R-Local Unlabeled Sensing: A Novel Graph Matching Approach for Multiview Unlabeled Sensing Under Local Permutations

AHMED ALI ABBASI [ID]**, ABIY TASISSA, AND SHUCHIN AERON** [ID]

Tufts University, Medford, MA 02155 USA

CORRESPONDING AUTHOR: AHMED ABBASI (e-mail: ahmed.abbasi@tufts.edu).

ABSTRACT Unlabeled sensing is a linear inverse problem where the measurements are scrambled under an unknown permutation leading to loss of correspondence between the measurements and the rows of the sensing matrix. Motivated by practical tasks such as mobile sensor networks, target tracking and the pose and correspondence estimation between point clouds, we study a special case of this problem restricting the class of permutations to be local and allowing for multiple views. In this setting, namely multi-view unlabeled sensing under local permutations, previous results and algorithms are not directly applicable. In this paper, we propose a computationally efficient algorithm, R-local unlabeled sensing (RLUS), that creatively exploits the machinery of indefinite relaxations of the graph matching problem to estimate the local permutations. Simulation results on synthetic data sets indicate that the proposed algorithm is scalable and applicable to the challenging regimes of low to moderate SNR.

INDEX TERMS Graph matching, multi-view, unlabeled sensing.

I. INTRODUCTION

Motivated by several practical problems such as sampling in the presence of clock jitter, mobile sensor networks and multiple target tracking in radar, the problem of unlabeled sensing was first considered in [1]. There, the authors derived information theoretic results for identification of unknown signal under linear measurements when the measurement correspondence is lost. Several generalizations of this problem as well as specific cases have been considered in [2]–[13]. A detailed survey of these works vis-a-vis our work is outlined in the related work Section I-B. Below we first introduce and motivate the model considered here.

In this paper, we study a specific case of the unlabeled sensing problem where the set of permutations, modeling the loss of correspondence, is local. Specifically, we consider the task of estimating the unknown permutation $\mathbf{P}^* \in \Pi_{n,r}$ from the views $\mathbf{Y} \in \mathbb{R}^{n \times m}$ such that

$$\mathbf{Y} = \mathbf{P}^* \mathbf{B} \mathbf{X}^* + \mathbf{N}. \tag{1}$$

In the representation above, $\mathbf{B} \in \mathbb{R}^{n \times d}$ is the known measurement matrix, $\mathbf{N} \in \mathbb{R}^{n \times m}$ is additive noise with i.i.d entries
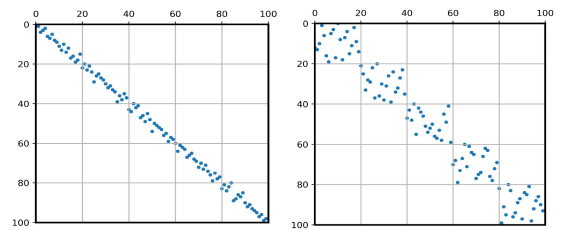

FIGURE 1. Left: The sparse permutation with *r*-local structure considered in [11] with $r = 20$. A permutation is sparse [10] if it has a small number of off diagonal elements. Right: 20-local permutation without sparsity.

from the Gaussian distribution $\mathcal{N}(0, \sigma^2)$ and $\mathbf{P}^* \in \Pi_{n,r}$ is an *r*-local permutation defined below.

Definition 1: Permutation $\mathbf{P}$ is *r*-local, $\mathbf{P} \in \Pi_{n,r}$, if it is composed of n/r blocks along the diagonal, with each block being a permutation matrix of size $r \times r$. More precisely, $\mathbf{P} = \text{diag}(\mathbf{P}_1, \ldots, \mathbf{P}_{n/r})$, where $\mathbf{P}_i \in \Pi_r$ denotes an $r \times r$ permutation matrix.

Fig. 1 shows examples of *r*-local permutations for $r = 20$.

TABLE 1. Summary of Existing Works. Π_n Denotes the General Set of Permutations of Order n. $\Pi_{n,s}$ Denotes the Set of Sparse Permutations. This Paper Considers the Set of r-Local Permutations, $\Pi_{n,r}$, See *Definition* 1, Fig. 1

Paper	Model $\left(\mathbf{P}^* \mid \mathbf{X}^*_{d \times m} \mid \sigma^2\right)$	Approach
[2]	$\mathbf{P}^* \in \Pi_n \mid m = 1 \mid \sigma^2 > 0$	Likelihood maximization
[3]	$\mathbf{P}^* \in \Pi_n \mid m = 1 \mid \sigma^2 > 0$	Dynamic programming
[4]	$\mathbf{P}^* \in \Pi_n \mid m = 1 \mid \sigma^2 > 0$	Symmetric Polynomials
[5]	$\mathbf{P}^* \in \Pi_n \mid d = 1, m = 1 \mid \sigma^2 > 0$	One dimensional sort
[6]	$\mathbf{P}^* \in \Pi_{n,s} \mid m = 1 \mid \sigma^2 > 0$	Linear program for $\mathbf{P}^*$, convex program for $\mathbf{X}^*$
[7]	$\mathbf{P}^* \in \Pi_n \mid m = 1$, fixed $\boldsymbol{x}^* \mid \sigma^2 = 0$	Lattice reduction
[8]	$\mathbf{P}^* \in \Pi_n \mid m = 1$, sparse $\boldsymbol{x}^* \mid \sigma^2 = 0$	Branch and bound
[9]	$\mathbf{P}^* \in \Pi_n \mid m = d \mid \sigma^2 = 0$	Subspace matching
[10]	$\mathbf{P}^* \in \Pi_{n,s} \mid m = d \mid \sigma^2 > 0$	Subspace clustering
[11]	$\mathbf{P}^* \in \Pi_{n,s} \mid m = d, \mathbf{X}^*(\mathbf{X}^*)^\mathsf{T} = \mathbf{I} \mid \sigma^2 > 0$	Orthogonal least squares
[12]	$\mathbf{P}^* \in \Pi_n \mid m > 1 \mid \sigma^2 > 0$	Linear program for $\mathbf{P}^*$, closed form solution for $\mathbf{X}^*$
[13]	$\mathbf{P}^* \in \Pi_{n,s} \mid m > 1 \mid \sigma^2 > 0$	Linear program for $\mathbf{P}^*$, convex program for $\mathbf{X}^*$
This paper	$\mathbf{P}^* \in \Pi_{n,r} \mid m > 1 \mid \sigma^2 > 0$	Graph matching via quadratic assignment problem
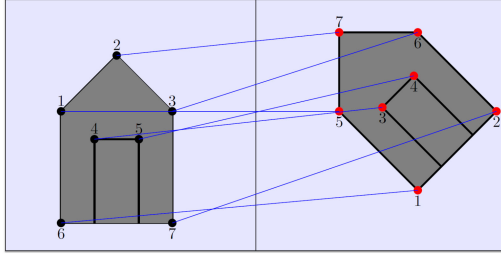
FIGURE 2. Left: An object with 7 keypoints, **Right:** The object after linear transformation (a rotation of 45 degrees) with 7 keypoints. In the pose and correspondence problem, given the keypoints corresponding to the left and right images, the task is to find the underlying permutation.

A. MOTIVATING APPLICATIONS

We outline two applications that can fall under the proposed model. Consider samples $y(i) = x(iT)$ of an analog signal $x(t)$. Jittered sampling [14], defined as $y(i) = x(iT + T_{\text{skew}})$, in interleaved ADC systems [15] scrambles the order of the r samples output at each cycle.

The second application is the pose and correspondence problem and is illustrated in Fig. 2.

B. RELATION TO EXISTING WORK

The problem of unlabeled sensing has received considerable attention. In addition to the fully shuffled case, several variants with structure on either $\mathbf{X}^*$ and/or the underlying permutation $\mathbf{P}^*$ have been studied. The single view case is a specific instance of the problem in (1), with $m = 1$, where the view matrix $\mathbf{Y}$ reduces to a single column vector. In Table 1, we summarize the various models, based on different structures on $\mathbf{X}^*$ and $\mathbf{P}^*$, and the corresponding algorithms. The rest of the literature survey is focused on the multi-view unlabeled sensing problem, the subject of the present paper.

The LEVSORT algorithm in [9] recovers the unknown permutation by aligning the leverage scores corresponding to the measurements matrix $\mathbf{Y}$ and sensing matrix $\mathbf{B}$. A drawback of this algorithm is its sensitivity to additive noise. In [10], another structure on $\mathbf{P}^*$ is considered for the multiview setting, namely that the number of shuffled elements is much smaller compared to n. By observing that the inliers lie in the span of a lower d-dimensional subspace of $\mathbb{R}^n$, the authors in [10] consider a sparse subspace clustering approach. For the same model proposed in [10], namely sparse $\mathbf{P}^*$, the works in [6], [13] proposed ℓ_1 constrained robust regression formulations.

In addition to assuming structure on the unknown permutation $\mathbf{P}^*$, previous techniques are applicable when dimension d is small [3], [4]. The results in [9]–[11] require number of views m equal to dimension d. By contrast, we consider recovery under a more challenging regime where n scales moderately with d, a small number of views with $m < d$ and additive measurement noise. Finally, the block diagonal permutation structure we consider has also been discussed in [11] but with the following additional assumptions on the model in (1): (a) The data matrix $\mathbf{X}^* \in \mathbb{R}^{d \times d}$ is orthonormal, and (b) The block diagonal permutation $\mathbf{P}^*$ is sparse. The discussion in Section II shows that the assumptions in [11] result in a simpler instance of the graph alignment problem compared to the problem considered in this paper. A related problem in graph signal processing (GSP) is the recovery of a band-limited graph signal [16], [17]. However, in the setup we consider here, there is no feature associated with nodes that will help us in graph matching. Since we are agnostic to what the nodes represent and the network topology, the GSP framework is not applicable here.

a) Major Contributions: Our main contributions can be summarized as follows.

1) For the unlabeled sensing problem, we present a novel graph matching formulation and propose an algorithm, R-local unlabeled sensing (RLUS), that creatively exploits the machinery of graph matching to resolve local permutations via multiple views.

IEEE
Signal
Processing
Society

IEEE Open Journal of
Signal Processing

2) We propose an initialization that utilizes the r-local structure based on collapsing the measurements. This principled initialization shows competitive performance in numerical simulations.

3) Motivated by several applications, we analyze a novel permutation model that we refer to as the *r*-local model (see Def. 1). A very restrictive version of the *r*-local model was considered in [11] and the setup that we consider here significantly expands upon the model therein.

4) One of the main conclusions of this paper is that, for the *r*-local model, multiple views significantly help in permutation recovery.

b) Outline: The rest of the paper is organized as follows: Section II gives background on graph matching and outlines our proposed approach. In Section III-A, our proposed algorithm is discussed. In Section IV, detailed simulation results are provided. We conclude in Section V with the main findings of the paper.

C. NOTATION

The following is a summary of the notation used in this paper. Let $b_i^{\mathsf{T}} \ \forall i \in [n]$ denote the rows (measurement vectors) of the matrix $\mathbf{B} \in \mathbb{R}^{n \times d}$, $\mathbf{B}^{\mathsf{T}} = [b_1^{\mathsf{T}} \,|, \ldots, |\, b_n^{\mathsf{T}}]$. Similarly $\mathbf{Y}^{\mathsf{T}} = [y_1^{\mathsf{T}} \,|, \ldots, |\, y_n^{\mathsf{T}}]$. Let $\mathbf{Y}_k \in \mathbb{R}^{r \times d} \ \forall k \in [n/r]$ denote blocks of $\mathbf{Y}$, $\mathbf{Y}^{\mathsf{T}} = [\mathbf{Y}_1^{\mathsf{T}} \,|\, \cdots \,|\, \mathbf{Y}_{n/r}^{\mathsf{T}}]$. Let $\mathbf{Y}_{P_k,:}$ denote the submatrix of $\mathbf{Y}_k$ comprising rows with indices in P_k. $\mathbb{1}_r \in \mathbb{R}^r$ denotes the *r*-dimensional vector of all ones. Π_n denotes the set of permutations in $\mathbb{R}^{n \times n}$. $\mathbb{D}_n \in \mathbb{R}^{n \times n}$ is the set of doubly stochastic matrices: $\mathbb{D}_n = \{\mathbf{P} \in \mathbb{R}^{n \times n} \mid \mathbf{P}\mathbb{1}_n = 1, \mathbf{P}^T \mathbb{1}_n = 1, P_{ij} \geq 0\}$. $\mathbf{I}$ denotes the identity matrix. The Moore-Penrose pseudoinverse of a matrix is denoted by $(\cdot)^{\dagger}$. $\| \cdot \|_F^2$ denotes the square of the Frobenius matrix norm defined as the sum of the squares of the elements of the matrix.

II. PROPOSED APPROACH

A central part of the proposed algorithm is based on theory and methods for graph matching, which we briefly discuss here. Given two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ each with *n* vertices, the graph matching problem is concerned with finding the permutation or map that best aligns them. If we choose the Frobenius norm as a metric of alignment, the general graph matching problem can be formulated as follows

$$\min_{\mathbf{P} \in \Pi_n} \quad ||\mathbf{A}\mathbf{P} - \mathbf{P}\mathbf{B}||_F^2, \tag{2}$$

where $\mathbf{A}$ and $\mathbf{B}$ are weighted adjacency matrices of G_1 and G_2 respectively. Using the fact that $\mathbf{P}$ is an orthonormal matrix, the objective can be expanded as follows: $||\mathbf{A}\mathbf{P} - \mathbf{P}\mathbf{B}||_F^2 = ||\mathbf{A}||_F^2 + ||\mathbf{B}||_F^2 - 2\,\text{trace}(\mathbf{A}\mathbf{P}\mathbf{B}^{\mathsf{T}}\mathbf{P}^{\mathsf{T}})$. With this, (2) can also be equivalently written as

$$\min_{\mathbf{P} \in \Pi_n} \quad - \text{trace}(\mathbf{A}\mathbf{P}\mathbf{B}^{\mathsf{T}}\mathbf{P}^{\mathsf{T}}) \tag{3}$$

This latter problem is the well known quadratic assignment problem (QAP) [18].

Since the minimization problem over the set of permutations is NP hard, a common relaxation of the graph matching programs in (2) and (3) is based on replacing the set of permutation matrices with the set of doubly stochastic matrices $\mathbb{D}_n$. Using the Birkhoff-Von Neumann theorem [19], [20], $\mathbb{D}_n$ is known to be the convex hull for the set of permutation matrices. While the objectives in (2) and (3) are equivalent over the set of permutation matrices, they are markedly different after relaxation. The relaxation of (2) is a convex program while the relaxation of (3) is typically indefinite and generally a nonconvex program. Here on, we refer to this relaxation as the indefinite relaxation. The theoretical analysis on correlated random Bernouilli graphs in [21] shows that the indefinite relaxation recovers the true permutation of exact graph matching problem with high probability while the convex program almost always fails. In addition, the indefinite relaxation is shown to be empirically successful for fast graph matching [22]. This relaxation is utilized in this paper.

A. COVARIANCE MATCHING: ORTHONORMAL X*

Given permuted observations $\mathbf{Y}$, the sensing matrix $\mathbf{B}$ and with $\mathbf{C}_Y$ and $\mathbf{C}_{\widehat{Y}}$ denoting covariance matrices, the unknown permutation is estimated as follows

$$\widehat{\mathbf{P}} = \underset{\widehat{\mathbf{Y}} \in \mathbb{R}^{n \times m}, \mathbf{P} \in \Pi_n}{\arg\min} \quad \|\mathbf{C}_Y \mathbf{P} - \mathbf{P}\mathbf{C}_{\widehat{Y}}\|_F^2, \tag{4}$$

where $\widehat{\mathbf{Y}} = \mathbf{B}\widehat{\mathbf{X}}$, $\mathbf{C}_{\widehat{Y}} = \widehat{\mathbf{Y}}\widehat{\mathbf{Y}}^{\mathsf{T}}$, $\mathbf{C}_Y = \mathbf{Y}\mathbf{Y}^{\mathsf{T}}$ and $\widehat{\mathbf{P}}$ is an estimate of $\mathbf{P}^*$. For a fixed $\widehat{\mathbf{Y}}$, the minimization problem in (4) corresponds to a QAP with similarity matrices $\mathbf{C}_Y$ and $\mathbf{C}_{\widehat{Y}}$. However, unlike (3) where the weighted adjacency matrices are explicitly given, $\mathbf{C}_{\widehat{Y}}$ in (4) has to be estimated as $\mathbf{X}^*$ is unknown. When $\mathbf{X}^*$ is an orthonormal matrix, as considered in [11], $\mathbf{C}_{\widehat{Y}} = \mathbf{B}\mathbf{B}^{\mathsf{T}}$ does not depend on $\mathbf{X}^*$. In this case, it is easy to see that the alignment in (4) reduces to:

$$\widehat{\mathbf{P}} = \underset{\mathbf{P} \in \Pi_n}{\arg\min} \quad \|\mathbf{C}_Y \mathbf{P} - \mathbf{P}\mathbf{C}_B\|_F^2, \tag{5}$$

with $\mathbf{C}_B = \mathbf{B}\mathbf{B}^{\mathsf{T}}$.

B. COVARIANCE MATCHING: GENERAL X*

For general $\mathbf{X}^*$, we could consider an alternating minimization between $\widehat{\mathbf{P}}$, $\widehat{\mathbf{X}}$. Since minimizing (4) is an NP-hard problem [18], we propose a two-stage algorithm to first estimate $\widehat{\mathbf{X}}$. We set $\widehat{\mathbf{Y}} = \mathbf{B}\widehat{\mathbf{X}}$ and let $\widehat{\mathbf{P}}$ be given via the indefinite relaxation of (4).

Stage-A: In stage-A of the algorithm, we employ alternating minimization over $\mathbf{X}$, Π_n on the following bi-convex objective

$$(\widehat{\mathbf{X}}, \widehat{\mathbf{P}}) = \underset{\mathbf{X}, \mathbf{P} \in \Pi_n}{\arg\min} \quad \|\mathbf{Y} - \mathbf{P}\mathbf{B}\mathbf{X}\|_F^2 \tag{6}$$

We leverage the *r*-local assumption on $\mathbf{P}^*$ for initializing $\widehat{\mathbf{X}}$ for the optimization in (6). This will be done via the collapsed linear system in (8) to be discussed later. Unlike the QAP in (4), the optimization over Π_n in (6) is a linear programming problem over $\mathbb{D}_n$. As outlined in Section III-A, the *r*-local assumption on $\mathbf{P}^*$ allows us to minimize over Π_n locally by
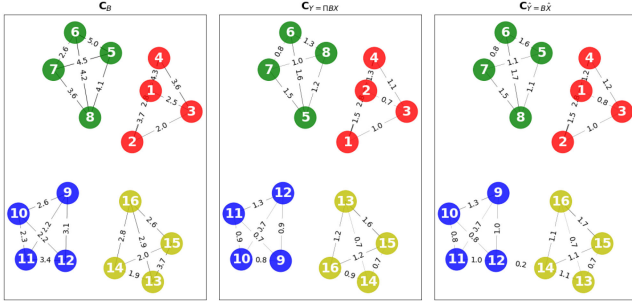
FIGURE 3. A block of r consecutive rows in matrices $\mathbf{B}, \mathbf{Y}, \widehat{\mathbf{Y}}$ corresponds to a cluster of r nodes in the corresponding graph. Left: The graph of known measurement matrix $\mathbf{B}$. Middle: The graph of given measurements $\mathbf{Y} = \mathbf{P}^*\mathbf{B}\mathbf{X}^*$. The r-local permutation, $\mathbf{P}^*$ distorts correspondence of nodes within each cluster, and the unknown linear transformation $\mathbf{X}^*$ distorts the edge weights. Right: The graph of estimate $\widehat{\mathbf{Y}} = \mathbf{B}\widehat{\mathbf{X}}$. The proposed approach first estimates $\widehat{\mathbf{Y}}$ and estimates $\mathbf{P}^*$ by matching $\mathbf{C}_{\widehat{Y}}$ to $\mathbf{C}_Y$.

using permutations of size r.

$$\widehat{\mathbf{P}}_k = \underset{\mathbf{P}_k \in \Pi_r}{\arg\min} \quad \|\mathbf{Y}_k - \mathbf{P}_k \mathbf{B}_k \widehat{\mathbf{X}}\|_F^2$$

$$= \underset{\mathbf{P}_k \in \Pi_r}{\arg\min} - \operatorname{trace}(\mathbf{B}_k \widehat{\mathbf{X}} \mathbf{Y}_k^\mathsf{T} \mathbf{P}_k),$$

$\forall k \in [n/r]$ with $\mathbf{B}_k$ and $\mathbf{Y}_k$ denoting the blocks of matrices $\mathbf{B}$ and $\mathbf{Y}$ respectively. [1]

Stage-B: Given estimate $\widehat{\mathbf{Y}} = \mathbf{B}\widehat{\mathbf{X}}$ from stage-A, we employ the indefinite relaxation of graph matching to estimate $\widehat{\mathbf{P}}$. This will be done via estimating the blocks of $\widehat{\mathbf{P}}$ as follows.

$$\widehat{\mathbf{P}}_k = \underset{\mathbf{P}_k \in \mathbb{D}_r}{\arg\min} \quad - \operatorname{trace}(\mathbf{C}_k \mathbf{P}_k \widehat{\mathbf{C}}_k^\mathsf{T} \mathbf{P}_k^\mathsf{T}), \tag{7}$$

$\forall k \in [n/r]$ with $\widehat{\mathbf{C}}_k = \widehat{\mathbf{Y}}_k \widehat{\mathbf{Y}}_k^\mathsf{T}$ and $\mathbf{C}_k = \mathbf{Y}_k \mathbf{Y}_k^\mathsf{T}$ denoting the block covariance matrices.

Fig. 3 is a visual summary of the proposed approach. With this perspective, in Section III-A, we propose and outline the algorithm in detail for the r-local permutation model. In Section IV, we provide simulation results, and focus particularly on quantifying permutation recovery via the proposed algorithm under an increasing number of views.

III. R-LOCAL UNLABELED SENSING (RLUS) ALGORITHM

RLUS has three main components: (a) an initialization via collapsing, (b) estimating $\widehat{\mathbf{Y}}$ and (c) estimating $\widehat{\mathbf{P}}$. This will be discussed in detail below.

A. INITIALIZATION

For our problem setup, we are given a sensing matrix $\mathbf{B}$, locally permuted observations $\mathbf{Y}$ and radius r. We note that, for $\mathbf{X}^* \in \mathbb{R}^{d \times m}$, each $r \times d$ block of measurement vectors and $r \times m$ block of permuted observations, $[\,\mathbf{P}_k^* \mathbf{B}_k \,\vert\, \mathbf{Y}_k\,] \; \forall \; k \in [n/r]$, is an r-dimensional unlabeled sensing problem. We can extract n/r labeled measurements on $\mathbf{X}^*$ by collapsing each

---

Algorithm 1.1: *Stage-A.*

Input: Radius $r \in \mathbb{R}$, $\mathbf{B} \in \mathbb{R}^{n \times d}$, $\mathbf{Y} \in \mathbb{R}^{n \times m}$
Output: $\widehat{\mathbf{Y}}, \widehat{\mathbf{P}}_{\text{LP}}$

1: $P \leftarrow \{1, \ldots, n\}$ $Q \leftarrow \{1, \ldots, n\}$ $K^{(0)} \leftarrow \{1, \ldots, n/r\}$
2: $\tilde{b}_i^{\mathsf{T}^{(0)}} \leftarrow \mathbb{1}_i^\mathsf{T} \mathbf{B}$ $\tilde{y}_i^{\mathsf{T}^{(0)}} \leftarrow \mathbb{1}_i^\mathsf{T} \mathbf{Y}$ $\forall i \in [n/r]$
3: $\widehat{\mathbf{Y}}^{(0)} \leftarrow \mathbf{B}(\widetilde{\mathbf{B}}^{(0)\dagger} \widetilde{\mathbf{Y}}^{(0)})$
4: **for** $t \in 0 \ldots d - n/r - 1$ **do**
5:      **for** $k \in K^{(t)}$ **do**
6:          $k_r \leftarrow \{(k-1)r + 1, \ldots, kr\}$
7:          $P_k \leftarrow k_r \cap P$ $Q_k \leftarrow k_r \cap Q$
8:          $\widehat{\mathbf{P}}_k = \underset{\mathbf{P} \in \Pi_{|P_k|}}{\arg\min} - \operatorname{trace}(\widehat{\mathbf{Y}}_{P_k,:}^{(t)} \mathbf{Y}_{Q_k,:}^\mathsf{T} \mathbf{P})$
9:          $P_k \leftarrow \widehat{\mathbf{P}}_k P_k$
10:     **end for**
11:     $(p^{*(t)}, q^{*(t)}) = \underset{p \in P_k, q \in Q_k}{\arg\min} \left\| \mathbf{Y} - \mathbf{B}([\frac{\widetilde{\mathbf{B}}^{(t)}}{b_p^\mathsf{T}}]^\dagger [\frac{\widetilde{\mathbf{Y}}^{(t)}}{y_q^\mathsf{T}}]) \right\|_F$
12:     $\widetilde{\mathbf{B}}^{(t+1)} \leftarrow [\frac{\widetilde{\mathbf{B}}^{(t)}}{b_{p^*}^\mathsf{T}}]$ $\widetilde{\mathbf{Y}}^{(t+1)} \leftarrow [\frac{\widetilde{\mathbf{Y}}^{(t)}}{y_{q^*}^\mathsf{T}}]$
13:     $\widehat{\mathbf{Y}}^{(t+1)} \leftarrow \mathbf{B}(\widetilde{\mathbf{B}}^{(t+1)\dagger} \widetilde{\mathbf{Y}}^{(t+1)})$
14:     $P \leftarrow P \setminus p^*$ $Q \leftarrow Q \setminus q^*$
15:     **for** $k \in K^{(t)}$ **do**
16:         $k_r \leftarrow \{(k-1)r + 1, \ldots, kr\}$
17:         **if**$(|k_r \cap P| = 1)$ $K^{(t+1)} \leftarrow K^{(t)} \setminus k$
18:     **end for**
19: **end for**
20: **for** $k \in 1, \ldots, n/r$ **do**
21:     $\widehat{\mathbf{P}}_k = \underset{\mathbf{P} \in \Pi_r}{\arg\min} - \operatorname{trace}(\widehat{\mathbf{Y}}_k^{(t)} \mathbf{Y}_k^\mathsf{T} \mathbf{P}_k)$
22: **end for**
23: $\widehat{\mathbf{P}}_{\text{LP}} \leftarrow \operatorname{diag}(\widehat{\mathbf{P}}_1, \ldots, \widehat{\mathbf{P}}_{n/r})$
24: $\widehat{\mathbf{Y}} \leftarrow \mathbf{B}(\mathbf{B}^\dagger \widehat{\mathbf{P}}_{\text{LP}}^\mathsf{T} \mathbf{Y})$

Algorithm 1.2 *Stage-B*

Input: Radius r, $\mathbf{B}, \mathbf{Y}, \widehat{\mathbf{Y}}, \widehat{\mathbf{P}}_{\text{LP}} = \operatorname{diag}(\bar{\mathbf{P}}_1, \ldots, \bar{\mathbf{P}}_{n/r})$
Output: $\widehat{\mathbf{P}}_r$

for $k \in 1 \ldots n/r$ **do**
    $\widehat{\mathbf{C}}_k \leftarrow \widehat{\mathbf{Y}}_k \widehat{\mathbf{Y}}_k^\mathsf{T}$ $\mathbf{C}_k \leftarrow \mathbf{Y}_k \mathbf{Y}_k^\mathsf{T}$
    $\widehat{\mathbf{P}}_k = \underset{\mathbf{P}_k \in \mathbb{D}_r}{\arg\min} - \operatorname{trace}(\mathbf{C}_k \mathbf{P}_k \widehat{\mathbf{C}}_k^\mathsf{T} \mathbf{P}_k^\mathsf{T})$ //*Initialize to* $\bar{\mathbf{P}}_k$
    $\widehat{\mathbf{P}}_k = \underset{\mathbf{P}_k \in \Pi_r}{\arg\min} - \operatorname{trace}(\widehat{\mathbf{P}}^\mathsf{T} \mathbf{P}_k)$
end for
$\widehat{\mathbf{P}} \leftarrow \operatorname{diag}(\widehat{\mathbf{P}}_1, \ldots, \widehat{\mathbf{P}}_{n/r})$

---

block as follows

$$\mathbf{P}_1^* \mathbf{B}_1 \mathbf{X}^* = \mathbf{Y}_1 \qquad \Rightarrow \qquad \underbrace{\mathbb{1}_r^\mathsf{T} \mathbf{P}_1^* \mathbf{B}_1}_{\tilde{b}_1^\mathsf{T}} \mathbf{X}^* = \underbrace{\mathbb{1}_r^\mathsf{T} \mathbf{Y}_1}_{\tilde{y}_1^\mathsf{T}}$$

$$\mathbf{P}_2^* \mathbf{B}_2 \mathbf{X}^* = \mathbf{Y}_2 \qquad \Rightarrow \qquad \underbrace{\mathbb{1}_r^\mathsf{T} \mathbf{P}_2^* \mathbf{B}_2}_{\tilde{b}_2^\mathsf{T}} \mathbf{X}^* = \underbrace{\mathbb{1}_r^\mathsf{T} \mathbf{Y}_2}_{\tilde{y}_2^\mathsf{T}}$$

$$\vdots \qquad\qquad \vdots$$

$$\mathbf{P}^*_{n/r}\mathbf{B}_{n/r}\mathbf{X}^* = \mathbf{Y}_{n/r} \quad \Rightarrow \quad \underbrace{\mathbb{1}_r^{\mathsf{T}}\mathbf{P}^*_{n/r}\mathbf{B}_{n/r}}_{\tilde{b}_{n/r}^{\mathsf{T}}}\mathbf{X}^* = \underbrace{\mathbb{1}_r^{\mathsf{T}}\mathbf{Y}_{n/r}}_{\tilde{y}_{n/r}^{\mathsf{T}}}$$

to form the *collapsed* system

$$\left[\,\widetilde{\mathbf{B}}\,\big|\,\widetilde{\mathbf{Y}}\,\right], \tag{8}$$

$\widetilde{\mathbf{B}} \in \mathbb{R}^{n/r \times d}$, $\widetilde{\mathbf{Y}} \in \mathbb{R}^{n/r \times m}$. The permuted linear system of equations (L.S.E.) $[\mathbf{P}^*\mathbf{B}\,|\,\mathbf{Y}]$ has $n > d$ permuted measurements, whereas the collapsed L.S.E. $[\widetilde{\mathbf{B}}\,|\,\widetilde{\mathbf{Y}}]$ is underdetermined with n/r measurements. We consider recovery of $\mathbf{P}^*$ in the case $d > n/r$.

B. STAGE-A

Stage-A of the algorithm estimates $\widehat{\mathbf{Y}}$ by alternating minimization of the following objective:

$$\widehat{\mathbf{X}}, \widehat{\mathbf{P}}_k = \underset{\mathbf{X}, \mathbf{P}_k \in \Pi_r}{\arg\min} \quad \|\mathbf{Y}_k - \mathbf{P}_k \mathbf{B}_k \mathbf{X}\|_F^2, \tag{9}$$

$\forall k \in [n/r]$. The minimization is initialized by setting $\widehat{\mathbf{X}}^{(0)}$ to the least-squares solution of the collapsed linear system of equations in (8) with $\widehat{\mathbf{X}}^{(0)} = \widetilde{\mathbf{B}}^{\dagger}\widetilde{\mathbf{Y}}$. The update for $\widehat{\mathbf{P}}^{(t+1)} = \text{diag}(\widehat{\mathbf{P}}_1, \ldots, \widehat{\mathbf{P}}_{n/r})$ is given by the solution to n/r linear programs:

$$\widehat{\mathbf{P}}_k = \underset{\mathbf{P}_k \in \Pi_r}{\arg\min} \quad -\text{trace}(\widehat{\mathbf{Y}}_k^{(t)}\mathbf{Y}_k^{\mathsf{T}}\mathbf{P}_k),$$

$\forall k \in [n/r]$, and $\widehat{\mathbf{Y}}_k^{(t)} = \mathbf{B}_k \widehat{\mathbf{X}}^{(t)}$.

The update for $\widehat{\mathbf{X}}$ is based on augmenting the collapsed L.S.E. $[\widetilde{\mathbf{B}}\,|\,\widetilde{\mathbf{Y}}]$ to full rank. At each iteration t, a measurement vector (row of matrix $\mathbf{B}$), b_{p*}^{T}, and a measurement (row of matrix $\mathbf{Y}$), y_{q*}^{T}, are appended to matrices $\widetilde{\mathbf{B}}^{(t)} \in \mathbb{R}^{n/r+t}$ and $\widetilde{\mathbf{Y}}^{(t)} \in \mathbb{R}^{n/r+t}$ respectively (see Line 12 of Algorithm 1.1). The indices $(p^{*(t)}, q^{*(t)})$ are selected (Line 11) such that the following forward error is minimized.

$$(p^{*(t)}, q^{*(t)}) = \underset{p \in P_k, q \in Q_k}{\arg\min} \left\| \mathbf{Y} - \mathbf{B}\left(\begin{bmatrix} \widetilde{\mathbf{B}}^{(t)} \\ b_p^{\mathsf{T}} \end{bmatrix}^{\dagger} \begin{bmatrix} \widetilde{\mathbf{Y}}^{(t)} \\ y_q^{\mathsf{T}} \end{bmatrix} \right) \right\|_F, \tag{10}$$

where $P_k(Q_k)$ are sets containing indices of unmatched measurement vectors (measurements) in block k. The indices $p^{*(t)}, q^{*(t)}$ selected at iteration t are excluded from subsequent iterations so that the minimization in (10) is over $n - t$ pairs of indices (Line 14 of Algorithm 1.1). Since a collapsed block by its construction depends on the rows of the block in consideration, if $r - 1$ indices of a block have been matched based on the above forward error, we exclude the single remaining index in that block from subsequent iterations (Lines 15–18). This step is optional but it promotes the augmented system to have full row-rank.

To augment the LSE in (8) to be full-rank, we set the number of iterations in Algorithm 1.1 equal to $d - n/r$. Algorithm 1.1 outputs $\widehat{\mathbf{Y}}$ and $\widehat{\mathbf{P}}_{\text{LP}}$.

C. STAGE-B

The output $\widehat{\mathbf{Y}}$ from Algorithm 1.1 is used to build the covariance matrix $\widehat{\mathbf{C}}_k = \widehat{\mathbf{Y}}_k \widehat{\mathbf{Y}}_k^{\mathsf{T}}$ for the graph matching stage in Algorithm 1.2. The output $\widehat{\mathbf{P}}_{\text{LP}}$ from Algorithm 1.1 is used to initialize the indefinite relaxation of the covariance matching program:

$$\widehat{\mathbf{P}}_k = \underset{\mathbf{P}_k \in \mathbb{D}_r}{\arg\min} \quad -\text{trace}(\mathbf{C}_k \mathbf{P}_k \widehat{\mathbf{C}}_k^{\mathsf{T}} \mathbf{P}_k^{\mathsf{T}}), \tag{11}$$

$\forall\, k \in [n/r]$, $\widehat{\mathbf{C}}_k = \widehat{\mathbf{Y}}_k \widehat{\mathbf{Y}}_k^{\mathsf{T}}$ and $\mathbf{C}_k = \mathbf{Y}_k \mathbf{Y}_k^{\mathsf{T}}$. Since the unknown permutation is assumed to be r-local, the optimization in (11) is over the set of $r \times r$ doubly stochastic matrices. We use the Frank-Wolfe algorithm [23] to minimize the objective in (11). A detailed description of the FW algorithm is given in [22]. The doubly stochastic solution to (11), $\widehat{\mathbf{P}}_k$, is projected onto the set of $r \times r$ permutation matrices (Line 4 of Algorithm 1.2).

A natural question is the exactness of the indefinite relaxation of the local covariance matchings in stage-B. We will argue below that the indefinite relaxation is exact under some conditions on the covariance matrices. A similar result appears in [24] where the authors utilize graph Laplacian matrices for the indefinite graph matching problem. Let the local covariance matrices $\mathbf{C}_k$ and $\widehat{\mathbf{C}}_k$ be positive definite. Assume that there exists a permutation $\bar{\mathbf{P}}$ such that $\mathbf{C}_k\bar{\mathbf{P}} = \bar{\mathbf{P}}\widehat{\mathbf{C}}_k$. Then, we claim that

$$\underset{\mathbf{P} \in \mathbb{D}_r}{\arg\max} \quad \text{trace}(\mathbf{C}_k \mathbf{P}\widehat{\mathbf{C}}_k^{\mathsf{T}}\mathbf{P}^{\mathsf{T}}) = \bar{\mathbf{P}}.$$

We prove this result as follows. Let $f(\mathbf{P})$ denote the objective. We first show that $f(\mathbf{P})$ is strictly convex on $\mathbb{D}$. A direct computation yields that,

$$\nabla^2 f(\mathbf{P}) = 2\,\widehat{\mathbf{C}}_k \otimes \mathbf{C}_k.$$

Using the fact that the covariances matrices are positive definite and the properties of the Kronecker product, all eigenvalues of the Hessian are positive. Therefore, $f(\mathbf{P})$ is strictly convex. Under the assumption that there exists a unique solution $\bar{\mathbf{P}}$ to the graph matching problem, we have

$$\bar{\mathbf{P}} = \underset{\mathbf{P} \in \Pi_r}{\arg\min} \, ||\mathbf{C}_k\mathbf{P} - \mathbf{P}\,\widehat{\mathbf{C}}_k||_F^2 = \underset{\mathbf{P} \in \Pi_r}{\arg\max} \, \text{trace}(\mathbf{P}^T\mathbf{C}_k^T\mathbf{P}\widehat{\mathbf{C}}_k)$$

A convex function over a compact convex set attains a maximizer at an extreme point [25]. It then follows that

$$\underset{\mathbf{P} \in \mathbb{D}_r}{\arg\max} \, f(\mathbf{P}) = \underset{\mathbf{P} \in \Pi_r}{\arg\max} \quad \text{trace}(\mathbf{P}^T\mathbf{C}_k\mathbf{P}\widehat{\mathbf{C}}_k) = \bar{\mathbf{P}}$$

The implication of the aforementioned result is that the indefinite relaxation obtains the underlying solution to the graph matching problem under structural assumptions on the covariance matrices. We conjecture that if $m > r$, $\widehat{\mathbf{C}}_k, \mathbf{C}_k$ in (11) are positive definite with high probability implying that the assumptions of the theorem hold true with the same high probability. While we have observed this empirically, a detailed analysis is left for future investigation.

D. COMPUTATIONAL COMPLEXITY

In this section, we estimate the worst-case complexity of RLUS.

Stage-A: Stage-A has at most d iterations. The cost of the linear program per iteration is $O(r^3)$. Minimizing the forward error is based on the pseudoinverse of an augmented matrix that varies in each iteration. For a streamlined analysis, we assume the worst-case augmentation of $\tilde{\mathbf{B}}$ denoted by $\bar{\mathbf{B}} \in \mathbb{R}^{d \times d}$. To compute the product $\bar{\mathbf{B}}^\dagger \mathbf{Z}$, for an $d \times m$ matrix $\mathbf{Z}$, we use the Cholesky decomposition of $\bar{\mathbf{B}}^\dagger$ followed by backward and forward substitutions. The total cost of these operations (Line 11) is $O(d^3) + O(md^2)$. The matrix product with $\mathbf{B} \in \mathbb{R}^{n \times d}$ and computing the Frobenius norm costs $O(nm) + O(ndm)$. Finally, the least squares problem to obtain $\widehat{\mathbf{Y}}$ (Line 24) is $O(n^3)$. With this, the worst-case complexity of stage-A is $O(d^4) + O(md^3) + O(nmd) + O(nd^2m) + O(n^3d)$.

Stage-B: Stage-B only requires forming covariance matrices and graph matching via indefinite relaxations. The cost of these operations is $O(r^3) + O(mr^2)$. The dominating cost of $O(r^3)$ is the cost of the Hungarian algorithm which projects a given matrix onto the set of doubly stochastic matrices during the Frank Wolfe updates.

Total cost: The total worst-case complexity of the proposed algorithm is quartic with respect to d and cubic with respect to n. In the regime of interest, $d \ll n$, the complexity in d is less than cubic complexity in n. We remark here that the computational complexity estimate here is loose and can be sharpened using the relation of the Cholesky factorization of an augmented matrix with the original matrix [26].

IV. RESULTS

In this section, we show the performance of RLUS, implemented in MATLAB, on synthetic data sets. All of our numerical experiments are evaluated with respect to the ground truth permutation using the fractional Hamming distortion.

Synthetic data: For the model in (1), we draw entries of the measurement matrix $\mathbf{B} \in \mathbb{R}^{n \times d}$ and the signal $\mathbf{X}^* \in \mathbb{R}^{d \times m}$ i.i.d from the standard normal distribution. The additive noise $\mathbf{N}$ is drawn i.i.d from Gaussian distribution $\mathcal{N}(0, \sigma^2)$, and the permutation $\mathbf{P}^*$ is picked uniformly at random from the set $\Pi_{n,r}$. Following other works in the unlabeled sensing literature [10], [12], [13], the signal to noise ratio is defined as

$$\text{SNR} = \frac{\|\mathbf{X}^*\|_F^2}{m\sigma^2}. \qquad (12)$$

All results are averaged over 25 Monte Carlo runs.

Baselines: We compare the proposed approach to 4 baseline algorithms:

i) Subspace matching: The *levsort* algorithm in [9] attempts to recover the unknown permutation by matching the subspace spanned by $\mathbf{U}_B$, the left singular vectors of the measurement matrix $\mathbf{B}$, to $\mathbf{U}_Y$. For $m = d$ views in the noiseless setting, the *levsort* algorithm recovers the unknown permutation exactly.

ii) Alt-min: We compare the proposed approach against an alternating minimization approach for the biconvex objective

in (6). The updates on $\widehat{\mathbf{P}}, \widehat{\mathbf{X}}$ are given as:

$$\widehat{\mathbf{X}}^{(t)} = \mathbf{B}^\dagger (\widehat{\mathbf{P}}^{(t)\mathsf{T}} \mathbf{Y}) \qquad (\text{alt} - \min)$$

$$\widehat{\mathbf{P}}_k^{(t+1)} = \underset{\mathbf{P}_k \in \Pi_r}{\arg\min} \quad - \text{trace}(\mathbf{B}\widehat{\mathbf{X}}^{(t)}\mathbf{Y}\mathbf{P}_k)$$

$\forall k \in [n/r]$, and $\widehat{\mathbf{P}}_k^{(0)} = \mathbf{I}_{r \times r}$, $\widehat{\mathbf{P}} = \text{diag}(\widehat{\mathbf{P}}_1, \ldots, \widehat{\mathbf{P}}_{n/r})$.

iii) Covariance matching via GW alignment: We compare the QAP based graph matching in Algorithm 1.2 to the Gromov-Wasserstein alignment. Given two graphs with adjacency matrices $\mathbf{A}$ and $\mathbf{B}$, the Gromov Wasserstein alignment consists in minimizing the square of the Gromov-Wasserstein distance [27], $\text{GW}_2^2(\mathbf{A}, \mathbf{B})$, between them:

$$\text{GW}^2(\mathbf{A}, \mathbf{B}) = \min_{\mathbf{P} \in \mathbb{D}_n} \sum_{p,'q'} \sum_{p,q} \left(\mathbf{B}_{p'q'} - \mathbf{A}_{pq}\right)^2 \mathbf{P}_{p'p} \mathbf{P}_{q'q} \quad (13)$$

The GW alignment problem in (13) is a non-convex quadratic optimization problem. We use the algorithm proposed in [28] to solve (13) with entropic regularization:

$$\text{GW}_\epsilon^2(\mathbf{A}, \mathbf{B}) = \text{GW}^2(\mathbf{A}, \mathbf{B}) - \epsilon H(\mathbf{P}),$$

where entropy $H(\mathbf{P}) = -\sum P_{pq}(\log(P_{pq}) - 1)$, and ϵ is the amount of entropic regularization. A higher value of ϵ (high entropic regularization) returns a diffused/soft assignment between the weighted adjacency matrices $\mathbf{A}, \mathbf{B}$. In contrast, lower values of ϵ return more precise assignments.

iv) Covariance matching for orthonormal $\mathbf{X}^*$: For an orthonormal signal matrix, $\mathbf{X}^*(\mathbf{X}^*)^\mathsf{T} = \mathbf{I}$, as considered in [11], we recover $\widehat{\mathbf{P}}$ by minimizing the objective in (5).

A. COMPARISONS WITH BASELINES

For all comparisons in Fig. 4, we set the block size $r = 50$, number of measurements $n = 600$, views $m = d = 25$, and plot the fractional Hamming distortion d_H/n against SNR.

Comparison with levsort: In Fig. 4(a), we compare our algorithm to the subspace matching algorithm *levsort* in [9]. We observe that the algorithm in [9] is not robust to noise as the distortion in the estimate returned by *levsort* (green) is high even for diminishing levels of noise power. This is in contrast to the proposed algorithm which returns $d_H(\widehat{\mathbf{P}}, \mathbf{P}^*) = 0$ for SNR = 1000.

Comparison with alt-min: In Fig. 4(b), we observe that the alternating minimization approach in (9) fails to recover the unknown permutation $\mathbf{P}^*$. This is because, unlike the models considered in [10], [12], [13], the r-local model in this work does not assume partially shuffled measurements. In particular, the initialization $\widehat{\mathbf{P}}^{(0)} = \mathbf{I}_{r \times r}$ is suitable only when the measurements are partially shuffled.

Comparison with GW alignment: In Fig. 4(c), we observe that the proposed QAP alignment outperforms GW, especially in the low SNR regime. For both QAP and GW, we let $\mathbf{A} = \mathbf{Y}_k \mathbf{Y}_k^\mathsf{T}$ and $\mathbf{B} = \widehat{\mathbf{Y}}_k \widehat{\mathbf{Y}}_k^\mathsf{T}$. We set the amount of entropic regularization for GW to the lowest value that avoided numerical instability in the experiments.

$\mathbf{X}^*(\mathbf{X}^*)^\mathsf{T} = \mathbf{I}$: In Fig. 4(d), we observe that, for orthonormal $\mathbf{X}^*$ (blue), the unknown permutation $\mathbf{P}^*$ can
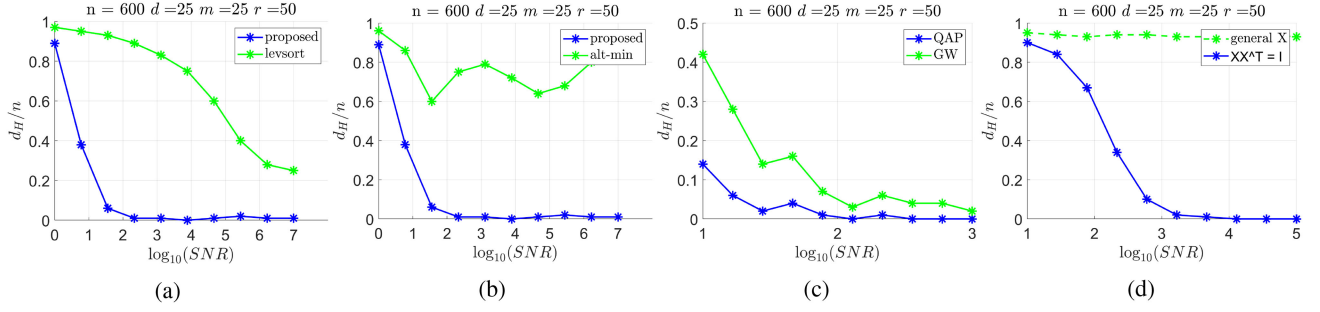
FIGURE 4. $\mathbf{Y} = \mathbf{P}^*\mathbf{B}_{n \times d}\mathbf{X}^*_{d \times m} + \mathbf{N}$. Figure plots the fractional Hamming distortion, $d_H(\mathbf{P}^*, \widehat{\mathbf{P}})/n$, against SNR.

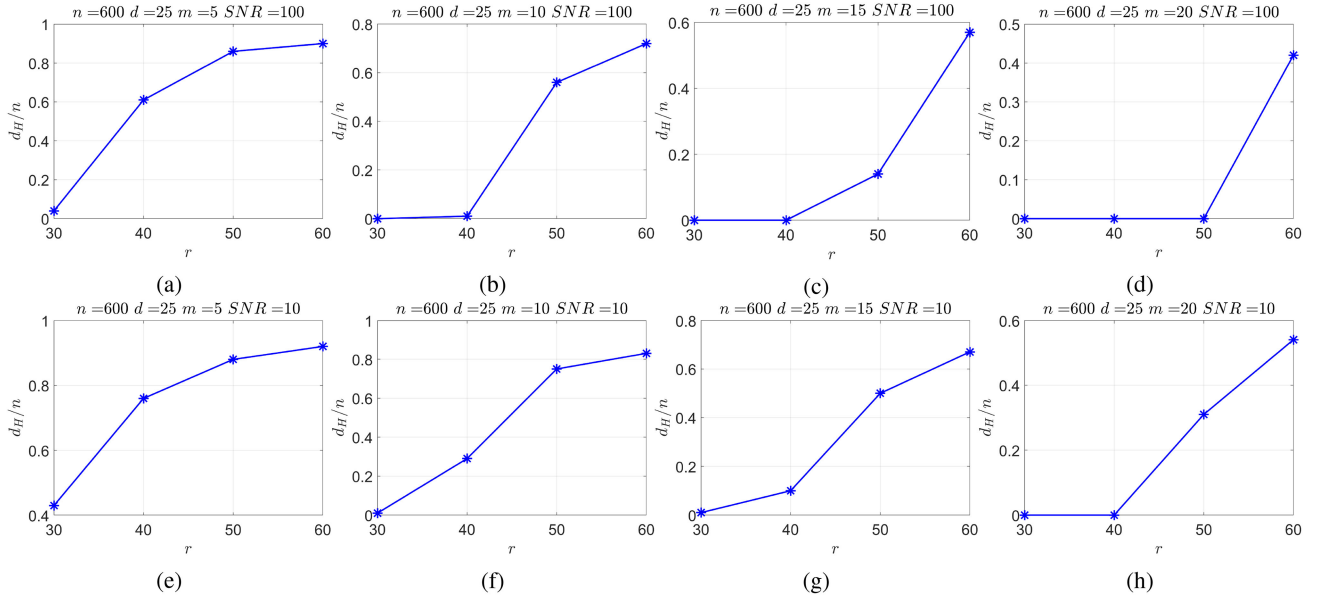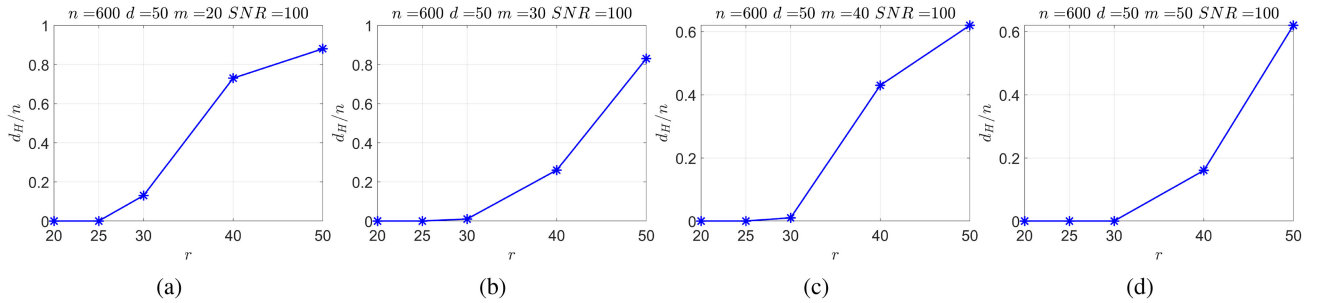**FIGURE 5.** $\mathbf{Y} = \mathbf{P}^*\mathbf{B}_{n \times d}\mathbf{X}^*_{d \times m} + \mathbf{N}, d = 25$. Figure plots the fractional Hamming distortion, $d_H(\mathbf{P}^*, \widehat{\mathbf{P}})/n$, against block size r for SNR $\in$ {10 100}.

FIGURE 6. $\mathbf{Y} = \mathbf{P}^*\mathbf{B}_{n \times d}\mathbf{X}^*_{d \times m} + \mathbf{N}, d = 50$. Figure plots the fractional Hamming distortion, $d_H(\mathbf{P}^*, \widehat{\mathbf{P}})/n$, against block size r for SNR = 100.

be recovered by aligning the covariance blocks $\mathbf{B}_k\mathbf{B}_k^\mathsf{T}$ to $\mathbf{Y}_k\mathbf{Y}_k^\mathsf{T}$. However, this approach fails for general $\mathbf{X}^*$ (green). The results corroborate the claim in Section II-A that, for orthonormal $\mathbf{X}^*$ considered in [11], the permutation can be recovered without estimating the unknown signal $\mathbf{X}^*$.

B. OTHER SIMULATIONS

In Figs. 5, 6, we plot the fractional Hamming distortion $d_H(\widehat{\mathbf{P}}, \mathbf{P}^*)/n$ against block size r.

For the simulations in Fig. 5, we set signal dimension $d = 25$. For $m \in \{5, 10, 15, 25\}$, we plot $d_H(\widehat{\mathbf{P}}, \mathbf{P}^*)/n$ against $r \in \{30, 40, 50, 60\}$. We observe the following:

i) Hamming distortion in the returned estimate is lower for small values of r and increases with increasing r. This is because for fixed n, as r increases, the collapsed system in (8) becomes increasingly underdetermined. Secondly, the size of each local permutation, $r!$, increases with increasing r.

ii) Permutation recovery is significantly improved with increasing views. For example, for $r = 50$, $m = 5$ in Fig. 5(a), the Hamming distortion is nearly 1. In contrast, for $m = 20$ views in Fig. 5(d), the Hamming distortion is 0. As shown in [12], the necessary SNR required for recovering $\mathbf{P}^*$ reduces from $\mathrm{SNR} = \Omega(n^2)$ for the single-view case to $\mathrm{SNR} \simeq n^{1/\rho(\mathbf{X})}$, for the multi-view setting. [2] The stable rank $\rho(\mathbf{X})$ of matrix $\mathbf{X}$ is defined as the ratio $\sum \sigma^2(\mathbf{X})/\sigma_{\max}^2(\mathbf{X})$, where $\sigma(\mathbf{X})$ are the singular values of $\mathbf{X}$. For $\mathbf{X} \in \mathbb{R}^{d \times m}$, the stable rank is upper bounded by $\rho(\mathbf{X}) \leq \mathrm{rank}(\mathbf{X}) = \min(m, d)$. Increasing the number of views m increases the stable rank of signal matrix $\mathbf{X}^*$ and reduces the necessary SNR for permutation recovery. In Fig. 5(e)–(h), we decrease SNR from $\mathrm{SNR} = 100$ in Fig. 5(a)–(d) to $\mathrm{SNR} = 10$.

For the simulations in Fig. 6, we set dimension $d = 50$, $\mathrm{SNR} = 100$, and plot the Hamming distortion against $r \in \{20, 25, 30, 40, 50\}$. We can draw similar conclusions as from Fig. 5. The distortion in the estimated permutation increases with increasing r and decreases with increasing m. However, for the same block size r, the distortion in $\widehat{\mathbf{P}}$ is higher for $d = 50$ than for $d = 25$. The collapsed system in (8) used to initialize our algorithm is underdetermined by $d - n/r$ measurements. A higher value of d corresponds to a poor initialization of the non-convex optimization problem in Stage-A. Since the permutation estimate $\widehat{\mathbf{P}}$ in stage-B is obtained from aligning $\widehat{\mathbf{C}}_k$ to $\mathbf{C}_k$, the distortion in this estimate is higher when $\widehat{\mathbf{C}}_k$ is not estimated well.

V. CONCLUSION

We study a linear inverse problem with scrambled measurements well known as the unlabeled sensing problem. In this paper, we focus on a novel setup of this problem, the r-local model, that considers local permutations, and, in contrast to existing work, does not assume the measurements to be partially shuffled. We propose a novel algorithm based on insights from graph matching and recent advances in alignment via the quadratic assignment problem. Numerical experiments on synthetic data show that the proposed algorithm is robust to additive noise and exploits multiple views for improved recovery. As part of future work, we will explore the theoretical performance bounds for the r-local model.

REFERENCES

[1] J. Unnikrishnan, S. Haghighatshoar, and M. Vetterli, "Unlabeled sensing with random linear measurements," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3237–3253, 2018.

[2] A. Abid and J. Zou, "A. stochastic expectation-maximization approach to shuffled linear regression," in *Proc. 56th Annu. Allerton Conf. Commun., Control, Comput.*, 2018, pp. 470–477.

[3] M. Tsakiris and L. Peng, "Homomorphic sensing," in *Proc. the 36th Int. Conf. Mach. Learn., Proc. Mach. Learn. Res.*, Long Beach, CA, USA, Jun. 2019, pp. 6335–6344.

[4] M. C. Tsakiris, L. Peng, A. Conca, L. Kneip, Y. Shi, and H. Choi, "An algebraic-geometric approach for linear regression without correspondences," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 5130–5144, Aug. 2020.

[5] A. Pananjady, M. J. Wainwright, and T. A. Courtade, "Linear regression with shuffled data: Statistical and computational limits of permutation recovery," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3286–3300, May 2018.

[6] M. Slawski and E. Ben-David, "Linear regression with sparsely permuted data," *Electron. J. Statist.*, vol. 13, no. 1, pp. 1–36, 2019.

[7] D. J. Hsu, K. Shi, and X. Sun, "Linear regression without correspondence," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1531–1540.

[8] V. Emiya, A. Bonnefoy, L. Daudet, and R. Gribonval, "Compressed sensing with unknown sensor permutation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2014, pp. 1040–1044.

[9] A. Pananjady, M. J. Wainwright, and T. A. Courtade, "Denoising linear models with permuted data," in *Proc. IEEE Int. Symp. Inf. Theory*, 2017, pp. 446–450.

[10] M. Slawski, M. Rahmani, and P. Li, "A sparse representation-based approach to linear regression with partially shuffled labels," *Uncertainty Artif. Intell.*, 2020, pp. 38–48.

[11] X. Shi, X. Li, and T. Cai, "Spherical regression under mismatch corruption with application to automated knowledge translation," *J. Amer. Stat. Assoc.*, pp. 1–12, 2020.

[12] H. Zhang and P. Li, "Optimal estimator for unlabeled linear regression," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, pp. 11153–11162.

[13] M. Slawski, E. Ben-David, and P. Li, "A two-stage approach to multivariate linear regression with sparsely mismatched data," *J. Mach. Learn Res.*, vol. 21, no. 204, pp. 1–42, 2020.

[14] A. Balakrishnan, "On the problem of time jitter in sampling," *IRE Trans. Inf. Theory*, vol. 8, no. 3, pp. 226–236, 1962.

[15] B. Razavi, "Design considerations for interleaved ADCs," *IEEE J. Solid-State Circuits*, vol. 48, no. 8, pp. 1806–1817, Aug. 2013.

[16] P. Di Lorenzo, S. Barbarossa, and P. Banelli, "Sampling and recovery of graph signals," in *Cooperative Graph Signal Process.* Amsterdam, The Netherlands: Elsevier, 2018, pp. 261–282.

[17] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4845–4860, Sep. 2016.

[18] E. M. Loiola, N. M. M. de Abreu, P. O. Boaventura-Netto, P. Hahn, and T. Querido, "A survey for the quadratic assignment problem," *Eur. J. Oper. Res.*, vol. 176, no. 2, pp. 657–690, 2007.

[19] G. Birkhoff, "Three observations on linear algebra," *Univ. Nac. Tacuman, Rev. Ser. A*, vol. 5, pp. 147–151, 1946.

[20] J. Von Neumann, "A certain zero-sum two-person game equivalent to the optimal assignment problem," *Contributions Theory Games*, vol. 2, pp. 5–12, 1953.

[21] V. Lyzinski *et al.*, "Graph matching: Relax at your own risk," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 60–73, Jan. 2016.

[22] J. T. Vogelstein *et al.*, "Fast approximate quadratic programming for graph matching," *PLOS One*, vol. 10, no. 4, 2015, Art. no. e0121002.

[23] M. Jaggi, "Revisiting frank-wolfe: Projection-free sparse convex optimization," in *Proc. 30th Int. Conf. Mach. Learn. Res.*, Atlanta, Georgia, USA, Jun 2013, pp. 427–435.

[24] M. Zaslavskiy, F. Bach, and J.-P. Vert, "A. path following algorithm for the graph matching problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2227–2242, Dec. 2009.

[25] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*. Hoboken, Wiley, 2013.

[26] P. E. Gill, G. H. Golub, W. Murray, and M. A. Saunders, "Methods for modifying matrix factorizations," *Math. Comput.*, vol. 28, no. 126, pp. 505–535, 1974.

[27] F. Mémoli, "Gromov-wasserstein distances and the metric approach to object matching," *Found. Comput. Math.*, vol. 11, no. 4, pp. 417–487, 2011.

[28] G. Peyré, M. Cuturi, and J. Solomon, "Gromov-wasserstein averaging of kernel and distance matrices," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 2664–2672.

[2]Eq. (7) in [12]

AHMED ALI ABBASI received the B.E. degree in electrical engineering from the National University of Sciences and Technology, Islamabad, Pakistan. He is currently working toward the Ph.D. degree in electrical engineering as a Member of Professor Shuchin Aeron's Statistical Inference Learning and Computation Group with Tufts University, Medford, MA, USA. His research interests include optimal transport, dictionary learning, unassigned distance geometry, and applications of machine learning. He was awarded the Greenough Fellowship in 2018. He was also one of three inaugural recipients of the Tufts Tripods Fellowship in 2019.

ABIY TASISSA received the B.Sc. degree in mathematics and the M.Sc. degree in aeronautics and astronautics from the Massachusetts Institute of Technology, Cambridge, MA, USA, and the Ph.D. degree in applied mathematics from Rensselaer Polytechnic Institute, Troy, NY, USA. He is currently a Norbert Wiener Assistant Professor with the Department of Mathematics, Tufts University, Medford, MA, USA and a Research Associate with CRISP Group, Harvard University, Cambridge, MA, USA. His research interests include distance geometry, compressive sensing, matrix completion, graph matching, and more recently structured deep learning and manifold learning.

SHUCHIN AERON received the B.Tech. degree in electrical engineering from the Indian Institute of Technology Kanpur, Kanpur, India, in 2002, and the M.S. and Ph.D. degrees in electrical engineering from Boston University, Boston, MA, USA, in 2004 and 2009, respectively. He is currently an Associate Professor with the Department of ECE, Tufts University, Medford, MA, USA. From 2009 to 2011, he was a Postdoctoral Fellow with Schlumberger Doll Research and has a number of patents with applications to acoustic logging while drilling. His area of expertise and research interests include information theory, mathematical statistics, tensor data analytics, compressed sensing, statistical machine learning, and optimal transport. He was awarded the Best Ph.D. Thesis Award from the School of Engineering and Department of Electrical and Computer Engineering. He was the recipient of the NSF CAREER Award in 2016. He is currently an Associate Editor for the IEEE Transactions on Geosciences and Remote Sensing and is on the technical committee of Machine Learning for Signal Processing, IEEE Signal Processing Society. He is also the Co-Director of the Data Science programs at the School of Engineering, Tufts University.