Distributed Computation of Graph Matching in Multi-Agent Networks

Quoc Van Tran, Zhiyong Sun, Brian D. O. Anderson, and Hyo-Sung Ahn

Abstract—This work considers the distributed computation of the one-to-one vertex correspondences between two undirected and connected graphs, which is called graph matching, over multi-agent networks. Given two isomorphic and asymmetric graphs, there is a unique permutation matrix that maps the vertices in one graph to the vertices in the other. Based on a convex relaxation of graph matching in Aflalo et al. [1], we propose a distributed computation of graph matching as a distributed convex optimization problem subject to equality constraints and a global set constraint, using a network of multiple agents whose interaction graph is connected. Each agent in the network only knows one column of each of the adjacency matrices of the two graphs, and all agents collaboratively learn the graph matching by exchanging information with their neighbors. The proposed algorithm employs a projected primal-dual gradient method to handle equality constraints and a set constraint. Under the proposed algorithm, the agents’ estimates of the permutation matrix converge to the optimal permutation globally and exponentially fast. Finally, simulation results are given to illustrate the effectiveness of the method.

I. INTRODUCTION

A graph consists of a set of vertices and a set of edges. The vertices might represent abstract entities such as features in an image, point patterns or users in social networks, or represent physical agents such as body parts, mobile robots or unmanned aerial vehicles, and the edges in the graph represent the relations between the vertices. Given two connected graphs, the problem of finding an optimal permutation matrix that minimizes the disagreement between two corresponding adjacency matrices \( A \) and \( B \in \mathbb{R}^{n \times n} \) is referred to as graph matching. Graph matching has a wide range of applications in different science and engineering disciplines such as computer vision and pattern recognition [2]–[4], neuroscience [5], and formation control [6], [7], to name a few.

Graph matching has been studied extensively in the last few decades. Though there are numerous graph matching algorithms developed in the literature, graph matching remains computationally intractable [8]. Graph matching with zero adjacency disagreement is said to be exact matching. In the presence of noise, it is referred to as inexact matching with the minimal adjacency disagreement. Heuristic algorithms, e.g., ones based on some forms of tree search [9], [10], have no theoretical guarantee of the convergence to the global minimizer of the graph matching. Spectral methods rely on the fact that two adjacency matrices of two isomorphic graphs share the same spectrum [11], [12]. A popular approach for addressing graph matching is the continuous optimization based on relaxations of the discrete graph matching problem [1], [8]. Two common relaxations of graph matching are the indefinite relaxation by minimizing \(-\langle AP, PB \rangle\) [5], [8], where \( \langle \cdot, \cdot \rangle \) denotes the inner product, and the convex relaxation by minimizing \(\|AP−PB\|_F^2\) [1], [8], where \(\|\cdot\|_F\) denotes the Frobenius norm, over \( P \) in the set of doubly stochastic matrices. The former in general has multiple local minima and hence depending on the initialization, a gradient-based algorithm will converge to a local minimum of the objective function, but not necessarily the global minimum. The convex relaxation has a unique least-squares solution in the doubly stochastic matrix set. However, the actual permutation that matches two isomorphic graphs can be only recovered when the norm of the perturbed adjacency matrix is less than a small value [8]. By using friendliness properties of adjacency matrices, which are characterized by their spectral properties, the graph matching can be further relaxed by replacing the set of doubly stochastic matrices by the set of pseudo-stochastic matrices \( \{P : P^T_1 = 1_n\} \) [1]. We refer the readers to [3], [8] for more comprehensive reviews of graph matching algorithms.

This paper considers the distributed computation of graph matching over an \( n \)-agent network in which each agent in the network only knows one column of each of the adjacency matrices \( A \) and \( B \). The agents in the network collaboratively learn the graph matching by exchanging information with their neighbors. The distributed setup is commonly used in distributed algorithms to solve linear algebraic equations [13]–[15] and linear matrix equations [16], [17]. Distributed algorithms and distributed optimizations over networked systems in particular have attracted lot of research interest in different areas of science and engineering [18]–[22], partly due to the increasing scale of the underlying problems, the distributed nature of networked systems, and the privacy of individual information. Motivated by these facts, we present a distributed algorithm to compute the graph matching between two isomorphic graphs based on a convex relaxation of graph matching, and by using distributed optimization over multi-agent networks.

The contributions of this paper are as follows. First, based on the centralized algorithm in [1], we formulate the graph matching as a multi-agent convex optimization problem subject to equality constraints and a global set constraint,
in which each agent in the network only knows one column of each of the adjacency matrices of the two graphs to be matched. Then as a development of [1], we prove that almost all adjacency matrices have friendliness properties, which allows us to focus on graph matching of asymmetric graphs without loss of generality. As the second contribution, we describe the geometric interpretation of the constrained set and derive an orthogonal projection operator associated with it. We then propose a distributed optimization algorithm to compute the permutation matrix that matches two isomorphic and asymmetric graphs, over the multi-agent network. Further, we establish the globally exponential convergence of the agents’ estimates of the permutation matrix to the actual one, assuming that the interaction graph of the agents is connected. Finally, we illustrate the theoretical analysis through simulations.

The rest of the paper is organized as follows. Preliminaries and the graph matching problem are presented in Section II. In Section III, we investigate a projected optimization algorithm over a multi-agent system to compute the graph matching. An example of matching two isomorphic and asymmetric graphs is presented in Section IV. Finally, Section V concludes this paper.

II. PRELIMINARIES AND PROBLEM FORMULATION

Notation: Let $G_1 = (V, E_1, A)$ and $G_2 = (V, E_2, B)$ be two undirected graphs of $n$ vertices whose index set is $V = \{1, \ldots, n\}$ and edge sets are $E_1, E_2 \subseteq V \times V$, respectively. In addition, $A$ and $B \in [0, \infty)^{n \times n}$ denote the symmetric adjacency matrices of the undirected graphs $G_1$ and $G_2$, respectively, whose entries are non-negative scalar weights characterizing the interactions between the vertices. Let $a_i$ and $b_j \in \mathbb{R}^n$ be the $i$th column vectors of $A$ and $B$, respectively. The space of vertex permutations is denoted by $P = \{\pi : V \rightarrow V\}$, which is characterized by the set of permutation matrices $\{\Pi : \Omega = \Pi^T 1_n = 1_n\}$, where $1_n$ is the vector of all ones. Let $\text{col}(x_1, \ldots, x_n) = [x_1, \ldots, x_n]^\top$ be the stack vector of $x_1, \ldots, x_n \in \mathbb{R}^n$. Denote the inner product of two vectors or two matrices of the same size as $\langle \Phi, X \rangle = \sum_{i,j} \Phi_{ij} X_{ij}$, where $(\cdot)_{ij}$ is the $(i,j)$-th entry. Let $(X)_i^C$ and $(X)_i^R$ be the $i$th column vector and the $i$th row vector of a matrix $X$, respectively. The notation $|| \cdot ||$ denotes the Euclidean norm. Let $\mathbb{R}^n_{+}$ be the nonnegative orthant of $\mathbb{R}^n$. Denote by $B_\epsilon(x), x \in \mathbb{R}^n$ as the open ball centered at a point $x$ with radius $\epsilon > 0$.

A. Convex Analysis

A set $\Omega \subseteq \mathbb{R}^n$ is convex if for any $x$ and $y \in \Omega$ and $\alpha \in [0, 1]$, it holds $\alpha x + (1-\alpha) y \in \Omega$. A function $f : \Omega \rightarrow \mathbb{R}$ is convex if $f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$, for any $x$ and $y \in \Omega$ and $\alpha \in [0, 1]$.

1In this work, graphs are considered to be weighted without further explicit mention.

B. Convergence Analysis

Consider the autonomous system

$$\dot{x} = f(x), \quad x(0) = x_0,$$

where $f : \Omega \rightarrow \mathbb{R}^n$ is a Lipschitz continuous map from a set $\Omega \subset \mathbb{R}^n$ into $\mathbb{R}^n$. Let $\bar{x} \in \Omega$ be an equilibrium point of $f$. Consider a solution trajectory $x(t)$ of $f$. A point $p$ is said to be a positive limit point of $x(t)$ if there exists a sequence $\{t_m\}$, with $t_m \rightarrow \infty$ as $m \rightarrow \infty$, such that $x(t_m) \rightarrow p$ as $m \rightarrow \infty$. A set $M$ is said to be a positively invariant set if $x(0) \in M \Rightarrow x(t) \in M, \forall t \geq 0$. The equilibrium point $\bar{x}$ of $f$ is stable if, for each $\epsilon > 0$, there is $\delta = \delta(\epsilon) > 0$ such that $x(0) \in B_\epsilon(\bar{x}) \Rightarrow x(t) \in B_\delta(\bar{x}), \forall t \geq 0$.

Lemma 1: [23, Theorem 4.4] Let $\Omega \subseteq \mathbb{R}^n$ be a compact set that is positively invariant with respect to $f$. Let $V : \Omega \rightarrow \mathbb{R}$ be a continuous differentiable positive definite function such that $\nabla V(x) \leq 0$ in $\Omega$. Let $S = \{x \in \Omega : \nabla V(x) = 0\}$. Let $M$ be the largest invariant set in $S$. Then, every solution starting in $\Omega$ approaches $M$ as $t \rightarrow \infty$.

Lemma 2: [17, Lemma 2.2] Suppose that the system $f$ is a linear time-invariant system, i.e., $f(x) = Mx + b$, where $M \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$, and $\Omega = \mathbb{R}^n$. Then, if the system $f$ converges to an equilibrium for any initial condition, its convergence is exponentially fast.

C. Graph Matching

A permutation $\pi$ maps a vertex $i$ in $G_1$ to a vertex $\pi_i$ in $G_2$, and associates each $(i,j)$-th entry of $A$ to an entry $(B)_{\pi_i,\pi_j}$ in $G_2$. Let $\Pi$ be the permutation matrix associated with this permutation $\pi$. Then, $A = \Pi^T B \Pi$ for exact matching. We denote by

$$\text{dis}_{G_1 \leftrightarrow G_2} (\Pi) = ||A - \Pi^T B \Pi||_F$$

the distortion function specifying the adjacency disagreement between $A$ and $B$. Two graphs $G_1$ and $G_2$ are said to be isomorphic if their adjacency disagreement is zero in the sense of Eq. (2), for some permutations $\Pi$. Denote by $\text{Iso}(G_1 \leftrightarrow G_2) = \{\Pi \in P : \text{dis}_{G_1 \leftrightarrow G_2}(\Pi) = 0\}$ the set of all permutations, or, i.e., isomorphisms, matching $G_1$ and $G_2$.

A graph $\mathcal{G} = (V, A)$ is symmetric if there exists a nontrivial permutation $\Pi \in P$ such that $\text{dis}_{\mathcal{G} \rightarrow \mathcal{G}} (\Pi) = 0$, or, i.e., $\mathcal{G}$ has a nontrivial automorphism group $\text{Iso}(\mathcal{G} \rightarrow \mathcal{G})$. The graph $\mathcal{G}$ is asymmetric if it is not symmetric, or equivalently the only permutation matrix satisfying $\text{dis}_{\mathcal{G} \rightarrow \mathcal{G}} (\Pi) = 0$ is the trivial identity matrix. If two asymmetric graphs $G_1$ and $G_2$ are isomorphic, they are related by a unique permutation, denoted by $\Pi^*$, which is the global solution of problem (3) below. On the other hand, given two isomorphic and symmetric graphs $G_1$ and $G_2$, there are two or more permutation matrices which satisfy $\text{dis}_{G_1 \leftrightarrow G_2} = 0$.

Remark 1: The symmetry/asymmetry property defined above for a graph, which is characterized by the uniqueness of its automorphisms, is also different from structural symmetry/structural asymmetry of the associated graph topology.

2The symmetry or asymmetry of a graph should be distinguished from the symmetry of its associated adjacency matrix. The latter itself is symmetric simply when the graph $\mathcal{G}$ is undirected and connected.
In Fig. 1, \(G_1\) and \(G_3\) are \textit{structurally symmetric} with respect to their vertical axes. That is, the permutations of the corresponding vertices on the two sides of the vertical dashed line leave the graphs unchanged. As a comparison, the only permutation under which the graphs \(G_2\) and \(G_4\) are invariant is the identity mapping. As a result, \(G_2\) and \(G_4\) in Fig. 1 are \textit{structurally asymmetric}. A structurally asymmetric graph is also asymmetric for all the edge weights in its associated adjacency matrices. A symmetric graph has a structurally symmetric graph topology and has some pairs of edges in which edges in each pair have the same weight. For instance, \(G_3\) in Fig. 1 is symmetric when the weights of every two corresponding edges on the two sides of the vertical dashed line are equal.

We now define a subclass of graphs which both covers almost all graphs with a given number of nodes, and is the subclass which we will prove to be as large as the set of asymmetric graphs.

**Definition 1 (Friendly Graphs):** [1] A graph \(G\) is said to be \textit{friendly} if its adjacency matrix \(A\) has simple spectrum (all eigenvalues are distinct) and eigenvectors satisfy \(u_i^\top 1_n \neq 0\), for all \(i = 1, \ldots, n\).

The properties of adjacency matrices of friendly graphs are important in convexly relaxing the graph matching problem in the next subsection. A friendly graph is necessarily asymmetric [1, Lemma 1]. In addition, it will be shown below that almost all adjacency matrices of asymmetric graphs are also friendly. That is, the properties of the adjacency matrices of unfriendly graphs are nongeneric. In fact, it is shown in [24] that almost all adjacency matrices of random graphs have simple spectrum. In addition, in the sequel, we show that almost all adjacency matrices with nonnegative entries have eigenvectors that are not orthogonal to a given nonzero vector, e.g., \(1_n\).

**Theorem 1:** The set of weighted adjacency matrices which have nonnegative entries and an eigenvector orthogonal to \(1_n\) is a set of measure zero.

**Proof:** See Appendix A.

We, therefore, focus on matching two asymmetric graphs by implicitly assuming friendliness for each, which holds almost surely according to above results.

### D. Convex relaxation of graph matching

The graph matching problem is to find an optimal permutation matrix \(\Pi^*\) satisfying the following optimization problem:

\[
(P) \quad \Pi^* = \arg\min_{\Pi \in \mathcal{P}} \|	ext{tr}(B^\top A - \Pi^\top B^\top A)\|_F.
\]

where \(\text{tr}\) stands for the trace of a matrix, and \(\mathcal{P}\) is the symmetric positive definite matrices. The polytope \(\mathcal{P}\) is non-convex and so the problem \((P)\) is not a convex problem. Although the set is finite, it has \(n!\) elements, which means that for large \(n\), exhaustive search will become infeasible. Thus, it is desirable to replace \(\mathcal{P}\) with a larger convex set. For example, consider the space of \textit{pseudo-stochastic} matrices \(\mathcal{D}_n := \{P : P1_n = 1_n\}\) and the following relaxation of the graph matching problem (RGM):

\[
(RGM) \quad P^* = \arg\min_{P \in \mathcal{D}_n} \|PA - B^\top P\|_F.
\]

We focus in this paper just on the issue of finding the solution to the \textit{isomorphic graph matching} problem, i.e., when the two graphs in question are isomorphic. We now have the following lemma.

**Lemma 2:** [1] Let \(G_1\) and \(G_2\) be two asymmetric and isomorphic graphs. Then, the problems \((P)\) and \((RGM)\) are equivalent.

**Remark 2:** In some circumstances, we might deal with isomorphic graph matching in the presence of noise. Let \(G_1\) and \(G_2\) be isomorphic and asymmetric graphs with spectrum radius \(\sigma = \max_i |\lambda_i|\), related by the unique isomorphism \(\Pi^*\). The adjacency matrix \(A(G_1)\) has the spectrum gap \(\min_{i \neq j} |\lambda_i - \lambda_j| > \delta\) and all eigenvectors satisfy \(\varepsilon < |u_i^\top 1_n| < \frac{\varepsilon}{\delta}, \delta > 0\). Let \(B\) be a perturbed adjacency matrix of \(B(G_2)\) such that \(B = B + \rho R\), where \(R\) is symmetric with \(\|R\|_F \leq 1\) and \(\rho \leq \min \{\sqrt{2\delta}, \sqrt{\frac{\varepsilon^2}{\delta^2 \min_{i \neq j} |\lambda_i - \lambda_j|}}\}\). Then, the optimal solution \(P^*\) of the problem \((RGM)\) is unique and satisfies \(\|P^* - \Pi^*\|_F < 1/2\) [1, Lemma 2]. The perturbation \(\rho R\) characterizes the total adjacency disagreement that still allows the optimal permutation to be recovered. The conclusion of course does not actually depend on \(B^\top\) obtained by the addition of noise to \(B\) for which isomorphism holds, but simply on having a \(B\) suitably close to \(B\). By way of a side remark, we note that if \(B\) contains zero entries, some entries of \(R\) might need to be non-negative so that \(B\) is a proper adjacency matrix. Consequently, we can solve \((RGM)\) for \(P^*\) and project it onto \(\mathcal{P}\) to get \(\tilde{\Pi}\). If \(\text{dis}_{G_1 \to G_2} (\Pi)\) is small enough the graphs are isomorphic. The orthogonal projection \(P^*\) onto \(\mathcal{P}\) can be obtained by optimizing the Euclidean inner product

\[
\tilde{\Pi} = \text{Proj}_\mathcal{P} P^* = \arg\max_{\Pi \in \mathcal{P}} \text{tr}(\Pi^\top P^*),
\]

which is solvable in polynomial time using the Hungarian method [25]. In addition, it follows from \(\|P^* - \Pi^*\|_F < 1/2\)
Algorithm 1 Centralized Computation of Graph Matching.

Require: A, B adjacency matrices of isomorphic and asymmetric graphs.
1: Solve the convex optimization problem (4) for $P^*$.
2: Project $P^*$ onto $P$.
3: return $\Pi$.

in the preceding remark that $|(P^*)_{ij} - (\Pi)_{ij}^*| < 1/2$, for all $i,j$ entries. Thus, the projection $\text{Proj}_P P^*$ can be simply obtained by rounding up/down the entries of $P^*$ to the nearest integer numbers in $\{0,1\}$, i.e., $\arg\min_{x\in\{0,1\}} |(P^*)_{ij} - x|$ entry-wise for every $i,j$.

In the absence of noise, the optimal solution of (4) is identical to $\Pi^*$, and hence a distributed algorithm with asymptotic stability property can compute $\Pi^*$ as $t \to \infty$. Further, such a permutation can be obtained after a finite time using projection when $||P(t) - \Pi^*||_F < 1/2$, as discussed above. In summary, there are two steps in solving the (GM) problem as summarized in Algorithm 1.

III. DISTRIBUTED OPTIMIZATION TO SOLVE RGM

In this part, we formulate the RGM problem (4) as a distributed optimization problem and propose a distributed optimization over a multi-agent network of $n$ agents to solve it.

Multi-agent network: We assume that each agent $i$ in a network of $n$ agents only knows $a_i$ and $b_i \in \mathbb{R}^n$, the $i$th column vectors of $A$ and $B$, respectively, and can exchange information with some neighboring agents. This exchange process itself, which effectively defines the way calculations determining the optimum $P^*$ are distributed, can be modelled by a graph. To differentiate with the two graphs $G_1$ and $G_2$ to be matched, we denote the interaction graph of the agents as $\mathcal{H} = (\mathcal{I}, \mathcal{E}_H)$, where $\mathcal{I} = \{1, \ldots, n\}$ and $\mathcal{E}_H \subseteq \mathcal{I} \times \mathcal{I}$ denote the index set and edge set of the agents, respectively. When agents $i$ and $j$ are neighbors, i.e., $(i,j) \in \mathcal{E}_H$, we associate with this edge an arbitrary positive weight $w_{ij} = w_{ji} > 0$; when $(i,j) \notin \mathcal{E}_H$, $w_{ij} = w_{ji} = 0$. The graph $\mathcal{H}$ is assumed to be undirected and connected.

A. Geometric interpretation of the pseudo-stochastic matrix set $D_n$ and projection operator

The sum of the elements in any row vector of a matrix in $D_n = \{P : P1_n = 1_n\}$ is one. Thus, pseudo-stochastic matrices contain $n$ rows which, if each row vector is considered as a point in the $n$-dimensional Cartesian space, belong to a hyperplane in $\mathbb{R}^n$, i.e., the plane $\sum_{k=1}^n x_k = 1$, where $x_k, k = 1, \ldots, n$ are the coordinates of a point vector $x \in \mathbb{R}^n$. Let $S_n$ denote this plane. Then, the normal vector of the hyperplane $S_n$ is $1_n$ (See Fig. 2). For simplicity, we say a row of a matrix $X \in D_n$ belongs to the hyperplane $S_n$ when there is no risk of confusion.

The orthogonal projection of a vector $v \in \mathbb{R}^n$ onto the hyperplane $S_n$ is given as

$$\text{Proj}_{S_n}(v) := (I_n - (1/n)1_n1_n^\top)v.$$  

Fig. 2: Geometric interpretation of the pseudo-stochastic matrix set $D_n$. A point, whose coordinates are the elements of any row vector of a matrix in $D_n$, lies in the hyperplane $S_n$ in the $n$-dimensional space.

Given a $n \times n$ matrix $V$, we denote by $\text{Proj}_{D_n}(V)$ the orthogonal projection of $V$ onto $D_n$, i.e.,

$$\text{Proj}_{D_n}(V) := V(I_n - (1/n)1_n1_n^\top),$$

which consists of $n$ orthogonal projections of the $n$ corresponding row vectors of $V$ onto $S_n$. The matrix $V$ is said to be parallel to $D_n$ if $\text{Proj}_{D_n}(V) = V$, and is orthogonal to $D_n$ if $\text{Proj}_{D_n}(V) = 0$.

Lemma 4: Let an arbitrary vector $u \in \mathbb{R}^n$ and any two points $x,y \in S_n$. Then, there holds

$$(x - y)^\top \text{Proj}_{S_n}(u) = (x - y)^\top u.$$  

Proof: We have

$$(x - y)^\top \text{Proj}_{S_n}(u) = (x - y)^\top (I_n - (1/n)1_n1_n^\top)u
= (x - y)^\top u - (1/n)(x - y)^\top 1_n1_n^\top u
= (x - y)^\top u,$$

where the last equality follows from $(x - y) \perp 1_n$, for any two points $x,y \in S_n$. \hfill \square

B. GM as a distributed optimization problem

Assume that each agent $i \in \mathcal{I}$ knows $a_i$ and $b_i$ and holds a local estimate of the common optimal matrix $P^*$ of the problem (RGM), denoted by $P_i \in \mathbb{R}^{n \times n}$. The agents cooperatively estimate $P^*$ such that $P_i \to P^*, \forall i \in \mathcal{I}$, as $t \to \infty$. In the case of exact matching, due to $A = A^\top$ the optimal solution to the problem (4) is equivalent to finding a matrix $P^* \in D_n$ satisfying the system of equations

$$\begin{cases}
P^*a_i = y_i^* , \; i = 1, \ldots, n, \\
b_i^\top P^* = z_i^* , \; i = 1, \ldots, n, \\
y^* \triangleq [y_1^*, \ldots, y_n^*] = Z^* \triangleq \text{col}(z_1^\top, \ldots, z_n^\top), \\
P^* \in D_n,
\end{cases}$$

where $y_i^*, z_i^* \in \mathbb{R}^n$ for all $i \in \mathcal{I}$. As a result, in addition to $P_i$ each agent $i \in \mathcal{I}$ maintains two variables $y_i$ and $z_i$. The third relation in the preceding equation is introduced to impose equality constraints involving $y_i$ and $z_i$, i.e., $Y = Z$. While the consensus constraints $P_i = P_j, \forall i,j \in \mathcal{I}$, can be easily treated, i.e., through distributed averaging, the coupling constraint $Y = Z$ is not separable to each agent $i, \forall i \in \mathcal{I}$ since it only knows the $i$th column of $Y$ and the $i$th row of $Z$. To deal with such coupling constraint, we use
the following transformation, for all $i \in I$ (see e.g., [16], [17]):
\[
Y = Z \iff \exists \{K_i\}_{i=1}^n, \text{ s.t.}
\]
\[
[Y]_i - [Z]_i - \sum_{j=1}^n w_{ij}(K_i - K_j) = 0_{n \times n},
\]
where
\[
[Y]_i := [0_{n \times (i-1)}, y_i, 0_{n \times (n-i)}] \in \mathbb{R}^{n \times n}
\]
is an $n \times n$ matrix whose $i$th column is $y_i$ and other columns are zeros,
\[
[Z]_i := \text{col}(0_{(i-1) \times n}, z_i^T, 0_{(n-i) \times n}) \in \mathbb{R}^{n \times n}
\]
is a matrix of all zero row vectors except the $i$th row is being $z_i^T$, and $K_i \in \mathbb{R}^{n \times n}, i = 1, \ldots, n$ are used to compensate for the inconsistencies between $[Y]_i$ and $[Z]_i$. Note that $Y = \sum_{i=1}^n [Y]_i$ and $Z = \sum_{i=1}^n [Z]_i$. In addition, by summing up both sides of (10) over $i$ from 1 to $n$ we obtain $Y = Z$.

Let $X = \text{col}(P_1, \ldots, P_n)$ be the stack of all local estimates of $P^*$ at some intermediate point in the execution of the algorithm. Using the relations (9) and (10), we reformulate the problem (4) as a distributed optimization problem subject to a global set constraint and equality constraints, over the multi-agent network:
\[
\begin{align*}
\arg & \min_{X, Y, Z, K} \frac{1}{2} \sum_{i=1}^n ||P_i a_i - y_i||^2 \\
\text{s.t.} & \quad b_i^T P_i = z_i^T, \quad P_i \in \mathcal{D}_n, \\
& \quad \sum_{j=1}^n w_{ij}(P_i - P_j) = 0_{n \times n}, \\
& \quad [Y]_i - [Z]_i - \sum_{j=1}^n w_{ij}(K_i - K_j) = 0_{n \times n},
\end{align*}
\]
for all $i \in I$, where $K = \text{col}(K_1, \ldots, K_n)$. In addition, $P_i a_i$ is the agent $i$'s estimate of the $i$th component of $P^* A$ and is assigned to the vector $y_i$, while $b_i^T P_i$ is its estimate of the $i$th row of $BP^*$ and is stored at the row vector $z_i^T$. These two vectors satisfy the last coupling equality constraint in (11). Consequently, the agents cooperatively learn the graph matching using knowledge of $(a_i, b_i)$ and the auxiliary state variables $(P_i, y_i, z_i, K_i)$, for all $i \in I$. In the sequel, we develop a projected multi-agent optimization algorithm to solve the constrained optimization problem (11).

C. Distributed learning scheme

Consider the Lagrangian function of the problem (11)
\[
\mathcal{L}_1 = \frac{1}{2} \sum_{i=1}^n ||P_i a_i - y_i||^2 + \sum_{i=1}^n \langle \lambda_i^T, b_i^T P_i - z_i^T \rangle \\
+ \sum_{i=1}^n \langle \Theta_i, \sum_{j=1}^n w_{ij}(P_i - P_j) \rangle \\
+ \sum_{i=1}^n \langle \Upsilon_i, [Y]_i - [Z]_i - \sum_{j=1}^n w_{ij}(K_i - K_j) \rangle,
\]
where $\lambda_i \in \mathbb{R}^n$, $\Theta_i \in \mathbb{R}^{n \times n}$ and $\Upsilon_i \in \mathbb{R}^{n \times n}$ are Lagrange multipliers of agent $i$ associated with the equality constraints in problem (11), for all $i \in I$. We first define
\[
\lambda^T = (\lambda_1^T, \ldots, \lambda_n^T) \in \mathbb{R}^{n \times n}, \\
\Theta = (\Theta_1, \ldots, \Theta_n) \in \mathbb{R}^{n \times n \times n}, \\
\Upsilon = (\Upsilon_1, \ldots, \Upsilon_n) \in \mathbb{R}^{n \times n \times n}, \\
Q = \text{col}(X, Y, Z, K, \lambda, \Theta, \Upsilon).
\]

In addition, let $Q^* = \text{col}(X^*, Y^*, Z^*, K^*, \lambda^*, \Theta^*, \Upsilon^*)$ be an optimal solution of (11). Then, we have the following lemma.

Lemma 5: The optimal states $(P_i^*, y_i^*, z_i^*, K_i^*)$ and the Lagrange multipliers (or the dual optimal variables) $(\lambda_i^*, \Theta_i^*, \Upsilon_i^*)$, for all $i \in I$, satisfy the following necessary and sufficient optimality condition:
\[
\begin{align*}
\text{Proj}_{\mathcal{D}_n} \left( \{P_i^* a_i - y_i^*\} a_i^T + b_i^T \lambda_i^T + \right. \\
\left. + \sum_{j=1}^n w_{ij}(\Theta_i^* - \Theta_j^*) \right) = 0, \\
\nabla_{\lambda_i^T} \mathcal{L}_1(Q^*) = b_i^T P_i^* - z_i^T = 0, \\
\nabla_{\Theta_i} \mathcal{L}_1(Q^*) = \sum_{j=1}^n w_{ij}(P_i^* - P_j^*) = 0, \\
\nabla_{\Upsilon_i} \mathcal{L}_1(Q^*) = [Y_i^*]_i - [Z_i^*]_i \\
- \sum_{j=1}^n w_{ij}(K_i^* - K_j^*) = 0, \\
\n\nabla_{y_i} \mathcal{L}_1(Q^*) = -(P_i^* a_i - y_i^*) + (Y_i^*)_i = 0, \\
\n\nabla_{z_i} \mathcal{L}_1(Q^*) = -\lambda_i^T + (Y_i^*)_i = 0, \\
\n\nabla_{K_i} \mathcal{L}_1(Q^*) = [Y_i^*]_i - [Z_i^*]_i = 0.
\end{align*}
\]

Proof: $(P_i^* a_i - y_i^*) a_i^T + b_i^T \lambda_i^T + \sum_{j=1}^n w_{ij}(\Theta_i^* - \Theta_j^*) + N_{\mathcal{D}_n}(P_i^*) = 0, \quad (13)$
\]
\]
\]
\]
\]
\]
\]
\]
\]
\]
\]

for an $n \times n$ matrix $N_{\mathcal{D}_n}(P_i^*) = x_1^T$ with an arbitrary $x \in \mathbb{R}^n$, whose row vectors are in span$(1_n^1)$, or, i.e., orthogonal to the hyperplane $S_n$ defined in Section III-A. Since the subspaces $\mathcal{D}_n$ and $N_{\mathcal{D}_n}$ are orthogonal in the sense that $\langle \Delta, x_1^T \rangle = 0$ for any $\Delta \in \mathbb{R}^{n \times n}$ parallel to $\mathcal{D}_n$, (13) is equivalent to (12a).

Consider the augmented Lagrangian:
\[
\mathcal{L}_2 = \mathcal{L}_1 + \frac{1}{2} \sum_{i=1}^n \langle P_i, \sum_{j=1}^n w_{ij}(P_i - P_j) \rangle \\
+ \frac{1}{2} \sum_{i=1}^n ||b_i^T P_i - z_i^T||^2 - \frac{1}{2} \sum_{i=1}^n \langle \Upsilon_i, \sum_{j=1}^n w_{ij}(Y_i - Y_j) \rangle, \\
\]
where the last three terms are augmented terms. The additional (quadratic) augmented terms in $\mathcal{L}_2$, which vanish at an optimal solution $Q^*$ of problem (11) due to Eq. (12a), are used to impose further constraints in the positively invariant set of the system (15), as will be shown in Eq. (18).
Primal-dual gradient methods using augmented Lagrangian functions can be found in [17], [20], [27]. We employ the primal-dual gradient method for the problem (11) that evolves on the manifold \(D_n\), as described in what follows.

We propose the following distributed algorithm to solve the problem (11) based on the saddle-point dynamics of the augmented Lagrangian function \(\mathcal{L}_2\).

**Algorithm 2** Distributed Algorithm for Solving (11).

1. **Initialize**: \(P_i(0) \in D_n, y_i(0) \in \mathbb{R}^n, z_i(0) \in \mathbb{R}^n, K_i(0) \in \mathbb{R}^{n \times n}; \lambda_i(0) \in \mathbb{R}^n, \Theta_i(0) \in \mathbb{R}^{n \times n}\) and \(Y_i(0) \in \mathbb{R}^{n \times n}, \forall i \in \mathcal{I}\).

2. **Update rules**:

\[
\begin{align*}
\dot{P}_i(t) &= \text{Proj}_{D_n} \left\{ - (P_i a_i - y_i) a_i^T - b_i \lambda_i^T \right\} \\
&
- \sum_{j=1}^n w_{ij} (\Theta_i - \Theta_j) + \sum_{j=1}^n w_{ij} (P_i - P_j) \quad (15a) \\
\dot{\lambda}_i^T(t) &= b_i^T P_i - z_i^T \\
\dot{\Theta}_i(t) &= \sum_{j=1}^n w_{ij} (P_i - P_j) \quad (15c) \\
\dot{Y}_i(t) &= [Y_i - K_i^T - \sum_{j=1}^n w_{ij} (K_i - K_j)] \\
& \quad - \sum_{j=1}^n w_{ij} (Y_i - Y_j) \quad (15d) \\
\dot{y}_i(t) &= (P_i a_i - y_i) - (Y_i)_i^T \\
\dot{z}_i^T(t) &= \lambda_i^T + (Y_i)_i^T + b_i^T P_i - z_i^T \\
\dot{K}_i(t) &= \sum_{j=1}^n w_{ij} (Y_i - Y_j) \quad (15g)
\end{align*}
\]

for all \(i = 1, \ldots, n\).

Then, in the light of Lemma 3, the optimal solution of the problem (11) has a unique \(\Pi^*\).

**Lemma 7**: Let \(G_1\) and \(G_2\) be two asymmetric and isomorphic graphs, related by an unique permutation \(\Pi^* \in \mathcal{P}\). Then, \((X^*, Y^*, Z^*, K^*)\) is an optimal solution of the problem (11) if and only if there exist \(\lambda^* \in \mathbb{R}^n, \Theta^* \in \mathbb{R}^{n \times n}, Y^* \in \mathbb{R}^{n \times n}\) such that \(Q^*\) is an equilibrium point of (15). Moreover, such \(\Pi^*\) is \(\Pi^*, \forall i \in \mathcal{I}\).

**Proof**: The proof follows from the necessary and sufficient condition for optimality (12) and Lemma 2.

**Theorem 2**: Let \(G_1\) and \(G_2\) be two isomorphic and asymmetric graphs. Then, every trajectory \(Q(t)\) of the system (15) in Algorithm 2 with an initial condition \(Q(0)\), converges globally asymptotically to an equilibrium of (15). In addition, \(P_i \to \Pi^*\) as \(t \to \infty\), and the orthogonal projection of \(P_i\) onto \(\Pi\) is identical to \(\Pi^*\) after a finite time \(T > 0\) when \(\|P_i(T) - \Pi^*\| < 1/2\), for all \(i = 1, \ldots, n\).

**Proof**: It can be verified that the optimal solution \(Q^* = \text{col}(X^*, Y^*, Z^*, K^*, \lambda^*, \Theta^*, Y^*)\) satisfying the optimality condition (12) is an equilibrium of the system (15). Consider the Lyapunov function \(V = \frac{1}{2} \|Q - Q^*\|_F^2\) which is positive definite and radially unbounded. The time derivative of \(V\) along the trajectory of (15) is given as

\[
\dot{V} = \text{tr}(Q^T (Q - Q^*))
\]

\[
= \text{tr} \left\{ \sum_{i=1}^n \dot{P}_i^T (P_i - P_i^*) + \sum_{i=1}^n \dot{y}_i^T (y_i - y_i^*) \right\}
\]

\[
+ \sum_{i=1}^n \dot{z}_i^T (z_i - z_i^*)^T + \sum_{i=1}^n \dot{K}_i^T (K_i - K_i^*)
\]

\[
+ \sum_{i=1}^n \sum_{j=1}^n \dot{\psi}_i^T (\psi_i - \psi_i^*) \right\}. \quad (16)
\]

Let \(\Delta_i \in \mathbb{R}^{n \times n}\) be the expression inside the projection operator corresponding to agent \(i\) in (15a). Then, it follows from Lemmas 2 and 6 and the properties of the trace function that, for all \(i \in \mathcal{I}\),

\[
\text{tr}\{P_i^T (P_i - P_i^*)\} = \text{tr}\{P_i^T (P_i^* + \Delta_i) - P_i^T (P_i^* + \Delta_i)\}
\]

\[
= \text{tr}\{(P_i - P_i^* + \Delta_i)\text{Proj}_{D_i}((\Delta_i)^T)\}
\]

\[
= \sum_{j=1}^n (P_i - P_i^* + \Delta_i)^j \text{Proj}_{D_i}((\Delta_i)^j)^T
\]

\[
\equiv \sum_{j=1}^n (P_i - P_i^* + \Delta_i)^j (\Delta_i)^j
\]

\[
= \text{tr}\{(P_i - P_i^*) (\Delta_i)^T\}
\]

\[
= \text{tr}\{(P_i - P_i^*) \Delta_i^T\}
\]

\[
= \text{tr}\left\{ (P_i^T (P_i - P_i^*)) \right\} \right\} = \text{tr}\left\{ (P_i^T (P_i - P_i^*)) \Delta_i^T\right\}, \quad (17)
\]

where \((\cdot)^j\) denotes the \(j\)th row vector of the associated matrix.

**Step 1**: Negative semidefiniteness of \(\dot{V}(t)\)

By using the optimality condition (12) and the relation (17), it can be shown that \(\dot{V}\) in (16) along the trajectory...
of (15) has the following more concise expression (see Appendix B):

\[
\dot{V} = -\sum_{i=1}^{n} \left\{ \left( (P_i - P_i^*)a_i - (y_i - y_i^*) \right)^2 + \left( b_i^T P_i - z_i \right)^2 \right\} - \sum_{(i,j) \in E_n} w_{ij} \left\{ \|P_i - P_j\|_F^2 + \|Y_i - Y_j\|_F^2 \right\},
\]

(18)

which is negative semidefinite. Consequently, \(V(t)\) is bounded, i.e., \(V(t) \leq V(0)\). Since \(V(Q)\) is radially unbounded, every level set \(\Omega_c = \{ Q : V(Q) \leq c \}\) with a positive \(c \in \mathbb{R}\), is a compact, positively invariant set. It follows that any trajectory \(Q(t)\) of the system (15) is bounded and converges to the largest invariant set that contains \(Q\) such that \(\dot{V}(Q) = 0\) due to the LaSalle’s invariance principle (Lemma 1).

**Step 2: Globally Asymptotic convergence of the optimal solution**

Let the invariant set \(S_Q \triangleq \{ Q : \dot{V}(Q) = 0 \} = \{ Q : P_i = P_j, P_i \in D_n, Y_i = Y_j, (P_i - P_i^* a_i = y_i - y_i^*, b_i^T P_i = z_i)^i, \forall i, j \in I, i \neq j \}\). We consider a solution trajectory of (15) that satisfies \(Q(t) \in S_Q\). Then, there holds:

i) \(\dot{\lambda}_i(t) = 0, \dot{y}_i = 0\) and \(\dot{K}_i = 0\) for all \(i = 1, \ldots, n\).

Thus, \(\dot{\lambda}_i, \dot{\lambda}_j, K_i\) are constants for all \(i = 1, \ldots, n\).

ii) \(\Delta_i := -(P_i a_i - y_i^* a_i^T) + b_i^T \lambda_i = -(\sum_{j=1}^{n} w_{ij} (\Theta_j - \Theta_i) - \sum_{j=1}^{n} w_{ij} (\Theta_j - \Theta_i) - \sum_{j=1}^{n} w_{ij} (\Theta_j - \Theta_i))^i - b_i^T P_i - z_i = 0\) is constant because \(\dot{\lambda}_i, \dot{\lambda}_j\) are constants. As a result, \(P_i = \text{Proj}_{D_n}(\Delta_i)\) is constant. Since \(P_i, \forall i \in I\) is bounded, we have \(P_i = 0\), for all \(i \in I\). It follows that \(P_i, \forall i \in I\) are constants.

iii) \(\dot{z}_i = b_i^T P_i\) is constant and hence \(\dot{\lambda}_i = 0\) for all \(i = 1, \ldots, n\). Similarly, \(\dot{y}_i = (P_i - P_i^* a_i = y_i^* a_i^T)\) is constant and consequently \(\dot{y}_i = 0\), for all \(i = 1, \ldots, n\).

As a result, \(\dot{Y}_i = [Y_i - Z_i]^i - \sum_{j=1}^{n} w_{ij} (\Theta_j - \Theta_i)^i\) is constant. It follows that \(\dot{Y}_i = 0\) due to boundedness of \(Y_i\), for all \(i = 1, \ldots, n\).

It follows that the largest invariant set in \(S_Q\) contains only the equilibrium point of (15), which is globally asymptotically stable. It follows from Lemma 7 that any trajectory \((X, Y, Z, K)\) with initial condition \(Q(0)\) converges globally and asymptotically to an optimal solution of problem (11).

**Theorem 3:** Let \(G_1\) and \(G_2\) are two isomorphic and asymmetric graphs. Then, the trajectory \(Q(t)\) of the system (15) in Algorithm 2 with an initial condition \(Q(0)\) converges globally exponentially to an equilibrium of (15). In addition, \(P_i \rightarrow \Pi^*\) exponentially fast as \(t \rightarrow \infty\).

**Proof:** The projected update law (15a) can be simply written as

\[
P_i = \Delta_i P_{\text{proj}},
\]

(19)

where \(\Delta_i\) is the expression inside the projection operator in (15a). It is noted that the preceding equation is a linear system, since \(\text{Proj}_{D_n}\) is just a linear operation. By using the preceding equation and the following relations

\[
(I_n \otimes X^T) \text{vec}(Y^T) = XY, \quad (X \otimes I_n) \text{vec}(Y^T) = XY,
\]

for any two matrices \(X, Y\) of suitable dimensions, the system (15) can be expressed as a linear invariant system of \(\text{vec}(Q^T)\). Then, the conclusion on exponential convergence follows from Lemma 2 and Theorem 2.

**Remark 4:** Though the time \(T > 0\) associated with convergence into the ball around \(\Pi^*\) has been shown to exist under the conditions in Theorem 2 it is not straightforward to estimate even an upper bound for this convergence time nor to determine in the course of executing the algorithm that the time has been reached. However, the evolution of \(P_i(t)\) with respect to time \(t\) might give some hint on whether an optimal permutation has been obtained by an agent \(i \in I\). In particular, every entry \((P_i)_{k,l}(t)\) will change relatively slowly after a sufficiently large time \(T_1 > 0\) due to the exponential convergence of \(P_i(t)\) to \(\Pi^*\). In addition, \((P_i)_{k,l}(t) - (\Pi^*)_{k,l}\) will also remain sufficiently small for all entries \(k, l\), and for a proper projected permutation \(\Pi^* = \text{Proj}_{\Pi^*} P_i(t)\), after a sufficiently large time.

**IV. SIMULATION**

Consider two weighed and connected graphs \(G_1, G_2\) of six vertices given in Fig. 3. In addition, the adjacency matrices associated with \(G_1, G_2\) are explicitly given as

\[
A = \begin{bmatrix}
0 & 1 & 0 & 0 & 0.95 & 0.05 \\
0 & 0 & 0.9 & 0.96 & 0.05 & 0.05 \\
0.9 & 0 & 0.9 & 0.9 & 0.7 & 0.0 \\
0.96 & 0.05 & 0 & 0.9 & 0.9 & 0.05 \\
0.05 & 0.05 & 0.9 & 0 & 0.9 & 0.9 \\
0.05 & 0.05 & 0.0 & 0.9 & 0 & 0.9
\end{bmatrix},
\]

(20)

\[
B = \begin{bmatrix}
0 & 0 & 0.95 & 1.75 & 0.8 & 0.85 \\
0 & 0 & 0 & 1.75 & 0.8 & 0.85 \\
0 & 0 & 0 & 0 & 1.75 & 0.85 \\
0 & 0 & 0 & 0 & 0 & 1.75 \\
0 & 0 & 0 & 0 & 0 & 0.85 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}.
\]

(21)

It is noted that the two graphs are asymmetric and isomorphic. The computation graph \(\mathcal{G}\) of the multi-agent network is chosen to be identical to \(G_1\). The vertex-to-vertex matchings between \(G_1\) and \(G_2\) are illustrated in Fig. 4a. The optimal permutation matrix is given as

\[
\Pi^* = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}.
\]

(22)

It is observed from Fig. 4a that \(P_i \rightarrow \Pi^*\) asymptotically as \(t \rightarrow \infty\), for every \(i = 1, \ldots, 6\). In addition, the adjacency
and δ eigenvectors of ˜ real scalars otherwise, so that A

\[ A \delta v_i + \delta A v_i = \lambda_i \delta v_i + \delta \lambda_i v_i \]

\[ A \sum_{k=1}^{n} \beta_{ik} v_k + \delta A v_i = \lambda_i \sum_{k=1}^{n} \beta_{ik} v_k + \delta \lambda_i v_i, \]

(23)

where the second equality follows from A v_i = \lambda_i v_i and by neglecting second-order terms. Left multiplying by v_i^T on both sides of the preceding relation gives

\[ \delta \lambda_i = v_i^T \delta A v_i \leq |\lambda_{max}(A)|. \]

(24)

Thus, by choosing ||δA|| sufficiently small ˜A will still have simple spectrum. Left multiplying by v_i^T, j ≠ i on both sides of (23) yields

\[ \beta_{ij} \lambda_j + v_i^T \delta A v_i = \beta_{ij} \lambda_i \]

\[ \iff \beta_{ij} = f_{ij}(\delta A) \equiv v_j^T \delta A v_i/(\lambda_i - \lambda_j), \forall j \neq i. \]

(25)

From the unity condition ˜v_i^T ˜v_i = 1 \iff (v_i + \delta v_i)^T (v_i + \delta v_i) = 1 \iff 1 + 2v_i^T \delta v_i = 1 \iff \beta_{ii} = 0, where we use ||δv_i||^2 = 0. As a result, suppose that v_i ⊥ 1_n, then ˜v_i ⊥ 1_n when \delta v_i = \sum_{k=1}^{n} \beta_{ik} v_k = 0 for sufficiently small ||δA||. It follows that \beta_{ik} = 0 \iff f_{ij}(\delta A) = 0, for all j = 1, . . . , n, j ≠ i, due to the mutually linear independence of \{v_k\}. Let δA ∈ R^m×n be a vector containing m_i nonnegative upper-diagonal terms and the other p_i = (n(n−1)/2 − m_i) real upper-diagonal terms of δA, respectively.

Then, for a small open ball B_r(0) the set

\[ \Omega_{δA} = \{ δA ∈ B_r(0) \cap (R^m×n) : \delta v_i = 0 \} \]

is either a set of measure zero or the entire set B_r(0)∩(R^m×n) [28]. It follows from δA v_i ⊥ v_j, \forall j = 1, . . . , n, j ≠ i that δA v_i = \gamma v_i for a scalar \gamma. Suppose that for all δA ∈ B_r(0) \cap (R^m×n) there holds δA v_i = \gamma v_i for a scalar \gamma and nonzero v_i = [v_{i1}, . . . , v_{in}]^T . Choose δA = \begin{bmatrix} c & 0 \\ 0 & 0 \end{bmatrix} for a nonzero scalar c and 0 matrices of proper dimensions, then δA v_i = \gamma v_i leads to cv_{i2} = \gamma v_{i1} and \delta v_{i2} = 0, \forall j = 3, . . . , n. Similarly, select (δA)_{23} = (δA)_{32} = c and the other entries are zeros, for a nonzero scalar c, then δA v_i = \gamma v_i leads to v_{i1} = 0. Consequently, v_i ≡ 0, which is a contradiction, and hence \Omega_{δA} is a set of measure zero.

Thus, ˜v_i is not orthogonal to 1_n for all A outside a set of measure zero. This completes the proof. ■
B. Proof of negative semi-definiteness of $\dot{V}$

**Proof:** To proceed, we consider the following relations:

$$
\text{tr}\left\{ -\sum_{i=1}^n w_{ij} (\Theta_i - \Theta_j) (P_i - P_j)^T \right\}
\leq \text{tr}\left\{ -a_i^T ((P_i - P_j^*) a_i - (y_i - y_i^*))^T \right\}
- \sum_{i=1}^n \sum_{j=1}^n w_{ij} (\Theta_i - \Theta_j) (\Theta_i^* - \Theta_j^*) (P_i - P_j)
= \sum_{i=1}^n \sum_{j=1}^n w_{ij} (\Theta_i - \Theta_j) (\Theta_i^* - \Theta_j^*) (P_i - P_j)
- \sum_{i=1}^n \sum_{j=1}^n w_{ij} (\Theta_i - \Theta_j) (\Theta_i^* - \Theta_j^*) (P_i - P_j)
\leq \sum_{i=1}^n \sum_{j=1}^n w_{ij} (\Theta_i - \Theta_j) (\Theta_i^* - \Theta_j^*) (P_i - P_j)
\leq \sum_{i=1}^n \sum_{j=1}^n w_{ij} (\Theta_i - \Theta_j) (\Theta_i^* - \Theta_j^*) (P_i - P_j).
$$

Furthermore,

$$
\text{tr}\left\{ \left( \dot{y}_i^T (y_i - y_i^*) \right)^T \left( \dot{y}_i^* (y_i - y_i^*) \right) \right\}
= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n w_{ij} (\Theta_i - \Theta_j) (\Theta_i^* - \Theta_j^*) (P_i - P_j)
\leq \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n w_{ij} (\Theta_i - \Theta_j) (\Theta_i^* - \Theta_j^*) (P_i - P_j).
$$

Combining the first term in (26) with the first term in (27) gives

$$
\text{tr}\left\{ -(P_i - P_j^*) a_i (y_i - y_i^*)^T (P_i - P_j^*)^T a_i
+ [(P_i - P_j^*) a_i (y_i - y_i^*)]^T (y_i - y_i^*) \right\}
= -\| (P_i - P_j^*) a_i (y_i - y_i^*) \|^2.
$$

Summing the two terms in the sum (16) corresponding to the last term in (15a) and the last term in (15f) we have

$$
\text{tr}\left\{ -(b_i^T P_i - z_i^*)^T b_i^T (P_i - P_j^*)
+ (b_i^T P_i - z_i^*)^T (z_i - z_i^*)^T \right\}
= -\| b_i^T P_i - z_i^* \|^2.
$$

The last term in (15a) in the sum (16) is given as

$$
\text{tr}\left\{ -\sum_{i=1}^n \sum_{j=1}^n w_{ij} (P_i - P_j)^T (P_i - P_j^*) \right\}
\leq \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n w_{ij} (\Theta_i - \Theta_j) (\Theta_i^* - \Theta_j^*) (P_i - P_j)
\leq \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n w_{ij} (\Theta_i - \Theta_j) (\Theta_i^* - \Theta_j^*) (P_i - P_j).
$$

By (15d) we have

$$
\text{tr}\left\{ \sum_{i=1}^n (K_i - K_i^*) \right\}
= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n w_{ij} (y_i - y_j) (K_i - K_j)
\leq \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n w_{ij} (y_i - y_j) (K_i - K_j).
$$

Following (15d) we have

$$
\text{tr}\left\{ \sum_{i=1}^n \dot{y}_i^T (y_i - y_i^*) \right\}
= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n w_{ij} (y_i - y_j) (y_i - y_j)
\leq \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n w_{ij} (y_i - y_j) (y_i - y_j).
$$
Lastly, using the relation \((12b)\) and \((12f)\), we have
\[
\text{tr}\left(\sum_{i=1}^{n} (\tilde{Y}_i - \tilde{Y}_i^*)^\top (\tilde{Y}_i - \tilde{Y}_i^*)\right) = \text{tr}\left(\sum_{i=1}^{n} (\tilde{Z}_i^\top - \tilde{Z}_i^* )^\top (\tilde{Z}_i^\top - \tilde{Z}_i^* )\right),
\]
Combining \((32)\) and \((33)\) gives
\[
\sum_{i=1}^{n} (\tilde{Y}_i - \tilde{Y}_i^*)^\top (\tilde{Y}_i - \tilde{Y}_i^*) + \text{tr}\left(\sum_{i=1}^{n} (\tilde{Y}_i - \tilde{Y}_i^*)^\top (\tilde{Y}_i - \tilde{Y}_i^*)\right) - \sum_{i=1}^{n} (\tilde{Z}_i^\top - \tilde{Z}_i^* )^\top (\tilde{Z}_i^\top - \tilde{Z}_i^* ) = \sum_{i=1}^{n} (\tilde{Y}_i - \tilde{Y}_i^*)^\top (\tilde{Y}_i - \tilde{Y}_i^*) + \text{tr}\left(\sum_{i=1}^{n} (\tilde{Y}_i - \tilde{Y}_i^*)^\top (\tilde{Y}_i - \tilde{Y}_i^*)\right) - \sum_{i=1}^{n} (\tilde{Z}_i^\top - \tilde{Z}_i^* )^\top (\tilde{Z}_i^\top - \tilde{Z}_i^* ) = (34).
\]
By combining all the preceding relations, we obtain \((18)\).

ACKNOWLEDGMENTS

The work of Q. V. Tran and H.-S. Ahn was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2019R1A4A1029003).

B. D. O. Anderson is supported by the Australian Research Council (ARC) under grant DP-160104500 and DP-190100887, also by Data61 - CSIRO.

REFERENCES


