Multi-task Representation Learning for Pure Exploration in Bilinear Bandits

Subhojyoti Mukherjee
ECE Department
UW-Madison
Wisconsin, Madison
smukherjee27@wisc.edu

Qiaomin Xie
ISyE Department
UW-Madison
Wisconsin, Madison

Josiah P. Hanna
CS Department
UW-Madison
Wisconsin, Madison

Robert Nowak
ECE Department
UW-Madison
Wisconsin, Madison

Abstract

We study multi-task representation learning for the problem of pure exploration in bilinear bandits. In bilinear bandits, an action takes the form of a pair of arms from two different entity types and the reward is a bilinear function of the known feature vectors of the arms. In the *multi-task bilinear bandit problem*, we aim to find optimal actions for multiple tasks that share a common low-dimensional linear representation. The objective is to leverage this characteristic to expedite the process of identifying the best pair of arms for all tasks. We propose the algorithm GOBLIN that uses an experimental design approach to optimize sample allocations for learning the global representation as well as minimize the number of samples needed to identify the optimal pair of arms in individual tasks. To the best of our knowledge, this is the first study to give sample complexity analysis for pure exploration in bilinear bandits with shared representation. Our results demonstrate that by learning the shared representation across tasks, we achieve significantly improved sample complexity compared to the traditional approach of solving tasks independently.

1 Introduction

Bilinear bandits (Jun et al., 2019; Lu et al., 2021; Kang et al., 2022) are an important class of sequential decision-making problems. In bilinear bandits (as opposed to the standard linear bandit setting) we are given a pair of arms $\mathbf{x}_t \in \mathbb{R}^{d_1}$ and $\mathbf{z}_t \in \mathbb{R}^{d_2}$ at every round t and the interaction of this pair of arms with a low-rank hidden parameter, $\mathbf{\Theta}_* \in \mathbb{R}^{d_1 \times d_2}$ generates the noisy feedback (reward) $r_t = \mathbf{x}_t^\top \mathbf{\Theta}_* \mathbf{z}_t + \eta_t$. The η_t is random 1-subGaussian noise.

A lot of real-world applications exhibit the above bilinear feedback structure, particularly applications that involve selecting pairs of items and evaluating their compatibility. For example, in a drug discovery application, scientists may want to determine whether a particular (drug, protein) pair interacts in the desired way (Luo et al., 2017). Likewise, an online dating service might match a pair of people and gather feedback about their compatibility (Shen et al., 2023). A clothing website's recommendation system may suggest a pair of items (top, bottom) for a customer based on their likelihood of matching (Reyes et al., 2021). In all of these scenarios, the two items are considered as a single unit, and the system must utilize available feature vectors $(\mathbf{x}_t, \mathbf{z}_t)$ to learn which features of the pairs are most indicative of positive feedback in order to make effective recommendations. All the previous works in this setting (Jun et al., 2019; Lu et al., 2021; Kang et al., 2022) exclusively

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

focused on maximizing the number of pairs with desired interactions discovered over time (regret minimization). However, in many real-world applications where obtaining a sample is expensive and time-consuming, e.g., clinical trials (Zhao et al., 2009; Zhang et al., 2012), it is often desirable to identify the optimal option using as few samples as possible, i.e., we face the pure exploration scenario (Fiez et al., 2019; Katz-Samuels et al., 2020) rather than regret minimization.

Moreover, in various decision-making scenarios, we may encounter multiple interrelated tasks such as treatment planning for different diseases (Bragman et al., 2018) and content optimization for multiple websites (Agarwal et al., 2009). Often, there exists a shared representation among these tasks, such as the features of drugs or the representations of website items. Therefore, we can leverage this shared representation to accelerate learning. This area of research is called multi-task representation learning and has recently generated a lot of attention in machine learning (Bengio et al., 2013; Li et al., 2014; Maurer et al., 2016; Du et al., 2020; Tripuraneni et al., 2021). There are many applications of this multi-task representation learning in real-world settings. For instance, in clinical treatment planning, we seek to determine the optimal treatments for multiple diseases, and there may exist a low-dimensional representation common to multiple diseases. To avoid the time-consuming process of conducting clinical trials for individual tasks and collecting samples, we utilize the shared representation and decrease the number of required samples.

The above multi-task representation learning naturally shows up in bilinear bandit setting as follows: Let there be M tasks indexed as $m = 1, 2, \ldots, M$ with each task having its own hidden parameter $\boldsymbol{\Theta}_{m,*} \in \mathbb{R}^{d_1 \times d_2}$. Let each $\boldsymbol{\Theta}_{m,*}$ has a decomposition of $\boldsymbol{\Theta}_{m,*} = \mathbf{B}_1 \mathbf{S}_{m,*} \mathbf{B}_2^\top$, where $\mathbf{B}_1 \in \mathbb{R}^{d_1 \times k_1}$ and $\mathbf{B}_2 \in \mathbb{R}^{d_2 \times k_2}$ are shared across tasks, but $\mathbf{S}_{m,*} \in \mathbb{R}^{k_1 \times k_2}$ is specific for task m. We assume that $k_1, k_2 \ll d_1, d_2$ and $M \gg d_1, d_2$. Thus, $\mathbf{B}_1$ and $\mathbf{B}_2$ provide a means of dimensionality reduction. Furthermore, we assume that each $\mathbf{S}_{m,*}$ has rank $r \ll \min\{k_1, k_2\}$. In the terminology of multi-task representation learning $\mathbf{B}_1, \mathbf{B}_2$ are called *feature extractors* and $\mathbf{x}_{m,t}, \mathbf{z}_{m,t}$ are called *rich observations* (Yang et al., 2020, 2022; Du et al., 2023). The reward for the task $m \in \{1, 2, \ldots, M\}$ at round t is

$$r_{m,t} = \mathbf{x}_{m,t}^\top \boldsymbol{\Theta}_{m,*} \mathbf{z}_{m,t} + \eta_{m,t} = \underbrace{\mathbf{x}_{m,t}^\top \mathbf{B}_1}_{\mathbf{g}_{m,t}^\top} \mathbf{S}_{m,*} \underbrace{\mathbf{B}_2^\top \mathbf{z}_{m,t}}_{\mathbf{v}_{m,t}} + \eta_{m,t} = \mathbf{g}_{m,t}^\top \mathbf{S}_{m,*} \mathbf{v}_{m,t} + \eta_{m,t}. \quad (1)$$

Observe that similar to the learning procedure in Yang et al. (2020, 2022), at each round $t = 1, 2, \cdots$, for each task $m \in [M]$, the learner selects a left and right action $\mathbf{x}_{m,t} \in \mathcal{X}$ and $\mathbf{z}_{m,t} \in \mathcal{Z}$. After the player commits the batch of actions for each task $\{\mathbf{x}_{m,t}, \mathbf{z}_{m,t} : m \in [M]\}$, it receives the batch of rewards $\{r_{m,t} : m \in [M]\}$. Also note that in (1) we define the $\widetilde{\mathbf{g}}_{m,t} \in \mathbb{R}^{k_1}, \widetilde{\mathbf{v}}_{m,t} \in \mathbb{R}^{k_2}$ as the latent features, and both $\widetilde{\mathbf{g}}_{m,t}, \widetilde{\mathbf{v}}_{m,t}$ are unknown to the learner and needs to be learned for each task m (hence the name multi-task representation learning).

In this paper, we focus on pure exploration for multi-task representation learning in bilinear bandits where the goal is to find the optimal left arm $\mathbf{x}_{m,*}$ and right arm $\mathbf{z}_{m,*}$ for each task m with a minimum number of samples (fixed confidence setting). First, consider a single-task setting and let $\boldsymbol{\Theta}_*$ have low rank r. Let the SVD of the $\boldsymbol{\Theta}_* = \mathbf{U}\mathbf{D}\mathbf{V}^\top$. Prima-facie, if $\mathbf{U}$ and $\mathbf{V}$ are known then one might want to project all the left and right arms in the $r \times r$ subspace of $\mathbf{U}$ and $\mathbf{V}$ and reduce the bilinear bandit problem into a r^2 dimension linear bandit setting. Then one can apply one of the algorithms from Soare et al. (2014); Fiez et al. (2019); Katz-Samuels et al. (2020) to solve this r^2 dimensional linear bandit pure exploration problem. Following the analysis of this line of work (in linear bandits) (Mason et al., 2021; Mukherjee et al., 2022, 2023) one might conjecture that a sample complexity bound of $\widetilde{O}(r^2/\Delta^2)$ is possible where Δ is the minimum reward gap and $\widetilde{O}(\cdot)$ hides log factors. Similarly, for the multi-task setting one might be tempted to use the linear bandit analysis of Du et al. (2023) to convert this problem into M concurrent r^2 dimensional linear bandit problems with shared representation and achieve a sample complexity bound of $\widetilde{O}(Mr^2/\Delta^2)$. However, these matrices (subspaces) are unknown and so there is a model mismatch as noted in the regret analysis of bilinear bandits (Jun et al., 2019; Lu et al., 2021; Kang et al., 2022). Thus it is difficult to apply the r^2 dimensional linear bandit sample complexity analysis. Following the regret analysis of bilinear bandit setting by Jun et al. (2019); Lu et al. (2021); Kang et al. (2022) we know that the effective dimension is actually $(d_1 + d_2)r$. Similarly for the multi-task representation learning the effective dimension should scale with the learned latent features $(k_1 + k_2)r$. Hence the natural questions to ask are these:

> **1) Can we design a single-task pure exploration bilinear bandit algorithm whose sample complexity scales as $\widetilde{O}((d_1 + d_2)r/\Delta^2)$?**

2

2) Can we design an algorithm for multi-task pure exploration bilinear bandit problem that can learn the latent features and has sample complexity that scales as $\widetilde{O}(M(k_1 + k_2)r/\Delta^2)$?

In this paper, we answer both these questions affirmatively. In doing so, we make the following novel contributions to the growing literature of multi-task representation learning in online settings:

1) We formulate the multi-task bilinear representation learning problem. To our knowledge, this is the first work that explores pure exploration in a multi-task bilinear representation learning setting.

2) We proposed the algorithm GOBLIN for a single-task pure exploration bilinear bandit setting whose sample complexity scales as $\widetilde{O}((d_1 + d_2)r/\Delta^2)$. This improves over RAGE (Fiez et al., 2019) whose sample complexity scales as $\widetilde{O}((d_1 d_2)/\Delta^2)$.

3) Our algorithm GOBLIN for multi-task pure exploration bilinear bandit problem learns the latent features and has sample complexity that scales as $\widetilde{O}(M(k_1 + k_2)r/\Delta^2)$. This improves over DouExpDes (Du et al., 2023) whose samples complexity scales as $\widetilde{O}(M(k_1 k_2)/\Delta^2)$.

Preliminaries: We assume that $\|\mathbf{x}\|_2 \leq 1$, $\|\mathbf{z}\|_2 \leq 1$, $\|\boldsymbol{\Theta}_*\|_F \leq S_0$ and the r-th largest singular value of $\boldsymbol{\Theta}_* \in \mathbb{R}^{d_1 \times d_2}$ is S_r. Let $p := d_1 d_2$ denote the ambient dimension, and $k = (d_1 + d_2)r$ denote the effective dimension. Let $[n] := \{1, 2, \ldots, n\}$. Let $\mathbf{x}_*, \mathbf{z}_* := \arg\max_{\mathbf{x},\mathbf{z}} \mathbf{x}^\top \boldsymbol{\Theta}_* \mathbf{z}$. For any $\mathbf{x}, \mathbf{z}$ define the gap $\Delta(\mathbf{x}, \mathbf{z}) := \mathbf{x}_*^\top \boldsymbol{\Theta}_* \mathbf{z}_* - \mathbf{x}^\top \boldsymbol{\Theta}_* \mathbf{z}$ and furthermore $\Delta = \min_{\mathbf{x} \neq \mathbf{x}_*, \mathbf{z} \neq \mathbf{z}_*} \Delta(\mathbf{x}, \mathbf{z})$. Similarly, for any arbitrary vector $\mathbf{w} \in \mathcal{W}$ define the gap of $\mathbf{w} \in \mathbb{R}^p$ as $\Delta(\mathbf{w}) := (\mathbf{w}_* - \mathbf{w})^\top \boldsymbol{\theta}_*$, for some $\boldsymbol{\theta}_* \in \mathbb{R}^p$ and furthermore, $\Delta = \min_{\mathbf{w} \neq \mathbf{w}_*} \Delta(\mathbf{w})$. If $\mathbf{A} \in \mathbb{R}^{d \times d}_{\geq 0}$ is a positive semidefinite matrix, and $\mathbf{w} \in \mathbb{R}^p$ is a vector, let $\|\mathbf{w}\|_\mathbf{A}^2 := \mathbf{w}^\top \mathbf{A} \mathbf{w}$ denote the induced semi-norm. Given any vector $\mathbf{b} \in \mathbb{R}^{|\mathcal{W}|}$ we denote the $\mathbf{w}$-th component as $\mathbf{b}_\mathbf{w}$. Let $\Delta_\mathcal{W} := \{\mathbf{b} \in \mathbb{R}^{|\mathcal{W}|} : \mathbf{b}_\mathbf{w} \geq 0, \sum_{\mathbf{w} \in \mathcal{W}} \mathbf{b}_\mathbf{w} = 1\}$ denote the set of probability distributions on $\mathcal{W}$. We define $\mathcal{Y}(\mathcal{W}) = \{\mathbf{w} - \mathbf{w}' : \forall \mathbf{w}, \mathbf{w}' \in \mathcal{W}, \mathbf{w} \neq \mathbf{w}'\}$ as the directions obtained from the differences between each pair of arms and $\mathcal{Y}^*(\mathcal{W}) = \{\mathbf{w}_* - \mathbf{w} : \forall \mathbf{w} \in \mathcal{W} \backslash \mathbf{w}_*\}$ as the directions obtained from the differences between the optimal arm and each suboptimal arm.

2 Pure Exploration in Single-Task Bilinear Bandits

In this section, we consider pure exploration in a single-task bilinear bandit setting as a warm-up to the main goal of learning representations for the multi-task bilinear bandit. To our knowledge, this is the first study of pure exploration in single-task bilinear bandits. We first recall the single-task bilinear bandit setting as follows: At every round $t = 1, 2, \ldots$ the learner observes the reward $r_t = \mathbf{x}_t^\top \boldsymbol{\Theta}_* \mathbf{z}_t + \eta_t$ where the low rank hidden parameter $\boldsymbol{\Theta}_* \in \mathbb{R}^{d_1 \times d_2}$ is unknown to the learner, $\mathbf{x}_t \in \mathbb{R}^{d_1}$, $\mathbf{z}_t \in \mathbb{R}^{d_2}$ are visible to the learner, and η_t is a 1-sub-Gaussian noise. We assume that the matrix $\boldsymbol{\Theta}_*$ has a low rank r which is known to the learner and $d_1, d_2 \gg r$. Finally recall that the goal is to identify the optimal left and right arms $\mathbf{x}_*, \mathbf{z}_*$ with a minimum number of samples.

We propose a phase-based, two-stage arm elimination algorithm called **G**-**O**ptimal Design for **Bilin**ear Bandits (abbreviated as GOBLIN). GOBLIN proceeds in phases indexed by $\ell = 1, 2, \ldots$. As this is a pure-exploration problem, the total number of samples is controlled by the total phases which depends on the intrinsic problem complexity. Each phase ℓ of GOBLIN consists of two stages; the estimation of $\boldsymbol{\Theta}_*$ stage, which runs for τ_ℓ^E rounds, and pure exploration in rotated arms stage that runs for τ_ℓ^G rounds. We will define τ_ℓ^E in Section 2.1, while rotated arms and τ_ℓ^G are defined in Section 2.2. At the end of every phase, GOBLIN eliminates sub-optimal arms to build the active set for the next phase and stops when only the optimal left and right arms are remaining. Now we discuss the individual stages that occur at every phase ℓ of GOBLIN.

2.1 Estimating Subspaces of $\boldsymbol{\Theta}_*$ (Stage 1 of the ℓ-th phase)

In the first stage of phase ℓ, GOBLIN estimates the row and column sub-spaces $\boldsymbol{\Theta}_*$. Then GOBLIN uses these estimates to reduce the bilinear bandit problem in the original ambient dimension $p := d_1 d_2$ to a lower effective dimension $k := (d_1 + d_2)r$. To do this, GOBLIN first vectorizes the $\mathbf{x} \in \mathbb{R}^{d_1}$, $\mathbf{z} \in \mathbb{R}^{d_2}$ into a new vector $\overline{\mathbf{w}} \in \mathbb{R}^p$ and then solves the E-optimal design in Step 3 of

3

Algorithm 1 (Pukelsheim, 2006; Jun et al., 2019; Du et al., 2023). Let the solution to the E-optimal design problem at the stage 1 of ℓ-th phase be denoted by $\mathbf{b}_\ell^E$. Then GOBLIN samples each $\overline{\mathbf{w}}$ for $\lceil \tau_\ell^E \mathbf{b}_{\ell,\overline{\mathbf{w}}}^E \rceil$ times, where $\tau_\ell^E = \widetilde{O}(\sqrt{d_1 d_2 r}/S_r)$ (step 7 of Algorithm 1). In this paper, we sample an arm $\lceil \tau_\ell^E \mathbf{b}_{\ell,\overline{\mathbf{w}}}^E \rceil$ number of times. However, this may lead to over-sampling of an arm than what the design (G or E-optimal) is actually suggesting. However, we can match the number of allocations of an arm to the design using an *efficient Rounding Procedures* (see Pukelsheim (2006); Fiez et al. (2019)). Let $\widehat{\boldsymbol{\Theta}}_\ell$ be estimate of $\boldsymbol{\Theta}_*$ in stage 1 of phase ℓ. GOBLIN estimates this by solving the following well-defined regularized minimization problem with nuclear norm penalty:

$$\widehat{\boldsymbol{\Theta}}_\ell = \underset{\boldsymbol{\Theta} \in \mathbb{R}^{d_1 \times d_2}}{\arg\min} \ L_\ell(\boldsymbol{\Theta}) + \gamma_\ell \|\boldsymbol{\Theta}\|_{\mathrm{nuc}}, \quad L_\ell(\boldsymbol{\Theta}) = \langle \boldsymbol{\Theta}, \boldsymbol{\Theta} \rangle - \frac{2}{\tau_\ell^E} \sum_{s=1}^{\tau_\ell^E} \langle \widetilde{\psi}_\nu(r_s \cdot Q(\mathbf{x}_s \mathbf{z}_s^\top)), \boldsymbol{\Theta} \rangle \quad (2)$$

where $Q(\cdot)$, $\widetilde{\psi}_\nu(\cdot)$, are appropriate functions stated in Definition 1, 3 respectively in Appendix A.3. The $Q(\cdot)$ function takes as input the rank-one matrix $\mathbf{x}_s \mathbf{z}_s^\top$ which is obtained after reshaping $\overline{\mathbf{w}}_s$. Note that $\mathbf{x}_s$, and $\mathbf{z}_s$ are the observed vectors in d_1 and d_2 dimension and $\widehat{\boldsymbol{\Theta}}_\ell \in \mathbb{R}^{d_1 \times d_2}$ Finally, set the regularization parameter $\gamma_\ell := 4\sqrt{\frac{2(4+S_0^2)Cd_1d_2\log(2(d_1+d_2)/\delta)}{\tau_\ell^E}}$. This is in step 8 of Algorithm 1.

2.2 Optimal Design for Rotated Arms (Stage 2 of ℓ-th phase)

In stage 2 of phase ℓ, GOBLIN leverages the information about the learned sub-space of $\boldsymbol{\Theta}_*$ to rotate the arm set and then run the optimal design on the rotated arm set. Once we recover $\widehat{\boldsymbol{\Theta}}_\ell$, one might be tempted to run a pure exploration algorithm (Soare et al., 2014; Fiez et al., 2019; Katz-Samuels et al., 2020; Zhu et al., 2021) to identify $\mathbf{x}_*$ and $\mathbf{z}_*$. However, then the sample complexity will scale with $d_1 d_2$. In contrast GOBLIN uses the information about the learned sub-space of $\boldsymbol{\Theta}_*$ to reduce the problem from ambient dimension $d_1 d_2$ to effective dimension $(d_1 + d_2)r$. This reduction is done as follows: Let $\widehat{\boldsymbol{\Theta}}_\ell = \widehat{\mathbf{U}}_\ell \widehat{\mathbf{D}}_\ell \widehat{\mathbf{V}}_\ell^\top$ be the SVD of $\widehat{\boldsymbol{\Theta}}_\ell$ in the ℓ-th phase. Let $\widehat{\mathbf{U}}_\perp^\ell$ and $\widehat{\mathbf{V}}_\perp^\ell$ be orthonormal bases of the complementary subspaces of $\widehat{\mathbf{U}}_\ell$ and $\widehat{\mathbf{V}}_\ell$ respectively. Let $\mathcal{X}_\ell$ and $\mathcal{Z}_\ell$ be the active set of arms in the stage 2 of phase ℓ. Then rotate the arm sets such that new rotated arm sets are as follows:

$$\underline{\mathcal{X}}_\ell = \{\underline{\mathbf{x}} = [\widehat{\mathbf{U}}_\ell \widehat{\mathbf{U}}_\ell^\perp]^\top \mathbf{x} \mid \mathbf{x} \in \mathcal{X}_\ell\}, \underline{\mathcal{Z}}_\ell = \{\underline{\mathbf{z}} = [\widehat{\mathbf{V}}_\ell \widehat{\mathbf{V}}_\ell^\perp]^\top \mathbf{z} \mid \mathbf{z} \in \mathcal{Z}_\ell\}. \quad (3)$$

Let $\widehat{\mathbf{H}}_\ell = [\widehat{\mathbf{U}}_\ell \widehat{\mathbf{U}}_\ell^\perp]^\top \widehat{\boldsymbol{\Theta}}_\ell [\widehat{\mathbf{V}}_\ell \widehat{\mathbf{V}}_\ell^\perp]$. Then define vectorized arm set so that the last $(d_1 - r) \cdot (d_2 - r)$ components are from the complementary subspaces as follows:

$$\underline{\mathcal{W}}_\ell = \big\{ \big[\mathbf{vec}\left(\underline{\mathbf{x}}_{1:r}\underline{\mathbf{z}}_{1:r}^\top\right); \mathbf{vec}\left(\underline{\mathbf{x}}_{r+1:d_1}\underline{\mathbf{z}}_{1:r}^\top\right); \mathbf{vec}\left(\underline{\mathbf{x}}_{1:r}\underline{\mathbf{z}}_{r+1:d_2}^\top\right);$$
$$\mathbf{vec}\left(\underline{\mathbf{x}}_{r+1:d_1}\underline{\mathbf{z}}_{r+1:d_2}^\top\right) \big] \in \mathbb{R}^{d_1 d_2} : \underline{\mathbf{x}} \in \mathcal{X}_\ell, \underline{\mathbf{z}} \in \mathcal{Z}_\ell \big\}$$
$$\widehat{\boldsymbol{\theta}}_{\ell,1:k} = [\mathbf{vec}(\widehat{\mathbf{H}}_{\ell,1:r,1:r}); \mathbf{vec}(\widehat{\mathbf{H}}_{\ell,r+1:d_1,1:r}); \mathbf{vec}(\widehat{\mathbf{H}}_{\ell,1:r,r+1:d_2})],$$
$$\widehat{\boldsymbol{\theta}}_{\ell,k+1:p} = \mathbf{vec}(\widehat{\mathbf{H}}_{\ell,r+1:d_1,r+1:d_2}). \quad (4)$$

which implies $\|\widehat{\boldsymbol{\theta}}_{k+1:p}\|_2 = O\left(d_1 d_2 r/\tau_\ell^E\right)$ by Lemma 3 in Appendix A.1. So the last $p - k$ components of $\widehat{\boldsymbol{\theta}}_\ell$ are very small compared to the first k components. Hence, GOBLIN has now reduced the $d_1 d_2$ dimensional linear bandit to $(d_1 + d_2)r$ dimensional linear bandit using (3), (4). This is shown in step 10 of Algorithm 1.

Now in stage 2 of phase ℓ, GOBLIN implements G-optimal design (Pukelsheim, 2006; Fiez et al., 2019) in the rotated arm set $\underline{\mathcal{X}}_\ell, \underline{\mathcal{Z}}_\ell$ defined in (3). To do this, first GOBLIN defines the rotated vector $\underline{\mathbf{w}} = [\underline{\mathbf{x}}_{1:d_1}; \underline{\mathbf{z}}_{1:d_2}] \in \mathbb{R}^p$ that belong to the set $\underline{\mathcal{W}}_\ell$. Then GOBLIN solves the G-optimal design (Pukelsheim, 2006) as follows:

$$\widehat{\mathbf{b}}_\ell^G = \underset{\mathbf{b}_{\underline{\mathbf{w}}}}{\arg\min} \ \max_{\underline{\mathbf{w}},\underline{\mathbf{w}}' \in \underline{\mathcal{W}}_\ell} \|\underline{\mathbf{w}} - \underline{\mathbf{w}}'\|_{(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}} \mathbf{b}_{\underline{\mathbf{w}}}\underline{\mathbf{w}}\ \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell/n)^{-1}}^2. \quad (5)$$

This is shown in step 11 of Algorithm 1 and $\boldsymbol{\Lambda}_\ell$ is defined in (6). It can be shown that sampling according to $\widehat{\mathbf{b}}_\ell^G$ leads to the optimal sample complexity. This is discussed in Remark 1 in Appendix A.2. The key point to note from (5) is that due to the estimation in the rotated arm space $\underline{\mathcal{W}}_\ell$ we are guaranteed that the support of $\mathbf{supp}(\widehat{\mathbf{b}}_\ell^G) \le \widetilde{O}(k(k+1)/2)$ (Pukelsheim, 2006). On the other hand,

4

if the G-optimal design of Fiez et al. (2019); Katz-Samuels et al. (2020) are run in $d_1 d_2$ dimension then the support of $\widehat{\mathbf{b}}_\ell^G$ will scale with $d_1 d_2$ which will lead to higher sample complexity. Then GOBLIN samples each $\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell$ for $\lceil \tau_\ell^G \mathbf{b}_{\ell,\underline{\mathbf{w}}}^G \rceil$ times, where $\tau_\ell^G := \lceil \frac{8 B_*^\ell \rho^G(\mathcal{Y}(\mathcal{W}_\ell)) \log(4\ell^2 |\mathcal{W}|/\delta)}{\epsilon_\ell^2} \rceil$. Note that the total length of phase ℓ, combining stages 1 and 2 is $(\tau_\ell^E + \tau_\ell^G)$ rounds. Observe that the stage 1 design is on the whole arm set $\overline{\mathcal{W}}$ whereas stage 2 design is on the refined active set $\underline{\mathcal{W}}_\ell$.

Let the observed features in stage 2 of phase ℓ be denoted by $\underline{\mathbf{W}}_\ell \in \mathbb{R}^{\tau_\ell^G \times p}$, and $\mathbf{r}_\ell \in \mathbb{R}^{\tau_\ell^G}$ be the observed rewards. Define the diagonal matrix $\mathbf{\Lambda}_\ell$ as

$$\mathbf{\Lambda}_\ell = \mathbf{diag}[\underbrace{\lambda, \ldots, \lambda}_{k}, \underbrace{\lambda_\ell^\perp, \ldots, \lambda_\ell^\perp}_{p-k}] \tag{6}$$

where, $\lambda_\ell^\perp := \tau_{\ell-1}^G / 8k \log(1 + \tau_{\ell-1}^G/\lambda) \gg \lambda$. Deviating from Soare et al. (2014); Fiez et al. (2019) GOBLIN constructs a regularized least square estimator at phase ℓ as follows

$$\widehat{\boldsymbol{\theta}}_\ell = \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \frac{1}{2} \|\underline{\mathbf{W}}_\ell \boldsymbol{\theta} - \mathbf{r}_\ell\|_2^2 + \frac{1}{2} \|\boldsymbol{\theta}\|_{\mathbf{\Lambda}_\ell}^2. \tag{7}$$

This regularized least square estimator in (7) forces the last $p - k$ components of $\widehat{\boldsymbol{\theta}}_\ell$ to be very small compared to the first k components. Then GOBLIN builds the estimate $\widehat{\boldsymbol{\theta}}_\ell$ from (7) only from the observations from this phase (step 13 in Algorithm 1) and eliminates sub-optimal actions in step 14 in Algorithm 1 using the estimator $\widehat{\boldsymbol{\theta}}_\ell$. Finally GOBLIN eliminates sub-optimal arms to build the next phase active set $\underline{\mathcal{W}}_\ell$ and stops when $|\underline{\mathcal{W}}_\ell| = 1$. GOBLIN outputs the arm in $\underline{\mathcal{W}}_\ell$ and reshapes it to get the $\widehat{\mathbf{x}}_*$ and $\widehat{\mathbf{z}}_*$. The full pseudocode is presented in Algorithm 1.

Algorithm 1 G-Optimal Design for Bilinear Bandits (GOBLIN) for single-task setting

1: **Input:** arm set $\mathcal{X}, \mathcal{Z}$, confidence δ, rank r of $\mathbf{\Theta}_*$, spectral bound S_r of $\mathbf{\Theta}_*$, $S, S_\ell^\perp := \frac{8 d_1 d_2 r}{\tau_\ell^E S_r^2} \log\left(\frac{d_1 + d_2}{\delta_\ell}\right)$, $\lambda, \lambda_\ell^\perp := \tau_{\ell-1}^G / 8(d_1 + d_2) r \log(1 + \frac{\tau_{\ell-1}^G}{\lambda})$. Let $p := d_1 d_2$, $k := (d_1 + d_2) r$.

2: Let $\underline{\mathcal{W}}_1 \leftarrow \underline{\mathcal{W}}, \ell \leftarrow 1, \tau_0^G := \log(4\ell^2 |\mathcal{X}|/\delta)$. Define $\mathbf{\Lambda}_\ell$ as in (6), $B_*^\ell := (8\sqrt{\lambda} S + \sqrt{\lambda_\ell^\perp S_\ell^\perp})$.

3: Define a vectorized arm $\overline{\mathbf{w}} := [\mathbf{x}_{1:d_1}; \mathbf{z}_{1:d_2}]$ and $\overline{\mathbf{w}} \in \overline{\mathcal{W}}$. Let $\tau_\ell^E := \frac{\sqrt{8 d_1 d_2 r \log(4\ell^2 |\mathcal{W}|/\delta_\ell)}}{S_r}$. Let the E-optimal design be $\mathbf{b}_\ell^E := \arg\min_{\mathbf{b} \in \triangle_{\overline{\mathcal{W}}}} \|(\sum_{\overline{\mathbf{w}} \in \overline{\mathcal{W}}} \mathbf{b}_{\overline{\mathbf{w}}} \overline{\mathbf{w}} \overline{\mathbf{w}}^\top)^{-1}\|$.

4: **while** $|\underline{\mathcal{W}}_\ell| > 1$ **do**

5: $\quad$ $\epsilon_\ell = 2^{-\ell}, \delta_\ell = \delta/\ell^2$.

6: $\quad$ **(Stage 1:) Explore the Low-Rank Subspace**

7: $\quad$ Pull arm $\overline{\mathbf{w}} \in \overline{\mathcal{W}}$ exactly $\left\lceil \widehat{\mathbf{b}}_{\ell,\overline{\mathbf{w}}}^E \tau_\ell^E \right\rceil$ times and observe rewards r_t, for $t = 1, \ldots, \tau_\ell^E$.

8: $\quad$ Compute $\widehat{\mathbf{\Theta}}_\ell$ using (2).

9: $\quad$ **(Stage 2:) Reduction to low dimensional linear bandits**

10: $\quad$ Let the SVD of $\widehat{\mathbf{\Theta}}_\ell = \widehat{\mathbf{U}}_\ell \widehat{\mathbf{D}}_\ell \widehat{\mathbf{V}}_\ell^\top$. Rotate arms in active set $\underline{\mathcal{W}}_{\ell-1}$ to build $\underline{\mathcal{W}}_\ell$ following (4).

11: $\quad$ Let $\widehat{\mathbf{b}}_\ell^G := \arg\min_{\mathbf{b}_{\underline{\mathbf{w}}}} \max_{\underline{\mathbf{w}}, \underline{\mathbf{w}}' \in \underline{\mathcal{W}}_\ell} \|\underline{\mathbf{w}} - \underline{\mathbf{w}}'\|_{(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}} \underline{\mathbf{w}}^\top + \mathbf{\Lambda}_\ell/n)^{-1}}^2$.

12: $\quad$ Define $\rho^G(\mathcal{Y}(\underline{\mathcal{W}}_\ell)) := \min_{\mathbf{b}_{\underline{\mathbf{w}}}} \max_{\underline{\mathbf{w}}, \underline{\mathbf{w}}' \in \underline{\mathcal{W}}_\ell} \|\underline{\mathbf{w}} - \underline{\mathbf{w}}'\|_{(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}} \underline{\mathbf{w}}^\top + \mathbf{\Lambda}_\ell/n)^{-1}}^2$.

13: $\quad$ Set $\tau_\ell^G := \lceil \frac{64 B_*^\ell \rho^G(\mathcal{Y}(\mathcal{W}_\ell)) \log(4\ell^2 |\mathcal{W}|/\delta_\ell)}{\epsilon_\ell^2} \rceil$. Then pull arm $\underline{\mathbf{w}} \in \underline{\mathcal{W}}$ exactly $\left\lceil \widehat{\mathbf{b}}_{\ell,\underline{\mathbf{w}}}^G \tau_\ell^G \right\rceil$ times and construct the least squares estimator $\widehat{\boldsymbol{\theta}}_\ell$ using only the observations of this phase where $\widehat{\boldsymbol{\theta}}_\ell$ is defined in (7). Note that $\widehat{\boldsymbol{\theta}}_\ell$ is also rotated following (4).

14: $\quad$ Eliminate arms such that $\underline{\mathcal{W}}_{\ell+1} \leftarrow \underline{\mathcal{W}}_\ell \setminus \{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell : \max_{\underline{\mathbf{w}}' \in \underline{\mathcal{W}}_\ell} \langle \underline{\mathbf{w}}' - \underline{\mathbf{w}}, \widehat{\boldsymbol{\theta}}_\ell \rangle > 2\epsilon_\ell\}$

15: $\quad$ $\ell \leftarrow \ell + 1$

16: Output the arm in $\underline{\mathcal{W}}_\ell$ and reshape to get the $\widehat{\mathbf{x}}_*$ and $\widehat{\mathbf{z}}_*$

2.3 Sample Complexity Analysis of Single-Task GOBLIN

We now analyze the sample complexity of GOBLIN in the single-task setting through the following theorem.

5

Theorem 1. *(informal)* *With probability at least* $1 - \delta$, *GOBLIN returns the best arms* $\mathbf{x}_*$, $\mathbf{z}_*$, *and the number of samples used is bounded by* $\widetilde{O}\left(\frac{(d_1+d_2)r}{\Delta^2} + \frac{\sqrt{d_1 d_2 r}}{S_r}\right)$.

Discussion 1. In Theorem 1 the first quantity is the number of samples needed to identify the best arms $\mathbf{x}_*$, $\mathbf{z}_*$ while the second quantity is the number of samples to learn $\mathbf{\Theta}_*$ (which is required to find the best arms). Note that the magnitude of S_r would be free of d_1, d_2 since $\mathbf{\Theta}_*$ contains only r nonzero singular values and $\|\mathbf{\Theta}_*\| \leq 1$, and hence we assume that $S_r = \Theta(1/\sqrt{r})$ (Kang et al., 2022). So the sample complexity of single-task GOBLIN scales as $\widetilde{O}(\frac{(d_1+d_2)r}{\Delta^2})$. However, if one runs RAGE (Fiez et al., 2019) on the arms in $\mathcal{X}, \mathcal{Z}$ then the sample complexity will scale as $\widetilde{O}(\frac{d_1 d_2}{\Delta^2})$.

Proof (Overview) of Theorem 1: Step 1 (Subspace estimation in high dimension): We denote the vectorized arms in high dimension as $\overline{\mathbf{w}} \in \overline{\mathcal{W}}$. We run the E-optimal design to sample the arms in $\overline{\mathcal{W}}$. Note that this E-optimal design satisfies the distribution assumption of Kang et al. (2022) which enables us to apply the Lemma 3 in Appendix A.1. This leads to $\|\widehat{\mathbf{\Theta}}_\ell - \mathbf{\Theta}_*\|_F^2 \leq \frac{C_1 d_1 d_2 r \log(2(d_1+d_2)/\delta)}{\tau_\ell^E}$ for some $C_1 > 0$. Also, note that in the first stage of the ℓ-th phase by setting $\tau_\ell^E = \frac{\sqrt{8 d_1 d_2 r \log(4\ell^2 |\mathcal{W}|/\delta_\ell)}}{S_r}$ and sampling each arm $\overline{\mathbf{w}} \in \overline{\mathcal{W}}$ exactly $\lceil \widehat{\mathbf{b}}_{\ell,\overline{\mathbf{w}}}^E \tau_\ell^E \rceil$ times we are guaranteed that $\|\boldsymbol{\theta}_{k+1:p}^*\|_2 = O(d_1 d_2 r / \tau_\ell^E)$. Summing up over $\ell = 1$ to $\lceil \log_2\left(4\Delta^{-1}\right) \rceil$ we get that the total sample complexity of the first stage is bounded by $\widetilde{O}(\sqrt{d_1 d_2 r}/S_r)$.

Step 2 (Effective dimension for rotated arms): We rotate the arms $\overline{\mathbf{w}} \in \overline{\mathcal{W}}$ in high dimension to get the rotated arms $\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell$ in step 10 of Algorithm 1. Then we show that the effective dimension of $\underline{\mathbf{w}}$ scales $8k \log\left(1 + \tau_{\ell-1}^G/\lambda\right)$ when $\lambda_\ell^\perp = \frac{\tau_{\ell-1}^G}{8k \log\left(1+\tau_{\ell-1}^G/\lambda\right)}$ in Lemma 7 of Appendix A.4. Note that this requires a different proof technique than Valko et al. (2014) where the budget n is given apriori and effective dimension scales with $\log(n)$. This step also diverges from the pure exploration proof technique of Fiez et al. (2019); Katz-Samuels et al. (2020) as there is no parameter $\lambda_\ell^\perp$ to control during phase ℓ, and the effective dimensions in those papers do not depend on phase length.

Step 3 (Bounded Support): For any phase ℓ, we can show that $1 \leq \rho^G(\mathcal{Y}(\underline{\mathcal{W}}_\ell)) \leq p/\gamma_{\mathcal{Y}}^2$ where, $\gamma_{\mathcal{Y}} = \max\{c > 0 : c\mathcal{Y} \subset \text{conv}(\underline{\mathcal{W}} \cup -\underline{\mathcal{W}})\}$ is the gauge norm of $\mathcal{Y}$ (Rockafellar, 2015). Note that this is a worst-case dependence when $\rho^G(\mathcal{Y}(\underline{\mathcal{W}}_\ell))$ scales with p. Substituting this value of $\rho^G(\mathcal{Y}(\underline{\mathcal{W}}_\ell))$ in the definition of $\lambda_\ell^\perp$ we can show that $\mathbf{\Lambda}_\ell$ does not depend on $\underline{\mathbf{w}}$ or $\mathbf{y} = \underline{\mathbf{w}} - \underline{\mathbf{w}}'$. Then following Theorem 21.1 in Lattimore and Szepesvári (2020) we can show that the G-optimal design $\widehat{\mathbf{b}}_\ell^G$ is equivalent to D-optimal design $\widehat{\mathbf{b}}_\ell^D = \arg\max_{\mathbf{b}} \log \frac{\left| \sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}} \, \underline{\mathbf{w}}^\top + \mathbf{\Lambda}_\ell \right|}{|\mathbf{\Lambda}_\ell|}$. Then using Frank-Wolfe algorithm (Jamieson and Jain, 2022) we can show the support $\widehat{\mathbf{b}}_\ell^G$ or equivalently $\widehat{\mathbf{b}}_\ell^D$ is bounded by at most $\frac{8k \log(1+\tau_{\ell-1}^G/\lambda)(8k\log(1+\tau_{\ell-1}^G/\lambda)+1)}{2}$. This is shown in Lemma 9 (Appendix A.4).

Step 4 (Phase length and Elimination): Using the Lemma 9, concentration Lemma 5, and using the log determinant inequality in Lemma 7 and Proposition 1 (Appendix A.4) we show that the phase length in the second stage is given by $\tau_\ell^G = \lceil \frac{8 B_*^\ell \rho(\mathcal{Y}(\underline{\mathcal{W}}_\ell)) \log(2|\mathcal{W}|/\delta)}{(\mathbf{x}^\top(\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}^*))^2} \rceil$. This is discussed in Discussion 3 (Appendix A.4). We show in Lemma 10 (Appendix A.4) that setting this phase length and sampling each active arm in $\underline{\mathcal{W}}_\ell$ exactly $\lceil \widehat{\mathbf{b}}_{\ell,\underline{\mathbf{w}}} \tau_\ell^G \rceil$ times results in the elimination of sub-optimal actions with high probability.

Step 5 (Total Samples): We first show that the total samples in the second phase are bounded by $O(\frac{k}{\gamma_{\mathcal{Y}}^2} \log(\frac{k \log_2(\Delta^{-1})|\mathcal{W}|}{\delta}) \lceil \log_2(\Delta^{-1}) \rceil)$ where the effective dimension $k = (d_1 + d_2)r$. Finally, we combine the total samples of phase ℓ as $(\tau_\ell^E + \tau_\ell^G)$. The final sample complexity is given by summing over all phases from $\ell = 1$ to $\lceil \log_2\left(4\Delta^{-1}\right) \rceil$. The claim of the theorem follows by noting $\widetilde{O}(k/\gamma_{\mathcal{Y}}^2) \leq \widetilde{O}(k/\Delta^2)$.

3   Multi-task Representation Learning

In this section, we extend GOBLIN to multi-task representation learning for the bilinear bandit setting. In the multi-task setting, we now have M tasks, where each task $m \in [M]$ has a reward model stated

6

in (1). The learning proceeds as follows: At each round $t = 1, 2, \cdots$, for each task $m \in [M]$, the learner selects a left and right action $\mathbf{x}_{m,t} \in \mathcal{X}$ and $\mathbf{z}_{m,t} \in \mathcal{Z}$. After the player commits the batch of actions for each task $\{\mathbf{x}_{m,t}, \mathbf{z}_{m,t} : m \in [M]\}$, it receives the batch of rewards $\{r_{m,t} : m \in [M]\}$. Finally recall that the goal is to identify the optimal left and right arms $\mathbf{x}_{m,*}, \mathbf{z}_{m,*}$ for each task m with a minimum number of samples. We now state the following assumptions to enable representation learning across tasks.

Assumption 1. *(Low-rank Tasks) We assume that the hidden parameter $\boldsymbol{\Theta}_{m,*}$ for all the $m \in [M]$ have a decomposition $\boldsymbol{\Theta}_{m,*} = \mathbf{B}_1 \mathbf{S}_{m,*} \mathbf{B}_2^\top$ and each $\mathbf{S}_{m,*}$ has rank r.*

This is similar to the assumptions in Yang et al. (2020, 2022); Du et al. (2023) ensuring the feature extractors are shared across tasks in the bilinear bandit setting.

Assumption 2. *(Diverse Tasks) We assume that $\sigma_{\min}(\frac{1}{M} \sum_{m=1}^{M} \boldsymbol{\Theta}_{m,*}) \geq \frac{c_0}{S_r}$, for some $c_0 > 0$, S_r is the r-th largest singular value of $\boldsymbol{\Theta}_{m,*}$ and $\sigma_{\min}(\mathbf{A})$ denotes the minimum eigenvalue of matrix $\mathbf{A}$.*

This assumption is similar to the diverse tasks assumption of Yang et al. (2020, 2022); Tripuraneni et al. (2021); Du et al. (2023) and ensures the possibility of recovering the feature extractors $\mathbf{B}_1$ and $\mathbf{B}_2$ shared across tasks.

Our extension of GOBLIN to the multi-task setting is now a phase-based, *three-stage* arm elimination algorithm. In GOBLIN each phase $\ell = 1, 2, \ldots$ consists of three stages; the stage for estimation of feature extractors $\mathbf{B}_1, \mathbf{B}_2$, which runs for τ_ℓ^E rounds, the stage for estimation of $\mathbf{S}_{m,*}$ which runs for $\sum_m \widetilde{\tau}_{m,\ell}^E$ rounds, and a stage of pure exploration with rotated arms that runs for $\sum_m \tau_{m,\ell}^G$ rounds. We will define $\tau_{m,\ell}^E$ in Section 3.1, $\widetilde{\tau}_{m,\ell}^E$ in Section 3.2, while the rotated arms and $\tau_{m,\ell}^G$ are defined in Section 3.3. At the end of every phase, GOBLIN eliminates sub-optimal arms to build the active set for the next phase and stops when only the optimal left and right arms are remaining. Now we discuss the individual stages that occur at every phase $\ell = 1, 2, \ldots$ for multi-task GOBLIN.

3.1 Estimating Feature Extractors $\mathbf{B}_1$ and $\mathbf{B}_2$ (Stage 1 of Phase ℓ)

In the first stage of phase ℓ, GOBLIN leverages the batch of rewards $\{r_{m,t} : m \in [M]\}$ at every round t from M tasks to learn the feature extractors $\mathbf{B}_1$ and $\mathbf{B}_2$. To do this, GOBLIN first vectorizes the $\mathbf{x} \in \mathcal{X}, \mathbf{z} \in \mathcal{Z}$ into a new vector $\overline{\mathbf{w}} = [\mathbf{x}_{1:d_1}; \mathbf{z}_{1:d_2}] \in \overline{\mathcal{W}}_m$ and then solves the E-optimal design in step 3 of Algorithm 2. Similar to the single-task setting (Section 2) GOBLIN samples each $\overline{\mathbf{w}} \in \overline{\mathcal{W}}_m$ for $\lceil \tau_\ell^E \mathbf{b}_{\ell, \overline{\mathbf{w}}}^E \rceil$ times for each task m, where $\tau_\ell^E = \widetilde{O}(\sqrt{d_1 d_2 r} / S_r)$ and $\mathbf{b}_{\ell, \overline{\mathbf{w}}}^E$ is the solution to E-optimal design on $\overline{\mathbf{w}}$. Let the sampled arms for each task m at round s be denoted by $\mathbf{x}_{m,s}, \mathbf{z}_{m,s}$ which is obtained after reshaping $\overline{\mathbf{w}}_s$. Then it builds the estimator $\widehat{\mathbf{Z}}_\ell$ as follows:

$$\widehat{\mathbf{Z}}_\ell = \underset{\boldsymbol{\Theta} \in \mathbb{R}^{d_1 \times d_2}}{\arg\min}\ L_\ell(\boldsymbol{\Theta}) + \gamma_\ell \|\boldsymbol{\Theta}\|_{\text{nuc}},$$

$$L_\ell(\boldsymbol{\Theta}) = \langle \boldsymbol{\Theta}, \boldsymbol{\Theta} \rangle - \frac{2}{M \tau_\ell^E} \sum_{m=1}^{M} \sum_{s=1}^{\tau_\ell^E} \langle \widetilde{\psi}_\nu(r_{m,s} \cdot Q(\mathbf{x}_{m,s} \mathbf{z}_{m,s}^\top)), \boldsymbol{\Theta} \rangle \tag{8}$$

Then it performs SVD decomposition on $\widehat{\mathbf{Z}}_\ell$, and let $\widehat{\mathbf{B}}_1, \widehat{\mathbf{B}}_2$ be the top-k_1 and top-k_2 left and right singular vectors of $\widehat{\mathbf{Z}}_\ell$ respectively. These are the estimation of the feature extractors $\mathbf{B}_1$ and $\mathbf{B}_2$.

3.2 Estimating Hidden Parameter $\mathbf{S}_{m,*}$ per Task (Stage 2 of phase ℓ)

In the second stage of phase ℓ, the goal is to recover the hidden parameter $\mathbf{S}_{m,*}$ for each task m. GOBLIN proceeds as follows: First, let $\widetilde{\mathbf{g}}_m = \mathbf{x}^\top \widehat{\mathbf{B}}_{1,\ell}$ and $\widetilde{\mathbf{v}}_m = \mathbf{z}^\top \widehat{\mathbf{B}}_{2,\ell}$ be the latent left and right arm respectively for each m. Then GOBLIN defines the vector $\widetilde{\mathbf{w}} = [\widetilde{\mathbf{g}}_m; \widetilde{\mathbf{v}}_m] \in \widetilde{\mathcal{W}}_m$ and then solves the E-optimal design in step 11 of Algorithm 2. It then samples for each task m, the latent arm $\widetilde{\mathbf{w}} \in \widetilde{\mathcal{W}}_m$ for $\lceil \widetilde{\tau}_{m,\ell}^E \widetilde{\mathbf{b}}_{m,\ell,\widetilde{\mathbf{w}}}^E \rceil$ times, where $\widetilde{\tau}_{m,\ell}^E := \widetilde{O}(\sqrt{k_1 k_2 r} / S_r)$ and $\widetilde{\mathbf{b}}_{m,\ell,\widetilde{\mathbf{w}}}^E$ is the solution to

7

E-optimal design on $\widetilde{\mathbf{w}}$. Then it builds estimator $\widehat{\mathbf{S}}_{m,\ell}$ for each task m in step 12 as follows:

$$\widehat{\mathbf{S}}_{m,\ell} = \underset{\boldsymbol{\Theta} \in \mathbb{R}^{k_1 \times k_2}}{\arg \min} \ L'_\ell(\boldsymbol{\Theta}) + \gamma_\ell \|\boldsymbol{\Theta}\|_{\text{nuc}},$$

$$L'_\ell(\boldsymbol{\Theta}) = \langle \boldsymbol{\Theta}, \boldsymbol{\Theta} \rangle - \frac{2}{\widetilde{\tau}^E_{m,\ell}} \sum_{s=1}^{\widetilde{\tau}^E_{m,\ell}} \langle \widetilde{\psi}_\nu(r_{m,s} \cdot Q(\widetilde{\mathbf{g}}_{m,s} \widetilde{\mathbf{v}}^\top_{m,s})), \boldsymbol{\Theta} \rangle \tag{9}$$

Once GOBLIN recovers the $\widehat{\mathbf{S}}_{m,\ell}$ for each task m it has reduced the $d_1 d_2$ bilinear bandit to a $k_1 k_2$ dimension bilinear bandit where the left and right arms are $\widetilde{\mathbf{g}}_m \in \mathcal{G}_m, \widetilde{\mathbf{v}}_m \in \mathcal{V}_m$ respectively.

3.3   Optimal Design for Rotated Arms per Task (Stage 3 of phase ℓ)

In the third stage of phase ℓ, similar to Algorithm 1, the multi-task GOBLIN defines the rotated arm set $\underline{\mathcal{G}}_m, \underline{\mathcal{V}}_m$ for each task m for these $k_1 k_2$ dimensional bilinear bandits. Let the SVD of $\widehat{\mathbf{S}}_{m,\ell} = \widehat{\mathbf{U}}_{m,\ell} \widehat{\mathbf{D}}_{m,\ell} \widehat{\mathbf{V}}^\top_{m,\ell}$. Define $\widehat{\mathbf{H}}_{m,\ell} = [\widehat{\mathbf{U}}_{m,\ell} \widehat{\mathbf{U}}^\perp_{m,\ell}]^\top \widehat{\mathbf{S}}_{m,\ell} [\widehat{\mathbf{V}}_{m,\ell} \widehat{\mathbf{V}}^\perp_{m,\ell}]$. Then define the vectorized arm set so that the last $(k_1 - r) \cdot (k_2 - r)$ components are from the complementary subspaces as follows:

$$\underline{\mathcal{W}}_{m,\ell} = \big\{ \big[\mathbf{vec}\left(\widetilde{\mathbf{g}}_{m,1:r} \widetilde{\mathbf{v}}^\top_{m,1:r}\right); \mathbf{vec}\left(\widetilde{\mathbf{g}}_{m,r+1:k_1} \widetilde{\mathbf{v}}^\top_{m,1:r}\right); \mathbf{vec}\left(\widetilde{\mathbf{g}}_{m,1:r} \widetilde{\mathbf{v}}^\top_{m,r+1:k_2}\right);$$

$$\mathbf{vec}\left(\widetilde{\mathbf{g}}_{m,r+1:k_1} \widetilde{\mathbf{v}}^\top_{m,r+1:k_2}\right) \big] \big\}$$

$$\widehat{\boldsymbol{\theta}}_{m,\ell,1:k} = [\mathbf{vec}(\widehat{\mathbf{H}}_{m,\ell,1:r,1:r}); \mathbf{vec}(\widehat{\mathbf{H}}_{m,\ell,r+1:k_1,1:r}); \mathbf{vec}(\widehat{\mathbf{H}}_{m,\ell,1:r,r+1:k_2})],$$

$$\boldsymbol{\theta}_{\ell,k+1:p} = \mathbf{vec}(\widehat{\mathbf{H}}_{m,\ell,r+1:k_1,r+1:k_2}). \tag{10}$$

This is shown in step 14 of Algorithm 2. Now we proceed similarly to Section 2.2. We construct a per-task optimal design for the rotated arm set $\underline{\mathcal{V}}_m, \underline{\mathcal{G}}_m$ and define the $\underline{\mathbf{w}} = [\widetilde{\mathbf{g}}_{m,1:d_1}; \widetilde{\mathbf{v}}_{m,1:d_2}]$ and $\widetilde{\mathbf{w}} \in \widetilde{\mathcal{W}}_m$ where $\widetilde{\mathbf{g}}_m \in \underline{\mathcal{G}}_m$ and $\widetilde{\mathbf{v}}_m \in \underline{\mathcal{V}}_m$ respectively. Following (5) we know that to minimize the sample complexity for the m-th bilinear bandit we need to sample according to G-optimal design

$$\widehat{\mathbf{b}}^G_{m,\ell} = \underset{\mathbf{b}_{m,\underline{\mathbf{w}}}}{\arg \min} \ \underset{\underline{\mathbf{w}}, \underline{\mathbf{w}}' \in \underline{\mathcal{W}}_{m,\ell}}{\max} \ \|\underline{\mathbf{w}} - \underline{\mathbf{w}}'\|^2_{(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_m} \mathbf{b}_{m,\underline{\mathbf{w}}} \underline{\mathbf{w}} \ \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_{m,\ell}/n)^{-1}} \tag{11}$$

Then GOBLIN runs G-optimal design on the arm set $\underline{\mathcal{W}}_\ell$ following the (11) and then samples each $\underline{\mathbf{w}} \in \underline{\mathcal{W}}_{m,\ell}$ for $\lceil \tau^G_{m,\ell} \widehat{\mathbf{b}}^G_{m,\ell,\underline{\mathbf{w}}} \rceil$ times where $\widehat{\mathbf{b}}^G_{m,\ell,\underline{\mathbf{w}}}$ is the solution to the G-optimal design, and τ^G_ℓ is defined in step 17 of Algorithm 2. So the total length of phase ℓ, combining stages 1, 2 and 3 is $(\tau^E_\ell + \sum_m \widetilde{\tau}^E_{m,\ell} + \sum_m \tau^G_{m,\ell})$ rounds. Observe that the stage 1 and 2 design is on the whole arm set $\overline{\mathcal{W}}, \widetilde{\mathcal{W}}_m$ whereas the stage 3 design is on the refined active set $\underline{\mathcal{W}}_{m,\ell}$. Let at the stage 3 of ℓ-th phase the actions sampled be denoted by the matrix $\underline{\mathbf{W}}_{m,\ell} \in \mathbb{R}^{\tau^G_{m,\ell} \times k_1 k_2}$ and observed rewards $\mathbf{r}_m \in \mathbb{R}^{\tau^G_{m,\ell} \times k_1 k_2}$. Define the positive diagonal matrix $\boldsymbol{\Lambda}_{m,\ell}$ according to (6) but set $p = k_1 k_2$ and $k = (k_1 + k_2)r$. Then similar to Section 2.2 we can build for each task m only from the observations from this phase

$$\widehat{\boldsymbol{\theta}}_{m,\ell} = \arg \min_{\boldsymbol{\theta}} \tfrac{1}{2} \|\underline{\mathbf{W}}_{m,\ell} \boldsymbol{\theta} - \mathbf{r}_m\|^2_2 + \tfrac{1}{2} \|\boldsymbol{\theta}\|^2_{\boldsymbol{\Lambda}_{m,\ell}} \tag{12}$$

Finally GOBLIN eliminates the sub-optimal arms using the estimator $\widehat{\boldsymbol{\theta}}_{m,\ell}$ to build the next phase active set $\underline{\mathcal{W}}_{m,\ell}$ and stops when $|\underline{\mathcal{W}}_{m,\ell}| = 1$. The full pseudo-code is given in Algorithm 2.

3.4   Sample Complexity analysis of Multi-task GOBLIN

We now present the sample complexity of GOBLIN for the multi-task setting.

Theorem 2. *(informal)* *With probability at least $1 - \delta$, GOBLIN returns the best arms $\mathbf{x}_{m,*}$, $\mathbf{z}_{m,*}$ for each task m, and the total number of samples is bounded by* $\widetilde{O}\left(\frac{M(k_1 + k_2)r}{\Delta^2} + \frac{M\sqrt{k_1 k_2 r}}{S_r} + \frac{\sqrt{d_1 d_2 r}}{S_r} \right)$.

Discussion 2. In Theorem 2 the first quantity is the sample complexity to identify the best arms $\mathbf{x}_{m,*}$, $\mathbf{z}_{m,*}$ and the second quantity is the number of samples to learn $\mathbf{S}_{m,*}$ for each task m. This is required to rotate the arms to reach the effective dimension of $(k_1 + k_2)r$. Finally, the third quantity is the number of samples needed to learn $\boldsymbol{\Theta}_{m,*}$ (which in turn is used to estimate the feature extractors $\mathbf{B}_1$ and $\mathbf{B}_2$ to learn the $\mathbf{S}_{m,*}$). Again we assume that $S_r = \Theta(1/\sqrt{r})$ (Kang et al., 2022). So the sample complexity of multi-task GOBLIN scales as $\widetilde{O}(M(k_1 + k_2)r/\Delta^2)$. However, if one runs DouExpDes (Du et al., 2023) then the sample complexity will scale as $\widetilde{O}(M(k_1 k_2)/\Delta^2)$ which is worse than GOBLINwhen $r \ll k_1$ or k_2.

8

Algorithm 2 G-Optimal Design for Bilinear Bandits (GOBLIN) for multi-task setting

---

1: Input: arm set $\mathcal{X}, \mathcal{Z}$, confidence δ, rank r of $\boldsymbol{\Theta}_*$, spectral bound S_r of $\boldsymbol{\Theta}_*$, $S, S_{m,\ell}^\perp = \frac{8k_1k_2r}{\widetilde{\tau}_{m,\ell}^E S_r^2}\log(\frac{k_1+k_2}{\delta_\ell}), \lambda, \lambda_{m,\ell}^\perp = \frac{\tau_{m,\ell-1}^G}{(8(k_1+k_2)r\log(1+\tau_{m,\ell-1}^G/\lambda))}$. Let $p = k_1k_2$, $k = (k_1+k_2)r$.

2: Let $\underline{\mathcal{W}}_{m,1} \leftarrow \underline{\mathcal{W}}_m, \ell \leftarrow 1, \tau_0^G = \log(4\ell^2|\mathcal{X}|/\delta)$. Define $\boldsymbol{\Lambda}_{m,\ell}$ as in (6), $B_{m,*}^\ell := (8\sqrt{\lambda}S + \sqrt{\lambda_{m,\ell}^\perp S_{m,\ell}^\perp})$

3: Define arm $\overline{\mathbf{w}} = [\mathbf{x}_{1:d_1}; \mathbf{z}_{1:d_2}]$ and $\overline{\mathbf{w}} \in \overline{\mathcal{W}}_m$. Let $\tau_\ell^E = \frac{\sqrt{8d_1d_2r\log(4\ell^2|\mathcal{W}|/\delta_\ell)}}{S_r}$. Let E-optimal design be $\mathbf{b}_\ell^E = \arg\min_{\mathbf{b}\in\triangle_{\overline{\mathcal{W}}}}\|(\sum_{\overline{\mathbf{w}}\in\overline{\mathcal{W}}}\mathbf{b}_{\overline{\mathbf{w}}}\overline{\mathbf{w}}\,\overline{\mathbf{w}}^\top)^{-1}\|$.

4: **while** $\exists m \in [M], |\underline{\mathcal{W}}_{m,\ell}| > 1$ **do**

5:    $\epsilon_\ell = 2^{-\ell}, \delta_\ell = \delta/\ell^2$.

6:    **(Stage 1:) Explore the Low-Rank Subspace**

7:    Pull arm $\overline{\mathbf{w}} \in \overline{\mathcal{W}}$ exactly $\lceil\mathbf{b}_{\ell,\overline{\mathbf{w}}}^E\tau_\ell^E\rceil$ times for each task m and observe rewards $\{r_{m,t}\}_{t=1}^{\tau_\ell^E}$.

8:    Compute $\widehat{\mathbf{Z}}_\ell$ using (8).

9:    **(Stage 2:) Build $\widehat{\mathbf{S}}_{m,\ell}$ for each task m**

10:    Let $\widehat{\mathbf{B}}_{1,\ell}, \widehat{\mathbf{B}}_{2,\ell}$ be the top-k_1 left and top-k_2 right singular vectors of $\widehat{\mathbf{Z}}_\ell$ respectively. Build $\widetilde{\mathbf{g}}_m = \mathbf{x}^\top\widehat{\mathbf{B}}_{1,\ell}$ and $\widetilde{\mathbf{v}}_m = \mathbf{z}^\top\widehat{\mathbf{B}}_{2,\ell}$ for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{z} \in \mathcal{Z}$ for each m.

11:    Define a vectorized arm $\widetilde{\mathbf{w}} = [\widetilde{\mathbf{g}}_{m,1:k_1}; \widetilde{\mathbf{v}}_{m,1:k_2}]$ and $\widetilde{\mathbf{w}} \in \widetilde{\mathcal{W}}_m$ for each m. Let $\widetilde{\tau}_{m,\ell}^E = \frac{\sqrt{8k_1k_2r\log(4\ell^2|\mathcal{W}|/\delta_\ell)}}{S_r}$, and $\widetilde{\mathbf{b}}_{m,\ell}^E = \arg\min_{\mathbf{b}_m\in\triangle_{\widetilde{\mathcal{W}}_m}}\|(\sum_{\widetilde{\mathbf{w}}\in\widetilde{\mathcal{W}}_m}\mathbf{b}_{m,\widetilde{\mathbf{w}}}\widetilde{\mathbf{w}}\,\widetilde{\mathbf{w}}^\top)^{-1}\|$.

12:    Pull arm $\widetilde{\mathbf{w}} \in \widetilde{\mathcal{W}}_m$ exactly $\lceil\widetilde{\mathbf{b}}_{m,\ell,\widetilde{\mathbf{w}}}^E\widetilde{\tau}_{m,\ell}^E\rceil$ times and observe rewards $r_{m,t}$, for $t = 1, \ldots, \widetilde{\tau}_{m,\ell}^E$, for each task m. Then compute $\widehat{\mathbf{S}}_{m,\ell}$ using (9) for each m.

13:    **(Stage 3:) Reduction to low dimensional linear bandits for each task m**

14:    SVD of $\widehat{\mathbf{S}}_{m,\ell} = \widehat{\mathbf{U}}_{m,\ell}\widehat{\mathbf{D}}_{m,\ell}\widehat{\mathbf{V}}_{m,\ell}^\top$. Rotate arms in active set $\underline{\mathcal{W}}_{m,\ell-1}$ to build $\underline{\mathcal{W}}_{m,\ell}$ using (10).

15:    Let $\widehat{\mathbf{b}}_{m,\ell}^G = \arg\min_{\mathbf{b}_{m,\underline{\mathbf{w}}}}\max_{\underline{\mathbf{w}},\underline{\mathbf{w}}'\in\underline{\mathcal{W}}_{m,\ell}}\|\underline{\mathbf{w}} - \underline{\mathbf{w}}'\|_{(\sum_{\underline{\mathbf{w}}_m\in\underline{\mathcal{W}}_m}\mathbf{b}_{m,\underline{\mathbf{w}}}\underline{\mathbf{w}}_m\,\underline{\mathbf{w}}_m^\top+\boldsymbol{\Lambda}_{m,\ell}/n)^{-1}}^2$.

16:    Define $\rho^G(\mathcal{Y}(\underline{\mathcal{W}}_{m,\ell})) = \min_{\mathbf{b}_{m,\underline{\mathbf{w}}}}\max_{\underline{\mathbf{w}},\underline{\mathbf{w}}'\in\underline{\mathcal{W}}_{m,\ell}}\|\underline{\mathbf{w}} - \underline{\mathbf{w}}'\|_{(\sum_{\underline{\mathbf{w}}\in\underline{\mathcal{W}}_m}\mathbf{b}_{m,\underline{\mathbf{w}}}\underline{\mathbf{w}}\,\underline{\mathbf{w}}^\top+\frac{\boldsymbol{\Lambda}_{m,\ell}}{n})^{-1}}^2$.

17:    Set $\tau_{m,\ell}^G = \frac{64B_{m,*}^\ell\rho^G(\mathcal{Y}(\mathcal{W}_{m,\ell}))\log(4\ell^2|\mathcal{W}_m|/\delta_\ell)}{\epsilon_\ell^2}$. Then pull arm $\underline{\mathbf{w}} \in \underline{\mathcal{W}}_m$ for each task m exactly $\lceil\widehat{\mathbf{b}}_{m,\ell,\underline{\mathbf{w}}}\tau_{m,\ell}^G\rceil$ times and construct the least squares estimator $\widehat{\boldsymbol{\theta}}_{m,\ell}$ using only the observations of this phase where $\widehat{\boldsymbol{\theta}}_{m,\ell}$ is defined in (12).

18:    Eliminate arms such that $\underline{\mathcal{W}}_{m,\ell+1} \leftarrow \underline{\mathcal{W}}_{m,\ell}\setminus\{\underline{\mathbf{w}}_m \in \underline{\mathcal{W}}_{m,\ell} : \max_{\underline{\mathbf{w}}_m'\in\underline{\mathcal{W}}_{m,\ell}}\langle\underline{\mathbf{w}}_m' - \underline{\mathbf{w}}_m, \widehat{\boldsymbol{\theta}}_{m,\ell}\rangle > 2\epsilon_{m,\ell}\}$

19:    $\ell \leftarrow \ell + 1$

20: Output the arm in $\underline{\mathcal{W}}_{m,\ell}$ and reshape to get the $\widehat{\mathbf{x}}_{m,*}$ and $\widehat{\mathbf{z}}_{m,*}$ for each task m.

Proof (Overview) of Theorem 2: Step 1 (Subspace estimation in high dimension): The first steps diverge from the proof technique of Theorem 1. We now build the average estimator $\widehat{\mathbf{Z}}_\ell$ to estimate the quantity $\mathbf{Z}_* = \frac{1}{M}\sum_{m=1}^M\boldsymbol{\Theta}_{*,m}$ using (8). This requires us to modify the Lemma 3 in Appendix A.1 and apply Stein's lemma (Lemma 1) to get a bound of $\|\widehat{\mathbf{Z}}_\ell - \mathbf{Z}_*\|_F^2 \leq \frac{C_1d_1d_2r\log(2(d_1+d_2)/\delta)}{\tau_\ell^E}$ for some $C_1 > 0$. This is shown in Lemma 12 in Appendix A.6. Summing up over $\ell = 1$ to $\lceil\log_2(4\Delta^{-1})\rceil$ we get that the total samples complexity of the first stage is bounded by $\widetilde{O}(\sqrt{d_1d_2r}/S_r)$.

Step 2 (Estimation of left and right feature extractors): Now using the estimator in (8) we get a good estimation of the feature extractors $\mathbf{B}_1$ and $\mathbf{B}_2$. Let $\widehat{\mathbf{B}}_{1,\ell}, \widehat{\mathbf{B}}_{2,\ell}$ be the top-k_1 left and top-k_2 right singular vectors of $\widehat{\mathbf{Z}}_\ell$ respectively. Then using the Davis-Kahan $\sin\theta$ Theorem (Bhatia, 2013) in Lemma 14, 15 (Appendix A.6) we have $\|(\widehat{\mathbf{B}}_{1,\ell}^\perp)^\top\mathbf{B}_1\|, \|(\widehat{\mathbf{B}}_{2,\ell}^\perp)^\top\mathbf{B}_2\| \leq \widetilde{O}(\sqrt{(d_1+d_2)r/M\tau_\ell^E})$.

Step 3 (Estimation of $\widehat{\mathbf{S}}_{m,\ell}$ in low dimension): Now we estimate the quantity $\widehat{\mathbf{S}}_{m,\ell} \in \mathbb{R}^{k_1\times k_2}$ for each task m. To do this we first build the latent arms $\widetilde{\mathbf{g}}_m = \mathbf{x}^\top\widehat{\mathbf{U}}_\ell$ and $\widetilde{\mathbf{v}}_m = \mathbf{z}^\top\widehat{\mathbf{V}}_\ell$ for all $\mathbf{x} \in \mathcal{X}$

9

and $\mathbf{z} \in \mathcal{Z}$ for each m, and sample them following the E-optimal design in step 12 of Algorithm 2. We also show in Lemma 16 (Appendix A.6) that $\sigma_{\min}(\sum_{\widetilde{\mathbf{w}} \in \widetilde{\mathcal{W}}} \mathbf{b}_{\widetilde{\mathbf{w}}} \widetilde{\mathbf{w}} \widetilde{\mathbf{w}}^\top) > 0$ which enables us to sample following E-optimal design. Then use the estimator in (9). Then in Lemma 19 we show that $\|\widehat{\mathbf{S}}_{m,\ell} - \mu^* \mathbf{S}_{m,*}\|_F^2 \le C_1 k_1 k_2 r \log\left(\frac{2(k_1+k_2)}{\delta_\ell}\right)/\tau_{m,\ell}^E$ holds with probability greater than $(1 - \delta)$. Also, note that in the second phase by setting $\widetilde{\tau}_{m,\ell}^E = \sqrt{8k_1 k_2 r \log(4\ell^2|\mathcal{W}|/\delta_\ell)}/S_r$ and sampling each arm $\overline{\mathbf{w}} \in \overline{\mathcal{W}}$ exactly $\lceil \widehat{\mathbf{b}}_{\ell,\overline{\mathbf{w}}}^E \widetilde{\tau}_{m,\ell}^E \rceil$ times we are guaranteed that $\|\boldsymbol{\theta}_{k+1:p}^*\|_2 = O(k_1 k_2 r / \widetilde{\tau}_{m,\ell}^E)$ in the ℓ-th phase. Summing up over $\ell = 1$ to $\lceil \log_2\left(4\Delta^{-1}\right) \rceil$ across each task M we get that the total samples complexity of the second stage is bounded by $\widetilde{O}(M\sqrt{k_1 k_2 r}/S_r)$.

Step 4 (Convert to $k_1 k_2$ bilinear bandits): Once GOBLIN recovers $\widehat{\mathbf{S}}_{m,\tau_\ell^E}$ it rotates the arm set following (10) to build $\underline{\mathcal{W}}_m$ to get the $k_1 k_2$ bilinear bandits. The rest of the steps follow the same way as in steps $2, 3$ and 4 of proof of Theorem 1.

Step 5 (Total Samples): We show the total samples in the third phase are bounded by $O(\frac{k}{\gamma_{\mathcal{Y}}^2} \log(\frac{k \log_2(\Delta^{-1})|\mathcal{W}|}{\delta})\lceil\log_2(\Delta^{-1})\rceil)$ where the effective dimension $k = (k_1 + k_2)r$. The total samples of phase ℓ is given by $\tau_\ell^E + \sum_m (\widetilde{\tau}_{m,\ell}^E + \tau_{m,\ell}^G)$. Finally, we get the total sample complexity by summing over all phases from $\ell = 1$ to $\lceil \log_2\left(4\Delta^{-1}\right) \rceil$. The claim of the theorem follows by noting $\widetilde{O}(k/\gamma_{\mathcal{Y}}^2) \le \widetilde{O}(k/\Delta^2)$.

4 Experiments

In this section, we conduct proof-of-concept experiments on both single and multi-task bilinear bandits. In the single-task experiment, we compare against the state-of-the-art RAGE algorithm (Fiez et al., 2019). We show in Figure 1 (left) that GOBLIN requires fewer samples than the RAGE with an increasing number of arms. In the multi-task experiment, we compare against the state-of-the-art DouExpDes algorithm (Du et al., 2023). We show in Figure 1 (right) that GOBLIN requires fewer samples than DouExpDes with an increasing number of tasks. As experiments are not a central contribution, we defer a fuller description of the experimental set-up to Appendix A.8.
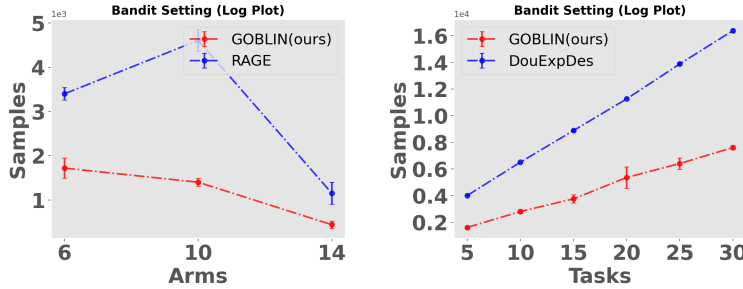
Figure 1: (Left) Single-task experiment: results show the number of samples required to identify the optimal action pair for differing numbers of actions. (Right) Multi-task experiment: results show the number of samples required to identify the optimal action pair for varying numbers of tasks.

5 Conclusions and Future Directions

In this paper, we formulated the first pure exploration multi-task representation learning problem. We introduce an algorithm, GOBLIN that achieves a sample complexity bound of $\widetilde{O}((d_1 + d_2)r/\Delta^2)$ which improves upon the $\widetilde{O}((d_1 d_2)/\Delta^2)$ sample complexity of RAGE (Fiez et al., 2019) in a single-task setting. We then extend GOBLIN for multi-task pure exploration bilinear bandit problems by learning latent features which enables sample complexity that scales as $\widetilde{O}(M(k_1 + k_2)r/\Delta^2)$ which improves over the $\widetilde{O}(M(k_1 k_2)/\Delta^2)$ sample complexity of DouExpDes (Du et al., 2023). Our analysis opens an exciting opportunity to analyze representation learning in the kernel and neural bandits (Zhu et al., 2021; Mason et al., 2021). We can leverage the fact that this type of optimal design does not require the arm set to be an ellipsoid (Du et al., 2023) which enables us to extend our analysis to non-linear representations.

10

References

Agarwal, D., Chen, B.-C., and Elango, P. (2009). Explore/exploit schemes for web content optimiza­tion. In *2009 Ninth IEEE International Conference on Data Mining*, pages 1–10. IEEE.

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828.

Bhatia, R. (2013). *Matrix analysis*, volume 169. Springer Science & Business Media.

Bragman, F. J., Tanno, R., Eaton-Rosen, Z., Li, W., Hawkes, D. J., Ourselin, S., Alexander, D. C., McClelland, J. R., and Cardoso, M. J. (2018). Uncertainty in multitask learning: joint representa­tions for probabilistic mr-only radiotherapy planning. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11*, pages 3–11. Springer.

Du, S. S., Hu, W., Kakade, S. M., Lee, J. D., and Lei, Q. (2020). Few-shot learning via learning the representation, provably. *arXiv preprint arXiv:2002.09434*.

Du, Y., Huang, L., and Sun, W. (2023). Multi-task representation learning for pure exploration in linear bandits. *arXiv preprint arXiv:2302.04441*.

Fiez, T., Jain, L., Jamieson, K. G., and Ratliff, L. (2019). Sequential experimental design for transductive linear bandits. *Advances in neural information processing systems*, 32.

Jamieson, K. and Jain, L. (2022). Interactive machine learning.

Jun, K.-S., Willett, R., Wright, S., and Nowak, R. (2019). Bilinear bandits with low-rank structure. In *International Conference on Machine Learning*, pages 3163–3172. PMLR.

Kang, Y., Hsieh, C.-J., and Lee, T. C. M. (2022). Efficient frameworks for generalized low-rank matrix bandit problems. *Advances in Neural Information Processing Systems*, 35:19971–19983.

Katz-Samuels, J., Jain, L., Jamieson, K. G., et al. (2020). An empirical process approach to the union bound: Practical algorithms for combinatorial and linear bandits. *Advances in Neural Information Processing Systems*, 33:10371–10382.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366.

Lattimore, T. and Szepesvári, C. (2020). *Bandit algorithms*. Cambridge University Press.

Li, J., Zhang, H., Zhang, L., Huang, X., and Zhang, L. (2014). Joint collaborative representation with multitask learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5923–5936.

Lu, Y., Meisami, A., and Tewari, A. (2021). Low-rank generalized linear bandit problems. In *International Conference on Artificial Intelligence and Statistics*, pages 460–468. PMLR.

Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., Peng, J., Chen, L., and Zeng, J. (2017). A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nature communications*, 8(1):573.

Mason, B., Camilleri, R., Mukherjee, S., Jamieson, K., Nowak, R., and Jain, L. (2021). Nearly optimal algorithms for level set estimation. *arXiv preprint arXiv:2111.01768*.

Maurer, A., Pontil, M., and Romera-Paredes, B. (2016). The benefit of multitask representation learning. *Journal of Machine Learning Research*, 17(81):1–32.

Minsker, S. (2018). Sub-gaussian estimators of the mean of a random matrix with heavy-tailed entries. *The Annals of Statistics*, 46(6A):2871–2903.

Mukherjee, S., Tripathy, A. S., and Nowak, R. (2022). Chernoff sampling for active testing and extension to active regression. In *International Conference on Artificial Intelligence and Statistics*, pages 7384–7432. PMLR.

11

Mukherjee, S., Xie, Q., Hanna, J., and Nowak, R. (2023). Speed: Experimental design for policy evaluation in linear heteroscedastic bandits. *arXiv preprint arXiv:2301.12357*.

Pukelsheim, F. (2006). *Optimal design of experiments*. SIAM.

Reyes, L. J. P., Oviedo, N. B., Camacho, E. C., and Calderon, J. M. (2021). Adaptable recommendation system for outfit selection with deep learning approach. *IFAC-PapersOnLine*, 54(13):605–610.

Rockafellar, R. (2015). Convex analysis. princeton landmarks in mathematics and physics.

Shamir, O. (2011). A variant of azuma's inequality for martingales with subgaussian tails. *arXiv preprint arXiv:1110.2392*.

Shen, Q., Han, S., Han, Y., and Chen, X. (2023). User review analysis of dating apps based on text mining. *Plos one*, 18(4):e0283896.

Soare, M., Lazaric, A., and Munos, R. (2014). Best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 27.

Stein, C., Diaconis, P., Holmes, S., and Reinert, G. (2004). Use of exchangeable pairs in the analysis of simulations. *Lecture Notes-Monograph Series*, pages 1–26.

Tripuraneni, N., Jin, C., and Jordan, M. (2021). Provable meta-learning of linear representations. In *International Conference on Machine Learning*, pages 10434–10443. PMLR.

Valko, M., Munos, R., Kveton, B., and Kocák, T. (2014). Spectral bandits for smooth graph functions. In *International Conference on Machine Learning*, pages 46–54. PMLR.

Yang, J., Hu, W., Lee, J. D., and Du, S. S. (2020). Impact of representation learning in linear bandits. *arXiv preprint arXiv:2010.06531*.

Yang, J., Lei, Q., Lee, J. D., and Du, S. S. (2022). Nearly minimax algorithms for linear bandits with shared representation. *arXiv preprint arXiv:2203.15664*.

Zhang, D., Shen, D., Initiative, A. D. N., et al. (2012). Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease. *NeuroImage*, 59(2):895–907.

Zhao, Y., Kosorok, M. R., and Zeng, D. (2009). Reinforcement learning design for cancer clinical trials. *Statistics in medicine*, 28(26):3294–3315.

Zhu, Y., Zhou, D., Jiang, R., Gu, Q., Willett, R., and Nowak, R. (2021). Pure exploration in kernel and neural bandits. *Advances in neural information processing systems*, 34:11618–11630.

12

A Appendix

A.1 Probability Tools and Previous Results

In this section, we state useful lemmas we use in our proofs and previous results.

Lemma 1. *(Generalized Stein's Lemma, ([Stein et al., 2004](#))) For a random variable X with continuously differentiable density function $\mathbf{p} : \mathbb{R}^d \to \mathbb{R}$, and any continuously differentiable function $f : \mathbb{R}^d \to \mathbb{R}$. Let $Q(\cdot)$ be a scoring function defined in Definition [1](#). If the expected values of both $\nabla f(X)$ and $f(X) \cdot Q(X)$ regarding the density p exist, then they are identical, i.e.*

$$\mathbb{E}[f(X) \cdot Q(X)] = \mathbb{E}[\nabla f(X)].$$

Lemma 2. *([Minsker, 2018](#)) Define $\|\mathbf{A}\|_{op}$ as the operator norm of $\mathbf{A}$. Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \in \mathbb{R}^{d_1 \times d_2}$ be a sequence of independent real random matrices, and assume that*

$$\sigma_n^2 \geq \max \left(\left\| \sum_{j=1}^n \mathbb{E} \left(\mathbf{Y}_j \mathbf{Y}_j^\top \right) \right\|_{op}, \left\| \sum_{j=1}^n \mathbb{E} \left(\mathbf{Y}_j^\top \mathbf{Y}_j \right) \right\|_{op} \right).$$

Then for any $t \in \mathbb{R}^+$ and $\nu \in \mathbb{R}^+$, it holds that,

$$P \left(\left\| \sum_{j=1}^n \widetilde{\psi}_\nu \left(\mathbf{Y}_j \right) - \sum_{j=1}^n \mathbb{E} \left(\mathbf{Y}_j \right) \right\|_{op} \geq t\sqrt{n} \right) \leq 2 \left(d_1 + d_2 \right) \exp \left(\nu t \sqrt{n} + \frac{\nu^2 \sigma_n^2}{2} \right)$$

Lemma 3. *(Restatement of Theorem 4.1 in [Kang et al.](#) ([2022](#))) For any low-rank linear model with samples $\mathbf{X}_1 \ldots, \mathbf{X}_{n_1}$ drawn from $\mathcal{X}$ according to $\mathcal{D}$ then for the optimal solution to the nuclear norm regularization problem in ([2](#)) with $\nu = \sqrt{2 \log \left(2 \left(d_1 + d_2 \right) / \delta \right) / \left(\left(4 + S_0^2 \right) M n_1 d_1 d_2 \right)}$ and*

$$\gamma_{n_1} = 4 \sqrt{\frac{2 \left(4 + S_0^2 \right) C d_1 d_2 \log \left(2 \left(d_1 + d_2 \right) / \delta \right)}{n_1}},$$

with probability at least $1 - \delta$ it holds that:

$$\left\| \widehat{\boldsymbol{\Theta}} - \mu^* \boldsymbol{\Theta}_* \right\|_F^2 \leq \frac{C_1 d_1 d_2 r \log \left(\frac{2(d_1 + d_2)}{\delta} \right)}{n_1},$$

for $C_1 = 36 \left(4 + S_0^2 \right) C$, $\|\mathbf{X}\|_F, \|\boldsymbol{\Theta}_*\|_F \leq S_0$, some nonzero constant μ^*, and $\mathbb{E} \left[\left(S^\mathbf{p}(\mathbf{X}) \right)_{ij}^2 \right] \leq C, \forall i, j$.

A.2 G-optimal design on rotated arms

Remark 1. (G-optimal design on rotated arms:) Using the concentration inequality in Proposition [1](#) we can show that for any arbitrary vector $\mathbf{y} \in \mathbb{R}^p$:

$$|\mathbf{y}^\top (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}_*)| \leq \|\mathbf{y}\|_{\mathbf{V}_\ell^{-1}} 2\sqrt{14 \log(2/\delta)} + \|\boldsymbol{\theta}_*\|_{\boldsymbol{\Lambda}_\ell} \leq \|\mathbf{y}\|_{\mathbf{V}_\ell^{-1}} \sqrt{8 B_*^\ell \log(2/\delta)}$$

where the co-variance matrix $\mathbf{V}_\ell := \sum_{s=1}^{\tau_\ell^E} \underline{\mathbf{w}}_s \underline{\mathbf{w}}_s^\top + \boldsymbol{\Lambda}_\ell$ and B_*^ℓ is defined in Proposition [1](#). Now we want this to hold for all $\mathbf{y} \in \mathcal{Y}^*(\underline{\mathcal{W}}_\ell)$, and so we need to union bound over $\mathcal{W} \supseteq \underline{\mathcal{W}}_\ell$ replacing δ with $\delta/|\mathcal{W}|$. Set the phase length $\tau_\ell^G := \left\lceil \frac{64 B_*^\ell \rho^G(\mathcal{Y}(\underline{\mathcal{W}}_\ell)) \log(4\ell^2 |\mathcal{W}|/\delta)}{\epsilon_\ell^2} \right\rceil$ where $\rho^G(\mathcal{Y}(\underline{\mathcal{W}}_\ell))$ is defined in step 14 of Algorithm [1](#).

Then for the allocation $2\lfloor \mathbf{b}_{\underline{\mathbf{w}}}^G \tau_\ell^G \rfloor$ for each $\mathbf{b}_{\underline{\mathbf{w}}} \in \mathcal{W}_\ell$, we have for each $\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell \backslash \underline{\mathbf{w}}_*$ that with probability at least $1 - \delta$,

$$(\underline{\mathbf{w}}_* - \underline{\mathbf{w}})^\top \widehat{\boldsymbol{\theta}}_\ell \geq (\underline{\mathbf{w}}_* - \underline{\mathbf{w}})^\top \boldsymbol{\theta}_* - \|\underline{\mathbf{w}}_* - \underline{\mathbf{w}}\|_{(\sum_{\underline{\mathbf{w}} \in \mathcal{W}} \lceil 2\tau_\ell^G \mathbf{b}_{\underline{\mathbf{w}}}^* \rceil \underline{\mathbf{w}} \, \underline{\mathbf{w}}^\top + 2\tau_\ell^G \boldsymbol{\Lambda}_\ell / \tau_\ell^G)^{-1}} \sqrt{8 B_*^\ell \log(2|\mathcal{W}|/\delta)}$$

13

since for every $\underline{\mathbf{w}}_* - \underline{\mathbf{w}} \in \mathcal{Y}^*(\mathcal{W})$ we have

$$(\underline{\mathbf{w}}_* - \underline{\mathbf{w}})^\top \big(2 \sum_{\underline{\mathbf{w}} \in \mathcal{W}} \lceil \tau_\ell^G \mathbf{b}_{\underline{\mathbf{w}}}^* \rceil \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + 2\tfrac{\tau_\ell^G \mathbf{\Lambda}_\ell}{\tau_\ell^G} \big)^{-1} (\underline{\mathbf{w}}_* - \underline{\mathbf{w}}) \leq \tfrac{1}{\tau_\ell^G} \|\underline{\mathbf{w}}_* - \underline{\mathbf{w}}\|^2_{(\sum_{\underline{\mathbf{w}} \in \mathcal{W}_\ell} \mathbf{b}_{\underline{\mathbf{w}}}^* \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \frac{\mathbf{\Lambda}_\ell}{\tau_\ell^G})^{-1}}$$

$$\leq \frac{((\underline{\mathbf{w}}_* - \underline{\mathbf{w}})^\top \boldsymbol{\theta}_*)^2}{\sqrt{8B_*^\ell \log(2|\mathcal{W}|/\delta)}}.$$

The last inequality follows by plugging in the value of τ_ℓ^G and $\rho^G(\mathcal{Y}(\underline{\mathcal{W}}_\ell))$. Hence to minimize the number of samples τ_ℓ^G in phase ℓ we can re-arrange the above equation to show that

$$\tau_\ell^G \geq \sqrt{8B_*^\ell \log(2|\mathcal{W}|/\delta)} \max_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell \setminus \underline{\mathbf{w}}_*} \frac{\|\underline{\mathbf{w}}_* - \underline{\mathbf{w}}\|^2_{(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell} \mathbf{b}_{\underline{\mathbf{w}}}^* \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \mathbf{\Lambda}/n)^{-1}}}{(\underline{\mathbf{w}}_* - \underline{\mathbf{w}})^\top \boldsymbol{\theta}_*}$$

Hence, to minimize the sample complexity for the bilinear setting we need to sample according to

$$\mathbf{b}_\ell^G = \arg\min_{\mathbf{b}} \max_{\underline{\mathbf{w}}} \frac{\|\underline{\mathbf{w}}_* - \underline{\mathbf{w}}\|^2_{(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \mathbf{\Lambda}/n)^{-1}}}{(\underline{\mathbf{w}}_* - \underline{\mathbf{w}})^\top \boldsymbol{\theta}_*} \tag{13}$$

However, note that we do know the identity of $\mathbf{w}_*$ or the gaps $(\underline{\mathbf{w}}_* - \underline{\mathbf{w}})^\top \boldsymbol{\theta}_*$. So we replace the gaps with a lower bound of $\epsilon = 2^{-t}$ and compare against every pair of arms $\underline{\mathbf{w}}$ and $\underline{\mathbf{w}}'$ as follows:

$$\mathbf{b}_\ell^G = \arg\min_{\mathbf{b}_{\underline{\mathbf{w}}}} \max_{\underline{\mathbf{w}}, \underline{\mathbf{w}}' \in \underline{\mathcal{W}}_\ell} \|\underline{\mathbf{w}} - \underline{\mathbf{w}}'\|^2_{(\sum_{\underline{\mathbf{w}} \in \mathcal{W}} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \mathbf{\Lambda}_\ell/n)^{-1}} \tag{14}$$

This is shown in step 12 of Algorithm 1.

A.3 Application of Stein's Lemma

We also present the following two definitions from Kang et al. (2022) to facilitate analysis via Stein's method:

Definition 1. (Score Function) Let $\mathbf{p} : \mathbb{R} \to \mathbb{R}$ be a univariate probability density function defined on $\mathbb{R}$. The score function $Q^{\mathbf{p}} : \mathbb{R} \to \mathbb{R}$ regarding density $\mathbf{p}(\cdot)$ is defined as:

$$Q^{\mathbf{p}}(x) = -\nabla_x \log(\mathbf{p}(x)) = -\nabla_x \mathbf{p}(x)/\mathbf{p}(x), \quad x \in \mathbb{R}.$$

In particular, for a random matrix with its entrywise probability density $\mathbf{p} = (p_{ij}) : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^{d_1 \times d_2}$, we define its score function $Q^{\mathbf{p}} = (Q_{ij}^{\mathbf{p}}) : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^{d_1 \times d_2}$ as $Q_{ij}^{\mathbf{p}}(x) = Q^{p_{ij}}(x)$ by applying the univariate score function to each entry of $\mathbf{p}$ independently.

Assumption 3. *The norm of true parameter $\mathbf{\Theta}^*$ and feature matrices in $\mathcal{X}$ is bounded: there exists $S \in \mathbb{R}^+$ such that for all arms $\mathbf{X} \in \mathcal{X}, \|\mathbf{X}\|_F, \|\mathbf{\Theta}^*\|_F \leq S_0$*

Assumption 4. *(Finite second-moment score)* *There exists a sampling distribution $\mathcal{D}$ over $\mathcal{X}$ such that for the random matrix $\mathbf{X}$ drawn from $\mathcal{D}$ with its associated density $\mathbf{p} : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^{d_1 \times d_2}$, we have $\mathbb{E}\left[(Q^{\mathbf{p}}(\mathbf{X}))_{ij}^2\right] \leq C, \forall i, j$*

Definition 2. Given a rectangular matrix $\mathbf{A} \in \mathbb{R}^{d_1 \times d_2}$, the (Hermitian) dilation $\mathcal{H} : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^{(d_1+d_2) \times (d_1+d_2)}$ is defined as:

$$\mathcal{H}(\mathbf{A}) = \begin{pmatrix} 0 & \mathbf{A} \\ \mathbf{A}^\top & 0 \end{pmatrix}$$

Definition 3. (The function $\widetilde{\psi}_\nu$) To explore the valid subspace of the parameter matrix $\mathbf{\Theta}_*$, we define a function $\psi : \mathbb{R} \to \mathbb{R}$ in (15). Let $\mathcal{H}(\cdot)$ be as defined in Definition 2. Then define $\widetilde{\psi}_\nu : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^{d_1 \times d_2}$ as $\widetilde{\psi}_\nu(\mathbf{A}) = \psi(\nu \mathcal{H}(\mathbf{A}))_{1:d_1,(d_1+1):(d_1+d_2)}/\nu$ for some parameter $\nu \in \mathbb{R}^+$

$$\psi(x) = \begin{cases} \log\left(1 + x + x^2/2\right), & x \geq 0 \\ -\log\left(1 - x + x^2/2\right), & x < 0 \end{cases} \tag{15}$$

14

A.4 Single-task Pure Exploration Proofs

Good Event: Define the good event $\mathcal{F}_\ell$ in phase ℓ that GOBLIN has a good estimate of $\boldsymbol{\Theta}_*$ as follows:

$$\mathcal{F}_\ell = \left\|\widehat{\boldsymbol{\Theta}}_\ell - \mu^* \boldsymbol{\Theta}_* \right\|_F^2 \leq \frac{C_1 d_1 d_2 r \log\left(\frac{2(d_1+d_2)}{\delta_\ell}\right)}{\tau_\ell^E}, \tag{16}$$

where, $C_1 = 36\left(4 + S_0^2\right)C$, $\|\mathbf{X}\|_F, \|\boldsymbol{\Theta}_*\|_F \leq S_0$, some nonzero constant μ^*, $\mathbb{E}\left[\left(S^{\mathbf{P}}(\mathbf{X})\right)_{ij}^2\right] \leq C, \forall i, j$, and $\widehat{\boldsymbol{\Theta}}_\ell$ is the estimate from (2). Then define the good event

$$\mathcal{F} := \bigcap_{\ell=1}^{\infty} \mathcal{F}_\ell. \tag{17}$$

Lemma 4. *The event $\mathcal{F}$ holds with probability greater than $(1 - \delta/2)$.*

Proof. From Lemma 3 we know the event $\mathcal{F}_\ell$ in (16) holds with probability $(1 - \delta_\ell)$. Taking a union bound over all phases $\ell \geq 1$ and recalling $\delta_\ell := \frac{\delta}{2\ell^2}$, we obtain

$$\mathbb{P}(\mathcal{F}) \geq 1 - \sum_{\ell=1}^{\infty} \mathbb{P}\left(\mathcal{F}_\ell^c\right)$$

$$\geq 1 - \sum_{\ell=1}^{\infty} \frac{\delta_\ell}{2}$$

$$= 1 - \sum_{\ell=1}^{\infty} \frac{\delta}{4\ell^2}$$

$$\geq 1 - \frac{\delta}{2}.$$

This concludes our proof. $\qquad\qquad\square$

Now we move to the second stage for the rotated arm set $\underline{\mathbf{w}} \in \underline{\mathcal{W}}$ and prove the following concentration event.

Lemma 5. *For any fixed $\underline{\mathbf{w}} \in \mathbb{R}^p$ and any $\delta > 0$, we have that if $\beta\left(\boldsymbol{\theta}_*, \delta\right) = 2\sqrt{14 \log(2/\delta)} + \|\boldsymbol{\theta}_*\|_{\boldsymbol{\Lambda}}$, then at time $\tau_{\ell-1} + 1$ (beginning of phase ℓ):*

$$\mathbb{P}\left(\left|\underline{\mathbf{w}}^\top \left(\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}_*\right)\right| \leq \|\underline{\mathbf{w}}\|_{\mathbf{V}_\ell^{-1}} \beta\left(\boldsymbol{\theta}_*, \delta\right)\right) \geq 1 - \delta$$

where, $\mathbf{V}_\ell := \sum_{s=\tau_{\ell-1}+1}^{\tau_\ell} \underline{\mathbf{w}}_s \underline{\mathbf{w}}_s^\top + \boldsymbol{\Lambda}_\ell$.

Proof. We follow the proof technique of Lemma 7 of Valko et al. (2014). Defining $\xi_\ell = \sum_{s=\tau_{\ell-1}+1}^{\tau_\ell} \underline{\mathbf{w}}_s \eta_s$, we have:

$$\left|\underline{\mathbf{w}}^\top \left(\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}_*\right)\right| \stackrel{(a)}{=} \left|\underline{\mathbf{w}}^\top \left(-\mathbf{V}_\ell^{-1} \boldsymbol{\Lambda} \boldsymbol{\theta}_* + \mathbf{V}_\ell^{-1} \xi_\ell\right)\right|$$

$$\stackrel{(b)}{\leq} \left|\underline{\mathbf{w}}^\top \mathbf{V}_\ell^{-1} \boldsymbol{\Lambda}_\ell \boldsymbol{\theta}_*\right| + \left|\underline{\mathbf{w}}^\top \mathbf{V}_\ell^{-1} \xi_\ell\right| \tag{18}$$

where (a) follows from Woodbury matrix identity and rearranging the terms, and (b) follows from the triangle inequality.

The first term in the right-hand side of (18) is bounded as:

$$\left|\underline{\mathbf{w}}^\top \mathbf{V}_\ell^{-1} \boldsymbol{\Lambda}_\ell \boldsymbol{\theta}_*\right| \leq \left\|\underline{\mathbf{w}}^\top \mathbf{V}_\ell^{-1} \boldsymbol{\Lambda}_\ell^{1/2}\right\| \left\|\boldsymbol{\Lambda}_\ell^{1/2} \boldsymbol{\theta}_*\right\|$$

$$\stackrel{(a)}{=} \|\boldsymbol{\theta}_*\|_{\boldsymbol{\Lambda}_\ell} \sqrt{\underline{\mathbf{w}}^\top \mathbf{V}_\ell^{-1} \boldsymbol{\Lambda}_\ell \mathbf{V}_\ell^{-1} \underline{\mathbf{w}}}$$

$$\leq \|\boldsymbol{\theta}_*\|_{\boldsymbol{\Lambda}_\ell} \sqrt{\underline{\mathbf{w}}^\top \mathbf{V}_\ell^{-1} \underline{\mathbf{w}}} = \|\boldsymbol{\theta}_*\|_{\boldsymbol{\Lambda}_\ell} \|\underline{\mathbf{w}}\|_{\mathbf{V}_\ell^{-1}}$$

15

where, (a) follows as $\|\boldsymbol{\theta}_*\|_{\boldsymbol{\Lambda}_\ell} = \sqrt{\boldsymbol{\theta}_*^\top \boldsymbol{\Lambda}_\ell \boldsymbol{\theta}_*} = \|\boldsymbol{\Lambda}_\ell^{1/2} \boldsymbol{\theta}_*\|$ and similarly for $\left\|\underline{\mathbf{w}}^\top \mathbf{V}_\ell^{-1} \boldsymbol{\Lambda}_\ell^{1/2}\right\|$. Now consider the second term in the r.h.s. of (18). We have:

$$\left|\underline{\mathbf{w}}^\top \mathbf{V}_\ell^{-1} \xi_\ell\right| = \left|\sum_{s=\tau_{\ell-1}+1}^{\tau_\ell} \left(\underline{\mathbf{w}}^\top \mathbf{V}_\ell^{-1} \underline{\mathbf{w}}_s\right) \eta_s\right|.$$

Now note that the arms $(\underline{\mathbf{w}}_s)$ selected by the algorithm during phase ℓ only depend on the proportion $\mathbf{b}_*^G$ (the G-optimal design) and do not depend on the rewards received during the phase $\ell - 1$. Thus, given $\mathcal{F}_{j-2}$, the sequence $(\underline{\mathbf{w}}_s)_{\tau_{\ell-1}+1 \le s < \tau_\ell}$ is deterministic. Consequently, one may use a variant of Azuma's inequality (Shamir (2011)) with a 1-sub Gaussian assumption:

$$\mathbb{P}\left(\left|\underline{\mathbf{w}}^\top \mathbf{V}_\ell^{-1} \xi_\ell\right|^2 \le 28 \times 2 \log(2/\delta) \times \underline{\mathbf{w}}^\top \mathbf{V}_\ell^{-1} \left(\sum_{s=\tau_{\ell-1}+1}^{\tau_\ell} \underline{\mathbf{w}}_s \underline{\mathbf{w}}_s^\top\right) \mathbf{V}_\ell^{-1} \underline{\mathbf{w}} \mid \mathcal{F}_{\ell-2}\right) \ge 1 - \delta,$$

from which we deduce:

$$\mathbb{P}\left(\left|\underline{\mathbf{w}}^\top \mathbf{V}_\ell^{-1} \xi_\ell\right|^2 \le 56 \underline{\mathbf{w}}^\top \mathbf{V}_\ell^{-1} \underline{\mathbf{w}} \log(2/\delta) \mid \mathcal{F}_{\ell-2}\right) \ge 1 - \delta,$$

since $\sum_{s=\tau_{\ell-1}+1}^{\tau_\ell} \underline{\mathbf{w}}_s \underline{\mathbf{w}}_s^\top \prec \mathbf{V}_\ell$. Thus:

$$\mathbb{P}\left(\left|\underline{\mathbf{w}}^\top \mathbf{V}_\ell^{-1} \xi_\ell\right| \le 2\|\underline{\mathbf{w}}\|_{\mathbf{V}_\ell^{-1}} \sqrt{14 \log(2/\delta)}\right) \ge 1 - \delta$$

Combining everything we get that

$$\mathbb{P}\left(\left|\underline{\mathbf{w}}^\top \left(\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}_*\right)\right| \le 2\sqrt{14 \log(2/\delta)} + \|\boldsymbol{\theta}_*\|_{\boldsymbol{\Lambda}_\ell}\right) \le 1 - \delta.$$

$\square$

We need to change Lemma 6 of Valko et al. (2014) in the following way so that the dependence on horizon n is replaced by $\tau_{\ell-1}^G$. Note that $\tau_{\ell-1}^G$ is the phase length in the $\ell - 1$-th phase and is determined before the start of phase $\tau_{\ell-1}^G$. Also, note that using the standard analysis of phase-based algorithms in Fiez et al. (2019); Lattimore and Szepesvári (2020) we do not re-use data between phases. First, we need the following support lemma from Valko et al. (2014).

Lemma 6. *(Restatement of Lemma 4 from Valko et al. (2014)) Let $\boldsymbol{\Lambda}_\ell = \mathrm{diag}\left(\lambda_1, \ldots, \lambda_\ell^\perp\right)$ be any diagonal matrix with strictly positive entries. Define $\mathbf{V}_\ell = \boldsymbol{\Lambda}_\ell + \sum_{s=\tau_{\ell-1}^G+1}^{\tau_\ell^G} \underline{\mathbf{w}}_s \underline{\mathbf{w}}_s^\top$. Then for any vectors $(\underline{\mathbf{w}}_s)_{\tau_{\ell-1}^G+1 \le s \le \tau_\ell^G}$, such that $\|\underline{\mathbf{w}}_s\|_2 \le 1$ for all rounds s such that $\tau_{\ell-1}^G + 1 \le s \le \tau_\ell^G$, we have that the determinant $|\mathbf{V}_\ell|$ is maximized when all $\underline{\mathbf{w}}_s$ are aligned with the axes.*

Lemma 7. *Let k be the effective dimension. Then*

$$\log \frac{|\mathbf{V}_\ell|}{|\boldsymbol{\Lambda}_\ell|} \le 8k \log\left(1 + \frac{\tau_{\ell-1}^G}{\lambda}\right)$$

when $\lambda_\ell^\perp = \frac{\tau_{\ell-1}^G}{k \log\left(1 + \tau_{\ell-1}^G/\lambda\right)}$.

Proof. We want to bound the determinant $|\mathbf{V}_\ell|$ under the coordinate constraints $\|\underline{\mathbf{w}}_t\|_2 \le 1$. Let:

$$M\left(\underline{\mathbf{w}}_1, \ldots, \underline{\mathbf{w}}_t\right) = \left|\boldsymbol{\Lambda}_\ell + \sum_{s=\tau_{\ell-1}^G+1}^{\tau_\ell^G} \underline{\mathbf{w}}_s \underline{\mathbf{w}}_s^\top\right|$$

From Lemma 6 we deduce that the maximum of M is reached when all $\underline{\mathbf{w}}_s$ are aligned with the axes. Let the number of samples of these axes-aligned $\underline{\mathbf{w}}_s$'s during the ℓ-th phase be denoted as

16

$t_1^\ell, t_2^\ell, \ldots, t_p^\ell$ such that $\sum_{i=1}^p t_i^\ell = \tau_\ell^G$. Then we can show that

$$M \stackrel{(a)}{=} \max_{\mathbf{w}_1,\ldots,\mathbf{w}_t;\underline{\mathbf{w}}_s \in \{\mathbf{e}_1,\ldots,\mathbf{e}_p\}} \left| \mathbf{\Lambda}_\ell + \sum_{s=\tau_{\ell-1}^G+1}^{\tau_\ell^G} \underline{\mathbf{w}}_s \underline{\mathbf{w}}_s^\top \right|$$

$$\stackrel{(b)}{=} \max_{t_1^\ell,\ldots,t_p^\ell, \text{ positive integers, } \sum_{i=1}^p t_i^\ell = \tau_\ell^G} \left| \mathrm{diag}\left(\lambda_i + t_i^\ell\right) \right|$$

$$\stackrel{(c)}{\leq} \max_{t_1^\ell,\ldots,t_p^\ell, \text{ positive integers, } \sum_{i=1}^p t_i^\ell = \tau_\ell^G} \prod_{i=1}^p \left(\lambda_i + t_i^\ell\right)$$

where, (a) follows from Lemma 6, (b) follows as the $\mathrm{diag}\left(\lambda_i + t_i^\ell\right)$ contains the number of times axis-aligned $\underline{\mathbf{w}}_s \in \{\mathbf{e}_1,\ldots,\mathbf{e}_p\}$ are observed, and (c) follows as the determinant of a diagonal matrix is the product of the diagonal elements. Now we can show that

$$\log \frac{|\mathbf{V}_\ell|}{|\mathbf{\Lambda}_\ell|} \leq \sum_{i=1}^k \log\left(1 + \frac{t_i^\ell}{\lambda}\right) + \sum_{i=k+1}^p \log\left(1 + \frac{t_i^\ell}{\lambda_i}\right)$$

$$\stackrel{(a)}{\leq} k \log\left(1 + \frac{t_i^\ell}{\lambda}\right) + \sum_{i=1}^p \frac{t_i^{\ell-1}}{\lambda_\ell^\perp}$$

$$\stackrel{(b)}{\leq} k \log\left(1 + \frac{t_i^\ell}{\lambda}\right) + \frac{\tau_{\ell-1}^G}{\lambda_\ell^\perp}$$

$$\stackrel{(c)}{\leq} k \log\left(1 + \frac{t_i^\ell}{\lambda}\right) + k \log\left(1 + \frac{\tau_{\ell-1}^G}{\lambda}\right)$$

$$\stackrel{(d)}{\leq} 8k \log\left(1 + \frac{\tau_{\ell-1}^G}{\lambda}\right)$$

where, (a) follows as $\log(1 + t_i^\ell/\lambda_i) \leq t_i^{\ell-1}/\lambda_\ell^\perp$, (b) follows as $\sum_{i=1}^p t_i^{\ell-1} = \tau_{\ell-1}^G$, and (c) follows for $\lambda_\ell^\perp = \frac{\tau_{\ell-1}^G}{k \log\left(1 + \tau_{\ell-1}^G/\lambda\right)}$ and (d) follows from Lemma 8. $\qquad\square$

Lemma 8. *Let* $\rho^G(\mathcal{Y}(\mathcal{W}_\ell)) = \min_{\mathbf{b}_{\mathbf{w}}} \max_{\mathbf{w},\mathbf{w}' \in \mathcal{W}_\ell} \|\mathbf{w} - \mathbf{w}'\|^2_{(\sum_{\mathbf{w} \in \mathcal{W}} \mathbf{b}_{\mathbf{w}} \mathbf{w} \mathbf{w}^\top + \mathbf{\Lambda}_\ell/n)^{-1}}$. *Recall that* $\tau_\ell^G = \frac{8 B_*^\ell \rho^G(\mathcal{Y}(\mathcal{W}_\ell)) \log(4\ell^2 |\mathcal{W}|/\delta)}{\epsilon_\ell^2}$. *Assume* $\log(p) \leq k$. *Then we can show that*

$$\log\left(1 + \frac{\tau_\ell^G}{\lambda}\right) \leq 8k \log\left(1 + \frac{\tau_{\ell-1}^G}{\lambda}\right).$$

Proof. We start by first recalling the definition of

$$\tau_\ell^G = 2\epsilon_\ell^{-2} 8k \log(1 + \frac{\tau_{\ell-1}^G}{\lambda}) \rho^G(\mathcal{Y}(\mathcal{W}_\ell)) \log\left(4\ell^2 |\mathcal{W}|/\delta_\ell\right).$$

Then we can show the following

$$\frac{\tau_\ell^G}{\tau_{\ell-1}^G} = \frac{2\epsilon_\ell^{-2} B_*^\ell \rho^G(\mathcal{Y}(\mathcal{W}_\ell)) \log\left(4\ell^2 |\mathcal{W}|/\delta_\ell\right)}{2\epsilon_{\ell-1}^{-2} B_*^{\ell-1} \rho^G(\mathcal{Y}(\mathcal{W}_{\ell-1})) \log\left(4(\ell-1)^2 |\mathcal{W}|/\delta_\ell\right)} \leq \frac{4\rho^G(\mathcal{Y}(\mathcal{W}_\ell)) \left(64(\lambda S^2 + \lambda_\ell^\perp S_\perp^{(2),\ell})\right)}{\rho^G(\mathcal{Y}(\mathcal{W}_{\ell-1})) \left(64(\lambda S^2 + \lambda_{\ell-1}^\perp S_\perp^{(2),\ell-1})\right)}$$

$$\leq \frac{4\rho^G(\mathcal{Y}(\mathcal{W}_\ell)) \tau_{\ell-1}^G / \log(1 + \frac{\tau_{\ell-1}^G}{\lambda})}{\rho^G(\mathcal{Y}(\mathcal{W}_{\ell-1})) \tau_{\ell-2}^G / \log(1 + \frac{\tau_{\ell-2}^G}{\lambda})}$$

$$\stackrel{(a)}{\leq} \frac{4\rho^G(\mathcal{Y}(\mathcal{W}_\ell)) \tau_{\ell-1}^G \log(1 + \frac{\tau_{\ell-1}^G}{\lambda})}{\rho^G(\mathcal{Y}(\mathcal{W}_{\ell-1}))}$$

$$\stackrel{(b)}{\leq} \frac{4p}{\gamma_{\mathcal{Y}}^2} \frac{\max_{\mathbf{w} \in \mathcal{W}} \|\mathbf{w}\|_2}{\max_{\mathbf{y} \in \mathcal{Y}(\mathcal{W}_\ell)} \|\mathbf{y}\|_2^2} \tau_{\ell-1}^G \log(1 + \frac{\tau_{\ell-1}^G}{\lambda}) = \frac{4p}{C\gamma_{\mathcal{Y}}^2}$$

17

where, (a) follows as $\log(1 + \frac{\tau_{\ell-2}^G}{\lambda}) \geq 1$ and $\log(1 + \frac{\tau_{\ell-1}^G}{\lambda}) \geq \log(1 + \frac{\tau_{\ell-2}^G}{\lambda})$. The (b) follows using Lemma 1 from Fiez et al. (2019) such that

$$\max_{\mathbf{y} \in \mathcal{Y}(\underline{\mathcal{W}}_\ell)} \|\mathbf{y}\|_2^2 / \left(\max_{\underline{\mathbf{w}} \in \mathcal{W}} \|\mathbf{w}\|_2 \right) \leq \rho^G(\mathcal{Y}(\underline{\mathcal{W}}_\ell)) \leq p/\gamma_\mathcal{Y}^2 \stackrel{(a_1)}{\Longrightarrow} 1 \leq \rho^G(\mathcal{Y}(\underline{\mathcal{W}}_\ell)) \leq p/\gamma_\mathcal{Y}^2.$$

where, (a_1) follows as $\|\mathbf{x}\| \leq 1$, $\|\mathbf{z}\| \leq 1$. This implies that for a constant $C > 0$

$$\tau_\ell^G \leq \frac{4p}{C\gamma_\mathcal{Y}^2}(\tau_{\ell-1}^G)^2 \log(1 + \frac{\tau_{\ell-1}^G}{\lambda}) \implies \log\left(1 + \frac{\tau_\ell^G}{\lambda}\right) \leq \log\left(1 + \frac{4p}{C\gamma_\mathcal{Y}^2}(\tau_{\ell-1}^G)^2 \log(1 + \frac{\tau_{\ell-1}^G}{\lambda})\right)$$

$$\stackrel{(a)}{\implies} \log\left(1 + \frac{\tau_\ell^G}{\lambda}\right) \leq 4k \log\left(1 + \frac{\tau_{\ell-1}^G}{C\gamma_\mathcal{Y}^2\lambda}\right)$$

$$\stackrel{(b)}{\implies} \log\left(1 + \frac{\tau_\ell^G}{\lambda}\right) \leq 8k \log\left(1 + \frac{\tau_{\ell-1}^G}{\lambda}\right)$$

where, in (a) follows for $\log(p) \leq k$, $\log(a^2 \log(a)) \leq 4\log(a)$. The (b) follows as $4k \log\left(1 + \frac{\tau_{\ell-1}^G}{C\gamma_\mathcal{Y}^2\lambda}\right) \leq 8k \log\left(1 + \frac{\tau_{\ell-1}^G}{\lambda}\right)$. $\qquad\square$

Lemma 9. *The G-optimal design in* (5) *is equivalent to solving the D-optimal design*

$$\mathbf{b}_\ell^D = \arg\max_{\mathbf{b}} \log \frac{\left|\sum_{\underline{\mathbf{w}} \in \mathcal{W}_\ell} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell\right|}{|\boldsymbol{\Lambda}_\ell|}.$$

Furthermore, the support of $|\mathbf{b}_\ell^D| \leq \frac{8k \log(1 + \frac{\tau_{\ell-1}^G}{\lambda})(8k \log(1 + \frac{\tau_{\ell-1}^G}{\lambda}) + 1)}{2}$, *where* $k = (d_1 + d_2)r$.

Proof. To prove the equivalence between $\mathbf{b}_*^G$ and $\mathbf{b}_*^D$ we need to first show that the regularization matrix $\boldsymbol{\Lambda}_\ell$ does not depend on $\underline{\mathbf{w}}$ or $\mathbf{y} = \underline{\mathbf{w}} - \underline{\mathbf{w}}'$, where $\underline{\mathbf{w}} \in \mathcal{W}_\ell$. Define $\text{conv}(\mathcal{X} \cup -\mathcal{X})$ as the convex hull of $\mathcal{X} \cup -\mathcal{X}$. Now recall we have from Lemma 1 of Fiez et al. (2019) that

$$1 \leq \rho^G(\mathcal{Y}(\underline{\mathcal{W}}_\ell)) \leq p/\gamma_\mathcal{Y}^2 \qquad (19)$$

where, $\gamma_\mathcal{Y} = \max\{c > 0 : c\mathcal{Y} \subset \text{conv}(\underline{\mathcal{W}}_\ell \cup -\underline{\mathcal{W}}_\ell)\}$ as the gauge norm of $\mathcal{Y}$ (Rockafellar, 2015). We can consider the gauge norm $\gamma_\mathcal{Y}$ as a problem-dependent constant. Now recall that

$$\lambda_\perp^\ell = \frac{\tau_{\ell-1}^G}{8k \log(1 + \frac{\tau_{\ell-1}^G}{\lambda})} \leq 2\tau_{\ell-1}^G \leq \frac{64 B_*^{\ell-1} \rho^G(\mathcal{Y}(\mathcal{W}_{\ell-1})) \log(4(\ell-1)^2 |\mathcal{W}|/\delta_\ell)}{\epsilon_{\ell-1}^2}$$

$$\stackrel{(a)}{\leq} \frac{64(B_*^{\ell-1})^2 p \log(4(\ell-1)^2 |\mathcal{W}|/\delta_\ell)}{\gamma_\mathcal{Y}^2 \epsilon_{\ell-1}^2}$$

$$\stackrel{(b)}{\leq} \frac{\left(256(\lambda S^2 + 8p^2 r)\right) \log(4(\ell-1)^2 p |\mathcal{W}|/\delta_\ell)}{S_r \gamma_\mathcal{Y}^2 \epsilon_{\ell-1}^2}$$

where, (a) follows from (19) and noting that $B_*^{\ell-1} \leq (B_*^{\ell-1})^2$. The (b) follows as $S_\ell^\perp := \frac{8pr}{\tau_\ell^E S_r^2} \log\left(\frac{d_1 + d_2}{\delta_\ell}\right)$, $p = d_1 d_2$ and substituting this value and $\lambda_\perp^\ell$ in B_*^ℓ we get

$$B_*^\ell \leq (B_*^\ell)^2 \leq \left(256(\lambda S^2 + \lambda_\perp^\ell S_\perp^{(2),\ell})\right) \leq \left(256\left(\lambda S^2 + \frac{8pr}{S_r^2}\frac{\tau_{\ell-1}^G}{\tau_\ell^E} \cdot \log(4(\ell-1)^2 |\mathcal{W}|/\delta_\ell) \log(d_1 + d_2/\delta_\ell)\right)\right)$$

$$\leq \left(256\left(\lambda S^2 + \frac{8pr}{S_r^2}\rho_{\ell-1}^G(\underline{\mathcal{W}}) \log(p|\mathcal{W}|/\delta_\ell)\right)\right) \stackrel{(a)}{\leq} \left(512\left(\lambda S^2 + 8p^2 r\right)/S_r\right).$$

Here S_r is the r-th largest eigenvalue of matrix $\boldsymbol{\Theta}_*$. Substituting this value of $\lambda_\ell^\perp$ we can show that $\boldsymbol{\Lambda}_\ell$ does not depend on $\underline{\mathbf{w}}$ or $\mathbf{y} = \underline{\mathbf{w}} - \underline{\mathbf{w}}'$. The rest of the proof to show equivalence follows the same way as in Theorem 21.1 in Lattimore and Szepesvári (2020).

18

To bound the support of $\mathbf{b}_*^D$ we proceed as follows: Define the set $\mathcal{Y}(\underline{\mathcal{W}}_\ell)$ as the set of all arms containing $\mathbf{y} = \underline{\mathbf{w}} - \underline{\mathbf{w}}' \in \mathbb{R}^p$. Then we can use Lemma 7 of Soare et al. (2014) to show that the solution to

$$\max_{\mathbf{y} \in \mathcal{Y}(\underline{\mathcal{W}}_\ell)} \|\mathbf{y}\|^2_{\left(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell\right)^{-1}} = \max_{\underline{\mathbf{w}}, \underline{\mathbf{w}}'} \|\underline{\mathbf{w}} - \underline{\mathbf{w}}'\|^2_{\left(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}}\underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell\right)^{-1}}$$

has a support of atmost $(k_1 + 1)k_1/2$ where $k_1 = 8k \log(1 + \tau^G_{\ell-1}/\lambda)$. The proof follows from the fact that for any pair $(\underline{\mathbf{w}}, \underline{\mathbf{w}}')$ we can show that

$$\|\underline{\mathbf{w}} - \underline{\mathbf{w}}'\|_{\left(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}}\underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}\right)^{-1}} \leq 2 \max_{\underline{\mathbf{w}}'' \in \underline{\mathcal{W}}_\ell} \|\underline{\mathbf{w}}''\|_{\left(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell\right)^{-1}} \cdot$$

Then following the work of Jamieson and Jain (2022) Frank-Wolfe algorithm (in section 2.3.1) with the

$$g(\mathbf{b}) = \log \frac{\left|\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell\right|}{|\boldsymbol{\Lambda}_\ell|} = \log \left|\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell\right| - \log|\boldsymbol{\Lambda}_\ell|,$$

and setting for the j-th iteration of the Frank-Wolfe the

$$I_j = \arg\max_{\mathbf{y} \in \mathcal{Y}(\underline{\mathcal{W}}_\ell)} \|\mathbf{y}\|^2_{\left(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell} \mathbf{b}^j_{\underline{\mathbf{w}}} \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell\right)^{-1}},$$

and stopping condition

$$\max_{\mathbf{y} \in \mathcal{Y}(\underline{\mathcal{W}}_\ell)} \|\mathbf{y}\|^2_{\left(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell} \mathbf{b}^j_{\underline{\mathbf{w}}} \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell\right)^{-1}} \leq 8k \log(1 + \frac{\tau^G_{\ell-1}}{\lambda}) \tag{20}$$

This can be done because note that for any $\mathbf{b} \in \triangle_{\underline{\mathcal{W}}_\ell}$ we have by Kiefer-Wolfowitz Theorem (Kiefer and Wolfowitz, 1960) that $[\nabla g(\mathbf{b})]_{\mathbf{y}} = \|\mathbf{y}\|^2_{\left(\sum_{\mathbf{x} \in \mathcal{X}} \mathbf{b}_x \mathbf{x}\mathbf{x}^\top + \boldsymbol{\Lambda}\right)^{-1}} \geq 8k \log(1 + \frac{\tau^G_{\ell-1}}{\lambda})$. This is because $\boldsymbol{\Lambda}_\ell$ does not depend on $\underline{\mathbf{w}}$ or $\mathbf{y}$ by the same logic as discussed before. The rest of the proof follows by the same way as in section 2.3.1 in Jamieson and Jain (2022). This will result in a support size of $\mathbf{b}$ at most $8k \log(1 + \frac{\tau^G_{\ell-1}}{\lambda})(8k \log(1 + \frac{\tau^G_{\ell-1}}{\lambda}) + 1)/2$ following Lemma 7 of Soare et al. (2014). Hence, it follows that solving the Equation (5) will result in a support of $|\mathbf{b}^D_\ell| \leq \frac{8k \log(1 + \frac{\tau^G_{\ell-1}}{\lambda})(8k \log(1 + \frac{\tau^G_{\ell-1}}{\lambda}) + 1)}{2}$. $\qquad\square$

Proposition 1. *If* $\mathbf{b}^G_\ell$ *is the G-optimal design for* $\underline{\mathcal{W}}_\ell$ *then if we pull arm* $\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell$ *exactly* $\lceil \tau^G \mathbf{b}^G_\ell \rceil$ *times for some* $\tau^G_\ell > 0$ *and compute the least squares estimator* $\widehat{\boldsymbol{\theta}}_\ell$. *Then for each* $\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell$ *we have with probability at least* $1 - \delta$

$$\mathbb{P}\left(\bigcup_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell} \left\{ \left|\langle \underline{\mathbf{w}}, \widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}^* \rangle\right| \leq \sqrt{\frac{64 B^\ell_* k \log(1 + \frac{\tau^G_{\ell-1}}{\lambda}) \log(2|\mathcal{W}|/\delta_\ell)}{\tau^G_\ell}} \right\}\right) \geq 1 - \delta_\ell.$$

where, $\|\boldsymbol{\theta}^*\|_\Lambda \leq \sqrt{\lambda \|\boldsymbol{\theta}_{1:k}\|^2_2 + \lambda^\perp_\ell \|\boldsymbol{\theta}_{k+1:p}\|^2_2} \leq \sqrt{\lambda} S + \sqrt{\lambda^\perp_\ell} S^\ell_\perp$, $B^\ell_* = \sqrt{\lambda} S + \sqrt{\lambda^\perp_\ell} S^\perp_\ell$.

Proof. From Woodbury Matrix Identity we know that for any arbitrary matrix $\mathbf{A}$ and $\mathbf{B}$, we have the following identity $(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} - \left(\mathbf{A} + \mathbf{A}\mathbf{B}^{-1}\mathbf{A}\right)^{-1}$. It follows then that

$$\underline{\mathbf{w}}^\top (\mathbf{A} + \mathbf{B})^{-1} \underline{\mathbf{w}} = \underline{\mathbf{w}}^\top \left(\mathbf{A}^{-1} - \left(\mathbf{A} + \mathbf{A}\mathbf{B}^{-1}\mathbf{A}\right)^{-1}\right) \underline{\mathbf{w}} \leq \underline{\mathbf{w}}^\top \mathbf{A}^{-1} \underline{\mathbf{w}} = \|\underline{\mathbf{w}}\|_{\mathbf{A}^{-1}}.$$

Hence we can show that,

$$\|\underline{\mathbf{w}}\|_{\left(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell} \lceil \tau^G_\ell \mathbf{b}^G_\ell \rceil \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell\right)^{-1}} \leq \underline{\mathbf{w}}^\top \left(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell} \lceil \tau^G_\ell \mathbf{b}^G_\ell \rceil \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top\right)^{-1} \underline{\mathbf{w}}. \tag{21}$$

From Lemma 9 we know the support of $\mathbf{b}^G_\ell$ is less than

$$\frac{8k \log(1 + \frac{\tau^G_{\ell-1}}{\lambda})(8k \log(1 + \frac{\tau^G_{\ell-1}}{\lambda}) + 1)}{2} \leq (8k \log(1 + \frac{\tau^G_{\ell-1}}{\lambda}))^2.$$

19

Also note that $\|\boldsymbol{\theta}^*\|_{\boldsymbol{\Lambda}_\ell} \leq \sqrt{\lambda \|\boldsymbol{\theta}_{1:k}\|_2^2 + \lambda_\perp^\ell \|\boldsymbol{\theta}_{k+1:p}\|_2^2} \leq \sqrt{\lambda} S + \sqrt{\lambda_\perp^\ell} S_\perp^\ell$. Then we can show that

$$
\begin{aligned}
\left\langle \underline{\mathbf{w}}, \widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}^* \right\rangle &\leq \|\underline{\mathbf{w}}\|_{\left(\sum_{\underline{\mathbf{w}} \in \mathcal{W}_\ell} \lceil \tau_\ell^G \mathbf{b}_\ell^G \rceil \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}\right)^{-1}} \left(2\sqrt{14 \log(2/\delta_\ell)} + \sqrt{\lambda} S + \sqrt{\lambda_\perp^\ell} S_\perp^\ell \right) \\
&\leq \frac{1}{\sqrt{\tau_\ell^G}} \|\underline{\mathbf{w}}\|_{\left(\sum_{\underline{\mathbf{w}} \in \mathcal{W}_\ell} \mathbf{b}_{\ell^*} \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}\right)^{-1}} \left(2\sqrt{14 \log(2/\delta_\ell)} + \sqrt{\lambda} S + \sqrt{\lambda_\perp^\ell} S_\perp^\ell \right) \\
&\overset{(a)}{\leq} \sqrt{\frac{56 \times 8k \log(1 + \frac{\tau_{\ell-1}^G}{\lambda}) \log(2/\delta_\ell)}{\tau_\ell^G}} + \sqrt{\frac{28k \log(1 + \frac{\tau_{\ell-1}^G}{\lambda}) \log(2/\delta_\ell)}{\tau_\ell^G}} (\sqrt{\lambda} S + \sqrt{\lambda_\perp^\ell} S_\perp^\ell) \\
&= \sqrt{\frac{8k \log(1 + \frac{\tau_{\ell-1}^G}{\lambda}) \log(2/\delta_\ell)}{\tau_\ell^G}} \left(\sqrt{56} + \underbrace{\sqrt{\lambda} S + \sqrt{\lambda_\perp^\ell} S_\perp^\ell}_{B_*^\ell} \right) \\
&\leq \sqrt{\frac{64 B_*^\ell k \log(1 + \frac{\tau_{\ell-1}^G}{\lambda}) \log(2/\delta_\ell)}{\tau_\ell^G}}
\end{aligned}
$$

where, (a) follows as $\|\underline{\mathbf{w}}\|_{\left(\sum_{\underline{\mathbf{w}} \in \mathcal{W}_\ell} \mathbf{b}_{\ell,\underline{\mathbf{w}}}^G \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell\right)^{-1}} \leq (8k \log(1 + \tau_{\ell-1}^G/\lambda))^2$. Thus we have taken at most $\tau_\ell^G + \frac{8k \log(1 + \frac{\tau_{\ell-1}^G}{\lambda})(8k \log(1 + \frac{\tau_{\ell-1}^G}{\lambda}) + 1)}{2}$ pulls. Thus, for any $\delta \in (0, 1)$ we have

$$
\mathbb{P}\left(\bigcup_{\underline{\mathbf{w}} \in \mathcal{W}_\ell} \left\{ \left| \left\langle \underline{\mathbf{w}}, \widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}^* \right\rangle \right| \geq \sqrt{\frac{64k B_*^\ell \log(1 + \frac{\tau_{\ell-1}^G}{\lambda}) \log(2|\mathcal{W}|/\delta_\ell)}{\tau_\ell^G}} \right\} \right) \leq \delta_\ell.
$$

The claim of the lemma follows. $\qquad\square$

Discussion 3. (Phase Length) It follows from Proposition 1 that if the gaps are known, one can set the phase length as

$$
\tau_\ell^G = \frac{64 B_*^\ell \rho(\mathcal{Y}(\mathcal{W}_\ell)) \log(2|\mathcal{W}|/\delta)}{(\mathbf{w}^\top (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}^*))^2}
$$

since $8k \log(1 + \frac{\tau_{\ell-1}^G}{\lambda}) \leq \rho(\mathcal{Y}(\mathcal{W}_\ell)) := 8k \log(1 + \frac{\tau_\ell^G}{\lambda})$ to guarantee that the event $\bigcup_{\underline{\mathbf{w}} \in \mathcal{W}_\ell} \left\{ \left| \left\langle \underline{\mathbf{w}}, \widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}^* \right\rangle \right| \right.$ holds with probability greater than $1 - \delta$.

However, since in practice, the gaps are not known, for an agnostic algorithm that does not know the gaps, one can set a proxy for the gap as ϵ_ℓ (for some $\epsilon_\ell > 0$) and get the phase length as follows:

$$
\tau_\ell^G = \frac{64 B_*^\ell \rho(\mathcal{Y}(\mathcal{W}_\ell)) \log(2|\mathcal{W}|/\delta)}{\epsilon_\ell^2}.
$$

This gives us the desired phase length so that the event $\bigcup_{\underline{\mathbf{w}} \in \mathcal{W}_\ell} \left\{ \left| \left\langle \underline{\mathbf{w}}, \widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}^* \right\rangle \right| \right.$ holds with probability greater than $1 - \delta$.

Lemma 10. *Assume that $\max_{\underline{\mathbf{w}} \in \mathcal{W}} \langle \underline{\mathbf{w}}_* - \underline{\mathbf{w}}, \boldsymbol{\theta}^* \rangle \leq 2$. With probability at least $1 - \delta$, we have $\underline{\mathbf{w}}_* \in \mathcal{W}_\ell$ and $\max_{\underline{\mathbf{w}} \in \mathcal{W}_\ell} \langle \underline{\mathbf{w}}_* - \underline{\mathbf{w}}, \boldsymbol{\theta}^* \rangle \leq 4\epsilon_\ell$ for all $\ell \in \mathbb{N}$.*

Proof. For any $\mathcal{V} \subseteq \mathcal{W}_\ell$ be the active set and $\underline{\mathbf{w}} \in \mathcal{V}$ define

$$
\mathcal{E}_{\underline{\mathbf{w}}, \ell}(\mathcal{V}) = \left\{ \left| \left\langle \underline{\mathbf{w}} - \underline{\mathbf{w}}_*, \widehat{\boldsymbol{\theta}}_\ell(\mathcal{V}) - \boldsymbol{\theta}^* \right\rangle \right| \leq \epsilon_\ell \right\} \tag{22}
$$

where it is implicit that $\widehat{\boldsymbol{\theta}}_\ell := \widehat{\boldsymbol{\theta}}_\ell(\mathcal{V})$ is the design constructed in the algorithm at stage ℓ with respect to $\mathcal{W}_\ell = \mathcal{V}$. Also note that $\delta_\ell = \frac{\delta}{4\ell^2}$. Given $\mathcal{W}_\ell$, with probability at least $1 - 2 \cdot \frac{\delta}{4\ell^2 |\mathcal{W}|}$

20

$$\left|\left\langle \underline{\mathbf{w}} - \underline{\mathbf{w}}_*, \widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}^* \right\rangle\right| \overset{(a)}{\leq} \sqrt{64 B_*^\ell k \log(1 + \frac{\tau_{\ell-1}^G}{\lambda}) \log\left(4\ell^2 |\mathcal{W}|/\delta\right)}$$

$$\leq \frac{(k \log(1 + \frac{\tau_{\ell-1}^G}{\lambda}))^2}{\sqrt{\tau_\ell^G}} \sqrt{64 B_*^\ell \log\left(4\ell^2 |\mathcal{W}|/\delta\right)}$$

$$\overset{(b)}{\leq} \sqrt{\frac{\|\underline{\mathbf{w}} - \underline{\mathbf{w}}_*\|^2_{\left(\sum_{\underline{\mathbf{w}} \in \mathcal{V}} \mathbf{b}_{\ell,\underline{\mathbf{w}}}^G(\mathcal{V}) \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell\right)^{-1}}}{64 B_*^\ell \epsilon_\ell^{-2} \rho(\mathcal{Y}(\mathcal{W}_\ell)) \log\left(4\ell^2 |\mathcal{W}|/\delta\right)}} \sqrt{64 B_*^\ell \log\left(4\ell^2 |\mathcal{W}|/\delta\right)}$$

$$\leq \sqrt{\frac{\|\underline{\mathbf{w}} - \underline{\mathbf{w}}_*\|^2_{\left(\sum_{\underline{\mathbf{w}} \in \mathcal{V}} \mathbf{b}_{\ell,\underline{\mathbf{w}}}^G(\mathcal{V}) \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell\right)^{-1}}}{\epsilon_\ell^{-2} \rho(\mathcal{Y}(\mathcal{W}_\ell)) \log\left(4\ell^2 |\mathcal{W}|/\delta\right)}} \sqrt{\log\left(4\ell^2 |\mathcal{W}|/\delta\right)}$$

$$\overset{(c)}{=} \epsilon_\ell$$

where, (a) follows Proposition 1. The (b) follows as

$$(k \log(1 + \frac{\tau_{\ell-1}^G}{\lambda}))^2 \leq (k \log(1 + \frac{\tau_\ell^G}{\lambda}))^2 := \rho(\mathcal{Y}(\mathcal{W}_\ell)) = \|\underline{\mathbf{w}} - \underline{\mathbf{w}}_*\|^2_{\left(\sum_{\underline{\mathbf{w}} \in \mathcal{V}} \mathbf{b}_{\ell,\underline{\mathbf{w}}}^G(\mathcal{V}) \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell\right)^{-1}} \cdot$$

The (c) follows as $\rho(\mathcal{Y}(\mathcal{W}_\ell)) = \|\underline{\mathbf{w}} - \underline{\mathbf{w}}_*\|^2_{\left(\sum_{\underline{\mathbf{w}} \in \mathcal{V}} \mathbf{b}_{\ell,\underline{\mathbf{w}}}^G(\mathcal{V}) \underline{\mathbf{w}}\, \underline{\mathbf{w}}^\top + \boldsymbol{\Lambda}_\ell\right)^{-1}}$. By exactly the same sequence of steps as above, we have

$$\mathbb{P}\left(\bigcap_{\ell=1}^\infty \bigcap_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell} \left\{\left|\left\langle \underline{\mathbf{w}} - \underline{\mathbf{w}}_*, \widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* \right\rangle\right| > \epsilon_t\right\}\right) = \mathbb{P}\left(\bigcap_{\underline{\mathbf{w}} \in \mathcal{W}_\ell} \bigcap_{\ell=1}^\infty \mathcal{E}_{\underline{\mathbf{w}},\ell}\left(\mathcal{W}_\ell\right)\right) \geq 1 - \delta,$$

so assume these events hold. Consequently, for any $\underline{\mathbf{w}}' \in \underline{\mathcal{W}}_\ell$

$$\left\langle \underline{\mathbf{w}}' - \underline{\mathbf{w}}_*, \widehat{\boldsymbol{\theta}}_\ell \right\rangle = \left\langle \underline{\mathbf{w}}' - \underline{\mathbf{w}}_*, \widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}^* \right\rangle + \left\langle \underline{\mathbf{w}}' - \underline{\mathbf{w}}_*, \boldsymbol{\theta}^* \right\rangle$$

$$\leq \left\langle \underline{\mathbf{w}}' - \underline{\mathbf{w}}_*, \widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}^* \right\rangle$$

$$\leq \epsilon_\ell$$

so that $\underline{\mathbf{w}}_*$ would survive to round $\ell + 1$. And for any $\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell$ such that $\langle \underline{\mathbf{w}}_* - \underline{\mathbf{w}}, \boldsymbol{\theta}^* \rangle > 2\epsilon_\ell$ we have

$$\max_{\underline{\mathbf{w}}' \in \underline{\mathcal{W}}_\ell} \left\langle \underline{\mathbf{w}}' - \underline{\mathbf{w}}, \widehat{\boldsymbol{\theta}}_\ell \right\rangle \geq \left\langle \underline{\mathbf{w}}_* - \underline{\mathbf{w}}, \widehat{\boldsymbol{\theta}}_\ell \right\rangle$$

$$= \left\langle \underline{\mathbf{w}}_* - \underline{\mathbf{w}}, \widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}^* \right\rangle + \left\langle \underline{\mathbf{w}}_* - \underline{\mathbf{w}}, \boldsymbol{\theta}^* \right\rangle$$

$$> -\epsilon_\ell + 2\epsilon_\ell$$

$$= \epsilon_\ell$$

which implies this $\underline{\mathbf{w}}$ would be eliminated. Note that this implies that $\max_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_{\ell+1}} \langle \underline{\mathbf{w}}_* - \underline{\mathbf{w}}, \boldsymbol{\theta}^* \rangle \leq 2\epsilon_\ell = 4\epsilon_{\ell+1}$. Hence, the claim of the lemma follows. $\qquad\square$

A.5 Final Sample Complexity Bound for Single Task Setting

Theorem 1. *(Restatement)* *With probability at least $1 - \delta$,* GOBLIN *returns the best arms $\mathbf{x}_*$, $\mathbf{z}_*$, and the number of samples used is bounded by*

$$\widetilde{O}\left(\frac{(d_1 + d_2)r}{\Delta^2} + \frac{\sqrt{d_1 d_2 r}}{S_r}\right)$$

where, $\Delta = \min_{\mathbf{x} \in \mathcal{X} \setminus \{\mathbf{x}_*\}, \mathbf{z} \in \mathcal{Z} \setminus \{\mathbf{z}_*\}} (\mathbf{x}_*^\top \boldsymbol{\Theta}_* \mathbf{z}_* - \mathbf{x}^\top \boldsymbol{\Theta}_* \mathbf{z})$, and S_r is the r-th largest singular value of $\boldsymbol{\Theta}_*$.

21

Proof. For the rest of the proof we have that the good events $\mathcal{F}_\ell \bigcap \mathcal{E}_{\mathbf{w},\ell}(\mathcal{W}_\ell)$ holds true for each phase ℓ with probability greater than $(1-\delta)$. The two events are defined in (16) and (22).

Second Stage: Define $\underline{A}_\ell = \{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_\ell : \langle \underline{\mathbf{w}}_* - \underline{\mathbf{w}}, \boldsymbol{\theta}^* \rangle \le 4\epsilon_\ell\}$. Note that by assumption $\underline{\mathcal{W}} = \underline{\mathcal{W}}_1 = \underline{\mathcal{S}}_1$. The above lemma implies that with probability at least $1 - \delta$ we have $\bigcap_{\ell=1}^\infty \{\underline{\mathcal{W}}_\ell \subseteq \underline{A}_\ell\}$. This implies that

$$
\begin{aligned}
\rho^G\left(\underline{\mathcal{W}}_\ell\right) &= \min_{\mathbf{b} \in \Delta_{\underline{\mathcal{W}}}} \max_{\underline{\mathbf{w}},\underline{\mathbf{w}}' \in \underline{\mathcal{W}}_\ell} \|\underline{\mathbf{w}} - \underline{\mathbf{w}}'\|^2_{\left(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}}\,\underline{\mathbf{w}}^\top + \mathbf{\Lambda}\right)^{-1}} \\
&\le \min_{\mathbf{b} \in \Delta_{\underline{\mathcal{W}}}} \max_{\underline{\mathbf{w}},\underline{\mathbf{w}}' \in \underline{A}_\ell} \|\underline{\mathbf{w}} - \underline{\mathbf{w}}'\|^2_{\left(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}}\,\underline{\mathbf{w}}^\top + \mathbf{\Lambda}\right)^{-1}} \\
&= \rho^G\left(\underline{A}_\ell\right).
\end{aligned}
$$

Define $k_1^\ell = 8k \log(1 + \tau_{\ell-1}^G/\lambda)$. For $\ell \ge \lceil \log_2\left(4\Delta^{-1}\right) \rceil$ we have that $\underline{A}_\ell = \{\underline{\mathbf{w}}_*\}$, thus, the sample complexity to identify $\underline{\mathbf{w}}_*$ is equal to

$$
\begin{aligned}
\sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} \sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}} \left\lceil \tau_\ell^G \widehat{\mathbf{b}}_{\ell,\underline{\mathbf{w}}}^G \right\rceil &= \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} \left(\frac{(k_1^\ell + 1)k_1^\ell}{2} + \tau_\ell^G \right) \\
&= \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} \left(\frac{(k_1^\ell + 1)k_1^\ell}{2} + 2\epsilon_\ell^{-2} \rho^G(\underline{\mathcal{W}}_\ell) B_*^\ell \log\left(4k_1^\ell \ell^2 |\underline{\mathcal{W}}|/\delta\right) \right) \\
&\overset{(a)}{\le} 2 \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} \left(\frac{(k+1)k}{2} \log^2(1 + \tau_{\ell-1}^G) + 2\epsilon_\ell^{-2} \rho^G(\underline{\mathcal{W}}_\ell) B_*^\ell \log\left(4k_1^\ell \ell^2 |\underline{\mathcal{W}}|/\delta\right) \right) \\
&\overset{(b)}{\le} 2\frac{(k+1)k}{2} \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} \left(\log^2(1 + \tau_{\ell-1}^G) + 8\epsilon_\ell^{-2} \rho^G(\underline{\mathcal{W}}_\ell) B_*^\ell \log\left(4k_1^\ell \ell^2 |\underline{\mathcal{W}}|/\delta\right) \right) \\
&\overset{(c)}{\le} (k+1)k \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} \left(1 + 16\epsilon_\ell^{-2} \rho^G(\underline{\mathcal{W}}_\ell) B_*^\ell \log^2\left(4k_1^\ell \ell^2 |\underline{\mathcal{W}}|/\delta\right) \right) \\
&\overset{(d)}{\le} (k+1)k \lceil \log_2\left(4\Delta^{-1}\right) \rceil + \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} 32\epsilon_\ell^{-2} f\left(\underline{A}_\ell\right) B_*^\ell \log\left(4k\ell^2 |\underline{\mathcal{W}}|/\delta\right) \\
&\overset{(e)}{\le} (k+1)k \lceil \log_2\left(4\Delta^{-1}\right) \rceil + \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} 32\epsilon_\ell^{-2} f\left(\underline{A}_\ell\right) (64\lambda S^2 + 64\tau_{\ell-1}^G) \log\left(4k\ell^2 |\underline{\mathcal{W}}|/\delta\right) \\
&= (k+1)k \lceil \log_2\left(4\Delta^{-1}\right) \rceil + \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} 32\epsilon_\ell^{-2} f\left(\underline{A}_\ell\right) (64\lambda S^2) \log\left(4k\ell^2 |\underline{\mathcal{W}}|/\delta\right) \\
&\quad + (k+1)k \lceil \log_2\left(4\Delta^{-1}\right) \rceil + \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} 32\epsilon_\ell^{-2} f\left(\underline{A}_\ell\right) (64\tau_{\ell-1}^G) \log\left(4k\ell^2 |\underline{\mathcal{W}}|/\delta\right) \\
&\overset{(f)}{\le} (k+1)k \lceil \log_2\left(4\Delta^{-1}\right) \rceil + \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} 64\epsilon_\ell^{-2} f\left(\underline{A}_\ell\right) (64\lambda S^2) \log\left(4k\ell^2 |\underline{\mathcal{W}}|/\delta\right) \\
&\le (k+1)k \lceil \log_2\left(4\Delta^{-1}\right) \rceil + 2048\lambda S^2 \log\left(\frac{4k \log_2^2\left(8\Delta^{-1}\right) |\underline{\mathcal{W}}|}{\delta} \right) \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} 2^{2\ell} f\left(\underline{A}_\ell\right)
\end{aligned}
$$

where, (a) follows as $\log^2(1 + \tau_{\ell-1}^G/\lambda) \le \log^2(1 + \tau_{\ell-1}^G)$, (b) follows by noting that $\log(x \log(1 + x)) \le 2\log(x)$ for any $x > 1$. The (c) follows by subsuming the $\log^2(1 + \tau_{\ell-1}^G)$ into $2\tau_\ell^G$. The (d) follows as $\log(1 + \tau_{\ell-1}^G) < \tau_\ell^G$ which enables us to replace the k_1^ℓ inside the log with an additional

22

factor of 2. The (e) follows by noting that

$$B_*^\ell \leq 64(\sqrt{\lambda}S + \sqrt{\lambda_\ell^\perp}S_\ell^\perp)$$

$$\leq 64\lambda S^2 + \left(\frac{64\tau_{\ell-1}^G}{8(d_1+d_2)r\log(1+\frac{\tau_{\ell-1}^G}{\lambda})}\right) \cdot \left(\frac{8d_1d_2r}{\tau_\ell^E S_r^2}\log\left(\frac{d_1+d_2}{\delta_\ell}\right)\right)$$

$$\overset{(a_1)}{\leq} 64\lambda S^2 + 64\tau_{\ell-1}^G. \tag{23}$$

where, (a_1) follows by first substituting the value of $\tau_\ell^E := \frac{\sqrt{8d_1d_2r\log(4\ell^2|\mathcal{W}|/\delta_\ell)}}{S_r}$ and noting that $\sqrt{(d_1d_2r)} \leq (d_1+d_2)r$ and cancelling out the other terms. Finally the (f) follows by subsuming the $\tau_{\ell-1}^G$ with a factor of 2 into the quantity of τ_ℓ^G. Then it follows that

$$\rho_*^G = \inf_{\mathbf{b}\in\triangle_{\mathcal{W}}} \max_{\underline{\mathbf{w}}\in\underline{\mathcal{W}}} \frac{\|\underline{\mathbf{w}} - \underline{\mathbf{w}}_*\|^2_{(\sum_{\underline{\mathbf{w}}\in\mathcal{W}} \mathbf{b}_{\underline{\mathbf{w}}}\underline{\mathbf{w}}\,\underline{\mathbf{w}}^\top+\mathbf{\Lambda})^{-1}}}{(\langle\underline{\mathbf{w}} - \underline{\mathbf{w}}_*, \boldsymbol{\theta}^*\rangle)^2}$$

$$= \inf_{\mathbf{b}\in\triangle_{\mathcal{W}}} \max_{\ell\leq\lceil\log_2(4\Delta^{-1})\rceil} \max_{\underline{\mathbf{w}}\in\underline{\mathcal{A}}_\ell} \frac{\|\underline{\mathbf{w}} - \underline{\mathbf{w}}_*\|^2_{(\sum_{\underline{\mathbf{w}}\in\mathcal{W}} \mathbf{b}_{\underline{\mathbf{w}}}\underline{\mathbf{w}}\,\underline{\mathbf{w}}^\top+\mathbf{\Lambda})^{-1}}}{(\langle\underline{\mathbf{w}} - \underline{\mathbf{w}}_*, \boldsymbol{\theta}^*\rangle)^2}$$

$$\geq \frac{1}{\lceil\log_2(4\Delta^{-1})\rceil} \inf_{\mathbf{b}\in\triangle_{\mathcal{W}}} \sum_{\ell=1}^{\lceil\log_2(4\Delta^{-1})\rceil} \max_{\underline{\mathbf{w}}\in\underline{\mathcal{A}}_\ell} \frac{\|\underline{\mathbf{w}} - \underline{\mathbf{w}}_*\|^2_{(\sum_{\underline{\mathbf{w}}\in\mathcal{W}} \mathbf{b}_{\underline{\mathbf{w}}}\underline{\mathbf{w}}\,\underline{\mathbf{w}}^\top+\mathbf{\Lambda})^{-1}}}{(\langle\underline{\mathbf{w}} - \underline{\mathbf{w}}_*, \boldsymbol{\theta}^*\rangle)^2}$$

$$\geq \frac{1}{16\lceil\log_2(4\Delta^{-1})\rceil} \sum_{\ell=1}^{\lceil\log_2(4\Delta^{-1})\rceil} 2^{2\ell} \inf_{\mathbf{b}\in\triangle_{\mathcal{W}}} \max_{\underline{\mathbf{w}}\in\underline{\mathcal{A}}_\ell} \|\underline{\mathbf{w}} - \underline{\mathbf{w}}_*\|^2_{(\sum_{\underline{\mathbf{w}}\in\mathcal{W}} \mathbf{b}_{\underline{\mathbf{w}}}\underline{\mathbf{w}}\,\underline{\mathbf{w}}^\top+\mathbf{\Lambda})^{-1}}$$

$$\geq \frac{1}{64\lceil\log_2(4\Delta^{-1})\rceil} \sum_{\ell=1}^{\lceil\log_2(4\Delta^{-1})\rceil} 2^{2\ell} \inf_{\mathbf{b}\in\triangle_{\mathcal{W}}} \max_{\underline{\mathbf{w}},\underline{\mathbf{w}}'\in\underline{\mathcal{A}}_\ell} \|\underline{\mathbf{w}} - \underline{\mathbf{w}}'\|^2_{(\sum_{\underline{\mathbf{w}}\in\mathcal{W}} \mathbf{b}_{\underline{\mathbf{w}}}\underline{\mathbf{w}}\,\underline{\mathbf{w}}^\top+\mathbf{\Lambda})^{-1}}$$

$$\geq \frac{1}{64\lceil\log_2(4\Delta^{-1})\rceil} \sum_{\ell=1}^{\lceil\log_2(4\Delta^{-1})\rceil} 2^{2\ell} f(\underline{\mathcal{A}}_\ell).$$

This implies that

$$\sum_{\ell=1}^{\lceil\log_2(4\Delta^{-1})\rceil} 2^{2\ell} f(\underline{\mathcal{A}}_\ell) \leq \rho_*^G 64 \lceil\log_2(4\Delta^{-1})\rceil$$

Plugging this back we get

$$\sum_{\ell=1}^{\lceil\log_2(4\Delta^{-1})\rceil} \sum_{\underline{\mathbf{w}}\in\underline{\mathcal{W}}} \lceil\tau_\ell^G\widehat{\mathbf{b}}_{\ell,\underline{\mathbf{w}}}\rceil \leq (k+1)k\lceil\log_2(4\Delta^{-1})\rceil + 2048\lambda S^2 \log\left(\frac{8k\log_2^2(8\Delta^{-1})|\mathcal{W}|}{\delta}\right)\rho_*^G 64\lceil\log_2(4\Delta^{-1})\rceil$$

$$\leq (k+1)k\lceil\log_2(4\Delta^{-1})\rceil + C_2\lambda S^2 \log\left(\frac{8k\log_2^2(8\Delta^{-1})|\mathcal{W}|}{\delta}\right)\rho_*^G\lceil\log_2(4\Delta^{-1})\rceil$$

for some constant $C_2 > 0$. Now to understand the bound we need the following: Let $\mathrm{conv}(\underline{\mathcal{W}}\cup-\underline{\mathcal{W}})$ denote the convex hull of $\underline{\mathcal{W}}\cup-\underline{\mathcal{W}}$, and for any set $\mathcal{Y}\subset\mathbb{R}^p$ define the gauge of $\mathcal{Y}$

$$\gamma_{\mathcal{Y}} = \max\{c > 0 : c\mathcal{Y}\subset\mathrm{conv}(\underline{\mathcal{W}}\cup-\underline{\mathcal{W}})\} \tag{24}$$

23

In the case where $\mathcal{Y}$ is a singleton $\mathcal{Y} = \{y\}$, $\gamma(y) := \gamma_{\mathcal{Y}}$ is the gauge norm of $\mathbf{y}$ with respect to $\mathrm{conv}(\mathcal{W} \cup -\mathcal{W})$. We can provide a natural upper bound for $\rho(\mathcal{Y})$ in terms of the gauge. Observe that

$$
\begin{aligned}
\rho_*^G &= \inf_{\mathbf{b} \in \triangle_{\mathcal{W}}} \max_{\mathbf{y} \in \mathcal{Y}} \|\mathbf{y}\|^2_{(\sum_{\mathbf{w} \in \mathcal{W}} b_{\mathbf{w}} \underline{\mathbf{w}} \, \underline{\mathbf{w}}^\top + \Lambda)^{-1}} \\
&= \frac{1}{\gamma_{\mathcal{Y}}^2} \inf_{\mathbf{b} \in \triangle_{\mathcal{W}}} \max_{\mathbf{y} \in \mathcal{Y}} \|\mathbf{y}\gamma_{\mathcal{Y}}\|^2_{(\sum_{\mathbf{w} \in \mathcal{W}} b_{\mathbf{w}} \underline{\mathbf{w}} \, \underline{\mathbf{w}}^\top + \Lambda)^{-1}} \\
&\leq \frac{1}{\gamma_{\mathcal{Y}}^2} \inf_{\mathbf{b} \in \triangle_{\mathcal{W}}} \max_{\mathbf{w} \in \mathrm{conv}(\mathcal{W} \cup -\mathcal{W})} \|\mathbf{w}\|^2_{(\sum_{\mathbf{w} \in \mathcal{W}} b_{\mathbf{w}} \underline{\mathbf{w}} \, \underline{\mathbf{w}}^\top + \Lambda)^{-1}} \\
&\overset{(a)}{=} \frac{1}{\gamma_{\mathcal{Y}}^2} \inf_{\mathbf{b} \in \triangle_{\mathcal{W}}} \max_{\mathbf{w} \in \mathcal{W}} \|\mathbf{w}\|^2_{(\sum_{\mathbf{w} \in \mathcal{W}} b_{\mathbf{w}} \underline{\mathbf{w}} \, \underline{\mathbf{w}}^\top + \Lambda)^{-1}} \overset{(b)}{\leq} k^{3/2} \\
&\leq \frac{k}{\gamma_{\mathcal{Y}}^2} O\left(B_* \log\left(\frac{k \log_2\left(\Delta^{-1}\right) |\mathcal{W}|}{\delta} \right) \lceil \log_2\left(\Delta^{-1}\right) \rceil \right)
\end{aligned}
$$

The (a) follows from the fact that the maximum value of a convex function on a convex set must occur at a vertex. The (b) follows from Kiefer-wolfowitz theorem for $\underline{\mathbf{w}} \in \mathbb{R}^p$ such that $\inf_{\mathbf{b} \in \triangle_{\mathcal{W}}} \max_{\underline{\mathbf{w}} \in \mathcal{W}} \|\mathbf{w}\|^2_{(\sum_{\mathbf{w} \in \mathcal{W}} b_{\mathbf{w}} \underline{\mathbf{w}} \, \underline{\mathbf{w}}^\top + \Lambda)^{-1}} \leq k \log(1 + \frac{\tau^G_{\ell-1}}{\lambda})$. The simplified sample complexity for the second stage is given by

$$
N_2 \leq O\left(\frac{k}{\Delta^2} \log\left(\frac{k \log_2\left(\Delta^{-1}\right) |\mathcal{W}|}{\delta} \right) \right) = \widetilde{O}\left(\frac{(d_1 + d_2)r}{\Delta^2} \right)
$$

where $\Delta = \min_{\underline{\mathbf{w}} \in \mathcal{W}} (\underline{\mathbf{w}}_* - \underline{\mathbf{w}})^\top \boldsymbol{\theta}_* \overset{(a_1)}{=} \min_{\mathbf{x} \in \mathcal{X} \setminus \{\mathbf{x}_*\}, \mathbf{z} \in \mathcal{Z} \setminus \{\mathbf{z}_*\}} (\mathbf{x}_*^\top \boldsymbol{\Theta}_* \mathbf{z}_* - \mathbf{x}^\top \boldsymbol{\Theta}_* \mathbf{z})$. The (a_1) follows by reshaping the arms in $\underline{\mathcal{W}}$ to recover the arms in $\mathcal{X}$ and $\mathcal{Z}$.

1st Stage: First recall that the E-optimal design in step 3 of Algorithm 1 satisfies the Assumption 4 as the sample distribution $\mathcal{D}$ has finite second order moments. For the first stage first observe that by plugging in the definition of $\tau_\ell^E = \frac{\sqrt{8 d_1 d_2 r \log(4\ell^2 |\mathcal{W}|/\delta)}}{S_r}$ we get

$$
\begin{aligned}
\left\|\boldsymbol{\theta}^*_{k+1:p}\right\|_2^2 &= \sum_{i > r \wedge j > r} H_{ij}^2 = \left\| (\widehat{\mathbf{U}}_\ell^\perp)^\top \left(\mathbf{U}^* \mathbf{S}^* \mathbf{V}^{*\top}\right) \widehat{\mathbf{V}}_\ell^\perp \right\|_F^2 \\
&\leq \left\| (\widehat{\mathbf{U}}_\ell^\perp)^\top \mathbf{U}^* \right\|_F^2 \|\mathbf{S}^*\|_2^2 \left\| (\widehat{\mathbf{V}}_\ell^\perp)^\top \mathbf{V}^* \right\|_F^2 \leq O\left(\frac{d_1 d_2 r}{\tau_\ell^E S_r^2} \log\left(\frac{d_1 + d_2}{\delta} \right) \right) \\
&= O\left(\frac{\sqrt{d_1 d_2 r}}{S_r} \log\left(\frac{d_1 + d_2}{\delta} \right) \right)
\end{aligned}
$$

which implies $\left\|\boldsymbol{\theta}^*_{k+1:p}\right\|_2 = \widetilde{O}\left(\sqrt{d_1 d_2 r}/S_r \right)$. Now we bound the sample complexity from the first stage. From the first stage we can show that we have for the arm set $\overline{\mathcal{W}}$

$$
\begin{aligned}
N_1 &= \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} \sum_{\overline{\mathbf{w}} \in \overline{\mathcal{W}}} \left\lceil \tau_\ell^E \widehat{\mathbf{b}}^E_{\ell, \overline{\mathbf{w}}} \right\rceil \\
&= \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} \left(\frac{(p+1)p}{2} + \tau_\ell^E \right) \\
&= \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} \left(\frac{(p+1)p}{2} + \frac{\sqrt{8 d_1 d_2 r \log\left(4\ell^2 |\mathcal{W}|/\delta\right)}}{S_r} \right) \\
&\leq (p+1)p \lceil \log_2\left(4\Delta^{-1}\right) \rceil + 32 \frac{\sqrt{d_1 d_2 r}}{S_r} \log\left(\frac{4 \log_2^2\left(8\Delta^{-1}\right) |\mathcal{W}|}{\delta} \right) \lceil \log_2\left(4\Delta^{-1}\right) \rceil \\
&\overset{(a)}{=} O\left(\frac{\sqrt{d_1 d_2 r}}{S_r} \log\left(\frac{4 \log_2^2\left(8\Delta^{-1}\right) |\mathcal{W}|}{\delta} \right) \right) = \widetilde{O}\left(\frac{\sqrt{d_1 d_2 r}}{S_r} \right)
\end{aligned}
$$

where, (a) follows as $p = d_1 d_2$. Combining N_1 and N_2 gives the claim of the theorem. $\qquad \square$

24

A.6 Multi-Task Pure Exploration Proofs

Remark 2. (Comparison with Du et al. (2023)) In this remark, we discuss a key comparison of DouExpDes (Du et al., 2023) with GOBLIN. Note that DouExpDes does not implement the second stage of finding the $\mathbf{S}_{m,*} \in \mathbb{R}^{k_1 \times k_2}$ for each of the m bilinear bandits. Hence, DouExpDes does not rotate the arms so that the last $(k_1 - r) \cdot (k_2 - r)$ components are from the complementary subspaces of the left and right eigenvectors of $\mathbf{S}_{m,*}$. This results in DouExpDes suffering a sample complexity of $\widetilde{O}(k_1 k_2 / \Delta^2)$ even though it learns the common feature extractors shared across the tasks. In contrast GOBLIN uses the second stage to learn $\mathbf{S}_{m,*} \in \mathbb{R}^{k_1 \times k_2}$ and reduces the latent bilinear bandits problem of $k_1 k_2$ dimension to $(k_1 + k2)r$ dimension by rotating the arms so that the last $(k_1 - r) \cdot (k_2 - r)$ components are from the complementary subspaces of the left and right eigenvectors of $\mathbf{S}_{m,*}$. Hence, GOBLIN suffers a sample complexity of $\widetilde{O}((k_1 + k_2)r / \Delta^2)$.

Remark 3. (Arm set) The observable left and right arm sets $\mathcal{X}$ and $\mathcal{Z}$ are common across the M tasks. This leads to each task estimating the same E-optimal design in line 3 of Algorithm 2 of stage 1. Note that Du et al. (2023) also uses a similar idea of the same arm set $\mathcal{X}$ shared across tasks in the linear bandit setting. Observe that if each task has access to its own separate arm sets $\mathcal{X}_m$ and $\mathcal{Z}_m$, then each of the m-tasks has to estimate a separate E-optimal design for the stage 1. This will lead to the sample complexity of the first stage scaling as $\widetilde{O}(M\sqrt{d_1 d_2}/S_r)$ instead of $\widetilde{O}(\sqrt{d_1 d_2}r/S_r)$.

Good Event: We first recall the total stage 1 length as

$$\tau_\ell^E := \frac{\sqrt{8 d_1 d_2 r \log(4\ell^2 |\mathcal{W}|/\delta_\ell)}}{S_r}.$$

Then define the good event $\mathcal{F}_\ell$ in phase ℓ that GOBLIN has a good estimate of $\mathbf{Z}_* = \frac{1}{M}\sum_{m=1}^{M} \mathbf{\Theta}_m$ as follows: For any phase $\ell > 0$

$$\mathcal{F}_\ell := \left\{ \left\| \widehat{\mathbf{Z}}_\ell - \mu^* \mathbf{Z}_* \right\|_F^2 \leq \frac{C_1 d_1 d_2 r \log\left(\frac{2(d_1 + d_2)}{\delta_\ell}\right)}{\tau_\ell^E} \right\}, \tag{25}$$

where, $C_1 = 36 \left(4 + S_0^2\right) C$, $\|\mathbf{X}\|_F, \|\mathbf{\Theta}_*\|_F \leq S_0$, some nonzero constant μ^*, $\mathbb{E}\left[(S\mathbf{P}(\mathbf{X}))_{ij}^2\right] \leq C, \forall i, j$, and $\widehat{\mathbf{\Theta}}_\ell$ is the estimate from (8). Then define the event

$$\mathcal{F} := \bigcap_{\ell=1}^{\infty} \mathcal{F}_\ell \tag{26}$$

Then we start by modifying Lemma 3 for the multi-task setting. We first prove this support lemma for the loss function defined in (8).

Lemma 11. *Let* $L : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}$ *is the loss function defined in* (8). *Then by setting*

$$t = \sqrt{2 d_1 d_2 C \left(4 + S_0^2\right) \log\left(\frac{2\left(d_1 + d_2\right)}{\delta_\ell}\right)},$$

$$\nu = \frac{t}{\left(4 + S_0\right) C d_1 d_2 \sqrt{M \tau_\ell^E}} = \sqrt{\frac{2 \log\left(\frac{2(d_1+d_2)}{\delta_\ell}\right)}{M \tau_\ell^E d_1 d_2 C \left(4 + S_0^2\right)}},$$

we have with probability at least $1 - \delta_\ell$, *it holds that*

$$\mathbb{P}\left(\left\| \nabla L \left(\mu^* \mathbf{Z}_*\right) \right\|_{op} \geq \frac{2t}{\sqrt{M \tau_\ell^E}} \right) \leq \delta_\ell,$$

where $\mu^* = \frac{2}{M}\mathbb{E}\left[\langle \mathbf{X}_m, \mathbf{Z}_* \rangle\right] > 0$, *and* $\mathbf{X}_m = \mathbf{x}_m \mathbf{z}_m^\top$.

25

Proof. Let $\mathbf{Z}_* = \frac{1}{M} \sum_{m=1}^{M} \mathbf{\Theta}_{m,*}$. Let $\mathbf{X}_{m,i} = \mathbf{x}_{m,i} \mathbf{z}_{m,i}^\top$ for $i \in [\tau_\ell^E]$. Based on the definition of our loss function $L(\cdot)$ in (8), we have that

$$\nabla_{x_m} L(\mathbf{Z}_*) = \mu^* \mathbf{Z}_* - \frac{2}{M \tau_\ell^E} \sum_{m=1}^{M} \sum_{i=1}^{\tau_\ell^E} \widetilde{\psi}_\nu (r_{m,i} \cdot Q(x_m))$$

$$= \frac{2}{M} \mathbb{E}\left[\langle \mathbf{X}_{m,1}, \mathbf{Z}_* \rangle\right] \mathbf{Z}_* - \frac{2}{M \tau_\ell^E} \sum_{m=1}^{M} \sum_{i=1}^{\tau_\ell^E} \widetilde{\psi}_\nu (r_{m,i} \cdot Q(\mathbf{X}_{m,i}))$$

$$\overset{(a)}{=} \frac{2}{M} \mathbb{E}\left[\langle \mathbf{X}_{m,1}, \mathbf{Z}_* \rangle Q(\mathbf{X}_{m,1})\right] - \frac{2}{M \tau_\ell^E} \sum_{m=1}^{M} \sum_{i=1}^{\tau_\ell^E} \widetilde{\psi}_\nu (r_{m,i} \cdot Q(\mathbf{X}_{m,i}))$$

$$\overset{(b)}{=} \frac{2}{M}\left[\mathbb{E}(r_{m,1} \cdot Q(\mathbf{X}_{m,1})) - \frac{1}{M \tau_\ell^E} \sum_{m=1}^{M} \sum_{i=1}^{\tau_\ell^E} \widetilde{\psi}_\nu (r_{m,i} \cdot Q(\mathbf{X}_{m,i}))\right]$$

where we have (a) due to the generalized Stein's Lemma stated in Lemma 1, and (b) comes from the fact that the random noise $\eta_1 = y_1 - \langle \mathbf{X}_1, \mathbf{Z}_* \rangle$ is zero-mean and independent from $\mathbf{X}_1$. Therefore, in order to implement the Lemma 2, we can see that it suffices to get σ^2 defined as:

$$\sigma^2 = \max\left(\left\|\frac{2}{M} \sum_{m=1}^{M} \sum_{j=1}^{\tau_\ell^E} \mathbb{E}\left[r_{m,j}^2 Q(\mathbf{X}_{m,j}) Q(\mathbf{X}_{m,j})^\top\right]\right\|_{\mathrm{op}}, \left\|\frac{2}{M} \sum_{m=1}^{M} \sum_{j=1}^{\tau_\ell^E} \mathbb{E}\left[r_{m,j}^2 Q(\mathbf{X}_{m,j})^\top Q(\mathbf{X}_{m,j})\right]\right\|_{\mathrm{op}}\right).$$

$$\left\|\frac{2}{M} \sum_{m=1}^{M} \sum_{j=1}^{\tau_\ell^E} \mathbb{E}\left[r_{m,j}^2 Q(\mathbf{X}_{m,j}) Q(\mathbf{X}_{m,j})^\top\right]\right\|_{\mathrm{op}} \leq \tau_\ell^E \times \left\|\mathbb{E}\left[r_{m,1}^2 Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top\right]\right\|_{\mathrm{op}}$$

$$\overset{(a)}{=} \tau_\ell^E \times \left\|\mathbb{E}\left[(\eta_{m,1} + \langle \mathbf{X}_{m,1}, \mathbf{Z}_* \rangle)^2 Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top\right]\right\|_{\mathrm{op}}$$

$$\overset{(b)}{=} \tau_\ell^E \times \left\|\mathbb{E}\left[\eta_{m,1}^2 Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top\right] + \mathbb{E}\left[\langle \mathbf{X}_{m,1}, \mathbf{Z}_* \rangle)^2 Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top\right]\right\|_{\mathrm{op}}$$

$$\overset{(c)}{=} \tau_\ell^E \times \left\|\mathbb{E}(\eta_{m,1}^2) \mathbb{E}\left[Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top\right] + \mathbb{E}\left[\langle \mathbf{X}_{m,1}, \mathbf{Z}_* \rangle)^2 Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top\right]\right\|_{\mathrm{op}}$$

$$\overset{(d)}{\leq} \tau_\ell^E \times \left\|4\mathbb{E}\left[Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top\right] + S_0^2 \mathbb{E}\left[Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top\right]\right\|_{\mathrm{op}}$$

$$= (4 + S_0^2) \tau_\ell^E \times \left\|\mathbb{E}\left[Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top\right]\right\|_{\mathrm{op}}$$

where the (a) follows by plugging in the definition for reward, (b) follows by the linearity of expectation, (c) follows as noises are independent, and the inequality (d) comes from the fact that $|\langle \mathbf{X}_{m,1}, \mathbf{Z}_* \rangle| \leq S_0$, and $Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top$ is always positive semidefinite. Next, since we know that $\mathbb{E}\left[Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top\right]$ is always symmetric and positive semidefinite, and hence we have

$$\left\|\mathbb{E}\left[Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top\right]\right\|_{\mathrm{op}} \overset{(a)}{\leq} \left\|\mathbb{E}\left[Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top\right]\right\|_{\mathrm{nuc}} = \mathrm{trace}\left(\mathbb{E}\left[Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top\right]\right)$$

$$= \mathbb{E}\left[\mathrm{trace}\left(Q(\mathbf{X}_{m,1}) Q(\mathbf{X}_{m,1})^\top\right)\right] = \mathbb{E}\left(\frac{2}{M} \sum_{m=1}^{M} \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} Q_{ij}(\mathbf{X}_{m,1})^2\right)$$

$$\leq d_1 d_2 C.$$

where, in (a) $\|\cdot\|_{\mathrm{nuc}}$ denotes the nuclear norm. Therefore, we have that under 1-subGaussian assumption

$$\left\|\frac{2}{M} \sum_{m=1}^{M} \sum_{j=1}^{\tau_\ell^E} \mathbb{E}\left[r_{m,j}^2 Q(\mathbf{X}_{m,j}) Q(\mathbf{X}_{m,j})^\top\right]\right\|_{\mathrm{op}} \leq (4 + S_0^2) d_1 d_2 \tau_\ell^E C.$$

26

And similarly, we can prove that

$$\left\| \frac{2}{M} \sum_{m=1}^{M} \sum_{j=1}^{\tau_\ell^E} \mathbb{E}\left[r_{m,j}^2 Q\left(\mathbf{X}_{m,j}\right)^\top Q\left(\mathbf{X}_{m,j}\right) \right] \right\|_{\mathrm{op}} \leq \left(4 + S_0^2\right) d_1 d_2 \tau_\ell^E C.$$

Therefore, we can take $\sigma^2 = \left(4 + S_0^2\right) d_1 d_2 \tau_\ell^E C$ consequently. By using Lemma 2, we have

$$\mathbb{P}\left(\|\nabla L\left(\mu^* \mathbf{Z}_*\right)\|_{\mathrm{op}} \geq \frac{2t}{\sqrt{M\tau_\ell^E}} \right) \leq 2\left(d_1 + d_2\right) \exp\left(-\nu t \sqrt{M\tau_\ell^E} + \frac{\nu^2 \left(4 + S_0^2\right) C d_1 d_2 \tau_\ell^E}{2} \right)$$

By plugging the values of t and ν in Lemma 11, we finish the proof. $\qquad\square$

Lemma 12. *For any low-rank linear model with samples $\mathbf{X}_1 \ldots, \mathbf{X}_{\tau_\ell^E}$ drawn from $\mathcal{X}$ according to $\mathcal{D}$ then for the optimal solution to the nuclear norm regularization problem in (2) with $\nu = \sqrt{2 \log\left(2\left(d_1 + d_2\right)/\delta_\ell\right) / \left(\left(4 + S_0^2\right) M\tau_\ell^E d_1 d_2\right)}$ and*

$$\gamma_\ell = 4\sqrt{\frac{2\left(4 + S_0^2\right) C d_1 d_2 \log\left(2\left(d_1 + d_2\right)/\delta_\ell\right)}{M\tau_\ell^E}},$$

with probability at least $1 - \delta_\ell$ it holds that:

$$\left\| \widehat{\mathbf{Z}}_\ell - \mu^* \mathbf{Z}_* \right\|_F^2 \leq \frac{C_1 d_1 d_2 r \log\left(\frac{2(d_1 + d_2)}{\delta_\ell} \right)}{M\tau_\ell^E},$$

for $C_1 = 36\left(4 + S_0^2\right) C$, $\|\mathbf{X}\|_F, \|\mathbf{Z}_*\|_F \leq S_0$, some nonzero constant μ^*, and $\mathbb{E}\left[\left(S^\mathbf{P}(\mathbf{X})\right)_{ij}^2 \right] \leq C, \forall i, j$. Summing over all phases $\ell \geq 1$ it follows that $\mathbb{P}(\mathcal{F}) \geq 1 - \delta/2$.

Proof. Since the estimator $\widehat{\mathbf{Z}}_\ell$ minimizes the regularized loss function defined in (8), we have

$$L(\widehat{\mathbf{Z}}_\ell) + \gamma_\ell \|\widehat{\mathbf{Z}}_\ell\|_{\mathrm{nuc}} \leq L\left(\mu^* \mathbf{Z}_*\right) + \gamma_\ell \left\| \mu^* \mathbf{Z}_* \right\|_{\mathrm{nuc}}$$

And due to the fact that $L(\cdot)$ is a quadratic function, we have the following expression based on multivariate Taylor's expansion:

$$L(\widehat{\mathbf{Z}}_\ell) - L\left(\mu^* \mathbf{Z}_*\right) = \langle \nabla L\left(\mu^* \mathbf{Z}_*\right), \mathbf{\Theta} \rangle + 2\|\mathbf{\Theta}\|_F^2, \quad \text{where } \mathbf{\Theta} = \widehat{\mathbf{Z}}_\ell - \mu^* \mathbf{Z}_*$$

By rearranging the above two results, we can deduce that

$$
\begin{aligned}
2\|\mathbf{\Theta}\|_F^2 &\leq -\langle \nabla L\left(\mu^* \mathbf{Z}_*\right), \mathbf{\Theta} \rangle + \gamma_\ell \left\| \mu^* \mathbf{Z}_* \right\|_{\mathrm{nuc}} - \gamma_\ell \|\widehat{\mathbf{Z}}_\ell\|_{\mathrm{nuc}} \\
&\overset{(a)}{\leq} \|\nabla L\left(\mu^* \mathbf{Z}_*\right)\|_{\mathrm{op}} \|\mathbf{\Theta}\|_{\mathrm{nuc}} + \gamma_\ell \left\| \mu^* \mathbf{Z}_* \right\|_{\mathrm{nuc}} - \gamma_\ell \|\widehat{\mathbf{Z}}_\ell\|_{\mathrm{nuc}},
\end{aligned}
\tag{27}
$$

where (a) comes from the duality between matrix operator norm and nuclear norm. Next, we represent the saturated SVD of $\mathbf{Z}_*$ in the main paper as $\mathbf{Z}_* = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ where $\mathbf{U} \in \mathbb{R}^{d_1 \times r}$ and $\mathbf{V} \in \mathbb{R}^{d_2 \times r}$, and here we would work on its full version, i.e.

$$\mathbf{Z}_* = \left(\mathbf{U}, \mathbf{U}_\perp\right) \begin{pmatrix} \mathbf{D} & 0 \\ 0 & 0 \end{pmatrix} \left(\mathbf{V}, \mathbf{V}_\perp\right)^\top = \left(\mathbf{U}, \mathbf{U}_\perp\right) \mathbf{D}^* \left(\mathbf{V}, \mathbf{V}_\perp\right)^\top$$

where we have $\mathbf{U}_\perp \in \mathbb{R}^{d_1 \times (d_1 - r)}, \mathbf{D}^* \in \mathbb{R}^{d_1 \times d_2}$ and $\mathbf{V}_\perp \in \mathbb{R}^{d_2 \times (d_2 - r)}$. Furthermore, we define

$$\mathbf{\Lambda} = \left(\mathbf{U}, \mathbf{U}_\perp\right)^\top \mathbf{\Theta} \left(\mathbf{V}, \mathbf{V}_\perp\right) = \begin{pmatrix} \mathbf{U}^\top \mathbf{\Theta} \mathbf{V} & \mathbf{U}^\top \mathbf{\Theta} \mathbf{V}_\perp \\ \mathbf{U}_\perp^\top \mathbf{\Theta} \mathbf{V} & \mathbf{U}_\perp^\top \mathbf{\Theta} V_\perp \end{pmatrix} = \mathbf{\Lambda}_1 + \mathbf{\Lambda}_2$$

where we write

$$\mathbf{\Lambda}_1 = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{U}_\perp^\top \mathbf{\Theta} \mathbf{V}_\perp \end{pmatrix}, \quad \mathbf{\Lambda}_2 = \begin{pmatrix} \mathbf{U}^\top \mathbf{\Theta} \mathbf{V} & \mathbf{U}^\top \mathbf{\Theta} \mathbf{V}_\perp \\ \mathbf{U}_\perp^\top \mathbf{\Theta} \mathbf{V} & 0 \end{pmatrix}.$$

27

Afterward, it holds that

$$\|\widehat{\mathbf{Z}}_\ell\|_{\mathrm{nuc}} = \|\mu^*\mathbf{Z}_* + \mathbf{\Theta}\|_{\mathrm{nuc}} \overset{(a)}{=} \left\|(\mathbf{U},\mathbf{U}_\perp)(\mu^*\mathbf{D}^* + \mathbf{\Lambda})(\mathbf{V},\mathbf{V}_\perp)^\top\right\|_{\mathrm{nuc}}$$

$$\overset{(b)}{=} \|\mu^*\mathbf{D}^* + \mathbf{\Lambda}\|_{\mathrm{nuc}} + \|\mu^*\mathbf{D}^* + \mathbf{\Lambda}_1 + \mathbf{\Lambda}_2\|_{\mathrm{nuc}}$$
$$\geq \|\mu^*\mathbf{D}^* + \mathbf{\Lambda}_1\|_{\mathrm{nuc}} - \|\mathbf{\Lambda}_2\|_{\mathrm{nuc}}$$
$$= \|\mu^*\mathbf{D}\|_{\mathrm{nuc}} + \|\mathbf{\Lambda}_1\|_{\mathrm{nuc}} - \|\mathbf{\Lambda}_2\|_{\mathrm{nuc}}$$
$$= \|\mu^*\mathbf{Z}_*\|_{\mathrm{nuc}} + \|\mathbf{\Lambda}_1\|_{\mathrm{nuc}} - \|\mathbf{\Lambda}_2\|_{\mathrm{nuc}}, \tag{28}$$

where, (a) follows from the definition of $\mathbf{Z}_*$, and (b) follows from the definition of $\mathbf{\Lambda}$. This implies that

$$\|\mu^*\mathbf{Z}_*\|_{\mathrm{nuc}} - \|\widehat{\mathbf{Z}}_\ell\|_{\mathrm{nuc}} \leq \|\mathbf{\Lambda}_2\|_{\mathrm{nuc}} - \|\mathbf{\Lambda}_1\|_{\mathrm{nuc}}.$$

Combining (27) and (28), we have that

$$2\|\mathbf{\Theta}\|_F^2 \leq \left(\|\nabla L(\mu^*\mathbf{Z}_*)\|_{\mathrm{op}} + \gamma_\ell\right)\|\mathbf{\Lambda}_2\|_{\mathrm{nuc}} + \left(\|\nabla L(\mu^*\mathbf{Z}_*)\|_{\mathrm{op}} - \gamma_\ell\right)\|\mathbf{\Lambda}_1\|_{\mathrm{nuc}}.$$

Then, we refer to the setting in our Lemma 11, and we choose $\gamma_\ell = 4t/\sqrt{M\tau_\ell^E}$ where the value of t is determined in Lemma 11, i.e.

$$\gamma_\ell = 4\sqrt{\frac{2(4 + S_0^2)Cd_1d_2\log(2(d_1 + d_2)/\delta_\ell)}{M\tau_\ell^E}},$$

we know that $\lambda_{T-1} \geq 2\|\nabla L(\mu^*\mathbf{Z}_*)\|_{op}$ with probability at least $1-\delta_\ell$ for any $\delta_\ell \in (0,1)$. Therefore, with a probability at least $1 - \delta_\ell$, we have

$$2\|\mathbf{\Theta}\|_F^2 \leq \frac{3}{2}\gamma_\ell\|\mathbf{\Lambda}_2\|_{\mathrm{nuc}} - \frac{1}{2}\gamma_\ell\|\mathbf{\Lambda}_1\|_{\mathrm{nuc}} \leq \frac{3}{2}\gamma_\ell\|\mathbf{\Lambda}_2\|_{\mathrm{nuc}}.$$

Since we can easily verify that the rank of $\mathbf{\Lambda}_2$ is at most $2r$, and by using Cauchy-Schwarz Inequality we have that

$$2\|\mathbf{\Theta}\|_F^2 \leq \frac{3}{2}\gamma_\ell\sqrt{2r}\|\mathbf{\Lambda}_2\|_F \leq \frac{3}{2}\gamma_\ell\sqrt{2r}\|\mathbf{\Lambda}\|_F = \frac{3}{2}\gamma_\ell\sqrt{2r}\|\mathbf{\Theta}\|_F$$

which implies that

$$\|\mathbf{\Theta}\|_F \leq \frac{3}{4}\sqrt{2r}\gamma_\ell = 6\sqrt{\frac{(4 + S_0^2)Cd_1d_2r\log\left(\frac{2(d_1+d_2)}{\delta_\ell}\right)}{M\tau_\ell^E}}.$$

This implies that $\mathbb{P}(\mathcal{F}_\ell) \geq 1-\delta_\ell$. Taking a union bound over all phases $\ell \geq 1$ and recalling $\delta_\ell := \frac{\delta}{2\ell^2}$, we obtain

$$\mathbb{P}(\mathcal{F}) \geq 1 - \sum_{\ell=1}^\infty \mathbb{P}(\mathcal{F}_\ell^c)$$

$$\geq 1 - \sum_{\ell=1}^\infty \frac{\delta_\ell}{2}$$

$$= 1 - \sum_{\ell=1}^\infty \frac{\delta}{4\ell^2}$$

$$\geq 1 - \frac{\delta}{2}.$$

This concludes our proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Define $\mathbf{X}_{\mathrm{batch}}^+ := \left(\mathbf{X}_{\mathrm{batch}}^\top\mathbf{X}_{\mathrm{batch}}\right)^{-1}\mathbf{X}_{\mathrm{batch}}^\top$ where $\mathbf{X}_{\mathrm{batch}}^+$ is constructed through the E-optimal design. Using Lemma C.1 from Du et al. (2023) it holds that

$$\left\|\mathbf{X}_{\mathrm{batch}}^+\right\| \leq \sqrt{\frac{(1+\beta)\rho_\ell^E}{\overline{p}}}.$$

where $\overline{p} = 180d_1d_2/\beta^2$ is the batch size to control the rounding procedure and $\rho_\ell^E = \min_{\mathbf{b}\in\triangle_{\overline{\mathcal{W}}}}\|(\sum_{\overline{\mathbf{w}}\in\overline{\mathcal{W}}}\mathbf{b}_{\overline{\mathbf{w}}}\overline{\mathbf{w}}\,\overline{\mathbf{w}}^\top)^{-1}\|$. It follows then that $\left\|\mathbf{X}_{\mathrm{batch}}^+\right\|^2 \leq 4\rho_\ell^E$.

28

Lemma 13. *(Expectation of $\widehat{\mathbf{Z}}_\ell$). It holds that $\mathbb{E}\left[\widehat{\mathbf{Z}}_\ell\right] = \mathbf{Z} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{\Theta}_m.$*

Proof. Note that we can re-write

$$\widehat{\mathbf{Z}}_\ell = \underset{\mathbf{\Theta}\in\mathbb{R}^{d_1\times d_2}}{\arg\min} \ L_\ell(\mathbf{\Theta}) + \gamma_\ell\|\mathbf{\Theta}\|_{\mathrm{nuc}}, L_\ell(\mathbf{\Theta}) = \langle\mathbf{\Theta},\mathbf{\Theta}\rangle - \frac{2}{M\tau_\ell^E}\sum_{m=1}^{M}\sum_{s=1}^{\tau_\ell^E}\langle\widetilde{\psi}_\nu(r_{m,s}\cdot Q(\mathbf{x}_{m,s}\mathbf{z}_{m,s}^\top)),\mathbf{\Theta}\rangle$$

such that

$$\widehat{\mathbf{Z}}_\ell = \frac{2}{M\tau_\ell^E}\sum_{m=1}^{M}\sum_{s=1}^{\tau_\ell^E}\widehat{\mathbf{\Theta}}_{m,s} - \boldsymbol{X}_{\mathrm{batch}}^+\left(\boldsymbol{X}_{\mathrm{batch}}^+\right)^\top$$

where $\widehat{\mathbf{\Theta}}_{m,s} = \langle\mathbf{\Theta},\mathbf{\Theta}\rangle - \frac{2}{Ms}\sum_{m=1}^{M}\langle\widetilde{\psi}_\nu(r_{m,s}\cdot Q(\mathbf{x}_{m,s}\mathbf{z}_{m,s}^\top)),\mathbf{\Theta}\rangle$. Now using Lemma C.2 from Du et al. (2023) we can prove the result of the lemma. $\qquad\square$

Lemma 14. *(Concentration of $\widehat{\mathbf{B}}_{1,\ell}$). Suppose that event $\mathcal{F}_\ell$ holds. Then, for any phase $\ell > 0$,*

$$\left\|(\widehat{\mathbf{B}}_{1,\ell}^\perp)^\top\mathbf{B}_1\right\| \le \frac{c'\rho_\ell^E\sqrt{(d_1+d_2)r}}{S_r\sqrt{M\tau_\ell^E}}\log\left(\frac{16(d_1+d_2)rM\tau_\ell^E}{\delta_\ell}\right),$$

for some constant $c' > 0$ and $\rho_\ell^E = \min_{\mathbf{b}\in\triangle_{\overline{\mathcal{W}}}}\left\|(\sum_{\overline{\mathbf{w}}\in\overline{\mathcal{W}}}\mathbf{b}_{\overline{\mathbf{w}}}\,\overline{\mathbf{w}}\,\overline{\mathbf{w}}^\top)^{-1}\right\|.$

Proof. Using the Davis-Kahan $\sin\theta$ Theorem (Bhatia, 2013) and letting τ_ℓ^E be large enough to satisfy $\left\|\widehat{\mathbf{Z}}_\ell - \mu^*\mathbf{Z}_*\right\|_F^2 \le \frac{C_1 d_1 d_2 r\log\left(\frac{2(d_1+d_2)}{\delta_\ell}\right)}{M\tau_\ell^E}$, we have

$$
\begin{aligned}
\left\|(\widehat{\mathbf{B}}_{1,\ell}^\perp)^\top\mathbf{B}_1\right\| &\le \frac{\left\|\widehat{\mathbf{Z}}_\ell - \mathbb{E}\left[\widehat{\mathbf{Z}}_\ell\right]\right\|}{\sigma_r\left(\mathbb{E}\left[\widehat{\mathbf{Z}}_\ell\right]\right) - \sigma_{r+1}\left(\mathbb{E}\left[\widehat{\mathbf{Z}}_\ell\right]\right) - \left\|\widehat{\mathbf{Z}}_\ell - \mathbb{E}\left[\widehat{\mathbf{Z}}_\ell\right]\right\|} \\
&\overset{(a)}{\le} \frac{c_0}{S_r}\left\|\widehat{\mathbf{Z}}_\ell - \mathbb{E}\left[\widehat{\mathbf{Z}}_\ell\right]\right\| \\
&\overset{(b)}{\le} \frac{cc_0\left\|\mathbf{X}_{\mathrm{batch}}^+\right\|^2\sqrt{(d_1+d_2)r}}{S_r\sqrt{M\tau_\ell^E}}\log\left(\frac{16(d_1+d_2)rM\tau_\ell^E}{\delta_\ell}\right) \\
&\overset{(c)}{\le} \frac{c'\rho_\ell^E\sqrt{(d_1+d_2)r}}{S_r\sqrt{M\tau_\ell^E}}\log\left(\frac{16(d_1+d_2)rM\tau_\ell^E}{\delta_\ell}\right).
\end{aligned}
$$

where, (a) follows from Assumption 2, the (b) follows from event $\mathcal{F}_\ell$ and (b) follows as $\left\|\mathbf{X}_{\mathrm{batch}}^+\right\|^2 \le 4\rho_\ell^E$. The claim of the lemma follows. $\qquad\square$

Lemma 15. *(Concentration of $\widehat{\mathbf{B}}_{2,\ell}$). Suppose that event $\mathcal{F}_\ell$ holds. Then, for any phase $\ell > 0$,*

$$\left\|(\widehat{\mathbf{B}}_{2,\ell}^\perp)^\top\mathbf{B}_2\right\| \le \frac{c\rho_\ell^E\sqrt{(d_1+d_2)r}}{S_r\sqrt{M\tau_\ell^E}}\log\left(\frac{16(d_1+d_2)rM\tau_\ell^E}{\delta_\ell}\right),$$

for some constant $c' > 0$ and $\rho_\ell^E = \min_{\mathbf{b}\in\triangle_{\overline{\mathcal{W}}}}\left\|(\sum_{\overline{\mathbf{w}}\in\overline{\mathcal{W}}}\mathbf{b}_{\overline{\mathbf{w}}}\,\overline{\mathbf{w}}\,\overline{\mathbf{w}}^\top)^{-1}\right\|.$

Proof. The proof follows the same way as Lemma 14 and using the Davis-Kahan $\sin\theta$ Theorem (Bhatia, 2013)

$$\left\|(\widehat{\mathbf{B}}_{2,\ell}^\perp)^\top\mathbf{B}_2\right\| \le \frac{c\rho_\ell^E\sqrt{(d_1+d_2)r}}{S_r\sqrt{M\tau_\ell^E}}\log\left(\frac{16(d_1+d_2)rM\tau_\ell^E}{\delta_\ell}\right).$$

The claim of the lemma follows. $\qquad\square$

29

Good Event per Task: We now define the good event $\mathcal{F}'_\ell$ in phase ℓ that GOBLIN has a good estimate of $\mathbf{S}_{m,*}$ as follows: For any phase $\ell > 0$

$$\mathcal{F}'_\ell := \left\{ \left\| \widehat{\mathbf{S}}_{m,\ell} - \mu^* \mathbf{S}_{m,*} \right\|_F^2 \leq \frac{C_1 k_1 k_2 r \log \left(\frac{2(k_1+k_2)}{\delta_\ell} \right)}{\tau^E_{m,\ell}} \right\}, \tag{29}$$

where, $C, \mu^* > 0$ are constants and $\widehat{\mathbf{S}}_{m,\ell}$ is the estimate from (9). Then define the event

$$\mathcal{F}' := \bigcap_{\ell=1}^\infty \mathcal{F}'_\ell. \tag{30}$$

We now prove the following lemmas to show the good event $\mathcal{F}'_\ell$ holds with probability $(1 - \delta_\ell)$. Before proving the concentration of $\widehat{\mathbf{S}}_{m,\ell}$ we first need to show that $\sigma_{\min}(\sum_{\widetilde{\mathbf{w}} \in \widetilde{\mathcal{W}}} \mathbf{b}_{\widetilde{\mathbf{w}}} \widetilde{\mathbf{w}} \widetilde{\mathbf{w}}^\top) > 0$. If this holds true then we can sample following E-optimal design.

Lemma 16. *For any phase $\ell > 0$ and task $m \in [M]$, let $\left\| \widehat{\mathbf{B}}_{1,\ell}^\top \mathbf{B}_1^\perp \right\| \leq c_1$ and $\left\| \widehat{\mathbf{B}}_{2,\ell}^\top \mathbf{B}_2^\perp \right\| \leq c_2$, for some $c_1, c_2 > 0$. Then we have*

$$\sigma_{\min}(\sum_{\widetilde{\mathbf{w}} \in \widetilde{\mathcal{W}}} \mathbf{b}_{\widetilde{\mathbf{w}}} \widetilde{\mathbf{w}} \widetilde{\mathbf{w}}^\top) > 0$$

Proof. We can show that

$$\sum_{\widetilde{\mathbf{w}} \in \widetilde{\mathcal{W}}} \mathbf{b}_{\widetilde{\mathbf{w}}} \widetilde{\mathbf{w}} \widetilde{\mathbf{w}}^\top \overset{(a)}{=} \sum_{\mathbf{x} \in \mathcal{X}_m, \mathbf{z} \in \mathcal{Z}_m} \mathbf{b}_{x,z} \underbrace{\widehat{\mathbf{U}}_\ell^\top \mathbf{x}_m}_{\widetilde{\mathbf{g}}_m} \underbrace{\mathbf{z}_m \widehat{\mathbf{V}}_\ell^\top}_{\widetilde{\mathbf{v}}_m^\top}$$

where, in (a) the $\mathbf{b}_{x,z}$ is the sampling proportion for the arms $\mathbf{x}$ and $\mathbf{z}$ (they are allocated the same proportion, as they are pulled the same number of times). Also note that from Lemma 12 we know that $\left\| \widehat{\mathbf{B}}_{1,\ell}^\top \mathbf{B}_1^\perp \right\| \leq c_1$ and $\left\| \widehat{\mathbf{B}}_{2,\ell}^\top \mathbf{B}_2^\perp \right\| \leq c_2$ for some $c_1, c_2 > 0$ holds with high probability. This helps us to apply Lemma 17 to get the claim of the lemma. $\square$

Lemma 17. *(Restatement of Lemma C.5 from Du et al. (2023)) For any phase $\ell > 0$ and task $m \in [M]$, if $\left\| \widehat{\mathbf{U}}_\ell^\top \mathbf{U}^\perp \right\| \leq c$ for some $c > 0$, then we have*

$$\sigma_{\min} \left(\sum_{i=1}^n \mathbf{b}_m^* (\mathbf{x}_i) \widehat{\mathbf{U}}_\ell^\top \mathbf{x}_i \mathbf{x}_i^\top \widehat{\mathbf{U}}_\ell \right) > 0$$

where $\mathbf{b}_m^*$ is a sampling proportion on $\mathbf{x}$.

Lemma 18. *Let $L' : \mathbb{R}^{k_1 \times k_2} \to \mathbb{R}$ is the loss function defined in (9). Then by setting*

$$t = \sqrt{2k_1 k_2 C (4 + S_0^2) \log \left(\frac{2(k_1+k_2)}{\delta_\ell} \right)},$$

$$\nu = \frac{t}{(4 + S_0) C k_1 k_2 \sqrt{\tau^E_\ell}} = \sqrt{\frac{2 \log \left(\frac{2(k_1+k_2)}{\delta_\ell} \right)}{\tau^E_\ell k_1 k_2 C (4 + S_0^2)}},$$

we have with probability at least $1 - \delta_\ell$, it holds that

$$\mathbb{P} \left(\left\| \nabla L' (\mu^* \mathbf{S}_{m,*}) \right\|_{op} \geq \frac{2t}{\sqrt{\tau^E_\ell}} \right) \leq \delta_\ell,$$

where $\mu^* = \mathbb{E} \left[\langle \mathbf{X}_m, \mathbf{S}_{m,*} \rangle \right] > 0$, and $\mathbf{X}_m = \widetilde{\mathbf{g}}_m \widetilde{\mathbf{v}}_m^\top$.

30

Proof. Let $\mathbf{X}_{m,i} = \widetilde{\mathbf{g}}_{m,i} \widetilde{\mathbf{v}}_{m,i}^\top$ Based on the definition of our loss function $L'(\cdot)$ in (9), we have that

$$\nabla_{x_m} L'(\mathbf{S}_{m,*}) = \mu^* \mathbf{S}_{m,*} - \frac{2}{\tau_\ell^E} \sum_{i=1}^{\tau_\ell^E} \widetilde{\psi}_\nu(r_{m,i} \cdot Q(x_m))$$

$$\stackrel{(a)}{=} \left[\mathbb{E}(r_{m,1} \cdot Q(\mathbf{X}_{m,1})) - \frac{1}{\tau_\ell^E} \sum_{i=1}^{\tau_\ell^E} \widetilde{\psi}_\nu(r_{m,i} \cdot Q(\mathbf{X}_{m,i})) \right]$$

where (a) follows using the same steps as in Lemma 11. Similarly, using the same steps for a single task as in Lemma 11 we have

$$\mathbb{P}\left(\|\nabla L'(\mu^* \mathbf{S}_{m,*})\|_{\mathrm{op}} \geq \frac{2t}{\sqrt{\tau_\ell^E}} \right) \leq 2(k_1 + k_2) \exp\left(-\nu t \sqrt{\tau_\ell^E} + \frac{\nu^2 (4 + S_0^2) C k_1 k_2 \tau_\ell^E}{2} \right)$$

By plugging the values of t and ν in Lemma 11, we finish the proof. $\qquad\square$

Lemma 19. *(**Concentration of** $\widehat{\mathbf{S}}_{m,\ell}$) For any low-rank linear model with samples $\mathbf{X}_1 \ldots, \mathbf{X}_{\tau_\ell^E}$ drawn from $\mathcal{X}$ according to $\mathcal{D}$ then for the optimal solution to the nuclear norm regularization problem in (2) with $\nu = \sqrt{2 \log(2(k_1 + k_2)/\delta_\ell)/((4 + S_0^2) \tau_\ell^E k_1 k_2)}$ and*

$$\gamma_{m,\ell} = 4 \sqrt{\frac{2(4 + S_0^2) C k_1 k_2 \log(2(k_1 + k_2)/\delta_\ell)}{\tau_{m,\ell}^E}},$$

with probability at least $1 - \delta_\ell$ it holds that:

$$\left\| \widehat{\mathbf{S}}_{m,\ell} - \mu^* \mathbf{S}_{m,*} \right\|_F^2 \leq \frac{C_1 k_1 k_2 r \log\left(\frac{2(k_1 + k_2)}{\delta_\ell} \right)}{\tau_{m,\ell}^E},$$

for $C_1 = 36(4 + S_0^2) C$, $\|\mathbf{X}\|_F, \|\mathbf{S}_{m,*}\|_F \leq S_0$, some nonzero constant μ^*, and $\mathbb{E}\left[(S^\mathbf{P}(\mathbf{X}))_{ij}^2 \right] \leq C, \forall i, j$. Summing over all phases $\ell \geq 1$ it follows that $\mathbb{P}(\mathcal{F}') \geq 1 - \delta/2$.

Proof. Since the estimator $\widehat{\mathbf{S}}_{m,\ell}$ minimizes the regularized loss function defined in Eqn. (6), we have

$$L(\widehat{\mathbf{S}}_{m,\ell}) + \gamma_\ell \|\widehat{\mathbf{S}}_{m,\ell}\|_{\mathrm{nuc}} \leq L(\mu^* \mathbf{S}_{m,*}) + \gamma_\ell \|\mu^* \mathbf{S}_{m,*}\|_{\mathrm{nuc}}.$$

And due to the fact that $L'(\cdot)$ is a quadratic function, we have the following expression based on multivariate Taylor's expansion:

$$L'(\widehat{\mathbf{S}}_{m,\ell}) - L'(\mu^* \mathbf{S}_{m,*}) = \langle \nabla L'(\mu^* \mathbf{S}_{m,*}), \mathbf{\Theta} \rangle + 2\|\mathbf{\Theta}\|_F^2, \quad \text{where } \mathbf{\Theta} = \widehat{\mathbf{S}}_{m,\ell} - \mu^* \mathbf{S}_{m,*}.$$

By rearranging the above two results, we can deduce that

$$2\|\mathbf{\Theta}\|_F^2 \leq -\langle \nabla L'(\mu^* \mathbf{S}_{m,*}), \mathbf{\Theta} \rangle + \gamma_\ell \|\mu^* \mathbf{S}_{m,*}\|_{\mathrm{nuc}} - \gamma_\ell \|\widehat{\mathbf{S}}_{m,*}\|_{\mathrm{nuc}}$$

$$\stackrel{(i)}{\leq} \|\nabla L'(\mu^* \mathbf{S}_{m,*})\|_{\mathrm{op}} \|\mathbf{\Theta}\|_{\mathrm{nuc}} + \gamma_\ell \|\mu^* \mathbf{S}_{m,*}\|_{\mathrm{nuc}} - \gamma_\ell \|\widehat{\mathbf{S}}_{m,\ell}\|_{\mathrm{nuc}}, \qquad (31)$$

where (i) comes from the duality between matrix operator norm and nuclear norm. Next, we represent the saturated SVD of $\mathbf{S}_{m,*}$ as $\mathbf{S}_{m,*} = \mathbf{U} \mathbf{D} \mathbf{V}^\top$ where $\mathbf{U} \in \mathbb{R}^{k_1 \times r}$ and $\mathbf{V} \in \mathbb{R}^{k_2 \times r}$, and here we would work on its full version, i.e.

$$\mathbf{S}_{m,*} = (\mathbf{U}, \mathbf{U}_\perp) \begin{pmatrix} \mathbf{D} & 0 \\ 0 & 0 \end{pmatrix} (\mathbf{V}, \mathbf{V}_\perp)^\top = (\mathbf{U}, \mathbf{U}_\perp) \mathbf{D}^* (\mathbf{V}, \mathbf{V}_\perp)^\top$$

where we have $\mathbf{U}_\perp \in \mathbb{R}^{k_1 \times (k_1 - r)}, \mathbf{D}^* \in \mathbb{R}^{k_1 \times k_2}$ and $\mathbf{V}_\perp \in \mathbb{R}^{k_2 \times (k_2 - r)}$. Furthermore, we define

$$\mathbf{\Lambda} = (\mathbf{U}, \mathbf{U}_\perp)^\top \mathbf{\Theta} (\mathbf{V}, \mathbf{V}_\perp) = \begin{pmatrix} \mathbf{U}^\top \mathbf{\Theta} \mathbf{V} & \mathbf{U}^\top \mathbf{\Theta} \mathbf{V}_\perp \\ \mathbf{U}_\perp^\top \mathbf{\Theta} \mathbf{V} & \mathbf{U}_\perp^\top \mathbf{\Theta} \mathbf{V}_\perp \end{pmatrix} = \mathbf{\Lambda}_1 + \mathbf{\Lambda}_2$$

31

where we write

$$\Lambda_1 = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{U}_\perp^\top \Theta \mathbf{V}_\perp \end{pmatrix}, \quad \Lambda_2 = \begin{pmatrix} \mathbf{U}^\top \Theta \mathbf{V} & \mathbf{U}^\top \Theta \mathbf{V}_\perp \\ \mathbf{U}_\perp^\top \Theta \mathbf{V} & 0 \end{pmatrix}.$$

Afterward, it holds that

$$
\begin{aligned}
\|\widehat{\mathbf{S}}_\ell\|_{\mathrm{nuc}} = \|\mu^* \mathbf{S}_{m,*} + \Theta\|_{\mathrm{nuc}} &\overset{(a)}{=} \left\| (\mathbf{U}, \mathbf{U}_\perp)(\mu^* \mathbf{D}^* + \Lambda)(\mathbf{V}, \mathbf{V}_\perp)^\top \right\|_{\mathrm{nuc}} \\
&\overset{(b)}{=} \|\mu^* \mathbf{D}^* + \Lambda\|_{\mathrm{nuc}} + \|\mu^* \mathbf{D}^* + \Lambda_1 + \Lambda_2\|_{\mathrm{nuc}} \\
&\geq \|\mu^* \mathbf{D}^* + \Lambda_1\|_{\mathrm{nuc}} - \|\Lambda_2\|_{\mathrm{nuc}} \\
&= \|\mu^* \mathbf{D}\|_{\mathrm{nuc}} + \|\Lambda_1\|_{\mathrm{nuc}} - \|\Lambda_2\|_{\mathrm{nuc}} \\
&= \|\mu^* \mathbf{S}_{m,*}\|_{\mathrm{nuc}} + \|\Lambda_1\|_{\mathrm{nuc}} - \|\Lambda_2\|_{\mathrm{nuc}},
\end{aligned}
\tag{32}
$$

where, (a) follows from the definition of $\mathbf{S}_{m,*}$, and (b) follows the definition of Λ. This implies that

$$\|\mu^* \mathbf{S}_{m,*}\|_{\mathrm{nuc}} - \|\widehat{\mathbf{S}}_{m,\ell}\|_{\mathrm{nuc}} \leq \|\Lambda_2\|_{\mathrm{nuc}} - \|\Lambda_1\|_{\mathrm{nuc}}.$$

Combining (31) and (32), we have that

$$2\|\Theta\|_F^2 \leq \left(\|\nabla L(\mu^* \mathbf{B})\|_{\mathrm{op}} + \gamma_\ell \right) \|\Lambda_2\|_{\mathrm{nuc}} + \left(\|\nabla L(\mu^* \mathbf{B})\|_{\mathrm{op}} - \gamma_\ell \right) \|\Lambda_1\|_{\mathrm{nuc}}.$$

Then, we refer to the setting in our Lemma 11, and we choose $\gamma_\ell = 4t/\sqrt{M\tau_\ell^E}$ where the value of t is determined in Lemma 11, i.e.

$$\gamma_\ell = 4\sqrt{\frac{2(4 + S_0^2) C k_1 k_2 \log(2(d_1 + d_2)/\delta_\ell)}{M\tau_\ell^E}},$$

we know that $\lambda_{T-1} \geq 2\|\nabla L(\mu^* \mathbf{S}_{m,*})\|_{\mathrm{op}}$ with probability at least $1 - \delta_\ell$ for any $\delta_\ell \in (0,1)$. Therefore, with a probability at least $1 - \delta_\ell$, we have

$$2\|\Theta\|_F^2 \leq \frac{3}{2}\gamma_\ell \|\Lambda_2\|_{\mathrm{nuc}} - \frac{1}{2}\gamma_\ell \|\Lambda_1\|_{\mathrm{nuc}} \leq \frac{3}{2}\gamma_\ell \|\Lambda_2\|_{\mathrm{nuc}}.$$

Since we can easily verify that the rank of Λ_2 is at most $2r$, and by using Cauchy-Schwarz Inequality we have that

$$2\|\Theta\|_F^2 \leq \frac{3}{2}\gamma_\ell \sqrt{2r} \|\Lambda_2\|_F \leq \frac{3}{2}\gamma_\ell \sqrt{2r}\|\Lambda\|_F = \frac{3}{2}\gamma_\ell \sqrt{2r}\|\Theta\|_F$$

which implies that

$$\|\Theta\|_F \leq \frac{3}{4}\sqrt{2r}\gamma_\ell = 6\sqrt{\frac{(4 + S_0^2) C k_1 k_2 r \log\left(\frac{2(k_1 + k_2)}{\delta_\ell}\right)}{\tau_\ell^E}}$$

This implies that $\mathbb{P}(\mathcal{F}'_\ell) \geq 1 - \delta_\ell$. Taking a union bound over all phases $\ell \geq 1$ and recalling $\delta_\ell := \frac{\delta}{2\ell^2}$, we obtain

$$
\begin{aligned}
\mathbb{P}(\mathcal{F}') &\geq 1 - \sum_{\ell=1}^\infty \mathbb{P}\left((\mathcal{F}')_\ell^c\right) \\
&\geq 1 - \sum_{\ell=1}^\infty \frac{\delta_\ell}{2} \\
&= 1 - \sum_{\ell=1}^\infty \frac{\delta}{4\ell^2} \\
&\geq 1 - \frac{\delta}{2}.
\end{aligned}
$$

This concludes our proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

32

A.7 Final Sample Complexity Bound

We first define the arm elimination event similar to Theorem 1. For any $\mathcal{V} \subseteq \underline{\mathcal{W}}$ be the active set and $\underline{\mathbf{w}} \in \mathcal{V}$ define

$$\mathcal{E}_{\underline{\mathbf{w}},\ell}(\mathcal{V}) = \left\{ \left| \left\langle \underline{\mathbf{w}} - \underline{\mathbf{w}}^\star, \widehat{\boldsymbol{\theta}}_\ell(\mathcal{V}) - \boldsymbol{\theta}^* \right\rangle \right| \leq \epsilon_\ell \right\} \tag{33}$$

where it is implicit that $\widehat{\boldsymbol{\theta}}_\ell := \widehat{\boldsymbol{\theta}}_\ell(\mathcal{V})$ is the design constructed in the algorithm at stage ℓ with respect to $\underline{\mathcal{W}}_\ell = \mathcal{V}$.

Theorem 2. *(Restatement)* *With probability at least* $1 - \delta$, *multi-task* GOBLIN *returns the best arms* $\mathbf{x}_*$, $\mathbf{z}_*$, *and the number of samples used is bounded by*

$$\widetilde{O}\left(\frac{M(k_1 + k_2)r}{\Delta^2} + \frac{M\sqrt{k_1 k_2 r}}{S_r} + \frac{\sqrt{d_1 d_2 r}}{S_r} \right).$$

Proof. For the rest of the proof we have that the good events $\mathcal{F}_\ell \bigcap \mathcal{F}'_\ell \bigcap \mathcal{E}_{\mathbf{w},\ell}(\underline{\mathcal{W}}_\ell)$ holds true for each phase ℓ with probability greater than $(1 - \delta)$. The three events are defined in (25), (29) and (33).

Third Stage: Define $\underline{\mathcal{A}}_{m,\ell} = \{\underline{\mathbf{w}} \in \mathcal{W}_\ell : \langle \underline{\mathbf{w}}^\star - \underline{\mathbf{w}}, \boldsymbol{\theta}^* \rangle \leq 4\epsilon_{m,\ell}\}$. Note that by assumption $\underline{\mathcal{W}} = \underline{\mathcal{W}}_1 = \underline{\mathcal{A}}_1$. The above lemma implies that with probability at least $1 - \delta$ we have $\bigcap_{\ell=1}^\infty \{\underline{\mathcal{W}}_{m,\ell} \subseteq S_{m,\ell}\}$. This implies that

$$\rho^G\left(\underline{\mathcal{W}}_{m,\ell}\right) = \min_{\mathbf{b} \in \Delta_{\underline{\mathcal{W}}_m}} \max_{\mathbf{w},\mathbf{w}' \in \underline{\mathcal{W}}_{m,\ell}} \|\mathbf{w} - \mathbf{w}'\|^2_{\left(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}} \, \underline{\mathbf{w}}^\top + \mathbf{\Lambda}\right)^{-1}}$$

$$\leq \min_{\mathbf{b} \in \Delta_{\underline{\mathcal{W}}_m}} \max_{\mathbf{w},\mathbf{w}' \in S_{m,\ell}} \|\mathbf{w} - \mathbf{w}'\|^2_{\left(\sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}} \mathbf{b}_{\underline{\mathbf{w}}} \underline{\mathbf{w}} \, \underline{\mathbf{w}}^\top + \mathbf{\Lambda}\right)^{-1}}$$

$$= \rho^G\left(\underline{\mathcal{A}}_{m,\ell}\right).$$

Let the effective dimension be $k = (k_1 + k_2)r$. Define $k_1^\ell = 8k\log(1 + \tau^G_{m,\ell-1}/\lambda)$. For $\ell \geq \lceil \log_2\left(4\Delta_m^{-1}\right) \rceil$ we have that $S_{m,\ell} = \{\underline{\mathbf{w}}^\star\}$, thus, the sample complexity to identify $\underline{\mathbf{w}}^\star_m$ is equal to

$$\sum_{\ell=1}^{\lceil \log_2\left(4\Delta_m^{-1}\right) \rceil} \sum_{\underline{\mathbf{w}} \in \underline{\mathcal{W}}_m} \left\lceil \tau^G_{m,\ell} \widehat{\mathbf{b}}^G_{m,\ell,\underline{\mathbf{w}}} \right\rceil = \sum_{\ell=1}^{\lceil \log_2\left(4\Delta_m^{-1}\right) \rceil} \left(\frac{(k_1^\ell + 1)k_1^\ell}{2} + \tau^G_{m,\ell} \right)$$

$$= \sum_{\ell=1}^{\lceil \log_2\left(4\Delta_m^{-1}\right) \rceil} \left(\frac{(k_1^\ell + 1)k_1^\ell}{2} + 2\epsilon_{m,\ell}^{-2}\rho^G(\underline{\mathcal{W}}_{m,\ell})B^\ell_{m,*}\log\left(4k_1^\ell \ell^2 |\underline{\mathcal{W}}_m|/\delta\right) \right)$$

$$\overset{(a)}{\leq} 2\sum_{\ell=1}^{\lceil \log_2\left(4\Delta_m^{-1}\right) \rceil} \left(\frac{(k+1)k}{2}\log^2(1 + \tau^G_{m,\ell-1}) + 2\epsilon_{m,\ell}^{-2}\rho^G(\underline{\mathcal{W}}_{m,\ell})B^\ell_{m,*}\log\left(4k_1^\ell \ell^2 |\underline{\mathcal{W}}_m|/\delta\right) \right)$$

$$\overset{(b)}{\leq} 2\frac{(k+1)k}{2}\sum_{\ell=1}^{\lceil \log_2\left(4\Delta_m^{-1}\right) \rceil} \left(\log^2(1 + \tau^G_{m,\ell-1}) + 8\epsilon_{m,\ell}^{-2}\rho^G(\underline{\mathcal{W}}_{m,\ell})B^\ell_{m,*}\log\left(4k_1^\ell \ell^2 |\underline{\mathcal{W}}_m|/\delta\right) \right)$$

$$\overset{(c)}{\leq} (k+1)k\sum_{\ell=1}^{\lceil \log_2\left(4\Delta_m^{-1}\right) \rceil} \left(1 + 16\epsilon_{m,\ell}^{-2}\rho^G(\underline{\mathcal{W}}_{m,\ell})B^\ell_{m,*}\log^2\left(4k_1^\ell \ell^2 |\underline{\mathcal{W}}_m|/\delta\right) \right)$$

$$\overset{(d)}{\leq} (k+1)k\left\lceil \log_2\left(4\Delta_m^{-1}\right) \right\rceil + \sum_{\ell=1}^{\lceil \log_2\left(4\Delta_m^{-1}\right) \rceil} 32\epsilon_{m,\ell}^{-2}f\left(\underline{\mathcal{A}}_{m,\ell}\right)B^\ell_{m,*}\log\left(4k\ell^2 |\underline{\mathcal{W}}_m|/\delta\right)$$

$$\overset{(e)}{\leq} (k+1)k\left\lceil \log_2\left(4\Delta^{-1}\right) \right\rceil + \sum_{\ell=1}^{\lceil \log_2\left(4\Delta^{-1}\right) \rceil} 32\epsilon_\ell^{-2}f\left(\underline{\mathcal{A}}_{m,\ell}\right)(64\lambda S^2 + 64\tau^G_{m,\ell-1})\log\left(4k\ell^2 |\underline{\mathcal{W}}|/\delta\right)$$

33

$$= (k+1)k \left\lceil \log_2\left(4\Delta^{-1}\right)\right\rceil + \sum_{\ell=1}^{\left\lceil \log_2\left(4\Delta^{-1}\right)\right\rceil} 32\epsilon_\ell^{-2} f\left(\underline{\mathcal{A}}_{m,\ell}\right)(64\lambda S^2)\log\left(4k\ell^2|\mathcal{W}|/\delta\right)$$

$$+ (k+1)k \left\lceil \log_2\left(4\Delta^{-1}\right)\right\rceil + \sum_{\ell=1}^{\left\lceil \log_2\left(4\Delta^{-1}\right)\right\rceil} 32\epsilon_\ell^{-2} f\left(\underline{\mathcal{A}}_{m,\ell}\right)(64\tau_{m,\ell-1}^G)\log\left(4k\ell^2|\mathcal{W}|/\delta\right)$$

$$\overset{(f)}{\leq} (k+1)k \left\lceil \log_2\left(4\Delta_m^{-1}\right)\right\rceil + 2048\lambda S^2 \log\left(\frac{4k\log_2^2\left(8\Delta_m^{-1}\right)|\mathcal{W}_m|}{\delta}\right) \sum_{\ell=1}^{\left\lceil \log_2\left(4\Delta_m^{-1}\right)\right\rceil} 2^{2\ell} f\left(\underline{\mathcal{A}}_{m,\ell}\right).$$

where, (a) follows as $\log^2(1+\tau_{m,\ell-1}^G/\lambda) \leq \log^2(1+\tau_{m,\ell-1}^G)$, (b) follows by noting that $\log(x\log(1+x)) \leq 2\log(x)$ for any $x > 1$. The (c) follows by subsuming the $\log^2(1+\tau_{m,\ell-1}^G)$ into $2\tau_\ell^G$. The (d) follows as $\log(1+\tau_{m,\ell-1}^G) < \tau_\ell^G$ which enables us to replace the k_1^ℓ inside the log with an addtional factor of 2. The (e) follows similarly to (23) by noting that

$$B_{m,*}^\ell \leq 64(\sqrt{\lambda}S + \sqrt{\lambda_{m,\ell}^\perp}S_{m,\ell}^\perp)$$

$$\leq 64\lambda S^2 + \left(\frac{64\tau_{m,\ell-1}^G}{8(d_1+d_2)r\log(1+\frac{\tau_{m,\ell-1}^G}{\lambda})}\right) \cdot \left(\frac{8d_1 d_2 r}{\tau_\ell^E S_r^2}\log\left(\frac{d_1+d_2}{\delta_\ell}\right)\right)$$

$$\overset{(a_1)}{\leq} 64\lambda S^2 + 64\tau_{m,\ell-1}^G.$$

Finally the (f) follows by subsuming the $\tau_{\ell-1}^G$ with a factor of 2 into the quantity of τ_ℓ^G. Then it follows that

$$\rho_{m,*}^G = \inf_{\mathbf{b}\in\triangle_{\mathcal{W}_m}} \max_{\underline{\mathbf{w}}\in\mathcal{W}_m} \frac{\|\underline{\mathbf{w}} - \underline{\mathbf{w}}^\star\|_{\left(\sum_{\underline{\mathbf{w}}\in\mathcal{W}_m} \mathbf{b}_{\underline{\mathbf{w}}}\underline{\mathbf{w}}\,\underline{\mathbf{w}}^\top + \mathbf{\Lambda}\right)^{-1}}^2}{(\langle\underline{\mathbf{w}} - \underline{\mathbf{w}}^\star, \boldsymbol{\theta}^*\rangle)^2}$$

$$= \inf_{\mathbf{b}\in\triangle_{\mathcal{W}_m}} \max_{\ell\leq\lceil\log_2(4\Delta_m^{-1})\rceil} \max_{\underline{\mathbf{w}}\in S_{m,\ell}} \frac{\|\underline{\mathbf{w}} - \underline{\mathbf{w}}^\star\|_{\left(\sum_{\underline{\mathbf{w}}\in\mathcal{W}_m} \mathbf{b}_{\underline{\mathbf{w}}}\underline{\mathbf{w}}\,\underline{\mathbf{w}}^\top + \mathbf{\Lambda}\right)^{-1}}^2}{(\langle\underline{\mathbf{w}} - \underline{\mathbf{w}}^\star, \boldsymbol{\theta}^*\rangle)^2}$$

$$\geq \frac{1}{\lceil\log_2\left(4\Delta_m^{-1}\right)\rceil} \inf_{\mathbf{b}\in\triangle_{\mathcal{W}_m}} \sum_{\ell=1}^{\lceil\log_2(4\Delta_m^{-1})\rceil} \max_{\underline{\mathbf{w}}\in\mathcal{A}_{m,\ell}} \frac{\|\underline{\mathbf{w}} - \underline{\mathbf{w}}^\star\|_{\left(\sum_{\underline{\mathbf{w}}\in\mathcal{W}_m} \mathbf{b}_{\underline{\mathbf{w}}}\underline{\mathbf{w}}\,\underline{\mathbf{w}}^\top + \mathbf{\Lambda}\right)^{-1}}^2}{(\langle\underline{\mathbf{w}} - \underline{\mathbf{w}}^\star, \boldsymbol{\theta}^*\rangle)^2}$$

$$\geq \frac{1}{16\lceil\log_2\left(4\Delta_m^{-1}\right)\rceil} \sum_{\ell=1}^{\lceil\log_2(4\Delta_m^{-1})\rceil} 2^{2\ell} \inf_{\mathbf{b}\in\triangle_{\mathcal{W}_m}} \max_{\underline{\mathbf{w}}\in\mathcal{A}_{m,\ell}} \|\underline{\mathbf{w}} - \underline{\mathbf{w}}^\star\|_{\left(\sum_{\underline{\mathbf{w}}\in\mathcal{W}_m} \mathbf{b}_{\underline{\mathbf{w}}}\underline{\mathbf{w}}\,\underline{\mathbf{w}}^\top + \mathbf{\Lambda}\right)^{-1}}^2$$

$$\geq \frac{1}{64\lceil\log_2\left(4\Delta_m^{-1}\right)\rceil} \sum_{\ell=1}^{\lceil\log_2(4\Delta_m^{-1})\rceil} 2^{2\ell} \inf_{\mathbf{b}\in\triangle_{\mathcal{W}_m}} \max_{\underline{\mathbf{w}},\underline{\mathbf{w}}'\in\mathcal{A}_{m,\ell}} \|\underline{\mathbf{w}} - \underline{\mathbf{w}}'\|_{\left(\sum_{\underline{\mathbf{w}}\in\mathcal{W}_m} \mathbf{b}_{\underline{\mathbf{w}}}\underline{\mathbf{w}}\,\underline{\mathbf{w}}^\top + \mathbf{\Lambda}\right)^{-1}}^2$$

$$\geq \frac{1}{64\lceil\log_2\left(4\Delta_m^{-1}\right)\rceil} \sum_{\ell=1}^{\lceil\log_2(4\Delta_m^{-1})\rceil} 2^{2\ell} f\left(\underline{\mathcal{A}}_{m,\ell}\right).$$

This implies that

$$\sum_{\ell=1}^{\lceil\log_2(4\Delta_m^{-1})\rceil} 2^{2\ell} f\left(\underline{\mathcal{A}}_{m,\ell}\right) \leq \rho_{m,*}^G 64\left\lceil\log_2\left(4\Delta_m^{-1}\right)\right\rceil.$$

34

Plugging this back we get

$$\sum_{\ell=1}^{\lceil \log_2(4\Delta_m^{-1}) \rceil} \sum_{\mathbf{w} \in \mathcal{W}_m} \left\lceil \tau_{m,\ell}^G \widehat{\mathbf{b}}_{\ell,\mathbf{w}} \right\rceil \leq (k+1)k \left\lceil \log_2\left(4\Delta_m^{-1}\right) \right\rceil$$

$$+ 2048\lambda S^2 \log\left(\frac{8k \log_2^2\left(8\Delta_m^{-1}\right) |\mathcal{W}_m|}{\delta}\right) 64\rho_{m,*}^G \left\lceil \log_2\left(4\Delta_m^{-1}\right) \right\rceil$$

$$\leq (k+1)k \left\lceil \log_2\left(4\Delta_m^{-1}\right) \right\rceil$$

$$+ C_2\lambda S^2 \log\left(\frac{8k \log_2^2\left(8\Delta_m^{-1}\right) |\mathcal{W}_m|}{\delta}\right) \rho_{m,*}^G \left\lceil \log_2\left(4\Delta_m^{-1}\right) \right\rceil$$

where, $C_2 > 0$ is a constant. Summing over each task m, the simplified sample complexity for the third stage is given by

$$N_3 \leq O\left(\frac{Mk}{\Delta^2} \log\left(\frac{k \log_2\left(\Delta^{-1}\right) |\mathcal{W}_m|}{\delta}\right)\right) = \widetilde{O}\left(\frac{M(k_1+k_2)r}{\Delta^2}\right)$$

where $\Delta = \min_{\mathbf{w} \in \mathcal{W}}(\mathbf{w}_* - \mathbf{w})^\top \boldsymbol{\theta}_* \stackrel{(a1)}{=} \min_{\mathbf{x} \in \mathcal{X} \setminus \{\mathbf{x}_*\}, \mathbf{z} \in \mathcal{Z} \setminus \{\mathbf{z}_*\}}(\mathbf{x}_*^\top \boldsymbol{\Theta}_* \mathbf{z}_* - \mathbf{x}^\top \boldsymbol{\Theta}_* \mathbf{z})$. The $(a1)$ follows by reshaping the arms in $\underline{\mathcal{W}}$ to recover the arms in $\mathcal{X}$ and $\mathcal{Z}$.

2nd Stage: Again recall that the E-optimal design in stage 2 of Algorithm 2 satisfies the Assumption 4 as the sample distribution $\mathcal{D}$ has finite second order moments.

For the second stage first observe that by plugging in the definition of $\widetilde{\tau}_E^\ell$ we get

$$\left\| \boldsymbol{\theta}_{m,k+1:p}^* \right\|_2^2 = \sum_{i>r \wedge j>r} H_{ij}^2 = \left\| (\widehat{\mathbf{U}}_\ell^\perp)^\top \left(\mathbf{U}^*\mathbf{S}^*\mathbf{V}^{*\top}\right) \widehat{\mathbf{V}}_\ell^\perp \right\|_F^2$$

$$\leq \left\| (\widehat{\mathbf{U}}_\ell^\perp)^\top \mathbf{U}^* \right\|_F^2 \|\mathbf{S}^*\|_2^2 \left\| (\widehat{\mathbf{V}}_\ell^\perp)^\top \mathbf{V}^* \right\|_F^2 \leq O\left(\frac{k_1 k_2 r}{\tau_\ell^E} \log\left(\frac{k_1+k_2}{\delta}\right)\right)$$

$$= O\left(\frac{\sqrt{d_1 d_2 r}}{S_r} \log\left(\frac{d_1+d_2}{\delta_\ell}\right)\right),$$

which implies $\left\| \boldsymbol{\theta}_{k+1:p}^* \right\|_2 = \widetilde{O}\left(\sqrt{k_1 k_2 r}/S_r\right)$. We also set $\frac{8k_1 k_2 r}{\tau_{m,\ell}^E S_r^2} \log\left(\frac{k_1+k_2}{\delta_\ell}\right) := S_{m,\ell}^\perp$. Now we bound the sample complexity from the second stage. From the second stage we can show that we have for the arm set $\widetilde{\mathcal{W}}_m$

$$N_2 = \sum_{\ell=1}^{\lceil \log_2(4\Delta^{-1}) \rceil} \sum_{\widetilde{\mathbf{w}} \in \widetilde{\mathcal{W}}_m} \left\lceil \widetilde{\tau}_{m,\ell}^E \widehat{\mathbf{b}}_{m,\ell,\widetilde{\mathbf{w}}}^E \right\rceil$$

$$= \sum_{\ell=1}^{\lceil \log_2(4\Delta^{-1}) \rceil} \left(\frac{(p+1)p}{2} + \widetilde{\tau}_{m,\ell}^E\right)$$

$$= \sum_{\ell=1}^{\lceil \log_2(4\Delta^{-1}) \rceil} \left(\frac{(p+1)p}{2} + \frac{\sqrt{8k_1 k_2 r \log\left(4\ell^2 |\mathcal{W}|/\delta\right)}}{S_r}\right)$$

$$\leq (p+1)p \left\lceil \log_2\left(4\Delta^{-1}\right) \right\rceil + 32\frac{\sqrt{k_1 k_2 r}}{S_r} \log\left(\frac{4 \log_2^2\left(8\Delta^{-1}\right) |\mathcal{W}|}{\delta}\right) \left\lceil \log_2\left(4\Delta^{-1}\right) \right\rceil$$

$$= O\left(\frac{\sqrt{k_1 k_2 r}}{S_r} \log\left(\frac{4 \log_2^2\left(8\Delta^{-1}\right) |\mathcal{W}|}{\delta}\right)\right) \stackrel{(a)}{=} \widetilde{O}\left(\frac{\sqrt{k_1 k_2 r}}{S_r}\right)$$

1st Stage: Finally we also use the E-optimal design in first stage of Algorithm 2. Note that this design satisfies the Assumption 4 as the sample distribution $\mathcal{D}$ has finite second-order moments. Now

35

we bound the sample complexity from the first stage. From the first stage we can show that we have for the arm set $\overline{\mathcal{W}}$

$$
\begin{aligned}
N_1 &= \sum_{\ell=1}^{\lceil \log_2(4\Delta^{-1}) \rceil} \sum_{m=1}^{M} \sum_{\overline{\mathbf{w}} \in \overline{\mathcal{W}}_m} \left\lceil \tau_{m,\ell}^E \widehat{\mathbf{b}}_{m,\ell,\overline{\mathbf{w}}}^E \right\rceil \\
&= \sum_{\ell=1}^{\lceil \log_2(4\Delta^{-1}) \rceil} \left(\frac{M(p+1)p}{2} + \sum_{m} \tau_{m,\ell}^E \right) \\
&= \sum_{\ell=1}^{\lceil \log_2(4\Delta^{-1}) \rceil} \left(\frac{M(p+1)p}{2} + \frac{\sqrt{8d_1 d_2 r \log\left(4\ell^2 |\mathcal{W}|/\delta\right)}}{S_r} \right) \\
&\leq M(p+1)p \left\lceil \log_2\left(4\Delta^{-1}\right) \right\rceil + 32 \frac{\sqrt{d_1 d_2 r}}{S_r} \log\left(\frac{4 \log_2^2\left(8\Delta^{-1}\right) |\mathcal{W}|}{\delta} \right) \left\lceil \log_2\left(4\Delta^{-1}\right) \right\rceil \\
&\overset{(a)}{=} O\left(\frac{\sqrt{d_1 d_2 r}}{S_r} \log\left(\frac{4 \log_2^2\left(8\Delta^{-1}\right) |\mathcal{W}|}{\delta} \right) \right) \overset{(a)}{=} \widetilde{O}\left(\frac{\sqrt{d_1 d_2 r}}{S_r} \right)
\end{aligned}
$$

where, (a) follows as $p = d_1 d_2$. Combining N_1, N_2 and N_3 gives the claim of the theorem. $\qquad\square$

A.8 Additional Experimental Details

Single Task Unit Ball: This experiment consists of a set of $\{6, 10, 14\}$ left and right arms that are arranged in a unit ball in $\mathbb{R}^6$, and $\|\mathbf{x}\| = 1$, $\|\mathbf{z}\| = 1$ for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{z} \in \mathcal{Z}$. Hence, we have $d_1 \in \mathbb{R}^6$ and $d_2 \in \mathbb{R}^6$. We choose a random $\mathbf{\Theta}_* \in \mathbb{R}^{d_1 \times d_2}$ which has rank $r = 2$. We set $\delta = 0.1$. We compare against RAGE (Fiez et al., 2019) that treats this $d_1 d_2$ bilinear bandit as a linear bandit setting and suffers a sample complexity that scales as $\widetilde{O}(d_1 d_2 / \Delta^2)$. We do a continuous relaxation of the algorithm when implementing it to make this more tractable.

Multi-task Unit Ball: This experiment consists of a set of $\{5, 10, 15, 20, 25, 30\}$ tasks. For each task, we choose left and right arms that are arranged in a unit ball in $\mathbb{R}^8$, and $\|\mathbf{x}\| = 1$, $\|\mathbf{z}\| = 1$ for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{z} \in \mathcal{Z}$. Hence, we have $d_1 \in \mathbb{R}^8$ and $d_2 \in \mathbb{R}^8$. We choose $k_1 = k_2 = 4$, and feature extractors $\mathbf{B}_1 \in \mathbb{R}^{d_1 \times k_1}$, $\mathbf{B}_2 \in \mathbb{R}^{d_2 \times k_2}$ shared across tasks. We choose a random matrix $\mathbf{S}_{m,*} \in \mathbb{R}^{k_1 \times k_2}$ for each task m such that $\mathbf{S}_{m,*}$ has rank $r = 2$. We set $\delta = 0.1$. We compare against DouExpDes (Du et al., 2023) that treats this setting as M $k_1 k_2$ bilinear bandits (after learning the feature extractors) and suffers a sample complexity that scales as $\widetilde{O}(M k_1 k_2 / \Delta^2)$ (see Remark 2). Again we do a continuous relaxation of the algorithm when implementing it to make this more tractable.

36

B Table of Notations

Notations	Definition
$\mathcal{X}$	Left arm set
$\mathcal{Z}$	Right arm set
M	Number of tasks
ℓ	Phase number
$\boldsymbol{\Theta}_{m,*}$	Hidden parameter matrix for
$\mathbf{b}_\ell^E$	E-optimal design at the ℓ-th phase
$\mathbf{b}_{m,\ell}^G$	G-optimal design at the ℓ-th phase for the m-th task
$S_{m,\ell}^\perp$	$\frac{8d_1 d_2 r}{\tau_\ell^E S_r^2} \log\left(\frac{d_1 + d_2}{\delta_\ell}\right)$
$\lambda_{m,\ell}^\perp$	$\tau_{m,\ell-1}^G / 8(d_1 + d_2) r \log(1 + \frac{\tau_{m,\ell-1}^G}{\lambda})$
$B_{m,*}^\ell$	$(8\sqrt{\lambda} S + \sqrt{\lambda_{m,\ell}^\perp S_{m,\ell}^\perp})$
$\mathbf{B}_1$	Left feature extractor
$\mathbf{B}_2$	Right feature extractor
S_r	r-th largest singular value of $\boldsymbol{\Theta}_*$
$\Delta(\mathbf{x}, \mathbf{z})$	$\mathbf{x}_*^\top \boldsymbol{\Theta}_* \mathbf{z}_* - \mathbf{x}^\top \boldsymbol{\Theta}_* \mathbf{z}$
Δ	$\min_{\mathbf{x} \neq \mathbf{x}_*, \mathbf{z} \neq \mathbf{z}_*} \Delta(\mathbf{x}, \mathbf{z})$
$\mathcal{Y}(\mathcal{W})$	$\{\mathbf{w} - \mathbf{w}' : \forall \mathbf{w}, \mathbf{w}' \in \mathcal{W}, \mathbf{w} \neq \mathbf{w}'\}$
$\mathcal{Y}^*(\mathcal{W})$	$\{\mathbf{w}_* - \mathbf{w} : \forall \mathbf{w} \in \mathcal{W} \backslash \mathbf{w}_*\}$
δ	confidence level

Table 1: Table of Notations

37