UNLABELED SENSING WITH LOCAL PERMUTATIONS

Ahmed Abbasi, Shoaib Bin Masud, Boyang Lyu, Shuchin Aeron
Department of ECE, Tufts University, Medford, MA 02155

ABSTRACT

Unlabeled sensing is a linear inverse problem where the measurements are scrambled with an unknown permutation resulting in a loss of correspondence to the measurement matrix. In this paper, we consider a special case of the unlabeled sensing problem where we restrict the class of permutations to be local and allow for multiple views. This setting is motivated via some practical problems. In this setting, we consider a regime where none of the previous results and algorithms are applicable and provide a computationally efficient algorithm, which exploits the Gromov-Wasserstein alignment framework locally. Simulation results are provided on synthetic data sets.

Index Terms — Unlabeled Sensing, Gromov-Wasserstein Alignment, Timing mismatch, ADC

1. 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 [14, 5] where information theoretic results were derived for identification of unknown signal under linear measurements when the measurement correspondence is lost. Several generalizations of the set-up as well as specific cases have been considered in [7, 3, 4, 6, 8, 12, 2]. Relation to these existing works are explained in section 2.1.

Our contributions: In contrast to the specific cases considered recently in [8, 12, 3] , in this paper we consider a local permutation model and derive an algorithm that is robust under some mild conditions. In this context we are motivated by the applications of interleaved ADC sampling schemes. An interleaved ADC system of order $r$ outputs $r$ samples at each clock cycle, effectively increasing the sampling rate by a factor of $r$, Fig[13]. A natural extension of the unlabeled sampling problem to an interleaved system is to consider signal recovery from observations where the exact time indices corresponding to each of the $r$ samples output at every clock are not known, Fig[14]. The problem is well motivated in that it can arise as a consequence of timing mismatch [11, 10] in interleaved ADC sampling. Mismatch, or skew, occurs as the sampling clocks drift out of phase with respect to one another, resulting in $r$ skewed samples at each clock cycle. Fig[15] shows a representative example of timing mismatch. We model unlabeled sensing in such interleaved systems via $r$—local permutations, as defined in Definition 1. A local structure on the unknown permutation can also be used to model other practical problems, such as automated knowledge translation [13] or spatial mapping via mobile sensors with random Brownian motion trajectories [9]. Unlabeled sensing in source synchronous signalling schemes, such as dual data rate (DDR) or quad data rate (QDR) interfaces, is also exactly modelled by $2$—local and $4$—local permutations, respectively. In this context a novel aspect of our proposed algorithm is that it exploits the machinery of Graph Alignment/Gromov-Wasserstein (GW) to resolve local permutations.

The rest of the paper is organized as follows: In section 2, we present the mathematical set-up followed by a brief literature survey on related work. In section 3 we outline in detail our proposed algorithm for the model in 2. In section 4 we provide detailed simulations results.

2. PROBLEM FORMULATION

The problem is to estimate the unknown signal, $X \in \mathbb{R}^{d \times m}$, from the observations, $Y \in \mathbb{R}^{n \times m}$:

$$Y = \pi_o B X + N \quad (1)$$

where $B \in \mathbb{R}^{n \times d}$ is the known measurement matrix, $N \in \mathbb{R}^{n \times m}$ is additive noise with each entry being IID $\mathcal{N}(0, \sigma^2)$, and $\pi_o$ is an $r$—local permutation, as defined below:

Definition 1 A permutation, $\pi_o$, is $r$—local, i.e. $\pi_o \in \mathcal{P}_n^r$, if it is composed of $n/r$ blocks along the diagonal, with each block being a permutation matrix of size $r \times r$. That is, $\pi = \text{blkdiag}[\pi_1, \ldots, \pi_{n/r}]$ where $\pi_i \in \mathcal{P}_r \forall i \in [n/r]$ and $\mathcal{P}_r$ denotes the set of $r \times r$ permutation matrices.

Fig.2 shows examples of $r$—local permutations, where $r = 20$.

2.1. Relation to existing work

Algorithms for unlabeled sensing, with specific structures on the unknown permutation have been proposed in [8, 12, 13], [2] considers the multi-view setting and , given $m = d$ noiseless observations, where $m = \text{rank}(X), d = \text{rank}(B)$, proposes the LeVSort algorithm that recovers the unknown permutation exactly. In this work, we consider signal recovery from $m < d$ noisy, locally permuted observations. The

Ahmed Abbasi and Shuchin Aeron are supported by NSF CAREER award. Boyang Lyu is supported through a grant from AFOSR. Shaib Bin Masud is supported through a grant from Army Natick Research Center.
block diagonal permutation structure we consider, has also been considered in [13]; however, with the following additional assumptions: (a) The data, \( X \in \mathbb{R}^{d \times d} \) is orthogonal, i.e. \( X^\top X = I \); (b) The number of observations, \( m = d \), where \( d \) is the dimension of the underlying signal; (c) The unknown block diagonal permutation, \( \pi_n \), is also assumed to be sparse there, in contrast to our assumptions. We next outline how our proposed approach and algorithm derives from and compares with existing work.

**Proposed Approach:** Given a single permuted observation, \( y \), where \( y = \pi B x \), [8] proposes the following iterative system of equations to recover the problem parameters \((\pi, x)\):

\[
\begin{align}
    x_{i+1} & \in \operatorname{arg \ min}_{x \in \mathbb{R}^d} \| y - \pi_i B x \|^2 \tag{2a} \\
    \pi_{i+1} & \in \operatorname{arg \ min}_{\pi \in P_n} \| y - \pi_i B x_{i+1} \|^2 \tag{2b}
\end{align}
\]

The formulation in (2) is justified by the RIP Lemma [12], but involves a combinatorial search over the set, \( P_n \). In contrast, we extend the setup in (2) to account for multiple observations, \( Y \) and relax (2b) as follows.

\[
\pi^* \in \arg \min_{\pi \in P_n} \| YY^\top - \left( U'U'^\top Y \right) \left( U'U'^\top Y \right)^\top \|^2 \tag{3}
\]

Note that \( U' = \pi U_B \) where \( U_B \) is the matrix of the left singular vectors of the measurement matrix, \( B \). As detailed in section 3, the novelty of our approach lies in approximating (3) and computing a soft assignment at each iteration \( t \) by the iterative Quadratic Assignment Problem or the Gromov-Wasserstein (GW) formulation:

\[
\pi_{i+1} \in \arg \min_{\pi \in P_n} \| YY^\top - \left( \pi(U_iU_i^\top)Y \right) \left( \pi(U_iU_i^\top)Y \right)^\top \|^2,
\]

where \( U_{i+1} = \pi_{i+1} U_i, U_1 = U_B \). Further, in contrast to other existing algorithms such as [4], our algorithm is stable in the presence of noise.

**3. ALGORITHM**

The algorithm, **Depermute**, under the problem set up of the previous section is outlined in Algorithm 1 and is described below in detail. We let \( \tilde{n} \doteq n/r, \tilde{m} \doteq m/r, \Delta \doteq d - \tilde{n} \), and \( 1_i \in \mathbb{R}^n \) denote a column vector defined via:

\[
1_i(p) = \begin{cases} 
    1 & \forall p = (i - 1)r + 1, \ldots, ir \\
    0 & \text{else}
\end{cases}
\]

The operator \( \operatorname{norm}_c(A) \) scales the columns of the matrix \( A \) to unit magnitude. \( \Gamma \in GW(C,C',\epsilon) \) is the coupling returned between the similarity matrices \( C, C' \), with the entropic regularization parameter set to \( \epsilon \). The detailed description of the Gromov-Wasserstein algorithm can be found in [1].

**Initial Estimate:** We consider the following collapsed system of linear equations, determined by \( \tilde{B} \in \mathbb{R}^{\tilde{n} \times d}, \tilde{Y} \in \mathbb{R}^{\tilde{n} \times m} \) and the unknown \( \tilde{X} \in \mathbb{R}^{d \times m} \):

\[
\tilde{B} \tilde{X} = \tilde{Y}
\]

where each row in \( \tilde{B}, \tilde{Y} \) is the sum of \( r \) corresponding rows in \( B, Y \) respectively:

\[
\tilde{b}(i) = \tilde{1}_i^\top B \quad \forall i = 1, \ldots, \tilde{n}, \quad \tilde{y}(i) = \tilde{1}_i^\top Y \quad \forall i = 1, \ldots, \tilde{n}
\]
Algorithm 1 Depermute

Input: Radius \( r \in \mathbb{R} \), measurement matrix \( B \in \mathbb{R}^{n \times d} \),
Observations \( Y \in \mathbb{R}^{n \times m} \), Number of iterations \( \ell \leq n - \bar{n} \).

// Set Fixed Target Distributions
\( Y_{im} \leftarrow S_iYS_{im}^* \forall i \in [\bar{n}], \ m \in [\bar{m}] \)
\( Y_{im} \leftarrow \text{normc}(Y_{im}) \forall i \in [\bar{n}], \ m \in [\bar{m}] \)

\[
\begin{align*}
\bar{b}(i) & \leftarrow \bar{B}^T \ \forall i \in [\bar{n}] \\
\bar{y}(i) & \leftarrow \bar{Y} \ \forall i \in [\bar{n}] \\
\bar{J}_t & \leftarrow \{ \} \ \text{or} \ \bar{J}_t \leftarrow \{ \}
\end{align*}
\]

for \( t = 1 \) : \( \ell \) do

// Block Diagonal GW Coupling

\( \bar{Y} = B \left( \bar{B}^T \bar{Y} \right) \)

\( \bar{Y}_{im} = S_i \bar{Y}S_{im}^* \forall i \in [\bar{n}], \ m \in [\bar{m}] \)
\( \bar{Y}_{im} \leftarrow \text{normc}(\bar{Y}_{im}) \forall i \in [\bar{n}], \ m \in [\bar{m}] \)
\( \Gamma_{im} \in \text{GW}(\bar{Y}_{im}, \bar{Y}_{im}^T) \forall i \in [\bar{n}], \ m \in [\bar{m}] \)
\( \Gamma_t \leftarrow \Gamma_{i1} + \ldots \Gamma_{im} \forall i \in [\bar{n}] \)
\( \Gamma \in \text{diag} \left[ \Gamma_{i1}, \ldots, \Gamma_{in} \right] \)

// Augment Linear System of Equations

\[
S = \left\{ (i,j) \mid \frac{\text{rank} \left[ B_{b(i)} \right]}{\text{rank}(\bar{B}) + 1, d}, \right\}
\]

\((i_t, j_t) \in \arg\max_{i,j \in S} \Gamma(i,j)\)

\( \bar{B} = \left[ \bar{b}(i_t) \right] \bar{Y} = \left[ \bar{Y}_{im} \right] \)

// Update Sets of Matched Indices

\( \bar{I}_{i+1} = \bar{I}_i \cup i_t \)
\( \bar{J}_{i+1} = \bar{J}_i \cup j_t \)

end for

return \( \bar{X} = B^T \bar{Y} \)

Note that \( \bar{b}(i), \bar{y}(i) \) denote the \( i \)th row of \( \bar{B}, \bar{Y} \) respectively. Let \( \bar{X} \in \mathbb{R}^{d \times \bar{m}} \) be the least squares solution of the LSE \([2]\), i.e. \( \bar{X} = B^T \bar{Y} \). For a sufficiently overdetermined, noiseless instantiation of the model in \( [2] \) i.e. \( \bar{n} \geq d, \sigma^2 = 0 \), the LSE estimate, \( \bar{X} \) is equal to the unknown \( X \) in \( [2] \), i.e. \( \bar{X} = X \). We do not consider the fully determined, noiseless case but propose an iterative algorithm for underdetermined \( \bar{n} < d \) and noisy \( \sigma^2 > 0 \) case.

**Block Diagonal GW Coupling:** As shown in Fig. 3 we approximate each unknown sub-permutation \( \pi_i \in \mathbb{R}^{r \times r} \forall i \in [\bar{n}] \), where \( \pi_i = \text{blkdiag} [\pi_1, \ldots, \pi_{\bar{n}}] \), via the average of \( \bar{m} \) independent GW couplings \( \Gamma_{im} \in \mathbb{R}^{r \times r} \):

\[
\begin{align*}
\Gamma_{im} & \in \text{GW}(\bar{Y}_{im}, \bar{Y}_{im}^T) \forall i \in [\bar{n}], \ m \in [\bar{m}] \quad (5a) \\
\Gamma_t & = \Gamma_{i1} + \ldots + \Gamma_{\bar{m}} \\
\Gamma & = \text{blkdiag} [\Gamma_{i1}, \ldots, \Gamma_{in}] \quad (5c)
\end{align*}
\]

Note that \( \Gamma_{im} \) is a coupling between a similarity matrix \( \bar{Y}_{im} \) and the noisy instantiation of the model in \( [2] \) i.e. \( \bar{Y}_{im} \) is matched, via the coupling \( \Gamma_{im} \), to the corresponding target distribution \( Y_{im} \) of the sub-matrix, \( Y_{im} \in \mathbb{R}^{r \times r} \)

\[
\begin{align*}
\bar{Y}_{im} & \in \mathbb{R}^{r \times r} \text{ of the sub-matrix, } \bar{Y}_{im} \text{ and the corresponding target similarity matrix, } Y_{im}Y_{im}^T \in \mathbb{R}^{r \times r} \text{ of the sub-matrix, } Y_{im} \in \mathbb{R}^{r \times r}
\end{align*}
\]

\[
\begin{align*}
\bar{Y} & = \left[ \bar{Y}_{11} \mid \bar{Y}_{12} \mid \ldots \mid \bar{Y}_{1m} \right] \quad (5d) \\
Y & = \left[ Y_{11} \mid Y_{12} \mid \ldots \mid Y_{1m} \right] \\
\end{align*}
\]

where \( \bar{Y}_{11} \) denotes the column concatenation of \( \bar{Y}_{11} \in \mathbb{R}^{r \times r} \)
\( \bar{Y}_{12} \in \mathbb{R}^{r \times r} \) and \( \bar{Y}_{11} \mid \ldots \mid \bar{Y}_{1m} \) denotes the row concatenation of the matrix \( \bar{Y}_{11} \mid \ldots \mid \bar{Y}_{1m} \in \mathbb{R}^{r \times m} \) with \( \bar{Y}_{11} \mid \ldots \mid \bar{Y}_{1m} \in \mathbb{R}^{r \times m} \). Each sub-matrix \( \bar{Y}_{im} \) is extracted from \( Y(Y) \) via pre and post multiplication with selection matrices \( S_i, S_{im}^* \) respectively:

\( \bar{Y}_{im} = S_i \bar{Y}S_{im}^* \), \( Y_{im} = S_iYS_{im}^* \forall i \in [\bar{n}], \ m \in [\bar{m}] \)

where the construction of \( S_i \in \mathbb{R}^{r \times r} \) is given below:

\[
\begin{align*}
S_1 & = [0_{r \times r} \mid 0_{r \times r} \mid \cdots \mid 0_{r \times r}] \\
S_2 & = [0_{r \times r} \mid I_{r \times r} \mid \cdots \mid 0_{r \times r}] \\
S_{\bar{n}} & = [0_{r \times r} \mid 0_{r \times r} \mid \cdots \mid I_{r \times r}]
\end{align*}
\]

Note that \( 0_{r \times r} \in \mathbb{R}^{r \times r} \) denotes the matrix of all zeros and \( I_{r \times r} \in \mathbb{R}^{r \times r} \) is the \( r \times r \) identity matrix. \( S' \in \mathbb{R}^{m \times m} \) can be similarly constructed to extract \( r \) columns form \( \left( S_i, \bar{Y} \right) \in \mathbb{R}^{r \times m} \).

**Augment Linear System of Equations:** We use the global coupling, \( \Gamma \) computed in \( [5c] \) to augment the linear system of equations \([3]\) with the measurement vector, \( b_2 \) and the obser-
Fig. 4: Probability of successful recovery, $\Pr[A]$, via Depermute algorithm for $\triangle \in [32, 40, 64]$ against the number of given observations, $m$. The underlying signal has dimension, $d = n/r + \Delta$. Compared to the proposed algorithm, successful signal recovery as defined in metric 1, via the least squares solution to the collapsed LSE (4) requires an additional $\Delta r$ measurements. For $n = 1024$, $\Delta = 64$, $r = 4$, this would correspond to a 25 percent increase in the number of required measurements.

4. SIMULATION RESULTS

4.1. Data Generation

For all synthetic data experiments, the entries of the measurement matrix, $B \in \mathbb{R}^{m \times d}$, are drawn IID from the $\mathcal{N}(0, 1)$ distribution. Each entry of the underlying signal, $X \in \mathbb{R}^{d \times m}$, is drawn IID from $\mathcal{N}(0, 1)$, the standard normal distribution. To normalize for SNR, we first generate $Y = BX$, and subsequently scale both sides by $\frac{1}{\text{diag}(\tilde{Y}^\top \tilde{Y})}$ before adding noise. Note that for this case, following [12], the signal to noise ratio, SNR, is given by $\text{SNR} = \frac{1}{\langle \sigma^2 \rangle}$. We now describe the metrics with respect to which we will evaluate the performance of the algorithm. For all our simulations, we set the $GW$ entropic regularizer, $\epsilon = 0.5$.

Metric 1: Depermutation. The collapsed LSE in (4) is under determined and requires $d - \bar{n}$ de-permuted measurements to form a fully determined system. We define $\triangle \triangleq d - \bar{n}$, and set the number of iterations, $\ell = \triangle$. Let $E_1, \cdots, E_\ell$ be binary error random variables such that $E_t = 0$ if iteration $#t$ of the algorithm correctly matches the measurement vector, $b(i_t)$ with the observation, $y(j_t)$ i.e.

$$E_t = \begin{cases} 0 & \text{if } \pi^*_t(i_t) = j_t \\ 1 & \text{otherwise.} \end{cases}$$

We consider recovery by the Depermute algorithm successful if $E_t = 0 \forall t \in [\ell]$ and denote the successful event by $A = \{E_1 = 0, E_2 = 0, \cdots, E_\ell = 0\}$.

For $n = 1024$, $\text{SNR} = 30\, \text{db}$, $d = n/r + \Delta$, Fig. 4 plots the probability of event $A$, $\Pr[A]$, averaged over twenty five repetitions of the Depermute algorithm for $\triangle \in [32, 40, 64]$ against the number of observations, $m$. The results match well with expectation as the initial estimate, $\tilde{X}$ depends on on how under determined the collapsed system is and the averaging across $\bar{m} = m/r$ couplings in (55) improves performance in the presence of noise.

Metric 2: Mean Square Error in Signal Recovery

For $n = 1200$, $r \in [3, 4, 6]$, $m = 8r$, $d = n/r + 12r$, $\ell = 20r$, $\text{SNR} \in [20, 30, 40]$, the scaled mean square error (M.S.E), $\Sigma \triangleq \frac{1}{\bar{m}} \|X - \widehat{X}\|^2$ is given in Table 1.

5. CONCLUSION

For unlabeled sensing with local permutations, we have proposed a robust algorithm that iteratively de-depermutes the measurements. Our results show accurate signal recovery from (a) a small number of observations (b) in varying levels of noise, over (c) a range of problem parameters.

![Graph](image-url)

Table 1: Scaled Mean Square Error, $\Sigma$ as a function of $r$, $\text{SNR}$

| $\text{SNR}$ | 20 dB | 30 dB | 40 dB |
|-------------|-------|-------|-------|
| 3           | 4.41e-06 | 4.26e-07 | 2.77e-08 |
| 4           | 2.70e-06 | 1.92e-06 | 1.25e-06 |
| 6           | 1.33e-05 | 1.12e-05 | 8.31e-06 |
6. REFERENCES

[1] David Alvarez-Melis and Tommi Jaakkola. Gromov-Wasserstein alignment of word embedding spaces. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 1881–1890, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

[2] Thomas A. Courtade Ashwin Pananjady, Martin J. Wainwright. Denoising Linear Models with Permuted Data. 2017 IEEE International Symposium on Information Theory (ISIT), pages 446–450, 2017.

[3] Thomas A. Courtade Ashwin Pananjady, Martin J. Wainwright. Linear Regression With Shuffled Data: Statistical and Computational Limits of Permutation Recovery. IEEE Transactions on Information Theory, 64:3286–3300, 2018.

[4] Xiaorui Sun Daniel Hsu, Kevin Shi. Linear regression without correspondence. 2017.

[5] Martin Vetterli Jayakrishnan Unnikrishnan, Saeid Haghighatshoar. Unlabeled Sensing With Random Linear Measurements. IEEE Transactions on Information Theory, 64:3237–3253, 2018.

[6] Liangzu Peng Manolis C Tsakiris. Homomorphic Sensing. 2019.

[7] Emanuel Ben-David Martin Slawski. Linear Regression with Sparsely Permuted Data. 2017.

[8] Ping Li Martin Slawski, Emanuel Ben-David. A Two-Stage Approach to Multivariate Linear Regression with Sparsely Mismatched Data. 2019.

[9] Y. Mostofi. Compressive cooperative sensing and mapping in mobile networks. IEEE Transactions on Mobile Computing, 10(12):1769–1784, Dec 2011.

[10] D. G. Nairn. Time-interleaved analog-to-digital converters. In 2008 IEEE Custom Integrated Circuits Conference, pages 289–296, Sep. 2008.

[11] B. Razavi. Design Considerations for Interleaved ADCs. IEEE Journal of Solid-State Circuits, 48(8):1806–1817, Aug 2013.

[12] Giuseppe Caire Saeid Haghighatshoar. Signal Recovery From Unlabeled Samples. IEEE Transactions on Signal Processing, 66:1242–1257, 2017.

[13] Xu Shi, Xiaowu Li, and Tianxi Cai. Spherical Regression under Mismatch Corruption with Application to Automated Knowledge Translation. arXiv e-prints, page arXiv:1810.05679, Oct 2018.

[14] J. Unnikrishnan, S. Haghighatshoar, and M. Vetterli. Unlabeled sensing: Solving a linear system with unordered measurements. In 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), pages 786–793, Sep. 2015.