r-local Unlabeled Sensing: Improved algorithm and applications

Ahmed Ali Abbasi[1], Abiy Tasissa[2], and Shuchin Aeron[1]

[1] Department of ECE, Tufts University, Medford, MA 02155
[2]Department of Mathematics, Tufts University, Medford, MA 02155

Abstract

The unlabeled sensing problem is to solve a noisy linear system of equations under unknown permutation of the measurements. We study a particular case of the problem where the permutations are restricted to be r-local, i.e. the permutation matrix is block diagonal with $r \times r$ blocks. Assuming a Gaussian measurement matrix, we argue that the r-local permutation model is more challenging compared to a recent sparse permutation model. We propose a proximal alternating minimization algorithm for the general unlabeled sensing problem that provably converges to a first order stationary point. Applied to the r-local model, we show that the resulting algorithm is efficient. We validate the algorithm on synthetic and real datasets. We also formulate the 1-d unassigned distance geometry problem as an unlabeled sensing problem with a structured measurement matrix.

1 Introduction

The standard least squares problem is to recover signal $\mathbf{X} \in \mathbb{R}^{d\times m}$ given measurements $\mathbf{Y} \in \mathbb{R}^{n\times m}$, and a measurement matrix $\mathbf{B}$. The signal matrix $\mathbf{X}$ is estimated by solving the least squares minimization problem $\min_{\mathbf{X}} \|\mathbf{Y}-\mathbf{B}\mathbf{X}\|_F^2$. This formulation assumes that there is one to one correspondence between measurements (the rows of $\mathbf{Y}$) and the measurement vectors (the rows of $\mathbf{B}$) i.e. we know which measurement corresponds to what measurement vector. However, in many applied problems of interest such as header free communication for mobile wireless networks [16], sampling in the presence of clock jitter [4], linear regression without correspondence [3] and point cloud registration [12], this mapping is not explicitly given. This motivates the study of the unlabeled sensing problem with the prototypical form given by

$$\mathbf{Y} = \mathbf{P}\mathbf{B}\mathbf{X} + \mathbf{W}, \tag{1}$$

where $\mathbf{P} \in \mathbb{R}^{n\times n}$ denotes the unknown permutation matrix and $\mathbf{W} \in \mathbb{R}^{n\times m}$ denotes an additive Gaussian noise with $\mathbf{W}_{ij} = \mathcal{N}(0, \sigma^2)$. Given this set-up, the unlabeled sensing problem is to estimate $\mathbf{X}$ and $\mathbf{P}$ given the measurement matrix $\mathbf{B}$ and measurements $\mathbf{Y}$.

1.1 Related work

The unlabeled sensing problem was first considered with a Gaussian measurement matrix $\mathbf{B}$ in the single-view setup , where $m = 1$, in [20]. There, the authors showed that $n = 2d$ noiseless measurements are sufficient for recovery of $\mathbf{x}$. The same result was subsequently proven in [8] using an algebraic argument. For the case of Gaussian $\mathbf{B}$, the work in [15] showed that with SNR $\triangleq \|\mathbf{X}\|_F^2/m\sigma^2$, SNR $= \Omega(n^2)$ is necessary for recovering $\mathbf{P}$ exactly. For the multi-view setup, where $m > 1$, the result in [23] shows that the necessary SNR for recovery decreases as m increases, with necessary SNR $= O(1)$ for $m = \Omega(\log n)$.

Several algorithms have been proposed for the single view unlabeled sensing problem [16, 15, 17]. The multi-view problem setup has also been considered in several works [23, 18, 14, 22, 1]. As the maximum likelihood estimator for a general permutation $\mathbf{P}$ is NP-hard [14], the works in [23, 18, 22] impose a k-sparse structure on $\mathbf{P}$ (see Figure 1a) where k elements of the permutation are on the diagonal. In a recent work [1], the authors introduced the r-local model (see section 2) along with an algorithm that is based on graph-matching. For the problem in (1), the r-local model imposes the following structure on $\mathbf{P}$. A permutation matrix of size $n \times n$ is r-local if it is composed of n/r blocks. Formally, $\mathbf{P} = \text{diag}\left(\mathbf{P}_1, \cdots, \mathbf{P}_{n/r}\right)$, where $\mathbf{P}_i \in \Pi_r$ denotes an $r \times r$ permutation matrix (see Figure 1b).

1.2 Contributions

(a) We compare the r-local model with the k-sparse model and argue that the r-local problem is more challenging under the most widely studied case of Gaussian measurement matrix. (b) We show for the first time that a specific case of another class of unlabeled problems, namely the 1-d unassigned distance geometry problem (uDGP) [11] is an unlabeled sensing problem but with a structured measurement matrix. Further, we discuss the setting in which the r-local model can be applied to this problem. (c) We propose a proximal alternating minimization algorithm for the r-local unlabeled sensing problem. Simulation results show that our algorithm outperforms the proposed algorithms in [23, 14, 1].

2 r-local model vs k-sparse model

A permutation matrix is k-sparse if k of its elements are on the diagonal with $\langle \mathbf{I}, \mathbf{P}\rangle = k$. Figures 1a and 1b show a k-sparse and an r-local permutation matrix respectively. We compare the two models via the difficulty of recovering $\mathbf{P}$ under each model when the matrix $\mathbf{B}$ is assumed to be random Gaussian. Under this assumption, for any permutation $\mathbf{P}$, and estimator $\widehat{\mathbf{P}}$, the probability of error, $\Pr\{\widehat{\mathbf{P}} \neq \mathbf{P}\}$, is lower bounded by the result in [23]

$$\Pr\{\widehat{\mathbf{P}} \neq \mathbf{P}\} \geq 1 - \frac{1 + \frac{nm}{2}\log_2(1+\text{SNR})}{H(\mathbf{P})}, \tag{2}$$

where $H(\mathbf{P})$ denotes the entropy of the random permutation matrix $\mathbf{P}$. For $\mathbf{P} \in \mathbb{R}^{n\times n}$, $H_r(\mathbf{P}) = \log_2 r!^{n/r}$ for r-local permutations and $H_k(\mathbf{P}) = \log_2(n!/k!)$ for k-sparse permutations. Performance guarantees for proposed algorithms are given for $k = 7n/8$ in [23] and $k = 3n/4$ in [22]. All simulation results for permutation recovery in [18, 19] are for

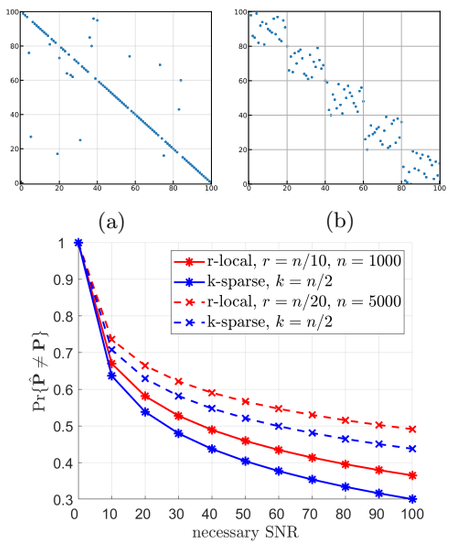


Figure 1: (a). A sparse (or partially shuffled) permutation conisdered in [23, 18, 22, 19]. (b). The r-local permutation structure considered in this paper, with $r = 20$. (c). Figure plots $\Pr\{\widehat{\mathbf{P}} \neq \mathbf{P}\}$ against SNR for the 2 models, see (2).

Figure 2: The 1-d unassigned distance geometry problem (uDGP) is to recover the point coordinates $(0, 3, 7, 9)$ from their unlabeled pairwise distances $\{3, 2, 4, 6, 7, 9\}$. uDGP can be formulated as an unlabeled sensing problem with a deterministic measurement matrix $\mathbf{B}_u$ (see (4))

.

$k \geq n/2$. Figure 1c shows that, for the same SNR, the lower bound in (2) is higher for the r-local model, $r = \{n/10, n/20\}$, than for the k-sparse model with $k = n/2$. This suggests that permutation recovery under the r-local model is a challenging problem, especially in the large n regime.

3 1-d Unassigned Distance Geometry as an unlabeled sensing problem

The one-dimensional unassigned distance geometry problem (1-d uDGP) (see Figure 2) is to recover point coordinates from their unlabeled pairwise distances [11, 9, 6]. The distances are unlabeled as the pair (i, j) corresponding to the distance $|x_i - x_j|$ is not known. Let $\overline{\mathbf{x}} \in \mathbb{R}^d = [\overline{x}_1, \cdots, \overline{x}_d]^\intercal$ be the vector of unknown coordinates. Without loss of generality, we assume that the unknown point coordinates are sorted in decreasing order. It follows that each pairwise distance $\overline{x}_i - \overline{x}_j$ is a linear measurement of $\overline{\mathbf{x}}$. As the distances are translation invariant, we can also assume that the left most point is at the origin. The distance vector $\mathbf{y} \in \mathbb{R}^{d(d-1)/2}$ is the matrix-vector product $\mathbf{y} = \mathbf{B}_u\mathbf{x}$ where $\mathbf{x} = [x_1, \cdots, x_{d-1}] \in \mathbb{R}^{d-1}$. It can be verified that $\mathbf{B}_u$ has the following form.

$$\mathbf{B}_u = [\mathbf{I} \mid \mathbf{B}_2 \mid \cdots \mid \mathbf{B}_{d-1}]^\intercal, \tag{3}$$

$$[\mathbf{B}_i \in \mathbb{R}^{(d-i)\times(d-1)}]_{pq} = \begin{cases} 1 & \text{if } q = i-1 \\ -1 & \text{if } q = p+i-1 \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

where $p \leq d - i$. The unlabeled sensing formulation for the example in Figure 2 is given below.

$$\underbrace{\begin{pmatrix} 3 \\ 2 \\ 4 \\ 6 \\ 7 \\ 9 \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} 0&0&1&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&0&1 \\ 0&0&0&0&1&0 \\ 0&1&0&0&0&0 \\ 1&0&0&0&0&0 \end{pmatrix}}_{\mathbf{P}} \underbrace{\begin{pmatrix} 1&0&0 \\ 0&1&0 \\ 0&0&1 \\ 1&-1&0 \\ 1&0&-1 \\ 0&1&-1 \end{pmatrix}}_{\mathbf{B}_u} \underbrace{\begin{pmatrix} 9 \\ 7 \\ 3 \end{pmatrix}}_{\mathbf{x}}. \tag{5}$$

We note the following facts that follow from (5): (1) The point coordinates $[0 \mid \mathbf{x}] \in \mathbb{R}^d$ is equal to $\bar{\mathbf{x}}$ up to a translation by 2 units. (2) The matrix $\mathbf{B}_u$ has full column rank, as is typical for i.i.d. Gaussian measurement matrices in unlabeled sensing and (3) The matrix $\mathbf{B}_u$ has block structure, with the first block comprising the first 3 rows corresponding to the identity matrix, the next 2 rows forming the second block, and the last block composed of the final row. Next, the condition number of the matrix $\mathbf{B}_u$ is given in Proposition 1.

Proposition 1. *The condition number of the measurement matrix* $\mathbf{B}_u \in \mathbb{R}^{d(d-1)/2\times d-1}$ *is* $\kappa(\mathbf{B}_u) = \sqrt{d}$.

The proof is provided in section 7

Why r-local model for uDGP? An r-local structure on $\mathbf{P}$ has a practical interpretation for the uDGP problem. The problem where, in addition to the distances $|x_i - x_j|$, one of the two corresponding indices (i, j) is known, is modelled by an r-local $\mathbf{P} = \text{blkdiag}(d-1, \cdots, 2)$. The r-local structure also models the problem where distance assignments are known up to a cluster of points. For example, for two clusters, each distance is known to be between a pair of points in cluster 1 or cluster 2 (intra-cluster) or a point in cluster 1 and a point in cluster 2 (inter-cluster) but the pair of points corresponding to each distance is still unknown.

4 Proposed approach and algorithm

In this section, we present a new algorithm for the r-local unlabeled sensing problem. To motivate our algorithm, we note that for an i.i.d Gaussian noise $\mathbf{W}$ in (1), the maximum likelihood estimate for $\mathbf{P}$ is

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

where Π_n denotes the set of permutations of order n. The optimization problem in (6) is non-convex since the set of permutations is discrete. A natural optimization scheme for the problem in (6) is the alternating minimization algorithm which alternates between updating the signal matrix $\mathbf{X}$ and the permutation $\mathbf{P}$. The works in [18, 14, 22, 19] propose one step estimators for $\mathbf{X}, \mathbf{P}$. We consider complete alternating minimization for (6) yielding the following updates

$$\mathbf{P}^{(t)} = \arg\min_{\mathbf{P}\in\Pi_n} \langle -\mathbf{Y}(\mathbf{X}^{(t)})^\intercal\mathbf{B}^\intercal, \mathbf{P}\rangle \tag{7}$$

$$\mathbf{X}^{(t+1)} = \arg\min_{\mathbf{X}} F(\mathbf{X}, \mathbf{P}^{(t)}) = (\mathbf{P}^{(t)}\mathbf{B})^\dagger\mathbf{Y}. \tag{8}$$

The iterates $(\mathbf{P}^{(t)}, \mathbf{X}^{(t)})$ in (7), (8) are monotonically decreasing $F(\mathbf{P}^{(t+1)}, \mathbf{X}^{(t+1)}) \leq F(\mathbf{P}^{(t)}, \mathbf{X}^{(t)})$. However, this is not sufficient to establish convergence of the iterates. Specifically, we can not claim that $\lim_{t\to\infty}\{(\mathbf{P}^{(t)}, \mathbf{X}^{(t)})\}$ exists. We propose to use proximal alternating minimization (PAM) [2] instead, which as we show in section 4.1, converges to a first

order stationary point of the objective in (6). PAM adds a regularization term that encourages the new iterate to be close to the current iterate. Formally, for the objective in (6), the PAM updates are given by

$$\mathbf{P}^{(t+1)} = \underset{\mathbf{P}\in\Pi_n}{\arg\min}\ F(\mathbf{X}^{(t)}, \mathbf{P}) + \lambda^{(t)}\|\mathbf{P} - \mathbf{P}^{(t)}\|_F^2 \quad (9)$$

$$\mathbf{X}^{(t+1)} = \underset{\mathbf{X}}{\arg\min}\ F(\mathbf{X}, \mathbf{P}^{(t+1)}) + \lambda^{(t)}\|\mathbf{X} - \mathbf{X}^{(t)}\|_F^2, \quad (10)$$

where $\lambda_t > 0$. Our algorithm follows from the updates in (9) and (10). Similar to (7) and (8), the updates are a linear program and a regularized least squares problem respectively. Assuming an r-local structure on $\mathbf{P}$, the update in (9) is equivalent to the simpler update of n/r blocks of size r. A summary of our algorithm is given in Algorithm 1.

Algorithmic details: We use the *collapsed* initialization, as also done in [1], to initialize $\widehat{\mathbf{Y}}^{(0)} = \mathbf{B}\widehat{\mathbf{X}}^{(0)}$ in our algorithm. The initialization estimates $\mathbf{X}$ from n/r measurements given by summing r consecutive rows in $\mathbf{B}$ and $\mathbf{Y}$. The convergence criteria is based on the relative change in the objective $(F^{(t-1)} - F^{(t)})/F^{(t-1)}$. Each iteration of Algorithm 1 has two main steps. The total cost of updating all the local permutations via linear programs is $O(nr^2)$. The regularized least squares problem can be solved in $O(2nd^2 - \frac{2}{3}d^3)$.

Algorithm 1 Proximal Alt-Min for r-local unlabeled sensing

Require: $\mathbf{B}, \mathbf{Y}, \widehat{\mathbf{Y}}^{(0)}, \widehat{\mathbf{X}}^{(0)} r.\lambda, \epsilon$
Ensure: $\widehat{\mathbf{P}}$
$\widehat{\mathbf{P}}_k^{(0)} \leftarrow \mathbf{1}_{r\times r}/r$
while relative change $> \epsilon$ **do**
  for $k \in 1\cdots n/r$ **do**
    $\widehat{\mathbf{P}}_k = \underset{\mathbf{P}_k\in\Pi_r}{\arg\min} -\langle \mathbf{Y}\widehat{\mathbf{Y}}^{(t)} + \lambda\widehat{\mathbf{P}}_k^{(t)}, \mathbf{P}_k\rangle$
  end for
  $\widehat{\mathbf{P}}^{(t+1)} \leftarrow \mathrm{diag}(\widehat{\mathbf{P}}_1, \cdots, \widehat{\mathbf{P}}_{n/r})$
  $\widehat{\mathbf{X}}^{(t+1)} \leftarrow \left[\frac{\mathbf{B}_{n\times d}}{\sqrt{\lambda}\mathbf{I}_{d\times d}}\right]^{\dagger}\left[\frac{\widehat{\mathbf{P}}^{(t+1)\intercal}\mathbf{Y}}{\sqrt{\lambda}\widehat{\mathbf{X}}^{(t)}}\right]$
  $\widehat{\mathbf{Y}}^{(t+1)} \leftarrow \mathbf{B}\widehat{\mathbf{X}}^{(t+1)}$
  $t \leftarrow t+1$
end while

4.1 Convergence analysis

Proposition 2. *The sequence of iterates* $\{(\mathbf{X}^{(t)}, \mathbf{P}^{(t)})\}$ *generated by* (9), (10) *converges to a first order stationary point of the objective.*

Proof. The result follows from Theorem 9 in [2]. To use Theorem 9, we need to verify that the Kurdyka-Lojasiewicz (KL) inequality holds for the objective $F(\mathbf{X}, \mathbf{P}) = \|\mathbf{Y} - \mathbf{P}\mathbf{B}\mathbf{X}\|_F^2 + \mathbf{I}_C(\mathbf{P})$ where $\mathbf{I}_C(\mathbf{P})$ is the indicator function of the set of permutations. First, note that the set of permutations of order n is semi-algebraic because each permutation can be specified via n^2 polynomial (linear) constraints. The indicator function of a semi-algebraic function is a semi-algebraic function [2]. The first term in the objective is polynomial, and therefore real-analytic. The sum of a real-analytic function and a semi-algebraic function is sub-analytic, and sub-analytic functions satisfy the KL inequality [21].

5 Experiments

For all experiments using Algorithm 1, we set the tolerance $\epsilon = 10^{-3}$ and the proximal regularization is set to $\lambda^{(1)} = 100$ and scaled as $\lambda^{(t+1)} = \lambda^{(t)}/10$.

5.1 Synthetic simulations

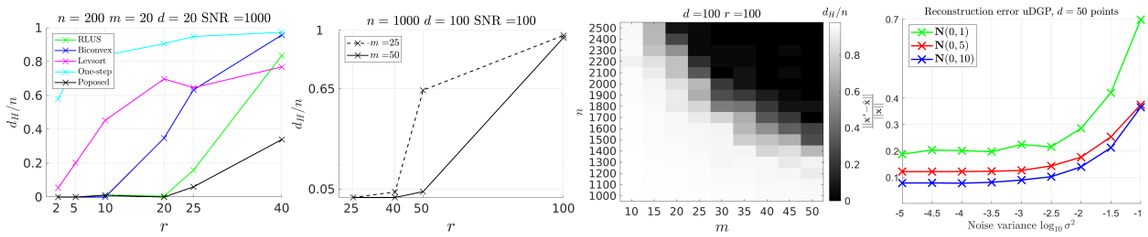
Data generation: The entries of the signal matrix $\mathbf{X}$, the sensing matrix $\mathbf{B}$, and the noise $\mathbf{W}$ are sampled i.i.d. from the normal distribution. The matrix $\mathbf{W}$ is subsequently scaled by σ to set a specific SNR. The permutation $\mathbf{P}$ is picked uniformly at random from the set of r-local permutations. All results are averages of 75 Monte Carlo runs.

Baselines: We compare our algorithm to four algorithms proposed in [23, 14, 22, 1]. First is a biconvex relaxation of (6) proposed in [23] and solved via the alternating direction method of multipliers (ADMM) algorithm. Second is the Levsort algorithm in [14]. For $m = d$ views and $\sigma^2 = 0$, the Levsort algorithm provably recovers $\mathbf{P}$ exactly. The third algorithm we compare to is the estimator in [22] which approximates $\mathbf{X} \simeq \mathbf{B}^{\intercal}\mathbf{Y}$ in (7) and solves $\min_{\mathbf{P}\in\Pi}\langle \mathbf{Y}\mathbf{Y}^{\intercal}\mathbf{B}\mathbf{B}^{\intercal}, \mathbf{P}\rangle$. The approximation is justified when matrix $\mathbf{B}$ is Gaussian i.i.d. and the rows of $\mathbf{B}$ are partially shuffled. In order to ensure a fair comparison with the algorithms in [23, 14, 22], we specify an additional constraint on the optimization problem to constrain the permutation solutions to be r-local. Finally, we compare to the algorithm in [1] that solves a quadratic assignment problem to estimate $\mathbf{P}$. The ADMM penalty parameter ρ for [23] was tuned in the range 10^{-3} to 10^5 for each Monte Carlo run and the $\widehat{\mathbf{P}}$ with the lowest distortion was retained. We refer to the algorithms in [23, 14, 22, 1] as 'Biconvex', 'Levsort, 'One-step', 'RLUS' respectively.

Results: In Figure 3a, we compare our proposed algorithm to the baselines. The results show that the Levsort algorithm is sensitive to noise as the algorithm fails to recover $\mathbf{P}$ even for $m = d$ views. Biconvex [23] only recovers $\mathbf{P}$ for $r \leq 10$. One-step [22] fails to recover $\mathbf{P}$ for even small values of r, underscoring that algorithms to recover sparse permutations are not suited to the r-local model. RLUS [1] gives comparable recovery of $\mathbf{P}$ when $r \leq 20$ but is outperformed by the proposed algorithm at higher values of r. In Figure 3b, we set $(d, n) = (100, 1000)$ and observe that the algorithm recovers $\mathbf{P}$ for $r \leq 50$. In Figure 3c, we set $(d, r) = (100, 100)$ and vary (n, m). We observe that the Hamming distortion decreases as n increases. This is because as n increases, the initialization to the algorithm via n/r measurements is less under-determined (see section 4). We note that for all experiments the estimate $\widehat{\mathbf{P}}$ improves as the value of m increases. This observation agrees with the result in (2) and the discussion in section 1.

5.2 Scrambled image restoration

In this experiment, we consider a variation of the problem in [13]. Therein, given scrambled images $\mathbf{P}\mathbf{y}$ and unscrambled training data $\mathbf{y}$, a neural network is trained to unscramble $\mathbf{P}\mathbf{y}$. For our experiment, the images are drawn from the YALE B dataset [10] and the MNIST dataset [7]. For each class in the MNIST (YALE B) dataset, the matrix $\mathbf{B} \in \mathbb{R}^{n\times d}$ is constituted from the first $d = 5$ (10) principal components of the unscrambled data. A class in the MNIST



(a) (b) (c) (d)

Figure 3: **Synthetic simulations** for $\mathbf{Y} = \mathbf{P}\mathbf{B}_{n\times d}\mathbf{X}_{d\times m} + \mathbf{W}$. (a.b). Figures plot the fractional Hamming distortion d_H/n against block size r. The Hamming distortion d_H/n is the fraction of mismatches in estimate $\widehat{\mathbf{P}}$ of $\mathbf{P}$. (c). Figure plots d_H/n against (m, n) for fixed (d, r). (d). **uDGP**: Figure plots the relative error in the estimated point coordinates. The $d = 50$ points are sampled i.i.d. from the normal distribution with variances $\{1, 5, 10\}$. The permutation $\mathbf{P}$ is r-local with blocks of sizes $(d-1, \cdots, 2)$.

	Proposed	RLUS
MNIST	$\mathbf{12.31} \pm 2.57$	12.03 ± 2.47
YALE	$\mathbf{28.30} \pm 2.87$	26.43 ± 2.95

Table 1: For each dataset, the PSNR values (dB) are averaged over 10 classes.

(a) PSNR = 27.04 dB (b) PSNR = 11.79 dB

Figure 4: (a). Left: Unscrambled image $\mathbf{y} \in \mathbb{R}^n$, $n = 48\times 42$ from the YALE B dataset. Middle. Scrambled input image $\mathbf{P}\mathbf{y}$ via an r-local permutation, $r = 96$. Right. Reconstructed image $\hat{\mathbf{y}} = \widehat{\mathbf{P}}^\intercal\mathbf{P}\mathbf{y}$. (b). Left: Unscrambled image $\mathbf{y} \in \mathbb{R}^n$, $n = 28 \times 28$ from the MNIST dataset. Middle: Scrambled input image via an r-local permutation, $r = 28$. Right: Reconstructed image. The peak signal to noise ratio PSNR is defined in terms of the mean square reconstruction error $e = \frac{1}{n}\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2$ as $\text{PSNR} = 10\log_{10}(\frac{1}{e^2})$.

(YALE B) dataset comprises all images corresponding to a single digit (particular face). The problem is to recover the unknown permutation $\mathbf{P}$ given the scrambled image $\mathbf{P}\mathbf{y}$ and the matrix $\mathbf{B}$. The setup is summarized in Figure 4. We compare our algorithm to RLUS [1] and the results are given in Table 1. The results show that the proposed algorithm outperforms RLUS [1].

5.3 1-d uDGP

We consider the 1-dimensional uDGP problem (Section 3) where the distances are corrupted with i.i.d. Gaussian noise of variance σ^2. The results in Figure 3d show that the relative error in the recovered points is low even for high noise variance. It can be shown that the condition number of the sensing matrix $\mathbf{B}_u$ is small, $\kappa_2(\mathbf{B}_u) = \sqrt{d}$. Numerical stability of the proposed formulation is important because noisy uDGP is NP-hard [5].

6 Conclusion

We propose a proximal alternating minimization (PAM) algorithm for the unlabeled sensing problem and apply it to the setting where the unknown permutation is r-local. The resulting algorithm is efficient and theoretically converges to a first order stationary point. Experiments on synthetic and real data show that the algorithm outperforms competing baselines. We formulate the 1-d unassigned distance geometry problem as an unlabeled sensing problem with a structured measurement matrix. In future work, we would like to explore information-theoretic inachievability results for the r-local model.

7 Appendix

Let us recall the definition of the measurement matrix $\mathbf{B}_u$ below.

$$\mathbf{B}_u = [\mathbf{I} \mid \mathbf{B}_2 \mid \cdots \mid \mathbf{B}_{d-1}]^\intercal,$$
$$[\mathbf{B}_i \in \mathbb{R}^{(d-i)\times(d-1)}]_{pq} = \begin{cases} 1 & \text{if } q = i-1 \\ -1 & \text{if } q = p+i-1 \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 1. *The condition number of the measurement matrix* $\mathbf{B}_u \in \mathbb{R}^{d(d-1)/2\times d-1}$ *in* (4) *is* $\kappa(\mathbf{B}_u) = \sqrt{d}$.

The matrix $\mathbf{B}_u$ in (4) has block structure. Specifically, $\mathbf{B}_u^\intercal = [\mathbf{I} \mid \mathbf{B}_2 \mid \mathbf{B}_3 \mid \cdots \mathbf{B}_{d-1}]$, where $\mathbf{I}$ is the $(d-1)\times(d-1)$ identity matrix, and $\mathbf{B}_i$ is an $d-i \times d-1$ matrix as follows. The matrix $\mathbf{B}_u^\intercal\mathbf{B}_u$ can be written in terms of its blocks

$$\mathbf{B}_u^\intercal\mathbf{B}_u = \sum_{i=2}^{i=d-1} \mathbf{B}_i^\intercal\mathbf{B}_i + \mathbf{I} \tag{11}$$

Next we expres $\mathbf{B}_u^\intercal\mathbf{B}_u$. Let us define the following matrix $\mathbf{E}_{pq}$ for block i.

$$\mathbf{E}_{pq} \in \mathbb{R}^{(d-i)\times(d-1)} \begin{cases} 1 & \text{if } p = q \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Each block $\mathbf{B}_i$ of the matrix $\mathbf{B}_u$ can be represented in terms of the matrices $\mathbf{E}$.

$$\mathbf{B}_i = \sum_{p=1}^{p=d-i} \mathbf{E}_{p,i-1} - \mathbf{E}_{p,i-1+p} \tag{13}$$

To calculate $\mathbf{B}_i^\intercal\mathbf{B}_i$, we first note the following

$$\mathbf{E}_{pq}^\intercal\mathbf{E}_{rs} \begin{cases} \mathbf{E}_{qs} & \text{if } p = r \\ 0_{d-i\times d-i} & \text{otherwise.} \end{cases} \tag{14}$$

Now, we can write $\mathbf{B}_i^\intercal\mathbf{B}_i$ in terms of $\mathbf{E}$

$$\begin{aligned}
\mathbf{B}_i^\intercal\mathbf{B}_i &= (\sum_{p=1}^{p=d-i} \mathbf{E}_{p,i-1} - \mathbf{E}_{p,i-1+p})^\intercal(\sum_{p=1}^{p=d-i} \mathbf{E}_{p,i-1} - \mathbf{E}_{p,i-1+p}) \\
&= \sum_{p=1}^{p=d-i} \mathbf{E}_{p,i-1}^\intercal\mathbf{E}_{p,i-1} - \mathbf{E}_{p,i-1}^\intercal\mathbf{E}_{p,i-1+p} - \mathbf{E}_{p,i-1+p}^\intercal\mathbf{E}_{p,i-1} + \mathbf{E}_{p,i-1+p}^\intercal\mathbf{E}_{p,i-1+p} \quad (\mathbf{E}_{p1,\cdot}^\intercal.\mathbf{E}_{p2,\cdot} = 0 \text{ if } p1 \neq p2\ ) \\
&= \sum_{p=1}^{p=d-i} \mathbf{E}_{i-1,i-1} - \mathbf{E}_{i-1,i-1+p} - \mathbf{E}_{i-1+p,i-1} + \mathbf{E}_{i-1+p,i-1+p} \qquad (\ (14)\ ) \\
&= (d-i)\mathbf{E}_{i-1,i-1} + \sum_{p=1}^{p=d-i} \mathbf{E}_{i-1+p,i-1+p} + \sum_{p=1}^{p=d-i} -\mathbf{E}_{i-1,i-1+p} - \mathbf{E}_{i-1+p,i-1}
\end{aligned}$$

Diagonal terms. The first 2 terms in the summation above, $(d-i)\mathbf{E}_{i-1,i-1} + \sum_{p=1}^{p=d-i} \mathbf{E}_{i-1+p,i-1+p}$, are the diagonal elements of $\mathbf{B}_i^\intercal\mathbf{B}_i$. Summing over the $d-2$ blocks, see (11), the diagonal elements of the matrix $\sum_{i=2}^{i=d-2} \mathbf{B}_i^\intercal\mathbf{B}_i$ are the sum of the following matrices.

$$\begin{aligned}
&\sum_{i=2}^{i=d-1} (d-i)\mathbf{E}_{i-1,i-1} + \sum_{i=2}^{i=d-1}\sum_{p=1}^{p=d-i} \mathbf{E}_{i-1+p,i-1+p} \\
&= \sum_{i=2}^{i=d-1} (d-i)\mathbf{E}_{i-1,i-1} + \sum_{i=2}^{i=d-1}\sum_{p=i}^{p=d-1} \mathbf{E}_{p,p}
\end{aligned}$$

The diagonal terms are equal to $d-2$. This can be seen from the above by realizing that the t^{th} diagonal entry, for $t \in [1, \cdots, d-1]$, is the sum of t matrices. For example,

$$\begin{aligned}
&t = 1 && (d-2)\mathbf{E}_{1,1} \\
&t = 2 && (d-3)\mathbf{E}_{2,2} + 1\mathbf{E}_{2,2} \\
&t = 3 && (d-4)\mathbf{E}_{2,2} + 2\mathbf{E}_{2,2} \\
&\vdots && \vdots \\
&t = d-1 && 0 + (d-2)\mathbf{E}_{d-1,d-1}
\end{aligned}$$

The diagonal terms of the matrix $\sum_{i=2}^{i=d-2} \mathbf{B}_i^\intercal\mathbf{B}_i$ are therefore equal to $d-2$, and the diagonal elements of $\mathbf{B}_u^\intercal\mathbf{B}_u = \sum_{i=2}^{i=d-2} \mathbf{B}_i^\intercal\mathbf{B}_i + \mathbf{I}^\intercal\mathbf{I}$ are therefore $d-1$.

Off-diagonal terms. The off diagonal terms are the following.

$$\sum_{i=2}^{i=d-1} \sum_{p=1}^{p=d-i} -\mathbf{E}_{i-1,i-1+p} - \mathbf{E}_{i-1+p,i-1}$$

.

$$= \sum_{i=2}^{i=d-1} \sum_{p=i}^{p=d-1} -\mathbf{E}_{i-1,p} - \mathbf{E}_{p,i-1}$$

Each off-diagonal entry is the sum of one matrix, and therefore equal to -1. The matrix $\mathbf{B}_u^\intercal \mathbf{B}_u \in \mathbb{R}^{n-1 \times n-1}$,

$$[\mathbf{B}_u^\intercal \mathbf{B}_u]_{pq} = \begin{cases} d-1 & \text{if } p = q \\ -1 & \text{otherwise.} \end{cases} \tag{15}$$

The matrix $\mathbf{B}_u^\intercal \mathbf{B}_u$ can also be written as

$$\mathbf{B}_u^\intercal \mathbf{B}_u = -\mathbf{1}_{d-1 \times d-1} + d\mathbf{I} \tag{16}$$

The eigenvalues of the matrix $-\mathbf{1}_{d-1 \times d-1}$ are $d-1$ and 0. The eigenvalues of matrix $-\mathbf{1}_{d-1 \times d-1} + d\mathbf{I}$ are 1 and d. From this we conclude that

$$\sigma^2_{\max}(\mathbf{B}_u) = d \quad \sigma^2_{\min}(\mathbf{B}_u) = 1 \tag{17}$$

and therefore the condition number squared $\kappa_2^2(\mathbf{B}_u) = d$.

References

[1] Ahmed Ali Abbasi, Abiy Tasissa, and Shuchin Aeron. R-local unlabeled sensing: A novel graph matching approach for multiview unlabeled sensing under local permutations. *IEEE Open Journal of Signal Process.*, 2:309–317, 2021.

[2] Hédy Attouch, Jérôme Bolte, Patrick Redont, and Antoine Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010.

[3] Zhidong Bai and Tailen Hsing. The broken sample problem. *Probability theory and related fields*, 131(4):528–552, 2005.

[4] A. Balakrishnan. On the problem of time jitter in sampling. *IRE Trans. on Inf. Theory*, 8(3):226–236, 1962.

[5] Mark Cieliebak and Stephan Eidenbenz. Measurement errors make the partial digest problem np-hard. In *Latin American Symposium on Theoretical Informatics*, pages 379–390. Springer, 2004.

[6] Tamara Dakic. *On the turnpike problem.* Simon Fraser University BC, Canada, 2000.

[7] Li Deng. The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.

[8] I. Dokmanić. Permutations unlabeled beyond sampling unknown. *IEEE Signal Processing Letters*, 26(6):823–827, 2019.

[9] Phil Duxbury, Carlile Lavor, Leo Liberti, and Luiz Leduino de Salles-Neto. Unassigned distance geometry and molecular conformation problems. *Journal of Global Optimization*, pages 1–10, 2021.

[10] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[11] Shuai Huang and Ivan Dokmanic. Reconstructing point sets from distance distributions. *IEEE Trans. Signal Process.*, 69:1811–1827, 2021.

[12] Manuel Marques, Marko Stošić, and João Costeira. Subspace matching: Unique solution to point matching with geometric constraints. In *2009 IEEE 12th Int. Conference on Computer Vision*, pages 1288–1294, 2009.

[13] Gonzalo Mena, David Belanger, Scott Linderman, and Jasper Snoek. Learning latent permutations with gumbel-sinkhorn networks, 2018.

[14] A. Pananjady, M. J. Wainwright, and T. A. Courtade. Denoising linear models with permuted data. In *2017 IEEE Int. Symposium on Inf. Theory (ISIT)*, pages 446–450, 2017.

[15] 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*, 64(5):3286–3300, 2018.

[16] Liangzu Peng, Xuming Song, Manolis C Tsakiris, Hayoung Choi, Laurent Kneip, and Yuamming Shi. Algebraically-initialized expectation maximization for header-free communication. In *IEEE Int. Conf. on Acous., Speech and Signal Process. (ICASSP)*, pages 5182–5186. IEEE, 2019.

[17] Liangzu Peng and Manolis C. Tsakiris. Linear regression without correspondences via concave minimization. *IEEE Signal Proces. Letters*, 27:1580–1584, 2020.

[18] Martin Slawski, Emanuel Ben-David, and Ping Li. Two-stage approach to multivariate linear regression with sparsely mismatched data. *J. Mach. Learn. Res.*, 21(204):1–42, 2020.

[19] Martin Slawski, Mostafa Rahmani, and Ping Li. A sparse representation-based approach to linear regression with partially shuffled labels. In *Uncertainty in Artificial Intelligence*, pages 38–48. PMLR, 2020.

[20] Jayakrishnan Unnikrishnan, Saeid Haghighatshoar, and Martin Vetterli. Unlabeled sensing with random linear measurements. *IEEE Trans. Inf. Theory*, 64(5):3237–3253, 2018.

[21] Yangyang Xu and Wotao Yin. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on imaging sciences*, 6(3):1758–1789, 2013.

[22] Hang Zhang and Ping Li. Optimal estimator for unlabeled linear regression. In *Proceedings of the 37th Int. Conference on Machine Learning (ICML)*, 2020.

[23] Hang Zhang, Martin Slawski, and Ping Li. Permutation recovery from multiple measurement vectors in unlabeled sensing. In *2019 IEEE Int. Symposium on Inf. Theory (ISIT)*, pages 1857–1861, 2019.