Match Made with Matrix Completion: Efficient Offline and Online Learning in Matching Markets

Zhiyuan Tang
Naveen Jindal School of Management, University of Texas at Dallas, zhiyuan.tang@utdallas.edu

Wanning Chen
Foster School of Business, University of Washington, wnchen@uw.edu

Kan Xu
W. P. Carey School of Business, Arizona State University, kanxu1@asu.edu

Online matching markets face increasing needs to accurately learn the matching qualities between demand and supply for effective design of matching policies. However, the growing diversity of participants introduces a *high-dimensional* challenge in practice, as there are a substantial number of *unknown* matching rewards and learning all rewards requires a large amount of data. We leverage a natural low-rank matrix structure of the matching rewards in these two-sided markets, and propose to utilize *matrix completion* (specifically the nuclear norm regularization approach) to accelerate the reward learning process with only a small amount of offline data. A key challenge in our setting is that the matrix entries are observed with *matching interference*, distinct from the independent sampling assumed in existing matrix completion literature. We propose a new proof technique and prove a near-optimal average accuracy guarantee with improved dependence on the matrix dimensions. Furthermore, to guide matching decisions, we develop a novel "double-enhancement" procedure that refines the nuclear norm regularized estimates and further provides near-optimal entry-wise estimations. Our paper makes the first investigation into adopting matrix completion techniques for matching problems. We also extend our approach to online learning settings for optimal matching and stable matching by incorporating matrix completion in multi-armed bandit algorithms. We present improved regret bounds in matrix dimensions through reduced costs during the exploration phase. Finally, we demonstrate the practical value of our methods using both synthetic data and real data of labor markets.

Key words: matrix completion, matching interference, multi-armed bandit, two-sided market

1. Introduction

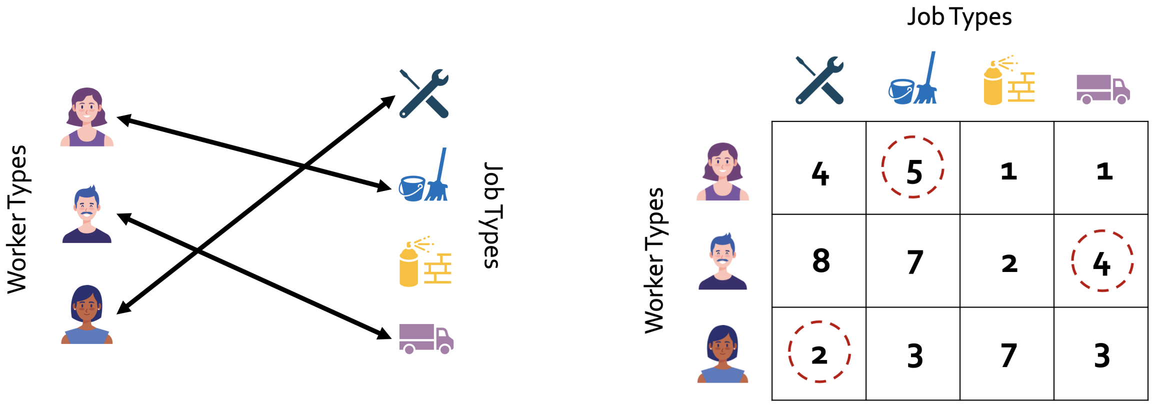
Online matching markets have become increasingly essential to facilitate the matching efficiency across many domains. For instance, freelance service platforms, such as Upwork and Taskrabbit, provide new opportunities to help businesses secure temporary labor (Belavina et al. 2020); dating apps, such as Tinder and Bumble, have become the most popular way for couples to meet (Rosenfeld et al. 2019); ride-hailing companies, such as Uber and Lyft, thrive through bridging demand and supply of transportation in a timely fashion (Yan et al. 2020); volunteer crowdsourcing platforms, such as Food Rescue US and Food Rescue Hero, rely on volunteers to deliver donated food from local businesses (Lo et al. 2024). The majority of the matching markets, including all previous examples, exhibit a *two-sided* structure, e.g., jobs and workers in the labor market, or riders and drivers in the ride-hailing sector. For convenience, we will use the aforementioned online labor marketplaces as our primary setting; thus, we denote the two sides of the market, i.e., the demand side and supply side, as *jobs* and *workers* respectively.

While the matching literature typically focuses on designing matching policies given *known* matching qualities between any types of jobs and workers, we consider the problem of learning *unknown* matching qualities through noisy *data*. To illustrate, consider the following example of one-to-one matching in a centralized online labor marketplace:

Example 1. Online labor platforms, such as Upwork and Taskrabbit, want to match each unfilled job with an available worker. In such a case, the set of job and worker types makes up the two sides of the online labor market. The design of any matching policies relies on accurate estimates of the quality or reward of any match between each job and worker type from, e.g., ratings of worker performance (Belavina et al. 2020) or certain quality measure of productivity (Kaynar and Siddiq 2023). To that end, the platform tests any possible matches of jobs and workers of different types over time, collects observable feedback of all matched pairs, and finally uses these data to estimate the matching qualities. Then, the platform can decide how to match the two sides with regard to the updated knowledge about matching qualities.

The increasing variety of both jobs and workers introduces significant challenges into learning the matching market both practically and theoretically. The current literature on two-sided matching typically assumes a small number of demand and supply types to ensure theoretical guarantees, yet limited in model dimensionality and hence practicality (Chen et al. 2023). In practice, the number of matching qualities to learn between job and worker types can be *high-dimensional.* For example, Upwork has more than a hundred job categories according to their website description; therefore, even tens of worker types will result in thousands of matching qualities to learn in such an online labor platform. As a result, learning accurate matching qualities statistically requires a large number of samples of matched pairs and their feedback. In other words, it can be very costly to collect enough observations to learn all matching qualities until the downstream decision making of matching can become precise.

In this paper, we propose to leverage the matrix structure of the matching market and speed up the learning process of matching qualities using the technique of *matrix completion* (Candes and Recht 2008, Candes and Plan 2010, Candès and Tao 2010). Note that the two-sided structure of the matching market (see Figure 1a) naturally possesses the form of a matrix — i.e., the job and worker types form the rows and columns of the matrix (see Figure 1b). For instance, in Figure 1b, each row of the matrix denotes one worker type, and each column denotes one job type; the value of each entry equals the unknown quality or reward of matching jobs and workers from the corresponding types. Our goal is to learn every entry of a large reward matrix (as discussed in the previous paragraph) using only a small number of observed noisy matching outcomes, which is typically much less than the total number of matrix entries. Yet, this is hardly likely without assuming some underlying structure of the reward matrix that relates different entries. Intuitively, if the entries of the reward matrix are related, then the observed matching outcomes of one pair of job and worker types can help learn the matching value of another pair.



4	5	1	1
8	7	2	4
2	3	7	3

(a) One-to-one Match

(b) Matrix Representation

Figure 1 **Toy example of one-to-one matching for 3 worker types and 4 job types. (a) presents one matching of this labor market. (b) shows the matrix representation of the matching problem, where each entry denotes the reward of matching the corresponding worker and job pair; the matched pairs in (a) are circled in red.**

To that end, we impose a natural *low-rank* structure of the reward matrix. The low-rank property has been demonstrated for general high-dimensional matrices (Udell and Townsend 2019) and has proven to be practically efficient across various domains, such as recommendation systems (Bell and Koren 2007), healthcare (Schuler et al. 2016), and textual analytics (Pennington et al. 2014). Formally, consider a matching market with N types of workers and K types of jobs. The reward matrix $\mathbf{\Theta}^* \in \mathbb{R}^{N\times K}$ represents the true matching qualities between jobs and workers; particularly, the $(i,j)^{\text{th}}$ entry of the matrix, $\mathbf{\Theta}^{*(i,j)}$, denotes the reward of matching workers of type $i \in [N] = \{1,\cdots,N\}$ and jobs of type $j \in [K] = \{1,\cdots,K\}$. We expect that, in a large matching market with large values of N and K, the reward matrix $\mathbf{\Theta}^*$ has a low rank r much smaller than the number of worker and job types, i.e., $\mathsf{rank}(\mathbf{\Theta}^*) = r \ll \min\{N,K\}$. In other words, we have $\mathbf{\Theta}^* = \mathbf{U}^*\mathbf{V}^{*\top}$ for some $\mathbf{U}^* \in \mathbb{R}^{N\times r}$ and $\mathbf{V}^* \in \mathbb{R}^{K\times r}$, of which the i^{th} row $\mathbf{U}^{*(i,\cdot)}$ and the j^{th} row $\mathbf{V}^{*(j,\cdot)}$ are r-dimensional latent vectors of worker type i and job type j respectively. These latent vectors can be interpreted as latent features of the corresponding worker and job types. For example, $\mathbf{U}^{*(i,\cdot)}$ might include worker type i's information such as education background and working experience; similarly, the

$\mathbf{V}^{*(j,\cdot)}$ might contain job type j's information such as income level and job responsibilities. Then, the matching reward $\mathbf{\Theta}^{*(i,j)}$ is determined by the interaction between the latent features of worker type i and job type j. Consequently, we are able to cross-learn all the matching rewards of workers with type i (or jobs with type j) since they share a common latent feature vector $\mathbf{U}^{*(i,\cdot)}$ (or $\mathbf{V}^{*(j,\cdot)}$). This allows us to estimate a much smaller number of $r(N+K)$ unknown parameters of $\mathbf{U}^*$ and $\mathbf{V}^*$ (Candes and Plan 2010) to learn the whole matching market, in contrast to learning NK unknown rewards in $\mathbf{\Theta}^*$ directly.

One state-of-the-art approach of matrix completion is to incorporate a *nuclear norm* penalty on $\mathbf{\Theta}^*$ (Koltchinskii et al. 2011, Negahban and Wainwright 2012). In particular, the nuclear norm penalty (which we will introduce in Section 3) only requires few samples to identify the *sparse* set of non-zero singular values of the matrix accurately and hence estimate the low-rank matrix efficiently.

However, we face two key challenges when applying this approach to learn the matching reward matrix. First, the current literature on matrix completion primarily provides improved performance bounds under the assumption that the entries of the matrix are observed or sampled *independently*, a condition not met by our setting due to *matching interference*. To illustrate, consider a one-to-one matching problem as in Figure 1, which shows one matching of the market. The three observed entries, circled in red in Figure 1b, are not independently observed since the same worker cannot choose multiple jobs and vice versa (i.e., there can be at most one entry in each row and column). Thus, it introduces additional correlation among observations and makes the current proof techniques not applicable. Second, to guide the downstream matching decision in certain scenarios, it is necessary to quantify the statistical uncertainty of the estimate of *each* matrix entry. For instance, given informative entry-wise uncertainty, we can find the matching with the maximum total matching reward (e.g., optimal matching) or learn the preference rankings of the demand side (e.g., stable matching) with high probability. However, the mainstream of the literature can only provide a theoretical guarantee of the *average* accuracy over all matrix entries. A few recent studies establish entry-wise error bounds for low-rank matrices (e.g., Chen et al. 2020, 2021); yet, these works heavily build on the assumptions of independent sampling and cannot be extended to our settings with sampling interference.

Our work addresses these two challenges through new analytical techniques. First, most proof techniques (see, e.g., Klopp 2014, Hamidi and Bayati 2022) rely on a contraction inequality. Yet, in our context with matching interference, this approach results in a sub-optimal error rate as if matrix completion were not employed; that is, the error rate scales similarly as that of estimating every entry using its sample average individually. To overcome this issue, we utilize a new "linearization" trick that leverages the sampling property of matrix completion problems, and show that a near-optimal bound with improved dependence on matrix dimensions (i.e., N and K) is achievable. Second, we develop a novel "double-enhancement" procedure that refines entry-wise estimates of the reward matrix and improves the entry-wise accuracy atop the nuclear norm regularized estimation. Specifically, this procedure estimates more precisely each row of the latent feature matrices $\mathbf{U}^*$ and $\mathbf{V}^*$ respectively — and thus each entry of $\boldsymbol{\Theta}^*$ — using linear regression based on the initial nuclear norm regularized estimate of $\boldsymbol{\Theta}^*$. We provide a tighter theoretical guarantee on the entry-wise errors of matrix completion by analyzing this enhanced estimator in our matching setting, while the existing proof techniques only work for independent sampling.

Furthermore, we extend our approach to online learning problems, where we efficiently learn the matching reward matrix in a dynamic fashion and perform both optimal matching and stable matching via bandit algorithms (Gai et al. 2010, Liu et al. 2020). Analogous to our offline setting, our bandit policy incorporates a natural low-rank structure of the unknown reward matrix to effectively balance the exploration of new worker-job pairs and the exploitation of the learned matching rewards. Specifically, it leverages low-rank matrix completion to accelerate the exploration phase during online matching. We derive regret upper bounds of our algorithms with improved dependence on the matrix dimension for bandit learning of both optimal matching and stable matching. Intuitively, our low-rank reward estimators allow us to collect only few samples while still obtain accurate reward estimates across all worker-job pairs during the exploration stage. Accordingly, our approaches provide sample-efficient solutions that substantially reduce the exploration costs, which is especially crucial for online matching problem with a large number of participants or with a relatively short time horizon.

Finally, we empirically evaluate the performance of our approaches using both synthetic and real data in the labor market. Our findings indicate that our approaches based on matrix completion significantly improve the learning process in the offline setting and facilitate the overall matching performance in the online scenario.

1.1. Related Literature

Our work relates to the literature of learning through matrix completion, and bandits for matching problems.

Matrix Completion. Matrix completion involves learning missing entries of a matrix from a small sample of observed ones, particularly for high-dimensional data that naturally forms a matrix structure. Due to its efficiency, there has been significant interest from the operations research and machine learning communities in applying matrix completion to various domains, such as learning preferences in recommendation systems (Farias and Li 2019), personalizing assortment planning

(Kallus and Udell 2020), detecting inventory inaccuracies (Farias et al. 2024), and enhancing textual analytics (Xu et al. 2024). As aforementioned, our reward learning problem in the matching markets can also be formulated as a matrix completion task, where workers and jobs define the two dimensions of the reward matrix. To the best of our knowledge, we propose the first framework of using matrix completion to facilitate reward learning in the matching problems.

Typically, the low-rank matrix completion literature assumes the entries of the unknown matrix $\mathbf{\Theta}^*$ of interest are sampled with the same probability independently (i.e., uniformly at random) (Candes and Recht 2008, Candès and Tao 2010, Koltchinskii et al. 2011). These studies show theoretically that the nuclear norm regularization can give rise to a minimax optimal error rate up to logarithmic terms in *Frobenius norm*, which measures the average estimation accuracy. Negahban and Wainwright (2012), Klopp (2014), Hamidi and Bayati (2022) later consider more general random sampling schemes, where the matrix entries are still sampled independently but with different probabilities (i.e., non-uniformly at random). Recently, Athey et al. (2021) conduct one of the first investigations into the sampling scheme with dependence, where the entries within the same row are observed dependently due to the panel data structure in a causal inference problem. In contrast, the dependence of the entry sampling in our problem is driven by the aforementioned matching interference, where we can at most observe one entry from each row and column in a one-to-one matching (see Figure 1b). As a result, existing proof techniques (Klopp 2014, Athey et al. 2021, Hamidi and Bayati 2022) that exploit contraction inequality (e.g., Theorem 2.3 in Koltchinskii (2011) or Theorem 4.4 in Ledoux and Talagrand (2013)) lead to sub-optimal error bounds in Frobenius norm in our setting. Rather, we build on a "linearization" trick that leverages the sampling property of matrix completion, and show that we can achieve a near-optimal error bound with improved dependence on matrix dimensions (i.e., the number of worker/job types).

For matching problems, it is also important to provide entry-wise error control in ℓ_∞ *norm* for downstream decision making, which is a harder problem than controlling Frobenius norm error. Chen et al. (2020, 2021) are among the first to provide a near-optimal entry-wise error bound for matrix completion; however, their analyses rely on the assumption that the entries are sampled uniformly at random, and thus do not apply to our sampling scheme with matching interference. We develop a novel double-enhancement algorithm that improves the standard nuclear norm regularized estimator and ensures an entry-wise accuracy guarantee in our setting. Our approach is inspired by the row-enhancement technique from Hamidi et al. (2019); however, they consider a different matrix factorization problem for contextual bandits, and only provide row-wise error guarantees[1].

[1] Row-wise error control is simpler than entry-wise analysis, but harder than Frobenius norm error control.

We also note that our matrix completion problem is different from those in Candes and Recht (2008), Candes and Plan (2010), Chen et al. (2020), which only allow each entry to be observed at most once. In turn, we follow the settings in an alternative line of matrix completion literature (Koltchinskii et al. 2011, Negahban and Wainwright 2012, Klopp 2014, Hamidi and Bayati 2022), where we might observe each entry repeatedly throughout sampling.

Multi-Armed Bandits. Our online learning algorithms contribute to the literature of bandit algorithm design for optimal matching and stable matching. In the bandit problem for optimal matching, the goal is to learn the unknown rewards of matching workers with jobs over time and maximize the cumulative rewards of matched pairs. This can be formulated as a combinatorial semi-bandit problem, where each arm is one pair of worker and job types in our setting and a set of arms that form one matching is played simultaneously in each round (Gai et al. 2010, Chen et al. 2013, Kveton et al. 2015). Chen et al. (2013) and Wang and Chen (2018) propose Combinatorial Upper Confidence Bound (CUCB) and Combinatorial Thomspon Sampling (CTS), adapted from the classic bandit literature (Auer et al. 2002, Chapelle and Li 2011). Kveton et al. (2015) improve the CUCB algorithm; they derive near-optimal upper bounds of their algorithm, matching the lower bounds.

Recently there has also been growing interest in learning stable matching (Gale and Shapley 1962) from bandit feedback in a centralized platform (Liu et al. 2020, Jagadeesan et al. 2021, Cen and Shah 2022). Liu et al. (2020) first introduce this bandit problem, where one side of the market (i.e., agents) has no prior knowledge about its ranking preference over the other side (i.e., arms) and needs to learn from stochastic rewards of matched pairs. They propose an Explore-then-Commit (ETC) algorithm to minimize the agent-optimal stable regret — i.e., to find an optimal policy that is a stable matching and optimal for the specific agent side. Cen and Shah (2022) extend the model to incorporate predetermined costs and transfers, and provide a guarantee for both stability and low regret using Upper Confidence Bound (UCB). Jagadeesan et al. (2021) consider a different matching with transfers problem (Shapley and Shubik 1971), and introduce a new notion of regret that captures the deviation of a market outcome from equilibrium. We adopt the bandit setting from Liu et al. (2020) for our stable matching problem with workers being the agent side and jobs being the arms, as its regret definition aligns closely with the traditional bandit literature.

However, the regret rates of existing bandit algorithms exhibit an unfavorable dependence on the number of arms for both optimal matching and stable matching. This can degrade the algorithm performance, especially when the number of arms is large or the time horizon is relatively short. As aforementioned, our approach uses matrix completion to expedite the exploration process, and achieve improved regret upper bounds with regard to the number of arms (i.e., the number of worker/job types).

There is another stream of literature that considers a multi-armed bandit or linear contextual bandit problem, where the unknown parameter of the reward has a low-rank structure (see, e.g., Jun et al. 2019, Bayati et al. 2022, Zhou et al. 2024). However, their settings are different from our matching problem, and thus their algorithms do not apply to our setting with matching interference.

Dynamic Matching. Our online learning part is also related to the dynamic matching literature in operations research that attempts to incorporate learning in a two-sided market (Massoulié and Xu 2016, Bimpikis and Markakis 2019, Shah et al. 2020, Johari et al. 2021, Hsu et al. 2022). However, the majority of this literature focuses on learning unknown types of one market side (e.g., worker/job side), and thus requires knowledge of type-dependent matching rewards. Hsu et al. (2022) provide the first study with both unknown matching payoffs and unknown types of one side, but consider a different problem with queueing and do not aim to improve the dependence on market size (e.g., the number of worker/job types) as we do. Sauré and Zeevi (2013) propose an ETC algorithm and present improved regret dependence on market size (i.e., the number of products); however, they address a different demand learning problem in dynamic assortment and hence their technique cannot be extended to our setting. In contrast, we focus on improving the performance bounds regarding their dependence on market size given unknown rewards and known types of participants.

1.2. Contributions

We highlight the main contributions of our paper as follows:

1. We propose the first framework that utilizes matrix completion to expedite reward learning in matching problems. We theoretically justify the efficiency of matrix completion via nuclear norm regularization in our context with matching interference. Specifically, we present a new proof technique through which a near-optimal error bound with improved dependence on matrix dimensions is achievable. We also prove that our error bound is minimax optimal up to logarithmic terms regarding matrix dimension through a lower bound analysis (Section 3).

2. We further design a novel double-enhancement procedure that significantly refines the accuracy of entry-wise estimates of low-rank matrix completion. Our approach works for broader sampling distributions, including our setting of sampling with matching interference, while the existing guarantees typically depend on the property of independent sampling and do not apply to our setting (Section 4).

3. We also extend our offline approach to the online learning setting, and design bandit algorithms for both stable matching and optimal matching problems. Particularly, we leverage low-rank matrix completion to learn the matching reward matrix efficiently to facilitate exploration. We demonstrate the improvement of our algorithms in matrix dimensions with two regret upper bound results (Section 5).

4. We provide experimental results on both synthetic and real data to show that our algorithms substantially outperform the benchmarks in both offline reward learning problem and online bandit problems for optimal and stable matching (Section 6).

2. Problem Formulation

This section formalizes the problem of learning the matching rewards between workers and jobs using offline matching data as a matrix completion problem. In Section 2.1, we formulate the unknown rewards as a low-rank matrix and introduce the observation model for this reward matrix. In Section 2.2, we highlight the key challenge of applying matrix completion to matching problems — i.e., dependent entry-sampling due to matching interference, in contrast to canonical independent sampling.

Notation. We use regular capital letters for vectors, lowercase letters for scalars, and bold capital letters for matrices, unless otherwise specified. For any positive integer k, let $[k]$ denote the index set $\{1, 2, \cdots, k\}$. For any vector V, let $V^{(i)}$ denote its i^{th} entry, and $\|V\|$ denote its ℓ_2 norm. For any matrix $\mathbf{\Theta} \in \mathbb{R}^{d_1 \times d_2}$, we use $\mathbf{\Theta}^{(i,j)}$ to represent entry (i, j) of a matrix $\mathbf{\Theta}$, $\mathbf{\Theta}^{(i,\cdot)}$ the i^{th} row of the matrix, and $\mathbf{\Theta}^{(\cdot,j)}$ the j^{th} column. For a matrix $\mathbf{\Theta}$ of rank r, we denote its non-zero singular values by $\sigma_{\max}(\mathbf{\Theta}) = \sigma_1(\mathbf{\Theta}) \geq \sigma_2(\mathbf{\Theta}) \geq \cdots \geq \sigma_r(\mathbf{\Theta}) = \sigma_{\min}(\mathbf{\Theta}) > 0$, its Frobenius norm by $\|\mathbf{\Theta}\|_F = \sqrt{\sum_{i=1}^r \sigma_i^2(\mathbf{\Theta})}$, its operator norm by $\|\mathbf{\Theta}\|_{\text{op}} = \sigma_1(\mathbf{\Theta})$, its nuclear norm by $\|\mathbf{\Theta}\|_* = \sum_{i=1}^r \sigma_i(\mathbf{\Theta})$, its $\ell_{2,\infty}$ norm by $\|\mathbf{\Theta}\|_{2,\infty} = \max_{i \in [d_1]} \|\mathbf{\Theta}^{(i,\cdot)}\|$, and its vector ℓ_∞ norm by $\|\mathbf{\Theta}\|_\infty = \max_{i,j} |\mathbf{\Theta}^{(i,j)}|$. Given any two matrices $\mathbf{\Theta}, \mathbf{\Theta}' \in \mathbb{R}^{d_1 \times d_2}$, we denote their trace inner product by $\langle \mathbf{\Theta}, \mathbf{\Theta}' \rangle = \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} \mathbf{\Theta}^{(i,j)} \mathbf{\Theta}'^{(i,j)}$, and their Hadamard product by $\mathbf{\Theta} \circ \mathbf{\Theta}' \in \mathbb{R}^{d_1 \times d_2}$ with $(\mathbf{\Theta} \circ \mathbf{\Theta}')^{(i,j)} = \mathbf{\Theta}^{(i,j)} \cdot \mathbf{\Theta}'^{(i,j)}$. Let $e_i(d) \in \mathbb{R}^d$ denote a basis vector with value 1 in its i^{th} entry and 0 otherwise, i.e., $e_i(d)^{(j)} = 1$ for $j = i$ and 0 otherwise.

2.1. Matching in a Two-Sided Market

Consider a two-sided online labor platform (see, e.g., Example 1) with N available types of workers to be matched with K unfilled types of jobs. Assume $N \leq K$ without loss of generality. The platform is centralized — i.e., it has full control over job assignments.

Reward Matrix. The true qualities or rewards of matching the worker and job sides can be naturally collected into a matrix form, according to the two-sided structure of the market (see, e.g., Figure 1). We use $\mathbf{\Theta}^* \in \mathbb{R}^{N \times K}$ to denote the reward matrix, where the worker types $i \in [N]$ and job types $j \in [K]$ define one dimension of the matrix respectively. Particularly, each row i of the matrix corresponds to the matching rewards of one worker type i, and each column j corresponds to one job type j; the value of each entry (i, j), i.e., $\mathbf{\Theta}^{*(i,j)}$, indicates the expected reward the platform receives when a worker of type i is matched with a job of type j, for $i \in [N]$ and $j \in [K]$.

We make two standard assumptions on the reward matrix $\mathbf{\Theta}^*$, which is directly adapted from the matrix completion literature (Koltchinskii et al. 2011, Negahban and Wainwright 2012, Klopp 2014, Farias and Li 2019, Chen et al. 2020, Athey et al. 2021). First, the true reward matrix $\mathbf{\Theta}^*$ is *entry-wise bounded* by 1, i.e., $\|\mathbf{\Theta}^*\|_\infty \leq 1$. Note that we choose an upper bound of 1 just for simplicity — our results hold for any constant upper bound.

Second, $\mathbf{\Theta}^*$ has *low rank*, i.e., $\mathsf{rank}(\mathbf{\Theta}^*) = r \ll \min\{N, K\}$. As discussed in Section 1, Udell and Townsend (2019) demonstrate that any sufficiently large matrix has low-rank property in general, while typical matching markets oftentimes face a large number of worker and job types (i.e., the matrix dimensions N and K are large) in practice. Particularly, as mentioned in Section 1, there exist two low-dimensional matrices $\mathbf{U}^* \in \mathbb{R}^{N\times r}$ and $\mathbf{V}^* \in \mathbb{R}^{K\times r}$ such that $\mathbf{\Theta}^* = \mathbf{U}^*\mathbf{V}^{*\top}$. The matching reward of worker type i and job type j is then jointly determined by their latent features, i.e., $\mathbf{\Theta}^{*(i,j)} = \mathbf{U}^{*(i,\cdot)}\mathbf{V}^{*(j,\cdot)\top}$. Intuitively, low-rankness suggests that the true reward matrix depends on very few parameters, which helps reduce the number of parameters to learn from NK of $\mathbf{\Theta}^*$ to $r(N+K)$ of $\mathbf{U}^*$ and $\mathbf{V}^*$.

Matching. We consider the *one-to-one matching* scheme for simplicity, i.e., each worker can be matched with at most one job and vice versa; our technical results can be naturally extended to the general many-to-many matching setting. Let $\mathbb{M}$ denote one such *matching*, defined as a set of pairs of worker and job types:

$$\mathbb{M} = \{(i, j(i)) \mid i \in [N];\, j(i) \neq j(i'),\, \forall i \neq i'\},$$

where, with slight abuse of notation, let $j(i)$ denote the job type matched with the i^{th} worker type, and no two worker types $i \neq i'$ share the same job types $j(i) \neq j(i')$ and vice versa. Note that the platform clears the market in any matching.

We can also denote a matching $\mathbb{M}$ using a matrix $\mathbf{X} \in \{0,1\}^{N\times K}$, where the $(i,j)^{\text{th}}$ entry $\mathbf{X}^{(i,j)}$ takes value 1 if $(i,j) \in \mathbb{M}$ and 0 otherwise. Any matching should belong to the following set of matchings

$$\mathcal{M} = \left\{ \mathbf{X} \in \{0,1\}^{N\times K} \,\middle|\, \sum_{j=1}^{K} \mathbf{X}^{(i,j)} = 1, \forall i \in [N]; \sum_{i=1}^{N} \mathbf{X}^{(i,j)} \leq 1, \forall j \in [K] \right\}. \tag{2.1}$$

Particularly, $\sum_{i=1}^{N} \mathbf{X}^{(i,j)} \leq 1$ for a job type j since j can be assigned to at most one worker type; in other words, the j^{th} column of $\mathbf{X}$, i.e., $\mathbf{X}^{(\cdot,j)}$, contains at most one entry of 1 and elsewhere 0. Similarly, $\sum_{j=1}^{K} \mathbf{X}^{(i,j)} = 1$ since the market clears and each worker type can take exactly one job type. When no ambiguity arises, we will use $\mathbb{M}$ and $\mathbf{X}$ interchangeably in the subsequent sections.

We further define the *matched pair* of worker type i in a matching $\mathbf{X}$ using a superscript i

$$\mathbf{X}^i = e_i(N) e_{j(i)}(K)^\top \in \mathbb{R}^{N\times K}, \tag{2.2}$$

that is, a basis matrix with the $(i, j(i))^{\text{th}}$ entry being 1 and 0 otherwise (recall that $e_i(d) \in \mathbb{R}^d$ is a basis vector with value 1 in its i^{th} entry and 0 otherwise). By our definitions, the matched pairs of all worker types make up the matching $\mathbf{X}$:

$$\sum_{i=1}^{N} \mathbf{X}^i = \mathbf{X} \in \mathcal{M}. \tag{2.3}$$

Observation Model. Our offline matching data consists of n independently observed matchings on the platform. Let $\mathbb{M}_t$ (and $\mathbf{X}_t$) denote each matching for $t \in [n]$, and $\mathbf{X}_t^i$ denote the matched pair of worker type i in the matching $\mathbf{X}_t$. We use $(i, j_t(i)) \in \mathbb{M}_t$ to represent a matched pair of worker type i and its corresponding job type from $\mathbb{M}_t$. Then, each matching $\mathbf{X}_t$ is sampled independently from the set $\mathcal{M}$ defined in (2.1); formally, $\mathbf{X}_t$ is drawn from a distribution Π, where Π is a discrete uniform distribution over $\mathcal{M}$[2].

For each matching $\mathbb{M}_t$, the platform observes only a noisy signal of the true reward $\boldsymbol{\Theta}^{*(i, j_t(i))}$ for any pair $(i, j_t(i)) \in \mathbb{M}_t$. We concatenate the N noisy rewards of the N matched pairs in $\mathbb{M}_t$ into a vector $Y_t \in \mathbb{R}^N$, and denote the N corresponding noises as $\varepsilon_t \in \mathbb{R}^N$. Specifically, the i^{th} entry of Y_t, denoted by $Y_t^{(i)}$, corresponds to the observed reward of the pair $(i, j_t(i))$; $\varepsilon_t^{(i)}$ represents the unobserved noise of that pair. Then, each reward $Y_t^{(i)}$ of worker type i in a matching $\mathbb{M}_t$ has

$$Y_t^{(i)} = \langle \mathbf{X}_t^i, \boldsymbol{\Theta}^* \rangle + \varepsilon_t^{(i)} \tag{2.4}$$

for $t \in [n]$ and $i \in [N]$. The noises $\varepsilon_t^{(i)}$ are σ-subgaussian (see Definition 1) and independent across matchings $t \in [n]$; we want to mention that the noises are not required to be independent across matched pairs for $i \in [N]$ within the same matching $t \in [n]$.

Definition 1. A random variable $X \in \mathbb{R}$ is σ-subgaussian if $\mathbb{E}[X] = 0$ and $\mathbb{E}[\exp(sX)] \leq \exp\left(\frac{\sigma^2 s^2}{2}\right)$, $\forall s \in \mathbb{R}$.

To simplify notation, we define an *observation operator* $\mathcal{X}_t : \mathbb{R}^{N \times K} \to \mathbb{R}^N$ such that

$$\mathcal{X}_t(\boldsymbol{\Theta}) = \left[\langle \mathbf{X}_t^1, \boldsymbol{\Theta} \rangle, \cdots, \langle \mathbf{X}_t^N, \boldsymbol{\Theta} \rangle\right]^\top \tag{2.5}$$

for any $\boldsymbol{\Theta} \in \mathbb{R}^{N \times K}$ and $t \in [n]$.

Remark 1. As we will discuss in Section 2.2, the entries of $\boldsymbol{\Theta}^*$ are not independently sampled, i.e., $\{\mathbf{X}_t^i\}_{i \in [N], t \in [n]}$ are not independent, even though the matchings $\{\mathbf{X}_t\}_{t \in [n]}$ are independent. Particularly, $\{\mathbf{X}_t^i\}_{i \in [N]}$ are correlated for every $t \in [n]$ due to the one-to-one matching constraint.

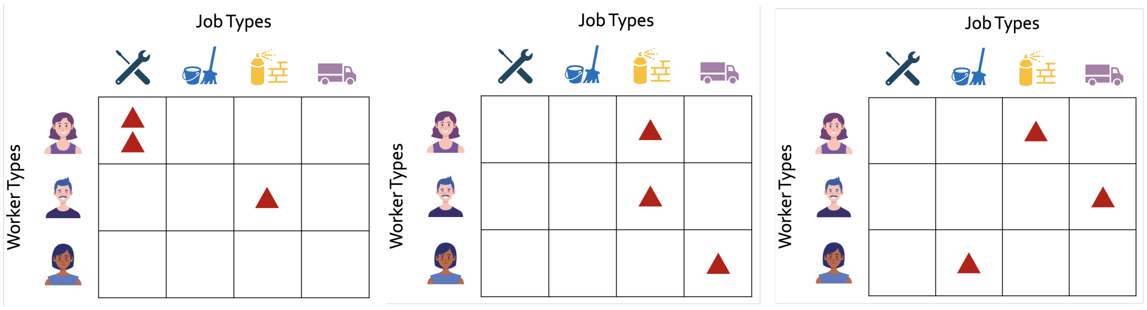
[2] Sampling one matching $\mathbf{X}$ from the combinatorial set $\mathcal{M}$ can be efficiently implemented in practice by randomly generating a N-permutation of K, i.e., an ordered sequence of N randomly selected numbers from the index set $[K]$ without replacement. Then, the i^{th} element in this sequence represents the column index $j(i)$ where the i^{th} row of $\mathbf{X}$ takes the value 1.

Estimation. Our problem of recovering the unknown low-rank matrix $\mathbf{\Theta}^*$ using entry-wise observations is an instance of the matrix completion problem (Candes and Recht 2008, Candès and Tao 2010). In total, we observe nN samples for the model (2.4) from n matchings and N matched pairs in each matching, i.e., $\{(\mathbf{X}_t^i, Y_t^{(i)}) \mid t \in [n], i \in [N]\}$. Note that we might face a high-dimensional problem where the sample size nN could be limited relative to the total number of unknown parameters in $\mathbf{\Theta}^*$, i.e., NK. As a result, we propose to exploit the low-rank structure of $\mathbf{\Theta}^*$ and efficiently learn the unknown matrix using matrix completion. We describe the details of the estimation procedure in Section 3.1.

We measure the estimation accuracy of an estimator $\widehat{\mathbf{\Theta}}$ by its Frobenius norm, i.e., $\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_F$, in Section 3. As aforementioned, we require an entry-wise error guarantee for certain downstream matching problems. Thus, we also measure an entry-wise estimation error of $\widehat{\mathbf{\Theta}}$ by its ℓ_∞ norm, i.e., $\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_\infty$, in Section 4.

2.2. Sampling with Matching Interference

A key challenge of our matrix completion problem is the dependence of entry sampling due to interference in the matching markets. We provide detailed descriptions of this problem, and compare our matching sampling scheme to two standard independent sampling schemes in the literature, i.e., independent sampling and independent row sampling. For simplicity, we take the one-to-one matching setting as an example; our argument applies to many-to-many matching setting as well.



(a) Independent Sampling (b) Indepedent Row Sampling (c) Matching Sampling

Figure 2 **Toy example of different sampling schemes for $N = 3$ worker types and $K = 4$ job types. One red triangle indicates that the corresponding entry is sampled once. (a) shows an independent sampling scheme, where all entries are sampled at random with replacement. (b) shows an independent row sampling scheme, where the column is selected at random for each row independently. (c) shows a dependent sampling scheme with matching interference in our matching setting.**

Independent Sampling. Independent sampling is the most widely studied sampling scheme in the matrix completion literature. Typically, it refers to the setting where the entries of a matrix are sampled at random either uniformly (see, e.g., Candes and Recht 2008, Candès and Tao 2010) or non-uniformly (see, e.g., Negahban and Wainwright 2012, Klopp 2014). The majority of the literature derives an improved estimation error bound for an unknown low-rank matrix under this assumption, given its tractability. For instance, consider sampling uniformly at random in the observation model (2.4). In this case, the entry sampling or the matched pair $\mathbf{X}_t^i$ has

$$\mathbf{X}_t^i = e_\iota(N) e_j(K)^\top, \tag{2.6}$$

where ι and j are randomly sampled from the index sets $[N]$ and $[K]$ respectively for any $i \in [N]$ and $t \in [n]$. Note that here both ι and j do not depend on the indices i and t. Therefore, $\{\mathbf{X}_t^i\}_{i\in[N],t\in[n]}$ are all independent from each other.

We illustrate that in Figure 2a; for simplicity, we consider the three samples of a specific t but of three different i's with $i \in [N]$ and $N = 3$, i.e., $\{(\mathbf{X}_t^i, Y_t^{(i)}) \mid i \in [3]\}$. Particularly, one red triangle indicates that the corresponding entry is sampled once (or the corresponding worker-job pair is matched once). Then, all entries of the matrix are sampled three times at random with replacement given our definition in (2.6); since the location of one red triangle does not affect that of another red triangle, the entries are sampled independently. Consequently, the same pair of worker and job types might be observed multiple times, while some worker or job types might not be matched through the sampling process. Apparently, independent sampling cannot guarantee a one-to-one matching for our matching setting.

Independent Row Sampling. An alternative sampling scheme is the independent row sampling (Jain and Pal 2022, Baby and Pal 2024). Particularly, sampling takes place within each row of the matrix, and a column is randomly selected for each row independently. Now, consider independent row sampling in the observation model (2.4). In such a case, the entry sampling or the matched pair $\mathbf{X}_t^i$ has

$$\mathbf{X}_t^i = e_i(N) e_j(K)^\top,$$

where j is randomly sampled from the index set $[K]$ for any $i \in [N]$ and $t \in [n]$. Different from our example of independent sampling, there is no longer randomness in sampling rows for a given $i \in [N]$. Yet, $\{\mathbf{X}_t^i\}_{i\in[N],t\in[n]}$ are still independent from each other, since the columns are uniformly sampled at random regardless of the row index i. The independent row sampling is closely related to the independent sampling, since the entry samplings or matched pairs $\mathbf{X}_t^i$'s are still independent from each other.

This is illustrated in Figure 2b. Similarly, we consider three samples of a specific t but of three different i's so matrix entries are observed exactly three times through $\{\mathbf{X}_t^i\}_{i\in[N]}$. Under independent row sampling, the entries in each row are randomly sampled once; the entry sampling is independent across rows, since the column of one red triangle in a given row does not interfere with those of different rows. As a result, multiple worker types might be matched with the same job type, though no worker type remains unmatched. The independent row sampling still cannot ensure a one-to-one matching for our matching purpose.

Matching Sampling. In contrast, the one-to-one matching constraint in our matching problem raises a unique challenge for adopting matrix completion techniques. Specifically, our matching sampling scheme introduces dependence of entry sampling or matched pairs within a matching due to matching interference. Recall that $\{\mathbf{X}_t^i\}_{i\in[N],t\in[n]}$ are not independent from each other in our observation model (2.4). In particular, the N matched pairs $\{\mathbf{X}_t^i\}_{i\in[N]}$ that comprise any matching $\mathbf{X}_t$ are correlated, since they have to satisfy

$$1=\sum_{j=1}^{K}\mathbf{X}_t^{(i,j)}=\sum_{j=1}^{K}\mathbf{X}_t^{i(i,j)},\,\forall i\in[N]\,,\quad\text{and}\quad 1\geq\sum_{i=1}^{N}\mathbf{X}_t^{(i,j)}=\sum_{i=1}^{N}\mathbf{X}_t^{i(i,j)},\,\forall j\in[K]$$

according to (2.1) and (2.3) (recall that the superscript (i,j) indicates the entry (i,j) of a matrix). Basically, these inequalities enforce the one-to-one matching constraint, ensuring that two worker types $i\neq i'$ cannot be matched with a same job type j in the same matching $\mathbf{X}_t$, i.e., $j_t(i)\neq j_t(i')$. Nevertheless, any matched pairs $\mathbf{X}_t^i$ and $\mathbf{X}_{t'}^{i'}$ across different matchings $\mathbf{X}_t$ and $\mathbf{X}_{t'}$ with $t\neq t'$ are still independent for any $i,i'\in[N]$, since the matchings are independently sampled from the distribution Π given our model setup in Section 2.1.

In Figure 2c, we present one possible matching $\mathbf{X}_t$ of three observed entries or matched pairs given a specific t and $i\in[N]$. The entries are sampled according to the one-to-one matching constraint; thus, the entry sampling is dependent in our case, as any two of the three red triangles cannot fall into the same row or column. In other words, no two worker types share the same job type and vice versa, and all worker types are fully occupied.

Our matching sampling scheme involves dependence across entry sampling due to matching interference, and is hence intrinsically different from both the independent sampling and independent row sampling schemes. Therefore, the current proof techniques based on these assumptions cannot be extended to our setting, and new proof techniques are required to achieve an optimal sample efficiency for our matching sampling. In the next section, we first specify our matrix completion approach, and then provide a new proof technique to demonstrate the optimality of matrix completion in such a setting.

3. Matrix Completion for Offline Matching Data

We first describe the standard matrix completion approach of nuclear norm regularization, and utilize it to estimate the reward matrix in our matching setting in Section 3.1. Then, in Section 3.2, we prove that this estimator can achieve a near-optimal error rate in Frobenius norm under dependent sampling with matching interference. Finally, we provide a minimax lower bound for our estimator to show its sample efficiency in Section 3.3.

3.1. Estimation via Nuclear Norm Regularization

We estimate the unknown reward matrix $\mathbf{\Theta}^*$ using the standard nuclear norm regularization defined as follows:

$$\widehat{\mathbf{\Theta}} = \underset{\|\mathbf{\Theta}\|_\infty \leq 1, \, \mathsf{rank}(\mathbf{\Theta}) \leq r}{\arg\min} \left\{ \frac{1}{n} \sum_{t=1}^{n} \|Y_t - \mathcal{X}_t(\mathbf{\Theta})\|^2 + \lambda \|\mathbf{\Theta}\|_* \right\}, \tag{3.1}$$

where $\mathcal{X}_t$ is the observation operator constructed with respect to the N matched pairs $\{\mathbf{X}_t^i\}_{i \in [N]}$ as defined in (2.5), Y_t is a vector of N noisy rewards observed for the N matched pairs in one matching $\mathbb{M}_t$, and λ is a hyperparameter. In total, we have nN samples of entries observed through n matchings with N matched pairs in each matching. We follow (Chen and Wainwright 2015, Ma et al. 2018) and search our estimate in $\mathsf{rank}(\mathbf{\Theta}) \leq r$ to ensure that our estimate has rank no greater than r. Previously, Chen and Wainwright (2015), Ma et al. (2018) demonstrate that such a rank constraint can lead to improved estimation accuracy. Indeed, enforcing the low-rank structure is crucial for obtaining an optimal error rate in our Theorem 1[3].

Our objective function (3.1) is directly adapted from the existing matrix completion literature (Koltchinskii et al. 2011, Negahban and Wainwright 2012). The first part of (3.1) is the mean squared error of the matrix estimation. Its notation slightly departs from the literature since we calculate the loss in two steps — we first estimate the errors of the N matched pairs within each matching, and then accumulate the errors of all the n matchings.

In the second part of (3.1), we use the nuclear norm penalty to efficiently estimate the reward matrix. Note that a low-rank matrix has only a sparse set of non-zero singular values. Considering that, we implement the nuclear norm regularization to identify such a sparse set accurately and thus estimate our low-rank reward matrix efficiently. Intuitively, the nuclear norm (defined as the sum of the matrix singular values) regularization for matrix estimation is comparable to the LASSO estimator (Bühlmann and Van De Geer 2011, Negahban et al. 2012) for linear regression — the LASSO penalty can efficiently identify and estimate a sparse unknown vector parameter.

The hyperparameter λ trades off bias and variance. When $\lambda \to 0$, we perform a least square estimation using the observed data; the estimator is unbiased but has high variance due to the scarcity of the samples. Alternatively, when λ is large, we might regularize the estimator too much towards zero; in other words, we obtain an estimator with high bias despite low variance. Our main result in Theorem 1 will provide a theoretically optimal value for λ, which appropriately balances bias and variance in our setting.

[3] We want to note that such a constraint is only for theoretical purpose; in practice, the nuclear norm regularization remains effective without the constraints.

3.2. Error Bound in Frobenius Norm

To measure the accuracy of our matrix completion approach in (3.1), we provide an estimation error bound in Frobenius norm in Theorem 1. Intuitively, the Frobenius norm error quantifies the average accuracy of the estimator across all matrix entries. To derive Theorem 1, we first need to show that an intermediary result of restricted strong convexity (RSC) holds with high probability for our matching model. This condition is common in the matrix completion literature (Negahban and Wainwright 2012, Negahban et al. 2012, Klopp 2014, Athey et al. 2021). To proceed, we define the $L^2(\Pi)$ norm of any matrix $\mathbf{\Theta} \in \mathbb{R}^{N\times K}$ as

$$\|\mathbf{\Theta}\|_{L^2(\Pi)} = \sqrt{\mathbb{E}\left[\sum_{i=1}^{N} \langle \mathbf{X}_t^i, \mathbf{\Theta}\rangle^2\right]},$$

where the expectation is taken over $\{\mathbf{X}_t^i\}_{i\in[N]}$ defined in (2.2) (recall that $\mathbf{X}_t$'s are sampled from Π).

Proposition 1. *Define for any $\alpha > 0$*

$$\mathcal{C}_\alpha(r) = \left\{ \mathbf{\Delta} \in \mathbb{R}^{N\times K} \;\middle|\; \|\mathbf{\Delta}\|_\infty \leq 1, \|\mathbf{\Delta}\|^2_{L^2(\Pi)} > \frac{c_0 N\alpha}{n}, \mathit{rank}(\mathbf{\Delta}) \leq r \right\}. \tag{3.2}$$

Then, for any $\mathbf{\Delta} \in \mathcal{C}_\alpha(r)$, we have

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{N} \langle \mathbf{X}_t^i, \mathbf{\Delta}\rangle^2 \geq c_2 \|\mathbf{\Delta}\|^2_{L^2(\Pi)} - c_3\left(\frac{r^2 K \log[(N+K)n]}{n}\right), \tag{3.3}$$

with probability greater than $1 - \exp(-\alpha)$, where c_0, c_2 and c_3 are positive constants.

The proof is provided in Appendix A.5. Essentially, our RSC condition shows that the loss function in the first part of (3.1) is almost strongly convex, since the last term on the right hand side of (3.3) is minor with moderate size of n. This condition is similar to the minimum eigenvalue condition for linear regression or compatibility condition for LASSO estimation (Bühlmann and Van De Geer 2011, Negahban et al. 2012), which provides enough convexity guarantee on the loss function to identify the optimal estimates. Intuitively, our matching sampling scheme, though limited by the one-to-one matching constraint, still captures and reveals a substantial proportion of the true matrix entries.

Based on the RSC condition, we now present our main result that provides a theoretical guarantee on the nuclear norm regularized estimator in (3.1). We will further show its minimax optimality (up to logarithmic terms) in the upcoming Section 3.3. In essence, our result demonstrates that the standard nuclear norm regularization approach can still achieve an optimal performance even under a harder matching sampling scheme than the canonical independent sampling.

Theorem 1. *The estimator $\widehat{\boldsymbol{\Theta}}$ in* (3.1) *satisfies*

$$\left\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\right\|_F = \widetilde{\mathcal{O}}\left(\frac{rK}{\sqrt{n}}\right)$$

with probability at least $1 - 4\exp(-\alpha)$ for any $\alpha > 0$, where $\lambda = c_\lambda \sigma(\alpha + \log(N+K))/\sqrt{n}$ for a positive constant c_λ.

We provide the proof in Appendix A.2. The nuclear norm regularized estimator yields an estimation error of $\widetilde{\mathcal{O}}(rK/\sqrt{n})$ in the Frobenius norm in our setting. In contrast, we will obtain a worse rate of $\mathcal{O}(K\sqrt{N/n})$, if we estimate each entry of the matrix independently using sampling average without sharing any latent information. Our result shows that matrix completion can be especially helpful in the high-dimensional setting when N and K are much larger than the rank r. Note that under independent sampling scheme, Negahban and Wainwright (2012) provide the minimax-optimal error bound in Frobenius norm as $\widetilde{\mathcal{O}}(K\sqrt{r/n})$[4]. Our error bound is comparable with that under independent sampling; we have an additional factor of $\sqrt{r}$, which is rather insignificant since $r \ll \min\{N, K\}$, due to the complexity introduced by the sampling interference in our matching setting.

Indeed, the presence of certain dependent structure among sampling can give rise to poor error rates in a matrix completion problem (Athey et al. 2021). Existing proof techniques in Klopp (2014), Hamidi and Bayati (2022) use contraction inequality; however, this proof strategy only works under their independent sampling scheme. In our matching sampling scheme with dependent structure, the same technique will induce a loose bound of $\widetilde{\mathcal{O}}(K\sqrt{rN/n})$ with an extra factor $\sqrt{N/r}$ ($\gg 1$); in other words, the error bound is similar to that of using sampling average to estimate matrix entries individually without matrix completion. Particularly, a key step in proving Proposition 1 (see Appendix A.5) is to bound an error term that captures the degree of non-convexity of the loss (i.e., the first part of (3.1)). Such an upper bound leads to the last term on the right hand side of our RSC condition in (3.3), and eventually the error bound in Theorem 1. Specifically, this small error term is

$$\mathbb{E}\left[\sup_{\boldsymbol{\Delta}\in\widetilde{\mathcal{C}}} \frac{1}{n}\sum_{t=1}^{n} \xi_t \Big(\sum_{i=1}^{N} \langle \mathbf{X}_t^i, \boldsymbol{\Delta}\rangle^2\Big)\right] \tag{3.4}$$

for some restricted set $\widetilde{\mathcal{C}}$ ($\subseteq \mathcal{C}_\alpha(r)$, defined in (3.2)), where $\mathbf{X}_t^i$ is defined in (2.2) and $\{\xi_t\}_{t\in[n]}$ is a sequence of Rademacher random variables. A contraction inequality (Maurer 2016) results in an upper bound with a non-trivial extra factor of N — i.e.,

$$N \cdot \mathbb{E}\left[\sup_{\boldsymbol{\Delta}\in\widetilde{\mathcal{C}}} \frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{N} \xi_{t,i} \langle \mathbf{X}_t^i, \boldsymbol{\Delta}\rangle\right],$$

[4] The rate is adjusted according to two differences in our setting. First, we observe nN samples in total. Second, we assume $\|\boldsymbol{\Theta}^*\|_\infty \leq 1$ while Negahban and Wainwright (2012) assume $\|\boldsymbol{\Theta}^*\|_F$ is upper bounded by a constant.

where $\{\xi_{t,i}\}_{t\in[n],i\in[N]}$ is a sequence of Rademacher random variables — and thus degrades the final Frobenius norm error bound.

To tackle this challenge, we propose a new "linearization" trick to refine the upper bound of (3.4) and thus obtain a near-optimal error guarantee. Particularly, our linearization technique leverages the sampling property of matrix completion to linearize and upper bound (3.4) — i.e., as $\mathbf{X}_t^i$ is a basis matrix with only one entry being 1, we have

$$\langle \mathbf{X}_t^i, \boldsymbol{\Delta}\rangle^2 = (\boldsymbol{\Delta}^{(i,j_t(i))})^2 = \langle \mathbf{X}_t^i, \boldsymbol{\Delta}\circ\boldsymbol{\Delta}\rangle,$$

where $\boldsymbol{\Delta}\circ\boldsymbol{\Delta}\in\mathbb{R}^{N\times K}$ is the Hadamard square of $\boldsymbol{\Delta}$ with $(\boldsymbol{\Delta}\circ\boldsymbol{\Delta})^{(i,j)} = (\boldsymbol{\Delta}^{(i,j)})^2$. Then, we can simply upper bound (3.4) with

$$\mathbb{E}\left[\sup_{\boldsymbol{\Delta}\in\widetilde{\mathcal{C}}}\frac{1}{n}\sum_{t=1}^{n}\xi_t\sum_{i=1}^{N}\langle \mathbf{X}_t^i, \boldsymbol{\Delta}\circ\boldsymbol{\Delta}\rangle\right].$$

We further manage to control this upper bound without introducing additional significant factors by using a property of Hadamard square, i.e., the Hadamard square of a low-rank matrix is also low-rank. This simple trick, in place of contraction inequality, provides a tighter guarantee for the convexity of our loss function in our RSC condition, and hence improves our error bound by replacing a significant factor of $\sqrt{N}$ with only a much smaller $\sqrt{r}$ ($\ll\sqrt{N}$).

3.3. Minimax Lower Bound

We establish the minimax lower bound in Frobenius norm for our matrix completion approach for matching problems. We show that our nuclear norm regularized estimator is minimax optimal since its upper bound in Theorem 1 matches the lower bound we provide below in Theorem 2.

We first define the minimax risk of our problem in the Frobenius norm as

$$\ell(\boldsymbol{\Theta}^*, \|\cdot\|_F) = \inf_{\widetilde{\boldsymbol{\Theta}}}\sup_{\boldsymbol{\Theta}^*\in\mathcal{C}}\mathbb{E}\left[\|\widetilde{\boldsymbol{\Theta}}-\boldsymbol{\Theta}^*\|_F\right],$$

where $\mathcal{C} = \{\boldsymbol{\Theta}\in\mathbb{R}^{N\times K} \mid \mathsf{rank}(\boldsymbol{\Theta}) = r, \|\boldsymbol{\Theta}\|_\infty \le 1\}$ and the infimum ranges over all possible estimators. The minimax risk measures in principle the complexity of estimating any unknown matrix $\boldsymbol{\Theta}^*$ satisfying our assumptions. We provide a lower bound for the minimax risk as follows.

THEOREM 2. *The minimax risk of $\boldsymbol{\Theta}^*$ satisfies*

$$\ell(\boldsymbol{\Theta}^*, \|\cdot\|_F) = \Omega\left(K\sqrt{\frac{r}{n}}\right).$$

We provide the proof in Appendix B. The proof strategy is similar to that of Theorem 3 in Negahban and Wainwright (2012) and Theorem 5 in Koltchinskii et al. (2011), which provide a minimax lower bound under the independent sampling scheme. The result shows that the minimax lower bound of the matrix completion problem in our matching setting scales as $\Omega(K\sqrt{r/n})$; note that this bound is the same as the lower bound provided in Negahban and Wainwright (2012) for independent sampling scheme[5]. Compared to our upper bound in Theorem 1, this lower bound suggests that our estimator in (3.1) is minimax optimal up to logarithmic terms and insignificant factors such as $\sqrt{r}$ ($\ll \min\{N, K\}$).

4. Double-Enhancement Procedure

In this section, we show how to obtain a desired entry-wise guarantee, which is essential to some downstream matching decision-making processes such as online stable matching discussed in Section 5.2. We first design a double-enhancement procedure to produce an enhanced estimator (Section 4.1). We then provide an entry-wise error bound for this enhanced estimator (Section 4.2), which is typically harder than the Frobenius error bound in Section 3.

4.1. Double-Enhancement Design

As aforementioned, to guide certain downstream decision making of matching, it is important to control the statistical uncertainty of the entry-wise estimates in ℓ_∞ norm. For instance, entry-wise estimates enable learning the total reward of one matching or the preference rankings of the market sides, which can be crucial for decision making of matching discussed in Section 5. To that end, we provide a double-enhancement technique atop the nuclear norm regularization to further refine the entry-wise estimation. The procedure is summarized in Algorithm 1.

Our double-enhancement procedure is motivated by a row-enhancement design in Hamidi et al. (2019). However, our matching context differs from their problem in two aspects. First, we consider a matrix completion problem, where our observation operator in (2.5) only reveals entry-wise information of matching rewards. In contrast, Hamidi et al. (2019) consider a matrix factorization problem for contextual bandits; their observation operator reveals information induced by Gaussian contexts and thus provides more information than ours. Second, they only provide row-wise error controls, while we need an entry-wise error guarantee, which is much harder than obtaining row-wise or Frobenius norm error bounds (Chen et al. 2020). Typically, the standard analytical techniques quantify the uncertainty of the matrix estimates as a whole in Frobenius norm, such as our bound in our Theorem 1. This motivates us to design a new analytical approach to sharpen the entry-wise estimation through carefully characterizing and enhancing the estimated row and column spaces of the matrix.

Particularly, our double-enhancement procedure achieves this goal through the following steps based on the nuclear norm regularized estimates. First, we split the whole n matchings into two

[5] Analogous to Footnote 4, the rate is adjusted according to our sample size of nN and our assumption $\|\boldsymbol{\Theta}^*\|_\infty \leq 1$.

Algorithm 1 Double-Enhancement

```
  Inputs: $\lambda$
  Set $n_1 = \lfloor n/2 \rfloor$, $\mathcal{J}_1 = [n_1]$, $\mathcal{J}_2 = [n] \setminus [n_1]$
  Calculate $\widehat{\boldsymbol{\Theta}}$ in (3.1) using the data $\{(\mathbf{X}_t^i, Y_t^{(i)}) \mid t \in \mathcal{J}_1, i \in [N]\}$
  Compute the SVD $\widehat{\boldsymbol{\Theta}} = \widehat{\mathbf{U}}\widehat{\mathbf{D}}\widehat{\mathbf{V}}^\top$
  for $i \in [N]$ do
      Set $\mathcal{R}_i = \{(\mathbf{X}_t^i, Y_t^{(i)}) \mid t \in \mathcal{J}_2\}$
      Compute $\widetilde{\beta}_i = \arg\min_{\gamma \in \mathbb{R}^r} \left\{ \sum_{(\mathbf{X}, y) \in \mathcal{R}_i} (y - \mathbf{X}^{(i,\cdot)} \widehat{\mathbf{V}} \gamma)^2 \right\}$
  end for
  for each $j \in [K]$ do
      Set $\mathcal{C}_j = \{(\mathbf{X}_t^i, Y_t^{(i)}) \mid t \in \mathcal{J}_2, i \in [N], j_t(i) = j\}$
      Compute $\widetilde{\alpha}_j = \arg\min_{\gamma \in \mathbb{R}^r} \left\{ \sum_{(\mathbf{X}, y) \in \mathcal{C}_j} (y - \widehat{\mathbf{U}}^\top \mathbf{X}^{(\cdot,j)} \gamma^\top)^2 \right\}$
  end for
  Let $\widetilde{\mathbf{U}} = \left[\widetilde{\beta}_1 \ \widetilde{\beta}_2 \ \cdots \ \widetilde{\beta}_N\right]^\top$, $\widetilde{\mathbf{V}} = \left[\widetilde{\alpha}_1 \ \widetilde{\alpha}_2 \ \cdots \ \widetilde{\alpha}_K\right]^\top$
  Compute the SVD $\widetilde{\mathbf{U}} = \mathbf{U}_1 \mathbf{D}_1 \mathbf{Q}_1$
  Compute $\widetilde{\boldsymbol{\Theta}} = \mathbf{U}_1 \mathbf{D}_1 \widetilde{\mathbf{V}}^\top$
  Outputs: $\widetilde{\boldsymbol{\Theta}}$
```

subsets $\mathcal{J}_1$ and $\mathcal{J}_2$, where we estimate a nuclear norm regularized estimator $\widehat{\boldsymbol{\Theta}}$ via (3.1) using the samples in $\mathcal{J}_1$. Next, we refine the estimation of the row and column spaces of $\boldsymbol{\Theta}^*$ alternatively using the remaining samples in $\mathcal{J}_2$. Let $\boldsymbol{\Theta}^* = \mathbf{U}^* \mathbf{D}^* \mathbf{V}^{*\top}$ be the singular value decomposition (SVD) of the true matrix $\boldsymbol{\Theta}^*$ with $\mathbf{U}^*$ and $\mathbf{V}^*$ being two orthogonal matrices; then, the row and column spaces of $\boldsymbol{\Theta}^*$ refer to the subspace spanned by the columns of $\mathbf{V}^*$ and columns of $\mathbf{U}^*$ respectively. Note that our matching model (2.4)

$$Y_t^{(i)} = \langle \mathbf{X}_t^i, \boldsymbol{\Theta}^* \rangle + \varepsilon_t^{(i)} = \langle \mathbf{X}_t^{i(i,\cdot)}, \boldsymbol{\Theta}^{*(i,\cdot)} \rangle + \varepsilon_t^{(i)} = \underbrace{\mathbf{X}_t^{i(i,\cdot)} \mathbf{V}^*}_{\text{features}} \underbrace{(\mathbf{U}^{*(i,\cdot)} \mathbf{D}^*)^\top}_{\text{parameters}} + \varepsilon_t^{(i)} \tag{4.1}$$

can be represented by a standard linear regression model, where $\mathbf{X}_t^{i(i,\cdot)} \mathbf{V}^*$ is the feature vector and $\mathbf{U}^{*(i,\cdot)} \mathbf{D}^*$ is the unknown parameter vector. Since we have no direct access to $\mathbf{V}^*$ in the feature vector, we approximate it with the orthogonal matrix $\widehat{\mathbf{V}}$ from the SVD of the estimator $\widehat{\boldsymbol{\Theta}} = \widehat{\mathbf{U}}\widehat{\mathbf{D}}\widehat{\mathbf{V}}^\top$. Now, we can enhance (i.e., first enhancement) the row space of the nuclear norm regularized estimate $\widehat{\boldsymbol{\Theta}}$ by estimating $\mathbf{U}^{*(i,\cdot)} \mathbf{D}^*$ using linear regression; the corresponding least square estimate $\widetilde{\mathbf{U}}$ enjoys a tighter row-wise guarantee in $\ell_{2,\infty}$ norm, i.e., $\|\widetilde{\mathbf{U}} - \mathbf{U}^* \mathbf{D}^*\|_{2,\infty} \approx \|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_F / \sqrt{N}$. Similarly, we can also enhance (i.e., second enhancement) the column space with an estimate $\widetilde{\mathbf{V}}$ that satisfies $\|\widetilde{\mathbf{V}} - \mathbf{V}^* \mathbf{D}^*\|_{2,\infty} \approx \|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_F / \sqrt{K}$. Finally, our enhanced estimator $\widetilde{\boldsymbol{\Theta}}$

is built upon $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$, obtained through our double-enhancement procedure. The specific design of $\widetilde{\mathbf{\Theta}}$ allows us to obtain an entry-wise error bound through the row-wise bounds of $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$, considering that $\|\mathbf{A}\mathbf{B}^\top\|_\infty \le \|\mathbf{A}\|_{2,\infty} \cdot \|\mathbf{B}\|_{2,\infty}$ for any matrices $\mathbf{A}, \mathbf{B}$. In the next section, we provide a theoretical guarantee on the entry-wise error bound for $\widetilde{\mathbf{\Theta}}$; we will show that the entry-wise error bound scales as $1/\sqrt{NK}$ of the Frobenius norm error $\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_F$ provided in Theorem 1.

4.2. Entry-Wise Estimation Error Bound

Our entry-wise estimation error bound of the enhanced estimator $\widetilde{\mathbf{\Theta}}$, introduced in Section 4.1, holds under a standard *spikiness* condition in the matrix completion literature (Negahban and Wainwright 2012, Hamidi and Bayati 2022).

Assumption 1 **(Spikiness)**. *There exists a constant $\eta \ge 1$ such that*

$$\frac{\sqrt{NK}\|\mathbf{\Theta}^*\|_\infty}{\|\mathbf{\Theta}^*\|_F} \le \eta.$$

Intuitively, the spikiness condition excludes matrices with overly large values in a few entries. If an unknown matrix is spiky, i.e., it has a few entries with extremely large values, then we cannot accurately estimate all matrix entries using matrix completion without observing all the entries (Candes and Plan 2010, Negahban and Wainwright 2012, Hamidi and Bayati 2022). For example, consider a matrix $\mathbf{\Theta}^*$ with $\mathbf{\Theta}^{*(1,1)} = 1$ and other entries 0. This matrix does not satisfy our assumption since $\sqrt{NK}\|\mathbf{\Theta}^*\|_\infty / \|\mathbf{\Theta}^*\|_F = \sqrt{NK}$, which cannot be bounded by a constant. Note that it is impossible to recover the entry $\mathbf{\Theta}^{*(1,1)}$ and hence maintain small entry-wise error without observing rewards from this entry. In other words, the spikiness condition ensures the entry-wise identifiability of the matrix given any random data samples.

Remark 2. The spikiness condition is related to another popular incoherence condition (Candes and Recht 2008, Candes and Plan 2010, Chen et al. 2020), defined as

$$\|\mathbf{U}^*\|_{2,\infty} \le \mu\frac{\sqrt{r}}{N}, \quad \|\mathbf{V}^*\|_{2,\infty} \le \mu\frac{\sqrt{r}}{K},$$

for some constant μ in our setting, where $\mathbf{U}^*$ and $\mathbf{V}^*$ are from the SVD $\mathbf{\Theta}^* = \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\top}$. Intuitively, the incoherence condition implies that all rows of $\mathbf{U}^*$ and $\mathbf{V}^*$ are of similar scales, and thus prevents $\mathbf{\Theta}^*$ from being spiky. We note that our results still hold under the incoherence condition.

Next, we state the following result of an entry-wise error bound of the enhanced estimator $\widetilde{\mathbf{\Theta}}$ under the spikiness assumption.

Theorem 3. *Suppose $n = \widetilde{\Omega}(\max\{(r^4K)/N, (K/N)^2\})$. The enhanced estimator $\widetilde{\mathbf{\Theta}}$ in Algorithm 1 satisfies*

$$\|\widetilde{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_\infty = \widetilde{\mathcal{O}}\left(r^2\sqrt{\frac{K}{Nn}}\right),$$

with probability at least $1-(3N+3K+5)\exp(-\alpha)$ *for any* $\alpha > 0$*, where* $\lambda = c_\lambda(\alpha+\log(N+K))/\sqrt{n}$ *for some positive constant* c_λ*.*

The proof is provided in Appendix C. According to Theorem 3, the entry-wise estimation error $\|\widetilde{\mathbf{\Theta}}-\mathbf{\Theta}^*\|_\infty$ for the enhanced estimator $\widetilde{\mathbf{\Theta}}$ in our setting is of order $\widetilde{\mathcal{O}}(r^2\sqrt{K/(nN)})$. In comparison, the nuclear norm regularized estimator $\widehat{\mathbf{\Theta}}$ has no entry-wise guarantee other than a trivial one using Theorem 1, i.e., $\|\widehat{\mathbf{\Theta}}-\mathbf{\Theta}^*\|_\infty \le \|\widehat{\mathbf{\Theta}}-\mathbf{\Theta}^*\|_F = \widetilde{\mathcal{O}}(rK/\sqrt{n})$. Thus, our double-enhancement design saves us a significant factor of $\sqrt{NK}/r$ ($\ll 1$) in the entry-wise error control. As aforementioned in Section 1.1, some existing literature, such as Chen et al. (2020, 2021), have also derived an entry-wise error bound. However, their approach exploits the unique property of the independent sampling scheme, and thus does not apply to our setting with sampling interference. Our double-enhancement procedure in Algorithm 1 might be of independent interest; it can be readily applied to enhance existing estimators and obtain entry-wise guarantees in broader sampling schemes, including the independent sampling scheme in Chen et al. (2020, 2021).

Remark 3. We note that our entry-wise error bound in Theorem 3 also matches its minimax lower bound up to logarithmic terms and insignificant factors. The proof of an entry-wise minimax risk is similar to that of our Theorem 2.

5. Online Learning in the Matching Market

In this section, we extend our offline matrix completion approach to the online learning setting, where a centralized platform needs to learn from adaptively collected data and make sequential matching decisions with no prior information. We discuss both the optimal matching (Section 5.1) and the stable matching (Section 5.2). We propose two algorithms respectively that speed up the learning process by reducing exploration cost, which can be especially useful under short horizons and in large matching markets.

5.1. Online Optimal Matching

We first discuss an online optimal matching problem, which is usually formulated as a combinatorial semi-bandit problem in the literature (Gai et al. 2010, Chen et al. 2013, Kveton et al. 2015).

Problem Formulation. Analogous to our offline setting in Section 2.1, we consider a two-sided matching platform with N worker types and K job types, and $\mathbf{\Theta}^* \in \mathbb{R}^{N\times K}$ represents their reward matrix. In each step t of a time horizon T, the platform chooses a matching (an arm) $\pi_t = \mathbf{X}_t$ from the set of all matchings $\mathcal{M}$ (defined in (2.1)) based on all historical information, and receives noisy rewards Y_t from all N matched pairs (defined in (2.4)). We want to learn the unknown reward matrix $\mathbf{\Theta}^*$ and maximize the total expected reward in each matching. We compare our policy π to an optimal matching $\mathbf{X}^*$ that obtains the maximum total reward among all matchings, i.e.,

$\mathbf{X}^* = \arg\max_{\mathbf{X}\in\mathcal{M}}\langle \mathbf{X}, \mathbf{\Theta}^*\rangle$[6]. Our goal is to learn a policy π to minimize the cumulative regret over time

$$R(T) = \sum_{t=1}^{T}\Big(\langle \mathbf{X}^*, \mathbf{\Theta}^*\rangle - \langle \mathbf{X}_t, \mathbf{\Theta}^*\rangle\Big), \tag{5.1}$$

where $\langle \mathbf{X}^*, \mathbf{\Theta}^*\rangle - \langle \mathbf{X}_t, \mathbf{\Theta}^*\rangle$ is the regret at time t.

Algorithm Design. We propose a **Comb**inatorial **L**ow-**R**ank **B**andit (CombLRB) algorithm in Algorithm 2 for the online optimal matching problem. CombLRB exploits the low-rank structure of the reward matrix $\mathbf{\Theta}^*$ to accelerate reward learning in the exploration phase; it incorporates the nuclear norm regularization approach formulated in Section 3.1. Our algorithm has a two-stage design of exploration and exploitation. First, it explores for E_h time periods in the early stage. For each $t \in [E_h]$, it draws a matching $\mathbf{X}_t$ following the uniform distribution Π over $\mathcal{M}$ (defined in Section 2.1). Then, at the end of the exploration, we calculate a nuclear norm regularized estimator $\widehat{\mathbf{\Theta}}$ based on the data we have collected in the exploration phase, i.e., $\{(\mathbf{X}_t^i, Y_t^{(i)}) \mid t \in [E_h], i \in [N]\}$. Our algorithm identifies a matching $\mathbf{X}_c \in \mathcal{M}$ that maximizes the total reward using $\widehat{\mathbf{\Theta}}$ as a surrogate for $\mathbf{\Theta}^*$, i.e., $\mathbf{X}_c = \arg\max_{\mathbf{X}\in\mathcal{M}}\langle \mathbf{X}, \widehat{\mathbf{\Theta}}\rangle$. Finally, our algorithm commits to the matching $\mathbf{X}_c$ and keeps playing this arm for the remaining time periods.

Algorithm 2 Combinatorial Low-Rank Bandit (CombLRB)

```
Inputs: $E_h$, $\lambda$
for $t \in [E_h]$ do
    Choose matching $\pi_t = \mathbf{X}_t \sim \Pi$
    Observe rewards $Y_t^{(i)} = \langle \mathbf{X}_t^i, \mathbf{\Theta}^*\rangle + \varepsilon_t^{(i)}$ for all $i \in [N]$
end for
Calculate $\widehat{\mathbf{\Theta}}$ in (3.1) using the data $\{(\mathbf{X}_t^i, Y_t^{(i)}) \mid t \in [E_h], i \in [N]\}$
Compute $\mathbf{X}_c = \arg\max_{\mathbf{X}\in\mathcal{M}}\langle \mathbf{X}, \widehat{\mathbf{\Theta}}\rangle$
for $t \in [T] \setminus [E_h]$ do
    Choose matching $\pi_t = \mathbf{X}_c$
end for
```

Our next theorem provides a regret upper bound on our CombLRB Algorithm in the online optimal matching setting.

THEOREM 4. *The regret of CombLRB in Algorithm 2 has*

$$\mathbb{E}[R(T)] = \widetilde{\mathcal{O}}\left(r(N+K)T^{2/3}\right),$$

[6] Given $\mathbf{\Theta}^*$, $\mathbf{X}^*$ can be efficiently calculated by many well-established algorithms such as Hungarian algorithm (Kuhn 1955) and Munkres algorithm (Munkres 1957).

where $E_h = \mathcal{O}(rT^{2/3})$ and $\lambda = c_\lambda \sigma \log(N(N+K)T)/\sqrt{E_h}$ for some positive constant c_λ.

We provide a proof in Appendix D. We compare the regret bound of our approach with that of a state-of-the-art CUCB algorithm Kveton et al. (2015). CUCB builds on the upper confidence bound insight from the bandit literature and estimates each matrix entry using its sample average individually. In other words, CUCB does not exploit the low-rank structure of the rewards and thus cannot gain efficiency through information sharing as we do. This approach yields a regret upper bound of $\widetilde{\mathcal{O}}(\sqrt{N^2KT})$. In contrast, our CombLRB algorithm achieves an improved regret bound for short time horizon T or large market with large values of N and K. Specifically, when $T = \mathcal{O}(K^3)$, our bound is strictly better than that of CUCB since we obtain a favorable dependency on the matrix dimensions N and K. Indeed, the low-rankness of the reward matrix facilitates the reward learning and hence the matching decisions through only few explorations. It can be especially helpful for short horizons or high-dimensional contexts, when reward learning is very costly.

It is also worth noting that our algorithm can run for any extremely short time horizon $T < K$ as well, while CUCB requires at least K initial random matchings for exploration and only works for relatively long time horizons.

5.2. Online Stable Matching

Next, we discuss an online stable matching problem; our problem setting is the same as that formulated by Liu et al. (2020).

Problem Formulation. Consider a two-sided platform with N worker types and K job types. Unlike all previous settings, now the matrix $\mathbf{\Theta}^*$ represents the rewards of the worker side, which further implies the worker preference rankings over jobs. Particularly, its $(i,j)^{\text{th}}$ entry $\mathbf{\Theta}^{*(i,j)}$ denotes the reward received by a worker of type i if they are matched with a job of type j. Then, the preference ranking of worker type i over the K job types is determined by the i^{th} row of $\mathbf{\Theta}^*$; that is, they prefer a job of type j over j' if $\mathbf{\Theta}^{*(i,j)} > \mathbf{\Theta}^{*(i,j')}$. We encode the job side preferences in columns of another matrix $\mathbf{\Phi}^* \in \mathbb{R}^{N\times K}$. The platform initially does not know the worker rewards (i.e., $\mathbf{\Theta}^*$), hence has to learn their preferences online, whereas job preferences (i.e., $\mathbf{\Phi}^*$) are known in advance.

At each time t of a time horizon T, the platform chooses a matching $\pi_t = \mathbf{X}_t \in \mathcal{M}$ and observes the noisy rewards Y_t received by the N worker types formulated by (2.4). We want to learn the unknown matrix $\mathbf{\Theta}^*$ of worker preferences and find the worker-optimal stable matching $\mathbf{X}^*$; particularly, $\mathbf{X}^*$ is the stable matching returned by the Gale-Shapley (GS) Algorithm (Gale and Shapley 1962) when the workers are the proposing side. We note that $\mathbf{X}^*$ is optimal among all stable matchings $\mathcal{S}$ ($\subseteq \mathcal{M}$) for all worker types (Knuth 1997), i.e., $\langle \mathbf{X}^{*i}, \mathbf{\Theta}^* \rangle \geq \langle \mathbf{X}^i, \mathbf{\Theta}^* \rangle, \forall \mathbf{X} \in \mathcal{S}$ with $\mathbf{X}^i$ defined in (2.2). We design a policy π to minimize the worker-optimal stable regret

$$R_i(T) = \sum_{t=1}^{T} \Big(\langle \mathbf{X}^{*i}, \mathbf{\Theta}^* \rangle - \langle \mathbf{X}_t^i, \mathbf{\Theta}^* \rangle \Big), \tag{5.2}$$

for every worker $i \in [N]$, where $\langle \mathbf{X}^{*i}, \mathbf{\Theta}^* \rangle - \langle \mathbf{X}_t^i, \mathbf{\Theta}^* \rangle$ is the regret for workers of type i at time t.

Algorithm Design. We develope a **Comp**eting **L**ow-**R**ank **B**andit (CompLRB) algorithm as in Algorithm 3. The design of CompLRB closely follows our CombLRB algorithm for online optimal matching but with two distinctions. Note that the regret (5.2) captures the entry-wise value difference of $\mathbf{\Theta}^*$, and thus can be bounded more tightly using the enhanced estimator $\widetilde{\mathbf{\Theta}}$ from our double-enhancement procedure in Section 4.1. Additionally, since our goal is to find an optimal policy among stable matchings, our algorithm identifies a worker-optimal stable matching $\mathbf{X}_c$ using GS algorithm based on $\widetilde{\mathbf{\Theta}}$ and $\mathbf{\Phi}^*$.

Algorithm 3 Competing Low-Rank Bandit (CompLRB)

```
  Input: $E_h$, $\lambda$, $\mathbf{\Phi}^*$
for $t \in [E_h]$ do
    Choose matching $\pi_t = \mathbf{X}_t \sim \Pi$
    Observe rewards $Y_t^{(i)} = \langle \mathbf{X}_t^i, \mathbf{\Theta}^* \rangle + \varepsilon_t^{(i)}$ for all $i \in [N]$
end for
Calculate $\widetilde{\mathbf{\Theta}}$ in Algorithm 1 using the data $\{(\mathbf{X}_t^i, Y_t^{(i)}) \mid t \in [E_h], i \in [N]\}$
Compute a stable matching $\mathbf{X}_c$ via GS algorithm with inputs $\widetilde{\mathbf{\Theta}}$ and $\mathbf{\Phi}^*$
for $t \in [T] \setminus [E_h]$ do
    Choose matching $\pi_t = \mathbf{X}_c$
end for
```

We state the regret upper bound of our CompLRB algorithm for online stable matching as follows.

Theorem 5. *Let $\Delta_{\min} = \min_{i \in [N]} \{\min_{j \neq j'} |\mathbf{\Theta}^{*(i,j)} - \mathbf{\Theta}^{*(i,j')}|\}$. Then, the regret of CompLRB in Algorithm 3 for worker type i has*

$$\mathbb{E}[R_i(T)] = \mathcal{O}\left(\frac{r^3 K \max\{\log^2[(N+K)T], r\log[(N+K)T]\}}{N\Delta_{\min}^2} \right),$$

where $E_h = \mathcal{O}\left(\frac{r^3 K \max\{\log^2[(N+K)T], r\log[(N+K)T]\}}{N\Delta_{\min}^2} \right)$ and $\lambda = c_\lambda \sigma \log\big[(N+K)(3N+3K+5)T\big]/\sqrt{E_h}$ for some positive constant c_λ.

We provide a proof in Appendix E. In comparison, Liu et al. (2020) provide an upper bound of $\mathcal{O}(K\log(NT)/\Delta_{\min}^2)$ for any worker type $i \in [N]$. Our regret bound improves upon that in Liu et al. (2020) in the matrix dimensions N and K by a factor of at least N/r^4 up to logarithmic terms (recall that $N \gg r$). Specifically, our algorithm exploits the tightness of the enhanced low-rank estimator on entry-wise errors, and achieves an improved performance through reduced explorations. Instead, the algorithm in Liu et al. (2020) estimates every entry of $\mathbf{\Theta}^*$ using a naive sample average estimate. Our theoretical result shows that CompLRB can be very useful for large markets with many participants, i.e., large N and K.

Remark 4. In practice, the job side preferences, i.e., $\mathbf{\Phi}^*$, might be unknown to the matching platform as well. Then, we need to learn both matrices $\mathbf{\Theta}^*$ and $\mathbf{\Phi}^*$ online. Our algorithm can be easily adapted to this setting, and will result in a regret bound of the same scale as in Theorem 5.

6. Experiments

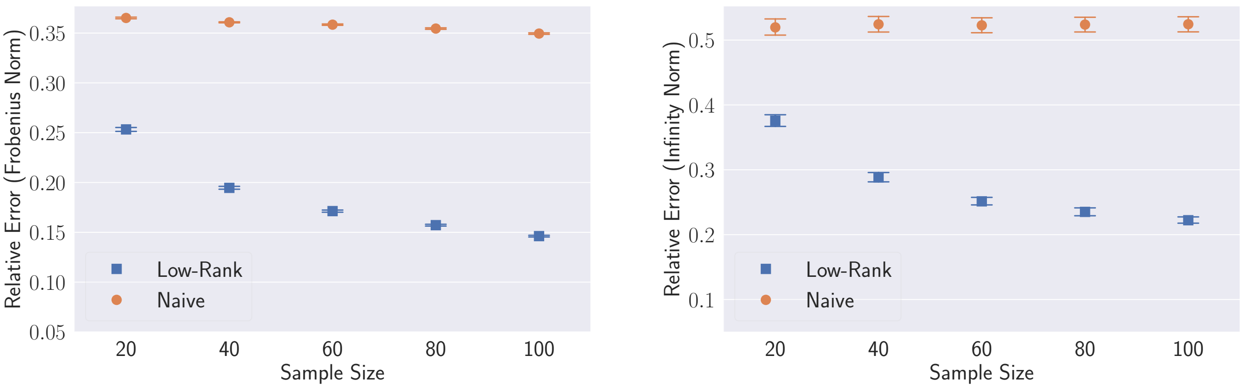
We now demonstrate the practical relevance and effectiveness of our proposed approaches in both offline and online settings using synthetic data and real data of labor market.

In the offline setting, we compare the following methods: (i) Low Rank: the nuclear norm regularization approach we propose, and (ii) Naive: an entry-wise sample-average estimator. In the online setting, we compare the following online optimal matching algorithms: (i) CombLRB: combinatorial low-rank bandit algorithm we propose, (ii) CUCB: combinatorial upper confidence bound algorithm developed by (Chen et al. 2013), and (iii) CTS: combinatorial thompson sampling proposed by (Wang and Chen 2018). For online stable matching, we compare the following algorithms: (i) CompLRB: competing low-rank bandit algorithm we propose, and (ii) CompB: competing bandit algorithm proposed by Liu et al. (2020).

6.1. Synthetic Data

Offline Matching Data. We synthetically generate the matching reward matrix $\mathbf{\Theta}^*$, and the details are provided in Appendix G.1. Figure 3 shows the relative estimation errors of learning the matching reward matrix $\mathbf{\Theta}^*$ in Frobenius norm and entry-wise norm respectively. The relative estimation error of a specified norm is equal to the estimation error of the matrix in that norm divided by the specified norm of the ground-truth matrix $\mathbf{\Theta}^*$.

We find that our proposed estimator Low-Rank substantially outperforms the benchmarks in both Frobenius norm and infinity norm. The entry-wise sample-average estimator Naive takes the entry-wise empirical mean as estimates for observed entries, and the average of observations from the same row for unobserved entries. Thus, Naive does not fully utilize the low-rank structure of the reward matrix, and might introduce additional bias for unobserved entries. Instead, our approach captures the low-rank structure of $\mathbf{\Theta}^*$ efficiently, and thus delivers much smaller estimation errors through shared information across entries. Our results are consistent over varying matching sample size n. Matching our theory, the estimation error of Low-Rank decreases with increasing sample size; in contrast, Naive converges slowly due to insufficient samples entry-wise and potential bias for unobserved entries.



(a) Frobenius Norm Error ($\|\cdot\|_F$)

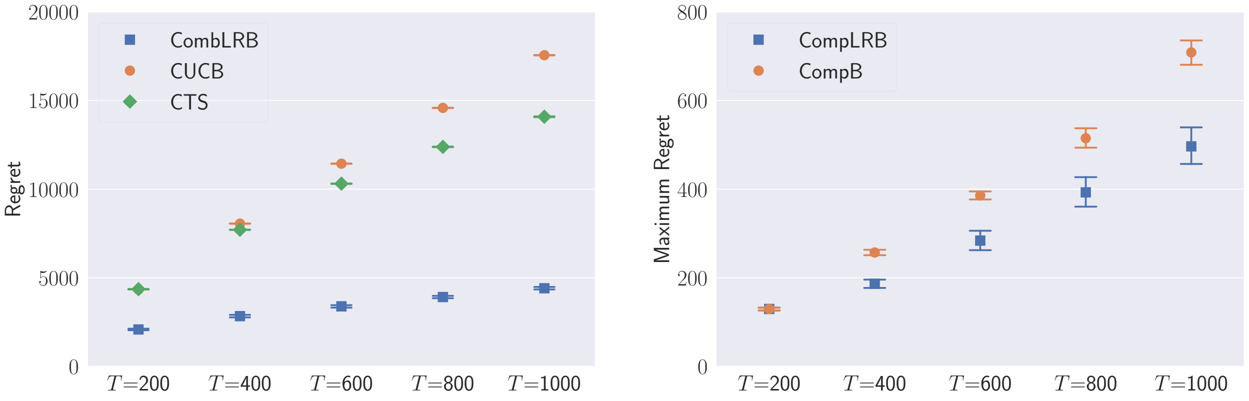
(b) Entry-Wise Error ($\|\cdot\|_\infty$)

Figure 3 Relative estimation errors of the matching reward matrix Θ^* in Frobenius norm (left) and infinity norm (right) averaged over 50 trials. Error bars represent 95% confidence intervals. We consider $N = 100$ worker types, $K = 100$ job types, and the matrix rank $r = 3$. Sample size on the x-axis refers to the number of matchings n. 'Low-Rank' represents our nuclear norm regularization approach.

Online Matching Algorithms. Figure 4 compares the cumulative regrets over varying time horizon T for optimal matching (Figure 4a) and stable matching (Figure 4b) respectively. Appendix G.1 provides more details.

Similar to the offline setting, we find in Figure 4a that our low-rank approach CombLRB significantly outperforms other benchmarks for optimal matching. CUCB and CTS are based on the ideas of upper confidence bound and Thompson sampling; they do not exploit the low-rank structure but instead learn the true reward of each arm individually. Thus, these algorithms cannot efficiently learn and identify the optimal matching given relatively short time horizons.

Figure 4b similarly shows that our algorithm CompLRB achieves much smaller cumulative regret for stable matching over other benchmarks. Note that here we consider the worst-case scenario with "maximum regret" — i.e., we compare the maximum of the N per-worker cumulative regrets for different algorithms. Since our algorithm has better performance in maximum regret, our algorithm also obtains a significant improvement over other benchmarks in total cumulative regrets of all workers in the market. As expected, our CompLRB obtains such an improvement through leveraging the underlying low-rank structure of the worker reward matrix, compared to CompB, which uses sample average to estimate the worker rewards and hence worker preferences.



(a) Optimal Matching

(b) Stable Matching

Figure 4 Regret for optimal matching (left) and maximum per-worker regret for stable matching (right) averaged over 50 trials. Error bars represent 95% confidence intervals. We consider $N = 100$ worker types, $K = 100$ job types, and the matrix rank $r = 3$. 'CombLRB' and 'CompLRB' represent our combinatorial low-rank bandit algorithm and competing low-rank bandit algorithm respectively.

In summary, the empirical results obtained from synthetic experiments align with our theory, given the low-rank nature of the true reward matrix. Next, we further explore the robustness of our algorithms on a real data, where the low-rank assumption might not hold.

6.2. Real Data of Labor Market

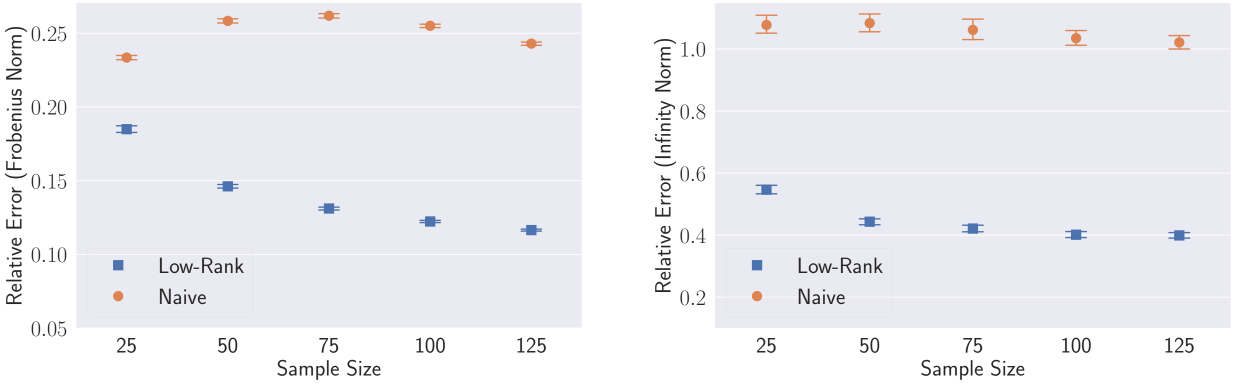
We further evaluate the real-world performance of all our approaches on one of the largest workforce dataset provided by Revelio Lab[7], which collects matching information for a diverse set of job and candidate profiles. We use a subset of the individual-level data of employment duration of mid-level software engineers in the United States from 2010 to 2015. In this experiment, we cluster the engineers (i.e., the workers) and companies (i.e., the jobs) into 50 groups respectively, i.e., $N = K = 50$. Then, we create the true reward matrix $\mathbf{\Theta}^*$, where the value of each entry takes the average of an indicator of whether an employment exceeds six months over all workers who belong to the corresponding worker and job group. That is, the $(i, j)^{\text{th}}$ entry $\mathbf{\Theta}^{*(i,j)}$ represents the probability of a worker employment from group i lasting long than six months in a company from group j. Intuitively, we measure the matching reward by the worker satisfaction; the higher the probability is, the more satisfied the worker is with the corresponding company.

It is worth noting that, in this real-world setting, our reward matrix might not satisfy the low-rank assumption, which is different from our synthetic experiments with low-rankness imposed. The details of data pre-processing and experimental setup are provided in Appendix G.2.

Offline Matching Data. Figure 5 presents the results of an offline reward learning experiment on our real data of labor market. The results again exhibit the superb performance of our low-rank

[7] See `https://wrds-www.wharton.upenn.edu/pages/about/data-vendors/revelio-labs/`.

matrix completion approach, similar to our synthetic experiments. Notably, our approach improves upon the best benchmark in the Frobenius norm and entry-wise norm by 44% and 43% respectively on average across all different sample sizes. Basically, our approach can learn the worker satisfaction with only few samples, and thus gain early insights into a company's employment condition.



(a) Frobenius Norm Error ($\|\cdot\|_F$)

(b) Entry-Wise Error ($\|\cdot\|_\infty$)

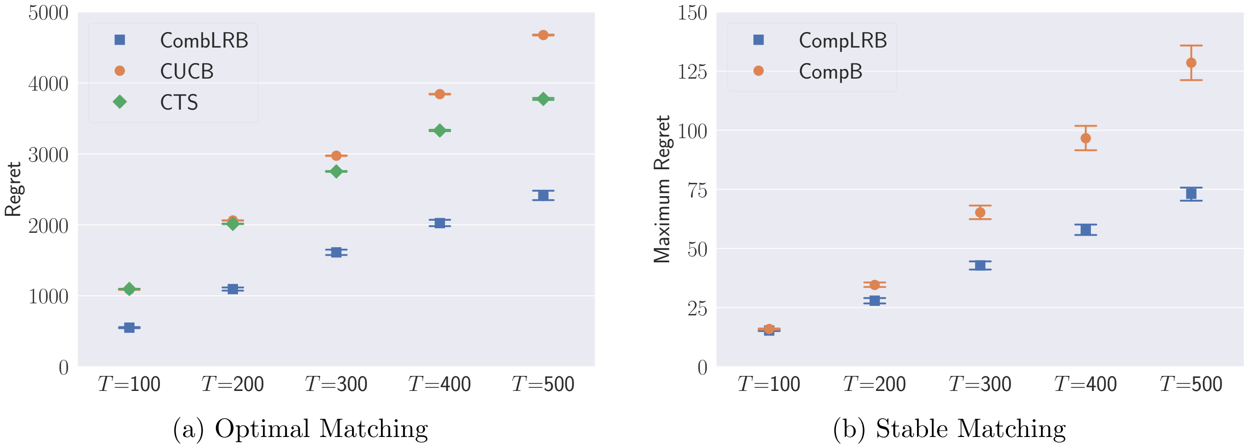
Figure 5 Relative estimation errors of the matching reward matrix Θ^* in Frobenius norm (left) and infinity norm (right) averaged over 50 trials. Error bars represent 95% confidence intervals. We consider $N = 50$ worker types, and $K = 50$ job types. Sample size on the x-axis refers to the number of matchings n. 'Low-Rank' represents our nuclear norm regularization approach.

Online Matching Algorithms. We also aim to learn the matching decisions directly in an online manner through our matrix completion approach. Note that the total number of rewards (i.e., $N \times K = 2500$) to learn is much larger than the time horizon T considered in our experiments; that is, we have relatively short time horizons and large matching markets.

The results are presented in Figure 6. Figure 6a shows that, even when the reward matrix might not be low-rank, our CombLRB algorithm outperforms other two benchmark algorithms CUCB and CTS over varying time horizons for online optimal matching. Specifically, CombLRB improves on the regret by 41% compared to the best performing CTS among the benchmarks. Figure 6b further confirms the efficiency of using low-rank matrix completion in the online stable matching setting. Our algorithm CompLRB beats the benchmark algorithm by 36% on average across all different time horizons.

7. Conclusion

In this paper, we focus on efficiently learning matching qualities from a small amount of offline matching data for large-scale centralized matching platforms. Motivated by the natural low-rank matrix structure of two-sided markets, we utilize a matrix completion approach via nuclear norm



(a) Optimal Matching

(b) Stable Matching

Figure 6 Regret for optimal matching (left) and maximum per-worker regret for stable matching (right) averaged over 50 trials. Error bars represent 95% confidence intervals. We consider $N = 50$ worker types, and $K = 50$ job types. 'CombLRB' and 'CompLRB' represent our combinatorial low-rank bandit algorithm and competing low-rank bandit algorithm respectively.

regularization to estimate the matching rewards efficiently. To the best of our knowledge, we propose the first matrix completion framework to address a reward learning problem in the matching setting. Our matching problem involves a challenging dependent sampling scheme due to matching interference; we develop a new proof strategy based on a linearization trick and establish a near-optimal error bound in Frobenius norm. Furthermore, we propose a novel double-enhancement procedure that refines entry-wise estimation atop the nuclear norm regularized estimates and ensures an entry-wise guarantee. In the online setting, we propose two algorithms CombLRB and CompLRB to efficiently learn optimal matching and stable matching policies respectively, thereby improving regret bounds in the matrix dimensions. Finally, our empirical experiments show that our matrix completion approach can indeed boost both offline reward learning and online decision making in the matching problems. Both our theoretical and empirical findings underscore the importance of adopting matrix completion methods in matching markets.

References

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, Khashayar Khosravi. 2021. Matrix completion methods for causal panel data models. *Journal of the American Statistical Association* **116**(536) 1716–1730.

Auer, Peter, Nicolo Cesa-Bianchi, Paul Fischer. 2002. Finite-time analysis of the multiarmed bandit problem. *Machine learning* **47** 235–256.

Baby, Dheeraj, Soumyabrata Pal. 2024. Online matrix completion: A collaborative approach with hott items. *arXiv preprint arXiv:2408.05843* .

Bayati, Mohsen, Junyu Cao, Wanning Chen. 2022. Speed up the cold-start learning in two-sided bandits with many arms. *arXiv preprint arXiv:2210.00340* .

Belavina, Elena, Karan Girotra, Ken Moon, Jiding Zhang. 2020. Matching in labor marketplaces: The role of experiential information. *Available at SSRN 3543906* .

Bell, Robert M, Yehuda Koren. 2007. Lessons from the netflix prize challenge. *Acm Sigkdd Explorations Newsletter* **9**(2) 75–79.

Bimpikis, Kostas, Mihalis G Markakis. 2019. Learning and hierarchies in service systems. *Management Science* **65**(3) 1268–1285.

Bühlmann, Peter, Sara Van De Geer. 2011. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Candes, Emmanuel J, Yaniv Plan. 2010. Matrix completion with noise. *Proceedings of the IEEE* **98**(6) 925–936.

Candes, Emmanuel J, Benjamin Recht. 2008. Exact low-rank matrix completion via convex optimization. *2008 46th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 806–812.

Candès, Emmanuel J, Terence Tao. 2010. The power of convex relaxation: Near-optimal matrix completion. *IEEE transactions on information theory* **56**(5) 2053–2080.

Cen, Sarah H, Devavrat Shah. 2022. Regret, stability & fairness in matching markets with bandit learners. *International Conference on Artificial Intelligence and Statistics*. PMLR, 8938–8968.

Chapelle, Olivier, Lihong Li. 2011. An empirical evaluation of thompson sampling. *Advances in neural information processing systems* **24**.

Chen, Wei, Yajun Wang, Yang Yuan. 2013. Combinatorial multi-armed bandit: General framework and applications. *International conference on machine learning*. PMLR, 151–159.

Chen, Yilun, Yash Kanoria, Akshit Kumar, Wenxin Zhang. 2023. Feature based dynamic matching. *Available at SSRN 4451799* .

Chen, Yudong, Martin J Wainwright. 2015. Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees. *arXiv preprint arXiv:1509.03025* .

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

Chen, Yuxin, Jianqing Fan, Cong Ma, Yuling Yan. 2021. Bridging convex and nonconvex optimization in robust pca: Noise, outliers, and missing data. *Annals of statistics* **49**(5) 2948.

Cuvelier, Thibaut, Richard Combes, Eric Gourdin. 2021. Statistically efficient, polynomial-time algorithms for combinatorial semi-bandits. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* **5**(1) 1–31.

Farias, Vivek F, Andrew A Li. 2019. Learning preferences with side information. *Management Science* **65**(7) 3131–3149.

Farias, Vivek F, Andrew A Li, Tianyi Peng. 2024. Fixing inventory inaccuracies at scale. *Manufacturing & Service Operations Management* **26**(3) 1102–1118.

Gai, Yi, Bhaskar Krishnamachari, Rahul Jain. 2010. Learning multiuser channel allocations in cognitive radio networks: A combinatorial multi-armed bandit formulation. *2010 IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN)*. IEEE, 1–9.

Gale, David, Lloyd S Shapley. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly* **69**(1) 9–15.

Hamidi, Nima, Mohsen Bayati. 2022. On low-rank trace regression under general sampling distribution. *Journal of Machine Learning Research* **23**(321) 1–49.

Hamidi, Nima, Mohsen Bayati, Kapil Gupta. 2019. Personalizing many decisions with high-dimensional covariates. *Advances in Neural Information Processing Systems* **32**.

Hsu, Wei-Kang, Jiaming Xu, Xiaojun Lin, Mark R Bell. 2022. Integrated online learning and adaptive control in queueing systems with uncertain payoffs. *Operations Research* **70**(2) 1166–1181.

Jagadeesan, Meena, Alexander Wei, Yixin Wang, Michael Jordan, Jacob Steinhardt. 2021. Learning equilibria in matching markets from bandit feedback. *Advances in Neural Information Processing Systems* **34** 3323–3335.

Jain, Prateek, Soumyabrata Pal. 2022. Online low rank matrix completion. *arXiv preprint arXiv:2209.03997* .

Jin, Chi, Praneeth Netrapalli, Rong Ge, Sham M Kakade, Michael I Jordan. 2019. A short note on concentration inequalities for random vectors with subgaussian norm. *arXiv preprint arXiv:1902.03736* .

Johari, Ramesh, Vijay Kamble, Yash Kanoria. 2021. Matching while learning. *Operations Research* **69**(2) 655–681.

Jun, Kwang-Sung, Rebecca Willett, Stephen Wright, Robert Nowak. 2019. Bilinear bandits with low-rank structure. *International Conference on Machine Learning*. PMLR, 3163–3172.

Kallus, Nathan, Madeleine Udell. 2020. Dynamic assortment personalization in high dimensions. *Operations Research* **68**(4) 1020–1037.

Kaynar, Nur, Auyon Siddiq. 2023. Estimating effects of incentive contracts in online labor platforms. *Management Science* **69**(4) 2106–2126.

Keshavan, Raghunandan H, Andrea Montanari, Sewoong Oh. 2010. Matrix completion from a few entries. *IEEE transactions on information theory* **56**(6) 2980–2998.

Klopp, Olga. 2014. Noisy low-rank matrix completion with general sampling distribution. *Bernoulli* **20**(1) 282 – 303. doi:10.3150/12-BEJ486. URL `https://doi.org/10.3150/12-BEJ486`.

Knuth, Donald Ervin. 1997. *Stable marriage and its relation to other combinatorial problems: An introduction to the mathematical analysis of algorithms*, vol. 10. American Mathematical Soc.

Koltchinskii, Vladimir. 2011. *Oracle inequalities in empirical risk minimization and sparse recovery problems: École D'Été de Probabilités de Saint-Flour XXXVIII-2008*, vol. 2033. Springer Science & Business Media.

Koltchinskii, Vladimir, Karim Lounici, Alexandre B Tsybakov. 2011. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion .

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

Kveton, Branislav, Zheng Wen, Azin Ashkan, Csaba Szepesvari. 2015. Tight regret bounds for stochastic combinatorial semi-bandits. *Artificial Intelligence and Statistics*. PMLR, 535–543.

Ledoux, Michel, Michel Talagrand. 2013. *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media.

Li, Wen, Weiwei Sun. 2006. Some remarks on the perturbation of polar decompositions for rectangular matrices. *Numerical Linear Algebra with Applications* **13**(4) 327–338.

Liu, Lydia T, Horia Mania, Michael Jordan. 2020. Competing bandits in matching markets. *International Conference on Artificial Intelligence and Statistics*. PMLR, 1618–1628.

Lo, Irene, Vahideh Manshadi, Scott Rodilitz, Ali Shameli. 2024. Commitment on volunteer crowdsourcing platforms: Implications for growth and engagement. *Manufacturing & Service Operations Management* .

Ma, Cong, Kaizheng Wang, Yuejie Chi, Yuxin Chen. 2018. Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval and matrix completion. *International Conference on Machine Learning*. PMLR, 3345–3354.

Massart, Pascal. 2000. About the constants in talagrand's concentration inequalities for empirical processes. *The Annals of Probability* **28**(2) 863–884.

Massoulié, Laurent, Kuang Xu. 2016. On the capacity of information processing systems. *Conference on Learning Theory*. PMLR, 1292–1297.

Maurer, Andreas. 2016. A vector-contraction inequality for rademacher complexities. *Algorithmic Learning Theory: 27th International Conference, ALT 2016, Bari, Italy, October 19-21, 2016, Proceedings 27*. Springer, 3–17.

Million, Elizabeth. 2007. The hadamard product. *Course Notes* **3**(6) 1–7.

Munkres, James. 1957. Algorithms for the assignment and transportation problems. *Journal of the society for industrial and applied mathematics* **5**(1) 32–38.

Negahban, Sahand, Martin J Wainwright. 2012. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *The Journal of Machine Learning Research* **13**(1) 1665–1697.

Negahban, Sahand N, Pradeep Ravikumar, Martin J Wainwright, Bin Yu. 2012. A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers .

Pennington, Jeffrey, Richard Socher, Christopher D Manning. 2014. Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

Rosenfeld, Michael J, Reuben J Thomas, Sonia Hausen. 2019. Disintermediating your friends: How online dating in the united states displaces other ways of meeting. *Proceedings of the National Academy of Sciences* **116**(36) 17753–17758.

Sauré, Denis, Assaf Zeevi. 2013. Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management* **15**(3) 387–404.

Scarlett, Jonathan, Volkan Cevher. 2019. An introductory guide to fano's inequality with applications in statistical estimation. *arXiv preprint arXiv:1901.00555* .

Schuler, Alejandro, Vincent Liu, Joe Wan, Alison Callahan, Madeleine Udell, David E Stark, Nigam H Shah. 2016. Discovering patient phenotypes using generalized low rank models. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, vol. 21. NIH Public Access, 144.

Shah, Virag, Lennart Gulikers, Laurent Massoulié, Milan Vojnović. 2020. Adaptive matching for expert systems with uncertain task types. *Operations Research* **68**(5) 1403–1424.

Shapley, Lloyd S, Martin Shubik. 1971. The assignment game i: The core. *International Journal of game theory* **1**(1) 111–130.

Tropp, Joel A. 2012. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics* **12** 389–434.

Tropp, Joel A, et al. 2015. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning* **8**(1-2) 1–230.

Udell, Madeleine, Alex Townsend. 2019. Why are big data matrices approximately low rank? *SIAM Journal on Mathematics of Data Science* **1**(1) 144–160.

Wang, Siwei, Wei Chen. 2018. Thompson sampling for combinatorial semi-bandits. *International Conference on Machine Learning*. PMLR, 5114–5122.

Weyl, Hermann. 1912. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen* **71**(4) 441–479.

Xu, Kan, Xuanyi Zhao, Hamsa Bastani, Osbert Bastani. 2024. Group-sparse matrix factorization for transfer learning of word embeddings. *Available at SSRN 4730050* .

Yan, Chiwei, Helin Zhu, Nikita Korolko, Dawn Woodard. 2020. Dynamic pricing and matching in ride-hailing platforms. *Naval Research Logistics (NRL)* **67**(8) 705–724.

Zhou, Jie, Botao Hao, Zheng Wen, Jingfei Zhang, Will Wei Sun. 2024. Stochastic low-rank tensor bandits for multi-dimensional online decision making. *Journal of the American Statistical Association* 1–14.

Appendix A: Proof of Theorem 1

A.1. Major Steps of the Proof

Our proof strategy is adapted from the three steps in the proof of Theorem 2 in Athey et al. (2021) (see Remark 6.4 of Athey et al. (2021) for more details). We summarize our three steps in Lemma A.1, Lemma A.2 and Proposition 1 (stated in Section 3.2) respectively. For simplicity, we define $\mathbf{\Delta} = \mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}$, and let $\mathfrak{C} = \sum_{t=1}^n \sum_{i=1}^N \varepsilon_t^{(i)} \mathbf{X}_t^i$.

The operator norm of the noise matrix $\mathfrak{C}$ indicates the "scale" of the noises; the larger the operator norm, the higher the noise level. Lemma A.1 upper bounds the error "collected" by $\mathbf{X}_t$ for deterministic $\mathfrak{C}$. As long as the regularized hyperparameter λ take a sufficiently large value, the error collected by $\mathbf{X}_t$ can be upper bounded with respect to the corresponding error term's Frobenius norm.

Lemma A.1. *For any $\lambda \geq 3\|\mathfrak{C}\|_{\mathrm{op}}/n$, we have*

$$\frac{\sum_{t=1}^n \sum_{i=1}^N \langle \mathbf{X}_t^i, \mathbf{\Delta}\rangle^2}{n} \leq 4\lambda\sqrt{r}\|\mathbf{\Delta}\|_F.$$

Next, Lemma A.2 provides a probabilistic bound on the operator norm of the noise matrix $\mathfrak{C}$. In other words, the condition $\lambda \geq 3\|\mathfrak{C}\|_{\mathrm{op}}/n$ in Lemma A.1 holds with high probability, given our choice of λ in Theorem 1. Similar to Proposition 1, the key challenge of Lemma A.2 lies in the dependent structure of the observational data.

Lemma A.2. *Suppose $\sqrt{n} \geq \max\{\alpha, \log(NKn^2)\}$. Then, there exists a constant c_1 such that*

$$\|\mathfrak{C}\|_{\mathrm{op}} \leq c_1\sigma\big(\alpha + \log(N+K)\big)\sqrt{n}$$

with probability greater than $1 - 3\exp(-\alpha)$ for any $\alpha > 0$.

As summarized before, our last step will be the RSC condition previously mentioned in Proposition 1. With these three steps, we are now ready to prove Theorem 1.

A.2. Proof of Theorem 1

Proof of Theorem 1. By the constraint in (3.1), the rank of $\widehat{\mathbf{\Theta}}$ is less than or equal to r. Since $\mathsf{rank}(\mathbf{\Theta}^*) \leq r$, it is straightforward to see that $\mathsf{rank}(\mathbf{\Delta}) \leq 2r$.

First note that if $\|\mathbf{\Delta}\|_F^2 \leq c_0 NK\alpha/n$, then our argument goes. Otherwise, since $\|\mathbf{\Delta}\|_{L^2(\Pi)}^2 = \|\mathbf{\Delta}\|_F^2/K > c_0 N\alpha/n$, by the definition of the set $\mathcal{C}_\alpha(2r)$ in Proposition 1, we have $\frac{1}{2}\mathbf{\Delta} \in \mathcal{C}_\alpha(2r)$. Then we can apply Proposition 1 and have with probability greater than $1 - \exp(-\alpha)$ that,

$$\frac{1}{n}\sum_{t=1}^n \sum_{i=1}^N \big\langle \mathbf{X}_t^i, \frac{1}{2}\mathbf{\Delta}\big\rangle^2 > c_2 \left\|\frac{1}{2}\mathbf{\Delta}\right\|_{L^2(\Pi)}^2 - c_3\left(\frac{r^2 K \log[(N+K)n]}{n}\right). \tag{A.1}$$

By Lemma A.2 and our choice of $\lambda = c_\lambda \sigma \frac{\alpha + \log(N+K)}{\sqrt{n}}$ where c_λ is large enough, we have $\lambda \geq 3\|\mathfrak{C}\|_{\mathrm{op}}/n$ holds with probability greater than $1 - 3\exp(-\alpha)$. Then by Lemma A.1, we have

$$\frac{1}{n}\sum_{t=1}^n \sum_{i=1}^N \langle \mathbf{X}_t^i, \mathbf{\Delta}\rangle^2 \leq 4\lambda\sqrt{r}\|\mathbf{\Delta}\|_F = \frac{4c_\lambda\sigma\sqrt{r}\big(\alpha + \log(N+K)\big)}{\sqrt{n}}\|\mathbf{\Delta}\|_F \tag{A.2}$$

with probability higher than $1-3\exp(-\alpha)$. By combining (A.1) and (A.2), we obtain with probability greater than $1-4\exp(-\alpha)$ that,

$$c_2\|\boldsymbol{\Delta}\|^2_{L^2(\Pi)} - 4c_3\left(\frac{r^2K\log[(N+K)n]}{n}\right) \leq \frac{4c_\lambda\sigma\sqrt{r}\,(\alpha+\log(N+K))}{\sqrt{n}}\|\boldsymbol{\Delta}\|_F. \tag{A.3}$$

On the RHS, the basic inequality yields that

$$\frac{4c_\lambda\sigma\sqrt{r}\big(\alpha+\log(N+K)\big)}{\sqrt{n}}\|\boldsymbol{\Delta}\|_F \leq \frac{c_2}{2K}\|\boldsymbol{\Delta}\|_F^2 + \frac{8Kc_\lambda^2\sigma^2r\big(\alpha+\log(N+K)\big)^2}{c_2n}.$$

On the LHS, we have

$$\|\boldsymbol{\Delta}\|^2_{L^2(\Pi)} = \mathbb{E}\left[\sum_{i=1}^N\langle\mathbf{X}_t^i,\boldsymbol{\Delta}\rangle^2\right] = \frac{\|\boldsymbol{\Delta}\|_F^2}{K}.$$

Plugging the above into (A.3), we have

$$\begin{aligned}c_2\|\boldsymbol{\Delta}\|^2_{L^2(\Pi)} - 4c_3\left(\frac{r^2K\log[(N+K)n]}{n}\right) &= c_2\frac{\|\boldsymbol{\Delta}\|_F^2}{K} - 4c_3\left(\frac{r^2K\log[(N+K)n]}{n}\right)\\ &\leq \frac{c_2}{2K}\|\boldsymbol{\Delta}\|_F^2 + \frac{8Kc_\lambda^2\sigma^2r\big(\alpha+\log(N+K)\big)^2}{c_2n}.\end{aligned}$$

Rearranging the above inequality gives us

$$\frac{c_2}{2K}\|\boldsymbol{\Delta}\|_F^2 \leq 4c_3\left(\frac{r^2K\log[(N+K)n]}{n}\right) + \frac{8Kc_\lambda^2\sigma^2r\big(\alpha+\log(N+K)\big)^2}{c_2n},$$

thus

$$\|\boldsymbol{\Delta}\|_F^2 \leq \frac{8c_3r^2K^2\log[(N+K)n]}{c_2n} + \frac{16K^2c_\lambda^2\sigma^2r\big(\alpha+\log(N+K)\big)^2}{c_2^2n}.$$

So (A.3) implies that

$$\frac{\|\boldsymbol{\Delta}\|_F}{\sqrt{NK}} \leq c_4\max\left\{c_\lambda\sigma\left(\alpha+\log(N+K)\right)\sqrt{\frac{rK}{Nn}}, r\sqrt{\frac{K\log[(N+K)n]}{Nn}}\right\}, \tag{A.4}$$

where c_4 is a universal constant that depends on c_2 and c_3. Since (A.3) holds with probability greater than $1-4\exp(-\alpha)$, our argument goes. □

A.3. Proof of Lemma A.1

Proof of Lemma A.1. By (3.1) we have

$$\frac{1}{n}\sum_{t=1}^n\|Y_t-\mathcal{X}_t(\widehat{\boldsymbol{\Theta}})\|^2+\lambda\|\widehat{\boldsymbol{\Theta}}\|_* \leq \frac{1}{n}\sum_{t=1}^n\|Y_t-\mathcal{X}_t(\boldsymbol{\Theta}^*)\|^2+\lambda\|\boldsymbol{\Theta}^*\|_*,$$

which implies

$$\frac{1}{n}\sum_{t=1}^n\|Y_t-\mathcal{X}_t(\widehat{\boldsymbol{\Theta}})\|^2-\frac{1}{n}\sum_{t=1}^n\|Y_t-\mathcal{X}_t(\boldsymbol{\Theta}^*)\|^2 \leq \lambda\|\boldsymbol{\Theta}^*\|_*-\lambda\|\widehat{\boldsymbol{\Theta}}\|_*. \tag{A.5}$$

By the definition of $\mathcal{X}_t$, we have

$$\begin{aligned}Y_t-\mathcal{X}_t(\widehat{\boldsymbol{\Theta}}) &= Y_t-\mathcal{X}_t(\boldsymbol{\Theta}^*)+\mathcal{X}_t(\boldsymbol{\Theta}^*)-\mathcal{X}_t(\widehat{\boldsymbol{\Theta}})\\ &= Y_t-\mathcal{X}_t(\boldsymbol{\Theta}^*)+\mathcal{X}_t(\boldsymbol{\Delta}).\end{aligned}$$

Then

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\|Y_t-\mathcal{X}_t(\widehat{\mathbf{\Theta}})\|^2-\frac{1}{n}\sum_{t=1}^{n}\|Y_t-\mathcal{X}_t(\mathbf{\Theta}^*)\|^2\\
&=\frac{1}{n}\sum_{t=1}^{n}\|Y_t-\mathcal{X}_t(\mathbf{\Theta}^*)+\mathcal{X}_t(\mathbf{\Delta})\|^2-\frac{1}{n}\sum_{t=1}^{n}\|Y_t-\mathcal{X}_t(\mathbf{\Theta}^*)\|^2\\
&=\frac{2}{n}\sum_{t=1}^{n}\langle Y_t-\mathcal{X}_t(\mathbf{\Theta}^*),\mathcal{X}_t(\mathbf{\Delta})\rangle+\frac{1}{n}\sum_{t=1}^{n}\|\mathcal{X}_t(\mathbf{\Delta})\|^2.
\end{aligned}$$

By (2.4), we have

$$Y_t-\mathcal{X}_t(\mathbf{\Theta}^*)=\left[\varepsilon_t^{(1)}\ \varepsilon_t^{(2)}\ \cdots\ \varepsilon_t^{(N)}\right].$$

Then by the definition of the noise matrix $\mathfrak{C}$,

$$\begin{aligned}
\sum_{t=1}^{n}\langle Y_t-\mathcal{X}_t(\mathbf{\Theta}^*),\mathcal{X}_t(\mathbf{\Delta})\rangle&=\sum_{t=1}^{n}\sum_{i=1}^{N}\varepsilon_t^{(i)}\langle\mathbf{X}_t^i,\mathbf{\Delta}\rangle\\
&=\sum_{t=1}^{n}\sum_{i=1}^{N}\langle\varepsilon_t^{(i)}\mathbf{X}_t^i,\mathbf{\Delta}\rangle\\
&=\left\langle\sum_{t=1}^{n}\sum_{i=1}^{N}\varepsilon_t^{(i)}\mathbf{X}_t^i,\mathbf{\Delta}\right\rangle=\langle\mathfrak{C},\mathbf{\Delta}\rangle.
\end{aligned}$$

Plugging the above into (A.5) yields that

$$\begin{aligned}
\frac{2}{n}\sum_{t=1}^{n}\langle Y_t-\mathcal{X}_t(\mathbf{\Theta}^*),\mathcal{X}_t(\mathbf{\Delta})\rangle+\frac{1}{n}\sum_{t=1}^{n}\|\mathcal{X}_t(\mathbf{\Delta})\|^2&=\frac{2}{n}\langle\mathfrak{C},\mathbf{\Delta}\rangle+\frac{1}{n}\sum_{t=1}^{n}\|\mathcal{X}_t(\mathbf{\Delta})\|^2\\
&=\frac{2}{n}\langle\mathfrak{C},\mathbf{\Delta}\rangle+\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{N}\langle\mathbf{X}_t^i,\mathbf{\Delta}\rangle^2\le\lambda\|\mathbf{\Theta}^*\|_*-\lambda\|\widehat{\mathbf{\Theta}}\|_*
\end{aligned}$$

where the second equality is given by the definition of $\mathcal{X}_t$. Then by $\lambda\ge 3\|\mathfrak{C}\|_{\mathrm{op}}/n$, the above inequality gives

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{N}\langle\mathbf{X}_t^i,\mathbf{\Delta}\rangle^2&\le-\frac{2}{n}\langle\mathfrak{C},\mathbf{\Delta}\rangle+\lambda\|\mathbf{\Theta}^*\|_*-\lambda\|\widehat{\mathbf{\Theta}}\|_*\\
&\le\frac{2}{n}\|\mathbf{\Delta}\|_*\|\mathfrak{C}\|_{\mathrm{op}}+\lambda\|\mathbf{\Delta}\|_*\\
&\le\frac{5}{3}\lambda\|\mathbf{\Delta}\|_*. \qquad \text{(A.6)}
\end{aligned}$$

In Section A.2 we have shown that $\mathsf{rank}(\mathbf{\Delta})\le 2r$. Then we can derive the following inequality: $\|\mathbf{\Delta}\|_*\le\sqrt{2r}\|\mathbf{\Delta}\|_F$. By substituting this inequality into the previous inequality (A.6), we obtain:

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{N}\langle\mathbf{X}_t^i,\mathbf{\Delta}\rangle^2\le 4\lambda\sqrt{r}\|\mathbf{\Delta}\|_F.$$

Since $\mathbf{\Delta}=\mathbf{\Theta}^*-\widehat{\mathbf{\Theta}}$, our argument goes. □

A.4. Proof of Lemma A.2

Proof of Lemma A.2. For every $t\in[n]$, define

$$\mathbf{B}_t=\sum_{i=1}^{N}\varepsilon_t^{(i)}\mathbf{X}_t^i.$$

Thus we have $\mathfrak{C} = \sum_{t=1}^{n} \mathbf{B}_t$, $\mathbb{E}[\mathbf{B}_t] = \mathbf{0}$, and each $\mathbf{B}_t$ is independent. Let $\vartheta = \sigma\sqrt{2\log\big(NKn\big) + 2\alpha}$. For each $t \in [n]$ and $i \in [N]$ define the truncation of $\varepsilon_t^{(i)}$ as $\bar{\varepsilon}_t^{(i)} = \varepsilon_t^{(i)} \mathbb{1}\{|\varepsilon_t^{(i)}| \le \vartheta\}$. Also define

$$\bar{\mathbf{B}}_t = \sum_{i=1}^{N} \bar{\varepsilon}_t^{(i)} \mathbf{X}_t^i.$$

Notice that for any $\gamma \ge 0$,

$$\begin{aligned}
\mathbb{P}[\|\mathfrak{C}\|_{\text{op}} \ge \gamma + \vartheta] &= \mathbb{P}\Bigg[\Big\|\sum_{t=1}^{n} \mathbf{B}_t\Big\|_{\text{op}} \ge \gamma + \vartheta\Bigg] \\
&\le \mathbb{P}\Bigg[\Big\|\sum_{t=1}^{n} \bar{\mathbf{B}}_t\Big\|_{\text{op}} \ge \gamma\Bigg] + \mathbb{P}\Bigg[\bigcup_{t=1}^{n}\bigcup_{i=1}^{N}\{|\varepsilon_t^{(i)}| > \vartheta\}\Bigg]. \qquad \text{(A.7)}
\end{aligned}$$

Since $\varepsilon_t^{(i)}$ are σ-subgaussian noises (defined in Definition 1), we have

$$\mathbb{P}\Bigg[\bigcup_{t=1}^{n}\bigcup_{i=1}^{N}\{|\varepsilon_t^{(i)}| > \vartheta\}\Bigg] \le \sum_{t=1}^{n}\sum_{i=1}^{N} \mathbb{P}\big[|\varepsilon_t^{(i)}| > \vartheta\big] \le Nn \cdot 2\exp\left(\frac{-\vartheta^2}{2\sigma^2}\right) = 2\exp(-\alpha)\cdot K^{-1} \le 2\exp(-\alpha). \qquad \text{(A.8)}$$

Next, we provide an upper bound for $\mathbb{P}\Bigg[\Big\|\sum_{t=1}^{n} \bar{\mathbf{B}}_t\Big\|_{\text{op}} \ge \gamma\Bigg]$. For each $t \in [n]$, define

$$\mathbf{G}_t = \bar{\mathbf{B}}_t - \sum_{i=1}^{N} \mathbb{E}[\bar{\varepsilon}_t^{(i)}]\mathbf{X}_t^i.$$

Then

$$\begin{aligned}
\Big\|\sum_{t=1}^{n} \bar{\mathbf{B}}_t\Big\|_{\text{op}} &= \Big\|\sum_{t=1}^{n}\Big(\mathbf{G}_t + \sum_{i=1}^{N} \mathbb{E}[\bar{\varepsilon}_t^{(i)}]\mathbf{X}_t^i\Big)\Big\|_{\text{op}} \\
&\le \Big\|\sum_{t=1}^{n} \mathbf{G}_t\Big\|_{\text{op}} + \Big\|\sum_{t=1}^{n}\sum_{i=1}^{N} \mathbb{E}[\bar{\varepsilon}_t^{(i)}]\mathbf{X}_t^i\Big\|_{\text{op}} \\
&\le \Big\|\sum_{t=1}^{n} \mathbf{G}_t\Big\|_{\text{op}} + \Big\|\sum_{t=1}^{n}\sum_{i=1}^{N} \mathbb{E}[\bar{\varepsilon}_t^{(i)}]\mathbf{X}_t^i\Big\|_{F} \\
&\le \underbrace{\Big\|\sum_{t=1}^{n} \mathbf{G}_t\Big\|_{\text{op}}}_{=h_0} + \underbrace{\sqrt{NK}\Big\|\sum_{t=1}^{n}\sum_{i=1}^{N} \mathbb{E}[\bar{\varepsilon}_t^{(i)}]\mathbf{X}_t^i\Big\|_{\infty}}_{=h_1}. \qquad \text{(A.9)}
\end{aligned}$$

For the term h_1 on the RHS, we have

$$\begin{aligned}
h_1 = \sqrt{NK}\Big\|\sum_{t=1}^{n}\sum_{i=1}^{N} \mathbb{E}[\bar{\varepsilon}_t^{(i)}]\mathbf{X}_t^i\Big\|_{\infty} &\le \sqrt{NK}\sum_{t=1}^{n}\Big\|\sum_{i=1}^{N} \mathbb{E}[\bar{\varepsilon}_t^{(i)}]\mathbf{X}_t^i\Big\|_{\infty} \\
&\le \sqrt{NK}n \max_{i\in[N],t\in[n]} |\mathbb{E}[\bar{\varepsilon}_t^{(i)}]|.
\end{aligned}$$

Since $\mathbb{E}[\varepsilon_t^{(i)}] = 0$, we have for any $t \in [n], i \in [N]$

$$\begin{aligned}
\big|\mathbb{E}[\bar{\varepsilon}_t^i]\big| = \Big|\mathbb{E}\Big[\varepsilon_t^i \mathbb{1}\{|\varepsilon_t^{(i)}| \le \vartheta\}\Big]\Big| &= \Big|\mathbb{E}\Big[\varepsilon_t^{(i)} \mathbb{1}\{|\varepsilon_t^{(i)}| > \vartheta\}\Big]\Big| \\
&\le \sqrt{\mathbb{E}\big[(\varepsilon_t^{(i)})^2\big]\mathbb{P}[|\varepsilon_t^{(i)}| > \vartheta]} \\
&\le \sqrt{2\sigma^2 \exp\big[-\vartheta^2/(2\sigma^2)\big]} \\
&= \frac{\sqrt{2}\sigma\exp(-\alpha/2)}{\sqrt{NKn}} \le \frac{\sqrt{2}\sigma}{\sqrt{NKn}}
\end{aligned}$$

where the last inequality is given by $\alpha > 0$. Thus

$$h_1 \le \sqrt{NK} n \max_{i\in[N],t\in[n]} |\mathbb{E}[\bar{\varepsilon}_t^{(i)}]| \le \sqrt{2}\sigma. \tag{A.10}$$

Now we will bound the term h_0 via matrix Berstein inequality. First, we provide an upper bound for $\|\mathbf{G}_t\|_{\mathrm{op}}$ and

$$\sigma_Z^2 = \max\left\{ \left\| \mathbb{E}\Big[\sum_{t=1}^n \mathbf{G}_t \mathbf{G}_t^\top \Big] \right\|_{\mathrm{op}}, \left\| \mathbb{E}\Big[\sum_{t=1}^n \mathbf{G}_t^\top \mathbf{G}_t \Big] \right\|_{\mathrm{op}} \right\}.$$

We have the following lemma which is proved at the end of this subsection:

Lemma A.3. *We have*

$$\|\mathbf{G}_t\|_{\mathrm{op}} \le 2\vartheta, \forall t \in [n], \quad \textit{and} \quad \sigma_Z^2 = \max\left\{ \left\| \mathbb{E}\Big[\sum_{t=1}^n \mathbf{G}_t \mathbf{G}_t^\top \Big] \right\|_{\mathrm{op}}, \left\| \mathbb{E}\Big[\sum_{t=1}^n \mathbf{G}_t^\top \mathbf{G}_t \Big] \right\|_{\mathrm{op}} \right\} \le n\sigma^2.$$

Due to the special dependent structure we have under the matching sampling scheme, this crucial lemma sets us apart from Lemma 2 in Athey et al. (2021), where their σ_Z^2 is of order $\mathcal{O}(n\sqrt{K})$ given their dependent structure.

Given Lemma A.3, we can bound h_0 via Lemma F.1. By Lemma F.1, for any $\gamma > 0$,

$$\begin{aligned} \mathbb{P}\Big[h_0 \ge \gamma \Big] = \mathbb{P}\Big[\Big\| \sum_{t=1}^n \mathbf{G}_t \Big\|_{\mathrm{op}} \ge \gamma \Big] &\le (N+K) \exp\Big\{ \frac{-\gamma^2}{2\sigma_Z^2 + (4\vartheta\gamma)/3} \Big\} \\ &\le (N+K) \exp\Big\{ \frac{-\gamma^2}{2n\sigma^2 + (4\vartheta\gamma)/3} \Big\}. \end{aligned}$$

By choosing

$$\gamma \ge \max\left\{ 2\sigma\sqrt{n}\sqrt{\alpha + \log(N+K)}, \frac{8\vartheta}{3}\big(\alpha + \log(N+K)\big) \right\},$$

we have

$$\begin{aligned} (N+K) \exp\Big\{ \frac{-\gamma^2}{2n\sigma^2 + (4\vartheta\gamma)/3} \Big\} &= (N+K) \exp\Big\{ \frac{-\frac{\gamma^2}{2} - \frac{\gamma^2}{2}}{2n\sigma^2 + (4\vartheta\gamma)/3} \Big\} \\ &\le (N+K) \exp\Big\{ \frac{-2n\sigma^2(\alpha + \log(N+K)) - \frac{4\vartheta\gamma}{3}\big(\alpha + \log(N+K)\big)}{2n\sigma^2 + (4\vartheta\gamma)/3} \Big\} \\ &= (N+K) \exp\left(-\alpha - \log(N+K)\right) \\ &= \exp(-\alpha). \end{aligned}$$

Since

$$\begin{aligned} \frac{8\vartheta}{3}\big(\alpha + \log(N+K)\big) &= \frac{8\sqrt{2}\sigma}{3}\left(\alpha + \log(N+K)\right)\left(\alpha + \log(NKn)\right) \\ &\le 5\sigma\left(\alpha + \log(N+K)\right)\left(\alpha + \log(NKn)\right), \end{aligned}$$

we can choose

$$\gamma = 5\sigma\big(\alpha + \log(N+K)\big)\left(\sqrt{n} + \alpha + \log(NKn^2)\right),$$

which yields that

$$\mathbb{P}\left[h_0 \ge 5\sigma\big(\alpha + \log(N+K)\big)\left(\sqrt{n} + \alpha + \log(NKn^2)\right)\right] \le \exp(-\alpha).$$

Combining with (A.9) and $h_1 \le \sqrt{2}\sigma$ in (A.10) gives us

$$\begin{aligned}&\mathbb{P}\left[\Big\|\sum_{t=1}^{n}\bar{\mathbf{B}}_t\Big\|_{\text{op}} \ge 5\sigma\big(\alpha+\log(N+K)\big)\left(\sqrt{n}+\alpha+\log(NKn^2)\right)+\sqrt{2}\sigma\right]\\ &\le \mathbb{P}\left[h_0 \ge 5\sigma\big(\alpha+\log(N+K)\big)\left(\sqrt{n}+\alpha+\log(NKn^2)\right)\right] \le \exp(-\alpha).\end{aligned}$$

By (A.7) and (A.8), we have that, with probability larger than $1-3\exp(-\alpha)$,

$$\begin{aligned}\|\mathfrak{C}\|_{\text{op}} &\le 5\sigma\big(\alpha+\log(N+K)\big)\left(\sqrt{n}+\alpha+\log(NKn^2)\right)+\sqrt{2}\sigma+\sigma\sqrt{2\log\big(NKn\big)+2\alpha}\\ &\le 8\sigma\big(\alpha+\log(N+K)\big)\left(\sqrt{n}+\alpha+\log(NKn^2)\right)\\ &\le 24\sigma\big(\alpha+\log(N+K)\big)\sqrt{n}\end{aligned}$$

where the last inequality is given by $\sqrt{n} \ge \max\{\alpha, \log(NKn^2)\}$, which completes the proof. □

Proof of Lemma A.3 Given any $t \in [n]$, we can write $\mathbf{G}_t$ as

$$\mathbf{G}_t = \bar{\mathbf{B}}_t - \sum_{i=1}^{N}\mathbb{E}[\bar{\varepsilon}_t^{(i)}]\mathbf{X}_t^i = \sum_{i=1}^{N}\tilde{\varepsilon}_t^{(i)}\mathbf{X}_t^i,$$

where $\tilde{\varepsilon}_t^{(i)} = \bar{\varepsilon}_t^{(i)} - \mathbb{E}[\bar{\varepsilon}_t^{(i)}]$. We first bound $\|\mathbf{G}_t\|_{\text{op}}$. By the definition of $\tilde{\varepsilon}_t^{(i)}$ and that $\tilde{\varepsilon}_t^{(i)}$ are independent with $\mathbf{X}_t^i$, we have

$$\|\mathbf{G}_t\|_{\text{op}} = \Big\|\sum_{i=1}^{N}\tilde{\varepsilon}_t^{(i)}\mathbf{X}_t^i\Big\|_{\text{op}} \le \max_{k\in[n]}|\tilde{\varepsilon}_t^k|\Big\|\sum_{i=1}^{N}\mathbf{X}_t^i\Big\|_{\text{op}} \le 2\vartheta\Big\|\sum_{i=1}^{N}\mathbf{X}_t^i\Big\|_{\text{op}}.$$

We can then treat $\sum_{i=1}^{N}\mathbf{X}_t^i$ as a binary matrix obtained by permuting the columns of a binary diagonal matrix in $\mathbb{R}^{N\times K}$. Since the norm of a matrix does not depend on the column order, we have $\Big\|\sum_{i=1}^{N}\mathbf{X}_t^i\Big\|_{\text{op}} = 1$ and thus $\|\mathbf{G}_t\|_{\text{op}} \le 2\vartheta$. By an analogous argument, we can also bound σ_Z^2. We have

$$\mathbf{G}_t\mathbf{G}_t^\top = \sum_{i=1}^{N}\sum_{k=1}^{N}\tilde{\varepsilon}_t^{(i)}\tilde{\varepsilon}_t^{(k)}\mathbf{X}_t^i\mathbf{X}_t^{k\top}.$$

Notice that for any $i \ne k$, we have $\mathbf{X}_t^i\mathbf{X}_t^{k\top} = \mathbf{0}$. Additionally, for any realization of $\mathbf{X}_t$, we have

$$\begin{aligned}\sum_{i=1}^{N}\mathbf{X}_t^i\mathbf{X}_t^{i\top} &= \sum_{i=1}^{N}e_i(N)e_{j_t(i)}^\top(K)e_{j_t(i)}(K)e_i^\top(N)\\ &= \sum_{i=1}^{N}e_i(N)e_i^\top(N)\\ &= \mathbf{I}_{N\times N}.\end{aligned}$$

Thus

$$\Big\|\sum_{i=1}^{N}\mathbf{X}_t^i\mathbf{X}_t^{i\top}\Big\|_{\text{op}} = 1, \tag{A.11}$$

and

$$\Big\|\sum_{i=1}^{N}\mathbf{X}_t^{i\top}\mathbf{X}_t^i\Big\|_{\text{op}} = 1. \tag{A.12}$$

Then

$$\mathbf{G}_t\mathbf{G}_t^\top = \sum_{i=1}^{N}\sum_{k=1}^{N}\tilde{\varepsilon}_t^{(i)}\tilde{\varepsilon}_t^{(k)}\mathbf{X}_t^i\mathbf{X}_t^{k\top} = \sum_{i=1}^{N}(\tilde{\varepsilon}_t^{(i)})^2\mathbf{X}_t^i\mathbf{X}_t^{i\top}$$

and

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{G}_t\mathbf{G}_t^\top\|_{\mathrm{op}}\right] &= \mathbb{E}\left[\Big\|\sum_{i=1}^N (\widetilde{\varepsilon}_t^{(i)})^2\mathbf{X}_t^i\mathbf{X}_t^{i\top}\Big\|_{\mathrm{op}}\right] \\ &\le \max_{i\in[N]}\{\mathbb{E}[(\widetilde{\varepsilon}_t^{(i)})^2]\}\mathbb{E}\left[\Big\|\sum_{i=1}^N \mathbf{X}_t^i\mathbf{X}_t^{i\top}\Big\|_{\mathrm{op}}\right] \\ &\le \sigma^2\mathbb{E}\left[\Big\|\sum_{i=1}^N \mathbf{X}_t^i\mathbf{X}_t^{i\top}\Big\|_{\mathrm{op}}\right] \\ &\le \sigma^2,\end{aligned}$$

where the first inequality is due to that $\widetilde{\varepsilon}_t^{(i)}$ and $\mathbf{X}_t^i$ are independent with each other, the second inequality is given by $\mathbb{E}[(\widetilde{\varepsilon}_t^{(i)})^2] \le \mathbb{E}[(\bar{\varepsilon}_t^{(i)})^2] \le \mathbb{E}[(\varepsilon_t^{(i)})^2] \le \sigma^2$. Similarly we have $\mathbb{E}\left[\|\mathbf{G}_t^\top\mathbf{G}_t\|_{\mathrm{op}}\right] \le \sigma^2$. Thus,

$$\sigma_Z^2 \le n\max_{t\in[n]}\{\mathbb{E}[\|\mathbf{G}_t\mathbf{G}_t^\top\|_{\mathrm{op}}], \mathbb{E}[\|\mathbf{G}_t^\top\mathbf{G}_t\|_{\mathrm{op}}]\} \le n\sigma^2. \quad \square$$

A.5. Proof of Proposition 1

Proof of Proposition 1. Recall that, for any $t \in [n]$,

$$\mathbf{X}_t = \sum_{i=1}^N \mathbf{X}_t^i. \tag{A.13}$$

We will prove Proposition 1 by a standard peeling argument. Let

$$w_\zeta = \sup_{\mathbf{\Delta}\in\mathcal{C}_\alpha(r,\zeta)}\left|\frac{1}{n}\sum_{t=1}^n\sum_{i=1}^N\langle\mathbf{X}_t^i,\mathbf{\Delta}\rangle^2 - \|\mathbf{\Delta}\|_{L^2(\Pi)}^2\right|,$$

where

$$\mathcal{C}_\alpha(r,\zeta) = \left\{\mathbf{\Delta}\in\mathcal{C}_\alpha(r)\,\middle|\,\frac{\zeta}{2} \le \|\mathbf{\Delta}\|_{L^2(\Pi)}^2 < \zeta\right\}.$$

First notice that

$$\sum_{t=1}^n\sum_{i=1}^N\langle\mathbf{X}_t^i,\mathbf{\Delta}\rangle^2 = \sum_{t=1}^n\langle\mathbf{X}_t,\mathbf{\Delta}\circ\mathbf{\Delta}\rangle. \tag{A.14}$$

Then by Lemma F.2, we have

$$\mathbb{P}\left[w_\zeta > 2\mathbb{E}[w_\zeta] + \frac{7\zeta}{24}\right] \le \exp\left(-\frac{n\zeta}{288N}\right) \tag{A.15}$$

since

$$\mathcal{C}_\alpha(r,\zeta) \subseteq \left\{\mathbf{\Delta}\in\mathcal{C}_\alpha(r)\,\middle|\,\|\mathbf{\Delta}\|_{L^2(\Pi)}^2 < \zeta\right\},$$

Now we provide an upper bound for $\mathbb{E}[w_\zeta]$. For $t \in [n]$, let ξ_t be independent Rademacher random variables. By the symmetrization inequality,

$$\begin{aligned}\mathbb{E}[w_\zeta] &= \mathbb{E}\left[\sup_{\mathbf{\Delta}\in\mathcal{C}(r,\zeta)}\left|\frac{1}{n}\sum_{t=1}^n\sum_{i=1}^N\langle\mathbf{X}_t^i,\mathbf{\Delta}\rangle^2 - \|\mathbf{\Delta}\|_{L^2(\Pi)}^2\right|\right] \\ &\le 2\mathbb{E}\left[\sup_{\mathbf{\Delta}\in\mathcal{C}(r,\zeta)}\frac{1}{n}\sum_{t=1}^n\xi_t(\sum_{i=1}^N\langle\mathbf{X}_t^i,\mathbf{\Delta}\rangle^2)\right] \\ &= 2\mathbb{E}\left[\sup_{\mathbf{\Delta}\in\mathcal{C}(r,\zeta)}\frac{1}{n}\sum_{t=1}^n\xi_t\langle\mathbf{X}_t,\mathbf{\Delta}\circ\mathbf{\Delta}\rangle\right] \\ &= 2\mathbb{E}\left[\sup_{\mathbf{\Delta}\in\mathcal{C}(r,\zeta)}\langle\mathbf{\Sigma}_R,\mathbf{\Delta}\circ\mathbf{\Delta}\rangle\right],\end{aligned}$$

where $\mathbf{\Sigma}_R = \frac{1}{n}\sum_{t=1}^{n}\xi_t \mathbf{X}_t$. Notice that Lemma F.3 gives $\mathsf{rank}(\mathbf{\Delta}\circ\mathbf{\Delta}) \le r^2$, and thus we have $\|\mathbf{\Delta}\circ\mathbf{\Delta}\|_* \le \sqrt{\mathsf{rank}(\mathbf{\Delta}\circ\mathbf{\Delta})}\|\mathbf{\Delta}\|_F \le \sqrt{r^2}\|\mathbf{\Delta}\|_F$. In addition, since $\|\mathbf{\Delta}\|_\infty \le 1$, we have $\|\mathbf{\Delta}\circ\mathbf{\Delta}\|_F \le \|\mathbf{\Delta}\|_F$. Then it holds that

$$\begin{aligned}
\mathbb{E}\big[\sup_{\mathbf{\Delta}\in\mathcal{C}(r,\zeta)} \langle \mathbf{\Sigma}_R, \mathbf{\Delta}\circ\mathbf{\Delta}\rangle\big] &\le \sup_{\mathbf{\Delta}\in\mathcal{C}(r,\zeta)} \|\mathbf{\Delta}\circ\mathbf{\Delta}\|_* \mathbb{E}\left[\|\mathbf{\Sigma}_R\|_{\mathrm{op}}\right] \\
&\le \sup_{\mathbf{\Delta}\in\mathcal{C}(r,\zeta)} \sqrt{r^2}\|\mathbf{\Delta}\circ\mathbf{\Delta}\|_F \mathbb{E}\left[\|\mathbf{\Sigma}_R\|_{\mathrm{op}}\right] \\
&\le \sup_{\mathbf{\Delta}\in\mathcal{C}(r,\zeta)} \sqrt{r^2}\|\mathbf{\Delta}\|_F \mathbb{E}\left[\|\mathbf{\Sigma}_R\|_{\mathrm{op}}\right] \\
&= r\sqrt{K\|\mathbf{\Delta}\|^2_{L^2(\Pi)}}\mathbb{E}\left[\|\mathbf{\Sigma}_R\|_{\mathrm{op}}\right] \\
&\le r\sqrt{K\zeta}\mathbb{E}\left[\|\mathbf{\Sigma}_R\|_{\mathrm{op}}\right] \\
&\le 4r^2K\left(\mathbb{E}\left[\|\mathbf{\Sigma}_R\|_{\mathrm{op}}\right]\right)^2 + \frac{\zeta}{16},
\end{aligned}$$

where the first inequality is given by the duality between nuclear norm and operator norm (i.e., $|\langle\mathbf{P},\mathbf{Q}\rangle| \le \|\mathbf{P}\|_*\cdot\|\mathbf{Q}\|_{\mathrm{op}}, \forall \mathbf{P},\mathbf{Q}\in\mathbb{R}^{N\times K}$). Plugging the above into (A.15) gives us

$$\mathbb{P}\left[w_\zeta > 4r^2K\left(\mathbb{E}\left[\|\mathbf{\Sigma}_R\|_{\mathrm{op}}\right]\right)^2 + \frac{11\zeta}{24}\right] \le \exp\left(\frac{-n\zeta}{288N}\right). \tag{A.16}$$

Lemma A.4 shows that

$$\mathbb{E}[\|\mathbf{\Sigma}_R\|_{\mathrm{op}}] \le \frac{6\sqrt{\log[(N+K)n]}}{\sqrt{n}},$$

and thus

$$4r^2K\big[\mathbb{E}[\|\mathbf{\Sigma}_R\|_{\mathrm{op}}]\big]^2 \le \frac{144r^2K\log[(N+K)n]}{n}. \tag{A.17}$$

Now set

$$\varphi = \frac{144r^2K\log[(N+K)n]}{n}$$

and define the bad event

$$\mathcal{B} = \left\{\exists\mathbf{\Delta}\in\mathcal{C}_\alpha(r) \quad \text{s.t.} \quad \left|\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{N}\langle\mathbf{X}_t^i,\mathbf{\Delta}\rangle^2 - \|\mathbf{\Delta}\|^2_{L^2(\Pi)}\right| \ge \frac{11}{24}\|\mathbf{\Delta}\|^2_{L^2(\Pi)} + \varphi\right\}.$$

Notice that if event $\mathcal{B}$ holds with small probability, then our argument goes. For any $l\in\mathbb{N}^+$, define

$$\mathcal{B}_l = \left\{\exists\mathbf{\Delta}\in\mathcal{C}\Big(r, \frac{576N\alpha}{n}\cdot 2^l\Big) \quad \text{s.t.} \quad \left|\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{N}\langle\mathbf{X}_t^i,\mathbf{\Delta}\rangle^2 - \|\mathbf{\Delta}\|^2_{L^2(\Pi)}\right| \ge \frac{11}{24}\|\mathbf{\Delta}\|^2_{L^2(\Pi)} + \varphi\right\}.$$

According to (A.16) and (A.17), we have

$$\mathbb{P}[\mathcal{B}_l] \le \exp\big(-2\alpha\cdot 2^l\big).$$

Note that $\mathcal{B}\subseteq\bigcup_{l=1}^{\infty}\mathcal{B}_l$, which implies

$$\mathbb{P}[\mathcal{B}] \le \sum_{l=1}^{\infty}\mathbb{P}[\mathcal{B}_l] \le \sum_{l=1}^{\infty}\exp\big(-2\alpha\cdot 2^l\big) \le \sum_{l=1}^{\infty}\exp\big(-2\alpha\cdot l\big) \le \frac{\exp(-2\alpha)}{1-\exp(-2\alpha)} \le \exp(-\alpha). \quad \square$$

LEMMA A.4. *For any $t\in[n]$, let ξ_t be i.i.d. Rademacher random variables and let*

$$\mathbf{\Sigma}_R = \frac{1}{n}\sum_{t=1}^{n}\xi_t\mathbf{X}_t$$

where $\mathbf{X}_t$ satisfies (A.13). *Then we have*

$$\mathbb{E}\left[\|\mathbf{\Sigma}_R\|_{\mathrm{op}}\right] \le \frac{6\sqrt{\log[(N+K)n]}}{\sqrt{n}}.$$

Proof of Lemma A.4 Our proof techniques rely on the concentration inequality for matrix Rademacher series in Lemma F.4. Thus we need to first calculate the value of σ_Z^2. Our approach is similar to the argument in the proof of Lemma A.3. First note that

$$\|\mathbf{X}_t\mathbf{X}_t^\top\|_{\text{op}} = \left\|\sum_{i=1}^N\sum_{j=1}^N \mathbf{X}_t^i\mathbf{X}_t^{j\top}\right\|_{\text{op}} = \left\|\sum_{i=1}^N \mathbf{X}_t^i\mathbf{X}_t^{i\top}\right\|_{\text{op}}$$

since $\mathbf{X}_t^i\mathbf{X}_t^{j\top} = \mathbf{0}, \forall i \neq j$. By (A.11) we have

$$\left\|\sum_{t=1}^n \mathbf{X}_t\mathbf{X}_t^\top\right\|_{\text{op}} \leq n \cdot \max_{t\in[n]} \|\mathbf{X}_t\mathbf{X}_t^\top\|_{\text{op}} = n.$$

Similarly by (A.12) we have

$$\left\|\sum_{t=1}^n \mathbf{X}_t^\top\mathbf{X}_t\right\|_{\text{op}} \leq n \cdot \max_{t\in[n]} \|\mathbf{X}_t^\top\mathbf{X}_t\|_{\text{op}} = n.$$

By applying Lemma F.4 with $\sigma_Z^2 = n$, we have

$$\mathbb{P}\left[\left\|\sum_{t=1}^n \xi_t\mathbf{X}_t\right\|_{\text{op}} \geq \rho\right] \leq (N+K)\exp\left(\frac{-\rho^2}{2n}\right), \forall \rho > 0.$$

Set $\rho = \sqrt{2n\log[(N+K)n^{3/2}]}$ and we obtain

$$\mathbb{P}\left[\left\|\sum_{t=1}^n \xi_t\mathbf{X}_t\right\|_{\text{op}} \geq \sqrt{2n\log[(N+K)n^{3/2}]}\right] \leq \frac{1}{n^{3/2}}.$$

We also have that $\left\|\sum_{t=1}^n \xi_t\mathbf{X}_t\right\|_{\text{op}} \leq n$ almost surely. Therefore, we have

$$\begin{aligned}
\mathbb{E}[\|\mathbf{\Sigma}_R\|_{\text{op}}] &= \mathbb{E}\left[\|\mathbf{\Sigma}_R\|_{\text{op}}\mathbb{1}\left\{\|\mathbf{\Sigma}_R\|_{\text{op}} \geq \sqrt{2\log[(N+K)n^{3/2}]/n}\right\}\right] \\
&\quad + \mathbb{E}\left[\|\mathbf{\Sigma}_R\|_{\text{op}}\mathbb{1}\left\{\|\mathbf{\Sigma}_R\|_{\text{op}} < \sqrt{2\log[(N+K)n^{3/2}]/n}\right\}\right] \\
&\leq \mathbb{P}\left[\|\mathbf{\Sigma}_R\|_{\text{op}} \geq \sqrt{2\log[(N+K)n^{3/2}]/n}\right] \times n + \sqrt{2\log[(N+K)n^{3/2}]/n} \\
&\leq \frac{1}{\sqrt{n}} + \frac{\sqrt{2\log[(N+K)n^{3/2}]}}{\sqrt{n}} \leq \frac{3\sqrt{\log[(N+K)n^{3/2}]}}{\sqrt{n}} \leq \frac{6\sqrt{\log[(N+K)n]}}{\sqrt{n}}. \quad \square
\end{aligned}$$

Appendix B: Proof of Theorem 2

Proof of Theorem 2. Consider a set $\{\mathbf{\Theta}_1, \mathbf{\Theta}_2, \cdots, \mathbf{\Theta}_{n(\varsigma)}\}$ where $\mathbf{\Theta}_i \in \mathcal{C}, \|\mathbf{\Theta}_i\|_F \leq \varsigma, \forall i \in [n(\varsigma)]$ and $\|\mathbf{\Theta}_i - \mathbf{\Theta}_j\|_F \geq \varsigma, \forall i \neq j$. We first choose index $m \in [n(\varsigma)]$ uniformly at random, and we are given the observations $\mathcal{S} = \{(\mathbf{X}_t^i, Y_t^{(i)}), t \in [n], i \in [N]\}$ sampled according to (2.4) with $\mathbf{\Theta}^* = \mathbf{\Theta}_m$. Suppose we obtain an estimation of m by samples in $\mathcal{S}$, denoted by $\widehat{m}$. Then we have

$$\mathbb{P}\left[\|\widetilde{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_F \geq \frac{\varsigma}{2}\right] \geq \mathbb{P}[\widehat{m} \neq m]$$

by the triangle inequality. The Fano's inequality (see, e.g., Theorem 9 of Scarlett and Cevher (2019)) yields that

$$\mathbb{P}[\widehat{m} \neq m \mid \mathcal{X}_1, \cdots, \mathcal{X}_n] \geq 1 - \frac{\max_{i,j\in[n(\varsigma)], i\neq j} \mathrm{D}(\mathbf{\Theta}_i \,\|\, \mathbf{\Theta}_j) + \log 2}{\log\big(n(\varsigma)\big)}, \tag{B.1}$$

where $\mathrm{D}(\mathbf{\Theta}_i \,\|\, \mathbf{\Theta}_j)$ is the KL divergence between the distributions of

$$(Y_1, Y_2, \cdots, Y_n \mid \mathcal{X}_1, \cdots, \mathcal{X}_n, \mathbf{\Theta}_i) \quad \text{and} \quad (Y_1, Y_2, \cdots, Y_n \mid \mathcal{X}_1, \cdots, \mathcal{X}_n, \mathbf{\Theta}_j).$$

For Gaussian noises with variance σ^2, we have, for any $i \neq j$,

$$\mathrm{D}(\mathbf{\Theta}_i \parallel \mathbf{\Theta}_j) = \frac{1}{2\sigma^2}\sum_{t=1}^{n} \|\mathcal{X}_t(\mathbf{\Theta}_i) - \mathcal{X}_t(\mathbf{\Theta}_j)\|^2$$

and thus

$$\mathbb{E}[\mathrm{D}(\mathbf{\Theta}_i \parallel \mathbf{\Theta}_j)] = \frac{n}{2K\sigma^2}\|\mathbf{\Theta}_i - \mathbf{\Theta}_j\|_F^2$$

where the expectation is taken over $(\mathcal{X}_1, \cdots, \mathcal{X}_n)$. Then by (B.1), we have

$$\begin{aligned}\mathbb{P}[\widehat{m} \neq m] &\geq 1 - \frac{\frac{n}{2K\sigma^2}\|\mathbf{\Theta}_i - \mathbf{\Theta}_j\|_F^2 + \log 2}{\log\big(n(\varsigma)\big)} \\ &\geq 1 - \frac{\frac{2n\varsigma^2}{K\sigma^2} + \log 2}{\log\big(n(\varsigma)\big)}\end{aligned}$$

where the last inequality is given by $\|\mathbf{\Theta}_i - \mathbf{\Theta}_j\|_F \leq \|\mathbf{\Theta}_i\|_F + \|\mathbf{\Theta}_j\|_F \leq 2\varsigma$. By the proof of Theorem 5 in Koltchinskii et al. (2011), for any constant $0 < \upsilon \leq 1$, when

$$\varsigma = \upsilon(\sigma \wedge 1)\sqrt{\frac{rK^2}{n}},$$

there exists $\{\mathbf{\Theta}_1, \mathbf{\Theta}_2, \cdots, \mathbf{\Theta}_{n(\varsigma)}\}$ with $n(\varsigma) \geq 2^{rK/8} + 1$ that satisfies $\|\mathbf{\Theta}_i\|_F = \varsigma, \mathbf{\Theta}_i \in \mathcal{C}, \forall i \in [n(\varsigma)]$ and $\|\mathbf{\Theta}_i - \mathbf{\Theta}_j\|_F \geq \varsigma, \forall i \neq j$. Then choosing $\upsilon = \sqrt{\log 2}/8$ yields

$$\mathbb{P}[\widehat{m} \neq m] \geq 1 - \frac{\frac{2\upsilon^2(\sigma^2 \wedge 1)rK^2}{K\sigma^2} + \log 2}{\frac{rK\log 2}{8}} \geq 1 - \frac{2\upsilon^2 rK + \log 2}{\frac{rK\log 2}{8}} = 1 - \frac{rK/4 + 8}{rK} \geq \frac{1}{2},$$

when $rK \geq 32$. Thus, we have

$$\begin{aligned}\ell(\mathbf{\Theta}^*, \|\cdot\|_F) &= \inf_{\widetilde{\mathbf{\Theta}}} \sup_{\mathbf{\Theta}^* \in \mathcal{C}} \mathbb{E}\left[\|\widetilde{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_F\right] \\ &\geq \inf_{\widetilde{\mathbf{\Theta}}} \sup_{\mathbf{\Theta}^* \in \mathcal{C}} \frac{\varsigma}{2}\mathbb{P}\left[\|\widetilde{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_F \geq \frac{\varsigma}{2}\right] \\ &\geq \inf_{\widetilde{\mathbf{\Theta}}} \frac{\varsigma}{2}\mathbb{P}\left[\hat{m} \neq m\right] \\ &\geq \frac{\varsigma}{4} = \frac{\sqrt{\log 2}(\sigma \wedge 1)}{8}\sqrt{\frac{rK^2}{n}},\end{aligned}$$

which completes our proof.

Appendix C: Proof of Theorem 3

C.1. Major Steps of the Proof

We first define the conditional number for any matrix $\mathbf{\Theta}$, as $\kappa = \sigma_{\max}(\mathbf{\Theta})/\sigma_{\min}(\mathbf{\Theta})$. The very first step of our proof will be showing that the number of samples collected for each row and column in the enhancement procedure is proportional to the total number of enhancement samples $n_2 = |\mathcal{J}_2|$. To establish this, we let

$$n_{\min} = \min\left\{\min_{i \in [N]} |\mathcal{R}_i|, \min_{j \in [K]} |\mathcal{C}_j|\right\}$$

denote the minimum sample size of each row $i \in [N]$ and each column $j \in [K]$ in the enhancement procedure, and we have the following lemma which is proved in Appendix C.3:

Lemma C.1. *If*

$$n_2 \geq \frac{\alpha K^2}{N^2},$$

then

$$n_{\min} \geq \frac{n_2 N}{2K}$$

with probability greater than $1-\exp(-\alpha)$ *for any* $\alpha > 0$.

Given the minimum sample size requirement, we are now able to derive Theorem 3. We break our proof into two steps.

The first step (Lemma C.2) shows that with sufficient samples at each row and column, the $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$ returned by the row enhancement procedure in Algorithm 1 will better approximate $\mathbf{U}^*\mathbf{D}^*$ and $\mathbf{V}^*\mathbf{D}^*$ (recall the SVD $\mathbf{\Theta}^* = \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\top}$); specifically, the row-wise distance, i.e., $\ell_{2,\infty}$ norm error, is improved. This improvement is enabled by two least square regressions in the double-enhancement procedure Algorithm 1. The proof technique for this lemma is similar to that in Hamidi et al. (2019). However, as aforementioned, they are only interested in the row-wise error $\|\widetilde{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_{2,\infty}$, and hence only discuss how to enhance $\widehat{\mathbf{V}}$ to better approximate the row space of $\mathbf{\Theta}^*$ but not the column space (i.e., $\widehat{\mathbf{U}}$). In contrast, our analysis requires a simultaneous enhancement of both $\widetilde{\mathbf{V}}$ and $\widetilde{\mathbf{U}}$, which is more involved. The proof of Lemma C.2 is in Appendix C.4.

Lemma C.2. *Suppose there exists* $\epsilon_f > 0$ *such that the first stage estimator* $\widehat{\mathbf{\Theta}}$ *has*

$$\frac{\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\|_F}{\sqrt{NK}} \leq \epsilon_f \leq \frac{\|\mathbf{\Theta}^*\|_\infty}{2\eta\kappa\sqrt{r}}; \tag{C.1}$$

besides, assume that the minimum row and column sample size has

$$n_{\min} \geq \frac{64\kappa^2\eta^2 r \log r \log(N+K)}{\|\mathbf{\Theta}^*\|_\infty^2}. \tag{C.2}$$

Then, the $\widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}$ *returned by Algorithm 1 satisfy*

$$\|\widetilde{\mathbf{U}} - \mathbf{U}^*\mathbf{D}^*\|_{2,\infty} \leq \left(\frac{15\epsilon_f\kappa\eta\sqrt{r}}{2\|\mathbf{\Theta}^*\|_\infty} + \frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\mathbf{\Theta}^*\|_\infty^2} \right) \|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty},$$

and

$$\|\widetilde{\mathbf{V}} - \mathbf{V}^*\mathbf{D}^*\|_{2,\infty} \leq \left(\frac{15\epsilon_f\kappa\eta\sqrt{r}}{2\|\mathbf{\Theta}^*\|_\infty} + \frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\mathbf{\Theta}^*\|_\infty^2} \right) \|\mathbf{V}^*\mathbf{D}^*\|_{2,\infty}$$

with probability greater than $1 - 3(N+K)\exp(-\alpha)$ *for any* $\alpha > 0$.

Under the sample size requirement in Theorem 3, ϵ_f in the above lemma is of order $\widetilde{\mathcal{O}}(r\sqrt{K/Nn})$ according to Theorem 1 given our choice of λ, which will be helpful in establishing the condition in our second key step.

The second step (Lemma C.3) shows that, by setting $\widetilde{\mathbf{\Theta}} = \mathbf{U}_1\mathbf{Q}_1\widetilde{\mathbf{V}}^\top$, the entry-wise error $\|\widetilde{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_\infty$ is controlled by the row-wise bounds $\|\widetilde{\mathbf{U}} - \mathbf{U}^*\mathbf{D}^*\|_{2,\infty}$ and $\|\widetilde{\mathbf{V}} - \mathbf{V}^*\mathbf{D}^*\|_{2,\infty}$ derived in Lemma C.2. This bound is deterministic, leading towards our final probabilistic bound in Theorem 3. The proof of Lemma C.3 is provided in Appendix C.5.

Lemma C.3. *Suppose there exists $\epsilon < 1/(2\sqrt{r}\eta\kappa)$ such that*

$$\|\widetilde{\mathbf{U}} - \mathbf{U}^*\mathbf{D}^*\|_{2,\infty} \le \epsilon\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty} \quad \textit{and} \quad \|\widetilde{\mathbf{V}} - \mathbf{V}^*\mathbf{D}^*\|_{2,\infty} \le \epsilon\|\mathbf{V}^*\mathbf{D}^*\|_{2,\infty}. \tag{C.3}$$

Then, $\widetilde{\mathbf{\Theta}} = \mathbf{U}_1\mathbf{Q}_1\widetilde{\mathbf{V}}^\top$ satisfies

$$\|\widetilde{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_\infty \le \frac{33\epsilon\sqrt{r}\eta\kappa\|\mathbf{\Theta}^*\|_\infty}{4},$$

where $\mathbf{U}_1, \mathbf{Q}_1$ are from the SVD $\widetilde{\mathbf{U}} = \mathbf{U}_1\mathbf{D}_1\mathbf{Q}_1$.

The ϵ in Lemma C.3 exists with probability higher than $1 - 3(N+K)\exp(-\alpha)$ for any $\alpha > 0$ if we set

$$\epsilon = \frac{15\epsilon_f\kappa\eta\sqrt{r}}{2\|\mathbf{\Theta}^*\|_\infty} + \frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\mathbf{\Theta}^*\|_\infty^2},$$

according to our results in Lemma C.2. As aforementioned, ϵ_f is of order $\widetilde{\mathcal{O}}(r\sqrt{K/Nn})$, thus ϵ is of order $\widetilde{\mathcal{O}}(r^{3/2}\sqrt{K/Nn})$. By Lemma C.3, this yields our final result order of $\widetilde{\mathcal{O}}(r^2\sqrt{K/Nn})$.

C.2. Proof of Theorem 3

Proof of Theorem 3. First we provide the explicit form of our condition $n = \widetilde{\Omega}\left(\max\{(r^4K)/N, (K/N)^2\}\right)$; that is,

$$n \ge \max\left\{\frac{2\alpha K^2}{N^2} + 1, \frac{256\kappa^2\eta^2 rK\log r\log(N+K)}{\|\mathbf{\Theta}^*\|_\infty^2 N} + 1, \frac{1800c_4^2c_\lambda^2\sigma^2\eta^4\kappa^4r^3K\left(\alpha + \log(N+K)\right)^2}{N\|\mathbf{\Theta}^*\|_\infty^2},\right.$$
$$\left.\frac{1800c_4^2\eta^4\kappa^4r^4K\log[(N+K)n]}{N\|\mathbf{\Theta}^*\|_\infty^2}, \frac{2048\alpha\sigma^2\eta^6\kappa^4r^3K}{N\|\mathbf{\Theta}^*\|_\infty^2} + 1\right\},$$

where c_4 is the constant from (A.4). This requirement for sample size is to ensure that the conditions in Lemma C.2 and Lemma C.3 hold with high probability. Specifically, we will show that if the sample size requirement is met, then

$$n_{\min} \ge \frac{n_2N}{2K} \ge \frac{(n-1)N}{4K} \ge \frac{64\kappa^2\eta^2r\log r\log(N+K)}{\|\mathbf{\Theta}^*\|_\infty^2} \tag{C.4}$$

with probability higher than $1 - \exp(-\alpha)$ and

$$\frac{\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\|_F}{\sqrt{NK}} \le \epsilon_f \le \frac{\|\mathbf{\Theta}^*\|_\infty}{30r\eta^2\kappa^2} \le \frac{\|\mathbf{\Theta}^*\|_\infty}{2\eta\kappa\sqrt{r}} \tag{C.5}$$

with probability higher than $1 - 4\exp(-\alpha)$ by choosing

$$\epsilon_f = c_4\max\left\{c_\lambda\sigma\left(\alpha + \log(N+K)\right)\sqrt{\frac{rK}{Nn}}, r\sqrt{\frac{K\log[(N+K)n]}{Nn}}\right\}.$$

We first show (C.4) holds with high probability. By Lemma C.1 and our assumption that

$$n \ge \frac{256\kappa^2\eta^2rK\log r\log(N+K)}{N\|\mathbf{\Theta}^*\|_\infty^2} + 1,$$

we immediately have (C.4) holds with probability higher than $1 - \exp(-\alpha)$. Then we will show that (C.5) holds with high probability via Theorem 1. The explicit form of Theorem 1 in (A.4) tells us with probability higher than $1 - 4\exp(-\alpha)$,

$$\frac{\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\|_F}{\sqrt{NK}} \le c_4\max\left\{c_\lambda\sigma\left(\alpha + \log(N+K)\right)\sqrt{\frac{rK}{Nn_1}}, r\sqrt{\frac{K\log[(N+K)n_1]}{Nn_1}}\right\}.$$

Since

$$n_1 \geq \frac{n}{2} \geq \max\left\{\frac{900c_4^2c_\lambda^2\sigma^2\eta^4\kappa^4r^3K\big(\alpha+\log(N+K)\big)^2}{N\|\boldsymbol{\Theta}^*\|_\infty^2}, \frac{900c_4^2\eta^4\kappa^4r^4K\log[(N+K)n]}{N\|\boldsymbol{\Theta}^*\|_\infty^2}\right\},$$

by choosing

$$\epsilon_f = c_4\max\left\{c_\lambda\sigma\left(\alpha+\log(N+K)\right)\sqrt{\frac{rK}{Nn_1}}, r\sqrt{\frac{K\log[(N+K)n_1]}{Nn_1}}\right\},$$

we have

$$\frac{\big\|\boldsymbol{\Theta}^*-\widehat{\boldsymbol{\Theta}}\big\|_F}{\sqrt{NK}} \leq \epsilon_f \leq \frac{\|\boldsymbol{\Theta}^*\|_\infty}{30r\eta^2\kappa^2} \leq \frac{\|\boldsymbol{\Theta}^*\|_\infty}{2\eta\kappa\sqrt{r}}$$

with probability higher than $1-4\exp(-\alpha)$, where the last inequality is given by $\eta,\kappa,r\geq 1$.

By our choice of ϵ_f and sufficient sample size established in (C.4) and (C.5), we have the conditions in Lemma C.2 hold. Therefore,

$$\|\widetilde{\mathbf{U}}-\mathbf{U}^*\mathbf{D}^*\|_{2,\infty} \leq \left(\frac{15\epsilon_f\kappa\eta\sqrt{r}}{2\|\boldsymbol{\Theta}^*\|_\infty}+\frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\boldsymbol{\Theta}^*\|_\infty^2}\right)\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty}$$

$$\|\widetilde{\mathbf{V}}-\mathbf{V}^*\mathbf{D}^*\|_{2,\infty} \leq \left(\frac{15\epsilon_f\kappa\eta\sqrt{r}}{2\|\boldsymbol{\Theta}^*\|_\infty}+\frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\boldsymbol{\Theta}^*\|_\infty^2}\right)\|\mathbf{V}^*\mathbf{D}^*\|_{2,\infty}$$

with probability higher than $1-(3N+3K)\exp(-\alpha)$. This further implies that, as long as

$$\frac{15\epsilon_f\kappa\eta\sqrt{r}}{2\|\boldsymbol{\Theta}^*\|_\infty}+\frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\boldsymbol{\Theta}^*\|_\infty^2} \leq \frac{1}{2\sqrt{r}\eta\kappa}, \tag{C.6}$$

then the condition in Lemma C.3 is satisfied with probability higher than $1-(3N+3K)\exp(-\alpha)$ by choosing

$$\epsilon = \frac{15\epsilon_f\kappa\eta\sqrt{r}}{2\|\boldsymbol{\Theta}^*\|_\infty}+\frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\boldsymbol{\Theta}^*\|_\infty^2}.$$

Now we show that (C.6) holds. By (C.4) and

$$n \geq \frac{2048\alpha\sigma^2\eta^6\kappa^4r^3K}{N\|\boldsymbol{\Theta}^*\|_\infty^2}+1,$$

we have

$$n_{\min} \geq \frac{(n-1)N}{4K} \geq \frac{512\alpha\sigma^2\eta^6\kappa^4r^3}{\|\boldsymbol{\Theta}^*\|_\infty^2};$$

Thus, it gives

$$\frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\boldsymbol{\Theta}^*\|_\infty^2} \leq \frac{1}{4\sqrt{r}\eta\kappa}.$$

By combining this argument with (C.5), we have

$$\frac{15\epsilon_f\kappa\eta\sqrt{r}}{2\|\boldsymbol{\Theta}^*\|_\infty}+\frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\boldsymbol{\Theta}^*\|_\infty^2} \leq \frac{15\frac{\|\boldsymbol{\Theta}^*\|_\infty}{30r\eta^2\kappa^2}\kappa\eta\sqrt{r}}{2\|\boldsymbol{\Theta}^*\|_\infty}+\frac{1}{4\sqrt{r}\eta\kappa} \leq \frac{1}{2\sqrt{r}\eta\kappa},$$

which implies that (C.6) holds.

Now by our choice of ϵ, the condition in Lemma C.3 holds with probability higher than $1-(3N+3K)\exp(-\alpha)$. We have

$$\epsilon = \frac{15\epsilon_f\kappa\eta\sqrt{r}}{2\|\boldsymbol{\Theta}^*\|_\infty}+\frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\boldsymbol{\Theta}^*\|_\infty^2}$$

$$
\begin{aligned}
&= \frac{15\kappa\eta\sqrt{r}}{2\|\boldsymbol{\Theta}^*\|_\infty}\left(c_4 \max\left\{c_\lambda\sigma\left(\alpha+\log(N+K)\right)\sqrt{\frac{rK}{Nn_1}}, r\sqrt{\frac{\log[(N+K)n_1]}{Nn_1}}\right\}\right)+\frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\boldsymbol{\Theta}^*\|_\infty^2}\\
&\le \frac{15\kappa\eta\sqrt{r}}{2\|\boldsymbol{\Theta}^*\|_\infty}\left(c_4 \max\left\{c_\lambda\sigma\left(\alpha+\log(N+K)\right)\sqrt{\frac{2rK}{Nn}}, r\sqrt{\frac{2K\log[(N+K)n1]}{Nn}}\right\}\right)+\frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\|\boldsymbol{\Theta}^*\|_\infty^2}\sqrt{\frac{4K}{N(n-1)}}\\
&\le c_5\frac{\kappa\eta r}{\|\boldsymbol{\Theta}^*\|_\infty}\left(\max\left\{c_\lambda\sigma\left(\alpha+\log(N+K)\right)\sqrt{\frac{K}{Nn}}, \sqrt{\frac{rK\log[(N+K)n]}{Nn}}\right\}+\frac{\eta\sigma\sqrt{\alpha}}{\|\boldsymbol{\Theta}^*\|_\infty}\sqrt{\frac{K}{Nn}}\right),
\end{aligned}
$$

where c_5 is an absolute constant. Plugging the above into Lemma C.3 yields

$$
\begin{aligned}
&\|\widetilde{\boldsymbol{\Theta}}-\boldsymbol{\Theta}^*\|_\infty\\
&\le \frac{33\epsilon\sqrt{r}\eta\kappa\|\boldsymbol{\Theta}^*\|_\infty}{4}\\
&\le \frac{33\sqrt{r}\eta\kappa\|\boldsymbol{\Theta}^*\|_\infty}{4}\cdot c_5\frac{\kappa\eta r}{\|\boldsymbol{\Theta}^*\|_\infty}\left(\max\left\{c_\lambda\sigma\left(\alpha+\log(N+K)\right)\sqrt{\frac{rK}{Nn}}, \sqrt{\frac{rK\log[(N+K)n]}{Nn}}\right\}+\frac{\eta\sigma\sqrt{\alpha}}{\|\boldsymbol{\Theta}^*\|_\infty}\sqrt{\frac{K}{Nn}}\right)\\
&\le c_6\eta^2\kappa^2 r^{3/2}\max\left\{c_\lambda\sigma\left(\alpha+\log(N+K)\right)\sqrt{\frac{K}{Nn}}, \sqrt{\frac{rK\log[(N+K)n]}{Nn}}, \frac{\eta\sigma\sqrt{\alpha}}{\|\boldsymbol{\Theta}^*\|_\infty}\sqrt{\frac{K}{Nn}}\right\}, \qquad \text{(C.7)}
\end{aligned}
$$

where c_6 is an absolute constant.

Since (C.4) and (C.5) hold with probability higher than $1-5\exp(-\alpha)$, we have (C.7) holds with probability higher than $1-(3N+3K+5)\exp(-\alpha)$, which completes the proof. □

C.3. Proof of Lemma C.1

Proof of Lemma C.1. Let $r_i=|\mathcal{R}_i|$ and $c_i=|\mathcal{C}_j|$ where $\mathcal{R}_i$ and $\mathcal{C}_j$ are set in Algorithm 1 and $i\in[N], j\in[K]$. By the definition of matching, we have $r_i=n_2, \forall i\in[N]$. Then, it suffices to show that

$$\min_{j\in[K]} c_j \ge \frac{N}{2Kn_2}$$

with high probability. We have for any $j\in[K]$,

$$c_j=|\{t\in\mathcal{J}_2, j_t(i)=j\}|=\sum_{t\in\mathcal{J}_2}\sum_{i\in[N]}\mathbb{I}\{j_t(i)=j\}.$$

By our matching structure, for any $t\in\mathcal{J}_2$, $\sum_{i\in[N]}\mathbb{I}\{j_t(i)=j\}$ are i.i.d. Bernoulli random variables with expectation N/K. Then, by Hoeffding inequality, we have

$$\mathbb{P}\left[\left|c_j-\frac{n_2N}{K}\right|\ge\frac{n_2N}{2K}\right]\le 2\exp\left(\frac{-2\left(\frac{n_2N}{2K}\right)^2}{n_2}\right)\le 2\exp\left(-\frac{n_2N^2}{2K^2}\right)\le 2\exp(-\alpha),$$

where the last inequality is given by the condition $n_2\ge 2K^2\alpha/N^2$ of this lemma.

C.4. Proof of Lemma C.2

Proof of Lemma C.2. We will prove the result for $\widetilde{\mathbf{U}}$, and the result for $\widetilde{\mathbf{V}}$ can be obtained via an analogous argument. We first recall some notations for this proof. We let $e_i(N)$ for $i\in[N]$ denote the canonical basis vector in $\mathbb{R}^N$; that is, $e_i(N)$ is a column vector with 1 in the i-th entry and 0 in the other entries. Similarly, $e_j(K)$ for $j\in[K]$ denotes the basis vector in $\mathbb{R}^K$.

Let $\beta^\top=\mathbf{U}^{*(i,\cdot)}\mathbf{D}^*=e_i(N)^\top\mathbf{U}^*\mathbf{D}^*\in\mathbb{R}^r$ for abbreviation. Recall that in Algorithm 1, we use $\mathcal{R}_i$ to denote $\{(\mathbf{X}_t^i, Y_t^{(i)})\mid t\in\mathcal{J}_2\}$, which is a set of enhancement samples from row i. Define $r_i=|\mathcal{R}_i|$ and index the elements in $\mathcal{R}_i$ as

$$(\mathbf{Z}_k, y_k), k\in[r_i].$$

Then, define

$$\mathbf{Z} = \begin{bmatrix} e_i(N)^\top \mathbf{Z}_1 \\ \vdots \\ e_i(N)^\top \mathbf{Z}_{r_i} \end{bmatrix} \in \mathbb{R}^{r_i \times K},$$

and the response vector

$$Y = \begin{bmatrix} y_1 \cdots y_{r_i} \end{bmatrix}^\top.$$

By (2.4), we have

$$Y = \mathbf{Z}\mathbf{V}^* \beta + \varepsilon,$$

where ε is a vector in $\mathbb{R}^{r_i}$ consisting of independent σ-subgaussian noises. Then the solution $\tilde{\beta}_i$ in Algorithm 1 is equivalent to

$$\widetilde{\beta}_i \in \operatorname*{arg\,min}_{\gamma \in \mathbb{R}^r} \|Y - \mathbf{Z}\widehat{\mathbf{V}}\gamma\|^2. \tag{C.8}$$

Define

$$\mathbf{H} = \mathbf{Z}\widehat{\mathbf{V}}.$$

Assume that $\mathbf{H}^\top \mathbf{H}$ is invertible, which implies that the unique solution for (C.8) is

$$\operatorname*{arg\,min}_{\gamma \in \mathbb{R}^r} \|Y - \mathbf{Z}\widehat{\mathbf{V}}\gamma\|^2 = \left(\mathbf{H}^\top \mathbf{H}\right)^{-1} \mathbf{H}^\top Y.$$

We use $\widetilde{\beta} = \left(\mathbf{H}^\top \mathbf{H}\right)^{-1} \mathbf{H}^\top Y$ to denote this solution for ease of notation. We will show later in (C.11) that $\mathbf{H}^\top \mathbf{H}$ is invertible with high probability. Now we proceed to analyze $\|\widetilde{\beta} - \beta\|$. We have

$$\begin{aligned} \widetilde{\beta} &= (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top Y \\ &= (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top (\mathbf{Z}\mathbf{V}^* \beta + \varepsilon) \\ &= (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top (\mathbf{Z}\widehat{\mathbf{V}}\beta + \mathbf{Z}\mathbf{V}^*\beta - \mathbf{Z}\widehat{\mathbf{V}}\beta + \varepsilon) \\ &= (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{H}\beta + (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{Z}(\mathbf{V}^* - \widehat{\mathbf{V}})\beta + (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \varepsilon \\ &= \beta + (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{Z}(\mathbf{V}^* - \widehat{\mathbf{V}})\beta + (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \varepsilon. \end{aligned}$$

So

$$\begin{aligned} \|\widetilde{\beta} - \beta\| &= \left\|(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{Z}(\mathbf{V}^* - \widehat{\mathbf{V}})\beta + (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \varepsilon\right\| \\ &\le \left\|(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{Z}(\mathbf{V}^* - \widehat{\mathbf{V}})\beta\right\| + \left\|(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \varepsilon\right\| \\ &\le \left\|(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{Z}\right\|_{\mathrm{op}} \left\|\mathbf{V}^* - \widehat{\mathbf{V}}\right\|_{\mathrm{op}} \|\beta\| + \left\|(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \varepsilon\right\| \\ &\le \underbrace{\left\|(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{Z}\right\|_{\mathrm{op}}}_{=h_1} \left\|\mathbf{V}^* - \widehat{\mathbf{V}}\right\|_F \|\beta\| + \underbrace{\left\|(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \varepsilon\right\|}_{=h_2}. \end{aligned} \tag{C.9}$$

We first bound h_2. We have

$$\begin{aligned} h_2 &= \left\|\left(\mathbf{H}^\top \mathbf{H}\right)^{-1} \mathbf{H}^\top \varepsilon\right\| \\ &\le \left\|\left(\mathbf{H}^\top \mathbf{H}\right)^{-1}\right\|_{\mathrm{op}} \|\mathbf{H}^\top \varepsilon\| \\ &= \frac{\|\mathbf{H}^\top \varepsilon\|}{\lambda_{\min}(\mathbf{H}^\top \mathbf{H})}. \end{aligned} \tag{C.10}$$

So we will upper bound h_2 by upper bounding $\mathbf{H}^\top \varepsilon$ and lower bounding $\lambda_{\min}(\mathbf{H}^\top \mathbf{H})$. We will use Lemma F.5 on the norm of the weighted sum of subgaussian vectors to upper bound $\|\mathbf{H}^\top \varepsilon\|$. Notice that

$$\|\mathbf{H}^\top \varepsilon\| = \|\varepsilon^\top \mathbf{H}\| = \left\| \sum_{k=1}^{r_i} \varepsilon^{(k)} e_i(N)^\top \mathbf{Z}_k \widehat{\mathbf{V}} \right\|$$

and each random vector $e_i(N)^\top \mathbf{Z}_k \widehat{\mathbf{V}} \in \mathbb{R}^r$ follows an independent and uniform distribution over the set $\{e_j(K)^\top \widehat{\mathbf{V}}, j \in [K]\}$, thus $\|e_i(N)^\top \mathbf{Z}_k \widehat{\mathbf{V}}\| \leq \|\widehat{\mathbf{V}}\|_{2,\infty}, \forall k \in [r_i]$. Then by Lemma F.5, we have for any $\rho > 0$,

$$\mathbb{P}\left[\left\| \sum_{k=1}^{r_i} \varepsilon^{(k)} e_i(N)^\top \mathbf{Z}_k \widehat{\mathbf{V}} \right\| \geq \rho \right] \leq 2\exp\left(\frac{-\rho^2}{2 r_i r \|\widehat{\mathbf{V}}\|_{2,\infty}^2 \sigma^2} \right) \leq 2\exp\left(\frac{-\rho^2}{2 n_{\min} r \|\widehat{\mathbf{V}}\|_{2,\infty}^2 \sigma^2} \right)$$

where the second inequality is given by the definition $n_{\min} = \min\{\min_{i\in[N]} |\mathcal{R}_i|, \min_{j\in[K]} |\mathcal{C}_j|\}$. Taking $\rho = \sigma\sqrt{2 r \alpha n_{\min}} \|\widehat{\mathbf{V}}\|_{2,\infty}$ yields

$$\mathbb{P}\left[\left\| \mathbf{H}^\top \varepsilon \right\| \geq \sigma \sqrt{2 r \alpha n_{\min}} \|\widehat{\mathbf{V}}\|_{2,\infty} \right] \leq 2\exp(-\alpha).$$

We will then use the matrix Chernoff bound in Lemma F.6 to lower bound $\lambda_{\min}(\mathbf{H}^\top \mathbf{H})$. Note that by Weyl's inequality (Lemma F.7),

$$\begin{aligned}
\lambda_{\min}(\mathbf{H}^\top \mathbf{H}) &= \lambda_{\min}\left(\sum_{k=1}^{r_i} \widehat{\mathbf{V}}^\top \mathbf{Z}_k^\top e_i(N) e_i(N)^\top \mathbf{Z}_k \widehat{\mathbf{V}} \right) \\
&\geq \lambda_{\min}\left(\sum_{k=1}^{n_{\min}} \widehat{\mathbf{V}}^\top \mathbf{Z}_k^\top e_i(N) e_i(N)^\top \mathbf{Z}_k \widehat{\mathbf{V}} \right).
\end{aligned}$$

By our previous argument that each random vector $e_i(N)^\top \mathbf{Z}_k \widehat{\mathbf{V}} \in \mathbb{R}^r$ follows an independent and uniform distribution over the set $\{e_j(K)^\top \widehat{\mathbf{V}}, j \in [K]\}$, we have

$$\lambda_{\max}(\widehat{\mathbf{V}}^\top \mathbf{Z}_k^\top e_i(N) e_i(N)^\top \mathbf{Z}_k \widehat{\mathbf{V}}) = \|e_i(N)^\top \mathbf{Z}_k \widehat{\mathbf{V}}\|^2 \leq \|\widehat{\mathbf{V}}\|_{2,\infty}^2, \forall k \in [r_i]$$

and

$$\begin{aligned}
\lambda_{\min}\left(\mathbb{E}\left[\widehat{\mathbf{V}}^\top \mathbf{Z}_k^\top e_i(N) e_i(N)^\top \mathbf{Z}_k \widehat{\mathbf{V}} \right] \right) &= \lambda_{\min}\left(\sum_{j=1}^{K} \frac{1}{K} \widehat{\mathbf{V}}^\top e_j(K) e_j(K)^\top \widehat{\mathbf{V}} \right) \\
&= \frac{1}{K} \lambda_{\min}\left[\widehat{\mathbf{V}}^\top \left(\sum_{j=1}^{K} e_j(K) e_j(K)^\top \right) \widehat{\mathbf{V}} \right] \\
&= \frac{1}{K} \lambda_{\min}\left(\widehat{\mathbf{V}}^\top \mathbf{I}_{K\times K} \widehat{\mathbf{V}} \right) \\
&= \frac{1}{K} \lambda_{\min}(\mathbf{I}_{r\times r}) = \frac{1}{K}.
\end{aligned}$$

Thus, plugging $\mu_{\min} = n_{\min}/K$ and $\omega = \|\widehat{\mathbf{V}}\|_{2,\infty}^2$ into Lemma F.6, we obtain

$$\mathbb{P}\left[\lambda_{\min}\Big(\sum_{k=1}^{n_{\min}} \widehat{\mathbf{V}}^\top \mathbf{Z}_k^\top e_i(N) e_i(N)^\top \mathbf{Z}_k \widehat{\mathbf{V}} \Big) \leq \frac{\rho n_{\min}}{K} \right] \leq r \exp\left(\frac{-(1-\rho)^2 n_{\min}}{2K \|\widehat{\mathbf{V}}\|_{2,\infty}^2} \right), \forall \rho \in (0,1).$$

Then, set $\rho = 1/2$. If $n_{\min} \geq 8\alpha K \|\widehat{\mathbf{V}}\|_{2,\infty}^2 \log r$, we have

$$\begin{aligned}
\mathbb{P}\left[\lambda_{\min}(\mathbf{H}^\top \mathbf{H}) \leq \frac{n_{\min}}{2K} \right] &\leq \mathbb{P}\left[\lambda_{\min}\left(\sum_{k=1}^{n_{\min}} \widehat{\mathbf{V}}^\top \mathbf{Z}_k^\top e_i(N) e_i(N)^\top \mathbf{Z}_k \widehat{\mathbf{V}} \right) \leq \frac{n_{\min}}{2K} \right] \\
&\leq r \exp\left(\frac{-n_{\min}}{8K \|\widehat{\mathbf{V}}\|_{2,\infty}^2} \right) \leq \exp(-\alpha), \qquad \text{(C.11)}
\end{aligned}$$

which implies that $\mathbf{H}^\top\mathbf{H}$ is invertible with probability higher than $1-\exp(-\alpha)$, and furthermore,

$$\begin{aligned}
h_2 \le \frac{\|\mathbf{H}^\top\varepsilon\|}{\lambda_{\min}(\mathbf{H}^\top\mathbf{H})} &\le \frac{\sigma\sqrt{2r\alpha n_{\min}}\|\widehat{\mathbf{V}}\|_{2,\infty}}{\frac{n_{\min}}{2K}} \\
&= \frac{2K\sigma\sqrt{2r\alpha}\|\widehat{\mathbf{V}}\|_{2,\infty}}{\sqrt{n_{\min}}\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty}}\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty} \\
&= \frac{2K\sigma\sqrt{2r\alpha}\|\widehat{\mathbf{V}}\|_{2,\infty}}{\sqrt{n_{\min}}\|\mathbf{\Theta}^*\|_{2,\infty}}\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty} \\
&\le \frac{2K\sigma\sqrt{2r\alpha}\|\widehat{\mathbf{V}}\|_{2,\infty}}{\sqrt{n_{\min}}\frac{\|\mathbf{\Theta}^*\|_F}{\sqrt{N}}}\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty} \\
&\le \frac{2K\sigma\sqrt{2r\alpha}\|\widehat{\mathbf{V}}\|_{2,\infty}}{\sqrt{n_{\min}}\frac{\sqrt{NK}\|\mathbf{\Theta}^*\|_\infty}{\eta\sqrt{N}}}\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty} \\
&= \frac{2\eta\sigma\sqrt{2Kr\alpha}\|\widehat{\mathbf{V}}\|_{2,\infty}}{\sqrt{n_{\min}}\|\mathbf{\Theta}^*\|_\infty}\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty} \qquad \text{(C.12)}
\end{aligned}$$

with probability higher than $1-3\exp(-\alpha)$, where the second equation is given by (F.1) in Lemma F.8, and the last inequality is given by Assumption 1. In the following lemma, we provide an upper bound on $\|\widehat{\mathbf{V}}\|_{2,\infty}$ and thus $n_{\min} \ge 8\alpha K\|\widehat{\mathbf{V}}\|_{2,\infty}^2 \log r$ must hold as long as

$$n_{\min} \ge \frac{64\kappa^2\eta^2 r\log r\log(N+K)}{\|\mathbf{\Theta}^*\|_\infty^2} \quad \text{and} \quad \epsilon_f \le \frac{\|\mathbf{\Theta}^*\|_\infty}{2\eta\kappa\sqrt{r}}.$$

Lemma C.4. *If*

$$\epsilon_f \le \frac{\|\mathbf{\Theta}^*\|_\infty}{2\eta\kappa\sqrt{r}}, \qquad \text{(C.13)}$$

then

$$\sigma_{\min}(\widehat{\mathbf{\Theta}}) \ge \frac{1}{2\sigma_{\min}(\mathbf{\Theta}^*)}, \quad \textit{and} \quad \|\widehat{\mathbf{V}}\|_{2,\infty} \le \frac{2\kappa\eta}{\|\mathbf{\Theta}^*\|_\infty}\sqrt{\frac{r}{K}}.$$

We leave the proof of this lemma at the end of this subsection. Combining Lemma C.4 and (C.12), when (C.13) holds, we have

$$\begin{aligned}
h_2 &\le \frac{2\eta\sigma\sqrt{2Kr\alpha}\|\widehat{\mathbf{V}}\|_{2,\infty}}{\sqrt{n_{\min}}\|\mathbf{\Theta}^*\|_\infty}\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty} \\
&\le \frac{2\eta\sigma\sqrt{2Kr\alpha}\frac{2\kappa\eta}{\|\mathbf{\Theta}^*\|_\infty}\sqrt{\frac{r}{K}}}{\sqrt{n_{\min}}\|\mathbf{\Theta}^*\|_\infty}\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty} \\
&= \frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\mathbf{\Theta}^*\|_\infty^2}\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty}
\end{aligned}$$

with probability higher than $1-3\exp(-\alpha)$.

The term h_1 can be bounded via basic algebra. Notice that

$$\begin{aligned}
h_1 = \left\|(\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{Z}\right\|_{\mathrm{op}} &= \left\|(\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{Z}\widehat{\mathbf{V}}\right\|_{\mathrm{op}} \\
&= \left\|(\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{H}\right\|_{\mathrm{op}} \\
&= 1,
\end{aligned}$$

where the first equality is due to Lemma F.9. Plugging the above bounds of h_1 and h_2 into (C.9) gives us

$$\begin{aligned}\|\widetilde{\beta}-\beta\| &\le h_1\left\|\mathbf{V}^*-\widehat{\mathbf{V}}\right\|_F\left\|\beta\right\|+h_2\\ &\le \left\|\mathbf{V}^*-\widehat{\mathbf{V}}\right\|_F\left\|\beta\right\|+h_2\\ &\overset{(a)}{\le} 5\|\mathbf{\Theta}^*-\widehat{\mathbf{\Theta}}\|_F\left(\frac{1}{\sigma_{\min}(\mathbf{\Theta}^*)}+\frac{1}{\sigma_{\min}(\widehat{\mathbf{\Theta}})}\right)\|\beta\|+h_2\\ &\overset{(b)}{\le} \frac{15\epsilon_f\sqrt{NK}}{2\sigma_{\min}(\mathbf{\Theta}^*)}\|\beta\|+h_2\\ &\overset{(c)}{\le} \frac{15\epsilon_f\kappa\eta\sqrt{r}}{2\|\mathbf{\Theta}^*\|_\infty}\|\beta\|+h_2\\ &\le \left(\frac{15\epsilon_f\kappa\eta\sqrt{r}}{2\|\mathbf{\Theta}^*\|_\infty}+\frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\mathbf{\Theta}^*\|_\infty^2}\right)\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty} \qquad (C.14)\end{aligned}$$

with probability higher than $1-3\exp(-\alpha)$, if (C.1) and (C.2) hold. In detail, (a) is given by Lemma F.10, (b) is given by (C.1) and Lemma C.4, and (c) is given by (F.2) in Lemma F.8. By our definition, $\widetilde{\beta}-\beta=\left(\widetilde{\mathbf{U}}-\mathbf{U}^*\mathbf{D}^*\right)^{(i,\cdot)}$. Extending (C.14) to every $i\in[N]$, we have

$$\left\|\left(\widetilde{\mathbf{U}}-\mathbf{U}^*\mathbf{D}^*\right)^{(i,\cdot)}\right\|\le\left(\frac{15\epsilon_f\kappa\eta\sqrt{r}}{2\|\mathbf{\Theta}^*\|_\infty}+\frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\mathbf{\Theta}^*\|_\infty^2}\right)\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty}$$

with probability higher than $1-3\exp(-\alpha),\forall i\in[N]$. Since $\|\widetilde{\mathbf{U}}-\mathbf{U}^*\mathbf{D}^*\|_{2,\infty}=\max_{i\in[N]}\left\|\left(\widetilde{\mathbf{U}}-\mathbf{U}^*\mathbf{D}^*\right)^{(i,\cdot)}\right\|$, by union probability, we have

$$\mathbb{P}\left[\|\widetilde{\mathbf{U}}-\mathbf{U}^*\mathbf{D}^*\|_{2,\infty}\le\left(\frac{15\epsilon_f\kappa\eta\sqrt{r}}{2\|\mathbf{\Theta}^*\|_\infty}+\frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\mathbf{\Theta}^*\|_\infty^2}\right)\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty}\right]\ge 1-3N\exp(-\alpha).$$

Similarly, we can also show that

$$\mathbb{P}\left[\|\widetilde{\mathbf{V}}-\mathbf{V}^*\mathbf{D}^*\|_{2,\infty}\le\left(\frac{15\epsilon_f\kappa\eta\sqrt{r}}{2\|\mathbf{\Theta}^*\|_\infty}+\frac{4\sqrt{2\alpha}\eta^2\kappa r\sigma}{\sqrt{n_{\min}}\|\mathbf{\Theta}^*\|_\infty^2}\right)\|\mathbf{V}^*\mathbf{D}^*\|_{2,\infty}\right]\ge 1-3K\exp(-\alpha).$$

Therefore our argument goes. □

Proof of Lemma C.4. We first prove the bound for $\sigma_{\min}(\widehat{\mathbf{\Theta}})$. By the condition in the claim statement and (F.2), we have

$$\epsilon_f\le\frac{\|\mathbf{\Theta}^*\|_\infty}{2\eta\kappa\sqrt{r}}\le\frac{\sqrt{KN}\sigma_{\min}(\mathbf{\Theta}^*)}{2}.$$

By Lemma F.11, we have

$$\begin{aligned}\sigma_{\min}(\widehat{\mathbf{\Theta}})\ge\sigma_{\min}(\mathbf{\Theta}^*)-\|\mathbf{\Theta}^*-\widehat{\mathbf{\Theta}}\|_F&\ge\sigma_{\min}(\mathbf{\Theta}^*)-\epsilon_f\sqrt{KN}\\ &\ge\sigma_{\min}(\mathbf{\Theta}^*)-\frac{\sigma_{\min}(\mathbf{\Theta}^*)}{2}=\frac{\sigma_{\min}(\mathbf{\Theta}^*)}{2}.\end{aligned}$$

Hence by Lemma F.12 and (F.2), we have

$$\|\widehat{\mathbf{V}}\|_{2,\infty}\le\frac{\sqrt{N}\|\widehat{\mathbf{\Theta}}\|_\infty}{\sigma_{\min}(\widehat{\mathbf{\Theta}})}\le\frac{2\sqrt{N}\|\widehat{\mathbf{\Theta}}\|_\infty}{\sigma_{\min}(\mathbf{\Theta}^*)}\le\frac{2\sqrt{N}\|\widehat{\mathbf{\Theta}}\|_\infty}{\frac{\sqrt{NK}\|\mathbf{\Theta}^*\|_\infty}{\sqrt{r}\eta\kappa}}\le\frac{2\kappa\eta}{\|\mathbf{\Theta}^*\|_\infty}\sqrt{\frac{r}{K}},$$

where the last inequality is given by the constraint in (3.1). □

C.5. Proof of Lemma C.3

Proof of Lemma C.3. First we show that as long as $\epsilon \le \frac{1}{2\sqrt{r}\eta\kappa}$, we have $\mathsf{rank}(\widetilde{\mathbf{U}}) = r$. Then it suffices to show that $\sigma_r(\widetilde{\mathbf{U}}) > 0$. By Weyl's inequality (Lemma F.7) and the definition of ϵ in (C.3), we have $\forall i \in [r]$

$$\begin{aligned}|\sigma_i(\mathbf{D}_1) - \sigma_i(\mathbf{\Theta}^*)| = |\sigma_i(\widetilde{\mathbf{U}}) - \sigma_i(\mathbf{U}^*\mathbf{D}^*)| &\le \|\widetilde{\mathbf{U}} - \mathbf{U}^*\mathbf{D}^*\|_F \\ &\le \sqrt{N}\|\widetilde{\mathbf{U}} - \mathbf{U}^*\mathbf{D}^*\|_{2,\infty} \\ &\le \epsilon\sqrt{NK}\|\mathbf{\Theta}^*\|_\infty.\end{aligned}$$

Then by (F.2), we have

$$\begin{aligned}\sigma_r(\widetilde{\mathbf{U}}) = \sigma_r(\mathbf{D}_1) &\ge \sigma_{\min}(\mathbf{\Theta}^*) - |\sigma_r(\mathbf{D}_1) - \sigma_{\min}(\mathbf{\Theta}^*)| \\ &\ge \frac{\sqrt{NK}\|\mathbf{\Theta}^*\|_\infty}{\sqrt{r}\eta\kappa} - \epsilon\sqrt{NK}\|\mathbf{\Theta}^*\|_\infty \\ &= \left(\frac{1}{\sqrt{r}\eta\kappa} - \epsilon\right)\sqrt{NK}\|\mathbf{\Theta}^*\|_\infty \ge \max\left\{\frac{\epsilon}{2}, \frac{1}{2\sqrt{r}\eta\kappa}\right\}\sqrt{NK}\|\mathbf{\Theta}^*\|_\infty, \qquad \text{(C.15)}\end{aligned}$$

where the last inequality is given by the condition $\epsilon \le \frac{1}{2\sqrt{r}\eta\kappa}$. (C.15) shows that $\mathbf{D}_1$ is a full rank diagonal matrix in $\mathbb{R}^{r\times r}$, and thus $\mathbf{D}_1^{-1}$ is well-defined. Then we can write

$$\begin{aligned}\widetilde{\mathbf{\Theta}} &= \mathbf{U}_1\mathbf{Q}_1\widetilde{\mathbf{V}}^\top \\ &= (\mathbf{U}_1\mathbf{D}_1\mathbf{Q}_1)\mathbf{Q}_1^\top\mathbf{D_1}^{-1}\mathbf{Q}_1\widetilde{\mathbf{V}}^\top \\ &= (\mathbf{U}^*\mathbf{D}^* + \underbrace{\widetilde{\mathbf{U}} - \mathbf{U}^*\mathbf{D}^*}_{=\mathbf{\Delta}_U})(\mathbf{D}^{*-1} + \underbrace{\mathbf{Q}_1^\top\mathbf{D}_1^{-1}\mathbf{Q}_1 - \mathbf{D}^{*-1}}_{=\mathbf{\Delta}_D})(\mathbf{D}^*\mathbf{V}^{*\top} + \underbrace{\widetilde{\mathbf{V}}^\top - \mathbf{D}^*\mathbf{V}^{*\top}}_{=\mathbf{\Delta}_V^\top}).\end{aligned}$$

We first bound the three terms $\mathbf{\Delta}_U$, $\mathbf{\Delta}_D$ and $\mathbf{\Delta}_V$ respectively. Lemma F.13 and (C.15) yield

$$\begin{aligned}\|\mathbf{\Delta}_D\|_{\mathrm{op}} &\le \frac{\max_{i\in[r]}|\sigma_i(\mathbf{D}_1) - \sigma_i(\mathbf{D}^*)|}{\sigma_{\min}(\mathbf{D}_1)\sigma_{\min}(\mathbf{D}^*)} \\ &\le \frac{\epsilon\sqrt{NK}\|\mathbf{\Theta}^*\|_\infty}{\max\left\{\frac{\epsilon}{2}, \frac{1}{2\sqrt{r}\eta\kappa}\right\}\sqrt{NK}\|\mathbf{\Theta}^*\|_\infty\sigma_{\min}(\mathbf{\Theta}^*)} \\ &\le \min\left\{\frac{2}{\sigma_{\min}(\mathbf{\Theta}^*)}, \frac{2\sqrt{r}\eta\kappa\epsilon}{\sigma_{\min}(\mathbf{\Theta}^*)}\right\}. \qquad \text{(C.16)}\end{aligned}$$

Additionally, given our conditions in the lemma statement, we have

$$\|\mathbf{\Delta}_U\|_{2,\infty} = \|\widetilde{\mathbf{U}} - \mathbf{U}^*\mathbf{D}^*\|_{2,\infty} \le \epsilon\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty}$$

and

$$\|\mathbf{\Delta}_V\|_{2,\infty} = \|\widetilde{\mathbf{V}} - \mathbf{V}^*\mathbf{D}^*\|_{2,\infty} \le \epsilon\|\mathbf{V}^*\mathbf{D}^*\|_{2,\infty}. \qquad \text{(C.17)}$$

Now we decompose $\widetilde{\mathbf{\Theta}} - \mathbf{\Theta}^*$ as

$$\begin{aligned}\widetilde{\mathbf{\Theta}} - \mathbf{\Theta}^* &= (\mathbf{U}^*\mathbf{D}^* + \mathbf{\Delta}_U)(\mathbf{D}^{*-1} + \mathbf{\Delta}_D)(\mathbf{D}^*\mathbf{V}^{*\top} + \mathbf{\Delta}_V^\top) - \mathbf{\Theta}^* \\ &= \mathbf{U}^*\mathbf{D}^*(\mathbf{D}^{*-1} + \mathbf{\Delta}_D)(\mathbf{D}^*\mathbf{V}^{*\top} + \mathbf{\Delta}_V^\top) + \mathbf{\Delta}_U(\mathbf{D}^{*-1} + \mathbf{\Delta}_D)(\mathbf{D}^*\mathbf{V}^{*\top} + \mathbf{\Delta}_V^\top) - \mathbf{\Theta}^* \\ &= \mathbf{U}^*\mathbf{D}^*\mathbf{D}^{*-1}\mathbf{D}^*\mathbf{V}^{*\top} - \mathbf{\Theta}^* + \mathbf{U}^*\mathbf{D}^*(\mathbf{D}^{*-1} + \mathbf{\Delta}_D)\mathbf{\Delta}_V^\top + \mathbf{U}^*\mathbf{D}^*\mathbf{\Delta}_D(\mathbf{D}^*\mathbf{V}^{*\top} + \mathbf{\Delta}_V^\top) \\ &\quad + \mathbf{\Delta}_U(\mathbf{D}^{*-1} + \mathbf{\Delta}_D)(\mathbf{D}^*\mathbf{V}^{*\top} + \mathbf{\Delta}_V^\top) \\ &= \underbrace{\mathbf{U}^*\mathbf{D}^*(\mathbf{D}^{*-1} + \mathbf{\Delta}_D)\mathbf{\Delta}_V^\top}_{=\mathbf{\Delta}_1} + \underbrace{\mathbf{U}^*\mathbf{D}^*\mathbf{\Delta}_D(\mathbf{D}^*\mathbf{V}^{*\top} + \mathbf{\Delta}_V^\top)}_{=\mathbf{\Delta}_2} + \underbrace{\mathbf{\Delta}_U(\mathbf{D}^{*-1} + \mathbf{\Delta}_D)(\mathbf{D}^*\mathbf{V}^{*\top} + \mathbf{\Delta}_V^\top)}_{=\mathbf{\Delta}_3}. \qquad \text{(C.18)}\end{aligned}$$

In what follows, we bound the infinity norm of the three terms $\boldsymbol{\Delta}_1$, $\boldsymbol{\Delta}_2$ and $\boldsymbol{\Delta}_3$ respectively.

For the term $\boldsymbol{\Delta}_1$, we have the following inequalities:

$$\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty} = \|\boldsymbol{\Theta}^*\|_{2,\infty} \le \|\boldsymbol{\Theta}^*\|_\infty\sqrt{K}, \quad \text{and} \quad \|\mathbf{V}^*\mathbf{D}^*\|_{2,\infty} = \|\boldsymbol{\Theta}^{*\top}\|_{2,\infty} \le \|\boldsymbol{\Theta}^*\|_\infty\sqrt{N}, \tag{C.19}$$

which are given by (F.1), and

$$\begin{aligned}\|\mathbf{D}^{*-1} + \boldsymbol{\Delta}_D\|_{\mathrm{op}} \le \|\mathbf{D}^{*-1}\|_{\mathrm{op}} + \|\boldsymbol{\Delta}_D\|_{\mathrm{op}} &= \frac{1}{\sigma_{\min}(\mathbf{D}^*)} + \|\boldsymbol{\Delta}_D\|_{\mathrm{op}} \\ &\le \frac{1}{\sigma_{\min}(\mathbf{D}^*)} + \frac{2}{\sigma_{\min}(\mathbf{D}^*)} \\ &= \frac{3}{\sigma_{\min}(\mathbf{D}^*)} \le \frac{3\sqrt{r}\eta\kappa}{\sqrt{NK}\|\boldsymbol{\Theta}^*\|_\infty},\end{aligned} \tag{C.20}$$

where the first inequality is given by (C.16) and the last inequality is given by (F.2). Then by Lemma F.14 we have

$$\begin{aligned}\|\boldsymbol{\Delta}_1\|_\infty &\le \|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty}\|\mathbf{D}^{*-1} + \boldsymbol{\Delta}_D\|_{\mathrm{op}}\|\boldsymbol{\Delta}_V\|_{2,\infty} \\ &\le \|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty}\|\mathbf{D}^{*-1} + \boldsymbol{\Delta}_D\|_{\mathrm{op}}\epsilon\|\mathbf{D}^*\mathbf{V}^{*\top}\|_{2,\infty} \\ &\le \|\boldsymbol{\Theta}^*\|_\infty\sqrt{K}\cdot\frac{3\sqrt{r}\eta\kappa}{\sqrt{NK}\|\boldsymbol{\Theta}^*\|_\infty}\cdot\epsilon\|\boldsymbol{\Theta}^*\|_\infty\sqrt{N} \\ &= 3\epsilon\sqrt{r}\kappa\eta\|\boldsymbol{\Theta}^*\|_\infty.\end{aligned}$$

where the second inequality is given by (C.17), (C.19) and (C.20). Similarly, we have

$$\begin{aligned}\|\boldsymbol{\Delta}_3\|_\infty &= \left\|\boldsymbol{\Delta}_U(\mathbf{D}^{*-1} + \boldsymbol{\Delta}_D)(\mathbf{D}^*\mathbf{V}^{*\top} + \boldsymbol{\Delta}_V^\top)\right\|_\infty \\ &\le \|\boldsymbol{\Delta}_U\|_{2,\infty}\|\mathbf{D}^{*-1} + \boldsymbol{\Delta}_D\|_{\mathrm{op}}\|\mathbf{D}^*\mathbf{V}^{*\top} + \boldsymbol{\Delta}_V^\top\|_{2,\infty} \\ &\le \epsilon\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty}\|\mathbf{D}^{*-1} + \boldsymbol{\Delta}_D\|_{\mathrm{op}}(1+\epsilon)\|\mathbf{D}^*\mathbf{V}^{*\top}\|_{2,\infty} \\ &\le (1+\epsilon)\frac{3\epsilon\sqrt{r}\eta\kappa\|\boldsymbol{\Theta}^*\|_\infty}{2} \\ &\le \left(1 + \frac{1}{2\sqrt{r}\eta\kappa}\right)\frac{3\epsilon\sqrt{r}\eta\kappa\|\boldsymbol{\Theta}^*\|_\infty}{2} \\ &\le \frac{9\epsilon\sqrt{r}\eta\kappa\|\boldsymbol{\Theta}^*\|_\infty}{4},\end{aligned}$$

where the last inequality is given by $\kappa, r, \eta > 1$. For the term $\boldsymbol{\Delta}_2$, we have an analogous argument

$$\begin{aligned}\|\boldsymbol{\Delta}_2\|_\infty &\le \|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty}\|\boldsymbol{\Delta}_D\|_{\mathrm{op}}\|\mathbf{D}^*\mathbf{V}^{*\top} + \boldsymbol{\Delta}_V^\top\|_{2,\infty} \\ &\le \|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty}\|\boldsymbol{\Delta}_D\|_{\mathrm{op}}(1+\epsilon)\|\mathbf{D}^*\mathbf{V}^{*\top}\|_{2,\infty} \\ &\le (1+\epsilon)\|\boldsymbol{\Theta}^*\|_\infty\sqrt{K}\cdot\min\left\{\frac{2}{\sigma_{\min}(\boldsymbol{\Theta}^*)}, \frac{2\sqrt{r}\eta\kappa\epsilon}{\sigma_{\min}(\boldsymbol{\Theta}^*)}\right\}\cdot\|\boldsymbol{\Theta}^*\|_\infty\sqrt{N} \\ &\le (1+\epsilon)\|\boldsymbol{\Theta}^*\|_\infty\sqrt{K}\cdot\frac{2\sqrt{r}\eta\kappa}{\sqrt{NK}\|\boldsymbol{\Theta}^*\|_\infty}\cdot\|\boldsymbol{\Theta}^*\|_\infty\sqrt{N} \\ &\le 3\epsilon\sqrt{r}\eta\kappa\|\boldsymbol{\Theta}^*\|_\infty\end{aligned}$$

where the third inequality is given by (C.16). Thus we conclude

$$\begin{aligned}\|\widetilde{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_\infty &\le \|\boldsymbol{\Delta}_1\|_\infty + \|\boldsymbol{\Delta}_2\|_\infty + \|\boldsymbol{\Delta}_3\|_\infty \\ &\le \frac{33\epsilon\sqrt{r}\eta\kappa\|\boldsymbol{\Theta}^*\|_\infty}{4}. \quad \square\end{aligned}$$

Appendix D: Proofs of Theorem 4

D.1. Major Steps of the Proof

To prove Theorem 4, we will decompose the regret of Algorithm 2 into the regret from the exploration phase and the regret from the exploitation phase. The regret in the exploration phase can be easily upper bounded by NE_h since the per-period regret is less than N for any period $t \in [T]$, regardless of which $\mathbf{X}_t$ we choose.

To show the regret in the exploitation phase, we first prove Proposition 2, which says that the per-period regret for the exploitation phase $\langle \mathbf{X}^*, \mathbf{\Theta}^* \rangle - \langle \mathbf{X}_c, \mathbf{\Theta}^* \rangle = \widetilde{O}(r^{3/2} K E_h^{-1/2})$, with probability higher than $1 - 4(NT)^{-1}$. We provide the proof in Appendix D.3.

Proposition 2. *Given* $\lambda = c_\lambda \sigma (\log(NT) + \log(N+K))/\sqrt{E_h}$ *in* (3.1) *for a constant* c_λ*,* $\mathbf{X}_c$ *obtained in Algorithm 2 after* E_h *exploration steps satisfies*

$$\langle \mathbf{X}^*, \mathbf{\Theta}^* \rangle - \langle \mathbf{X}_c, \mathbf{\Theta}^* \rangle \le 2\sqrt{2r}\,\delta_F(E_h)$$

with probability higher than $1 - 4(NT)^{-1}$*, where*

$$\delta_F(E_h) = c_4 \max\left\{ c_\lambda \sigma K \left(\alpha + \log(N+K)\right) \sqrt{\frac{r}{E_h}}, rK\sqrt{\frac{\log[(N+K)E_h]}{NE_h}} \right\}$$

for the constant c_4 *from* (A.4)*.*

Equipped with Proposition 2, we can show the following lemma that combines the regret from the two phases, which is sufficient for proving Theorem 4. The proof for this lemma is in Appendix D.4.

Lemma D.1. *Set*

$$\lambda = c_\lambda \sigma \frac{\log(NT) + \log(N+K)}{\sqrt{E_h}}$$

in (3.1) *where* c_λ *is a constant. Then the regret of Algorithm 2 satisfies*

$$\mathbb{E}[R(T)] \le NE_h + 2\sqrt{2r}(T - E_h)\delta_F(E_h) + 4.$$

D.2. Proof of Theorem 4

Proof of Theorem 4. By Lemma D.1, choosing $E_h = c_h r T^{2/3}$ where c_h is a suitable constant yields

$$\begin{aligned}
\mathbb{E}[R(T)] &\le NE_h + 2\sqrt{2r}(T - E_h)\delta_F(E_h) + 4 \\
&= c_h r N T^{2/3} + 2\sqrt{2r}(T - c_h r T^{2/3})\delta_F(c_h r T^{2/3}) + 4 \\
&\le c_h r N T^{2/3} + 2\sqrt{2r}T\delta_F(c_h r T^{2/3}) + 4 \\
&= \widetilde{\mathcal{O}}\left(NrT^{3/2} + r^{1/2} T r K (rT^{2/3})^{-1/2}\right) \\
&= \widetilde{\mathcal{O}}\left(r(N+K)T^{3/2}\right),
\end{aligned}$$

which completes our proof. □

D.3. Proof of Proposition 2

Proof of Proposition 2. We first invoke the explicit form of Theorem 1 in (A.4) with $n = E_h$ to get

$$\mathbb{P}[\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\|_F \geq \delta_F(E_h)] \leq 4(NT)^{-1},$$

by our choice of λ. Thus

$$\begin{aligned}
\langle \mathbf{X}^*, \boldsymbol{\Theta}^* \rangle - \langle \mathbf{X}_c, \boldsymbol{\Theta}^* \rangle &= \langle \mathbf{X}^*, \boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}} \rangle + \langle \mathbf{X}_c, \boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}} \rangle \\
&\leq \left| \langle \mathbf{X}^*, \boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}} \rangle \right| + \left| \langle \mathbf{X}_c, \boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}} \rangle \right| \\
&\leq \|\mathbf{X}^*\|_{\text{op}} \|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\|_* + \|\mathbf{X}_c\|_{\text{op}} \|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\|_* \\
&= 2\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\|_* \\
&\leq 2\sqrt{2r}\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}\|_F \leq 2\sqrt{2r}\delta_F(E_h)
\end{aligned}$$

with probability higher than $1 - 4(NT)^{-1}$ where the second equality is given by the matching structure of $\mathbf{X}^*$ and $\mathbf{X}_c$, and the last inequality is given by $\mathsf{rank}(\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}) \leq \mathsf{rank}(\boldsymbol{\Theta}^*) + \mathsf{rank}(\widehat{\boldsymbol{\Theta}}) \leq 2r$. □

D.4. Proof of Lemma D.1

Proof of Lemma D.1. For the first E_h rounds, we have

$$\sum_{t=1}^{E_h} \langle \mathbf{X}^*, \boldsymbol{\Theta}^* \rangle - \langle \mathbf{X}_t, \boldsymbol{\Theta}^* \rangle \leq \sum_{t=1}^{E_h} \langle \mathbf{X}^*, \boldsymbol{\Theta}^* \rangle \leq \sum_{t=1}^{E_h} N\|\boldsymbol{\Theta}^*\|_\infty \leq NE_h.$$

By Proposition 2, we have $\langle \mathbf{X}^*, \boldsymbol{\Theta}^* \rangle - \langle \mathbf{X}_c, \boldsymbol{\Theta}^* \rangle \leq 2\sqrt{2r}\delta_F(E_h)$ with probability higher than $1 - 4(NT)^{-1}$. This implies that

$$\sum_{t=E_h}^{T} \langle \mathbf{X}^*, \boldsymbol{\Theta}^* \rangle - \langle \mathbf{X}_t, \boldsymbol{\Theta}^* \rangle = \sum_{t=E_h}^{T} \langle \mathbf{X}^*, \boldsymbol{\Theta}^* \rangle - \langle \mathbf{X}_c, \boldsymbol{\Theta}^* \rangle \leq (T - E_h) 2\sqrt{2r}\delta_F(E_h)$$

with probability higher than $1 - 4(NT)^{-1}$. Therefore, the total regret

$$R(T) = \sum_{t=1}^{T} \langle \mathbf{X}^*, \boldsymbol{\Theta}^* \rangle - \langle \mathbf{X}_t, \boldsymbol{\Theta}^* \rangle \leq NE_h + (T - E_h) 2\sqrt{2r}\delta_F(E_h)$$

with probability higher than $1 - 4(NT)^{-1}$. Now notice that for any $t \in [T]$ and any $\mathbf{X}_t \in \mathcal{M}$, we have

$$\langle \mathbf{X}^*, \boldsymbol{\Theta}^* \rangle - \langle \mathbf{X}_t, \boldsymbol{\Theta}^* \rangle \leq N,$$

which implies that

$$R(T) = \sum_{t=1}^{T} \left(\langle \mathbf{X}^*, \boldsymbol{\Theta}^* \rangle - \langle \mathbf{X}_t, \boldsymbol{\Theta}^* \rangle \right) \leq NT$$

almost surely. Then we can decompose the regret expectation

$$\begin{aligned}
\mathbb{E}\left[R(T)\right] &= \mathbb{E}\left[\sum_{t=1}^{T} \left(\langle \mathbf{X}^*, \boldsymbol{\Theta}^* \rangle - \langle \mathbf{X}_t, \boldsymbol{\Theta}^* \rangle \right)\right] \\
&= \mathbb{E}\left[R(T) \cdot \mathbb{I}\left\{R(T) \leq NE_h + (T - E_h) 2\sqrt{2r}\delta_F(E_h)\right\}\right] \\
&\quad + \mathbb{E}\left[R(T) \cdot \mathbb{I}\left\{R(T) > NE_h + (T - E_h) 2\sqrt{2r}\delta_F(E_h)\right\}\right] \\
&\leq NE_h + (T - E_h) 2\sqrt{2r}\delta_F(E_h) + NT \cdot \mathbb{P}\left[R(T) > NE_h + (T - E_h) 2\sqrt{2r}\delta_F(E_h)\right] \\
&\leq NE_h + (T - E_h) 2\sqrt{2r}\delta_F(E_h) + NT \cdot 4(NT)^{-1} \\
&= NE_h + (T - E_h) 2\sqrt{2r}\delta_F(E_h) + 4,
\end{aligned}$$

which completes our proof. □

Appendix E: Proof of Theorem 5

E.1. Major Steps of the Proof

Following the literature (Gale and Shapley 1962), we assume there is no tie in preferences without loss of generality. A key ingredient for the proof of Theorem 5 is that a stable matching is determined solely based on the preference rankings of both workers and jobs; in other words, accurate reward estimation is not critical as long as the rankings derived from it are precise. When the minimum reward gap between pairs is sufficiently large, accurate rankings and hence a worker-optimal stable matching can still be achieved even under imprecise reward estimations. Liu et al. (2020) provide a similar observation in their Lemma 3. To this end, we let

$$\Delta_{\min} = \min_{i \in [N]} \left\{ \min_{j \neq j'} |\mathbf{\Theta}^{*(i,j)} - \mathbf{\Theta}^{*(i,j')}| \right\}$$

denote the minimum gap of rewards over all workers $i \in [N]$. Then, provided that the infinity-norm estimation error $\|\widetilde{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_\infty \leq \Delta_{\min}/2$ where $\widetilde{\mathbf{\Theta}}$ is obtained from Algorithm 3, we can ensure that the stable matching we commit to, i.e., $\mathbf{X}_c$, coincides with the worker-optimal stable matching $\mathbf{X}^*$. This main idea gives rise to the bound on the stable regret for every worker (as previously defined in (5.2)) in Theorem 5.

We introduce the following lemma before we prove Theorem 5. Its proof is provided in Appendix E.3.

Lemma E.1. *Given our choice of $\lambda = c_\lambda \sigma \log\big((N+K)(3N+3K+5)T\big)/\sqrt{E_h}$, the stable regret of Algorithm 3 for worker i satisfies*

$$\mathbb{E}[R_i(T)] \leq 2E_h + 2(T - E_h) \cdot \mathbb{I}\{\delta_\infty(E_h) > \Delta_{\min}/2\} + 2, \quad \forall i \in [N], \tag{E.1}$$

where

$$\delta_\infty(E_h) = c_6 \eta^2 \kappa^2 r^{3/2} \max\left\{ c_\lambda \sigma \log\big[(N+K)(3N+3K+5)T\big] \sqrt{\frac{K}{NE_h}}, \sqrt{\frac{rK\log[(N+K)E_h]}{NE_h}}, \frac{\eta\sigma\sqrt{\log\big[(3N+3K+5)T\big]}}{\|\mathbf{\Theta}^*\|_\infty} \sqrt{\frac{K}{NE_h}} \right\}$$

for the constant c_6 from (C.7).

E.2. Proof of Theorem 5

Proof of Theorem 5. Following Lemma E.1, choose

$$E_h = c_7 \frac{r^3 K}{N\Delta_{\min}^2} \max\left\{ c_\lambda^2 \sigma^2 \log^2\big[(N+K)(3N+3K+5)T\big], r\log\big[(N+K)T\big], \frac{\eta^2\sigma^2 \log\big[(3N+3K+5)T\big]}{\|\mathbf{\Theta}^*\|_\infty^2} \right\}$$

where $c_7 \geq 4c_6^2\eta^4\kappa^4$ is a constant and and suppose that T is large enough such that $E_h \leq T$. Then we have

$$\begin{aligned} E_h &= c_7 \frac{r^3 K}{N\Delta_{\min}^2} \max\left\{ c_\lambda^2 \sigma^2 \log^2\big[(N+K)(3N+3K+5)T\big], r\log\big[(N+K)T\big], \frac{\eta^2\sigma^2 \log\big[(3N+3K+5)T\big]}{\|\mathbf{\Theta}^*\|_\infty^2} \right\} \\ &\geq \frac{4c_6^2\eta^4\kappa^4 r^3 K}{N\Delta_{\min}^2} \max\left\{ c_\lambda^2 \sigma^2 \log^2\big[(N+K)(3N+3K+5)T\big], r\log\big[(N+K)T\big], \frac{\eta^2\sigma^2 \log\big[(3N+3K+5)T\big]}{\|\mathbf{\Theta}^*\|_\infty^2} \right\} \\ &\geq \frac{4c_6^2\eta^4\kappa^4 r^3 K}{N\Delta_{\min}^2} \max\left\{ c_\lambda^2 \sigma^2 \log^2\big[(N+K)(3N+3K+5)T\big], r\log\big[(N+K)E_h\big], \frac{\eta^2\sigma^2 \log\big[(3N+3K+5)T\big]}{\|\mathbf{\Theta}^*\|_\infty^2} \right\}, \end{aligned}$$

which implies

$$\delta_\infty(E_h) \le \frac{\Delta_{\min}}{2}.$$

By Lemma E.1, we have

$$\begin{aligned}\mathbb{E}[R_i(T)] &\le 2E_h + 2(T-E_h)\cdot \mathbb{I}\{\delta_\infty(E_h) > \Delta_{\min}/2\} + 1\\ &= \mathcal{O}\left(\frac{r^3 K \max\{\log^2[(N+K)T], r\log[(N+K)T]\}}{N\Delta_{\min}^2}\right),\end{aligned}$$

which completes our proof. □

E.3. Proof of Lemma E.1

Proof of Lemma E.1. First notice that, as long as the estimation error $\|\boldsymbol{\Theta}^* - \widetilde{\boldsymbol{\Theta}}\|_\infty \le \Delta_{\min}/2$, we will have $\mathbf{X}_c = \mathbf{X}^*$. This is because, as aforementioned, the Gale-Shapley algorithm only considers the preference rankings from both sides when deriving the worker-optimal stable matching. In other words, as long as the preference rankings derived from $\widetilde{\boldsymbol{\Theta}}$ are the same as those derived from $\boldsymbol{\Theta}^*$, the result returned by Gale-Shapley algorithm will be the same. The condition $\|\boldsymbol{\Theta}^* - \widetilde{\boldsymbol{\Theta}}\|_\infty \le \Delta_{\min}/2$ will ensure that each worker's preferences will be the same under $\boldsymbol{\Theta}^*$ and $\widetilde{\boldsymbol{\Theta}}$. Now with this claim, we can bound the regret in the rounds from E_h+1 to T. For any $i \in [N]$, we have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=E_h+1}^{T}\Big(\langle \mathbf{X}^{*i}, \boldsymbol{\Theta}^*\rangle - \langle \mathbf{X}_t^i, \boldsymbol{\Theta}^*\rangle\Big)\right]\\
&= \mathbb{E}\left[\sum_{t=E_h+1}^{T}\Big(\langle \mathbf{X}_c^{i}, \boldsymbol{\Theta}^*\rangle - \langle \mathbf{X}_t^i, \boldsymbol{\Theta}^*\rangle\Big)\right]\\
&= \mathbb{E}\left[\sum_{t=E_h+1}^{T}\Big(\langle \mathbf{X}_c^{i}, \boldsymbol{\Theta}^*\rangle - \langle \mathbf{X}_t^i, \boldsymbol{\Theta}^*\rangle\Big)\cdot \mathbb{I}\{\mathbf{X}_c \ne \mathbf{X}^*\}\right]\\
&= \mathbb{E}\left[\sum_{t=E_h+1}^{T}\Big(\langle \mathbf{X}_c^{i}, \boldsymbol{\Theta}^*\rangle - \langle \mathbf{X}_t^i, \boldsymbol{\Theta}^*\rangle\Big)\cdot \mathbb{I}\{\mathbf{X}_c \ne \mathbf{X}^*\}\cdot \mathbb{I}\{\|\widetilde{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_\infty \le \delta_\infty(E_h)\}\cdot \mathbb{I}\{\delta_\infty(E_h) \le \Delta_{\min}/2\}\right]\\
&\quad + \mathbb{E}\left[\sum_{t=E_h+1}^{T}\Big(\langle \mathbf{X}_c^{i}, \boldsymbol{\Theta}^*\rangle - \langle \mathbf{X}_t^i, \boldsymbol{\Theta}^*\rangle\Big)\cdot \mathbb{I}\{\mathbf{X}_c \ne \mathbf{X}^*\}\cdot \mathbb{I}\{\|\widetilde{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_\infty \le \delta_\infty(E_h)\}\cdot \mathbb{I}\{\delta_\infty(E_h) > \Delta_{\min}/2\}\right]\\
&\quad + \mathbb{E}\left[\sum_{t=E_h+1}^{T}\Big(\langle \mathbf{X}_c^{i}, \boldsymbol{\Theta}^*\rangle - \langle \mathbf{X}_t^i, \boldsymbol{\Theta}^*\rangle\Big)\cdot \mathbb{I}\{\mathbf{X}_c \ne \mathbf{X}^*\}\cdot \mathbb{I}\{\|\widetilde{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_\infty > \delta_\infty(E_h)\}\right]\\
&= \mathbb{E}\left[\sum_{t=E_h+1}^{T}\Big(\langle \mathbf{X}_c^{i}, \boldsymbol{\Theta}^*\rangle - \langle \mathbf{X}_t^i, \boldsymbol{\Theta}^*\rangle\Big)\cdot \mathbb{I}\{\mathbf{X}_c \ne \mathbf{X}^*\}\cdot \mathbb{I}\{\|\widetilde{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_\infty \le \delta_\infty(E_h)\}\cdot \mathbb{I}\{\delta_\infty(E_h) > \Delta_{\min}/2\}\right]\\
&\quad + \mathbb{E}\left[\sum_{t=E_h+1}^{T}\Big(\langle \mathbf{X}_c^{i}, \boldsymbol{\Theta}^*\rangle - \langle \mathbf{X}_t^i, \boldsymbol{\Theta}^*\rangle\Big)\cdot \mathbb{I}\{\mathbf{X}_c \ne \mathbf{X}^*\}\cdot \mathbb{I}\{\|\widetilde{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_\infty > \delta_\infty(E_h)\}\right]\\
&\le 2(T-E_h)\cdot \mathbb{I}\{\delta_\infty(E_h) > \Delta_{\min}/2\} + 2(T-E_h)\mathbb{P}\left[\|\widetilde{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_\infty > \delta_\infty(E_h)\right]\\
&\le 2(T-E_h)\cdot \mathbb{I}\{\delta_\infty(E_h) > \Delta_{\min}/2\} + 2(T-E_h)T^{-1} \le 2(T-E_h)\cdot \mathbb{I}\{\delta_\infty(E_h) > \Delta_{\min}/2\} + 2
\end{aligned}$$

where the first inequality is given by $\langle \mathbf{X}_c^i, \boldsymbol{\Theta}^*\rangle - \langle \mathbf{X}_t^i, \boldsymbol{\Theta}^*\rangle \le |\langle \mathbf{X}_c^i, \boldsymbol{\Theta}^*\rangle| + |\langle \mathbf{X}_t^i, \boldsymbol{\Theta}^*\rangle| \le 2$ and the second inequality comes from $\|\widetilde{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_\infty \le \delta_\infty(E_h)$ with probability higher than $1-(T)^{-1}$ given Theorem 3. Then we have

$$\mathbb{E}\left[R_i(T)\right] = \mathbb{E}\left[\sum_{t=1}^{E_h}\Big(\langle \mathbf{X}^{*i}, \boldsymbol{\Theta}^*\rangle - \langle \mathbf{X}_t^i, \boldsymbol{\Theta}^*\rangle\Big)\right] + \mathbb{E}\left[\sum_{t=E_h+1}^{T}\big(\langle \mathbf{X}^{*i}, \boldsymbol{\Theta}^*\rangle - \langle \boldsymbol{\Theta}\mathbf{X}_t^i, \boldsymbol{\Theta}^*\rangle\big)\right]$$

$$
\begin{aligned}
&\leq \mathbb{E}\left[\sum_{t=1}^{E_h}\Big(|\langle \mathbf{X}^{*i}, \mathbf{\Theta}^*\rangle| + |\langle \mathbf{X}_t^i, \mathbf{\Theta}^*\rangle|\Big)\right] + \mathbb{E}\left[\sum_{t=E_h+1}^{T}\big(\langle \mathbf{X}^{*i}, \mathbf{\Theta}^*\rangle - \langle \mathbf{X}_t^i, \mathbf{\Theta}^*\rangle\big)\right] \\
&\leq 2E_h + 2(T-E_h)\cdot \mathbb{I}\{\delta_\infty(E_h) > \Delta_{\min}/2\} + 2,
\end{aligned}
$$

which completes the proof. □

Appendix F: Technical Lemmas

LEMMA F.1. *Let $\mathbf{Z}_1, \cdots, \mathbf{Z}_n$ be i.i.d. matrices in $\mathbb{R}^{N\times K}$ with $\mathbb{E}[\mathbf{Z}_i] = \mathbf{0}$ and $\|\mathbf{Z}_i\|_{\text{op}} \leq \omega$ almost surely for all $i \in [n]$; let σ_Z be a parameter such that*

$$
\sigma_Z^2 \geq \max\left\{\left\|\sum_{i=1}^n \mathbb{E}\left[\mathbf{Z}_i^\top \mathbf{Z}_i\right]\right\|_{\text{op}}, \left\|\sum_{i=1}^n \mathbb{E}\left[\mathbf{Z}_i \mathbf{Z}_i^\top\right]\right\|_{\text{op}}\right\}.
$$

Then for any $\rho > 0$,

$$
\mathbb{P}\left[\left\|\sum_{i=1}^n \mathbf{Z}_i\right\|_{\text{op}} \geq \rho\right] \leq (N+K)\exp\left[\frac{-\rho^2}{2\sigma_Z^2 + (2\omega\rho)/3}\right].
$$

Proof of Lemma F.1. See Proposition 1 in Athey et al. (2021). □

LEMMA F.2. *For any $\zeta > 0$, let*

$$
\mathcal{C}(\zeta) = \{\mathbf{A} \in \mathbb{R}^{N\times K} \mid \|\mathbf{A}\|_\infty \leq 1, \mathbb{E}_\Pi[\|\mathbf{A}\circ\mathbf{X}\|_F^2] \leq \zeta\},
$$

where $\mathbf{X}$ is a random matrix following distribution Π (defined in Section 2). Let $\mathbf{X}_k, k \in [n]$ be i.i.d. copies of $\mathbf{X}$ and $z_\zeta = \sup_{\mathbf{A}\in\mathcal{C}(\zeta)} \left|\frac{1}{n}\sum_{k=1}^n \|\mathbf{A}\circ\mathbf{X}_k\|_F^2 - \mathbb{E}_\Pi[\|\mathbf{A}\circ\mathbf{X}\|_F^2]\right|$. Then we have

$$
\mathbb{P}\left[z_\zeta \geq 2\mathbb{E}[z_\zeta] + \frac{7\zeta}{24}\right] \leq \exp\left(\frac{-n\zeta}{288N}\right).
$$

Proof of Lemma F.2. Our result relies on the Massart's concentration inequality (see Theorem 3 in Massart (2000)). First note that

$$
\|\mathbf{A}\circ\mathbf{X}\|_F^2 \leq N, \forall \mathbf{A} \in \mathcal{C}(\zeta).
$$

Then we need to provide an upper bound for

$$
\sigma_\zeta^2 = \sup_{\mathbf{A}\in\mathcal{C}(\zeta)} \sum_{k=1}^n \text{Var}(\|\mathbf{A}\circ\mathbf{X}\|_F^2).
$$

We have

$$
\begin{aligned}
\sigma_\zeta^2 &= \sup_{\mathbf{A}\in\mathcal{C}(\zeta)} \sum_{k=1}^n \text{Var}(\|\mathbf{A}\circ\mathbf{X}\|_F^2) \\
&\leq \sup_{\mathbf{A}\in\mathcal{C}(\zeta)} \sum_{k=1}^n \mathbb{E}[\|\mathbf{A}\circ\mathbf{X}\|_F^4] \\
&\leq \sup_{\mathbf{A}\in\mathcal{C}(\zeta)} \sum_{k=1}^n \mathbb{E}[\|\mathbf{A}\circ\mathbf{X}\|_F^2]\mathbb{E}[\|\mathbf{A}\circ\mathbf{X}\|_F^2] \\
&\leq nN\zeta,
\end{aligned}
$$

where the second inequality is given by the definition of variance and the third inequality is given by the definition of $\mathcal{C}(\zeta)$. Thus we choose $\varepsilon' = 1, \sigma_\zeta = \sqrt{nN\zeta}, b = N$ and $x = n\zeta/(288N)$ in (11) of Massart (2000), and we obtain

$$
\mathbb{P}\left[z_\zeta \geq 2\mathbb{E}[z_\zeta] + \frac{7\zeta}{24}\right] \leq \exp\left(\frac{-n\zeta}{288N}\right). \quad \square
$$

LEMMA F.3. *For any two matrices* $\mathbf{A}, \mathbf{B}$ *we have*

$$\mathit{rank}(\mathbf{A} \circ \mathbf{B}) \leq \mathit{rank}(\mathbf{A}) \cdot \mathit{rank}(\mathbf{B})$$

where $\mathbf{A} \circ \mathbf{B}$ *denotes the Hadamard product of matrix* $\mathbf{A}$ *and* $\mathbf{B}$.

Proof of Lemma F.3. See Theorem 4.5 of Million (2007). □

LEMMA F.4. *Let* $\mathbf{Z}_1, \cdots, \mathbf{Z}_n$ *be fixed matrices in* $\mathbb{R}^{N \times K}$ *and* $\xi_1, \xi_2, \cdots, \xi_n$ *be independent Rademacher random variables; let* σ_Z *be a parameter such that*

$$\sigma_Z^2 \geq \max\left\{ \left\| \sum_{i=1}^n \mathbf{Z}_i^\top \mathbf{Z}_i \right\|_{\text{op}}, \left\| \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^\top \right\|_{\text{op}} \right\}.$$

Then for any $\rho > 0$,

$$\mathbb{P}\left[\left\| \sum_{i=1}^n \xi_i \mathbf{Z}_i \right\|_{\text{op}} \geq \rho \right] \leq (N+K) \exp\left[\frac{-\rho^2}{2\sigma_Z^2} \right].$$

Proof of Lemma F.4. See Theorem 4.1.1 of Tropp et al. (2015). □

LEMMA F.5. *Consider* n *fixed vectors* $X_1, X_2, \cdots, X_n \in \mathbb{R}^d$ *where* $\|X_i\| \leq S, \forall i \in [n]$. *Then for* n *i.i.d.* σ*-subgaussian random variables* ε_i, *we have*

$$\mathbb{P}\left[\left\| \sum_{i=1}^n \varepsilon_i X_i \right\| \geq \rho \right] \leq 2 \exp\left(\frac{-\rho^2}{2ndS^2\sigma^2} \right), \forall \rho > 0.$$

Proof of Lemma F.5. It is straightforward to see that $\varepsilon_i X_i$ is a σS-subgaussian random vector for each $i \in [n]$. Thus $\sum_i \varepsilon_i X_i$ is a $\sqrt{n}\sigma S$-subgaussian random vector. By Lemma F.15, we immediately have

$$\mathbb{P}\left[\left\| \sum_{i=1}^n \varepsilon_i X_i \right\| \geq \rho \right] \leq 2 \exp\left(\frac{-\rho^2}{2ndS^2\sigma^2} \right). \quad \square$$

LEMMA F.6 (**Matrix Chernoff**). *Consider independent positive-semidefinite matrices* $\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_k \in \mathbb{R}^{d \times d}$ *that satisfy*

$$\|\mathbf{Z}_i\|_{\text{op}} \leq \omega, \forall i \in [k] \quad \textit{almost surely.}$$

Then we have

$$\mathbb{P}\left[\lambda_{\min}\left(\sum_{i=1}^k \mathbf{Z}_i \right) \leq \rho \mu_{\min} \right] \leq d \cdot \mathrm{e}^{-(1-\rho)^2 \mu_{\min}/2\omega}, \forall \rho \in [0,1]$$

and

$$\mathbb{P}\left[\lambda_{\max}\left(\sum_{i=1}^k \mathbf{Z}_i \right) \geq \rho \mu_{\max} \right] \leq d \cdot \left[\frac{\mathrm{e}}{\rho} \right]^{\rho \mu_{\max}/\omega}, \forall \rho \geq \mathrm{e},$$

where

$$\mu_{\min} = \lambda_{\min}\left(\sum_{i=1}^k \mathbb{E}\mathbf{Z}_i \right) \quad \textit{and} \quad \mu_{\max} = \lambda_{\max}\left(\sum_{i=1}^k \mathbb{E}\mathbf{Z}_i \right).$$

Proof of Lemma F.6. See Remark 5.3 in Tropp (2012). □

LEMMA F.7 (**Weyl's Inequality**). *For any matrix* $\mathbf{A}, \mathbf{A}' \in \mathbb{R}^{N \times K}$, *it holds for all* $i \in [\min\{N, K\}]$ *that*

$$|\sigma_i(\mathbf{A}) - \sigma_i(\mathbf{A}')| \leq \|\mathbf{A} - \mathbf{A}'\|_{\text{op}}.$$

Proof of Lemma F.7. See Weyl (1912). □

LEMMA F.8. *For any $\boldsymbol{\Theta}^* \in \mathbb{R}^{N\times K}$ with SVD $\boldsymbol{\Theta}^* = \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\top}$, we have*

$$\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty} = \|\boldsymbol{\Theta}^*\|_{2,\infty}. \tag{F.1}$$

Further, if $\boldsymbol{\Theta}^*$ satisfies Assumption 1, we have

$$\sigma_{\min}(\boldsymbol{\Theta}^*) = \frac{\|\boldsymbol{\Theta}^*\|_{\rm op}}{\kappa} \geq \frac{\|\boldsymbol{\Theta}^*\|_F}{\sqrt{r}\kappa} \geq \frac{\sqrt{NK}\|\boldsymbol{\Theta}^*\|_\infty}{\sqrt{r}\eta\kappa}, \tag{F.2}$$

where κ is the conditional number of $\boldsymbol{\Theta}^*$.

Proof of Lemma F.8. For (F.1), notice that

$$\|\mathbf{U}^*\mathbf{D}^*\|_{2,\infty} = \max_{i\in[N]} \|e_i(N)^\top \mathbf{U}^*\mathbf{D}^*\| = \max_{i\in[N]} \|e_i(N)^\top \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\top}\| = \max_{i\in[N]} \|e_i(N)^\top \boldsymbol{\Theta}^*\| = \|\boldsymbol{\Theta}^*\|_{2,\infty},$$

where the second equation is given by that $\mathbf{V}^*$ is orthogonal and multiplying a vector by an orthogonal matrix does not change its length.

For (F.2), the first equation and the first inequality are straightfoward. The last inequality directly follows Assumption 1, i.e.,

$$\|\boldsymbol{\Theta}^*\|_F \geq \frac{\sqrt{NK}\|\boldsymbol{\Theta}^*\|_\infty}{\eta}. \quad \square$$

LEMMA F.9. *Let $\mathbf{V} \in \mathbb{R}^{K\times r}$ be an orthogonal matrix (i.e., $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_{r\times r}$). We have for any matrix $\mathbf{A} \in \mathbb{R}^{r\times K}$, $\|\mathbf{V}\mathbf{A}\|_{\rm op} \leq \|\mathbf{A}\|_{\rm op} = \|\mathbf{A}\mathbf{V}\|_{\rm op}$.*

Proof of Lemma F.9. By the property of operator norm, we have

$$\|\mathbf{V}\mathbf{A}\|_{\rm op} \leq \|\mathbf{V}\|_{\rm op}\|\mathbf{A}\|_{\rm op} = \|\mathbf{A}\|_{\rm op}.$$

Let the SVD of $\mathbf{A}$ be $\mathbf{A} = \mathbf{U}_A\mathbf{D}\mathbf{V}_A$, where $\mathbf{V}_A$ is an $r \times r$ orthogonal matrix. Thus we have $\mathbf{A}\mathbf{V} = \mathbf{U}_A\mathbf{D}(\mathbf{V}_A\mathbf{V})$, which is a SVD since both $\mathbf{U}_A$ and $\mathbf{V}_A\mathbf{V}$ are orthogonal matrices. Hence the singular values of $\mathbf{A}$ and $\mathbf{A}\mathbf{V}$ are the same, which yields $\|\mathbf{A}\|_{\rm op} = \|\mathbf{A}\mathbf{V}\|_{\rm op}$. □

LEMMA F.10. *For any rank r matrices $\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}} \in \mathbb{R}^{N\times K}$, let $\widehat{\boldsymbol{\Theta}} = \widehat{\mathbf{U}}\widehat{\mathbf{D}}\widehat{\mathbf{V}}^\top$ be an arbitrary SVD of $\widehat{\boldsymbol{\Theta}}$, then there exists SVD $\boldsymbol{\Theta} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$, where*

$$\|\mathbf{U} - \widehat{\mathbf{U}}\|_F \leq 5\left(\frac{1}{\sigma_r(\boldsymbol{\Theta})} + \frac{1}{\sigma_r(\widehat{\boldsymbol{\Theta}})}\right)\|\boldsymbol{\Theta} - \widehat{\boldsymbol{\Theta}}\|_F$$

and

$$\|\mathbf{V} - \widehat{\mathbf{V}}\|_F \leq 5\left(\frac{1}{\sigma_r(\boldsymbol{\Theta})} + \frac{1}{\sigma_r(\widehat{\boldsymbol{\Theta}})}\right)\|\boldsymbol{\Theta} - \widehat{\boldsymbol{\Theta}}\|_F.$$

Proof of Lemma F.10. In this proof, let $\boldsymbol{\Theta} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ be an arbitrary SVD of $\boldsymbol{\Theta}$. Moreover, let

$$\mathbf{F}' = \begin{bmatrix}\mathbf{U}' \\ \mathbf{V}'\end{bmatrix} \quad \text{and} \quad \mathbf{F} = \begin{bmatrix}\mathbf{U} \\ \mathbf{V}\end{bmatrix}.$$

Then it remains to bound

$$d(\mathbf{F}, \mathbf{F}') = \min_{\mathbf{Q}\in\mathcal{O}_r} \|\mathbf{F} - \mathbf{F}'\mathbf{Q}\|_F = \min_{\mathbf{Q}\in\mathcal{O}_r} \|\mathbf{F}\mathbf{Q}^\top - \mathbf{F}'\|_F.$$

By Remark 6.1 in Keshavan et al. (2010), we have

$$d(\mathbf{F},\mathbf{F}') \le \sqrt{2}\|\mathbf{F}'\mathbf{F}'^\top - \mathbf{F}\mathbf{F}^\top\|_F = \sqrt{2}(\|\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top - \mathbf{U}\mathbf{U}^\top\|_F + \|\widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top\|_F + 2\|\widehat{\mathbf{U}}\widehat{\mathbf{V}}^\top - \mathbf{U}\mathbf{V}^\top\|_F).$$

By Lemma F.16, we have that

$$\|\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top - \mathbf{U}\mathbf{U}^\top\|_F \le \frac{\sqrt{2}\|\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\|_F}{\sigma_r(\mathbf{\Theta})} \quad \text{and} \quad \|\widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top\|_F \le \frac{\sqrt{2}\|\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\|_F}{\sigma_r(\widehat{\mathbf{\Theta}})}.$$

Furthermore, notice that $\widehat{\mathbf{U}}\widehat{\mathbf{V}}^\top = \operatorname{sgn}(\widehat{\mathbf{\Theta}})$ and $\mathbf{U}\mathbf{V}^\top = \operatorname{sgn}(\mathbf{\Theta})$, which are defined in Lemma F.17. Then by Lemma F.17, we have

$$\|\widehat{\mathbf{U}}\widehat{\mathbf{V}}^\top - \mathbf{U}\mathbf{V}^\top\|_F \le \frac{3}{2}\left(\frac{1}{\sigma_r(\mathbf{\Theta})} + \frac{1}{\sigma_r(\widehat{\mathbf{\Theta}})}\right)\|\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\|_F.$$

Thus

$$\begin{aligned} d(\mathbf{F},\mathbf{F}') &\le \left(\frac{3\sqrt{2}}{2} + 2\right)\left(\frac{1}{\sigma_r(\mathbf{\Theta})} + \frac{1}{\sigma_r(\widehat{\mathbf{\Theta}})}\right)\|\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\|_F \\ &\le 5\left(\frac{1}{\sigma_r(\mathbf{\Theta})} + \frac{1}{\sigma_r(\widehat{\mathbf{\Theta}})}\right)\|\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\|_F, \end{aligned}$$

which completes the proof. □

Lemma F.11. *For any rank r matrices $\mathbf{\Theta}, \widehat{\mathbf{\Theta}} \in \mathbb{R}^{N\times K}$, we have*

$$\left|\sigma_{\min}(\mathbf{\Theta}) - \sigma_{\min}(\widehat{\mathbf{\Theta}})\right| \le \|\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\|_F.$$

Proof of Lemma F.11. By Lemma F.7, we have

$$\left|\sigma_r(\mathbf{\Theta}) - \sigma_r(\widehat{\mathbf{\Theta}})\right| \le \|\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\|_{\mathrm{op}} \le \|\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\|_F. \quad \square$$

Lemma F.12. *For any matrix $\mathbf{\Theta} \in \mathbb{R}^{N\times K}$, let the SVD of $\mathbf{\Theta} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$, we have*

$$\|\mathbf{U}\|_{2,\infty} \le \frac{\sqrt{K}\|\mathbf{\Theta}\|_\infty}{\sigma_{\min}(\mathbf{\Theta})} \quad \textit{and} \quad \|\mathbf{V}\|_{2,\infty} \le \frac{\sqrt{N}\|\mathbf{\Theta}\|_\infty}{\sigma_{\min}(\mathbf{\Theta})}.$$

Proof of Lemma F.12. We have

$$\|\mathbf{U}\|_{2,\infty}\sigma_{\min}(\mathbf{\Theta}) \le \|\mathbf{\Theta}\|_{2,\infty} \le \sqrt{K}\|\mathbf{\Theta}\|_\infty,$$

which gives us the first inequality. The second inequality can be obtained via an analogous argument. □

Lemma F.13. *For any positive definite matrices $\mathbf{A}, \mathbf{D} \in \mathbb{R}^{r\times r}$ where $\mathbf{D}$ is a diagonal matrix, we have*

$$\|\mathbf{A}^{-1} - \mathbf{D}^{-1}\|_{\mathrm{op}} \le \frac{\max_{i\in[r]} |\sigma_i(\mathbf{A}) - \mathbf{D}_{ii}|}{\sigma_{\min}(\mathbf{A})\sigma_{\min}(\mathbf{D})}$$

Proof of Lemma F.13. For any positive definite matrix $\mathbf{A}$, there exists an orthogonal matrix $\mathbf{Q}$ and a diagonal matrix $\mathbf{D}_A$ such that

$$\mathbf{A} = \mathbf{Q}\mathbf{D}_A\mathbf{Q}^\top,$$

where the i^{th} element on the diagonal of $\mathbf{D}_A$ is $\sigma_i(\mathbf{A})$. Thus we have $\mathbf{A}^{-1} = \mathbf{Q}\mathbf{D}_A^{-1}\mathbf{Q}^\top$, and

$$\begin{aligned} \mathbf{A}^{-1} - \mathbf{D}^{-1} &= \mathbf{Q}\mathbf{D}_A^{-1}\mathbf{Q}^\top - \mathbf{D}^{-1} \\ &= \mathbf{Q}\mathbf{D}_A^{-1}\mathbf{Q}^\top - \mathbf{Q}\mathbf{D}^{-1}\mathbf{Q}^\top \\ &= \mathbf{Q}(\mathbf{D}_A^{-1} - \mathbf{D}^{-1})\mathbf{Q}^\top. \end{aligned}$$

By the definition of operator norm, we know that

$$\begin{aligned}\|\mathbf{A}^{-1}-\mathbf{D}^{-1}\|_{\mathrm{op}} &= \max_{i\in[r]}|\mathbf{D}_A^{-1}(i,i)-\mathbf{D}^{-1}(i,i)| \\ &= \max_{i\in[r]}|\sigma_i(\mathbf{A})^{-1}-\mathbf{D}_{ii}^{-1}| \\ &= \max_{i\in[r]}\left|\frac{\mathbf{D}_{ii}-\sigma_i(\mathbf{A})}{\sigma_i(\mathbf{A})\mathbf{D}_{ii}}\right| \\ &\le \frac{\max_{i\in[r]}|\sigma_i(\mathbf{A})-\mathbf{D}_{ii}|}{\sigma_{\min}(\mathbf{A})\sigma_{\min}(\mathbf{D})}. \quad \square\end{aligned}$$

Lemma F.14. *For any matrices* $\mathbf{A},\mathbf{B},\mathbf{C}$*, we have*

$$\|\mathbf{A}\mathbf{B}\mathbf{C}^\top\|_\infty \le \|\mathbf{A}\mathbf{B}\|_{2,\infty}\|\mathbf{C}\|_{2,\infty} \le \|\mathbf{A}\|_{2,\infty}\|\mathbf{B}\|_{\mathrm{op}}\|\mathbf{C}\|_{2,\infty}.$$

Proof of Lemma F.14. The first inequality is straightforward. Here we show that

$$\|\mathbf{A}\mathbf{B}\|_{2,\infty} \le \|\mathbf{A}\|_{2,\infty}\|\mathbf{B}\|_{\mathrm{op}}$$

for any matrices $\mathbf{A}$ and $\mathbf{B}$. Let d_1 denote the row-dimension of matrix $\mathbf{A}$ and we have

$$\|\mathbf{A}\mathbf{B}\|_{2,\infty} = \max_{i\in[d_1]}\|e_i(d_1)^\top\mathbf{A}\mathbf{B}\| \le \max_{i\in[d_1]}\|e_i(d_1)^\top\mathbf{A}\|\|\mathbf{B}\|_{\mathrm{op}} = \|\mathbf{A}\|_{2,\infty}\|\mathbf{B}\|_{\mathrm{op}}. \quad \square$$

Lemma F.15. *For any* σ*-subgaussian random vector* $X\in\mathbb{R}^d$*, we have*

$$\mathbb{P}\left[\|X\|\ge\rho\right] \le 2\exp\left(\frac{-\rho^2}{2d\sigma^2}\right), \forall\rho>0.$$

Proof of Lemma F.15. See Lemma 1 in Jin et al. (2019). □

Lemma F.16. *For any rank* r *matrices* $\mathbf{\Theta},\widehat{\mathbf{\Theta}}\in\mathbb{R}^{N\times K}$*, let* $\widehat{\mathbf{\Theta}}=\widehat{\mathbf{U}}\widehat{\mathbf{D}}\widehat{\mathbf{V}}^\top$ *and* $\mathbf{\Theta}=\mathbf{U}\mathbf{D}\mathbf{V}^\top$ *be their SVDs. Then we have*

$$\|\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top-\mathbf{U}\mathbf{U}^\top\|_F \le \frac{\sqrt{2}\|\mathbf{\Theta}-\widehat{\mathbf{\Theta}}\|_F}{\sigma_r(\mathbf{\Theta})}.$$

Proof of Lemma F.16. First notice that

$$\|\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top-\mathbf{U}\mathbf{U}^\top\|_F = \sqrt{2}\cdot\inf_{\mathbf{Q}\in\mathbb{R}^{r\times r}}\|\mathbf{U}-\widehat{\mathbf{U}}\mathbf{Q}\|_F.$$

To show this equality, we first note that $\|\mathbf{U}-\widehat{\mathbf{U}}\mathbf{Q}\|_F$ is minimized when $\mathbf{Q}=\widehat{\mathbf{U}}^\top\mathbf{U}$. Thus we have

$$\begin{aligned}\inf_{\mathbf{Q}\in\mathbb{R}^{r\times r}}\|\mathbf{U}-\widehat{\mathbf{U}}\mathbf{Q}\|_F &= \|\mathbf{U}-\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top\mathbf{U}\|_F \\ &= \sqrt{\mathrm{tr}\left(\left(\mathbf{U}-\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top\mathbf{U}\right)\left(\mathbf{U}-\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top\mathbf{U}\right)^\top\right)} \\ &= \sqrt{\mathrm{tr}\left(\mathbf{U}\mathbf{U}^\top\right)-2\,\mathrm{tr}\left(\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top\mathbf{U}\mathbf{U}^\top\right)+\mathrm{tr}\left(\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top\mathbf{U}\mathbf{U}^\top\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top\right)} \\ &= \sqrt{\mathrm{tr}\left(\mathbf{U}\mathbf{U}^\top\right)-\mathrm{tr}\left(\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top\mathbf{U}\mathbf{U}^\top\right)} \\ &= \sqrt{r-\mathrm{tr}\left(\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top\mathbf{U}\mathbf{U}^\top\right)}.\end{aligned}$$

On the other hand, we have

$$\begin{aligned}\|\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top - \mathbf{U}\mathbf{U}^\top\|_F &= \sqrt{\operatorname{tr}\left(\left(\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top - \mathbf{U}\mathbf{U}^\top\right)\left(\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top - \mathbf{U}\mathbf{U}^\top\right)^\top\right)} \\ &= \sqrt{2r - 2\operatorname{tr}\left(\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top\mathbf{U}\mathbf{U}^\top\right)}.\end{aligned}$$

With this inequality, now we can derive

$$\begin{aligned}\|\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top - \mathbf{U}\mathbf{U}^\top\|_F &= \sqrt{2}\cdot \inf_{\mathbf{Q}\in\mathbb{R}^{r\times r}} \|\mathbf{U} - \widehat{\mathbf{U}}\mathbf{Q}\|_F \\ &= \sqrt{2}\cdot \inf_{\mathbf{Q}\in\mathbb{R}^{r\times r}} \|\mathbf{U}^\top - \mathbf{Q}^\top\widehat{\mathbf{U}}^\top\|_F \\ &= \sqrt{2}\cdot \inf_{\mathbf{Q}\in\mathbb{R}^{r\times r}} \|\mathbf{D}^{-1}(\mathbf{D}\mathbf{U}^\top - \mathbf{D}\mathbf{Q}^\top\widehat{\mathbf{U}}^\top)\|_F \\ &\leq \sqrt{2}\cdot \|\mathbf{D}^{-1}\left(\mathbf{D}\mathbf{U}^\top - \mathbf{D}(\mathbf{D}^{-1}\mathbf{V}^\top\widehat{\mathbf{V}}\widehat{\mathbf{D}})\widehat{\mathbf{U}}^\top\right)\|_F \\ &= \sqrt{2}\cdot \|\mathbf{D}^{-1}\left(\mathbf{V}^\top\mathbf{V}\mathbf{D}\mathbf{U}^\top - \mathbf{V}^\top\widehat{\mathbf{V}}\widehat{\mathbf{D}}\widehat{\mathbf{U}}^\top\right)\|_F \\ &= \sqrt{2}\cdot \left\|\mathbf{D}^{-1}\mathbf{V}^\top\left(\mathbf{V}\mathbf{D}\mathbf{U}^\top - \widehat{\mathbf{V}}\widehat{\mathbf{D}}\widehat{\mathbf{U}}^\top\right)\right\|_F \\ &\leq \sqrt{2}\left\|\mathbf{V}\mathbf{D}\mathbf{U}^\top - \widehat{\mathbf{V}}\widehat{\mathbf{D}}\widehat{\mathbf{U}}^\top\|_F\right\|\mathbf{D}^{-1}\mathbf{V}^\top\|_{\mathrm{op}} \\ &\leq \frac{\sqrt{2}\|\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\|_F}{\sigma_{\min}(\mathbf{\Theta})},\end{aligned}$$

which completes our proof. □

Lemma F.17. *For any rank r matrices $\mathbf{\Theta}, \widehat{\mathbf{\Theta}} \in \mathbb{R}^{N\times K}$, we have*

$$\|\operatorname{sgn}(\mathbf{\Theta}) - \operatorname{sgn}(\widehat{\mathbf{\Theta}})\|_F \leq \frac{3}{2}\left(\frac{1}{\sigma_r(\mathbf{\Theta})} + \frac{1}{\sigma_r(\widehat{\mathbf{\Theta}})}\right)\|\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\|_F$$

where $\operatorname{sgn}(\cdot)$ *denotes the matrix sign function, i.e.,* $\operatorname{sgn}(\mathbf{\Theta}) = \mathbf{U}\mathbf{V}^\top$ *for a matrix* $\mathbf{\Theta}$ *with SVD* $\mathbf{U}\mathbf{D}\mathbf{V}^\top$.

Proof of Lemma F.17. See Theorem 2.1 of Li and Sun (2006). □

Appendix G: Experiment Details

G.1. Synthetic Data

We create our synthetic data through the following steps. In particular, we consider $N = 100$ worker types and $K = 100$ job types — i.e., there are $N \times K = 10,000$ number of worker-job pairs. For each of the 50 trials, we generate the ground-truth matching reward matrix[8] $\mathbf{\Theta}^* \in \mathbb{R}^{N\times K}$ with rank $r = 3$ as follows. First, we create two matrices $\mathbf{U}^* \in \mathbb{R}^{N\times r}$ and $\mathbf{V}^* \in \mathbb{R}^{K\times r}$, of which the entries are independently drawn from the uniform distribution on $[0,1]$. Then, we multiply the two matrices together and obtain $\mathbf{\Theta}^* = \mathbf{U}^*\mathbf{V}^{*\top}$.

Offline. We simulate n number of offline matching samples by drawing the matchings $\mathbf{X}_t$'s uniformly at random from the set of all matchings $\mathcal{M}$. n takes the values $20, 40, 60, 80, 100$ respectively. Each matching sample $t \in [n]$ contains N noisy rewards from its N matched pairs following (2.4). We draw the noises $\varepsilon_t^{(i)}$'s i.i.d. from a Gaussian distribution $\mathcal{N}(0, \sigma^2)$ with $\sigma^2 = 0.1$. For our algorithm, we tune c_λ (i.e., a hyperparameter of λ stated in Theorem 1) on a pre-specified grid and choose c_λ equal to 0.001.

[8] For the online stable matching experiment, $\mathbf{\Theta}^*$ represents the worker reward matrix that indicates worker preference rankings.

Online. We set our total time horizon T equal to $200, 400, 600, 800, 1000$ respectively. For the online optimal matching experiments, we set the number of exploration steps in our CombLRB to be $E_h = qT^{2/3}$ as advised by Theorem 4 for some hyperparameter q. We tune the hyperparameter q and choose $q = 1$; λ is set in the same way as the offline setting. The implementations of both CUCB and CTS algorithms follow that described in Section 6 of Cuvelier et al. (2021).

For the online stable matching experiments, we generate the job preference rankings over workers through the matrix $\mathbf{\Phi}^*$. Each column of $\mathbf{\Phi}^*$ is a random permutation of $[N]$. For the 50 trials, we draw 10 such $\mathbf{\Phi}^*$ independently and then conduct 5 independent trials with each $\mathbf{\Phi}^*$. we set the length of exploration phase for our CompLRB to be $E_h = q \log T$ for some hyperparameter q. We tune the hyperparameter q and choose $q = 40$. λ is set in the same way as the offline setting.

G.2. Real Data of Labor Market

We use an individual-level workforce dataset provided by Revelio Labs, which includes comprehensive employment histories of individuals, including their roles, skills, activities, education, seniority, geographic location, etc. For our analysis, we focus on software engineers with mid-level seniority employed in the United States between 2010 and 2015. This subset contains observations from 468,807 software engineers employed by 89,365 different companies.

We group the engineers by their education background. Specifically, engineers who graduate from the same school are put in the same group. We group all the 37,926 schools into 236 clusters by their sizes, approximated by the total number of the graduates (i.e., employees) observed in the data. Schools with similar sizes are grouped into the same cluster. Similarly, companies with similar sizes are grouped into 100 clusters. We finally keep the top $N = 50$ engineer clusters and $K = 50$ company clusters with the highest number of observations to leave out pairs with missing data and very few observations. This procedure results in a 50×50 ground-truth matrix $\mathbf{\Theta}^*$, where each entry represents the empirical probability that an engineer from a given cluster stays in a company from a corresponding cluster for more than six months.

Our matching observations are generated following (2.4), where the noises are i.i.d. Gaussian with mean 0 and variance 0.1. The variance is set to be the empirical variance across all entries. For online stable matching experiment, the job preference ranking matrix $\mathbf{\Phi}^*$ is generated in the same way as our synthetic experiment. Similar to our synthetic experiments, we tune all the hyperparameters on pre-specified grids.