Greed is good: correspondence recovery for unlabeled linear regression

Hang Zhang, Ping Li

Cognitive Computing Lab
Baidu Research
10900 NE 8th St, Bellevue, WA 98004, USA
{zhanghanghitomi, pingli98}@gmail.com

Abstract

We consider the unlabeled linear regression reading as $\mathbf{Y} = \mathbf{\Pi}^*\mathbf{X}\mathbf{B}^* + \mathbf{W}$, where $\mathbf{\Pi}^*, \mathbf{B}^*$ and $\mathbf{W}$ represents missing (or incomplete) correspondence information, signals, and additive noise, respectively. Our goal is to perform data alignment between $\mathbf{Y}$ and $\mathbf{X}$, or equivalently, reconstruct the correspondence information encoded by $\mathbf{\Pi}^*$. Based on whether signal $\mathbf{B}^*$ is given a prior, we separately propose two greedy-selection-based estimators, which both reach the mini-max optimality. Compared with previous works, our work (i) supports partial recovery of the correspondence information; and (ii) applies to a general matrix family rather than the permutation matrices, to put more specifically, selection matrices, where multiple rows of $\mathbf{X}$ can correspond to the same row in $\mathbf{Y}$. Moreover, numerical experiments are provided to corroborate our claims.

1 INTRODUCTION

Starting under the name "broken sample" problem in 1970s [DeGroot and Goel, 1976, 1980, Goel, 1975], data alignment has received increasing attention nowadays due to its wide spectrum of applications, which span from computer vision to curve registration to natural language processing to data privacy to linkage record [Unnikrishnan et al., 2015, Pananjady et al., 2018, Hsu et al., 2017, Slawski and Ben-David, 2019, Dokmanic, 2019, Zhang et al., 2019, Slawski et al., 2020, Zhang and Li, 2020, Tsakiris et al., 2020]. Among the numerous applications, two prominent examples are the linkage attack and database merging.

In linkage attacks, intruders aim at the disclosure of sensitive data by using public data. This can be viewed as the inverse problem of data de-anonymization. Usually, these attacks involve direct comparison between the sensitive data and the public data, where their correspondence information is formulated as an unknown selection matrix to be reconstructed. In the task of database merging, the goal is to merge multiple databases, which contain data of the same identity, into one comprehensive database. In practice, these databases may not be well aligned due to the data formatting and data quality issues. How to reconstruct the correspondence information across the databases and properly align their data constitute a technical challenge. For more applications, we refer the interested readers to Pananjady et al. [2018], Slawski and Ben-David [2019], Unnikrishnan et al. [2015], Slawski et al. [2020], Zhang et al. [2022].

In this paper, we formulate the above-mentioned problem as an unlabeled linear regression reading as

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

where $\mathbf{Y} \in \mathbb{R}^{n_1\times m}$ denotes the sensing results, $\mathbf{\Pi}^* \in \{0,1\}^{n_1\times n_2}$ is the (unknown) selection matrix, $\mathbf{X} \in \mathbb{R}^{n_2\times p}$ is the sensing matrix, $\mathbf{B}^* \in \mathbb{R}^{p\times m}$ represents the signal of interest, and $\mathbf{W} \in \mathbb{R}^{n_1\times m}$ is the additive sensing noise. Our goal is to reconstruct the correct correspondence of rows, which are sabotaged by the unknown matrix $\mathbf{\Pi}^*$. To start with, we briefly review the previous works.

Related work. As mentioned before, research on regression without correspondence has a long history that can at least date back to 1970s dubbed "broken sample problem" [DeGroot and Goel, 1976, 1980, Goel, 1975, Bai and Hsing, 2005]. Due to their large volume, we restrict ourselves to recent work on this area, which begins with Unnikrishnan et al. [2015] and a noiseless setting ($\mathbf{W} = \mathbf{0}$) with single observation ($m = 1$) is adopted. Provided that entries $\mathbf{X}_{ij}$ in the sensing matrix are drawn from continuous distributions, they establish the necessary condition $n \geq 2p$ for correct correspondence reconstruction. Similar results can also be found in Dokmanic [2019], Tsakiris et al. [2020]. Later, Pananjady et al. [2018], Slawski and Ben-David [2019] extend the noiseless setting to a noisy setting and discover a phase transition phenomenon when the *signal-to-noise-ratio* (SNR) exceeds a certain threshold. Apart from the above

Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023), PMLR 216:2509–2518.

work, there are other works focusing on the single observation model, i.e., $m = 1$ [Hsu et al., 2017, Haghighatshoar and Caire, 2018, Tsakiris and Peng, 2019, Peng and Tsakiris, 2020, Zhang and Li, 2021]. Apart from the linear sensing model, there are also some research on the generalized linear sensing relation such as Fang and Li [2022]. Since the single observation model is not our focus, we only leave their names without further discussion.

Then we discuss the research on the multiple observation model, i.e., $m \gg 1$. Albeit there are some previous works, we find the theoretical analysis first appears in Pananjady et al. [2017], whose focus is to denoise the product $\mathbf{\Pi}^*\mathbf{X}\mathbf{B}^*$. Adopting vastly different strategies, Slawski et al. [2020], Zhang et al. [2019], Zhang and Li [2020] later propose three estimators for the correspondence recovery. Zhang et al. [2019] designs the estimator to approximately compute the *maximum likelihood* (ML) estimator. In Slawski et al. [2020], they take the viewpoint of robustness and view the permuted rows as outliers. The basic idea of their estimator is to first detect the outliers and then reconstruct the correspondence based on these identified outliers. The same idea is also used in Slawski et al. [2019]. Meanwhile, Zhang and Li [2020] tackles the problem from the perspective of non-convex optimization. In contrast to our paper, all these works need to reconstruct the whole selection matrix even if only one row's correspondence is desired. Notably, our estimators designed for recovering one single row's correspondence can automatically yield estimators to recover the whole selection matrix, which are completed by iterative applications to each row. For a detailed comparison, we defer it to Table 1 as the technical details yet to be presented are inevitably involved.

In addition to the literature on unlabeled linear regression, other work on graph matching, generalized linear regression without correspondence, linear assignment problem, quadratic assignment problem are also related to ours. Since their connections are in a rather loose manner, we only mention a few of their names without further discussion [Mézard and Parisi, 1986, 1985, Caracciolo et al., 2017, Malatesta et al., 2019, Chertkov et al., 2010, Semerjian et al., 2020, Koopmans and Beckmann, 1957, Umeyama, 1988].

Our contribution can be summarized as follows:

- We propose two optimal estimators for the correspondence recovery with selection matrix. Compared with **the previous work focusing on the permutation matrix**, our estimators apply to a broader class of matrices and do not enforce bijection between $\mathbf{Y}$ and $\mathbf{\Pi}^*\mathbf{X}\mathbf{B}^*$.
- We first propose estimators to recover the correspondence for one single row. In this paper, we separately consider the oracle case, where $\mathbf{B}^*$ is known, and the non-oracle case, where $\mathbf{B}^*$ is unknown. Numerical experiments suggest both estimators can reliably reconstruct the correspondence under certain conditions.
- We provide a theoretical guarantee of our estimators' performance. In this part, the major technical difficulty comes from the heavy tails inherent in the non-oracle estimator. To handle such an issue, we tailor the *leave-one-out* technique [Karoui et al., 2013, Karoui, 2018, Chen et al., 2020, Sur et al., 2019]. Focusing on the high stable rank regime, i.e., $\mathsf{srank}(\mathbf{B}^*) \gg \log^4 n$, we show that reliable correspondence recovery can be guaranteed for both oracle and non-oracle case once SNR exceeds certain positive constants.

Notations. We denote c, c_0, and $c^{'}$ as some fixed positive constants. We write $a \lesssim b$ if there exists some positive constant c such that $a \leq cb$. Similarly, we define $a \gtrsim b$. Provided $a \lesssim b$ and $b \lesssim a$ hold simultaneously, we write $a \asymp b$ to indicate a and b are of the same order. In addition, we denote the inner product between two vectors (and matrices) as $\langle \cdot, \cdot \rangle$. For an arbitrary vector $\boldsymbol{v}$, we denote its Euclidean norm as $\|\boldsymbol{v}\|_2$.

For an arbitrary matrix $\mathbf{M}$, we define its stable rank $\mathsf{srank}(\mathbf{M})$ as $\|\mathbf{M}\|_{\mathrm{F}}^2/\|\mathbf{M}\|_{\mathrm{OP}}^2$, where $\|\cdot\|_{\mathrm{F}}$ and $\|\cdot\|_{\mathrm{OP}}$ denote the Frobenius norm and operator norm, respectively, and their definitions can be found in Horn and Johnson [1990]. For an arbitrary row index i ($1 \leq i \leq n$), we define $\pi(i)$ as the correspondence index for i associated with the selection matrix $\mathbf{\Pi}$. For the ground-truth selection matrix $\mathbf{\Pi}^*$, we denote $\pi^*(i)$ as the correct correspondence index for i. Moreover, we define the *signal-to-noise-ratio* (SNR) as $\|\mathbf{B}^*\|_{\mathrm{F}}^2/(m\sigma^2)$, where $\mathbf{B}^* \in \mathbb{R}^{p\times m}$ denotes the signal and σ^2 represents the variance of the sensing noise.

2 PROBLEM SETTING

This section starts with a formal restatement of the problem

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

where $\mathbf{Y} \in \mathbb{R}^{n_1\times m}$ denotes the sensing result, $\mathbf{\Pi}^* \in \mathbb{R}^{n_1\times n_2}$ is the (unknown) selection matrix such that $\mathbf{\Pi}^* = \{0,1\}^{n_1\times n_2}, \sum_j \mathbf{\Pi}^*_{ij} = 1$, $\mathbf{X} \in \mathbb{R}^{n_2\times p}$ is the sensing matrix with each entry being i.i.d. standard normal random variable, $\mathbf{B}^* \in \mathbb{R}^{p\times m}$ is the signal of interests, and $\mathbf{W} \in \mathbb{R}^{n_1\times m}$ denotes the additive sensing noise such that its entries are i.i.d. Gaussian distributed random variables with zero mean and σ^2 variance, namely, $\mathbf{W}_{ij} \overset{\text{i.i.d}}{\sim} \mathsf{N}(0,\sigma^2)$. For the clarify of presentation, we assume $n_1 = n_2 = n$. Notice that this condition is not enforced by the selection matrix. In fact, our current analysis can be generalized to the $n_1 \neq n_2$ case effortlessly.

Compared with previous works [Pananjady et al., 2018, Hsu et al., 2017, Slawski and Ben-David, 2019, Zhang et al., 2019, Slawski et al., 2020, Zhang and Li, 2020, Zhang et al., 2022], our work has the following two noticeable characteristics:

Table 1: Comparison with the prior art. All results are w.r.t. the exact correspondence recovery in the non-oracle case ($\mathbf{B}^*$ is unknown). Besides, these results are presented in their best orders, which only hold in certain regimes. Notation $\mathsf{SNR}_{\min}$, $n_{\min}$, and $h_{\max}$ denotes the minimum required SNR, minimum sample number, and maximum allowed number of mismatched rows, respectively. Moreover, the logarithmic term is omitted in the notation $\widetilde{\Omega}(\cdot)$ and $\widetilde{O}(\cdot)$. The notation $r(\cdot)$ denotes the rank of the corresponding matrix.

	$\mathsf{SNR}_{\min}$ $(\geq)$		${n_{\min}}/{p}$ $(\geq)$		${h_{\max}}/{n}$ $(\leq)$		Partial Recover.	Select. Matrix
	$m=1$	$m \gg 1$	$m=1$	$m \gg 1$	$m=1$	$m \gg 1$		
[Pananjady et al., 2018]	$\widetilde{\Omega}(n^c)$		$\widetilde{\Omega}(1)$		$\widetilde{O}(1)$			
[Hsu et al., 2017]	∞		$\widetilde{\Omega}(1)$		$\widetilde{O}(1)$			
[Slawski and Ben-David, 2019]	$\widetilde{\Omega}(n^c)$		$\widetilde{\Omega}(1)$		$\widetilde{O}(\log^{-1} n)$			
[Zhang et al., 2019]		$\widetilde{\Omega}(n^{\frac{c}{\mathsf{srank}(\mathbf{B}^*)}})$		$\widetilde{\Omega}(1)$		$\widetilde{O}\left(\log^{-1} r(\mathbf{B}^*)\right)$		
[Slawski et al., 2020]		$\widetilde{\Omega}(n^{\frac{c}{\mathsf{srank}(\mathbf{B}^*)}})$		$\widetilde{\Omega}(p)$		$\widetilde{O}(\log^{-1} n)$		✓
[Zhang and Li, 2020]	$\widetilde{\Omega}(n^c)$	$\widetilde{\Omega}(n^{\frac{c}{\mathsf{srank}(\mathbf{B}^*)}})$	$\widetilde{\Omega}(1)$	$\widetilde{\Omega}(\sqrt{p})$	$\widetilde{O}(1)$	$\widetilde{O}(1)$		
Our Estimator		$\widetilde{\Omega}(n^{\frac{c}{\mathsf{srank}(\mathbf{B}^*)}})$		$\widetilde{\Omega}(1)$		$\widetilde{O}(1)$	✓	✓

- We can support partial correspondence recovery. In certain applications, simultaneously recovering all correspondence is unnecessary. One particular example is merging databases while only a small proportion of the correspondence information is desired. While previous works such as [Pananjady et al., 2018, Hsu et al., 2017, Slawski and Ben-David, 2019, Zhang et al., 2019, Slawski et al., 2020, Zhang and Li, 2020] all focus on reconstructing the whole permutation (selection) matrix, our work is the first work that can support partial correspondence recovery. This can lead to significant computational savings, especially when the sample number n is sufficiently large.
- Our estimators apply to a general family of matrices, to put it more specifically, selection matrices, where multiple rows in $\mathbf{\Pi}^*\mathbf{X}\mathbf{B}^*$ can correspond to the same row in $\mathbf{Y}$, rather than the permutation matrices, where the correspondences between $\mathbf{\Pi}^*\mathbf{X}\mathbf{B}^*$ and $\mathbf{Y}$ are bijective. [1] The only work focusing on the selection matrix is Slawski et al. [2020] however it (i) only allows a limited number of mismatched rows, (ii) requires a much larger sample size, and (iii) is with a much higher computational cost.

A detailed comparison between our work and previous works is put in Table 1. [2]

2.1 MINI-MAX LOWER BOUND

First, we present the mini-max lower bound for the permuted linear regression, which is a subset of our problem, as the baseline for comparison.

[1]The family of permutation matrices is a subset of the selection matrices. Compared with permutation matrices, selection matrices (i) are not necessarily a square matrix, or equivalently, $n_1 \neq n_2$; (ii) allow multiple rows to have the same correspondence.

[2]Notice that an algorithm with a larger $h_{\max}$ covers a boarder class of selection matrices, as more mismatched rows are allowed.

Theorem 1 (Theorem 1 in Zhang et al. [2019])**.** *Provided that* $\log\det\left(\mathbf{I} + {\mathbf{B}^{*\top}\mathbf{B}^*}/{\sigma^2}\right) < \frac{\log n! - 2}{n}$*, we conclude*

$$\inf_{\widehat{\mathbf{\Pi}}} \sup_{\mathbf{\Pi}^*} \mathbb{P}_{\mathbf{X},\mathbf{W}}(\widehat{\mathbf{\Pi}} \neq \mathbf{\Pi}^*) \geq \frac{1}{2}, \tag{1}$$

where the probability $\mathbb{P}_{\mathbf{X},\mathbf{W}}(\cdot)$ *is w.r.t.* $\mathbf{X}$ *and* $\mathbf{W}$*, and the infimum is over all possible permutation estimators* $\widehat{\mathbf{\Pi}}$.

According to the above theorem, we conclude that correct correspondence recovery requires $\log\det(\mathbf{I} + {\mathbf{B}^{*\top}\mathbf{B}^*}/{\sigma^2})$ to be at least of order $\log n$. With the relation such that $\log\det(\mathbf{I} + {\mathbf{B}^{*\top}\mathbf{B}^*}/{\sigma^2}) \simeq \mathsf{srank}(\mathbf{B}^*)\log(1 + \mathsf{SNR})$, we can approximately write the SNR requirement as $\log(1 + \mathsf{SNR}) \gtrsim \frac{n}{\mathsf{srank}(\mathbf{B}^*)}$, or equivalently, $\mathsf{SNR} \gtrsim c_0 n^{\frac{c_1}{\mathsf{srank}(\mathbf{B}^*)}} - 1$.

In the following context, we separately present partial correspondence recovery estimators for the oracle and non-oracle case, whose behaviors all match the mini-max lower bound in Theorem 1.

3 ORACLE CASE ESTIMATOR

As a warm-up example, this section considers the oracle scenario, where $\mathbf{B}^*$ is given a priori. In this scenario, it is well known that *maximum likelihood* (ML) estimator can be recast as a *linear assignment problem* (LAP) [Kuhn, 1955, Bertsekas and Castañón, 1992]. However, we have to solve the whole selection matrix even if we only need the correspondence of one single row.

To handle such an issue, we modify the LAP formulation and propose a greedy-selection-based estimator. A formal statement is put in Algorithm 1. Then we conclude

Theorem 2. *Consider the oracle case. Assume that* (i) $\mathsf{srank}(\mathbf{B}^*) \gg \log^2 n$*,* (ii) $n \geq 2p$*, and* (iii) $\mathsf{SNR} \geq c$*. For an arbitrary row index* i*, we conclude Algorithm 1 ob-*

Algorithm 1 Oracle greedy estimator for correspondence recovery.

Input: observation $\mathbf{Y}$, sensing matrix $\mathbf{X}$, and matrix $\mathbf{B}^*$.
Output: Reconstruct the correspondence $\widehat{\pi}(i)$ as

$$\widehat{\pi}(i) = \text{argmax}_j \left\langle \mathbf{Y}_{i,:}, \mathbf{B}^{*\top}\mathbf{X}_{j,:} \right\rangle,$$

where $\mathbf{Y}_{i,:}^\top$ denotes the ith row of the matrix $\mathbf{Y}$ and $\mathbf{X}_{j,:}^\top$ denotes the jth row of the matrix $\mathbf{X}$.

tains its correct correspondence, i.e., $\widehat{\pi}(i) = \pi^*(i)$*, with probability at least* $1 - c_0 n^{-c_1}$ *when* n *is sufficiently large.*

Invoking Theorem 2, we can prove that the ground-truth selection matrix $\mathbf{\Pi}^*$ can be correctly reconstructed once $\mathsf{SNR} \geq c$. Comparing our estimator in Algorithm 1 with the statistical lower bound in Theorem 1, we conclude that our greedy-selection-based estimator almost reaches the minimax optimality. In addition, we can automatically obtain an estimator for the entire selection matrix $\mathbf{\Pi}^*$, which is to iterative apply Algorithm 1 to each row.

Remark 3. *Notice that the assumption on the stable rank is quite common in permutation recovery literature [Slawski et al., 2020, Zhang et al., 2019, Zhang and Li, 2020]. Roughly speaking, it is used to describe the data diversity. For an arbitrary matrix* $\mathbf{B}^*$*, we have its stable rank* $\text{srank}(\mathbf{B}^*) = \text{rank}(\mathbf{B}^*) = \min(p, m)$ *when its energy are uniformly distributed among all eigenvalues. When its principal eigenvalue dominates the signal strength, in other words, the energy of the rest eigenvalues is negligible, we have* $\text{srank}(\mathbf{B}^*)$ *to be approximately one.*

3.1 PROOF OUTLINES

Denote the correct correspondence index associated with index i to be $\pi^*(i)$. To begin with, we notice that the correct correspondence $\pi^*(i)$ is obtained provided the following inequality holds for all indices j except $\pi^*(i)$, i.e.,

$$\begin{aligned} \left\|\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}\right\|_2^2 > &\left\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}, \mathbf{B}^{*\top}\mathbf{X}_{j,:}\right\rangle \\ &+ \left\langle \mathbf{W}_{i,:}, \mathbf{B}^{*\top}\left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^*(i),:}\right)\right\rangle, \end{aligned} \quad (2)$$

for all $j \neq \pi^*(i)$. This inequality is a restatement of the condition $\pi^*(i) = \text{argmax}_j \langle \mathbf{Y}_{i,:}, \mathbf{B}^{*\top}\mathbf{X}_{j,:}\rangle$.

Then, we separately prove the following relations hold with a high probability

- $\|\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}\|_2 \gtrsim \|\mathbf{B}^*\|_{\mathrm{F}}$;
- $\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}, \mathbf{B}^{*\top}\mathbf{X}_{j,:}\rangle \lesssim \log n \left\|\mathbf{B}^*\mathbf{B}^{*\top}\right\|_{\mathrm{F}}$;
- $\langle \mathbf{W}_{i,:}, \mathbf{B}^{*\top}(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^*(i),:})\rangle \lesssim \sigma \log n \|\mathbf{B}^*\|_{\mathrm{F}}$.

Afterward, we complete the proof by verifying (2), which proceeds as

$$\begin{aligned} \|\mathbf{B}^*\|_{\mathrm{F}}^2 &\overset{①}{\gtrsim} \frac{\log n}{\sqrt{\text{srank}(\mathbf{B}^*)}}\|\mathbf{B}^*\|_{\mathrm{F}}^2 + \sigma(\log n)\|\mathbf{B}^*\|_{\mathrm{F}} \\ &\overset{②}{=} \|\mathbf{B}^*\|_{\mathrm{F}}\|\mathbf{B}^*\|_{\mathrm{OP}} + \sigma(\log n)\|\mathbf{B}^*\|_{\mathrm{F}} \\ &\overset{③}{\geq} \left\|\mathbf{B}^*\mathbf{B}^{*\top}\right\|_{\mathrm{F}} + \sigma(\log n)\|\mathbf{B}^*\|_{\mathrm{F}}. \end{aligned}$$

In ①, we use the assumptions $\mathsf{SNR} \geq c$ and $\text{srank}(\mathbf{B}^*) \gg \log^2 n$, in ② we use the definition of $\text{srank}(\mathbf{B}^*)$, and in ③ we use the relation $\left\|\mathbf{B}^*\mathbf{B}^{*\top}\right\|_{\mathrm{F}} \leq \|\mathbf{B}^*\|_{\mathrm{OP}}\|\mathbf{B}^*\|_{\mathrm{F}}$. The technical details are left in the supplementary material.

Having presented the algorithm for the oracle case, in the subsequent section, we will move on to the non-oracle case.

4 NON-ORACLE CASE ESTIMATOR

This section designs an estimator for the non-oracle case, where information $\mathbf{B}^*$ is unavailable. As explained in Pananjady et al. [2018], even for the unlabeled linear regression, [3] reconstructing the correspondence is NP-hard for a general sensing matrix $\mathbf{X}$.

To reconstruct the correspondence information in polynomial time, we need to exploit the statistical properties of $\mathbf{X}$. Our design insight starts with the fact such that $\mathbb{E}\mathbf{X}^\top\mathbf{Y} = \mathbb{E}\mathbf{X}^\top\mathbf{\Pi}^*\mathbf{X}\mathbf{B}^* = (n-h)\mathbf{B}^*$, where h is the number of mismatched rows. This implies that we can obtain the direction of $\mathbf{B}^*$, i.e., $\mathbf{B}^*/\|\mathbf{B}^*\|_{\mathrm{F}}$, from the product $\mathbb{E}\mathbf{X}^\top\mathbf{Y}$. With the belief such that $\mathbf{X}^\top\mathbf{Y}^*$ should be close to $\mathbb{E}\mathbf{X}^\top\mathbf{Y}^*$, we would like to approximate the value of $\mathbf{B}^*$ by $\mathbf{X}^\top\mathbf{Y}$. The proposed algorithm is summarized in Algorithm 2. Numerical results show that this greedy selection estimator can restore the correct correspondence with high probability, even when SNR is with modest value.

Algorithm 2 Non-Oracle greedy estimator for correspondence recovery.

Input: observation $\mathbf{Y}$ and sensing matrix $\mathbf{X}$.
Output: Reconstruct the corrspondence $\widehat{\pi}(i)$ as

$$\widehat{\pi}(i) = \text{argmax}_j \left\langle \mathbf{Y}_{i,:}, \mathbf{Y}^\top\mathbf{X}\mathbf{X}_{j,:}\right\rangle,$$

where $\mathbf{Y}_{i,:}^\top$ denotes the ith row of the matrix $\mathbf{Y}$ and $\mathbf{X}_{j,:}^\top$ denotes the jth row of the matrix $\mathbf{X}$.

4.1 MAIN RESULTS FOR NON-ORACLE ESTIMATOR

Regarding its theoretical performance, we have that Algorithm 2 can yield the ground-truth correspondence $\pi^*(i)$

[3] $\mathbf{\Pi}^*$ is a permutation matrix rather than a selection matrix.

when $\mathsf{SNR} \geq c$ in certain regime. A formal statement is given as the following.

Theorem 4. *Consider the non-oracle case. Assume that* (i) $\mathrm{srank}(\mathbf{B}^*) \gg \log^4 n$, (ii) $n \gtrsim p \log^6 n$, (iii) $h \leq c_0 \cdot n$, *and* (iv) $\mathsf{SNR} \geq c$. *Then for an arbitrary row index* i, *we conclude Algorithm 2 obtains its correct correspondence, i.e.,* $\widehat{\pi}(i) = \pi^*(i)$, *with probability at least* $1 - c_0 \cdot p^{-c_1} - c_2 \cdot n^{-c_3}$ *when* n *and* p *are sufficiently large.*

Similar to the oracle case, we can design the algorithm for the whole selection matrix recovery by iteratively applying Algorithm 2 to each row.

Comparing with Theorem 1, we conclude Algorithm 2 reaches the minimax optimal convergence rate as the lower bound in Theorem 1 becomes

$$\Omega\left(n^{\frac{c}{\mathrm{srank}(\mathbf{B}^*)}}\right) = \Omega\left[\exp\left(\frac{c \log n}{\mathrm{srank}(\mathbf{B}^*)}\right)\right] \overset{①}{=} \Omega(1),$$

where in ① we use the assumption $\mathrm{srank}(\mathbf{B}^*) \gg \log^4 n$.

Remark 5. *Notice that we do not require most rows to be matched. In fact, we allow the maximum allowed number of mismatched rows to be in the same order of* n *(optimal order), i.e.,* $h_{\max} \asymp n$. *A numerical experiment (c.f. the bottom right of Figure 3) suggests our estimator can reconstruct the correspondence even when half of the rows are permuted.*

Comparison with prior work. In addition, we would like to compare our results with the previous works. We put a detailed comparison in Table 1 and briefly discuss our advantages over Zhang and Li [2020], Slawski et al. [2020], whose settings are of the most similarity to ours.

The most noticeable characteristic of our work is its unique ability to perform partial correspondence recovery. Moreover, we notice certain improvements in the number of required samples n and that of maximum allowed mismatched rows $h_{\max}$. For the minimum sample number, Slawski et al. [2020] require $n \geq \widetilde{\Omega}(p^2)$ and Zhang and Li [2020] require $n \geq \widetilde{\Omega}(p^{1.5})$. Meanwhile, our work improves it to $\widetilde{\Omega}(p)$. Turning to the maximum allowed mismatched rows, Slawski et al. [2020] limits it to $O(n/\log n)$, while our work reaches the optimal order, i.e., $\widetilde{O}(n)$. In addition, we have a slight advantage over Zhang and Li [2020] in terms of the SNR requirement. While they require $\mathsf{SNR} \gtrsim \log n$ even when the stable rank $\mathrm{srank}(\mathbf{B}^*)$ is sufficiently large, we reduce the requirement to $\mathsf{SNR} \geq \Omega(1)$.

The only loss is its more stringent requirement on $\mathrm{srank}(\mathbf{B}^*)$: we require the stable rank to satisfy $\mathrm{srank}(\mathbf{B}^*) \gg \log^4 n$ while most of the prior arts only require $\mathrm{srank}(\mathbf{B}^*) \gg \log n$. We conjecture the stringent requirement on $\mathrm{srank}(\mathbf{B}^*)$ is due to the proof artifacts rather than inherent in the estimator, which hopefully will be fixed with a more delicate analysis.

4.2 PROOF OUTLINES

In addition, we would like to discuss the proof technique, which is based on a modified version of the *leave-one-out* technique [Karoui et al., 2013, Karoui, 2018, Chen et al., 2020, Sur et al., 2019] and may serve independent technical interests.

Denote matrix $\widetilde{\mathbf{B}}$ as $(n-h)^{-1}\,\mathbf{X}^\top \mathbf{\Pi}^* \mathbf{X}\mathbf{B}^*$. The proof of Theorem 4 lies in showing

$$\begin{aligned}&\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:} + \mathbf{W}_{i,:}, (\widetilde{\mathbf{B}} + (n-h)^{-1}\mathbf{X}^\top\mathbf{W})^\top \mathbf{X}_{\pi^*(i),:}\rangle \\ \geq\ &\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:} + \mathbf{W}_{i,:}, (\widetilde{\mathbf{B}} + (n-h)^{-1}\mathbf{X}^\top\mathbf{W})^\top \mathbf{X}_{j,:}\rangle.\end{aligned} \tag{3}$$

For notational conciseness, we define

$$\begin{aligned}\mathsf{Term}_1 &= \left\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}, \mathbf{W}^\top\mathbf{X}\left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^*(i),:}\right)\right\rangle/(n-h);\\ \mathsf{Term}_2 &= \left\langle \mathbf{W}_{i,:}, \widetilde{\mathbf{B}}^\top\left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^*(i),:}\right)\right\rangle;\\ \mathsf{Term}_3 &= (n-h)^{-1}\left\langle \mathbf{W}_{i,:}, \mathbf{W}^\top\mathbf{X}\left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^*(i),:}\right)\right\rangle;\\ \mathsf{Term}_{\mathsf{tot}} &= \left\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}, \widetilde{\mathbf{B}}^\top\left(\mathbf{X}_{\pi^*(i),:} - \mathbf{X}_{j,:}\right)\right\rangle.\end{aligned}$$

Then (3) is equivalent to $\mathsf{Term}_{\mathsf{tot}} \geq \mathsf{Term}_1 + \mathsf{Term}_2 + \mathsf{Term}_3$. The technical challenge comes from the correlation between $\widetilde{\mathbf{B}}$ and the rows of $\mathbf{X}$. Take $\mathsf{Term}_{\mathsf{tot}}$ as an example. With the definition of $\widetilde{\mathbf{B}}$, we conclude this term involves Gaussian random variables of form $(\cdot)^4$, whose behavior are difficult to capture. Similar problems exist in the other three terms too.

To decouple the correlation thereof, we propose a tailored version of the leave-one-out technique [Karoui et al., 2013, Karoui, 2018, Chen et al., 2020, Sur et al., 2019]. Compared with the prior works using the **leave-one-out technique**, which creates independence by replacing fixed number of rows/columns with their i.i.d. substitutes, our method is rather adaptive and requires simultaneous replacement of rows ranging from two to four (The specific number is determined by the relations of $i, j, \pi^*(i)$, and $\pi^*(j)$).

The analysis can be divided into the following three stages.

- **Stage I.** First, we create i.i.d. copies $\mathbf{X}'_{i,:} \in \mathbb{R}^p$ of the rows in $\mathbf{X}$. Then, for each row π^*_i, we construct $\{\widetilde{\mathbf{B}}_{\setminus(\pi^*(i))}\}_{\pi^*(i)=1}^n$ as

$$\begin{aligned}\widetilde{\mathbf{B}}_{\setminus(\pi^*(i))} = (n-h)^{-1}\Bigg(&\sum_{\substack{k\neq\pi^*(i)\\ k\neq i}} \mathbf{X}_{\pi(k),:}\mathbf{X}_{k,:}^\top \\ &+ \sum_{\substack{k=i \text{ or}\\ k=\pi^*(i)}} \mathbf{X}'_{\pi(k),:}\mathbf{X}'^{\top}_{k,:}\Bigg)\mathbf{B}^*.\end{aligned}$$

  Easily, we can verify that $\widetilde{\mathbf{B}}_{\setminus(\pi^*(i))}$ is independent of the $\pi^*(i)$th row $\mathbf{X}_{\pi^*(i),:}$ as $\mathbf{X}_{\pi^*(i),:}$ is not contained in the perturbed sample $\widetilde{\mathbf{B}}_{\setminus(\pi^*(i))}$.

Then, we construct perturbed copies $\{\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}\}_{\pi^*(i)\neq j}$ for every possible pair $(\pi^*(i), j)$. In formulae:

$$\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)} = (n-h)^{-1}\Bigg(\sum_{\substack{k\neq \pi^*(i),j \\ \pi^*(k)\neq \pi^*(i),j}} \mathbf{X}_{\pi(k),:}\mathbf{X}_{k,:}^{\top} + \sum_{\substack{k=\pi^*(i) \text{ or } k=j \text{ or} \\ k=i \text{ or } \pi^*(k)=j}} \mathbf{X}'_{\pi(k),:}\mathbf{X}_{k,:}^{'\top}\Bigg)\mathbf{B}^*,$$

Same as above, we can verify that $\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}$ is independent of the rows $\mathbf{X}_{\pi^*(i),:}$ and $\mathbf{X}_{j,:}$, $1\leq \pi^*(i)\neq j\leq n$.

- **Stage II.** We analyze the separate behavior of $\mathsf{Term}_{\text{tot}}$, Term_1, Term_2, and Term_3. The difficulties incurred by the correlations between $\widetilde{\mathbf{B}}$ and the rows of $\mathbf{X}$ is tackled via the perturbed samples created above. To illustrate the procedure, we consider $\mathsf{Term}_{\text{tot}}$ without loss of generality. First, we rewrite it as

$$\mathsf{Term}_{\text{tot}} = \underbrace{\left\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}, \widetilde{\mathbf{B}}^{\top}_{\setminus(\pi^*(i),j)}\left(\mathbf{X}_{\pi^*(i),:}-\mathbf{X}_{j,:}\right)\right\rangle}_{\mathsf{Term}_{\text{tot},1}} + \underbrace{\left\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}, \left(\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}-\widetilde{\mathbf{B}}\right)^{\top}\left(\mathbf{X}_{\pi^*(i),:}-\mathbf{X}_{j,:}\right)\right\rangle}_{\mathsf{Term}_{\text{tot},2}}.$$

For $\mathsf{Term}_{\text{tot},1}$, we exploit the independence across the rows in $\mathbf{X}$. We first condition on the rows $\mathbf{X}_{k,:}$ ($k\neq \pi^*(i), j$) and can view it as a random variable determined by $\mathbf{X}_{\pi^*(i),:}$ and $\mathbf{X}_{j,:}$, which can be analyzed by the standard results such as Hanson-wright inequality [Vershynin, 2018], etc.

For $\mathsf{Term}_{\text{tot},2}$, we notice that $\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}$ only differs in $\widetilde{\mathbf{B}}$ in a finite number of terms, which are all related to $\mathbf{X}_{\pi^*(i),:}, \mathbf{X}_{j,:}$. This suggests that almost all terms in the difference $(\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}-\widetilde{\mathbf{B}})$ have been crossed out. Thus, we analyze $\mathsf{Term}_{\text{tot},2}$ by separately considering each nonzero term in $(\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}-\widetilde{\mathbf{B}})$ and complete the analysis of $\mathsf{Term}_{\text{tot}}$. For other terms, we follow a similar approach and can show

$$\begin{aligned}
&\mathsf{Term}_1 \leq c_1\sigma(\log n)^{5/2}\sqrt{\frac{p}{n}}\|\mathbf{B}^*\|_{\text{F}} \triangleq \Delta_1;\\
&\mathsf{Term}_2 \leq c_2\sigma(\log^2 n)\|\mathbf{B}^*\|_{\text{F}} \triangleq \Delta_2;\\
&\mathsf{Term}_3 \leq c_3\left[\frac{mp(\log n)^2\sigma^2}{n} + \sigma^2(\log n)^2\sqrt{\frac{mp}{n}}\right] \triangleq \Delta_3;\\
&\mathsf{Term}_{\text{tot}} \geq c_4\|\mathbf{B}^*\|_{\text{F}}^2 + (\log^2 n)(\log n^2 p^3)\sqrt{\frac{p}{n}}\|\mathbf{B}^*\|_{\text{F}}^2\\
&\qquad + \frac{p(\log n)^{3/2}}{n}\|\mathbf{B}^*\|_{\text{F}}^2 + \frac{(\log n)\|\mathbf{B}^*\|_{\text{F}}^2}{\sqrt{\text{srank}(\mathbf{B}^*)}}.
\end{aligned} \tag{4}$$

- **Stage III.** Under the settings of Theorem 4, we complete the proof by showing the right-hand side of $\mathsf{Term}_{\text{tot}}$ in (4) is no less than $\Delta_1+\Delta_2+\Delta_3$, which further leads to the relation such that

$$\mathsf{Term}_{\text{tot}} \geq \Delta_1+\Delta_2+\Delta_3,$$

holds with high probability.

Due to the space limit, we omit the technical details and defer them to the supplementary material. In the next section, we will present numerical experiments to verify our theorems.

5 NUMERICAL EXPERIMENTS

This section presents the numerical results to validate our claims. To evaluate the correspondence recovery, we adopt the recovery rate of the whole selection matrix, namely, $\mathbb{P}(\widehat{\mathbf{\Pi}}=\mathbf{\Pi}^*)$, rather than that of the rows, i.e., $\mathbb{P}(\pi^*(i)=\widehat{\pi}(i))$. The underlying reason is that one single row's correspondence may still be recovered correctly even when SNR is zero since numerous rows remain matched. However, due to the lack of signal strength, or equivalently, small SNR, the whole selection matrix cannot be reconstructed in such a case. This suggests $\mathbb{P}(\widehat{\mathbf{\Pi}}=\mathbf{\Pi}^*)$ may be a better quantity to measure the performance.

In the following context, we separately study the impact of signal length p, the ratio n/p between sample number and signal length, and the number of mismatched rows h on the correspondence recovery. Note that plots have varying X-axis as the phase transitions happen at different points.

5.1 IMPACT OF SIGNAL LENGTH p

We separately investigate the impact of signal length on the oracle estimator and non-oracle estimator when p increases.

Experiment setup. We set the signal length p to be $\{100, 150, 200\}$ and let $n=5p$. The number of mismatched rows h is set to be $n/4$ and the stable rank $\text{srank}(\mathbf{B}^*)$ is set to be $\{0.09n, 0.12n, 0.16n, 0.2n\}$.

Results discussion. The numerical results are put in Figure 1, from which we observe a sharp transition of the correspondence recovery once SNR exceeds a certain threshold. Comparing the thresholds for the oracle case, we find the thresholds for the non-oracle case are much larger.

In addition, we notice the threshold shrinks with the increasing n, p, and $\text{srank}(\mathbf{B}^*)$. Take the oracle estimator for example. When n increases from 500 to 1000, the corresponding phase transition threshold reduces from 1 to 0.7. A similar phenomenon also appears in the non-oracle case.

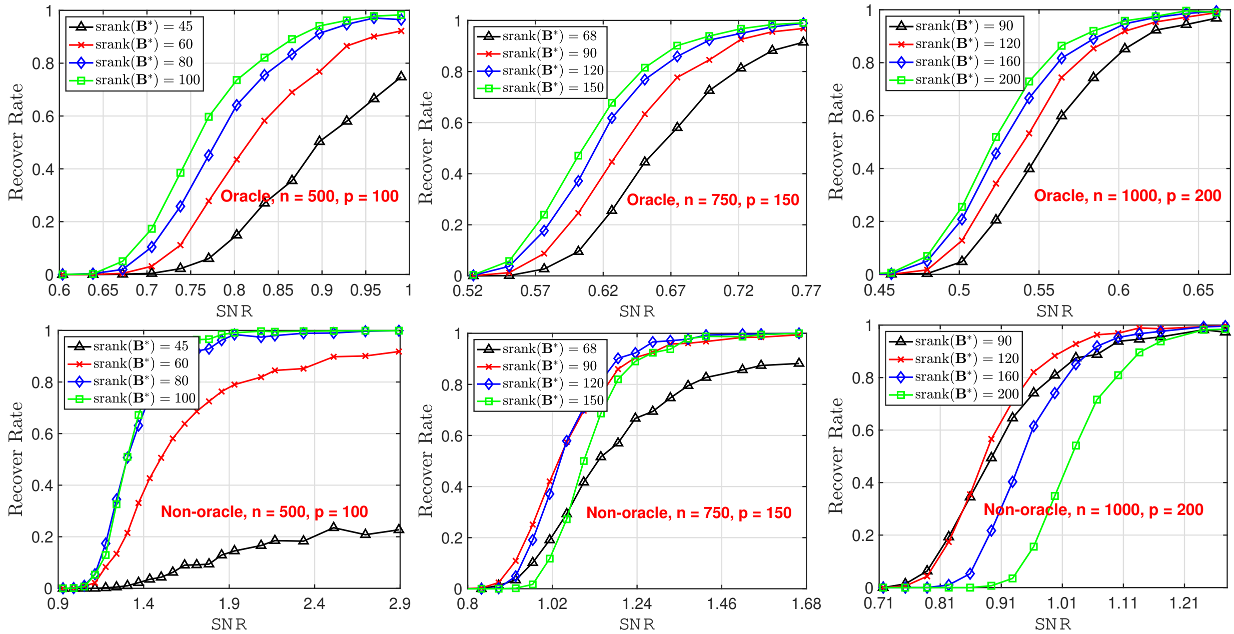

Figure 1: Impact of length p: the number of mismatched rows is fixed as $h = {}^{n}/_{4}$.

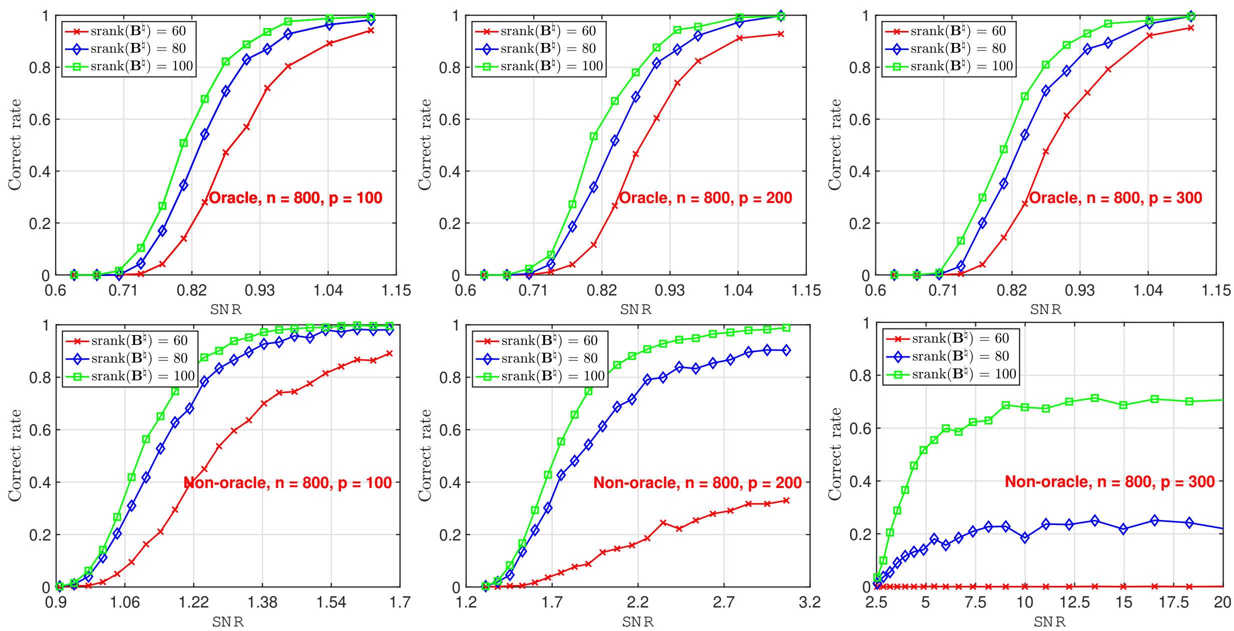

Figure 2: Impact of ${}^{n}/_{p}$ ratio: the number of mismatched rows is fixed as $h = {}^{n}/_{4}$.

5.2 IMPACT OF RATIO n/p

In this subsection, we investigate the impact of the ratio n/p on the reconstruction performance.

Experiment setup. We fix the sample number n to be 800 and vary p within $\{100, 200, 300\}$. The stable rank srank$(\mathbf{B}^*)$ is set $\{60, 80, 100\}$ and the number of mismatched rows h is fixed as $n/4$.

Results discussion. Corresponding numerical results are put in Figure 2. For the oracle case, we find the ratio ${}^{n}/_{p}$ hardly has any impact on the reconstruction performance: in all cases, the selection matrix can be recovered with positive probability when SNR ≥ 0.7 and is reliably reconstructed (with almost 100% correctness) when SNR ≥ 1.08. Meanwhile, for the non-oracle case, we can see that a lower ${}^{n}/_{p}$ ratio makes it harder to reconstruct the selection matrix. For example, when ${}^{n}/_{p} = 8$, we have that the correct rate becomes positive once SNR ≥ 0.9, a little larger than the corresponding value for the oracle case. When ${}^{n}/_{p}$ decreases

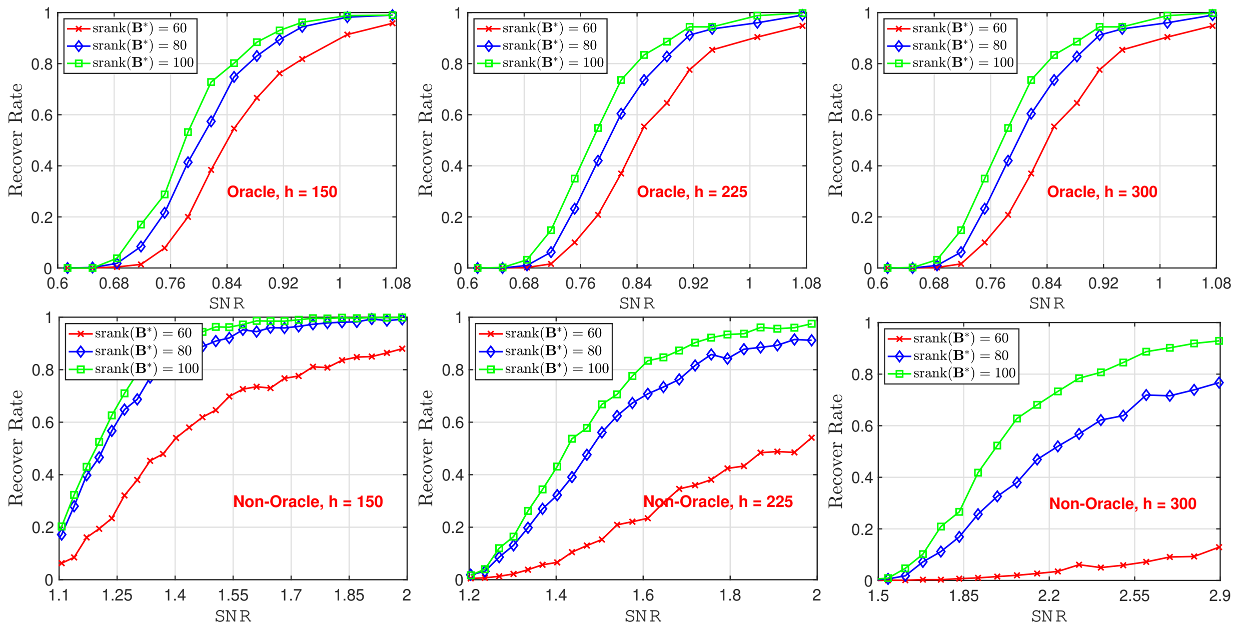

Figure 3: Impact of mismatched rows: the sample number n is fixed as 600 while the signal length p is fixed as 100.

to ${}^{n}/_{p} = 4$, this value increases to 1.4. When ${}^{n}/_{p}$ decreases to ${}^{n}/_{p} = {}^{8}/_{3}$, this value further increases to 2.5. When ${}^{n}/_{p}$ further decreases below 2 (not plotted), we find it impossible to reconstruct the selection matrix even with infinite SNR, i.e., noiseless sensing relation, which is rigorously proved in Unnikrishnan et al. [2018].

In addition, we notice that a lower ${}^{n}/_{p}$ ratio will put more stringent conditions on $\text{srank}(\mathbf{B}^*)$. When ${}^{n}/_{p} = 8$, we can reliably recover the selection matrix with $\text{srank}(\mathbf{B}^*)$ being 60. When ${}^{n}/_{p}$ decreases to $8/3$, we find the selection matrix can hardly be reconstructed.

5.3 IMPACT OF MISMATCHED ROWS

This subsection studies the impact of mismatched rows.

Experiment setup. We fix the signal length p to be 100, the sample number n to be 600, and the stable rank $\text{srank}(\mathbf{B}^*)$ to be $\{60, 80, 100\}$. Then we vary the number of mismatched rows h to be $\{n/4, 3n/8, n/2\}$.

Results discussion. Corresponding results are shown in Figure 3. Similar to the discussion w.r.t. the ratio ${}^{n}/_{p}$, in the oracle case we find the performances are almost identical for a different number of mismatched rows. While for the non-oracle case, the number of mismatched rows has a negative influence on the correspondence recovery: more mismatched rows lead to poorer performance. When ${}^{h}/_{n} = {}^{1}/_{4}$, we have the recovery rate become positive when $\mathsf{SNR} \geq 1.1$; when ${}^{h}/_{n} = {}^{3}/_{8}$, we have this threshold value increase to 1.2; and when ${}^{h}/_{n} = {}^{1}/_{2}$, we have this threshold value further jump to 1.5, which exhibits a similar trend when ${}^{n}/_{p}$ decreases.

Moreover, we notice a higher ${}^{h}/_{n}$ ratio puts more stringent requirements on the stable rank $\text{srank}(\mathbf{B}^*)$ for a reliable recovery of permutation. A similar phenomenon has also been observed in Figure 2.

In addition, we notice the allowed number of mismatched rows are affected by the ${}^{n}/_{p}$ ratio: a larger ${}^{n}/_{p}$ allows more mismatched rows, in other words, a larger proportion ${}^{h_{\max}}/_{n}$.

6 CONCLUSION

This paper considers the correspondence recovery for the unlabeled linear regression. Depending on whether the signal $\mathbf{B}^*$ is known or not, we propose separate estimators for each case. To the best of our knowledge, these are the first estimators that can support partial correspondence recovery, where only a proportion of the rows' correspondences rather than the whole selection matrix are to be reconstructed. Compared with the previous works on permuted linear regression, our estimators apply to a broader family of matrices, i.e., selection matrices rather than permutation matrices. Moreover, we prove both estimators are mini-max optimal. Notably, in analyzing the non-oracle estimator, we tailor the leave-one-out technique to an adaptive "leave-multiple-out" technique, which involves the simultaneous replacement of multiple (un-deterministic) rows and may serve as independent technical interests. Moreover, numerical experiments are presented to confirm our claims.

References

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

Dimitri P. Bertsekas and David A. Castañón. A forward/reverse auction algorithm for asymmetric assignment problems. *Comput. Optim. Appl.*, 1(3):277–297, 1992.

Sergio Caracciolo, Matteo P D'Achille, Enrico M Malatesta, and Gabriele Sicuro. Finite-size corrections in the random assignment problem. *Physical Review E*, 95(5):052129, 2017.

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

Michael Chertkov, Lukas Kroc, F Krzakala, M Vergassola, and L Zdeborová. Inference in particle tracking experiments by passing messages between images. *Proceedings of the National Academy of Sciences*, 107(17):7663–7668, 2010.

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.

Ivan Dokmanic. Permutations unlabeled beyond sampling unknown. *IEEE Signal Process. Lett.*, 26(6):823–827, 2019.

Guanhua Fang and Ping Li. Regression with label permutation in generalized linear model. *arXiv preprint arXiv:2206.11775*, 2022.

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

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

Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge university press, 1990.

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

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

Noureddine El Karoui, Derek Bean, Peter J Bickel, Chinghway Lim, and Bin Yu. On robust regression with high-dimensional predictors. *Proceedings of the National Academy of Sciences*, 110(36):14557–14562, 2013.

Tjalling C Koopmans and Martin Beckmann. Assignment problems and the location of economic activities. *Econometrica: journal of the Econometric Society*, pages 53–76, 1957.

Harold W Kuhn. The Hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

Enrico M Malatesta, Giorgio Parisi, and Gabriele Sicuro. Fluctuations in the random-link matching problem. *Physical Review E*, 100(3):032102, 2019.

Marc Mézard and Giorgio Parisi. Replicas and optimization. *Journal de Physique Lettres*, 46(17):771–778, 1985.

Marc Mézard and Giorgio Parisi. Mean-field equations for the matching and the travelling salesman problems. *EPL (Europhysics Letters)*, 2(12):913, 1986.

Ashwin Pananjady, Martin J Wainwright, and Thomas A Courtade. Denoising linear models with permuted data. In *Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT)*, pages 446–450, Aachen, Germany, 2017.

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.

Liangzu Peng and Manolis C. Tsakiris. Linear regression without correspondences via concave minimization. *IEEE Signal Process. Lett.*, 27:1580–1584, 2020.

Guilhem Semerjian, Gabriele Sicuro, and Lenka Zdeborová. Recovery thresholds in the sparse planted matching problem. *Physical Review E*, 102(2):022304, 2020.

Martin Slawski and Emanuel Ben-David. Linear regression with sparsely permuted data. *Electronic Journal of Statistics*, 1:1–36, 2019.

Martin Slawski, Mostafa Rahmani, and Ping Li. A sparse representation-based approach to linear regression with partially shuffled labels. In *Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI)*, page 7, Tel Aviv, Israel, 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. *Probability Theory and Related Fields*, 175(1):487–558, 2019.

Manolis C. Tsakiris and Liangzu Peng. Homomorphic sensing. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 6335–6344, Long Beach, CA, 2019.

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.

Shinji Umeyama. An eigendecomposition approach to weighted graph matching problems. *IEEE transactions on pattern analysis and machine intelligence*, 10(5):695–703, 1988.

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)*, pages 786–793, Monticello, IL, 2015.

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

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Hang Zhang and Ping Li. Optimal estimator for unlabeled linear regression. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 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)*, pages 1760–1765, Melbourne, Australia, 2021.

Hang Zhang, Martin Slawski, and Ping Li. Permutation recovery from multiple measurement vectors in unlabeled sensing. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 1857–1861, Paris, France, 2019.

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.