Asymptotically Optimal Linear Best Feasible Arm Identification with Fixed Budget

Jie Bian[1] **Vincent Y. F. Tan**[2,1]

[1]Department of Electrical and Computer Engineering, National University of Singapore, Singapore
[2]Department of Mathematics, National University of Singapore, Singapore

Abstract

The challenge of identifying the best feasible arm within a fixed budget has attracted considerable interest in recent years. However, a notable gap remains in the literature: the exact exponential rate at which the error probability approaches zero has yet to be established, even in the relatively simple setting of K-armed bandits with Gaussian noise. In this paper, we address this gap by examining the problem within the context of linear bandits. We introduce a novel algorithm for best feasible arm identification that guarantees an exponential decay in the error probability. Remarkably, the decay rate—characterized by the exponent—matches the theoretical lower bound derived using information-theoretic principles. Our approach leverages a posterior sampling framework embedded within a game-based sampling rule involving a min-learner and a max-learner. This strategy shares its foundations with Thompson sampling, but is specifically tailored to optimize the identification process under fixed-budget constraints. Furthermore, we validate the effectiveness of our algorithm through comprehensive empirical evaluations across various problem instances with different levels of complexity. The results corroborate our theoretical findings and demonstrate that our method outperforms several benchmark algorithms in terms of both accuracy and efficiency.

1 INTRODUCTION

Many real-world decision-making systems, such as recommendation platforms, clinical trials, and autonomous agents, must make selections under feasibility constraints that reflect resource limitations, safety thresholds, or business requirements. For instance, a system may wish to recommend actions that maximize utility while adhering to pricing limits, fairness constraints, or operational risks. These constraints are often modeled as linear inequalities over contextual features, with parameters—such as cost or risk scores—that are unknown and must be inferred from data. Moreover, these estimates are inherently noisy due to model approximation, user behavior variability, and environmental uncertainty. We formalize this as follows: an arm $z \in \mathbb{R}^d$ is considered feasible if its expected cost satisfies $\langle \theta^c, z \rangle \leq \tau$, where $\theta^c \in \mathbb{R}^d$ is an unknown cost vector and $\tau \in \mathbb{R}$ is a known threshold. The goal is to identify the best feasible arm—the one with the highest expected reward among all feasible options—based on noisy observations of both reward and cost. This setting captures a wide range of constrained decision-making problems and offers a tractable framework for inference and optimization. It can be viewed as a constrained variant of the best arm identification (BAI) problem in multi-armed bandits (MAB). In classical BAI, the objective is to identify the arm with the highest expected reward under a fixed budget or confidence. However, feasibility constraints introduce new challenges, especially under a fixed sampling budget where confidence intervals cannot be reliably maintained. This motivates the development of new algorithms that jointly reason over reward and constraint uncertainty within limited resources.

The MAB problem is a central challenge in machine learning and statistical decision theory (Lattimore and Szepesvári [2020]). Here, a decision-maker, often referred to as the player, repeatedly selects one of several available arms. Each arm corresponds to a distinct action, and upon selection, the player receives a reward drawn from an unknown probability distribution associated with that arm.

The MAB problem can be categorized into two main settings: regret minimization and best arm identification. In regret minimization, the goal is to maximize cumulative reward, or equivalently, minimize cumulative regret, defined as the gap between the algorithm's cumulative reward and that of always pulling the best arm at each time step. In contrast, best arm identification focuses on maximizing the probability of correctly identifying the arm with the

highest expected reward (Maron and Moore [1997]), either within a predetermined budget (Gabillon et al. [2012], Komiyama et al. [2022]) or with a specified level of confidence (Kalyanakrishnan et al. [2012], Kuroki et al. [2020]).

MAB problems have broad applications, including online advertising, clinical trials, and recommendation systems (Kuleshov and Precup [2014], Bouneffouf and Rish [2019]). Despite their utility, these problems present key challenges, such as managing the exploration-exploitation trade-off, scalability across large action spaces, and incorporating contextual information.

1.1 REGRET MINIMIZATION

Regret minimization focuses on reducing the cumulative difference between the rewards obtained by the algorithm and those of the optimal strategy, requiring a careful balance between exploration and exploitation. In the classic K-armed bandit problem, algorithms such as UCB1 Auer et al. [2002] and Thompson Sampling Thompson [1933] are widely used. UCB1 selects arms based on an upper confidence bound that prioritizes arms with higher uncertainty until sufficient data is gathered, while Thompson Sampling maintains a posterior distribution over rewards and selects arms probabilistically based on reward samples. In the linear bandit setting, where rewards are linear functions of a known context vector with unknown parameters, LinUCB Abbasi-Yadkori et al. [2011] extends UCB1 by incorporating parameter uncertainty into the confidence bounds, and LinTS Agrawal and Goyal [2013] applies a similar approach by sampling from the posterior distribution of the linear parameters to select arms.

1.2 FIXED BUDGET BEST ARM IDENTIFICATION

The BAI problem seeks to identify the arm with the highest expected reward, either under a fixed budget of arm pulls or by minimizing the number of trials required to achieve a specified level of confidence. Unlike regret minimization, this objective places a greater emphasis on exploration to ensure reliable identification of the optimal arm.

In the finite fixed budget setting, two algorithms are particularly noteworthy: Successive Rejects and Sequential Halving. The Successive Rejects algorithm Audibert and Bubeck [2010] iteratively eliminates the arm with the lowest empirical mean reward after allocating a predefined number of trials to all arms in each phase. This process continues until only the best arm remains. The Sequential Halving algorithm Karnin et al. [2013] adopts a similar phase-based approach but eliminates half of the remaining arms with the lowest observed rewards after each round. Subsequent rounds concentrate trials on the more promising arms, improving the efficiency of arm elimination.

Within the infinite fixed-budget setting, the primary objective of algorithms is to asymptotically achieve an exponential rate for the upper bound on regret as the time horizon T approaches infinity. Algorithms in this category often prioritize efficient exploration-exploitation trade-offs to maximize long-term performance. One notable example is the Top-Two Thompson Sampling (TTTS) algorithm, introduced by Russo [2016]. The Top-Two Thompson Sampling algorithm introduces a critical tuning parameter, denoted by β. Achieving optimal rate of posterior convergence under this algorithm requires that β precisely match the optimal allocation rate associated with the best arm. Only under this condition does the algorithm attain the optimal rate of posterior convergence. This dependency underscores the importance of parameter selection and adaptive tuning mechanisms for Bayesian algorithms operating in infinite horizon settings. Recently, new algorithm (Li et al. [2024]) called Pure Exploration with Projection-Free Sampling (PEPS) was developed; this algorithm does not need tuning of β and achieves the optimal rate of posterior convergence when the parameter set is convex and bounded (Kone et al. [2024]).

1.3 RELATED WORK

The literature on MABs has evolved significantly across three key dimensions: (1) unconstrained vs. constrained settings, (2) regret minimization vs. best arm identification objectives, and (3) fixed-confidence vs. fixed-budget paradigms. Our work bridges a critical gap in constrained BAI under fixed budgets by proposing a tuning-free algorithm with matching asymptotic bounds. Below, we contextualize our contributions within these axes of research.

Unconstrained Best Arm Identification. Early work in linear bandits focused on unconstrained settings. Recent breakthroughs, such as the min-max game framework by Li et al. [2024], achieve minimax optimal error rates without tuning parameters. For structured bandits, Azizi et al. [2021] proposed an early fixed-budget method for linear BAI, though their theoretical guarantees fall short of tightness. Yang and Tan [2022] offered a more refined analysis, but their upper bound retains a constant suboptimality gap compared to the lower bound. These advances, however, do not account for resource constraints.

Constrained Bandits: Regret Minimization. In constrained regret minimization, confidence intervals for rewards and costs drive algorithm design. Works such as Amani et al. [2019], Moradipari et al. [2021] and kernelized extensions [Zhou and Ji, 2022] ensure sublinear regret. Contextual settings further refine this approach [Pacchiano et al., 2024]. While effective for minimizing cumulative regret, these methods are ill-suited for identifying the optimal feasible arm under fixed budgets.

Constrained Best Arm Identification. Under fixed-confidence settings, safety constraints dominate. Katz-Samuels and Scott [2019], Camilleri et al. [2022] maintain conservative safe arm sets via confidence intervals, while Wang et al. [2022], Shang et al. [2023] address pricing-

specific constraints. These methods prioritize statistical efficiency over budget allocation, limiting their applicability to fixed-budget scenarios. Hou et al. [2023] proposed an optimal BAI algorithm that accounts for variance constraints.

Fixed-Budget Challenges. The fixed-budget setting demands optimal resource allocation. Early efforts in K-armed bandits [Faizal and Nair, 2022] and combinatorial settings [Tang et al., 2024] lack matching error bounds. Recent progress in linear bandits [Yang and Tan, 2022] achieves minimax optimality, yet these ignore constraints entirely. A key unresolved challenge lies in balancing cost-awareness with asymptotic optimality under limited samples.

Infinite-Budget and Parameter Sensitivity. For infinite horizons, Yang et al. [2025] proposed Top-Two Thompson Sampling with β-optimality, requiring careful parameter tuning. Parallel advances in unconstrained settings [Li et al., 2024] eliminate tuning parameters via posterior sampling, but extending this to constrained problems remains open.

Feasibility vs. Optimization. Prior work such as Katz-Samuels and Scott [2018] identifies entire feasible arm sets, diverging from our goal of pinpointing the best feasible arm. Similarly, Pareto set identification [Kone et al., 2024] focuses on multi-objective trade-offs rather than constrained single-objective optimization.

1.4 MAIN CONTRIBUTIONS

Our contributions in this paper are fourfold:

1. We propose a novel algorithm BLFAIPS that achieves matching upper and lower bounds for the best feasible arm identification problem under a fixed budget. This is the first algorithm in this domain to achieve optimality in the exponent of the error probability.
2. We derive two distinct forms of the exponential rate in the lower bound (Theorems 4.1 and 4.2). One form is derived from a Bayesian perspective, while the other arises from a frequentist viewpoint. We further prove the equivalence of these two formulations, thus unifying both theoretical perspectives.
3. We improve the structure introduced in Li et al. [2024] by replacing the exponential weights algorithm with AdaHedge and eliminating the need for the doubling trick. Additionally, we relax the assumption that the reward is bounded, which was a requirement in their theoretical guarantees and proofs.
4. We conduct empirical studies on synthetic and real-world datasets to evaluate the performance of our algorithm against several baseline methods on representative problem instances. Our results demonstrate that the proposed algorithm consistently outperforms or remains competitive with other state-of-the-art algorithms across different scenarios.

2 PROBLEM SETUP

We consider the BAI problem for linear bandits in the fixed budget setting with a budget T. In contrast to the conventional setting in which there are no constraints, we impose a constraint on the permissible arm to be selected at the end of the horizon T. The precise problem setting is as follows. There are two finite sets of arms – a *training set* $\mathcal{X} \subset \mathbb{R}^d$ and a *testing set* $\mathcal{Z} \subset \mathbb{R}^d$. At each time $t \in [T]$, based on the learner's past selections and their corresponding outcomes, the learner chooses an arm $X_t \in \mathcal{X}$ and the environment reveals noisy versions of a reward (r) and a cost (c) which are respectively denoted as

$$Y_t^{\mathrm{r}} = \langle \theta^{\mathrm{r}}, X_t \rangle + \epsilon_t, \quad \text{and} \tag{1}$$

$$Y_t^{\mathrm{c}} = \langle \theta^{\mathrm{c}}, X_t \rangle + \eta_t, \tag{2}$$

where ε_t and η_t are zero mean independent Gaussian noises with variances σ^2 and γ^2 respectively. In Eqns. (1) and (2), $\theta^{\mathrm{r}} \in \Theta^{\mathrm{r}}$ and $\theta^{\mathrm{c}} \in \Theta^{\mathrm{c}}$ are *unknown* d-dimensional vectors. The decision X_t at each time t is based on the history $\mathcal{H}_{t-1} := \{(X_s, Y_s^{\mathrm{r}}, Y_s^{\mathrm{c}})\}_{s=1}^{t-1}$. Our goal is to select the arm that maximizes the mean reward subject to a constraint on the cost, measured according to the test vectors in $\mathcal{Z}$. More precisely, we denote $\mu^{\mathrm{r}}(z) := \langle \theta^{\mathrm{r}}, z \rangle$ and $\mu^{\mathrm{c}}(z) := \langle \theta^{\mathrm{c}}, z \rangle$ as the expected reward and cost of the testing arm vector $z \in \mathcal{Z}$. We define the *best feasible arm as*

$$z^* := \underset{z \in \mathcal{Z}: \mu^{\mathrm{c}}(z) \le \tau}{\arg\max} \ \mu^{\mathrm{r}}(z),$$

where $\tau \in \mathbb{R}$ is the known *threshold* on the cost. At the end of the horizon, the learner recommends an arm $z_{\mathrm{out}} \in \mathcal{Z}$ which is its best guess of z^*. For ease of exposition later, we define the following three sets.

- Suboptimal arm set
  $\mathcal{S} := \{z \in \mathcal{Z} : \mu^{\mathrm{r}}(z) \le \langle \theta^{\mathrm{r}}, z^* \rangle\}$;
- Superoptimal arm set
  $\overline{\mathcal{S}} := \{z \in \mathcal{Z} : \mu^{\mathrm{r}}(z) > \langle \theta^{\mathrm{r}}, z^* \rangle\}$;
- Feasible arm set
  $\mathcal{F} := \{z \in \mathcal{Z} : \mu^{\mathrm{c}}(z) \le \tau\}$;
- Infeasible arm set
  $\overline{\mathcal{F}} := \{z \in \mathcal{Z} : \mu^{\mathrm{c}}(z) > \tau\}$

For convenient, denote $\mathcal{A}_1 := \overline{\mathcal{F}} \cap \overline{\mathcal{S}}, \mathcal{A}_2 := \mathcal{F} \cap \mathcal{S}, \mathcal{A}_3 := \overline{\mathcal{F}} \cap \mathcal{S}$. For any $z \in \mathcal{Z}$, define $\overline{\Theta}_z \subset \Theta^{\mathrm{r}} \times \Theta^{\mathrm{c}}$ such that for any $(\theta_1, \theta_2) \in \overline{\Theta}_z$, the arm z is not the best feasible arm while $\theta^{\mathrm{r}} = \theta_1, \theta^{\mathrm{c}} = \theta_2$.

Furthermore we have the following assumptions:

Assumption 2.1. $\Theta^{\mathrm{r}}, \Theta^{\mathrm{c}}$ are both bounded and closed with non-empty interiors such that $\max_{\theta \in \Theta^{\mathrm{r}}} \|\theta\|_2 \le R_1, \max_{\theta \in \Theta^{\mathrm{c}}} \|\theta\|_2 \le R_2$.

Assumption 2.2. The training and testing sets are bounded such that $\max_{x \in \mathcal{X}} \|x\|_2 \le L_1, \max_{z \in \mathcal{Z}} \|z\|_2 \le L_2$.

Assumption 2.3. The best feasible arm z_1 is unique and $\mathrm{span}(\mathcal{Z}) \subset \mathrm{span}(\mathcal{X})$.

Notations: We define $\|x\|_A := \sqrt{x^\top A x}$ for any d-dimensional column vector x and positive definite $d \times d$ matrix A. Given a finite set $\mathcal{X} \subset \mathbb{R}^d$, we define the probability simplex as $\Delta_{\mathcal{X}} := \left\{\lambda \in \mathbb{R}_{\geq 0}^{|\mathcal{X}|} : \sum_{i=1}^{|\mathcal{X}|} \lambda_i = 1\right\}$. For any $\lambda \in \Delta_{\mathcal{X}}$, we define the matrix $A(\lambda)$ as $A(\lambda) := \sum_{x \in \mathcal{X}} \lambda_x x x^\top$. Denote two sequences of positive real numbers $\{a_n\}$ and $\{b_n\}$ to be equal to first order in the exponent as $a_n \doteq b_n$ if $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$.

3 BEST LINEAR FEASIBLE ARM IDENTIFICATION WITH POSTERIOR SAMPLING (BLFAIPS)

We introduce a novel approach for linear feasible arm identification utilizing posterior sampling. This algorithm, called BLFAIPS (and whose pseudocode is in Algorithm 1), employs posterior sampling for the min-learner and AdaHedge for the max-learner. We call them *min-learner* and *max-learner* since by Sion's minimax theorem, the exponential rate Γ in the lower bound, as we will see in Theorem 4.2, can be represented by:

$$\Gamma := \max_{w \in \Delta_{\mathcal{X}}} \inf_{(\theta_1,\theta_2) \in \overline{\Theta}_{z^*}} \frac{1}{2}\left(\frac{\|\theta_1 - \theta^{\mathrm{r}}\|_{A(w)}^2}{\sigma^2} + \frac{\|\theta_2 - \theta^{\mathrm{c}}\|_{A(w)}^2}{\gamma^2}\right)$$
$$= \max_{w \in \Delta_{\mathcal{X}}} \min_{p \in \Delta(\overline{\Theta}_{z^*})} \frac{1}{2}\mathbb{E}\left[\frac{\|\theta_1 - \theta^{\mathrm{r}}\|_{A(w)}^2}{\sigma^2} + \frac{\|\theta_2 - \theta^{\mathrm{c}}\|_{A(w)}^2}{\gamma^2}\right]$$
$$= \min_{p \in \Delta(\overline{\Theta}_{z^*})} \max_{w \in \Delta_{\mathcal{X}}} \frac{1}{2}\mathbb{E}\left[\frac{\|\theta_1 - \theta^{\mathrm{r}}\|_{A(w)}^2}{\sigma^2} + \frac{\|\theta_2 - \theta^{\mathrm{c}}\|_{A(w)}^2}{\gamma^2}\right],$$

where the expectations above are with respect to $(\theta_1, \theta_2) \sim p$. The quantity Γ captures the hardness of the best feasible arm identification problem—it characterizes the exponential rate at which the optimal arm can be distinguished from all other (possibly infeasible) arms under reward and cost uncertainty. To match this fundamental limit, our algorithm is designed to mirror the structure of this bound: the max-learner adaptively selects sampling distributions to maximize distinguishability, while the min-learner simulates adversarial parameters that minimize it. Since both components achieve sublinear regret over time (though our focus is not on regret), the overall algorithm converges to the optimal error exponent in the long run. A comprehensive description of the proposed algorithm is provided below.

Initialization In Line 1, we define the exploration distribution using the G-optimal design, which is specifically designed to determine the weights of the arms by minimizing the maximum confidence over the arms. This design ensures that the exploration process effectively gathers sufficient information about the unknown linear coefficients, θ^{r} and θ^{c}. In Lines 2-3, we initialize the weight for the AdaHedge algorithm with a uniform distribution, set the covariance matrix to the identity matrix, and initialize S_0 as

Algorithm 1 Best Linear Feasible arm identification with Posterior Sampling (BLFAIPS)

Require: Finite set of arms $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{Z} \subset \mathbb{R}^d$, time horizon T, posterior reward and cost distribution variance parameter $\eta_{\mathrm{r}}, \eta_{\mathrm{c}}$, shrinking rate of exploration $\alpha = \frac{1}{4}$, AdaHedge initial learning rate $\eta_1 := \infty$, cost threshold constant $\tau \in \mathbb{R}$, initial cumulative mixability gap $\Delta_0 = 0$.

1: **Define** $\lambda^G = \arg\min_{\lambda \in \Delta_{\mathcal{X}}} \max_{x \in \mathcal{X}} \|x\|_{\mathcal{A}(\lambda)^{-1}}^2$
2: $\lambda_1 = \frac{1}{|\mathcal{X}|}\mathbf{1}$
3: Initialize $V_0 = I, S_0 = 0, p_1 = \mathcal{N}(0, V_0), \hat{\theta}_1$ arbitrarily
4: **for** $t = 1, 2, \ldots, T$ **do**
5: $\quad \gamma_t = t^{-\alpha}$ ▷ Posterior Sampling
6: $\quad \hat{\mathcal{F}} := \{z \in \mathcal{Z} \mid z^\top \hat{\theta}_t^{\mathrm{c}} \leq \tau\}$
7: $\quad$ **if** $|\hat{\mathcal{F}}| \neq \emptyset$ **then**
8: $\quad\quad$ Compute $\hat{z}_t = \arg\max_{z \in \hat{\mathcal{F}}} z^\top \hat{\theta}_t^{\mathrm{r}}$
9: $\quad$ **else**
10: $\quad\quad$ Uniformly sample $\hat{z}_t$ from $\mathcal{Z}$
11: $\quad$ **end if**
12: $\quad$ Sample $(\theta_t^{\mathrm{r}}, \theta_t^{\mathrm{c}}) \sim \mathcal{N}(\hat{\theta}_t^{\mathrm{r}}, \eta_r^{-1} V_{t-1}^{-1}) \otimes \mathcal{N}(\hat{\theta}_t^{\mathrm{c}}, \eta_c^{-1} V_{t-1}^{-1}) \mid \overline{\Theta}_{\hat{z}_t}$
13: $\quad$ Sample $X_t \sim \lambda_t$ where $\lambda_t = (1 - \gamma_t)\lambda_t + \gamma_t \lambda^G$
14: $\quad$ Observe $y_t^{\mathrm{r}} = \langle \theta_t^{\mathrm{r}}, X_t \rangle + \epsilon_t$ where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$
15: $\quad$ Observe $y_t^{\mathrm{c}} = \langle \theta_t^{\mathrm{c}}, X_t \rangle + \gamma_t$ where $\gamma_t \sim \mathcal{N}(0, \gamma^2)$ ▷ AdaHedge
16: $\quad$ Receive the loss vector $l_t \in \mathbb{R}^{|\mathcal{X}|}$ where $l_{t,x} := -\left(\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|_{xx^\top}^2}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|_{xx^\top}^2}{\gamma^2}\right)$
17: $\quad$ Compute the Hedge loss $h_t = \sum_{i=1}^{|\mathcal{X}|} \lambda_{t,i} l_{t,i}$
18: $\quad$ Update $V_t = V_{t-1} + X_t X_t^\top$
19: $\quad$ $S_t^{\mathrm{c}} = S_{t-1}^{\mathrm{c}} + x_t y_t^{\mathrm{c}}$, $S_t^{\mathrm{r}} = S_{t-1}^{\mathrm{r}} + x_t y_t^{\mathrm{r}}$
20: $\quad$ $\hat{\theta}_{t+1}^{\mathrm{r}} = V_t^{-1} S_t^{\mathrm{r}}$, $\hat{\theta}_{t+1}^{\mathrm{c}} = V_t^{-1} S_t^{\mathrm{c}}$
21: $\quad$ Update $\lambda_{t+1} \propto \lambda_t e^{-\eta_t l_t}$
22: $\quad$ Compute mixed loss

$$m_t := -\frac{1}{\eta_t} \log\left(\sum_{i \in \mathcal{X}} \lambda_{t,i} e^{-\eta_t l_{t,i}}\right)$$

23: $\quad$ Compute the mixability gap $\delta_t := h_t - m_t$, $\Delta_t = \Delta_{t-1} + \delta_t$
24: $\quad$ Update the AdaHedge learning rate $\eta_{t+1} = \frac{\log(|\mathcal{X}|)}{\Delta_t}$
25: **end for**
26: Sample $(\theta_{T+1}^{\mathrm{r}}, \theta_{T+1}^{\mathrm{c}}) \sim \mathcal{N}(\hat{\theta}_{T+1}^{\mathrm{r}}, \sigma^2 V_T^{-1}) \otimes \mathcal{N}(\hat{\theta}_{T+1}^{\mathrm{c}}, \gamma^2 V_T^{-1}) \mid \Theta$
27: $\hat{\mathcal{F}} \leftarrow \{z \in \mathcal{Z} \mid z^\top \theta^{\mathrm{c}} \leq \tau\}$
28: **if** $|\hat{\mathcal{F}}| \neq \emptyset$ **then**
29: $\quad$ Compute $\hat{z}_t \leftarrow \arg\max_{z \in \hat{\mathcal{F}}} z^\top \hat{\theta}_{T+1}^{\mathrm{r}}$
30: **else**
31: $\quad$ Uniformly sample $\hat{z}_t$ from $\mathcal{Z}$
32: **end if**
33: **Output:** $\hat{z}_{\mathrm{out}} \leftarrow \arg\max_{z \in \hat{\mathcal{F}}} z^\top \theta^{\mathrm{r}}$

the $d \times 1$ zero matrix, following the standard initialization procedure for Ridge regression.

Min-learner From Lines 7 to 16, we perform posterior sampling for the min-learner. In Line 7, we first compute the empirical feasible arm set, and if this set is empty, we randomly select the best empirical feasible arm. If the set is not empty, we proceed in Line 9 to compare the empirical rewards to identify the empirical best feasible arm. In Line 13, we then perform posterior sampling of $(\theta_t^{\mathrm{r}}, \theta_t^{\mathrm{c}})$ from the posterior distribution $\mathcal{N}(\hat{\theta}_t^{\mathrm{r}}, \eta_r^{-1} V_{t-1}^{-1}) \otimes \mathcal{N}(\hat{\theta}_t^{\mathrm{c}}, \eta_c^{-1} V_{t-1}^{-1})$ within the space $\overline{\Theta}_{\hat{z}_t}$, ensuring that $\hat{z}_t$ is not the best feasible arm. For the learning parameters, we choose $\eta_r := \frac{\eta}{\sigma^2}, \eta_c := \frac{\eta}{\gamma^2}$ where $\eta := \min\left\{\frac{\sigma^2}{8L^2R_1^2}, \frac{\gamma^2}{8L^2R_2^2}\right\}$. Following this, in Line 14, we draw the arm under the distribution λ_t, which combines the AdaHedge distribution and the G-optimal distribution, with the proportion of the G-optimal distribution decreasing exponentially over time. Finally, in Lines 15-16, we observe the reward and cost corresponding to the selected arm, which are utilized for updating the ridge estimator in Line 20.

Max-learner From Lines 18 to 24, we employ AdaHedge for the max-learner instead of the traditional exponential weighting algorithm, enabling the learning rate parameter to automatically adapt to the environment without the need for further tuning. Specifically, in Line 18, we compute the loss for each arm x at time step t as

$$l_{t,x} := -\left(\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|_{xx^\top}^2}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|_{xx^\top}^2}{\gamma^2}\right),$$

where the first term represents the loss from the reward and the second term represents the loss from the cost. This choice of loss function is derived from our lower bound analysis in Section 4. Following this, in the subsequent lines, we apply the AdaHedge algorithm to determine the distribution for the next time step and update the necessary parameters to calculate the learning rate for the next round.

Recommendation From Lines 26 to 33, upon accumulating the relevant information, the algorithm recommends the arm that is most likely to be the optimal feasible arm. This procedure adheres to the principles of posterior sampling, closely following the approach outlined in Lines 7 to 13.

Min-max intuition Our algorithm is guided by a game-theoretic interpretation of the information-theoretic lower bound. Intuitively, this corresponds to a zero-sum game between two players: the learner (max-player) selects a sampling distribution over arms to maximize information gain, while the adversary (min-player) chooses alternative parameters that are hardest to distinguish from the true ones. This formulation focuses exploration on the most informative regions of the parameter space. The structure mirrors the lower bound in Theorem 4.2, where the exponent is expressed as a max–min over sampling distributions and perturbations within the feasible set $\overline{\Theta}_{z^*}$. We include a diagram in the appendix to illustrate this interaction and support readers less familiar with game-based bandit formulations.

Novelty of our algorithm Compared to the linear best arm identification algorithm without constraints proposed by Li et al. [2024], our algorithm, BLFAIPS, incorporates several key innovations:

First, we replace the exponential weights algorithm with AdaHedge (Lines 17-24). By utilizing AdaHedge, our algorithm does not require the doubling trick employed in Li et al. [2024] to achieve anytime adaptivity. Unlike standard Hedge-style algorithms that need manually scheduled learning rates or restarts, AdaHedge automatically adapts to the loss sequence, which leads to smoother convergence and more effective use of the fixed sampling budget.

Second, we define a novel loss function (Line 18), derived from our lower bound theorem (Theorem 4.2). This loss function generalizes the one used in Li et al. [2024], allowing for greater flexibility and improved theoretical guarantees within constrained settings.

Finally, in Line 13, our algorithm incorporates constraints directly into the posterior sampling process. This enhancement contrasts with the min-learner approach of Li et al. [2024], enabling our algorithm to better account for feasibility constraints during decision-making.

These improvements collectively enhance both the theoretical and practical performance of our algorithm under resource-constrained scenarios.

4 THEORETICAL GUARANTEES

In this section, we state an information-theoretic lower bound on the error probability. We provide intuitive explanations for each of the terms in the bound.

4.1 LOWER BOUND

Theorem 4.1. *Under the environment $\Theta^{\mathrm{r}} = \Theta^{\mathrm{c}} = \mathbb{R}^d$, let $\Pi_t^{\mathrm{r}} := \mathcal{N}(\hat{\theta}_t^{\mathrm{r}}, \Sigma_t^{\mathrm{r}}), \Pi_t^{\mathrm{c}} := \mathcal{N}(\hat{\theta}_t^{\mathrm{c}}, \Sigma_t^{\mathrm{c}})$ be the posterior distribution of the unknown θ^{r} and θ^{c} at time step t, for any sampling rule where $V_w := \sum_{i=1}^K w_i x_i x_i^\top$ and $w_i := \lim_{n\to\infty} \frac{T_{i,n}}{n}$, where $W := \{w = (w_1, w_2, \ldots, w_K) : \sum_i w_i = 1, w_i \geq 0, \forall\, i \in [K]\}$,*

$$\limsup_{T\to\infty} -\frac{1}{T} \log \mathbb{P}_{\Pi_T^{\mathrm{r}}, \Pi_T^{\mathrm{c}}}(z_{out} \neq z^*) \leq \Gamma$$

where the hardness term is

$$\Gamma := \max_{w\in W} \min_{z\in\mathcal{Z}} \min_{i=1,2,3,4} \{f_i(w, z)\},$$

and

$$f_1(w, z) := \frac{(\Delta^{\mathrm{c}}(z))^2}{2\gamma^2 \|z\|_{V_w^{-1}}^2} \cdot \mathbb{1}\{z \in \mathcal{A}_1\},$$

$$f_2(w,z) := \frac{(\Delta^{\mathrm{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_w^{-1}}} \cdot \mathbb{1}\{z \in \mathcal{A}_2\},$$

$$f_3(w,z) := \left(\frac{(\Delta^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_w^{-1}}} + \frac{(\Delta^{\mathrm{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_w^{-1}}}\right) \cdot \mathbb{1}\{z \in \mathcal{A}_3\},$$

$$f_4(w,z) := \frac{(\Delta^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_w^{-1}}} \cdot \mathbb{1}\{z = z^*\}.$$

The complexity of the linear BAI problem under feasibility constraints is characterized by the hardness parameter Γ. This parameter is composed of four terms, denoted by $f_1(w,z)$, $f_2(w,z)$, $f_3(w,z)$, and $f_4(w,z)$, each reflecting a specific source of difficulty in distinguishing arms based on feasibility and optimality conditions.

The term $f_1(w,z)$ captures the challenge of misidentifying superoptimal and infeasible arms as the best feasible arm. This occurs when arms with superior rewards, but violating feasibility constraints, are incorrectly favored due to insufficient exploration of feasibility conditions. The term $f_2(w,z)$ quantifies the difficulty arising from confusing feasible yet suboptimal arms with the optimal feasible arm. In such cases, the agent may exploit arms that satisfy constraints but do not yield the highest reward due to inadequate exploitation of known reward estimates.

Similarly, $f_3(w,z)$ corresponds to the hardness induced by incorrectly identifying infeasible and suboptimal arms as optimal. This issue arises when infeasibility is not adequately recognized during exploration, leading to wasted budget on arms that neither satisfy constraints nor provide optimal rewards. Finally, $f_4(w,z)$ represents the difficulty associated with mistakenly classifying the true best feasible arm as being infeasible.

Theorem 4.2. *Under the same assumptions as Theorem 4.1, we establish the equivalence between two distinct expressions of the hardness term in the lower bound:*

$$\Gamma := \max_{w\in\Delta_{\mathcal{X}}} \inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}} \frac{1}{2}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{A(w)}}{\sigma^2} + \frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{A(w)}}{\gamma^2}\right)$$

Compared to the equivalence of the two hardness terms without constraint, the difference here is that the KL-divergence in the reward and cost are summarized with the respective rates since the variances of the noise in reward and cost are different. If we set $\tau \to \infty$, then our lower bound particularizes to the hardness term as in the sample complexity of pure exploration for linear bandits (Jedra and Proutiere [2020]). This form of the exponential rate in the lower bound is especially useful, as it aligns with the structure of our upper bound, which is derived in the same form.

4.2 UPPER BOUND

We now state the upper bound on the error probability. Theorem 4.3 establishes that the error probability of our proposed algorithm BLFAIPS decays exponentially with rate (exponent) at least Γ, where Γ is the same hardness constant appearing in the lower bound of Theorem 4.1. This implies that BLFAIPS is asymptotically optimal: no strategy can achieve faster exponential decay rate.

Theorem 4.3. *Under Assumptions 2.1, 2.2, and 2.3, with probability 1,*

$$\liminf_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{(\theta^{\mathrm{r}}_{T+1},\theta^{\mathrm{c}}_{T+1})\sim p_{T+1}}(\hat{z}_{out} \neq z^*) \geq \Gamma,$$

where $p_{T+1} := \mathcal{N}(\hat{\theta}^{\mathrm{r}}_{T+1}, \sigma^2 V_T^{-1}) \otimes \mathcal{N}(\hat{\theta}^{\mathrm{c}}_{T+1}, \gamma^2 V_T^{-1})|\Theta$.

To the best of our knowledge, our algorithm BLFAIPS is the first to demonstrate matching upper and lower bounds in the literature on BAI with constraints under a fixed budget, both in the K-armed bandit setting (Faizal and Nair [2022]) and the linear bandit setting (Tang et al. [2024]).

5 PROOF SKETCH OF THEOREM 4.3

We provide a proof sketch for Theorem 4.3, which is the central result of this paper. To prove this theorem, we first introduce the following lemmas concerning good events:

Lemma 5.1. *Define the good event as*

$$E_{1,\delta} := \bigcap_{t=1}^{T}\left\{\|\hat{\theta}^r_t - \theta^r\|_{V_{t-1}} \leq \sqrt{\beta_1\left(t, \frac{1}{\delta^2}\right)}, \textit{and} \quad \|\hat{\theta}^c_t - \theta^c\|_{V_{t-1}} \leq \sqrt{\beta_2\left(t, \frac{1}{\delta^2}\right)}\right\}$$

where $\beta_1(t, \frac{1}{\delta^2}) := (S_1 + \sigma\sqrt{2\log(\frac{1}{\delta^2}) + d\log\left(\frac{d+tL^2}{d}\right)})^2$, $\beta_2(t, \frac{1}{\delta^2}) := (S_2 + \gamma\sqrt{2\log(\frac{1}{\delta^2}) + d\log\left(\frac{d+tL^2}{d}\right)})^2$.

Then with probability $1-2\delta$*, good event* $E_{1,\delta}$ *holds.*

This lemma, which establishes the concentration property of the ridge regression estimator over the entire time horizon, plays a crucial role in the linear bandit literature.

Lemma 5.2. *Define the good event:*

$$E_{2,\delta} := \bigcap_{t=1}^{T}\left\{\max_{x\in\mathcal{X}}|\langle\hat{\theta}^r_t, x\rangle| \leq B_1, \max_{x\in\mathcal{X}}|\langle\hat{\theta}^c_t, x\rangle| \leq B_2\right\}$$

where

$$B_1 = LR_1 + L\sqrt{\beta_1\left(T, \frac{1}{\delta^2}\right)}, \quad B_2 = LR_2 + L\sqrt{\beta_2\left(T, \frac{1}{\delta^2}\right)}$$

Then,

$$E_{1,\delta} \subseteq E_{2,\delta}$$

This lemma demonstrates that, under the conditions of Lemma 5.1 and the bounded Θ assumption, the empirical reward and cost at each time step are also bounded. This result is essential for applying the Azuma-Hoeffding inequality and deriving an upper bound on regret in our analysis.

Lemma 5.3 (Lemma C.14 in Li et al. [2024])**.** *Define the event* $E_{3,\delta} := \left\{V_t \geq \frac{3}{4}A(\lambda^G), \forall t \geq T_1(\delta), x \in \mathcal{X}\right\}$ *where* $T_1(\delta) := \max_{x\in\mathcal{X}} \left(\frac{6\sqrt{\log(|\mathcal{X}|\frac{T}{\delta})}}{\lambda_x^G}\right)^4$.[1] *Then with probability* $1-\delta$*, the event* $E_{3,\delta}$ *holds.*

This lemma establishes that, under the mixed G-optimal design allocation, the algorithm gathers sufficient information about the arms by a time point that is on the order of $O(\log T)$. This allows us to determine the time at which the empirically best feasible arm coincides with the true one.

Lemma 5.4. *Define the good event*

$$E_{4,\delta} := \{\hat{z}_t = z^*, \quad \forall t > \max\{T_1(\delta)+1, T_2(\delta)\}\},$$

where

$$T_2(\delta) := \max\left\{\left(\frac{m_{max}|\mathcal{X}|}{\Delta_r^{min}}\sqrt{d\beta_1(t,\frac{1}{\delta^2})}\right)^{\frac{8}{3}}, \left(\frac{n_{max}|\mathcal{X}|}{\Delta_c^{min}}\sqrt{d\beta_2(t,\frac{1}{\delta^2})}\right)^{\frac{8}{3}}\right\}.$$

Then

$$E_{1,\delta} \cap E_{3,\delta} \subseteq E_{4,\delta}.$$

This lemma demonstrates that, with the support of Lemmas 5.1 and 5.3, after a time point in the order of $O(\log T)$, the empirically best feasible arm is indeed the true best feasible arm. This result is crucial to our proof, as both the min-learner strategy (posterior sampling) and the max-learner strategy (AdaHedge) in our algorithm rely on the accuracy of identifying the empirically best feasible arm.

Lemma 5.5. *Define*

$$E_{5,\infty} := \left\{\lim_{T\to\infty}\sup_{\theta_1\in\Theta_1,\theta_2\in\Theta_2}\frac{1}{T}\left|\log\frac{p_{T+1}(\theta_1,\theta_2)}{p_{T+1}(\theta^{\mathrm{r}},\theta^{\mathrm{c}})} + \frac{T}{2}\left(\frac{\|\theta^{\mathrm{r}}-\theta_1\|^2_{\overline{V}_T}}{\sigma^2}+\frac{\|\theta^{\mathrm{c}}-\theta_2\|^2_{\overline{V}_T}}{\gamma^2}\right)\right| = 0\right\},$$

with probability 1, the good event $E_{5,\infty}$ *holds. This also implies that*

$$\frac{p_{T+1}(\theta_1,\theta_2)}{p_{T+1}(\theta^{\mathrm{r}},\theta^{\mathrm{c}})} \doteq \exp\left(-\frac{1}{2}\left(\frac{\|\theta^{\mathrm{r}}-\theta_1\|^2_{V_T}}{\sigma^2}+\frac{\|\theta^{\mathrm{c}}-\theta_2\|^2_{V_T}}{\gamma^2}\right)\right).$$

[1]We note that $A(\lambda^G)$ depends on x.

This lemma shows that as $T\to\infty$, the logarithm of the ratio between $p_{T+1}(\theta_1,\theta_2)$ and $p_{T+1}(\theta^{\mathrm{r}},\theta^{\mathrm{c}})$ is asymptotically equal to the hardness term in the exponential lower bound, namely, Γ. This result is crucial for applying the Laplace approximation to analyze the error probability in our proof.

Lemma 5.6. *Define the good event*

$$E_{6,\delta} := \left\{\left|\max_{w\in\Delta_{\mathcal{X}}}\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{A(w)}}{2\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{A(w)}}{2\gamma^2}\right) - \inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{\overline{V}_T}}{2\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{\overline{V}_T}}{2\gamma^2}\right)\right| \leq o(1)\right\}$$

Then event $E_{6,\frac{1}{T}}$ *holds with probability at least* $1-\frac{28}{T}$*, conditioned on events* $E_{1,\frac{1}{T}}, E_{2,\frac{1}{T}}, E_{3,\frac{1}{T}}, E_{4,\frac{1}{T}}$.

This lemma shows that when events $E_{1,\frac{1}{T}}$ $E_{2,\frac{1}{T}}$ $E_{3,\frac{1}{T}}$, and $E_{4,\frac{1}{T}}$ hold, the exponent in the lower bound Γ is asymptotically equal to the Laplace approximation of the integral of the log-ratio, as stated in Lemma 5.5.

Then the error probability of our algorithm is

$$\begin{aligned}
&\mathbb{P}_{(\theta^{\mathrm{r}}_{T+1},\theta^{\mathrm{c}}_{T+1})\sim p_{T+1}}(\hat{z}_{\text{out}} \neq z^*)\\
&= \frac{\int_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}} p_{T+1}(\theta_1,\theta_2)\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2}{\int_{(\theta_1,\theta_2)\in\Theta_{z^*}} p_{T+1}(\theta_1,\theta_2)\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2}\\
&= \frac{\int_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}} p_{T+1}(\theta_1,\theta_2)/p_{T+1}(\theta^{\mathrm{r}},\theta^{\mathrm{c}})\,\mathrm{d}\theta_1\mathrm{d}\theta_2}{\int_{(\theta_1,\theta_2)\in\Theta_{z^*}} p_{T+1}(\theta_1,\theta_2)/p_{T+1}(\theta^{\mathrm{r}},\theta^{\mathrm{c}})\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2}\\
&\overset{(a)}{\doteq} \frac{\int_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}} \exp(M)\,\mathrm{d}\theta_1\mathrm{d}\theta_2}{\int_{(\theta_1,\theta_2)\in\Theta_{z^*}} \exp(M)\,\mathrm{d}\theta_1\mathrm{d}\theta_2},
\end{aligned}$$

where $M = M(\theta_1,\theta_2) := -\frac{1}{2}\left(\frac{\|\theta^{\mathrm{r}}-\theta_1\|^2_{V_T}}{\sigma^2}+\frac{\|\theta^{\mathrm{c}}-\theta_2\|^2_{V_T}}{\gamma^2}\right)$ and (a) comes from good event $E_{5,\infty}$. Furthermore, from the Laplace approximation (Lemma G.3) and the fact that

$$\inf_{(\theta_1,\theta_2)\in\Theta}\frac{\|\theta^{\mathrm{r}}-\theta_1\|^2_{V_T}}{\sigma^2}+\frac{\|\theta^{\mathrm{c}}-\theta_2\|^2_{V_T}}{\gamma^2} = 0,$$

we have

$$\mathbb{P}_{(\theta^{\mathrm{r}}_{T+1},\theta^{\mathrm{c}}_{T+1})\sim p_{T+1}}(\hat{z}_{\text{out}} \neq z^*) \doteq \exp\left(-\frac{T}{2}\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta^{\mathrm{r}}-\theta_1\|^2_{\overline{V}_T}}{\sigma^2}+\frac{\|\theta^{\mathrm{c}}-\theta_2\|^2_{\overline{V}_T}}{\gamma^2}\right)\right).$$

Then combining with the good event $E_{6,\delta}$, we have

$$\mathbb{P}_{(\theta^{\mathrm{r}}_{T+1},\theta^{\mathrm{c}}_{T+1})\sim p_{T+1}}(\hat{z}_{\text{out}} \neq z^*) \doteq \exp(-T\Gamma).$$

To summarize, with the choice $\delta = \frac{1}{T}$, with probability 1,

$$\lim_{T\to\infty} -\frac{1}{T}\log\mathbb{P}_{(\theta^{\mathrm{r}}_{T+1},\theta^{\mathrm{c}}_{T+1})\sim p_{T+1}}(\hat{z}_{\text{out}} \neq z^*) = \Gamma.$$

6 EMPIRICAL STUDIES

For the empirical studies, we compare our proposed algorithm BLFAIPS against three baselines: a modified Linear Thompson Sampling algorithm (Agrawal and Goyal [2013]), which selects the empirically best feasible arm at each round; the Linear β-Top-Two Thompson Sampling algorithm, where β is set to match the allocation rate of the best feasible arm; and an Oracle baseline, which pulls arms according to the optimal allocation rate derived from the lower bound established in Theorem 4.1; the Linear Top-Two Thompson Sampling algorithm with the optimal β under constraints, which can be viewed as an extension from K-armed bandits to linear bandits of the algorithm in Yang and Tan [2022]. Due to space constraints, we show one representative plot here in the main paper. Other plots for other parameters are shown in Appendix H.

6.1 "END OF OPTIMISM" INSTANCE

We perform our first experiment on the ubiquitous "End of Optimism" instance (Lattimore and Szepesvari [2017]) plus an additional feasibility constraint. This can be viewed as the Soare's Instance as considered in Li et al. [2024] with linear constraints. To be more specific, in this instance, we choose $\theta^{\mathrm{r}} = [1, 0]^\top$, $\theta^{\mathrm{c}} = [0, 1]^\top$, $\tau = 0.5$, there are five arms: $[1, 0]^\top, [0, 0.15]^\top, [0, 1]^\top, [1.2, 1.2]^\top, [\cos(\alpha), \sin(\alpha)]^\top$, where $[1, 0]^\top$ is the best feasible arm, $[0, 0.15]^\top$ is the suboptimal feasible arm, $[0, 1]^\top$ is the suboptimal and infeasible arm, $[1.2, 1.2]^\top$ is the superoptimal and infeasible arm. Figures 1, 7, and 8 (in Appendix H.1) show the accuracies of identifying the best feasible arm over time for these algorithms when $\alpha = 0.1, 0.2, 0.3$. The plots also present confidence intervals, represented by a range of plus or minus two standard deviations from the average. We choose the time horizon $T = 2,000$ and run each algorithm over 50 repetitions.

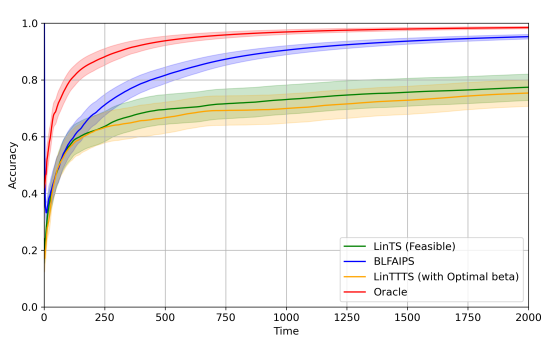

Figure 1: $\alpha = 0.1$

The experimental results, as illustrated by the plots, demonstrate that our proposed algorithm, BLFAIPS, consistently outperforms alternative approaches, including the constrained versions of Linear Thompson Sampling and Linear β-Top-two Thompson Sampling. This performance advantage persists even when the β parameter is fine-tuned to match the true allocation rate of the given instance.

6.2 RANDOM INSTANCES

We next evaluate the performance of our algorithm on random instances within a d-dimensional unit ball. For these experiments, we randomly sampled $K = 5, 20, 50$ arms from $d = 2, 20, 50$-dimensional spaces. The parameters were set as follows: $\theta^{\mathrm{r}} = [1, 0, \ldots, 0]^\top \in \mathbb{R}^d$, $\theta^{\mathrm{c}} = [0, \ldots, 0, 1]^\top \in \mathbb{R}^d$, and $\tau = 0.5$. Figures 2, 3, and 9 (in Appendix H.2) present the accuracies over time for varying values of K, while Figures 4, 9, 10 illustrate the accuracies over time for different dimensionalities d.

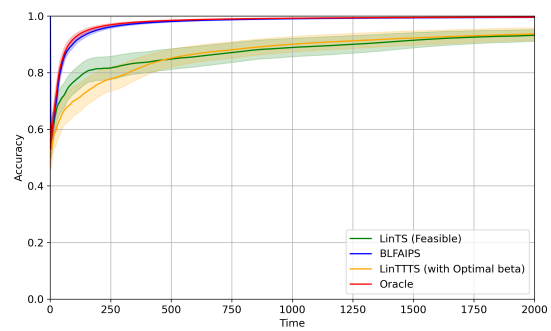

Figure 2: $d = 2, K = 5$

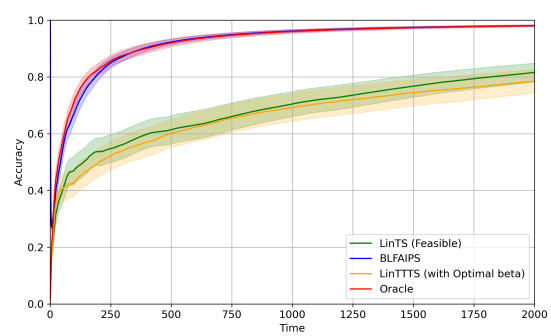

Figure 3: $d = 2, K = 50$

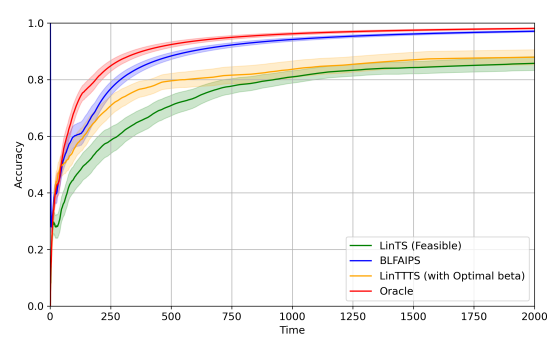

Figure 4: $d = 50, K = 20$

The results indicate that our proposed algorithm, BLFAIPS, consistently exhibits faster convergence compared to competing algorithms across varying numbers of arms K and dimensionalities d in random instances.

6.3 COMPARISON WITH PEPS

We also compare our algorithm with the PEPS algorithm introduced by Li et al. [2024], which uses AdaHedge but does not incorporate the doubling trick. While their algorithm PEPS performs well empirically, they did not provide a formal upper bound guaranteeing its asymptotic optimality. In contrast, our Theorem 4.3 establishes such a guarantee, matching the information-theoretic lower bound (Theorem 4.1) and closing a key theoretical gap in the literature.

We evaluate both algorithms (and others) in the unconstrained setting using the same configuration as in the random instance experiments with $d = 2$ and $K = 5$, except that the constraint threshold τ is set to ∞. When the sampling budget T is known in advance, PEPS and BLFAIPS achieve comparable performance. However, when T is *unknown*, as is common in many real-world applications, PEPS tends to waste almost half its budget during its initial exploration phase, leading to strictly inferior results. In such cases, BLFAIPS achieves superior performance as shown in Figure 5.

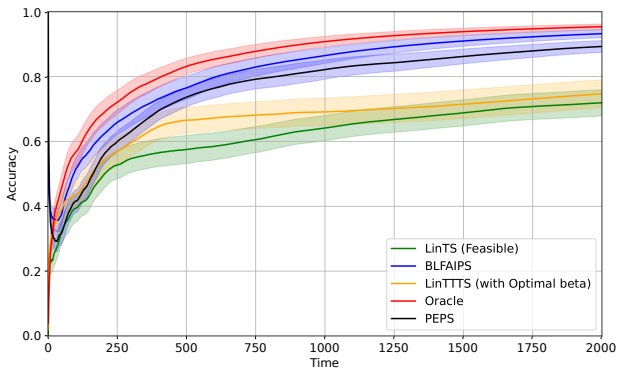

Figure 5: Comparison of BLFAIPS to PEPS for unknown T

6.4 EVALUATION ON A REAL-WORLD DATASET

We also conducted an experiment utilizing a real-world dataset: the MovieLens 10M dataset augmented with IMDb ratings, based on an open-source preprocessing script Lab [2017]. Our objective was to identify the highest-rated movie (on a 1-to-5 scale) among the top 20 movies in MovieLens, subject to the constraint that the IMDb rating as of April 2025 remains below $\tau = 7.5$ (on a 1-to-10 scale).

To assess the feasibility-aware performance of our algorithm, we calculated the cumulative accuracies over 5,000 time steps for various algorithms. Accuracy is defined as the proportion of instances where the selected movie satisfies the IMDb constraint and is simultaneously the highest-rated based on user ratings.

Since this is a real-world dataset, the true problem instance is unknown. Thus, it not possible to determine an optimal solution or a lower bound. We compared our approach against baseline configurations using the same parameters, which were set as follows: $\sigma = 1$, $\gamma = 1$, $L = \sqrt{3}$, $R_1 = \sqrt{10}$, and $R_2 = \sqrt{20}$. Each algorithm was executed 50 times over the 5,000 time steps.

As illustrated in Figure 6, our proposed method BLFAIPS significantly outperforms both the constrained versions of Linear Feasible Thompson Sampling and Linear TTTS with $\beta = 0.5$. It consistently achieves higher accuracy while maintaining similar error bars, showing the accuracy, stability, and robustness of BLFAIPS.

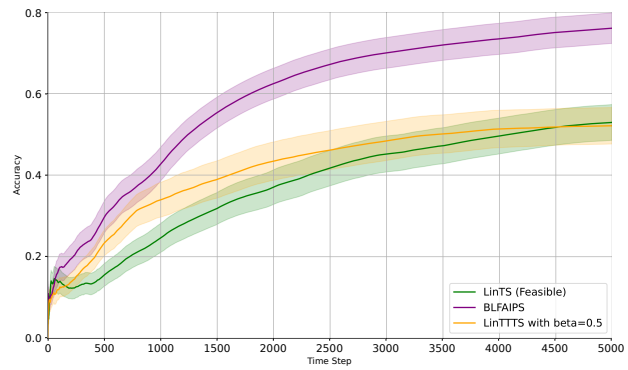

Figure 6: Comparison of accuracies of various algorithms on the MovieLens dataset

7 CONCLUSION AND FUTURE WORK

In this paper, we addressed the best feasible arm identification problem under a fixed budget in linear bandits. We proposed a novel algorithm that matches upper and lower bounds on the exponential rate of the error probability. To the best of our knowledge, it is the first to attain such optimality in this domain. Our contributions include two equivalent formulations of the lower bound rate from Bayesian and frequentist perspectives, structural improvements to existing methods, and empirical results showing the superiority of our approach over several baselines.

Our theoretical findings provide a basis for further exploration in constrained bandit problems. While we have established matching bounds under specific conditions, extending these results to more general settings, such as non-convex or high-dimensional parameter spaces, remains valuable. Refining guarantees under weaker assumptions, especially regarding noise and cost constraints, could further broaden applicability.

We hope our work will inspire advances in constrained best arm identification and related problems within the multi-armed bandit framework.

Acknowledgements

This research is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG2-PhD-2023-08-044T-J), and is part of the programme DesCartes which is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in Neural Information Processing Systems*, 24, 2011.

Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*, pages 127–135. PMLR, 2013.

Sanae Amani, Mahnoosh Alizadeh, and Christos Thrampoulidis. Linear stochastic bandits under safety constraints. *Advances in Neural Information Processing Systems*, 32, 2019.

Jean-Yves Audibert and Sébastien Bubeck. Best arm identification in multi-armed bandits. In *Proc. of 23rd Conference on Learning Theory*, page 13, 2010.

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

Mohammad Javad Azizi, Branislav Kveton, and Mohammad Ghavamzadeh. Fixed-budget best-arm identification in structured bandits. *arXiv preprint arXiv:2106.04763*, 2021.

Djallel Bouneffouf and Irina Rish. A survey on practical applications of multi-armed and contextual bandits. *arXiv preprint arXiv:1904.10040*, 2019.

Romain Camilleri et al. Active learning with safety constraints. *Advances in Neural Information Processing Systems*, 35:33201–33214, 2022.

Steven De Rooij, Tim Van Erven, Peter D Grünwald, and Wouter M Koolen. Follow the leader if you can, hedge if you must. *The Journal of Machine Learning Research*, 15(1):1281–1316, 2014.

Fathima Zarin Faizal and Jayakrishnan Nair. Constrained pure exploration multi-armed bandits with a fixed budget. *arXiv preprint arXiv:2211.14768*, 2022.

Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. *Advances in Neural Information Processing Systems*, 25, 2012.

Yunlong Hou, Vincent Y. F. Tan, and Zixin Zhong. Almost optimal variance-constrained best arm identification. *IEEE Transactions on Information Theory*, 69(4):2603–2634, 2023.

Yassir Jedra and Alexandre Proutiere. Optimal best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 33:10007–10017, 2020.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. PAC subset selection in stochastic multi-armed bandits. In *International Conference on Machine Learning*, volume 12, pages 655–662, 2012.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *International Conference on Machine Learning*, pages 1238–1246. PMLR, 2013.

Julian Katz-Samuels and Clay Scott. Feasible arm identification. In *International Conference on Machine Learning*, pages 2535–2543. PMLR, 2018.

Julian Katz-Samuels and Clayton Scott. Top feasible arm identification. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1593–1601. PMLR, 2019.

Junpei Komiyama, Taira Tsuchiya, and Junya Honda. Minimax optimal algorithms for fixed-budget best arm identification. *Advances in Neural Information Processing Systems*, 35:10393–10404, 2022.

Cyrille Kone, Marc Jourdan, and Emilie Kaufmann. Pareto set identification with posterior sampling. *arXiv preprint arXiv:2411.04939*, 2024.

Volodymyr Kuleshov and Doina Precup. Algorithms for multi-armed bandit problems. *arXiv preprint arXiv:1402.6028*, 2014.

Yuko Kuroki, Liyuan Xu, Atsushi Miyauchi, Junya Honda, and Masashi Sugiyama. Polynomial-time algorithms for multiple-arm identification with full-bandit feedback. *Neural Computation*, 32(9):1733–1773, 2020.

NTU Computational Lab. Striatum: A contextual bandit library. `https://github.com/ntucllab/striatum`, 2017. Accessed: 2025-06-02.

Tor Lattimore and Csaba Szepesvari. The end of optimism? an asymptotic analysis of finite-armed linear bandits. In *Artificial Intelligence and Statistics*, pages 728–737. PMLR, 2017.

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

Zhaoqi Li, Kevin Jamieson, and Lalit Jain. Optimal exploration is no harder than Thompson sampling. In *International Conference on Artificial Intelligence and Statistics*, pages 1684–1692. PMLR, 2024.

Oden Maron and Andrew W Moore. The racing algorithm: Model selection for lazy learners. *Artificial Intelligence Review*, 11:193–225, 1997.

Ahmadreza Moradipari, Sanae Amani, Mahnoosh Alizadeh, and Christos Thrampoulidis. Safe linear Thompson sampling with side information. *IEEE Transactions on Signal Processing*, 69:3755–3767, 2021.

Aldo Pacchiano, Mohammad Ghavamzadeh, and Peter Bartlett. Contextual bandits with stage-wise constraints. *arXiv preprint arXiv:2401.08016*, 2024.

Daniel Russo. Simple Bayesian algorithms for best arm identification. In *Conference on Learning Theory*, pages 1417–1418. PMLR, 2016.

Xuedong Shang, Igor Colin, Merwan Barlier, and Hamza Cherkaoui. Price of safety in linear best arm identification. *arXiv preprint arXiv:2309.08709*, 2023.

Dengwang Tang, Rahul Jain, Ashutosh Nayyar, and Pierluigi Nuzzo. Pure exploration for constrained best mixed arm identification with a fixed budget. *arXiv preprint arXiv:2405.15090*, 2024.

William R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933. ISSN 00063444. URL `http://www.jstor.org/stable/2332286`.

Zhenlin Wang, Andrew J Wagenmaker, and Kevin Jamieson. Best arm identification with safety constraints. In *International Conference on Artificial Intelligence and Statistics*, pages 9114–9146. PMLR, 2022.

Junwen Yang and Vincent Tan. Minimax optimal fixed-budget best arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 35:12253–12266, 2022.

Le Yang, Siyang Gao, Cheng Li, and Yi Wang. Stochastically constrained best arm identification with Thompson sampling. *Automatica*, 176:112223, Jun 2025.

Xingyu Zhou and Bo Ji. On kernelized multi-armed bandits with constraints. *Advances in Neural Information Processing Systems*, 35:14–26, 2022.

Asymptotically Optimal Linear Best Feasible Arm Identification with Fixed Budget
(Supplementary Material)

Jie Bian[1] **Vincent Y. F. Tan**[2,1]

[1]Department of Electrical and Computer Engineering, National University of Singapore, Singapore
[2]Department of Mathematics, National University of Singapore, Singapore

A TABLE OF NOTATION

$B_1 = 2LR_1 + L\sqrt{\beta_1(T, \frac{1}{\delta^2})}$	Upper bound of $\max_{x\in\mathcal{X}} \max_{t\le T} \lvert\langle x, \hat{\theta}^{\mathrm{r}}_t\rangle\rvert$
$B_2 = 2LR_2 + L\sqrt{\beta_2(T, \frac{1}{\delta^2})}$	Upper bound of $\max_{x\in\mathcal{X}} \max_{t\le T} \lvert\langle x, \hat{\theta}^{\mathrm{c}}_t\rangle\rvert$
$T_1(\delta)$	Time while the algorithm collecting enough information for the exploration
$\Delta^{\mathrm{r}}_{\min} := \min_{z\in\mathcal{Z}} (z^* - z)^\top \theta^{\mathrm{r}}$	minimum gap of the reward for $\mathcal{Z}$
$\Delta^{\mathrm{c}}_{\min} := \min_{z\in\mathcal{Z}} \lvert\tau - z^\top \theta^{\mathrm{c}}\rvert$	minimum gap of the cost for $\mathcal{Z}$
$\overline{V}_T := \frac{1}{T}(I + \sum_{t=1}^T X_t X_t^\top)$	Empirical variance matrix after T rounds
$p_{T+1} := \mathcal{N}\left(\hat{\theta}^{\mathrm{r}}_{T+1}, \eta_{\mathrm{r}}^{-1} V_T^{-1}\right) \otimes$ $\mathcal{N}\left(\hat{\theta}^{\mathrm{c}}_{T+1}, \eta_{\mathrm{c}}^{-1} V_T^{-1}\right) \Big\vert \Theta$	posterior distribution for recommendation at time T
η_{r}	learning rate of reward for min-learner
η_{c}	learning rate of cost for min-learner
$p_t := \mathcal{N}(\hat{\theta}^{\mathrm{r}}_t, \eta_r^{-1} V_{t-1}^{-1}) \otimes$ $\mathcal{N}(\hat{\theta}^{\mathrm{c}}_t, \eta_c^{-1} V_{t-1}^{-1}) \vert \overline{\Theta}_{\hat{z}_t}$	posterior distribution for sampling at time t
$\overline{p}_T := \frac{1}{T}\sum_{t=1}^T p_t$	Average posterior distribution for sampling over T rounds
$\tilde{p}_t := \mathcal{N}(\hat{\theta}^{\mathrm{r}}_t, \eta_r^{-1} V_{t-1}^{-1}) \otimes$ $\mathcal{N}(\hat{\theta}^{\mathrm{c}}_t, \eta_c^{-1} V_{t-1}^{-1}) \vert \overline{\Theta}_{z^*}$	optimal posterior distribution for sampling at time t

B PROOF OF THEOREM 4.1

Proof. We use the standard posterior convergence analysis to derive this theorem. Without loss of generality, we assume that the training and testing arm sets $\mathcal{X}, \mathcal{Z}$ both span $\mathbb{R}^d$. Denote the data as $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$ and $X_i \in \mathbb{R}^d$. Define

$$X = \begin{bmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{bmatrix} \in \mathbb{R}^{n\times d}$$

We assume the prior distribution for $\theta^{\mathrm{r}}, \theta^{\mathrm{c}}$ are multivariate Gaussian:

$$\theta^{\mathrm{r}} \sim \mathcal{N}(\theta^{\mathrm{r}}_0, \lambda^{-1} I) \quad \text{and} \quad \theta^{\mathrm{c}} \sim \mathcal{N}(\theta^{\mathrm{c}}_0, \lambda^{-1} I)$$

The likelihood for each reward and cost $Y_i^{\mathrm{r}}, Y_i^{\mathrm{c}}$ are:

$$P(Y_i^{\mathrm{r}} \mid \theta^{\mathrm{r}}) = \mathcal{N}(Y_i^{\mathrm{r}} \mid \langle \theta^{\mathrm{r}}, X_i \rangle, \sigma^2) \quad \text{and} \quad P(Y_i^{\mathrm{c}} \mid \theta^{\mathrm{c}}) = \mathcal{N}(Y_i^{\mathrm{c}} \mid \langle \theta^{\mathrm{c}}, X_i \rangle, \gamma^2)$$

For n observations, the likelihood is:

$$P(\mathcal{D} \mid \theta^{\mathrm{r}}) = \prod_{i=1}^{n} \mathcal{N}(Y_i^{\mathrm{r}} \mid \langle \theta^{\mathrm{r}}, X_i \rangle, \sigma^2) \quad \text{and} \quad P(\mathcal{D} \mid \theta^{\mathrm{c}}) = \prod_{i=1}^{n} \mathcal{N}(Y_i^{\mathrm{c}} \mid \langle \theta^{\mathrm{c}}, X_i \rangle, \gamma^2)$$

Let $X \in \mathbb{R}^{n \times d}$ be the feature matrix where each row is $X_i^\top$. Then the vector of rewards and costs $Y^{\mathrm{r}} = (Y_1^{\mathrm{r}}, Y_2^{\mathrm{r}}, \ldots, Y_n^{\mathrm{r}})^\top, Y^{\mathrm{c}} = (Y_1^{\mathrm{c}}, Y_2^{\mathrm{c}}, \ldots, Y_n^{\mathrm{c}})^\top$ follows:

$$Y^{\mathrm{r}} \mid \theta^{\mathrm{r}} \sim \mathcal{N}(X\theta^{\mathrm{r}}, \sigma^2 I) \quad \text{and} \quad Y^{\mathrm{c}} \mid \theta^{\mathrm{c}} \sim \mathcal{N}(X\theta^{\mathrm{c}}, \gamma^2 I)$$

Using Bayes' theorem, the posterior distribution is:

$$P(\theta^* \mid \mathcal{D}) \propto P(\mathcal{D} \mid \theta^*) P(\theta^*)$$

Since both the prior and likelihood are Gaussian, the posterior is also Gaussian, the posterior mean and covariance of the reward are

$$\Sigma_n^{\mathrm{r}} = \left(\lambda I + \frac{1}{\sigma^2} X^\top X\right)^{-1} \quad \text{and} \quad \theta_n^{\mathrm{r}} = \Sigma_n^{\mathrm{r}} \left(\lambda \theta_0^{\mathrm{r}} + \frac{1}{\sigma^2} X^\top Y^{\mathrm{r}}\right)$$

where the posterior distribution is

$$\theta^{\mathrm{r}} \mid \mathcal{D} \sim \mathcal{N}(\theta_n^{\mathrm{r}}, \Sigma_n^{\mathrm{r}})$$

The posterior mean and covariance of the cost are:

$$\Sigma_n^{\mathrm{c}} = \left(\lambda I + \frac{1}{\gamma^2} X^\top X\right)^{-1} \quad \text{and} \quad \theta_n^{\mathrm{c}} = \Sigma_n^{\mathrm{c}} \left(\lambda \theta_0^{\mathrm{c}} + \frac{1}{\gamma^2} X^\top Y^{\mathrm{c}}\right)$$

where the posterior distribution is

$$\theta^{\mathrm{c}} \mid \mathcal{D} \sim \mathcal{N}(\theta_n^{\mathrm{c}}, \Sigma_n^{\mathrm{c}}).$$

Assume $\theta_0^{\mathrm{r}}, \theta_0^{\mathrm{c}}$ are zero vectors, then

$$\begin{aligned}
\Sigma_n^{\mathrm{r}} &= \sigma^2 \left(\lambda \sigma^2 I + X^\top X\right)^{-1}, \\
\theta_n^{\mathrm{r}} &= \left(\lambda \sigma^2 I + X^\top X\right)^{-1} X^\top Y^{\mathrm{r}}, \\
\Sigma_n^{\mathrm{c}} &= \gamma^2 \left(\lambda \gamma^2 I + X^\top X\right)^{-1}, \\
\theta_n^{\mathrm{c}} &= \left(\lambda \gamma^2 I + X^\top X\right)^{-1} X^\top Y^{\mathrm{c}}.
\end{aligned}$$

Hence the posterior distributions of the expected reward and cost of arm $z \in \mathcal{Z}$ are:

$$\mu^{\mathrm{r}}(z) \mid \mathcal{D} \sim \mathcal{N}(\langle \theta_n^{\mathrm{r}}, z \rangle, \sigma^2 \|z\|^2_{V_{nr}^{-1}}) \quad \text{and} \quad \mu^{\mathrm{c}}(z) \mid \mathcal{D} \sim \mathcal{N}(\langle \theta_n^{\mathrm{c}}, z \rangle, \gamma^2 \|z\|^2_{V_{nc}^{-1}}),$$

where $V_{nr} := \lambda \sigma^2 I + X^\top X, V_{nc} := \lambda \gamma^2 I + X^\top X$. Let $\mu_n^{\mathrm{r}}(z), \mu_n^{\mathrm{c}}(z)$ denote the sample reward and cost of arm $z \in \mathcal{Z}$ at step n drawn from the posterior distribution $\mu^{\mathrm{r}}(z)|\mathcal{D}, \mu^{\mathrm{c}}(z)|\mathcal{D}$. Therefore for any suboptimal arm $z \neq z^*$,

$$\mu^{\mathrm{r}}(z) - \mu(z^*) \mid \mathcal{D} \sim \mathcal{N}(\langle \theta_n^{\mathrm{r}}, z - z^* \rangle, \sigma^2 \|z - z^*\|^2_{V_{nr}^{-1}}).$$

For any arm $z \in \mathcal{Z}$,

$$\tau - \mu^{\mathrm{c}}(z) \mid \mathcal{D} \sim \mathcal{N}(\tau - \langle \theta_n^{\mathrm{c}}, z \rangle, \gamma^2 \|z\|^2_{V_{nc}^{-1}}).$$

Denote the reward gap and sample reward gap for suboptimal arm $z \neq z^*$ as $\Delta^{\mathrm{r}}(z) := \langle \theta^{\mathrm{r}}, z^* - z \rangle, \Delta_n^{\mathrm{r}}(z) := \langle \theta_n^{\mathrm{r}}, z^* - z \rangle$ respectively, define the cost gap and sample cost gap for arm $z \in \mathcal{Z}$ by $\Delta^{\mathrm{c}}(z) := |\tau - \langle \theta^{\mathrm{c}}, z \rangle|, \Delta_n^{\mathrm{c}}(z) := |\tau - \langle \theta_n^{\mathrm{c}}, z \rangle|$ respectively.

Since

$$1-\mathbb{P}(z_{\text{out}} \neq z^*) = \mathbb{P}(\{\mu_n^{\text{c}}(z^*) > \tau\} \cup \{\exists z \neq z^*, \mu_n^{\text{r}}(z) \geq \mu_n^{\text{r}}(z^*) \text{ and } \mu_n^{\text{c}}(z) \leq \tau\})$$
$$\doteq \max_{z \in \mathcal{Z}} \left\{\mathbb{P}(\mu_n^{\text{r}}(z) \geq \mu_n^{\text{r}}(z^*), \mu_n^{\text{c}}(z) \leq \tau) \cdot \mathbb{1}\{z \neq z^*\}, \mathbb{P}(\mu_n^{\text{c}}(z) > \tau) \cdot \mathbb{1}\{z = z^*\}\right\},$$

then the problem reduces to derive the asymptotic expression within the maximum term. The following lemma about Gaussian distribution will be useful in our proofs:

Lemma B.1. *For Gaussian distribution $X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu \leq 0$,*

$$\frac{x}{1+x^2}\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{\mu^2}{2\sigma^2}\Big) \leq \mathbb{P}(X \geq 0) \leq \frac{1}{2}\exp\Big(-\frac{\mu^2}{2\sigma^2}\Big)$$

where $x = -\frac{\mu}{\sigma}$.

With the above lemma, first if $z = z^*$ we obtain the following lemma.

Lemma B.2. $\mathbb{P}(\mu_n^{\text{c}}(z^*) > \tau) \doteq \exp\left(-\frac{(\Delta_n^{\text{c}}(z^*))^2}{2\gamma^2\|z^*\|^2_{V_{nc}^{-1}}}\right)$.

Proof. Since $\mu_n^{\text{c}}(z^*) \sim \mathcal{N}(\langle\theta_n^{\text{c}}, z^*\rangle, \gamma^2\|z^*\|^2_{V_{nc}^{-1}})$, applying Lemma B.1 we have

$$\frac{x}{1+x^2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(\langle\theta_n^{\text{c}}, z^*\rangle - \tau)^2}{2\gamma^2\|z^*\|^2_{V_{nc}^{-1}}}\right) \leq \mathbb{P}(\mu_n^{\text{c}}(z^*) > \tau) \leq \frac{1}{2}\exp\left(-\frac{(\langle\theta_n^{\text{c}}, z^*\rangle - \tau)^2}{2\gamma^2\|z^*\|^2_{V_{nc}^{-1}}}\right)$$

where $x = \frac{\tau - \langle\theta_n^{\text{c}}, z^*\rangle}{\gamma\|z^*\|_{V_{nc}^{-1}}}$ here. Then

$$\frac{1}{n}\log\frac{1}{\sqrt{2\pi}} + \frac{1}{n}\log\frac{x}{1+x^2} \leq \frac{1}{n}\log\left(\frac{\mathbb{P}(\mu_n^{\text{c}}(z^*) > \tau)}{\exp\left(-\frac{(\langle\theta_n^{\text{c}}, z^*\rangle - \tau)^2}{2\gamma^2\|z^*\|^2_{V_{nc}^{-1}}}\right)}\right) \leq \frac{1}{n}\log\frac{1}{2}.$$

Notice $\|z^*\|_{V_{nc}^{-1}} = \frac{\|z^*\|_{V_\pi^{-1}}}{\sqrt{n}}$ where $V_{\pi c} := \frac{\lambda\gamma^2}{n}I + \sum_{i=1}^K \frac{T_n(z)}{n} zz^\top$ and $T_n(z)$ is the total number of pulls of arm z over n rounds. Since there must exist one arm j^* such that $T_n(j^*) \geq \frac{n}{K}$, then

$$V_{\min} := \frac{\lambda\gamma^2}{n}I + \frac{j^* j^{*\top}}{K} \lesssim V_{\pi c} \lesssim \frac{\lambda\gamma^2}{n}I + \sum_{i=1}^K zz^\top := V_{\max}.$$

Hence $\lim_{n\to\infty}\frac{1}{n}\log\frac{x}{1+x^2} = 0$ since $\lim_{n\to\infty}\frac{1}{n}\log n = 0$, we have

$$\lim_{n\to\infty}\frac{1}{n}\log\left(\frac{\mathbb{P}(\mu_n^{\text{c}}(z^*) > \tau)}{\exp\left(-\frac{(\langle\theta_n^{\text{c}}, z^*\rangle - \tau)^2}{2\gamma^2\|z^*\|^2_{V_{nc}^{-1}}}\right)}\right) = 0.$$

□

Lemma B.3. *For $z \in \mathcal{F} \cap \mathcal{S}$, $\mathbb{P}(\mu_n^{\text{r}}(z) \geq \mu_n^{\text{r}}(z^*), \mu_n^{\text{c}}(z) \leq \tau) \doteq \exp\left(-\frac{(\Delta_n^{\text{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_{nr}^{-1}}}\right)$.*

Proof. First $\mathbb{P}(\mu_\infty^{\text{c}}(z) \leq \tau) = 1$ since $z \in \mathcal{F}$, then by Lemma B.1,

$$\frac{x}{1+x^2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(\Delta_n^{\text{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_{nr}^{-1}}}\right) \leq \mathbb{P}(\mu_n^{\text{r}}(z) \geq \mu_n^{\text{r}}(z^*)) \leq \frac{1}{2}\exp\left(-\frac{(\Delta_n^{\text{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_{nr}^{-1}}}\right),$$

where $x = \frac{\Delta_n^{\mathrm{r}}(z)}{\sigma\|z-z^*\|_{V_{nr}^{-1}}}$.

Then

$$\frac{1}{n}\log\frac{1}{\sqrt{2\pi}} + \frac{1}{n}\log\frac{x}{1+x^2} \le \frac{1}{n}\log\left(\frac{\mathbb{P}(\mu_n^{\mathrm{r}}(z) \ge \mu_n^{\mathrm{r}}(z^*) \mid \mathcal{D})}{\exp\left(-\frac{(\langle\theta_n^{\mathrm{r}}, z-z^*\rangle)^2}{2\sigma^2\|z-z^*\|^2_{V_n^{-1}}}\right)}\right) \le \frac{1}{n}\log\frac{1}{2}.$$

Notice that $\|z-z^*\|_{V_{nr}^{-1}} = \frac{\|z-z^*\|_{V_\pi^{-1}}}{\sqrt{n}}$ where $V_{\pi r} := \frac{\lambda\sigma^2}{n}I + \sum_{i=1}^K \frac{T_n(z)}{n} zz^\top$ and $T_n(z)$ is the total number of pulls of arm z over n rounds. Since there must exist one arm j^* such that $T_n(j^*) \ge \frac{n}{K}$, then

$$V_{\min} := \frac{\lambda\sigma^2}{n}I + \frac{j^* j^{*\top}}{K} \lesssim V_{\pi r} \lesssim \frac{\lambda\sigma^2}{n}I + \sum_{i=1}^K zz^\top := V_{\max}.$$

Hence $\lim_{n\to\infty}\frac{1}{n}\log\frac{x}{1+x^2} = 0$ since $\lim_{n\to\infty}\frac{1}{n}\log n = 0$, we have

$$\lim_{n\to\infty}\frac{1}{n}\log\left(\frac{\mathbb{P}(\mu_n^{\mathrm{r}}(z) \ge \mu_n^{\mathrm{r}}(z) \mid \mathcal{D})}{\exp\left(-\frac{(\langle\theta_n^{\mathrm{r}}, z-z^*\rangle)^2}{2\sigma^2\|z-z^*\|^2_{V_{nr}^{-1}}}\right)}\right) = 0.$$

Hence

$$\mathbb{P}(\mu_n^{\mathrm{r}}(z) \ge \mu_n^{\mathrm{r}}(z^*), \mu_n^{\mathrm{c}}(z) \le \tau) \doteq \exp\left(-\frac{(\Delta_n^{\mathrm{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_{nr}^{-1}}}\right).$$

□

Lemma B.4. *For $z \in \overline{\mathcal{F}} \cap \overline{\mathcal{S}}$, $\mathbb{P}(\mu_n^{\mathrm{r}}(z) \ge \mu_n^{\mathrm{r}}(z^*), \mu_n^{\mathrm{c}}(z) \le \tau) \doteq \exp\left(-\frac{(\Delta_n^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_{nc}^{-1}}}\right)$*

Proof. First $\mathbb{P}(\mu_\infty^{\mathrm{r}}(z) \ge \mu_\infty^{\mathrm{r}}(z^*)) = 1$ since $i \in \mathcal{S}$, then with the similar procedure as in Lemma B.3,

$$\mathbb{P}(\mu_n^{\mathrm{c}}(z) \le \tau) \doteq \frac{1}{2}\exp\left(-\frac{(\Delta_n^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_{nc}^{-1}}}\right).$$

Hence

$$\mathbb{P}(\mu_n^{\mathrm{r}}(z) \ge \mu_n^{\mathrm{r}}(z^*), \mu_n^{\mathrm{c}}(z) \le \tau) \doteq \exp\left(-\frac{(\Delta_n^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_{nc}^{-1}}}\right).$$

□

Lemma B.5. *For $i \in \overline{\mathcal{F}} \cap \mathcal{S}$, $\mathbb{P}(\mu_{n,i}^{\mathrm{r}} \ge \mu_{n,1}^{\mathrm{r}}, \mu_{n,i}^{\mathrm{c}} \le \tau) \doteq \exp\left(-\frac{(\Delta_n^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_{nc}^{-1}}}\right) \cdot \exp\left(-\frac{(\Delta_n^{\mathrm{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_{nr}^{-1}}}\right)$*

Proof. Since $\mathbb{P}(\mu_{n,i}^{\mathrm{r}} \ge \mu_{n,1}^{\mathrm{r}}, \mu_{n,i}^{\mathrm{c}} \le \tau) = \mathbb{P}(\mu_{n,i}^{\mathrm{r}} \ge \mu_{n,1}^{\mathrm{r}}) \cdot \mathbb{P}(\mu_{n,i}^{\mathrm{c}} \le \tau)$, then with the similar procedure as in Lemma B.3, this lemma can be proved. □

For convenience, denote $\mathcal{A}_1 := \overline{\mathcal{F}} \cap \overline{\mathcal{S}}, \mathcal{A}_2 := \mathcal{F} \cap \mathcal{S}, \mathcal{A}_3 := \overline{\mathcal{F}} \cap \mathcal{S}$.

To summarize,

$$\begin{aligned}
&1 - \mathbb{P}(z_{\text{out}} \ne z^*) \\
&\doteq \max_{z\in\mathcal{Z}} \left\{\mathbb{P}(\mu_n^{\mathrm{r}}(z) \ge \mu_n^{\mathrm{r}}(z^*), \mu_n^{\mathrm{c}}(z) \le \tau) \cdot \mathbb{1}\{z \ne z^*\}, \mathbb{P}(\mu_n^{\mathrm{c}}(z) > \tau) \cdot \mathbb{1}\{z = z^*\}\right\}
\end{aligned}$$

$$\doteq \max_{z\in\mathcal{Z}} \Bigg\{ \exp\left(-\frac{(\Delta_n^{\mathrm{c}}(z^*))^2}{2\gamma^2\|z^*\|^2_{V_{nc}^{-1}}}\right), \exp\left(-\frac{(\Delta_n^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_{nc}^{-1}}}\right)\mathbb{1}\{z\in\mathcal{A}_1\},$$
$$\exp\left(-\frac{(\Delta_n^{\mathrm{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_{nr}^{-1}}}\right)\mathbb{1}\{z\in\mathcal{A}_2\}, \exp\left(-\frac{(\Delta_n^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_{nc}^{-1}}}\right)\cdot\exp\left(-\frac{(\Delta_n^{\mathrm{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_{nr}^{-1}}}\right)\mathbb{1}\{z\in\mathcal{A}_3\}\Bigg\}.$$

Hence

$$\begin{aligned}
1-\mathbb{P}(z_{\text{out}}\neq z^*) &\doteq \exp\Bigg(-n\min_{z\in\mathcal{Z}}\Bigg\{\frac{(\Delta_n^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_w^{-1}}}\mathbb{1}\{z\in\mathcal{A}_1\}, \frac{(\Delta_n^{\mathrm{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_w^{-1}}}\mathbb{1}\{z\in\mathcal{A}_2\},\\
&\qquad \left(\frac{(\Delta_n^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_w^{-1}}}+\frac{(\Delta_n^{\mathrm{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_w^{-1}}}\right)\mathbb{1}\{i\in\mathcal{A}_3\}, \frac{(\Delta_n^{\mathrm{c}}(z))^2}{2\gamma^2\|z^*\|^2_{V_w^{-1}}}\Bigg\}\Bigg)\\
&\geq \exp\Bigg(-n\max_{w\in W}\min_{z\in\mathcal{Z}}\Bigg\{\frac{(\Delta^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_w^{-1}}}\mathbb{1}\{z\in\mathcal{A}_1\}, \frac{(\Delta^{\mathrm{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_w^{-1}}}\mathbb{1}\{z\in\mathcal{A}_2\},\\
&\qquad \left(\frac{(\Delta^{\mathrm{c}}(z))^2}{2\gamma^2\|x_i\|^2_{V_w^{-1}}}+\frac{(\Delta^{\mathrm{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_w^{-1}}}\right)\mathbb{1}\{z\in\mathcal{A}_3\}, \frac{(\Delta^{\mathrm{c}}(z^*))^2}{2\gamma^2\|z^*\|^2_{V_w^{-1}}}\Bigg\}\Bigg)\\
&= \exp(-n\Gamma)
\end{aligned}$$

for any sampling rule where $V_w := \sum_{i=1}^K w_i x_i x_i^\top$ and $w_i := \lim_{n\to\infty}\frac{T_{i,n}}{n}$.

Let $W := \{w=(w_1,w_2,\ldots,w_K) : \sum_i w_i = 1, w_i\geq 0, \forall\, i\in[K]\}$. The exponential rate in the lower bound can be written as

$$\begin{aligned}
\Gamma := \max_{w\in W}\min_{z\in\mathcal{Z}}\Bigg\{&\frac{(\Delta^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_w^{-1}}}\mathbb{1}\{z\in\mathcal{A}_1\}, \frac{(\Delta^{\mathrm{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_w^{-1}}}\mathbb{1}\{z\in\mathcal{A}_2\},\\
&\left(\frac{(\Delta^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_w^{-1}}}+\frac{(\Delta^{\mathrm{r}}(z))^2}{2\sigma^2\|z-z^*\|^2_{V_w^{-1}}}\right)\mathbb{1}\{z\in\mathcal{A}_3\}, \frac{(\Delta^{\mathrm{c}}(z^*))^2}{2\gamma^2\|z^*\|^2_{V_w^{-1}}}\Bigg\}.
\end{aligned}$$

□

C PROOF OF THEOREM 4.2

Proof. We start from the right-hand side of the equation in Theorem 4.2, this optimization problem is equivalent to the following:

$$\begin{aligned}
&\min_{\theta_1,\theta_2}\frac{1}{2}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{V_w}}{\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{V_w}}{\gamma^2}\right)\\
&\textbf{s.t.}\quad (\theta_1,\theta_2)\in\overline{\Theta}_{z^*}
\end{aligned}$$

By the definition of feasibility, this optimization problem is equal to:

$$\begin{aligned}
&\min_{\theta_1,\theta_2}\frac{1}{2}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{V_w}}{\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{V_w}}{\gamma^2}\right)\\
&\textbf{s.t.}\quad \theta_2^\top z^* > \tau \text{ or } \exists z\in\mathcal{Z}\backslash z^* : \theta_1^\top z\geq\theta_1^\top z^* \text{ and } \theta_2^\top z\leq\tau
\end{aligned}$$

Furthermore this problem can be divided into two optimization problem and the minimum of the solutions of this two problem will equal to the solution of the original optimization problem, the two subproblems are:

$$\begin{aligned}
&\min_{\theta_1,\theta_2}\frac{1}{2}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{V_w}}{\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{V_w}}{\gamma^2}\right)\\
&\textbf{s.t.}\quad \exists z\in\mathcal{Z}\backslash z^* : \theta_1^\top z\geq\theta_1^\top z^* \text{ and } \theta_2^\top z\leq\tau
\end{aligned}$$

and

$$\min_{\theta_1,\theta_2} \frac{1}{2}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|_{V_w}^2}{\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|_{V_w}^2}{\gamma^2}\right) \tag{3}$$

$$\textbf{s.t.} \quad \theta_2^\top z^* > \tau \tag{4}$$

Then the Lagrange multiplier of the first suboptimization problem is

$$\mathcal{L}(\theta_1,\theta_2,\mu,\nu)=\frac{1}{2}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|_{V_w}^2}{\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|_{V_w}^2}{\gamma^2}\right)-\mu(\theta_1^\top z-\theta_1^\top z^*)-\nu(\tau-\theta_2^\top z).$$

Derivative with respect to θ_1:

$$\frac{\partial\mathcal{L}}{\partial\theta_1}=\frac{1}{\sigma^2}V_w(\theta_1-\theta^r)-\mu(z-z^*),$$

and solving $\frac{\partial\mathcal{L}}{\partial\theta_1}=0$ gives:

$$\theta_1-\theta^{\mathrm{r}}=\mu\sigma^2V_w^{-1}(z-z^*). \tag{5}$$

Derivative with respect to θ_2:

$$\frac{\partial\mathcal{L}}{\partial\theta_2}=\frac{1}{\gamma^2}V_w(\theta_2-\theta^{\mathrm{c}})+\nu z,$$

and solving $\frac{\partial\mathcal{L}}{\partial\theta_2}=0$ gives:

$$\theta_2-\theta^{\mathrm{c}}=-\nu\gamma^2V_w^{-1}z.$$

Derivative with respect to μ:

$$\frac{\partial\mathcal{L}}{\partial\mu}=-\theta_1^\top z+\theta_1^\top z^*,$$

and the condition is:

$$\theta_1^\top z=\theta_1^\top z^*.$$

Derivative with respect to ν:

$$\frac{\partial\mathcal{L}}{\partial\nu}=-\tau+\theta_2^\top z,$$

and the condition is:

$$\theta_2^\top z=\tau.$$

From Eqn. (5), we have

$$\|\theta^{\mathrm{r}}-\theta_1\|_{V_w}^2=\mu^2\sigma^4\|z-z^*\|_{V_w^{-1}}^2,$$

then

$$\mu\sigma^2=\frac{\|\theta^{\mathrm{r}}-\theta_1\|_{V_w}}{\|z-z^*\|_{V_w^{-1}}}\ . \tag{6}$$

Plugging Eqn. (6) into Eqn. (5), we have

$$(z-z^*)^\top(\theta_1-\theta^{\mathrm{r}})=\|z-z^*\|_{V_w^{-1}}\|\theta^{\mathrm{r}}-\theta_1\|_{V_w}$$

or

$$\|\theta^{\mathrm{r}}-\theta_1\|_{V_w}^2=\frac{((z-z^*)^\top(\theta_1-\theta^{\mathrm{r}}))^2}{\|z-z^*\|_{V_w^{-1}}^2} \tag{7}$$

Similarly, we also have

$$z^\top(\theta^{\mathrm{c}} - \theta_2) = \|\theta_2 - \theta^{\mathrm{c}}\|_{V_w}\|z\|_{V_w^{-1}}.$$

Substitute $\tau = \theta_2^\top z$ into it,

$$z^\top\theta^{\mathrm{c}} - \tau = \|\theta_2 - \theta^{\mathrm{c}}\|_{V_w}\|z\|_{V_w^{-1}}$$

or

$$\|\theta_2 - \theta^{\mathrm{c}}\|_{V_w}^2 = \frac{(\tau - z^\top\theta^{\mathrm{c}})^2}{\|z\|_{V_w^{-1}}^2}. \tag{8}$$

With Eqns. (7) and (8), the optimization problem becomes

$$\begin{aligned} &\min_{\theta_1,\theta_2} \frac{1}{2}\left(\frac{((z-z^*)^\top(\theta^{\mathrm{r}} - \theta_1))^2}{\sigma^2\|z-z^*\|_{V_w^{-1}}^2} + \frac{(\tau - z^\top\theta^{\mathrm{c}})^2}{\gamma^2\|z\|_{V_w^{-1}}^2}\right) \\ &\textbf{s.t.} \quad \exists z \in \mathcal{Z}\backslash z^* : \theta_1^\top z \geq \theta_1^\top z^* \text{ and } \theta_2^\top z \leq \tau. \end{aligned} \tag{9}$$

If $z \in \mathcal{A}_1$, to minimize the objective function, $\theta_1 = \theta^{\mathrm{r}}$, the minimum objective function value will then be $\frac{(\Delta^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|_{V_w^{-1}}^2}\mathbb{1}\{z \in \mathcal{A}_1\}$;

if $z \in \mathcal{A}_2$, to minimize the objective function, $\theta_2 = \theta^{\mathrm{c}}$ since in this case $\tau = \theta_2^\top z = (\theta^{\mathrm{c}})^\top z$, which will make the second term in Eqn. (9) to be zero. On the other hand for the first term, since

$$(z-z^*)^\top(\theta_1 - \theta^{\mathrm{r}}) = (z-z^*)^\top\theta_1 + (z^*-z)^\top\theta^{\mathrm{r}} = (z-z^*)^\top\theta_1 + \Delta^{\mathrm{r}}(z),$$

then with the constraint $\theta_1^\top z \geq \theta_1^\top z^*$, the minimum will occur at θ_1 such that $\theta_1^\top z = \theta_1^\top z^*$. Hence the minimum objective value will then be $\frac{(\Delta^{\mathrm{r}}(z))^2}{2\sigma^2\|z-z^*\|_{V_w^{-1}}^2}\mathbb{1}\{z \in \mathcal{A}_2\}$. Similarly, we can see if $z \in \mathcal{A}_3$, the minimum will be $(\frac{(\Delta^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|_{V_w^{-1}}^2} + \frac{(\Delta^{\mathrm{r}}(z))^2}{2\sigma^2\|z-z^*\|_{V_w^{-1}}^2})\mathbb{1}\{z \in \mathcal{A}_3\}$.

For the second optimization problem, recall that it is

$$\begin{aligned} &\min_{\theta_1,\theta_2} \frac{1}{2}\left(\frac{\|\theta_1 - \theta^{\mathrm{r}}\|_{V_w}^2}{\sigma^2} + \frac{\|\theta_2 - \theta^{\mathrm{c}}\|_{V_w}^2}{\gamma^2}\right) \\ &\textbf{s.t.} \quad \theta_2^\top z^* > \tau. \end{aligned}$$

First since the first term and second term are independent, the minimum occurs when $\theta_1 = \theta^{\mathrm{r}}$, then we assume $\theta_2^\top z^* = \tau + \alpha$, where $\alpha \geq 0$, then the Langrangian is

$$\mathcal{L}(\theta_2, \mu) = \frac{\|\theta_2 - \theta^{\mathrm{c}}\|_{V_w}^2}{2\gamma^2} + \mu(\tau - \theta_2^\top z^* + \alpha).$$

Then

$$\frac{\partial\mathcal{L}}{\partial\theta_2} = \frac{1}{\gamma^2}V_w(\theta_2 - \theta^{\mathrm{c}}) - \mu z^* = 0.$$

therefore

$$\theta_2 - \theta^{\mathrm{c}} = \gamma^2\mu V_w^{-1}z^*,$$

which implies that

$$\|\theta_2 - \theta^{\mathrm{c}}\|_{V_w}^2 = \gamma^4\mu^2\|z^*\|_{V_w^{-1}}^2,$$

and which leads to

$$\gamma^2\mu = \frac{\|\theta_2 - \theta^{\mathrm{c}}\|_{V_w}}{\|z^*\|_{V_w{}^{-1}}}.$$

Hence

$$(\theta_2 - \theta^{\mathrm{c}})^\top z^* = \|\theta_2 - \theta^{\mathrm{c}}\|_{V_w}\|z^*\|_{V_w{}^{-1}}.$$

The optimization problem then becomes

$$\begin{aligned}\min_{\theta_2} \quad & \frac{((\theta_2 - \theta^{\mathrm{c}})^\top z^*)^2}{2\gamma^2\|z^*\|^2_{V_w{}^{-1}}} \\ \textbf{s.t.} \quad & \theta_2^\top z^* = \tau + \alpha,\end{aligned}$$

since $(\theta_2 - \theta^{\mathrm{c}})^\top z^* = \tau - (\theta^{\mathrm{c}})^\top z^* + \alpha = \Delta^{\mathrm{c}}(z^*) + \alpha$, then when we choose $\alpha \to 0$, we get the minimum value $\frac{(\Delta^{\mathrm{c}}(z^*))^2}{2\gamma^2\|z^*\|^2_{V_w{}^{-1}}}$.

To summarize, the original optimization problem is equivalent to

$$\begin{aligned}\min_{z\in\mathcal{Z}} \Bigg\{ & \frac{(\Delta^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_w^{-1}}}\mathbb{1}\left\{z \in \mathcal{A}_1\right\}, \frac{(\Delta^{\mathrm{r}}(z))^2}{2\sigma^2\|z - z^*\|^2_{V_w^{-1}}}\mathbb{1}\left\{z \in \mathcal{A}_2\right\}, \\ & \left(\frac{(\Delta^{\mathrm{c}}(z))^2}{2\gamma^2\|z\|^2_{V_w^{-1}}} + \frac{(\Delta^{\mathrm{r}}(z))^2}{2\sigma^2\|z - z^*\|^2_{V_w^{-1}}}\right)\mathbb{1}\left\{z \in \mathcal{A}_3\right\}, \frac{(\Delta^{\mathrm{c}}(z^*))^2}{2\gamma^2\|z^*\|^2_{V_w^{-1}}}\Bigg\}.\end{aligned}$$

Hence the theorem is proved.

D PROOF OF THEOREM 4.3

During the proof of Theorem 4.3, we need the help of the good event lemmas in Section E. When these good events hold, the error probability of our algorithm is

$$\begin{aligned}&\mathbb{P}_{(\theta^{\mathrm{r}}_{T+1},\theta^{\mathrm{c}}_{T+1})\sim p_{T+1}}(\hat{z}_{\text{out}} \neq z^*) \\ &\doteq \frac{\int_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}} p_{T+1}(\theta_1,\theta_2)\,\mathrm{d}\theta_1\mathrm{d}\theta_2}{\int_{(\theta_1,\theta_2)\in\Theta_{z^*}} p_{T+1}(\theta_1,\theta_2)\,\mathrm{d}\theta_1\mathrm{d}\theta_2} \\ &= \frac{\int_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}} p_{T+1}(\theta_1,\theta_2)/p_{T+1}(\theta^{\mathrm{r}},\theta^{\mathrm{c}})\,\mathrm{d}\theta_1\mathrm{d}\theta_2}{\int_{(\theta_1,\theta_2)\in\Theta_{z^*}} p_{T+1}(\theta_1,\theta_2)/p_{T+1}(\theta^{\mathrm{r}},\theta^{\mathrm{c}})\,\mathrm{d}\theta_1\mathrm{d}\theta_2} \\ &\overset{(a)}{\doteq} \frac{\int\limits_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}} \exp\left(M\right)\,\mathrm{d}\theta_1\mathrm{d}\theta_2}{\int\limits_{(\theta_1,\theta_2)\in\Theta_{z^*}} \exp\left(M\right)\,\mathrm{d}\theta_1\mathrm{d}\theta_2}\end{aligned}$$

where $M := -\frac{1}{2}\left(\frac{\|\theta^{\mathrm{r}}-\theta_1\|^2_{V_T}}{\sigma^2} + \frac{\|\theta^{\mathrm{c}}-\theta_2\|^2_{V_T}}{\gamma^2}\right)$ and (a) comes from the definition of the good event $E_{5,\infty}$. Notice that when T is enough large, at the last round the probability that all arms are empirically identified as infeasible is asymptotically equal to zero according to the good event in Lemma 5.4. Furthermore, using the Laplace approximation (stated in Lemma G.3) and the fact that

$$\inf_{(\theta_1,\theta_2)\in\Theta} \frac{\|\theta^{\mathrm{r}} - \theta_1\|^2_{V_T}}{\sigma^2} + \frac{\|\theta^{\mathrm{c}} - \theta_2\|^2_{V_T}}{\gamma^2} = 0\,,$$

we have

$$\mathbb{P}_{(\theta^{\mathrm{r}}_{T+1},\theta^{\mathrm{c}}_{T+1})\sim p_{T+1}}(\hat{z}_{\text{out}} \neq z^*) \doteq \exp\left(-\frac{T}{2}\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta^{\mathrm{r}} - \theta_1\|^2_{\overline{V}_T}}{\sigma^2} + \frac{\|\theta^{\mathrm{c}} - \theta_2\|^2_{\overline{V}_T}}{\gamma^2}\right)\right).$$

Then combining with the good event $E_{6,\delta}$, we have

$$\mathbb{P}_{(\theta^{\mathrm{r}}_{T+1},\theta^{\mathrm{c}}_{T+1})\sim p_{T+1}}(\hat{z}_{\text{out}} \neq z^*) \doteq \exp\left(-T\Gamma\right).$$

To summarize, with the choice $\delta = \frac{1}{T}$, with probability 1,

$$\lim_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{(\theta^{\mathrm{r}}_{T+1},\theta^{\mathrm{c}}_{T+1})\sim p_{T+1}}(\hat{z}_{\text{out}} \neq z^*) = \Gamma\ .$$

E GOOD EVENT LEMMAS

Lemma E.1. *Define the good event 1 as*

$$E_{1,\delta} := \left\{ \|\hat{\theta}^{\mathrm{r}}_t - \theta^{\mathrm{r}}\|_{V_{t-1}} \leq \sqrt{\beta_1(t, \frac{1}{\delta^2})} \quad \textit{and} \quad \|\hat{\theta}^{\mathrm{c}}_t - \theta^{\mathrm{c}}\|_{V_{t-1}} \leq \sqrt{\beta_2(t, \frac{1}{\delta^2})} \quad \forall t \in [T] \right\}$$

where $\beta_1(t, \frac{1}{\delta^2}) := (S_1 + \sigma\sqrt{2\log(\frac{1}{\delta^2}) + d\log\left(\frac{d+tL^2}{d}\right)})^2$, $\beta_2(t, \frac{1}{\delta^2}) := (S_2 + \gamma\sqrt{2\log(\frac{1}{\delta^2}) + d\log\left(\frac{d+tL^2}{d}\right)})^2$. *Then with probability* $1-2\delta$, *good event* $E_{1,\delta}$ *holds.*

Proof. This lemma is proved directly with union bound with Lemmas G.1 and G.2. □

Lemma E.2. *Define good event*

$$E_{2,\delta} := \left\{ \max_{x\in\mathcal{X}} |\langle \hat{\theta}^{\mathrm{r}}_t, x\rangle| \leq B_1 \quad \textit{and} \quad \max_{x\in\mathcal{X}} |\langle \hat{\theta}^{\mathrm{c}}_t, x\rangle| \leq B_2 \quad \forall t \in [T] \right\}$$

where $B_1 = LR_1 + L\sqrt{\beta_1(T, \frac{1}{\delta^2})}$, $B_2 = LR_2 + L\sqrt{\beta_2(T, \frac{1}{\delta^2})}$. *Then*

$$E_{1,\delta} \subseteq E_{2,\delta}$$

Proof. Follow the proof of Lemma D.2 in Li et al. [2024], when event $E_{1,\delta}$ holds, for any $x \in \mathcal{X}$,

$$\begin{aligned}
|\langle x, \hat{\theta}^{\mathrm{r}}_t\rangle| &= |\langle x, \theta^{\mathrm{r}}\rangle| + |\langle x, \hat{\theta}^{\mathrm{r}}_t - \theta^{\mathrm{r}}\rangle| \\
&\leq LR_1 + \|x\|_{V^{-1}_{t-1}} \|\hat{\theta}^{\mathrm{r}}_t - \theta^{\mathrm{r}}\|_{V_{t-1}} \\
&\leq LR_1 + \|x\|_{V^{-1}_{t-1}} \sqrt{\beta_1(T, \frac{1}{\delta^2})} \\
&\leq LR_1 + L\sqrt{\beta_1(T, \frac{1}{\delta^2})}
\end{aligned}$$

and

$$\begin{aligned}
|\langle x, \hat{\theta}^{\mathrm{c}}_t\rangle| &= |\langle x, \theta^{\mathrm{c}}\rangle| + |\langle x, \hat{\theta}^{\mathrm{c}}_t - \theta^{\mathrm{c}}\rangle| \\
&\leq LR_2 + \|x\|_{V^{-1}_{t-1}} \|\hat{\theta}^{\mathrm{c}}_t - \theta^{\mathrm{c}}\|_{V_{t-1}} \\
&\leq LR_2 + \|x\|_{V^{-1}_{t-1}} \sqrt{\beta_2(T, \frac{1}{\delta^2})} \\
&\leq LR_2 + L\sqrt{\beta_2(T, \frac{1}{\delta^2})}.
\end{aligned}$$

Hence $\max_{x\in\mathcal{X}} |\langle \hat{\theta}^{\mathrm{r}}_t, x\rangle| \leq B_1$ and $\max_{x\in\mathcal{X}} |\langle \hat{\theta}^{\mathrm{c}}_t, x\rangle| \leq B_2$ for all $t \in [T]$. □

Lemma E.3 (Lemma C.14 in Li et al. [2024])**.** *Define the event* $E_{3,\delta} := \left\{ V_t \geq \frac{3}{4}A(\lambda^G), \forall t \geq T_1(\delta), x \in \mathcal{X} \right\}$ *where* $T_1(\delta) := \max_{x\in\mathcal{X}} \left(\frac{6\sqrt{\log(|\mathcal{X}|\frac{T}{\delta})}}{\lambda^G_x} \right)^4$. *Then with probability* $1-\delta$, *the event* $E_{3,\delta}$ *holds.*

Lemma E.4. *Define the good event*

$$E_{4,\delta} := \{\hat{z}_t = z^*, \quad \forall t > \max\{T_1(\delta)+1, T_2(\delta)\}\}$$

where $T_2(\delta) := \max\left\{(\frac{m_{max}|\mathcal{X}|}{\Delta^{\mathrm{r}}_{min}}\sqrt{d\beta_1(t,\frac{1}{\delta^2})})^{\frac{8}{3}}, (\frac{n_{max}|\mathcal{X}|}{\Delta^{\mathrm{c}}_{min}}\sqrt{d\beta_2(t,\frac{1}{\delta^2})})^{\frac{8}{3}}\right\}$. *Then* $E_{1,\delta}\cap E_{3,\delta}\subseteq E_{4,\delta}$.

Proof. First from Lemma C.15 in Li et al. [2024] we have under $E_{1,\delta}\cap E_{3,\delta}$, for any $t > T_1(\delta)+1$, we have for any $x\in\mathcal{X}$,

$$|\langle x, \hat{\theta}^{\mathrm{r}}_t - \theta^{\mathrm{r}}\rangle| \le \sqrt{\frac{d}{t^{3/4}}\beta_1(t,\frac{1}{\delta^2})}. \tag{10}$$

and

$$|\langle x, \hat{\theta}^{\mathrm{c}}_t - \theta^{\mathrm{c}}\rangle| \le \sqrt{\frac{d}{t^{3/4}}\beta_2(t,\frac{1}{\delta^2})}.$$

Since $\mathcal{X}$ spans $\mathbb{R}^d$, there exists $z^* - z = \sum_{x\in\mathcal{X}} m(x)x$, $z = \sum_{x\in\mathcal{X}} n(x)x$ for any $z\in\mathcal{Z}$ where $0\le m(x), n(x) < \infty$. Then

$$\begin{aligned}
|\langle z^* - z, \hat{\theta}^{\mathrm{r}}_t - \theta^{\mathrm{r}}\rangle| &= \left|\Big\langle \sum_{x\in\mathcal{X}} m(x)x, \hat{\theta}^{\mathrm{r}}_t - \theta^{\mathrm{r}}\Big\rangle\right| \\
&\le \sum_{x\in\mathcal{X}} m(x)\frac{d}{t^{3/4}}\beta_1(t,\frac{1}{\delta^2}) \\
&\le m_{\max}|\mathcal{X}|\sqrt{\frac{d}{t^{3/4}}\beta_1(t,\frac{1}{\delta^2})}
\end{aligned}$$

and

$$\begin{aligned}
|\langle z, \hat{\theta}^{\mathrm{r}}_t - \theta^{\mathrm{r}}\rangle| &= \left|\Big\langle \sum_{x\in\mathcal{X}} n(x)x, \hat{\theta}^{\mathrm{r}}_t - \theta^{\mathrm{r}}\Big\rangle\right| \\
&\le \sum_{x\in\mathcal{X}} n(x)\frac{d}{t^{3/4}}\beta_2(t,\frac{1}{\delta^2}) \\
&\le n_{\max}|\mathcal{X}|\sqrt{\frac{d}{t^{3/4}}\beta_2(t,\frac{1}{\delta^2})},
\end{aligned}$$

where $m_{\max} := \max_{x\in\mathcal{X}} m(x), n_{\max} := \max_{x\in\mathcal{X}} n(x)$.

Therefore when $t > \max\left\{(\frac{m_{\max}|\mathcal{X}|}{\Delta^{\mathrm{r}}_{\min}}\sqrt{d\beta_1(t,\frac{1}{\delta^2})})^{\frac{8}{3}}, (\frac{n_{\max}|\mathcal{X}|}{\Delta^{\mathrm{c}}_{\min}}\sqrt{d\beta_2(t,\frac{1}{\delta^2})})^{\frac{8}{3}}\right\}$, for all $x\in\mathcal{X}$,

$$|\langle z^* - z, \hat{\theta}^{\mathrm{r}}_t - \theta^{\mathrm{r}}\rangle| < \Delta^{\mathrm{r}}_{\min}$$

and

$$|\langle z, \hat{\theta}^{\mathrm{r}}_t - \theta^{\mathrm{r}}\rangle| < \Delta^{\mathrm{c}}_{\min}\ .$$

Hence $\hat{z}_t = z^*$ and $E_{1,\delta}\cap E_{3,\delta}\subseteq E_{4,\delta}$. □

Lemma E.5. *Define* $E_{5,\infty} := \lim_{T\to\infty}\sup_{\theta_1\in\Theta_1,\theta_2\in\Theta_2}\frac{1}{T}\left|\log\frac{p_{T+1}(\theta_1,\theta_2)}{p_{T+1}(\theta^{\mathrm{r}},\theta^{\mathrm{c}})} + \frac{T}{2}(\frac{\|\theta^{\mathrm{r}}-\theta_1\|^2_{V_T}}{\sigma^2} + \frac{\|\theta^{\mathrm{c}}-\theta_2\|^2_{V_T}}{\gamma^2})\right| = 0$, *with probability 1, the good event* $E_{5,\infty}$ *holds, which also means* $\frac{p_{T+1}(\theta_1,\theta_2)}{p_{T+1}(\theta^{\mathrm{r}},\theta^{\mathrm{c}})} \doteq \exp\left(-\frac{1}{2}(\frac{\|\theta^{\mathrm{r}}-\theta_1\|^2_{V_T}}{\sigma^2} + \frac{\|\theta^{\mathrm{c}}-\theta_2\|^2_{V_T}}{\gamma^2})\right)$.

Proof. Since $p_{T+1} := \mathcal{N}(\hat{\theta}^{\mathrm{r}}_{T+1}, \sigma^2 V_T^{-1})\otimes\mathcal{N}(\hat{\theta}^{\mathrm{c}}_{T+1}, \gamma^2 V_T^{-1})|\Theta$, then

$$\frac{p_{T+1}(\theta_1,\theta_2)}{p_{T+1}(\theta^{\mathrm{r}},\theta^{\mathrm{c}})} = \exp\left(-\frac{1}{2}\Big(\frac{\|\theta_1-\hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T}}{\sigma^2} + \frac{\|\theta_2-\hat{\theta}^{\mathrm{c}}_{T+1}\|^2_{V_T}}{\gamma^2} - \frac{\|\theta^{\mathrm{r}}-\hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T}}{\sigma^2} - \frac{\|\theta^{\mathrm{c}}-\hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T}}{\gamma^2}\Big)\right).$$

Hence,

$$\log \frac{p_{T+1}(\theta_1,\theta_2)}{p_{T+1}(\theta^{\mathrm{r}},\theta^{\mathrm{c}})} + \frac{T}{2}\left(\frac{\|\theta^{\mathrm{r}}-\theta_1\|^2_{\overline{V}_T}}{\sigma^2} + \frac{\|\theta^{\mathrm{c}}-\theta_2\|^2_{\overline{V}_T}}{\gamma^2}\right)$$
$$-\frac{1}{2}\left(\frac{\|\theta_1-\hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T} - \|\theta^{\mathrm{r}}-\hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T} - \|\theta^{\mathrm{r}}-\theta_1\|^2_{V_T}}{\sigma^2} + \frac{\|\theta_2-\hat{\theta}^{\mathrm{c}}_{T+1}\|^2_{V_T} - \|\theta^{\mathrm{c}}-\hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T} - \|\theta^{\mathrm{c}}-\theta_2\|^2_{V_T}}{\gamma^2}\right).$$

Following the proof of Lemma C.3 in Li et al. [2024], when $T \to \infty$, with probability $1-2\delta$,

$$\begin{aligned}
&\left|\|\theta^{\mathrm{r}}-\hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T} - \|\theta_1-\hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T} + \|\theta^{\mathrm{r}}-\theta_1\|^2_{V_T}\right| \\
&= \left|-2\sum_{s=1}^{T}\epsilon_s x_s^\top(\theta^{\mathrm{r}}-\theta_1)\right| \\
&\le O\left(\sigma L R_1\sqrt{T}\sqrt{2d\log\left(\frac{d+TL^2}{d\delta}\right)}\right).
\end{aligned}$$

Similarly, we have:

$$\begin{aligned}
&\left|\|\theta^{\mathrm{c}}-\hat{\theta}^{\mathrm{c}}_{T+1}\|^2_{V_T} - \|\theta_2-\hat{\theta}^{\mathrm{c}}_{T+1}\|^2_{V_T} + \|\theta^{\mathrm{c}}-\theta_2\|^2_{V_T}\right| \\
&= \left|-2\sum_{s=1}^{T}\eta_s x_s^\top(\theta^{\mathrm{c}}-\theta_2)\right| \\
&\le O\left(\gamma L R_2\sqrt{T}\sqrt{2d\log\left(\frac{d+TL^2}{d\delta}\right)}\right).
\end{aligned}$$

With the choice $\delta = \frac{1}{T}$, with probability 1, the event $E_{5,\infty}$ holds. □

Lemma E.6. *Define the good event*

$$\begin{aligned}
E_{6,\delta} := \Bigg\{ \Bigg| &\max_{w\in\Delta_{\mathcal{X}}} \inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}} \frac{1}{2}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{A(w)}}{\sigma^2} + \frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{A(w)}}{\gamma^2}\right) \\
&- \inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}} \frac{1}{2}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{\overline{V}_T}}{\sigma^2} + \frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{\overline{V}_T}}{\gamma^2}\right) \Bigg| \le o(1) \Bigg\}.
\end{aligned}$$

Event $E_{6,\frac{1}{T}}$ holds with probability at least $1-\frac{28}{T}$, conditioned on events $E_{1,\frac{1}{T}}, E_{2,\frac{1}{T}}, E_{3,\frac{1}{T}}, E_{4,\frac{1}{T}}$.

Proof. We decompose the simple regret with the following terms:

$$\begin{aligned}
&\max_{\lambda\in\Delta_{\mathcal{X}}} \mathbb{E}_{\theta\sim\tilde{p}_T}\left[\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{A(\lambda)}}{\sigma^2} + \frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{A(\lambda)}}{\gamma^2}\right] - \min_{p\in\mathcal{P}(\overline{\Theta}_{z^*})} \mathbb{E}_{\theta\sim p}\left[\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{\overline{V}_T}}{\sigma^2} + \frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{\overline{V}_T}}{\gamma^2}\right] \\
&= \max_{\lambda\in\Delta_{\mathcal{X}}} \mathbb{E}_{\theta\sim\tilde{p}_T}\left[\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{A(\lambda)}}{\sigma^2} + \frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{A(\lambda)}}{\gamma^2}\right] - \inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{\overline{V}_T}}{\sigma^2} + \frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{\overline{V}_T}}{\gamma^2}\right) \\
&= \max_{\lambda\in\Delta_{\mathcal{X}}} \mathbb{E}_{\theta\sim\tilde{p}_T}\left[\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{A(\lambda)}}{\sigma^2} + \frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{A(\lambda)}}{\gamma^2}\right] - \frac{1}{T}\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{V_T}}{\sigma^2} + \frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{V_T}}{\gamma^2}\right) + F_6 \\
&= \underbrace{\max_{\lambda\in\Delta_{\mathcal{X}}} \mathbb{E}_{\theta\sim\tilde{p}_T}\left[\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{A(\lambda)}}{\sigma^2} + \frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{A(\lambda)}}{\gamma^2}\right] - \max_{\lambda\in\Delta_{\mathcal{X}}} \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t}\left[\frac{\|\theta_1-\hat{\theta}^{\mathrm{r}}_t\|^2_{A(\lambda)}}{\sigma^2} + \frac{\|\theta_2-\hat{\theta}^{\mathrm{c}}_t\|^2_{A(\lambda)}}{\gamma^2}\right]}_{F_1}
\end{aligned}$$

$$+\underbrace{\max_{\lambda\in\Delta_{\mathcal{X}}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|_{A(\lambda)}^2}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|_{A(\lambda)}^2}{\gamma^2}\right]-\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|_{A(\tilde{\lambda}_t)}^2}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|_{A(\tilde{\lambda}_t)}^2}{\gamma^2}\right]}_{F_2}$$

$$+\underbrace{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|_{A(\tilde{\lambda}_t)}^2}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|_{A(\tilde{\lambda}_t)}^2}{\gamma^2}\right]-\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim \tilde{p}_t}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|_{A(\tilde{\lambda}_t)}^2}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|_{A(\tilde{\lambda}_t)}^2}{\gamma^2}\right]}_{F_3}$$

$$+\underbrace{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim \tilde{p}_t}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|_{A(\tilde{\lambda}_t)}^2}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|_{A(\tilde{\lambda}_t)}^2}{\gamma^2}\right]-\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim \tilde{p}_t}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|_{X_tX_t^\top}^2}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|_{X_tX_t^\top}^2}{\gamma^2}\right]}_{F_4}$$

$$+\underbrace{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim \tilde{p}_t}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|_{X_tX_t^\top}^2}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|_{X_tX_t^\top}^2}{\gamma^2}\right]-\frac{1}{T}\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta_1-\hat{\theta}_{T+1}^{\mathrm{r}}\|_{V_T}^2}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_{T+1}^{\mathrm{c}}\|_{V_T}^2}{\gamma^2}\right)}_{F_5}+F_6.$$

where

$$F_6:=\frac{1}{T}\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta_1-\hat{\theta}_{T+1}\|_{V_T}^2}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_{T+1}\|_{V_T}^2}{\gamma^2}\right)-\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|_{\overline{V}_T}^2}{\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|_{\overline{V}_T}^2}{\gamma^2}\right).$$

By combining Lemmas F.1, F.2, F.6, F.7, F.8, and F.9 with the union bound, this completes the proof of the lemma. □

□

F LEMMAS FOR THE PROOF OF LEMMA 5.6

Lemma F.1. *When the good events $E_{1,\delta}, E_{3,\delta}$ both hold,*

$$\max_{\lambda\in\Delta_{\mathcal{X}}}\mathbb{E}_{\theta\sim\overline{p}_T}\left[\frac{\|\theta_1-\theta^{\mathrm{r}}\|_{A(\lambda)}^2}{\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|_{A(\lambda)}^2}{\gamma^2}\right]$$
$$-\max_{\lambda\in\Delta_{\mathcal{X}}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|_{A(\lambda)}^2}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|_{A(\lambda)}^2}{\gamma^2}\right]\le o(1).$$

Proof. Consider,

$$\begin{aligned}
F_1:=&\max_{\lambda\in\Delta_{\mathcal{X}}}\mathbb{E}_{\theta\sim\overline{p}_T}\left[\frac{\|\theta_1-\theta^{\mathrm{r}}\|_{A(\lambda)}^2}{\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|_{A(\lambda)}^2}{\gamma^2}\right]\\
&-\max_{\lambda\in\Delta_{\mathcal{X}}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|_{A(\lambda)}^2}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|_{A(\lambda)}^2}{\gamma^2}\right]\\
=&\max_{\lambda\in\Delta_{\mathcal{X}}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\theta\sim p_t}\left[\frac{\|\theta_1-\theta^{\mathrm{r}}\|_{A(\lambda)}^2}{\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|_{A(\lambda)}^2}{\gamma^2}\right]\\
&-\max_{\lambda\in\Delta_{\mathcal{X}}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|_{A(\lambda)}^2}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|_{A(\lambda)}^2}{\gamma^2}\right]\\
\le&\max_{\lambda\in\Delta_{\mathcal{X}}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\theta\sim p_t}\left[\frac{\|\theta_1-\theta^{\mathrm{r}}\|_{A(\lambda)}^2-\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|_{A(\lambda)}^2}{\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|_{A(\lambda)}^2-\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|_{A(\lambda)}^2}{\gamma^2}\right].
\end{aligned}$$

When the event $E_{1,\delta}$ and $E_{3,\delta}$ both hold,

$$\begin{aligned}
\|\theta_1 - \theta^{\mathrm{r}}\|_{A(\lambda)}^2 - \|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|_{A(\lambda)}^2 &= (\theta_1 - \theta^{\mathrm{r}})^\top A(\lambda)(\theta_1 - \theta^{\mathrm{r}}) - (\theta_1 - \hat{\theta}_t^{\mathrm{r}})^\top A(\lambda)(\theta_1 - \hat{\theta}_t^{\mathrm{r}}) \\
&= (\theta^{\mathrm{r}} + \hat{\theta}_t^{\mathrm{r}} - 2\theta_1)^\top A(\lambda)(\theta^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}) \\
&= \sum_{x\in\mathcal{X}} \lambda_x (\theta^{\mathrm{r}} + \hat{\theta}_t^{\mathrm{r}} - 2\theta_1)^\top x x^\top (\theta^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}) \\
&\le \max_{x\in\mathcal{X}} (\theta^{\mathrm{r}} + \hat{\theta}_t^{\mathrm{r}} - 2\theta_1)^\top x x^\top (\theta^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}) \\
&\le (3LR_1 + B_1) \max_{x\in\mathcal{X}} x^\top (\theta^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}).
\end{aligned}$$

From Eqn. (10), we have $\max_{x\in\mathcal{X}} |\langle x, \hat{\theta}_t^{\mathrm{r}} - \theta^{\mathrm{r}}\rangle| \le \sqrt{\frac{d}{t^{3/4}}\beta_1(t, \frac{1}{\delta^2})}$ for $t > T_1(\delta) + 1$; from event $E_{1,\delta}$, we have $\max_{x\in\mathcal{X}} |\langle x, \hat{\theta}_t^{\mathrm{r}} - \theta^{\mathrm{r}}\rangle| \le L\sqrt{\beta_1(t, \frac{1}{\delta^2})}$ for $t \ge 1$. Hence

$$\begin{aligned}
\max_{x\in\mathcal{X}} \frac{1}{T}\sum_{t=1}^{T} x^\top(\theta^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}) &= \max_{x\in\mathcal{X}} \frac{1}{T}\sum_{t=1}^{T_1(\delta)+1} x^\top(\theta^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}) + \max_{x\in\mathcal{X}} \frac{1}{T}\sum_{t=T_1(\delta)+2}^{T} x^\top(\theta^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}) \\
&\le \frac{(T_1(\delta)+1)L\sqrt{\beta_1(t, \frac{1}{\delta^2})}}{T} + \frac{1}{T}\sum_{t=T_1(\delta)+2}^{T} \sqrt{\frac{d}{t^{3/4}}\beta_1(t, \frac{1}{\delta^2})} \\
&\le \frac{(T_1(\delta)+1)L\sqrt{\beta_1(t, \frac{1}{\delta^2})}}{T} + \frac{8}{5}\frac{(T+1)^{\frac{5}{8}}}{T}.
\end{aligned}$$

Hence when we choose $\delta = \frac{1}{T}$,

$$\lim_{T\to\infty} \max_{\lambda\in\Delta_{\mathcal{X}}} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{\theta\sim p_t} \frac{\|\theta_1 - \theta^{\mathrm{r}}\|_{A(\lambda)}^2 - \|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|_{A(\lambda)}^2}{\sigma^2} \le o(1).$$

Similarly we can also prove

$$\lim_{T\to\infty} \max_{\lambda\in\Delta_{\mathcal{X}}} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{\theta\sim p_t} \frac{\|\theta_2 - \theta^{\mathrm{c}}\|_{A(\lambda)}^2 - \|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|_{A(\lambda)}^2}{\gamma^2} \le o(1).$$

Hence

$$F_1 \le o(1)\ .$$

□

Lemma F.2. *With probability* $1 - \frac{12}{T}$,

$$F_2 \le o(1).$$

Proof. Consider,

$$\begin{aligned}
F_2 :=& \max_{\lambda\in\Delta_{\mathcal{X}}} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim p_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|_{A(\lambda)}^2}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|_{A(\lambda)}^2}{\gamma^2}\right] \\
&- \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim p_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|_{A(\tilde{\lambda}_t)}^2}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|_{A(\tilde{\lambda}_t)}^2}{\gamma^2}\right]
\end{aligned}$$

$$= \max_{\lambda \in \Delta_{\mathcal{X}}} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t \sim \lambda} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$
$$- \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t \sim \tilde{\lambda}_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right].$$

For any λ,

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t \sim \lambda} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t \sim \tilde{\lambda}_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$
$$= \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t \sim \lambda} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t \sim \tilde{\lambda}_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$
$$- \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \lambda} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] + \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \tilde{\lambda}_t} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$
$$+ \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \lambda} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \tilde{\lambda}_t} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$
$$= \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t \sim \lambda} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t \sim \tilde{\lambda}_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$
$$- \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \lambda} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] + \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \tilde{\lambda}_t} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$
$$+ \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \lambda} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \lambda_t} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$
$$+ \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \lambda_t} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \tilde{\lambda}_t} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$
$$= F_{2,1} + F_{2,2} + F_{2,3}$$

where

$$F_{2,1} := \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t \sim \lambda} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t \sim \tilde{\lambda}_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$
$$- \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \lambda} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] + \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \tilde{\lambda}_t} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$

and

$$F_{2,2} := \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \lambda} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \lambda_t} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$

and

$$F_{2,3} := \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \lambda_t} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \tilde{\lambda}_t} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$

Then with Lemma F.3, F.4, and F.5, we have for any λ, with probability $1-\frac{6}{T}$,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t\sim\lambda}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_tx_t^\top}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_tx_t^\top}}{\gamma^2}\right]\\
&\quad-\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t\sim\tilde{\lambda}_t}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_tx_t^\top}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_tx_t^\top}}{\gamma^2}\right]\\
&=F_{2,1}+F_{2,2}+F_{2,3}\\
&\le o(1)\ .
\end{aligned}$$

Define $\overline{\lambda}:=\arg\max_{\lambda\in\Delta_{\mathcal{X}}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t\sim\lambda}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_tx_t^\top}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_tx_t^\top}}{\gamma^2}\right]$. Next, denote the ε-covering of $\Delta_{\mathcal{X}}$ as $\mathcal{B}_\varepsilon:=\left\{\lambda_1:\|\lambda_1-\overline{\lambda}\|_1\le\varepsilon\right\}$. Then there exists $\lambda_2\in\mathcal{B}_\varepsilon$ such that

$$\begin{aligned}
&\max_{\lambda\in\Delta_{\mathcal{X}}}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t\sim\lambda}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_tx_t^\top}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_tx_t^\top}}{\gamma^2}\right]\\
&\quad-\max_{\lambda\in\mathcal{B}_\varepsilon}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t\sim\lambda}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_tx_t^\top}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_tx_t^\top}}{\gamma^2}\right]\\
&\le\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t\sim\overline{\lambda}}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_tx_t^\top}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_tx_t^\top}}{\gamma^2}\right]-\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t\sim\lambda_2}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_tx_t^\top}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_tx_t^\top}}{\gamma^2}\right]\\
&\overset{(a)}{\le}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t}\left(\frac{(LR_1+B_1)^2}{\sigma^2}+\frac{(LR_2+B_2)^2}{\gamma^2}\right)\|\overline{\lambda}-\lambda_2\|_1\\
&\le T\varepsilon\left(\frac{(LR_1+B_1)^2}{\sigma^2}+\frac{(LR_2+B_2)^2}{\gamma^2}\right)
\end{aligned}$$

The term (a) follows from the fact that

$$\mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t\sim\lambda}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_tx_t^\top}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_tx_t^\top}}{\gamma^2}\right]$$

is $\left(\frac{(LR_1+B_1)^2}{\sigma^2}+\frac{(LR_2+B_2)^2}{\gamma^2}\right)$-Lipschitz for any t under the event $E_{2,\delta}$.

Additionally, since $|\mathcal{B}_\varepsilon|\le\left(\frac{3}{\varepsilon}\right)^{|\mathcal{X}|}$, with probability $1-\frac{6}{T}$, and based on the proof of Lemmas F.3, F.4, and F.5 with $\delta=\frac{1}{T|\mathcal{B}_\varepsilon|}$ and $\varepsilon=\frac{1}{\sqrt{T}}$, we can also conclude that

$$\begin{aligned}
&\max_{\lambda\in\mathcal{B}_\varepsilon}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t\sim\lambda}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_tx_t^\top}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_tx_t^\top}}{\gamma^2}\right]\\
&\quad-\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t\sim\tilde{\lambda}_t}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_tx_t^\top}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_tx_t^\top}}{\gamma^2}\right]\le o(1)
\end{aligned}$$

Hence with probability $1-\frac{12}{T}$,

$$F_2\le\frac{1}{\sqrt{T}}\left(\frac{(LR_1+B_1)^2}{\sigma^2}+\frac{(LR_2+B_2)^2}{\gamma^2}\right)+o(1)=o(1)\ .$$

□

Lemma F.3. *With probability* $1 - \frac{2}{T}$,

$$F_{2,1} \leq o(1) \ .$$

Proof. Denote $\mathcal{F}_t$ as the history up to t as in the algorithm. Since

$$\begin{aligned} F_{2,1} :=& \frac{1}{T}\sum_{t=1}^{T}\Bigg(\mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t\sim\lambda}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2}\right] \\ &- \mathbb{E}_{(\theta_1,\theta_2)\sim p_t, x_t\sim\tilde{\lambda}_t}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2}\right] \\ &- \mathbb{E}_{x_t\sim\lambda}\left[\frac{\|\theta_t^{\mathrm{r}}-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2}+\frac{\|\theta_t^{\mathrm{c}}-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2}\right] \\ &+ \mathbb{E}_{x_t\sim\tilde{\lambda}_t}\left[\frac{\|\theta_t^{\mathrm{r}}-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2}+\frac{\|\theta_t^{\mathrm{c}}-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2}\right]\Bigg) =: \frac{1}{T}\sum_{t=1}^{T} Y_t \end{aligned}$$

One can see that $\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}] = 0$ is a martingale difference since p_t is determined under $\mathcal{F}_{t-1}$.

Furthermore, under the good event $E_{2,\delta}$,

$$\|\theta_t^{\mathrm{r}}-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top} \leq (B_1+LR_1)^2, \qquad \|\theta_t^{\mathrm{c}}-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top} \leq (B_2+LR_2)^2 \ ,$$

then

$$|Y_t| \leq 4\left(\frac{(B_1+LR_1)^2}{\sigma^2}+\frac{(B_2+LR_2)^2}{\gamma^2}\right)$$

Then by Azuma–Hoeffding's inequality, with probabaility at least $1-\frac{2}{T}$,

$$F_{2,1} \leq \sqrt{\frac{8(\frac{(B_1+LR_1)^2}{\sigma^2}+\frac{(B_2+LR_2)^2}{\gamma^2})\log(T)}{T}} \leq o(1) \ .$$

□

Lemma F.4. *With probability* $1 - \frac{2}{T}$,

$$F_{2,2} \leq o(1) \ .$$

Proof. Apply the upper bound of the AdaHedge as in De Rooij et al. [2014], we have

$$\begin{aligned} &\max_{\lambda\in\Delta_{\mathcal{X}}}\sum_{t=1}^{T}\mathbb{E}_{x_t\sim\lambda}\left[\frac{\|\theta_t^{\mathrm{r}}-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2}+\frac{\|\theta_t^{\mathrm{c}}-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2}\right] - \sum_{t=1}^{T}\mathbb{E}_{x_t\sim\lambda_t}\left[\frac{\|\theta_t^{\mathrm{r}}-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2}+\frac{\|\theta_t^{\mathrm{c}}-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2}\right] \\ &\leq 2r_T\sqrt{T\log|\mathcal{X}|} + r_T\Big(\frac{16}{3}\log|\mathcal{X}|+2\Big) \end{aligned}$$

where $r_T := \max_{t\in[T]}\max_{x\in\mathcal{X}} \frac{\|\theta_t^{\mathrm{r}}-\hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2}+\frac{\|\theta_t^{\mathrm{c}}-\hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2}$.

Then with the choice $\delta = \frac{1}{T}$, with probability $1-\frac{2}{T}$, good events $E_{1,\delta}, E_{2,\delta}$ hold for any $t \in [T]$, hence by Lemma 5.1 and Lemma 5.2,

$$r_T \leq \frac{(LR_1+B_1)^2}{\sigma^2}+\frac{(LR_2+B_2)^2}{\gamma^2} = O(\log T)$$

Hence

$$F_{2,2} \le \frac{1}{T} \max_{\lambda \in \Delta_{\mathcal{X}}} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \lambda} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \lambda_t} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$

$$\le O\left(\frac{2 \log T \sqrt{T \log |\mathcal{X}|} + \log T (\frac{16}{3} \log |\mathcal{X}| + 2)}{T} \right) \le o(1) \ .$$

□

Lemma F.5. *With probability* $1 - \frac{2}{T}$,

$$F_{2,3} \le o(1) \ .$$

Proof. Consider,

$$F_{2,3} := \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \lambda_t} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{x_t \sim \tilde{\lambda}_t} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]$$

By the definition of $\tilde{\lambda}_t := (1 - \gamma_t)\lambda_t + \gamma_t \lambda^G$, we have

$$F_{2,3} = \frac{1}{T} \sum_{t=1}^{T} \gamma_t \left(\mathbb{E}_{x_t \sim \lambda_t} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] - \mathbb{E}_{x_t \sim \lambda^G} \left[\frac{\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] \right) .$$

Furthermore, under good event $E_{2,\delta}$,

$$\|\theta_t^{\mathrm{r}} - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top} \le (B_1 + LR_1)^2, \|\theta_t^{\mathrm{c}} - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top} \le (B_2 + LR_2)^2 \ ,$$

Hence with $\gamma_t = t^{-\frac{1}{4}}$, $\delta := \frac{1}{T}$,

$$\begin{aligned} F_{2,3} &\le \frac{2}{T} \left(\frac{(LR_1 + B_1)^2}{\sigma^2} + \frac{(LR_2 + B_2)^2}{\gamma^2} \right) \sum_{t=1}^{T} t^{-\frac{1}{4}} \\ &\le \frac{8}{3T} \left(\frac{(LR_1 + B_1)^2}{\sigma^2} + \frac{(LR_2 + B_2)^2}{\gamma^2} \right) T^{\frac{3}{4}} \\ &\le o(1) \ . \end{aligned}$$

□

Lemma F.6. *With probability* $1 - \frac{3}{T}$,

$$F_3 \le o(1).$$

Proof. When the good events $E_{2,\delta}$ and $E_{4,\delta}$ both hold, by the definition of $\tilde{p}_t$, for $t > \max\{T_1(\delta) + 1, T_2(\delta)\}$, we have $\hat{z}_t = z^*$, and thus $\tilde{p}_t = p_t$.

Hence, with probability $1 - 3\delta$ and $\delta = \frac{1}{T}$, we can write:

$$\begin{aligned} F_3 := \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{(\theta_1, \theta_2) \sim p_t} &\left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{A(\tilde{\lambda}_t)}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{A(\tilde{\lambda}_t)}}{\gamma^2} \right] \\ - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{(\theta_1, \theta_2) \sim \tilde{p}_t} &\left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{A(\tilde{\lambda}_t)}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{A(\tilde{\lambda}_t)}}{\gamma^2} \right] \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{T} \sum_{t=1}^{\max\{T_1(\delta)+1, T_2(\delta)\}} \mathbb{E}_{(\theta_1,\theta_2)\sim p_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{A(\tilde{\lambda}_t)}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{A(\tilde{\lambda}_t)}}{\gamma^2} \right] \\
&\quad - \frac{1}{T} \sum_{t=1}^{\max\{T_1(\delta)+1, T_2(\delta)\}} \mathbb{E}_{(\theta_1,\theta_2)\sim \tilde{p}_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{A(\tilde{\lambda}_t)}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{A(\tilde{\lambda}_t)}}{\gamma^2} \right] \\
&\leq \frac{2\max\{T_1(\delta)+1, T_2(\delta)\} \left(\frac{(LR_1+B_1)^2}{\sigma^2} + \frac{(LR_2+B_2)^2}{\gamma^2} \right)}{T} \\
&= o(1).
\end{aligned}
$$

□

Lemma F.7. *With probability* $1 - \frac{2}{T}$,

$$
F_4 \leq o(1).
$$

Proof. We define F_4 as follows:

$$
\begin{aligned}
F_4 := &\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim \tilde{p}_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{A(\tilde{\lambda}_t)}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{A(\tilde{\lambda}_t)}}{\gamma^2} \right] \\
&- \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim \tilde{p}_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right].
\end{aligned}
$$

Since

$$
\mathbb{E}_{x_t} \left[\mathbb{E}_{(\theta_1,\theta_2)\sim \tilde{p}_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] \middle| \mathcal{F}_{t-1} \right] = \mathbb{E}_{(\theta_1,\theta_2)\sim \tilde{p}_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{A(\tilde{\lambda}_t)}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{A(\tilde{\lambda}_t)}}{\gamma^2} \right],
$$

we observe that:

$$
\mathbb{E}_{(\theta_1,\theta_2)\sim \tilde{p}_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{A(\tilde{\lambda}_t)}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{A(\tilde{\lambda}_t)}}{\gamma^2} \right] - \mathbb{E}_{(\theta_1,\theta_2)\sim \tilde{p}_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right]
$$

is a martingale difference sequence.

When the good event $E_{2,\delta}$ holds, we have:

$$
\begin{aligned}
&\left| \mathbb{E}_{(\theta_1,\theta_2)\sim \tilde{p}_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{A(\tilde{\lambda}_t)}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{A(\tilde{\lambda}_t)}}{\gamma^2} \right] - \mathbb{E}_{(\theta_1,\theta_2)\sim \tilde{p}_t} \left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{x_t x_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{x_t x_t^\top}}{\gamma^2} \right] \right| \\
&\qquad \leq 2 \left(\frac{(LR_1+B_1)^2}{\sigma^2} + \frac{(LR_2+B_2)^2}{\gamma^2} \right).
\end{aligned}
$$

Then, with probability at least $1 - \frac{2}{T}$ and $\delta = \frac{1}{T}$, we have:

$$
F_4 \leq \sqrt{\frac{2\left(\frac{(LR_1+B_1)^2}{\sigma^2} + \frac{(LR_2+B_2)^2}{\gamma^2} \right) \log T}{T}} \leq o(1).
$$

□

Lemma F.8. *With probability* $1 - \frac{4}{T}$,

$$F_5 := \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{(\theta_1,\theta_2)\sim\tilde{p}_t}\left[\frac{\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{X_t X_t^\top}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{X_t X_t^\top}}{\gamma^2}\right] - \frac{1}{T}\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta_1 - \hat{\theta}_{T+1}^{\mathrm{r}}\|^2_{V_T}}{\sigma^2} + \frac{\|\theta_2 - \hat{\theta}_{T+1}^{\mathrm{c}}\|^2_{V_T}}{\gamma^2}\right) \le o(1).$$

Proof. Since $\tilde{p}_t := \mathcal{N}(\hat{\theta}_t^{\mathrm{r}}, \eta_r^{-1}V_{t-1}^{-1}) \otimes \mathcal{N}(\hat{\theta}_t^{\mathrm{c}}, \eta_c^{-1}V_{t-1}^{-1})|\overline{\Theta}_{z^*}$, we define the normalization term at time t as

$$W_t := \int_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}} \exp\left(-\frac{1}{2}\left[\eta_r\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{V_{t-1}} + \eta_c\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{V_{t-1}}\right]\right)\mathrm{d}\theta_1\mathrm{d}\theta_2$$

Then we have for $t > 1$,

$$\begin{aligned}\log\frac{W_t}{W_{t-1}} &:= \log\frac{\int_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}} \exp\left(-\frac{1}{2}\left[\eta_r\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{V_{t-1}} + \eta_c\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{V_{t-1}}\right]\right)\mathrm{d}\theta_1\mathrm{d}\theta_2}{W_{t-1}} \\ &= \log\frac{\int_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}} \exp\left(-\frac{1}{2}\left[\eta_r\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{V_{t-1}} + \eta_c\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{V_{t-1}} + w_{t-1} - w_{t-1}\right]\right)\mathrm{d}\theta_1\mathrm{d}\theta_2}{W_{t-1}}\end{aligned}$$

where $w_t := \eta_r\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{V_{t-1}} + \eta_c\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{V_{t-1}}$

Then

$$\begin{aligned}\log\frac{W_t}{W_{t-1}} &= \log\mathbb{E}_{(\theta_1,\theta_2)\sim\tilde{p}_t}\exp\left(-\frac{1}{2}\left[\eta_r\|\theta_1 - \hat{\theta}_t^{\mathrm{r}}\|^2_{V_{t-1}} + \eta_c\|\theta_2 - \hat{\theta}_t^{\mathrm{c}}\|^2_{V_{t-1}} - w_{t-1}\right]\right) \\ &= \log\mathbb{E}_{(\theta_1,\theta_2)\sim\tilde{p}_t}\exp\left(-\frac{1}{2}\Gamma_1 + \frac{1}{2}\Gamma_2\right) \\ &\overset{(a)}{\le} \frac{1}{2}\log\mathbb{E}_{(\theta_1,\theta_2)\sim\tilde{p}_t}\exp(-\Gamma_1) + \frac{1}{2}\log\mathbb{E}_{(\theta_1,\theta_2)\in\tilde{p}_t}\exp(\Gamma_2)\end{aligned}$$

where

$$\begin{aligned}\Gamma_1 &:= \eta_r\|\hat{\theta}_{t-1}^{\mathrm{r}} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} + \eta_c\|\hat{\theta}_{t-1}^{\mathrm{c}} - \theta_2\|^2_{x_{t-1}x_{t-1}^\top} \\ \Gamma_2 &:= \eta_r\|\hat{\theta}_{t-1}^{\mathrm{r}} - \theta_1\|^2_{V_{t-1}} - \eta_r\|\hat{\theta}_t^{\mathrm{r}} - \theta_1\|^2_{V_{t-1}} + \eta_c\|\hat{\theta}_{t-1}^{\mathrm{c}} - \theta_2\|^2_{V_{t-1}} - \eta_c\|\hat{\theta}_t^{\mathrm{c}} - \theta_2\|^2_{V_{t-1}}\end{aligned}$$

and (a) follows from the Cauchy–Schwarz inequality. Next,

$$\begin{aligned}-\|\hat{\theta}_{t-1}^{\mathrm{r}} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} &\le -\|\theta^{\mathrm{r}} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} + 2(\theta^{\mathrm{r}} - \hat{\theta}_{t-1}^{\mathrm{r}})^\top(x_{t-1}x_{t-1}^\top)(\theta^{\mathrm{r}} - \theta_1) \\ &\le -\|\theta^{\mathrm{r}} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} + 2\|\theta^{\mathrm{r}} - \hat{\theta}_{t-1}^{\mathrm{r}}\|_{V_{t-2}}\|x_{t-1}\|_{V_{t-2}^{-1}}\|\theta^{\mathrm{r}} - \theta_1\|_{x_{t-1}x_{t-1}^\top} \\ &\le -\|\theta^{\mathrm{r}} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} + 2\|\theta^{\mathrm{r}} - \hat{\theta}_{t-1}^{\mathrm{r}}\|_{V_{t-2}}\|x_{t-1}\|_{V_{t-2}^{-1}}D_1 \\ &= -\|\theta^{\mathrm{r}} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} + a_1\end{aligned}$$

where $a_1 := 2\|\theta^{\mathrm{r}} - \hat{\theta}_{t-1}^{\mathrm{r}}\|_{V_{t-2}}\|x_{t-1}\|_{V_{t-2}^{-1}}D_1$ and $D_1 := \max_{x\in\mathcal{X}} x^\top(\theta_1 - \theta^{\mathrm{r}})$.

On the other hand, since

$$\begin{aligned}&-\|\theta^{\mathrm{r}} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} \le -\|\hat{\theta}_{t-1}^{\mathrm{r}} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} - 2(\theta^{\mathrm{r}} - \hat{\theta}_{t-1}^{\mathrm{r}})^\top(x_{t-1}x_{t-1}^\top)(\hat{\theta}_{t-1}^{\mathrm{r}} - \theta_1) \\ &\le -\|\hat{\theta}_{t-1}^{\mathrm{r}} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} + 2\|\theta^{\mathrm{r}} - \hat{\theta}_{t-1}^{\mathrm{r}}\|_{V_{t-2}}\|x_{t-1}\|_{V_{t-2}^{-1}}\|\hat{\theta}_{t-1}^{\mathrm{r}} - \theta_1\|_{x_{t-1}x_{t-1}^\top}. \\ &\le -\|\hat{\theta}_{t-1}^{\mathrm{r}} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} + 2\|\theta^{\mathrm{r}} - \hat{\theta}_{t-1}^{\mathrm{r}}\|_{V_{t-2}}\|x_{t-1}\|_{V_{t-2}^{-1}}(\|\theta^{\mathrm{r}} - \hat{\theta}_{t-1}^{\mathrm{r}}\|_{x_{t-1}x_{t-1}^\top} + \|\theta^{\mathrm{r}} - \theta_1\|_{x_{t-1}x_{t-1}^\top})\end{aligned}$$

$$
\begin{aligned}
&\leq -\|\hat{\theta}^{\mathrm{r}}_{t-1} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} + 2\|\theta^{\mathrm{r}} - \hat{\theta}^{\mathrm{r}}_{t-1}\|_{V_{t-2}}\|x_{t-1}\|_{V_{t-2}^{-1}}(\|\theta^{\mathrm{r}} - \hat{\theta}^{\mathrm{r}}_{t-1}\|_{x_{t-1}x_{t-1}^\top} + D_1) \\
&\leq -\|\hat{\theta}^{\mathrm{r}}_{t-1} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} + 2\|\theta^{\mathrm{r}} - \hat{\theta}^{\mathrm{r}}_{t-1}\|_{V_{t-2}}\|x_{t-1}\|_{V_{t-2}^{-1}}(\|\theta^{\mathrm{r}} - \hat{\theta}^{\mathrm{r}}_{t-1}\|_{V_{t-2}}\|x_{t-1}\|_{V_{t-2}^{-1}} + D_1) \\
&= -\|\hat{\theta}^{\mathrm{r}}_{t-1} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} + a_1 + \frac{a_1^2}{2D_1^2}
\end{aligned}
$$

then

$$
-\|\theta^{\mathrm{r}} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} \leq -\|\hat{\theta}^{\mathrm{r}}_{t-1} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} + 2a_1 + \frac{a_1^2}{2D_1^2}.
$$

Similarly we also have

$$
-\|\theta^{\mathrm{c}} - \theta_2\|^2_{x_{t-1}x_{t-1}^\top} \leq -\|\hat{\theta}^{\mathrm{c}}_{t-1} - \theta_2\|^2_{x_{t-1}x_{t-1}^\top} + 2a_2 + \frac{a_2^2}{2D_2^2},
$$

where $a_2 := 2\|\theta^{\mathrm{c}} - \hat{\theta}^{\mathrm{c}}_{t-1}\|_{V_{t-2}}\|x_{t-1}\|_{V_{t-2}^{-1}} D_1$ and $D_2 := \max_{x\in\mathcal{X}} x^\top(\theta_2 - \theta^{\mathrm{c}})$.

Furthermore, from Corollary 1 in Kone et al. [2024], when we choose

$$
\eta_r \leq \frac{1}{8L^2R_1^2} \quad \text{and} \quad \eta_c \leq \frac{1}{8L^2R_2^2},
$$

whence, $\exp(-\Gamma_1)$ is concave in terms of (θ_1, θ_2).

Then

$$
\begin{aligned}
\log\frac{W_t}{W_{t-1}} &\leq \frac{1}{2}\log \mathbb{E}_{(\theta_1,\theta_2)\sim\tilde{p}_t}\exp(-\Gamma_1) + \frac{1}{2}\log \mathbb{E}_{(\theta_1,\theta_2)\sim\tilde{p}_t}\exp(\Gamma_2) \\
&\leq \frac{1}{2}\left(-\mathbb{E}_{(\theta_1,\theta_2)\sim\tilde{p}_t}\eta_r\|\hat{\theta}^{\mathrm{r}}_{t-1} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} + \eta_c\|\hat{\theta}^{\mathrm{c}}_{t-1} - \theta_2\|^2_{x_{t-1}x_{t-1}^\top} + \eta_r\Big(2a_1 + \frac{a_1^2}{2D_1^2}\Big) + \eta_c\Big(2a_2 + \frac{a_2^2}{2D_2^2}\Big)\right) \\
&\quad + \frac{1}{2}\log \mathbb{E}_{(\theta_1,\theta_2)\in\tilde{p}_t}\exp(\Gamma_2)
\end{aligned}
$$

Then by telescoping

$$
\begin{aligned}
\log\frac{W_{T+1}}{W_1} &= \sum_{t=2}^{T+1}\log\frac{W_t}{W_{t-1}} \\
&\leq \frac{1}{2}\sum_{t=2}^{T+1}\left(-\mathbb{E}_{(\theta_1,\theta_2)\sim\tilde{p}_t}\eta_r\|\hat{\theta}^{\mathrm{r}}_{t-1} - \theta_1\|^2_{x_{t-1}x_{t-1}^\top} + \eta_c\|\hat{\theta}^{\mathrm{c}}_{t-1} - \theta_2\|^2_{x_{t-1}x_{t-1}^\top} + \eta_r\Big(2a_1 + \frac{a_1^2}{2D_1^2}\Big) + \eta_c\Big(2a_2 + \frac{a_2^2}{2D_2^2}\Big)\right) \\
&\quad + \frac{1}{2}\sum_{t=2}^{T+1}\log \mathbb{E}_{(\theta_1,\theta_2)\sim\tilde{p}_t}\exp(\Gamma_2) \qquad (11)
\end{aligned}
$$

On the other hand, let $z > 0$, and $(\tilde{\theta}^{\mathrm{r}}_T, \tilde{\theta}^{\mathrm{c}}_T) = \arg\min_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\Big(\eta_r\|\theta_1 - \hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T} + \eta_c\|\theta_2 - \hat{\theta}^{\mathrm{c}}_{T+1}\|^2_{V_T}\Big)$.

Given that $\overline{\Theta}_{z^*}$ constitutes a union of convex sets, there exists a convex subset $\mathcal{C} \subseteq \overline{\Theta}_{z^*}$ such that $(\tilde{\theta}^{\mathrm{r}}_T, \tilde{\theta}^{\mathrm{c}}_T) \in \mathcal{C}$ and $\mathrm{vol}(\mathcal{C}) > 0$.

Next, define the set $\mathcal{N}_z$ as follows:

$$
\mathcal{N}_z := \big\{(1-z)(\tilde{\theta}^{\mathrm{r}}_T, \tilde{\theta}^{\mathrm{c}}_T) + z(\theta_1,\theta_2) \mid (\theta_1,\theta_2) \in \mathcal{C}\big\} = (1-z)(\tilde{\theta}^{\mathrm{r}}_T, \tilde{\theta}^{\mathrm{c}}_T) + z\mathcal{C},
$$

which represents a convex combination of the point $(\tilde{\theta}^{\mathrm{r}}_T, \tilde{\theta}^{\mathrm{c}}_T)$ and elements from $\mathcal{C}$, parameterized by $z \in [0,1]$.

Hence,

$$
\log\frac{W_{T+1}}{W_1} \geq \log\frac{\int_{\mathcal{C}}\exp\left(-\frac{\eta_r\|\theta_1 - \hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T} + \eta_c\|\theta_2 - \hat{\theta}^{\mathrm{c}}_{T+1}\|^2_{V_T}}{2}\right)\mathrm{d}\theta_1\,\mathrm{d}\theta_2}{W_1}
$$

$$
\begin{aligned}
&\geq \log \frac{\int_{\mathcal{N}_z} \exp\left(-\frac{\eta_r\|\theta_1-\hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T}+\eta_c\|\theta_2-\hat{\theta}^{\mathrm{c}}_{T+1}\|^2_{V_T}}{2}\right)\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2}{W_1} \\
&= \log \frac{\int_{z\mathcal{C}} \exp\left(-\frac{\eta_r\|\hat{\theta}^{\mathrm{r}}_{T+1}-(1-z)\tilde{\theta}^{\mathrm{r}}_{T}-\theta_1\|^2_{V_T}+\eta_c\|\hat{\theta}^{\mathrm{c}}_{T+1}-\tilde{\theta}^{\mathrm{c}}_{T}-\theta_2\|^2_{V_T}}{2}\right)\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2}{W_1} \\
&= \log \frac{\int_{\mathcal{C}} z^{2d} \exp\left(-\frac{\eta_r\|(1-z)(\hat{\theta}^{\mathrm{r}}_{T+1}-\tilde{\theta}^{\mathrm{r}}_{T})+z\hat{\theta}^{\mathrm{r}}_{T+1}-z\theta_1\|^2_{V_T}+\eta_c\|(1-z)(\hat{\theta}^{\mathrm{c}}_{T+1}-\tilde{\theta}^{\mathrm{c}}_{T})+z\hat{\theta}^{\mathrm{c}}_{T+1}-z\theta_2\|^2_{V_T}}{2}\right)\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2}{W_1} \\
&\geq \log \frac{\int_{\mathcal{C}} z^{2d} \exp\left(-\frac{\eta_r\|(1-z)(\hat{\theta}^{\mathrm{r}}_{T+1}-\tilde{\theta}^{\mathrm{r}}_{T})\|^2_{V_T}+z\|\hat{\theta}^{\mathrm{r}}_{T+1}-\theta_1\|^2_{V_T}+\eta_c\|(1-z)(\hat{\theta}^{\mathrm{c}}_{T+1}-\tilde{\theta}^{\mathrm{c}}_{T})\|^2_{V_T}+z\|\hat{\theta}^{\mathrm{c}}_{T+1}-\theta_2\|^2_{V_T}}{2}\right)\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2}{W_1} \\
&= 2d\log(z) - \frac{\eta_r(1-z)}{2}\|\hat{\theta}^{\mathrm{r}}_{T+1}-\tilde{\theta}^{\mathrm{r}}_{T}\|^2_{V_T} - \frac{\eta_c(1-z)}{2}\|\hat{\theta}^{\mathrm{c}}_{T+1}-\tilde{\theta}^{\mathrm{c}}_{T}\|^2_{V_T} \\
&\qquad + \log \frac{\int_{\mathcal{C}} \exp\left(-\frac{\eta_r z\|\theta_1-\hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T}+\eta_c z\|\theta_2-\hat{\theta}^{\mathrm{c}}_{T+1}\|^2_{V_T}}{2}\right)\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2}{W_1} \\
&\geq 2d\log(z) - \frac{\eta_r(1-z)}{2}\|\hat{\theta}^{\mathrm{r}}_{T+1}-\tilde{\theta}^{\mathrm{r}}_{T}\|^2_{V_T} - \frac{\eta_c(1-z)}{2}\|\hat{\theta}^{\mathrm{c}}_{T+1}-\tilde{\theta}^{\mathrm{c}}_{T}\|^2_{V_T} \\
&\qquad - \mathbb{E}_{(\theta_1,\theta_2)\sim\mathrm{Uni}(\mathcal{C})}\frac{\eta_r z\|\theta_1-\hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T}+\eta_c z\|\theta_2-\hat{\theta}^{\mathrm{c}}_{T+1}\|^2_{V_T}}{2} + \log\frac{\mathrm{vol}(\mathcal{C})}{W_1}.
\end{aligned}
$$

Then with Eqn. (11), we have

$$
\begin{aligned}
&\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim\tilde{p}_t}\left(\eta_r\|\hat{\theta}^{\mathrm{r}}_t-\theta_1\|^2_{x_tx_t^\top}+\eta_c\|\hat{\theta}^{\mathrm{c}}_t-\theta_2\|^2_{x_tx_t^\top}\right) - \frac{1-z}{2}\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\eta_r\|\theta_1-\hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T}+\eta_c\|\theta_2-\hat{\theta}^{\mathrm{c}}_{T+1}\|^2_{V_T}\right) \\
&\leq \frac{1}{2}\sum_{t=2}^{T+1}\left[\eta_r\left(2a_1+\frac{a_1^2}{2D_1^2}\right)+\eta_c\left(2a_2+\frac{a_2^2}{2D_2^2}\right)\right] + \mathbb{E}_{(\theta_1,\theta_2)\sim\mathrm{Uni}(\mathcal{C})}\frac{\eta_r z\|\theta_1-\hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T}+\eta_c z\|\theta_2-\hat{\theta}^{\mathrm{c}}_{T+1}\|^2_{V_T}}{2} \\
&- 2d\log z - \log\frac{\mathrm{vol}(\mathcal{C})}{W_1} + \frac{1}{2}\sum_{t=2}^{T+1}\log\mathbb{E}_{(\theta_1,\theta_2)\in\tilde{p}_t}\exp\left(\Gamma_2\right).
\end{aligned}
$$

Then with the choice $z=\frac{1}{T}$,

$$
\eta_r := \frac{\eta}{\sigma^2}, \quad \eta_c := \frac{\eta}{\gamma^2} \quad \text{and} \quad \eta := \min\left\{\frac{\sigma^2}{8L^2R_1^2}, \frac{\gamma^2}{8L^2R_2^2}\right\}.
$$

From Lemma 12 in Kone et al. [2024], we have with probability $1-\frac{4}{T}$,

$$
\begin{aligned}
\frac{1}{2}\sum_{t=2}^{T+1}\left[\eta_r\left(2a_1+\frac{a_1^2}{2D_1^2}\right)+\eta_c\left(2a_2+\frac{a_2^2}{2D_2^2}\right)\right] &= O(\sqrt{T\log^2 T}), \quad \text{and} \\
\frac{1}{2}\sum_{t=2}^{T+1}\log\mathbb{E}_{(\theta_1,\theta_2)\sim\tilde{p}_t}\exp\left(\Gamma_2\right) &= O(\sqrt{T\log^2 T}).
\end{aligned}
$$

Also, under good event $E_{1,\frac{1}{T}}$, we have

$$
\begin{aligned}
\mathbb{E}_{(\theta_1,\theta_2)\sim\mathrm{Uni}(\mathcal{C})}\frac{\eta_r\gamma\|\theta_1-\hat{\theta}^{\mathrm{r}}_{T+1}\|^2_{V_T}+\eta_c\gamma\|\theta_2-\hat{\theta}^{\mathrm{c}}_{T+1}\|^2_{V_T}}{2} &= O(1), \quad \text{and} \\
-2d\log z = 2d\log T &= O(\log T).
\end{aligned}
$$

Since $\overline{\Theta}_{z^*}$ is finite and $\mathrm{vol}(\mathcal{C})>0$, observe from the definition of W_t that $W_1 \leq \mathrm{vol}(\overline{\Theta}_{z^*})$. Thus, $-\log\frac{\mathrm{vol}(\mathcal{C})}{W_1} = O(1)$.

To summarize, with probability $1-\frac{4}{T}$,

$$F_5 := \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{(\theta_1,\theta_2)\sim\tilde{p}_t}\left[\frac{\|\theta_1-\hat{\theta}_t^{\mathrm{r}}\|^2_{X_tX_t^\top}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_t^{\mathrm{c}}\|^2_{X_tX_t^\top}}{\gamma^2}\right]-\frac{1}{T}\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta_1-\hat{\theta}_{T+1}^{\mathrm{r}}\|^2_{V_T}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_{T+1}^{\mathrm{c}}\|^2_{V_T}}{\gamma^2}\right)$$
$$\leq O\Big(\sqrt{\frac{\log^2 T}{T}}\Big)=o(1).$$

□

Lemma F.9. *With probability* $1-\frac{4}{T}$,

$$F_6 := \frac{1}{T}\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta_1-\hat{\theta}_{T+1}\|^2_{V_T}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_{T+1}\|^2_{V_T}}{\gamma^2}\right)-\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{V_T}}{\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{V_T}}{\gamma^2}\right)\leq o(1).$$

Proof. Denote $(\theta_3,\theta_4) := \arg\min_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{V_T}}{\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{V_T}}{\gamma^2}\right)$. Then,

$$\begin{aligned}
F_6 &= \frac{1}{T}\Bigg[\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta_1-\hat{\theta}_{T+1}\|^2_{V_T}}{\sigma^2}+\frac{\|\theta_2-\hat{\theta}_{T+1}\|^2_{V_T}}{\gamma^2}\right)\\
&\quad-\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_{z^*}}\left(\frac{\|\theta_1-\theta^{\mathrm{r}}\|^2_{V_T}}{\sigma^2}+\frac{\|\theta_2-\theta^{\mathrm{c}}\|^2_{V_T}}{\gamma^2}\right)\Bigg]\\
&\leq \frac{1}{T}\left[\left(\frac{\|\theta_3-\hat{\theta}_{T+1}\|^2_{V_T}}{\sigma^2}+\frac{\|\theta_4-\hat{\theta}_{T+1}\|^2_{V_T}}{\gamma^2}\right)-\left(\frac{\|\theta_3-\theta^{\mathrm{r}}\|^2_{V_T}}{\sigma^2}+\frac{\|\theta_4-\theta^{\mathrm{c}}\|^2_{V_T}}{\gamma^2}\right)\right]\\
&= \frac{1}{T}\Bigg[\frac{(\|\theta_3-\hat{\theta}_{T+1}\|_{V_T}-\|\theta_3-\theta^{\mathrm{r}}\|_{V_T})(\|\theta_3-\hat{\theta}_{T+1}\|_{V_T}+\|\theta_3-\theta^{\mathrm{r}}\|_{V_T})}{\sigma^2}\\
&\quad+\frac{(\|\theta_4-\hat{\theta}_{T+1}\|_{V_T}-\|\theta_4-\theta^{\mathrm{c}}\|_{V_T})(\|\theta_4-\hat{\theta}_{T+1}\|_{V_T}-\|\theta_4+\theta^{\mathrm{c}}\|_{V_T})}{\gamma^2}\Bigg]\\
&= \frac{1}{T}\Bigg[\frac{\|\theta^{\mathrm{r}}-\hat{\theta}_{T+1}\|_{V_T}(\|\theta_3-\hat{\theta}_{T+1}\|_{V_T}+\|\theta_3-\theta^{\mathrm{r}}\|_{V_T})}{\sigma^2}\\
&\quad+\frac{\|\theta^{\mathrm{c}}-\hat{\theta}_{T+1}\|_{V_T}(\|\theta_4-\hat{\theta}_{T+1}\|_{V_T}-\|\theta_4+\theta^{\mathrm{c}}\|_{V_T})}{\gamma^2}\Bigg]\\
&\overset{(a)}{\leq} \frac{1}{T}\Bigg[\frac{\|\theta^{\mathrm{r}}-\hat{\theta}_{T+1}\|_{V_T}(3LR_1+B_1)\sqrt{T}}{\sigma^2}\\
&\quad+\frac{\|\theta^{\mathrm{c}}-\hat{\theta}_{T+1}\|_{V_T}(3LR_2+B_2)\sqrt{T}}{\gamma^2}\Bigg]\\
&\overset{(b)}{\leq} \frac{1}{\sqrt{T}}\Bigg[\frac{\sqrt{\beta_1(T+1,\frac{1}{\delta^2})}(3LR_1+B_1)}{\sigma^2}\\
&\quad+\frac{\sqrt{\beta_2(T+1,\frac{1}{\delta^2})}(3LR_2+B_2)}{\gamma^2}\Bigg]
\end{aligned}$$

where (a) comes from the good event $E_{2,\delta}$, and (b) comes from Lemmas G.1 and G.2.

Then, with probability $1-4\delta$ and with the choice $\delta=\frac{1}{T}$,

$$F_6\leq\frac{1}{\sqrt{T}}\left[\frac{\sqrt{\beta_1(T+1,T^2)}(3LR_1+B_1)}{\sigma^2}+\frac{\sqrt{\beta_2(T+1,T^2)}(3LR_2+B_2)}{\gamma^2}\right]=o(1).$$

□

G USEFUL LEMMAS

In this section we present some useful existing lemmas for our proof.

Lemma G.1 (Theorem 2 in Abbasi-Yadkori et al. [2011]). *For all $t > 0$, with probability $1-\delta$,*

$$\|\hat{\theta}_t^{\mathrm{r}} - \theta^{\mathrm{r}}\|_{V_{t-1}} \leq \sqrt{\beta_1(t, \frac{1}{\delta^2})}$$

where $\beta_1(t, \frac{1}{\delta^2}) := (S_1 + \sigma\sqrt{2\log(\frac{1}{\delta^2}) + d\log\left(\frac{d+tL^2}{d}\right)})^2$.

Lemma G.2 (Theorem 2 in Abbasi-Yadkori et al. [2011]). *For all $t > 0$, with probability $1-\delta$,*

$$\|\hat{\theta}_t^{\mathrm{c}} - \theta^{\mathrm{c}}\|_{V_{t-1}} \leq \sqrt{\beta_2(t, \frac{1}{\delta^2})}$$

where $\beta_2(t, \frac{1}{\delta^2}) := (S_2 + \gamma\sqrt{2\log(\frac{1}{\delta^2}) + d\log\left(\frac{d+tL^2}{d}\right)})^2$.

Lemma G.3 (Laplace Approximation). *Denote $\overline{V}_T := I + \frac{1}{T}\sum_{t=1}^T X_t X_t^\top$, for any bounded open sets $\Theta_1, \Theta_2 \subseteq \mathbb{R}^d$, we have*

$$\int_{\Theta_1}\int_{\Theta_2} \exp\left(-\frac{T}{2}(\frac{\|\theta^{\mathrm{r}} - \theta_1\|_{\overline{V}_T}^2}{\sigma^2} + \frac{\|\theta^{\mathrm{c}} - \theta_2\|_{\overline{V}_T}^2}{\gamma^2})\right) \mathrm{d}\theta_1\, \mathrm{d}\theta_2 \doteq \exp\left(-\frac{T}{2}\inf_{\theta_1\in\Theta_1, \theta_2\in\Theta_2} \frac{\|\theta^{\mathrm{r}} - \theta_1\|_{\overline{V}_T}^2}{\sigma^2} + \frac{\|\theta^{\mathrm{c}} - \theta_2\|_{\overline{V}_T}^2}{\gamma^2}\right).$$

Proof. First, since Θ_1 and Θ_2 are bounded,

$$\text{LHS} \leq \mathrm{Vol}(\Theta_1)\cdot\mathrm{Vol}(\Theta_2)\exp\left(-\frac{T}{2}\inf_{\theta_1\in\Theta_1, \theta_2\in\Theta_2} \frac{\|\theta^{\mathrm{r}} - \theta_1\|_{\overline{V}_T}^2}{\sigma^2} + \frac{\|\theta^{\mathrm{c}} - \theta_2\|_{\overline{V}_T}^2}{\gamma^2}\right).$$

Second, define $(\theta_1^*, \theta_2^*) := \arg\min_{(\theta_1,\theta_2)\in\Theta_1\times\Theta_2} \frac{\|\theta^{\mathrm{r}} - \theta_1\|_{\overline{V}_T}^2}{\sigma^2} + \frac{\|\theta^{\mathrm{c}} - \theta_2\|_{\overline{V}_T}^2}{\gamma^2}$

For any $\theta_1 \in \Theta_1, \theta_2 \in \Theta_2$, since

$$\begin{aligned}
&\left|\left(\frac{\|\theta^{\mathrm{r}} - \theta_1\|_{\overline{V}_T}^2}{\sigma^2} + \frac{\|\theta^{\mathrm{c}} - \theta_2\|_{\overline{V}_T}^2}{\gamma^2}\right) - \left(\frac{\|\theta^{\mathrm{r}} - \theta_1^*\|_{\overline{V}_T}^2}{\sigma^2} + \frac{\|\theta^{\mathrm{c}} - \theta_2^*\|_{\overline{V}_T}^2}{\gamma^2}\right)\right| \\
&\leq \frac{1}{\sigma^2}\left|\,\|\theta^{\mathrm{r}} - \theta_1\|_{\overline{V}_T}^2 - \|\theta^{\mathrm{r}} - \theta_1^*\|_{\overline{V}_T}^2\,\right| + \frac{1}{\gamma^2}\left|\,\|\theta^{\mathrm{c}} - \theta_2\|_{\overline{V}_T}^2 - \|\theta^{\mathrm{c}} - \theta_2^*\|_{\overline{V}_T}^2\,\right| \\
&\leq \frac{4L^2R_1}{\sigma^2}\|\theta_1 - \theta_1^*\|_2 + \frac{4L^2R_2}{\gamma^2}\|\theta_2 - \theta_2^*\|_2
\end{aligned}$$

then when $\|\theta_1 - \theta_1^*\|_2 \leq \epsilon_1, \|\theta_2 - \theta_2^*\|_2 \leq \epsilon_2$,

$$\begin{aligned}
&\left|\left(\frac{\|\theta^{\mathrm{r}} - \theta_1\|_{\overline{V}_T}^2}{\sigma^2} + \frac{\|\theta^{\mathrm{c}} - \theta_2\|_{\overline{V}_T}^2}{\gamma^2}\right) - \left(\frac{\|\theta^{\mathrm{r}} - \theta_1^*\|_{\overline{V}_T}^2}{\sigma^2} + \frac{\|\theta^{\mathrm{c}} - \theta_2^*\|_{\overline{V}_T}^2}{\gamma^2}\right)\right| \\
&\leq \frac{4L^2R_1}{\sigma^2}\epsilon_1 + \frac{4L^2R_2}{\gamma^2}\epsilon_2.
\end{aligned}$$

Define $\Theta_1^* := \{\theta_1 \mid \|\theta_1 - \theta_1^*\| \leq \epsilon_1\}, \Theta_2^* := \{\theta_2 \mid \|\theta_2 - \theta_2^*\| \leq \epsilon_2\}$, then

$$\text{LHS} \geq \int_{\Theta_1^*}\int_{\Theta_2^*} \exp\left(-\frac{T}{2}(\frac{\|\theta^{\mathrm{r}} - \theta_1\|_{\overline{V}_T}^2}{\sigma^2} + \frac{\|\theta^{\mathrm{c}} - \theta_2\|_{\overline{V}_T}^2}{\gamma^2})\right) \mathrm{d}\theta_1\, \mathrm{d}\theta_2$$

$$\geq \mathrm{Vol}(\Theta_1^*) \cdot \mathrm{Vol}(\Theta_2^*) \exp\left(-\frac{T}{2}\left(\frac{4L^2R_1}{\sigma^2}\epsilon_1 + \frac{4L^2R_2}{\gamma^2}\epsilon_2\right)\right) \exp\left(-\frac{T}{2} \inf_{\theta_1\in\Theta_1,\theta_2\in\Theta_2} \frac{\|\theta^{\mathrm{r}} - \theta_1\|^2_{\overline{V}_T}}{\sigma^2} + \frac{\|\theta^{\mathrm{c}} - \theta_2\|^2_{\overline{V}_T}}{\gamma^2}\right)$$

Then when we choose $\epsilon_1 \to 0, \epsilon_2 \to 0$, we can have

$$\int_{\Theta_1}\int_{\Theta_2} \exp\left(-\frac{T}{2}\left(\frac{\|\theta^{\mathrm{r}} - \theta_1\|^2_{\overline{V}_T}}{\sigma^2} + \frac{\|\theta^{\mathrm{c}} - \theta_2\|^2_{\overline{V}_T}}{\gamma^2}\right)\right) \mathrm{d}\theta_1\, \mathrm{d}\theta_2 \doteq \exp\left(-\frac{T}{2} \inf_{\theta_1\in\Theta_1,\theta_2\in\Theta_2} \frac{\|\theta^{\mathrm{r}} - \theta_1\|^2_{\overline{V}_T}}{\sigma^2} + \frac{\|\theta^{\mathrm{c}} - \theta_2\|^2_{\overline{V}_T}}{\gamma^2}\right).$$

□

H MORE EMPIRICAL PLOTS

H.1 MORE PLOTS FOR "END OF OPTIMISM" INSTANCE

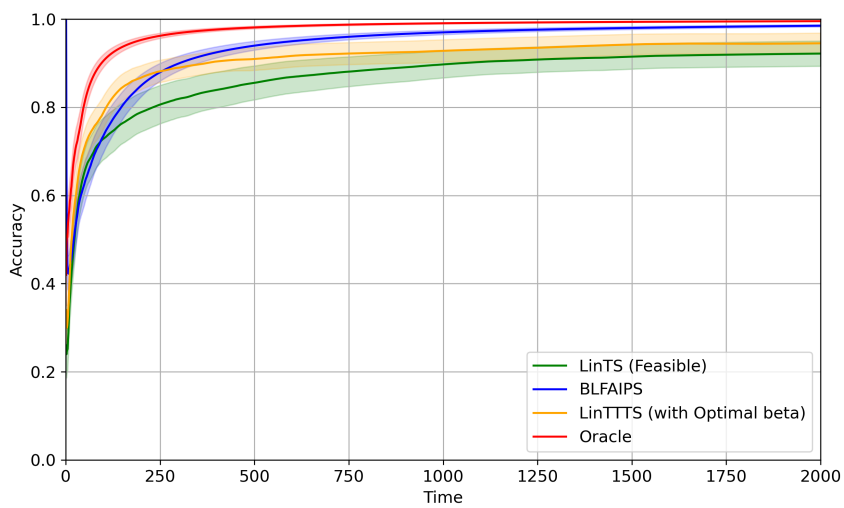

Figure 7: $\alpha = 0.2$

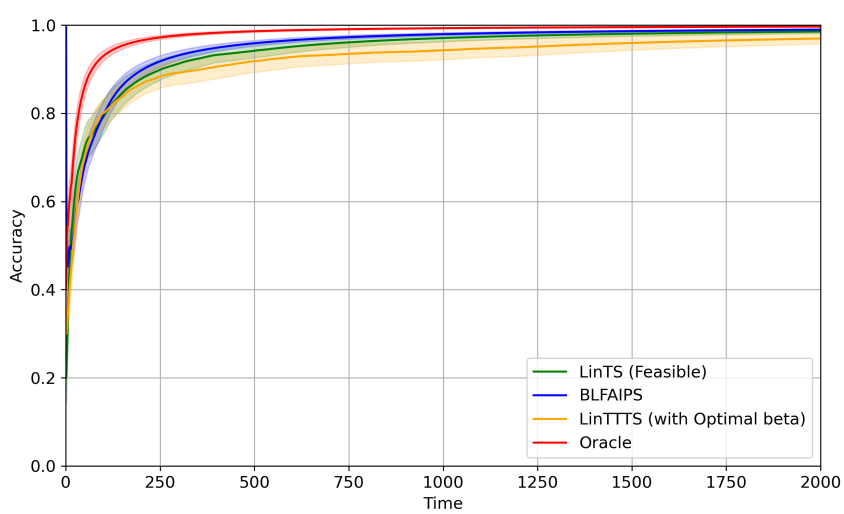

Figure 8: $\alpha = 0.3$

H.2 MORE PLOTS FOR RANDOM INSTANCE

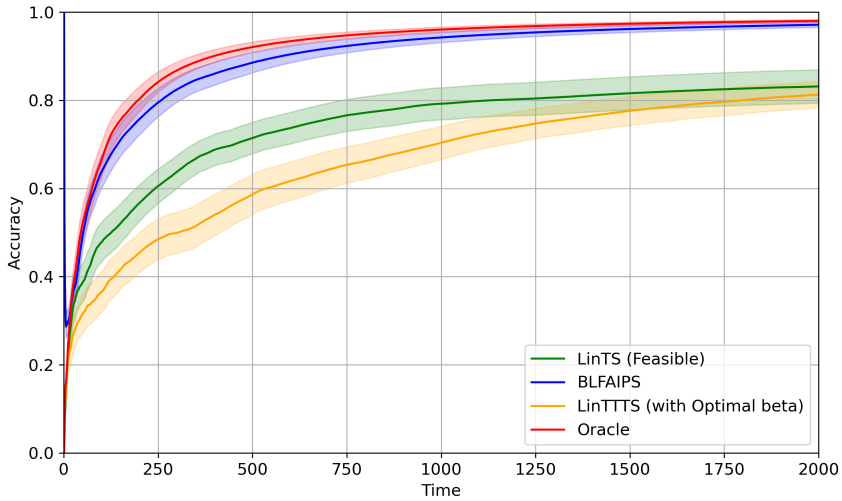

Figure 9: $d = 2, K = 20$

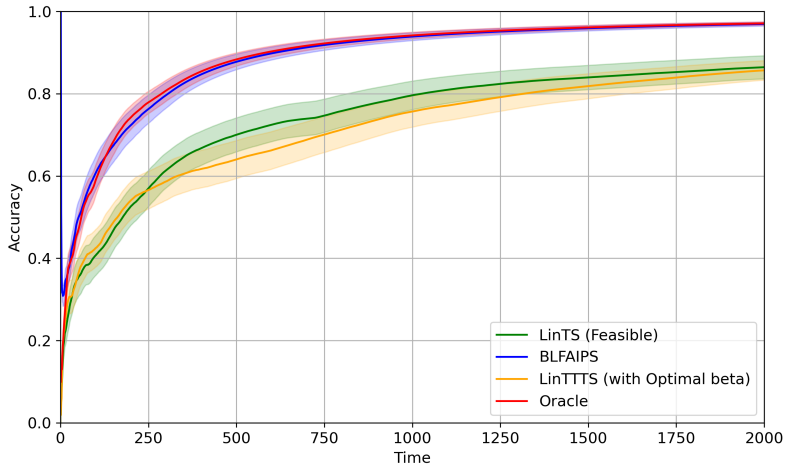

Figure 10: $d = 20, K = 20$