Pareto Set Identification With Posterior Sampling

Cyrille Kone
Univ. Lille, CNRS, Inria, Centrale Lille,
UMR 9189-CRIStAL, F-59000 Lille, France

Marc Jourdan
EPFL,
Lausanne, Switzerland

Emilie Kaufmann
Univ. Lille, CNRS, Inria, Centrale Lille,
UMR 9189-CRIStAL, F-59000 Lille, France

Abstract

The problem of identifying the best answer among a collection of items having real-valued distribution is well-understood. Despite its practical relevance for many applications, fewer works have studied its extension when multiple and potentially conflicting metrics are available to assess an item's quality. Pareto set identification (PSI) aims to identify the set of answers whose means are not uniformly worse than another. This paper studies PSI in the transductive linear setting with potentially correlated objectives. Building on posterior sampling in both the stopping and the sampling rules, we propose the PSIPS algorithm that deals simultaneously with structure and correlation without paying the computational cost of existing oracle-based algorithms. Both from a frequentist and Bayesian perspective, PSIPS is asymptotically optimal. We demonstrate its good empirical performance in real-world and synthetic instances.

1 INTRODUCTION

When facing a pure exploration problem, the decision maker aims to answer a question about a set of unknown distributions (e.g. modeling the treatment's effects) from which she gathers observations (e.g., measure of its outcomes) while providing theoretical guarantees on the candidate's answer. Practitioners might have multiple, potentially conflicting, metrics to optimize simultaneously (Zuluaga et al., 2013). When there is no clear trade-off between those metrics, aggregation into a unique objective is unrealistic despite being appealing. For instance, in clinical trials, both the effectiveness and safety of a treatment should be high (Jennison and Turnbull, 1993), but specifying the safety cost of an increased efficacy would be unethical. Likewise, in chip design, the runtime of the hardware and its energy consumption should be low (Almer et al., 2011). Moreover, there may exist a statistical correlation among the objectives (e.g., toxicity-efficacy in clinical trials) or observable features that characterize each item (e.g., chip area and design parameters in chip design (Zuluaga et al., 2013)).

To further our investigations, we adopt the well-studied stochastic bandit model (Audibert et al., 2010; Lattimore and Szepesvári, 2020; Drugan and Nowe, 2013) in the multi-dimensional setting with dimension $d \in \mathbb{N}$. In this decision-making game, the learner interacts sequentially with an environment of $K \in \mathbb{N}$ *arms*. Each arm $i \in [K] := \{1, \dots, K\}$ is associated with an unknown probability distribution over $\mathbb{R}^d$ denoted by $\nu_i \in \mathcal{D}$, where $\mathcal{D}$ is the set of possible distributions $\mathcal{D} \subseteq \mathcal{P}(\mathbb{R}^d)$, and having an unknown finite mean $\mu_i := (\mu_i(c))_{c\in[d]} \in \mathbb{R}^d$, where $\mu_i(c) := \mathbb{E}_{X\sim\nu_i}[X](c)$ denotes the mean of the objective c for the arm i. A *bandit instance* is uniquely characterized by its vector of distributions $\boldsymbol{\nu} := (\nu_i)_{i\in[K]} \in \mathcal{D}^K$ admitting $\boldsymbol{\mu} := (\mu_i)_{i\in[K]} \in \mathcal{M}$ as its matrix of means, where $\mathcal{M} \subseteq \mathbb{R}^{K\times d}$ is the set of possible vectors of means.

We consider the set $\mathcal{D} = \{\mathcal{N}(\lambda, \Sigma) \mid \lambda \in \mathbb{R}^d\}$ of Gaussian distributions with known covariance matrix Σ. Σ models the correlation between objectives (e.g., toxicity-efficacy in dose-finding trials). To model the dependency between the arms (e.g. treatments having similar active ingredients), we consider the *linear* setting (Soare et al., 2014; Degenne et al., 2020) in dimension $h \in \mathbb{N}$, where $\boldsymbol{\mu}$ is fully characterized by the set of arms vectors $\mathcal{A} := \{a_i\}_{i\in[K]} \subseteq \mathbb{R}^h$ and the regression matrix $\boldsymbol{\theta} \in \Theta$, where $\Theta \subseteq \mathbb{R}^{h\times d}$ is the set of possible regression matrices, namely $\mathcal{M} := \{\boldsymbol{\mu} \in \mathbb{R}^{K\times d} \mid \exists \boldsymbol{\theta} \in \Theta,\ \boldsymbol{\mu} = \boldsymbol{A}\boldsymbol{\theta}\}$ where $\boldsymbol{A} = (a_i)_{i\in[K]} \in \mathbb{R}^{K\times h}$. We will use $i \in [K]$ and $a \in \mathcal{A}$ to denote arms and $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$ to denote means. The above setting encompasses two well-studied models: *unstructured multi-dimensional* ($h = K$ and $\boldsymbol{A} = I_K$) and *linear one-dimensional* ($d = 1$).

Pareto Set Identification We focus on Pareto set identification (PSI) for *transductive* linear bandits, in which there exists a finite set of *answers* $\mathcal{Z} \subseteq \mathbb{R}^h$ which can be dif-

Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS) 2025, Mai Khao, Thailand. PMLR: Volume 258.

ferent from the set of arms $\mathcal{A}$ (Fiez et al., 2019; Li et al., 2024). For example, the treatments of interests $\mathcal{Z}$ might be too numerous (e.g. $\mathcal{A} \subseteq \mathcal{Z}$) or too costly (e.g. $\mathcal{A} \cap \mathcal{Z} = \emptyset$) compared to the treatments $\mathcal{A}$ that could be administered during clinical trials. Given an answer $z \in \mathcal{Z}$, its vector of means is denoted by $\mu_z := \boldsymbol{\theta}^\intercal z \in \mathbb{R}^d$. In that task, the goal of the learner is to identify the set of answers whose means are not uniformly worse coordinate-wise than another, known as the *Pareto set* and denoted as $S^\star(\boldsymbol{\theta}) \subseteq \mathcal{Z}$ (or $S^\star$ when $\boldsymbol{\theta}$ is clear from the context). The Pareto set contains the answers that satisfy optimal trade-offs in the objectives. The linear setting is the case $\mathcal{Z} = \mathcal{A}$.

An answer $z \in \mathcal{Z}$ is said to be *Pareto dominated* by an answer $x \in \mathcal{Z}$, which we denote by $\mu_z \prec \mu_x$ or $z \prec x$, if

$$\forall c \in [d],\ \mu_z(c) \leq \mu_x(c) \ \text{ and } \ \exists c \in [d],\ \mu_z(c) < \mu_x(c)\ .$$

The Pareto set of a matrix of means $\boldsymbol{\lambda} \in \mathbb{R}^{h \times d}$ is the set of answers that are not Pareto dominated, i.e.

$$S^\star(\boldsymbol{\lambda}) := \{z \in \mathcal{Z} \mid \nexists x \in \mathcal{Z} \backslash \{z\},\ \lambda_z \prec \lambda_x\}\ .$$

In the one-dimensional case ($d = 1$), we recover best answer identification (BAI) since the Pareto set reduces to the set of best answers, i.e. $S^\star(\boldsymbol{\lambda}) = \operatorname{argmax}_{z \in \mathcal{Z}} \lambda_z$.

Identification Strategy At each time t, the learner chooses an arm $a_t \in \mathcal{A}$ based on the observations previously collected and obtains a sample $X_t \in \mathbb{R}^d$, random variable with conditional distribution ν_{a_t} given a_t. It then proceeds to the next stage. An *algorithm* for the learner in this interaction should specify a *sampling rule* that determines a_t based on previously observed samples and some exogenous randomness. Formally, a_t is $\mathcal{F}_t$-measurable with the σ-algebra $\mathcal{F}_t := \sigma(U_1, a_1, X_1, \cdots, a_{t-1}, X_{t-1}, U_t)$, called *history* before time t, where $U_t \sim \mathcal{U}([0,1])$ materializes the possible independent randomness used by the algorithm at time t. The empirical allocation over arms is $\boldsymbol{N}_t := (N_{t,a})_{a \in \mathcal{A}}$ where $N_{t,a} := \sum_{s \in [t-1]} \mathbb{1}(a_s = a)$ and $\frac{\boldsymbol{N}_t}{t-1} \in \triangle_K := \{w \in \mathbb{R}_+^K \mid \sum_{a \in \mathcal{A}} w_a = 1\}$ where $\triangle_K$ denotes the probability simplex of dimension $K-1$.

As the learner aims at identifying $S^\star(\boldsymbol{\theta})$, its algorithm should include a *recommendation rule*. At time t, the agent recommends a *candidate answer* $\widehat{S}_t \subseteq \mathcal{Z}$ for the Pareto set before pulling arm a_t, therefore, $\widehat{S}_t$ is $\mathcal{F}_t$-measurable.

Fixed-confidence PSI There are several ways to evaluate the performance of a PSI algorithm. The two major theoretical frameworks are the *fixed-confidence* setting (Even-Dar et al., 2006; Jamieson and Nowak, 2014; Garivier and Kaufmann, 2016), which will be the focus of this paper, and the *fixed-budget* setting (Audibert et al., 2010; Gabillon et al., 2012). In the fixed-confidence setting (Kone et al., 2023; Crepon et al., 2024), the agent aims at minimizing the number of samples used to identify $S^\star(\boldsymbol{\theta})$ with confidence $1 - \delta \in (0,1)$. In the fixed-budget setting (Kone et al., 2024), the objective is to minimize the probability of misidentifying $S^\star(\boldsymbol{\theta})$ with a fixed number of samples.

In the fixed-confidence setting, the learner should define a *stopping rule* that is a stopping time for the filtration $(\mathcal{F}_t)_t$. The *sample complexity* of an algorithm corresponds to its stopping time τ_δ, which counts the number of rounds before termination. An algorithm is said to be δ*-correct* on the problem class $\mathcal{D}^K$ with a set of regression matrices Θ if its probability of stopping and not recommending a correct answer is upper bounded by δ, i.e. $\mathbb{P}_{\boldsymbol{\nu}}(\tau_\delta < +\infty, \widehat{S}_{\tau_\delta} \neq S^\star(\boldsymbol{\theta})) \leq \delta$ for all instances $\boldsymbol{\nu} \in \mathcal{D}^K$ having regression matrix $\boldsymbol{\theta} \in \Theta$. A fixed-confidence PSI algorithm is judged based on its expected sample complexity $\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta]$, i.e., the expected number of samples collected before it stops and returns a correct answer with confidence $1-\delta$. The learner should design a δ-correct algorithm minimizing $\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta]$.

Notation For any $\boldsymbol{w} \in \mathbb{R}_+^K$, let $V_{\boldsymbol{w}} := \sum_{a \in \mathcal{A}} w_a a a^\intercal$ be the *design matrix*, which is symmetric and positive semi-definite, and definite if and only if $\mathrm{Span}(\{a \in \mathcal{A} \mid w_a \neq 0\}) = \mathbb{R}^h$. For any symmetric positive semi-definite matrix V, we define the semi-norm $\|x\|_V := \sqrt{x^\intercal V x}$. It is a norm if V is positive definite. $|||A|||$ is the operator norm of A. Let $A \otimes B$ denote the Kronecker product of A, B and let $vec(A)$ denote the concatenation of the columns of A.

Lower Bound A δ-correct algorithm should distinguish $\boldsymbol{\theta}$ from all the instances $\boldsymbol{\lambda} \in \Theta$ having a different Pareto set. To better disentangle the structure induced by PSI from the underlying structure Θ, we define the set of alternative instances as $\mathrm{Alt}(S^\star) := \{\boldsymbol{\lambda} \in \mathbb{R}^{h \times d} \mid S^\star(\boldsymbol{\lambda}) \neq S^\star\}$ (i.e. without Θ). The requirement to distinguish $\boldsymbol{\theta}$ from $\Theta \cap \mathrm{Alt}(S^\star)$ leads to a lower bound on the expected sample complexity on any instance (Garivier and Kaufmann, 2016; Garivier et al., 2019; Crepon et al., 2024).

Lemma 1. *An algorithm which is δ-correct on all problems in $\mathcal{D}^K$ satisfies that, for all $\boldsymbol{\nu} \in \mathcal{D}^K$ with regression matrix $\boldsymbol{\theta} \in \Theta$, $\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta] \geq T^\star(\boldsymbol{\theta}) \log(1/(2.4\delta))$ where $T^\star(\boldsymbol{\theta})$ is a* characteristic time *whose inverse is defined as*

$$2T^\star(\boldsymbol{\theta})^{-1} := \sup_{\boldsymbol{w} \in \triangle_K} \inf_{\boldsymbol{\lambda} \in \Theta \cap \mathrm{Alt}(S^\star(\boldsymbol{\theta}))} \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\Sigma^{-1} \otimes V_{\boldsymbol{w}}},$$

and its maximizer set of optimal allocations *is $\boldsymbol{w}^\star(\boldsymbol{\theta})$.*

We say that an algorithm is *asymptotically optimal* if its sample complexity matches that lower bound, namely if $\limsup_{\delta \to 0} \mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta] / \log(1/\delta) \leq T^\star(\boldsymbol{\theta})$.

1.1 Algorithms

We use the ξ-regularized least-square estimator for $\boldsymbol{\theta}$, i.e. $\widehat{\boldsymbol{\theta}}_t := (V_{\boldsymbol{N}_t} + \xi I_h)^{-1} \sum_{s \in [t-1]} a_s X_s^\intercal$ with $\xi \geq 0$. Given the known covariance Σ for the objectives, we use the ξ-regularized empirical covariance $\boldsymbol{\Sigma}_t := \Sigma \otimes (V_{\boldsymbol{N}_t} + \xi I_h)^{-1}$. We recommend the empirical Pareto set $\widehat{S}_t := S^\star(\widehat{\boldsymbol{\theta}}_t)$.

Stopping Rules Before designing a sampling rule matching the asymptotic lower bound, one should specify a good stopping rule. Elimination-based (Even-Dar et al., 2006; Karnin et al., 2013) and gap-based (Kalyanakrishnan et al., 2012) stopping rules are sub-optimal in their δ-dependency. Hence, they fail to reach asymptotic optimality even for known $\boldsymbol{w}^\star(\boldsymbol{\theta})$. The well-studied Chernoff stopping times (Garivier and Kaufmann, 2016) use the generalized loglikelihood ratio (GLR) statistic. The GLR stopping rule is defined as $\tau_\delta^{\text{GLR}} := \inf\{t \mid \text{GLR}(t) > \beta(t-1,\delta)\}$ where β is a threshold. The ξ-regularized GLR statistic is

$$2\text{GLR}(t) := \inf_{\boldsymbol{\lambda} \in \Theta \cap \text{Alt}(\widehat{S}_t)} \|vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\lambda})\|^2_{\boldsymbol{\Sigma}_t^{-1}} . \quad (1)$$

Provided β is well chosen, the GLR stopping rule ensures δ-correctness regardless of the sampling rule. Moreover, combining a well-chosen sampling rule with the GLR stopping rule yields an asymptotically optimal algorithm.

Due to its δ-correctness and simplicity, the GLR stopping rule has become the standard choice. However, in some specific pure exploration problems, the computational cost of the GLR statistic becomes a major bottleneck. PSI for transductive linear bandits is an example of a setting where evaluating $\text{GLR}(t)$ is numerically challenging. In transductive linear BAI ($d = 1$), the computational cost of (1) heavily depends on Θ, and there are no procedure for general Θ. While there is a closed form solution in $\mathcal{O}(h^2|\mathcal{Z}|)$ when $\Theta = \mathbb{R}^h$, a more involved one-dimensional optimization procedure should be solved $|\mathcal{Z}|$ times when Θ is a ball (Degenne et al., 2020). In the unstructured multi-dimensional setting, to the best of our knowledge, computing (1) is only tractable for the setting with independent objectives, i.e. Σ diagonal, $\mathcal{Z} = \mathcal{A}$, $h = K$, $\boldsymbol{A} = I_K$ and $\Theta = \mathbb{R}^{K\times d}$. Crepon et al. (2024) proposed the best-known procedure in that restrictive setting, which requires to solve $\mathcal{O}(Kd^3|\widehat{S}_t|^d)$ convex problems at each round. Their computationally costly procedure is difficult to extend to correlated objectives or structured settings.

We introduce the posterior sampling (PS) stopping rule. Based on at most $M(t-1,\delta)$ independent realizations from a posterior distribution inflated by $c(t-1,\delta)$, it stops when

$$\tau_\delta^{\text{PS}} := \inf\{t \mid \forall m \in [M(t-1,\delta)],\ \boldsymbol{\theta}_t^m \notin \Theta \cap \text{Alt}(\widehat{S}_t)\} , \quad (2)$$

with $vec(\boldsymbol{\theta}_t^m - \widehat{\boldsymbol{\theta}}_t) \mid \mathcal{F}_t \sim \mathcal{N}(0_{hd}, c(t-1,\delta)\boldsymbol{\Sigma}_t)$ for all $m \in [M(t-1,\delta)]$. For precise choice of $M(t,\delta)$ and $c(t,\delta)$ to ensure δ-correctness, we refer the reader to Section 2.1. The PS stopping rule allows tackling structured settings with correlated objectives and has a lower computational cost than the GLR stopping rule.

Sampling Rules Given a stopping rule, the sampling rule should make it stop as soon as possible. While gap-based sampling rules will be sub-optimal in their asymptotic allocation, the asymptotic optimality of Track-and-Stop (Garivier and Kaufmann, 2016) will come at the cost of a computational intractability since it computes $\boldsymbol{w}_t \in \boldsymbol{w}^\star(\widehat{\boldsymbol{\theta}}_t)$ at time t. The game-based approach came from seeing the lower bound as the solution of a two-player game (Degenne et al., 2019). Using two no-regret learning algorithms against each other yields a saddle-point algorithm that sequentially approximates $T^\star(\boldsymbol{\theta})^{-1}$. The popularized instance uses the best response as min learner, i.e.

$$\boldsymbol{\lambda}^{\text{BR}} : \boldsymbol{w} \mapsto \operatorname*{argmin}_{\boldsymbol{\lambda} \in \Theta \cap \text{Alt}(\widehat{S}_t)} \|vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\lambda})\|^2_{\Sigma^{-1} \otimes (V_{\boldsymbol{w}} + \xi I_h)} , \quad (3)$$

whose computational cost is the same as the GLR statistic (1). As for the PS stopping rule, we rely on an inflated posterior sampling (PS) learner for the min learner. Therefore, our algorithm can tackle structured settings with correlated objectives and has a low computational cost.

1.2 Contributions

As main contribution, we propose the PSIPS algorithm, which is the first computationally efficient algorithm for transductive linear PSI. PSIPS builds on posterior sampling in both the PS stopping rule and the min learner of the game-based sampling rule. By removing the Oracle calls to (1) or (3), PSIPS deals with the transductive linear PSI structure in a computationally efficient way.

For the PS stopping rule and regardless of the sampling rule, we exhibit choices of $M(t,\delta)$ and $c(t,\delta)$ ensuring δ-correctness in the unstructured setting with independent or correlated objectives and in transductive linear BAI ($d = 1$ and $\Theta = \mathbb{R}^h$). When $\lim_{\delta \to 0} \frac{c(t,\delta) \log M(t,\delta)}{\log(1/\delta)} \leq 1$ (satisfied by our choices), PSIPS is shown to be asymptotically optimal for transductive linear PSI (Theorem 1). From a Bayesian perspective, the posterior probability that PSIPS misidentifies the Pareto set decays exponentially fast (as a function of t) with the optimal rate $T^\star(\boldsymbol{\theta})$ (Theorem 2). Our experiments on both real-world and synthetic instances showcase the superior performance of our algorithm both in terms of sample complexity and computational cost.

1.3 Related Work

Pioneering the PSI problem for bandits, Auer et al. (2016) proposed a δ-correct algorithm based on uniform sampling with an accept/reject mechanism. They proved a high-probability lower bound on the sample complexity of any δ-correct algorithm for PSI, essentially scaling as $H(\boldsymbol{\mu})\log(1/\delta)$ with $H(\boldsymbol{\mu}) := \sum_{i \in [K]} \Delta_i^{-2}$ where Δ_i's are PSI's "sub-optimality" gaps introduced therein. The sample complexity of the proposed algorithm is

$\mathcal{O}(H(\boldsymbol{\mu})\log(Kd\log(H(\boldsymbol{\mu}))/\delta))$. Kone et al. (2023) generalized the LUCB algorithm (Kalyanakrishnan et al., 2012) to the PSI setting, proving guarantees similar to those of Auer et al. (2016). Still in the unstructured setting and for $\Sigma = \sigma^2 I_d$, Crepon et al. (2024) proposed a δ-correct gradient-based algorithm. Their sampling rule requires computing a (super)-gradient of $\boldsymbol{N}_t \mapsto \mathrm{GLR}(t)$, whose per-round computational cost is $\mathcal{O}(Kd^3|\widehat{S}_t|^d)$. The GLR stopping rule is obtained by solving $\mathcal{O}(Kd^3|\widehat{S}_t|^d)$ convex problems per round. This gradient computation procedure does not apply to the non-isotropic case or the structured setting. For fixed-budget PSI, Kone et al. (2024) proposed an algorithm whose error probability is upper bounded by $\mathcal{O}(\exp(-T/(H(\boldsymbol{\mu})\log K)))$. The authors showed that up to the $\log K$ terms and some constants, this is tight on a set of non-trivial instances.

Gaussian processes were used to study PSI. Each arm has a feature whose mean is estimated by Gaussian process regression. Zuluaga et al. (2013, 2016) introduced a racing algorithm with correctness guarantees. The linear setting is also addressed by Kim et al. (2025); Kone et al. (2025), who introduced PSI algorithms leveraging G-optimal design approaches (Fiez et al., 2019). They proved gap-based guarantees comparable to those in Auer et al. (2016), which remain unaffected by off-diagonal correlation terms. Although efficient in practice, these approaches are known to be sub-optimal, in particular when δ is small.

Other multi-objective problems have been studied. Ararat and Tekin (2023) introduced the identification of the non-dominated set of any cone-induced partial order, of which PSI is a special case. The authors proposed an extension of the algorithm of Auer et al. (2016). In the fixed-budget setting, Katz-Samuels and Scott (2018) aims at identifying the arms belonging to a given polyhedron. For multi-objective regret minimization, we refer to Drugan and Nowe (2013); Xu and Klabjan (2023).

In the pure exploration literature, there is a surge of interest in algorithms based on posterior sampling. This literature aims at extending the success of Thompson Sampling (TS) for regret minimization (Thompson, 1933; Kaufmann et al., 2012; Agrawal and Goyal, 2013). Russo (2016) proposed the Top Two TS (TTTS) sampling rule for BAI and showed posterior contraction at the optimal rate. Having only Bayesian guarantees, TTTS is not paired with a stopping rule. Shang et al. (2020) considered a variant of the Top Two algorithm for BAI. They propose a Bayesian stopping rule that requires computing the posterior probability for an arm's optimality, which is computationally inefficient. In combinatorial pure exploration (a subcase of our setting), the TS-Explore algorithm (Wang and Zhu, 2022) is the first to leverage posterior samples in the sampling and the stopping rules. While their stopping rule is similar to the PS stopping rule defined in (2), the authors use a gap-based sampling rule that is asymptotically sub-optimal. In transductive linear BAI (subcase of our setting), the PEPS algorithm (Li et al., 2024) is a game-based algorithm using posterior sampling for the min learner. PEPS relies on phases to tune the learning rates that depend on an upper bound on the stochastic losses. PEPS achieves the optimal convergence rate of the posterior, yet it lacks a stopping rule and fixed-confidence or fixed-budget guarantees.

Algorithm 1: PSIPS

Input: regularization $\xi \geq 0$, budget M and inflation c functions, exploration allocation $\boldsymbol{w}_{\exp}$ and rate $\alpha > 0$, inflation rate function η, AdaHedge learner denoted by $\mathcal{L}^{\mathcal{A}}$;

Get $(\boldsymbol{Z}_1, V_1)$ and $\widehat{\boldsymbol{\theta}}_1 = V_1^{-1}\boldsymbol{Z}_1$ and $\boldsymbol{\Sigma}_1 = \Sigma \otimes V_1^{-1}$;

for $t \geq 1$ **do**
- Get $\widehat{S}_t = S^\star(\widehat{\boldsymbol{\theta}}_t)$;
- Set $m = 0$, $m_t = +\infty$ and $m_{t,\delta} = +\infty$;
- **while** $\max\{m_t, m_{t,\delta}\} = +\infty$ **do**
  - Set $m \hookleftarrow m+1$ and get $\boldsymbol{v}_t^m \sim \mathcal{N}(\boldsymbol{0}_{hd}, \boldsymbol{\Sigma}_t)$;
  - `// Stopping Rule`
  - **if** $m_{t,\delta} = +\infty$ **then**
    - $vec(\boldsymbol{\theta}_t^m) = vec(\widehat{\boldsymbol{\theta}}_t) + \sqrt{c(t-1,\delta)}\boldsymbol{v}_t^m$;
    - **if** $\boldsymbol{\theta}_t^m \in \Theta \cap \mathrm{Alt}(\widehat{S}_t)$ **then** $m_{t,\delta} \hookleftarrow m$;
    - **else if** $m \geq M(t-1,\delta)$ **then break** and **return** $\widehat{S}_t$;
  - **end**
  - `// Min Learner`
  - **if** $m_t = +\infty$ **then**
    - $vec(\boldsymbol{\lambda}_t^m) = vec(\widehat{\boldsymbol{\theta}}_t) + \eta_t^{-1/2}\boldsymbol{v}_t^m$;
    - **if** $\boldsymbol{\lambda}_t^m \in \Theta_t \cap \mathrm{Alt}(\widehat{S}_t)$ **then** $m_t \hookleftarrow m$;
  - **end**
- **end**
- Get $\boldsymbol{w}_t$ from learner $\mathcal{L}^{\mathcal{A}}$ and set $\boldsymbol{\lambda}_t = \boldsymbol{\lambda}_t^{m_t}$;
- Set $\widetilde{\boldsymbol{w}}_t = (1-\gamma_t)\boldsymbol{w}_t + \gamma_t \boldsymbol{w}_{\exp}$ with $\gamma_t = t^{-\alpha}$;
- Get arm $a_t \sim \widetilde{\boldsymbol{w}}_t$ and collect $X_t \sim \nu_{a_t}$;
- Feed gain $g_t(\boldsymbol{w}) = \|vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\lambda}_t)\|^2_{\Sigma^{-1}\otimes V_{\boldsymbol{w}}}$ to $\mathcal{L}^{\mathcal{A}}$;
- Update $\widehat{\boldsymbol{\theta}}_{t+1} = V_{t+1}^{-1}\boldsymbol{Z}_{t+1}$ and $\Sigma_{t+1} = \Sigma \otimes V_{t+1}^{-1}$ with $\boldsymbol{Z}_{t+1} = \boldsymbol{Z}_t + a_t X_t^\intercal$ and $V_{t+1} = V_t + a_t a_t^\intercal$;

end

2 PSIPS ALGORITHM

We propose the PSIPS (PSI with Posterior Sampling) algorithm for PSI in the transductive linear setting with correlated objectives. PSIPS combines the PS stopping rule and the game-based sampling rule using a PS min learner.

2.1 The Posterior Sampling (PS) Stopping Rule

We introduce the posterior sampling (PS) stopping rule. To specify the PS stopping rule, a budget function M :

$\mathbb{N} \times (0,1) \to \mathbb{R}_+$ and an inflation function $c : \mathbb{N} \times (0,1) \to \mathbb{R}_+$ are given. The PS stopping rule defined in (2) stops when $M(t-1,\delta)$ independent realizations from a posterior distribution inflated by $c(t-1,\delta)$ agrees with the evidence suggesting that $S^\star(\boldsymbol{\theta}) = \widehat{S}_t$. Conditioned on $\mathcal{F}_t$, let $(\boldsymbol{v}_t^m)_{m \in [M(t,\delta)]}$ be i.i.d. draws from the centered posterior distribution $\Pi_t := \mathcal{N}(0_{hd}, \boldsymbol{\Sigma}_t)$. For all $m \in [M(t-1,\delta)]$, let $vec(\boldsymbol{\theta}_t^m) := vec(\widehat{\boldsymbol{\theta}}_t) + \sqrt{c(t-1,\delta)}\boldsymbol{v}_t^m$. Then,

$$\tau_\delta^{\text{PS}} := \inf\{t \mid \forall m \in [M(t-1,\delta)],\ \boldsymbol{\theta}_t^m \notin \Theta \cap \operatorname{Alt}(\widehat{S}_t)\} .$$

When $\tau_\delta^{\text{PS}} < +\infty$, we recommend $\widehat{S}_{\tau_\delta^{\text{PS}}} := S^\star(\widehat{\boldsymbol{\theta}}_{\tau_\delta^{\text{PS}}})$. When $t < \tau_\delta^{\text{PS}}$, let $m_{t,\delta} := \inf\{m \mid \boldsymbol{\theta}_t^m \in \Theta \cap \operatorname{Alt}(\widehat{S}_t)\}$. Intuitively, $\boldsymbol{\theta}_t^{m_{t,\delta}}$ is a randomized approximation of the best-response Oracle $\boldsymbol{\lambda}^{\text{BR}}(\boldsymbol{N}_t)$ as in (3).

Computational Cost To reduce the computational cost of PSIPS, we draw $(\boldsymbol{v}_t^m)_m$ sequentially as the PS stopping condition is often infringed before $M(t,\delta)$ realizations are drawn (see Figure 6). We re-use the realizations $(\boldsymbol{v}_t^m)_m$ in the PS min learner of our sampling rule with a different inflation parameter (see Section 2.2).

Compared to the Oracle call to (1), testing that $\boldsymbol{\theta}_t^m \notin \Theta \cap \operatorname{Alt}(\widehat{S}_t)$ has a lower computational cost. It scales as the sum of the costs of membership to Θ and to $\operatorname{Alt}(\widehat{S}_t)^\mathsf{c}$, which is at most $\mathcal{O}(dh|\widehat{S}_t| \max\{|\widehat{S}_t|, |\mathcal{Z}| - |\widehat{S}_t|\})$ since

$$\begin{aligned}\operatorname{Alt}(\widehat{S}_t)^\mathsf{c} = &\bigcap_{(z,x) \in \widehat{S}_t^2, x \neq z} \bigcup_{c \in [d]} \{\boldsymbol{\lambda} \mid \langle E_{z,x}(c), vec(\boldsymbol{\lambda})\rangle \geq 0\} \\ &\cap \bigcap_{z \notin \widehat{S}_t} \bigcup_{x \in \widehat{S}_t} \bigcap_{c \in [d]} \{\boldsymbol{\lambda} \mid \langle E_{x,z}(c), vec(\boldsymbol{\lambda})\rangle \geq 0\} ,\end{aligned} \quad (4)$$

where $E_{z,x}(c) = e_c \otimes (z - x)$ and $e_c = (\mathbb{1}(c' = c))_{c' \in [d]}$.

To update the candidate answer, we check whether $\widehat{\boldsymbol{\theta}}_t \in \operatorname{Alt}(\widehat{S}_{t-1})^\mathsf{c}$, in which case $\widehat{S}_t = \widehat{S}_{t-1}$. Otherwise, we compute $\widehat{S}_t = S^\star(\widehat{\boldsymbol{\theta}}_t)$. While a naive implementation is at most in $\mathcal{O}(dh|\mathcal{Z}|^2)$, Kung et al. (1975) proposed an algorithm having a cost scaling as $\mathcal{O}(h|\mathcal{Z}|(\log|\mathcal{Z}|)^{\max\{1,d-2\}})$.

Correctness Lemma 2 exhibits choices of budget $M(t,\delta)$ and inflation $c(t,\delta)$ ensuring δ-correctness in the unstructured setting with independent or correlated objectives and in transductive linear BAI ($d = 1$ and $\Theta = \mathbb{R}^h$). Letting $X \sim \mathcal{N}(0,1)$, we denote by $R(x) := \frac{\mathbb{P}(X > x)}{f_X(x)}$, the Mills ratio (Mills, 1926) of X, with f_X, the density of X.

Lemma 2. *Let $\delta \in (0,1)$, $s > 1$, ζ be the Riemann ζ function. Define* $r(\delta, n) := \left(\frac{1}{\sqrt{2\pi}} R\left(\sqrt{\frac{2}{n}\log(1/\delta)}\right)\right)^n$. *Let $\beta(t,\delta)$ be an anytime upper bound on $\frac{1}{2}\|\theta - \theta_t\|^2_{\boldsymbol{\Sigma}_t^{-1}}$ with probability at least $1-\delta$, as in Lemma 3. Regardless of the sampling rule, the PS stopping rule with $c(t,\delta) = \beta(t, \frac{\delta}{2}) / \log\frac{1}{\delta}$ and $M(t,\delta) = \left\lceil \frac{\log(2t^s\zeta(s)/\delta)}{\delta q(t,\delta)} \right\rceil$ ensures that the algorithm is δ-correct for any $\delta \in (0,1)$,*

***1)** when $\Theta = \mathbb{R}^{K \times d}$, $\mathcal{Z} = \mathcal{A}$, $h = K$, $\boldsymbol{A} = I_K$, by taking $q(t,\delta) = \min\{r(\delta,d), r(\delta, d + |\widehat{S}_{t+1}|)\}$ for Σ diagonal, and otherwise $q(t,\delta) = \det(\Sigma\bar{\sigma})^{-1/2} \min\{r(\delta^{\frac{d_\Sigma}{d}}, d)$, $r(\delta^{\frac{d_\Sigma + |\widehat{S}_{t+1}|}{d + |\widehat{S}_{t+1}|}}, d + |\widehat{S}_{t+1}|)\}$ with $\bar{\sigma} = |||\Sigma^{-1}|||$ and $d_\Sigma = \|1_d\|^2_{(\bar{\sigma}\Sigma)^{-1}}$.*
***2)** when $d = 1$ and $\Theta = \mathbb{R}^h$, by taking $q(t,\delta) = r(\delta, 1)$.*
The above choices satisfy $\limsup_{\delta \to 0} \frac{c(t,\delta)\log M(t,\delta)}{\log(1/\delta)} \leq 1$.

Lemma 2 does not propose choices of (M, c) for transductive linear PSI with general Θ. In Section 3.1.1, our proof sketch highlights the challenge arising due to the structure.

2.2 Game-based Sampling Rule

We introduce a game-based sampling rule (Degenne and Koolen, 2019) using a PS min learner. By combining a max learner playing $\boldsymbol{w}_t$ and a min learner playing $\boldsymbol{\lambda}_t$, it yields a saddle-point algorithm which approximates $T^\star(\boldsymbol{\theta})^{-1}$.

Initialization In the unstructured setting, we pull each arm once and observe $X_i \sim \nu_i$ for all $i \in [K]$. We take $\xi = 0$ and set $(\boldsymbol{Z}_1, V_1) = (\boldsymbol{X}_1, I_K)$ where $\boldsymbol{X}_1 = (X_i^\intercal)_{i \in [K]}$. In the structured setting, there is no initial pull of arms. We use $\xi > 0$ and set $(\boldsymbol{Z}_1, V_1) = (\boldsymbol{0}_{h \times d}, \xi I_h)$.

Max Learner As in Degenne and Koolen (2019), we opt for AdaHedge (De Rooij et al., 2014) as the max learner. We add forced exploration by mixing the played $\boldsymbol{w}_t$ with an exploration allocation $\boldsymbol{w}_{\text{exp}} \in \triangle_K$. Formally, we pull $a_t \sim \widetilde{\boldsymbol{w}}_t$ with $\widetilde{\boldsymbol{w}}_t := (1-\gamma_t)\boldsymbol{w}_t + \gamma_t \boldsymbol{w}_{\text{exp}}$ where $\gamma_t = 1/t^\alpha$ with $\alpha \in (0,1)$. The allocation $\boldsymbol{w}_{\text{exp}}$ should be chosen such $\lambda_{\min}(V_{\boldsymbol{w}_{\text{exp}}}) > 0$. For example, we can use a uniform allocation on a set of arms spanning $\mathbb{R}^h$ or the G-optimal design (Li et al., 2024). Forced exploration ensures that $\widehat{S}_t$ converges to $S^\star(\boldsymbol{\theta})$ despite the initial fluctuation of the min learning space $\Theta \cap \operatorname{Alt}(\widehat{S}_t)$.

Min Learner As in Li et al. (2024), we propose a PS min learner. The PS min learner draws independent realization from a posterior distribution inflated by $\eta_t > 0$ until one disagrees with $\widehat{S}_t$. Let $(\boldsymbol{v}_t^m)_m$ be i.i.d. draws from Π_t. For all $m \geq 1$, let $vec(\boldsymbol{\lambda}_t^m) := vec(\widehat{\boldsymbol{\theta}}_t) + \eta_t^{-1/2}\boldsymbol{v}_t^m$. Then,

$$\boldsymbol{\lambda}_t := \boldsymbol{\lambda}_t^{m_t} \text{ with } m_t = \inf\{m \mid \boldsymbol{\lambda}_t^m \in \Theta_t \cap \operatorname{Alt}(\widehat{S}_t)\}, \quad (5)$$

where $\Theta_t := \Theta$ when Θ is bounded, and Θ_t is a bounded confidence region otherwise. Equivalently, we could draw $vec(\boldsymbol{\lambda}_t)$ from the distribution $\mathcal{N}(vec(\widehat{\boldsymbol{\theta}}_t), \eta_t^{-1/2}\boldsymbol{\Sigma}_t)$ truncated to $\Theta_t \cap \operatorname{Alt}(\widehat{S}_t)$. The inflation η_t is chosen as an upper bound on the magnitude of the stochastic loss. We refer the reader to Appendix D for more details on how η_t and Θ_t are defined sequentially.

Computational Cost At time $t < \tau_\delta^{\text{PS}}$, we re-use sequentially the realizations $(\boldsymbol{v}_t^m)_{m \in [m_{t,\delta}]}$ drawn by the PS stopping rule with a different inflation parameter. When $m_{t,\delta} < m_t$, we draw fresh realization sequentially. Despite the different inflations, m_t is expected to be close to $m_{t,\delta}$, i.e. $\mathcal{O}(M(t,\delta))$ (see Figure 6). The computational cost of (5) is lower than a best-response Oracle to (1).

3 THEORETICAL GUARANTEES

We show that under a generic condition on (c, M), which is satisfied in Lemma 2, the expected sample complexity of PSIPS is asymptotically optimal (Theorem 1). From a Bayesian perspective, the rate of decay for its posterior probability of misidentifying the Pareto set is shown to be asymptotically tight (Theorem 2). Theorems 1 and 2 hold for unstructured PSI ($\Theta = \mathbb{R}^{K\times d}$, $\mathcal{Z} = \mathcal{A}$, $h = K$, $\boldsymbol{A} = I_K$) and transductive linear PSI (bounded convex Θ).

Expected Sample Complexity Theorem 1 shows that PSIPS is asymptotically optimal when using suitable choices of budget and inflation. In comparison, known upper bounds on $\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta]$ were either sub-optimal (Auer et al., 2016), or restricted to the unstructured setting with independent objectives (Crepon et al., 2024) or transductive linear BAI ($d = 1$). The proof is sketched in Section 3.1.2.

Theorem 1. *Using budget M and inflation c such that $\limsup_{\delta\to 0} \frac{c(t,\delta)\log M(t,\delta)}{\log(1/\delta)} \le 1$, the PSIPS algorithm satisfies that $\limsup_{\delta\to 0} \mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta^{PS}]/\log(1/\delta) \le T^\star(\boldsymbol{\theta})$ for all $\boldsymbol{\nu} \in \mathcal{D}^K$ with regression matrix $\boldsymbol{\theta} \in \Theta$, both for unstructured PSI ($\Theta = \mathbb{R}^{K\times d}$, $\mathcal{Z} = \mathcal{A}$, $h = K$, $\boldsymbol{A} = I_K$) and transductive linear PSI (bounded convex Θ).*

Lemma 2 gives choices of (c, M) satisfying the above condition and ensuring δ-correctness in unstructured PSI. Theorem 1 also holds on the set $\widetilde{\mathcal{D}}$ of multi-variate Σ-sub-Gaussian, where $\kappa \in \widetilde{\mathcal{D}}$ with mean λ implies that $\mathbb{E}_{X\sim\kappa}[e^{u^\mathsf{T}(X-\lambda)}] \le e^{\frac{1}{2}u^\mathsf{T}\Sigma^{-1}u}$ for all $u \in \mathbb{R}^d$ (Degenne and Perchet, 2016). This includes distributions with Bernoulli marginals or with bounded support. Optimality is only achieved for multi-variate Gaussian distributions.

Posterior Probability of Misidentification Theorem 2 shows that the posterior probability that PSIPS misidentifies the Pareto set decays exponentially fast (as a function of t) with a rate $T^\star(\boldsymbol{\theta})$, which is shown to be asymptotically optimal by a lower bound. In comparison, known similar Bayesian guarantees were restricted to BAI ($d = 1$) in the unstructured and structured settings (e.g. TTTS in Russo (2016) and PEPS in Li et al. (2024)).

Theorem 2. *Let $\widetilde{\Pi}_t := \mathcal{N}(\widehat{\boldsymbol{\theta}}_t, \Sigma \otimes V_{\boldsymbol{N}_t}^{-1})$ be the posterior distribution under a flat Gaussian prior (without inflation).For all $\boldsymbol{\nu} \in \mathcal{D}^K$ with regression matrix $\boldsymbol{\theta} \in \Theta$, it almost surely holds that $\limsup_{t\to+\infty} -t^{-1}\log \mathbb{P}_{\widetilde{\Pi}_t|\mathcal{F}_t}(\Theta \cap \mathrm{Alt}(S^\star)) \le T^\star(\boldsymbol{\theta})^{-1}$ for any algorithm, and PSIPS almost surely satisfies that $\liminf_{t\to+\infty} -t^{-1}\log \mathbb{P}_{\widetilde{\Pi}_t|\mathcal{F}_t}(\Theta \cap \mathrm{Alt}(S^\star)) \ge T^\star(\boldsymbol{\theta})^{-1}$, both for unstructured PSI ($\Theta = \mathbb{R}^{K\times d}$, $\mathcal{Z} = \mathcal{A}$, $h = K$, $\boldsymbol{A} = I_K$) and transductive linear PSI (bounded convex Θ).*

3.1 Proof Sketches

Before sketching the proofs of Lemma 2 and Theorem 1, we introduce the concentration event $\mathcal{E}_\delta := \bigcap_{t\in\mathbb{N}} \mathcal{E}_{t,\delta}$ with

$$\mathcal{E}_{t,\delta} := \{\|vec(\theta - \theta_t)\|^2_{\boldsymbol{\Sigma}_t^{-1}} \le 2\beta(t-1,\delta)\} \, . \qquad (6)$$

Lemma 3 in Appendix C gives choices of $\beta(t,\delta)$ such that $\mathbb{P}_{\boldsymbol{\nu}}(\mathcal{E}_\delta^\complement) \le \delta$. They satisfy that $\lim_{\delta\to 0} \beta(\cdot,\delta)/\log(1/\delta) = 1$ and $\beta(t,\cdot) =_{+\infty} \mathcal{O}(\log\log t)$. Using $\beta(t,\delta)$ in the GLR stopping rule yields δ-correctness for any sampling rule.

3.1.1 Proof Sketch of Lemma 2

Under $\mathcal{E}_{\delta/2}$, the budget M and the inflation c are chosen to ensure δ-correctness of the PS stopping rule, regardless of the sampling rule. Formally, we need to prove that

$$\mathbb{P}_{\boldsymbol{\nu}}(\mathcal{E}_{\delta/2} \cap \{\tau_\delta^{\mathrm{PS}} < +\infty, \widehat{S}_{\tau_\delta^{\mathrm{PS}}} \neq S^\star\}) \le \delta/2 \, .$$

Let $\widehat{\Pi}_t := \mathcal{N}(\widehat{\boldsymbol{\theta}}_t, c(t-1,\delta)\boldsymbol{\Sigma}_t)$. By union bound, the stopping in (2) and $1 - x \le \exp(-x)$, it suffices to upper bound

$$\mathbb{1}_{\mathcal{E}_{t,\delta/2}\cap\{\widehat{S}_t\neq S^\star\}} \exp\left(-M(t-1,\delta)\mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\Theta \cap \mathrm{Alt}(\widehat{S}_t))\right)$$

by $\frac{\delta}{2\zeta(s)(t-1)^s}$ since their sum is smaller than $\delta/2$. Therefore, we should lower bound $\mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\Theta \cap \mathrm{Alt}(\widehat{S}_t))$ under the event $\mathcal{E}_{t,\delta/2} \cap \{\widehat{S}_t \neq S^\star\}$ to conclude the proof.

Unstructured Setting Having $\widehat{S}_t \neq S^\star$ implies that $\boldsymbol{\theta} \in \mathrm{Alt}(\widehat{S}_t)$, see (4). Either (1) there exists $(i,j) \in \widehat{S}_t$ with $i \neq j$ such that $\mu_i(c) < \mu_j(c)$ for all $c \in [d]$, and we define $\mathcal{J}_t = \{j\}$. Or (2) there exists $i \notin \widehat{S}_t$ and $c \in [d]^{|\widehat{S}_t|}$ such that $\mu_j(c_j) < \mu_i(c_j)$ for all $j \in \widehat{S}_t$, and $\mathcal{J}_t = \widehat{S}_t$. Since $\Theta = \mathbb{R}^{K\times d}$, we can show that $\mathbb{1}_{\widehat{S}_t\neq S^\star}\mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\mathrm{Alt}(\widehat{S}_t)) \ge$

$$\prod_{k\in\{i\}\cup\mathcal{J}_t} \mathbb{P}_{X\sim\mathcal{N}(0_d,\Sigma)}\left(X > \sqrt{\frac{N_{t,k}}{c(t-1,\delta)}}(\mu_k - \hat{\mu}_{t,k})\right) ,$$

where $X > x$ denotes $X(c) > x(c)$ for all $c \in [d]$.

When Σ is diagonal, we have $\mathbb{P}_{X\sim\mathcal{N}(0_d,\Sigma)}(X > x) = \prod_{c\in[d]} \mathbb{P}_{X\sim\mathcal{N}(0,\Sigma_{c,c})}(X > x(c))$. When Σ is not diagonal, we derive a lower bound on $\mathbb{P}_{X\sim\mathcal{N}(0_d,\Sigma)}(X > x)$ for any vector $x \in \mathbb{R}^d$ (Lemma 7). To further lower bound those quantities, we introduce the ratio of the tail distribution to the density function for Gaussian, known as the Mills ratio (Mills, 1926). Under the event $\mathcal{E}_{t,\delta/2}$, we have

$N_{t,k}\|\mu_k - \hat{\mu}_{t,k}\|_{\Sigma^{-1}} \leq 2\beta(t-1, \delta/2)$. Using that the Mills ratio is non-increasing and log-convex, we obtain that

$$\mathbb{1}_{\mathcal{E}_{t,\delta/2} \cap \{\widehat{S}_t \neq S^\star\}} \mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\mathrm{Alt}(\widehat{S}_t)) \geq \delta q(t-1, \delta)$$

by taking $c(t, \delta) = \beta(t, \delta/2)/\log(1/\delta)$. This concludes the proof by considering $M(t, \delta) = \left\lceil \frac{\log(2t^s \zeta(s)/\delta)}{\delta q(t,\delta)} \right\rceil$.

Structured Setting Both Θ and the arms $\mathcal{A} \subseteq \mathbb{R}^h$ will introduce correlations between arms. Therefore, one cannot lower bound $\mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\Theta \cap \mathrm{Alt}(\widehat{S}_t))$ by considering arms separately. Consequently, choosing M and c to ensure δ-correctness for PSI in the transductive linear setting is a challenging open problem.

When $d = 1$ (BAI) and $\Theta = \mathbb{R}^h$, we can show that $\mathbb{1}_{\widehat{S}_t \neq S^\star} \mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\mathrm{Alt}(\widehat{S}_t)) \geq \mathbb{P}_{X\sim\mathcal{N}(0,1)}(X > \|vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta})\|_{\boldsymbol{\Sigma}_t^{-1}}/\sqrt{c(t-1, \delta)})$. Then, under $\mathcal{E}_{t,\delta/2}$, we conclude similarly using the Mills ratio properties.

Technical Challenge The δ-correctness of the GLR stopping rule is obtained quite simply by concentration. We have $\mathrm{GLR}(t) \leq \frac{1}{2}\|vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta})\|^2_{\boldsymbol{\Sigma}_t^{-1}} \leq \beta(t-1, \delta)$ under $\mathcal{E}_{t,\delta} \cap \{\widehat{S}_t \neq S^\star\}$, hence $\mathbb{P}_{\boldsymbol{\nu}}(\mathcal{E}_\delta \cap \{\tau_\delta^{\mathrm{GLR}} < +\infty, \widehat{S}_{\tau_\delta^{\mathrm{GLR}}} \neq S^\star\}) = 0$. In contrast, under a similar event, proving the δ-correctness of the PS stopping rule requires to control the randomness of $\widehat{\Pi}_t \mid \mathcal{F}_t$ by deriving anti-concentration results on $\mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\Theta \cap \mathrm{Alt}(\widehat{S}_t))$.

3.1.2 Proof Sketch of Theorem 1

Studying the expected sample complexity of an algorithm using the PS stopping rule requires to control the randomness of $\widehat{\Pi}_t \mid \mathcal{F}_t$ by deriving concentration results on $\mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\Theta \cap \mathrm{Alt}(\widehat{S}_t))$ since a direct union bound yields that

$$\mathbb{P}_{\boldsymbol{\nu}}(\tau_\delta^{\mathrm{PS}} > t) \leq M(t-1, \delta)\mathbb{E}_{\boldsymbol{\nu}}[\mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\Theta \cap \mathrm{Alt}(\widehat{S}_t))] \,.$$

Since $\mathrm{Alt}(\widehat{S}_t)$ is a union of convex sets (Lemma 26), we use a Gaussian concentration result on convex sets (Lemma 27) for $\widehat{\Pi}_t \mid \mathcal{F}_t$. This relates $\mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\Theta \cap \mathrm{Alt}(\widehat{S}_t))$ to the GLR statistic defined in (1), namely we show that

$$\mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\Theta \cap \mathrm{Alt}(\widehat{S}_t)) \leq \alpha_0 \exp(-\mathrm{GLR}(t)/c(t-1, \delta)) \,,$$

where $\alpha_0 \leq |\mathcal{Z}|(|\mathcal{Z}| + d^{|\mathcal{Z}|})/2$ (Lemma 17). Once the event $\{\tau_\delta^{\mathrm{PS}} > t\}$ is linked to the GLR statistic, it suffices to lower bound $\mathrm{GLR}(t)$ to conclude the proof. The rest of the proof would be conducted similarly when using the GLR stopping rule.

Studying the saddle-point convergence of our algorithm yields that, there exists $(\mathcal{E}_t)_{t\geq 1}$ and $T_0 \in \mathbb{N}$ such that for all $t > T_0$, when $\mathcal{E}_t$ holds, $\mathrm{GLR}(t) \geq (t-1)/T^\star(\boldsymbol{\theta}) - f(t)$ where $f(t) =_{+\infty} o(t)$ and $\sum_{t\in\mathbb{N}} \mathbb{P}(\mathcal{E}_t^c) = C_0 < +\infty$. Then, we have $\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta^{\mathrm{PS}}] \leq T_0 + C_0 +$

$$\sum_{t>T_0} \alpha_0 \exp\left(-\frac{t-1-T^\star(\boldsymbol{\theta})f(t)}{T^\star(\boldsymbol{\theta})c(t-1,\delta)} + \log M(t-1,\delta)\right) .$$

Using that $\limsup_{\delta\to 0} \frac{c(t,\delta)\log M(t,\delta)}{\log(1/\delta)} \leq 1$, a direct inversion of the above sum concludes the proof.

4 EXPERIMENTS

We evaluate the performance of PSIPS on real-world-inspired and synthetic instances. As benchmarks, we consider the following PSI algorithms: APE (Kone et al., 2023), the gradient-based algorithm of Crepon et al. (2024) denoted as GAPSI, the *oracle* algorithm which samples arm following optimal weights, i.e $a_t \sim \boldsymbol{w}^\star(\boldsymbol{\theta})$, and round-robin *uniform* sampling (RR).

APE relies on confidence bounds to pull an arm at each round and to tailor its stopping rule. While APE can be extended to the structured setting, it does not exploit the correlation between objectives as its confidence intervals only consider the marginals' variance. GAPSI computes a supergradient of the GLR (1) and uses the GLR stopping rule. Similarly, GAPSI has no efficient implementation or guarantees for correlated objectives and the structured setting. Since GAPSI is computationally expensive, we exclude it from the experiment in the linear setting (due to the large number of arms), and we run it as heuristic in experiments with correlated objectives.

In our experiments, we use the heuristics $c(t, \delta) = 1 + \frac{\log\log t}{\log(1/\delta)}$ and $M(t, \delta) = \frac{1}{\delta}\log\frac{t}{\delta}$. For a fair comparison, APE uses $\beta(t, \delta) = \log(1/\delta) + \log\log t$ in its bounds, hence improving its performance. We report the averaged results over 500 independent runs with boxplots or shaded areas for standard deviation. The observed empirical error is order of magnitude lower than δ. Appendix J.3 contains supplementary experiments (e.g. higher (K, d)).

Cov-Boost Trial As prior work in PSI, we use the dataset from Munro et al. (2021) to simulate a bandit instance for PSI. The dataset consists of the average response of patients cohorts in a covid19 trial including 20 vaccine strategies. Among the indicators recorded to measure the efficacy of each strategy, following Kone et al. (2023); Crepon et al. (2024), we keep three of them: titres of neutralizing antibodies and immunoglobuline G and the wild-type cellular response (cf Table 3). The ideal vaccine strategy would maximize all three indicators, but the average response for the considered metrics reveals a Pareto set of two vaccine strategies in the trial.

Figure 1 shows that PSIPS has a high variability of stopping time, which was expected due to the use of posterior

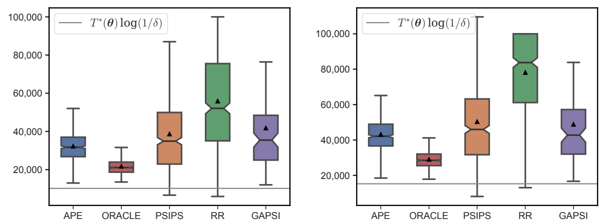

Figure 1: Empirical stopping times on the covid19 experiment with $\delta = 0.01$ (left) and $\delta = 0.001$ (right).

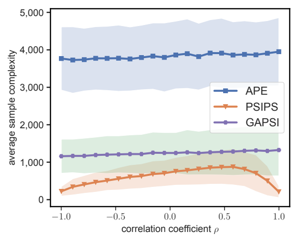

Figure 2: Impact of the correlation ρ on the stopping time.

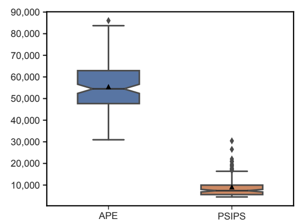

Figure 3: Empirical stopping time in the NoC experiment.

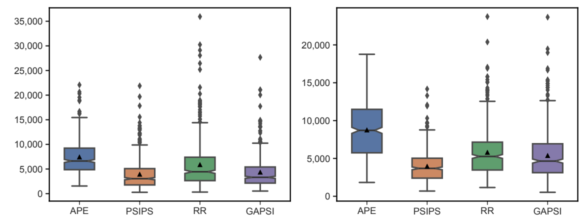

Figure 4: Empirical stopping time on random Gaussian (left) and Bernoulli (right) instances.

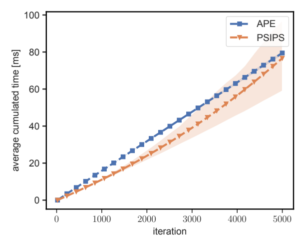

Figure 5: Average runtime for the first T iterations in the covid19 experiment.

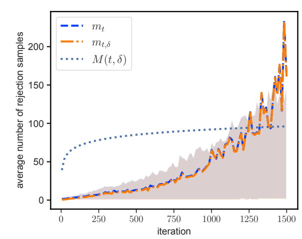

Figure 6: Average number of per-round rejection samples $m_t, m_{t,\delta}$.

sampling. While PSIPS outperforms uniform sampling, it performed on average slightly worse than APE and Oracle on the Cov-Boost instance. For $\delta = 0.1$, we reported an average sample complexity of 20456 for PSIPS compared to 17909 reported for GAPSI by Crepon et al. (2024). However, due to the high computational cost of their algorithm, we averaged its performance over 100 runs.

Correlated Objectives To assess the performance of PSIPS when the objectives are correlated (Σ not diagonal), we choose a Gaussian instance with five arms in dimension two (see Appendix J for the values). The covariance matrix Σ_ρ is chosen with unit variance and an off-diagonal correlation coefficient $\rho \in (-1, 1)$: ρ close to 1 (resp. -1) means the two objectives are strongly positively (resp. negatively) correlated and $\rho = 0$ means independent objectives.

Figure 2 shows the sample complexity of PSIPS versus ρ, including GAPSI as heuristic. It reveals that PSIPS has a decreasing sample complexity for stronger (negative) correlation between objectives; reducing by up to factor 3 w.r.t the uncorrelated case ($\rho = 0$). Due to its asymptotic optimality, PSIPS inherits the properties of $T^\star(\boldsymbol{\theta})$, which is most likely decreasing with negative ρ on this specific instance. Both GAPSI and APE performance are independent of ρ.

Robustness on Random Instances To evaluate the robustness of PSIPS, we measure its performance on 250 randomly picked Bernoulli (marginals) and Gaussian instances with $K = 5, d = 2$ and $\Sigma = I_2/2$.

Figure 4 showcases the competitive performance of PSIPS, which remains good when used on sub-Gaussian instances.

Structured Instance Network on Chip (NoC, Almer et al. (2011); Zuluaga et al. (2013)) is a bi-objective optimization dataset designed for hardware development. The objective is to minimize two key performance criteria: energy consumption and runtime of NoC implementations. The dataset includes 259 designs, each characterized by four descriptive features, and its Pareto set comprises 4 arms (see Appendix J).

Figure 3 shows the superior performance of PSIPS on the structured instance as it can exploit the linear structure to reduce the sample complexity of PSI. We did not include RR as it was more than five times worse than PSIPS.

Computational Cost We evaluate the number of rejection samples used by PSIPS on a Gaussian instance with $\Sigma = I_2/2$ defined as $\mu_1 = (1, 1)^\intercal$ and $\mu_i = R_{\pi/5}\mu_{i-1}$ for $i \in \{2, 3, 4, 5\}$ where $R_{\pi/5}$ is the $\pi/5$ rotation matrix. Without actually stopping, we record both m_t and $m_{t,\delta}$ at each round, averaged over 1000 runs. We also report the cumulated runtime of PSIPS and APE on the covid19 experiment.

Figure 5 shows that the computational cost of PSIPS is comparable to APE, which is order of magnitudes smaller than GAPSI (see Figure 4 in Crepon et al. (2024)). Figure 6 reveals that, while finding an alternative is initially faster, more rejections are needed before finding an alternative when the posterior concentrates. It is precisely at this time that the PS stopping rule triggers.

5 PERSPECTIVES

We proposed the PSIPS algorithm for Pareto set identification in the transductive linear setting with correlated objectives. By leveraging posterior sampling in both the stopping and the sampling rules, PSIPS is a computationally efficient algorithm that deals with structure and correlation, without using costly oracle calls as existing algorithms. We show that PSIPS is asymptotically optimal both from a frequentist and Bayesian perspective. Moreover, PSIPS has competitive empirical performance.

While our choice of c and M ensures δ-correctness of the PS stopping rule in the unstructured setting with independent or correlated objectives, we lack δ-correctness guarantees in the structured setting. Even for transductive linear BAI ($d = 1$), we obtain δ-correctness only when $\Theta = \mathbb{R}^h$. Therefore, it fails to account for the boundedness assumption, which is a key assumption to derive the upper bound on the expected sample complexity in the linear setting. To address this open problem, we need tight non-asymptotic lower bounds on the posterior probability of drawing a realization in $\Theta \cap \mathrm{Alt}(\widehat{S}_t)$.

Despite being a theoretically convenient distributional assumption, using Gaussian distributions with known covariance matrix is too restrictive for many applications. Therefore, generalizing our results to other classes of distributions is another interesting direction for future work. For example, when the covariance matrices are unknown and different for each arm, the estimation of the Pareto set is intertwined with estimating the correlation structure.

Acknowledgements

This work has been partially supported by the French National Research Agency (ANR) in the framework of the PEPR IA FOUNDRY project (ANR-23-PEIA-0003). Cyrille Kone is funded by an Inria/Inserm PhD grant. The project was conducted while Marc Jourdan was a PhD student in the Scool team of Inria Lille, partially supported by the THIA ANR program "AI_PhD@Lille".

References

S. Agrawal and N. Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International conference on machine learning*. PMLR, 2013.

O. Almer, N. Topham, and B. Franke. A learning-based approach to the automated design of mpsoc networks. In *Proceedings of the 24th International Conference on Architecture of Computing Systems*. Springer-Verlag, 2011.

C. Ararat and C. Tekin. Vector optimization with stochastic bandit feedback. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*. PMLR, 2023.

J.-Y. Audibert, S. Bubeck, and R. Munos. Best arm identification in multi-armed bandits. In *23rd Conference on Learning Theory*, 2010.

P. Auer, C.-K. Chiang, R. Ortner, and M.-M. Drugan. Pareto front identification from stochastic bandit feedback. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*. PMLR, 2016.

A. Baricz. Mills' ratio: Monotonicity patterns and functional inequalities. *Journal of Mathematical Analysis and Applications*, 2008.

Z. W. Birnbaum. An Inequality for Mill's Ratio. *The Annals of Mathematical Statistics*, 1942.

E. Crepon, A. Garivier, and W. M Koolen. Sequential learning of the Pareto front for multi-objective bandits. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research. PMLR, 2024.

S. De Rooij, T. Van Erven, P. D. Grünwald, and W. M. Koolen. Follow the leader if you can, hedge if you must. *J. Mach. Learn. Res.*, 2014.

R. Degenne. *Impact of structure on the design and analysis of bandit algorithms*. PhD thesis, 2019.

R. Degenne and W. Koolen. Pure exploration with multiple correct answers. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2019.

R. Degenne and V. Perchet. Combinatorial semi-bandit with known covariance. *Advances in Neural Information Processing Systems*, 2016.

R. Degenne, W. M. Koolen, and P. Ménard. Non-asymptotic pure exploration by solving games. *Advances in Neural Information Processing Systems*, 2019.

R. Degenne, P. Ménard, X. Shang, and M. Valko. Gamification of pure exploration for linear bandits. In *International Conference on Machine Learning*. PMLR, 2020.

M.-M. Drugan and A. Nowe. Designing multi-objective multi-armed bandits algorithms: A study. In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, 2013.

E. Even-Dar, S. Mannor, and Y. Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, 2006.

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

V. Gabillon, M. Ghavamzadeh, and A. Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2012.

A. Garivier and E. Kaufmann. Optimal best arm identification with fixed confidence. In *29th Annual Conference on Learning Theory*. PMLR, 2016.

A. Garivier, P. Ménard, and G. Stoltz. Explore first, exploit next: The true shape of regret in bandit problems. *Mathematics of Operations Research*, 2019.

K. Jamieson and R. Nowak. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *48th Annual Conference on Information Sciences and Systems (CISS)*, 2014.

C. Jennison and B. W. Turnbull. Group sequential tests for bivariate response: interim analyses of clinical trials with both efficacy and safety endpoints. *Biometrics*, 1993.

M. Jourdan, D. Rémy, and K. Emilie. Dealing with unknown variances in best-arm identification. In *Proceedings of The 34th International Conference on Algorithmic Learning Theory*. PMLR, 2023.

S. Kalyanakrishnan, A. Tewari, P. Auer, and P. Stone. PAC subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th International Conference on International Conference on Machine Learning*. Omnipress, 2012.

Z. Karnin, T. Koren, and O. Somekh. Almost optimal exploration in multi-armed bandits. In *Proceedings of the 30th International Conference on International Conference on Machine Learning*. JMLR, 2013.

J. Katz-Samuels and C. Scott. Feasible arm identification. In *Proceedings of the 35th International Conference on Machine Learning*. PMLR, 2018.

E. Kaufmann, N. Korda, and R. Munos. Thompson sampling: An asymptotically optimal finite-time analysis. In *International conference on algorithmic learning theory*. Springer, 2012.

W. Kim, G. Iyengar, and A. Zeevi. Learning the pareto front using bootstrapped observation samples. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.

C. Kone, E. Kaufmann, and L. Richert. Adaptive algorithms for relaxed pareto set identification. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

C. Kone, E. Kaufmann, and L. Richert. Bandit Pareto set identification: the fixed budget setting. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*. PMLR, 2024.

C. Kone, E. Kaufmann, and L. Richert. Bandit pareto set identification in a multi-output linear model. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.

J. Kuelbs, W. Li, and W. Linde. The gaussian measure of shifted balls. *Probability Theory and Related Fields*, 1994.

H. T. Kung, F. Luccio, and F. P. Preparata. On finding the maxima of a set of vectors. *J. ACM*, 1975.

T. Lattimore and C. Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

Z. Li, K. Jamieson, and L. Jain. Optimal exploration is no harder than Thompson sampling. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*. PMLR, 2024.

D. Lu and W. V. Li. A note on multivariate gaussian estimates. *Journal of Mathematical Analysis and Applications*, 2009.

J. P. Mills. Table of the ratio: Area to bounding ordinate, for any portion of normal curve. *Biometrika*, 1926.

A.-P.-S. Munro, L. Janani, V. Cornelius, and et al. Safety and immunogenicity of seven COVID-19 vaccines as a third dose (booster) following two doses of ChAdOx1 nCov-19 or BNT162b2 in the UK (COV-BOOST): a blinded, multicentre, randomised, controlled, phase 2 trial. *The Lancet*, 2021.

D. Russo. Simple bayesian algorithms for best arm identification. In *29th Annual Conference on Learning Theory*. PMLR, 2016.

X. Shang, R. de Heide, P. Menard, E. Kaufmann, and M. Valko. Fixed-confidence guarantees for bayesian best-arm identification. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*. PMLR, 2020.

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

W. R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 1933.

S. Wang and J. Zhu. Thompson sampling for (combinatorial) pure exploration. In *International Conference on Machine Learning*. PMLR, 2022.

M. Xu and D. Klabjan. Pareto regret analyses in multi-objective multi-armed bandit. In *Proceedings of the 40th International Conference on Machine Learning*. JMLR, 2023.

M. Zuluaga, G. Sergent, A. Krause, and M. Püschel. Active learning for multi-objective optimization. In *Proceedings of the 30th International Conference on Machine Learning*. PMLR, 2013.

M. Zuluaga, A. Krause, and M. Püschel. e-pal: An active learning approach to the multi-objective optimization problem. *Journal of Machine Learning Research*, 2016.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. **Yes, see Sections 1, 2 and 3.**
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. **Yes, see Sections 2, 3 and 4, as well as Appendix J.**
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. **Yes, see Appendix J and the code attached in the supplementary materials.**

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. **Yes, see Sections 2 and 3.**
   (b) Complete proofs of all theoretical results. **Yes, see Section 3.1 and Appendices C, D, F, G, H and I.**
   (c) Clear explanations of any assumptions. **Yes, see Sections 1 and 3.**

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). **Yes, see Appendix J and the code attached in the supplementary materials.**
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). **Yes, see Appendix J and the code attached in the supplementary materials.**
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). **Yes, see Section 4 and Appendix J.**
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). **Yes, see Appendix J.**

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. **Yes, see section 4** and Appendix J.
   (b) The license information of the assets, if applicable. **Not Applicable.**
   (c) New assets either in the supplemental material or as a URL, if applicable. **Not Applicable.**
   (d) Information about consent from data providers/curators. **Not Applicable.**
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. **Not Applicable.**

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. **Not Applicable.**
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. **Not Applicable.**
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. **Not Applicable.**

A OUTLINE

The appendices are organized as follows:

- Notation are summarized in Appendix B.
- The δ-correctness of the stopping rule is proven in Appendix C.
- Appendix D shows that the game-based sampling rule yields a saddle-point algorithm.
- In Appendix E, we prove some results related to the generalized likelihood and the posterior probability of error.
- The asymptotic upper bound on the expected sample complexity is proven in Appendix F.
- The convergence of the posterior distribution is proven in Appendix G.
- Appendix H gathers concentration results.
- Appendix I gathers existing and new technical results which are used for our proofs.
- Implementation details and additional experiments are presented in Appendix J.

Table 1: Notation for the setting.

$K := \lvert\mathcal{A}\rvert$	number of arms
$\Theta \subset \mathbb{R}^{h\times d}$	set of parameters
$\mathcal{A} \subset \mathbb{R}^h$	set of measurements (arms)
$\mathcal{Z} \subset \mathbb{R}^h$	set of items
$\boldsymbol{A} \in \mathbb{R}^{K\times h}$	matrix of features
$\boldsymbol{\theta} \in \Theta$	true parameter of the instance
$\boldsymbol{\mu} := \boldsymbol{A}\boldsymbol{\theta} := (\mu_1 \dots \mu_K)^\intercal$	vector mean of each arm
$S^\star(\boldsymbol{\theta})$	Pareto set of $\mathcal{Z}$
$\mathcal{D}_\Theta := \max_{a\in\mathcal{A}} \max_{\boldsymbol{\theta},\boldsymbol{\theta}'} \lVert vec(\boldsymbol{\theta}-\boldsymbol{\theta}')\rVert_{\Sigma^{-1}\otimes aa^\intercal}$	diameter of the means space
$L_\mathcal{A} := \max_{a\in\mathcal{A}} \lVert a\rVert_2$	maximum feature norm
$L_* := 2\max_{a\in\mathcal{A}} \lVert a\rVert^2_{V(\boldsymbol{w}_{\text{exp}})^{-1}}$	squared feature norm due to forced exploration
$\Delta_{\min}$	the minimum Pareto gap (Section D.1)

B NOTATION

We recall some commonly used notation: the set of integers $[K] := \{1, \cdots, K\}$, the complement $X^\complement$ and interior $\mathring{X}$ and the closure $\overline{X}$ of a set X, the indicator function $\mathbb{1}(X)$ of an event, the probability $\mathbb{P}_{\boldsymbol{\nu}}$ and the expectation $\mathbb{E}_{\boldsymbol{\nu}}$ under distribution $\boldsymbol{\nu}$, Landau's notation o, $\mathcal{O}$, Ω and Θ, the $(K-1)$-dimensional probability simplex $\triangle_K := \left\{w \in \mathbb{R}^K_+ \mid w \geq 0, \sum_{i\in[K]} w_i = 1\right\}$. ζ is the Riemann ζ function. When $\boldsymbol{\lambda}$ is a matrix and π is a distribution supported on vectors, we write $\boldsymbol{\lambda} \sim \pi$ to denote $vec(\boldsymbol{\lambda}) \sim \pi$. When $\boldsymbol{M}$ is a set of matrices $vec(\boldsymbol{M}) := \{vec(\boldsymbol{m}) \mid \boldsymbol{m} \in \boldsymbol{M}\}$.

While Table 1 gathers problem-specific notation, Table 2 groups notation for the algorithms.

Table 2: Notation for algorithms.

$\mathrm{Alt}(S)$	set of parameters whose Pareto set is not S
$\widehat{S}_t$	Pareto set of the empirical mean at time t
$\widehat{\boldsymbol{\theta}}_t$	(regularized) least-squares estimator of $\boldsymbol{\theta}$
$a_t \in \mathcal{A}$	measurement (arm) pulled at time t
$V_t := \xi I_h + V_{\boldsymbol{N}_t}$	design matrix after t iterations
$L_\Theta := L_{\mathcal{M}} := \max_{\boldsymbol{\lambda}\in\Theta} \max_{c\in[d]} \lVert \boldsymbol{\lambda} e_c \rVert_2$	maximum regression norm
$\widetilde{a} := \Sigma^{-1/2} \otimes a$	extended feature (matrix)
$\widetilde{V}_t := \Sigma^{-1} \otimes V_t$	extended desing matrix
$\Lambda_t := \Theta_t \cap \mathrm{Alt}(\widehat{S}_t)$	truncated alternative space
$\pi_t := \mathcal{N}(vec(\boldsymbol{\theta}_t), \eta_t^{-1/2}\boldsymbol{\Sigma}_t; \Lambda_t)$	Gaussian truncated to Λ_t
$\boldsymbol{\lambda}_t \sim \pi_t$	strategy of the *min* player at time t
$\boldsymbol{w}_t$	strategy of the *max* player at time t
$\mathcal{F}_t := \sigma\left(a_1, y_1, \dots a_{t+1}, y_t\right)$	information gathered by the algorithm after t rounds
$X_t \sim \nu_{a_t}$	observation collected from ν_{a_t}
$\varepsilon_t := X_t - (I_d \otimes a_t) vec(\boldsymbol{\mu})$	centered (sub)Gaussian noise at time t
η_t	learning rate for the *min* learner (Section D.2)

C CORRECTNESS OF THE STOPPING RULE

In this section, we prove the correctness of the PS (Posterior Sampling) stopping rule. We recall the definition of $\tau_\delta^{\mathrm{PS}}$:

$$\tau_\delta^{\mathrm{PS}} := \inf\left\{ t \mid \forall m \in [M(t-1,\delta)],\ \boldsymbol{\theta}_t^m \notin \Theta \cap \mathrm{Alt}(\widehat{S}_t) \right\} ,$$

where, conditioned on $\mathcal{F}_t$, $\boldsymbol{\theta}_t, \boldsymbol{\theta}_t^1, \dots, \boldsymbol{\theta}_t^m$ are *i.i.d* samples from $\mathcal{N}(vec(\widehat{\boldsymbol{\theta}}_t), c(t-1,\delta)\boldsymbol{\Sigma}_t)$. To prove the correctness of this stopping rule, we will have to control the randomness in the posterior, which we do by introducing the concentration event $\mathcal{E}_\delta := \bigcap_{t\in\mathbb{N}} \mathcal{E}_{t,\delta}$ with

$$\mathcal{E}_{t,\delta} := \left\{ \left\lVert vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_t) \right\rVert^2_{\boldsymbol{\Sigma}_t^{-1}} \le 2\beta(t-1,\delta) \right\} ,$$

for which, correct values of the thresholds are given in the lemma below.

Lemma 3. *Let $s > 1$, ζ be the Riemann ζ function and $\overline{W}_{-1}(x) := -W_{-1}(-e^{-x})$ for all $x \ge 1$, where W_{-1} is the negative branch of the Lambert W function. Let $\mathcal{E}_\delta$ in (6). Then, we have $\mathbb{P}_{\boldsymbol{\nu}}(\mathcal{E}_\delta^c) \le \delta$ by taking*

$$\beta(t,\delta) = \frac{dK}{2} \overline{W}_{-1}\left(\frac{2}{dK} \log \frac{e^{Ks}\zeta(s)^K}{\delta} + \frac{2s}{d} \log\left(1 + \frac{d}{2s}\log\frac{t}{K}\right) + 1 \right)$$

in the unstructured setting, and taking

$$\sqrt{\beta(t,\delta)} = \sqrt{\log\left(\frac{1}{\delta}\left(\frac{L_{\mathcal{A}}^2}{h\xi}t + 1\right)^{\frac{dh}{2}}\right)} + \sqrt{\frac{dL_{\mathcal{M}}^2}{2\lambda_{\min}(\Sigma)\xi}}$$

in the transductive linear setting, where $L_{\mathcal{A}} := \max_{a\in\mathcal{A}} \lVert a \rVert_2$ and $L_{\mathcal{M}} := \max_{\boldsymbol{\lambda}\in\Theta} \max_{c\in[d]} \lVert \boldsymbol{\lambda} e_c \rVert_2$.

The following result shows that to prove the δ-correctness, it is sufficient to show some anti-concentration of the posterior under the event $\mathcal{E}$. $\mathbb{P}_{\boldsymbol{\nu}}(\mathcal{E}_{\delta/2} \cap \{\tau_\delta^{\mathrm{PS}} < +\infty, \widehat{S}_{\tau_\delta^{\mathrm{PS}}} \neq S^\star\}) \le \delta/2$.

Lemma 4. *For all c, M, the PS stopping rule satisfies*

$$\mathbb{P}_{\boldsymbol{\nu}}(\tau_\delta^{\mathrm{PS}} < +\infty, \widehat{S}_{\tau_\delta^{\mathrm{PS}}} \neq S^\star) \le \delta/2 + \mathbb{E}\left[\sum_{t\ge 1} \mathbb{1}_{\mathcal{E}_{t,\delta/2}} \mathbb{1}_{\widehat{S}_t \neq S^\star} \exp\left(-M(t-1,\delta) \mathbb{P}_{\boldsymbol{\nu}}\left(\boldsymbol{\theta}_t \in \Theta \cap \mathrm{Alt}(\widehat{S}_t) \mid \mathcal{F}_t \right) \right) \right] . \quad (7)$$

Proof. Using the definition of the PS stopping rule in (2), we obtain

$$
\begin{aligned}
\mathbb{P}_{\boldsymbol{\nu}}\left(\tau_\delta^{\mathrm{PS}} < \infty, \widehat{S}_{\tau_\delta^{\mathrm{PS}}} \neq S^\star\right) &\leq \delta/2 + \mathbb{P}_{\boldsymbol{\nu}}\left(\mathcal{E}_{\delta/2} \cap \{\tau_\delta^{\mathrm{PS}} < \infty, \widehat{S}_{\tau_\delta^{\mathrm{PS}}} \neq S^\star\}\right) \\
&\leq \delta/2 + \sum_{t\geq 1} \mathbb{P}_{\boldsymbol{\nu}}\left(\mathcal{E}_{t,\delta/2} \cap \{\tau_\delta^{\mathrm{PS}} = t, \widehat{S}_{\tau_\delta^{\mathrm{PS}}} \neq S^\star\}\right) \\
&= \delta/2 + \mathbb{E}\left[\sum_{t\geq 1} \mathbb{1}_{\mathcal{E}_{t,\delta/2}} \mathbb{1}_{\widehat{S}_t \neq S^\star} \mathbb{P}_{\boldsymbol{\nu}}\left(\tau_\delta^{\mathrm{PS}} = t \mid \mathcal{F}_t\right)\right] \\
&\leq \delta/2 + \mathbb{E}\left[\sum_{t\geq 1} \mathbb{1}_{\mathcal{E}_{t,\delta/2}} \mathbb{1}_{\widehat{S}_t \neq S^\star} \mathbb{P}_{\boldsymbol{\nu}}\left(\forall m \leq M(t-1,\delta), \boldsymbol{\theta}_t^m \notin \Theta \cap \mathrm{Alt}(\widehat{S}_t) \mid \mathcal{F}_t\right)\right] \\
&= \delta/2 + \mathbb{E}\left[\sum_{t\geq 1} \mathbb{1}_{\mathcal{E}_{t,\delta/2}} \mathbb{1}_{\widehat{S}_t \neq S^\star} \mathbb{P}_{\boldsymbol{\nu}}\left(\boldsymbol{\theta}_t \notin \Theta \cap \mathrm{Alt}(\widehat{S}_t) \mid \mathcal{F}_t\right)^{M(t-1,\delta)}\right] \\
&= \delta/2 + \mathbb{E}\left[\sum_{t\geq 1} \mathbb{1}_{\mathcal{E}_{t,\delta/2}} \mathbb{1}_{\widehat{S}_t \neq S^\star} \left(1 - \mathbb{P}_{\boldsymbol{\nu}}\left(\boldsymbol{\theta}_t \in \Theta \cap \mathrm{Alt}(\widehat{S}_t) \mid \mathcal{F}_t\right)\right)^{M(t-1,\delta)}\right]
\end{aligned}
$$

which follows since the samples $\boldsymbol{\theta}_t, \boldsymbol{\theta}_t^1, \dots, \boldsymbol{\theta}_t^m$ are *i.i.d* given $\mathcal{F}_t$. Further recalling that $1 - x \leq \exp(-x)$, it follows

$$
\mathbb{P}_{\boldsymbol{\nu}}\left(\tau_\delta^{\mathrm{PS}} < \infty, \widehat{S}_{\tau_\delta^{\mathrm{PS}}} \neq S^\star\right) \leq \delta/2 + \mathbb{E}\left[\sum_{t\geq 1} \mathbb{1}_{\mathcal{E}_{t,\delta/2}} \mathbb{1}_{\widehat{S}_t \neq S^\star} \exp\left(-M(t-1,\delta)\mathbb{P}_{\boldsymbol{\nu}}\left(\boldsymbol{\theta}_t \in \Theta \cap \mathrm{Alt}(\widehat{S}_t) \mid \mathcal{F}_t\right)\right)\right]. \tag{8}
$$

□

Lemma 4 shows that choosing c, M such that

$$
\mathbb{E}\left[\sum_{t\geq 1} \mathbb{1}_{\mathcal{E}_{t,\delta/2}} \mathbb{1}_{\widehat{S}_t \neq S^\star} \exp\left(-M(t-1,\delta)\mathbb{P}_{\boldsymbol{\nu}}\left(\boldsymbol{\theta}_t \in \Theta \cap \mathrm{Alt}(\widehat{S}_t) \mid \mathcal{F}_t\right)\right)\right]
$$

is not larger than $\delta/2$ will ensure the correctness of the PS stopping rule. However, it is not possible to pick $M(t-1,\delta) \propto \mathbb{P}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_t \in \Theta \cap \mathrm{Alt}(\widehat{S}_t) \mid \mathcal{F}_t)$ as the latter quantity is intractable due to multidimensional integration over a non-trivial domain. Instead, we will derive in the following sections, tractable lower bounds on $\mathbb{1}_{\mathcal{E}_{t,\delta/2}} \mathbb{1}_{S_t \neq S^\star} \mathbb{P}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_t \in \Theta \cap \mathrm{Alt}(\widehat{S}_t) \mid \mathcal{F}_t)$ with

$$
\begin{aligned}
\mathrm{Alt}(\widehat{S}_t) = &\left(\bigcup_{x \neq z \in \widehat{S}_t^2} \bigcap_{c \in [d]} \{\boldsymbol{\lambda} \in \mathbb{R}^{h \times d} \mid \langle E_{z,x}(c), vec(\boldsymbol{\lambda})\rangle \geq 0\}\right) \\
&\bigcup \left(\bigcup_{z \notin \widehat{S}_t} \bigcap_{x \in \widehat{S}_t} \bigcap_{c \in [d]} \{\boldsymbol{\lambda} \in \mathbb{R}^{h \times d} \mid \langle E_{x,z}(c), vec(\boldsymbol{\lambda})\rangle \geq 0\}\right),
\end{aligned} \tag{9}
$$

where $E_{z,x}(c) = e_c \otimes (z - x)$ and $e_c = (\mathbb{1}(c' = c))_{c' \in [d]}$.

C.1 Unstructured Setting

In this section, we calibrate c, M to ensure the correctness of the PS stopping rule in the unstructured setting. In this setting, we have $\Theta = \mathbb{R}^{K \times d}$, $\boldsymbol{A} = I_K$, so that the quantity to control is

$$
\mathbb{1}_{\mathcal{E}_{t,\delta/2}} \mathbb{1}_{\widehat{S}_t \neq S^\star} \mathbb{P}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_t \in \mathrm{Alt}(\widehat{S}_t) \mid \mathcal{F}_\infty) \quad \text{with} \quad vec(\boldsymbol{\theta}_t - \widehat{\boldsymbol{\theta}}_t) \mid \mathcal{F}_t \sim \mathcal{N}(\mathbf{0}_{Kd}, c(t-1,\delta)\boldsymbol{\Sigma}_t),
$$

and

$$
\mathrm{Alt}(\widehat{S}_t) = \left(\bigcup_{i \neq j \in \widehat{S}_t^2} \{\boldsymbol{\lambda} := (\lambda_1 \dots \lambda_K)^\intercal \in \Theta \mid \lambda_i \prec \lambda_j\}\right) \bigcup \left(\bigcup_{i \notin \widehat{S}_t} \left\{\boldsymbol{\lambda} := (\lambda_1 \dots \lambda_K)^\intercal \in \Theta \mid \forall j \in \widehat{S}_t, \lambda_i \not\prec \lambda_j\right\}\right).
$$

Indeed, recalling that $\mathrm{Alt}(S)$ is the set of parameters for which the Pareto set differs from S. To change the Pareto set, we either add an arm or remove one from it. For a parameter instance $\boldsymbol{\lambda} := (\lambda_1 \dots \lambda_K)^\intercal$, if $i \in S$ and there exists $j \in S$ such that $\lambda_i \prec \lambda_j$, then as i is a dominated arm in the instance $\boldsymbol{\lambda}$, we will have have $S^\star(\boldsymbol{\lambda}) \neq S$. Similarly, if for some $i \notin S$ it holds that for all $j \in S$, $\lambda_i \not\prec \lambda_j$, then we cannot have $S^\star(\boldsymbol{\lambda}) = S$, otherwise, as $i \notin S^\star(\boldsymbol{\lambda})$, an arm from $S^\star(\boldsymbol{\lambda}) = S$ would have dominated i. This is formally shown in Lemma 26, where it is further proven that

$$\mathrm{Alt}(S) = \left(\bigcup_{(i,j) \in S^2, i \neq j} \{\boldsymbol{\lambda} \in \mathbb{R}^{d\times K} \mid \lambda_i \prec \lambda_j) \right) \bigcup \left(\bigcup_{i \in S^c} \bigcup_{\bar{d}^i \in [d]^S} \{\boldsymbol{\lambda} \mid \forall j \in S, \lambda_i(d^i(\sigma(j))) \geq \lambda_j(d^i(\sigma(j)))) \right)$$

where σ is any permutation that maps S onto $\{1, \dots, |S|\}$. In the sequel, we let

$$\mathrm{Alt}^-(S) := \bigcup_{(i,j)\in S^2, i\neq j} \{\boldsymbol{\lambda} \mid \lambda_i \prec \lambda_j\} \text{ and } \mathrm{Alt}^+(S) := \bigcup_{i \in S^c} \bigcup_{\bar{d}^i \in [d]^S} \{\boldsymbol{\lambda} \mid \forall j \in S : \lambda_i(d^i(\sigma(j))) \geq \lambda_j(d^i(\sigma(j)))\}$$

so that $\mathrm{Alt}(S) = \mathrm{Alt}^+(S) \cup \mathrm{Alt}^-(S)$. We prove the following lemma.

Lemma 5. *It holds that*

$$\begin{aligned}
\mathbb{1}_{\widehat{S}_t \neq S^\star} \mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\mathrm{Alt}(\widehat{S}_t)) \geq \max \Bigg\{ & \max_{i \notin \widehat{S}_t, \bar{d} \in [d]^{|\widehat{S}_t|}} \mathbb{1}_{(\forall j \in \widehat{S}_t, \mu_i(\bar{d}^\sigma_j) \geq \mu_j(\bar{d}^\sigma_j))} \mathbb{P}_{Y \sim \mathcal{N}(0_d, \Sigma)} \left(Y > \sqrt{\frac{N_{t,i}}{c(t-1,\delta)}}(\mu_i - \widehat{\mu}_{t,i}) \right) \\
& \cdot \prod_{j \in \widehat{S}_t} \mathbb{P}_{X \sim \mathcal{N}(0,1)} \left(X > \sqrt{\frac{N_{t,j}}{c(t-1,\delta)}} \|\mu_j - \widehat{\mu}_{t,j}\|_{\Sigma^{-1}} \right), \\
& \max_{i \neq j \in \widehat{S}_t^2,} \mathbb{1}_{(\mu_i \prec \mu_j)} \mathbb{P}_{Y \sim \mathcal{N}(0_d, \Sigma)} \left(Y \geq \sqrt{\left(\frac{1}{N_{t,i}} + \frac{1}{N_{t,j}}\right)^{-1} \frac{1}{c(t-1,\delta)}} \left((\widehat{\mu}_{t,i} - \widehat{\mu}_{t,j}) - (\mu_i - \mu_j)\right) \right) \Bigg\}.
\end{aligned}$$

Proof.

$$\begin{aligned}
\mathbb{1}_{\widehat{S}_t \neq S^\star} \mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\mathrm{Alt}^-(\widehat{S}_t)) &\geq \mathbb{1}_{\widehat{S}_t \neq S^\star} \max_{(i,j)\in \widehat{S}_t, i\neq j} \mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\{\boldsymbol{\lambda} = (\lambda_1 \dots \lambda_K)^\intercal \in \mathbb{R}^{d\times K} : \lambda_i \prec \lambda_j\}) \\
&\geq \max_{(i,j)\in \widehat{S}_t, i\neq j} \mathbb{1}_{(\mu_i \prec \mu_j)} \mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\{\boldsymbol{\lambda} = (\lambda_1 \dots \lambda_K)^\intercal \in \mathbb{R}^{d\times K} : \lambda_i \prec \lambda_j\}) \\
&= \max_{(i,j)\in \widehat{S}_t, i\neq j} \mathbb{1}_{(\mu_i \prec \mu_j)} \mathbb{P}_{\boldsymbol{\nu}}(\theta_{t,i} \prec \theta_{t,j} \mid \mathcal{F}_t),
\end{aligned}$$

where $\theta_{t,i} = \boldsymbol{\theta}^T \widetilde{e}_i$ and $\widetilde{e}_1, \dots, \widetilde{e}_K$ is the cananical basis of $\mathbb{R}^K$. In the unstructured setting, it is simple to check that $\boldsymbol{\theta}_t^\intercal = (\theta_{t,1} \dots \theta_{t,K})$ where $\theta_{t,i} \mid \mathcal{F}_t \sim \mathcal{N}(\widehat{\mu}_{t,i}, c(t-1,\delta)\Sigma/N_{t,i})$. Therefore, we have

$$\begin{aligned}
\mathbb{1}_{\widehat{S}_t \neq S^\star} \mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\mathrm{Alt}^-(\widehat{S}_t)) &\geq \max_{(i,j)\in \widehat{S}_t, i\neq j} \mathbb{1}_{(\mu_i \prec \mu_j)} \mathbb{P}_{\boldsymbol{\nu}}((\theta_{t,i} - \theta_{t,j}) - (\widehat{\mu}_{t,i} - \widehat{\mu}_{t,j}) \prec -(\widehat{\mu}_{t,i} - \widehat{\mu}_{t,j}) \mid \mathcal{F}_t) \\
&\geq \max_{(i,j)\in \widehat{S}_t, i\neq j} \mathbb{1}_{(\mu_i \prec \mu_j)} \mathbb{P}_{\boldsymbol{\nu}}((\theta_{t,i} - \theta_{t,j}) - (\widehat{\mu}_{t,i} - \widehat{\mu}_{t,j}) \prec (\mu_i - \mu_j) - (\widehat{\mu}_{t,i} - \widehat{\mu}_{t,j}) \mid \mathcal{F}_t)
\end{aligned}$$

then, observe that $(\theta_{t,i} - \theta_{t,j}) \mid \mathcal{F}_t \sim \mathcal{N}\left(\widehat{\mu}_{t,i} - \widehat{\mu}_{t,j}, (N_{t,i}^{-1} + N_{t,j}^{-1})c(t-1,\delta)\Sigma\right)$, so that letting $Y \sim \mathcal{N}(0_d, \Sigma)$, we have

$$\begin{aligned}
&\mathbb{1}_{\widehat{S}_t \neq S^\star} \mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\mathrm{Alt}^-(\widehat{S}_t)) \\
&\geq \max_{i \neq j \in \widehat{S}_t^2} \mathbb{1}_{(\mu_i \prec \mu_j)} \mathbb{P}_{Y \sim \mathcal{N}(0_d, \Sigma)} \left(Y \prec \sqrt{\left(\frac{1}{N_{t,i}} + \frac{1}{N_{t,j}}\right)^{-1} \frac{1}{c(t-1,\delta)}} \left((\mu_i - \mu_j) - (\widehat{\mu}_{t,i} - \widehat{\mu}_{t,j})\right) \right) \\
&\max_{i \neq j \in S_t^2,} \mathbb{1}_{(\mu_i \prec \mu_j)} \mathbb{P}_{Y \sim \mathcal{N}(0_d, \Sigma)} \left(Y \geq \sqrt{\left(\frac{1}{N_{t,i}} + \frac{1}{N_{t,j}}\right)^{-1} \frac{1}{c(t-1,\delta)}} \left((\widehat{\mu}_{t,i} - \widehat{\mu}_{t,j}) - (\mu_i - \mu_j)\right) \right).
\end{aligned}$$

We now prove a similar result on $\mathrm{Alt}^+(\widehat{S}_t)$. We have

$$
\begin{aligned}
&\mathbb{1}_{\widehat{S}_t \neq S^\star} \mathbb{P}_{\widehat{\Pi}_t | \mathcal{F}_t}(\mathrm{Alt}^+(\widehat{S}_t)) \\
&\geq \max_{i \notin \widehat{S}_t, \bar{d} \in [d]^{|S_t|}} \mathbb{1}_{(\forall j \in \widehat{S}_t, \mu_i(\bar{d}_j^\sigma) \geq \mu_j(\bar{d}_j^\sigma))} \mathbb{P}_{\boldsymbol{\nu}}(\forall j \in \widehat{S}_t, \theta_{t,i}(\bar{d}_j^\sigma) \geq \theta_{t,j}(\bar{d}_j^\sigma) \mid \mathcal{F}_t) \\
&\geq \max_{i \notin \widehat{S}_t, \bar{d} \in [d]^{|\widehat{S}_t|}} \mathbb{1}_{(\forall j \in \widehat{S}_t, \mu_i(\bar{d}_j^\sigma) \geq \mu_j(\bar{d}_j^\sigma))} \mathbb{P}_{\boldsymbol{\nu}}(\forall j \in \widehat{S}_t, (\theta_{t,i} - \mu_i)(\bar{d}_j^\sigma) \geq (\theta_{t,j} - \mu_j)(\bar{d}_j^\sigma) \mid \mathcal{F}_t) \\
&\geq \max_{i \notin \widehat{S}_t, \bar{d} \in [d]^{|\widehat{S}_t|}} \mathbb{1}_{(\forall j \in \widehat{S}_t, \mu_i(\bar{d}_j^\sigma) \geq \mu_j(\bar{d}_j^\sigma))} \mathbb{P}_{\boldsymbol{\nu}}((\theta_{t,i} - \mu_i) > 0_d \mid \mathcal{F}_t) \prod_{j \in \widehat{S}_t} \mathbb{P}_{\boldsymbol{\nu}}((\theta_{t,j} - \mu_j)(\bar{d}_j^\sigma) > 0 \mid \mathcal{F}_t)
\end{aligned}
$$

then, noting again that $(\theta_{t,i} - \widehat{\mu}_{t,i}) \mid \mathcal{F}_t \sim \mathcal{N}(0, c(t-1,\delta)\Sigma / N_{t,i})$, it follows that

$$
\begin{aligned}
\mathbb{1}_{\widehat{S}_t \neq S^\star} \mathbb{P}_{\widehat{\Pi}_t | \mathcal{F}_t}(\mathrm{Alt}^+(\widehat{S}_t)) \geq \max_{i \notin \widehat{S}_t, \bar{d} \in [d]^{|\widehat{S}_t|}} \mathbb{1}_{(\forall j \in \widehat{S}_t, \mu_i(\bar{d}_j^\sigma) \geq \mu_j(\bar{d}_j^\sigma))} \mathbb{P}_{Y \sim \mathcal{N}(0_d, \Sigma)} \left(Y > \sqrt{\frac{N_{t,i}}{c(t-1,\delta)}} (\mu_i - \widehat{\mu}_{t,i}) \right) \\
\cdot \prod_{j \in \widehat{S}_t} \mathbb{P}_{\boldsymbol{\nu}}((\theta_{t,j} - \mu_j)(\bar{d}_j^\sigma) > 0 \mid \mathcal{F}_t). \qquad (10)
\end{aligned}
$$

Further noting that by Cauchy-Schwartz inequality,

$$
(\widehat{\mu}_{t,j} - \mu_j)(\bar{d}_j^\sigma) \quad \leq \quad \|\widetilde{e}_{\bar{d}_j^\sigma}\|_\Sigma \|\mu_j - \widehat{\mu}_{t,j}\|_{\Sigma^{-1}},
$$

it follows for any $j \in \widehat{S}_t$ that

$$
\begin{aligned}
\mathbb{P}_{\boldsymbol{\nu}}((\theta_{t,j} - \mu_j)(\bar{d}_j^\sigma) > 0 \mid \mathcal{F}_t) &= \mathbb{P}_{\boldsymbol{\nu}}((\theta_{t,j} - \widehat{\mu}_{t,j})(\bar{d}_j^\sigma) > (\mu_j - \widehat{\mu}_{t,j})(\bar{d}_j^\sigma) \mid \mathcal{F}_t) \\
&\geq \mathbb{P}_{\boldsymbol{\nu}}((\theta_{t,j} - \widehat{\mu}_{t,j})(\bar{d}_j^\sigma) > \|\widetilde{e}_{\bar{d}_j^\sigma}\|_\Sigma \|\mu_j - \widehat{\mu}_{t,j}\|_{\Sigma^{-1}} \mid \mathcal{F}_t) \\
&= \mathbb{P}_{X \sim \mathcal{N}(0,1)} \left(X > \sqrt{\frac{N_{t,j}}{c(t-1,\delta)}} \|\mu_j - \widehat{\mu}_{t,j}\|_{\Sigma^{-1}} \right),
\end{aligned}
$$

which follows from $(\theta_{t,j} - \widehat{\mu}_{t,j})(\bar{d}_j^\sigma) \mid \mathcal{F}_t \sim \mathcal{N}(0, c(t-1,\delta) \|\widetilde{e}_{\bar{d}_j^\sigma}\| / N_{t,\delta})$. Therefore, combining the last display with (10) yields

$$
\begin{aligned}
\mathbb{1}_{\widehat{S}_t \neq S^\star} \mathbb{P}_{\widehat{\Pi}_t | \mathcal{F}_t}(\mathrm{Alt}^+(\widehat{S}_t)) \geq \max_{i \notin \widehat{S}_t, \bar{d} \in [d]^{|\widehat{S}_t|}} \mathbb{1}_{(\forall j \in \widehat{S}_t, \mu_i(\bar{d}_j^\sigma) \geq \mu_j(\bar{d}_j^\sigma))} \mathbb{P}_{Y \sim \mathcal{N}(0_d, \Sigma)} \left(Y > \sqrt{\frac{N_{t,i}}{c(t-1,\delta)}} (\mu_i - \widehat{\mu}_{t,i}) \right) \\
\cdot \prod_{j \in \widehat{S}_t} \mathbb{P}_{X \sim \mathcal{N}(0,1)} \left(X > \sqrt{\frac{N_{t,j}}{c(t-1,\delta)}} \|\mu_j - \widehat{\mu}_{t,j}\|_{\Sigma^{-1}} \right),
\end{aligned}
$$

which achieves the proof. □

C.1.1 Special Case: PSI with Independent Covariate

In this section, we specialize Lemma 5 to the case where Σ is diagonal. In this case, $\Sigma = \mathrm{diag}(\{\sigma_c^2\}_{c \in [d]})$ and

$$
\begin{aligned}
(*) = \mathbb{P}_{Y \sim \mathcal{N}(0_d, \Sigma)} \left(Y \geq \sqrt{\left(\frac{1}{N_{t,i}} + \frac{1}{N_{t,j}} \right)^{-1} \frac{1}{c(t-1,\delta)}} \left((\widehat{\mu}_{t,i} - \widehat{\mu}_{t,j}) - (\mu_i - \mu_j) \right) \right) = \\
\prod_{c \in [d]} \mathbb{P}_{X \sim \mathcal{N}(0,1)} \left(X > \underbrace{\sqrt{\left(\frac{1}{N_{t,i}} + \frac{1}{N_{t,j}} \right)^{-1} \frac{1}{\sigma_c^2 c(t-1,\delta)}} \left((\widehat{\mu}_{t,i} - \widehat{\mu}_{t,j}) - (\mu_i - \mu_j) \right)(c)}_{Z_{i,j}(c)} \right),
\end{aligned}
$$

which we may rewrite with Mills ratio as $R(x) := \frac{\mathbb{P}(X > x)}{f_X(x)}$ so

$$
\mathbb{P}(X > x) = R(x) \exp(-\frac{1}{2}x^2) \frac{1}{\sqrt{2\pi}} = \widetilde{R}(x) \exp(-\frac{1}{2}x^2),
$$

and

$$(*) = \exp\left(-\frac{1}{2}\sum_{c\in[d]} Z_{i,j}(c)\right)\prod_{c\in[d]}\widetilde{R}(Z_{i,j}(c))$$

$$= \exp\left(-\left(\frac{1}{N_{t,i}}+\frac{1}{N_{t,j}}\right)^{-1}\frac{1}{c(t-1,\delta)}\sum_{c\in[d]}\frac{\left((\widehat{\mu}_{t,i}-\widehat{\mu}_{t,j})-(\mu_i-\mu_j)\right)(c)^2}{2\sigma_c^2}\right)\prod_{c\in[d]}\widetilde{R}(Z_{i,j}(c))$$

and by Cauchy-Swchartz inequality,

$$|\left((\widehat{\mu}_{t,i}-\widehat{\mu}_{t,j})-(\mu_i-\mu_j)\right)(c)| \leq \sqrt{\frac{1}{N_{t,i}}+\frac{1}{N_{t,j}}}\sqrt{N_{t,i}(\widehat{\mu}_{t,i}-\mu_i)(c)^2+N_{t,j}(\mu_j-\widehat{\mu}_{t,j})(c)^2}.$$

Therefore

$$(*) \geq \exp\left(-\frac{1}{c(t-1,\delta)}\sum_{k\in\{i,j\}}\frac{N_{t,k}}{2}\|\widehat{\mu}_{t,k}-\mu_k\|^2_{\Sigma^{-1}}\right)\prod_{c\in[d]}\widetilde{R}(Z_{i,j}(c)), \tag{11}$$

and using the following lemma,

Lemma 6. *The Mills ratio R is decreasing, log-convex and satisfies for all $(x_1,\dots,x_p)\in\mathbb{R}^p$,*

$$R\left(\frac{1}{p}\sum_{i=1}^p x_i\right)^p \leq \prod_{i=1}^p R(x_i).$$

we obtain

$$\begin{aligned}
\prod_{c\in[d]}\widetilde{R}(Z_{i,j}(c)) &\geq \widetilde{R}\left(\frac{1}{d}\sum_{c\in[d]}Z_{i,j}(c)\right)^d \\
&\overset{(a)}{\geq} \widetilde{R}\left(\frac{1}{d}\sum_{c\in[d]}\sqrt{\frac{1}{\sigma_c^2 c(t-1,\delta)}}\sqrt{N_{t,i}(\widehat{\mu}_{t,i}-\mu_i)(c)^2+N_{t,j}(\mu_j-\widehat{\mu}_{t,j})(c)^2}\right)^d \\
&\overset{(b)}{\geq} \widetilde{R}\left(\frac{\sqrt{2}}{\sqrt{c(t-1,\delta)d}}\sqrt{\sum_{k\in\{i,j\}}\frac{N_{t,k}}{2}\|\widehat{\mu}_{t,k}-\mu_k\|^2_{\Sigma^{-1}}}\right)^d,
\end{aligned}$$

where (a) follows since $\widetilde{R}$ is decreasing (Lemma 6) and (b) follows from this monotonicity and Cauchy-Schwartz. Combining the last display with (11), we prove that on the event $\mathcal{E}_{t,\delta/2}$, we have

$$(*) \geq \exp\left(-\frac{\beta(t-1,\delta/2)}{c(t-1,\delta)}\right)\widetilde{R}\left(\sqrt{\frac{2\beta(t-1,\delta/2)}{c(t-1,\delta)d}}\right)^d$$

Choosing

$$c(t,\delta) := \frac{\beta(t,\delta/2)}{\log(1/\delta)}$$

yields,

$$(*) \geq \delta\widetilde{R}\left(\sqrt{\frac{2\log(1/\delta)}{d}}\right)^d = r(\delta,d) \quad \text{with} \quad r(\delta,n) = \widetilde{R}\left(\sqrt{\frac{2\log(1/\delta)}{n}}\right)^n$$

and proceeding identically, we prove that under the event $\mathcal{E}_{t,\delta/2}$,

$$\mathbb{P}_{Y\sim\mathcal{N}(0_d,\Sigma)}\left(Y > \sqrt{\frac{N_{t,i}}{c(t-1,\delta)}}(\mu_i - \widehat{\mu}_{t,i})\right)\prod_{j\in\widehat{S}_t}\mathbb{P}_{X\sim\mathcal{N}(0,1)}\left(X > \sqrt{\frac{N_{t,j}}{c(t-1,\delta)}}\|\mu_j - \widehat{\mu}_{t,j}\|_{\Sigma^{-1}}\right)$$
$$\geq \delta\widetilde{R}\left(\sqrt{\frac{2\log(1/\delta)}{d+|\widehat{S}_t|}}\right)^{d+|\widehat{S}_t|}.$$

All put together, we proved that

$$\begin{aligned}\mathbb{1}_{\mathcal{E}_{t,\delta/2}}\mathbb{1}_{\widehat{S}_t\neq S^\star}\mathbb{P}_{\widetilde{\Pi}_t|\mathcal{F}_t}(\mathrm{Alt}(\widehat{S}_t)) &\geq \delta\max\left\{\mathbb{1}_{\boldsymbol{\mu}\in\mathrm{Alt}^-(\widehat{S}_t)}r(\delta,d),\mathbb{1}_{(\boldsymbol{\mu}\in\mathrm{Alt}^+(\widehat{S}_t))}r(\delta,d+|\widehat{S}_t|)\right\}\\ &\geq \delta\min(r(\delta,d),r(\delta,d+|\widehat{S}_t|))\end{aligned}$$

which follows since $\widehat{S}_t\neq S^\star \implies \boldsymbol{\mu}\in\mathrm{Alt}(\widehat{S}_t)=\mathrm{Alt}^+(\widehat{S}_t)\cup\mathrm{Alt}^-(\widehat{S}_t)$.

The conclusion is immediate by taking $M(t,\delta)=\left\lceil\frac{\log(2t^s\zeta(s)/\delta)}{\delta q(t,\delta)}\right\rceil$ where $q(t,\delta)=\min(r(\delta,d),r(\delta,d+|\widehat{S}_t|))$. Moreover, it is known (Birnbaum, 1942) that for $x\geq 0$,

$$R(x)\geq\frac{2}{x+\sqrt{x^2+4}}$$

and $R(x)\sim\frac{1}{x}, x\to+\infty$. The Mills ratio of the standard normal is implemented in common scientific libraries. To compute $M(t,\delta)$, we just need to compute at most the values $(r(\delta,d+k))_{k\in[K]\cup\{0\}}$ at the initialisation of the algorithm. Using the lower bound on the Mills ratio, one can easily check that

$$\frac{\log(M(t,\delta))c(t,\delta)}{\log(1/\delta)}\underset{\delta\to 0}{\leq}1.$$

Proof of Lemma 6. The monotonicity is well-known and simple to prove by noting that

$$\begin{aligned}R(x) &:= \exp\left(x^2/2\right)\int_x^\infty\exp\left(-t^2/2\right)dt\\ &= \exp\left(x^2/2\right)\int_0^\infty\exp\left(-(t+x)^2/2\right)dt\\ &= \int_0^\infty\exp\left(-tx-t^2/2\right)dt,\end{aligned}$$

from which we have $R(x+\alpha^2)=\int_0^\infty\exp(-t\alpha^2)\exp\left(-tx-t^2/2\right)dt<R(x)$. The log-convexity is proven in Theorem 2.5 of Baricz (2008). From the log-convexity and using Jensen inequality, we have

$$\begin{aligned}\log R\left(\frac{1}{p}\sum_{i=1}^p x_i\right) &\leq \frac{1}{p}\sum_{i=1}^p\log R(x_i)\\ &= \frac{1}{p}\log\left(\prod_{i=1}^p R(x_i)\right)\end{aligned}$$

which by monotonicity of the log proves the claimed statement. □

C.1.2 PSI with a Covariance Matrix

We specialise Lemma 5 to the case where the covariance Σ is non-diagonal. To derive this result, we will use the following lemma.

Lemma 7. *Let Σ a covariance matrix, V diagonal matrix such that $(V-\Sigma^{-1})$ is psd, and define $d_\Sigma := \|1_d\|^2_{(V^{-1/2}\Sigma^{-1}V^{-1/2})}$. Then, for all $x\in\mathbb{R}^d$, it holds that*

$$\mathbb{P}_{X\sim\mathcal{N}(0_d,\Sigma)}(X\geq x)\geq(2\pi)^{-d/2}\det(V\Sigma)^{-1/2}\exp\left(-\frac{1}{2}x^\intercal\Sigma^{-1}x\right)\prod_{c\in[d]}R(e_c^\intercal V^{-1/2}\Sigma^{-1}x),$$

and in particular, $\prod_{c\in[d]} R(e_c^\intercal V^{-1/2}\Sigma^{-1}x) \geq R\left(\|x\|_{\Sigma^{-1}}\sqrt{d_\Sigma}/d\right)^d$.

Applying this result with $V = \bar\sigma I_d$ where $\bar\sigma = |||\Sigma^{-1}|||$, we have

$$\mathbb{P}_{Y\sim\mathcal{N}(0_d,\Sigma)}\left(Y > \sqrt{\frac{N_{t,i}}{c(t-1,\delta)}}(\mu_i - \widehat\mu_{t,i})\right) \geq \frac{1}{\sqrt{(2\pi)^d\det(\bar\sigma\Sigma)}}\exp\left(-\frac{1}{2}\|\mu_i - \widehat\mu_{t,i}\|^2_{\Sigma^{-1}}\right)\prod_{c\in[d]} R(e_c^\intercal V^{-1/2}\Sigma^{-1}u_t),$$

with $u_t := \sqrt{\frac{N_{t,i}}{c(t-1,\delta)}}(\mu_i - \widehat\mu_{t,i})$, therefore, letting

$$(**) := \mathbb{P}_{Y\sim\mathcal{N}(0_d,\Sigma)}\left(Y > \sqrt{\frac{N_{t,i}}{c(t-1,\delta)}}(\mu_i - \widehat\mu_{t,i})\right)\prod_{j\in\widehat S_t}\mathbb{P}_{X\sim\mathcal{N}(0,1)}\left(X > \sqrt{\frac{N_{t,j}}{c(t-1,\delta)}}\|\mu_j - \widehat\mu_{t,j}\|_{\Sigma^{-1}}\right)$$

we have

$$(**) \geq \det(\bar\sigma\Sigma)^{-1/2}\exp\left(-\sum_{k\in\{i\}\cup\widehat S_t}\frac{1}{2}\|\mu_k - \widehat\mu_{t,k}\|^2_{\Sigma^{-1}}\right)\prod_{j\in\widehat S_t}\widetilde R\left(\sqrt{\frac{N_{t,j}\|\mu_j - \widehat\mu_{t,j}\|^2_{\Sigma^{-1}}}{c(t-1,\delta)}}\right)\prod_{c\in[d]}\widetilde R(e_c^\intercal V^{-1/2}\Sigma^{-1}u_t),$$

then note that by Lemma 6, we have

$$\prod_{j\in\widehat S_t}\widetilde R\left(\sqrt{\frac{N_{t,j}}{c(t-1,\delta)}}\|\mu_j - \widehat\mu_{t,j}\|_{\Sigma^{-1}}\right)\prod_{c\in[d]}\widetilde R(e_c^\intercal V^{-1/2}\Sigma^{-1}u_t/d) \geq \widetilde R\left(\frac{1_d^\intercal V^{-1/2}\Sigma^{-1}u_t + 1_{|\widehat S_t|}^\intercal h_t}{d+|\widehat S_t|}\right)^{d+|\widehat S_t|}$$

with $h_t := \left(\sqrt{\frac{N_{t,j}}{c(t-1,\delta)}}\|\mu_j - \widehat\mu_{t,j}\|_{\Sigma^{-1}}\right)_{j\in|\widehat S_t|}$ and by Cauchy-Swchartz inequality, replacing $V = \bar\sigma I_d$,

$$\begin{aligned}
1_d^\intercal V^{-1/2}\Sigma^{-1}u_t + 1_{|\widehat S_t|}^\intercal h_t &\leq \sqrt{\|1_d\|^2_{(\bar\sigma\Sigma)^{-1}} + \|1_{|\widehat S_t|}\|^2}\sqrt{\|u_t\|^2_{\Sigma^{-1}} + \|h_t\|^2} \\
&= \sqrt{d_\Sigma + |\widehat S_t|}\left(\sqrt{\frac{1}{c(t-1,\delta)}\sum_{k\in\{i\}\cup\widehat S_t} N_{k,t}\|\mu_k - \widehat\mu_{t,k}\|^2_{\Sigma^{-1}}}\right) \\
&\leq \sqrt{d_\Sigma + |\widehat S_t|}\sqrt{\frac{2\beta(t-1,\delta/2)}{c(t-1,\delta)}}
\end{aligned}$$

where the last inequality follows on the event $\mathcal{E}_{t,\delta/2}$. Comibining these displays with $c(t,\delta) := \frac{\beta(t,\delta/2)}{\log(1/\delta)}$, we have on the event $\mathcal{E}_{t,\delta/2}$,

$$\begin{aligned}
(**) &\geq \det(\bar\sigma\Sigma)^{-1/2}\delta\widetilde R\left(\frac{\sqrt{d_\Sigma + |\widehat S_t|}}{d+|\widehat S_t|}\sqrt{2\log(1/\delta)}\right) \\
&= \det(\bar\sigma\Sigma)^{-1/2}\delta r(\delta^{\frac{d_\Sigma + |\widehat S_t|}{d+|\widehat S_t|}}, d+|\widehat S_t|).
\end{aligned}$$

Following the same reasoning, it is simple to show that under the event $\mathcal{E}_{t,\delta/2}$, we have

$$\begin{aligned}
&\mathbb{P}_{Y\sim\mathcal{N}(0_d,\Sigma)}\left(Y \geq \sqrt{\left(\frac{1}{N_{t,i}} + \frac{1}{N_{t,j}}\right)^{-1}\frac{1}{c(t-1,\delta)}}\left((\widehat\mu_{t,i} - \widehat\mu_{t,j}) - (\mu_i - \mu_j)\right)\right) \\
&\geq \det(\bar\sigma\Sigma)^{-1/2}\delta\widetilde R(\sqrt{2\log(1/\delta)d_\Sigma}/d) = \det(\bar\sigma\Sigma)^{-1/2}\delta r(\delta^{\frac{d_\Sigma}{d}}, d).
\end{aligned}$$

All put together, we have

$$\begin{aligned}
\mathbb{1}_{\mathcal{E}_{t,\delta/2}}\mathbb{1}_{\widehat S_t\neq S^\star}\mathbb{P}_{\widetilde\Pi_t|\mathcal{F}_t}(\mathrm{Alt}(\widehat S_t)) &\geq \delta\det(\bar\sigma\Sigma)^{-1/2}\max\left\{\mathbb{1}_{\boldsymbol\mu\in\mathrm{Alt}^-(\widehat S_t)} r(\delta^{\frac{d_\Sigma}{d}}, d), \mathbb{1}_{(\boldsymbol\mu\in\mathrm{Alt}^+(\widehat S_t))} r(\delta^{\frac{d_\Sigma + |\widehat S_t|}{d+|\widehat S_t|}}, d+|\widehat S_t|)\right\} \\
&\geq \delta\det(\bar\sigma\Sigma)^{-1/2}\min(r(\delta^{\frac{d_\Sigma}{d}}, d), r(\delta^{\frac{d_\Sigma + |\widehat S_t|}{d+|\widehat S_t|}}, d+|\widehat S_t|))
\end{aligned}$$

which follows similarly to the Σ diagonal case, since $\widehat{S}_t \neq S^\star \implies \boldsymbol{\mu} \in \mathrm{Alt}(\widehat{S}_t) = \mathrm{Alt}^+(\widehat{S}_t) \cup \mathrm{Alt}^-(\widehat{S}_t)$. Remark that when $\Sigma = \sigma I_d$, we recover the results of section C.1.1.

Proof of Lemma 7. Σ is a $d \times d$ covariance matrix and f_Σ is the density function of $\mathcal{N}(0_d, \Sigma)$. We let $\mathcal{R}_\Sigma$ denote the Mills ratio of the distribution $\mathcal{N}(0_d, \Sigma)$, which is defined for a vector $x \in \mathbb{R}^d$ as

$$\mathcal{R}_\Sigma(x) := \frac{\mathbb{P}\left(\mathcal{N}(0_d, \Sigma) \geq x\right)}{f_\Sigma(x)} \tag{12}$$

where for two vectors x, y, the notation $x \leq y$ should be understood component-wise. Expanding (12) gives

$$\mathcal{R}_\Sigma(x) = \exp(\|x\|^2_{\Sigma^{-1}}/2) \int_{(x,\infty)} \exp\left(-\|u\|^2_{\Sigma^{-1}}/2\right) du,$$

with $(x, \infty) = (x(1), \infty) \times \cdots \times (x(d), \infty)$. By a simple translation, it follows that

$$\mathcal{R}_\Sigma(x) = \int_{(0_d,\infty)} \exp\left(-x^\intercal \Sigma^{-1} u - \|u\|^2_{\Sigma^{-1}}/2\right) du.$$

Since $(V - \Sigma^{-1})$ is psd, we have

$$\begin{aligned}
\mathcal{R}_\Sigma(x) &\geq \int_{(0_d,\infty)} \exp\left(-x^\intercal \Sigma^{-1} u - u^\intercal V u/2\right) du \\
&= \det(V)^{-1/2} \int_{(0_d,\infty)} \exp\left(-x^\intercal \Sigma^{-1} V^{-1/2} u - u^\intercal u/2\right) du
\end{aligned}$$

which follows by change of variable. Thus, letting $z = V^{-1/2}\Sigma^{-1}x$, we have

$$\begin{aligned}
\mathcal{R}_\Sigma(x) &\geq \det(V)^{-1/2} \int_{(0_d,\boldsymbol{\infty})} \exp\left(-u^\intercal z - u^\intercal u/2\right) du \\
&= \det(V)^{-1/2} \prod_{c \in [d]} R(z(c))
\end{aligned}$$

then using Lemma 6, it follows that

$$\begin{aligned}
\mathcal{R}_\Sigma(x) &\geq \det(V)^{-1/2} R\left(1_d^\intercal V^{-1/2} \Sigma^{-1} x/d\right)^d \\
&\geq \det(V)^{-1/2} R\left(\|1_d\|_{V^{-1/2}\Sigma^{-1}V^{-1/2}} \|x\|_{\Sigma^{-1}}/d\right)^d.
\end{aligned}$$

$\square$

C.2 Structured Setting: Transductive Linear BAI ($d = 1$ and $\Theta = \mathbb{R}^h$)

When $d = 1$, we have $S^\star = \{z^\star\}$ with $z^\star = z^\star(\boldsymbol{\theta})$ and $\widehat{S}_t = \{\hat{z}_t\}$ where $\hat{z}_t = z^\star(\hat{\boldsymbol{\theta}}_t)$ where $z^\star(\boldsymbol{\lambda}) := \mathrm{argmax}_{z \in \mathcal{Z}}\, \theta^\intercal z$. Let $\sigma^2 = \Sigma$ be the variance of the unique objective.

Let $\Pi_t = \mathcal{N}(0_h, \boldsymbol{\Sigma}_t)$ where $\boldsymbol{\Sigma}_t = \sigma^2(V_{\boldsymbol{N}_t} + \xi I_h)^{-1}$. For all $m \in [M(t-1, \delta)]$, let $\boldsymbol{v}_t^m \sim \Pi_t$ and $\boldsymbol{\theta}_t^m = \hat{\boldsymbol{\theta}}_t + \sqrt{c(t-1,\delta)}\boldsymbol{v}_t^m$. Therefore, using computation as above and $1 - x \leq e^{-x}$, we obtain

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\nu}}\left(\tau_\delta^{\mathrm{PS}} < +\infty, \hat{z}_{\tau_\delta^{\mathrm{PS}}} \neq z^\star\right) \leq \delta/2 + \\
&\quad \mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t \geq 1} \mathbb{1}\left(\mathcal{E}_{t,\delta/2} \cap \{\hat{z}_t \neq z^\star\}\right) \exp\left(-M(t-1,\delta)\mathbb{P}_{\boldsymbol{v}_t^m \sim \Pi_t}\left(\exists z \neq \hat{z}_t,\, (\hat{\boldsymbol{\theta}}_t + \sqrt{c(t-1,\delta)}\boldsymbol{v}_t^m)^\intercal (z - \hat{z}_t) > 0\right)\right)\right]
\end{aligned}$$

For all $m \in [M(t-1,\delta)]$, let $X_t^m = \frac{(\boldsymbol{v}_t^m)^\intercal (z^\star - \hat{z}_t)}{\|\hat{z}_t - z^\star\|_{\Sigma_t}}$. Then, we have $X_t^m \sim \mathcal{N}(0,1)$ since $\boldsymbol{v}_t^m \sim \Pi_t$. Under $\mathcal{E}_{t,\delta/2}$, we have $\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}\|_{\Sigma_t^{-1}} \le \sqrt{2\beta(t-1,\delta/2)}$. Using that $z^\star \neq \hat{z}_t$ and $\boldsymbol{\theta}^\intercal(z^\star - \hat{z}_t) > 0$, we obtain

$$\begin{aligned}
& \exists z \neq \hat{z}_t, \quad (\hat{\boldsymbol{\theta}}_t + \sqrt{c(t-1,\delta)}\boldsymbol{v}_t^m)^\intercal (z - \hat{z}_t) > 0 \\
\Longleftarrow \quad & \sqrt{c(t-1,\delta)}(\boldsymbol{v}_t^m)^\intercal (z^\star - \hat{z}_t) > -\boldsymbol{\theta}^\intercal (z^\star - \hat{z}_t) + (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta})^\intercal (\hat{z}_t - z^\star) \\
\Longleftarrow \quad & \sqrt{c(t-1,\delta)} \frac{(\boldsymbol{v}_t^m)^\intercal (z^\star - \hat{z}_t)}{\|\hat{z}_t - z^\star\|_{\Sigma_t}} \ge \|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}\|_{\Sigma_t^{-1}} \\
\Longleftarrow \quad & X_t^m \ge \sqrt{\frac{2\beta(t-1,\delta/2)}{c(t-1,\delta)}} .
\end{aligned}$$

In the following, we consider $c(t,\delta) = \frac{\beta(t,\frac{\delta}{2})}{\log\frac{1}{\delta}}$ and $M(t,\delta) = \left\lceil \frac{\log(2t^s\zeta(s)/\delta)}{\delta q(\delta)} \right\rceil$ where $q(\delta) = R(\sqrt{2\log(1/\delta)})/\sqrt{2\pi} = \mathbb{P}_{X\sim\mathcal{N}(0,1)}(X > \sqrt{2\log(1/\delta)})/\delta$. Therefore, we have shown that

$$\mathbb{P}_{\boldsymbol{\nu}}\left(\tau_\delta^{\text{PS}} < +\infty, \hat{z}_{\tau_\delta^{\text{PS}}} \neq z^\star\right) \le \delta/2 + \sum_{t\ge 1} \exp\left(- \left\lceil \frac{\log(2(t-1)^s\zeta(s)/\delta)}{\mathbb{P}_{X\sim\mathcal{N}(0,1)}(X > \sqrt{2\log(1/\delta)})} \right\rceil \mathbb{P}_{X\sim\mathcal{N}(0,1)}(X > \sqrt{2\log(1/\delta)}) \right) .$$

Using that $\lceil x \rceil \ge x$ and $\sum_{t\ge1} 1/t^s = \zeta(s)$ concludes the proof of δ-correctness, i.e. $\mathbb{P}_{\boldsymbol{\nu}}\left(\tau_\delta^{\text{PS}} < +\infty, \hat{z}_{\tau_\delta^{\text{PS}}} \neq z^\star\right) \le \delta$.

D SADDLE-POINT CONVERGENCE

In this section, we prove the following saddle-point convergence theorem.

Theorem 3. *There exists events $(\Xi_t)_{t\ge1}$ and $t_3 \in \mathbb{N}$ such that for all $t \ge t_3$,*

$$2GLR(t) \ge t \max_{\boldsymbol{w}\in\Delta_K} \inf_{\boldsymbol{\lambda}\in\Theta\cap\text{Alt}(S^\star)} \|vec(\boldsymbol{\theta}-\boldsymbol{\lambda})\|^2_{\widetilde{V}_{\boldsymbol{w}}} - 2f(t), \tag{13}$$

with $f(t) =_\infty o(t)$ and $\mathbb{P}_{\boldsymbol{\nu}}(\Xi_t) \ge 1 - 5/t^2$. In particular for $t \ge t_3$, $\widehat{S}_t = S^*(\boldsymbol{\theta})$.

To derive this result, we will first define some concentration events below :

- $\Xi_{1,t} := \left\{ \forall s \le t, \left\| vec(\widehat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}) \right\|^2_{\widetilde{V}_s} \le \beta(t, 1/t^2) =: f_1(t) \right\}$ where $\beta(t,\delta)$ is defined as in Lemma 3

We also define the following (high-probability) event:

- $\Xi_{2,t}$ as in Lemma 24.

D.1 Convergence of the Alternatives

We study the time after which the Pareto set is well estimated. The forced-exploration weight vector $\boldsymbol{w}_{\text{exp}}$ allows to define a deterministic time after which the Pareto set is well estimated on a good event. We prove the following result.

Lemma 8. *There exists $t_1 \in \mathbb{N}$ such that for all $t \ge t_1$, if the event $\Xi_{1,t} \cap \Xi_{2,t}$ holds then $\widehat{S}_t = S^\star(\boldsymbol{\theta})$.*

Proof. Let $e_1, \dots, e_d$ denote the canonical $\mathbb{R}^d$ basis and let $a \in \mathcal{A}$. And assume in this section that $\Xi_{1,t} \cap \Xi_{2,t}$ holds with $t \ge t_0(\alpha)$ (as in Lemma 24). For any $c \in [d]$ we have by Cauchy-Schwartz inequality,

$$\begin{aligned}
\left|(e_c \otimes a)^\intercal vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta})\right| \quad &\le \quad \left\|\Sigma^{1/2} \otimes V_t^{-1/2}(e_c \otimes a)\right\|_2 \left\|vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta})\right\|_{\Sigma^{-1}\otimes V_t} \\
&\le \quad \left\|\Sigma^{1/2} \otimes V_t^{-1/2}(e_c \otimes a)\right\|_2 \sqrt{f_1(t)},
\end{aligned}$$

on the event $\Xi_{1,t}$. From the inequality above, we have for all $a \in \mathcal{A}$,

$$\begin{aligned}
\|(I_d \otimes a)^\intercal vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta})\|_2^2 f_1(t)^{-1} &\leq \sum_{c\in[d]} (e_c \otimes a)^\intercal (\Sigma \otimes V_t^{-1})(e_c \otimes a) \\
&= \sum_{c\in[d]} \mathrm{Tr}\big((e_c \otimes a)^\intercal (\Sigma \otimes V_t^{-1})(e_c \otimes a)\big) \\
&= \mathrm{Tr}\left((\Sigma \otimes V_t^{-1}) \sum_{c\in[d]} (e_c \otimes a)(e_c \otimes a)^\intercal\right) \\
&= \mathrm{Tr}\left((\Sigma \otimes V_t^{-1}) \sum_{c\in[d]} (e_c e_c^T) \otimes (aa^T)\right) \\
&= \mathrm{Tr}\big((\Sigma \otimes V_t^{-1})(I_d \otimes (aa^T))\big) = \mathrm{Tr}\big(\Sigma \otimes V_t^{-1} aa^T\big) \\
&= \mathrm{Tr}(\Sigma) a^\intercal V_t^{-1} a,
\end{aligned}$$

which follows from $\mathrm{Tr}(A \otimes B) = \mathrm{Tr}(A)\mathrm{Tr}(B)$. Then Lemma 24 on forced exploration ensures that for any a, $\|a\|^2_{V_t^{-1}} \leq 2t^{\alpha-1}\|a\|^2_{V(\boldsymbol{w}_{\text{exp}})^{-1}}$, thus

$$\|(I_d \otimes a)^\intercal vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta})\|_2^2 f_1(t)^{-1} \leq 2t^{\alpha-1}\mathrm{Tr}(\Sigma)\|a\|^2_{V(\boldsymbol{w}_{\text{exp}})^{-1}}.$$

Now, letting μ_a denote the mean vector of arm a, remark that by writing the linear model we have $\mu_a = (I_d \otimes a)^\intercal vec(\boldsymbol{\theta})$, so that letting $\widehat{\mu}_{t,a} = (I_d \otimes a)^\intercal vec(\widehat{\boldsymbol{\theta}}_t)$ the estimated mean of arm a at time t, we have proved that

$$\|\mu_{t,a} - \mu_a\|_2 \leq \|a\|_{V(\boldsymbol{w}_{\text{exp}})^{-1}} \sqrt{2t^{\alpha-1}\mathrm{Tr}(\Sigma) f_1(t)}. \tag{14}$$

For any two arms a, a', let $\mathrm{M}(a,a') := \|(\mu_a - \mu_{a'})_+\|_2$ and $\mathrm{M}(a,a';t) := \|(\widehat{\mu}_{t,a} - \widehat{\mu}_{t,a'})_+\|_2$, its empirical counterpart, where $(w)_+ := \max(w, 0)$ for $w \in \mathbb{R}$ and it is applied component-wise for a vector. It is simple to prove that

$$\begin{aligned}
|\mathrm{M}(a,a') - \mathrm{M}(a,a';t)| &\leq \|\widehat{\mu}_{t,a} - \mu_a\|_2 + \|\widehat{\mu}_{t,a'} - \mu_{a'}\|_2 \\
&\leq 2\sqrt{L_* t^{\alpha-1}\mathrm{Tr}(\Sigma) f_1(t)}.
\end{aligned}$$

where we recall $L_* := 2\max_{a\in\mathcal{A}} \|a\|^2_{V(\boldsymbol{w}_{\text{exp}})^{-1}}$. Therefore, if t is such that

$$2\sqrt{L_* t^{\alpha-1}\mathrm{Tr}(\Sigma) f_1(t)} < \Delta_1 := \min_{a\in S^\star(\boldsymbol{\theta})} \min_{a'\in\mathcal{Z}\setminus\{a\}} \mathrm{M}(a,a')$$

then for all arms $a \in S^\star(\boldsymbol{\theta})$ and $a' \neq a \in \mathcal{Z}$, we will have $\mathrm{M}(a,a';t) > 0$, i.e. a is not dominated by a' : it is empirically optimal.

Similarly defining $\mathrm{m}(a,a') := \min_{c\in[d]}(\mu_a(c) - \mu_{a'}(c))$ and $\mathrm{m}(a,a';t) := \min_{c\in[d]}(\widehat{\mu}_{t,a}(c) - \widehat{\mu}_{t,a'}(c))$, one can also prove that

$$\begin{aligned}
|\mathrm{m}(a,a') - \mathrm{m}(a,a';t)| &\leq \|\widehat{\mu}_{t,a} - \mu_a\|_2 + \|\widehat{\mu}_{t,a'} - \mu_{a'}\|_2 \\
&\leq 2\sqrt{L_* t^{\alpha-1}\mathrm{Tr}(\Sigma) f_1(t)}
\end{aligned}$$

and observe that, if $\mathrm{m}(a,a';t) > 0$ then a is dominated by a', so that when

$$2\sqrt{L_* t^{\alpha-1}\mathrm{Tr}(\Sigma) f_1(t)} < \Delta_2 := \min_{a\in\mathcal{Z}\setminus S^\star(\boldsymbol{\theta})} \max_{a'\in S^\star(\boldsymbol{\theta})} \mathrm{m}(a,a'),$$

any arm in $\mathcal{Z}\setminus S^\star(\boldsymbol{\theta})$ will be empirically dominated. Therefore, we define a gap $\Delta_{\min}$ under which the Pareto set is well estimated as

$$\Delta_{\min} := \min(\Delta_1, \Delta_2),$$

and we define

$$t_1 := \inf\left\{n \geq t_0(\alpha) : \forall t \geq n, \frac{\sqrt{L_* t^{\alpha-1}\mathrm{Tr}(\Sigma) f_1(t)}}{\Delta_{\min}} < \frac{1}{2}\right\}, \tag{15}$$

which is well-defined as $f_1(t)$ is at most logarithmic and α is positive. Note that the gap $\Delta_{\min}$ is related to the gaps that appear in the finite-time analysis of PSI algorithms (Auer et al., 2016; Kone et al., 2023). $\square$

D.2 Learning on Unbounded Sets

In this section, we properly define Θ_t, η_t, and the results that allow this definition. Most game-based analyses will assume the parameter space is bounded (Degenne et al., 2019, 2020; Li et al., 2024). In the (transductive) linear setting, this assumption is also used to control the self-normalized deviations of the least-square estimate.

In the unstructured setting, we show that this assumption can be relaxed by learning the norm of $\boldsymbol{\mu}$ online. This relies on the following result, whose proof can be found in Appendix I.

Lemma 9. *Let $\boldsymbol{w} \in \mathbb{R}_+^K$. For any $\boldsymbol{\mu}' := (\mu'_1 \dots \mu'_K)^\intercal \in \mathbb{R}^{K\times d}$, the following statement holds in the unstructured setting*

$$\inf_{\boldsymbol{\lambda}\in \mathrm{Alt}(S^\star(\boldsymbol{\mu}'))} \|vec(\boldsymbol{\lambda}-\boldsymbol{\mu}')\|^2_{\Sigma^{-1}\otimes \mathrm{diag}(\boldsymbol{w})} = \inf_{\lambda\in \mathrm{Alt}(S^\star(\boldsymbol{\mu}'))\cap\{\boldsymbol{\lambda}|\ \max_i \|\mu'_i-\lambda_i\|_{\Sigma^{-1}}<\epsilon\}} \|vec(\boldsymbol{\lambda}-\boldsymbol{\mu}')\|^2_{\Sigma^{-1}\otimes \mathrm{diag}(\boldsymbol{w})}\,, \tag{16}$$

where $\epsilon := \max(2\max_{i\notin S}\max_{j\in S}\|\mu'_i-\mu'_j\|_{\Sigma^{-1}}, \max_{i,j\in S^2}\|\mu'_i-\mu'_j\|_{\Sigma^{-1}})$, and $S = S^\star(\boldsymbol{\mu}')$.

Intuitively, it shows that for any parameter, the best response lies inside a bounded region of the alternative space, and the bound depends on the true mean vectors. However, as this is unknown in practice, we then use the procedure described in Algorithm 2 to estimate the radius ϵ as stated in the previous lemma, and from this, we define Θ_t and η_t.

Algorithm 2: Estimate and Halve

Input: timestep $t+1, p_t, C_t$

```
// define ucb function
```

$h^t \leftarrow a \mapsto \sqrt{f_1(t)}\|a\|_{V_t^{-1}} + \|vec(\widehat{\boldsymbol{\mu}}_t)\|_{\Sigma^{-1}\otimes aa^\intercal}$;

```
// compute upper confidence bounds
```

$U_t \leftarrow \max\left\{\max\{h^t(a_i-a_j) \mid (i,j)\in \widehat{S}_t^2\}, 2\max\{h^t(a_i-a_j)\mid (i,j)\in(\widehat{S}_t^c, S_t)\}\right\}$;

$u_t \leftarrow \max_{a\in\mathcal{A}} h^t(a)$

Compute $p_{t+1} \leftarrow U_t \mathbb{1}_{(2U_t\leq p_t)} + p_t\mathbb{1}_{(2U_t>p_t)}$

Compute $C_{t+1} \leftarrow u_t \mathbb{1}_{(2u_t\leq C_t)} + C_t\mathbb{1}_{(2u_t>C_t)}$

Set $B_{t+1} \leftarrow C_t + p_t$

Set $\eta_{t+1} \leftarrow \frac{1}{8B_{t+1}^2}$

Set $\Theta_{t+1} = \{\boldsymbol{\lambda} := (\lambda_1,\dots,\lambda_K) \mid \max_{a\in\mathcal{A}}\|\lambda_a\|_{\Sigma^{-1}} < B_{t+1}\}$

return: $B_{t+1}, p_{t+1}, C_{t+1}, \eta_{t+1}, \Theta_{t+1}$

In the unstructured setting, the procedure initializes the parameters $B_1 = \infty, p_1, C_1 = \infty$. In particular, B_t will be updated whenever the confidence bound can be halved. At time $t+1$, Algorithm 2 is called to compute the parameters for the next round. It computes η_{t+1}, Θ_{t+1} along with the parameters $B_{t+1}, p_{t+1}, C_{t+1}$ that will be used in the next call.

We then show that on a good event, both B_t and η_t will stabilize. That is after a time $t_2\in\mathbb{N}$, we will have $B_t = B_{t_2}$ and $\Theta_t = \Theta_{t_2}$ for all $t\geq t_2$.

Lemma 10. *In the unstructured setting, there exists $t_2\in\mathbb{N}$ (as defined below) such that for all $t\geq t_2$, if the event $\Xi_{1,t}\cap\Xi_{2,t}$ holds then $\eta_t = \eta_{t_2}$ and $\Theta_t = \Theta_{t_2}$.*

Proof. We will first show that when $\Xi_{1,t}$ holds, U_t is an upper confidence bound on the quantity ϵ of Lemma 9. Let $t\geq t_1$ and assume $\Xi_{1,t}$ holds. By Lemma 8, we have $\widehat{S}_t = S^\star(\boldsymbol{\mu})$. Note that by keeping the notation of the linear setting, we have for all $i,j\in[K]$,

$$\begin{aligned}
\|\mu_i-\mu_j\|^2_{\Sigma^{-1}} &= \|I_d\otimes(a_i-a_j)vec(\boldsymbol{\mu})\|^2_{\Sigma^{-1}} \\
&= vec(\boldsymbol{\mu})^\intercal(I_d\otimes(a_i-a_j)^\intercal)\Sigma^{-1}(I_d\otimes(a_i-a_j))vec(\boldsymbol{\mu}) \\
&= vec(\boldsymbol{\mu})^\intercal\Sigma^{-1}\otimes(a_i-a_j)(a_i-a_j)^\intercal vec(\boldsymbol{\mu}),
\end{aligned}$$

which follows by properties of the Kronecker product. Let $\widetilde{a}_{i,j} = \Sigma^{-1/2}\otimes(a_i-a_j)$, so we have $\|\mu_i-\mu_j\|^2_{\Sigma^{-1}} = \|vec(\boldsymbol{\mu})\|^2_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal}$. Therefore, we should derive upper bound on $\|vec(\boldsymbol{\mu})\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal}$. We have by triangle inequality

$$\|vec(\boldsymbol{\mu})\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal} \leq \|vec(\widehat{\boldsymbol{\mu}}_t)\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal} + \|vec(\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}_t)\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal}, \tag{17}$$

then we rewrite the RHS. Note that

$$\begin{aligned} \|vec(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_t)\|^2_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal} &= vec(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_t)^\intercal \widetilde{V}_t^{1/2}\widetilde{V}_t^{-1/2}(\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal)vec(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_t) \\ &\leq \|vec(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_t)\|_{\widetilde{V}_t}\|vec(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_t)\|_{(\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal)\widetilde{V}_t^{-1}(\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal)}, \end{aligned} \tag{18}$$

which follows by Cauchy-Swchartz inequality, and we have

$$\begin{aligned} (\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal)\widetilde{V}_t^{-1}(\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal) &= (\Sigma^{-1}\otimes(a_{i,j}a_{i,j}^\intercal))(\Sigma\otimes V_t^{-1})(\Sigma^{-1}\otimes(a_{i,j}a_{i,j}^\intercal)) \\ &= (a_{i,j}^\intercal V_t^{-1}a_{i,j})\Sigma^{-1}\otimes(a_{i,j}a_{i,j}^\intercal) \\ &= \|a_{i,j}\|^2_{V_t^{-1}}\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal, \end{aligned}$$

so that combining with (18) yields

$$\|vec(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_t)\|^2_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal} \leq \|a_{i,j}\|_{V_t^{-1}}\|vec(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_t)\|_{\widetilde{V}_t}\|vec(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_t)\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal},$$

that is $\|vec(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_t)\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal} \leq \|a_{i,j}\|_{V_t^{-1}}\|vec(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_t)\|_{\widetilde{V}_t}$ and plugging back into (17), yields

$$\|vec(\boldsymbol{\mu})\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal} \leq h^t(a_i, a_j) := \|vec(\widehat{\boldsymbol{\mu}}_t)\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal} + \|a_{i,j}\|_{V_t^{-1}}\sqrt{f_1(t)},$$

on the event $\Xi_{1,t}$. Therefore, as $\widehat{S}_t = S^\star$, and by the line above, U_t is an upper bound on $\epsilon := \max(2\max_{i\notin S^\star}\max_{j\in S^\star}\|\mu_i - \mu_j\|_{\Sigma^{-1}}, \max_{i,j\in(S^\star)^2}\|\mu_i - \mu_j\|_{\Sigma^{-1}})$ and we have (due to the terms $\|vec(\widehat{\boldsymbol{\mu}}_t)\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal}$ in its definition),

$$U_t \geq \epsilon_t := \max\left(2\max_{i\notin\widehat{S}_t}\max_{j\in\widehat{S}_t}\|\widehat{\mu}_{t,i} - \widehat{\mu}_{t,j}\|_{\Sigma^{-1}}, \max_{i,j\in\widehat{S}_t^2}\|\widehat{\mu}_{t,i} - \widehat{\mu}_{t,j}\|_{\Sigma^{-1}}\right).$$

We now justify that after some rounds, B_t will not change anymore. For that, it suffices to show that both p_t and C_t will stabilize due to the estimate and halve procedure. To see this, note that

$$\begin{aligned} \|vec(\widehat{\boldsymbol{\mu}}_t)\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal} &\leq \|vec(\boldsymbol{\mu})\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal} + \|vec(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_t)\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal} \\ &\leq \|vec(\boldsymbol{\mu})\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal} + \|a_{i,j}\|_{V_t^{-1}}\|vec(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_t)\|_{\widetilde{V}_t} \quad \text{cf above} \\ &\leq \|vec(\boldsymbol{\mu})\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal} + \|a_{i,j}\|_{V_t^{-1}}\sqrt{f_1(t)}, \end{aligned}$$

on the event $\Xi_{1,t}$. Let $\mathcal{I} := \{(i,j)\in(S^\star)^2, i\neq j\}\cup\{(i,j)\in((S^\star)^c, S)\}$ and define t_2 as

$$\widetilde{t}_2 := \inf\left\{n : \forall t\geq n : \max_{(i,j)\in\mathcal{I}}\frac{\|a_{i,j}\|_{V_t^{-1}}\sqrt{f_1(t)}}{\|vec(\boldsymbol{\mu})\|_{\widetilde{a}_{i,j}\widetilde{a}_{i,j}^\intercal}} < \frac{1}{2}\right\}, \tag{19}$$

which is well defined as the numerator is of order $\log(t)/t^{1-\alpha}$ due to forced exploration (cf Lemma 24) and the denominator is positive for $(i,j)\in\mathcal{I}$; $\widetilde{t}_2$ is bounded by a deterministic time. From $\widetilde{t}_2$, we have $U_t \leq 2\epsilon$. For $t\geq\widetilde{t}_2$, when $\Xi_{1,t}$ holds, since $U_t\leq 2\epsilon$, if p_t is updated at some round $\hat{t}\geq\widetilde{t}_2$ then $p_{\hat{t}+1} = U_{\hat{t}} \leq 2\epsilon$ and because U_t is an upper bound on ϵ, it holds that $U_t > \epsilon$, at any time $t\geq t_1$. Thus $2U_t > p_t$ for any $t\geq\hat{t}$. That is, after time $\widetilde{t}_2$, p_t can only be changed once. We can proceed similarly for C_t to prove that there exists t_2 with finite expectation such that for $t\geq t_2$,

$$B_t = B_{t_2} \quad \eta_t = \eta_{t_2}, \quad \Theta_t = \Theta_{t_2}.$$

We recall that in the transductive linear setting $\Theta_t = \Theta$ (which is bounded) and

$$\eta_t := \eta = \frac{1}{8L_{\mathcal{A}}^2L_\Theta^2}. \tag{20}$$

□

The following result derives from Lemma 29 and the definition of Λ_t.

Corollary 1. *On the event $\Xi_{1,t}\cap\Xi_{2,t}$, for $t\geq t_2$, it holds that: in the unstructured setting or the transductive linear setting, for all $a\in\mathcal{A}$, $\boldsymbol{\lambda}\in\Lambda_t\mapsto\exp(-\eta_t\|vec(\boldsymbol{\theta}-\boldsymbol{\lambda})\|^2_{\Sigma^{-1}\otimes aa^\intercal})$ is concave.*

Proof. In the (transductive) linear setting, we have $\eta_t := \eta = \frac{1}{8L_{\mathcal{A}}^2 L_\Theta^2}$ then from the definition of $L_{\mathcal{A}}, L_\Theta$, the statement follows by Lemma 29. In the unstructured setting, we have by definition $\Lambda_t := \text{Alt}(\widehat{S}_t) \cap \Theta_t$ with $\Theta_t := \{\boldsymbol{\lambda} := (\lambda_1, \dots, \lambda_K) \mid \max_a \|\lambda_a\|_{\Sigma^{-1}} < B_t\}$ and $\eta_t := \frac{1}{8B_t^2}$. Note that in the unstructured setting we have

$$\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\Sigma^{-1} \otimes aa^T} = \|\mu_a - \lambda_a\|^2_{\Sigma^{-1}},$$

and, as $2B_t$ is an upper confidence bound on $\max_{a \in \mathcal{A}} \|\mu_a - \lambda_a\|_{\Sigma^{-1}}$, the conclusion follows again using Lemma 29.

□

D.3 Analysis of the Sampling Rule

Letting t_1, t_2 as defined in (15), (19) we let $t_3 = \max(t_1, t_2)$. When $t \geq t_3$ and $\Xi_{1,t} \cap \Xi_{2,t}$ holds, $\widehat{S}_t = S^\star$ (Lemma 8) and $\Lambda_t := \Theta_t \cap \text{Alt}(S_t) = \Theta_t \cap \text{Alt}(S^\star)$. As $\Theta_t \subset \Theta$ we have $\Lambda_t \subset \Theta \cap \text{Alt}(S^\star)$. In the sequel, assume $t > t_3$ and $\Xi_{1,t} \cap \Xi_{2,t}$ holds. As $t \geq t_3$, both η_t and Λ_t remain constant (Lemma 10); thus, in this section, we may simply refer to them as η and Λ respectively. In particular, Λ is bounded and we have

$$\Lambda := \Lambda_{t_3}, \eta := \eta_{t_3} \quad \text{and} \quad \max_{a \in \mathcal{A}, \boldsymbol{\lambda} \in \Lambda} \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{a}\widetilde{a}^\intercal} \leq B, \quad \text{with}$$

$$B := \begin{cases} 4\widetilde{L}_{\mathcal{A}}^2 L_\Theta^2 & \text{(transductive linear setting)} \\ (B_{t_3} + \max_{a \in \mathcal{A}} \|\mu_a\|_{\Sigma^{-1}})^2 & \text{(unstructured setting).} \end{cases} \tag{21}$$

Note that when $\Xi_{1,t} \cap \Xi_{2,t}$ holds and $t \geq t_3$, B_{t_3} is bounded by a constant that only depends on the problems' parameters.

Iterative saddle point convergence This section analyzes the regrets of the *min* and *max* learners. The proof of Theorem 3 will follow. In this regard, we relate the regret of each learner to the saddle-point value by:

$$\begin{aligned}
\inf_{\boldsymbol{\lambda} \in \Lambda_{t-1}} \left\|vec(\boldsymbol{\lambda} - \widehat{\boldsymbol{\theta}}_{t-1})\right\|^2_{\widetilde{V}_{t-1}} &\geq \sum_{s=t_3}^{t-1} \mathbb{E}_{\boldsymbol{\lambda} \sim \pi_{s-1}} \left[\left\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\right\|^2_{\widetilde{a}_s \widetilde{a}_s^\intercal}\right] - r_1(t) & (22) \\
&\geq \sum_{s=t_3}^{t-1} \mathbb{E}_{\boldsymbol{\lambda} \sim \pi_{s-1}, a \sim \widetilde{\boldsymbol{w}}_s} \left[\left\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\right\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] - r_1(t) - m(t) & (23) \\
&\geq \max_{\boldsymbol{w} \in \Delta_K} \sum_{s=t_3}^{t-1} \mathbb{E}_{\boldsymbol{\lambda} \sim \pi_{s-1}, a \sim \boldsymbol{w}} \left[\left\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\right\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] - r_1(t) - r_2(t) - m(t) & (24) \\
&\geq \max_{\boldsymbol{w} \in \Delta_K} \sum_{s=t_3}^{t-1} \mathbb{E}_{\boldsymbol{\lambda} \sim \pi_{s-1}, a \sim \boldsymbol{w}} \left[\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] - r_1(t) - r_2(t) - m(t) - z(t) & (25) \\
&\geq \max_{\boldsymbol{w} \in \Delta_K} \sum_{s=t_3}^{t-1} \inf_{p \in \mathcal{P}(\Lambda_s)} \mathbb{E}_{\boldsymbol{\lambda} \sim p, a \sim \boldsymbol{w}} \left[\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] - r_1(t) - r_2(t) - m(t) - z(t) \\
&\geq (t - t_3) \max_{\boldsymbol{w} \in \Delta_K} \inf_{\boldsymbol{\lambda} \in \Theta \cap \text{Alt}(S^\star)} \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{V}_{\boldsymbol{w}}} - r_1(t) - r_2(t) - m(t) - z(t),
\end{aligned}$$

where the last inequality follows since $\Lambda_s \subset \Theta \cap \text{Alt}(S^\star)$ for $s \geq t_3$ when $\Xi_{1,t} \cap \Xi_{2,t}$ holds. In the decomposition above, $r_1(t)$ (resp. $r_2(t)$) is the regret of the *min* (resp. *max*) learner, $m(t)$ is a martingale concentration term and $z(t)$ is an approximation error. Each of these terms will be controlled in the following subsections. In particular, we justify in the next paragraphs that they are sub-linear terms (i.e. $o_{t \to \infty}(t)$).

D.3.1 Bounding the approximation error : step (24) to (25)

We note that by convexity of the squared norm we have

$$\sum_{s=t_3}^{t-1} \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_s)\|^2_{\widetilde{a}\widetilde{a}^\intercal} \geq \sum_{s=t_3}^{t-1} \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda}_s)\|^2_{\widetilde{a}\widetilde{a}^\intercal} + 2 \sum_{s=t_3}^{t-1} vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})^\intercal (\widetilde{a}\widetilde{a}^\intercal) vec(\boldsymbol{\theta} - \boldsymbol{\lambda}_s)$$

then, let us define $y_s := (\widetilde{a}\widetilde{a}^\intercal)vec(\boldsymbol{\theta} - \boldsymbol{\lambda}_s)$ and observe that by Cauchy-Schwartz inequality, we have for $\boldsymbol{\lambda}_s \in \Lambda$,

$$\begin{aligned}
vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})^\intercal y_s &\leq \|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{V}_{s-1}} \|y_s\|_{\widetilde{V}_{s-1}^{-1}} \\
&= \|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{V}_{s-1}} \sqrt{vec(\boldsymbol{\theta} - \boldsymbol{\lambda}_s)^\intercal (\widetilde{a}\widetilde{a}^\intercal)\widetilde{V}_{s-1}^{-1}(\widetilde{a}\widetilde{a}^\intercal)vec(\boldsymbol{\theta} - \boldsymbol{\lambda}_s)} \\
&= \|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{V}_{s-1}} \|a\|_{V_{s-1}^{-1}} \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda}_s)\|_{\widetilde{a}\widetilde{a}^\intercal} \quad \text{(Kronecker product)} \\
&\leq \|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{V}_{s-1}} \|a\|_{V_{s-1}^{-1}} \sqrt{B}
\end{aligned}$$

where we recall that $\widetilde{a} := \Sigma^{-1/2} \otimes a$ and $\widetilde{V}_s := \Sigma^{-1} \otimes V_s$. From the previous inequality, it follows that for any $\boldsymbol{w} \in \Delta_K$ fixed, we have

$$\begin{aligned}
\sum_{s=t_3}^{t-1} \mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}, a\sim\boldsymbol{w}} \left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] &\geq \sum_{s=t_3}^{t-1} \mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}, a\sim\boldsymbol{w}} \left[\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] \\
&\quad - 2\sqrt{B}\mathbb{E}_{a\sim\boldsymbol{w}} \left[\sum_{s=t_3}^{t-1} \|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{V}_{s-1}} \|a\|_{V_{s-1}^{-1}}\right]. \quad (26)
\end{aligned}$$

We will bound the RHS of the equation and prove that it is sub-linear due to the tiny forced exploration. We define

$$z(t) := 2\sqrt{B} \max_{\boldsymbol{w}\in\Delta_K} \mathbb{E}_{a\sim\boldsymbol{w}} \left[\sum_{s=t_3}^{t-1} \|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{V}_{s-1}} \|a\|_{V_{s-1}^{-1}}\right] \quad (27)$$

which we claim to be sub-linear as from Lemma 24,

$$\begin{aligned}
\sum_{s=t_3}^{t-1} \|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{V}_{s-1}} \|a\|_{V_{s-1}^{-1}} &\leq \sqrt{t f_1(t)} \left(L_* \sum_{s=1}^{t} s^{\alpha-1}\right)^{1/2} \\
&\leq \sqrt{L_* t^{1+\alpha} f_1(t)/\alpha},
\end{aligned}$$

which is sub-linear for $\alpha \in (0, 1)$ so

$$z(t) \leq 2\sqrt{BL_* t^{1+\alpha} f_1(t)/\alpha} \quad (28)$$

is sub-linear. The next paragraph shows that $m(t)$ is a sub-linear term.

D.3.2 Martingale Concentration: step (22) to (23)

Let us define the s-indexed stochastic process (U_s) as

$$U_s := \mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}} \left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}\right] - \mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}, a\sim\widetilde{\boldsymbol{w}}_s} \left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right].$$

We recall that both $\boldsymbol{w}_s$ and $\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}} \left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right]$ (for fixed $\widetilde{a}$) are $\mathcal{F}_{s-1}$-measurable, and U_s is adapted to the filtration $(\mathcal{F}_s)_{s\geq 1}$. Moreover, simple calculation yields

$$\mathbb{E}[U_s \mid \mathcal{F}_{s-1}] = 0,$$

i.e $(U_s)_s$ is $\mathcal{F}$-martingale difference sequence and $|U_s| \leq f_2(s,t)$ on the event $\Xi_{1,t} \cap \Xi_{2,t}$ (bounded below). Let us define the event

$$\Xi_{3,t} := \left(\sum_{s=1}^{t} U_s \leq m(t) := \sqrt{2\log(t^2) \sum_{s=1}^{t} f_2(s,t)^2}\right) \quad (29)$$

and observe that by Azuma's inequality it holds that $\mathbb{P}(\Xi_{3,t}) \geq 1 - 1/t^2$. It remains to show that on $\Xi_{1,t} \cap \Xi_{2,t}$, $m(t)$ can be bounded by a sub-linear term. Indeed, we have

$$\begin{aligned}
\|vec(\widehat{\boldsymbol{\theta}}_s - \boldsymbol{\lambda})\|_{\widetilde{a}\widetilde{a}^\intercal} &\leq \|vec(\widehat{\boldsymbol{\theta}}_s - \boldsymbol{\theta})\|_{\widetilde{a}\widetilde{a}^\intercal} + \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|_{\widetilde{a}\widetilde{a}^\intercal} \\
&\leq \|vec(\widehat{\boldsymbol{\theta}}_s - \boldsymbol{\theta})\|_{\widetilde{V}_s} \|a\|_{V_s^{-1}} + \sqrt{B}
\end{aligned}$$

so that on $\Xi_{1,t} \cap \Xi_{2,t}$, $f_2(s,t) \leq f_1(t)\|a\|^2_{V_s^{-1}} + B + 2\sqrt{B}\|vec(\widehat{\boldsymbol{\theta}}_s - \boldsymbol{\theta})\|_{\widetilde{V}_s}\|a\|_{V_s^{-1}}$ then

$$f_2(s,t)^2 \leq \left(2L_* f_1(t)s^{\alpha-1} + B + 2\sqrt{L_* B f_1(t) s^{\alpha-1}}\right)^2$$

so that

$$\sum_{s=1}^{t} f_2(s,t)^2 \leq \mathcal{O}(f_1(t)^2 t^{2\alpha-1} + f_1(t)t^{\alpha} + t + \sqrt{f_1(t)}t^{(\alpha+1)/2} + f_1(t)\sqrt{f_1(t)}t^{(3\alpha-1)/2}),$$

which is at most linear, so that

$$m(t) = \mathcal{O}(\sqrt{f_1(t)^2 t \log(t^2)})\,, \tag{30}$$

which is sub-linear.

D.3.3 Regret of the *min* learner

We analyze the regret of the min player, which is akin to continuous exponential weights. The goal is to show that we can have a sub-linear regret with a constant learning rate for the min learner. In this section, we show the following guarantees for the min learner.

Lemma 11. *For any $t \geq t_3$, the following holds on Ξ_t:*

$$\sum_{s=t_3+1}^{t} \mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}\right] - \inf_{\boldsymbol{\lambda}\in\Lambda}\|vec(\boldsymbol{\lambda} - \widehat{\boldsymbol{\theta}}_t)\|^2_{\widetilde{V}_t} \leq \mathcal{O}\left(\sqrt{t\log^2(t)}\right).$$

Proof. At round s, the strategy $\boldsymbol{\lambda}_s$ of the *min* player is drawn from the distribution $\pi_{s-1} := \mathcal{N}(vec(\boldsymbol{\theta}_s), \eta_s^{-1/2}\boldsymbol{\Sigma}_s; vec(\Lambda_s))$ for which the density at λ is proportional to $\exp(-\frac{1}{2}\eta\|vec(\widehat{\boldsymbol{\theta}}_s) - \lambda\|^2_{\widetilde{V}_s})$, with a normalizing constant

$$W_s := \int_{vec(\Lambda_s)} \exp\left(-\frac{1}{2}\eta\|vec(\widehat{\boldsymbol{\theta}}_s) - \lambda\|^2_{\widetilde{V}_s}\right)d\lambda.$$

We have for $s \geq t_3$,

$$\begin{aligned}
\log\frac{W_s}{W_{s-1}} &= \log\frac{\int_{vec(\Lambda)}\exp\left(-\frac{\eta\|vec(\widehat{\boldsymbol{\theta}}_s)-\lambda\|^2_{\widetilde{V}_s}}{2}\right)d\lambda}{W_{s-1}} \\
&= \log\frac{\int_{vec(\Lambda)}\exp\left(-\frac{\eta\|vec(\widehat{\boldsymbol{\theta}}_s)-\lambda\|^2_{\widetilde{V}_s}}{2} + \frac{\eta\|vec(\widehat{\boldsymbol{\theta}}_{s-1})-\lambda\|^2_{\widetilde{V}_{s-1}}}{2} - \frac{\eta\|vec(\widehat{\boldsymbol{\theta}}_{s-1})-\lambda\|^2_{\widetilde{V}_{s-1}}}{2}\right)d\lambda}{W_{s-1}} \\
&= \log\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\exp\left(-\frac{\eta\|vec(\widehat{\boldsymbol{\theta}}_{s-1}-\boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}}{2} + \underbrace{\frac{\eta\|vec(\widehat{\boldsymbol{\theta}}_{s-1}-\boldsymbol{\lambda})\|^2_{\widetilde{V}_s}}{2} - \frac{\eta\|vec(\widehat{\boldsymbol{\theta}}_s-\boldsymbol{\lambda})\|^2_{\widetilde{V}_s}}{2}}_{\Gamma_s(\boldsymbol{\lambda})/2}\right)\right] \\
&\leq \frac{1}{2}\left(\log\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\exp\left(-\eta\|vec(\widehat{\boldsymbol{\theta}}_{s-1}-\boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}\right)\right] + \log\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\exp(\eta\Gamma_s(\boldsymbol{\lambda}))\right]\right) \quad \text{(Cauchy-Schwartz)}
\end{aligned}$$

At this step, the goal is to use the exp-concavity of the squared norm on bounded domains. However we will not use it directly with $\boldsymbol{\lambda} \mapsto \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}$, but using concentration, we relate the latter to $\boldsymbol{\lambda} \mapsto \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}$. For this purpose, observe that by convex inequality on the squared norm,

$$\begin{aligned}
-\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal} &\leq -\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal} + 2vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})^\intercal(\widetilde{a}_s\widetilde{a}_s^\intercal)vec(\boldsymbol{\theta} - \boldsymbol{\lambda}) \\
&\overset{(i)}{\leq} -\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal} + 2\|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{V}_{s-1}}\|a_s\|_{V_{s-1}^{-1}}\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|_{\widetilde{a}_s\widetilde{a}_s^\intercal} \\
&\overset{(ii)}{\leq} -\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal} + 2\|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{V}_{s-1}}\|a_s\|_{V_{s-1}^{-1}}\sqrt{B} \\
&= -\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal} + \alpha_s
\end{aligned}$$

where (i) follows from Cauchy-Schwartz, and (ii) follows from the definition of B and, we define

$$\alpha_s := 2\|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{V}_{s-1}} \|a_s\|_{V_{s-1}^{-1}} \sqrt{B}. \tag{31}$$

Thus, by taking the expectation on both sides, we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\exp\left(-\eta\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}\right)\right] &\leq \mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\exp\left(-\eta\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}\right)\right]\exp(\eta\alpha_s) \\
&\leq \exp\left(\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[-\eta)\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda}))\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}\right]\right)\exp(\eta\alpha_s)
\end{aligned}$$

where η ensures that the square loss ($\boldsymbol{\lambda} \mapsto \|vec(\boldsymbol{\lambda} - \boldsymbol{\theta})\|^2_{\widetilde{a}_t\widetilde{a}_t^\intercal}$) is η-exp-concave on Λ (Corollary 1). Therefore,

$$\log \mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\exp\left(-\eta\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}\right)\right] \leq -\eta\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}\right] + \eta\alpha_s. \tag{32}$$

Now we will return to the loss with $\widehat{\boldsymbol{\theta}}_{s-1}$ on the RHS of the above inequality. Similarly to the previous derivation, we have by convexity

$$\begin{aligned}
-\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal} &\leq -\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal} - 2vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{t-1})^\intercal(\widetilde{a}_s\widetilde{a}_s^\intercal)vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}) \\
&\leq -\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal} + 2\|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{V}_{s-1}}\|a_s\|_{V_{s-1}^{-1}}\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|_{\widetilde{a}_s\widetilde{a}_s^\intercal},
\end{aligned}$$

then observe that

$$\begin{aligned}
\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|_{\widetilde{a}_s\widetilde{a}_s^\intercal} &\leq \|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{a}_s\widetilde{a}_s^\intercal} + \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|_{\widetilde{a}_s\widetilde{a}_s^\intercal} \\
&\leq \|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{a}_s\widetilde{a}_s^\intercal} + \sqrt{B} \\
&\leq \|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{V}_{s-1}}\|a_s\|_{V_{s-1}^{-1}} + \sqrt{B}.
\end{aligned}$$

Putting these displays together yields

$$-\eta\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}\right] \leq -\eta\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}\right] + \eta(2\alpha_s + \alpha_s^2/(2B))$$

which combined with (32) yields

$$\log\frac{W_{s+1}}{W_s} \leq -\frac{\eta}{2}\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}\right] + \eta(2\alpha_s + \alpha_s^2/(2B)) + \frac{1}{2}\log\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\exp(\eta\Gamma_s(\boldsymbol{\lambda})\right]. \tag{33}$$

By telescoping, we have

$$\begin{aligned}
\log\frac{W_t}{W_{t_3}} &= \sum_{s=t_3+1}^{t}\log\frac{W_s}{W_{s-1}} && (34) \\
&\leq \sum_{s=t_3+1}^{t}\left(-\frac{\eta}{2}\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}\right] + \frac{1}{2}\log\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\exp(\eta\Gamma_s(\boldsymbol{\lambda})\right]\right) + \eta d_t && (35) \\
&= -\frac{\eta}{2}\sum_{s=t_3+1}^{t}\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}\right] + \frac{1}{2}\sum_{s=t_3+1}^{t}\log\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\exp(\eta\Gamma_s(\boldsymbol{\lambda})\right] + \eta d_t && (36)
\end{aligned}$$

where $d_t := \sum_{s=t_3+1}^{t}\eta(2\alpha_s + \alpha_s^2/(2B))$. We also define

$$\vartheta_t := \sum_{s=t_3+1}^{t}\log\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\exp(\eta\Gamma_s(\boldsymbol{\lambda})\right]. \tag{37}$$

On the other side, let $\gamma > 0, \widetilde{\boldsymbol{\lambda}}_t \in \operatorname{argmin}_{\boldsymbol{\lambda}\in\overline{\Lambda}}\|vec(\boldsymbol{\lambda} - \widehat{\boldsymbol{\theta}}_t)\|^2_{\widetilde{V}_t}$. Since Λ is a union of convex sets, there exists a convex set $\mathcal{C} \subset \overline{\Lambda}$ such that $\widetilde{\boldsymbol{\lambda}}_t \in \mathcal{C}$. Then letting $\mathcal{N}_\gamma := \{(1-\gamma)\widetilde{\boldsymbol{\lambda}}_t + \gamma\boldsymbol{\lambda}, \boldsymbol{\lambda} \in \mathcal{C}\} = (1-\gamma)\widetilde{\boldsymbol{\lambda}}_t + \gamma\mathcal{C}$, it follows

$$\begin{aligned}
\log \frac{W_t}{W_{t_3}} &\geq \log \frac{\int_{vec(\mathcal{C})} \exp\left(-\frac{\eta\|vec(\widehat{\boldsymbol{\theta}}_t)-\lambda\|^2_{\widetilde{V}_t}}{2}\right) d\lambda}{W_{t_3}} \\
&\geq \log \frac{\int_{vec(\mathcal{N}_\gamma)} \exp\left(-\frac{\eta\|vec(\widehat{\boldsymbol{\theta}}_t)-\lambda\|^2_{\widetilde{V}_t}}{2}\right) d\lambda}{W_{t_3}} \quad \text{(convexity of } \mathcal{C}) \\
&= \log \frac{\int_{\gamma vec(\mathcal{C})} \exp\left(-\frac{\eta\|vec(\widehat{\boldsymbol{\theta}}_t-(1-\gamma)\widetilde{\boldsymbol{\lambda}}_t)-\lambda\|^2_{\widetilde{V}_t}}{2}\right) d\lambda}{W_{t_3}} \\
&= \log \frac{\int_{vec(\mathcal{C})} \gamma^{dh} \exp\left(-\frac{\eta\|(1-\gamma)vec(\widehat{\boldsymbol{\theta}}_t-\widetilde{\boldsymbol{\lambda}}_t)+\gamma vec(\widehat{\boldsymbol{\theta}}_t)-\gamma\lambda\|^2_{\widetilde{V}_t}}{2}\right) d\lambda}{W_{t_3}} \\
&\geq \log \frac{\int_{vec(\mathcal{C})} \gamma^{dh} \exp\left(-\frac{\eta\left((1-\gamma)\|vec(\widehat{\boldsymbol{\theta}}_t-\widetilde{\boldsymbol{\lambda}}_t)\|^2_{\widetilde{V}_t}+\gamma\|\lambda-vec(\widehat{\boldsymbol{\theta}}_t)\|^2_{\widetilde{V}_t}\right)}{2}\right) d\lambda}{W_{t_3}} \quad \text{(convexity of squared norm)} \\
&= dh\log(\gamma) - \frac{\eta(1-\gamma)}{2}\|vec(\widehat{\boldsymbol{\theta}}_t-\widetilde{\boldsymbol{\lambda}}_t)\|^2_{\widetilde{V}_t} + \log \frac{\int_{vec(\mathcal{C})} \exp\left(-\frac{\eta\gamma\|\lambda-vec(\widehat{\boldsymbol{\theta}}_t)\|^2_{\widetilde{V}_t}}{2}\right) d\lambda}{W_{t_3}} \\
&\geq dh\log(\gamma) - \frac{\eta(1-\gamma)}{2}\|vec(\widehat{\boldsymbol{\theta}}_t-\widetilde{\boldsymbol{\lambda}}_t)\|^2_{\widetilde{V}_t} - \frac{\eta\gamma\mathbb{E}_{\boldsymbol{\lambda}\sim\mathcal{U}(\mathcal{C})}\left[\|vec(\boldsymbol{\lambda}-\widehat{\boldsymbol{\theta}}_t)\|^2_{\widetilde{V}_t}\right]}{2} + \log\frac{\mathrm{vol}(\mathcal{C}_*)}{W_{t_3}},
\end{aligned}$$

with $\Lambda = \bigcup_{i\in\mathcal{I}} \mathcal{C}_i$ where $(\mathcal{C}_i)_{i\in\mathcal{I}}$ are convex sets with non-empty interior and $\mathcal{C}_*$ is the set with minimum general volume. Combining the last inequality with (36) and letting $\gamma = 1/t$ yields

$$\sum_{s=t_3+1}^{t} \mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1}-\boldsymbol{\lambda})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal}\right] - \inf_{\boldsymbol{\lambda}\in\Lambda}\|vec(\boldsymbol{\lambda}-\widehat{\boldsymbol{\theta}}_t)\|^2_{\widetilde{V}_t} \leq \varrho_t,$$

where

$$\varrho_t = \frac{2}{\eta}\log(W_{t_3}) + 2d_t + \frac{\vartheta_t}{\eta} + \frac{2dh\log(t)}{\eta} - \frac{2}{\eta}\log\mathrm{vol}(\mathcal{C}_*) + \frac{\mathbb{E}_{\boldsymbol{\lambda}\sim\mathcal{U}(\mathcal{C})}[\|vec(\boldsymbol{\lambda}-\widehat{\boldsymbol{\theta}}_t)\|^2_{\widetilde{V}_t}]}{t}. \tag{38}$$

Then, note that $\mathrm{vol}(\mathcal{C}_*) > 0$, since $\mathcal{C}_*$ has non-empty interior ($\mathrm{Alt}(S^\star)$ is a union of convex sets with non-empty interior and Θ_t is a ball). Further remark that $W_{t_3} \leq \mathrm{vol}(\Lambda)$, which is finite by definition of Λ (cf (21)). Moreover, for $\boldsymbol{\lambda} \in \mathcal{C} \subset \Lambda$,

$$\begin{aligned}
\|vec(\boldsymbol{\lambda}-\widehat{\boldsymbol{\theta}}_t)\|_{\widetilde{V}_t} &\leq \|vec(\boldsymbol{\theta}-\boldsymbol{\lambda})\|_{\widetilde{V}_t} + \|vec(\widehat{\boldsymbol{\theta}}_t-\boldsymbol{\theta})\|_{\widetilde{V}_t} \\
&\leq \sqrt{tB} + \sqrt{f_1(t)},
\end{aligned}$$

which holds on the event $\Xi_{1,t}$. Summerizing the displays above, we have

$$\begin{aligned}
\varrho_t &\leq \frac{2}{\eta}\log\frac{\mathrm{vol}(\Lambda)}{\mathrm{vol}(\mathcal{C}_*)} + 2d_t + \frac{\vartheta_t}{\eta} + \frac{2dh\log(t)}{\eta} + (\sqrt{tB}+\sqrt{f_1(t)})^2/t \qquad (39) \\
&= \frac{\vartheta_t}{\eta} + 2d_t + \left(B + f_1(t)/t + \sqrt{Bf_1(t)/t} + \frac{2}{\eta}\log\frac{\mathrm{vol}(\Lambda)}{\mathrm{vol}(\mathcal{C}_*)} + \frac{2dh\log(t)}{\eta}\right), \qquad (40)
\end{aligned}$$

and recall that $f_1(t)$ is logarithmic. Bounding d_t and ϑ_t in Lemma 12 will conclude the proof by showing that $\varrho_t + d_t = \mathcal{O}(\sqrt{t\log^2 t})$. □

Lemma 12. *On the event $\Xi_{1,t} \cap \Xi_{4,t}$ (cf (44)) , it holds that*

$$d_t \leq 4hf_1(t)\log\left(\frac{Tr(V_0)+tL^2_{\mathcal{A}}}{d\det(V_0)^{1/d}}\right) + 2\sqrt{2hBtf_1(t)\log\left(\frac{Tr(V_0)+tL^2_{\mathcal{A}}}{d\det(V_0)^{1/d}}\right)} = \mathcal{O}\left(\sqrt{t\log^2(t)}\right), \textit{and}$$

$$\vartheta_t = \mathcal{O}\left(\sqrt{t\log^2(t)}\right).$$

Proof. The first statement follows by elliptic potential (Lemma 23). We have $d_t := \sum_{s=t_3+1}^t (\alpha_s + \alpha_s^2/(2B))$ with $\alpha_s := 2\|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})\|_{\widetilde{V}_{s-1}} \|a_s\|_{V_{s-1}^{-1}}\sqrt{B}$. We have on the event $\Xi_{1,t}$

$$\begin{aligned} d_t &\leq 2f_1(t)\sum_{s=t_3+1}^t \|a_s\|^2_{V_{s-1}^{-1}} + 2\sqrt{Btf_1(t)}\left(\sum_{s=t_3+1}^t \|a_s\|^2_{V_{s-1}^{-1}}\right)^{1/2} \\ &\leq 4hf_1(t)\log\left(\frac{\mathrm{Tr}(V_0)+tL_{\mathcal{A}}^2}{d\det(V_0)^{1/d}}\right) + 2\sqrt{2hBtf_1(t)\log\left(\frac{\mathrm{Tr}(V_0)+tL_{\mathcal{A}}^2}{d\det(V_0)^{1/d}}\right)} \\ &= \mathcal{O}\left(\sqrt{t\log^2(t)}\right). \end{aligned}$$

We now prove the second part of the statement. We recall that

$$\vartheta_t := \sum_{s=t_3+1}^t \log \mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\exp(\eta\Gamma_s(\boldsymbol{\lambda})\right].$$

Letting $\boldsymbol{\lambda}_s \sim \pi_s$, and using Fubini's lemma, ϑ_t rewrites as

$$\vartheta_t = \log\left(\mathbb{E}_{\boldsymbol{\lambda}_{t_3+1},\dots,\boldsymbol{\lambda}_t}\left[\exp\left(\sum_{s=t_3+1}^t \Gamma_s(\boldsymbol{\lambda}_{s-1})\right)\right]\right)$$

We will first bound $\sum_{s=t_3+1}^t \Gamma_s(\boldsymbol{\lambda}_{s-1})$ and derive a bound on ϑ_t. We recall that for any $\boldsymbol{\lambda}$,

$$\begin{aligned} \Gamma_s(\boldsymbol{\lambda}) &= \left\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\right\|^2_{\widetilde{V}_s} - \left\|vec(\widehat{\boldsymbol{\theta}}_s - \boldsymbol{\lambda})\right\|^2_{\widetilde{V}_s} \\ &\overset{(i)}{\leq} 2vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})^\intercal \widetilde{V}_s vec(\widehat{\boldsymbol{\theta}}_{s-1} - \widehat{\boldsymbol{\theta}}_s) \\ &\overset{(ii)}{=} -2(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})^\intercal[\Sigma^{-1}\otimes a_s](\varepsilon_s + (I_d\otimes a_s^\intercal)vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1})) \\ &= 2vec(\boldsymbol{\lambda} - \widehat{\boldsymbol{\theta}}_{s-1})^\intercal(\widetilde{a}_s\widetilde{a}_s^\intercal)vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1}) + 2vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})^\intercal[\Sigma^{-1}\otimes a_s]\varepsilon_s \end{aligned}$$

where (i) follows by convexity of the squared norm and (ii) is due to Lemma 30. We recall that ε_s is the centered (sub)Gaussian noise in the observations at time s, $\widetilde{a}_s = \Sigma^{-1/2}\otimes a_s$ with a_s the covariate of the arm pulled at time s. In the sequel, for any $\boldsymbol{\lambda}$, and at any time s, we define

$$\psi_s(\boldsymbol{\lambda}) := vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})^\intercal[\Sigma^{-1}\otimes a_s]\varepsilon_s \quad \text{and} \quad \phi_s(\boldsymbol{\lambda}) := vec(\boldsymbol{\lambda} - \widehat{\boldsymbol{\theta}}_{s-1})^\intercal(\widetilde{a}_s\widetilde{a}_s^\intercal)vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1}), \tag{41}$$

then, it follows that

$$\frac{1}{2}\sum_{s=t_3+1}^t \Gamma_s(\boldsymbol{\lambda}_{s-1}) \leq \sum_{s=t_3+1}^t \phi_s(\boldsymbol{\lambda}_{s-1}) + \sum_{s=t_3+1}^t \psi_s(\boldsymbol{\lambda}_{s-1}).$$

We further define

$$\Psi_t := \sum_{s=t_3+1}^t \psi_s(\boldsymbol{\lambda}_{s-1}) \quad \text{and} \quad \Phi_t := \sum_{s=t_3+1}^t \phi_s(\boldsymbol{\lambda}_{s-1}). \tag{42}$$

In the next step, we will bound each of these terms separately. In Lemma 13, we bound Φ_t and Lemma 14 uses martingale concentration to bound Ψ_t. ∎

Lemma 13. *The following statement holds on* $\Xi_{1,t}$

$$\Phi_t \leq \sqrt{2hBtf_1(t)\log\left(\frac{Tr(V_0)+tL_{\mathcal{A}}^2}{d\det(V_0)^{1/d}}\right)} + 2hf_1(t)\log\left(\frac{Tr(V_0)+tL_{\mathcal{A}}^2}{d\det(V_0)^{1/d}}\right) = \mathcal{O}\left(\sqrt{t\log^2(t)}\right).$$

Proof. First observe that for any $\boldsymbol{\lambda}$ and any time s

$$\phi_s(\boldsymbol{\lambda}) \leq \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|_{\widetilde{a}_s\widetilde{a}_s^\intercal} \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|_{\widetilde{V}_{s-1}} \|a_s\|_{V_{s-1}^{-1}} + \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|^2_{\widetilde{V}_{s-1}} \|a_s\|^2_{V_{s-1}^{-1}}, \tag{43}$$

which follows from

$$\begin{aligned}
\phi_s(\boldsymbol{\lambda}) &= vec(\boldsymbol{\lambda} - \widehat{\boldsymbol{\theta}}_{s-1})^\intercal (\widetilde{a}_s\widetilde{a}_s^\intercal) vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{s-1}) \\
&= \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|_{\widetilde{a}_s\widetilde{a}_s^\intercal} \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|_{\widetilde{a}_s\widetilde{a}_s^\intercal} \\
&\leq \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|_{\widetilde{a}_s\widetilde{a}_s^\intercal} \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|_{\widetilde{a}_s\widetilde{a}_s^\intercal} + \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|^2_{\widetilde{a}_s\widetilde{a}_s^\intercal},
\end{aligned}$$

and note that

$$\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|_{\widetilde{a}_s\widetilde{a}_s^\intercal} \leq \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|_{\widetilde{V}_{s-1}} \|a_s\|_{V_{s-1}^{-1}}.$$

Further, we have

$$\begin{aligned}
\mathit{\Phi}_t &:= \sum_{s=t_3+1}^{t} \phi_s(\boldsymbol{\lambda}_{s-1}) \\
&\leq \sum_{s=t_3+1}^{t} \left(\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda}_{s-1})\|_{\widetilde{a}_s\widetilde{a}_s^\intercal} \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|_{\widetilde{V}_{s-1}} \|a_s\|_{V_{s-1}^{-1}} + \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|^2_{\widetilde{V}_{s-1}} \|a_s\|^2_{V_{s-1}^{-1}} \right) \\
&\leq \sqrt{B \sum_{s=t_3+1}^{t} \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|^2_{\widetilde{V}_{s-1}}} \sqrt{\sum_{s=t_3+1}^{t} \|a_s\|^2_{V_{s-1}^{-1}}} + f_1(t) \sum_{s=t_3+1}^{t} \|a_s\|^2_{V_{s-1}^{-1}}
\end{aligned}$$

where the first term is due to Cauchy-Schwartz inequality, and the second term uses concentration on the event $\Xi_{1,t}$. Further applying elliptic potential (Lemma 23), we have

$$\mathit{\Phi}_t \leq \sqrt{2hBtf_1(t) \log\left(\frac{\mathrm{Tr}(V_0) + tL_{\mathcal{A}}^2}{d \det(V_0)^{1/d}}\right)} + 2hf_1(t) \log\left(\frac{\mathrm{Tr}(V_0) + tL_{\mathcal{A}}^2}{d \det(V_0)^{1/d}}\right).$$

□

Introducing the event

$$\Xi_{4,t} := \left(\sum_{s=t_3+1}^{t} \psi_s \leq f_4(t) := \sqrt{2 \sum_{s=t_3+1}^{t} v_s \log(t^2)} \right), \tag{44}$$

we prove the following result.

Lemma 14. *The following statement holds on the event $\Xi_{1,t} \cap \Xi_{4,t}$*

$$\mathit{\Psi}_t \leq \sqrt{2\log(t^2) \sum_{s=t_3+1}^{t} v_s},$$

with

$$\sum_{s=t_3+1}^{t} v_s \leq tB + 2hf_1(t) \log\left(\frac{Tr(V_0) + tL_{\mathcal{A}}^2}{d \det(V_0)^{1/d}}\right) + 2\sqrt{2hBtf_1(t) \log\left(\frac{Tr(V_0) + tL_{\mathcal{A}}^2}{d \det(V_0)^{1/d}}\right)},$$

so that $\mathit{\Psi}_t \leq \mathcal{O}\left(\sqrt{t \log(t)}\right)$.

Proof. To bound $\mathit{\Psi}_t$, we use martingale arguments by defining a slightly richer filtration. In the sequel, we denote by $\widetilde{\mathcal{F}}_s := \sigma(\mathcal{F}_s, \boldsymbol{\lambda}_s)$ the information accumulated up to time s including $\boldsymbol{\lambda}_s$, i.e $\boldsymbol{\lambda}_s$ is $\widetilde{\mathcal{F}}_s$-measurable. We simplify notation and let $\psi_s := \psi_s(\boldsymbol{\lambda}_{s-1}) = vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})^\intercal [\Sigma^{-1} \otimes a_s]\varepsilon_s$, where we recall that $\boldsymbol{\lambda}_s$ is the strategy of the *min* player at time s. Then observe that ψ_s is $\widetilde{\mathcal{F}}_s$-measurable and we have

$$\begin{aligned}
\mathbb{E}[\psi_s \mid \widetilde{\mathcal{F}}_{s-1}] &= vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})^\intercal \mathbb{E}[[\Sigma^{-1} \otimes a_s]\varepsilon_s \mid \widetilde{\mathcal{F}}_{s-1}] \\
&= 0
\end{aligned}$$

that is $(\psi_s)_s$ is a $\widetilde{\mathcal{F}}$-martingale difference sequence with conditionally (sub)Gaussian increment of variance (proxy)

$$v_s := \|[\Sigma^{-1} \otimes a_s^\intercal] vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\|_\Sigma^2.$$

Using Azuma's inequality, we have that $\mathbb{P}(\Xi_{4,t}) \geq 1 - 1/t^2$. We argue that $f_4(t)$ is a sub-linear term on $\Xi_{1,t}$. Indeed, observe that

$$\begin{aligned}
v_s &= vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})^\intercal [\Sigma^{-1} \otimes a_s] \Sigma [\Sigma^{-1} \otimes a_s^\intercal] vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1}) \\
&= (1/\|a_s\|^2) vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})^\intercal [\Sigma^{-1} \otimes a_s][\Sigma \otimes a_s^\intercal][I_d \otimes a_s][\Sigma^{-1} \otimes a_s^\intercal] vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1}) \\
&= (1/\|a_s\|^2) vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})^\intercal [I_d \otimes a_s a_s^\intercal][\Sigma^{-1} \otimes a_s a_s^\intercal] vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1}) \\
&= vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})^\intercal [\Sigma^{-1} \otimes a_s a_s^\intercal] vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1}) \\
&= \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\|^2_{\widetilde{a}_s \widetilde{a}_s^\intercal}.
\end{aligned}$$

We will now bound the right-hand side as :

$$\begin{aligned}
\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\|_{\widetilde{a}_s \widetilde{a}_s^\intercal} &\leq \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|_{\widetilde{a}_s \widetilde{a}_s^\intercal} + \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda}_{s-1})\|_{\widetilde{a}_s \widetilde{a}_s^\intercal} \\
&\leq \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|_{\widetilde{V}_{s-1}} \|a_s\|_{V_{s-1}^{-1}} + \sqrt{B},
\end{aligned}$$

thus

$$v_s \leq B + \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|^2_{\widetilde{V}_{s-1}} \|a_s\|^2_{V_{s-1}^{-1}} + 2\sqrt{B} \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|_{\widetilde{V}_{s-1}} \|a_s\|_{V_{s-1}^{-1}},$$

then, using Cauchy-Schwartz and the elliptic potential lemma, it follows that on the event $\Xi_{1,t}$,

$$\begin{aligned}
\sum_{s=t_3+1}^{t} v_s &\leq tB + f_1(t) \sum_{s=t_3+1}^{t} \|a_s\|^2_{V_{s-1}^{-1}} + 2\sqrt{B} \left(\sum_{s=t_3+1}^{t} \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|^2_{\widetilde{V}_{s-1}} \right)^{1/2} \left(\sum_{s=t_3+1}^{t} \|a_s\|^2_{V_{s-1}^{-1}} \right)^{1/2}, \\
&\leq tB + 2hf_1(t) \log\left(\frac{\mathrm{Tr}(V_0) + tL_{\mathcal{A}}^2}{d \det(V_0)^{1/d}} \right) + 2\sqrt{2hBtf_1(t) \log\left(\frac{\mathrm{Tr}(V_0) + tL_{\mathcal{A}}^2}{d \det(V_0)^{1/d}} \right)},
\end{aligned}$$

where the last inequality follows from the elliptic potential lemma, which achieves the proof. □

Combining Lemma 12 and Equation (40), we have proved that on $\Xi_{1,t} \cap \Xi_{2,t} \cap \Xi_{3,t} \cap \Xi_{4,t}$ for $t > t_3$, we have $r_1(t) = \mathcal{O}\left(\sqrt{t \log^2(t)}\right)$.

and this achieves the proof of the *min* player's regret, as we have bounded $r_1(t)$.

D.3.4 Regret of the *max*-learner

We study the *max*-leaner's regret, aiming to prove that $r_2(t)$ is a sub-linear term. Formally, introducing the event (with $Z_s(w)$ as in (46) and w_t^s defined in (47))

$$\Xi_{5,t} := \left(\sup_{\boldsymbol{w} \in \mathcal{V}(\varepsilon)} \sum_{s=t_3+1}^{t} Z_s(\boldsymbol{w}) \leq f_5(t) \right) \quad \text{with} \quad f_5(t) := \sqrt{2 \left(\sum_{s=t_3+1}^{t} (w_t^s(a))^2 \right) \log(|\mathcal{V}(\varepsilon)| t^2)}, \tag{45}$$

in this section, we prove that

Lemma 15. *On the event $\Xi_{1,t} \cap \Xi_{5,t}$, it holds that*

$$\sup_{\boldsymbol{w} \in \Delta_K} \sum_{s=t_3+1}^{t-1} (\widetilde{g}_s(\boldsymbol{w}) - \widetilde{g}_s(\boldsymbol{w}_s)) \leq \mathcal{O}\left(\sqrt{t \log^2(t)} + \log(t) t^{1-\alpha} \right),$$

with $\widetilde{g}_s(\boldsymbol{w}) := \mathbb{E}_{\boldsymbol{\lambda} \sim \pi_{s-1}, a \sim \boldsymbol{w}} \left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|^2_{\widetilde{a} \widetilde{a}^\intercal} \right]$.

Proof. We recall that $\boldsymbol{w} \in \Delta_K$ is considered a distribution over $\mathcal{A}$ and for any $a \in \mathcal{A}$, $\widetilde{a} := \Sigma^{-1/2} \otimes a$. Let us define for any $\boldsymbol{w} \in \Delta_K$,

$$Z_s(\boldsymbol{w}) := \mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}, a\sim\boldsymbol{w}}\left[\left\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\right\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] - \mathbb{E}_{a\sim\boldsymbol{w}}\left[\left\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\right\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] - \mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}, a\sim\widetilde{\boldsymbol{w}}_s}\left[\left\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\right\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] + \mathbb{E}_{a\sim\widetilde{\boldsymbol{w}}_s}\left[\left\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\right\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right]. \tag{46}$$

Then observe that $(Z_s(\boldsymbol{w}))_s$ is a $\widetilde{\mathcal{F}}$-adapted process and and it is a $\widetilde{\mathcal{F}}$-martingale difference sequence for any fixed $\boldsymbol{w}$ since by direct algebra, $\mathbb{E}[Z_s(\boldsymbol{w}) \mid \widetilde{\mathcal{F}}_{s-1}] = 0$. Further note that, for $\boldsymbol{\lambda} \in \Lambda_s$,

$$\begin{aligned} \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda})\|_{\widetilde{a}\widetilde{a}^\intercal} &\leq \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|_{\widetilde{a}\widetilde{a}^\intercal} + \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|_{\widetilde{a}\widetilde{a}^\intercal}, \\ &\leq \|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta})\|_{\widetilde{V}_{s-1}} \|a\|_{V_{s-1}^{-1}} + \sqrt{B}. \end{aligned}$$

Therefore, on the event $\Xi_{1,t}$, we have $|Z_s(\boldsymbol{w})| \leq 2\max_{a\in\mathcal{A}}(w_t^s(a))^2$, where

$$w_t^s(a) := 2\sqrt{f_1(t)}\|a\|_{V_{s-1}^{-1}} + \sqrt{B}. \tag{47}$$

Let $\varepsilon > 0$ and $\mathcal{V}(\varepsilon)$ be an ε-cover of Δ_K, $\mathcal{N}(\varepsilon) = |\mathcal{V}(\varepsilon)|$. Note that by Azuma-Hoeffding's inequality, it is simple to see that $\mathbb{P}(\Xi_{5,t}) \geq 1 - 1/t^2$. Direct development yields

$$\begin{aligned} (*) &:= \sum_{s=t_3+1}^{t-1} (\widetilde{g}_s(\boldsymbol{w}) - \widetilde{g}_s(\boldsymbol{w}_s)) \\ &= \sum_{s=t_3+1}^{t-1} Z_s(\boldsymbol{w}) + \sum_{s=t_3+1}^{t-1} \mathbb{E}_{a\sim\boldsymbol{w}}\left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] - \mathbb{E}_{a\sim\widetilde{\boldsymbol{w}}_s}\left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] \\ &= \sum_{s=t_3+1}^{t-1} \mathbb{E}_{a\sim\boldsymbol{w}}\left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] - \mathbb{E}_{a\sim\boldsymbol{w}_s}\left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] + \sum_{s=t_3+1}^{t-1} Z_s(\boldsymbol{w}) + \sum_{s=t_3+1}^{t-1} \varphi_s, \end{aligned}$$

where

$$\begin{aligned} \varphi_s &= \mathbb{E}_{a\sim\boldsymbol{w}_s}\left[\left\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\right\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] - \mathbb{E}_{a\sim\widetilde{\boldsymbol{w}}_s}\left[\left\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\right\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] \\ &= \gamma_s\left(\mathbb{E}_{a\sim\boldsymbol{w}_s}\left[\left\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\right\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] - \mathbb{E}_{a\sim\boldsymbol{w}_{\text{exp}}}\left[\left\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\right\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right]\right), \\ &\leq \gamma_s \max_{a\in\mathcal{A}}(w_t^s(a))^2 \end{aligned}$$

where the last inequality holds on the event $\Xi_{1,t}$. Therefore, on the event $\Xi_{5,t} \cap \Xi_{1,t}$, for all $\boldsymbol{w} \in \mathcal{V}(\varepsilon)$, it holds that

$$\begin{aligned} (*) \leq h(t) + f_5(t) + \sum_{s=2}^{t-1} \gamma_s \max_{a\in\mathcal{A}}(w_t^s(a))^2 + \sum_{s=2}^{t-1} \mathbb{E}_{a\sim\boldsymbol{w}}\left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] \\ - \mathbb{E}_{a\sim\boldsymbol{w}_s}\left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right], \end{aligned} \tag{48}$$

where the rightmost term of Equation (48) is related to the regret of AdaHedge and

$$\begin{aligned} h(t) &:= \sum_{s=2}^{t_3} \mathbb{E}_{a\sim\boldsymbol{w}_s}\left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right] - \mathbb{E}_{a\sim\boldsymbol{w}}\left[\|vec(\widehat{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\lambda}_{s-1})\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right], \\ &\leq \mathcal{O}(f_1(t)), \end{aligned} \tag{49}$$

which follows by expanding the sum, and since t_3 fixed.

Lemma 16 (De Rooij et al. (2014)). *AdaHedge run with gains $g_s(\boldsymbol{w}) := \mathbb{E}_{a\sim\boldsymbol{w}}[\|vec(\widehat{\boldsymbol{\theta}}_s - \boldsymbol{\lambda}_s)\|^2_{\widetilde{a}\widetilde{a}^\intercal}]$ satisfies the following regret bound*

$$\max_{\boldsymbol{w}\in\Delta_K}\sum_{s=2}^{t-1}\mathbb{E}_{x\sim\boldsymbol{w}}\left[\left\|vec(\widehat{\boldsymbol{\theta}}_{s-1}-\boldsymbol{\lambda}_{s-1})\right\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right]-\mathbb{E}_{a\sim\boldsymbol{w}_s}\left[\left\|vec(\widehat{\boldsymbol{\theta}}_{s-1}-\boldsymbol{\lambda}_{s-1})\right\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right]\leq 2\sigma_t\sqrt{\log(|\mathcal{A}|t)}+16\sigma_t\left(2+\frac{\log|\mathcal{A}|}{3}\right)$$

where $\sigma_t := \max_{s\leq t, a\in\mathcal{A}}\|vec(\widehat{\boldsymbol{\theta}}_{s-1}-\boldsymbol{\lambda}_{s-1})\|^2_{\widetilde{a}\widetilde{a}^\intercal}$.

We can bound σ_t as

$$\begin{aligned}\sigma_t &\leq \max_{a\in\mathcal{A},s\leq t}(2\sqrt{f_1(t)}\|a\|_{V_0^{-1}}+\sqrt{B_2})^2\\ &= (2\sqrt{f_1(t)}\max_{a\in\mathcal{A}}\|a\|_{V_0^{-1}}+\sqrt{B_2})^2. \end{aligned}\tag{50}$$

Thus, we obtain for all $\boldsymbol{w}\in\mathcal{V}(\varepsilon)$,

$$(*)\leq f_5(t)+h(t)+\sum_{s=1}^{t-1}\gamma_s(w_t^s(a))^2+2\sigma_t\sqrt{\log(|\mathcal{A}|t)}+16\sigma_t(2+\log(|\mathcal{A}|)/3).\tag{51}$$

In the next step, we relate the supremum over $\mathcal{V}(\varepsilon)$ to the supremum over Δ_K by using the covering argument and a Lipschitz condition. Letting $\boldsymbol{w},\boldsymbol{w}'\in\Delta_K$, observe that

$$\begin{aligned}\widetilde{g}_s(\boldsymbol{w})-\widetilde{g}_s(\boldsymbol{w}') &= \sum_{a\in\mathcal{A}}(\boldsymbol{w}_a-\boldsymbol{w}'_a)\mathbb{E}_{\boldsymbol{\lambda}\sim\pi_{s-1}}\left[\left\|vec(\widehat{\boldsymbol{\theta}}_{s-1}-\boldsymbol{\lambda})\right\|^2_{\widetilde{a}\widetilde{a}^\intercal}\right]\\ &\leq \|\boldsymbol{w}-\boldsymbol{w}'\|_1(w_t^s(a))^2\end{aligned}$$

thus, as $\mathcal{V}(\varepsilon)$ is an ε-cover of Δ_K, for any $\boldsymbol{w}\in\Delta_K$, there exists $\boldsymbol{w}'\in\mathcal{V}(\varepsilon)$ such that $\|\boldsymbol{w}-\boldsymbol{w}'\|_1\leq\varepsilon$ and $\widetilde{g}_s(\boldsymbol{w})\leq\widetilde{g}_s(\boldsymbol{w}')+\varepsilon(w_t^s(a))^2$. All put together, we have

$$\begin{aligned}\sup_{\boldsymbol{w}\in\Delta_K}\sum_{s=t_3+1}^{t-1}(\widetilde{g}_s(\boldsymbol{w})-\widetilde{g}_s(\boldsymbol{w}_s)) &= \sup_{\boldsymbol{w}\in\Delta_K}\min_{\boldsymbol{w}'\in\mathcal{V}(\varepsilon)}\sum_{s=t_3+1}^{t-1}(\widetilde{g}_s(\boldsymbol{w})-\widetilde{g}_s(\boldsymbol{w}')+\widetilde{g}_s(\boldsymbol{w}')-\widetilde{g}_s(\boldsymbol{w}_s))\\ &\leq \sup_{\boldsymbol{w}'\in\mathcal{V}(\varepsilon)}\sum_{s=t_3+1}^{t-1}(\widetilde{g}_s(\boldsymbol{w}')-\widetilde{g}_s(\boldsymbol{w}_s))+\varepsilon\sum_{s=t_3+1}^{t-1}(w_t^s(a))^2\\ &\leq f_5(t)+h(t)+\sum_{s=t_3+1}^{t-1}\gamma_s(w_t^s(a))^2+\sqrt{1/t}\sum_{s=t_3+1}^{t-1}(w_t^s(a))^2\\ &\quad+2\sigma_t\sqrt{\log(|\mathcal{A}|t)}+16\sigma_t(2+\log(|\mathcal{A}|)/3),\end{aligned}$$

where we take $\varepsilon=1/\sqrt{t}$ and recall that $\mathcal{N}(\varepsilon)\leq(3/\varepsilon)^K$. Replacing with the expression of each term, we have

$$\begin{aligned}f_5(t) &\leq \sqrt{4\left(\sum_{s=t_3+1}^{t}\max_{a\in\mathcal{A}}(w_t^s(a))^2\right)\log(|\mathcal{V}(\varepsilon)|t^2)}\\ &\leq \sqrt{4K\left(\sum_{s=t_3+1}^{t}\max_{a\in\mathcal{A}}(w_t^s(a))^2\right)\log(3t^2\sqrt{t})}\\ &\leq \sqrt{10Kt(2\sqrt{f_1(t)}\max_{a\in\mathcal{A}}\|a\|_{V_0^{-1}}+\sqrt{B})^2\log(3t)}=\mathcal{O}(\sqrt{t\log^2(t)}).\end{aligned}\tag{52}$$

Similarly, we have

$$\begin{aligned}\sum_{s=1}^{t-1}\gamma_s(w_t^s(a))^2 &\leq (2\sqrt{f_1(t)}\max_{a\in\mathcal{A}}\|a\|_{V_0^{-1}}+\sqrt{B})^2\sum_{s=1}^{t-1}s^{-\alpha}\\ &\leq (2\sqrt{f_1(t)}\max_{a\in\mathcal{A}}\|a\|_{V_0^{-1}}+\sqrt{B})^2\frac{t^{1-\alpha}}{1-\alpha}=\mathcal{O}(\log(t)t^{1-\alpha}).\end{aligned}\tag{53}$$

Combining Equations (49), (50), (52) and (53) yields the claimed statement. ∎

This achieves the proof of the sublinear regret for the *max* player.

D.4 Proof of Theorem 3

Introducing the event

$$\Xi_t = \bigcap_{i=1}^{5} \Xi_{i,t} \,, \tag{54}$$

we have proven in the sections above that there exists a time $t_3 \in \mathbb{N}$ such that when Ξ_t holds,

$$\begin{aligned} \inf_{\boldsymbol{\lambda} \in \Lambda_{t-1}} \|vec(\boldsymbol{\lambda} - \widehat{\boldsymbol{\theta}}_{t-1})\|^2_{\widehat{V}_{t-1}} &\geq (t-t_3) \max_{\boldsymbol{w} \in \Delta_K} \inf_{\boldsymbol{\lambda} \in \Theta \cap \mathrm{Alt}(S^\star)} \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widehat{V}_{\boldsymbol{w}}} - r_1(t) - r_2(t) - m(t) - z(t), \\ &= (t-1)T^\star(\boldsymbol{\theta})^{-1} - 2\underbrace{\left((1+t_3)T^\star(\boldsymbol{\theta})^{-1} + r_1(t) + r_2(t) + m(t) + z(t)\right)/2}_{f(t-1)}, \end{aligned} \tag{55}$$

and we further proved that $f(t)$ is sub-linear (Equation (28), Equation (30), Lemma 11, Lemma 15). To conclude the proof of Theorem 3, it remains to prove that the LHS of Equation (55) is the GLR at time $t-1$. Indeed, in the transductive setting, as $\Lambda_{t-1} = \Theta \cap \mathrm{Alt}(\widehat{S}_{t-1})$, the result is immediate. In the unstructured setting, the result follows by Lemma 10 applied to $\widehat{\boldsymbol{\mu}}_t$, for $t \geq t_3$ on the event $\Xi_{1,t} \cap \Xi_{2,t} \subset \Xi_t$. Further observe that by their definition we have for each $i \in [5]$, $\mathbb{P}_{\boldsymbol{\nu}}(\Xi_{i,t}) \leq 1/t^2$. Therefore,

$$\mathbb{P}_{\boldsymbol{\nu}}(\Xi_t) \geq 1 - 5/t^2$$

and the conclusion follows.

E LIKELIHOOD RATIO AND POSTERIOR PROBABILITY OF ERROR

In this section, we prove some results related to the generalized likelihood ratio, the lower bound, and the posterior probability of error.

E.1 Lower Bound

We discuss the proof of the lower bound of PSI in this section.

Lemma 1. *An algorithm which is δ-correct on all problems in $\mathcal{D}^K$ satisfies that, for all $\boldsymbol{\nu} \in \mathcal{D}^K$ with regression matrix $\boldsymbol{\theta} \in \Theta$, $\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta] \geq T^\star(\boldsymbol{\theta}) \log(1/(2.4\delta))$ where $T^\star(\boldsymbol{\theta})$ is a* characteristic time *whose inverse is defined as*

$$2T^\star(\boldsymbol{\theta})^{-1} := \sup_{\boldsymbol{w} \in \Delta_K} \inf_{\boldsymbol{\lambda} \in \Theta \cap \mathrm{Alt}(S^\star(\boldsymbol{\theta}))} \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\Sigma^{-1} \otimes V_{\boldsymbol{w}}} \,,$$

and its maximizer set of optimal allocations *is $\boldsymbol{w}^\star(\boldsymbol{\theta})$.*

Proof. The proof of this lemma follows the same lines as Theorem 1 of Garivier and Kaufmann (2016) from which it is simple to prove that the stopping time τ_δ of any δ-correct algorithm for PSI satisfies

$$\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta] \geq T^\star(\boldsymbol{\theta}) \log(1/(2.4\delta)),$$

where, for the problems in $\mathcal{D}^K$,

$$T^\star(\boldsymbol{\theta})^{-1} := \sup_{\boldsymbol{w} \in \Delta_K} \inf_{\boldsymbol{\lambda} \in \Theta \cap \mathrm{Alt}(S^\star(\boldsymbol{\theta}))} \sum_{a \in \mathcal{A}} \frac{1}{2} \boldsymbol{w}_a \|(\boldsymbol{\theta} - \boldsymbol{\lambda})^\intercal a\|^2_{\Sigma^{-1}} \,.$$

Then, observe that by the properties of vectorization and Kronecker product,

$$\begin{aligned} vec(\boldsymbol{\theta}^\intercal a) &= vec(a^\intercal \boldsymbol{\theta}), \\ &= (I_d \otimes a^\intercal) vec(\boldsymbol{\theta}), \end{aligned}$$

which follows from $vec(AB) = (I_d \otimes A)vec(B)$ for $A \in \mathbb{R}^{p,q}, B \in \mathbb{R}^{q,d}$. Therefore,

$$\begin{aligned} \|(\boldsymbol{\theta} - \boldsymbol{\lambda})^\intercal a\|^2_{\Sigma^{-1}} &= vec(\boldsymbol{\theta} - \boldsymbol{\lambda})^\intercal (I_d \otimes a^\intercal)^\intercal \Sigma^{-1} (I_d \otimes a^\intercal) vec(\boldsymbol{\theta} - \boldsymbol{\lambda}) \\ &= vec(\boldsymbol{\theta} - \boldsymbol{\lambda})^\intercal \left((I_d \otimes a)\Sigma^{-1}(I_d \otimes a^\intercal)\right) vec(\boldsymbol{\theta} - \boldsymbol{\lambda}), \end{aligned}$$

further note that

$$\begin{aligned}(I_d \otimes a)\Sigma^{-1}(I_d \otimes a^\intercal) &= \frac{1}{a^\intercal a}(I_d \otimes a)(\Sigma^{-1} \otimes a^\intercal a)(I_d \otimes a^\intercal) \\ &\overset{(i)}{=} \frac{1}{a^\intercal a}\Sigma^{-1} \otimes (aa^\intercal aa^\intercal) \\ &= \Sigma^{-1} \otimes (aa^\intercal),\end{aligned}$$

where (i) follows from $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$. All put together, we have

$$\begin{aligned}T^\star(\boldsymbol{\theta})^{-1} &:= \sup_{\boldsymbol{w}\in\Delta_K} \inf_{\boldsymbol{\lambda}\in\Theta\cap\mathrm{Alt}(S^\star(\boldsymbol{\theta}))} \sum_{a\in\mathcal{A}} \frac{1}{2}\boldsymbol{w}_a \left\|vec(\boldsymbol{\theta}-\boldsymbol{\lambda})\right\|^2_{\Sigma^{-1}\otimes aa^\intercal} \\ &= \frac{1}{2} \sup_{\boldsymbol{w}\in\Delta_K} \inf_{\boldsymbol{\lambda}\in\Theta\cap\mathrm{Alt}(S^\star(\boldsymbol{\theta}))} \left\|vec(\boldsymbol{\theta}-\boldsymbol{\lambda})\right\|^2_{\Sigma^{-1}\otimes V_{\boldsymbol{w}}},\end{aligned}$$

with $V_{\boldsymbol{w}} := \sum_{a\in\mathcal{A}} \boldsymbol{w}_a aa^\intercal$. □

E.2 Posterior Error Probability

Lemma 17. *Under the unstructured assumption or the linear transductive setting, and for any convex set Θ, it holds at each round that*

$$\mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\Theta \cap \mathrm{Alt}(\widehat{S}_t)) \le \alpha_t \exp\left(-\frac{GLR(t)}{c(t-1,\delta)}\right),$$

with $2\alpha_t = p_t(p_t-1) + (|\mathcal{Z}| - p_t)d^{p_t}$, with $p_t := |\widehat{S}_t|$, the size of the empirical Pareto set at time t.

Proof. The proof of this lemma relies on the properties of $\Theta \cap \mathrm{Alt}(\widehat{S}_t)$ and Gaussian concentration. Lemma 26 proves that there exists n_t and convex sets $C_1, \dots, C_{n_t}$ such that $\mathrm{Alt}(\widehat{S}_t) = \cup_{i\in[n_t]} C_i$. Since, $\widehat{\Pi}_t \mid \mathcal{F}_t = \mathcal{N}(vec(\boldsymbol{\theta}_t), c(t-1,\delta)\Sigma \otimes V_t^{-1})$, letting $vec(\boldsymbol{\lambda}_t) \mid \mathcal{F}_t \sim \mathcal{N}(vec(\boldsymbol{\theta}_t), c(t-1,\delta)\Sigma \otimes V_t^{-1})$, we have

$$\mathbb{P}(\boldsymbol{\lambda}_t \in C_i \mid \mathcal{F}_t) = \mathbb{P}(vec(\boldsymbol{\lambda}_t) \in vec(C_i) \mid \mathcal{F}_t)$$

and by Lemma 27, and convexity of $vec(C_i)$, it follows that

$$\mathbb{P}(\boldsymbol{\lambda}_t \in C_i \mid \mathcal{F}_t) \le \frac{1}{2}\exp\left(-\inf_{\boldsymbol{\lambda}\in C_i} \frac{\|vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\lambda})\|^2_{\Sigma\otimes V_t^{-1}}}{2c(t-1,\delta)}\right),$$

therefore, by union bound and convexity of $\Theta \cap C_i$ (since Θ is convex), it follows that

$$\begin{aligned}\mathbb{P}_{\widehat{\Pi}_t|\mathcal{F}_t}(\Theta \cap \mathrm{Alt}(\widehat{S}_t)) &:= \mathbb{P}\left(\boldsymbol{\lambda}_t \in \cup_{i\in[n_t]}(\Theta\cap C_i) \mid \mathcal{F}_t\right) \\ &\le \frac{1}{2}\sum_{i\in[n_t]} \exp\left(-\inf_{\boldsymbol{\lambda}\in\Theta\cap C_i} \frac{\|vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\lambda})\|^2_{\Sigma\otimes V_t^{-1}}}{2c(t-1,\delta)}\right) \\ &\le \frac{n_t}{2}\exp\left(-\inf_{\boldsymbol{\lambda}\in\Theta\cap\mathrm{Alt}(\widehat{S}_t)} \frac{\|vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\lambda})\|^2_{\Sigma\otimes V_t^{-1}}}{2c(t-1,\delta)}\right) \\ &= \frac{n_t}{2}\exp\left(-\frac{\mathrm{GLR}(t)}{c(t-1,\delta)}\right)\end{aligned}$$

and from Lemma 26, $n_t = p_t(p_t-1) + (|\mathcal{Z}| - p_t)d^{p_t}$, with $p_t := |\widehat{S}_t|$. □

F ASYMPTOTIC EXPECTED SAMPLE COMPLEXITY

We prove the asymptotic optimality of our algorithm.

Theorem 1. *Using budget M and inflation c such that $\limsup_{\delta\to 0} \frac{c(t,\delta)\log M(t,\delta)}{\log(1/\delta)} \leq 1$, the PSIPS algorithm satisfies that $\limsup_{\delta\to 0} \mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta^{PS}]/\log(1/\delta) \leq T^\star(\boldsymbol{\theta})$ for all $\boldsymbol{\nu} \in \mathcal{D}^K$ with regression matrix $\boldsymbol{\theta} \in \Theta$, both for unstructured PSI ($\Theta = \mathbb{R}^{K\times d}$, $\mathcal{Z} = \mathcal{A}$, $h = K$, $\boldsymbol{A} = I_K$) and transductive linear PSI (bounded convex Θ).*

Proof. In this section, $(\Xi_t)_{t\geq 1}$ denotes the sequence of events introduced in Equation (54) and $t_3 \in \mathbb{N}$. We have $\tau_\delta^{\text{PS}} = 1 + \sum_{t\geq 1} \mathbb{1}_{(\tau_\delta^{\text{PS}} > t)}$ then

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta^{\text{PS}}] &= 1 + \mathbb{E}\left[\sum_{t\geq 1} \mathbb{P}_{\boldsymbol{\nu}}\left(\tau_\delta^{\text{PS}} > t \mid \mathcal{F}_t\right)\right] \\
&= 1 + \mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t\geq 1} \mathbb{P}_{\boldsymbol{\nu}}\left(\exists m \in [M(t,\delta)] : \boldsymbol{\theta}_t^m \in \Theta \cap \text{Alt}(\widehat{S}_t) \mid \mathcal{F}_t\right)\right] \\
&\leq \mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t\geq t_3} \mathbb{1}_{\Xi_t} M(t,\delta) \mathbb{P}_{\boldsymbol{\nu}}\left(\boldsymbol{\theta}_t \in \Theta \cap \text{Alt}(\widehat{S}_t) \mid \mathcal{F}_t\right)\right] + \left[\sum_{t\geq 1} \mathbb{P}_{\boldsymbol{\nu}}(\Xi_t^c)\right] + (\widetilde{t}_3 + 1),
\end{aligned}$$

where $\boldsymbol{\theta}_t, \boldsymbol{\theta}_t^1 \ldots, \ldots, \boldsymbol{\theta}_t^m$ are *i.i.d* given $\mathcal{F}_t$. Moreover, by Lemma 17, we have

$$\mathbb{P}_{\boldsymbol{\nu}}\left(\boldsymbol{\theta}_t \in \Theta \cap \text{Alt}(\widehat{S}_t) \mid \mathcal{F}_t\right) \leq \alpha_t \exp\left(-\frac{\text{GLR}(t)}{c(t-1,\delta)}\right)$$

with $2\alpha_t = p_t(p_t-1) + (|\mathcal{Z}| - p_t)d^{p_t} \leq |\mathcal{Z}|(|\mathcal{Z}| + d^{|\mathcal{Z}|}) := \alpha_0$, therefore,

$$\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta^{\text{PS}}] \leq \underbrace{\mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t\geq t_3} \mathbb{1}_{\Xi_t} \alpha_0 M(t,\delta) \exp\left(-\frac{\text{GLR}(t)}{c(t-1,\delta)}\right)\right]}_{L_1(\delta)} + \underbrace{\left[\sum_{t\geq 1} \mathbb{P}_{\boldsymbol{\nu}}(\Xi_t^c)\right]}_{L_2} + (\widetilde{t}_3 + 1),$$

then, since $\mathbb{P}_{\boldsymbol{\nu}}(\Xi_t^c) \leq 5/t^2$ we immediately have

$$L_2 \leq 5\pi^2/6. \tag{56}$$

It remains to bound $L_1(\delta)$. Using the saddle-point convergence property on the event Ξ_t ensures that (Theorem 3) for $t \geq t_3$,

$$2\text{GLR}(t) := \inf_{\boldsymbol{\lambda}\in\Theta\cap\text{Alt}(\widehat{S}_t)} \left\|vec(\boldsymbol{\lambda} - \widehat{\boldsymbol{\theta}}_t)\right\|^2_{\Sigma^{-1}\otimes V_t} \geq t \max_{\boldsymbol{w}\in\Delta} \inf_{\boldsymbol{\lambda}\in\Theta\cap\text{Alt}(S^\star(\boldsymbol{\theta}))} \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\Sigma^{-1}\otimes V_{\boldsymbol{w}}} - 2f(t), \tag{57}$$

which further results in

$$\begin{aligned}
L_1(\delta) &\leq \sum_{t\geq t_3} \alpha_0 M(t,\delta) \exp\left(-\frac{t}{T^\star(\boldsymbol{\theta})c(t,\delta)} + f(t)/c(t,\delta)\right) \\
&= \sum_{t\geq t_3} \exp\left(-\frac{t}{T^\star(\boldsymbol{\theta})c(t,\delta)} + f(t)/c(t,\delta) + \log\left(\alpha_0 M(t,\delta)\right)\right).
\end{aligned}$$

To bound the above quantity, let us introduce

$$T(\delta) := \sup\left\{t \mid \frac{t}{T^\star(\boldsymbol{\theta})c(t,\delta)} - \log(M(t,\delta)) - f(t)/c(t,\delta) - \log(\alpha_0) \leq \log(t\log(t))\right\}, \tag{58}$$

then it follows that

$$L_1(\delta) \leq T(\delta) + \sum_{t\geq t_3} (t\log(t))^{-1}. \tag{59}$$

Further observe that $T(\delta)$ can be rewritten as

$$T(\delta) := \sup\left\{t \mid t \leq T^\star(\boldsymbol{\theta})c(t,\delta)\left(\log(M(t,\delta)) + f(t)/c(t,\delta) + \log(\alpha_0) + \log(t\log(t))\right)\right\}.$$

Bounding $T(\delta)$ Since f is sub-linear in t, there exists $\varepsilon \in (0,1)$ such that $f(t) = o_{t\to\infty}(t^\varepsilon)$ then $f(\log(1/\delta)^{1/\varepsilon}) = o(\log(1/\delta))$. Further observe that for $t_\delta = \log(1/\delta)^{1/\alpha}$,

$$\underbrace{c(t_\delta,\delta)\log(M(t_\delta,\delta)) + f(\log(1/\delta)^{1/\varepsilon}) + c(t_\delta,\delta)\left(\log(t_\delta\log(t_\delta)) + \alpha_0\right)}_{b(t_\delta)} \underset{\delta\to 0}{\sim} \log(1/\delta).$$

Let $\delta_{\min} \in (0,1)$ be defined as

$$\delta_{\min} := \inf\left\{\delta \in (0,1) \mid b(t_\delta) > \log(1/\delta)^{1/\varepsilon} T^\star(\boldsymbol{\theta})^{-1}\right\}$$

which is well defined as $b(t_\delta) \underset{\delta\to 0}{\sim} \log(1/\delta)$ and $\varepsilon \in (0,1)$. Letting $T_{\max} = \log(1/\delta_{\min})^{1/\varepsilon}$, remark that for all $t \geq T_{\max}$, there is $(0,1) \ni \delta' \leq \delta_{\min}$ $t'_\delta = t$ and $b(t_{\delta'}) < t_{\delta'}$. Therefore, for all $\delta \leq \delta_{\min}$

$$T(\delta) \leq \log(1/\delta)^{1/\varepsilon} \tag{60}$$

and further noting that by definition $T(\delta) \leq T^\star(\boldsymbol{\theta}) b(T(\delta))$ and b is increasing, it follows that

$$T(\delta) \leq T^\star(\boldsymbol{\theta}) b(\log(1/\delta)^{1/\varepsilon}). \tag{61}$$

Combining (56), (59) and (61), it follows that for $\delta \leq \delta_{\min}$,

$$\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta^{\text{PS}}] \quad \leq \quad T^\star(\boldsymbol{\theta}) b(\log(1/\delta)^{1/\varepsilon}) + 5\pi^2/6 + (\widetilde{t}_3 + 1) + \sum_{t \geq t_3} (t\log(t))^{-1}. \tag{62}$$

Finally, noting that $b(\log(1/\delta)^{1/\varepsilon}) \underset{\delta\to 0}{\sim} \log(1/\delta)$, we have proved that

$$\limsup_{\delta\to 0} \frac{\mathbb{E}[\tau_\delta^{\text{PS}}]}{\log(1/\delta)} \leq T^\star(\boldsymbol{\theta}). \tag{63}$$

□

G POSTERIOR CONVERGENCE

In this section, we prove the following result.

Theorem 2. *Let $\widetilde{\Pi}_t := \mathcal{N}(\widehat{\boldsymbol{\theta}}_t, \Sigma \otimes V_{\boldsymbol{N}_t}^{-1})$ be the posterior distribution under a flat Gaussian prior (without inflation).For all $\boldsymbol{\nu} \in \mathcal{D}^K$ with regression matrix $\boldsymbol{\theta} \in \Theta$, it almost surely holds that $\limsup_{t\to+\infty} -t^{-1}\log \mathbb{P}_{\widetilde{\Pi}_t|\mathcal{F}_t}(\Theta \cap \text{Alt}(S^\star)) \leq T^\star(\boldsymbol{\theta})^{-1}$ for any algorithm, and PSIPS almost surely satisfies that $\liminf_{t\to+\infty} -t^{-1}\log \mathbb{P}_{\widetilde{\Pi}_t|\mathcal{F}_t}(\Theta \cap \text{Alt}(S^\star)) \geq T^\star(\boldsymbol{\theta})^{-1}$, both for unstructured PSI ($\Theta = \mathbb{R}^{K\times d}$, $\mathcal{Z} = \mathcal{A}$, $h = K$, $\boldsymbol{A} = I_K$) and transductive linear PSI (bounded convex Θ).*

We prove that the posterior contraction rate of our algorithm is un-improvable for any adaptive algorithm. In BAI, Russo (2016) and Li et al. (2024) proved a similar result for truncated Gaussian (restricted to a bounded domain). We prove it more generally in the unbounded setting for Gaussian distribution.

We recall that when $\boldsymbol{\lambda}$ is a matrix and π is a distribution supported on vectors, $\boldsymbol{\lambda} \sim \pi$ denotes $vec(\boldsymbol{\lambda}) \sim \pi$.

Proof of Theorem 2. We first prove the upper bound. Similarly to Lemma 17, we can derive (using Lemma 26 and Lemma 27),

$$\mathbb{P}_{\widetilde{\Pi}_t|\mathcal{F}_t}\left(\Theta \cap \text{Alt}(S^\star)\right) \leq \alpha_0 \exp\left(-\frac{\inf_{\boldsymbol{\lambda}\in\Theta\cap\text{Alt}(S^\star)} \|vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\lambda})\|_{\boldsymbol{\Sigma}_t}^2}{2}\right)$$

with $\alpha_0 := |\mathcal{Z}|(|\mathcal{Z}| + d^{|\mathcal{Z}|})$.

From Theorem 3, there exists events $(\Xi_t)_{t\geq}$ and $t_3 \in \mathbb{N}$ such that for $t \geq t_3$, $\widehat{S}_t = S^\star$, and

$$2\text{GLR}(t) := \inf_{\boldsymbol{\lambda}\in\Theta\cap\text{Alt}(\widehat{S}_t)} \left\|vec(\boldsymbol{\lambda} - \widehat{\boldsymbol{\theta}}_t)\right\|_{\Sigma^{-1}\otimes V_t}^2 \quad \geq \quad t\max_{\boldsymbol{w}\in\Delta}\inf_{\boldsymbol{\lambda}\in\Theta\cap\text{Alt}(S^\star(\boldsymbol{\theta}))} \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|_{\Sigma^{-1}\otimes V_{\boldsymbol{w}}}^2 - 2f(t), \tag{64}$$

with $f(t) =_\infty o(t)$ and $\mathbb{P}_{\boldsymbol{\nu}}(\Xi_t) \geq 1 - 5/t^2$. Since $\sum_{t\geq 1} \mathbb{P}(\Xi_t^c) < \infty$, by Borel-Cantelli's lemma, with probability 1, there exists a finite time $\widetilde{t}_3$ (possibly) random such that for $t \geq \widetilde{t}_3$, Ξ_t holds. So for $t \geq \max(t_3, \widetilde{t}_3)$, we have $\widehat{S}_t = S^\star$ and

$$\frac{1}{2} \inf_{\boldsymbol{\lambda} \in \Theta \cap \mathrm{Alt}(S^*)} \left\| vec(\boldsymbol{\lambda} - \widehat{\boldsymbol{\theta}}_t) \right\|^2_{\Sigma^{-1} \otimes V_t} = \mathrm{GLR}(t) \geq tT^\star(\boldsymbol{\theta})^{-1} - f(t),$$

then

$$\mathbb{P}_{\widetilde{\Pi}_t | \mathcal{F}_t}(\Theta \cap \mathrm{Alt}(S^\star)) \leq \alpha_0 \exp\left(-tT^\star(\boldsymbol{\theta})^{-1} + f(t)\right),$$

so that

$$-\frac{1}{t} \log(\mathbb{P}_{\widetilde{\Pi}_t | \mathcal{F}_t}(\Theta \cap \mathrm{Alt}(S^\star))) \geq T^\star(\boldsymbol{\theta})^{-1} - (1/t)\log(\alpha_0) - f(t)/t,$$

thus, since $f(t) =_\infty o(t)$, put together, the above displays show that with probability 1,

$$\liminf_{t \to \infty} -\frac{1}{t} \log(\mathbb{P}_{\widetilde{\Pi}_t | \mathcal{F}_t}(\Theta \cap \mathrm{Alt}(S^\star))) \geq T^\star(\boldsymbol{\theta})^{-1}.$$

The proof of the lower bound uses Lemma 18. In the transductive linear setting, when Θ is convex $\mathrm{Alt}(S^*) \cap \Theta$ is countably convex (by Lemma 26) and bounded. Then taking $O = \mathrm{Alt}(S^*) \cap \Theta$ and applying Lemma 18 , it follows that with probability 1,

$$\begin{aligned} \limsup_{t \to \infty} -\frac{1}{t} \log \mathbb{P}_{\widetilde{\Pi}_t | \mathcal{F}_t}(\mathrm{Alt}(S^*) \cap \Theta) &\leq \frac{1}{2} \sup_{\boldsymbol{w} \in \Delta_K} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(S^*) \cap \Theta} \| vec(\boldsymbol{\theta} - \boldsymbol{\lambda}) \|^2_{\Sigma^{-1} \otimes V_{\boldsymbol{w}}} \\ &= T^*(\boldsymbol{\theta})^{-1}. \end{aligned}$$

In the unstrucutured setting, let $\mathbb{B}_\epsilon$ be the ball centered on $\boldsymbol{\mu}$ and with radius ε as in Lemma 9. We have $-\mathbb{P}(\mathrm{Alt}(S^*)) \leq -\mathbb{P}(\mathrm{Alt}(S^*) \cap \mathbb{B}_\epsilon)$, then $O = \mathrm{Alt}(S^*) \cap \mathbb{B}_\epsilon$ is countably convex (Lemma 26 and $\mathbb{B}_\epsilon$ is convex) and bounded. Applying Lemma 18, we obtain

$$\begin{aligned} \limsup_{t \to \infty} -\frac{1}{t} \log \mathbb{P}_{\widetilde{\Pi}_t | \mathcal{F}_t}(\mathrm{Alt}(S^*)) &\leq \limsup_{t \to \infty} -\frac{1}{t} \log \mathbb{P}_{\widetilde{\Pi}_t | \mathcal{F}_t}(\mathrm{Alt}(S^*) \cap \mathbb{B}_\epsilon) \\ &\leq \frac{1}{2} \sup_{\boldsymbol{w} \in \Delta_K} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(S^*) \cap \mathbb{B}_\epsilon} \| vec(\boldsymbol{\theta} - \boldsymbol{\lambda}) \|^2_{\Sigma^{-1} \otimes V_{\boldsymbol{w}}} \\ &= \frac{1}{2} \sup_{\boldsymbol{w} \in \Delta_K} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(S^*)} \| vec(\boldsymbol{\theta} - \boldsymbol{\lambda}) \|^2_{\Sigma^{-1} \otimes V_{\boldsymbol{w}}} \qquad \text{(Lemma 9)} \\ &= T^*(\boldsymbol{\theta})^{-1}, \end{aligned}$$

which concludes the proof. ◻

Lemma 18. *Let $O \subset \mathbb{R}^{h \times d}$ be countably convex bounded set, with non-empty interior. With probability 1, it holds that*

$$\limsup_{t \to \infty} -\frac{1}{t} \log \mathbb{P}_{\widetilde{\Pi}_t | \mathcal{F}_t}(O) \leq \frac{1}{2} \sup_{\boldsymbol{w} \in \Delta_K} \inf_{\boldsymbol{\lambda} \in O} \| vec(\boldsymbol{\theta} - \boldsymbol{\lambda}) \|^2_{\Sigma^{-1} \otimes V_{\boldsymbol{w}}}.$$

Proof. Since O is countably there exists convex sets $C_1 \dots C_n$ such that $O = \cup_{i \in [n]} C_i$.

We have

$$\mathbb{P}_{\widetilde{\Pi}_t | \mathcal{F}_t}(O) = (2\pi)^{-dh/2} \det(\boldsymbol{\Sigma}_t)^{-1/2} \int_{vec(O)} \exp\left(-\frac{\| vec(\widehat{\boldsymbol{\theta}}_t) - \lambda \|^2_{\widetilde{V}_t}}{2}\right) d\lambda.$$

Similarly to the proof of the regret of the min learner, let $\gamma > 0$ (to be defined) and $\widetilde{\boldsymbol{\lambda}}_t \in \mathrm{argmin}_{\boldsymbol{\lambda} \in \overline{O}} \| vec(\boldsymbol{\lambda} - \widehat{\boldsymbol{\theta}}_t) \|^2_{\widetilde{V}_t}$. Since O is a union of convex sets, there exists a convex set $\mathcal{C} \subset O$ such that $\widetilde{\boldsymbol{\lambda}}_t \in \mathcal{C}$. Then letting $\mathcal{N}_\gamma := \{(1-\gamma)\widetilde{\boldsymbol{\lambda}}_t + \gamma \boldsymbol{\lambda}, \boldsymbol{\lambda} \in$

$\mathcal{C}\} = (1-\gamma)\widetilde{\boldsymbol{\lambda}}_t + \gamma\mathcal{C}$, it follows

$$
\begin{aligned}
w_t &:= \int_{vec(O)} \exp\left(-\frac{\|vec(\widehat{\boldsymbol{\theta}}_t) - \lambda\|^2_{\widetilde{V}_t}}{2}\right) d\lambda \\
&\geq \int_{vec(\mathcal{C})} \exp\left(-\frac{\|vec(\widehat{\boldsymbol{\theta}}_t) - \lambda\|^2_{\widetilde{V}_t}}{2}\right) d\lambda \\
&\geq \int_{vec(\mathcal{N}_\gamma)} \exp\left(-\frac{\eta\|vec(\widehat{\boldsymbol{\theta}}_t) - \lambda\|^2_{\widetilde{V}_t}}{2}\right) d\lambda \quad \text{(convexity of } \mathcal{C}) \\
&= \int_{\gamma vec(\mathcal{C})} \exp\left(-\frac{\|vec(\widehat{\boldsymbol{\theta}}_t - (1-\gamma)\widetilde{\boldsymbol{\lambda}}_t) - \lambda\|^2_{\widetilde{V}_t}}{2}\right) d\lambda \\
&= \int_{vec(\mathcal{C})} \gamma^{dh} \exp\left(-\frac{\|(1-\gamma)vec(\widehat{\boldsymbol{\theta}}_t - \widetilde{\boldsymbol{\lambda}}_t) + \gamma vec(\widehat{\boldsymbol{\theta}}_t) - \gamma\lambda\|^2_{\widetilde{V}_t}}{2}\right) d\lambda \\
&\geq \int_{vec(\mathcal{C})} \gamma^{dh} \exp\left(-\frac{\left((1-\gamma)\|vec(\widehat{\boldsymbol{\theta}}_t - \widetilde{\boldsymbol{\lambda}}_t)\|^2_{\widetilde{V}_t} + \gamma\|\lambda - vec(\widehat{\boldsymbol{\theta}}_t)\|^2_{\widetilde{V}_t}\right)}{2}\right) d\lambda
\end{aligned}
$$

therefore,

$$
\log w_t \geq dh\log(\gamma) - \frac{(1-\gamma)}{2}\|vec(\widehat{\boldsymbol{\theta}}_t - \widetilde{\boldsymbol{\lambda}}_t)\|^2_{\widetilde{V}_t} - \frac{\gamma\mathbb{E}_{\boldsymbol{\lambda}\sim\mathcal{U}(\mathcal{C})}\left[\|vec(\boldsymbol{\lambda} - \widehat{\boldsymbol{\theta}}_t)\|^2_{\widetilde{V}_t}\right]}{2} + \log(\mathrm{vol}(vec(\mathcal{C}_*))) \tag{65}
$$

where $\mathcal{C}_*$ is the set of minimum volume among $C_1, \dots C_n$. We recall

$$
\Xi_{1,t} := \left\{\forall s \leq t, \left\|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_s)\right\|^2_{\widetilde{V}_s} \leq \beta(t, 1/t^2) =: f_1(t)\right\},
$$

where $\beta(t,\delta)$ is defined as in Lemma 3, and $\mathbb{P}_{\boldsymbol{\nu}}(\Xi_{1,t}) \geq 1 - 1/t^2$. Moreover, $f_1(t)$ is logarithmic in t. Since $\sum_{t\geq 1}\mathbb{P}(\Xi_t^c) < \infty$, by Borel-Cantelli's lemma, with probability 1, there exists a finite time $\widetilde{t}$ (possibly) random such that for $t \geq \widetilde{t}$, $\Xi_{1,t}$ holds.

Taking $\gamma = 1/t$ in Equation (65), and for $t \geq \widetilde{t}$, we get (after simplification)

$$
\log w_t \quad \geq \quad -dh\log(t) - \frac{(1-1/t)}{2}\|vec(\widehat{\boldsymbol{\theta}}_t - \widetilde{\boldsymbol{\lambda}}_t)\|^2_{\widetilde{V}_t} - (\sqrt{tL(O)} + \sqrt{f_1(t)})^2/t + \log(\mathrm{vol}(vec(\mathcal{C}_*))),
$$

where $L(O) = \max_{a\in\mathcal{A},\boldsymbol{\lambda}\in O}\|vec(\boldsymbol{\lambda} - \boldsymbol{\theta})\|^2_{\Sigma^{-1}\otimes aa^\intercal}$. Now, by convex inequality, we have for any $\boldsymbol{\lambda} \in O$,

$$
\begin{aligned}
\|vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\lambda})\|^2_{\widetilde{V}_t} &\leq \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{V}_t} + 2vec(\boldsymbol{\lambda} - \widehat{\boldsymbol{\theta}}_t)^T\widetilde{V}_t vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_t) \\
&\leq \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{V}_t} + 2\|vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\lambda})\|_{\widetilde{V}_t}\|vec(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_t)\|_{\widetilde{V}_t} \\
&\leq \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{V}_t} + 2f_1(t) + 2\sqrt{f_1(t)}\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|_{\widetilde{V}_t} \\
&\leq \|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{V}_t} + 2f_1(t) + 2\sqrt{f_1(t)}\sqrt{tL(O)},
\end{aligned}
$$

thus we have

$$
\|vec(\widehat{\boldsymbol{\theta}}_t - \widetilde{\boldsymbol{\lambda}}_t)\|^2_{\widetilde{V}_t} \quad \leq \quad \inf_{\boldsymbol{\lambda}\in O}\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{V}_t} + 2f_1(t) + 2\sqrt{f_1(t)}\sqrt{tL(O)}.
$$

Put together, we have proved that

$$
\begin{aligned}
-\log\mathbb{P}_{\widetilde{\Pi}_t|\mathcal{F}_t}(O) \leq \frac{1}{2}\inf_{\boldsymbol{\lambda}\in O}\|vec(\boldsymbol{\theta} - \boldsymbol{\lambda})\|^2_{\widetilde{V}_t} + 2f_1(t) + 2\sqrt{f_1(t)}\sqrt{tL(O)} + (dh/2)\log(2\pi) + \frac{1}{2}\log\det(\boldsymbol{\Sigma}_t) + dh\log(t) \\
+ (\sqrt{tL(O)} + \sqrt{f_1(t)})^2/t - \log(\mathrm{vol}(vec(\mathcal{C}_*))).
\end{aligned}
$$

We recall that $\boldsymbol{\Sigma}_t = \Sigma \otimes V_t^{-1}$ and $\widetilde{V}_t = \boldsymbol{\Sigma}_t^{-1}$ with $V_t = V_{\boldsymbol{N}_t} + V_0$ so $\log\det(\boldsymbol{\Sigma}_t) \le \log(\det(\Sigma \otimes V_0^{-1})) = cste$ (due to initialization or regularization). Combining the displays above yield,

$$-\log \mathbb{P}_{\widetilde{\Pi}_t|\mathcal{F}_t}(O) \le \frac{1}{2}\inf_{\boldsymbol{\lambda}\in O} \|vec(\boldsymbol{\theta}-\boldsymbol{\lambda})\|^2_{\widetilde{V}_t} + o(t),$$

which finally yields,

$$\begin{aligned}
-1/t \log \mathbb{P}_{\widetilde{\Pi}_t|\mathcal{F}_t}(O) &\le \frac{1}{2}\inf_{\boldsymbol{\lambda}\in O} \|vec(\boldsymbol{\theta}-\boldsymbol{\lambda})\|^2_{\widetilde{V}_t/t} + o(t)/t, \\
&= \frac{1}{2}\inf_{\boldsymbol{\lambda}\in O} \|vec(\boldsymbol{\theta}-\boldsymbol{\lambda})\|^2_{\Sigma^{-1}\otimes V_{\boldsymbol{N}_t/t}} + o(t)/t, \\
&\le \frac{1}{2}\sup_{\boldsymbol{w}\in\Delta_K}\inf_{\boldsymbol{\lambda}\in O} \|vec(\boldsymbol{\theta}-\boldsymbol{\lambda})\|^2_{\Sigma^{-1}\otimes V_{\boldsymbol{w}}} + o(t)/t,
\end{aligned}$$

and taking the limit yields the claimed statement. □

H CONCENTRATION RESULTS

H.1 Concentration of Good Event

Lemma 3. *Let $s > 1$, ζ be the Riemann ζ function and $\overline{W}_{-1}(x) := -W_{-1}(-e^{-x})$ for all $x \ge 1$, where W_{-1} is the negative branch of the Lambert W function. Let $\mathcal{E}_\delta$ in (6). Then, we have $\mathbb{P}_{\boldsymbol{\nu}}(\mathcal{E}_\delta^{\complement}) \le \delta$ by taking*

$$\beta(t,\delta) = \frac{dK}{2}\overline{W}_{-1}\left(\frac{2}{dK}\log\frac{e^{Ks}\zeta(s)^K}{\delta} + \frac{2s}{d}\log\left(1+\frac{d}{2s}\log\frac{t}{K}\right)+1\right)$$

in the unstructured setting, and taking

$$\sqrt{\beta(t,\delta)} = \sqrt{\log\left(\frac{1}{\delta}\left(\frac{L_{\mathcal{A}}^2}{h\xi}t+1\right)^{\frac{dh}{2}}\right)} + \sqrt{\frac{dL_{\mathcal{M}}^2}{2\lambda_{\min}(\Sigma)\xi}}$$

in the transductive linear setting, where $L_{\mathcal{A}} := \max_{a\in\mathcal{A}}\|a\|_2$ and $L_{\mathcal{M}} := \max_{\boldsymbol{\lambda}\in\Theta}\max_{c\in[d]}\|\boldsymbol{\lambda}e_c\|_2$.

H.1.1 Unstructured Setting

In the unstructured setting, we take $\xi = 0$ and denote the empirical mean by $\widehat{\boldsymbol{\mu}}_t$ instead of $\widehat{\boldsymbol{\theta}}_t$ with $\widehat{\mu}_{t,i} = \frac{1}{N_{t,i}}\sum_{s\in[t-1]}\mathbb{1}(I_s = i)X_s$ where I_t denote the index associated to the arm a_t pulled at time t.

The set of multivariate Gaussian distributions with known covariance matrix Σ is a d-dimensional canonical exponential family admitting $\mathbb{R}^d$ as set of possible natural parameters. Let us denote by $\theta_i := \Sigma^{-1}\mu_i$ the natural parameter associated to μ_i. The associated log-partition function is defined as $\phi_i(\theta_i) := \frac{1}{2}\theta_i^\intercal\Sigma\theta_i$ and satisfies that $\nabla\phi_i(\theta_i) = \Sigma\mu_i$ and $\nabla^2\phi_i(\mu_i) = \Sigma$. We note that $\widehat{\mu}_{t,i}$ is the mean parameter associated with the natural parameter $(\nabla\phi_i)^{-1}(\frac{1}{N_{t,i}}\sum_{s\in[t-1]}\mathbb{1}(I_s=i)X_s) = \Sigma^{-1}\widehat{\mu}_{t,i}$. Therefore, we can use the concentration results derived in Chapter 4 of Degenne (2019). In particular, we will be using the generalization to K arms derived in Jourdan et al. (2023).

Lemma 19 (Lemma 39 in Jourdan et al. (2023) based on Degenne (2019))**.** *Let $\{\rho_{0,i}\}_{i\in[K]} \subseteq \mathcal{P}(\mathbb{R}^d)$. With probability $1-\delta$, for all $t\in\mathbb{N}$,*

$$\sum_{i\in[K]}\frac{N_{t,i}}{2}\|\widehat{\mu}_{t,i}-\mu_i\|^2_{\Sigma^{-1}} \le \log(1/\delta) - \sum_{i\in[K]}\ln\mathbb{E}_{y\sim\rho_{0,i}}\exp\left(-\frac{N_{t,i}}{2}\|\widehat{\mu}_{t,i}-y\|^2_{\Sigma^{-1}}\right).$$

Lemma 20 gives the concentration threshold for the unstructured setting. The proof is obtained by using the method from Degenne (2019): peeling argument with a sequence of Gaussian priors.

Lemma 20. *Let $s > 1$, $\overline{W}_{-1}$ as in Lemma 25 and ζ be the Riemann ζ function. With probability $1-\delta$, for all $t\in\mathbb{N}$*

$$\sum_{i\in[K]}\frac{N_{t,i}}{2}\|\widehat{\mu}_{t,i}-\mu_i\|^2_{\Sigma^{-1}} \le \frac{dK}{2}\overline{W}_{-1}\left(\frac{2}{dK}\log\frac{e^{Ks}\zeta(s)^K}{\delta} + \frac{2s}{d}\log\left(1+\frac{d}{2s}\log\frac{t-1}{K}\right)+1\right).$$

Proof. Let $\gamma > 1$, $\eta > 0$ to be defined later. let $(k_i)_{i\in[K]} \in \mathbb{N}^K$ and $n_k := \gamma^k$ for $k \in \mathbb{N}$. Let $(M_{0,i})_{i\in[K]}$ be positive definite matrices to be defined later. Let us define the prior $\rho_{0,i} = \mathcal{N}(\mu_i, M_{0,i}^{-1})$ for all $i \in [K]$. Using the computations in the proof of Lemma 4.23 in Degenne (2019), we obtain, for all $i \in [K]$,

$$\mathbb{E}_{y\sim\rho_{0,i}} \exp\left(-\frac{N_{t,i}}{2}\|\widehat{\mu}_{t,i} - y\|^2_{\Sigma^{-1}}\right) = \sqrt{\frac{\det(M_{0,i})}{\det(N_{t,i}\Sigma^{-1} + M_{0,i})}} \exp\left(-\frac{1}{2}\|\widehat{\mu}_{t,i} - \mu_i\|^2_{(\Sigma/N_{t,i} + M_{0,i}^{-1})^{-1}}\right) .$$

For all $t \in \mathbb{N}$, for all i, there exists k_i such $N_{t,i} \in [n_{k_i}, n_{k_i+1})$. Let $(k_i)_{i\in[K]}$ be such vector of indices. In the following, we consider $M_{0,i}^{-1} := \frac{1}{n_{k_i}\eta}\Sigma$, hence we have

$$\sqrt{\frac{\det(M_{0,i})}{\det(N_{t,i}\Sigma^{-1} + M_{0,i})}} = \left(1 + \frac{N_{t,i}}{n_{k_i}\eta}\right)^{-d/2} \geq \left(1 + \frac{1}{\eta}\right)^{-d/2} \gamma^{-d/2} ,$$

$$(\Sigma/N_{t,i} + M_{0,i}^{-1})^{-1} = N_{t,i}(1 + \frac{N_{t,i}}{n_{k_i}\eta})^{-1}\Sigma^{-1} \preccurlyeq N_{t,i}\left(1 - \frac{1}{\eta + 1}\right)\Sigma^{-1} .$$

where we used that $N_{t,i} \geq n_{k_i}$, $N_{t,i} \leq n_{k_i}\gamma$ and $\gamma > 1$. Therefore, we obtain

$$\mathbb{E}_{y\sim\rho_{0,i}} \exp\left(-\frac{N_{t,i}}{2}\|\widehat{\mu}_{t,i} - y\|^2_{\Sigma^{-1}}\right) \geq \left(1 + \frac{1}{\eta}\right)^{-d/2} \gamma^{-d/2} \exp\left(-\frac{N_{t,i}}{2}\left(1 - \frac{1}{\eta + 1}\right)\|\widehat{\mu}_{t,i} - \mu_i\|^2_{\Sigma^{-1}}\right) .$$

We use Lemma 19 with prior $\rho_{0,i} = \mathcal{N}(\mu_i, M_{0,i}^{-1})$. By taking the logarithm, summing over $i \in [K]$ and re-ordering terms, we obtain that, with probability $1 - \delta$, for all $t \in \mathbb{N}$, if $\boldsymbol{N}_t \in \bigotimes_{i\in[K]}[n_{k_i}, n_{k_i+1})$, then

$$\sum_{i\in[K]} \frac{N_{t,i}}{2}\|\widehat{\mu}_{t,i} - \mu_i\|^2_{\Sigma^{-1}} \leq (\eta + 1)\left(\log(1/\delta) + \frac{dK}{2}\log\gamma + \frac{dK}{2}\log\left(1 + \frac{1}{\eta}\right)\right)$$

To choose an optimal value for η, we rely on Lemma 21 taken from Degenne (2019).

Lemma 21 (Lemma A.3 in Degenne (2019))**.** *For $a, b \geq 1$, the minimal value of $f(\eta) = (1+\eta)(a + \ln(b + \frac{1}{\eta}))$ is attained at $\eta^\star$ such that $f(\eta^\star) \leq 1 - b + \overline{W}_{-1}(a + b)$. If $b = 1$, then there is equality.*

Then, we have that, with probability $1 - \delta$, for all $t \in \mathbb{N}$, if $\boldsymbol{N}_t \in \bigotimes_{i\in[K]}[n_{k_i}, n_{k_i+1})$, then

$$\sum_{i\in[K]} \frac{N_{t,i}}{2}\|\widehat{\mu}_{t,i} - \mu_i\|^2_{\Sigma^{-1}} \leq \frac{dK}{2}\overline{W}_{-1}\left(1 + \frac{2}{dK}\log(1/\delta) + \log\gamma\right)$$

Let us define $w_k = \frac{1}{\zeta(s)(k+1)^s}$. Then, we have

$$\sum_{(k_i)_{i\in[K]}\in\mathbb{N}^K} \prod_{i\in[K]} w_{k_i} = \left(\sum_{k\in\mathbb{N}} w_k\right)^K = \left(\frac{1}{\zeta(s)}\sum_{k\in\mathbb{N}} \frac{1}{(k+1)^s}\right)^K = 1 .$$

Using that $\bigcup_{(k_i)_{i\in[K]}\in\mathbb{N}^K} \bigotimes_{i\in[K]}[n_{k_i}, n_{k_i+1}) = \mathbb{N}^K$, we can apply the above results for each $(k_i)_{i\in[K]} \in \mathbb{N}^K$ with probability $1 - \delta\prod_{i\in[K]} w_{k_i}$. When $N_{t,i} \in [n_{k_i}, n_{k_i+1})$, we have $k_i \leq \log(N_{t,i})/\log(\gamma)$. Therefore, a direct union bound yields that, with probability $1 - \delta$, for all $t \in \mathbb{N}$,

$$\begin{aligned}&\sum_{i\in[K]} \frac{N_{t,i}}{2}\|\widehat{\mu}_{t,i} - \mu_i\|^2_{\Sigma^{-1}} \\ &\leq \frac{dK}{2}\overline{W}_{-1}\left(1 + \frac{2}{dK}\log(1/\delta) + \frac{2}{d}\log\zeta(s) + \frac{2s}{dK}\sum_{i\in[K]}\log(\log(\gamma) + \log N_{t,i}) - \frac{2s}{d}\log\log(\gamma) + \log\gamma\right) .\end{aligned}$$

The function $f : \gamma \to \log\gamma - \frac{2s}{d}\log\log\gamma$ is minimized at $\gamma^\star = e^{2s/d}$ with value $f(\gamma^\star) = 2s(1 - \log(2s/d))/d$. Taking $\gamma^\star$ and using the concavity of $x \to \log\left(\frac{2s}{d} + \log x\right)$ concludes the proof since we have

$$\frac{2s}{dK}\sum_{i\in[K]}\log\left(\frac{2s}{d} + \log N_{t,i}\right) \leq \frac{2s}{d}\log\left(\frac{2s}{d} + \log\frac{t-1}{K}\right) .$$

$\square$

H.1.2 Structured Setting

We prove the following result.

Lemma 22. *Letting*

$$\sqrt{\beta(t,\delta)} = \sqrt{\log\left(\frac{1}{\delta}\left(\frac{L_{\mathcal{A}}^2}{h\xi}t+1\right)^{\frac{dh}{2}}\right)} + \sqrt{\frac{dL_{\mathcal{M}}^2}{2\lambda_{\min}(\Sigma)\xi}}\,,$$

where $L_{\mathcal{A}} := \max_{a\in\mathcal{A}} \|a\|_2$ *and* $L_{\mathcal{M}} := \max_{\boldsymbol{\lambda}\in\Theta}\max_{c\in[d]} \|\boldsymbol{\lambda} e_c\|_2$, *the event*

$$\mathcal{E}_\delta := \left\{\forall t\geq 1,\ \frac{1}{2}\left\|vec(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}_t)\right\|^2_{\Sigma^{-1}\otimes(V_{\boldsymbol{N}_t}+\xi I_h)} \leq \beta(t-1,\delta)\right\}$$

holds with probability at least $1-\delta$.

Proof. At time t, a_t is chosen from $[K]$ (adaptively) and a random vector $X_t := (I_d\otimes a_t)vec(\boldsymbol{\theta}) + \varepsilon_t$ is observed and $\varepsilon, \varepsilon_1, \dots \varepsilon_t$ are centered *i.i.d* Σ-subgaussian random vectors, i.e., for all $u\in\mathbb{R}^d$,

$$\mathbb{E}\left[\exp(u^\intercal\varepsilon)\right] \leq \exp\left(\frac{1}{2}\|u\|^2_\Sigma\right). \tag{66}$$

Let us define $S_t := \sum_{s=1}^t vec(a_s\varepsilon_s^\intercal)$, $\boldsymbol{A}_t := (a_1\dots a_t)^\intercal$, $\boldsymbol{X}_t := (X_1\dots X_t)$ and $\boldsymbol{H}_t := (\varepsilon_1\dots\varepsilon_t)^\intercal$. The regularized least-squares estimator of $\boldsymbol{\theta}$ (for the Frobenius norm) is given by

$$\widehat{\boldsymbol{\theta}}_t = (V_{\boldsymbol{N}_t}+\xi I_h)^{-1}\boldsymbol{A}_t^\intercal\boldsymbol{X}_t.$$

We introduce a slightly modified filtration that makes a_t predictable (as $\widetilde{\boldsymbol{w}}_{t+1}$ is known at time t, an agent can pull a_{t+1} at the end of round t). We define $\mathcal{F}_t = \sigma(\mathcal{F}_t, a_{t+1})$. We define the following t-indexed stochastic process for any $\lambda\in\mathbb{R}^{d\cdot h}$

$$M^\lambda(t) := \exp\left(\lambda^\intercal S_t - \frac{1}{2}\|\lambda\|^2_{\Sigma\otimes V_{\boldsymbol{N}_t}}\right).$$

We first justify that $M^\lambda(t)$ is a super-martingale. Indeed, $M^\lambda(t)$ is an $\{\widetilde{\mathcal{F}}_t\}_{t\geq1}$-adapted process and

$$\mathbb{E}\left[M^\lambda(t+1)\mid\widetilde{\mathcal{F}}_t\right] = M^\lambda(t)\mathbb{E}\left[\exp\left(\lambda^\intercal vec(a_{t+1}\varepsilon_{t+1}^\intercal) - \frac{1}{2}\|\lambda\|^2_{\Sigma\otimes a_{t+1}a_{t+1}^\intercal}\right)\mid\widetilde{\mathcal{F}}_t\right], \tag{67}$$

then, note that by properties of Kronecker product, $vec(x\varepsilon^\intercal) = (I_d\otimes x)\varepsilon$, so

$$\lambda^\intercal vec(a_{t+1}\varepsilon_{t+1}^\intercal) = \lambda^\intercal(I_d\otimes a_{t+1})\varepsilon_{t+1}, \tag{68}$$

and

$$\begin{aligned}
\lambda^\intercal(I_d\otimes a)\Sigma(\lambda^\intercal(I_d\otimes a))^\intercal &= \lambda^\intercal(I_d\otimes a)\Sigma(I_d\otimes a^\intercal)\lambda \\
&= \frac{1}{aa^\intercal}\lambda^\intercal(I_d\otimes a)(\Sigma\otimes a^\intercal a)(I_d\otimes a^\intercal)\lambda \\
&= \frac{1}{aa^\intercal}\lambda^\intercal(\Sigma\otimes(aa^\intercal x a^\intercal))\lambda = \lambda^\intercal\left(\Sigma\otimes aa^\intercal\right)\lambda
\end{aligned}$$

which follows from the property $(A\otimes B)(C\otimes D) = (AC)\otimes(BD)$, proving that

$$\|(I_d\otimes a_{t+1}{}^\intercal)\lambda\|^2_\Sigma = \|\lambda\|^2_{\Sigma\otimes(a_{t+1}a_{t+1}{}^\intercal)}.$$

Thus, combining the above display with (67) and the sub-gaussian property of ε_{t+1} shows that

$$\mathbb{E}\left[M^\lambda(t+1)\mid\widetilde{\mathcal{F}}_t\right] \leq M^\lambda(t),$$

so $(M^\lambda(t))_t$ is a super-martingale. For $\xi>0$ and $\Lambda\sim\mathcal{N}(0_{dh}, \Sigma^{-1}\otimes(\xi I_h)^{-1})$, we define

$$M(t) := \mathbb{E}\left[M^\Lambda(t)|\widetilde{\mathcal{F}}_\infty\right].$$

Letting $U_t := \Sigma \otimes V_{\boldsymbol{N}_t}$ and defining $c(P) := \sqrt{(2\pi)^{d \cdot h})/det(P)} = \int_{\mathbb{R}^{d \cdot h}} \exp(-\frac{1}{2}\lambda^\intercal P \lambda) d\lambda$, we have

$$\begin{aligned} M(t) &= \frac{1}{c(\Sigma \otimes \xi I_h)} \int \exp\left(\lambda^\intercal S_t - \frac{1}{2}\lambda^\intercal U_t \lambda - \frac{1}{2}\lambda^\intercal (\Sigma \otimes (\xi I_h))\lambda\right) d\lambda \\ &= \frac{1}{c(\Sigma \otimes \xi I_h)} \exp\left(\frac{1}{2} S_t^\intercal W_t^{-1} S_t\right) \int \exp\left(-\frac{1}{2}\left(\lambda - W_t^{-1}S_t\right)^\intercal W_t \left(\lambda - W_t^{-1} S_t\right)\right) d\lambda \end{aligned}$$

where $W_t := U_t + \Sigma \otimes (\xi I_h) = \Sigma \otimes (V_{\boldsymbol{N}_t} + \xi I_h)$ then direct algebra yields

$$\begin{aligned} M(t) &= \frac{c(\Sigma \otimes (V_{\boldsymbol{N}_t} + \xi I_h))}{c(\Sigma \otimes \xi I_h)} \exp\left(\frac{1}{2}\|S_t\|^2_{W_t^{-1}}\right) \\ &= \left(\frac{\det(\Sigma \otimes \xi I_h)}{\det(\Sigma \otimes (V_{\boldsymbol{N}_t} + \xi I_h))}\right)^{1/2} \exp\left(\frac{1}{2}\|S_t\|^2_{W_t^{-1}}\right), \end{aligned}$$

then it is known that M_t is also a super-martingale (by Fubini's theorem) with $\mathbb{E}[M_t] \leq 1$ and using classical technique similar to section 20.1 of Lattimore and Szepesvári (2020) on self-normalized concentration with super-martingales, it holds that with probability at least $1-\delta$ we have for all $t \geq 1$,

$$\begin{aligned} \|S_t\|^2_{W_t^{-1}} &\leq 2\log\left(\frac{det(\Sigma \otimes (V_{\boldsymbol{N}_t} + \xi I_h))^{1/2}}{\delta \det(\Sigma \otimes \xi I_h)^{1/2}}\right) \\ &= 2\log\left(\frac{\det(V_{\boldsymbol{N}_t} + \xi I_h)^{d/2}}{\delta \xi^{d \cdot h/2}}\right), \end{aligned} \tag{69}$$

which follows as for $A \in \mathbb{R}^{p \times p}$ and $B \in \mathbb{R}^{q \times q}$

$$\det(A \otimes B) = \det(A)^q \det(B)^p.$$

We further have $\widehat{\boldsymbol{\theta}}_t := (V_{\boldsymbol{N}_t} + \xi I_h)^{-1} \boldsymbol{A}_t^\intercal \boldsymbol{X}_t$ then

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_t &= (V_{\boldsymbol{N}_t} + \xi I_h)^{-1} \boldsymbol{A}_t^\intercal (\boldsymbol{A}_t \boldsymbol{\theta} + \boldsymbol{H}_t) \\ &= (V_{\boldsymbol{N}_t} + \xi I_h)^{-1} V_{\boldsymbol{N}_t} \boldsymbol{\theta} + (V_{\boldsymbol{N}_t} + \xi I_h)^{-1} \boldsymbol{A}_t^\intercal \boldsymbol{H}_t \\ &= \boldsymbol{\theta} - \xi (V_{\boldsymbol{N}_t} + \xi I_h)^{-1} \boldsymbol{\theta} + (V_{\boldsymbol{N}_t} + \xi I_h)^{-1} \boldsymbol{A}_t^\intercal \boldsymbol{H}_t. \end{aligned}$$

We recall that for any B, C well defined and $C \in \mathbb{R}^{p \times q}$

$$vec(BC) = (I_q \otimes B) vec(C).$$

Therefore, we can derive

$$vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}) = -\xi (I_d \otimes (V_{\boldsymbol{N}_t} + \xi I_h)^{-1}) vec(\boldsymbol{\theta}) + (I_d \otimes (V_{\boldsymbol{N}_t} + \xi I_h)^{-1}) S_t.$$

We recall that for any two invertible matrices A, B, $A \otimes B$ is invertible and

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

and for any matrices A, B, C, D when the product is possible

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$$

and $(A \otimes B)^\intercal = (A^\intercal \otimes B^\intercal)$. Letting $Z_t := I_d \otimes (V_{\boldsymbol{N}_t} + \xi I_h)$ and for any $x \in \mathbb{R}^{dh}$,

$$\begin{aligned} x^\intercal vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}) &= -\xi x^\intercal Z_t^{-1} vec(\boldsymbol{\theta}) + x^\intercal Z_t^{-1} S_t \\ &= \xi x^\intercal (\Sigma \otimes I_h) W_t^{-1} vec(\boldsymbol{\theta}) + x^\intercal (\Sigma \otimes I_h) W_t^{-1} S_t \end{aligned}$$

where $W_t = \Sigma \otimes (V_{\boldsymbol{N}_t} + \xi I_h)$, then thanks to Cauchy-Schwartz inequality

$$\begin{aligned} |x^\intercal vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta})| &\leq \left\| W_t^{-1/2} (\Sigma \otimes I_h) x \right\| \left(\xi \|vec(\boldsymbol{\theta})\|_{W_t^{-1}} + \|S_t\|_{W_t^{-1}}\right) \\ &= \left\| \Sigma^{1/2} \otimes (V_{\boldsymbol{N}_t} + \xi I_h)^{-1/2} x \right\| \left(\xi \|\theta\|_{W_t^{-1}} + \|S_t\|_{W_t^{-1}}\right). \end{aligned}$$

Letting

$$x := (\Sigma^{-1} \otimes (V_{\boldsymbol{N}_t} + \xi I_h)) vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}),$$

and plugging back into the last equation yields

$$\begin{aligned}
\left\| vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}) \right\|^2_{\Sigma^{-1}\otimes(V_{\boldsymbol{N}_t}+\xi I_h)} &\leq \left\| \Sigma^{-1/2} \otimes (V_{\boldsymbol{N}_t} + \xi I_h)^{1/2} vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}) \right\| \left(\xi \left\| vec(\boldsymbol{\theta}) \right\|_{W_t^{-1}} + \| S_t \|_{W_t^{-1}} \right), \\
&= \left\| vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}) \right\|_{\Sigma^{-1}\otimes(V_{N_n}+\xi I_h)} \left(\xi \left\| vec(\boldsymbol{\theta}) \right\|_{W_t^{-1}} + \| S_t \|_{W_t^{-1}} \right),
\end{aligned}$$

therefore

$$\left\| vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}) \right\|_{\Sigma^{-1}\otimes(V_{\boldsymbol{N}_t}+\xi I_h)} \leq \xi \left\| vec(\boldsymbol{\theta}) \right\|_{W_t^{-1}} + \| S_t \|_{W_t^{-1}},$$

with $W_t := \Sigma \otimes (V_{\boldsymbol{N}_t} + \xi I_h)$ and finally using the bound on (69), we prove that with probability at least $1 - \delta$, for all $t \geq 1$,

$$\left\| vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}) \right\|_{\Sigma^{-1}\otimes(V_{\boldsymbol{N}_t}+\xi I_h)} \leq \xi \left\| vec(\boldsymbol{\theta}) \right\|_{\Sigma^{-1}\otimes(V_{\boldsymbol{N}_t}+\xi I_h)^{-1}} + \sqrt{2 \log \left(\frac{det(V_{\boldsymbol{N}_t} + \xi I_h)^{d/2}}{\delta \xi^{d \cdot h/2}} \right)}, \tag{70}$$

that is

$$\frac{1}{2} \left\| vec(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}) \right\|^2_{\Sigma^{-1}\otimes(V_{\boldsymbol{N}_t}+\xi I_h)} \leq \left(\frac{\xi}{\sqrt{2}} \left\| vec(\boldsymbol{\theta}) \right\|_{\Sigma^{-1}\otimes(V_{\boldsymbol{N}_t}+\xi I_h)^{-1}} + \sqrt{\log \left(\frac{det(V_{\boldsymbol{N}_t} + \xi I_h)^{d/2}}{\delta \xi^{d \cdot h/2}} \right)} \right)^2. \tag{71}$$

The result then follows by simple algebra combined with the determinant-trace inequality (cf Lemma 19.4 of Lattimore and Szepesvári (2020)). □

H.2 Guarantees on the Sampling

The following result is well-known, and its proof can be found in Lattimore and Szepesvári (2020); Degenne et al. (2019).

Lemma 23 (Elliptic potential lemma). *For $V_0 = \xi I_h, \xi \geq 1$, it holds that*

$$\sum_{s=1}^{t} \| a_s \|^2_{V_{s-1}^{-1}} \leq 2h \log \left(\frac{Tr(V_0) + tL_{\mathcal{A}}^2}{d \det(V_0)^{1/d}} \right).$$

Lemma 24. *Let $\boldsymbol{w}_{exp}$ be a distribution on $\mathcal{A}$ supported on some atoms (not necessarily full support) and define the forced-exploration weights as $\widetilde{\boldsymbol{w}}_t := (1 - \gamma_t)\boldsymbol{w}_t + \gamma_t \boldsymbol{w}_{exp}$. For all $\alpha \in (0, 1/2)$, there exists $t_0(\alpha) \in \mathbb{N}$ and events $\Xi_{2,t}$ (cf (72)) such that for $t \geq t_0(\alpha)$, if $\Xi_{2,t}$ holds then $V_{\boldsymbol{N}_t} \geq \frac{t^{1-\alpha}}{2(1-\alpha)} V_{\boldsymbol{w}_{exp}}$. Moreover, if $V_{\boldsymbol{w}_{exp}}$ is non-singular then $V_{\boldsymbol{N}_t}^{-1} \leq 2(1-\alpha) t^{\alpha - 1} V_{\boldsymbol{w}_{exp}}^{-1}$.*

Proof. Let us introduce $D_{t-1}^a := N_{t,a} - \sum_{s=1}^{t-1} \widetilde{\boldsymbol{w}}_s(a) = \sum_{s=1}^{t-1} (\mathbb{1}_{(a_s = a)} - \widetilde{\boldsymbol{w}}_s(a))$ and the following event

$$\Xi_{2,t}^a := \left(|D_t| \leq \sqrt{2t \log(2|\mathcal{A}|t^2)} \right).$$

By simple algebra, D_t^a is $\mathcal{F}_t$ measurable and recalling that $\widetilde{\boldsymbol{w}}_t$ is $\mathcal{F}_{t-1}$ measurable (i.e., predictible), we have

$$\begin{aligned}
\mathbb{E}[D_t^a \mid \mathcal{F}_{t-1}] &= D_{t-1}^a + \mathbb{E}[\mathbb{1}_{(a_t = a)} - \widetilde{\boldsymbol{w}}_t(a) \mid \mathcal{F}_{t-1}] \\
&= 0,
\end{aligned}$$

so $(D_t^a)_{t \geq 1}$ is a $\mathcal{F}$ martingale which satisfies $|D_t^a - D_{t-1}^a| \leq 1$ almost surely. By Azuma's inequality, $\mathbb{P}_{\boldsymbol{\nu}}(\Xi_{2,t}^a) \geq 1 - \frac{1}{|\mathcal{A}|t^2}$ and defining

$$\Xi_{2,t} := \bigcap_{a \in \mathcal{A}} \Xi_{2,t}^a, \tag{72}$$

we have $\mathbb{P}_{\boldsymbol{\nu}}(\Xi_{2,t}) \geq 1 - 1/t^2$. When $\Xi_{2,t}$ holds, we have for any arm $a \in \mathcal{A}$,

$$\begin{aligned} N_{t,a} &\geq \sum_{s=1}^{t-1} \widetilde{\boldsymbol{w}}_s(a) - \sqrt{2t\log(2|\mathcal{A}|t^2)} \\ &\geq \sum_{s=1}^{t-1} \gamma_s \boldsymbol{w}_{\exp}(a) - \sqrt{2t\log(2|\mathcal{A}|t^2)}, \\ &\geq \frac{(t-1)^{1-\alpha}}{1-\alpha}\boldsymbol{w}_{\exp}(a) - \sqrt{2t\log(2|\mathcal{A}|t^2)}, \end{aligned}$$

so that introducing

$$t_0(\alpha) := \inf\left\{ n : \forall t \geq n, \forall a \in \mathcal{A} \mid \boldsymbol{w}_{\exp}(a) > 0, \frac{(t-1)^{1-\alpha}}{(1-\alpha)}\boldsymbol{w}_{\exp}(a) - \sqrt{2t\log(2|\mathcal{A}|t^2)} \geq \frac{t^{1-\alpha}}{2(1-\alpha)}\boldsymbol{w}_{\exp}(a) \right\}, \quad (73)$$

which is well defined for $\alpha \in (0, 1/2)$, we have for $t \geq t_0(\alpha)$, $N_{t,a} \geq \frac{t^{1-\alpha}}{2(1-\alpha)}\boldsymbol{w}_{\exp}(a)$, so that

$$\begin{aligned} V_{\boldsymbol{N}_t} &\geq \frac{t^{1-\alpha}}{2(1-\alpha)} \sum_{a\in\mathcal{A}} \boldsymbol{w}_{\exp}(a) a a^\intercal \\ &= \frac{t^{1-\alpha}}{2(1-\alpha)} V_{\boldsymbol{w}_{\exp}}, \end{aligned}$$

which is the claimed result, and the second part of the statement follows from Lemma 28. □

I TECHNICALITIES

Appendix I gathers existing and new technical results used for our proofs.

Lemma 25 gathers properties on the function $\overline{W}_{-1}$, which is used in the literature to obtain concentration results.

Lemma 25 (Jourdan et al. (2023)). *Let $\overline{W}_{-1}(x) := -W_{-1}(-e^{-x})$ for all $x \geq 1$, where W_{-1} is the negative branch of the Lambert W function. The function $\overline{W}_{-1}$ is increasing on $(1, +\infty)$ and strictly concave on $(1, +\infty)$. In particular, $\overline{W}'_{-1}(x) = \left(1 - \frac{1}{\overline{W}_{-1}(x)}\right)^{-1}$ for all $x > 1$. Then, for all $y \geq 1$ and $x \geq 1$, $\overline{W}_{-1}(y) \leq x$ if and only if $y \leq x - \ln(x)$. Moreover, for all $x > 1$, $x + \log(x) \leq \overline{W}_{-1}(x) \leq x + \log(x) + \min\left\{\frac{1}{2}, \frac{1}{\sqrt{x}}\right\}$.*

In general, $\mathrm{Alt}(\boldsymbol{\theta})$ is not convex, but the result below shows that it is countably convex i.e. union of convex sets, and their number depends on the size of the Pareto set, the dimension, and the number of arms.

Lemma 26. *For all parameter $\boldsymbol{\theta}$ letting $p := |S^*(\boldsymbol{\theta})|$, $\mathrm{Alt}(\boldsymbol{\theta})$ can be written as the union $(p(p-1) + (|\mathcal{Z}| - p)d^p)$ convex subsets.*

Proof. We let $E_{z,x}(c) = e_c \otimes (z - x)$ and $e_c = (\mathbb{1}(c' = c))_{c' \in [d]}$ for $z, x \in \mathcal{Z}$. To further ease notation, we let $\boldsymbol{E}_z = I_d \otimes z^\intercal$, and $\boldsymbol{E}_{z,x} = \boldsymbol{E}_z - \boldsymbol{E}_x$, which corresponds to row-wise stacking of vectors $(E_{z,x}(c))_c$. It yields $\boldsymbol{\theta}^\intercal z = \boldsymbol{E}_z vec(\boldsymbol{\theta})$ and $\boldsymbol{E}_z vec(\boldsymbol{\theta}) \prec \boldsymbol{E}_x vec(\boldsymbol{\theta})$ iff $\boldsymbol{E}_{z-x} vec(\boldsymbol{\theta}) \prec 0_d$.

To have a Pareto set different from S, either an arm of S should be made sub-optimal or an arm from $\mathcal{Z}\backslash S$ should be Pareto-optimal. We have

$$\begin{aligned} \boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\theta}) &\iff S^*(\boldsymbol{\lambda}) \neq S^*(\boldsymbol{\theta}) & (74)\\ &\iff \exists z \in S^*(\boldsymbol{\theta}) : z \notin S^*(\boldsymbol{\lambda}) \quad \text{or} \quad \exists z \in \mathcal{Z}\backslash S^*(\boldsymbol{\theta}) : z \in S^*(\boldsymbol{\lambda}) & (75)\\ &\iff (\exists z, x \in S^*(\boldsymbol{\theta}) : \boldsymbol{E}_{z-x} vec(\boldsymbol{\lambda}) \prec 0_d) \text{ or } (\exists z \in \mathcal{Z}\backslash S^*(\boldsymbol{\theta}) : \boldsymbol{E}_{z-x} vec(\boldsymbol{\lambda}) \not\prec 0_d \ \forall x \in S^*(\boldsymbol{\theta})) & (76) \end{aligned}$$

To see the last equivalence, assume $z \in S^*(\boldsymbol{\theta})\backslash S^*(\boldsymbol{\lambda})$, then there exists $x \in \mathcal{Z}$ such that $\boldsymbol{E}_z vec(\boldsymbol{\lambda}) \prec \boldsymbol{E}_x vec(\boldsymbol{\lambda})$, i.e. $\boldsymbol{E}_{z-x} vec(\boldsymbol{\lambda}) \prec 0_d$. If $x \in S^*(\boldsymbol{\theta})$ then (76) follows.

Further assume $x \notin S^*(\boldsymbol{\theta})$, then, either there exists $x' \in S^*(\boldsymbol{\theta})$ such that $\boldsymbol{E}_x vec(\boldsymbol{\lambda}) \prec \boldsymbol{E}_{x'} vec(\boldsymbol{\lambda})$ then by transitivity $\boldsymbol{E}_z vec(\boldsymbol{\lambda}) \prec \boldsymbol{E}_{x'} vec(\boldsymbol{\lambda})$ and $z, x' \in S^*(\boldsymbol{\theta})$ or for all $x' \in S^*(\boldsymbol{\theta})$, $\boldsymbol{E}_x vec(\boldsymbol{\lambda}) \not\prec \boldsymbol{E}_{x'} vec(\boldsymbol{\lambda})$. Put together we have

$(z \in S^*(\boldsymbol{\theta}) \backslash S^*(\boldsymbol{\lambda}))$ implies (76). Similarly, for $z \in S^*(\boldsymbol{\lambda}) \backslash S^*(\boldsymbol{\theta})$, it is direct by Pareto-optimality (in the instance $\boldsymbol{\lambda}$) that for all $x \in S^*(\boldsymbol{\theta})$, $\boldsymbol{E}_z vec(\boldsymbol{\lambda}) \not\prec \boldsymbol{E}_x vec(\boldsymbol{\lambda})$.

Reversely, if (76) holds, then in the case a), there exists $z \in S^*(\boldsymbol{\theta}) \backslash S^*(\boldsymbol{\lambda})$. If b) holds, then one cannot have $S^*(\boldsymbol{\theta}) = S^*(\boldsymbol{\lambda})$. Indeed if we had $S^*(\boldsymbol{\theta}) = S^*(\boldsymbol{\lambda})$, then it would follow that there exists $z \in \mathcal{Z} \backslash S^*(\boldsymbol{\theta}) = \mathcal{Z} \backslash S^*(\boldsymbol{\lambda})$ such that for all $z \in S^*(\boldsymbol{\lambda})$, $\boldsymbol{E}_z vec(\boldsymbol{\lambda}) \not\prec \boldsymbol{E}_z vec(\boldsymbol{\lambda})$, that in the instance $\boldsymbol{\lambda}$, there would be a sub-optimal arm that is not dominated by any other Pareto-optimal arm, which is not possible. This concludes the reverse inclusion. Therefore,

$$\mathrm{Alt}(\boldsymbol{\theta}) := \underbrace{\{\boldsymbol{\lambda} \mid z, x \in S^*(\boldsymbol{\theta}) : \boldsymbol{E}_{z-x} vec(\boldsymbol{\lambda}) \prec 0_d\}}_{A_{\boldsymbol{\theta}}} \cup \underbrace{\{\boldsymbol{\lambda} \mid \exists z \in \mathcal{Z} \backslash S^*(\boldsymbol{\theta}) : \forall x \in S^*(\boldsymbol{\theta}), \boldsymbol{E}_{z-x} vec(\boldsymbol{\lambda}) \not\prec 0_d\}}_{B_{\boldsymbol{\theta}}}, \tag{77}$$

then observe that

$$A_{\boldsymbol{\theta}} := \bigcup_{z \in S^*(\boldsymbol{\theta})} \bigcup_{x \in S^*(\boldsymbol{\theta}) \backslash \{z\}} \{\boldsymbol{\lambda} : \boldsymbol{E}_{z-x} vec(\boldsymbol{\lambda}) \prec 0_d\}, \tag{78}$$

and for z, x fixed, $\{\boldsymbol{\lambda} \mid \boldsymbol{E}_{z-x} vec(\boldsymbol{\lambda}) \prec 0_d\}$ is a convex set. On the other side, for z fixed, if $(\boldsymbol{E}_{z-x} vec(\boldsymbol{\lambda}) \not\prec 0_d, \forall x \in S^*(\boldsymbol{\theta}))$ holds then there exists $\bar{d}^z \in \{1, \dots, d\}^{S^*(\boldsymbol{\theta})}$ such that for all $x \in S^*(\boldsymbol{\theta})$, $\langle E_{z,x}(d^z(x)), vec(\boldsymbol{\lambda})\rangle \geq 0$. Putting these displays together yields

$$B_{\boldsymbol{\theta}} = \bigcup_{z \in \mathcal{Z} \backslash S^*(\boldsymbol{\theta})} \bigcup_{\bar{d}^z \in [d]^{S^*(\boldsymbol{\theta})}} \{\boldsymbol{\lambda} : \forall x \in S^*(\boldsymbol{\theta}) : \langle E_{z,x}(d^z(x)), vec(\boldsymbol{\lambda})\rangle \geq 0\}, \tag{79}$$

then remark that for $z, \bar{d}^z$ fixed, the set $\{\boldsymbol{\lambda} : \forall x \in S^*(\boldsymbol{\theta}) : \langle E_{z,x}(d^z(x)), vec(\boldsymbol{\lambda})\rangle \geq 0\}$ is convex. Finally, combining (79) with (78) yields the claimed result. □

The following lemma allows to upper bound the probability that a multi-variate normal vector belongs to a convex set. This is proven in Lu and Li (2009) where the authors attributed it to Kuelbs et al. (1994).

Lemma 27. *Let Σ be a variance-covariance matrix, $\mu \in \mathbb{R}^n$, let $X \sim \mathcal{N}(\mu, \Sigma)$ and $C \subset \mathbb{R}^n$, a convex set. Then, it holds that*

$$\mathbb{P}(X \in C) \leq \frac{1}{2} \exp\left(-\inf_{x \in C} \frac{1}{2} \|x - \mu\|^2_{\Sigma^{-1}}\right).$$

The followins lemma can be found in algebra books, we prove it for the sake of self-containedness. For two $n \times n$ matrices A, B we write $A \geq B$ iff $A - B$ is positive semi-definite i.e for all $x \in \mathbb{R}^n$, $x^\intercal (A - B) x \geq 0$.

Lemma 28. *Let A, B be two symmetric order n non-singular matrices. If $A - B \leq 0$ then $B^{-1} - A^{-1} \leq 0$.*

Proof. As A is symmetric, there exits P such that $A = P^2$ which is denoted by $A^{1/2} = P$. We have

$$\begin{aligned}
A - B \leq 0 \quad &\iff \quad \forall x \in \mathbb{R}^n, x^\intercal A x - x^\intercal B x \leq 0 \\
&\iff \quad \forall x \in \mathbb{R}^n, (A^{1/2}x)^\intercal (A^{1/2}x) - (A^{1/2}x)^\intercal A^{-1/2} B A^{-1/2} (A^{1/2}x) \leq 0 \\
&\iff \quad I_n - \underbrace{A^{-1/2} B A^{-1/2}}_{M} \leq 0 \quad (\text{as } A^{1/2} \text{ is invertible}) \\
&\iff \quad \forall x \in \mathbb{R}^n, (M^{1/2}x)^\intercal M^{-1} (M^{1/2}x) - (M^{1/2}x)^\intercal (M^{1/2}x) \leq 0 \quad (\text{as } M \text{ is symmetric and invertible}) \\
&\iff \quad M^{-1} - I_n \leq 0,
\end{aligned}$$

then by plugging in the expression of M, we have

$$\begin{aligned}
A - B \leq 0 \quad &\iff \quad (A^{-1/2} B A^{-1/2})^{-1/2})^{-1} - I_n \leq 0 \\
&\iff \quad \forall x \in \mathbb{R}^n, x^\intercal A^{1/2} B^{-1} A^{1/2} x - (A^{1/2}x)^\intercal A^{-1} (A^{1/2}x) \leq 0 \\
&\iff \quad B^{-1} - A^{-1} \leq 0,
\end{aligned}$$

where the last inequality follows again as $A^{1/2}$ is invertible which follows as A is symmetric non-singular. □

In the result below, we show that the Euclidean norm is α-exp concave over a bounded domain.

Lemma 29. *The function* $\boldsymbol{\lambda} \mapsto \exp\left(-\alpha \left\|vec(\boldsymbol{\theta}' - \boldsymbol{\lambda})\right\|^2_{\Sigma^{-1}\otimes aa^\intercal}\right)$ *defined a bounded domain* $\boldsymbol{D}$ *is concave for* $\alpha \in \left(0, \frac{1}{2\max_{\boldsymbol{\lambda}\in\boldsymbol{D}} \|vec(\boldsymbol{\theta}'-\boldsymbol{\lambda})\|^2_{\Sigma^{-1}\otimes aa^\intercal}}\right]$.

Proof. Let $\widetilde{V} = \Sigma^{-1} \otimes aa^\intercal$ and define $g(\lambda) := \exp(-\alpha\|\theta' - \lambda\|^2_{\Sigma^{-1}\otimes aa^\intercal})$ for $\lambda \in vec(\boldsymbol{D})$, with $\theta' := vec(\boldsymbol{\theta}')$. Direct calculation of the Hessian of g gives

$$\nabla^2 g(\lambda) = \left(-2\alpha\widetilde{V} + 4\alpha^2\widetilde{V}(\lambda - \theta')(\theta' - \theta')^\intercal\widetilde{V}\right) g(\lambda)$$

further, observe that,

$$\begin{aligned} u^T\left(-2\alpha\widetilde{V} + 4\alpha^2\widetilde{V}(\lambda - \theta')(\lambda - \theta')^\intercal\widetilde{V}\right) u &= -2\alpha\|u\|^2_{\widetilde{V}} + 4\alpha^2(u^\intercal\widetilde{V}(\lambda - \theta'))^2 \\ &\leq -2\alpha\|u\|^2_{\widetilde{V}} + 4\alpha^2\|u\|^2_{\widetilde{V}}\|\lambda - \theta'\|^2_{\widetilde{V}}, \end{aligned}$$

which proves the claimed statement as the Hessian is negative semi-definite for $\alpha \in (0, 1/(2\max_{\boldsymbol{\lambda}\in\boldsymbol{D}} \|vec(\boldsymbol{\theta}' - \boldsymbol{\lambda})\|^2_{\Sigma^{-1}\otimes aa^\intercal})]$. □

The following lemma expresses the difference in estimation of the empirical mean between two consecutive rounds.

Lemma 30. *Let* $\widetilde{V}_t := \Sigma^{-1} \otimes V_t$*, then, it holds that*

$$\widetilde{V}_t vec(\widehat{\boldsymbol{\theta}}_t - \widehat{\boldsymbol{\theta}}_{t-1}) = \left[\Sigma^{-1} \otimes a_t\right]\left(\varepsilon_t + [I_d \otimes a_t^\intercal]\, vec\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{t-1}\right)\right).$$

Proof. By Sherman-Morrison's formula, with $S_t = \sum_{s\in[t]} vec(a_s X_s^\intercal)$

$$\begin{aligned} vec(\widehat{\boldsymbol{\theta}}_t) &= \left[I_d \otimes (V_{t-1} + a_t a_t^\intercal)^{-1}\right] S_t \\ &= \left[I_d \otimes \left(V_{t-1}^{-1} - \frac{V_{t-1}^{-1}a_t a_t^\intercal V_{t-1}^{-1}}{1 + a_t^\intercal V_{t-1}^{-1}a_t}\right)\right] S_t \\ &= vec(\widehat{\boldsymbol{\theta}}_{t-1}) - \left[I_d \otimes \left(\frac{V_{t-1}^{-1}a_t a_t^\intercal V_{t-1}^{-1}}{1 + a_t^\intercal V_{t-1}^{-1}a_t}\right)\right] S_t + [I_d \otimes V_{t-1}^{-1}]vec(a_t X_t^\intercal) \\ &= vec(\widehat{\boldsymbol{\theta}}_{t-1}) - \frac{[I_d \otimes (V_{t-1}^{-1}a_t a_t^\intercal)][I_d \otimes V_{t-1}^{-1}]}{1 + a_t^\intercal V_{t-1}^{-1}a_t} S_t + [I_d \otimes V_{t-1}^{-1}]vec(a_t X_t^\intercal) \\ &= vec(\widehat{\boldsymbol{\theta}}_{t-1}) - \frac{[I_d \otimes (V_{t-1}^{-1}a_t a_t^\intercal)]vec(\widehat{\boldsymbol{\theta}}_{t-1}) + [I_d \otimes (V_{t-1}^{-1}a_t a_t^\intercal V_{t-1}^{-1})]vec(a_t X_t^\intercal)}{1 + a_t^\intercal V_{t-1}^{-1}a_t} + [I_d \otimes V_{t-1}^{-1}]vec(a_t X_t^\intercal) \end{aligned}$$

we recall that for two vector a, X,

$$vec(aX^\intercal) = [I_d \otimes a]X \quad \text{and} \quad [A \otimes B][C \otimes D] = (AC) \otimes (BD)$$

then

$$\begin{aligned} [I_d \otimes (V_{t-1}^{-1}a_t a_t^\intercal V_{t-1}^{-1})]vec(a_t X_t^\intercal) &= [I_d \otimes (V_{t-1}^{-1}a_t a_t^\intercal V_{t-1}^{-1})][I_d \otimes a_t]X_t \\ &= (a_t^\intercal V_{t-1}^{-1}a_t)[I_d \otimes (V_{t-1}^{-1}a_t)]X_t. \end{aligned}$$

Proceeding similarly yields

$$\begin{aligned} [I_d \otimes V_{t-1}^{-1}]vec(a_t X_t^\intercal) &= [I_d \otimes V_{t-1}^{-1}][I_d \otimes a_t]X_t \\ &= [I_d \otimes V_{t-1}^{-1}a_t]X_t, \end{aligned}$$

therefore,

$$vec(\widehat{\boldsymbol{\theta}}_t) = vec(\widehat{\boldsymbol{\theta}}_{t-1}) + \frac{[I_d \otimes (V_{t-1}^{-1}a_t)]X_t - [I_d \otimes (V_{t-1}^{-1}a_t a_t^\intercal)]vec(\widehat{\boldsymbol{\theta}}_{t-1})}{1 + a_t^\intercal V_{t-1}^{-1}a_t}.$$

By Kronecker product on the left, we have

$$\begin{aligned}[\Sigma^{-1}\otimes V_t]vec(\widehat{\boldsymbol{\theta}}_t-\widehat{\boldsymbol{\theta}}_{t-1}) &= \frac{[\Sigma^{-1}\otimes V_t][I_d\otimes V_{t-1}^{-1}a_t]X_t-[\Sigma^{-1}\otimes V_t][I_d\otimes (V_{t-1}^{-1}a_ta_t^\intercal)]vec(\widehat{\boldsymbol{\theta}}_{t-1})}{1+a_t^\intercal V_{t-1}^{-1}a_t}\\ &= \frac{[\Sigma^{-1}\otimes (V_tV_{t-1}^{-1}a_t)]X_t-[\Sigma^{-1}\otimes (V_tV_{t-1}^{-1}a_ta_t^\intercal)]vec(\widehat{\boldsymbol{\theta}}_{t-1})}{1+a_t^\intercal V_{t-1}^{-1}a_t}.\end{aligned}$$

Then observe that

$$\begin{aligned}[\Sigma^{-1}\otimes (V_tV_{t-1}^{-1}a_t)]X_t &= [\Sigma^{-1}\otimes ((I_h+a_ta_t^\intercal V_{t-1}^{-1})a_t)]X_t\\ &= \left([\Sigma^{-1}\otimes a_t]+(a_t^\intercal V_{t-1}^{-1}a_t)[\Sigma^{-1}\otimes a_t]\right)X_t\\ &= (1+a_t^\intercal V_{t-1}^{-1}a_t)[\Sigma^{-1}\otimes a_t]X_t,\end{aligned}$$

and similar developments yields

$$\begin{aligned}[\Sigma^{-1}\otimes (V_tV_{t-1}^{-1}a_ta_t^\intercal)]vec(\widehat{\boldsymbol{\theta}}_{t-1}) &= [\Sigma^{-1}\otimes ((I_h+a_ta_t^\intercal V_{t-1}^{-1})a_ta_t^\intercal)]vec(\widehat{\boldsymbol{\theta}}_{t-1})\\ &= \left([\Sigma^{-1}\otimes (a_ta_t^\intercal)]+(a_t^\intercal V_{t-1}^{-1}a_t)[\Sigma^{-1}\otimes (a_ta_t^\intercal)]\right)vec(\widehat{\boldsymbol{\theta}}_{t-1})\\ &= (1+a_t^\intercal V_{t-1}^{-1}a_t)[\Sigma^{-1}\otimes (a_ta_t^\intercal)]vec(\widehat{\boldsymbol{\theta}}_{t-1})\end{aligned}$$

so that we finally have

$$\begin{aligned}\widetilde{V}_tvec(\widehat{\boldsymbol{\theta}}_t-\widehat{\boldsymbol{\theta}}_{t-1}) &= [\Sigma^{-1}\otimes V_t]vec(\widehat{\boldsymbol{\theta}}_t-\widehat{\boldsymbol{\theta}}_{t-1})\\ &= [\Sigma^{-1}\otimes a_t]X_t-[\Sigma^{-1}\otimes (a_ta_t^\intercal)]vec(\widehat{\boldsymbol{\theta}}_{t-1})\\ &= [\Sigma^{-1}\otimes a_t]X_t-\left([\Sigma^{-1}\otimes a_t][I_d\otimes a_t^\intercal]\right)vec(\widehat{\boldsymbol{\theta}}_{t-1})\\ &= [\Sigma^{-1}\otimes a_t](X_t-(I_d\otimes a_t^\intercal)vec(\widehat{\boldsymbol{\theta}}_{t-1})).\end{aligned}$$

We recall that we have $X_t := (I_d\otimes a_t^\intercal)vec(\boldsymbol{\theta})+\varepsilon_t$ from the linear model. Thus, all put together, we recover the claimed result. □

Lemma 9. *Let $\boldsymbol{w}\in\mathbb{R}_+^K$. For any $\boldsymbol{\mu}' := (\mu_1'\dots\mu_K')^\intercal\in\mathbb{R}^{K\times d}$, the following statement holds in the unstructured setting*

$$\inf_{\boldsymbol{\lambda}\in\mathrm{Alt}(S^\star(\boldsymbol{\mu}'))}\|vec(\boldsymbol{\lambda}-\boldsymbol{\mu}')\|^2_{\Sigma^{-1}\otimes\mathrm{diag}(\boldsymbol{w})} = \inf_{\lambda\in\mathrm{Alt}(S^\star(\boldsymbol{\mu}'))\cap\{\boldsymbol{\lambda}\mid\max_i\|\mu_i'-\lambda_i\|_{\Sigma^{-1}}<\epsilon\}}\|vec(\boldsymbol{\lambda}-\boldsymbol{\mu}')\|^2_{\Sigma^{-1}\otimes\mathrm{diag}(\boldsymbol{w})}\,, \tag{16}$$

where $\epsilon := \max(2\max_{i\notin S}\max_{j\in S}\|\mu_i'-\mu_j'\|_{\Sigma^{-1}}, \max_{i,j\in S^2}\|\mu_i'-\mu_j'\|_{\Sigma^{-1}})$, and $S=S^\star(\boldsymbol{\mu}')$.

This result shows that the best response always exists in the unstructured setting as an inf over a compact subset of the alternative. Although its closed form is unknown in PSI, we show that it belongs to a ball centered at $\boldsymbol{\mu}'$ and whose radius depends also on $\boldsymbol{\mu}'$.

Proof. The idea of the proof is to show that when there is a ball $\mathbb{B}$ such that if an alternative parameter of $\boldsymbol{\mu}'$ does not belong to $\mathbb{B}$, then there is a parameter in $\mathrm{Alt}(S^\star(\boldsymbol{\mu}'))\cap\mathbb{B}$ for which the transportation cost will be smaller. Let $\boldsymbol{\mu}' := (\mu_1'\dots\mu_K')^\intercal$, where μ_i' denotes the vector mean of arm i and we let $\boldsymbol{\lambda} := (\lambda_1\dots\lambda_K)^\intercal$ where $\lambda_i\in\mathbb{R}^d$. To ease notation, we let $S=S^\star(\boldsymbol{\mu}')$. Introducing

$$\mathrm{Alt}^-(S) := \bigcup_{i,j\in S^2:i\neq j} W_{i,j}\quad\text{and}\quad\mathrm{Alt}^+(S) := \bigcup_{i\in S^c} V_i, \tag{80}$$

where $\Theta := \mathbb{R}^{K\times d}$, we define

$$W_{i,j} := \{\boldsymbol{\lambda} := (\lambda_1\dots\lambda_K)^\intercal\in\Theta\mid\lambda_i\leq\lambda_j\}\text{ and }V_i := \{\boldsymbol{\lambda} := (\lambda_1\dots\lambda_K)^\intercal\in\Theta\mid\exists\, i\in(S)^c:\forall\, j\in S,\lambda_i\not\prec\lambda_j\}\,.$$

By simply expanding the expression, we have in the unstructured case,

$$D(\boldsymbol{w},\boldsymbol{\lambda};\boldsymbol{\mu}') := \|vec(\boldsymbol{\lambda}-\boldsymbol{\mu}')\|^2_{\Sigma^{-1}\otimes\mathrm{diag}(\boldsymbol{w})} = \sum_{i=1}^K\boldsymbol{w}_i\|\lambda_i-\mu_i'\|^2_{\Sigma^{-1}}.$$

Let $(i,j) \in S^2$ be fixed and $\boldsymbol{\lambda} \in W_{i,j}$. Let $\alpha_{i,j} = \|\mu_i' - \mu_j'\|_{\Sigma^{-1}}$. If $\|\lambda_i - \mu_i'\|_{\Sigma^{-1}} > \alpha_{i,j}$ then define the instance $\widetilde{\boldsymbol{\lambda}}$ as

$$\widetilde{\lambda}_k := \begin{cases} \mu_k' & \text{if } k \notin \{i,j\}, \\ \mu_j' & \text{else }, \end{cases}$$

which satisfies $\widetilde{\lambda}_i \leq \widetilde{\lambda}_j$, so $\widetilde{\boldsymbol{\lambda}} \in W_{i,j}$, and

$$\begin{aligned} D(\boldsymbol{w}, \widetilde{\boldsymbol{\lambda}}; \boldsymbol{\mu}') &= \boldsymbol{w}_i \|\mu_i' - \mu_j'\|^2_{\Sigma^{-1}} + \sum_{k \notin \{i,j\}} \boldsymbol{w}_k \|\lambda_k - \mu_k'\|^2_{\Sigma^{-1}}, \\ &= \boldsymbol{w}_i \|\mu_i' - \mu_j'\|^2_{\Sigma^{-1}} < \boldsymbol{w}_i \|\lambda_i - \mu_i'\|^2_{\Sigma^{-1}} \\ &< D(\boldsymbol{w}, \boldsymbol{\lambda}, \boldsymbol{\mu}'), \end{aligned}$$

and further observe that $\max_{k \in [K]} \|\widetilde{\lambda}_k - \mu_k'\|_{\Sigma^{-1}} \leq \alpha_{i,j}$. We proceed similarly if $\|\lambda_j - \mu_j'\|_{\Sigma^{-1}} > \alpha_{i,j}$, by defining

$$\widetilde{\lambda}_k := \begin{cases} \mu_k' & \text{if } k \notin \{i,j\}, \\ \mu_i' & \text{else }, \end{cases}$$

and the same conclusion follows. So we have proved that there exists $\widetilde{\boldsymbol{\lambda}} \in W_{i,j}$ with $\max_{k \in [K]} \|\widetilde{\lambda}_k - \mu_k'\|_{\Sigma^{-1}} \leq \alpha_{i,j}$ and for which the transportation cost is not larger than that of $\boldsymbol{\lambda}$. We prove a similar property for $\mathrm{Alt}^+(S)$.

Now, fix $i \notin S$ and take $\boldsymbol{\lambda} \in V_i$. Let $b_i = \max_{k \in S} \|\mu_i' - \mu_k'\|_{\Sigma^{-1}}$. If $\|\lambda_i - \mu_i'\|_{\Sigma^{-1}} > b_i$ then it suffices to define $\widetilde{\boldsymbol{\lambda}}$ as

$$\widetilde{\lambda}_p := \begin{cases} \mu_p' & \text{if } p \neq i, \\ \mu_{\widetilde{i}}' & \text{else }, \end{cases}$$

for $\widetilde{i} \in S$, to ensure that $\widetilde{\boldsymbol{\lambda}} \in V_i$ and

$$\begin{aligned} D(\boldsymbol{w}, \widetilde{\boldsymbol{\lambda}}; \boldsymbol{\mu}') &= \boldsymbol{w}_i \|\mu_{\widetilde{i}}' - \mu_i'\|^2_{\Sigma^{-1}} + \sum_{k \neq i} \boldsymbol{w}_k \|\lambda_k - \mu_k'\|^2_{\Sigma^{-1}} \\ &\leq \boldsymbol{w}_i b_i^2 \\ &< \boldsymbol{w}_i \|\lambda_i - \mu_i'\|^2_{\Sigma^{-1}} \\ &< D(\boldsymbol{w}, \boldsymbol{\lambda}; \boldsymbol{\mu}'), \end{aligned}$$

where the second line also uses the fact that $\widetilde{i} \in S$. Assume $\|\lambda_i - \mu_i'\|_{\Sigma^{-1}} < b_i$ and that $H_i := \{k \in S : \|\lambda_k - \mu_k'\|_{\Sigma^{-1}} > 2b_i\}$ is non-empty. Let us define the instance $\widetilde{\boldsymbol{\lambda}}$:

$$\widetilde{\lambda}_k := \begin{cases} \lambda_i & \text{if } k \in H_i, \\ \lambda_k & \text{if } k \in (S \cup \{i\}) \backslash H_i, \\ \mu_k' & \text{else.} \end{cases}$$

Since $\boldsymbol{\lambda} \in V_i$, $\widetilde{\boldsymbol{\lambda}}$ as defined above satisfies $\widetilde{\lambda}_i \not\prec \widetilde{\lambda}_k, k \in S$, that is $\widetilde{\boldsymbol{\lambda}} \in V_i$ and

$$\begin{aligned} D(\boldsymbol{w}, \widetilde{\boldsymbol{\lambda}}; \boldsymbol{\mu}') &= \sum_{k \in (S \cup \{i\})} \boldsymbol{w}_k \|\widetilde{\lambda}_k - \mu_k'\|^2_{\Sigma^{-1}} \\ &= \sum_{k \in (S \cup \{i\}) \backslash H_i} \boldsymbol{w}_k \|\lambda_k - \mu_k\|^2_{\Sigma^{-1}} + \sum_{k \in H_i} \boldsymbol{w}_k \|\lambda_i - \mu_k'\|^2_{\Sigma^{-1}} \\ &< \sum_{k \in (S \cup \{i\}) \backslash H_i} \boldsymbol{w}_k \|\lambda_k - \mu_k'\|^2_{\Sigma^{-1}} + \sum_{k \in H_i} \boldsymbol{w}_k 4 b_i^2, \end{aligned}$$

which follows since $\|\lambda_i - \mu_k'\|_{\Sigma^{-1}} \leq \|\lambda_i - \mu_i'\|_{\Sigma^{-1}} + \|\mu_i' - \mu_k'\|_{\Sigma^{-1}} \leq 2b_i$. Recalling that for $k \in H_i$,

$$4b_i < \|\lambda_k - \mu_k'\|^2_{\Sigma^{-1}},$$

it follows that

$$D(\boldsymbol{w}, \widetilde{\boldsymbol{\lambda}}; \boldsymbol{\mu}') < D(\boldsymbol{w}, \boldsymbol{\lambda}; \boldsymbol{\mu}').$$

Put together, we have proved that for all $\boldsymbol{\lambda} \in V_i$, there exists $\widetilde{\boldsymbol{\lambda}} \in V_i$ whose transportation cost is not larger than that of $\boldsymbol{\lambda}$ and which additionally satisfies: $\max_{k\in[K]} \|\widetilde{\lambda}_k - \mu'_k\|_{\Sigma^{-1}} \leq 2b_i$.

To conclude, let us define $\epsilon := \max(2\max_{i\notin S} b_i, \max_{i,j\in S^2} \alpha_{i,j})$. Using what precedes, we have proved that

$$\inf_{\boldsymbol{\lambda}\in \mathrm{Alt}(S)\cap\{\boldsymbol{\lambda}:\max_k \|\lambda_k - \mu'_k\|_{\Sigma^{-1}} < \epsilon\}} D(\boldsymbol{w}, \boldsymbol{\lambda}; \boldsymbol{\mu}') \leq \inf_{\boldsymbol{\lambda}\in\mathrm{Alt}(S)} D(\boldsymbol{w}, \boldsymbol{\lambda}; \boldsymbol{\mu}'),$$

which proves the claimed result as we have $\mathrm{Alt}(S) \cap \{\boldsymbol{\lambda} : \max_k \|\lambda_k - \mu'_k\|_{\Sigma^{-1}} < \epsilon\} \subset \mathrm{Alt}(S)$. □

J IMPLEMENTATION DETAILS AND ADDITIONAL EXPERIMENTS

After presenting the implementation details and computational cost in Appendix J.1, we display supplementary experiments in Appendix J.3. The datasets are described in Appendix J.2.

J.1 Implementation Details and Complexity

J.1.1 Setup

The algorithm is implemented mainly in standard `C++17`. Our complete code is available on a git repository at `https://github.com/cyrille-kone/psips` with instructions to compile and run. We run our algorithms mainly on a personal laptop with 8GB RAM, 256GB SSD, and an Apple M2 CPU. The code is compiled with the `GCC13`. In the experiments, we include the "Oracle" algorithm, which computes the optimal weights of the underlying bandit instance. At each round, the oracle pulls an arm according to this optimal weights vector. Another algorithm included in our benchmark is the APE algorithm of Kone et al. (2023), which is an LUCB-type algorithm for PSI. GAPSI denotes the gradient-based algorithm of Crepon et al. (2024) for PSI.

In the following paragraphs, we analyze the time and memory complexity of the different parts of our algorithms, and we discuss their total time and complexity.

J.1.2 Time and memory complexity

Compute Pareto Set When $d = 2$, it is known that the Pareto set can be computed in an average $\mathcal{O}(|\mathcal{Z}|\log(|\mathcal{Z}|))$ time complexity and a worst-case $\mathcal{O}(|\mathcal{Z}|\log(|\mathcal{Z}|) + |\mathcal{Z}||p|)$ where p is the size of the Pareto set. For a general dimension $d > 3$, the algorithm of Kung et al. (1975) achieves a time complexity of $\mathcal{O}(|\mathcal{Z}|\log(|\mathcal{Z}|)^{d-2})$. Their algorithm is based on divide and conquer and in the case $d = 2$, it consists in sorting the arms along one coordinate, then by traversing the sorted items in decreasing order (for sorted dimension) each element is added to the set S if it is note dominated by an element of S. The resulting set is then the Pareto set, and combining the initial sorting and the construction of the Pareto set, the worst case complexity will be $\mathcal{O}(|\mathcal{Z}|\log(|\mathcal{Z}|) + |\mathcal{Z}|p + p(1-p)/2)$, p is the size of the Pareto set. The space complexity is $O(|\mathcal{Z}|)$.

To check if two parameters have the same Pareto set, in the worst case, one can compute their Pareto sets using the algorithms described earlier. However, it is possible (and more efficient) to check this condition without computing the Pareto sets of the parameters by simply checking from the definition of Alt.

Update $\widehat{\boldsymbol{\theta}}_{t+1}$ To compute $\widehat{\boldsymbol{\theta}}_{t+1}$ in the linear setting, we used Sherman-Morrison formula to avoid computing the inverse of V_t. Indeed as $V_t = V_{t-1} + a_t a_t^\intercal$ we have by Sherman-Morrison formula

$$V_t^{-1} = V_{t-1}^{-1} - \frac{V_{t-1}^{-1}a_t(V_{t-1}^{-1}a_t)^\intercal}{1 + \|a_t\|^2_{V_{t-1}^{-1}}},$$

which can be computed in time $\mathcal{O}(h^2)$. The computation of $\boldsymbol{Z}_{t+1}$ is done in time $\mathcal{O}(hd)$, so that the total cost of computing the least-squares estimate is dominated by the matrix product $V_{t+1}^{-1}\boldsymbol{Z}_{t+1}$, which can be done naively in $\mathcal{O}(h^2 d)$.

Generate Random Samples To pull from Π_t we generate samples from $\mathcal{N}(0_{dh}, I_{dh})$ (whose average cost is $\mathcal{O}(dh)$) and we compute a Cholesky decomposition of the $\Sigma \otimes V_t^{-1}$ which we do by computing the Cholesky of Σ and that of V_t^{-1}. Note that in the unstructured setting, the Cholesky of V_t^{-1} is trivial to compute. In summary, in the worst case, the total time complexity of this step is $\mathcal{O}(\max(d^2h^2, h^3))$ in the linear setting and $\mathcal{O}(d^2h)$ in the unstructured setting.

Miscellaneous For the other parts of the algorithm, the cost of a step of AdaHedge (De Rooij et al., 2014) is $\mathcal{O}(K)$, same for the memory complexity. When Θ is defined as a ball, the cost of the verification $\boldsymbol{\lambda} \in \Theta$ is simply the cost of computing the norm of Θ. For Euclidean norms, it will be $\mathcal{O}(dh)$. The computation of $M(t, \delta)$ is $\mathcal{O}(1)$, as we only need to compute $(q(\delta, i)_{i\in[K]})$ once at initialization.

Summary Combining the displays above, the time complexity of our algorithm at each step is

$$\mathcal{O}(|\mathcal{Z}| \log(|\mathcal{Z}|)^{\max(1,d-2)}) + \mathcal{O}(\max(d^2h^2, h^3)).$$

In the unstructured setting, it can be reduced to

$$\mathcal{O}(\max(K^2 d, K \log(K)^{\max(1,d-2)})).$$

J.2 Datasets

COV-BOOST This dataset has been publicly released in Munro et al. (2021) based on a phase 2 booster trial for Covid19 vaccine. Kone et al. (2023) further processed the data and extracted the parameters of a bandit instance with 20 arms and $d = 3$. We use the values tabulated by the authors to simulate our algorithms; see Tables 3 and 4.

Table 3: Table of the empirical arithmetic mean of the log-transformed immune response for three immunogenicity indicators. Each acronym corresponds to a vaccine. There are two groups of arms corresponding to the first 2 doses: one with prime BNT/BNT (BNT as first and second dose) and the second with prime ChAd/ChAd (ChAd as first and second dose). Each row in the table gives the values of the 3 immune responses for an arm (i.e., a combination of three doses).

Dose 1/Dose 2	Dose 3 (booster)	Immune response		
		Anti-spike IgG	NT_{50}	cellular response
Prime BNT/BNT	ChAd	9.50	6.86	4.56
	NVX	9.29	6.64	4.04
	NVX Half	9.05	6.41	3.56
	BNT	10.21	7.49	4.43
	BNT Half	10.05	7.20	4.36
	VLA	8.34	5.67	3.51
	VLA Half	8.22	5.46	3.64
	Ad26	9.75	7.27	4.71
	m1273	10.43	7.61	4.72
	CVn	8.94	6.19	3.84
Prime ChAd/ChAd	ChAd	7.81	5.26	3.97
	NVX	8.85	6.59	4.73
	NVX Half	8.44	6.15	4.59
	BNT	9.93	7.39	4.75
	BNT Half	8.71	7.20	4.91
	VLA	7.51	5.31	3.96
	VLA Half	7.27	4.99	4.02
	Ad26	8.62	6.33	4.66
	m1273	10.35	7.77	5.00
	CVn	8.29	5.92	3.87

Network on Chip We use the dataset studied by Almer et al. (2011) and publicly released by Zuluaga et al. (2013) and available at `http://www.spiral.net/software/pal.html`. We apply further preprocessing by normalizing the

Table 4: Pooled variance of each group.

	Immune response		
	Anti-spike IgG	NT_{50}	cellular response
Pooled sample variance	0.70	0.83	1.54

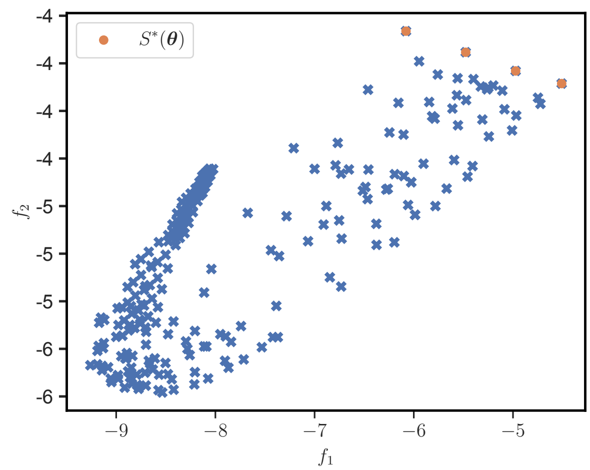

Figure 7: Means of each arm in the NoC instance with normalized features.

features to extract a linear instance with parameter $\boldsymbol{\theta}^*$ given below.

$$\boldsymbol{\theta}^* = \begin{pmatrix} -3.08665453 & -3.35487744 \\ -3.66027623 & 0.19333635 \\ -2.68963781 & -1.39779755 \\ -7.90670356 & -4.44360318 \end{pmatrix}.$$

In Figure 7 we plot the resulting PSI instance.

Reproducibility We provide below the parameters used in the experiments reported in the main paper. As explained in the experiment section, since their calibration relies on overly conservative union bounds, we used $M(t, \delta) = \frac{1}{\delta}\log(t/\delta)$ and $c(t, \delta) = 1 + \frac{\log(\log(t))}{\log(1/\delta)}$ for which we noticed a negligible empirical error for $\delta \leq 0.1$ in our experiments. For AdaHedge, we used the implementation of De Rooij et al. (2014), translating the pseudo-code provided in their paper. In the supplementary material, we provide our `C++` implementation together with the scripts to un each experiment. The complete scripts to reproduce our experiments are available at `https://github.com/cyrille-kone/psips`, with the instructions to compile and run the algorithms.

J.3 Supplementary Experiments

We provide additional experimental results. In particular, we conduct experiments to evaluate the performance of our algorithm in more complex scenarios. These experiments include instances in higher dimension d or with a larger number of arms K. Moreover, we also report the probability of error in the covid19 experiment. Each experiment is repeated on 100 independent runs, and the quantities reported are averaged statistics. Due to its high runtime, we omit GAPSI of Crepon et al. (2024) from these experiments.

Higher Dimension d In this experiment, we evaluate our algorithm in the unstructured setting on 100 random Gaussian instances with higher dimension $d \in \{3, 4, 5, 6\}$, $K = 10$ and $\Sigma = I_d/2$. Each arm is drawn from $\mathcal{U}([-1, 1]^d)$, and to have reasonable runtime, we reject instances whose complexity $H(\boldsymbol{\theta})$ is larger than 500. The average Pareto set size was respectively $4.42, 6.78, 8.57, 9.21$ in dimensions $d \in \{3, 4, 5, 6\}$. Using $\delta = 0.01$, we observe a negligible empirical error.

In Figure 8, we see that PSIPS has competitive empirical performance for higher dimension d. It consistently outperforms APE and performs on par with uniform sampling. The good performance of uniform sampling in this experiment can be

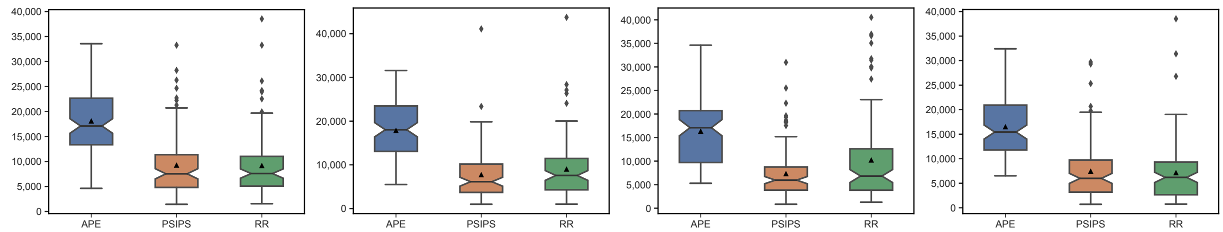

Figure 8: Empirical stopping time on 10-armed random Gaussian instances in dimension $d \in \{3, 4, 5, 6\}$ (left to right).

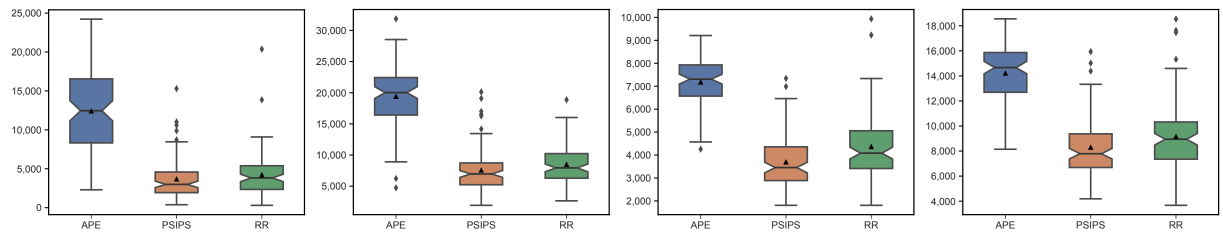

Figure 9: Empirical stopping time on 2-dimensional random Bernoulli instances with $K \in \{10, 15, 25, 40\}$ arms (left to right).

attributed to the large size of the Pareto set, which contains most of the arms. We conjecture that $\boldsymbol{w}^\star(\boldsymbol{\theta})$ is close to the uniform allocation on those instances.

Large Number of Arms K We benchmark the algorithms on instances with a larger number of arms. We report the results on 100 random Bernoulli instances with $K \in \{10, 15, 25, 40\}$ arms in dimension $d = 2$. The means of the instances are drawn from $\mathcal{U}([0.2, 0.9]^{K \times d})$, and we use $\Sigma = I_2/4$, and we reject instances whose complexity $H(\boldsymbol{\theta})$ is larger than 500. The average Pareto set size was respectively 1.73, 1.64, 1.37, 1.28 for the number of arms $K \in \{10, 15, 25, 40\}$. Using $\delta = 0.01$, we observe a negligible empirical error.

Figure 9 shows that PSIPS has competitive empirical performance for higher dimension d. It significantly outperforms APE and slightly outperforms uniform sampling. The worsening of uniform sampling's performance stems from the smaller size of the Pareto set, which almost contains only one arm.

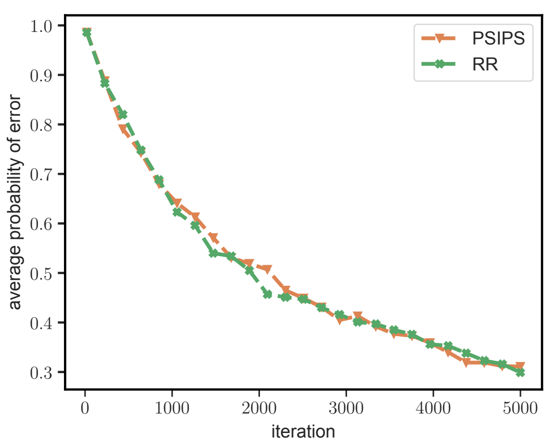

Figure 10: Average probability of misidentification of the Pareto set in the covid19 experiment. The results are averaged over 1000 runs.

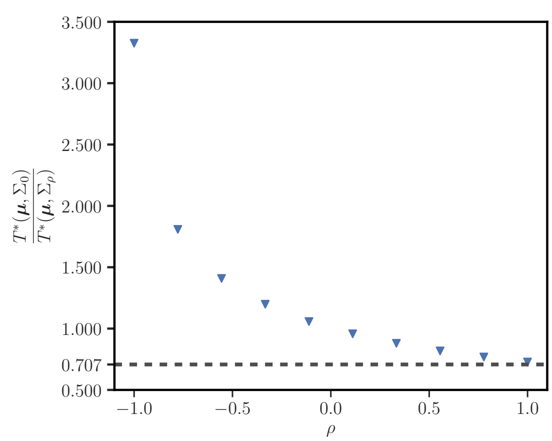

Figure 11: Theoretical impact of the correlation coefficient on the sample complexity for a 5-armed instance with $d = 2$.

Probability of Error Both the recommendation rule and the sampling rule of PSIPS are δ-independent and defined as any time t. Therefore, by disabling the stopping rule, it is possible to run PSIPS for $T = 5000$ time steps and report the empirical error $\mathbb{1}_{(\widehat{S}_t \neq S^\star)}$ at any time $t \in [K, 5000]$ and averaged over 1000 runs. Uniform sampling enjoys the same anytime property and has the same recommendation rule. Therefore, we use it as a benchmark.

Figure 10 reveals that the sampling rule from PSIPS yields an empirical error which is close to the one achieved by uniform sampling. Even though PSIPS has only theoretical guarantees in the fixed-confidence setting, it performs empirically well in the anytime setting.

Toward Understanding The Theoretical Impact of Correlation As a step toward understanding the theoretical effect of correlation on the sample complexity, we picked an instance defined by $\mu_1 := (0.73, 1.20)^T, \mu_2 := (0.45, -0.63)^T, \mu_3 := (0.63, 1.28)^T, \mu_4 := (0.94, 2.31)^T, \mu_5 := (2.08, 1.48)^T$ with a correlation matrix Σ_ρ: unit diagonal and an off-diagonal term $\rho \in (-1, 1)$ (correlation coefficient as in the setup of Section 4). Using an approximate convex solver and the exponentiated gradient algorithm, we approximate the theoretical complexity $T^*(\boldsymbol{\mu}, \Sigma_\rho)$ (as outlined in Lemma 1), which we report in Figure 11 w.r.t. ρ and $\rho = 0$ (uncorrelated setting) is used as reference. In this instance, we note that negative correlation decreases the sample complexity (w.r.t. the same instance with uncorrected objectives).