Multi-task Representation Learning for Fixed Budget Pure-Exploration in Linear and Bilinear Bandits

Subhojyoti Mukherjee, Qiaomin Xie, Robert Nowak

Keywords: Multi-task learning, Pure Exploration, Linear Bandits, Bilinear Bandits

Summary

We study fixed-budget pure exploration settings for multi-task representation learning (MTRL) in linear and bilinear bandits. In fixed budget MTRL linear bandit setting the goal is to find the optimal arm of each of the tasks with high probability within a pre-specified budget. Similarly, in a fixed budget MTRL bilinear setting the goal is to find the optimal left and right arms of each of the tasks with high precision within the budget. In both of these MTRL settings, the tasks share a common low-dimensional linear representation. Therefore, the goal is to leverage this underlying structure to expedite learning and identify the optimal arm(s) of each of the tasks with high precision.

We prove the first lower bound for the fixed-budget linear MTRL setting that takes into account the shared structure across the tasks. Motivated from the lower bound we propose the algorithm FB-DOE that uses a *double experimental design* approach to allocate samples optimally to the arms across the tasks, and thereby first learn the shared common representation and then identify the optimal arm(s) of each task. This is the first study on fixed-budget pure exploration of MTRL in linear and bilinear bandits. Our results show that learning the shared representation, jointly with allocating actions across the tasks following a double experimental design approach, achieves a smaller probability of error than solving the tasks independently.

Contribution(s)

1. We formulate the first fixed-budget MTRL problem for the linear and bilinear bandit settings and establish the first lower bound for the fixed-budget MTRL linear bandit setting.
Context: Previous work of MTRL setting studied fixed confidence linear (Du et al., 2023) and bilinear bandits (Mukherjee et al., 2023b). We establish the first lower bound for the fixed-budget MTRL in linear bandit setting and show that probability of error scales as $\widetilde{\Omega}(M \exp(-n\Delta^2/H_{2,\text{lin}} \log_2 k))$. Our bound contains the worst case hardness parameter $H_{2,\text{lin}}$ instead of the true hardness parameter $H_{1,\text{lin}}$. The work Du et al. (2023); Mukherjee et al. (2023b) provides no such lower bounds for the pure exploration MTRL setting.
2. We propose a double experimental design algorithm for fixed-budget MTRL linear bandits setting and prove a tight upper bound on the probability of error.
Context: Our proposed algorithm for fixed-budget MTRL linear bandits has the probability of error scaling as $\widetilde{O}(M \exp(-n\Delta^2/H_{2,\text{lin}} \log_2 k))$. Therefore, the upper bound on the probability of error of our proposed algorithm matches the lower bound with respect to the parameters k, d, M, and worst case hardness $H_{2,\text{lin}}$. Previous work (Du et al., 2023) studied fixed confidence MTRL linear bandit setting.
3. We also extend our work to fixed-budget bilinear bandit settings and again propose a double experimental design algorithm.
Context: Our proposed algorithm achieves a probability of error that scales as $\widetilde{O}(M(\exp(-n\Delta^2)/H_{2,\text{bilin}} \log_2(k_1 + k_2)r)$. Previous work (Mukherjee et al., 2023b) studied fixed confidence MTRL bilinear bandit setting. We show the first upper bound on the probability of error in bilinear setting that has the worst case hardness parameter $H_{2,\text{bilin}}$ in the bound.

Multi-task Representation Learning for Fixed Budget Pure-Exploration in Linear and Bilinear Bandits

Subhojyoti Mukherjee[1,†], Qiaomin Xie[2], Robert Nowak[3]
subhomuk@adobe.com, {qiaomin.xie,rdnowak}@wisc.edu

[1]**Adobe Research, San Jose, CA, USA**
[2]**ISYE, University of Wisconsin – Madison**
[3]**ECE Department, University of Wisconsin – Madison**

[†] Work was done as a PhD candidate at the ECE Department, University of Wisconsin-Madison

Abstract

In this paper, we study fixed-budget pure exploration settings for multi-task representation learning (MTRL) in linear and bilinear bandits. In fixed budget MTRL linear bandit setting the goal is to find the optimal arm of each of the tasks with high probability within a pre-specified budget. Similarly, in a fixed budget MTRL bilinear setting the goal is to find the optimal left and right arms of each of the tasks with high precision within the budget. In both of these MTRL settings, the tasks share a common low-dimensional linear representation. Therefore, the goal is to leverage this underlying structure to expedite learning and identify the optimal arm(s) of each of the tasks with high precision. We prove the first lower bound for the fixed-budget linear MTRL setting that takes into account the shared structure across the tasks. Motivated from the lower bound we propose the algorithm FB-DOE that uses a *double experimental design* approach to allocate samples optimally to the arms across the tasks, and thereby first learn the shared common representation and then identify the optimal arm(s) of each task. This is the first study on fixed-budget pure exploration of MTRL in linear and bilinear bandits. Our results show that learning the shared representation, jointly with allocating actions across the tasks following a double experimental design approach, achieves a smaller probability of error than solving the tasks independently.

1 Introduction

In this paper, we study Multi-task Representation Learning (MTRL) for fixed budget pure exploration settings in linear and bilinear bandits. Both linear and bilinear bandits are an important class of sequential decision-making problems. The linear bandit setting shows up in a lot of real-world settings such as news content recommendation (Li et al., 2010), ad recommendation (Chu et al., 2011), online safe decision making (Kazerouni et al., 2017). Similarly, the bilinear bandit setting shows up in applications that require interactions between pairs of items. 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; Jun et al., 2019). 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).

We focus on the multi-task representation learning setting (Bengio et al., 1990; Schaul & Schmidhuber, 2010). In many decision-making problems there exists several 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; Du et al., 2023; Mukherjee et al., 2023b). 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 total number of required samples.

Moreover, in many settings, it is expensive to collect samples and the learner wants to identify the optimal arm with high precision within a pre-specified number of samples n. This is termed the fixed budget setting (Bubeck et al., 2009; Audibert et al., 2010; Azizi et al., 2022; Lalitha et al., 2023) and the goal of the learner is to minimize the probability of error in identifying the optimal arm(s). Previously the Katz-Samuels et al. (2020); Yang & Tan (2021); Azizi et al. (2022) studied the setting under a single task linear bandit setting without representation learning component. Recent work (Du et al., 2023; Mukherjee et al., 2023b) focused on the fixed confidence setting for the MTRL linear and bilinear bandits. Note that Carpentier & Locatelli (2016) have shown that fixed budget setting requires a different approach than fixed confidence as the strategy that is optimal in fixed confidence may not be achievable in fixed budget setting. Therefore, the fixed budget MTRL in linear and bilinear bandits is an important area of study that has remained underexplored.

In particular, if we directly apply an existing approach from linear bandits, such as OD-LinBAI (Yang & Tan, 2021) or GSE (Azizi et al., 2022), to the linear MTRL fixed budget setting, the resulting probability of error scales as $\widetilde{O}(M \exp(-n\Delta^2/d \log_2 d))$, where $\widetilde{O}(\cdot)$ hides other smaller factors, d is the dimension of the feature of the arms, and Δ is the minimum reward gap. In this paper, for illustration purpose, we consider OD-LinBAI as a representative algorithm for single task fixed-budget linear bandits. Similarly, in the bilinear MTRL fixed budget setting, the probability of error of OD-LinBAI scales as $\widetilde{O}(M \exp(-n\Delta^2/d_1 d_2 \log_2 d_1 d_2))$ where d_1, d_2 are the dimensions of the feature of the left and right arms, respectively. Meanwhile, the power of MTRL lies in leveraging the underlying shared representation across tasks to expedite learning, which further reduces the individual task learning to a low dimensional latent space. Importantly, for linear bandits the low dimensional latent features scale with latent dimension $k \ll d$; for bilinear bandits, the latent dimensions of left and right arms $k_1, k_2 \ll d_1, d_2$, and the rank of hidden parameter matrix scales as $r \ll \min\{k_1, k_2\}$. The performance of OD-LinBAI suffers as it treats the task individually, and fails to learn the shared representation and the latent features in low dimension. Hence the two questions to ask are these:

> **1)** Can we design a MTRL algorithm for fixed-budget pure exploration in linear bandits whose probability of error scales as $\widetilde{O}(M \exp(-n\Delta^2/k \log_2 k))$?
> **2)** Can we design a MTRL algorithm for fixed-budget pure exploration in bilinear bandits whose probability of error scales as $\widetilde{O}(M \exp(-n\Delta^2)/(k_1 + k_2) r \log_2((k_1 + k_2) r))$?

In this paper, we answer positively to the above questions and make the following novel contributions to the MTRL decision-making setting:

1) We formulate the fixed-budget MTRL problem for the linear and bilinear bandit setting. To our knowledge, this is the first work that explores MTRL for fixed-budget pure exploration in linear and bilinear bandits.

2) We establish the first lower bound for the fixed-budget MTRL in linear bandit setting and show that probability of error scales as $\tilde{\Omega}(M \exp(-n\Delta^2/H_{2,\text{lin}} \log_2 k))$, where $H_{2,\text{lin}}$ is the worst case hardness of the problem. We leave getting a lower bound with respect to true hardness $H_{1,\text{lin}}$ for future works.

3) Motivated by the lower bound we propose the algorithm **F**ixed **B**udget **D**ouble **O**ptimal **DE**sign (abbreviated as FB-DOE) for the fixed-budget MTRL in linear bandits whose probability of error scales as $\widetilde{O}(M\exp(-n\Delta^2/H_{2,\mathrm{lin}}\log_2 k))$. Therefore, FB-DOE upperbound matches the lower bound in the linear MTRL setting with respect to the parameters k, d, M, and $H_{2,\mathrm{lin}}$. This improves over OD-LinBAI whose probability of error scales as $\widetilde{O}(M\exp(-n\Delta^2/H'_{2,\mathrm{lin}}\log_2 d))$ and $H'_{2,\mathrm{lin}} > H_{2,\mathrm{lin}}$.

4) Our algorithm FB-DOE for the fixed-budget MTRL in bilinear bandits achieves a probability of error that scales as $\widetilde{O}(M(\exp(-n\Delta^2)/H_{2,\mathrm{bilin}}\log(k_1+k_2)r)$. This improves over OD-LinBAI whose probability of error scales as $\widetilde{O}(M\exp(-n\Delta^2/H'_{2,\mathrm{bilin}}\log_2 d_1 d_2))$ and $H'_{2,\mathrm{bilin}} > H_{2,\mathrm{bilin}}$.

2 MTRL Fixed Budget Linear Bandit

In this section, we study the linear fixed-budget MTRL bandits. We first introduce the setting in Section 2.1. Recall that our goal is to devise an algorithm for the *fixed-budget* linear MTRL setting. To this effect, we first present the lower bound for fixed-budget linear MTRL bandits in Section 2.2. Motivated by the lower bound, we then introduce the MTRL algorithm for the fixed-budget linear bandits in Section 2.3.

2.1 Preliminaries

We now introduce the linear MTRL setting (Yang et al., 2020; 2022; Du et al., 2023). We denote $[n] = \{1, 2, \ldots, n\}$. We consider a setting with M tasks, indexed by $m \in [M]$. Each task m consists of a set of arms denoted by $\mathcal{X} \subset \mathbb{R}^d$ and an unknown parameter $\boldsymbol{\theta}_{m,*} \in \mathbb{R}^d$. For each $\mathbf{x} \in \mathcal{X}$, $\|\mathbf{x}\|_2 \leq L_x$ for some L_x. In the linear bandit setting, at each round t, the learner chooses an arm $\mathbf{x}_{m,t} \in \mathcal{X}$ for each task m, and the expected reward is $\mathbf{x}_{m,t}^\top \boldsymbol{\theta}_{m,*}$. We assume that each $\boldsymbol{\theta}_{m,*}$ can be decomposed as $\boldsymbol{\theta}_{m,*} = \mathbf{B}\mathbf{w}_m$, where $\mathbf{B} \in \mathbb{R}^{d\times k}$ is shared across tasks, while $\mathbf{w}_m \in \mathbb{R}^k$ is task-specific (Yang et al., 2020; 2022). Let $\|\mathbf{w}_m\|_2 \leq 1$. We assume that $k \ll d$, $k \geq 2$ and $M \gg d$, hence $\mathbf{B}$ facilitates dimensionality reduction. In the context of MTRL, $\mathbf{B}$ is referred to as *feature extractor*, while $\mathbf{x}_{m,t}$ is termed as *rich observations*. The reward for task $m \in [M]$ at round t is:

$$r_{m,t} = \mathbf{x}_{m,t}^\top \boldsymbol{\theta}_{m,*} + \eta_{m,t} = \mathbf{x}_{m,t}^\top \mathbf{B}\mathbf{w}_m + \eta_{m,t} \overset{(a)}{=} \mathbf{g}_{m,t}^\top \mathbf{w}_m + \eta_{m,t}. \tag{1}$$

where $\eta_{m,t}$ represents independent zero-mean 1-sub-Gaussian noise, and in (a), $\mathbf{g}_{m,t}^\top \triangleq \mathbf{x}_{m,t}^\top \mathbf{B} \in \mathbb{R}^k$ denotes the latent feature. After the learner commits the batch of actions $\{\mathbf{x}_{m,t} : m \in [M]\}$, they receive the batch of rewards $\{r_{m,t} : m \in [M]\}$. The latent feature $\mathbf{g}_{m,t}$ is unknown to the learner and needs to be learnt for each task m, hence the term MTRL. Let i_m^* be the optimal arm in task m and define the gap $\Delta_{m,i}=(\mathbf{x}_{i_m^*}^\top - \mathbf{x}_i)^\top \boldsymbol{\theta}_{m,*}$ for $i \neq i_m^*$. WLOG we assume $i_m^* = 1$. For simplicity, we assume that the expected rewards of the arms are in descending order and that the best arm is unique. The goal is to identify the optimal arm i_m^* for each task $m \in [M]$.

2.2 Lower Bound for Linear Fixed Budget MTRL

In this section, we present the first lower bound for the fixed-budget linear MTRL setting. The key idea is to formulate the linear MTRL linear setting as a hypothesis-testing problem. To this effect, we first define an environment model for task m as D_{ij}^m consisting of A actions and J hypotheses with true hypothesis $\boldsymbol{\theta}_*^m = \boldsymbol{\theta}_{i,j}^m$ (ij-th column). This is shown in (2) where, each ι_{ij} is distinct and satisfies $\iota_{ij} < \beta/4J + \Gamma/4N$ for some $\beta > 0$, $N > \max_{m\in[M]} \frac{kd\log_2 k}{\Delta_{m,\min}}$. The $\boldsymbol{\theta}_{11}^m$ is the optimal hypothesis in D_{11}^m, $\boldsymbol{\theta}_{12}^m$ is the optimal hypothesis in D_{12}^m and so on such that for each D_{ij}^m and $i \in [N], j \in [J]$ we have column (i,j) as the optimal hypothesis. This is a general hypothesis testing setting where the functions $\mu_a(\boldsymbol{\theta}^m)$ can be thought of as linear functions of $\boldsymbol{\theta}^m$ such that $\mu_a(\boldsymbol{\theta}^m) = \mathbf{x}_m(a)^\top \boldsymbol{\theta}^m = \mathbf{x}_m(a)\mathbf{B}_i\mathbf{w}_j^m$ for some $i \in [N]$ and $j \in [J]$. Note that this environment is different than previously studied for single-task linear bandit setting of Huang et al. (2017); Lattimore & Szepesvári (2020) as they do not consider the shared feature extractor $\mathbf{B}$ and the latent parameters $\mathbf{w}_m$.

$$\begin{array}{cccccccccc}
\boldsymbol{\theta}^m & = & \mathbf{B}_1\mathbf{w}_1^m & \mathbf{B}_1\mathbf{w}_2^m & \mathbf{B}_1\mathbf{w}_3^m & \ldots & \mathbf{B}_i\mathbf{w}_j^m & \ldots & \mathbf{B}_N\mathbf{w}_J^m \\
\hline
\mu_1(\boldsymbol{\theta}^m) & = & \beta+\Gamma & \beta+\Gamma-(\frac{\beta}{J}+\frac{\Gamma}{N}) & \beta+\Gamma-(\frac{2\beta}{J}+\frac{2\Gamma}{N}) & \ldots & \beta+\Gamma-(\frac{(j-1)\beta}{J}+\frac{(i-1)\Gamma}{N}) & \ldots & \beta+\Gamma-(\frac{(N-1)\beta}{J}+\frac{(N-1)\Gamma}{N}) \\
\mu_2(\boldsymbol{\theta}^m) & = & \iota_{211} & \iota_{212} & \iota_{213} & \ldots & \iota_{2ij} & \ldots & \iota_{2NJ} \\
\vdots & & & & & \vdots & & & \\
\mu_A(\boldsymbol{\theta}^m) & = & \iota_{A11} & \iota_{A12} & \iota_{A13} & \ldots & \iota_{Aij} & \ldots & \iota_{ANJ}
\end{array} \tag{2}$$

Theorem 1. ***(Lower Bound)*** *Let $|\boldsymbol{\Theta}| = 2^d$ and $\boldsymbol{\theta}_{m,*} \in \boldsymbol{\Theta}$. Then any δ-PAC policy π in the linear MTRL setting suffers a total probability of error as $\Omega(\exp(-\frac{Mn}{\log_2 d}) + M\exp(-\frac{n}{H_{2,\text{lin}}\log_2 k}))$ for the environment in (2), where $H_{2,\text{lin}} = \max_{m\in[M]}\max_{2\leq i\leq k}\frac{i}{\Delta_{m,i}^2}$ is the hardness parameter.*

Discussion 1. Observe that Theorem 1 has two terms in the bound. The first term is the probability of error in estimating the feature extractor $\mathbf{B}$ that increases as the number of tasks M increases and depends on the ambient dimension d. The second term is the probability of error of misidentifying the optimal arm in each task m. This term scales with the number of tasks M and latent dimension $k \ll d$,. The problem complexity parameter $H_{2,lin}$ is present in the term 2, which captures the worst-case difficulty of identifying the optimal arm across tasks. Note that we do not get the true hardness $H_{1,\text{lin}} = \max_{m\in[M]}\sum_{i=1}^{k}\frac{1}{\Delta_{m,i}^2}$ in the lower bound, and we leave this to future works.

Proof (Overview:) The proof differs from the lower bound proof techniques of Carpentier & Locatelli (2016); Huang et al. (2017) for the structured bandit settings. We reduce our MTRL linear bandit problem to the hypothesis testing setting and construct a worst-case environment as in (2). The key technical novelty lies in constructing the worst-case environment in (2), which jointly scales with the number of tasks and the latent parameter $\mathbf{w}_m$, whereas (Huang et al., 2017; Mukherjee et al., 2022) only consider a single-task setting. The proof is given in Appendix A.2.

2.3 Proposed Algorithm FB-DOE

We now present our algorithm for the fixed-budget linear MTRL setting. The Theorem 1 shows that an optimal agnostic algorithm should first estimate the shared feature extractor $\mathbf{B}$ and then estimate the optimal arm per task. Moreover, the budget n must carefully be divided to reach the optimal rate with respect to k, d, and M. Motivated by this we propose the FB-DOE, which is a phase-based, *two-stage* arm elimination algorithm. Recall that in the fixed budget setting the budget n is given. So we divide the algorithm into two stages. The first stage consists of $n/2$ rounds, where the FB-DOE estimates the feature extractor $\widehat{\mathbf{B}}_n$. Then the second stage consists of another $n/2$ rounds, where the FB-DOE eliminates sub-optimal arms in each task m and finally outputs the estimated optimal arm $\widehat{i}^*_m$ for each task m. Now we discuss each stage of FB-DOE.

2.3.1 Stage 1: Estimating B

In the first stage, FB-DOE leverages the batch of rewards $\{r_{m,t} : m \in [M]\}$ at every round t from M tasks to learn the feature extractor $\mathbf{B}$. To this end, FB-DOE first solves the E-optimal design in line 2 of Algorithm 1 in Appendix A.1. Note that E-optimal design minimizes the spectral norm of the inverse of the sample covariance matrix and is therefore the most suited strategy at the subspace recovery stage. For each task m, FB-DOE samples each arm $\mathbf{x}(i)$ for $\lceil \tau_m^E \mathbf{b}_{\mathbf{x}}^E(i)\rceil$ times, where $\tau_m^E = n/2M$, $\mathbf{b}_{\mathbf{x}}^E(i)$ is the allocation proportion of E-optimal design on $\mathbf{x}(i)$. With slight abuse of notation, we let $r_{m,t}(i)$ be the reward observed for the t-th pull of arm $\mathbf{x}(i)$. It then builds an estimator $\widehat{\mathbf{Z}}_n$ for the average hidden parameter $\mathbf{Z}_* := \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\theta}_{m,*}\boldsymbol{\theta}_{m,*}^\top$ as follows:

$$\widehat{\mathbf{Z}}_n = \frac{2}{Mn}\sum_{m=1}^{M}\widehat{\boldsymbol{\theta}}_m\widehat{\boldsymbol{\theta}}_m^\top - \left(\sum_{t=1}^{\tau_m^E}\mathbf{x}_{m,t}\mathbf{x}_{m,t}^\top\right)^{-1}, \qquad \widehat{\boldsymbol{\theta}}_m = \left(\sum_{t=1}^{\tau_m^E}\mathbf{x}_{m,t}\mathbf{x}_{m,t}^\top\right)^{-1}\sum_{t=1}^{\tau_m^E}\mathbf{x}_{m,t}r_{m,t} \tag{3}$$

where $\widehat{\boldsymbol{\theta}}_m \in \mathbb{R}^d$ serves as an estimator for $\theta_{m,*}$. Next, it performs SVD decomposition on $\widehat{\mathbf{Z}}_n$, and let the top-k left singular vectors of $\widehat{\mathbf{Z}}_n$ be $\widehat{\mathbf{B}}_n$, which serves as the estimator for the feature extractor $\mathbf{B}$. This is shown in lines 3-5 of the pseudocode in Algorithm 1 in Appendix A.1.

2.3.2 Stage 2: Per task arm elimination

In the second stage, FB-DOE aims to identify the optimal arm in each task m by reducing the original d-dimensional linear bandits to a lower k-dimension problem. This is done as follows: For each task m, define the dimension-reduced arm set $\mathcal{G}_m$ as $\mathcal{G}_m = \{\widetilde{\mathbf{g}}_m = \widehat{\mathbf{B}}_n^\top \mathbf{x}, \forall \mathbf{x} \in \mathcal{X}\}$. Note that $\widetilde{\mathbf{g}}_m \in \mathbb{R}^k$ and so we have reduced the original d-dimensional linear bandits to k-dimensional linear bandits for each task m. This step critically sets us apart from standard linear fixed-budget works (Yang & Tan, 2021; Azizi et al., 2022). Then FB-DOE runs the G-optimal design similar to OD-LinBAI. We use G-optimal design in this stage, as it minimizes the maximum prediction error for feature vectors. In particular, FB-DOE partitions the remaining $n/2$ rounds into $\lceil \log_2 k \rceil$ phases. It then maintains an active arm set $\mathcal{G}_{m,\ell}$ in each phase $\ell = 1, 2, \ldots, \lceil \log_2 k \rceil$. The length of each phase roughly equals $n_m(k)$, defined as

$$n_m(k) = \frac{\frac{n}{2M} - \min(A, \frac{k(k+1)}{2}) - \sum_{\ell=1}^{\lceil \log_2 k \rceil - 1} \left\lceil \frac{k}{2^\ell} \right\rceil}{\lceil \log_2 k \rceil}. \tag{4}$$

We use $n_m(k)$ to signify that phase length depends on the latent dimension k. Motivated by the equivalence of the original arm vectors and the dimension-reduced arm vectors, at the beginning of each phase ℓ, FB-DOE computes a set of dimension-reduced arm vectors $\{\widetilde{\mathbf{g}}_{m,\ell}(i) : i \in \mathcal{G}_{m,\ell-1}\} \subset \mathbb{R}^{k_{m,\ell}}$ that spans the $k_{m,\ell}$-dimensional Euclidean space $\mathbb{R}^{k_{m,\ell}}$. This can be implemented based on the arm vectors of the last phase $\{\widetilde{\mathbf{g}}_{m,\ell-1}(i) : i \in \mathcal{G}_{m,\ell-1}\}$ in an iterative manner (see lines 9-14 of Algorithm 1 in Appendix A.1).

Finally, FB-DOE finds a G-optimal design $\mathbf{b}^G_{m,\ell}$ for each task m in phase ℓ with the current dimension-reduced arm vectors, with a restriction on the cardinality of the support when $\ell = 1$. FB-DOE then pulls each arm in $\mathcal{G}_{m,\ell-1}$ according to $\mathbf{b}^G_{m,\ell}$. Specifically, it samples each arm $i \in \widetilde{\mathcal{G}}_{m,\ell-1}$ exactly $N_{m,\ell}(i) = \lceil \mathbf{b}^G_{m,\ell}(i) \cdot n_m(k) \rceil$ times, where $n_m(k)$ is defined in (4). This step stands in sharp contrast to prior fixed-confidence MTRL algorithm (Du et al., 2023), as the low dimensional elimination per task in every phase must be done carefully to reach the exponentially low probability of error (see lines 9-18 of Algorithm 1 in Appendix A.1).

Note that the support of the G-optimal design $\mathbf{b}^G_{m,\ell}$ must span $\mathbb{R}^{k_{m,\ell}}$ by Lemma A.1. Therefore, the ordinary least-square (OLS) estimator can be applied to estimate $\mathbf{w}_m$ (Line 21 of Algorithm 1 in Appendix A.1). Then for each arm $i \in \mathcal{G}_{m,\ell-1}$, an estimate of the expected reward is derived using only the observed rewards in that phase. At the end of each phase ℓ, FB-DOE eliminates a subset of possibly sub-optimal arms for each task m. In particular, $|\mathcal{G}_{m,0}| - \lceil k/2 \rceil$ arms are eliminated in the first phase, and about half of the active arms are eliminated in each of the following phases. Eventually, there is only a single arm $\widehat{i}^*_m$ in the active set for each task m, which is the output of FB-DOE. The full pseudo-code is given in Algorithm 1. We further discuss rounding procedures in Remark A.18 and additional insights on algorithm in Remark A.19.

2.4 Probability of error

In this section, we analyze FB-DOE and bound the probability of error in identifying the optimal arm i^*_m for each task $m \in [M]$. We first state assumptions required for our main results on linear setting.

Assumption 2.1. (Diverse Tasks) We assume that $\sigma_{\min}(\frac{1}{M}\sum_{m=1}^M \mathbf{w}_m \mathbf{w}_m^\top) \geq \frac{c_0}{k}$, for some $c_0 > 0$.

This assumption ensures that the parameters $\mathbf{w}_1, \ldots, \mathbf{w}_M$ are well-distributed in all directions of $\mathbb{R}^k$, which is necessary for recovering the feature extractor $\mathbf{B}$ (Yang et al., 2020; 2022; Du et al., 2023).

Assumption 2.2. (Eigenvalue of G-optimal Design Matrix) For any task $m \in [M]$, $\sigma_{\min}(\sum_i \mathbf{b}^G_m(i) \mathbf{B}^\top \mathbf{x}(i) \mathbf{x}(i)^\top \mathbf{B}) \geq \omega$ for some constant $\omega > 0$.
This assumption ensures that the covariance matrix $\sum_i \mathbf{b}^G_m(i) \mathbf{B}^\top \mathbf{x}(i) \mathbf{x}(i)^\top \mathbf{B}$ under the optimal sample allocation in the second stage is invertible, which is necessary for estimating $\mathbf{w}_m$.

Algorithm 1 Fixed Budget Double Optimal Design (FB-DOE) for Linear Bandits

Input: time budget n, arm set $\mathcal{X} \subset \mathbb{R}^d$.
Let E-optimal design be $\mathbf{b}_{\mathbf{x}}^E = \arg\min_{\mathbf{b}\in\triangle_{\mathcal{X}}} \left\|(\sum_i \mathbf{b}(i)\mathbf{x}\,\mathbf{x}^\top)^{-1}\right\|$. Set $\tau_m^E = \frac{n}{2M}$.
Stage 1 (Feature Recovery): Pull arm $\mathbf{x}(i) \in \mathcal{X}$ exactly $\lceil \mathbf{b}_{\mathbf{x}}^E(i)\tau_m^E \rceil$ times for each task m and observe rewards $\{r_{m,t}\}_{t=1}^{\tau_m^E}$.
Compute $\widehat{\mathbf{Z}}_n$ using (3). Let $\widehat{\mathbf{B}}_n$ be the top-k left singular vectors of $\widehat{\mathbf{Z}}_n$.
Build $\widetilde{\mathbf{g}}_m(i) = \mathbf{x}(i)^\top \widehat{\mathbf{B}}_n$ for all $\mathbf{x}(i) \in \mathcal{X}$ for each $m \in [M]$. Denote the set $\mathcal{G}_m$ containing $\widetilde{\mathbf{g}}_m$.
Initialize $t_{m,0} = 1, \mathcal{G}_{m,0} \leftarrow \mathcal{G}_m$ and $k_{m,0} = k$. For each arm $\widetilde{\mathbf{g}}_m(i) \in \mathcal{G}_{m,0}$, set $\widetilde{\mathbf{g}}_{m,0}(i) = \widetilde{\mathbf{g}}_m(i)$. Calculate $n_m(k)$ using (4).
Stage 2 (Low dimensional elimination)
for $\ell = 1$ to $\lceil \log_2 k \rceil$ **do**
  Set $k_{m,\ell} = \dim\left(\mathrm{span}\left(\{\widetilde{\mathbf{g}}_{m,\ell-1}(i) : i \in \mathcal{G}_{m,\ell-1}\}\right)\right)$.
  if $k_{m,\ell} = k_{m,\ell-1}$ **then**
    For each arm $i \in \mathcal{G}_{m,\ell-1}$, set $\widetilde{\mathbf{g}}_{m,\ell}(i) = \widetilde{\mathbf{g}}_{m,\ell-1}(i)$.
  else
    Find matrix $\mathbf{H}_{m,\ell} \in \mathbb{R}^{k_{m,\ell-1}\times k_{m,\ell}}$ whose columns form an orthonormal basis of the subspace spanned by $\{\widetilde{\mathbf{g}}_{m,\ell-1}(i) : i \in \mathcal{G}_{m,\ell-1}\}$. For each arm $i \in \mathcal{G}_{m,\ell-1}$, set $\widetilde{\mathbf{g}}_{m,\ell}(i) = \mathbf{H}_{m,\ell}^\top \widetilde{\mathbf{g}}_{m,\ell-1}(i)$
  end if
  if $\ell = 1$ **then**
    Find a G-optimal design $\mathbf{b}_{m,\ell}^G : \{\widetilde{\mathbf{g}}_{m,\ell}(i) : i \in \mathcal{G}_{m,\ell-1}\} \to [0,1]$ with $\left|\mathrm{Supp}\left(\mathbf{b}_{m,\ell}^G\right)\right| \le \frac{k(k+1)}{2}$.
  else
    Find a G-optimal design $\mathbf{b}_{m,\ell}^G : \{\widetilde{\mathbf{g}}_{m,\ell}(i) : i \in \mathcal{G}_{m,\ell-1}\} \to [0,1]$.
  end if
  Set $N_{m,\ell}(i) = \left\lceil \mathbf{b}_{m,\ell}^G\left(\widetilde{\mathbf{g}}_{m,\ell}(i)\right) \cdot n_m(k) \right\rceil$ and $N_{m,\ell} = \sum_{i\in\mathcal{G}_{m,\ell-1}} N_{m,\ell}(i)$. Choose each arm $i \in \mathcal{G}_{m,\ell-1}$ in each task m exactly $N_{m,\ell}(i)$ times.
  Calculate the OLS estimator for each task m:

$$\widehat{\mathbf{w}}_{m,\ell} = \mathbf{\Sigma}_{m,\ell}^{-1} \sum_{t=t_{m,\ell}}^{t_{m,\ell}+T_{m,\ell-1}} \widetilde{\mathbf{g}}_m\left(A_t\right) r_{m,t} \quad \text{with } \mathbf{\Sigma}_{m,\ell} = \sum_{i\in\mathcal{G}_{m,\ell-1}} N_{m,\ell}(i) \widetilde{\mathbf{g}}_{m,\ell}(i) \widetilde{\mathbf{g}}_{m,\ell}(i)^\top$$

  Set $\widehat{\boldsymbol{\theta}}_m = \widehat{\mathbf{B}}\widehat{\mathbf{w}}_m$ for each task m. For each arm $i \in \mathcal{G}_{m,\ell-1}$, estimate the expected reward: $\widehat{\mu}_{m,\ell}(i) = \langle \widehat{\boldsymbol{\theta}}_{m,\ell}, \mathbf{x}_m(i) \rangle$.
  Let $\mathcal{G}_{m,\ell}$ be the set of $\lceil k/2^\ell \rceil$ arms in $\mathcal{G}_{m,\ell-1}$ with the largest estimates of the expected rewards.
  Set $t_{m,\ell+1} = t_{m,\ell} + N_{m,\ell}$.
end for

Let $\widetilde{O}_{\omega,L_x}(\cdot)$ hide problem dependent factors ω and L_x. Then under Assumption 2.1 and Assumption 2.2, we have the following guarantee for FB-DOE in the MTRL linear bandit setting.

Theorem 2. ***(informal)*** *Define* $\Delta = \min_m \min_{i\in\mathcal{X}} \Delta_{m,i}$ *and* $H_{2,\,\textit{lin}} = \max_{m\in[M]} \max_{2\le i\le k} \frac{i}{\Delta_{m,i}^2}$. *If* $Mn \ge \lceil \frac{L_x^4 k^2 d^2 c'(\rho^E)^2 \log^2(2d)}{\omega^2\Delta^2} \rceil$, *then the total probability of error of Algorithm 1 is given by* $\widetilde{O}_{\omega,L_x}(\exp(\frac{-Mn}{\log_2 d}) + M\exp(-\frac{n}{H_{2,\,\textit{lin}}\log_2 k}))$.

Discussion 2. We have two terms in the bound of Theorem 2. The first term is the probability of error in estimating the feature extractor $\mathbf{B}$. Observe that as the number of tasks M increases, the first term decays faster, indicating that FB-DOE has a better estimation of the feature extractor $\mathbf{B}$. The second term is the probability of error that FB-DOE suffers in misidentifying the optimal arm in each task m. Observe that the second term scales with the number of tasks M and low dimension $k \ll d$, as FB-DOE runs an individual G-optimal design for each task in lower dimension k. The problem complexity parameter $H_{2,lin}$ is present in the term 2, which captures the worst-case difficulty of identifying the optimal arm across tasks. Note that this improves upon the bound of linear OD-LinBAI which scales as $\widetilde{O}(M\exp(-\frac{n}{\log_2 d H'_{2,\,\mathrm{lin}}}))$, where $H'_{2,\,\mathrm{lin}} = \max_{m\in[M]} \max_{2\le i\le d} \frac{i}{\Delta_{m,i}^2}$ and $H'_{2,\,\mathrm{lin}} > H_{2,\,\mathrm{lin}}$. We further discuss the bounds in Remark A.20 and theoretical comparison in Remark A.21. Also, observe that the condition $Mn \ge \lceil \frac{L_x^4 k^2 d^2 c'(\rho^E)^2 \log^2(2d)}{\omega^2\Delta^2} \rceil$ depends on the number of tasks and the given budget n. If budget n is small, a large number of tasks M can ensure the condition of Theorem 2 is satisfied, and it speeds up learning of shared representation across the tasks.

We remark that the upper bound on the probability of error in Theorem 2 matches the lower bound in Theorem 1 with respect to the parameters k, d, M and $H_{2m\text{lin}}$. However, note that $H_{2,\text{ lin}} \leq H_{1,\text{ lin}} = \max_{m\in[M]} \sum_{i=1}^{k} \frac{1}{\Delta_{m,i}^2} \leq H_{2,\text{ lin}} \log_2 k$. We leave getting a lower bound with true problem-dependent parameter $H_{1,\text{ lin}}$ for future works.

Proof (Overview): We divide the proof into three steps. In step 1 we bound the estimation error of the average estimator $\widehat{\mathbf{Z}}_n$. In step 2 we analyze the estimation error for feature extractor $\mathbf{B}$. Finally in step 3 we bound the probability of wrongly eliminating optimal arm in low dimension.

Step 1 (Estimation of average parameter, Stage 1): In the first stage FB-DOE builds the estimator $\widehat{\mathbf{Z}}_n$ for the average parameter $\mathbf{Z}_* = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{\theta}_{*,m}\boldsymbol{\theta}_{*,m}^\top$. We modify the proof technique of Du et al. (2023), and show in Lemma A.6 of Appendix A.3 that the total probability of error in the first stage is given by $\left(\frac{C(\rho^E)^2 d^2}{\sqrt{Mn}} \exp\left(-\frac{Mn}{2}\right)\right)$. Here, ρ^E is the optimization value of the E-optimal design in line 2 of Algorithm 1. Since the tasks share the same arm set $\mathcal{X}$, the $\rho^E = \rho^E_m$ for any $m \in [M]$. Observe that as the number of tasks M increases, the FB-DOE has better estimates of $\mathbf{Z}_*$.

Step 2 (Estimation of feature extractor, Stage 1): Now using the estimator in (3) we get a good estimation of the feature extractor $\mathbf{B}$. Let $\widehat{\mathbf{B}}_n$ be the top-k left singular vectors of $\widehat{\mathbf{Z}}_n$. Then using the Davis-Kahan $\sin\theta$ Theorem (Bhatia, 2013) in Lemma A.9, we have $\|(\widehat{\mathbf{B}}_n^\perp)^\top \mathbf{B}\| \leq \widetilde{O}\left(\rho^E\left(\frac{2ckd}{\sqrt{Mn}} \exp\left(-\frac{Mn}{2}\right)\right)\right)$. Recall that for task m, $\mathcal{G}_m$ consists of all latent arms $\widetilde{\mathbf{g}}_m(i) = \widehat{\mathbf{B}}_n^\top \mathbf{x}(i)$ for each $\mathbf{x}(i) \in \mathcal{X}$. Then we prove that $\sigma_{\min}(\sum_{\widetilde{\mathbf{g}}_m(i)\in\mathcal{G}_m} \mathbf{b}_m^G(i)\widetilde{\mathbf{g}}_m(i)\widetilde{\mathbf{g}}_m(i)^\top) > 0$ (Lemma A.10), which guarantees that the G-optimal design in stage 2 is valid. Next, Lemma A.12 states that the feature estimation error is low, such that for any task $m \in [M]$ and $\widetilde{\mathbf{g}}_m(j) \in \mathcal{G}_m$, $\|\widetilde{\mathbf{g}}_m(j)\|^2_{\boldsymbol{\Sigma}^{-1}_{m,\ell}} \leq \|\mathbf{g}_m(j)\|^2_{\boldsymbol{\Sigma}^{-1}_{m,\ell}} + \frac{cL_x^4}{k\omega^2}\exp(-Mn)$ for some constant $c > 0$. Finally using Lemma A.13, we show that the parameter estimation error is also low with the estimated feature $\widetilde{\mathbf{g}}_m(j)$. We remark that in all these steps the key challenge lies in deriving an exponentially decaying error bound under the budget n (Lemma A.10, Lemma A.12), which requires a significantly different analysis than the arguments in Du et al. (2023); Yang et al. (2020; 2022)—they only apply for fixed confidence or regret minimization setting.

Step 3 (Elimination in low dimension): In the final step we bound the probability of error in outputting i_m^* for individual tasks. Our key technical novelty lies in controlling the probability of error for each task m even with the noisy latent features in low dimension $\mathbb{R}^k$. Additionally, we have to account for feature and parameter estimation error for $Mn > \lceil \frac{L_x^4 k^2 d^2 c'(\rho^E)^2 \log^2(2d)}{\omega^2\Delta^2} \rceil$, which is not studied in Yang & Tan (2021). In Lemma A.14 we show that indeed the total budget used is at most n. Then in Lemma A.16, Lemma A.17 we ensure that the best arm i_m^* is eliminated in phase ℓ with an exponentially small probability with the right complexity parameter $H_{2,\text{ lin}}$ appearing in the bound. This parameter does not show up in the fixed confidence analysis of Du et al. (2023). We combine all steps to get the final claim in Theorem 2.

Technical challenge: Our key technical is to combine the proof technique of Du et al. (2023) with that of Yang & Tan (2021) to derive the upper bound. In the first stage, we derive the high confidence bounds that are exponentially decaying with budget n where we modify Lemma C.3 of Du et al. (2023) to take into account the fixed sample size of our phase (i.e $n/2$ rounds). This leads to a new estimation of the feature extractor $\mathbf{B}$ in Lemma A.9, and then for a sufficiently large $Mn > \lceil \frac{L_x^4 k^2 d^2 c'(\rho^E)^2 \log^2(2d)}{\omega^2\Delta^2} \rceil$ we have a non-vacuous solution to the G-optimal design in stage 2. These are shown in Lemma A.6-Lemma A.13. In the second stage, our technical novelty lies in controlling the probability of error for the noisy latent features in low-dimensional multi-task linear bandits. This is shown in Lemma A.14, Lemma A.16, and Lemma A.17. Note that this approach differs from the existing art of fixed budget linear bandit settings (Katz-Samuels et al., 2020; Yang

& Tan, 2021; Azizi et al., 2022) and significantly different than the fixed confidence linear bandit proofs in (Soare et al., 2014; Mason et al., 2021; Degenne & Koolen, 2019). This is because these works do not study the multi-task setting and therefore do not need to control the noisy latent feature estimation error in the bounds.

3 MTRL Fixed Budget Bilinear Bandit

In this section, we present the algorithm for the fixed-budget bilinear bandit setting. Similar as the linear MTRL setting, again we show that a double experimental design approach will lead to a lower probability of error than solving the tasks individually.

3.1 Preliminaries of MTRL for bilinear bandits

In the MTRL bilinear bandit setting, we again consider a scenario with M tasks, indexed as $m = [M]$. Each task is associated with a hidden parameter $\mathbf{\Theta}_{m,*} \in \mathbb{R}^{d_1 \times d_2}$. In the bilinear bandit setting, different from the conventional linear bandit framework, each task consists of left and right arm sets denoted by $\mathcal{X} \subset \mathbb{R}^{d_1}$ and $\mathcal{Z} \subset \mathbb{R}^{d_2}$ respectively. So the learner observes a pair of arms denoted by $\mathbf{x}_{m,t} \in \mathcal{X}$ and $\mathbf{z}_{m,t} \in \mathcal{Z}$ for each task m in each round t. The interaction of this arm pair with the hidden parameter, $\mathbf{\Theta}_{m,*} \in \mathbb{R}^{d_1 \times d_2}$, produces noisy feedback (reward) $r_{m,t} = \mathbf{x}_{m,t}^\top \mathbf{\Theta}_{m,*} \mathbf{z}_{m,t} + \eta_{m,t}$. The term $\eta_{m,t}$ represents independent zero-mean 1-sub-Gaussian noise.

Following the setting of Mukherjee et al. (2023b), we assume that each $\mathbf{\Theta}_{m,*}$ can be decomposed as $\mathbf{\Theta}_{m,*} = \mathbf{B}_1 \mathbf{S}_{m,*} \mathbf{B}_2^\top$, 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, while $\mathbf{S}_{m,*} \in \mathbb{R}^{k_1 \times k_2}$ is task-specific. We assume that $k_1, k_2 \ll d_1, d_2$, and $k_1, k_2 \geq 2$ as well as $M \gg d_1, d_2$. Thus, $\mathbf{B}_1$ and $\mathbf{B}_2$ serve as means of dimension reduction. Additionally, we assume each $\mathbf{S}_{m,*}$ has rank $r \ll \min\{k_1, k_2\}$. In the context of MTRL, $\mathbf{B}_1$ and $\mathbf{B}_2$ are referred to as *feature extractors*, while $\mathbf{x}_{m,t}$ and $\mathbf{z}_{m,t}$ are termed *rich observations*. The reward for task $m \in [M]$ at round t is:

$$r_{m,t} = \mathbf{x}_{m,t}^\top \mathbf{\Theta}_{m,*} \mathbf{z}_{m,t} + \eta_{m,t} = \mathbf{x}_{m,t}^\top \mathbf{B}_1 \mathbf{S}_{m,*} \mathbf{B}_2^\top \mathbf{z}_{m,t} + \eta_{m,t} \overset{(a)}{=} \mathbf{g}_{m,t}^\top \mathbf{S}_{m,*} \mathbf{v}_{m,t} + \eta_{m,t}. \quad (5)$$

where, (a) follows as $\mathbf{g}_{m,t}^\top \triangleq \mathbf{x}_{m,t}^\top \mathbf{B}_1$ and $\mathbf{v}_{m,t} \triangleq \mathbf{B}_2^\top \mathbf{z}_{m,t}$. Similar to the learning procedure in Yang et al. (2020; 2022), at each round $t \in [n]$, the learner chooses left and right actions $\mathbf{x}_{m,t} \in \mathcal{X}$ and $\mathbf{z}_{m,t} \in \mathcal{Z}$ for each task $m \in [M]$. After committing the batch of actions $\{\mathbf{x}_{m,t}, \mathbf{z}_{m,t} : m \in [M]\}$, the learner receives the batch of rewards $\{r_{m,t} : m \in [M]\}$. Furthermore, in (5), we refer $\mathbf{g}_{m,t} \in \mathbb{R}^{k_1}$ and $\mathbf{v}_{m,t} \in \mathbb{R}^{k_2}$ as the latent features. Both $\mathbf{g}_{m,t}$ and $\mathbf{v}_{m,t}$ are unknown to the learner and need to be learned for each task m. WLOG let $i_m^* = 1$ be the optimal arm in task m and define gap $\Delta_{m,i} = (\mathbf{x}_{i_m^*}^\top \mathbf{\Theta}_{m,*} \mathbf{z}_{i_m^*} - \mathbf{x}_i^\top \mathbf{\Theta}_{m,*} \mathbf{z}_i)$ for $i \neq i_m^*$. Let $\|\mathbf{x}\|, \|\mathbf{z}\| \leq L_x$, $\|\mathbf{S}_{m,*}\|_F \leq 1$. Again, for simplicity we assume that the expected rewards of the arms are in descending order and the best arm is unique. Let S_r be the minimum eigenvalue of $\mathbf{\Theta}_{m,*}$ for any $m \in [M]$.

3.2 Proposed algorithm: extension of FB-DOE

We now present an extension of FB-DOE to the bilinear bandit setting. The FB-DOE is a phase-based, *three-stage* arm elimination algorithm. The key difference from the linear bandit setting is that we need to have an extra stage to estimate the task-specific parameter $\mathbf{S}_{m,*}$. Specifically, the algorithm divides the fixed budget n into three stages. The first stage consists of $n/3$ rounds where FB-DOE estimates the left and the right feature extractors $\mathbf{B}_1$ and $\mathbf{B}_2$. The second stage consists of another $n/3$ rounds where FB-DOE aims to estimate the parameter $\mathbf{S}_{m,*}$ for each task m. The third stage consists of the last $n/3$ rounds. Here FB-DOE eliminates sub-optimal arms in each task m and finally outputs the estimated optimal arm $\widehat{i}_m^*$ for each task m. The full pseudo-code is given in Algorithm 2 in Appendix A.1. Now we discuss individual stages of FB-DOE.

3.2.1 Stage 1: Estimating $\mathbf{B}_1$ and $\mathbf{B}_2$

FB-DOE first 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, FB-DOE first vectorizes arms $\mathbf{x} \in \mathcal{X}, \mathbf{z} \in \mathcal{Z}$ into a new vector $\mathbf{w} = \text{vec}(\mathbf{x}; \mathbf{z}) \in \mathcal{W}$ and define $\mathbf{w}_{m,t} = \text{vec}(\mathbf{x}_{m,t}; \mathbf{z}_{m,t})$. The FB-DOE then solves the E-optimal design in line 2 of Algorithm 2. For each task m, FB-DOE samples each $\mathbf{w}(i) \in \mathcal{W}$ for $\lceil \tau^E \mathbf{b}^E_{\mathbf{w}}(i) \rceil$ times, where $\tau^E_m = n/3M$ and $\mathbf{b}^E_{\mathbf{w}}(i)$ is the allocation of E-optimal design on $\mathbf{w}(i)$. Then it builds the estimator $\widehat{\mathbf{Z}}_n$ for the average parameters $\mathbf{Z}_* = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{\theta}_{m,*}\boldsymbol{\theta}_{m,*}^\top$ as follows, where $\boldsymbol{\theta}_{m,*} \in \mathbb{R}^{d_1 d_2}$ is the vector of $\boldsymbol{\Theta}_{m,*}$:

$$\widehat{\mathbf{Z}}_n = \frac{3}{Mn}\sum_{m=1}^{M}\sum_{t=1}^{\tau^E_m} \widehat{\boldsymbol{\theta}}_m \widehat{\boldsymbol{\theta}}_m^\top - \Big(\sum_{t=1}^{\tau^E_m} \mathbf{w}_{m,t}\mathbf{w}_{m,t}^\top\Big)^{-1}, \widehat{\boldsymbol{\theta}}_m = \Big(\sum_{t=1}^{\tau^E_m} \mathbf{w}_{m,t}\mathbf{w}_{m,t}^\top\Big)^{-1} \sum_{t=1}^{\tau^E_m} \mathbf{w}_{m,t} r_{m,t} \quad (6)$$

where $\widehat{\boldsymbol{\theta}}_m \in \mathbb{R}^{d_1 d_2}$ is an estimator for $\boldsymbol{\theta}_{m,*}$. Then it performs SVD decomposition on $\widehat{\mathbf{Z}}_n$, and let $\widehat{\mathbf{B}}_{1,n}$, $\widehat{\mathbf{B}}_{2,n}$ be the top-k_1 left and top-k_2 right singular vectors of $\widehat{\mathbf{Z}}_n$, respectively, which are the estimations of the feature extractors $\mathbf{B}_1$ and $\mathbf{B}_2$. This is shown in lines 3-6 of Algorithm 2.

3.2.2 Stage 2: Estimating per task $\mathbf{S}_{m,*}$

In the second stage of phase ℓ, the goal is to recover the hidden parameter $\mathbf{S}_{m,*}$ for each task m. FB-DOE proceeds as follows: First, let $\widetilde{\mathbf{g}}_m = \mathbf{x}^\top \widehat{\mathbf{B}}_{1,n}$ and $\widetilde{\mathbf{v}}_m = \mathbf{z}^\top \widehat{\mathbf{B}}_{2,n}$ be the latent left and right arm respectively for each m. Then FB-DOE defines the vector $\widetilde{\mathbf{w}}_m = \text{vec}(\widetilde{\mathbf{g}}_m; \widetilde{\mathbf{v}}_m) \in \widetilde{\mathcal{W}}_m$ and then solves the E-optimal design in line 7 of Algorithm 2. For each task m, it then samples the latent arm $\widetilde{\mathbf{w}} \in \widetilde{\mathcal{W}}_m$ for $\lceil \widetilde{\tau}^E_m \mathbf{b}^E_{m,\widetilde{\mathbf{w}}} \rceil$ times, where $\widetilde{\tau}^E_m := \frac{n}{3M}$ and $\mathbf{b}^E_{m,\widetilde{\mathbf{w}}}$ is the solution to E-optimal design on $\widetilde{\mathbf{w}}$. Then it builds an estimator $\widehat{\mathbf{S}}_{m,n}$ for each task m in line 9 as follows:

$$\widehat{\mathbf{S}}_{m,n} = \underset{\boldsymbol{\Theta} \in \mathbb{R}^{k_1 \times k_2}}{\arg\min}\ L_n(\boldsymbol{\Theta}) + \lambda_n \|\boldsymbol{\Theta}\|_{\text{nuc}}, L_n(\boldsymbol{\Theta}) = \sum_{t=1}^{\tau^E_m} \left(r_{m,t} - \langle \widetilde{\mathbf{g}}_{m,t}\widetilde{\mathbf{v}}_{m,t}^\top, \boldsymbol{\Theta}\rangle\right)^2. \quad (7)$$

Once FB-DOE recovers the $\widehat{\mathbf{S}}_{m,n}$ for each task m, it reduces the $d_1 d_2$ bilinear bandit to a $k_1 k_2$ dimension bilinear bandit where the left and right arms are $\widetilde{\mathbf{g}}_m(i) \in \mathbb{R}^{k_1}$, $\widetilde{\mathbf{v}}_m(i) \in \mathbb{R}^{k_2}$ respectively for each $\mathbf{x}(i) \in \mathcal{X}$ and $\mathbf{z}(i) \in \mathcal{Z}$.

3.2.3 Stage 3: Rotated arm elimination per task

In the third stage, for each task m, FB-DOE defines the rotated arm set $\underline{\mathcal{G}}_m$ for these $k_1 k_2$ dimensional bilinear bandits. Consider the SVD of $\widehat{\mathbf{S}}_{m,n} = \widehat{\mathbf{U}}_{m,n}\widehat{\mathbf{D}}_{m,n}\widehat{\mathbf{V}}_{m,n}^\top$. Define $\widehat{\mathbf{H}}_{m,n} = [\widehat{\mathbf{U}}_{m,n}\widehat{\mathbf{U}}^\perp_{m,n}]^\top \widehat{\mathbf{S}}_{m,n}[\widehat{\mathbf{V}}_{m,n}\widehat{\mathbf{V}}^\perp_{m,n}]$ where $\widehat{\mathbf{U}}^\perp_{m,n}$ and $\widehat{\mathbf{V}}^\perp_{m,n}$ are the complementary subspaces of $\widehat{\mathbf{U}}_{m,n}$ and $\widehat{\mathbf{V}}_{m,n}$ respectively. Then define the vectorized arm set so that the last $(k_1 - r)\cdot(k_2 - r)$ components are from the complementary subspaces as:

$$\begin{aligned}
\underline{\mathcal{G}}_m &= \Big\{\big[\mathbf{vec}\left(\widetilde{\mathbf{g}}_{m,1:r}\widetilde{\mathbf{v}}_{m,1:r}^\top\right); \mathbf{vec}\left(\widetilde{\mathbf{g}}_{m,r+1:k_1}\widetilde{\mathbf{v}}_{m,1:r}^\top\right); \\
&\quad \mathbf{vec}\left(\widetilde{\mathbf{g}}_{m,1:r}\widetilde{\mathbf{v}}_{m,r+1:k_2}^\top\right); \mathbf{vec}\left(\widetilde{\mathbf{g}}_{m,r+1:k_1}\widetilde{\mathbf{v}}_{m,r+1:k_2}^\top\right)\big]\Big\} \\
\widehat{\mathbf{s}}_{m,n,1:\widetilde{k}} &= [\mathbf{vec}(\widehat{\mathbf{H}}_{m,n,1:r,1:r}); \mathbf{vec}(\widehat{\mathbf{H}}_{m,n,r+1:k_1,1:r}); \\
&\quad \mathbf{vec}(\widehat{\mathbf{H}}_{m,n,1:r,r+1:k_2})], \\
\widehat{\mathbf{s}}_{m,n,\widetilde{k}+1:k_1k_2} &= \mathbf{vec}(\widehat{\mathbf{H}}_{m,n,r+1:k_1,r+1:k_2}).
\end{aligned} \quad (8)$$

where $\widetilde{k} = (k_1 + k_2)r$ is the dimension of the rotated arm set. This is shown in line 9 of Algorithm 2. Now we implement a phase-based G-optimal design (like OD-LinBAI) where in the first phase $\ell = 0$ we construct a per-task optimal design for the rotated arm set $\underline{\mathcal{G}}_{m,0}$. Recall that to minimize the probability of error for the m-th bilinear bandit we need to sample according to G-optimal design:

$$\mathbf{b}^G_{m,\ell} = \arg\min_{\mathbf{b}} \max_{\underline{\mathbf{g}} \in \underline{\mathcal{G}}_{m,\ell}} \|\underline{\mathbf{g}}\|^2_{(\sum_i \mathbf{b}(i)\underline{\mathbf{g}}(i)\ \underline{\mathbf{g}}(i)^\top + \boldsymbol{\Lambda}_{m,\ell})^{-1}}. \quad (9)$$

Here $\mathbf{\Lambda}_{m,\ell}$ is a positive definite diagonal matrix defined as:

$$\mathbf{\Lambda}_{m,\ell} = \mathbf{diag}[\underbrace{\lambda, \ldots, \lambda}_{\widetilde{k}}, \underbrace{\lambda_{\ell}^{\perp}, \ldots, \lambda_{\ell}^{\perp}}_{k_1 k_2 - \widetilde{k}}] \tag{10}$$

where, $\lambda_{\ell}^{\perp} := n/24\widetilde{k}\log(1 + n/3\lambda) \gg \lambda$. Then FB-DOE runs G-optimal design on the arm set $\underline{\mathcal{G}}_{m,\ell}$ following the (9) and then samples each $\underline{\mathbf{w}} \in \underline{\mathcal{G}}_{m,\ell}$ for $N_{m,\ell}(i) = \lceil \mathbf{b}^{G}_{\underline{\mathbf{w}}_m,\ell}(i) \cdot n_m(\widetilde{k}) \rceil$ times where $\mathbf{b}^{G}_{m,\ell}$ is the solution to the G-optimal design, defined in step 19-23 of Algorithm 2. At the ℓ-th phase of stage 3, sample the actions according to the G-optimal design similar to Algorithm 1. This is shown in steps 11-30. The only difference with Algorithm 1 is the estimator $\widehat{\mathbf{s}}_{m,\ell} \in \mathbb{R}^{k_1 k_2}$. Then for each task m we can just use the observations from this phase to build the estimator $\widehat{\mathbf{s}}_{m,\ell}$ as shown in (11). Finally, FB-DOE eliminates the sub-optimal arms using the estimator $\widehat{\mathbf{s}}_{m,\ell}$, and builds the next phase active set $\underline{\mathcal{G}}_{m,\ell}$ and stops when $\ell = \lceil \log_2 \widetilde{k} \rceil$.

3.3 Probability of Error

In this section, we analyze FB-DOE for the bilinear bandits and bound the total probability of error in outputting the optimal arm i_m^* for each task $m \in [M]$. We first state our assumptions.

Assumption 3.1. (Diverse tasks) We assume that $\sigma_{\min}(\frac{1}{M}\sum_{m=1}^{M} \mathbf{S}_{m,*}) \geq \frac{c_0 S_r}{k_1 k_2}$, for some $c_0 > 0$ where S_r is the r-th largest singular value of $\mathbf{\Theta}_{m,*}$

This ensures the possibility of recovering the feature extractors $\mathbf{B}_1$ and $\mathbf{B}_2$ shared across tasks (Yang et al., 2020; 2022; Mukherjee et al., 2023b).

Assumption 3.2. (Eigenvalue of E-optimal design matrix) For the arm sets $\mathcal{X}, \mathcal{Z}$ we have $\sigma_{\min}(\sum_i \mathbf{b}^E_{\mathbf{w}}(i)\mathbf{B}_1^\top \mathbf{x}(i)\mathbf{x}(i)^\top \mathbf{B}_1) \geq \omega, \sigma_{\min}(\sum_i \mathbf{b}^E_{\mathbf{w}}(i)\mathbf{B}_2^\top \mathbf{z}(i)\mathbf{z}(i)^\top \mathbf{B}_2) \geq \omega$ for constant $\omega > 0$.

Assumption 3.3. (Eigenvalue of G-optimal design matrix) There exists a constant $\omega > 0$ such that for each task $m \in [M]$, $\sigma_{\min}(\sum_i \mathbf{b}^G_m(i)\mathbf{U}_m^\top \mathbf{g}(i)\mathbf{g}(i)^\top \mathbf{U}_m) \geq \omega$, and $\sigma_{\min}(\sum_i \mathbf{b}^G_m(i)\mathbf{V}_m^\top \mathbf{v}(i)\mathbf{v}(i)^\top \mathbf{V}_m) \geq \omega$.

Assumption 3.2 and Assumption 3.3 ensures that the covariance matrix in second and third stage is invertible under the E and G-optimal design, respectively. Then under Assumptions 3.1, 3.2, and 3.3, we have the following probability of error for FB-DOE in bilinear bandit setting.

Theorem 3. *Define* $\Delta = \min_m \min_{i \in \mathcal{X},\mathcal{Z}} \Delta_{m,i}$ *and* $H_{2,\text{ bilin}} = \max_{m\in[M]} \max_{2 \leq i \leq (k_1+k_2)r} \frac{i}{\Delta^2_{m,i}}$. *If* $Mn \geq \lceil \frac{(d_1 d_2)^2 (k_1 k_2)^2 c'(\rho^E)^2 \log^2(2d_1 d_2)}{S_r^2 \omega^2 \Delta^2} \rceil$, *then the total probability of error of Algorithm 2 is given by* $\widetilde{O}_{\omega, L_x, S_r}\big(\exp(-\frac{Mn}{\log_2 d_1 d_2}) + M\exp(\frac{-n}{\log_2(k_1+k_2)}) + M\exp(-\frac{n}{H_{2,\text{ bilin}}\log_2(k_1+k_2)r})\big)$.

Discussion 2. We have three terms in the bound of Theorem 3 . The first term is the probability of error in estimating the feature extractors $\mathbf{B}_1$ and $\mathbf{B}_2$. Observe that as the number of tasks M increases, the first term decays faster, indicating that FB-DOE has a better estimate of the feature extractors. The second term is the probability of error that FB-DOE suffers in estimating the hidden parameter $\mathbf{S}_{m,*}$ for each task m. This term scales with M and $(k_1 + k_2)$. Finally, the third term is the probability of error of mis-identifying the optimal left and right arm in each task m. The third term scales with the number of tasks M and rotated low dimension $(k_1 + k_2)r \ll d_1, d_2$ since FB-DOE runs an individual G-optimal design for each task in lower dimension $(k_1 + k_2)r$. The problem complexity parameter $H_{2,\text{bilin}}$ is present in term 3, which captures the worst-case difficulty of identifying the optimal left and right arm in each task. This improves upon the bound of bilinear OD-LinBAI which scales as $\widetilde{O}(M\exp(-\frac{n}{H'_{2,\text{ bilin}}\log_2 d_1 d_2}))$ where $H'_{2,\text{ bilin}} = \max_m \max_{2\leq i \leq d_1 d_2} \frac{i}{\Delta^2_{m,i}}$ and $H'_{2,\text{ bilin}} > H_{2,\text{ bilin}}$. We further discuss the bounds in Remark A.40.

Proof (Overview): The proof here follows similar arguments as that of the linear setting (Theorem 2), although more involved due to the bilinear structure. In particular, the proof now consists of four steps. In step 1 we again bound the error of the estimator $\widehat{\mathbf{Z}}_n$. In step 2 we analyze the estimation error of feature extractors $\mathbf{B}_1$ and $\mathbf{B}_2$, as well as left and right latent features. In step 3 we bound the error of estimator $\widehat{\mathbf{S}}_{m,n}$ for each task m, and further bound the estimation error of latent left and right features which now scale with $\widetilde{k}$. Finally, in step 4 we bound the probability of wrongly eliminating optimal arm in low dimension.

Step 1 (Estimation of average parameter, Stage 1): Note that FB-DOE builds the average estimator $\widehat{\mathbf{Z}}_n$ for the quantity $\mathbf{Z}_* = \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\theta}_{m,*}\boldsymbol{\theta}_{m,*}^\top$. We show that the total probability of error in first stage is given by $(\frac{C(\rho^E)^2 d^2}{\sqrt{Mn}}\exp(-\frac{Mn}{2}))$ in Lemma A.22 in Appendix A.4. Here, ρ^E is the optimization value of the E-optimal design in line 2. We modify the proof technique Mukherjee et al. (2023b) to account for the fixed budget setting.

Step 2 (Estimation of feature extractors, Stage 1): Now using the estimator $\widehat{\mathbf{Z}}_n$ in (6) we obtain estimators $\widehat{\mathbf{B}}_{1,n}$ and $\widehat{\mathbf{B}}_{2,n}$ for the feature extractors $\mathbf{B}_1$ and $\mathbf{B}_2$, respectively. Then again using the Davis-Kahan $\sin\theta$ Theorem (Bhatia, 2013), we bound the estimation error of $\widehat{\mathbf{B}}_{1,n}$ and $\widehat{\mathbf{B}}_{2,n}$ in Lemma A.24, and Lemma A.25. Let $\widetilde{\mathbf{g}}_m(i) = \widehat{\mathbf{B}}_{1,n}^\top\mathbf{x}(i)$ for each $\mathbf{x}(i)\in\mathcal{X}$ and $\widetilde{\mathbf{v}}_m(i) = \widehat{\mathbf{B}}_{2,n}^\top\mathbf{z}(i)$ for each $\mathbf{z}(i)\in\mathcal{Z}$ for task m. Let $\widetilde{\mathbf{w}}_m(i) = \mathrm{vec}(\widetilde{\mathbf{x}}(i);\widetilde{\mathbf{z}}(i))$. Then we show that $\sigma_{\min}(\sum_{\widetilde{\mathbf{w}}(i)}\mathbf{b}_m^E(i)\widetilde{\mathbf{w}}(i)\widetilde{\mathbf{w}}(i)^\top) > 0$ in Lemma A.26. This ensures that the E-optimal design in stage 2 is feasible and not vacuous. In Lemma A.27, we prove that the feature estimation error is low such that for each task $m\in[M]$ and any $\widetilde{\mathbf{g}}_m(j)\in\widetilde{\mathcal{G}}_m$, $\|\widetilde{\mathbf{g}}_m(j)\|^2_{\boldsymbol{\Sigma}^{-1}_{m,\ell}} \leq \|\mathbf{g}(j)\|^2_{\boldsymbol{\Sigma}^{-1}_{m,\ell}} + \frac{cL_x^4}{S_r^2 k_1 k_2\omega^2}\exp(-Mn)$ for some constant $c>0$. A similar result holds for $\widetilde{\mathbf{v}}_m(j)\in\mathcal{V}_m$ for each task m. In all these steps the key novelty lies in establishing an exponentially decaying error bound under the budget n.

Step 3 (Estimation of $\mathbf{S}_{m,*}$, Stage 2): Using the estimator in (21) we get a good estimation of the $\mathbf{S}_{m,*}$ for sufficiently large n. The key novelty in this step is to use Restricted String Convexity and Theorem 15 of Lu et al. (2021) to derive the exponentially decaying bound with the right dependence on k_1, k_2. Let the SVD of $\widehat{\mathbf{S}}_{m,n} = \widehat{\mathbf{U}}_{m,n}\widehat{\mathbf{D}}_{m,n}\widehat{\mathbf{V}}_{m,n}^\top$. Again using the Davis-Kahan $\sin\theta$ Theorem, we show in Lemma A.32, A.33 that we have good estimators $\widehat{\mathbf{U}}_{m,n}$ and $\widehat{\mathbf{V}}_{m,n}$. FB-DOE then rotates the arms following (8). Let $\underline{\mathbf{g}}_m(i) = \widehat{\mathbf{U}}_{m,n}^\top\mathbf{x}(i)$ for each $\widetilde{\mathbf{g}}_m(i)\in\mathcal{G}_m$ and $\underline{\mathbf{v}}_m(i) = \widehat{\mathbf{V}}_{m,n}^\top\mathbf{z}(i)$ for each $\widetilde{\mathbf{v}}_m(i)\in\mathcal{V}_m$ for task m. Then we ensure in Lemma A.36 that $\sigma_{\min}(\sum_{i\in\mathcal{G}_m}\mathbf{b}_m^G(i)\widetilde{\mathbf{g}}_m(i)\widetilde{\mathbf{g}}_m(i)^\top) > 0$. This ensures that the G-optimal design in stage 3 is valid. In Lemma A.37, we ensure that for any task $m\in[M]$, the estimation error of each $\widetilde{\mathbf{g}}_m(j)\in\mathcal{G}_m$ decays exponentially. A similar result holds for $\widetilde{\mathbf{v}}_j\in\mathcal{V}_m$ for each task m. Finally using Lemma A.38 we ensure that the estimation error is also low with the estimated features $\widetilde{\mathbf{g}}_m(j)$ and $\widetilde{\mathbf{v}}_m(j)$. Note that in all these steps the key novelty lies in deriving an exponentially decaying error bound under budget n with the right complexity parameter $H_{2,\,\text{bilin}}$ appearing in the bound. This parameter does not show up in the fixed confidence analysis of Mukherjee et al. (2023a).

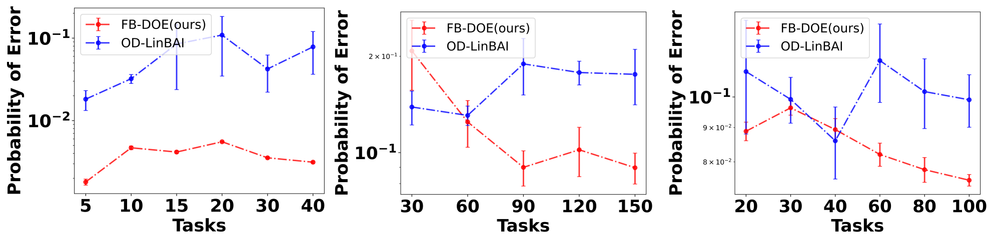
Step 4 (Elimination in low dimension): In the final step we follow the same steps as in step 3 of the proof of Theorem 2 for the rotated arm set $\underline{\mathcal{G}}_m$ (see line 10) for each task m. The final result follows by combining all the steps.

We leave proving the lower bound for fixed budget bilinear bandit setting to future works.

4 Experiments

In this section, we show two synthetic proof-of-concept experiments for MTRL linear and bilinear bandit settings and one-real world linear MTRL experiment on Nectar Dataset (Zhu et al., 2023).

In the MTRL linear bandit experiments (synthetic and Nectar), we compare against the OD-LinBAI (Yang & Tan, 2021). Figure 1a and Figure 1c show that FB-DOE achieves a lower probability of error than the OD-LinBAI with an increasing number of tasks. Note that the real-world experiment does not follow the linear MTRL Assumption 2.1, 2.2. In the MTRL bilinear bandit experiment, we compare against the fixed budget OD-LinBAI algorithm as there is no existing fixed budget algorithm for bilinear bandits. From Figure 1b, we see that FB-DOE achieves a lower probability of error than



(a) Linear setting (b) Bilinear setting (c) Nectar Dataset

Figure 1: Probability error with increasing task numbers.

OD-LinBAI with an increasing number of tasks. We defer a fuller description of the experimental setup to Appendix A.5.

5 Conclusions and Future Directions

In this paper, we formulate the first *fixed budget* pure exploration (bi)linear MTRL setting. We propose the first double and triple optimal design based algorithms for the fixed budget (bi)linear bandit setting. We show that our proposed algorithm FB-DOE in linear bandit setting achieves a probability of error scaling as $\widetilde{O}(M\exp(-n\Delta^2/k\log_2 k))$, which improves upon OD-LinBAI error of $\widetilde{O}(M\exp(-n\Delta^2/d\log_2 d))$. Similarly, in the bilinear bandits, FB-DOE achieves a probability of error scaling as $\widetilde{O}(M(\exp(-n\Delta^2/(k_1+k_2)r\log_2(k_1+k_2)r))$, which improves upon OD-LinBAI error of $\widetilde{O}(M\exp(-n\Delta^2/d_1d_2\log_2 d_1d_2))$. We also provide the first probability of error lower bound for the linear fixed budget MTRL setting and show that FB-DOE probability of error upper bound matches the lower bound with respect to k, d, M and worst case hardness paramater $H_{2,\text{lin}}$. In the future, we wish to extend our results to other structured bandit settings (Degenne & Koolen, 2019; Tirinzoni et al., 2020).

Acknowledgments: Q. Xie is supported in part by National Science Foundation (NSF) grants CNS-1955997 and EPCN-2339794 and EPCN-2432546.

References

Deepak Agarwal, Bee-Chung Chen, and Pradheep Elango. Explore/exploit schemes for web content optimization. In *2009 Ninth IEEE International Conference on Data Mining*, pp. 1–10. IEEE, 2009.

Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best arm identification in multi-armed bandits. In *COLT*, pp. 41–53, 2010.

Mohammad Javad Azizi, Branislav Kveton, and Mohammad Ghavamzadeh. Fixed-budget best-arm identification in structured bandits. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence*, 2022.

Yoshua Bengio, Samy Bengio, and Jocelyn Cloutier. *Learning a synaptic learning rule*. Université de Montréal, Département d'informatique et de recherche ..., 1990.

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

Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

Felix JS Bragman, Ryutaro Tanno, Zach Eaton-Rosen, Wenqi Li, David J Hawkes, Sebastien Ourselin, Daniel C Alexander, Jamie R McClelland, and M Jorge Cardoso. 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*, pp. 3–11. Springer, 2018.

Sebastien Bubeck, Remi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In *Proceedings of the 20th International Conference on Algorithmic Learning Theory*, pp. 23–37, 2009.

Alexandra Carpentier and Andrea Locatelli. Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Conference on Learning Theory*, pp. 590–604. PMLR, 2016.

Yifang Chen, Kevin Jamieson, and Simon Du. Active multi-task representation learning. In *International Conference on Machine Learning*, pp. 3271–3298. PMLR, 2022.

Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 208–214. JMLR Workshop and Conference Proceedings, 2011.

Rémy Degenne and Wouter M Koolen. Pure exploration with multiple correct answers. *Advances in Neural Information Processing Systems*, 32, 2019.

Simon S Du, Wei Hu, Sham M Kakade, Jason D Lee, and Qi Lei. Few-shot learning via learning the representation, provably. *arXiv preprint arXiv:2002.09434*, 2020.

Yihan Du, Longbo Huang, and Wen Sun. Multi-task representation learning for pure exploration in linear bandits. *arXiv preprint arXiv:2302.04441*, 2023.

Valerii Fedorov. Optimal experimental design. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(5):581–589, 2010.

Tanner Fiez, Lalit Jain, Kevin G Jamieson, and Lillian Ratliff. Sequential experimental design for transductive linear bandits. *Advances in neural information processing systems*, 32, 2019.

Matthew Hoffman, Bobak Shahriari, and Nando Freitas. On correlation and budget constraints in model-based bandit optimization with application to automatic machine learning. In *Artificial Intelligence and Statistics*, pp. 365–374. PMLR, 2014.

Ruitong Huang, Mohammad M. Ajallooeian, Csaba Szepesvári, and Martin Müller. Structured best arm identification with fixed confidence. In Steve Hanneke and Lev Reyzin (eds.), *International Conference on Algorithmic Learning Theory, ALT 2017, 15-17 October 2017, Kyoto University, Kyoto, Japan*, volume 76 of *Proceedings of Machine Learning Research*, pp. 593–616. PMLR, 2017. URL http://proceedings.mlr.press/v76/huang17a.html.

Kevin Jamieson and Lalit Jain. Interactive machine learning. 2022.

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

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

Abbas Kazerouni, Mohammad Ghavamzadeh, Yasin Abbasi Yadkori, and Benjamin Van Roy. Conservative contextual linear bandits. *Advances in Neural Information Processing Systems*, 30, 2017.

Anusha Lalitha Lalitha, Kousha Kalantari, Yifei Ma, Anoop Deoras, and Branislav Kveton. Fixed-budget best-arm identification with heterogeneous reward variances. In *Uncertainty in Artificial Intelligence*, pp. 1164–1173. PMLR, 2023.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Jiayi Li, Hongyan Zhang, Liangpei Zhang, Xin Huang, and Lefei Zhang. Joint collaborative representation with multitask learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5923–5936, 2014.

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pp. 661–670, 2010.

Yangyi Lu, Amirhossein Meisami, and Ambuj Tewari. Low-rank generalized linear bandit problems. In *International Conference on Artificial Intelligence and Statistics*, pp. 460–468. PMLR, 2021.

Yunan Luo, Xinbin Zhao, Jingtian Zhou, Jinglin Yang, Yanqing Zhang, Wenhua Kuang, Jian Peng, Ligong Chen, and Jianyang Zeng. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nature communications*, 8(1):573, 2017.

Blake Mason, Romain Camilleri, Subhojyoti Mukherjee, Kevin Jamieson, Robert Nowak, and Lalit Jain. Nearly optimal algorithms for level set estimation. *arXiv preprint arXiv:2111.01768*, 2021.

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

Subhojyoti Mukherjee, Ardhendu S Tripathy, and Robert Nowak. Chernoff sampling for active testing and extension to active regression. In *International Conference on Artificial Intelligence and Statistics*, pp. 7384–7432. PMLR, 2022.

Subhojyoti Mukherjee, Qiaomin Xie, Josiah Hanna, and Robert Nowak. Speed: Experimental design for policy evaluation in linear heteroscedastic bandits. *arXiv preprint arXiv:2301.12357*, 2023a.

Subhojyoti Mukherjee, Qiaomin Xie, Josiah P Hanna, and Robert Nowak. Multi-task representation learning for pure exploration in bilinear bandits. *arXiv preprint arXiv:2311.00327*, 2023b.

Friedrich Pukelsheim. *Optimal design of experiments*. SIAM, 2006.

Laura J Padilla Reyes, Natalia Bonifaz Oviedo, Edgar C Camacho, and Juan M Calderon. Adaptable recommendation system for outfit selection with deep learning approach. *IFAC-PapersOnLine*, 54 (13):605–610, 2021.

Tom Schaul and Jürgen Schmidhuber. Metalearning. *Scholarpedia*, 5(6):4650, 2010.

Qian Shen, Siteng Han, Yu Han, and Xi Chen. User review analysis of dating apps based on text mining. *Plos one*, 18(4):e0283896, 2023.

Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 27, 2014.

Hongjin Su, Weijia Shi, Jungo Kasai, Yizhong Wang, Yushi Hu, Mari Ostendorf, Wen-tau Yih, Noah A. Smith, Luke Zettlemoyer, and Tao Yu. One embedder, any task: Instruction-finetuned text embeddings. 2022. URL https://arxiv.org/abs/2212.09741.

Andrea Tirinzoni, Alessandro Lazaric, and Marcello Restelli. A novel confidence-based algorithm for structured bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 3175–3185. PMLR, 2020.

Nilesh Tripuraneni, Chi Jin, and Michael Jordan. Provable meta-learning of linear representations. In *International Conference on Machine Learning*, pp. 10434–10443. PMLR, 2021.

Alexandre B Tsybakov. *Introduction to nonparametric estimation*. Springer Science & Business Media, 2008.

Michal Valko, Rémi Munos, Branislav Kveton, and Tomáš Kocák. Spectral bandits for smooth graph functions. In *International Conference on Machine Learning*, pp. 46–54. PMLR, 2014.

Jiaqi Yang, Wei Hu, Jason D Lee, and Simon S Du. Impact of representation learning in linear bandits. *arXiv preprint arXiv:2010.06531*, 2020.

Jiaqi Yang, Qi Lei, Jason D Lee, and Simon S Du. Nearly minimax algorithms for linear bandits with shared representation. *arXiv preprint arXiv:2203.15664*, 2022.

Junwen Yang and Vincent YF Tan. Minimax optimal fixed-budget best arm identification in linear bandits. *arXiv preprint arXiv:2105.13017*, 2021.

Banghua Zhu, Evan Frick, Tianhao Wu, Hanlin Zhu, and Jiantao Jiao. Starling-7b: Improving llm helpfulness & harmlessness with rlaif, November 2023.

A Appendix

The G-optimal design (Pukelsheim, 2006; Fedorov, 2010) problem aims at finding a probability distribution $\mathbf{b} : \{\mathbf{x}(i) : i \in [A]\} \to [0,1]$ that minimises

$$g(\mathbf{b}^G) = \max_{i \in [A]} \|\mathbf{x}(i)\|^2_{\mathbf{V}(\mathbf{b}^G)^{-1}}$$

where $\mathbf{V}(\mathbf{b}) = \sum_{i \in [A]} \mathbf{b}(i)\mathbf{x}(i)\mathbf{x}(i)^\top$. Then the following lemma states the existence of a small-support G-optimal design and the minimum value of g.

Lemma A.1. *1 **(Restatement of Theorem 21.1 (Kiefer-Wolfowitz) from (Lattimore & Szepesvári, 2020)).** Assume that $\mathcal{X} \subset \mathbb{R}^d$ is compact and $\operatorname{span}(\mathcal{X}) = \mathbb{R}^d$. Then the following are equivalent:*

(a) $\mathbf{b}^G$ is a minimiser of g.

(b) $\mathbf{b}^G$ is a maximiser of $f(\mathbf{b}) = \log\det \mathbf{V}(\mathbf{b})$.

(c) $g\left(\mathbf{b}^G\right) = d$.

Furthermore, there exists a minimiser $\mathbf{b}^G$ of g such that $\left|\operatorname{Supp}\left(\mathbf{b}^G\right)\right| \leq d(d+1)/2$.

A.1 Pseudocode of Linear and Bilinear Algorithm

Now we present the bilinear FB-DOE in Algorithm 2.

A.2 Lower bounds for Linear Bandits

Theorem 1. ***(Lower Bound)** Let $|\boldsymbol{\Theta}| = 2^d$, $\boldsymbol{\theta}_{m,*} \in \boldsymbol{\Theta}$ and $M > \max_m \frac{kd\log_2 k}{\Delta_{m,\min}}$, where $\Delta_{m,\min} > 0$ is the minimum gap in task m. Then any δ-PAC policy π in the linear MTRL setting suffers a total probability of error as*

$$\Omega\left(\exp\left(-\frac{Mn}{\log_2 d}\right) + M\exp\left(-\frac{n}{H_{2,\text{lin}}\log_2 k}\right)\right)$$

for the environment in (2), where $H_{2,\text{lin}} = \max_{m\in[M]}\max_{2\le i\le k}\frac{i}{\Delta^2_{m,i}}$ is the hardness parameter.

Proof. **Step 1 (Define Environment):** We again define the environment model below for easier exposition to the reader. This is same as (2). Define the environment for the task m as D^m_{ij} consisting of A actions and J hypotheses with true hypothesis $\boldsymbol{\theta}^m_* = \boldsymbol{\theta}^m_{i,j}$ (ij-th column) as follows:

$$\begin{array}{ccccccccc}
\boldsymbol{\theta}^m & = & \mathbf{B}_1\mathbf{w}^m_1 & \mathbf{B}_1\mathbf{w}^m_2 & \mathbf{B}_1\mathbf{w}^m_3 & \dots & \mathbf{B}_i\mathbf{w}^m_j & \dots & \mathbf{B}_N\mathbf{w}^m_J \\
\hline
\mu_1(\boldsymbol{\theta}^m) & = & \beta+\Gamma & \beta+\Gamma-(\frac{\beta}{J}+\frac{\Gamma}{N}) & \beta+\Gamma-(\frac{2\beta}{J}+\frac{2\Gamma}{N}) & \dots & \beta+\Gamma-(\frac{(j-1)\beta}{J}+\frac{(i-1)\Gamma}{N}) & \dots & \beta+\Gamma-(\frac{(N-1)\beta}{J}+\frac{(N-1)\Gamma}{N}) \\
\mu_2(\boldsymbol{\theta}^m) & = & \iota_{211} & \iota_{212} & \iota_{213} & \dots & \iota_{2ij} & \dots & \iota_{2NJ} \\
 & \vdots & & & & \vdots & & & \\
\mu_A(\boldsymbol{\theta}^m) & = & \iota_{A11} & \iota_{A12} & \iota_{A13} & \dots & \iota_{Aij} & \dots & \iota_{ANJ}
\end{array} \tag{12}$$

where, each ι_{ij} is distinct and satisfies $\iota_{ij} < \beta/4J + \Gamma/4N$. $\boldsymbol{\theta}^m_{11}$ is the optimal hypothesis in D^m_{11}, $\boldsymbol{\theta}^m_{12}$ is the optimal hypothesis in D^m_{12} and so on such that for each D^m_{ij} and $i \in [N], j \in [J]$ we have column (i,j) as the optimal hypothesis.

This is a general hypothesis testing setting where the functions $\mu_a(\boldsymbol{\theta}^m)$ can be thought of as linear functions of $\boldsymbol{\theta}^m$ such that $\mu_a(\boldsymbol{\theta}^m) = \mathbf{x}_m(a)^\top\boldsymbol{\theta}^m = \mathbf{x}_m(a)\mathbf{B}_i\mathbf{w}^m_j$ for some $i \in [N]$ and $j \in [J]$. Assume that $0 < \mu_a(\boldsymbol{\theta}^m) \leq 1$. We also assume that all arms have the same variance σ^2 and $\sigma^2 > 1/4$. We will subsequently derive a suitable choice for N.

Algorithm 2 Fixed Budget Double Optimal Design (FB-DOE) for Bilinear Bandits

Input: time budget n, arm sets $\mathcal{X} \subset \mathbb{R}^{d_1}$, $\mathcal{Z} \subset \mathbb{R}^{d_2}$.
Define $\mathbf{w}(i) = \text{vec}(\mathbf{x}(i); \mathbf{z}(i)) \in \mathbb{R}^{d_1 d_2}$ for each $\mathbf{x}(i) \in \mathcal{X}$ and $\mathbf{z}(i) \in \mathcal{Z}$. Let $\mathcal{W}$ denote this new arm set.
Stage 1 (Feature Recovery): Let E-optimal design be $\mathbf{b}^E_{\mathbf{w}} = \arg\min_{\mathbf{b} \in \triangle_{\mathcal{W}}} \left\|(\sum_i \mathbf{b}(i)\mathbf{w}\,\mathbf{w}^\top)^{-1}\right\|$. Set $\tau^E_m = \frac{n}{3M}$.
Pull arm $\mathbf{w}_m(i) \in \mathcal{W}_m$ exactly $\lceil \mathbf{b}^E_{\mathbf{w}}(i)\tau^E_m \rceil$ times for each task m and observe rewards $\{r_{m,t}\}_{t=1}^{\tau^E_m}$.
Compute $\widehat{\mathbf{Z}}_n$ using (3). Let $\widehat{\mathbf{B}}_{1,n}$ be the top-k left singular vectors of $\widehat{\mathbf{Z}}_n$ and $\widehat{\mathbf{B}}_{2,n}$ be the top-k right singular vectors of $\widehat{\mathbf{Z}}_n$.
Build $\widetilde{\mathbf{g}}_m(i) = \mathbf{x}(i)^\top \widehat{\mathbf{B}}_{1,n}$ for all $\mathbf{x}(i) \in \mathcal{X}$ and $\widetilde{\mathbf{v}}_m(i) = \mathbf{z}(i)^\top \widehat{\mathbf{B}}_{2,n}$ for all $\mathbf{z}(i) \in \mathcal{Z}$ for each $m \in [M]$. Then define $\widetilde{\mathbf{w}}(i) = \text{vec}(\widetilde{\mathbf{g}}_m(i); \widetilde{\mathbf{v}}_m(i)) \in \mathbb{R}^{k_1 k_2}$ for each $\widetilde{\mathbf{g}}_m(i)$ and $\widetilde{\mathbf{v}}_m(i)$. Let $\widetilde{\mathcal{W}}_m$ denote this new arm set for each task m.
Stage 2 (Learn $\mathbf{S}_{m,*}$): Let E-optimal design be $\mathbf{b}^E_{\widetilde{\mathbf{w}}} = \arg\min_{\mathbf{b} \in \triangle_{\widetilde{\mathcal{W}}}} \left\|(\sum_i \mathbf{b}(i)\widetilde{\mathbf{w}}_m(i)\,\widetilde{\mathbf{w}}_m(i)^\top)^{-1}\right\|$. Set $\tau^E_m = \frac{n}{3M}$.
Pull arm $\widetilde{\mathbf{w}}_m(i) \in \widetilde{\mathcal{W}}_m$ exactly $\lceil \widehat{\mathbf{b}}^E_{\widetilde{\mathbf{w}}}(i)\tau^E_m \rceil$ times for each task m and observe rewards $\{r_{m,t}\}_{t=1}^{\tau^E_m}$.
Compute $\widehat{\mathbf{S}}_{m,n}$ using (3). Rotate the arms and build arm set $\underline{\mathcal{G}}_m$, s.t. each $\underline{\mathbf{g}}_m(i) \in \mathbb{R}^{\widetilde{k}}$ using (25) and $\widetilde{k} = (k_1 + k_2)r$.
Initialize $t_{m,0} = 1, \underline{\mathcal{G}}_{m,0} \leftarrow \underline{\mathcal{G}}_m$ and $\widetilde{k}_{m,0} = \widetilde{k}$. For each arm $\underline{\mathbf{g}}_m(i) \in \underline{\mathcal{G}}_{m,0}$, set $\underline{\mathbf{g}}_{m,0}(i) = \underline{\mathbf{g}}_m(i)$. Calculate $n_m(\widetilde{k})$ using (4).
Stage 3 (Low dimensional elimination)
for $\ell = 1$ to $\lceil \log_2 \widetilde{k} \rceil$ **do**
  Set $\widetilde{k}_{m,\ell} = \dim\left(\text{span}\left(\left\{\underline{\mathbf{g}}_{m,\ell-1}(i) : i \in \underline{\mathcal{G}}_{m,\ell-1}\right\}\right)\right)$.
  if $\widetilde{k}_{m,\ell} = \widetilde{k}_{m,\ell-1}$ **then**
    For each arm $i \in \underline{\mathcal{G}}_{m,\ell-1}$, set $\underline{\mathbf{g}}_m(i) = \underline{\mathbf{g}}_{m-1}(i)$.
  else
    Find matrix $\mathbf{H}_{m,\ell} \in \mathbb{R}^{\widetilde{k}_{m,\ell-1} \times \widetilde{k}_{m,\ell}}$ whose columns form an orthonormal basis of the subspace spanned by $\left\{\underline{\mathbf{g}}_{m,\ell-1}(i) : i \in \underline{\mathcal{G}}_{m,\ell-1}\right\}$. For each arm $i \in \underline{\mathcal{G}}_{m,\ell-1}$, set $\underline{\mathbf{g}}_{m,\ell}(i) = \mathbf{H}^\top_{m,\ell}\underline{\mathbf{g}}_{m,\ell-1}(i)$
  end if
  if $\ell = 1$ **then**
    Find a G-optimal design $\mathbf{b}^G_{m,\ell} : \left\{\underline{\mathbf{g}}_{m,\ell}(i) : i \in \underline{\mathcal{G}}_{m,\ell-1}\right\} \rightarrow [0,1]$ with $\left|\text{Supp}\left(\mathbf{b}^G_{m,\ell}\right)\right| \leq \frac{\widetilde{k}(\widetilde{k}+1)}{2}$.
  else
    Find a G-optimal design $\mathbf{b}^G_{m,\ell} : \left\{\underline{\mathbf{g}}_{m,\ell}(i) : i \in \underline{\mathcal{G}}_{m,\ell-1}\right\} \rightarrow [0,1]$.
  end if
  Set $N_{m,\ell}(i) = \lceil \mathbf{b}^G_{m,\ell}(\underline{\mathbf{g}}_{m,\ell}(i)) \cdot n_m(\widetilde{k}) \rceil$, $n_m(\widetilde{k})$ defined in (4), and $N_{m,\ell} = \sum_{i \in \underline{\mathcal{G}}_{m,\ell-1}} N_{m,\ell}(i)$. Choose each arm $i \in \underline{\mathcal{G}}_{m,\ell-1}$ exactly $N_{m,\ell}(i)$ times.
  Calculate the OLS estimator for each task m with the $\mathbf{\Lambda}_{m,\ell}$ defined in (10):

$$\widehat{\mathbf{s}}_{m,\ell} = \mathbf{\Sigma}^{-1}_{m,\ell} \sum_{t=t_{m,\ell}}^{t_{m,\ell}+T_{m,\ell-1}} \underline{\mathbf{g}}_m(A_t)\, r_{m,t} \quad \text{with } \mathbf{\Sigma}_{m,\ell} = \sum_{i \in \mathcal{G}_{m,\ell-1}} T_{m,\ell}(i)\underline{\mathbf{g}}_{m,\ell}(i)\underline{\mathbf{g}}_{m,\ell}(i)^\top + \mathbf{\Lambda}_{m,\ell} \tag{11}$$

  Reshape $\widehat{\mathbf{s}}_{m,\ell} \in \mathbb{R}^{k_1 k_2}$ into $\widehat{\mathbf{S}}_{m,\ell} \in \mathbb{R}^{k_1 \times k_2}$. Set $\widehat{\mathbf{\Theta}}_{m,\ell} = \widehat{\mathbf{B}}_{1,n}\widehat{\mathbf{S}}_{m,\ell}\widehat{\mathbf{B}}^\top_{2,n}$ for each task m.
  For each $i \in \underline{\mathcal{G}}_{m,\ell-1}$, estimate the expected reward: $\widehat{\mu}_{m,\ell}(i) = \mathbf{x}_m(i)^\top \widehat{\mathbf{\Theta}}_{m,\ell}\mathbf{z}_m(i)$.
  Let $\underline{\mathcal{G}}_{m,\ell}$ be the set of $\lceil \widetilde{k}/2^\ell \rceil$ arms in $\underline{\mathcal{G}}_{m,\ell-1}$ with the largest estimates of the expected rewards.
  Set $t_{m,\ell+1} = t_{m,\ell} + N_{m,\ell}$.
end for

Now observe that between any two hypothesis $\boldsymbol{\theta}^m$ and $\boldsymbol{\theta}^{m'}$ we have the following

$$\text{KL}\left(\mathcal{N}(\mu_i(\boldsymbol{\theta}^m), \sigma_i^2)) \middle|\middle| \mathcal{N}(\mu_i(\boldsymbol{\theta}^{m'}), \sigma^2))\right) = \frac{(\mu_i(\boldsymbol{\theta}^m) - \mu_i(\boldsymbol{\theta}^{m'}))^2}{2\sigma^2} \overset{(a)}{\geq} \frac{(\mu_i(\boldsymbol{\theta}^m) - \mu_i(\boldsymbol{\theta}^{m'}))^2}{8} \tag{13}$$

where, (a) follows from the condition that $\sigma^2 > 1/4$.

Step 2 (Minimum samples to verify $\boldsymbol{\theta}^m_*$): Let for the model D^m_{11} the optimal hypothesis be $\boldsymbol{\theta}^{m,*} = \boldsymbol{\theta}^m_{11}$. Let, for model D^m_{11} the Λ^m_{11} be the set of alternate models having a different optimal hypothesis than $\boldsymbol{\theta}^{m,*} = \boldsymbol{\theta}^m_{11}$ such that all models having different optimal hypothesis than $\boldsymbol{\theta}^m_{11}$ such as $D^m_{21}, D^m_{31}, \ldots D^m_{NJ}$ are in Λ^m_{11}. Let τ^m_δ be the stopping time for any δ-PAC policy π. That is τ_δ is the time that any algorithm stops and outputs its estimate $\widehat{\boldsymbol{\theta}}_{\tau_\delta}$. We will subsequently choose a suitable value of δ to satisfy the constraint of the budget n.

Let $T_{m,t}(a)$ denote the number of times the action a has been sampled till round t for the task m. Let $\widehat{\boldsymbol{\theta}}^m_{\tau_\delta}$ be the predicted optimal hypothesis at round τ^m_δ. We first consider the model D^m_{11}. Define the event $\xi = \{\widehat{\boldsymbol{\theta}}_{\tau^m_\delta} \neq \boldsymbol{\theta}^m_*\}$ as the error event in model D^m_{11}. Let event $\xi' = \{\widehat{\boldsymbol{\theta}}_{\tau_\delta} \neq \boldsymbol{\theta}'_{m,*}\}$ be the corresponding error event in model D^m_{12}. Note that $\xi^{\complement} \subset \xi'$. Since π is δ-PAC policy we have $\mathbb{P}_{D^m_{11},\pi}(\xi) \leq \delta$ and $\mathbb{P}_{D^m_{12},\pi}(\xi^{\complement}) \leq \delta$. Then

$$
\begin{aligned}
2\delta \geq \mathbb{P}_{D^m_{11},\pi}(\xi) + \mathbb{P}_{D^m_{12},\pi}(\xi^{\complement}) &\overset{(a)}{\geq} \frac{1}{2}\exp\left(-\mathrm{KL}\left(P_{D^m_{11},\pi} || P_{D^m_{12},\pi}\right)\right) \\
\mathrm{KL}\left(P_{D^m_{11},\pi} || P_{D^m_{12},\pi}\right) &\geq \log\left(\frac{1}{4\delta}\right) \\
\frac{1}{8}\sum_{i=1}^{A}\mathbb{E}_{D^m_{11},\pi}[T_{m,\tau_\delta}(i)]\cdot\left(\mu_i(\boldsymbol{\theta}_{m,*}) - \mu_i(\boldsymbol{\theta}'_{m,*})\right)^2 &\overset{(b)}{\geq} \log\left(\frac{1}{4\delta}\right) \\
\frac{1}{8}\left(\beta + \Gamma - \beta + \frac{\beta}{J} - \Gamma + \frac{\Gamma}{N}\right)^2\mathbb{E}_{D^m_{11},\pi}[T_{m,\tau_\delta}(1)] + \frac{1}{8}\sum_{i=2}^{A}(\iota_{i1} - \iota_{i2})^2\mathbb{E}_{D^m_{11},\pi}[T_{m,\tau_\delta}(i)] &\overset{(c)}{\geq} \log\left(\frac{1}{4\delta}\right) \\
\frac{1}{8}\left(\frac{\beta}{J} + \frac{\Gamma}{N}\right)^2\mathbb{E}_{D^m_{11},\pi}[T_{m,\tau_\delta}(1)] + \frac{1}{8}\sum_{i=2}^{A}(\iota_{i11} - \iota_{i12})^2\mathbb{E}_{D^m_{11},\pi}[T_{m,\tau_\delta}(i)] &\geq \log\left(\frac{1}{4\delta}\right) \\
\frac{1}{8}\left(\frac{\beta}{J} + \frac{\Gamma}{N}\right)^2\mathbb{E}_{D^m_{11},\pi}[T_{m,\tau_\delta}(1)] + \frac{1}{8}\sum_{i=2}^{A}\frac{1}{16}\left(\frac{\beta}{J} + \frac{\Gamma}{N}\right)^2\mathbb{E}_{D^m_{11},\pi}[T_{m,\tau_\delta}(i)] &\overset{(d)}{\geq} \log\left(\frac{1}{4\delta}\right)
\end{aligned}
$$

where, (a) follows from Lemma A.3, (b) follows from Lemma A.2, (c) follows from the construction of the bandit environments and (13), and (d) follows as $(\iota_{aij} - \iota_{aij'})^2 \leq \frac{1}{16}\left(\frac{\beta}{J} + \frac{\Gamma}{N}\right)^2$ for any i-th action and j-th hypothesis.

Now, we consider the alternate model D^m_{13}. Again define the event $\xi = \{\widehat{\boldsymbol{\theta}}_{\tau_\delta} \neq \boldsymbol{\theta}_{m,*}\}$ as the error event in model D^m_{11} and the event $\xi' = \{\widehat{\boldsymbol{\theta}}_{\tau_\delta} \neq \boldsymbol{\theta}''_{m,*}\}$ be the corresponding error event in model D^m_{31}. Note that $\xi^{\complement} \subset \xi'$. Now since π is δ-PAC policy we have $\mathbb{P}_{D^m_{11},\pi}(\xi) \leq \delta$ and $\mathbb{P}_{D^m_{13},\pi}(\xi^{\complement}) \leq \delta$. Following the same way as before we can show that,

$$
\frac{1}{8}\left(\frac{2\beta}{J} + \frac{2\Gamma}{N}\right)^2\mathbb{E}_{D^m_{13},\pi}[T_{m,\tau_\delta}(1)] + \frac{1}{8}\sum_{i=2}^{A}\frac{1}{16}\left(\frac{\beta}{J} + \frac{\Gamma}{N}\right)^2\mathbb{E}_{D^m_{13},\pi}[T_{m,\tau_\delta}(i)] \overset{(d)}{\geq} \log\left(\frac{1}{4\delta}\right). \tag{14}
$$

Similarly, we get the equations for all the other $(NJ-2)$ alternate models in Λ_{11}^m. Now consider an optimization problem (ignoring the constant factor of $\frac{1}{8}$ across all the constraints)

$$
\begin{aligned}
&\min_{t_i: i\in[A]} \sum t_i \\
s.t.\quad &\left(\frac{\beta}{J}+\frac{\Gamma}{N}\right)^2 t_1 + \frac{1}{16}\left(\frac{\beta}{J}+\frac{\Gamma}{N}\right)^2 \sum_{i=2}^{A} t_i \geq \log(1/4\delta) \\
&\left(\frac{2\beta}{J}+\frac{2\Gamma}{N}\right)^2 t_1 + \frac{1}{16}\left(\frac{\beta}{J}+\frac{\Gamma}{N}\right)^2 \sum_{i=2}^{A} t_i \geq \log(1/4\delta) \\
&\vdots \\
&\left(\frac{(J-1)\beta}{J}+\frac{(N-1)\Gamma}{N}\right)^2 t_1 + \frac{1}{16}\left(\frac{\beta}{J}+\frac{\Gamma}{N}\right)^2 \sum_{i=2}^{A} t_i \geq \log(1/4\delta) \\
&t_i \geq 0, \forall i \in [A]
\end{aligned}
$$

where the optimization variables are t_i. It can be seen that the optimum objective value is $\left(\frac{\beta}{J}+\frac{\Gamma}{N}\right)^2 \log(1/4\delta)$. Interpreting $t_i = \mathbb{E}_{D_{11}^m,\pi}[T_{m,\tau_\delta}(i)]$ for all i, we get that $\mathbb{E}_{D_{11}^m,\pi}[\tau_\delta] = \sum_i t_i = t_1$. Now we have that $t_1 \geq J^2\beta^{-2}\log(1/4\delta)$ which gives us the required lower bound to the number of pulls of action 1 for task m. Observe that the optimum objective value is reached by substituting $t_1 = \frac{1}{16}\left(\frac{\beta}{J}+\frac{\Gamma}{N}\right)^2 \log(1/4\delta)$ and $t_2 = \ldots = t_A = 0$. It follows that for verifying any hypothesis $\boldsymbol{\theta}_j^m \neq \boldsymbol{\theta}_*^m$ the verification proportion is given by $\pi_{\boldsymbol{\theta}_j^m} = (1, \underbrace{0, 0, \ldots, 0}_{\text{(A-1) zeros}})$.

Observe setting $\frac{1}{16}\left(\frac{\beta}{J}+\frac{\Gamma}{N}\right) \geq \sqrt{\log(1/4\delta)/n}$ recovers $\tau_\delta \geq n$ which implies that a budget of atmost n samples is required for verifying hypothesis $\boldsymbol{\theta}_j^m = \boldsymbol{\theta}_{m,*}$. For the remaining steps we take $\left(\frac{n\beta^2}{J^2}+\frac{n\Gamma^2}{N^2}\right) \geq \log(1/4\delta)/n$. This implies that

$$
\begin{aligned}
\log(1/4\delta)/n \leq \frac{1}{16}\left(\frac{\beta}{J}+\frac{\Gamma}{N}\right)^2 &\overset{(a)}{\implies} \log(1/4\delta) \leq \frac{n}{8}\left(\frac{\beta^2}{J^2}+\frac{\Gamma^2}{N^2}\right) \\
&\implies 1/4\delta \leq \exp\left(\frac{n\beta^2}{J^2}+\frac{n\Gamma^2}{N^2}\right) \\
&\implies \delta \geq \frac{1}{4}\exp\left(-\frac{n\beta^2}{J^2}-\frac{n\Gamma^2}{N^2}\right) \\
&\overset{(b)}{\implies} \delta \geq \frac{1}{4}\exp\left(-\frac{2n\beta^2}{J^2}\right)+\frac{1}{4}\exp\left(-\frac{2n\Gamma^2}{N^2}\right)
\end{aligned}
$$

where, (a) follows as $(a+b)^2 \leq 2(a^2+b^2)$ for $a, b > 0$, (b) follows as $\exp(-a-b) \geq \exp(-2a) + \exp(-2b)$ if $b < a$. This implies that $n\Gamma^2/N^2 < n\beta^2/J^2$ for sufficiently large N. This also shows a suitable lower bound to δ that depends on the budget n.

Then the total probability of error across all the M tasks is given by

$$
\begin{aligned}
M\delta &\geq \frac{M}{4}\exp\left(-\frac{2n\beta^2}{J^2}\right)+\frac{M}{4}\exp\left(-\frac{2n\Gamma^2}{N^2}\right) \\
&\geq \frac{M}{4}\exp\left(-\frac{2n\beta^2}{J^2}\right)+\frac{M}{4}\exp\left(-\frac{2n\Gamma^2}{N^2}\right). \qquad (15)
\end{aligned}
$$

Recall that $H_{1,\text{lin}} = \max_{m\in[M]} \sum_{i=1}^{k} \frac{1}{\Delta_{m,i}^2}$, and $\Delta_{m,\min}^2 = \min_i \Delta_{m,i}$. Then we can show that

$$
H_{1,\text{lin}} = \max_{m\in[M]} \sum_{i=1}^{k} \frac{1}{\Delta_{m,i}^2} \geq \max_{m\in[M]} \max_{i\in[k]} \frac{i}{\Delta_{m,(i)}^2} = H_{2,\text{lin}}
$$

It follows that $H_{2,\text{lin}} \leq H_{1,\text{lin}} \leq (\log_2 k) H_{2,\text{lin}}$. Now setting $\frac{J^2}{\beta^2} = \max_{m\in[M]} \frac{k \log_2 k}{\Delta^2_{m,\min}}$ we have that

$$-\max_{m\in[M]} \frac{\Delta^2_{m,\min}}{k\log_2 k} \geq -\frac{1}{H_{2,\text{lin}} \log_2 k}$$

and setting $\Gamma^2 = \frac{1}{\log_2 d}$ we have that

$$\frac{\Gamma^2}{N^2} < \frac{\beta^2}{J^2} \implies \frac{1}{N^2 \log_2 d} < \max_{m\in[M]} \frac{\Delta^2_{m,\min}}{k \log_2 k} \implies N^2 > \max_{m\in[M]} \frac{k\log_2 k}{\log_2 d \Delta^2_{m,\min}} \implies N > \max_{m\in[M]} \frac{kd\log_2 k}{\Delta_{m,\min}}$$

satisfies all the above conditions. Plugging everything back in (15) we have that

$$\begin{aligned}
M\delta &\geq \frac{M}{4}\exp\left(-\frac{2n}{N^2 \log_2 d}\right) + \frac{M}{4}\exp\left(-\frac{2n}{H_{2,\text{lin}}\log_2 k}\right) \\
&\overset{(a)}{\geq} \frac{1}{4}\exp\left(-\frac{2Mn}{\log_2 d}\right) + \frac{M}{4}\exp\left(-\frac{2n}{H_{2,\text{lin}}\log_2 k}\right)
\end{aligned}$$

where, (a) follows as for $N^2 > \frac{2n}{\log_2 d \log M + 2nM}$ we have that

$$\frac{M}{4}\exp\left(-\frac{2n}{N^2\log_2 d}\right) > \frac{1}{4}\exp\left(-\frac{2Mn}{\log_2 d}\right).$$

Note that as $\frac{2n}{\log_2 d \log M + 2nM} > 0$ the condition for N is satisfied by any budget $n \geq 1$ and number of tasks $M \geq 1$. Hence, for $M > \max_m \frac{kd\log_2 k}{\Delta_{m,\min}}$ we have all the conditions satisfied. The claim of the theorem follows. □

Lemma A.2. ***(Restatement of Lemma 15.1 in Lattimore & Szepesvári (2020), Divergence Decomposition)*** *Let B and B' be two bandit models having different optimal hypothesis $\boldsymbol{\theta}_*$ and $\boldsymbol{\theta'}^*$ respectively. Fix some policy π and round n. Let $\mathbb{P}_{B,\pi}$ and $\mathbb{P}_{B',\pi}$ be two probability measures induced by some n-round interaction of π with B and π with B' respectively. Then*

$$\mathrm{KL}\left(\mathbb{P}_{B,\pi} || \mathbb{P}_{B',\pi}\right) = \sum_{i=1}^{A} \mathbb{E}_{B,\pi}[T_n(i)] \cdot \mathrm{KL}(\mathcal{N}(\mu_i(\boldsymbol{\theta}),1) || \mathcal{N}(\mu_i(\boldsymbol{\theta}_*),1))$$

where, $\mathrm{KL}\left(.||.\right)$ *denotes the Kullback-Leibler divergence between two probability measures and* $T_n(i)$ *denotes the number of times action i has been sampled till round n.*

Lemma A.3. ***(Restatement of Lemma 2.6 in Tsybakov (2008))*** *Let $\mathbb{P}, \mathbb{Q}$ be two probability measures on the same measurable space $(\Omega, \mathcal{F})$ and let $\xi \subset \mathcal{F}$ be any arbitrary event then*

$$\mathbb{P}(\xi) + \mathbb{Q}\left(\xi^{\complement}\right) \geqslant \frac{1}{2}\exp\left(-\mathrm{KL}(\mathbb{P}||\mathbb{Q})\right)$$

where $\xi^{\complement}$ denotes the complement of event ξ and $\mathrm{KL}(\mathbb{P}||\mathbb{Q})$ *denotes the Kullback-Leibler divergence between $\mathbb{P}$ and $\mathbb{Q}$.*

A.3 Linear Bandit Fixed Budget Proofs

Define $\mathbf{X}^+_{\text{batch}} := \left(\mathbf{X}^\top_{\text{batch}} \mathbf{X}_{\text{batch}}\right)^{-1} \mathbf{X}^\top_{\text{batch}}$ where $\mathbf{X}^+_{\text{batch}} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{\tau^E_m}]^\top$ is constructed through the E-optimal design. Also note that $\rho^E_1 = \rho^E_2 = \ldots = \rho^E_M = \rho^E$ as the action set $\mathcal{X}$ is common across the tasks. Also, recall that

$$\widehat{\boldsymbol{\theta}}_{m,t} = \mathbf{X}^+_{\text{batch}} r_{m,t}.$$

Good Event: Define the good event $\mathcal{F}_n$ that the algorithm has a good estimate of $\mathbf{Z}_*$ as follows:

$$\mathcal{F}_n = \left\{ \left\| \widehat{\mathbf{Z}}_n - \mathbf{Z} \right\|_F \leq C \left\| \mathbf{X}^+_{\text{batch}} \right\|^2 \left(\frac{2cd^2}{\sqrt{Mn}} \exp\left(-\frac{Mn}{2}\right) \right) \right\} \tag{16}$$

where, $C_1 > 0$, some nonzero constant.

Lemma A.4. ***(Restatement of Lemma C.3 from (Du et al., 2023))*** *Define the event*

$$\xi_n := \left\{ \|\mathbf{Z}_n - \mathbb{E}[\mathbf{Z}_n]\| \leq \frac{c\left\|\mathbf{X}_{batch}^{+}\right\|^2 d \log\left(\frac{16d}{\delta_n}\right)}{\sqrt{Mn}} \log\left(\frac{16dMn}{\delta_n}\right) \right\}$$

Then it holds that $\Pr(\xi_n) \geq 1 - \frac{\delta_n}{2}$.

Lemma A.5. ***(Truncated Matrix Bernstern Inequality - Summation)*** *Consider a truncation level $U > 0$. If $\{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n\}$ is a sequence of $d_1 \times d_2$ independent random matrices, and $\boldsymbol{Z}_i' = \boldsymbol{Z}_i \cdot \mathbf{1}\{\|\boldsymbol{Z}_i\| \leq U\}$ and $\Delta \geq \|\mathbb{E}[\boldsymbol{Z}_i] - \mathbb{E}[\boldsymbol{Z}_i']\|$ for any $i \in [n]$, then for $\tau \geq 2n\Delta$,*

$$\Pr\left[\left\|\sum_{i=1}^{n}(\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i])\right\| \geq \tau\right] \leq (d_1+d_2)\exp\left(-\frac{1}{4}\cdot\frac{\tau^2}{2\sigma^2+\frac{U\tau}{3}}\right) + n\Pr\left[\|\boldsymbol{Z}_i\| \geq U\right],$$

where

$$\begin{aligned}\sigma^2 &= \max\left\{\left\|\sum_{i=1}^{n}\mathbb{E}\left[(\boldsymbol{Z}_i' - \mathbb{E}[\boldsymbol{Z}_i'])^\top(\boldsymbol{Z}_i' - \mathbb{E}[\boldsymbol{Z}_i'])\right]\right\|, \left\|\sum_{i=1}^{n}\mathbb{E}\left[(\boldsymbol{Z}_i' - \mathbb{E}[\boldsymbol{Z}_i'])(\boldsymbol{Z}_i' - \mathbb{E}[\boldsymbol{Z}_i'])^\top\right]\right\|\right\} \\ &\leq \max\left\{\left\|\sum_{i=1}^{n}\mathbb{E}\left[\boldsymbol{Z}_i'^\top\boldsymbol{Z}_i'\right]\right\|, \left\|\sum_{i=1}^{n}\mathbb{E}\left[\boldsymbol{Z}_i'\boldsymbol{Z}_i'^\top\right]\right\|\right\}\end{aligned}$$

Furthermore, we have

$$\Pr\left[\left\|\sum_{i=1}^{n}(\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i])\right\| \geq 4\sqrt{\sigma^2\log\left(\frac{d_1+d_2}{\delta}\right)} + 4U\log\left(\frac{d_1+d_2}{\delta}\right)\right] \leq \delta + n\Pr\left[\|\boldsymbol{Z}_i\| \geq U\right].$$

Lemma A.6. *Define the event*

$$\mathcal{F}_n := \left\{\|\mathbf{Z}_n - \mathbb{E}[\mathbf{Z}_n]\| \geq C\left\|\mathbf{X}_{batch}^{+}\right\|^2\left(\frac{2cd^2}{\sqrt{Mn}}\exp\left(-\frac{Mn}{2}\right)\right)\right\}$$

It follows then that

$$\mathbb{P}(\mathcal{F}_n) \leq 4d\exp(-Mn)$$

Proof. We use the truncated Matrix Bernstein inequality (Lemma A.5) to prove the exponentially low probability of error in the following way. Set $R = \sqrt{Mn}$ and define the truncation matrix $\mathbf{A}_n$ as follows:

$$\begin{aligned}\boldsymbol{A}_{m,t} &:= \frac{1}{M}\begin{bmatrix} 2\mathbf{x}_1^\top\boldsymbol{\theta}_m\eta_{m,1} & \cdots & \mathbf{x}_1^\top\boldsymbol{\theta}_m\eta_{m,1} + \mathbf{x}_{n/2}^\top\boldsymbol{\theta}_m\eta_{m,n/2} \\ \cdots & \cdots & \cdots \\ \mathbf{x}_1^\top\boldsymbol{\theta}_m\eta_{m,1} + \mathbf{x}_{n/2}^\top\boldsymbol{\theta}_m\eta_{m,n/2} & \cdots & 2\mathbf{x}_{n/2}^\top\boldsymbol{\theta}_m\eta_{m,n/2}\end{bmatrix} \\ \boldsymbol{A}_n &:= \sum_{m=1}^{M}\sum_{t=1}^{n/2}\boldsymbol{A}_{m,t}\end{aligned}$$

and truncation matrix $\mathbf{C}_n$ as:

$$\begin{aligned}\boldsymbol{C}_{m,t} &:= \frac{1}{M}\begin{bmatrix} (\eta_{m,1})^2 & \cdots & \eta_{m,1}\eta_{m,n/2} \\ \cdots & \cdots & \cdots \\ \eta_{m,1}\eta_{m,n/2} & \cdots & (\eta_{m,n/2})^2\end{bmatrix} \\ \boldsymbol{C}_n &:= \sum_{m=1}^{M}\sum_{j=1}^{n/2}\boldsymbol{C}_{m,t}.\end{aligned}$$

Then it can be shown easily using Lemma A.4 that the average estimation matrix $\mathbf{Z}_n$ can be upper bounded as

$$\|\mathbf{Z}_n - \mathbb{E}[\mathbf{Z}_n]\| \leq \left\|\mathbf{X}^{+}_{\text{batch}}\right\|^2 (\|\mathbf{A}_n - \mathbb{E}[\mathbf{A}_n]\| + \|\mathbf{C}_n - \mathbb{E}[\mathbf{C}_n]\|)$$

such that $\|\mathbf{A}_{m,t}\| \leq \frac{2}{Mn} \cdot 2dcR$, and $\|\mathbf{C}_{m,t}\| \leq \frac{2}{Mn} \cdot 2dc'R$ where $c, c' > 0$. Note that $\|\mathbf{C}_{m,t,i}\| \leq \frac{1}{Mn} \cdot 2dc'R$ because $\log(n/\delta) \leq \sqrt{Mn}$. Now using the truncated Matrix Bernstein inequality in Lemma A.5 we have that

$$\|\mathbf{Z}_n - \mathbb{E}[\mathbf{Z}_n]\| \leq \left\|\mathbf{X}^{+}_{\text{batch}}\right\|^2 \left(\frac{2dc}{Mn} \cdot 2d \cdot \left(R + \frac{1}{R}\right) \exp\left(-\frac{R^2}{2}\right) + \frac{d}{Mn} \cdot 2dc' \cdot \left(R + \frac{1}{R}\right) \exp\left(-\frac{R^2}{2}\right)\right)$$

holds as the noise $|\eta_{m,t}| \leq R$ with probability $1 - 4d \exp\left(-\frac{R^2}{2}\right)$ because $\eta_{m,t}$ is 1-sub Gaussian and $c, c' > 0$. Setting $R = \sqrt{Mn}$ we have that

$$\begin{aligned}
&\|\mathbf{Z}_n - \mathbb{E}[\mathbf{Z}_n]\| \\
&\leq \left\|\mathbf{X}^{+}_{\text{batch}}\right\|^2 \left(\frac{2dc}{Mn} \cdot 2d \cdot \left(\sqrt{Mn} + \frac{1}{\sqrt{Mn}}\right) \exp\left(-\frac{Mn}{2}\right) + \frac{d}{Mn} \cdot 2dc' \cdot \left(\sqrt{Mn} + \frac{1}{\sqrt{Mn}}\right) \exp\left(-\frac{Mn}{2}\right)\right) \\
&\leq \left\|\mathbf{X}^{+}_{\text{batch}}\right\|^2 \left(\frac{2cd^2}{\sqrt{Mn}} \exp\left(-\frac{Mn}{2}\right) + \frac{2cd^2}{(Mn)^{3/2}} \exp\left(-\frac{Mn}{2}\right) + \frac{2c'd^2}{\sqrt{Mn}} \exp\left(-\frac{Mn}{2}\right) + \frac{2c'd^2}{(Mn)^{3/2}} \exp\left(-\frac{Mn}{2}\right)\right) \\
&\leq C \left\|\mathbf{X}^{+}_{\text{batch}}\right\|^2 \left(\frac{2cd^2}{\sqrt{Mn}} \exp\left(-\frac{Mn}{2}\right)\right)
\end{aligned}$$

The claim of the lemma follows. □

Lemma A.7. ***(Restatement of Lemma C.2 in (Du et al., 2023))*** *Define the total number of samples*

$$T = \left\lceil \frac{C\left(\rho^E\right)^2 k^4}{M} \text{polylog}\left(\rho^E, d, k, \frac{1}{\delta}\right)\right\rceil$$

where C is an absolute constant. For a budget $n > 0$, task $m \in [M]$, round $t \in [T]$. we have that

$$\widehat{\boldsymbol{\theta}}_{m,t} = \mathbf{X}^{+}_{\textit{batch}}\, r_{m,t},$$

and

$$\boldsymbol{Z}_T = \frac{1}{M} \sum_{m=1}^{M} \sum_{t=1}^{T} \widehat{\boldsymbol{\theta}}_{m,t} \left(\widehat{\boldsymbol{\theta}}_{m,t}\right)^{\top} - \mathbf{X}^{+}_{\textit{batch}} \left(\mathbf{X}^{+}_{\textit{batch}}\right)^{\top}.$$

It holds then

$$\mathbb{E}[\boldsymbol{Z}_T] = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{\theta}_m \boldsymbol{\theta}_m^{\top}$$

Lemma A.8. *(Expectation of $\widehat{\mathbf{Z}}_n$). It holds that for $n > \frac{2L_x^4 k^2 d^2 c' (\rho^E)^2 \log^2(2d)}{\omega^2 M \Delta^2}$ the $\mathbb{E}\left[\widehat{\mathbf{Z}}_n\right] = \mathbf{Z} = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{\theta}_{m,*} (\boldsymbol{\theta}_{m,*})^{\top}$.*

Proof. First note that the total number of samples in stage 1 is sufficiently high such that

$$\frac{n}{2} > \frac{L_x^4 k^2 d^2 c' (\rho^E)^2 \log^2(2d)}{\omega^2 M \Delta^2} \geq \left\lceil \frac{Cn\left(\rho^E\right)^2 k^4}{M} \text{polylog}\left(\rho^E, d, k\right)\right\rceil$$

for some constant $C > 0$ and $\delta = c' \exp(-n)$ for som $c' > 0$. Then for the first stage after $\frac{n}{2}$ samples we can re-write

$$\widehat{\mathbf{Z}}_n = \frac{2}{M \sum_m \tau_m^E} \sum_{m=1}^{M} \sum_{t=1}^{\tau_m^E} \widehat{\boldsymbol{\theta}}_{m,t} \widehat{\boldsymbol{\theta}}_{m,t}^\top - \mathbf{X}_{\text{batch}}^{+} \left(\mathbf{X}_{\text{batch}}^{+}\right)^\top .$$

Now using Lemma A.7 we can prove the claim of the lemma. $\square$

Lemma A.9. *(Concentration of $\widehat{\mathbf{B}}_n$). Suppose that event $\mathcal{F}_n$ holds. Then, for any $n > 0$,*

$$\left\|(\widehat{\mathbf{B}}_n^{\perp})^\top \mathbf{B}\right\| \leq c' \rho^E \left(\frac{2ckd}{\sqrt{Mn}} \exp\left(-\frac{Mn}{2}\right)\right)$$

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

Proof. Using the Davis-Kahan $\sin\theta$ Theorem (Bhatia, 2013) and letting τ_m^E be large enough to satisfy $\left\|\widehat{\mathbf{Z}}_n - \mathbf{Z}\right\|_F \leq \frac{C_1 d \log(2d)}{\sqrt{M \sum_m \tau_m^E}}$, we have

$$\begin{aligned}
\left\|(\widehat{\mathbf{B}}_n^{\perp})^\top \mathbf{B}\right\| &\leq \frac{\left\|\widehat{\mathbf{Z}}_n - \mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]\right\|}{\sigma_r\left(\mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]\right) - \sigma_{r+1}\left(\mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]\right) - \left\|\widehat{\mathbf{Z}}_n - \mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]\right\|} \\
&\overset{(a)}{\leq} \frac{k}{c_0} \left\|\widehat{\mathbf{Z}}_n - \mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]\right\| \\
&\overset{(b)}{\leq} \frac{ck \left\|\mathbf{X}_{\text{batch}}^{+}\right\|^2 d}{\sqrt{M \sum_m \tau_m^E}} \exp\left(-\frac{Mn}{2}\right) \\
&\overset{(c)}{\leq} \frac{c' \rho^E kd}{\sqrt{M \sum_m \tau_m^E}} \exp\left(-\frac{Mn}{2}\right) = \frac{c' \rho^E kd}{\sqrt{Mn}} \exp\left(-\frac{Mn}{2}\right)
\end{aligned}$$

where, (a) follows from Assumption 2.1, the (b) follows from event $\mathcal{F}_n$ and (c) follows as $\left\|\mathbf{X}_{\text{batch}}^{+}\right\|^2 \leq 4\rho^E$, and $\tau_m^E = \frac{n}{2M}$. The claim of the lemma follows. $\square$

We now need to show that $\sigma_{\min}(\sum_{\widetilde{\mathbf{g}}_m(i) \in \mathcal{G}} \mathbf{b}_m(i) \widetilde{\mathbf{g}}(i) \widetilde{\mathbf{g}}(i)^\top) > 0$. If this holds true then we can sample the following G-optimal design and the solution to the G-optimal design in the second phase is not vacuous.

Lemma A.10. *For $Mn > \lceil \frac{L_x^4 k^2 d^2 c' (\rho^E)^2 \log^2(2d)}{\omega^2 \Delta^2} \rceil$ we have*

$$\sigma_{\min}\big(\sum_{i \in \mathcal{G}_m} \mathbf{b}_m^G(i) \widetilde{\mathbf{g}}_m(i) \widetilde{\mathbf{g}}_m(i)^\top\big) > 0$$

Proof. We can show that

$$\sum_{i \in \mathcal{G}_m} \mathbf{b}_m^G(i) \widetilde{\mathbf{g}}_m(i) \widetilde{\mathbf{g}}_m(i)^\top \overset{(a)}{=} \sum_{i \in \mathcal{G}_m} \mathbf{b}_m^G(i) \underbrace{\widehat{\mathbf{B}}_n^\top \mathbf{x}(i)}_{\widetilde{\mathbf{g}}_m(i)} \underbrace{\mathbf{x}(i) \widehat{\mathbf{B}}_n^\top}_{\widetilde{\mathbf{g}}_m(i)^\top}$$

where, in (a) the $\mathbf{b}_m^G(i)$ is the sampling proportion for the arm $\mathbf{x}(i)$ in second stage. Also note that from Lemma A.9 we know that

$$\left\|(\widehat{\mathbf{B}}_n^{\perp})^\top \mathbf{B}\right\| \leq \frac{c'\rho^E kd}{\sqrt{Mn}} \exp\left(-\frac{Mn}{2}\right) \overset{(a)}{\leq} \frac{\omega}{L_x^2} \frac{c'\rho^E kd\Delta}{c'kd\rho^E \log(2d)} \exp\left(-\frac{M}{2} \cdot \frac{L_x^4 d^2 c'(\rho^E)^2 \log^2(2d)}{\omega^2 M \Delta^2}\right)$$
$$= \frac{\omega}{L_x^2} \underbrace{\frac{\Delta}{\log(2d)}}_{\leq 1} \exp\left(-\frac{d^2 c'(\rho^E)^2 \log^2(2d)}{2\omega^2\Delta^2}\right)$$
$$\leq \frac{\omega}{L_x^2} \underbrace{\exp\left(-\frac{d^2 c'(\rho^E)^2 \log^2(2d)}{2\omega^2\Delta^2}\right)}_{\leq 1}$$

where (a) follows by substituting the value of n, and observe that the last inequality does not depend on the number of tasks M or budget n. Hence for $Mn \geq \lceil \frac{L_x^4 k^2 d^2 c'(\rho^E)^2 \log^2(2d)}{\omega^2\Delta^2} \rceil$ we have

$$\left\|\widehat{\mathbf{B}}_n^\top \mathbf{B}^\perp\right\| \leq \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}. \tag{17}$$

This holds with high probability as the event $\mathcal{F}_n$ holds true. This helps us to apply Lemma A.11 to get the claim of the lemma. ∎

Lemma A.11. ***(Restatement of Lemma C.5 from Du et al. (2023))*** *For any round $n > 0$ and task $m \in [M]$, if $\left\|\widehat{\mathbf{B}}_n^\top \mathbf{B}^\perp\right\| \leq \frac{\omega}{L_x^2}$ then we have*

$$\sigma_{\min}\left(\sum_{i=1}^{A} \mathbf{b}_m^G(i)\, \widehat{\mathbf{B}}_n^\top \mathbf{x}(i)\mathbf{x}(i)^\top \widehat{\mathbf{B}}_n\right) > 0$$

where $\mathbf{b}_m^G(i)$ is the sampling proportion of $\mathbf{x}(i)$.

Lemma A.12. *Suppose that event $\mathcal{F}_n$ holds and $Mn > \lceil \frac{L_x^4 k^2 d^2 c'(\rho^E)^2 \log^2(2d)}{\omega^2\Delta^2} \rceil$. Then define*

$$\mathbf{\Sigma}_{m,\ell} = \sum_{i \in \mathcal{G}_{m,\ell-1}} \mathbf{b}_{m,\ell}(i)\, \widetilde{\mathbf{g}}_{m,\ell}(i)\widetilde{\mathbf{g}}_{m,\ell}(i)^\top.$$

For any task $m \in [M]$ and $\widetilde{\mathbf{g}}_{m,\ell}(j) \in \mathbb{R}^k$,

$$\|\widetilde{\mathbf{g}}_{m,\ell}(j)\|^2_{\mathbf{\Sigma}_{m,\ell}^{-1}} \leq \|\mathbf{g}_m(j)\|^2_{\mathbf{\Sigma}_{m,\ell}^{-1}} + \frac{cL_x^4}{k\omega^2}\exp(-Mn)$$

for some constant $c > 0$.

Proof. Observe that we can rewrite the

$$\|\widetilde{\mathbf{g}}_{m,\ell}(j)\|^2_{\mathbf{\Sigma}_{m,\ell}^{-1}} = \left\|\widehat{\mathbf{B}}_n^\top \mathbf{x}_j\right\|^2_{\left(\sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{b}_{m,\ell}^G(i)\widehat{\mathbf{B}}_n^\top \mathbf{x}(i)\mathbf{x}(i)^\top \widehat{\mathbf{B}}_n\right)^{-1}}$$

Then we can show that

$$\sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{b}_{m,\ell}^G(i)\, \widehat{\mathbf{B}}_n^\top \mathbf{x}(i)\mathbf{x}(i)^\top \widehat{\mathbf{B}}_n = \sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{b}_{m,\ell}^G(i) \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}(i)\right) \cdot \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}(i)\right)^\top$$
$$+ \sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{b}_{m,\ell}^G(i) \left(\left(\widehat{\mathbf{B}}_n^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}(i)\right) \cdot \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}_\perp \mathbf{B}_\perp^\top \mathbf{x}(i)\right)^\top\right.$$
$$\left. + \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}_\perp \mathbf{B}_\perp^\top \mathbf{x}(i)\right) \cdot \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}(i)\right)^\top + \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}_\perp \mathbf{B}_\perp^\top \mathbf{x}(i)\right) \cdot \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}_\perp \mathbf{B}_\perp^\top \mathbf{x}(i)\right)^\top\right)$$

Then define the matrix

$$\mathbf{P}_n = \sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{b}^G_{m,\ell}(i) \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}(i)\right) \cdot \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}(i)\right)^\top$$

$$\mathbf{Q}_n = \sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{b}^G_{m,\ell}(i) \Bigg(\left(\widehat{\mathbf{B}}_n^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}(i)\right) \cdot \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}_\perp \mathbf{B}_\perp^\top \mathbf{x}(i)\right)^\top + \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}_\perp \mathbf{B}_\perp^\top \mathbf{x}(i)\right) \cdot \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}(i)\right)^\top + \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}_\perp \mathbf{B}_\perp^\top \mathbf{x}(i)\right) \cdot \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}_\perp \mathbf{B}_\perp^\top \mathbf{x}(i)\right)^\top \Bigg).$$

Then, we have $\sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{b}^G_{m,\ell}(i)\, \widehat{\mathbf{B}}_n^\top \mathbf{x}(i)\mathbf{x}(i)^\top \widehat{\mathbf{B}}_n = \mathbf{P}_n + \mathbf{Q}_n$.

From Assumption 2.2, we have that for any task $m \in [M]$, $\sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{b}^G_{m,\ell}(i)\, \mathbf{B}^\top \mathbf{x}(i)\mathbf{x}(i)^\top \mathbf{B}$ is invertible. Since $\widehat{\mathbf{B}}_n^\top \mathbf{B}$ is also invertible, we have that $\mathbf{P}_n$ is invertible. According to Lemma A.10, we have that $\sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{b}^G_{m,\ell}(i)\, \widehat{\mathbf{B}}_n^\top \mathbf{x}(i)\mathbf{x}(i)^\top \widehat{\mathbf{B}}_n$ is also invertible. Then we can write $\left(\sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{b}^G_{m,\ell}(i)\, \widehat{\mathbf{B}}_n^\top \mathbf{x}(i)\mathbf{x}(i)^\top \widehat{\mathbf{B}}_n\right)^{-1}$ as follows

$$\left(\sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{b}^G_{m,\ell}(i)\, \widehat{\mathbf{B}}_n^\top \mathbf{x}(i)\mathbf{x}(i)^\top \widehat{\mathbf{B}}_n\right)^{-1} = \mathbf{P}_n^{-1} - \left(\mathbf{P}_n + \mathbf{Q}_n\right)^{-1} \mathbf{Q}_n \mathbf{P}_n^{-1}$$

Hence, for any task $m \in [M]$ and $\mathbf{x}_j \in \mathbb{R}^d$, we have

$$\begin{aligned}
\left\|\widehat{\mathbf{B}}_n^\top \mathbf{x}_j\right\|^2_{\left(\sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{b}^G_{m,\ell}(i)\widehat{\mathbf{B}}_n^\top \mathbf{x}(i)\mathbf{x}(i)^\top \widehat{\mathbf{B}}_n\right)^{-1}} &= \left(\widehat{\mathbf{B}}_n^\top \mathbf{x}_j\right)^\top \left(\sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{b}^G_{m,\ell}(i)\, \widehat{\mathbf{B}}_n^\top \mathbf{x}(i)\mathbf{x}(i)^\top \widehat{\mathbf{B}}_n\right)^{-1} \widehat{\mathbf{B}}_n^\top \mathbf{x}_j \\
&= \underbrace{\left(\widehat{\mathbf{B}}_n^\top \mathbf{x}_j\right)^\top \mathbf{P}_n^{-1} \widehat{\mathbf{B}}_n^\top \mathbf{x}_j}_{\text{Term 1}} - \underbrace{\left(\widehat{\mathbf{B}}_n^\top \mathbf{x}_j\right)^\top \left(\mathbf{P}_n + \mathbf{Q}_n\right)^{-1} \mathbf{Q}_n \mathbf{P}_n^{-1} \widehat{\mathbf{B}}_n^\top \mathbf{x}_j}_{\text{Term 2}}.
\end{aligned}$$

From Lemma A.10, and (17) we have

$$\left\|\widehat{\mathbf{B}}_n^\top \mathbf{B}^\perp\right\| \le \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}$$

Now we can decompose the term 1 into the following 4 terms

$$\begin{aligned}
\text{Term } 1 &= \left(\widehat{\mathbf{B}}_n^\top \mathbf{x}_j\right)^\top \mathbf{P}_n^{-1} \widehat{\mathbf{B}}_n^\top \mathbf{x}_j \\
&= \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}_j + \widehat{\mathbf{B}}_n^\top \mathbf{B}_\perp \mathbf{B}_\perp^\top \mathbf{x}_j\right)^\top \mathbf{P}_n^{-1} \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}_j + \widehat{\mathbf{B}}_n^\top \mathbf{B}_\perp \mathbf{B}_\perp^\top \mathbf{x}_j\right) \\
&= \underbrace{\left(\widehat{\mathbf{B}}_n^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}_j\right)^\top \mathbf{P}_n^{-1} \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}_j\right)}_{\text{Term 1-1}} + \underbrace{\left(\widehat{\mathbf{B}}_n^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}_j\right)^\top \mathbf{P}_n^{-1} \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}_\perp \mathbf{B}_\perp^\top \mathbf{x}_j\right)}_{\text{Term 1-3}} \\
&\quad + \underbrace{\left(\widehat{\mathbf{B}}_n^\top \mathbf{B}_\perp \mathbf{B}_\perp^\top \mathbf{x}_j\right)^\top \mathbf{P}_n^{-1} \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}_j\right)}_{\text{Term 1-2}} + \underbrace{\left(\widehat{\mathbf{B}}_n^\top \mathbf{B}_\perp \mathbf{B}_\perp^\top \mathbf{x}_j\right)^\top \mathbf{P}_n^{-1} \left(\widehat{\mathbf{B}}_n^\top \mathbf{B}_\perp \mathbf{B}_\perp^\top \mathbf{x}_j\right)}_{\text{Term 1-4}}.
\end{aligned}$$

It follows using the steps similar to Lemma C.10 of (Du et al., 2023) and combining with our Lemma A.10, and (17) we have that

$$\text{Term } 1-1 = \left\|\widehat{\mathbf{B}}_n^\top \mathbf{x}_j\right\|^2_{\left(\sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{b}^G_{m,\ell}(i)\widehat{\mathbf{B}}_n^\top \mathbf{x}(i)\mathbf{x}(i)^\top \widehat{\mathbf{B}}_n\right)^{-1}}, \qquad \text{Term } 1-2 \le c_2 \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}$$

$$\text{Term } 1-3 \le c_3 \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}, \qquad \text{Term } 1-4 \le c_4 \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}$$

Combining the 4 terms above we get the upper bound to term 1 as follows

$$\text{Term } 1 = \left(\widehat{\mathbf{B}}_n^\top \mathbf{x}_j\right)^\top \mathbf{P}_n^{-1}\widehat{\mathbf{B}}_n^\top \mathbf{x}_j \leq \left\|\mathbf{B}^\top \mathbf{x}_j\right\|^2_{\left(\sum_{i\in\mathcal{G}_{m,\ell-1}} \mathbf{B}^\top \mathbf{x}(i)\mathbf{x}(i)^\top \mathbf{B}\right)^{-1}} + \frac{cL_x^4}{k\omega^2}\exp(-Mn).$$

for some constant $c > 0$. Similarly, we can show that

$$\text{Term } 2 = \left(\widehat{\mathbf{B}}_n^\top \mathbf{x}_j\right)^\top \left(\mathbf{P}_n + \mathbf{Q}_n\right)^{-1}\mathbf{Q}_n\mathbf{P}_n^{-1}\widehat{\mathbf{B}}_n^\top \mathbf{x}_j \leq \frac{c' L_x^4}{k\omega^2}\exp(-Mn)$$

for some constant $c' > 0$. Combining everything we have that

$$\|\widetilde{\mathbf{g}}_{m,\ell}(j)\|^2_{\mathbf{\Sigma}^{-1}_{m,\ell}} \leq \|\mathbf{g}(j)\|^2_{\mathbf{\Sigma}^{-1}_{m,\ell}} + \frac{cL_x^4}{k\omega^2}\exp(-Mn)$$

for some constant $c > 0$. The claim of the lemma follows. □

Lemma A.13. *Let $\mathcal{F}_n$ hold. Define $\widetilde{\Delta}_{m,i} = \widetilde{\mathbf{g}}_m(i)^\top\widehat{\mathbf{w}}_m - \widetilde{\mathbf{g}}_m(i_m^*)^\top\widehat{\mathbf{w}}_m$ and $\Delta_{m,i} = \mathbf{g}_m(i)^\top\mathbf{w}_m - \mathbf{g}_m(i_m^*)^\top\mathbf{w}_m$. Then the estimation error in the second stage is given by*

$$|\widetilde{\Delta}_{m,i} - \Delta_{m,i}| \leq 6kL_x \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}.$$

Further for $Mn > \lceil \frac{L_x^4 k^2 d^2 c'(\rho^E)^2\log^2(2d)}{\omega^2\Delta^2}\rceil$ we have that

$$|\widetilde{\Delta}_{m,i} - \Delta_{m,i}| \leq \frac{\Delta_{m,i}}{2}.$$

Proof. Combining our Lemma A.10, and (17) we can bound the estimation error for any pair of $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ for a task m as follows:

$$\left|(\mathbf{x}-\mathbf{x}')^\top\widehat{\boldsymbol{\theta}}_{m,n} - (\mathbf{x}-\mathbf{x}')^\top\boldsymbol{\theta}_{m,*}\right| \leq 2k\cdot L_xL_w\left\|\widehat{\mathbf{B}}_{n,\perp}^\top\mathbf{B}\right\| + \frac{\sqrt{\rho_m^G\cdot 2\log\left(\frac{4n^2M}{\delta}\right)}}{\sqrt{n}} + 2L_xL_w\left\|\widehat{\mathbf{B}}_{n,\perp}^\top\mathbf{B}\right\|$$

Setting $L_w = 1$, $\rho_m^G = k$ and $\log\left(\frac{4n^2M}{\delta}\right) = n$ and as the event $\mathcal{F}_n$ holds, we get that

$$\left|(\mathbf{x}-\mathbf{x}')^\top\widehat{\boldsymbol{\theta}}_{m,n} - (\mathbf{x}-\mathbf{x}')^\top\boldsymbol{\theta}_{m,*}\right| \leq 6kL_x\min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}$$

This implies that

$$|\widetilde{\Delta}_{m,i} - \Delta_{m,i}| \leq 6kL_x \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}$$

Now for $n > \frac{L_x^4k^2d^2c'(\rho^E)^2\log^2(2d)}{\omega^2M\Delta^2}$ we can show that

$$\begin{aligned}
6kL_x \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\} &= 6kL_x \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}\\
&\leq 6kL_x \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}(2d)^{-\frac{L_x^4k^2d^2c'(\rho^E)^2}{\omega^2\Delta^2}}\right\}\\
&\overset{(a)}{\leq} \frac{\Delta}{2} \overset{(b)}{\leq} \frac{\Delta_{m,i}}{2}
\end{aligned}$$

where, (a) holds as for any $\Delta > 0, d, k > 1, \omega > 0$ the following holds

$$\log(\frac{\Delta}{12}) + \log(\frac{L_x}{k\omega}) > -\frac{L_x^4k^2d^2c'(\rho^E)^2}{\omega^2\Delta^2}\log(2d).$$

The (b) holds as $\Delta_{m,i} \geq \Delta$. The claim of the lemma follows. □

Lemma A.14. *With parameter n_m defined in (4), Algorithm 1 terminates in phase $\lceil \log_2 k \rceil$ with no more than a total of n arm pulls.*

Proof. Proof. When $k = 2$, Algorithm 1 terminates in one phase. When $k > 2$, by the property of ceiling function, we have $\frac{1}{2} < \frac{k}{2^{\lceil \log_2 k \rceil}} \leq 1$. Thus, the number of arms in the active set for each task m is $\mathcal{G}_{m,\lceil \log_2 k \rceil - 1}$ is $\left\lceil \frac{k}{2^{\lceil \log_2 k \rceil - 1}} \right\rceil = 2$, in phase $\lceil \log_2 k \rceil$.

Now we bound the number of arm pulls. For any phase ℓ, $\left|\operatorname{Supp}\left(\mathbf{b}^G_{\ell,m}\right)\right|$ is always bounded by the cardinality of the active set $\mathcal{G}_{m,\ell-1}$. In particular, for the first phase, according to Lemma A.1, there exists a G-optimal design $\mathbf{b}^G_{m,\ell}$ with $\left|\operatorname{Supp}\left(\mathbf{b}^G_{m,\ell}\right)\right| \leq k(k+1)/2$. Altogether, we have

$$\left|\operatorname{Supp}\left(\mathbf{b}^G_{m,\ell}\right)\right| \leq \begin{cases} \min\left(A, \frac{k(k+1)}{2}\right) & \text{when } \ell = 1 \\ \left\lceil \frac{k}{2^{\ell-1}} \right\rceil & \text{when } \ell > 1 \end{cases}. \tag{18}$$

Then the number of total arm pulls for each task m is bounded as

$$\begin{aligned} \sum_{\ell=1}^{\lceil \log_2 k \rceil} N_{m,\ell} &= \sum_{\ell=1}^{\lceil \log_2 k \rceil} \sum_{i \in \mathcal{G}_{m,\ell}} N_{m,\ell}(i) \overset{(a)}{=} \sum_{\ell=1}^{\lceil \log_2 k \rceil} \sum_{i \in \mathcal{G}_{m,\ell}} \left\lceil \mathbf{b}^G_{m,\ell}\left(\widetilde{\mathbf{g}}_{m,\ell}(i)\right) \cdot n_m \right\rceil \\ &\overset{(b)}{\leq} \sum_{\ell=1}^{\lceil \log_2 k \rceil} \left(\left|\operatorname{Supp}\left(\mathbf{b}^G_{m,\ell}\right)\right| + \sum_{i \in \mathcal{G}_{m,\ell}} \mathbf{b}^G_{m,\ell}\left(\widetilde{\mathbf{g}}_{m,\ell}(i)\right) \cdot n_m \right) \\ &\overset{(c)}{\leq} \min\left(A, \frac{k(k+1)}{2}\right) + \sum_{\ell=2}^{\lceil \log_2 k \rceil} \left\lceil \frac{k}{2^{\ell-1}} \right\rceil + \lceil \log_2 k \rceil \cdot n_m \\ &\overset{(d)}{=} \frac{n}{2M} \end{aligned}$$

where, (a) follows as the allocation to each arm in task m is given by atmost $\left\lceil \mathbf{b}^G_{m,\ell}\left(\widetilde{\mathbf{g}}_{m,\ell}(i)\right) \cdot n_m \right\rceil$, (b) follows by using the two cases in (18), (c) follows by using Lemma A.1, and finally (d) follows plugging the value of n_m from (4). Therefore summing over all tasks $m \in [M]$ we get that the second stage is at most

$$\sum_{m=1}^{M} \tau^E_m = \sum_{m=1}^{M} \frac{n}{2M} = \frac{n}{2}.$$

For the first stage, for each phase $m \in [M]$ the algorithm uses atmost $\frac{n}{2}$ samples for the E-optimal design. Summing over all phases and stages we get that the total budget is used atmost n. ☐

Lemma A.15. *For an arbitrary constant Δ and $\mathbf{x} \in \mathbb{R}^d$ we can show that*

$$\mathbb{P}\left(\mathbf{x}^\top\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*\right) > \Delta\right) \leq \exp\left(-\frac{\Delta^2}{2\|\mathbf{x}\|^2_{\Sigma_n^{-1}}}\right)$$

where, $\boldsymbol{\Sigma}_n = \sum_{i=1}^{n} \sum_{j=1}^{K} \mathbf{x}_{i,j}(\mathbf{x}_{i,j})^\top$.

Proof. We follow the proof technique of section 2.2 of Jamieson & Jain (2022). Under the sub-Gaussian noise assumption, we can show that for any vector $\mathbf{x} \in \mathbb{R}^d$ the following holds

$$\mathbf{x}^\top\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*\right) = \underbrace{\mathbf{x}^\top\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top}_{\mathbf{w}}\eta = \mathbf{w}^\top\eta.$$

Then for an arbitrary constant Δ and $\mathbf{x} \in \mathbb{R}^d$, we can show that

$$\begin{aligned}
\mathbb{P}\left(\mathbf{x}^\top\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*\right) > \Delta\right) &= \mathbb{P}\left(\mathbf{w}^\top \eta > \Delta\right) \\
&\overset{(a)}{\leq} \exp(-\lambda\Delta)\mathbb{E}\left[\exp\left(\lambda \mathbf{w}^\top \eta\right)\right], \quad \text{let } \lambda > 0 \\
&= \exp(-\lambda\Delta)\mathbb{E}\left[\exp\left(\lambda \sum_{s=1}^{t} \mathbf{w}_s \eta_s\right)\right] \\
&\overset{(b)}{=} \exp(-\lambda\Delta) \prod_{s=1}^{t} \mathbb{E}\left[\exp\left(\lambda \mathbf{w}_s \eta_s\right)\right] \\
&\overset{(c)}{\leq} \exp(-\lambda\Delta) \prod_{s=1}^{t} \exp\left(\lambda^2 \mathbf{w}_s^2/2\right) \\
&= \exp(-\lambda\Delta)\exp\left(\frac{\lambda^2}{2}\|\mathbf{w}\|_2^2\right) \\
&\overset{(d)}{\leq} \exp\left(-\frac{\Delta^2}{2\|\mathbf{w}\|_2^2}\right) \\
&\overset{(e)}{=} \exp\left(-\frac{\Delta^2}{2\mathbf{x}^\top\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{x}}\right) = \exp\left(-\frac{\Delta^2}{2\|\mathbf{x}\|^2_{\Sigma_n^{-1}}}\right)
\end{aligned}$$

where, (a) follows from Chernoff Bound, (b) follows from independence of, (c) follows sub-Gaussian assumption, (d) follows by setting $\lambda = \frac{\Delta}{\|\mathbf{w}\|_2^2}$, and (e) follows from the equality

$$\|\mathbf{w}\|_2^2 = \mathbf{x}^\top\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{x} = \mathbf{x}^\top\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{x}.$$

The claim of the lemma follows. □

The following lemma bounds the probability that a certain arm has its estimate of the expected reward larger than that of the best arm in a single phase ℓ.

Lemma A.16. *Suppose $\mathcal{F}_n$ holds, and $Mn > \lceil \frac{L_x^4 k^2 d^2 c'(\rho^E)^2 \log^2(2d)}{\omega^2\Delta^2} \rceil$. For a fixed realization of $\widehat{\mathcal{X}}_{m,\ell-1}$ satisfying $i_m^* \in \widehat{\mathcal{X}}_{m,\ell-1}$, for any arm $i \in \widehat{\mathcal{X}}_{m,\ell-1}$,*

$$\mathbb{P}\left(\widehat{\mu}_{m,\ell}(i_m^*) < \widehat{\mu}_{m,\ell}(i)\right) \leq \exp\left(\frac{cL_x^4}{k\omega^2}\right)\exp\left(-\frac{n_m \Delta_{m,i}^2}{32\left\lceil \frac{k}{2^{\ell-1}} \right\rceil}\right)$$

Proof. Let $\boldsymbol{\theta}^*_{m,\ell}$ denote the corresponding unknown parameter vector for the task m and phase ℓ for the dimensionality-reduced arm vectors $\{\widetilde{\mathbf{g}}_{m,\ell}(i) : i \in \mathcal{G}_{m,\ell-1}\}$. Also, we set

$$\boldsymbol{\Sigma}_{m,\ell} = \sum_{i \in \mathcal{G}_{m,\ell-1}} \mathbf{b}_{m,\ell}(i)\, \widetilde{\mathbf{g}}_{m,\ell}(i)\widetilde{\mathbf{g}}_{m,\ell}(i)^\top.$$

Then we can show using the identities that $\widehat{\boldsymbol{\theta}}_{m,\ell} = \widehat{\mathbf{B}}_n \widehat{\mathbf{w}}_{m,\ell}$, $\boldsymbol{\theta}_m^* = \mathbf{B}\mathbf{w}_m$ and for $n > \frac{L_x^4 k^2 d^2 c'(\rho^E)^2 \log^2(2d)}{\omega^2 M \Delta^2}$ that

$$\|\widetilde{\mathbf{g}}_{m,\ell}(j)\|^2_{\boldsymbol{\Sigma}^{-1}_{m,\ell}} \leq \|\mathbf{g}(j)\|^2_{\boldsymbol{\Sigma}^{-1}_{m,\ell}} + \frac{cL_x^4}{k\omega^2}\exp(-Mn) \tag{19}$$

the following:

$$
\begin{aligned}
&\mathbb{P}\left(\widehat{\mu}_{m,\ell}(i_m^*) < \widehat{\mu}_{m,\ell}(i)\right) = \mathbb{P}\left((\widehat{\boldsymbol{\theta}}^*_{m,\ell})^\top \mathbf{x}_{m,\ell}(i_m^*) < (\widehat{\boldsymbol{\theta}}^*_{m,\ell})^\top \mathbf{x}_{m,\ell}(i)\right)\right) \\
=&\mathbb{P}\left((\widehat{\mathbf{B}}_n \widehat{\mathbf{w}}_{m,\ell})^\top \mathbf{x}_{m,\ell}(i_m^*) < (\widehat{\mathbf{B}}_n \widehat{\mathbf{w}}_{m,\ell})^\top \mathbf{x}_{m,\ell}(i)\right) \\
=&\mathbb{P}\left(\widehat{\mathbf{w}}_{m,\ell}^\top \widetilde{\mathbf{g}}_{m,\ell}(i_m^*) < \widehat{\mathbf{w}}_{m,\ell}^\top \widetilde{\mathbf{g}}_{m,\ell}(i)\right) \\
=&\mathbb{P}\left((\widehat{\mathbf{w}}_{m,\ell})^\top \widetilde{\mathbf{g}}_{m,\ell}(i_m^*) - (\widehat{\mathbf{w}}_{m,\ell})^\top \widetilde{\mathbf{g}}_{m,\ell}(i) - \widetilde{\Delta}_{m,i} < -\widetilde{\Delta}_{m,i}\right) \\
\overset{(a)}{\leq}&\mathbb{P}\left((\widehat{\mathbf{w}}_{m,\ell})^\top \widetilde{\mathbf{g}}_{m,\ell}(i_m^*) - (\widehat{\mathbf{w}}_{m,\ell})^\top \widetilde{\mathbf{g}}_{m,\ell}(i) - \left((\mathbf{w}_m)^\top \left(\widetilde{\mathbf{g}}_{m,\ell}(i_m^*) - \widetilde{\mathbf{g}}_{m,\ell}(i)\right)\right) < -\Delta_{m,i} + \frac{\Delta_{m,i}}{2}\right) \\
=&\mathbb{P}\left(\langle \widehat{\mathbf{w}}_{m,\ell} - \mathbf{w}_m, \widetilde{\mathbf{g}}_{m,\ell}(i_m^*) - \widetilde{\mathbf{g}}_{m,\ell}(i)\rangle < -\frac{3\Delta_{m,i}}{2}\right) \\
\overset{(b)}{\leq}& \exp\left(-\frac{\frac{9\Delta_{m,i}^2}{4}}{2\left\|\widetilde{\mathbf{g}}_{m,\ell}(i_m^*) - \widetilde{\mathbf{g}}_{m,\ell}(i)\right\|^2_{\boldsymbol{\Sigma}^{-1}_{m,\ell}}}\right) \\
\overset{(c)}{\leq}& \exp\left(-\frac{\frac{9\Delta_{m,i}^2}{4}}{8\max_{i\in\mathcal{G}_{m,\ell}}\left\|\widetilde{\mathbf{g}}_{m,\ell}(i)\right\|^2_{\boldsymbol{\Sigma}^{-1}_{m,\ell}}}\right) \\
\overset{(d)}{\leq}& \exp\left(-\frac{\frac{9\Delta_{m,i}^2}{4}\cdot n_m}{8\max_{i\in\mathcal{G}_{m,\ell}}\left\|\widetilde{\mathbf{g}}_{m,\ell}(i)\right\|^2_{\boldsymbol{\Sigma}^{-1}_{m,\ell}}}\right) \\
\overset{(e)}{\leq}& \exp\left(-\frac{\frac{9\Delta_{m,i}^2}{4}\cdot n_m}{8\max_{i\in\mathcal{G}_{m,\ell}}\left\|\widetilde{\mathbf{g}}_{m,\ell}(i)\right\|^2_{\boldsymbol{\Sigma}^{-1}_{m,\ell}} + \frac{cL_x^4}{k\omega^2}\exp(-Mn)}\right) \\
\overset{(f)}{\leq}& \exp\left(-\frac{n_m\frac{9\Delta_{m,i}^2}{4}}{8d_k + \frac{cL_x^4}{k\omega^2}}\right) \overset{(g)}{\leq} \exp\left(\frac{cL_x^4}{k\omega^2}\right)\exp\left(-\frac{n_m\frac{9\Delta_{m,i}^2}{4}}{8\left\lceil\frac{k}{2^{k-1}}\right\rceil}\right) \leq \exp\left(\frac{cL_x^4}{k\omega^2}\right)\exp\left(-\frac{n_m\Delta_{m,i}^2}{32\left\lceil\frac{k}{2^{k-1}}\right\rceil}\right).
\end{aligned}
$$

where, (a) follows from Lemma A.13 and $n > \frac{L_x^4 d^2 c'(\rho^E)^2\log^2(2d)}{\omega^2 M\Delta^2}$ (b) follows from Lemma A.15, (c) follows from triangle inequality. The inequality in (d) follows from

$$
\begin{aligned}
\left\|\widetilde{\mathbf{g}}_{m,\ell}(i)\right\|^2_{\boldsymbol{\Sigma}^{-1}_{m,\ell}} &= \widetilde{\mathbf{g}}_{m,\ell}(i)^\top \boldsymbol{\Sigma}^{-1}_{m,\ell}\widetilde{\mathbf{g}}_{m,\ell}(i) \\
&= \widetilde{\mathbf{g}}_{m,\ell}(i)^\top \left(\sum_{j\in\mathcal{G}_{m,\ell-1}} T_{m,\ell}(j)\widetilde{\mathbf{g}}_{m,\ell}(j)\widetilde{\mathbf{g}}_{m,\ell}(j)^\top\right)^{-1}\widetilde{\mathbf{g}}_{m,\ell}(i) \\
&\leq \widetilde{\mathbf{g}}_{m,\ell}(i)^\top \left(\sum_{j\in\mathcal{G}_{m,\ell-1}} n_m \mathbf{b}_{m,\ell}(j)\widetilde{\mathbf{g}}_{m,\ell}(j)\widetilde{\mathbf{g}}_{m,\ell}(j)^\top\right)^{-1}\widetilde{\mathbf{g}}_{m,\ell}(i) \\
&= \frac{1}{n_m}\widetilde{\mathbf{g}}_{m,\ell}(i)^\top \left(\sum_{j\in\mathcal{G}_{m,\ell-1}} \mathbf{b}_{m,\ell}(j)\widetilde{\mathbf{g}}_{m,\ell}(j)\widetilde{\mathbf{g}}_{m,\ell}(j)^\top\right)^{-1}\widetilde{\mathbf{g}}_{m,\ell}(i) \\
&= \frac{1}{n_m}\widetilde{\mathbf{g}}_{m,\ell}(i)^\top \boldsymbol{\Sigma}^{-1}_{m,\ell}\widetilde{\mathbf{g}}_{m,\ell}(i) \\
&= \frac{1}{n_m}\left\|\widetilde{\mathbf{g}}_{m,\ell}(i)\right\|^2_{\boldsymbol{\Sigma}^{-1}_{m,\ell}}.
\end{aligned}
$$

The equality in (f) follows from Lemma A.1 and the property of G-optimal design. Also we drop $\exp(-Mn) < 1$. The inequality in (g) follows from the fact that the dimension of the space spanned by the corresponding arm vectors of the active arm set $\mathcal{G}_{m,\ell-1}$ is not larger than the cardinality of

$\mathcal{G}_{m,\ell-1}$. Also note that the additional term $\exp\left(\frac{cL_x^4}{k\omega^2}\right)$ which results from latent feature estimation error. The claim of the lemma follows. $\square$

Lemma A.17. *Assume that the best arm i_m^* is not eliminated before phase ℓ, i.e., $i_m^* \in \mathcal{G}_{m,\ell-1}$. Then the probability that the best arm is eliminated in phase ℓ is bounded as*

$$\mathbb{P}\left(i_m^* \notin \mathcal{G}_{m,\ell} \mid i_m^* \in \mathcal{G}_{m,\ell-1}\right) \leq \begin{cases} \frac{4A}{k} \exp\left(\frac{cL_x^4}{k\omega^2}\right) \exp\left(-\frac{n_m \Delta_{m,i_\ell}^2}{32 i_{m,\ell}}\right) & \text{when } \ell = 1 \\ 3\exp\left(\frac{cL_x^4}{k\omega^2}\right) \exp\left(-\frac{n_m \Delta_{m,i_\ell}^2}{32 i_{m,\ell}}\right) & \text{when } \ell > 1 \end{cases}$$

where $i_{m,\ell} = \left\lceil \frac{k}{2^{\ell+1}} \right\rceil + 1$.

Proof. First, as Lemma A.16, we conditioned on the specific realization of $\mathcal{G}_{m,\ell-1}$ such that $1 \in \mathcal{G}_{m,\ell-1}$. Define $\mathcal{H}_{m,\ell}$ as the set of arms in $\mathcal{G}_{m,\ell-1}$ excluding the best arm and $\left\lceil \frac{k}{2^{\ell+1}} \right\rceil - 1$ suboptimal arms with the largest expected rewards. Therefore, we have $|\mathcal{H}_{m,\ell}| = |\mathcal{G}_{m,\ell-1}| - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil$ and $\min_{i\in\mathcal{H}_{m,\ell}} \Delta_{m,i} \geq \Delta_{m,\left\lceil \frac{k}{2^{\ell+1}} \right\rceil+1}$.

If the best arm for task m, i_m^* is eliminated in phase ℓ, then at least $\left\lceil \frac{k}{2^r} \right\rceil - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil + 1$ arms of $\mathcal{H}_{m,\ell}$ have their estimates of the expected rewards larger than that of the best arm.

Let $N_{m,\ell}$ denote the number of arms in $\mathcal{H}_{m,\ell}$ whose estimates of the expected rewards are larger than that of the best arm. By Lemma A.16, we have

$$\begin{aligned}
\mathbb{E}\left[N_{m,\ell}\right] = \sum_{i\in\mathcal{H}_{m,\ell}} \mathbb{P}\left(\widehat{\mu}_{m,\ell}(i_m^*) < \widehat{\mu}_{m,\ell}(i)\right) &\leq \exp\left(\frac{cL_x^4}{k\omega^2}\right) \sum_{i\in\mathcal{H}_{m,\ell}} \exp\left(-\frac{n_m \Delta_{m,i}^2}{32\left\lceil \frac{k}{2^{r-1}} \right\rceil}\right) \\
&\leq \exp\left(\frac{cL_x^4}{k\omega^2}\right) |\mathcal{H}_{m,\ell}| \max_{i\in\mathcal{H}_{m,\ell}} \exp\left(-\frac{n_m \Delta_{m,i}^2}{32\left\lceil \frac{k}{2^{\ell-1}} \right\rceil}\right) \\
&\leq \exp\left(\frac{cL_x^4}{k\omega^2}\right) \left(|\mathcal{G}_{m,\ell-1}| - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil\right) \exp\left(-\frac{n_m \Delta^2_{m,\left\lceil \frac{k}{2^{\ell+1}} \right\rceil+1}}{32\left\lceil \frac{k}{2^{\ell-1}} \right\rceil}\right) \\
&\leq \exp\left(\frac{cL_x^4}{k\omega^2}\right) \left(|\mathcal{G}_{m,\ell-1}| - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil\right) \exp\left(-\frac{n_m \Delta^2_{m,\left\lceil \frac{k}{2^{\ell+1}} \right\rceil+1}}{32\left(\left\lceil \frac{k}{2^{\ell-1}} \right\rceil+1\right)}\right).
\end{aligned}$$

Then, together with Markov's inequality, we obtain

$$\begin{aligned}
\mathbb{P}\left(i_m^* \notin \mathcal{G}_{m,\ell}\right) &\leq \mathbb{P}\left(T_{m,\ell} \geq \left\lceil \frac{k}{2^\ell} \right\rceil - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil + 1\right) \\
&\leq \frac{\mathbb{E}\left[T_{m,\ell}\right]}{\left\lceil \frac{k}{2^\ell} \right\rceil - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil + 1} \\
&\leq \frac{|\mathcal{G}_{m,\ell-1}| - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil}{\left\lceil \frac{k}{2^\ell} \right\rceil - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil + 1} \exp\left(-\frac{n_m \Delta^2_{m,\left\lceil \frac{k}{2^{\ell+1}} \right\rceil+1}}{32\left(\left\lceil \frac{k}{2^{\ell+1}} \right\rceil+1\right)}\right).
\end{aligned}$$

When $\ell = 1$, we have $|\mathcal{G}_{m,\ell-1}| = A$. Thus,

$$\begin{aligned}
\frac{|\mathcal{G}_{m,\ell-1}| - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil}{\left\lceil \frac{k}{2^r} \right\rceil - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil + 1} &= \frac{A - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil}{\left\lceil \frac{k}{2^r} \right\rceil - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil + 1} \\
&\leq \frac{A}{\frac{k}{2} - \frac{k}{2^2}} \\
&= \frac{4A}{k}.
\end{aligned}$$

When $\ell > 1$, we have $|\mathcal{G}_{m,\ell-1}| = \left\lceil \frac{k}{2^{\ell-1}} \right\rceil$. Thus,

$$\begin{aligned}
\frac{|\mathcal{G}_{m,\ell-1}| - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil}{\left\lceil \frac{k}{2^{r}} \right\rceil - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil + 1} &= \frac{\left\lceil \frac{k}{2^{\ell-1}} \right\rceil - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil}{\left\lceil \frac{k}{2^{r}} \right\rceil - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil + 1} \\
&\leq \frac{\frac{k}{2^{\ell-1}} + 1 - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil}{\frac{k}{2^{\ell}} - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil + 1} \\
&\leq \frac{3 \cdot \frac{k}{2^{\ell+1}} + \frac{k}{2^{\ell+1}} + 1 - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil}{\frac{k}{2^{\ell+1}} + \frac{k}{2^{\ell+1}} + 1 - \left\lceil \frac{k}{2^{\ell+1}} \right\rceil} \\
&\leq 3
\end{aligned}$$

where the last inequality results from the fact that for any $x, y > 0$, $\frac{3x+y}{x+y} \leq 3$. Therefore, for this specific realization of $\mathcal{G}_{m,\ell-1}$ satisfying $1 \in \mathcal{G}_{m,\ell-1}$,

$$\mathbb{P}\left(i_m^* \notin \mathcal{G}_{m,\ell}\right) \leq \begin{cases} \frac{4A}{k} \exp\left(\frac{cL_x^4}{k\omega^2}\right) \exp\left(-\frac{n_m \Delta_{m,i_\ell}^2}{32 i_{m,\ell}}\right) & \text{when } \ell = 1 \\ 3 \exp\left(\frac{cL_x^4}{k\omega^2}\right) \exp\left(-\frac{n_m \Delta_{m,i_\ell}^2}{32 i_{m,\ell}}\right) & \text{when } \ell > 1 \end{cases}$$

where $i_{m,\ell} = \left\lceil \frac{k}{2^{\ell+1}} \right\rceil + 1$. Finally, by the law of total probability, the error probability of phase ℓ conditioned on $i_m^* \in \mathcal{G}_{m,\ell-1}$ can be bounded as

$$\mathbb{P}\left[i_m^* \notin \mathcal{G}_{m,\ell} \mid i_m^* \in \mathcal{G}_{m,\ell-1}\right] \leq \begin{cases} \frac{4A}{k} \exp\left(\frac{cL_x^4}{k\omega^2}\right) \exp\left(-\frac{n_m \Delta_{m,i_\ell}^2}{32 i_{m,\ell}}\right) & \text{when } \ell = 1 \\ 3 \exp\left(\frac{cL_x^4}{k\omega^2}\right) \exp\left(-\frac{n_m \Delta_{m,i_\ell}^2}{32 i_{m,\ell}}\right) & \text{when } \ell > 1. \end{cases}$$

The claim of the lemma follows. □

Now we prove the main theorem for linear MTRL FB-DOE.

Theorem 2. *Define* $\Delta = \min_m \min_{i \in \mathcal{X}} \Delta_{m,i}$, $Mn \geq \left\lceil \frac{L_x^4 k^2 d^2 c' (\rho^E)^2 \log^2(2d)}{\omega^2 \Delta^2} \right\rceil$ *and* $\ell > 1$. *The total probability of error of the algorithm for* $\ell > 1$ *is given by*

$$8 \exp\left(\frac{-Mn}{\log_2 d}\right) + M\left(3 \log_2 k\right) \exp\left(-\frac{n}{64 H_{2,\,\textit{lin}}} + \frac{cL_x^4}{k\omega^2}\right)$$

and $\|\mathbf{x}\| \leq L_x$, $\omega > 0$ *is defined in Assumption 2.2 and* $H_{2,\,\textit{lin}} = \max_{m \in [M]} \max_{2 \leq i \leq k} \frac{i}{\Delta_{m,i}^2}$ *is the linear MTRL hardness parameter.*

Proof. **Stage 1:** Using Lemma A.6 we can show that the probability of error in the first stage is bounded by

$$8d \exp\left(-Mn\right) \overset{(a)}{\leq} 8 \exp\left(\frac{-Mn}{\log_2 d}\right).$$

where, (a) follows as

$$\exp\left(-Mn + \log d\right) \leq \exp\left(-\frac{Mn}{\log_2 d}\right).$$

The above inequality holds true because

$$\begin{aligned}
&- Mn + \log d \leq \left(-\frac{Mn}{\log_2 d}\right) \\
&\implies \left(\log_2 d\right) \log d - Mn\left(\log_2 d\right) \leq -Mn \\
&\implies \left(\log_2 d\right) \log d \leq Mn\left(\log_2 d\right) - Mn \\
&\overset{(b)}{\implies} \left(\log_2 d\right) \log d \leq Mn\left(\log_2 d - 1\right)
\end{aligned}$$

We can now substitute the lower bound value of $Mn \geq \left\lceil \frac{L_x^4 k^2 d^2 c' \left(\rho^5\right)^2 \log^2(2d)}{\omega^2 \Delta^2} \right\rceil$ and see that (b) holds true, and $d \gg k$ and $k \geq 2$. So we have $\log_2 d \gg 1$ and so $(\log_2 d - 1)$ is a positive quantity. Also we have shown in Lemma A.10 that if the good event $\mathcal{F}_n$ holds, then we get a valid G-optimal design and $\left\|(\widehat{\mathbf{B}}_n^{\perp})^\top \mathbf{B}\right\| \leq \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2} \exp\left(-Mn\right)\right\}$ for $Mn \geq \lceil \frac{L_x^4 k^2 d^2 c' (\rho^E)^2 \log^2(2d)}{\omega^2 \Delta^2} \rceil$.

Stage 2: By applying Lemma A.14 and Lemma A.17, we have

$$
\begin{aligned}
\mathbb{P}\left(\widehat{i^*}_m \neq i^*_m\right) &= \mathbb{P}\left[i^*_m \notin \mathcal{G}_{m,\lceil \log_2 k \rceil}\right] \\
&\leq \sum_{\ell=1}^{\lceil \log_2 k \rceil} \mathbb{P}\left[i^*_m \notin \mathcal{G}_{m,\ell} \mid i^*_m \in \mathcal{G}_{m,\ell-1}\right] \\
&\leq \exp\left(\frac{cL_x^4}{k\omega^2}\right) \sum_{\ell=2}^{\lceil \log_2 k \rceil} 3 \exp\left(-\frac{n_m \Delta^2_{m,i_\ell}}{32 i_{m,\ell}}\right) \\
&\leq \left(3\left(\lceil \log_2 k \rceil - 1\right)\right) \exp\left(\frac{cL_x^4}{k\omega^2}\right) \exp\left(-\frac{n_m}{32} \cdot \frac{1}{\max_{2\leq i\leq d} \frac{i}{\Delta_i^2}}\right) \\
&< \left(3 \log_2 k\right) \exp\left(\frac{cL_x^4}{k\omega^2}\right) \exp\left(-\frac{n_m}{32 H_{2,\text{ lin}}}\right)
\end{aligned}
$$

where $H_{2,\text{ lin}}$ is defined as

$$
H_{2,\text{ lin}} = \max_{m\in[M]} \max_{2\leq i\leq k} \frac{i}{\Delta^2_{m,i}}.
$$

Note that this is for a single task m. So the total probability of error in stage 2 is given by

$$
M\left(3 \log_2 k\right) \exp\left(\frac{cL_x^4}{k\omega^2}\right) \exp\left(-\frac{n_m}{32 H_{2,\text{ lin}}}\right)
$$

Combining both stage 1 and stage 2 and substituting the value of n_m we get that the total probability of error is given by

$$
8d \exp\left(-Mn\right) + M\left(3 \log_2 k\right) \exp\left(\frac{cL_x^4}{k\omega^2}\right) \exp\left(-\frac{n}{64 H_{2,\text{ lin}}}\right) \tag{20}
$$

The claim of the theorem follows. □

Remark A.18. **(Rounding Error)** Note that FB-DOE samples each arm $\lceil \tau_m^E \mathbf{b}_{\mathbf{x}}^E(i) \rceil$ in stage 1 and $\lceil \mathbf{b}^G_{m,\ell}(i) \cdot n_m(k) \rceil$ times in stage 2. However, this may lead to oversampling of an arm than what the design (G or E-optimal) is suggesting. However, we can match the number of allocations of an arm to the design using efficient Rounding Procedures (Pukelsheim, 2006; Fiez et al., 2019). This results in an estimation error of at most a multiplicative factor of (1_β), for some $\beta > 0$ (Lattimore & Szepesvári, 2020; Fiez et al., 2019; Du et al., 2023). For convenience and easier exposition of our result, we drop this factor of $(1 + \beta)$.

Remark A.19. **(Algorithmic Discussion)** Note that the allocation of $n/2$ total number of samples to each stage may seem arbitrary and one might be tempted to allocate total samples to the two stages more carefully. One such approach is shown in Chen et al. (2022) which studies the representation of learning in an active learning setting and minimizes the expected risk. However, we note that such an approach will only result in a linear scaling with $C'n$ for some $C' > 0$ while the scaling with the dimensions will remain unchanged which is the main theme of this paper.

Remark A.20. **(Discussion on Bound)** Observe that the probability of error depends on budget n, ambient dimension d, latent dimension k and linear hardness parameter $H_{2,\text{ lin}}$. The $H_{2,\text{ lin}}$ quantifies the difficulty of identifying the best arm in the linear bandit MTRL setting. In the single task setting, when $M = 1$, then the bounds scale with the ambient dimension d. Then the $H_{2,\text{ lin}} = \max_{2\leq i\leq d} \frac{i}{\Delta^2}$. This single task $H_{2,\text{ lin}}$ generalizes its stochastic bandit analogue $H_{2,\text{stoc}} = \max_{2\leq i\leq A} \frac{i}{\Delta_i^2}$ proposed by Audibert et al. (2010); Bubeck et al. (2009) for standard multi-armed bandits. Note that $H_{2,\text{ lin}}$ is never larger than $H_{2,\text{stoc}}$ since $H_{2,\text{ lin}}$ is a function of the first $d-1$ optimality gaps while $H_{2,\text{stoc}}$ considers all of the $d-1$ optimality gaps. In general, we have

$$H_{2,\text{ lin}} \leq H_{2,\text{stoc}} \leq \frac{A}{d} H_{2,\text{ lin}}$$

and both inequalities are essentially sharp, i.e., can be achieved by some linear bandit instances. This shows that the hardness in linear bandits due to their correlated structure should depend on d instead of A. Finally, note that when $M > 1$, it follows that $H_{2,\text{ lin}}$ should scale with the worst possible d gaps among all tasks.

Observe that the final probability of error in (20) consists of two terms. The first term is the probability of error in estimation of the feature extractor $\mathbf{B}$. The second term is the error in the estimation of the optimal arm in each task. Additionally, the factor $\exp\left(\frac{cL_x^4}{k\omega^2}\right)$ captures the error in estimating latent features. Also, note that from (20) we can show that

$$\begin{aligned}(3\log_2 k)\exp\left(\frac{cL_x^4}{k\omega^2}\right)\exp\left(-\frac{n}{64H_{2,\text{ lin}}}\right) &= \exp\left(-\frac{n}{64H_{2,\text{ lin}}} + \log(3\log_2 k) + \frac{cL_x^4}{k\omega^2}\right)\\ &\overset{(a)}{\leq} \exp\left(-\frac{n}{192H_{2,\text{ lin}}\log_2 k} + \frac{cL_x^4}{k\omega^2}\right)\end{aligned}$$

where, (a) follows as

$$\exp\left(-\frac{n}{64} + \log\log_2 k\right) \leq \exp\left(-\frac{n}{64} + \log_2 3k\right) \leq \exp\left(-\frac{n}{192\log_2 k}\right)$$

Then we introduce another novel lemma Lemma A.12 which shows using Lemma A.10 and (17) that the latent feature estimation is low. In Lemma A.13 we ensure that the estimation error with the latent parameter is low. This requires a different analysis than similar art in Du et al. (2023); Yang et al. (2020; 2022) as they only study fixed confidence or regret minimization setting. In the second stage, our technical novelty lies in controlling the probability of error for the noisy latent features in low dimensional multi-task linear bandits. This is shown in Lemma A.14, Lemma A.16, and Lemma A.17. Note that this approach differs from the existing art of fixed budget linear bandit settings (Katz-Samuels et al., 2020; Yang & Tan, 2021; Azizi et al., 2022) and significantly different than the fixed confidence linear bandit proofs in (Soare et al., 2014; Mason et al., 2021; Degenne & Koolen, 2019).

Remark A.21. **(Comparison with Peace, BayesGap, and GSE)** We now comment on the choice of OD-LinBAI in the second stage of FB-DOE as opposed to Peace (Katz-Samuels et al., 2020), BayesGap (Hoffman et al., 2014) or GSE (Azizi et al., 2022). In Yang & Tan (2021) they show that OD-LinBAI is minimax optimal in case of stochastic K-armed bandits, which is a special case of single task linear bandit setting. However, Yang & Tan (2021) also shows that Peace is not minimax optimal and suffers from an additional factor of $\log d$. This same argument also holds for FB-DOE. The BayesGap (Hoffman et al., 2014) algorithm works in the Bayesian linear bandit setting. It requires access to the problem-dependent parameter $H_1 = \sum_i \Delta_i^{-2}$ in a single task linear bandit setting. Note, that H_1 needs to be estimated using the true reward gap means, which is not practical. However, our algorithm FB-DOEdoes not require such access to the problem-dependent parameter H_1. Finally, we discuss the GSE algorithm (Azizi et al., 2022) which is also motivated by G-optimal design (Pukelsheim, 2006). Azizi et al. (2022) shows that GSE and OD-LinBAI outperform each other in some domains. In the case of single task linear bandits when $A < O(d^2)$ the OD-LinBAI has a lower probability of error, whereas in the case when $A = d^q$ for some $q > 2$, the GSE has a lower probability of error. The same argument also holds for FB-DOE. Nevertheless, our approach in stage 2 is quite general once the latent features have been estimated from stage 1 with exponentially decaying probability. After that, an algorithmic modification in stage 2 (similar to GSE) enables us to plug in the result of GSE to our bound. We leave this to future work.

A.4 Bi-Linear Bandit Fixed Budget Proofs

Stage 1 for FB-DOE

Define $\mathbf{W}_{\text{batch}}^{+} := \left(\mathbf{W}_{\text{batch}}^{\top} \mathbf{W}_{\text{batch}}\right)^{-1} \mathbf{W}_{\text{batch}}^{\top}$ where $\mathbf{W}_{\text{batch}}^{+} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{\tau_E^m}]^{\top}$ is constructed through the E-optimal design. Let $\mathbf{w} = \text{vec}(\mathbf{x}; \mathbf{z}) \in \mathbb{R}^{d_1 d_2}$. Also note that $\rho_1^E = \rho_2^E = \ldots = \rho_M^E = \rho^E$ as the action set $\mathcal{X}$, and $\mathcal{Z}$ are common across the tasks. Also rotate $\mathbf{\Theta}_{m,*} \in \mathbb{R}^{d_1 \times d_2}$ into the vector $\boldsymbol{\theta}_{m,*} \in \mathbb{R}^{d_1 d_2}$. Then recall that

$$\widehat{\boldsymbol{\theta}}_{t,m,j} = \mathbf{W}_{\text{batch}}^{+} r_{m,t},$$

where, $\widehat{\boldsymbol{\theta}}_{m,t} \in \mathbb{R}^{d_1 d_2}$. In stage 1 it builds the estimator $\widehat{\mathbf{Z}}_n$ as follows: The estimated parameter for task m at round t be denoted by $\widehat{\boldsymbol{\theta}}_{m,t} \in \mathbb{R}^{d_1 d_2}$ such that

$$\widehat{\boldsymbol{\theta}}_{m,t} = (\sum_{t=1}^{\tau_m^E} \mathbf{w}_{m,t} \mathbf{w}_{m,t}^{\top})^{-1} \sum_{t=1}^{\tau_m^E} \mathbf{w}_{m,t} r_{m,t}$$

Then calculate the estimate at round n as

$$\widehat{\mathbf{Z}}_n = \frac{3}{Mn} \sum_{m=1}^{M} \sum_{t=1}^{\tau_m^E} \widehat{\boldsymbol{\theta}}_{m,t} \widehat{\boldsymbol{\theta}}_{m,t}^{\top} - (\sum_{t=1}^{\tau_m^E} \mathbf{w}_{m,t} \mathbf{w}_{m,t}^{\top})^{-1} \tag{21}$$

Lemma A.22. *Define the event*

$$\mathcal{F}_n := \left\{ \|\mathbf{Z}_n - \mathbb{E}[\mathbf{Z}_n]\| \geq C \left\|\mathbf{W}_{batch}^{+}\right\|^2 \left(\frac{2c(d_1 d_2)^2}{\sqrt{Mn}} \exp\left(-\frac{Mn}{2}\right)\right)\right\}$$

It follows then that

$$\mathbb{P}(\mathcal{F}_n) \leq 4 d_1 d_2 \exp(-Mn)$$

Proof. We again proceed as Lemma A.6. Set $R = \sqrt{Mn}$ and define the truncation matrix $\mathbf{A}_n, \mathbf{C}_n$ as in the Lemma A.6. Then we can show that the quantity

$$\|\mathbf{Z}_n - \mathbb{E}[\mathbf{Z}_n]\| \leq \left\|\mathbf{W}_{\text{batch}}^{+}\right\|^2 (\|\mathbf{A}_n - \mathbb{E}[\mathbf{A}_n]\| + \|\mathbf{C}_n - \mathbb{E}[\mathbf{C}_n]\|)$$

such that $\|\mathbf{A}_{m,t,i}\| \leq \frac{3}{Mn} \cdot 2(d_1 d_2) c R$, and $\|\mathbf{C}_{m,t,i}\| \leq \frac{3}{Mn} \cdot 2(d_1 d_2) c' R$ where $c, c' > 0$. Note that $\|\mathbf{C}_{m,t,i}\| \leq \frac{1}{Mn} \cdot 2(d_1 d_2) c' R$ because $\log(n/\delta) \leq \sqrt{Mn}$. Now using the truncated Matrix Bernstein inequality we have that

$$\begin{aligned}\|\mathbf{Z}_n - \mathbb{E}[\mathbf{Z}_n]\| \leq \left\|\mathbf{W}_{\text{batch}}^{+}\right\|^2 \Bigg(&\frac{2(d_1 d_2) c}{Mn} \cdot 2(d_1 d_2) \cdot \left(R + \frac{1}{R}\right) \exp\left(-\frac{R^2}{2}\right) \\ &+ \frac{(d_1 d_2)}{Mn} \cdot 2(d_1 d_2) c' \cdot \left(R + \frac{1}{R}\right) \exp\left(-\frac{R^2}{2}\right)\Bigg)\end{aligned}$$

holds as the noise $|\eta_{m,t}| \leq R$ with probability $1-4(d_1d_2)\exp\left(-\frac{R^2}{2}\right)$ because $\eta_{m,t}$ is 1-sub Gaussian and $c, c' > 0$. Setting $R=\sqrt{Mn}$ we have that

$$
\begin{aligned}
\|\mathbf{Z}_n - \mathbb{E}[\mathbf{Z}_n]\| &\leq \left\|\mathbf{W}^+_{\text{batch}}\right\|^2 \left(\frac{2(d_1d_2)c}{Mn}\cdot 2(d_1d_2)\cdot\left(\sqrt{Mn}+\frac{1}{\sqrt{Mn}}\right)\exp\left(-\frac{Mn}{2}\right)\right.\\
&\qquad \left.+\frac{(d_1d_2)}{Mn}\cdot 2(d_1d_2)c'\cdot\left(\sqrt{Mn}+\frac{1}{\sqrt{Mn}}\right)\exp\left(-\frac{Mn}{2}\right)\right)\\
&\leq \left\|\mathbf{W}^+_{\text{batch}}\right\|^2\left(\frac{2c(d_1d_2)^2}{\sqrt{Mn}}\exp\left(-\frac{Mn}{2}\right)+\frac{2c(d_1d_2)^2}{(Mn)^{3/2}}\exp\left(-\frac{Mn}{2}\right)\right.\\
&\qquad \left.+\frac{2c'(d_1d_2)^2}{\sqrt{Mn}}\exp\left(-\frac{Mn}{2}\right)+\frac{2c'(d_1d_2)^2}{(Mn)^{3/2}}\exp\left(-\frac{Mn}{2}\right)\right)\\
&\leq C\left\|\mathbf{W}^+_{\text{batch}}\right\|^2\left(\frac{2c(d_1d_2)^2}{\sqrt{Mn}}\exp\left(-\frac{Mn}{2}\right)\right)
\end{aligned}
$$

The claim of the lemma follows. □

Lemma A.23. *(Expectation of $\widehat{\mathbf{Z}}_n$). It holds that $\mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]=\mathbf{Z}=\frac{1}{M}\sum_{m=1}^M \boldsymbol{\theta}_{m,*}\boldsymbol{\theta}_{m,*}^\top$.*

Proof. First note that the total number of samples in stage 1 is sufficiently high such that

$$
\frac{n}{3} > \frac{L_x^4(k_1k_2)^2(d_1d_2)^2c'(\rho^E)^2\log^2(2d_1d_2)}{S_r^2\omega^2M\Delta^2} \geq \left\lceil \frac{Cn\left(\rho^E\right)^2 k_1k_2^4}{M}\operatorname{polylog}\left(\rho^E, d_1d_2, k_1k_2\right)\right\rceil
$$

for some constant $C>0$ and $\delta=c'\exp(-n)$ for som $c'>0$. for some constant $C>0$ and $\delta=c'\exp(-n)$ for som $c'>0$. Then for the first stage after $\frac{n}{2}$ samples we can re-write

$$
\widehat{\mathbf{Z}}_n = \frac{2}{M\sum_m \tau_m^E}\sum_{m=1}^M\sum_{t=1}^{\tau_m^E}\boldsymbol{\theta}_{m,t}\boldsymbol{\theta}_{m,t}^\top - \mathbf{W}^+_{\text{batch}}\left(\mathbf{W}^+_{\text{batch}}\right)^\top.
$$

Now using Lemma A.7 we can prove the claim of the lemma. □

Lemma A.24. *(Concentration of $\widehat{\mathbf{B}}_{1,n}$). Suppose that event $\mathcal{F}_n$ holds. Then, for any $n>0$,*

$$
\left\|(\widehat{\mathbf{B}}_{1,n}^\perp)^\top\mathbf{B}_1\right\| \leq \frac{c'\rho^E(k_1k_2)(d_1d_2)}{S_r\sqrt{Mn}}\exp\left(-\frac{Mn}{2}\right)
$$

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

Proof. Using the Davis-Kahan $\sin\theta$ Theorem (Bhatia, 2013) and letting τ_m^E be large enough to satisfy $\left\|\widehat{\mathbf{Z}}_n-\mathbf{Z}\right\|_F \leq \frac{C_1(d_1d_2)^2}{\sqrt{M\sum_m\tau_m^E}}\exp\left(-\frac{Mn}{2}\right)$, we have

$$
\begin{aligned}
\left\|(\widehat{\mathbf{B}}_{1,n}^\perp)^\top\mathbf{B}_1\right\| &\leq \frac{\left\|\widehat{\mathbf{Z}}_n-\mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]\right\|}{\sigma_r\left(\mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]\right)-\sigma_{r+1}\left(\mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]\right)-\left\|\widehat{\mathbf{Z}}_n-\mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]\right\|}\\
&\overset{(a)}{\leq} \frac{k_1k_2}{S_rc_0}\left\|\widehat{\mathbf{Z}}_n-\mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]\right\|\\
&\overset{(b)}{\leq} \frac{ck_1k_2c_0\left\|\mathbf{W}^+_{\text{batch}}\right\|^2(d_1d_2)}{S_r\sqrt{M\sum_m\tau_m^E}}\exp\left(-\frac{Mn}{2}\right)\\
&\overset{(c)}{\leq} \frac{c'\rho^E(k_1k_2)(d_1d_2)}{S_r\sqrt{M\sum_m\tau_m^E}}\exp\left(-\frac{Mn}{2}\right) = \frac{c'\rho^E(k_1k_2)(d_1d_2)}{S_r\sqrt{Mn}}\exp\left(-\frac{Mn}{2}\right)
\end{aligned}
$$

where, (a) follows from Assumption 3.1, the (b) follows from event $\mathcal{F}_n$ and (c) follows as $\left\|\mathbf{W}^+_{\text{batch}}\right\|^2 \leq 4\rho_m^E$, and $\tau_m^E=\frac{n}{3M}$. The claim of the lemma follows. □

Lemma A.25. *(Concentration of $\widehat{\mathbf{B}}_{2,n}$). Suppose that event $\mathcal{F}_n$ holds. Then, for any $n > 0$,*

$$\left\|(\widehat{\mathbf{B}}_{2,n}^{\perp})^{\top}\mathbf{B}_2\right\| \leq \frac{c'\rho^E(k_1k_2)(d_1d_2)}{S_r\sqrt{Mn}}\exp\left(-\frac{Mn}{2}\right)$$

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

Proof. Using the Davis-Kahan $\sin\theta$ Theorem (Bhatia, 2013) and letting τ_m^E be large enough to satisfy $\left\|\widehat{\mathbf{Z}}_n - \mathbf{Z}\right\|_F \leq \frac{C_1(k_1k_2)(d_1d_2)}{\sqrt{M\sum_m \tau_m^E}}\exp\left(-\frac{Mn}{2}\right)$, we have following the same steps as in Lemma A.24 that

$$\begin{aligned}\left\|(\widehat{\mathbf{B}}_{2,n}^{\perp})^{\top}\mathbf{B}_2\right\| &\leq \frac{\left\|\widehat{\mathbf{Z}}_n - \mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]\right\|}{\sigma_r\left(\mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]\right) - \sigma_{r+1}\left(\mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]\right) - \left\|\widehat{\mathbf{Z}}_n - \mathbb{E}\left[\widehat{\mathbf{Z}}_n\right]\right\|} \\ &\leq \frac{c'\rho^E(k_1k_2)(d_1d_2)}{S_r\sqrt{Mn}}\exp\left(-\frac{Mn}{2}\right)\end{aligned}$$

where, (a) follows from Assumption 3.1, the (b) follows from event $\mathcal{F}_n$ and (c) follows as $\left\|\mathbf{W}_{\text{batch}}^{+}\right\|^2 \leq 4\rho_m^E$, and $\tau_m^E = \frac{n}{3M}$. The claim of the lemma follows. ☐

We now need to show that $\sigma_{\min}(\sum_{\widetilde{\mathbf{g}}_m(i)\in\mathcal{G}}\mathbf{b}_m(i)\widetilde{\mathbf{g}}(i)\widetilde{\mathbf{g}}(i)^{\top}) > 0$. If this holds true then we can sample following E-optimal design in the second stage and the solution to the E-optimal design in the second phase is not vacuous.

Lemma A.26. *For $Mn > \lceil\frac{(k_1k_2)^2(d_1d_2)^2c'(\rho_m^E)^2\log^2(2(d_1d_2))}{S_r^2\omega^2\Delta^2}\rceil$ we have*

$$\sigma_{\min}(\sum_i \mathbf{b}_{\widetilde{\mathbf{w}}}^E(i)\widetilde{\mathbf{w}}(i)\widetilde{\mathbf{w}}(i)^{\top}) > 0$$

Proof. We can show that

$$\sum_i \mathbf{b}_{\widetilde{\mathbf{w}}}^E(i)\widetilde{\mathbf{w}}(i)\widetilde{\mathbf{w}}(i)^{\top} \overset{(a)}{=} \sum_i \mathbf{b}^E(i)\left(\widehat{\mathbf{B}}_{1,n}^{\top}\mathbf{x}_m(i)\mathbf{x}_m(i)\widehat{\mathbf{B}}_{1,n}^{\top}\right)\widehat{\mathbf{S}}_{m,n}\left(\widehat{\mathbf{B}}_{2,n}^{\top}\mathbf{z}_m(i)\mathbf{z}_m(i)\widehat{\mathbf{B}}_{2,n}^{\top}\right)$$

where, in (a) the $\mathbf{b}^E(i)$ is the sampling proportion for the arms $\mathbf{x}$, and $\mathbf{z}$.

$$\begin{aligned}\left\|(\widehat{\mathbf{B}}_{1,n}^{\perp})^{\top}\mathbf{B}_1\right\| &\leq \frac{c'\rho^E(k_1k_2)(d_1d_2)}{S_r\sqrt{Mn}}\exp\left(-\frac{Mn}{2}\right) \\ &\overset{(a)}{\leq} \frac{\omega}{L_x^2}\frac{c'S_r\rho^E(k_1k_2)(d_1d_2)\Delta}{c'S_r(k_1k_2)(d_1d_2)\rho^E\log(2d)}\exp\left(-\frac{M}{2}\cdot\frac{L_x^4(k_1k_2)^2(d_1d_2)^2c'(\rho^E)^2\log^2(2(d_1d_2))}{S_r^2\omega^2M\Delta^2}\right) \\ &= \frac{\omega}{L_x^2}\underbrace{\frac{\Delta}{\log(2d)}}_{\leq 1}\exp\left(-\frac{(k_1k_2)^2(d_1d_2)^2c'(\rho^E)^2\log^2(2(d_1d_2))}{2S_r^2\omega^2\Delta^2}\right) \\ &\leq \frac{\omega}{L_x^2}\exp\left(-\frac{(k_1k_2)^2(d_1d_2)^2c'(\rho^E)^2\log^2(2(d_1d_2))}{2S_r^2\omega^2\Delta^2}\right)\end{aligned}$$

where (a) follows by substituting the value of n, and observe that the last inequality does not depend on the number of tasks M or budget n. Hence, for $Mn \geq \lceil\frac{(k_1k_2)^2(d_1d_2)^2c'(\rho_m^E)^2\log^2(2(d_1d_2))}{S_r^2\omega^2\Delta^2}\rceil$

$$\left\|\widehat{\mathbf{B}}_{1,n}^{\top}\mathbf{B}_1^{\perp}\right\| \leq \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\} \tag{22}$$

Similarly we can show for $n \geq \frac{(k_1 k_2)^2 (d_1 d_2)^2 c' (\rho_m^E)^2 \log^2(2(d_1 d_2))}{S_r^2 M \omega^2 \Delta^2}$

$$\left\| \widehat{\mathbf{B}}_{2,n}^\top \mathbf{B}_2^\perp \right\| \leq \min \left\{ \frac{\omega}{L_x^2}, \frac{\omega}{L_x^2} \exp\left(-Mn\right) \right\} \tag{23}$$

Then recall that for two positive semidefinite matrices $\mathbf{A}, \mathbf{B}$ we have that

$$\lambda_{\min}(\mathbf{A}) \lambda_{\min}(\mathbf{B}) \leq \lambda_{\min}(\mathbf{AB}).$$

Then we apply Lemma A.11 to show that

$$\sigma_{\min}\left(\sum_i \mathbf{b}_{\mathbf{w}}^E(i) \left(\widehat{\mathbf{B}}_{1,n}^\top \mathbf{x}_m(i) \mathbf{x}_m(i) \widehat{\mathbf{B}}_{1,n}^\top \right) \right) > 0, \quad \sigma_{\min}\left(\sum_i \mathbf{b}_{\mathbf{w}}^E(i) \left(\widehat{\mathbf{B}}_{2,n}^\top \mathbf{z}_m(i) \mathbf{z}_m(i) \widehat{\mathbf{B}}_{2,n}^\top \right) \right) > 0$$

and the $\sigma_{\min}(\widehat{\mathbf{S}}_{m,n}) > 0$ by the construction of (7). Hence we get the claim of the lemma. □

Lemma A.27. *Suppose that event $\mathcal{F}_n$ holds and $Mn > \lceil \frac{L_x^4 (k_1 k_2)^2 (d_1 d_2)^2 L_x^4 c' (\rho^E)^2 \log^2(2 d_1 d_2)}{S_r^2 \omega^2 \Delta^2} \rceil$. Then define*

$$\mathbf{V}_m = \sum_i \mathbf{b}_{\widetilde{\mathbf{w}}}^E\,(i)\, \widetilde{\mathbf{w}}_m(i) \widetilde{\mathbf{w}}_m(i)^\top.$$

where, $\widetilde{\mathbf{w}}_m(i) = \mathrm{vec}(\widetilde{\mathbf{g}}_m(i); \widetilde{\mathbf{v}}_m(i))$. For any task $m \in [M]$ and $\mathbf{x}_j \in \mathbb{R}^d$,

$$\|\widetilde{\mathbf{w}}(j)\|_{\mathbf{V}_m^{-1}}^2 \leq \|\mathbf{w}(j)\|_{\mathbf{V}_m^{-1}}^2 + \frac{c L_x^4}{k_1 k_2 S_r^2 \omega^2} \exp(-Mn)$$

for some constant $c > 0$

Proof. The proof of this lemma follows directly from Lemma A.12 and using the relation from (22) and (23)

$$\left\| \widehat{\mathbf{B}}_{1,n}^\top \mathbf{B}_1^\perp \right\| \leq \min \left\{ \frac{\omega}{L_x^2}, \frac{\omega}{L_x^2} \exp\left(-Mn\right) \right\}$$
$$\left\| \widehat{\mathbf{B}}_{2,n}^\top \mathbf{B}_2^\perp \right\| \leq \min \left\{ \frac{\omega}{L_x^2}, \frac{\omega}{L_x^2} \exp\left(-Mn\right) \right\}$$

Plugging the value of n and usin Assumption 3.2, we have that for any task $m \in [M]$, $\sum_i \mathbf{b}_{\mathbf{w}}^E\,(i)\, \mathbf{B}_1^\top \mathbf{x}(i) \mathbf{x}(i)^\top \mathbf{B}_1$ and $\sum_i \mathbf{b}_{\mathbf{w}}^E\,(i)\, \mathbf{B}_2^\top \mathbf{x}(i) \mathbf{x}(i)^\top \mathbf{B}_2$ is invertible we can get the claim of the lemma. □

Let $\widehat{\mathbf{s}}_{m,n} = \mathrm{vec}(\widehat{\mathbf{S}}_{m,n}) \in \mathbb{R}^{k_1 k_2}$ and $\mathbf{s}_{m,*} = \mathrm{vec}(\mathbf{S}_{m,*}) \in \mathbb{R}^{k_1 k_2}$.

Lemma A.28. *Let $\mathcal{F}_n$ hold. Define $\widetilde{\Delta}_{m,i} = \widetilde{\mathbf{w}}(i)^\top \widehat{\mathbf{s}}_{m,n} - \widetilde{\mathbf{w}}(i_m^*)^\top \widehat{\mathbf{s}}_{m,n}$ and $\Delta_{m,i} = \mathbf{w}(i)^\top \mathbf{s}_{m,*} - \mathbf{w}(i_m^*)^\top \mathbf{s}_{m,*}$. Then the estimation error in second stage is given by*

$$|\widetilde{\Delta}_{m,i} - \Delta_{m,i}| \leq 6 k L_x \min \left\{ \frac{\omega}{L_x^2}, \frac{\omega}{L_x^2} \exp\left(-Mn\right) \right\}.$$

Further for $Mn > \lceil \frac{L_x^4 (k_1 k_2)^2 (d_1 d_2)^2 c' (\rho^E)^2 \log^2(2 d_1 d_2)}{S_r \omega^2 \Delta^2} \rceil$ we have that

$$|\widetilde{\Delta}_{m,i} - \Delta_{m,i}| \leq \frac{\Delta_{m,i}}{2}.$$

Proof. The proof follows the same steps as in Lemma A.13 by first using the relation that under the event $\mathcal{F}_n$ the following holds,

$$\left\|\widehat{\mathbf{B}}_{1,n}^\top \mathbf{B}_1^\perp\right\| \leq \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}$$
$$\left\|\widehat{\mathbf{B}}_{2,n}^\top \mathbf{B}_2^\perp\right\| \leq \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}.$$

Then plugging in the value of n gives the claim of the lemma. This can be shown as follows:

$$\begin{aligned}\left|(\widetilde{\mathbf{w}}-\widetilde{\mathbf{w}}')^\top \widehat{\mathbf{s}}_{m,n} - (\widetilde{\mathbf{w}}-\widetilde{\mathbf{w}}')^\top \mathbf{s}_{m,*}\right| &\leq 2k_1k_2\cdot L_xL_w\left\|\widehat{\mathbf{B}}_{1,n}^\top \mathbf{B}_1^\perp\right\|\left\|\widehat{\mathbf{B}}_{2,n}^\top \mathbf{B}_2^\perp\right\| + \frac{\sqrt{\rho^E\cdot 2\log\left(\frac{4n^2M}{\delta}\right)}}{\sqrt{n}}\\ &+ 2L_xL_w\left\|\widehat{\mathbf{B}}_{1,n}^\top \mathbf{B}_1^\perp\right\|\left\|\widehat{\mathbf{B}}_{2,n}^\top \mathbf{B}_2^\perp\right\|\end{aligned}$$

Setting $L_w = 1$, $\rho^E = 2k_1k_2$ and $\log\left(\frac{4n^2M}{\delta}\right) = n$ and as the event $\mathcal{F}_n$ holds, we get that

$$\left|(\widetilde{\mathbf{w}}-\widetilde{\mathbf{w}}')^\top \widehat{\mathbf{s}}_{m,n} - (\widetilde{\mathbf{w}}-\widetilde{\mathbf{w}}')^\top \mathbf{s}_{m,*}\right| \leq 6k_1k_2L_x\min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}$$

This implies that

$$|\widetilde{\Delta}_{m,i} - \Delta_{m,i}| \leq 6k_1k_2L_x\min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}$$

Now for $Mn > \lceil\frac{L_x^4(k_1k_2)^2(d_1d_2)^2c'(\rho^E)^2\log^2(2d_1d_2)}{\omega^2\Delta^2}\rceil$ we can show that

$$\begin{aligned}6k_1k_2L_x\min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\} &= 6k_1k_2L_x\min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}\\ &\leq 6k_1k_2L_x\min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}(2d)^{-\frac{L_x^4(k_1k_2)^2(d_1d_2)^2c'(\rho^E)^2}{S_r^2\omega^2\Delta^2}}\right\}\\ &\overset{(a)}{\leq}\frac{\Delta}{2}\overset{(b)}{\leq}\frac{\Delta_{m,i}}{2}\end{aligned}$$

where, (a) holds as for any $\Delta > 0, d, k > 1, \omega > 0$ the following holds

$$\log(\frac{\Delta}{12}) + \log(\frac{L_x}{k\omega}) > -\frac{L_x^4(k_1k_2)^2(d_1d_2)^2c'(\rho^E)^2}{S_r^2\omega^2\Delta^2}\log(2d_1d_2).$$

The (b) holds as $\Delta_{m,i} \geq \Delta$. The claim of the lemma follows. □

Second Stage for FB-DOE

Good Event: Define the good event $\mathcal{F}_n'$ that the algorithm has a good estimate of $\mathbf{S}_{m,*}$ for each $m \in [M]$ as follows:

$$\mathcal{F}_n' = \left\{\left\|\widehat{\mathbf{S}}_{m,*} - \mathbf{S}_{m,*}\right\|_F \leq \frac{c(k_1+k_2)^{3/2}\sqrt{r}}{\sqrt{n}}\right\}, \tag{24}$$

where, $C_2 > 0$, some nonzero constant. Let the matrix $\widetilde{\mathbf{W}}_t = \widetilde{\mathbf{x}}_t\widetilde{\mathbf{z}}_t^\top$.

Lemma A.29. ***(Restatement of Lemma 23 of Lu et al. (2021), Converence under RSC, adapted from Proposition 10.1 in Wainwright** (2019)**)*** *Suppose the observations $\widetilde{\mathbf{W}}_1, \ldots, \widetilde{\mathbf{W}}_n \in \mathbb{R}^{k_1\times k_2}$ satisfies the non-scaled RSC condition, such that*

$$\frac{1}{n}\sum_{t=1}^n\left\langle\widetilde{\mathbf{W}}_t, \mathbf{S}\right\rangle^2 \geq \kappa\|\mathbf{S}\|_F^2 - \tau_n^2\|\mathbf{S}\|_{nuc}^2, \forall\mathbf{S}\in\mathbb{R}^{k_1\times k_2}.$$

Then under the event $G := \left\{ \left\| \frac{1}{n} \sum_{t=1}^{n} \eta_t \widetilde{\mathbf{W}}_t \right\| \leq \frac{\lambda_n}{2} \right\}$, *any optimal solution* $\widehat{\mathbf{S}}_n$ *to* (7) *satisfies the bound below:*

$$\left\| \widehat{\mathbf{S}}_n - \mathbf{S}_* \right\|_F^2 \leq 4.5 \frac{\lambda_n^2}{\kappa^2} r,$$

where $r = \operatorname{rank}(\mathbf{\Theta}^*)$ *and* $\frac{1}{\tau_n^2} \geq \frac{64r}{\kappa}$.

Lemma A.30. ***(Restatement of Theorem 15 of (Lu et al., 2021), Distribution*** $\mathbf{b}$ ***satisfies RSC)*** *Sample* $\widetilde{\mathbf{W}}_1, \ldots, \widetilde{\mathbf{W}}_{n/3} \in \mathbb{R}^{k_1 \times k_2}$ *from* $\widetilde{\mathcal{W}}$ *according to* $\mathbf{b}$, *and define* $\widetilde{\mathbf{w}}_i := \operatorname{vec}\left(\widetilde{\mathbf{W}}_i\right), \widetilde{\mathbf{Q}} = \left[\widetilde{\mathbf{w}}_1^T; \ldots; \widetilde{\mathbf{w}}_n^T\right] \in \mathbb{R}^{n/3 \times k_1 k_2}$ *and* $\widehat{\Gamma} := \frac{3}{n} \widetilde{\mathbf{Q}}^T \widetilde{\mathbf{Q}}$. *Let* $\mathbf{\Sigma}_n$ *be the covariance matrix after sampling* $\widetilde{\mathbf{W}}_t$ *using distribution* $\mathbf{b}$. *Then under the condition that the minimum eigenvalue of covariance matrix* $\mathbf{\Sigma}_n$ *is greater than* 0, *there exists constants* $c_1, c_2 > 0$, *such that with probability* $1 - \delta$,

$$\widetilde{\mathbf{S}}_m^T \widehat{\Gamma} \widetilde{\mathbf{S}}_m = \frac{3}{n} \sum_{t=1}^{n/3} \left\langle \widetilde{\mathbf{W}}_t, \mathbf{S}_m \right\rangle^2 \geq \frac{c_1}{k_1 k_2} \|\mathbf{S}_m\|_F^2 - \frac{c_2 (k_1 + k_2)}{n k_1 k_2} \|\mathbf{S}_m\|_{nuc}^2, \forall \mathbf{S}_m \in \mathbb{R}^{k_1 \times k_2},$$

for $n = \Omega\left((k_1 + k_2) \log\left(\frac{1}{\delta}\right)\right)$, *where* $\widetilde{\mathbf{S}}_m := \operatorname{vec}(\mathbf{S}_{m,*})$.

Lemma A.30 states that sampling $\widetilde{\mathbf{W}}_t$ from $\widetilde{\mathcal{W}}$ according to distribution $\mathbf{b}$ guarantees that the sampled arms satisfies Restricted String Convexity (RSC) condition. We further show that under RSC condition, the estimated $\widehat{\mathbf{S}}_{m,n}$ is guaranteed to converge to $\mathbf{S}_{m,*}$ at a fast rate in Lemma A.31. Using Lemma A.26, and Lemma A.28 we know that in the second stage for $Mn > \lceil \frac{L_x^4 (k_1 k_2)^2 (d_1 d_2)^2 c' (\rho^E)^2 \log^2(2 d_1 d_2)}{S_r^2 \omega^2 \Delta^2} \rceil$ the minimum eigenvalue is greater than 0, and the estimation error of features are small. We also know from Jun et al. (2019) that E-optimal design satisfies the property of the distribution D.

Lemma A.31. *The event* $\mathcal{F}_n \cap \mathcal{F}_n'$ *in* (24) *holds with probability greater than* $1 - 2(k_1 + k_2)^{3/2} \exp\left(\frac{c L_x^4}{k_1 k_2 S_r^2 \omega^2}\right) \exp\left(\frac{-n}{2}\right)$.

Proof. Define the rare event $\xi := \left\{\max_{t=1,\ldots,T_1} |\eta_t| > \sqrt{2n}\right\}$, so that $\mathbb{P}(\xi) \leq \exp(-n)$ can be proved by the definition of sub-Gaussian. Define By matrix Bernstein inequality, the probability of $G(\lambda_n)^c$ can be bounded in the following way using Lemma A.29 as follows:

$$\begin{aligned} \mathbb{P}\left(\left\| \frac{1}{n} \sum_{t=1}^{n} \eta_t \widetilde{\mathbf{W}}_t \right\| > \epsilon \right) &\overset{(a)}{\leq} \mathbb{P}\left(\left\| \frac{1}{n} \sum_{t=1}^{n} \eta_t \widetilde{\mathbf{W}}_t \right\| > \epsilon \mid \xi^c \right) + \mathbb{P}(\xi) \\ &\overset{(b)}{\leq} (k_1 + k_2) \exp\left(\frac{-n\epsilon^2/2}{2 \log\left(\frac{4n}{\delta}\right) \max\{1/k_1, 1/k_2\} + \epsilon \sqrt{2 \log\left(\frac{4n}{\delta}\right)}/3} \right) + \delta/2 \end{aligned}$$

where, in (a) the matrix $\widetilde{\mathbf{W}}_t = \widetilde{\mathbf{x}}_t \widetilde{\mathbf{z}}_t^\top$, and (b) follows from Matrix Bernstein inequality. Now setting $\log\left(\frac{k_1 + k_2}{\delta}\right) = n$, $p = k_1 + k_2 \geq \max\{1/k_1, 1/k_2\}$. This implies that $\log\left(\frac{4n}{\delta}\right) \leq (k_1 +$

$k_2)\log(4n) + n \leq 2np$. Then set

$$
\begin{aligned}
&(k_1+k_2)\exp\left(\frac{-n\epsilon^2/2}{4np^2+\epsilon\sqrt{4np}/3}\right) = \delta \\
\implies &\exp\left(\frac{-n\epsilon^2/2}{4np^2+\epsilon\sqrt{4np}/3}\right) = \frac{\delta}{k_1+k_2} \\
\implies &\frac{4np^2+\epsilon\sqrt{4np}/3}{n\epsilon^2/2} = \log\left(\frac{k_1+k_2}{\delta}\right) \\
\implies &\frac{4np^2+\epsilon\sqrt{4np}/3}{\log\left(\frac{k_1+k_2}{\delta}\right)} = n\epsilon^2/2 \\
\implies &\frac{4np^2+2\epsilon\sqrt{4np}/3}{n\log\left(\frac{k_1+k_2}{\delta}\right)} = \epsilon^2 \\
\implies &\frac{4p}{\log\left(\frac{k_1+k_2}{\delta}\right)} + \frac{\epsilon 2\sqrt{4p}}{3\sqrt{n}\log\left(\frac{k_1+k_2}{\delta}\right)} = \epsilon^2 \\
\implies &\epsilon^2 - \frac{\epsilon 2\sqrt{4p}}{3n\sqrt{n}} - \frac{4p}{n} = 0 \\
\implies &\epsilon = \frac{\frac{2\sqrt{4p}}{n\sqrt{n}} + \sqrt{\frac{16p}{9n^3} + 4\cdot 1\cdot\frac{4p}{n}}}{2} \\
\implies &\epsilon = \frac{\sqrt{4p}}{n\sqrt{n}} + 2\sqrt{\frac{p}{9n^3}+\frac{p}{n}}
\end{aligned}
$$

where the last equality follows by quadratic formula. Therefore by setting $\epsilon = \frac{c(k_1+k_2)}{\sqrt{n}}$ for some constant $c > 0$ we get that

$$
\mathbb{P}\left(\left\|\frac{1}{n}\sum_{t=1}^{n}\eta_t\widetilde{\mathbf{W}}_t\right\| > \frac{c(k_1+k_2)}{\sqrt{n}}\right) \leq C\left(k_1+k_2\right)\exp\left(\frac{-n}{2}\right)
$$

for some constant $C > 0$. Now set $\lambda_n = 2\epsilon$, we need $\lambda_n^2 = \frac{C(k_1+k_2)}{n}$ and under this condition we have $\mathbb{P}\left(G\left(\lambda_n\right)\right) \geq 1 - C\left(k_1+k_2\right)\exp\left(\frac{-n}{2}\right)$. Finally we complete the proof by noting that the scaling of the right hand side in Lemma A.29 under above choice of λ_n is less than $\frac{(k_1+k_2)^3 r}{n}$. This yields that

$$
\mathbb{P}(\mathcal{F}_n') = \mathbb{P}\left(\left\|\widehat{\mathbf{S}}_{m,*} - \mathbf{S}_{m,*}\right\|_F \geq \frac{c(k_1+k_2)^{3/2}\sqrt{r}}{\sqrt{n}}\right) \leq C\left(k_1+k_2\right)^{3/2}\exp\left(\frac{-n}{2}\right).
$$

Finally, note that the latent feature estimation error in the second stage results in an additional factor of $\exp\left(\frac{cL_x^4}{k_1k_2S_r^2\omega^2}\right)$. This yields that

$$
\mathbb{P}(\mathcal{F}_n') = \mathbb{P}\left(\left\|\widehat{\mathbf{S}}_{m,*} - \mathbf{S}_{m,*}\right\|_F \geq \frac{c(k_1+k_2)^{3/2}\sqrt{r}}{\sqrt{n}}\right) \leq C\left(k_1+k_2\right)^{3/2}\exp\left(\frac{cL_x^4}{k_1k_2S_r^2\omega^2}\right)\exp\left(\frac{-n}{2}\right).
$$

The claim of the lemma follows. □

Lemma A.32. *(Concentration of $\widehat{\mathbf{U}}_{m,n}$). Suppose that event $\mathcal{F}_n$ holds. Then, for any $n > 0$,*

$$
\left\|(\widehat{\mathbf{U}}_{m,n}^{\perp})^\top\mathbf{U}_m\right\| \leq \frac{c'(k_1k_2)^{2.5}\sqrt{r}}{S_r\sqrt{n}}\exp\left(\frac{-n}{2}\right),
$$

for some constant $c' > 0$.

Proof. Using the Davis-Kahan $\sin\theta$ Theorem (Bhatia, 2013) and letting $\tau_m^E = \frac{n}{3M}$ be large enough to satisfy $\left\|\widehat{\mathbf{S}}_{m,*} - \mathbf{S}_{m,*}\right\|_F \leq \frac{c(k_1+k_2)^{3/2}\sqrt{r}}{\sqrt{n}}$, we have

$$
\begin{aligned}
\left\|(\widehat{\mathbf{U}}_{m,n}^{\perp})^\top \mathbf{U}_m\right\| &\leq \frac{\left\|\widehat{\mathbf{S}}_{m,*} - \mathbb{E}\left[\widehat{\mathbf{S}}_{m,*}\right]\right\|}{\sigma_r\left(\mathbb{E}\left[\widehat{\mathbf{S}}_{m,*}\right]\right) - \sigma_{r+1}\left(\mathbb{E}\left[\widehat{\mathbf{S}}_{m,*}\right]\right) - \left\|\widehat{\mathbf{S}}_{m,*} - \mathbb{E}\left[\widehat{\mathbf{S}}_{m,*}\right]\right\|} \\
&\overset{(a)}{\leq} \frac{k_1 k_2}{S_r c_0}\left\|\widehat{\mathbf{S}}_{m,*} - \mathbb{E}\left[\widehat{\mathbf{S}}_{m,*}\right]\right\| \\
&\overset{(b)}{\leq} \frac{c'(k_1k_2)(k_1+k_2)^{3/2}\sqrt{r}}{S_r\sqrt{n}} \exp\left(\frac{-n}{2}\right) \\
&\leq \frac{c'(k_1k_2)^{2.5}\sqrt{r}}{S_r\sqrt{n}} \exp\left(\frac{-n}{2}\right)
\end{aligned}
$$

where, (a) follows from Assumption 3.1, the (b) follows from event $\mathcal{F}_n'$. The claim of the lemma follows. $\square$

Lemma A.33. *(Concentration of $\widehat{\mathbf{V}}_{m,n}$). Suppose that event $\mathcal{F}_n'$ holds. Then, for any $n > 0$,*

$$
\left\|(\widehat{\mathbf{V}}_{m,n}^{\perp})^\top \mathbf{V}_m\right\| \leq \frac{c'(k_1k_2)^{2.5}\sqrt{r}}{S_r\sqrt{n}} \exp\left(\frac{-n}{2}\right)
$$

for some constant $c' > 0$.

Proof. Using the Davis-Kahan $\sin\theta$ Theorem (Bhatia, 2013) and letting $\tau_m^E = \frac{n}{3M}$ be large enough to satisfy $\left\|\widehat{\mathbf{S}}_{m,*} - \mathbf{S}_{m,*}\right\|_F \leq \frac{c(k_1+k_2)^{3/2}\sqrt{r}}{\sqrt{n}}$, we have following the same steps as in Lemma A.32 that

$$
\begin{aligned}
\left\|(\widehat{\mathbf{V}}_{m,n}^{\perp})^\top \mathbf{V}_m\right\| &\overset{(a)}{\leq} \frac{\left\|\widehat{\mathbf{S}}_{m,*} - \mathbb{E}\left[\widehat{\mathbf{S}}_{m,*}\right]\right\|}{\sigma_r\left(\mathbb{E}\left[\widehat{\mathbf{S}}_{m,*}\right]\right) - \sigma_{r+1}\left(\mathbb{E}\left[\widehat{\mathbf{S}}_{m,*}\right]\right) - \left\|\widehat{\mathbf{S}}_{m,*} - \mathbb{E}\left[\widehat{\mathbf{S}}_{m,*}\right]\right\|} \\
&\overset{(b)}{\leq} \frac{c'(k_1k_2)^{2.5}\sqrt{r}}{S_r\sqrt{n}} \exp\left(\frac{-n}{2}\right)
\end{aligned}
$$

where, (a) follows from Assumption 3.1, the (b) follows from event $\mathcal{F}_n'$. The claim of the lemma follows. $\square$

Arm rotation in Stage 2 Recall that the SVD of $\widehat{\mathbf{S}}_{m,n} = \widehat{\mathbf{U}}_{m,n}\widehat{\mathbf{D}}_{m,n}\widehat{\mathbf{V}}_{m,n}^\top$. Define $\widehat{\mathbf{H}}_{m,\ell} = [\widehat{\mathbf{U}}_{m,n}\widehat{\mathbf{U}}_{m,n}^{\perp}]^\top \widehat{\mathbf{S}}_{m,n}[\widehat{\mathbf{V}}_{m,n}\widehat{\mathbf{V}}_{m,n}^{\perp}]$. 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:

$$
\begin{aligned}
\underline{\mathcal{G}}_{m,0} &= \Big\{\Big[\mathbf{vec}\left(\widetilde{\mathbf{g}}_{m,1:r}\widetilde{\mathbf{v}}_{m,1:r}^\top\right); \mathbf{vec}\left(\widetilde{\mathbf{g}}_{m,r+1:k_1}\widetilde{\mathbf{v}}_{m,1:r}^\top\right); \\
&\quad \mathbf{vec}\left(\widetilde{\mathbf{g}}_{m,1:r}\widetilde{\mathbf{v}}_{m,r+1:k_2}^\top\right); \mathbf{vec}\left(\widetilde{\mathbf{g}}_{m,r+1:k_1}\widetilde{\mathbf{v}}_{m,r+1:k_2}^\top\right)\Big]\Big\} \\
\widehat{\mathbf{s}}_{m,n,1:\widetilde{k}} &= [\mathbf{vec}(\widehat{\mathbf{H}}_{m,n,1:r,1:r}); \mathbf{vec}(\widehat{\mathbf{H}}_{m,n,r+1:k_1,1:r}); \\
&\quad \mathbf{vec}(\widehat{\mathbf{H}}_{m,n,1:r,r+1:k_2})], \\
\widehat{\mathbf{s}}_{m,n,\widetilde{k}+1:k_1k_2} &= \mathbf{vec}(\widehat{\mathbf{H}}_{m,n,r+1:k_1,r+1:k_2}).
\end{aligned} \tag{25}
$$

Finally we estimate the

$$
\widehat{\mathbf{s}}_{m,n} = \arg\min_{\mathbf{s}} \tfrac{1}{2}\|\underline{\mathbf{W}}_{m,n}\mathbf{s} - \mathbf{r}_m\|_2^2 + \tfrac{1}{2}\|\mathbf{s}\|_{\mathbf{\Lambda}_{m,n}}^2 \tag{26}
$$

Lemma A.34. ***(Restatement of Lemma 3 of Valko et al. (2014))*** *If $\lambda_\perp = \frac{n}{3k_1k_2\log\left(1+\frac{n}{3\lambda}\right)}$, then*

$$\log\frac{|\mathbf{V}_T|}{|\mathbf{\Lambda}|} \leq 2k_1k_2\log\left(1+\frac{n}{3\lambda}\right)$$

Lemma A.35. ***(Restatement of Lemma 1 of Jun et al. (2019))*** *Using Lemma A.34 we can show that*

$$\|\mathbf{s}_*\|_{\mathbf{\Lambda}} \leq \sqrt{\lambda\left\|\mathbf{s}_{1:\widetilde{k}}\right\|_2^2 + \lambda_\perp\left\|\mathbf{s}_{\widetilde{k}+1:k_1k_2}\right\|_2^2} \leq \sqrt{\lambda}B + \sqrt{\lambda_\perp}B_\perp$$

Setting $B_\perp = \frac{3}{n}$, and $\lambda_\perp = \frac{n}{3k_1k_2\log\left(1+\frac{n}{3\lambda}\right)}$ results in $\frac{1}{2}\|\mathbf{s}_*\|_{\mathbf{\Lambda}}^2 \leq \frac{1}{36n}$.

Lemma A.36. *For $Mn > \lceil\frac{(k_1k_2)^2(d_1d_2)^2L_x^4c'(\rho_m^E)^2\log^2(2(d_1d_2))}{S_r^2\omega^2\Delta^2}\rceil$ we have*

$$\sigma_{\min}\big(\sum_i \mathbf{b}_{\underline{\mathbf{w}}}^E(i)\underline{\mathbf{w}}(i)\underline{\mathbf{w}}(i)^\top\big) > 0$$

Proof. We can show that

$$\sum_i \mathbf{b}_{\underline{\mathbf{w}}}^E(i)\underline{\mathbf{w}}(i)\underline{\mathbf{w}}(i)^\top \overset{(a)}{=} \sum_i \mathbf{b}_{\underline{\mathbf{w}}}^E(i)\left(\widehat{\mathbf{U}}_{m,n}^\top\widetilde{\mathbf{x}}_m\widetilde{\mathbf{x}}_m\widehat{\mathbf{U}}_{1,n}^\top\right)\widehat{\mathbf{D}}_{m,n}\left(\widehat{\mathbf{V}}_{m,n}^\top\widetilde{\mathbf{z}}_m\widetilde{\mathbf{z}}_m\widehat{\mathbf{V}}_{m,n}^\top\right)$$

where, in (a) the $\mathbf{b}_m^G(i)$ is the sampling proportion for the arms $\widetilde{\mathbf{x}} \in \mathbb{R}^{k_1}$, $\widetilde{\mathbf{z}} \in \mathbb{R}^{k_2}$, $\widehat{\mathbf{U}} \in \mathbb{R}^{k_1\times r}$ and $\widehat{\mathbf{V}} \in \mathbb{R}^{k_1\times r}$. Also note that from Lemma A.32 and Lemma A.33 we know that

$$\begin{aligned}
\left\|(\widehat{\mathbf{U}}_{m,n}^\perp)^\top\mathbf{U}_m\right\| &\leq \frac{c'(k_1k_2)^2\sqrt{r}}{S_r\sqrt{n}}\exp\left(\frac{-n}{2}\right)\\
&\overset{(a)}{\leq} \frac{\omega}{L_x^2}\frac{c'S_r\rho^E(k_1k_2)(d_1d_2)\Delta}{c'S_r(k_1k_2)d\rho^E\log(2d)}\exp\left(-\frac{M}{2}\cdot\frac{L_x^4(k_1k_2)^2(d_1d_2)^2L_x^4c'(\rho^E)^2\log^2(2d)}{S_r^2\omega^2M\Delta^2}\right)\\
&= \frac{\omega}{L_x^2}\underbrace{\frac{\Delta}{\log(2d)}}_{\leq 1}\exp\left(-\frac{(k_1k_2)^2(d_1d_2)^2c'L_x^4(\rho^E)^2\log^2(2d)}{2S_r^2\omega^2\Delta^2}\right)\\
&\leq \frac{\omega}{L_x^2}\exp\left(-\frac{(k_1k_2)^2(d_1d_2)^2c'L_x^4(\rho^E)^2\log^2(2d)}{2S_r^2\omega^2\Delta^2}\right)
\end{aligned}$$

where (a) follows by substituting the value of n, and observe that the last inequality does not depend on the number of tasks M or budget n. Hence, for $Mn \geq \lceil\frac{(k_1k_2)^2(d_1d_2)^2L_x^4c'(\rho_m^E)^2\log^2(2(d_1d_2))}{S_r^2\omega^2\Delta^2}\rceil$ we have

$$\left\|\widehat{\mathbf{U}}_{m,n}^\top\mathbf{U}_m^\perp\right\| \leq \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-n\right)\right\}. \tag{27}$$

Similarly we can show that for $Mn \geq \lceil\frac{(k_1k_2)^2(d_1d_2)^2L_x^4c'(\rho_m^E)^2\log^2(2(d_1d_2))}{S_r^2\omega^2\Delta^2}\rceil$ we have

$$\left\|\widehat{\mathbf{V}}_{m,n}^\top\mathbf{V}_m^\perp\right\| \leq \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-n\right)\right\}, \quad \left\|\widehat{\mathbf{D}}_{m,n}^\top\mathbf{D}_m^\perp\right\| \leq \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-n\right)\right\}. \tag{28}$$

This holds with high probability as the event $\mathcal{F}_n'$ holds true. Then following the same steps as in Lemma A.26 and applying Lemma A.11 we get the claim of the lemma. □

If this holds true then we can sample the following G-optimal design and the solution to the G-optimal design in the third phase is not vacuous.

Lemma A.37. *Suppose that event $\mathcal{F}_n'$ holds and $Mn > \frac{L_x^4(k_1k_2)^2(d_1d_2)^2L_x^4c'(\rho^E)^2\log^2(2d_1d_2)}{S_r^2\omega^2\Delta^2}$. Then define*

$$\mathbf{\Sigma}_{m,\ell} = \sum_i \mathbf{b}^G_{m,\widetilde{\mathbf{g}}}(i)\, \widetilde{\mathbf{g}}_{m,\ell}(i)\widetilde{\mathbf{g}}_{m,\ell}(i)^\top.$$

For any task $m \in [M]$ and $\mathbf{x}_j \in \mathbb{R}^d$,

$$\|\widetilde{\mathbf{g}}(j)\|^2_{\mathbf{\Sigma}^{-1}_{m,\ell}} \leq \|\mathbf{g}(j)\|^2_{\mathbf{\Sigma}^{-1}_{m,\ell}} + \frac{cL_x^4}{(k_1+k_2)rS_r^2\omega^2}\exp(-Mn)$$

for some constant $c > 0$

Proof. The proof of this lemma follows using the same steps as in Lemma A.12 and using the relation from (27) and (28)

$$\left\|\widehat{\mathbf{U}}_{m,n}^\top \mathbf{U}_m^\perp\right\| \leq \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-n\right)\right\}.$$
$$\left\|\widehat{\mathbf{V}}_{m,n}^\top \mathbf{V}_m^\perp\right\| \leq \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-n\right)\right\}.$$

Plugging the value of n and using Assumption 3.3, we have that for any task $m \in [M]$, $\sum_i \mathbf{b}^G_{\mathbf{w}}(i)\, \mathbf{U}_m^\top \mathbf{g}(i)\mathbf{g}(i)^\top\mathbf{U}_m$ and $\sum_i \mathbf{b}^G_{\mathbf{w}}(i)\, \mathbf{V}_m^\top \mathbf{v}(i)\mathbf{v}(i)^\top\mathbf{V}_m$ is invertible we can get the claim of the lemma. ☐

Recall that $\widehat{\mathbf{s}}_{m,n} = \mathrm{vec}(\widehat{\mathbf{S}}_{m,n}) \in \mathbb{R}^{(k_1k_2}$ and $\mathbf{s}_{m,*} = \mathrm{vec}(\mathbf{S}_{m,*}) \in \mathbb{R}^{k_1k_2}$.

Lemma A.38. *Let $\widetilde{\Delta}_{m,i} = \mathbf{g}(i)^\top\widehat{\mathbf{s}}_{m,n} - \mathbf{g}(i_m^*)^\top\widehat{\mathbf{s}}_{m,n}$ and $\Delta_{m,i} = \mathbf{w}(i)^\top\mathbf{s}_{m,*} - \mathbf{w}(i_m^*)^\top\mathbf{s}_{m,*}$. Then the estimation error in second stage is given by*

$$|\widetilde{\Delta}_{m,i} - \Delta_{m,i}| \leq 6(k_1+k_2)rL_x\min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}.$$

Further for $Mn > \frac{L_x^4(k_1k_2)^2(d_1d_2)^2c'(\rho^E)^2\log^2(2d_1d_2)}{S_r^2\omega^2\Delta^2}$ we have that

$$|\widetilde{\Delta}_{m,i} - \Delta_{m,i}| \leq \frac{\Delta_{m,i}}{2}.$$

Proof. The proof follows the same steps as in Lemma A.13 by first using the relation that

$$\left\|\widehat{\mathbf{U}}_{m,n}^\top \mathbf{U}_m^\perp\right\| \leq \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}$$
$$\left\|\widehat{\mathbf{V}}_{m,n}^\top \mathbf{V}_m^\perp\right\| \leq \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}.$$

Then plugging in the value of n gives the claim of the lemma.

Again, for this stage using the same steps as in Lemma A.13 we can bound the estimation error for any pair of $\mathbf{g}, \mathbf{g}' \in \mathbb{R}^k$ for a task m as follows:

$$\left|(\mathbf{g}-\mathbf{g}')^\top\widehat{\mathbf{s}}_{m,n} - (\mathbf{g}-\mathbf{g}')^\top\mathbf{s}_{m,*}\right| \leq 2\widetilde{k}\cdot L_xL_w\left\|\widehat{\mathbf{U}}_{n,\perp}^\top\mathbf{U}\right\|\left\|\widehat{\mathbf{V}}_{n,\perp}^\top\mathbf{V}\right\| + \frac{\sqrt{\rho_m^E\cdot 2\log\left(\frac{4n^2M}{\delta}\right)}}{\sqrt{n}} + 2L_xL_w\left\|\widehat{\mathbf{U}}_{n,\perp}^\top\mathbf{U}\right\|\left\|\widehat{\mathbf{V}}_{n,\perp}^\top\mathbf{V}\right\|$$

Setting $L_w = 1$, $\rho_m^E = 2\widetilde{k}\log(1+\frac{n}{3\lambda})$ and $\log\left(\frac{4n^2M}{\delta}\right) = n$ and as the event $\mathcal{F}_n$ holds, we get that

$$\left|(\mathbf{g}-\mathbf{g}')^\top\widehat{\mathbf{s}}_{m,n} - (\mathbf{g}-\mathbf{g}')^\top\mathbf{s}_{m,*}\right| \leq 6k\log(1+\frac{n}{3\lambda})L_x\min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\}$$

This implies that

$$|\widetilde{\Delta}_{m,i} - \Delta_{m,i}| \leq 6kL_x \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn + \log\log(1+n/3\lambda)\right)\right\}$$

Now for $n > \frac{L_x^4 k^2 d^2 c'(\rho^E)^2 \log^2(2d)}{\omega^2 M \Delta^2}$ we can show that

$$\begin{aligned}
6\widetilde{k}L_x \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn\right)\right\} &= 6\widetilde{k}L_x \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}\exp\left(-Mn + \log\log(1+n/3\lambda)\right)\right\} \\
&\lesssim 6\widetilde{k}L_x \min\left\{\frac{\omega}{L_x^2}, \frac{\omega}{L_x^2}(2d)^{-\frac{L_x^4(k_1k_2)^2(d_1d_2)^2c'(\rho^E)^2}{S_r^2\omega^2\Delta^2}}\right\} \\
&\overset{(a)}{\leq} \frac{\Delta}{2} \overset{(b)}{\leq} \frac{\Delta_{m,i}}{2}
\end{aligned}$$

where, (a) holds as for any $\Delta > 0, d, k > 1, \omega > 0$ the following holds

$$\log(\frac{\Delta}{12}) + \log(\frac{L_x}{k\omega}) > -\frac{L_x^4 k^2 d^2 c'(\rho^E)^2}{\omega^2\Delta^2}\log(2d).$$

The (b) holds as $\Delta_{m,i} \geq \Delta$. The claim of the lemma follows. □

Third Stage for FB-DOE

Now we apply the G-optimal design to the rotated arm set.

Lemma A.39. *Assume that the best arm i_m^* is not eliminated before phase ℓ, i.e., $i_m^* \in \mathcal{G}_{m,\ell-1}$. Then the probability that the best arm is eliminated in phase ℓ is bounded as*

$$\mathbb{P}\left(i_m^* \notin \mathcal{G}_{m,\ell} \mid i_m^* \in \mathcal{G}_{m,\ell-1}\right) \leq \begin{cases} \frac{4A}{(k_1+k_2)r}\exp\left(\frac{cL_x^4}{(k_1+k_2)rS_r^2\omega^2}\right)\exp\left(-\frac{n_m\Delta_{m,i_\ell}^2}{32 i_{m,\ell}} + \log\log(1+n/3\lambda)\right) & \textit{when } \ell = 1 \\ 3\exp\left(\frac{cL_x^4}{(k_1+k_2)rS_r^2\omega^2}\right)\exp\left(-\frac{n_m\Delta_{m,i_\ell}^2}{32 i_{m,\ell}} + \log\log(1+n/3\lambda)\right) & \textit{when } \ell > 1 \end{cases}$$

where $i_{m,\ell} = \left\lceil \frac{(k_1+k_2)r}{2^{\ell+1}} \right\rceil + 1$.

Proof. We use the same proof technique as for the linear budget fixed arm setting in Lemma A.17. Note that we use the rotated arm set of dimension $(k_1+k_2)r$. Additionally observe that the latent feature estimation error factor $\exp\left(\frac{cL_x^4}{(k_1+k_2)rS_r^2\omega^2}\right)$ that shows up in the bound. □

We prove the main theorem for bilinear bandits.

Theorem 3. *Define* $\Delta = \min_m \min_{i\in\mathcal{X}} \Delta_{m,i}$, $Mn \geq \lceil \frac{(d_1d_2)^2(k_1k_2)^2c'(\rho^E)^2\log^2(2d_1d_2)}{\omega^2S_r^2\Delta^2} \rceil$ *and* $\ell > 1$. *Set* $\lambda_\perp = \frac{n}{3k_1k_2\log\left(1+\frac{n}{3\lambda}\right)}$ *and* $\lambda > 0$ *in* $\mathbf{\Lambda}_{m,\ell}$ *for each task m. Then the total probability of error of the algorithm is given by*

$$\begin{aligned}
8\exp\left(\frac{-Mn}{\log_2 d_1d_2}\right) + CM\exp\left(\frac{cL_x^4}{k_1k_2S_r^2\omega^2}\right)\exp\left(\frac{-n}{2\log_2(k_1+k_2)}\right) \\
+ M\left(6\log_2(k_1+k_2)r\right)\exp\left(\frac{cL_x^4}{(k_1+k_2)rS_r^2\omega^2}\right)\exp\left(-\frac{n_m}{32H_{2,\textit{bilin}}}\right).
\end{aligned}$$

Proof. **Stage 1:** Using Lemma A.22 we can show that the probability of error in the first stage is bounded by

$$8d_1d_2\exp\left(-Mn\right) \leq 8\exp\left(\frac{-Mn}{\log_2 d_1d_2}\right).$$

Also we have show in Lemma A.10 that if the good event $\mathcal{F}_n$ holds, then we get a valid G-optimal design and $\left\|(\widehat{\mathbf{B}}_n^{\perp})^\top \mathbf{B}\right\| \leq c$ for some constant c for $Mn \geq \lceil \frac{(d_1 d_2)^2 (k_1 k_2)^2 c' (\rho^E)^2 \log^2(2d_1 d_2)}{\omega^2 S_r^2 \Delta^2} \rceil$.

Stage 2: Using Lemma A.31 we can show that the probability of error in the second stage is bounded by

$$CM\,(k_1+k_2)^{3/2} \exp\left(\frac{cL_x^4}{k_1 k_2 S_r^2 \omega^2}\right) \exp\left(\frac{-n}{2}\right)$$

Also we have show in Lemma A.36 that if the good event $\mathcal{F}_n'$ holds, then we get a valid G-optimal design for the third stage.

Stage 3: Assume that $\mathcal{F}_n \cap \mathcal{F}_n'$ holds. First note that by rotation of the arms we have reduced the effective dimension to $\widetilde{k} = (k_1+k_2)r$. By applying Lemma A.14 and Lemma A.17, we have for $\ell > 1$

$$\begin{aligned}
\mathbb{P}\left(\widehat{i^*}_m \neq i_m^*\right) &= \mathbb{P}\left[i_m^* \notin \mathcal{G}_{m,\lceil \log_2(k_1+k_2)r\rceil}\right] \\
&\leq \sum_{\ell=1}^{\lceil \log_2(k_1+k_2)r\rceil} \mathbb{P}\left[i_m^* \notin \mathcal{G}_{m,\ell} \mid i_m^* \in \mathcal{G}_{m,\ell-1}\right] \\
&\leq \sum_{\ell=2}^{\lceil \log_2(k_1+k_2)r\rceil} 3\exp\left(\frac{cL_x^4}{(k_1+k_2)rS_r^2\omega^2}\right)\exp\left(-\frac{n_m \Delta_{m,i_\ell}^2}{32 i_{m,\ell}} + \log\log(1+n/3\lambda)\right) \\
&\leq \left(3\left(\lceil \log_2(k_1+k_2)r\rceil - 1\right)\right)\exp\left(\frac{cL_x^4}{(k_1+k_2)rS_r^2\omega^2}\right) \cdot \\
&\qquad \exp\left(-\frac{n_m}{32}\cdot\frac{1}{\max_{2\leq i\leq(k_1+k_2)r}\frac{i}{\Delta_{m,i}^2}} + \log\log(1+n/3\lambda)\right) \\
&< \left(3\log_2(k_1+k_2)r\right)\exp\left(\frac{cL_x^4}{(k_1+k_2)rS_r^2\omega^2}\right)\exp\left(-\frac{n_m}{32 H_{2,\text{ bilin}}} + \log\log(1+n/3\lambda)\right)
\end{aligned}$$

where $H_{2,\text{ bilin}}$ is defined as

$$H_{2,\text{ bilin}} = \max_{m\in[M]} \max_{2\leq i\leq (k_1+k_2)r} \frac{i}{\Delta_{m,i}^2}.$$

Note that this is for a single task m. So the total probability of error in stage 3 is given by

$$M\,\left(3\log_2(k_1+k_2)r\right)\exp\left(\frac{cL_x^4}{(k_1+k_2)rS_r^2\omega^2}\right)\exp\left(-\frac{n_m}{32 H_{2,\text{ bilin}}} + \log\log(1+n/3\lambda)\right)$$

Combining stages 1, 2, and 3, and substituting the value of n_m (and ignoring the $\log\log$ factor) we get that the total probability of error is given by

$$\begin{aligned}
\sum_m \mathbb{P}\left(\widehat{i^*}_m \neq i^*_m\right) &\leq 8\exp\left(\frac{-Mn}{\log_2 d_1 d_2}\right) + CM\left(k_1+k_2\right)^{3/2}\exp\left(\frac{cL_x^4}{k_1k_2S_r^2\omega^2}\right)\exp\left(\frac{-n}{2}\right)\\
&\quad + M\left(3\log_2(k_1+k_2)r\right)\exp\left(\frac{cL_x^4}{(k_1+k_2)rS_r^2\omega^2}\right)\exp\left(-\frac{n}{32H_{2,\,\text{bilin}}}\right)\\
&\overset{(a)}{\leq} 8\exp\left(\frac{-Mn}{\log_2 d_1 d_2}\right) + CM\left(k_1+k_2\right)\exp\left(\frac{cL_x^4}{k_1k_2S_r^2\omega^2}\right)\exp\left(\frac{-n}{2}\right)\\
&\quad + M\left(6\log_2(k_1+k_2)r\right)\exp\left(\frac{cL_x^4}{(k_1+k_2)rS_r^2\omega^2}\right)\exp\left(-\frac{n}{32H_{2,\,\text{bilin}}}\right)\\
&\overset{(b)}{\leq} 8\exp\left(\frac{-Mn}{\log_2 d_1 d_2}\right) + CM\exp\left(\frac{cL_x^4}{k_1k_2S_r^2\omega^2}\right)\exp\left(\frac{-n}{2\log_2(k_1+k_2)}\right)\\
&\quad + M\left(6\log_2(k_1+k_2)r\right)\exp\left(\frac{cL_x^4}{(k_1+k_2)rS_r^2\omega^2}\right)\exp\left(-\frac{n}{32H_{2,\,\text{bilin}}}\right) \qquad (29)
\end{aligned}$$

where, (a) follows as

$$\begin{aligned}
&\sqrt{(k_1+k_2)}\exp\left(-\frac{n}{2}\right) \leq \exp\left(-\frac{n}{32H_{2,\,\text{bilin}}}\right)\\
&\qquad \implies \exp\left(-\frac{n}{2}+\frac{3}{2}\log(k_1+k_2)\right) \leq \exp\left(-\frac{n}{32H_{2,\,\text{bilin}}}\right)
\end{aligned}$$

for $Mn \geq \frac{(d_1d_2)^2(k_1k_2)^2c'(\rho^E)^2\log^2(2d_1d_2)}{\omega^2S_r^2\Delta^2}$. The (b) follows as

$$\exp\left(-\frac{n}{2}+\log(k_1+k_2)\right) \leq \exp\left(-\frac{n}{2\log_2(k_1+k_2)}\right)$$

The claim of the theorem follows. □

Remark A.40. **(Discussion on Bound)** Observe that the probability of error depends on budget n, ambient dimension d_1, d_2, latent dimension k_1, k_2 and bilinear hardness parameter $H_{2,\,\text{bilin}}$. The $H_{2,\,\text{lin}}$ quantifies the difficulty of identifying the best arm in the bilinear bandit MTRL setting. Observe that the final probability of error in (29) consist of three terms. The first term is the probability of error in estimation of the feature extractors $\mathbf{B}_1, \mathbf{B}_2$. The second term is the error in the estimation of the hidden parameter $S_{m,*}$ in each task m. Additionally, the factor $\exp\left(\frac{cL_x^4}{k_1k_2S_r^2\omega^2}\right)$ captures the error in estimating latent features in second stage. The third term consist of the probability of error in identifying the pair of best arms in each task. Again, the factor $\exp\left(\frac{cL_x^4}{(k_1+k_2)rS_r^2\omega^2}\right)$ captures the error in estimating latent features in third stage. Finally, note that the $\log\log(1+n/3\lambda)$ term in the third factor is much smaller that $-\frac{n_m}{32H_{2,\,\text{bilin}}}$ and so can be effectively ignored.

Note that our key technical challenge in the fixed budget MTRL bilinear setting lies in carefully constructing the high confidence bounds that is exponentially decaying with budget n. In the stage 1 using Lemma A.22 we have to again modify Lemma C.3 of (Du et al., 2023) for the bilinear setting so that we get the exponentially decaying bound. This leads to a new estimation of the feature extractors $\mathbf{B}_1$, $\mathbf{B}_2$ in Lemma A.24, Lemma A.25, and then for a sufficiently large $Mn > \lceil\frac{(d_1d_2)^2(k_1k_2)^2c'(\rho^E)^2\log^2(2d_1d_2)}{\omega^2S_r^2\Delta^2}\rceil$ we have a non-vacuous solution to the E-optimal design in stage 2 (see Lemma A.26). Then we ensure in Lemma A.27 that the latent feature estimation is low and in Lemma A.28 we ensure that the estimation error with the latent feature is low in stage 2. This requires a different analysis than similar art in Du et al. (2023); Yang et al. (2020; 2022) as they only study fixed confidence or regret minimization setting. In the second stage our technical novelty lies in controlling the probability of error for the noisy latent features in low dimensional multi-task linear bandits. This is shown in Lemma A.14, Lemma A.16, and Lemma A.17.

In the second stage we also have to estimate the latent parameter $\mathbf{S}_{m,*}$ for each task m, and incorporate the noisy latent features into this. This requires a different approach than prior bilinear bandit proofs in Jun et al. (2019); Lu et al. (2021). We show this in Lemma A.31. Again we ensure for the third stage that the latent feature estimation is low (after rotation of arms) in Lemma A.37 and in Lemma A.38 we ensure that the estimation error with the latent feature is low in stage 3. Note that this approach differs from the existing art of fixed budget linear bandit settings (Katz-Samuels et al., 2020; Yang & Tan, 2021; Azizi et al., 2022) and significantly different than the fixed confidence linear bandit proofs proofs in (Soare et al., 2014; Mason et al., 2021; Degenne & Koolen, 2019).

A.5 Additional Experimental Details

MTRL linear bandit setting: This experiment consists of a set of $M \in \{5, 10, 15, 20, 30, 40\}$ tasks. We first fix the total number of tasks M from $\{5, 10, 15, 20, 25, 40\}$. Then in each of these tasks for this particular setting (for a particular M) the arm set $\mathcal{X}$ is selected from a unit ball in $\mathbb{R}^8$, and $\|\mathbf{x}\| \leq 1$ for all $\mathbf{x} \in \mathcal{X}$. So the dimension is $d = 8$. Then we choose a random common feature extractor $\mathbf{B} \in \mathbb{R}^{8\times 2}$. So $k = 2$. Then we choose a $\mathbf{w}_m \in \mathbb{R}^k$ for $m = 1, 2, \ldots, M$. This gives us the $\boldsymbol{\theta}_{*,m}$ for each task $m \in [M]$. We set $n = 5000$. We compare against OD-LinBAI (Yang & Tan, 2021) which was shown to be minimax optimal and performs better than PEACE in (Fiez et al., 2019). The OD-LinBAI treats the setting for each $M \in \{5, 10, 15, 20, 25, 40\}$ as a d dimensional linear bandit and suffers a probability of error that scales as $\widetilde{O}(M \exp(-n\Delta^2)/d \log_2 d)$.

MTRL bilinear bandit setting: This experiment consists of a set of $M \in \{30, 60, 90, 120, 150\}$ tasks. Then in each of these tasks for this particular setting (for a particular M) the left arm set $\mathcal{X}$ and the right arm set $\mathcal{Z}$ are selected from a unit ball in $\mathbb{R}^8$. Note that we ensure $\|\mathbf{x}\| \leq 1$ and $\|\mathbf{z}\| \leq 1$ for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{z} \in \mathcal{Z}$. So the dimension is $d_1 = d_2 = 8$. Then we choose random common feature extractors $\mathbf{B}_1 \in \mathbb{R}^{8\times 2}$, $\mathbf{B}_2 \in \mathbb{R}^{8\times 2}$. So $k_1 = k_2 = 2$. Then we choose a $\mathbf{S}_{m,*} \in \mathbb{R}^{k_1 \times k_2}$ for $m = 1, 2, \ldots, M$. This gives us the $\boldsymbol{\Theta}_{*,m}$ for each task $m \in [M]$. We set $n = 8000$. Again we compare against OD-LinBAI (Yang & Tan, 2021) as there are no fixed budget alternatives for the bilinear bandit setting. The OD-LinBAI treats the setting for each $\{20, 30, 40, 60, 80, 100\}$ as a $d_1 d_2$ dimensional linear bandit and suffers a probability of error that scales as $\widetilde{O}(M \exp(-n\Delta^2/d_1 d_2 \log_2 d_1 d_2))$.

MTRL linear Nectar setting: This is a real-world semi-synthetic experiment on the Nectar dataset (Zhu et al., 2023). This dataset consists of 100K prompts, where each prompt consists of 7 answers by Large Language models which are then ranked by humans. We select 20 prompts randomly from this dataset and obtain a 768 dimensional embedding using Instructor model (Su et al., 2022) which we denote as $\mathbf{q} \in \mathbb{R}^{768}$. Then we project this vector to $\mathbb{R}^6$ using a projection matrix. For each prompt, we also obtain a 768 dimensional embedding for each of the 7 answers and we denote this as $\mathbf{a} \in \mathbb{R}^{768}$. Then again we project this vector to $\mathbb{R}^6$ using a projection matrix. Finally, we obtain an arm $\mathbf{x} = \mathrm{vec}(\mathbf{q}\mathbf{a}^\top) \in \mathbb{R}^{36}$ and $d = 36$. So these 140 arms constitute the $\mathcal{X}$. Next, we fit the model $\boldsymbol{\theta}_*$ based on the original ranking in the dataset to these arms. Then for each task $m \in [M]$ we perturb the $\boldsymbol{\theta}_* + \epsilon$ with an $\epsilon \sim \mathcal{N}(0, 0.05 * I_d)$ to obtain $\boldsymbol{\theta}_{m,*}$.

Then in this experiment, we consider a set of $M \in \{20, 30, 40, 60, 80, 100\}$ tasks. We again first fix the total number of tasks M from $M \in \{20, 30, 40, 60, 80, 100\}$. Then in each of these tasks for this particular setting (for a particular M) the arm set $\mathcal{X}$ is selected as above. Then we choose a random common feature extractor $\mathbf{B} \in \mathbb{R}^{8\times 2}$. So $k = 2$. Then we choose a $\mathbf{w}_m \in \mathbb{R}^k$ for $m = 1, 2, \ldots, M$ such that $\mathbf{w}_m = \mathbf{B}^{-1}\boldsymbol{\theta}_{m,*}$. We set $n = 5000$. Again we compare against OD-LinBAI. The OD-LinBAI treats the setting for each $M \in \{20, 30, 40, 60, 80, 100\}$ as a d dimensional linear bandit and suffers a probability of error that scales as $\widetilde{O}(M \exp(-n\Delta^2)/d \log_2 d)$.

B Table of Notations

Notations	**Definition**
M	Number of tasks
$\mathcal{X}$	Left arm set
$\mathcal{Z}$	Right arm set
$\boldsymbol{\theta}_{m,*}$	Hidden parameter for linear bandit in ambient dimension
$\mathbf{w}_m$	Hidden low dimensional parameter for linear bandit
ℓ	Phase number
$\boldsymbol{\Theta}_{m,*}$	Hidden parameter matrix for bilinear bandits in ambient dimension
$\mathbf{S}_{m,*}$	Hidden low dimensional parameter matrix for bilinear bandits
$\mathbf{b}_{\mathbf{x}}^{E}$	E-optimal design
$\mathbf{b}_{m,\ell}^{G}$	G-optimal design at the ℓ-th phase for the m-th task
$\lambda_m^{\perp}$	$n/24(k_1+k_2)r\log(1+\frac{n}{3\lambda})$
$\mathbf{B}_1$	Left feature extractor
$\mathbf{B}_2$	Right feature extractor
S_r	r-th largest singular value of $\boldsymbol{\Theta}_*$
$\Delta = \min_m \min_{i\in\mathcal{X}} \Delta_{m,i}$	Linear bandit minimum gap
$H_{1,\text{lin}} = \min_{m\in[M]} \sum_{i=1}^{k} \frac{1}{\Delta_{m,i}^2}$	Linear bandit hardness parameter
$H_{2,\text{ lin}} = \max_{m\in[M]} \max_{2\leq i\leq k} \frac{i}{\Delta_{m,i}^2}$.	Linear bandit hardness parameter
$\Delta = \min_m \min_{i\in\mathcal{X},\mathcal{Z}} \Delta_{m,i}$	Bilinear bandit minimum gap
$H_{2,\text{ bilin}} = \max_{m\in[M]} \max_{2\leq i\leq (k_1+k_2)r} \frac{i}{\Delta_{m,i}^2}$.	Bilinear bandit hardness parameter
n	Total budget

Table 1: Table of Notations