCO problem.

Lemma III.1. The OT problem (1) is equivalent to

$$\min_{x_i} \sum_{i=1}^{n} c_i^{\mathsf{T}} x_i + \iota_{\geq \mathbf{0}}(x_i) \text{ s.t. } \sum_{i=1}^{n} M_i x_i = \begin{bmatrix} p \\ q \end{bmatrix}, \quad (10)$$

where the indicator function $\iota_{\geq \mathbf{0}}(y)$ is defined by

$$\iota_{\geq \mathbf{0}}(y) := \begin{cases} 0 & \text{if } y \geq 0 \\ \infty & \text{otherwise,} \end{cases} \quad (11)$$

and the matrix $M_i \in \mathbb{R}^{2n \times n}$ is the i^{th} column-block of

$$[M_1, \dots, M_n] = M := \begin{bmatrix} I_n & I_n & \cdots & I_n \\ \mathbf{1}_n^{\mathsf{T}} & \mathbf{0}_n^{\mathsf{T}} & \cdots & \mathbf{0}_n^{\mathsf{T}} \\ \mathbf{0}_n^{\mathsf{T}} & \mathbf{1}_n^{\mathsf{T}} & \cdots & \mathbf{0}_n^{\mathsf{T}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_n^{\mathsf{T}} & \mathbf{0}_n^{\mathsf{T}} & \cdots & \mathbf{1}_n^{\mathsf{T}} \end{bmatrix}. \quad (12)$$

Proof. The objective function and non-negativity constraint of (1) are encoded in the objective function of (10). The marginal constraint also follows directly from the definition of M and the linearity of matrix multiplication. □

Note that it follows immediately that an ϵ-optimal solution (in terms of objective function value) that is also ϵ-close to being feasible is an ϵ-approximate solution for D-OT.

Remark III.2. Since $\operatorname{rank}(M) = 2n-1$ we can remove the last row of both sides in the affine coupling constraint.

Proposition III.3. Problem (1) is equivalent to

$$\min_{x_i} \sum_{i=1}^{n} c_i^{\mathsf{T}} x_i + \iota_{\geq \mathbf{0}}(x_i) \text{ s.t. } \sum_{i=1}^{n} \tilde{M}_i x_i = \begin{bmatrix} p \\ \tilde{q} \end{bmatrix}, \quad (13)$$

where $\tilde{M}_i$ and $\tilde{q}$ are obtained from M_i and q with the last row removed.

B. Reformulating DE-OT as a DCCO problem

Similar to the notation changing in Section III-A, we will also do that for the EOT reformulation: Agents are indexed using i, with the local decision variable and the cost matrix of agent i denoted by $x_i \in \mathbb{R}_+^{n^2}$ and $c_i \in \mathbb{R}^{n^2}$ respectively (instead of $X^i \in \mathbb{R}^{n \times n}$ and $C^i \in \mathbb{R}^{n \times n}$). The reformulation for DE-OT is not as straightforward as compared to D-OT due to the equitable constraint. In particular, we must find a way to represent the equitable constraint into a coupling affine equality constraint. To achieve this, we use the Laplacian matrix $L \in \mathbb{R}^{N \times N}$ of the network to help us, thanks to the property: $Lx = 0 \iff x \in \operatorname{span}(\mathbf{1})$.

Furthermore, since $\operatorname{rank}(L) = N-1$, we can remove its last row, similar to the case in Proposition III.3. Let $\tilde{L} = \left[\ \tilde{L}_{*,1} \mid \cdots \mid \tilde{L}_{*,N}\ \right] \in \mathbb{R}^{(N-1) \times N}$ be the first $N-1$ rows of the Laplacian matrix L associated with the network, where $\tilde{L}_{*,i}$ is its i^{th} column. Note that $\tilde{L}$ has a full row rank.

Proposition III.4. Problem (4) is equivalent to

$$\min_{x_i} \sum_{i=1}^{N} c_i^{\mathsf{T}} x_i + \iota_{\geq \mathbf{0}}(x_i) \text{ s.t. } \sum_{i=1}^{N} \begin{bmatrix} \tilde{M} \\ E_i \end{bmatrix} x_i = \begin{bmatrix} p \\ \tilde{q} \\ \mathbf{0}_{N-1} \end{bmatrix}, \quad (14)$$

where the indicator function $\iota_{\geq \mathbf{0}}$ and and vector $\tilde{q}$ are as defined in Lemma III.1 and Propostion III.3, while matrix $\tilde{M}$ is the first $2n-1$ rows of matrix M from (12) and matrices E_i are defined as follows:

$$E_i := \left(\tilde{L}_{*,i}\right) c_i^{\mathsf{T}} = \tilde{L} \begin{bmatrix} \mathbf{0}_{(i-1) \times n^2} \\ c_i^{\mathsf{T}} \\ \mathbf{0}_{(N-i) \times n^2} \end{bmatrix} \in \mathbb{R}^{(N-1) \times n^2}. \quad (15)$$

Proof. The verification for the objective function, non-negativity constraint, and marginal constraint is similar to the proof of Lemma III.1. The equitable constraint is enforced since $\left[c_1^{\mathsf{T}} x_1; \cdots; c_N^{\mathsf{T}} x_N\right]$ is a constant vector if and only if

$$\sum_{i=1}^{N} E_i\, x_i = \tilde{L} \left(\sum_{i=1}^{N} \begin{bmatrix} \mathbf{0}_{i-1} \\ c_i^{\mathsf{T}} x_i \\ \mathbf{0}_{N-i} \end{bmatrix} \right) = \tilde{L} \left(\begin{bmatrix} c_1^{\mathsf{T}} x_1 \\ \vdots \\ c_N^{\mathsf{T}} x_N \end{bmatrix} \right) = \mathbf{0},$$

since $\operatorname{null}(\tilde{L}) = \operatorname{span}(\mathbf{1})$. □

We intentionally use the same notations to denote different things in the reformulations (13) and (14) because of their similarity. Indeed, we can abstract them into the form of

$$\min_{\mathbf{x}=(x_i) \in \mathbb{R}^{Nd}} f(\mathbf{x}) := \mathbf{c}^{\mathsf{T}} \mathbf{x} + \iota_{\geq \mathbf{0}}(\mathbf{x}) = \sum_{i=1}^{N} c_i^{\mathsf{T}} x_i + \iota_{\geq \mathbf{0}}(x_i)$$
$$\text{s.t.} \quad A\mathbf{x} = \sum_{i=1}^{N} A_i x_i = b. \quad (16)$$

where the number of agents N, the vector b and matrices A_i, and dimension d of x_i and c_i are problem-dependent.

C. DCCO algorithms to solve reformulated OT and EOT

Reformulating D-OT (3) and DE-OT (4) into (16) allows us to apply existing algorithms for DCCO (e.g., [19]–[25]) to solve it. In particular, we have corresponding $f_i(x_i) = c_i^{\mathsf{T}} x_i$ which is smooth and (non-strongly) convex, and $g_i = \iota_{\geq \mathbf{0}}$ which is non-smooth. Before describing the algorithms, we explicitly state the assumption of the network.

Network assumption. In problem (16), the network connecting the agents is an undirected graph $G = (V, E)$, where V is the set of agents and E is the set of edges. An edge $(i,j) \in E$ if and only if agents i and j are neighbors. For each agent i, we define the set of its neighbors as $\mathcal{N}_i = \{j \in V \mid (i,j) \in E\}$. We assume the network

TABLE I: Algorithms for Problem (16).

Paper	Convergence rate	Singe-loop
[19]	Asymptotic	No
[20], [21]	$O(1/k)$, ergodic	No
[22]	$O(1/k)$, non-ergodic	No
[23], [25]	Asymptotic	Yes

satisfies the following standard assumption in decentralized optimization, which implies that any two agents can influence each other in the long run.

Assumption III.5. The graph G is connected and static.

The algorithms in [20]–[22] can be used to solve problem (16), with a convergence rate of $O(1/k)$, where k is the iteration number. However, using algorithms in [20]–[22] to solve problem (16) requires solving a quadratic program at each iteration. In contrast, the primal-dual-based algorithms such as those in [23], [25] are single-loop (i.e., do not require an inner subroutine), but they can establish only an asymptotic convergence for non-smooth problems. The continuous-time algorithm in [24] is also single-loop, but its performance may not carry over to discretized implementation. In Table I, we summarize discrete-time algorithms that can be used to solve problem (16).

We adapt [21, Algorithm 1] for (16), which yields Algorithm 1. Note that setting $\tau_i = 0$ recovers [20, Algorithm 3], which has the following performance guarantee[1].

Proposition III.6. Let Assumption III.5 hold. For all $c_i \in \mathbb{R}^d$, the sequence $(\mathbf{x}^k)$ generated by Algorithm 1 with $\tau_i = 0$ and any parameter $\rho_i > 0$ converges to the optimal solution $\mathbf{x}^*$ of Problem (16). Furthermore,

$$|f(\bar{\mathbf{x}}^k) - f(\mathbf{x}^*)| + \left\|A\bar{\mathbf{x}}^k - b\right\|_2 = O(1/k), \qquad (17)$$

where $\bar{\mathbf{x}}^k = \frac{1}{k}\sum_{t=1}^k \mathbf{x}^t$.

Remark III.7. As in Remark II.1, we note that algorithms with "poor" theoretical properties may perform well in practice and vice versa. For instance, since most entries of a high-dimensional transportation plan are near-zero, performing a proximal gradient steps at every iteration (e.g., [23], [25]) might lead to numerical instability. In contrast, solving a quadratic program at each iteration (with off-the-shelf solvers) could be advantageous in practice.

IV. Euclidean regularized D-OT and DE-OT

Based on Section III-C, no existing discrete-time single-loop DCCO algorithm could directly solve problem (16) with an explicit non-asymptotic convergence rate. This is not unexpected since our objective function is non-strongly convex and non-smooth. In this section, we develop a single-loop algorithm for (16), which has also $O(1/\epsilon)$ iteration complexity. Our first step is to perturb (16) with a squared-Euclidean norm regularizer to make it strongly convex.

$$\min_{x_i \in \mathbb{R}^d} \sum_{i=1}^n c_i^\intercal x_i + \frac{\eta}{2}\|x_i\|^2 + \iota_{\geq \mathbf{0}}(x_i) \text{ s.t. } \sum_{i=1}^N A_i x_i = b. \quad (18)$$

This problem is an $\frac{\eta}{2}$-approximation of problem (16) in terms of the objective function value: If $\mathbf{x}^*$ and $\hat{\mathbf{x}}$ are minimizers of problems (16) and (18) respectively, then $f(\hat{\mathbf{x}}) - f(\mathbf{x}^*) \leq \frac{\eta}{2}$.

Remark IV.1. In the centralized OT literature, it is common to use entropic regularization (see Section II). The entropic-regularized problem can be written compactly as

$$\min_{\mathbf{x} \in \mathbb{R}^{Nd}} \quad \mathbf{c}^\intercal \mathbf{x} + H(\mathbf{x}) \qquad \text{s.t.} \qquad A\mathbf{x} = b, \qquad (19)$$

where $H(\cdot)$ is the entropy function. If $\|\mathbf{x}\|$ is unbounded, then problem (19) is not strongly convex. Consequently, the (decentralized) dual problem of (19) is not smooth. If we set $\|\mathbf{x}\| = 1$, then the corresponding smooth dual problem

$$\min_{\lambda_1 = \cdots = \lambda_N} \eta \log\left(\sum_{i=1}^N \left\|\exp\left(\frac{-\mathbf{c}_i - A_i^\intercal \lambda_i}{\eta}\right)\right\|_1\right) + \eta + \frac{1}{N}\lambda_i^\intercal \mathbf{q}, \qquad (20)$$

is not separable, making it unsuitable for current decentralized optimization methods.

A. PDC-ADMM

In this section, we provide an algorithm to solve (18) with an $O(1/k)$ ergodic convergence rate. The algorithm is an inexact variant of Algorithm 1, which is adapted from [21, Algorithm 1] from Section III-C. In [21], the author apply their Algorithm 1 to solve problems of the form

$$\begin{aligned} &\min_{x_i \in \mathcal{S}_i, y_i \geq 0} \quad \sum_{i=1}^N f_i(x_i) \\ &\text{s.t.} \sum_{i=1}^N A_i x_i = b \text{ and } B_i x_i + y_i - v_i = 0 \text{ for all } i, \end{aligned} \qquad (21)$$

which is a special case of Problem (9). Matching it with (18), we have $f_i(x_i) = c_i^\intercal x_i + \frac{\eta}{2}\|x_i\|^2$, which is strongly convex and smooth, $\mathcal{S}_i = \mathbb{R}^d$, $B_i = -I_d$, and $v_i = 0$:

$$\begin{aligned} &\min_{\substack{\mathbf{x}=(x_i) \in \mathbb{R}^{Nd}, \\ \mathbf{y}=(y_i) \geq 0}} \quad f(\mathbf{x}) := \sum_{i=1}^N c_i^\intercal x_i + \frac{\eta}{2}\|x_i\|^2 \\ &\text{s.t.} \quad A\mathbf{x} = \sum_{i=1}^N A_i x_i = b \text{ and } \mathbf{y} - \mathbf{x} = \mathbf{0}. \end{aligned} \qquad (22)$$

Compared to the formulation in (18) where the non-negativity constraint $x_i \in \mathbb{R}_+^d$ is represented using an indicator function $\iota_{\geq \mathbf{0}}(x_i)$ in the objective function, problem formulation (22) represents this constraint as an equality constraint $-x_i + y_i = 0$ using a non-negative variable $y_i \geq 0$.

In solving (21), Algorithm 1 of [21] solves an expensive subproblem at each iteration. In the same paper, the author gives an inexact update, which has a low-complexity implementation (see [21, Eq. (38)-(39)]). Adapting this for

[1]Setting $\tau_i = 0$ forces $y_i^k = x_i^k$, making y_i^k and z_i^k redundant.

problem (22) gives us a single-loop algorithm, which is obtained by replacing the update of (x_i^k, y_i^k) in Line 9 of Algorithm 1 with the following closed-form updates:

$$y_i^{k+1} := \Big(1 - \frac{1}{\beta_i \tau_i}\Big) y_i^k + \frac{1}{\beta_i \tau_i}\Big(x_i^k - \tau_i z_i^k\Big) \tag{23}$$

$$\begin{aligned} x_i^{k+1} := x_i^k - \frac{1}{\beta_i}\bigg(& c_i + \eta x_i^k + \frac{1}{\tau_i}(x_i^k - y_i^{k+1} - \tau_i z_i^k) + \\ & \frac{A_i^\intercal}{2\rho|\mathcal{N}_i|}\Big(A_i x_i^{k+1} - \frac{1}{N} b - p_i^k + \rho \sum_{j\in\mathcal{N}_i} \big(\lambda_j^k + \lambda_i^k\big)\Big)\bigg), \end{aligned} \tag{24}$$

where $\beta_i > 0$ is a penalty parameter. The performance guarantee of the inexact variant of Algorithm 1 is as given, which follows immediately from [21, Theorem 2].

Theorem IV.2. Let Assumption III.5 hold. Consider the inexact variant of Algorithm 1 obtained by replacing Line 9 with (23) and (24). Then the sequence $(\mathbf{x}^k, \mathbf{y}^k)$ generated by Algorithm 1 with parameters $\rho, \beta_i, \tau_i > 0$ satisfying

$$\beta_i \tau_i \geq 1 \text{ and } \Big(\beta_i - \frac{\beta_i}{\beta_i \tau_i - 1} - 1\Big) I_d - \frac{1}{2\rho|\mathcal{N}_i|} A_i^\intercal A_i \succ \mathbf{0}, \tag{25}$$

converges to the optimal solution $(\mathbf{x}^*, \mathbf{y}^* = \mathbf{x}^*)$ of Problem (22). Furthermore,

$$|f(\bar{\mathbf{x}}^k) - f(\mathbf{x}^*)| + \big\|A\bar{\mathbf{x}}^k - b\big\| + \|\bar{\mathbf{y}}^k - \bar{\mathbf{x}}^k\| = O(1/k), \tag{26}$$

where $(\bar{\mathbf{x}}^k, \bar{\mathbf{y}}^k) = \frac{1}{k}\sum_{t=1}^k (\mathbf{x}^t, \mathbf{y}^t)$.

Remark IV.3. While $\bar{\mathbf{x}}^k$ may not be non-negative, we can do a decentralized projection onto simplex, which has a linear convergence [26]. Furthermore, when $\big\|A\bar{\mathbf{x}}^k - b\big\| + \|\bar{\mathbf{y}}^k - \bar{\mathbf{x}}^k\| < \epsilon$, then the projected $\hat{\mathbf{x}}$ satisfies $\|\hat{\mathbf{x}} - \bar{\mathbf{x}}^k\| = O(\epsilon)$.

Corollary IV.4. Let Assumption III.5 hold, and we run the inexact variant of Algorithm 1 for Problems (3) and (4), both with parameters $\rho, \beta_i, \tau_i > 0$ satisfying (25). Then for any $\epsilon > 0$, there exists some $K = O(1/\epsilon)$ such that for all iteration $k \geq K$, the simplex projection $\hat{\mathbf{x}}(k)$ of $\bar{\mathbf{x}}^k$ is an ϵ-approximate solution. The arithmetic cost per iteration per agent are $O(|\mathcal{N}_i|n)$ and $O(n^2 + |\mathcal{N}_i|(2n+N))$ for Problems (3) and (4) respectively.

Proof. By Theorem IV.2, there exists some K such that for all $k \geq K$, we have $|f(\bar{\mathbf{x}}^k) - f(\mathbf{x}^*)| + \big\|A\bar{\mathbf{x}}^k - b\big\| + \|\bar{\mathbf{y}}^k - \bar{\mathbf{x}}^k\| = O(\epsilon)$. We now bound $|f(\hat{\mathbf{x}}) - f(\mathbf{x}^*)| + \|A\hat{\mathbf{x}} - b\|$. We have $\|A\hat{\mathbf{x}} - b\| = \|A\hat{\mathbf{x}} - A\bar{\mathbf{x}}^k + A\bar{\mathbf{x}}^k - b\| \leq \|A\|\|\hat{\mathbf{x}} - \bar{\mathbf{x}}^k\| + \big\|A\bar{\mathbf{x}}^k - b\big\| = O(\epsilon)$. Similarly, $|f(\hat{\mathbf{x}}) - f(\mathbf{x}^*)| \leq \mathbf{c}^\intercal(\hat{\mathbf{x}} - \bar{\mathbf{x}}^k + \bar{\mathbf{x}}^k - \mathbf{x}^*) \leq \|\mathbf{c}\|\|\hat{\mathbf{x}} - \bar{\mathbf{x}}^k\| + \mathbf{c}^\intercal(\bar{\mathbf{x}}^k - \mathbf{x}^*) = O(\epsilon)$. Therefore, we can find some $K = O(1/\epsilon)$ such that $\hat{\mathbf{x}}$ is an ϵ-optimal solution. We now bound the arithmetic cost per iteration per agent for Problem (3), but omit the details for Problem (4) since they are similar. The additions and subtractions of vectors are straightforward to bound, with the tricky part being the matrix multiplication with A_i and $A_i^\intercal$. But since $A_i = \tilde{M}_i$ is sparse and structured with $O(n)$ entries, multiplication takes $O(n)$ arithmetic operations. ☐

Remark IV.5. If the underlying network is a star graph, then the total arithmetic cost per iteration for all agents are $O(n\sum_i |\mathcal{N}_i|) = O(n^2)$ and $O(Nn^2 + (2n+N)\sum_i |\mathcal{N}_i|) = O(Nn^2 + N^2)$ for Problems (3) and (4) respectively.

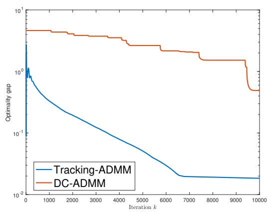
Algorithm 1 Exact PDC-ADMM for Problem (22)

1: **Input**: Each agent i is given a vector $c_i \in \mathbb{R}^d$ and vectors p and q
2: Each agent i creates matrix A_i and vector b
3: Initialize iteration counter $k = 0$.
4: Each agent i initialize variables $x_i^0 = y_i^0 = z_i^0 = \mathbf{0}_d$, and $\lambda_i^0 = p_i^0 = A_i x_i = \mathbf{0}$
5: Each agent i initialize penalty parameters $\rho_i > 0, \tau_i \geq 0$.
6: **Repeat** until a predefined stopping criterion is satisfied
7: For all agents $i \in [N]$ (in parallel)
8: Exchange λ_i^k with neighbors $\mathcal{N}_i$
9: Update $(x_i^{k+1}, y_i^{k+1}) := \operatorname*{argmin}_{\substack{x_i\in\mathbb{R}^d \\ y_i\in\mathbb{R}^d_+}} \Big\{ c_i^\intercal x_i + \frac{\rho}{4|\mathcal{N}_i|}\Big\| \frac{1}{\rho}(A_i x_i - \frac{1}{N}b) - \frac{1}{\rho}p_i^k + \sum_{j\in\mathcal{N}_i}\big(\lambda_j^k + \lambda_i^k\big)\Big\|_2^2 + \frac{1}{2\tau_i}\|y_i - x_i + \tau_i z_i^k\|_2^2 \Big\}$
10: Update $\lambda_i^{k+1} := \frac{1}{2|\mathcal{N}_i|}\Big(\sum_{j\in\mathcal{N}_i}\big(\lambda_j^k + \lambda_i^k\big) + \frac{1}{\rho}p_i^{k+1} + \frac{1}{\rho}\big(A_i x_i^{k+1} - \frac{1}{N}b\big)\Big)$
11: Update $z_i^{k+1} := z_i^k + \frac{1}{\tau_i}(y_i^{k+1} - x_i^{k+1})$
12: Update $p_i^{k+1} := p_i^k + \rho\sum_{j\in\mathcal{N}_i}\big(\lambda_i^{k+1} - \lambda_j^{k+1}\big)$
13: Set $k = k+1$
14: **Output** $\bar{x}_i^k := \frac{1}{k}\sum_{\ell=1}^{k-1} x_i^\ell$ for all agents i

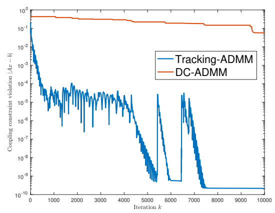
V. NUMERICAL SIMULATIONS

We simulate the performance of Algorithm 1 with $\tau_i = 0$ (i.e., DC-ADMM) and Tracking-ADMM [19, Algorithm 1] for the D-OT problem (3) and DE-OT problem (4). For D-OT, we consider a cost matrix and transport plan of size $n^2 = 2500$ with $N = n = 50$ agents. The probability distributions p and q are randomly generated. The cost matrix $C \in \mathbb{R}^{n\times n}$ is generated randomly with $C_{i,j} = \|x_i - y_j\|^2$ for some random $x_i \sim \mathcal{N}\left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}10 & 1\\ 1 & 10\end{pmatrix}\right)$ and $y_j \sim \mathcal{N}\left(\begin{pmatrix}2\\2\end{pmatrix}, \begin{pmatrix}2 & -0.2\\ -0.2 & 2\end{pmatrix}\right)$. The undirected network is also generated randomly. For DE-OT, we consider a cost matrix and transport plan of size $n^2 = 400$ with $N = 10$ agents. As before, the probability distributions p and q, and the undirected network are randomly generated. The only difference is the cost matrices generation: we first generate a base cost matrix C^{base} using the method from above, and then generate $C_{i,j}^k = C_{i,j}^{\text{base}} + \mathcal{N}(0, 10)$ for each agent k. For the computations, we first use MATLAB's built-in LP solver to solve the centralized problem to obtain an optimal objective function value $f^* = \mathbf{c}^\intercal\mathbf{x}^*$. We then use Tracking-ADMM and DC-ADMM to solve the same instances of D-OT and DE-OT. In Figure 1, we show the optimality gap $f(\mathbf{x}^k) - f^*$

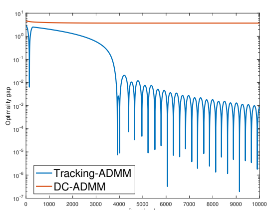
and feasibility violation $\|A\mathbf{x}^k - b\|$ across iterations for both the D-OT problem (3) and DE-OT problem (4) under the algorithms. The simulation results show that reformulation of Problems (3) and (4) into distributed constraint-coupled optimization (DCCO) allow them to be solved using existing DCCO algorithms. We also note that Tracking-ADMM has better performances than Algorithm 1 for both problems in practice, despite a poorer theoretical guarantee. This is not surprising, see Remark III.7.



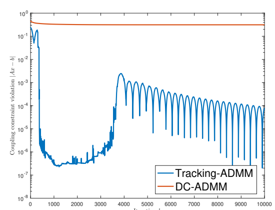
(a) Optimality gap for D-OT



(b) Feasibility violation for D-OT



(c) Optimality gap for DE-OT



(d) Feasibility violation for DE-OT

Fig. 1: Optimality gap $f(\mathbf{x}^k) - f^*$ and feasibility violation $\|A\mathbf{x}^k - b\|$ for D-OT and DE-OT under Tracking-ADMM and DC-ADMM.

VI. Conclusion

We studied the problem of decentralized optimal transport (D-OT) and decentralized equitable optimal transport (DE-OT), where a group of agents collaboratively compute an optimal transport plan. In the D-OT setup, each agent has access to a column of the cost matrix. Meanwhile, in the DE-OT, each agent has access to a private cost matrix, and agents try to cooperatively compute a transportation plan that ensures equity. We showed that these problems can be reformulated as distributed constrained-coupled problems and adapted existing work to provide a single-loop algorithm that has an iteration complexity of $O(1/\epsilon)$. Interestingly, our approach for DE-OT has a better iteration complexity than existing centralized methods.

References

[1] G. Monge, "Memory on the theory of cut and fill," *Mem. Math. Phys. Acad. Royal Sci.*, pp. 666–704, 1781.

[2] L. V. Kantorovich, "On the translocation of masses," in *Dokl. Akad. Nauk. USSR (NS)*, vol. 37, 1942, pp. 199–201.

[3] E. F. Montesuma, F. N. Mboula, and A. Souloumiac, "Recent advances in optimal transport for machine learning," *arXiv preprint arXiv:2306.16156*, 2023.

[4] M. Scetbon, L. Meunier, J. Atif, and M. Cuturi, "Equitable and optimal transport with multiple agents," in *AISTATS*, 2021.

[5] M. Huang, S. Ma, and L. Lai, "On the convergence of projected alternating maximization for equitable and optimal transport," *arXiv preprint arXiv:2109.15030*, 2021.

[6] J. Altschuler, J. Niles-Weed, and P. Rigollet, "Near-linear time approximation algorithms for optimal transport via sinkhorn iteration," *Advances in neural information processing systems*, vol. 30, 2017.

[7] G. Peyré, M. Cuturi, *et al.*, "Computational optimal transport: With applications to data science," *Foundations and Trends® in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.

[8] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," *Advances in neural information processing systems*, vol. 26, 2013.

[9] T. Lin, N. Ho, and M. I. Jordan, "On the efficiency of entropic regularized algorithms for optimal transport," *Journal of Machine Learning Research*, vol. 23, no. 137, pp. 1–42, 2022.

[10] P. Dvurechensky, A. Gasnikov, and A. Kroshnin, "Computational optimal transport: Complexity by accelerated gradient descent is better than by sinkhorn's algorithm," in *International conference on machine learning*. PMLR, 2018, pp. 1367–1376.

[11] A. Chambolle and J. P. Contreras, "Accelerated bregman primal-dual methods applied to optimal transport and wasserstein barycenter problems," *SIAM Journal on Mathematics of Data Science*, vol. 4, no. 4, pp. 1369–1395, 2022.

[12] A. Jambulapati, A. Sidford, and K. Tian, "A direct tilde {O}(1/epsilon) iteration parallel algorithm for optimal transport," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[13] G. Li, Y. Chen, Y. Chi, H. V. Poor, and Y. Chen, "Fast computation of optimal transport via entropy-regularized extragradient methods," *arXiv preprint arXiv:2301.13006*, 2023.

[14] D. A. Lorenz, P. Manns, and C. Meyer, "Quadratically regularized optimal transport," *Applied Mathematics & Optimization*, vol. 83, no. 3, pp. 1919–1949, 2021.

[15] D. A. Pasechnyuk, M. Persiianov, P. Dvurechensky, and A. Gasnikov, "Algorithms for euclidean regularised optimal transport," *arXiv preprint arXiv:2307.00321*, 2023.

[16] R. Zhang and Q. Zhu, "Consensus-based distributed discrete optimal transport for decentralized resource matching," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 5, no. 3, pp. 511–524, 2019.

[17] J. Hughes and J. Chen, "Fair and distributed dynamic optimal transport for resource allocation over networks," in *2021 55th Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2021, pp. 1–6.

[18] X. Wang, H. Xu, and M. Yang, "Decentralized entropic optimal transport for privacy-preserving distributed distribution comparison," *arXiv preprint arXiv:2301.12065*, 2023.

[19] A. Falsone, I. Notarnicola, G. Notarstefano, and M. Prandini, "Tracking-ADMM for distributed constraint-coupled optimization," *Automatica*, vol. 117, p. 108962, 2020.

[20] T.-H. Chang, M. Hong, and X. Wang, "Multi-agent distributed optimization via inexact consensus ADMM," *IEEE Transactions on Signal Processing*, vol. 63, no. 2, pp. 482–497, 2014.

[21] T.-H. Chang, "A proximal dual consensus ADMM method for multi-agent constrained optimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3719–3734, 2016.

[22] Y. Su, Q. Wang, and C. Sun, "Distributed primal-dual method for convex optimization with coupled constraints," *IEEE Transactions on Signal Processing*, vol. 70, pp. 523–535, 2021.

[23] S. A. Alghunaim, K. Yuan, and A. H. Sayed, "A proximal diffusion strategy for multiagent optimization with sparse affine constraints," *IEEE Transactions on Automatic Control*, vol. 65, no. 11, pp. 4554–4567, 2019.

[24] J. Li and H. Su, "Implicit tracking-based distributed constraint-coupled optimization," *IEEE Transactions on Control of Network Systems*, 2022.

[25] J. Li, Q. An, and H. Su, "Proximal nested primal-dual gradient algorithms for distributed constraint-coupled composite optimization," *Applied Mathematics and Computation*, vol. 444, p. 127801, 2023.

[26] F. Iutzeler and L. Condat, "Distributed projection on the simplex and ℓ_1 ball via ADMM and gossip," *IEEE Signal Processing Letters*, vol. 25, no. 11, pp. 1650–1654, 2018.