ONE-STEP ESTIMATOR FOR PERMUTED SPARSE RECOVERY

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper considers the unlabeled sparse recovery under multiple measurements, i.e., $\mathbf{Y} = \mathbf{\Pi}^\natural \mathbf{X}\mathbf{B}^\natural + \mathbf{W}$, where $\mathbf{Y} \in \mathbb{R}^{n\times m}, \mathbf{\Pi}^\natural \in \mathbb{R}^{n\times n}, \mathbf{X} \in \mathbb{R}^{n\times p}, \mathbf{B}^\natural \in \mathbb{R}^{p\times m}, \mathbf{W} \in \mathbb{R}^{n\times m}$ represents the observations, missing (or incomplete) correspondence information, sensing matrix, sparse signals, and additive sensing noise, respectively. Different from the previous works on multiple measurements ($m > 1$) which all focus on the sufficient samples regime, namely, $n > p$, we consider a sparse matrix $\mathbf{B}^\natural$ and investigate the insufficient samples regime (i.e., $n \ll p$) for the first time. To begin with, we establish the lower bound on the sample number and *signal-to-noise ratio* (SNR) for the correct permutation recovery. Moreover, we present a simple yet effective estimator. Under mild conditions, we show that our estimator can restore the correct correspondence information with high probability. Numerical experiments are presented to corroborate our theoretical claims.

1 INTRODUCTION

In recent years, linear regression with permuted correspondence has received increasing attention due to its wide applications in the field of machine learning, signal processing, and statistics. Among all these applications, two most prominent examples are (i) *linkage record*, which merges two datasets pertaining to the same objects into one comprehensive dataset; and (ii) *data de-anonymization*, which infers the hidden labels of private data with public datasets. Apart from these two applications, other applications include *correspondence estimation* between pose and estimation in graphics; *time-domain sampling* in the presence of clock jitter; *multi-target tracking*; *unsupervised data alignment*, etc (Pananjady et al., 2018; Slawski & Ben-David, 2019; Slawski et al., 2020; Zhang et al., 2018).

In this paper, we consider the canonical model, i.e., a linear sensing relation with permuted labels reading as

$$\mathbf{Y} = \mathbf{\Pi}^\natural \mathbf{X}\mathbf{B}^\natural + \mathbf{W},$$

where $\mathbf{Y} \in \mathbb{R}^{n\times m}$ is the sensing result, $\mathbf{\Pi}^\natural \in \mathbb{R}^{n\times n}$ is an unknown permutation matrix, $\mathbf{X} \in \mathbb{R}^{n\times p}$ is the design (sensing) matrix, $\mathbf{B}^\natural \in \mathbb{R}^{p\times m}$ represents the sparse signals of interests, and $\mathbf{W} \in \mathbb{R}^{n\times m}$ denotes the additive noise. Assuming the signal $\mathbf{B}^\natural$ is a sparse signal, to put more specifically, each column of $\mathbf{B}^\natural$ is k-sparse, we would like to (i) study the statistical limits of the permutation recovery under this scenario, e.g., the minimum sample number n and *signal-to-noise ratio* (SNR); and (ii) propose a practical estimator that can efficiently recover the permutation once the minimum requirements are met. To begin with, we briefly review the previous works.

Related Works. The study of permuted linear regression has a long history that can at least date back to DeGroot & Goel (1976; 1980); Goel (1975); Bai & Hsing (2005). Recent interests on this area start from Unnikrishnan et al. (2015). Focusing on the noiseless case $\mathbf{W} = \mathbf{0}$ with single measurement $(m = 1)$, Unnikrishnan et al. (2015) establish the necessary condition $n \geq 2p$ for the permutation recovery if $\mathbf{B}^\natural$ is an arbitrary vector residing within the linear space $\mathbb{R}^p$. Later, Pananjady et al. (2018) extend the analysis to the noisy scenario. They showed the minimum SNR should be at least the order of $\Omega(n^c)$, where $c > 0$ is some positive constant. Numerical experiments suggest c is within the region $[4, 5]$. Other works such as Hsu et al. (2017); Abid et al. (2017); Slawski & Ben-David (2019); Tsakiris et al. (2020); Haghighatshoar & Caire (2018) also focus on this regime and obtain the same answer. In Emiya et al. (2014), the setting with a sparse signal $\mathbf{B}^\natural$ is first studied. However, only empirical investigation is conducted without rigorous theoretical analysis. The first breakthrough

comes from Zhang & Li (2021), where both the statistical limits and practical estimators with almost optimal performance are presented for the permutation recovery. However, all these works suggest that $\mathsf{SNR} = \Omega(n^c)$ is inevitable for the permutation reconstruction if only one measurement is conducted, namely, $m = 1$.

On the other hand, numerous works suggest multiple measurements, i.e., $m > 1$, can greatly reduce the SNR requirement, even to some positive constant. This line of research starts from Zhang et al. (2022), where the information theoretic lower bounds and the *maximum likelihood* (ML) estimator are investigated. Later, Zhang & Li (2020) study this problem from the viewpoint of non-convex optimization and propose an optimal estimator for the permutation recovery. Independently, Slawski et al. (2020) investigate this problem from the viewpoint of denoising. Putting parsimonious constraints on the number of permuted rows, they view $(\mathbf{I} - \mathbf{\Pi}^\natural)\mathbf{X}\mathbf{B}^\natural$ as sparse outliers and design the permutation recovery algorithm accordingly. However, all the above works focus on the sufficient samples regime, namely, $n = \Omega(p)$. In this paper, we focus on the insufficient samples regime. Assuming $\mathbf{B}^\natural$ to be a sparse signal, we would like to show the correct permutation can be obtained with $n \ll p$ and $\mathsf{SNR} = O(1)$.

Our **contributions** are summarized as follows

- We study the lower bounds w.r.t the number n and *signal-to-noise ratio* (SNR) for the correct reconstruction of both the permutation matrix $\mathbf{\Pi}^\natural$ and the signal $\mathbf{B}^\natural$. Assuming each column $\mathbf{B}^\natural_{:,\ell}$ $(1 \le \ell \le m)$ is k-sparse, we show that the sample number n should be at least of the order $\Omega(k \log p)$; meanwhile the SNR should satisfy $\log\det(\mathbf{I} + {\mathbf{B}^\natural}^\top \mathbf{B}^\natural/\sigma^2) > \log n + \frac{m \log \binom{p}{k}}{n}$.
- We propose a one-step estimator for the correspondence recovery. By formulating the correspondence recovery as a *linear assignment problem* (LAP), we show that the correct permutation matrix can be obtained when SNR is above certain positive constant.

On top of the above contributions, we would like to briefly mention our proof strategy, which is based on a tailored version of the **leave-one-out** technique. Compared with the previous works that adopt the leave-out-out technique (Chen et al., 2020; Sur et al., 2019; El Karoui, 2013; 2018; Cai et al., 2021), our construction method bas the following characteristics

- Our construction method is adaptive, which replaces multiple rows ranging from 2 to 4 depending on each permuted row. Meanwhile, previous works such as Chen et al. (2020); Sur et al. (2019); El Karoui (2013; 2018); Cai et al. (2021) replace a fixed number of rows (or columns).
- We not only leave out the rows, but also modify the thresholding operator operated on the perturbed samples $\widetilde{\mathbf{B}}_{(\cdots)}$ from $\mathsf{thres}(\cdot)$ to $(\cdot)_{\mathsf{imax}}$ (its definition is deferred to Subsection 4.2). This step is essential in controlling the approximation error, since otherwise the non-zero elements in matrices $\widehat{\mathbf{B}}$ $(\propto \mathbf{X}^\top\mathbf{Y})$ and $\widetilde{\mathbf{B}}_{(\cdot)}$ may not share the same position and the approximation error can be considerably large. For a thorough understanding, we refer the interested readers to our proof of Theorem 3 in the supplementary material.

Notations. Denote c, $c^{'}$, c_i as some positive constants, whose values are not necessarily the same even for those with the same notations. We denote $a \lesssim b$ if there exists some positive constants $c_0 > 0$ such that $a \le c_0 b$. Similarly, we define $a \gtrsim b$ provided $a \ge c_0 b$ for some positive constant c_0. We write $a \asymp b$ when $a \lesssim b$ and $a \gtrsim b$ hold simultaneously.

For an arbitrary matrix $\mathbf{M}$, we denote $\mathbf{M}_{i,:}$ as its ith row, $\mathbf{M}_{:,i}$ as its ith column, and $\mathbf{M}_{ij}$ as its (i,j)th element. Its Frobenius norm is defined as $\|\mathbf{M}\|_{\mathrm{F}}$ while the operator norm is denoted as $\|\mathbf{M}\|_{\mathrm{OP}}$, whose definitions can be found in Section 2.3 of Golub & Loan (2013). In addition, we define its stable rank as $\mathrm{srank}(\cdot) \triangleq \|\cdot\|_{\mathrm{F}}^2/\|\cdot\|_{\mathrm{OP}}^2$ (Section 2.1.15 in Tropp (2015)). The inner product between matrices is denoted as $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ while the inner product between vectors is denoted as $\langle \cdot, \cdot \rangle$.

We define the set of all possible permutation matrices as $\mathcal{P}_n$, which is defined as $\{\mathbf{\Pi} \in \mathbb{R}^{n\times n} : \sum_{i=1}^n \mathbf{\Pi}_{ij} = 1, \sum_{j=1}^n \mathbf{\Pi}_{ij} = 1, \mathbf{\Pi}_{ij} \in \{0,1\}\}$. Associate with each permutation matrix $\mathbf{\Pi}$, we define the operator $\pi(\cdot)$ that transforms index i to $\pi(i)$ under $\mathbf{\Pi}$. The Hamming distance $\mathsf{d_H}(\mathbf{\Pi}_1, \mathbf{\Pi}_2)$ between two permutation matrices $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ is defined as $\mathsf{d_H}(\mathbf{\Pi}_1, \mathbf{\Pi}_2) = \sum_{i=1}^n \mathbb{1}(\pi_1(i) \neq \pi_2(i))$. The SNR is defined as $\|\mathbf{B}^\natural\|_{\mathrm{F}}^2/(m \cdot \sigma^2)$.

2 PROBLEM FORMULATION

We start this section with a formal restatement of the considered problem reading as

$$\mathbf{Y} = \mathbf{\Pi}^\natural \mathbf{X} \mathbf{B}^\natural + \mathbf{W}, \tag{1}$$

where $\mathbf{\Pi}^\natural \in \mathcal{P}_n$ denotes the (fixed but unknown) permutation matrix, $\mathbf{X} \in \mathbb{R}^{n\times p}$ is the sensing (design) matrix with its entries being i.i.d. standard normal random variables, i.e., $\mathbf{X}_{ij} \sim \mathsf{N}(0,1)$,[1] $\mathbf{B}^\natural \in \mathbb{R}^{p\times m}$ is a fixed sparse matrix awaiting to be reconstructed (corresponding to the sparse signal), and $\mathbf{W} \in \mathbb{R}^{n\times m}$ denote the noise with each entry $\mathbf{W}_{ij} \overset{\text{i.i.d}}{\sim} \mathsf{N}(0,\sigma^2)$. Here we put the separate sparse constraints on each column of $\mathbf{B}^\natural$, namely, $\|\mathbf{B}^\natural_{:,\ell}\|_0 \le k$ $(1 \le \ell \le m)$. In addition, we denote h as the number of permuted rows, or equivalently, the Hamming distance between the identify matrix $\mathbf{I}$ and the permutation matrix $\mathbf{\Pi}^\natural$, namely, $h \triangleq \mathsf{d_H}(\mathbf{I}, \mathbf{\Pi}^\natural)$.

Our goal is to reconstruct both the permutation matrix $\mathbf{\Pi}^\natural$ and sparse signal $\mathbf{B}^\natural$ from (1). Notice that we do not assume that different columns of $\mathbf{\Pi}^\natural$ share the same support set. Actually, we prefer each column to be with a different support set, since otherwise $\text{rank}(\mathbf{B}^\natural)$ will be bounded by k and will bring extra difficulties to the permutation recovery. A detailed explanation is deferred to Section 4.

Before proceeding, we briefly review the prior art. In Unnikrishnan et al. (2015), where $\mathbf{B}^\natural$ can reside within the entire linear space $\mathbb{R}^p$, it is proved that $n \ge 2p$ is required for the correct permutation recovery. As a result, subsequent works such as Pananjady et al. (2018); Zhang et al. (2022); Slawski & Ben-David (2019); Slawski et al. (2020); Zhang & Li (2020) all focus on the sufficient sample regime, i.e., $n = \Omega(p)$. Only until Zhang & Li (2021), the insufficient sample regime, i.e., $n = o(p)$, receives its first theoretical investigation. Similar to our setting, they put sparsity assumption on $\mathbf{B}^\natural$ however focus on the single measurement scenario, namely, $m = 1$. They show that the minimum SNR for correct $(\mathbf{\Pi}^\natural, \text{supp}(\mathbf{B}^\natural))$ to be $\Omega(n^{c_1/\text{srank}(\mathbf{B}^\natural)} p^{c_2 p/n})$. In the following context, we will show that the minimum SNR can be significantly reduced, to put more specifically, some positive constant provided multiple measurements are made $(m \gg 1)$.

3 INFORMATION THEORETIC LOWER BOUNDS

This section establishes the information theoretic lower bounds for the correct permutation recovery. Our goal is to ensure both $\mathbf{\Pi}^\natural$ and $\mathbf{B}^\natural$ can be reliably reconstructed. We investigate this problem from two perspectives: (i) sample number n and (ii) minimum SNR.

3.1 MINIMUM SAMPLE NUMBER n

We obtain the minimum sample number n such that sparse signal $\mathbf{B}^\natural$ can be reliably recovered with high probability. Here we consider the oracle situation where $\mathbf{\Pi}^\natural$ is given a prior. As each column in $\mathbf{B}^\natural$ does not necessarily share the same support set, we need to iteratively reconstruct each column $\mathbf{B}^\natural_{:,\ell}$, a k-sparse signal, from the corresponding readings $\mathbf{Y}_{:,\ell}$ $(1 \le \ell \le m)$. With the classical result in Donoho (2006); Candes et al. (2006); Candès et al. (2006), we obtain the lower bound on n, namely, $n \gtrsim k \log p$. Naturally, we can expect this bound applies to the non-oracle situation as well, since reliable estimation of $\mathbf{B}^\natural$ is hopeless provided it is out of reach in the oracle situation.

3.2 MINIMUM SNR

Then we turn to the minimum SNR requirement for the correct permutation recovery. To begin with, we restate the prior art in Zhang et al. (2022).

Theorem 1 (Theorem 1 in Zhang et al. (2022))**.** *Consider the oracle case where $\mathbf{B}^\natural$ is given a prior and let n approach infinity. For an arbitrary estimator $\widehat{\mathbf{\Pi}}$, we have*

$$\inf_{\widehat{\mathbf{\Pi}}} \sup_{\mathbf{\Pi}^\natural \in \mathcal{P}_n} \mathbb{P}_{\mathbf{X},\mathbf{W}}(\widehat{\mathbf{\Pi}} \ne \mathbf{\Pi}^\natural) \ge \frac{1}{2},$$

provided that $\log\det(\mathbf{I} + \mathbf{B}^{\natural\top}\mathbf{B}^\natural/\sigma^2) < \log n$.

[1] Our numerical experiments also suggest that we may relax this assumption to that $\mathbf{X}_{ij}$ are i.i.d. sub-Gaussian random variables with zero mean.

One drawback of this bound is the missing role of the sparsity number k. This is because Theorem 1 assumes $\mathbf{B}^\natural$ to be perfectly known while sparsity number k only kicks in when $\mathbf{B}^\natural$ needs to be reconstructed. To handle such issue, we take $\text{supp}(\mathbf{B}^\natural)$ into account as well. Then we obtain the following result

Theorem 2. *Consider the large-system limit where n approaches infinity. For arbitrary estimators $\widehat{\mathbf{\Pi}}$ and $\widehat{\mathbf{B}}$, we have*

$$\inf_{\widehat{\mathbf{\Pi}},\widehat{\mathbf{B}}} \sup_{\substack{\mathbf{\Pi}\in\mathcal{P}_n,\\ \mathbf{B}\in\mathcal{B}_{p,m}}} \mathbb{P}_{\mathbf{X},\mathbf{W}}[(\widehat{\mathbf{\Pi}}, \text{supp}(\widehat{\mathbf{B}})) \neq (\mathbf{\Pi}, \text{supp}(\mathbf{B}))] \geq \frac{1}{2},$$

where $\mathcal{P}_n$ denotes the set of all permutation matrices and $\mathcal{B}_{p,m}$ is defined as the set reading as $\{\mathbf{B} \in \mathbb{R}^{p\times m} : \log\det(\mathbf{I} + {}^{\mathbf{B}^\top\mathbf{B}}/_{\sigma^2}) \leq \log n + \frac{m\log\binom{p}{k}}{n}\}$.

This theorem suggests that for all possible estimators to reconstruct $\mathbf{\Pi}$ and $\text{supp}(\mathbf{B})$, there exists at least one pair $(\mathbf{\Pi}, \mathbf{B})$, $\mathbf{\Pi} \in \mathcal{P}_n, \mathbf{B} \in \mathcal{B}_{p,m}$ such that the reconstruction error rate will be at least $1/2$. [2] Hence, a reliable correspondence reconstruction requires $\mathbf{B}^\natural$ (a fixed but unknown matrix) to satisfy $\log\det\left(\mathbf{I} + {}^{\mathbf{B}^{\natural\top}\mathbf{B}^\natural}/_{\sigma^2}\right) > \log n + {}^{m\log\binom{p}{k}}/_{n}$.

We leave the rigorous proof to the supplementary material and only give an intuitive interpretation, which takes the viewpoint of coding theory. First, we assume each entry in $\mathbf{B}$ to be binary, i.e., $\mathbf{B}_{ij} \in \{0, 1\}$, $(1 \leq i \leq p,\ 1 \leq j \leq m)$. Thus, the information of $\mathbf{B}$ is fully incorporated in $\text{supp}(\mathbf{B})$. In the following, we will use $\text{supp}(\mathbf{B})$ and $\mathbf{B}$ interchangeably, as they are identical.

Afterwards, we view the sensing relation in (1) as the following transmission process: (i) pair $(\mathbf{\Pi}, \mathbf{B})$ is encoded into the codeword $\mathbf{\Pi X B}$; (ii) $\mathbf{\Pi X B}$ passes through a Gaussian additive channel; and (iii) one observes $\mathbf{Y}$, from which one would like to obtain $(\mathbf{\Pi}, \mathbf{B})$.

Using the terminology from coding theory, we can compute the corresponding code rate and channel capacity as $\frac{\log n! + m\log\binom{p}{k}}{n}$ and $\log\det(\mathbf{I} + {}^{\mathbf{B}^\top\mathbf{B}}/_{\sigma^2})$, respectively. Due to the Shannon theorem, we can expect non-negligible decoding error if the coding rate is greater than the channel capacity, which leads to $\log\det(\mathbf{I} + {}^{\mathbf{B}^\top\mathbf{B}}/_{\sigma^2}) < \log n + \frac{m\log\binom{p}{k}}{n}$, the formula in Theorem 2.

Having obtained the information theoretic lower bounds, we will propose a computationally-optimal estimator which matches the lower bounds thereof to a good extent.

4 Estimator Design

This section proposes a computationally efficient estimator for the permutation recovery. Denote $\mathsf{thres}(\cdot)$ as the operator which only keeps the element with the largest magnitude in each column, we reconstruct $\mathbf{\Pi}^\natural$ with the *linear assignment problem* (LAP) reading as

$$\mathbf{\Pi}^{\text{opt}} = \text{argmax}_{\mathbf{\Pi}\in\mathcal{P}_n} \left\langle \mathbf{\Pi}, \mathbf{Y}\cdot\ \mathsf{thres}(\mathbf{X}^\top\mathbf{Y})^\top \cdot \mathbf{X}^\top \right\rangle,$$

Once the permutation matrix $\mathbf{\Pi}^{\text{opt}}$ is obtained, we can transform (1) to the previous setting and iteratively recover each k-sparse column $\mathbf{B}^\natural_{:,i}$. A formal statement of the algorithm is in Algorithm 1. Here we use Lasso estimator to reconstruct the signal $\mathbf{B}^\natural$, which can be replaced with other estimators, say Dantzig estimator, etc.

Design intuition. The design of (2) shares a similar idea of the estimator in Zhang & Li (2020): we would like to approximate the direction of $\mathbf{B}^\natural$ by the product $\mathbf{X}^\top\mathbf{Y}$. However, due to insufficient samples, product $\mathbf{X}^\top\mathbf{Y}$ is poorly aligned with $\mathbf{B}^\natural$, or equivalently, large errors in approximating $\mathbf{B}^\natural$ with $\mathbf{X}^\top\mathbf{Y}$ and wide angle $\angle(\mathbf{B}^\natural, \mathbf{X}^\top\mathbf{Y})$. To reduce the approximation errors, we apply $\mathsf{thres}(\cdot)$ and set certain entries in $\mathbf{X}^\top\mathbf{Y}$ to zero.

Remark 1. *Notice that* ***we always keep one nonzero entry in the operation*** $\mathsf{thres}(\cdot)$ ***regardless of the sparsity number*** k. *This operation is different from almost all the previous works, which ranges from Blumensath & Davies (2009); Foucart (2011) in the literature of compressive sensing to Jain*

[2] This randomness is caused by the sensing matrix $\mathbf{X}$ and sensing noise $\mathbf{W}$.

Algorithm 1: One-step estimator.

Input: observation $\mathbf{Y}$ and sensing matrix $\mathbf{X}$.
Output: pair $(\mathbf{\Pi}^{\text{opt}},\ \mathbf{B}^{\text{opt}})$, which is written as

$$\mathbf{\Pi}^{\text{opt}} = \text{argmax}_{\mathbf{\Pi}\in\mathcal{P}_n} \left\langle \mathbf{\Pi}, \mathbf{Y}\cdot\ \mathsf{thres}(\mathbf{X}^\top\mathbf{Y})^\top\cdot\mathbf{X}^\top \right\rangle, \tag{2}$$

$$\mathbf{B}^{\text{opt}} = \text{argmin}_{\mathbf{B}} (2n)^{-1}\left\|\mathbf{\Pi}^{\text{opt}\top}\mathbf{Y} - \mathbf{X}\mathbf{B}\right\|_{\text{F}}^2 + \lambda_n\|\mathbf{B}\|_1, \tag{3}$$

where $\mathsf{thres}(\cdot)$ applies to each column and thresholds all entries to zero except the one with the largest magnitude, $\mathcal{P}_n$ denotes the set of all possible permutation matrices, and $\lambda_n > 0$ is some regularizer coefficient.

et al. (2013); Yuan et al. (2014); Li et al. (2016) in the literature of optimization, sine all these works suggest keeping at least $O(k)$ non-zero elements for a k-sparse signal.

More surprisingly, our numerical experiments suggest keeping more non-zero elements are detrimental to the permutation recovery. An illustration is put in Figure 1, from which we observe the SNR *required for correct correspondence recovery increases with the number of non-zero elements kept in $\mathbf{X}^\top\mathbf{Y}$. In the supplementary material, we give a rigorous explanation of this phenomenon.*

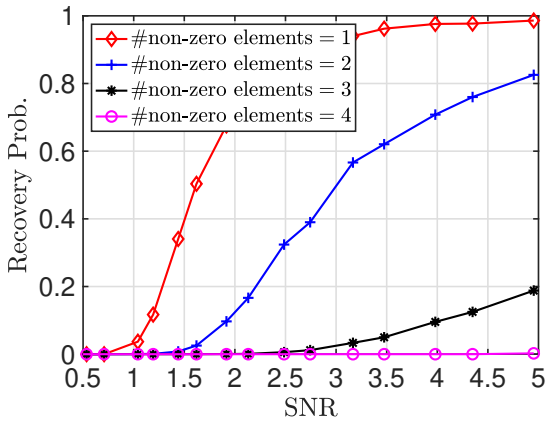

Figure 1: Illustration of the fact such that keeping more non-zero elements in $\mathbf{X}^\top\mathbf{Y}$ deteriorates the permutation recovery. We set $n = 100$, $p = 500$, $h = 25$, $k = 5$, and $\text{srank}(\mathbf{B}^\natural) = 100$. The #**non-zero elements** refers to the number of non-zero elements kept in each column of $\mathbf{X}^\top\mathbf{Y}$.

4.1 Main Results

This subsection formally presents our main results.

4.1.1 Results in Recovering $\mathbf{\Pi}^\natural$

First, we study the correspondence recovery. The formal statement is given as

Theorem 3. *Consider the large-system limit where n and p are sufficiently large. Provided that (i) $n \gg k^2(\log n)(\log^2 mnp)$, (ii) $\text{srank}(\mathbf{B}^\natural) \gg k^2\log^{2(1+\varepsilon)} n$, (iii) $h \le c_0\cdot n$, and (iv) $\mathsf{SNR} \ge c_1$, we have $\{\mathbf{\Pi}^{\text{opt}} = \mathbf{\Pi}^\natural\}$ with probability at least $1 - c_2\cdot n^{-c_3}$, where $\varepsilon > 0$ is an arbitrary positive constant.*

If we put a lower bound on the ratio $\inf_j \max_i |\mathbf{B}^\natural_{i,j}|/\|\mathbf{B}^\natural_{:,j}\|_2$ $(1 \le i \le p, 1 \le j \le m)$, we can relax the requirements on stable rank $\text{srank}(\mathbf{B}^\natural)$ and sample number n. The formal statement is given as

Corollary 1. *Consider the large-system limit where n and p are sufficiently large. Under the conditions (i) $\inf_j \max_i |\mathbf{B}^\natural_{i,j}|/\|\mathbf{B}^\natural_{:,j}\|_2 \ge c$, (ii) $n \gg k(\log n)(\log^2 mnp)$, (iii) $\text{srank}(\mathbf{B}^\natural) \gg \log^{2(1+\varepsilon)} n$,*

(iv) $h \leq c_0 \cdot n$, *and (v)* $\mathsf{SNR} \geq c_1$, *we have* $\{\mathbf{\Pi}^{\text{opt}} = \mathbf{\Pi}^{\natural}\}$ *with probability at least* $1 - c_2 \cdot n^{-c_3}$, *where* $\varepsilon > 0$ *is an arbitrary positive constant.*

Table 1: Comparison with previous works. All results are presented in their best orders, which only hold true in certain regimes. Here $\mathsf{SNR}_{\text{min}}$, n_{min} and h_{max} denotes the minimum SNR required for correct permutation recovery, the minimum required sample number n and maximum allowed number of permuted rows, respectively. Moreover, the logarithmic term is omitted in $\widetilde{\Omega}(\cdot)$.

	$\mathsf{SNR}_{\text{min}}$ $(\geq)$		n_{min}/p $(\geq)$		h_{max}/n $(\leq)$	
	$m = 1$	$m \gg 1$	$m = 1$	$m \gg 1$	$m = 1$	$m \gg 1$
(Pananjady et al., 2018)	$\widetilde{\Omega}(n^c)$		$\widetilde{\Omega}(1)$		$\widetilde{\Omega}(1)$	
(Slawski & Ben-David, 2019)	$\widetilde{\Omega}(n^c)$		$\widetilde{\Omega}(1)$		$\widetilde{\Omega}(\log^{-1} n)$	
(Zhang et al., 2022)		$\widetilde{\Omega}(1)$		$\widetilde{\Omega}(1)$		$\widetilde{\Omega}\left(\log^{-1} r(\mathbf{B}^{\natural})\right)$
(Slawski et al., 2020)		$\widetilde{\Omega}(1)$		$\widetilde{\Omega}(p)$		$\widetilde{\Omega}(\log^{-1} n)$
(Zhang & Li, 2020)	$\widetilde{\Omega}(n^c)$	$\widetilde{\Omega}(1)$	$\widetilde{\Omega}(1)$	$\widetilde{\Omega}(\sqrt{p})$	$\widetilde{\Omega}(1)$	$\widetilde{\Omega}(1)$
(Zhang & Li, 2021)	$\widetilde{\Omega}(n^c)$		$o(1)$		$\widetilde{\Omega}(1)$	
Our Estimator		$\widetilde{\Omega}(1)$		$o(1)$		$\widetilde{\Omega}(1)$

Discussion. Comparison with previous work is put in Table 1, from which we conclude that **our work gives the first affirmative answer such that** $\mathsf{SNR} \geq \Omega(1)$ **is sufficient to obtain the correct permutation matrix with insufficient samples**, i.e., $n \ll p$.

In addition, we would like to compare it with the lower bound. For the sample number n, the lower bound requires n to be at least of order $O(k \log p)$; while Corollary 1 requires n to be $\Omega\left(k(\log n)(\log^2 mnp)\right)$. [3] For the SNR, our estimator can obtain the correct permutation matrix when SNR is above some positive constant; meanwhile Theorem 2 requires $\mathsf{SNR} > 0$. This means our SNR requirement has at most a gap of some positive constant.

Moreover, our estimator allows the maximum number of permuted rows to be linearly proportional to the sample number, i.e., $h_{\text{max}} \asymp n$, which is order-optimal.

Furthermore, we should mention that our analysis prioritize the SNR requirement for the permutation recovery. We believe the minimum sample number n and minimum stable rank $\text{srank}(\mathbf{B}^{\natural})$ can be further reduced provided that the condition $\mathsf{SNR} \geq c_1$ is relaxed.

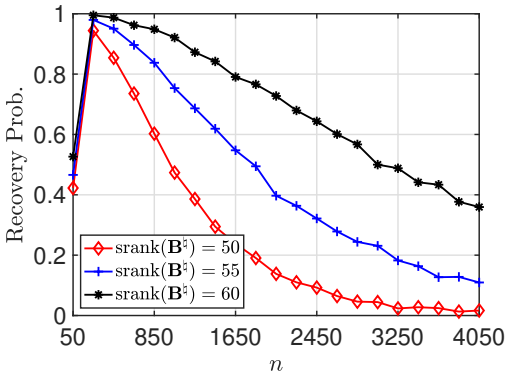

Figure 2: Illustration of the dual role of sample number n. We consider the noiseless case, or equivalently infinite SNR. We fix $p = 200$ (signal length), $k = 5$ (sparsity number), and $h = 25$ (number of permuted rows).

Remark 2. *Compared with Zhang & Li (2020) which only requires* $\text{srank}(\mathbf{B}^{\natural})$ *to be above certain positive constants, our estimator requires a larger stable rank* $\text{srank}(\mathbf{B}^{\natural})$. *Although we cannot claim that* $\text{srank}(\mathbf{B}^{\natural})$ *must be lower bounded by some polynomials of* $\log n$, *we have numerical evidences suggesting that* $\text{srank}(\mathbf{B}^{\natural})$ *may have to increase with sample number* n, *in other words, its lower bound may not be reduced to be some positive constant. One numerical evidence is put as follows.* [4]

Fixing the parameters p, k, h, *and* $\text{srank}(\mathbf{B}^{\natural})$, *we investigate the impact of sample number* n *on the permutation recovery and put the results in Figure 2. We observe that a larger sample size* n

[3] We conjecture that the required sample number n in both Theorem 3 and Theorem 1 can be further reduced with advanced analytical tools.

[4] To some extend, we can interpret that our estimator trades a high stable rank $\text{srank}(\mathbf{B}^{\natural})$ and the sparsity number k for a reduced sample number n.

has a negative impact on the permutation reconstruction once n exceeds certain threshold. One possible reason is that the stable rank $\text{srank}(\mathbf{B}^\natural)$ *is fixed as a constant and violates the requirement* $\text{srank}(\mathbf{B}^\natural) \gg \log^2 n$.

4.1.2 Results in recovering $\mathbf{B}^\natural$

Once the ground-truth $\mathbf{\Pi}^\natural$ is obtained, we can restore (1) to the traditional model in compressive sensing/sparse recovery (Candès et al., 2006; Candes et al., 2006; Donoho, 2006; Wainwright, 2019). One corollary is given as follows.

Corollary 2. *Consider the large-system limit where n and p are sufficiently large. Provided that (i) $n \gg k^2(\log n)(\log^2 mnp)$, (ii) $\text{srank}(\mathbf{B}^\natural) \gg k^2 \log^{2(1+\varepsilon)} n$, (iii) $h \leq c_0 \cdot n$, and (iv) $\mathsf{SNR} \geq c_1$. Setting λ_n in (3) as $c_2\sigma\sqrt{\log p/n}$, we conclude*

$$\left\|\mathbf{B}^\natural - \mathbf{B}^{\text{opt}}\right\|_{\text{F}} \lesssim \sigma\sqrt{\frac{mk\log p}{n}}$$

holds with probability exceeding $1 - c_3 n^{-c_4} - c_5 p^{-c_6}$.

Its proof is a simple combination of Theorem 3 and the previous results in Candès et al. (2006); Candes et al. (2006); Donoho (2006); Wainwright (2019). However, we observe a new phenomenon: the reconstruction error $\left\|\mathbf{B}^{\text{opt}} - \mathbf{B}^\natural\right\|_{\text{F}}$ is affected by the signal energy $\left\|\mathbf{B}^\natural\right\|_{\text{F}}$ on top of the sensing noise σ^2. Moreover, we should mention that the above difference will still exist even when we replace (3) with other estimators, say, Dantzig estimator.

At last, we can relax the constraints on $\text{srank}(\mathbf{B}^\natural)$ and n with an additional constraint on $\inf_j {}^{\max_i |\mathbf{B}^\natural_{i,j}|}/_{\|\mathbf{B}^\natural_{:,j}\|_2}$, the same as that in Corollary 1.

4.2 Proof outline

Due to the space limit, we only give a sketch of our proof ideas and leave the technical details to the supplementary material. Denote $\widehat{\mathbf{B}} = (n-h)^{-1}\mathbf{X}^\top\mathbf{Y}$, our goal is to show

$$\left\langle \mathbf{\Pi}^\natural\mathbf{X}\mathbf{B}^\natural + \mathbf{W}, \mathbf{\Pi}\mathbf{X}\cdot\mathsf{thres}(\widehat{\mathbf{B}})\right\rangle \geq \left\langle \mathbf{\Pi}^\natural\mathbf{X}\mathbf{B}^\natural + \mathbf{W}, \mathbf{\Pi}^\natural\mathbf{X}\cdot\mathsf{thres}(\widehat{\mathbf{B}})\right\rangle, \tag{4}$$

holds with high probability under the settings in Theorem 3.

Same as Zhang & Li (2020), our analysis faces the difficulties brought by (i) combinatorial nature of the problem and (ii) high-order moments of sub-Gaussian random variables. On top of these challenges, we are subject to insufficient samples, i.e., $n \ll p$. These issues are tackled by a combination of relaxation and a tailored leave-one-out analysis.

First, we relax (4) to its sufficient condition reading as

$$\langle \mathbf{Y}_{i,:}, \mathsf{thres}(\widehat{\mathbf{B}})^\top \mathbf{X}_{\pi^\natural(i),:}\rangle \geq \langle \mathbf{Y}_{i,:}, \mathsf{thres}(\widehat{\mathbf{B}})^\top \mathbf{X}_{j,:}\rangle, \ \ \forall\ j \neq \pi^\natural(i). \tag{5}$$

Then, we decompose the correlation between $\mathsf{thres}(\widehat{\mathbf{B}})$ and rows $\mathbf{X}_{\pi^\natural(i),:}$ and $\mathbf{X}_{j,:}$. This is accomplish by a tailored leave-one-out analysis. The basic idea is the same as the previous works using leave-one-out analysis, say Chen et al. (2020); Sur et al. (2019); El Karoui (2013; 2018); Cai et al. (2021): we replace the correlated components in $\mathsf{thres}(\widehat{\mathbf{B}})$ with its i.i.d. substitutes. However, our construction method has the following characteristics

- The number of replaced rows in our method varies for different pair of row indices. Meanwhile, the replacement number is fixed in the above mentioned works. For a better explanation, we refer the readers to our constructed *leave-one-out* samples $\widetilde{\mathbf{B}}_{(\cdot)}$ in the appendix.
- We modify the operator $\mathsf{thres}(\cdot)$ in approximating $\mathsf{thres}(\widehat{\mathbf{B}})$. While the previous works usually keep the operator $\mathsf{thres}(\cdot)$ intact, we approximate it with the operator $(\cdot)_{\mathsf{imax}}$, which denotes the positions of non-zero elements in $\mathsf{thres}(\widehat{\mathbf{B}})$. In other words, the positions of the non-zero elements we keep in the *leave-one-out* samples $\widetilde{\mathbf{B}}^{(\cdot)}$ is determined by $\mathsf{thres}(\widehat{\mathbf{B}})$ rather than $\mathsf{thres}(\widetilde{\mathbf{B}}^{(\cdot)})$. In

our analysis, we can see this step is essential in controlling the approximation errors. Otherwise, the approximation error can be considerably large, since $\mathsf{thres}(\widehat{\mathbf{B}})$ may not share the same support set with $\mathsf{thres}(\widetilde{\mathbf{B}}^{(\cdot)})$, let alone their ℓ_2 differences.

In analyzing the probability of (5), we find the key challenge to be lower bounding the inner product $\langle \mathbf{B}^\natural, \widehat{\mathbf{B}}_{\mathsf{imax}} \rangle$, where $(\cdot)_{\mathsf{imax}}$ denotes the indices of the non-zero elements in $\mathsf{thres}(\widehat{\mathbf{B}})$. [5] Expressing $\langle \mathbf{B}^\natural, \widehat{\mathbf{B}}_{\mathsf{imax}} \rangle$ as $\sum_{\boldsymbol{\beta}^\natural \in \{\mathbf{B}^\natural_{:,\ell}\}_{1 \le \ell \le m}} \langle \boldsymbol{\beta}^\natural, \widehat{\boldsymbol{\beta}}_{\mathsf{imax}} \rangle$, where $\widehat{\boldsymbol{\beta}}$ is the corresponding column in $\widehat{\mathbf{B}}$, we have

$$\langle \boldsymbol{\beta}^\natural, \widehat{\boldsymbol{\beta}}_{\mathsf{imax}} \rangle \ge |\boldsymbol{\beta}^\natural_{\mathsf{imax}}|^2 - \max_i |\boldsymbol{\beta}^\natural_i| \cdot \|\boldsymbol{\beta}^\natural - \widehat{\boldsymbol{\beta}}\|_\infty,$$

and separately (i) upper-bound $\max_i |\boldsymbol{\beta}^\natural_i|$ and $\|\boldsymbol{\beta}^\natural - \widehat{\boldsymbol{\beta}}\|_\infty$; and (ii) lower-bound $|\boldsymbol{\beta}^\natural_{\mathsf{imax}}|$. Part (i) is quite standard and part (ii) lies in analyzing the event $|\widehat{\boldsymbol{\beta}}_{\mathsf{imax}}| \ge \max_j |\widehat{\boldsymbol{\beta}}_j|$, which is due to the definition of $(\cdot)_{\mathsf{imax}}$. The technical details can be found in the supplementary material.

4.3 Computational complexities

Before proceeding to numerical experiments, we would like to mention our estimator is computationally optimal: in the first step for permutation recovery, our estimator only requires one matrix multiplication and thresholding operation on top of the operations in the oracle estimator; in the second step for sparse signal recovery, our estimator needs one additional matrix multiplication when compared with the work without permutation.

5 Numerical Results

This section presents the numerical experiments verifying our main theorem, to put more specifically, Theorem 3; we would like to prove the correct permutation can be obtained, i.e., $\{\mathbf{\Pi}^{\mathrm{opt}} = \mathbf{\Pi}^\natural\}$, with $n \ll p$ and $\mathsf{SNR} \ge c$.

Experiment setting with Gaussian distribution. We let $\mathbf{X}_{ij} \overset{\text{i.i.d}}{\sim} \mathsf{N}(0,1)$ and pick the sample number n to be $\{100, 150\}$ and set $h = n/4$. We vary the signal length p to be $\{500, 600\}$. Then we set the sparsity number k within the region $\{10, 15, 20\}$. And the stable rank $\mathrm{srank}(\mathbf{B}^\natural)$ is within the range $\{150, 200, 250\}$. The corresponding simulation results can be found in Figure 3.

Discussion w.r.t n/p. First, we confirm our theory such that correct permutation can be obtained with insufficient samples, i.e., $n \ll p$. In addition, we notice that the permutation recovery becomes easy with a larger n/p: the first row in Figure 3 is with $n/p = 0.2$ while the second row is with $n/p = 0/25$. We can verify the SNR for the correct permutation recovery is smaller in the second row than that for the first row. However, we should stress that this conclusion may not hold provided that $\mathrm{srank}(\mathbf{B}^\natural)$ is not sufficiently large. More details are referred to Figure 2.

Discussion w.r.t. sparsity number k. We vary the sparsity number k to be within $\{10, 15, 20\}$. We conclude a large sparsity number k makes the permutation recovery more difficult. For example, consider the case when $(n, p, \mathrm{srank}(\mathbf{B}^\natural)) = (100, 500, 200)$. When $k = 10$, correct permutation requires $\mathsf{SNR} \ge 1.4$; when $k = 15$, correct permutation needs $\mathsf{SNR} \ge 2.2$; and when $k = 20$, correct permutation requires $\mathsf{SNR} \ge 4$. The same conclusion holds for other cases as well.

Experiment setting with sub-Gaussian distribution. In addition to the Gaussian setting, we also evaluate our estimator's performance when $\mathbf{X}_{ij}$ being sub-Gaussian. Here, we pick $\mathbf{X}_{ij}$ to be i.i.d. Rademacher random variables such that $\mathbb{P}(\mathbf{X}_{ij} = \pm 1) = 1/2$. The corresponding results are put in Figure 4, from which we can observe a similar pattern as that of the Gaussian setting.

[5] Strictly speaking, our goal is to lower-bound $\langle \mathbf{B}^\natural, (\widetilde{\mathbf{B}}^{(\cdot)})_{\mathsf{imax}} \rangle$. For the sake of presentation, we write it as $\langle \mathbf{B}^\natural, \widehat{\mathbf{B}}_{\mathsf{imax}} \rangle$.

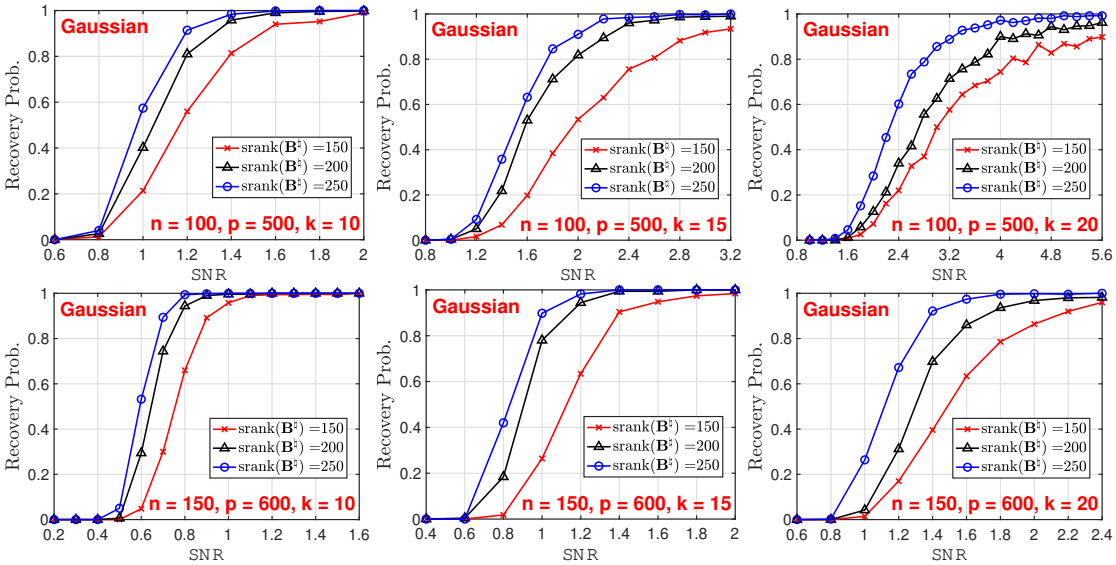

Figure 3: Simulated recovery rate $\mathbb{P}(\widehat{\mathbf{\Pi}} = \mathbf{\Pi}^{\natural})$ with $n = \{100, 150\}$, $p \in \{500, 600\}$, $h \in \{25, 37\}$, and $\mathbf{X}_{ij} \overset{\text{i.i.d}}{\sim} \mathsf{N}(0, 1)$, with respect to SNR.

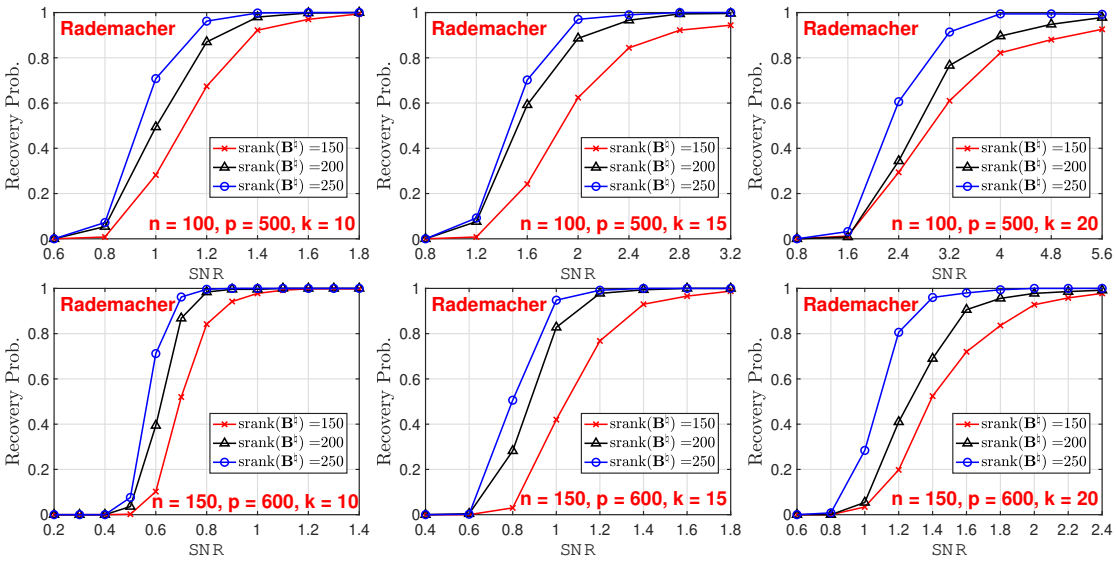

Figure 4: Simulated recovery rate $\mathbb{P}(\widehat{\mathbf{\Pi}} = \mathbf{\Pi}^{\natural})$ with $n = \{100, 150\}$, $p \in \{500, 600\}$, $h \in \{25, 37\}$, and $\mathbf{X}_{ij} \overset{\text{i.i.d}}{\sim}$ Rademacher, i.e., $\mathbb{P}(\mathbf{X}_{ij} = -1) = \mathbb{P}(\mathbf{X}_{ij} = 1) = 1/2$, w.r.t. SNR.

6 Conclusion

In this paper, we considered the unlabeled sparse recovery with multiple measurements for the first time. To begin with, we investigated the lower bounds on the sample number n and the SNR. Furthermore, we proposed a simple yet effective estimator, which restores the permutation matrix via a linear assignment problem. In addition, we proved that our estimator can obtain the correct correspondence information when SNR is above certain positive constant. In analyzing our estimator's performance, we developed a tailored version of the leave-one-out technique, which involves an adaptive number of replaced elements and simultaneous modification of the threshold operator. Moreover, we provided numerical experiments to confirm our claims and showed our estimator can reliably reconstruct the permutation matrix when the entries $\mathbf{X}_{ij}$ are sub-Gaussian random variables.

REFERENCES

Abubakar Abid, Ada Poon, and James Zou. Linear regression with shuffled labels. *arXiv preprint arXiv:1705.01342*, 2017.

Zhidong Bai and Tailen Hsing. The broken sample problem. *Probab. Theory Relat. Fields*, 131(4): 528–552, 2005.

Thomas Blumensath and Mike E Davies. Iterative hard thresholding for compressed sensing. *Applied and computational harmonic analysis*, 27(3):265–274, 2009.

Changxiao Cai, Gen Li, Yuejie Chi, H Vincent Poor, and Yuxin Chen. Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. *The Annals of Statistics*, 49 (2):944–967, 2021.

Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52 (2):489–509, 2006.

Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223, 2006.

Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, and Yuling Yan. Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM journal on optimization*, 30(4):3098–3121, 2020.

Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. John Wiley & Sons, 2nd edition, 2012.

Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Struct. Algorithms*, 22(1):60–65, 2003.

Morris H. DeGroot and Prem K. Goel. The matching problem for multivariate normal data. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*, 38(1):14–29, 1976.

Morris H. DeGroot and Prem K. Goel. Estimation of the correlation coefficient from a broken random sample. *The Annals of Statistics*, 8(2):264–278, 03 1980.

David L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.

Noureddine El Karoui. Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: rigorous results. *arXiv preprint arXiv:1311.2445*, 2013.

Noureddine El Karoui. On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. *Probab. Theory Relat. Fields*, 170(1-2): 95–175, 2018.

Valentin Emiya, Antoine Bonnefoy, Laurent Daudet, and Rémi Gribonval. Compressed sensing with unknown sensor permutation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1040–1044, Florence, Italy, 2014.

Simon Foucart. Hard thresholding pursuit: an algorithm for compressive sensing. *SIAM Journal on numerical analysis*, 49(6):2543–2563, 2011.

Prem K. Goel. On re-pairing observations in a broken random sample. *The Annals of Statistics*, 3(6): 1364–1369, 11 1975.

Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Johns Hopkins University Press Baltimore, 4th edition, 2013.

Saeid Haghighatshoar and Giuseppe Caire. Signal recovery from unlabeled samples. *IEEE Trans. Signal Process.*, 66(5):1242–1257, 2018.

Daniel J. Hsu, Kevin Shi, and Xiaorui Sun. Linear regression without correspondence. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 1531–1540, Long Beach, CA, 2017.

Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi. Low-rank matrix completion using alternating minimization. In *Symposium on Theory of Computing Conference (STOC)*, pp. 665–674, Palo Alto, CA, 2013.

Xingguo Li, Tuo Zhao, Raman Arora, Han Liu, and Jarvis D. Haupt. Stochastic variance reduced optimization for nonconvex sparse learning. In *Proceedings of the 33nd International Conference on Machine Learning (ICML)*, pp. 917–925, New York City, NY, 2016.

Ashwin Pananjady, Martin J. Wainwright, and Thomas A. Courtade. Linear regression with shuffled data: Statistical and computational limits of permutation recovery. *IEEE Trans. Inf. Theory*, 64(5): 3286–3300, 2018.

Martin Slawski and Emanuel Ben-David. Linear regression with sparsely permuted data. *Electron. J. Statist.*, 13(1):1–36, 2019.

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.

Pragya Sur, Yuxin Chen, and Emmanuel J Candès. The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square. *Probab. Theory Relat. Fields*, 175(1-2): 487–558, 2019.

Joel A. Tropp. An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.*, 8 (1-2):1–230, 2015.

Manolis C. Tsakiris, Liangzu Peng, Aldo Conca, Laurent Kneip, Yuanming Shi, and Hayoung Choi. An algebraic-geometric approach for linear regression without correspondences. *IEEE Trans. Inf. Theory*, 66(8):5130–5144, 2020.

Jayakrishnan Unnikrishnan, Saeid Haghighatshoar, and Martin Vetterli. Unlabeled sensing: Solving a linear system with unordered measurements. In *Proceedings of the 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 786–793, Monticello, IL, 2015.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

Martin J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.

Xiaotong Yuan, Ping Li, and Tong Zhang. Gradient hard thresholding pursuit for sparsity-constrained optimization. In *Proceedings of the 31th International Conference on Machine Learning (ICML)*, pp. 127–135, Beijing, China, 2014.

Hang Zhang and Ping Li. Optimal estimator for unlabeled linear regression. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pp. 11153–11162, Virtual Event, 2020.

Hang Zhang and Ping Li. Sparse recovery with shuffled labels: Statistical limits and practical estimators. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pp. 1760–1765, Melbourne, Australia, 2021.

Hang Zhang, Martin Slawski, and Ping Li. The benefits of diversity: Permutation recovery in unlabeled sensing from multiple measurement vectors. *IEEE Trans. Inf. Theory*, 68(4):2509–2529, 2022.

Huishuai Zhang, Yuejie Chi, and Yingbin Liang. Median-truncated nonconvex approach for phase retrieval with outliers. *IEEE Transactions on information Theory*, 64(11):7287–7310, 2018.