Covariance-Adaptive Least-Squares Algorithm for Stochastic Combinatorial Semi-Bandits

Julien Zhou, Pierre Gaillard, Thibaud Rahier, Houssam Zenati, Julyan Arbel

Covariance-Adaptive Least-Squares Algorithm for Stochastic Combinatorial Semi-Bandits

Julien Zhou*, [1, 2], Pierre Gaillard[2], Thibaud Rahier[1], Houssam Zenati[3], and Julyan Arbel[2]

[1]Criteo AI Lab, Paris, France
[2]Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France
[3]Inria, Saclay, France

Abstract

We address the problem of stochastic combinatorial semi-bandits, where a player can select from P subsets of a set containing d base items. Most existing algorithms (e.g. `CUCB`, `ESCB`, `OLS-UCB`) require prior knowledge on the reward distribution, like an upper bound on a sub-Gaussian proxy-variance, which is hard to estimate tightly. In this work, we design a variance-adaptive version of `OLS-UCB`, relying on an online estimation of the covariance structure. Estimating the coefficients of a covariance matrix is much more manageable in practical settings and results in improved regret upper bounds compared to proxy-variance-based algorithms. When covariance coefficients are all non-negative, we show that our approach efficiently leverages the semi-bandit feedback and provably outperforms bandit feedback approaches, not only in exponential regimes where $P \gg d$ but also when $P \leq d$, which is not straightforward from most existing analyses.

Keywords— Covariance-Adaptive; Bandits; Stochastic Combinatorial Semi-Bandit; Confidence Ellipsoid

1 Introduction

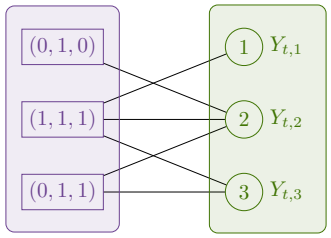
In sequential decision-making, the bandit framework has been extensively studied and was instrumental to several applications, e.g. A/B testing (Guo et al., 2020), online advertising and recommendation services (Zeng et al., 2016), network rooting (Tabei et al., 2023), demand-side management (Brégère et al., 2019), etc. Its popularity stems from its relative simplicity, allowing it to model and analyze a wide range of challenging real-world settings. Reference books like Bubeck and Cesa-Bianchi (2012) or Lattimore and Szepesvári (2020) offer a wide perspective on the subject.

In this framework, a *decision-maker* or *player* must make choices and receives associated rewards, but it lacks prior knowledge of its environment. This naturally leads to an exploration-exploitation trade-off: the player must explore different actions to determine the best one, but an inefficient exploration strategy harms the cumulative rewards. Efficient algorithms rely on exploiting the environment's structure, such as estimating parameters of a reward function rather than exploring every action.

In this paper, we focus on the stochastic combinatorial semi-bandit framework. In this setting, the player chooses a subset of *base items* and receives a feedback for each item chosen. The corresponding action set is included in the base items' power set, and can therefore be exponentially big and difficult to explore. However, it exhibits some structure that can be leveraged. The information collected by choosing different intersecting subsets can be shared, but the way to do it efficiently over time remains a challenging problem.

*julien.zhou@inria.fr
Preprint version, under review.

Problem formulation. We consider a set of $d \in \mathbb{N}^*$ *base items*, each item $i \in [d] = \{1, \dots, d\}$ yielding stochastic rewards. A *player*, the decision-maker, accesses these rewards through a set $\mathcal{A} \subseteq \{0,1\}^d$ of $P = |\mathcal{A}| \in \mathbb{N}^*$ *actions*, each corresponding to a subset of items.[1] We refer to actions $a \in \mathcal{A}$ using their components vectors $a = (a_i)_{i \in [d]} \in \{0,1\}^d$ where for all $j \in [d]$, $a_j = 1$ if and only if action a contains base item j (see Figure 1).



Actions $\mathcal{A} \subseteq \{0,1\}^d$ — Items $\{1, \dots, d\}$
Rewards $Y_t = (Y_{t,1}, \dots, Y_{t,d})$

Figure 1: Stochastic combinatorial semi-bandit setting where $d = 3$, $\mathcal{A} = \{(0,1,0), (1,1,1), (0,1,1)\}$, and $P = |\mathcal{A}| = 3$.

The player interacts with an *environment* over a sequence of $T \in \mathbb{N}^*$ *rounds*. At each round $t \in [T]$, the player chooses an action $A_t \in \mathcal{A}$, the environment samples a reward vector $Y_t \in \mathbb{R}^d$, the decision-maker observes the realization for every item contained in A_t, and receives their sum. The interactions between the player and the environment are summarized in Framework 1.

The objective of the decision-maker is to maximize the cumulative expected rewards, or equivalently, to minimize the expected cumulative regret defined as:

$$\mathbb{E}[R_T] = T\langle a^*, \mu \rangle - \sum_{t=1}^{T} \mathbb{E}\big[\langle A_t, Y_t \rangle\big] = \sum_{t=1}^{T} \mathbb{E}\big[\Delta_{A_t}\big]\,, \tag{1}$$

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $\mathbb{R}^d$, $a^* \in \arg\max_{a \in \mathcal{A}} \langle a, \mu \rangle$ is an optimal action, and $\Delta_a = \langle a^* - a, \mu \rangle$ is the *sub-optimality gap* for $a \in \mathcal{A}$.

Framework 1 Stochastic Combinatorial Semi-Bandit

For each $t \in \{1, \dots, T\}$:
- The player chooses an action $A_t \in \mathcal{A}$.
- The environment samples a vector of rewards $Y_t \in \mathbb{R}^d$ from a fixed unknown distribution.
- The player receives the reward $\langle A_t, Y_t \rangle = \sum_i A_{t,i} Y_{t,i}$.
- The player observes $Y_{t,i}$ for all $i \in [d]$ s.t. $A_{t,i} = 1$.

Assumptions. We make the following assumptions on the reward Y_t. For all $t \in [T]$, Y_t is independent of $\mathcal{F}_{t-1} = \sigma(Y_1, \dots, Y_{t-1})$ and A_t is $\mathcal{F}_{t-1}-$measurable. There exists a mean reward vector $\mu \in \mathbb{R}^d$ and a positive semi-definite covariance matrix $\mathbf{\Sigma}^* \in M_d(\mathbb{R})$ such that $\mathbb{E}[Y_t] = \mu \in \mathbb{R}^d$ and $\mathrm{Var}(Y_t) = \mathbf{\Sigma}^*$. There exists a known *bounds* vector $B \in \mathbb{R}_+^d$ such that for all $i \in [d]$, $|Y_{t,i} - \mu_i| \leq B_i$.

Contributions. One of the weaknesses of most existing algorithms for stochastic combinatorial semi-bandits is their reliance on information about the reward distribution. For example, it is common to assume the existence and knowledge of a positive semi-definite proxy-covariance matrix Γ, such that for all $t \geq 1$ and for all $u \in \mathbb{R}^d$, $\mathbb{E}[\exp(u^\top (Y_t - \mu))] \leq \exp(\frac{1}{2} u^\top \Gamma u)$, like for `OLS-UCB` in Degenne and Perchet (2016). As the performances of these algorithms explicitly involve this information, it is essential to use the tightest estimates possible. For example, when all the rewards are bounded by $B_{\max} > 0$, a rough proxy-covariance $\Gamma = dB_{\max}^2 \mathrm{I}_d$ can be provided. However, this bound is typically very loose: it poorly uses the structure of the reward distribution and results in sub-optimal guarantees. Estimating a good proxy-variance is actually a challenging problem especially when dealing with non-Gaussian distributions. We propose instead an algorithm relying on the "real" covariance matrix of the reward distribution which is easier to estimate with good theoretical guarantees.

[1] Throughout the paper, the term *item* (or *base item*) refers to an element in the set $[d]$, while an *action* denotes a subset of base items, see Figure 1.

Table 1: Asymptotic $\tilde{O}(\cdot)$ regret bounds for different types of feedback, up to poly-logarithmic terms in T and d, for the following algorithms: `UCB` (Auer et al., 2002a), `UCBV` (Audibert et al., 2009), `CUCB` (Kveton et al., 2015), `OLS-UCB` (Degenne and Perchet, 2016), `ESCB-C` (Perrault et al., 2020), and `OLS-UCBV` (ours). The top of the table concerns algorithms using bandit feedbacks, and the bottom algorithms exploiting semi-bandit ones. *Notations:* a refers to actions; i and j refer to items; m denotes the maximum number of items per action; Γ is a proxy-covariance matrix; γ is a maximum of "proxy-correlations"; we abbreviate $\max\{x, 0\}$ to $(x)_+$ for any $x \in \mathbb{R}$.

Feedback	Algorithm	Info.	Gap-Dependant Asymptotic Regret	Gap-Free Asymptotic Regret
Bandit	`UCB`	Γ	$\sum_a \frac{a^\top \Gamma a}{\Delta_a}$	$\sqrt{T \sum_a a^\top \Gamma a}$
	`UCBV`	$\varnothing$	$\sum_a \frac{a^\top \boldsymbol{\Sigma}^* a}{\Delta_a}$	$\sqrt{T \sum_a a^\top \boldsymbol{\Sigma}^* a}$
Semi-Bandit	`CUCB`	Γ	$m \sum_i \frac{\Gamma_{i,i}}{\min_{a/i\in a} \Delta_a}$	$\sqrt{mT \sum_i \Gamma_{i,i}}$
	`OLS-UCB`	Γ	$(1+\gamma m) \sum_i \frac{\Gamma_{i,i}}{\min_{a/i\in a} \Delta_a} + \frac{dm^2 \max_i \Gamma_{i,i}}{\Delta_{\min}^2}$	$(dm^2 \max_i \Gamma_{i,i})^{1/3} T^{2/3}$
	`ESCB-C`	$\varnothing$	$\sum_i \max_{a/i\in a} \frac{\sum_{j\in a} (\boldsymbol{\Sigma}^*_{i,j})_+}{\Delta_{\min}} + \frac{dm^2 \max_i \Gamma_{i,i}}{\Delta_{\min}^2}$	$(dm^2 \max_i \Gamma_{i,i})^{1/3} T^{2/3}$
	`OLS-UCBV`	$\varnothing$	$\sum_i \max_{a/i\in a} \frac{\sum_{j\in a} (\boldsymbol{\Sigma}^*_{i,j})_+}{\Delta_a}$	$\sqrt{T \sum_i \max_{a/i\in a} \sum_{j\in a} (\boldsymbol{\Sigma}^*_{i,j})_+}$

The literature concerning stochastic combinatorial semi-bandit settings also mostly focuses on cases where the action set is exponentially large, namely $P \gg d$, and the way to get quasi-optimal regret rates in these instances. However, outside of these regimes, the commonly derived regret bounds are too rough and fail to show the benefit of the semi-bandit feedback. Conventional combinatorial semi-bandit regret upper bounds grow as $O(\sqrt{mdT})$, where m is the maximum number of items per action (Kveton et al., 2015), while a rate of order $O(\sqrt{mPT})$ can be achieved using bandit feedback only (for multi-armed bandit with P arms and rewards having variance m). Intriguingly, the latter appears to outperform the semi-bandit rate as soon as $P < d$, making the extra information obtained through supplementary feedback seemingly useless. It is thus natural to look for more fine-grained analyses showing clear improvements when going from pure bandit to semi-bandit feedbacks, in all regimes.

Hereafter are listed our contributions.

- In Section 2, we show a gap-free lower bound on the regret for stochastic combinatorial semi-bandits, explicitly involving the structure of the problem (the items in each action) and the covariance matrix $\boldsymbol{\Sigma}^*$.
- In Section 3 and Section 4, we design and analyze `OLS-UCBV`, a least-squares-based algorithm estimating the covariance matrix online. We show that it satisfies logarithmic regret rates explicitly involving $\boldsymbol{\Sigma}^*$ and the structure of the action set $\mathcal{A}$. A good estimation of $\boldsymbol{\Sigma}^*$ is not only much simpler to get than a proxy-covariance for any reward distributions, but also provides improved regret guarantees. We show that this algorithm yields a similar gap-dependant regret bound as `ESCB-C` from Perrault et al. (2020), up to logarithmic factors, and an improved $\sqrt{T}$ gap-free regret bound. The corresponding regret bounds are summarized in blue in Table 1. We also show that under mild conditions, leveraging the semi-bandit feedback indeed consistently offers an advantage and outperforms bandit algorithms in all regimes of P/d.
- Lastly Section 5 compares the computational complexities of `OLS-UCBV` with existing algorithms. We particularly show that `OLS-UCBV` is computationally cheaper than `ESCB-C` as we circumvent the need to solve a convex optimization problem at each round with increasing accuracy, by relying on closed-form expressions. Algorithm performances are compared on synthetic environments in Appendix E.

Related works. The combinatorial bandit setting has been introduced in Chen et al. (2013) which proposes `CUCB` (Combinatorial `UCB`), later analyzed in Kveton et al. (2015). This algorithm plugs upper confidence bounds for the items' rewards into a linear maximization problem and picks the action with the highest upper bound. However this version does not exploit correlations between the different base items. Its regret bound has later been outperformed in Combes et al. (2015) and Degenne and Perchet (2016). These papers propose algorithms than rely on the knowledge of either the parametric form of the rewards (Bernoulli in Combes et al., 2015) or a sub-Gaussian proxy-covariance matrix (Degenne and Perchet, 2016). However,

needing a proxy-covariance matrix is a major limitation as getting tight estimates is difficult in practice, and using a loose one severely impacts the performance.

While Audibert et al. (2009) designs and analyses `UCBV` for multi-armed bandit, Perrault et al. (2020) recently proposed `ESCB-C`, a covariance-adaptive algorithm for stochastic combinatorial semi-bandit. It is based on estimations of the covariance matrix and solving linear optimization problems in convex sets to compute the best action. The algorithm notably satisfies gap-dependant regret bounds depending asymptotically on the covariance matrix and solves most limitations from the previous work of Degenne and Perchet (2016).

We propose and analyse a novel algorithm based on a least-squares estimator, closer to `OLS-UCB` (Degenne and Perchet, 2016) than to `ESCB-C` (Perrault et al., 2020). We manage to show similar regret rates as Perrault et al. (2020), up to logarithmic factors, but using a different approach. In contrast to the i.i.d assumption needed in Perrault et al. (2020), our analysis only needs sequential independence and uses tools from martingales theory, which could enable easier extensions to richer settings like Markov Decision Processes. We also manage to get an improved $\tilde{O}(\sqrt{T})$ gap-free regret rate, while a $1/\Delta_{\min}^2$ term in the analysis of both `OLS-UCB` and `ESCB-C` prevents it. Besides, we also gain on computational complexity over `ESCB-C` as we do not rely on the approximate resolution of an optimization problem to determine upper bounds on actions' rewards.

Our analysis of `OLS-UCBV` is close to the one in Degenne and Perchet (2016) but uses different concentration bounds. We draw inspiration from Zhou et al. (2021), who recently adapted martingale arguments from Dani et al. (2008). Our work however improves on Zhou et al. (2021) through the exploitation of a multi-dimensional noise covariance structure for Y_t. It requires us to tailor an analysis capable of coupling it with the peeling trick used in Degenne and Perchet (2016). Thanks to these concentration arguments we can bound a "noise" factor in the cumulative regret, and use the same kind of "deterministic" analysis as in Kveton et al. (2015) to get logarithmic regret rates depending on the covariance matrix and the structure of the action set.

2 Lower Bound

In their recent paper, Perrault et al. (2020) prove a gap-dependant lower bound on the cumulative regret, and provide `ESCB-C`, an algorithm satisfying it, up to some assumptions and constant terms.

Theorem 2.1 (Theorem 1 in Perrault et al., 2020)**.** *Let $d, m \in \mathbb{N}^*$ such that $d/m \geq 2$ is an integer, $\mathbf{\Sigma}^* \succeq 0$, $\delta \in (0,1)$ and a consistent policy π, such that for any combinatorial semi-bandit instance*

$$\mathbb{E}[R_T] = o(T^b) \quad \text{for any } b > 0\,,$$

as $T \to \infty$. Then, there exists a stochastic combinatorial semi-bandit with d base items, actions of size at most m, and a reward distribution with covariance matrix $\mathbf{\Sigma}^*$ where all suboptimal gaps are Δ, on which the regret satisfies

$$\liminf_{T\to\infty} \frac{\Delta}{\log(T)} R_T \geq 2 \sum_{i\in[d], i\notin a^*} \max_{a\in\mathcal{A}, i\in a} \sum_{j\in a} \mathbf{\Sigma}^*_{i,j}\,.$$

Their proof considers d/m disjoint actions, with Gaussian rewards, and a reduction to a multi-armed bandit's lower bound. For a similar type of instance, we introduce a novel gap-free lower bound.

Theorem 2.2. *Let $d, m \in \mathbb{N}^*$ such that $d/m \geq 2$ is an integer, $T \in \mathbb{N}^*$, and $\mathbf{\Sigma}^* \succeq 0$ a covariance matrix. Then, there exists a stochastic combinatorial semi-bandit with d base items, actions of size at most m, and a reward distribution with covariance matrix $\mathbf{\Sigma}^*$ on which for any policy π, the regret satisfies*

$$R_T \geq \frac{1}{8}\sqrt{T \sum_{i\in[d]} \max_{a\in\mathcal{A}, i\in a} \sum_{j\in a} \mathbf{\Sigma}^*_{i,j}}\,.$$

Proof. The proof is detailed in Appendix A.1. We follow the methodology of Auer et al. (2002b), modifying it to account for the different variances among actions. □

3 OLS-UCBV

In this section, we design a new algorithm that efficiently leverages the semi-bandit feedback by approximating the coefficients of the covariance matrix $\mathbf{\Sigma}^*$ online. This approximation is symmetric by construction and yields a coefficient-wise upper bound of $\mathbf{\Sigma}^*$. But it is not necessarily a semi-definite positive matrix, a constraint that can be extremely challenging to tightly impose in practical scenarios.

While Perrault et al. (2020) use an axis-realignment technique and a covering argument to derive their confidence region, our approach is closer to the least-squares methodology of Degenne and Perchet (2016). It uses ellipsoidal confidence regions in which we incorporate Bernstein-like concentration inequalities. This particularly simplifies the computation of an upper confidence bound for each action. While Perrault et al. (2020) need to solve linear programs in convex sets at each iteration to derive their bound, we have access to closed-form expressions. More details on these differences in complexity are given in Section 5.

3.1 Estimators for mean and covariance

Mean estimation. Let $a \in \mathcal{A}$, $t \in [T]$, we denote by $\mathbf{d}_a = \mathrm{diag}(a) \in M_d(\mathbb{R})$ the diagonal matrix where the non-null coefficients are the elements of a. The number of times two items $i, j \in [d]$ (with possibly $i = j$) have been chosen together after round t is denoted by $n_{t,(i,j)} = \sum_{s=1}^{t} \mathbb{1}\{(i,j) \in A_s\}$. We define $\mathbf{D}_t = \mathrm{diag}((n_{t,(i,i)})_{i\in[d]}) \in M_d(\mathbb{R})$ the diagonal matrix of item counts. Then the least-squares estimator dor the mean reward vector μ using all the data from the past rounds after round t is the empirical average

$$\hat{\mu}_t = \mathbf{D}_t^{-1} \sum_{s=1}^{t} \mathbf{d}_{A_s} Y_s = \mu + \mathbf{D}_t^{-1} \sum_{s=1}^{t} \mathbf{d}_{A_s} \eta_s \,, \tag{2}$$

where η_s denotes the deviation of reward Y_s from its mean μ. This average yields the estimator $\langle a, \hat{\mu}_t \rangle$ for the mean reward $\langle a, \mu \rangle$, to which a well-designed optimistic bonus should be added. We design a strategy inspired from `LinUCB` (Rusmevichientong and Tsitsiklis, 2010; Filippi et al., 2010). The ellipsoid-based bonus that we introduce enables the use of a peeling trick and Bernstein's style concentration inequalities.

Covariance estimation. Let $t \in \mathbb{N}^*$ and $i, j \in [d]$ such that $n_{t,(i,j)} \geq 2$. The coefficients of $\mathbf{\Sigma}^*$ can be estimated online by $\hat{\chi}_t$ with elements as follows

$$\hat{\chi}_{t,(i,j)} = \tfrac{1}{n_{t,(i,j)}} \sum_{s=1}^{t} A_{s,i} A_{s,j} \mathbb{1}\{n_{s,(i,j)} \geq 2\} (Y_{s,i} - \hat{\mu}_{s-1,i})(Y_{s,j} - \hat{\mu}_{s-1,j}) \,. \tag{3}$$

A major difference with the estimators used in Perrault et al. (2020) (or those used in Audibert et al., 2009) is the fact that the rewards are "centered" using the sample average of the past rewards at the time of their observation (s in our case), instead of the whole sample average at time t.

3.2 Upper confidence bounds for covariance and mean

Covariance coefficients upper confidence bound. The following result controls the error of the online covariance estimator $\hat{\chi}_t$ presented in Equation (3).

Proposition 3.1. *Let $T \geq 3$, $\delta \in (0,1)$. Then with probability $1-\delta$, for all $t \leq T$ and $(i,j) \in [d]^2$, such that $n_{t,(i,j)} \geq 2$,*

$$|\hat{\chi}_{t,(i,j)} - \mathbf{\Sigma}^*_{i,j}| \leq \mathcal{B}_{t,(i,j)}(\delta, T) \,,$$

where $\mathcal{B}_{t,(i,j)}(\delta,T) = 3B_iB_j \left(\frac{h_{T,\delta}}{\sqrt{n_{t,(i,j)}}} + \frac{h^2_{T,\delta}}{n_{t,(i,j)}} \log(T) \right)$ with $h_{T,\delta} = \log(5d^2T^2/\delta)$.

Proof. See Appendix B.1. □

This result suggests the following upper confidence bounds for $\mathbf{\Sigma}^*$ to be plugged into our algorithm:

$$\hat{\mathbf{\Sigma}}_{t,(i,j)} = \hat{\chi}_{t,(i,j)} + \mathcal{B}_{t,(i,j)}(\delta, T) \,. \tag{4}$$

Algorithm 2 `OLS-UCBV`

Input $\delta > 0$, $T \geq 1$, $B \in \mathbb{R}_+^d$.
for $t = 1, \dots, T$ **do**
  if $\left\{a \in \mathcal{A} \text{ s.t } \min_{(i,j)\in a} n_{t,(i,j)} \leq 1\right\} \neq \emptyset$ **then**
    Choose any A_t in the above set.
  else
    Choose $A_t \in \mathcal{A}$ from (7) using $\hat{\mu}_{t-1}, \hat{\mathbf{Z}}_{t-1}$.
    Environment samples $Y_t \in \mathbb{R}^d$.
    Receive reward $\langle A_t, Y_t \rangle = \sum_i A_{t,i} Y_{t,i}$.
    Compute $\hat{\mu}_t$ from (2).
    Compute $\hat{\mathbf{\Sigma}}_t$ from (3) and (4).
    Compute $\hat{\mathbf{Z}}_t$ from (5).
  end if
end for

Mean upper confidence bound. We propose an upper confidence bound for the average rewards of all actions $a \in \mathcal{A}$ at any round t.

Denoting by $\hat{\mathbf{\Sigma}}_t$ the coefficient-wise `UCB` covariance matrix given by (4), we introduce the following "regularized empirical design matrix":

$$\hat{\mathbf{Z}}_t = \sum_{s=1}^{t} \mathbf{d}_{A_s} \hat{\mathbf{\Sigma}}_t \mathbf{d}_{A_s} + \mathbf{d}_{\hat{\mathbf{\Sigma}}_t} \mathbf{D}_t + d\mathbf{d}_B \,, \tag{5}$$

where $\mathbf{d}_B = \mathrm{diag}((B_i^2)_{i\in[d]})$ and $\mathbf{d}_{\hat{\mathbf{\Sigma}}_t} = \mathrm{diag}(\hat{\mathbf{\Sigma}}_t)$ are matrices in $M_d(\mathbb{R})$. The regularization of the matrix $\hat{\mathbf{Z}}_t$ is engineered to enable the use of a peeling trick in the proof of the regret bound stated in Theorem 3.2.

As we show in the upcoming analysis (see proof of Proposition 4.5), with probability greater than $1-\delta$, for all $a \in \mathcal{A}$

$$|\langle a, \mu \rangle - \langle a, \hat{\mu}_t \rangle| \leq f_{t,\delta} \, \|\mathbf{D}_t^{-1} a\|_{\mathbf{Z}_t} \,,$$

where

$$f_{t,\delta} = 6d \log\big(\log(1+t)\big) + 3d \log(1+e) + \log(1/\delta) \,, \tag{6}$$

is an exploration factor depending on the desired uncertainty level $\delta \in (0,1)$ and $\mathbf{Z}_t$ is the "exact" counterpart of $\hat{\mathbf{Z}}_t$, using $\mathbf{\Sigma}^*$ instead of $\hat{\mathbf{\Sigma}}_t$.

3.3 Algorithm

We now present `OLS-UCBV` written in Algorithm 2.

The algorithm first performs an initial exploration by sampling every base item $i \in [d]$ and every "reachable" couple $(i,j) \in [d]^2$ at least twice. Then, for all subsequent rounds $t+1$, `OLS-UCBV` picks an action A_{t+1} such that:

$$A_{t+1} \in \operatorname*{arg\,max}_{a\in\mathcal{A}} \left\{ \langle a, \hat{\mu}_t \rangle + f_{t,\delta} \, \left\|\mathbf{D}_t^{-1} a\right\|_{\hat{\mathbf{Z}}_t} \right\}. \tag{7}$$

3.4 Regret upper bound

We establish the following regret upper bounds for `OLS-UCBV`. We denote $\tilde{O}$ for O when $T \to \infty$, up to poly-logarithmic terms.

Theorem 3.2. *Let $T \geq 5$, $B \in \mathbb{R}_+^d$, and $\delta = 1/T^2$. Let*

$$\sigma_{a,i}^2 = \sum_{j\in a} (\mathbf{\Sigma}^*_{i,j})_+ \,, \tag{8}$$

where $i \in [d]$, $a \in \mathcal{A}$ *and* $(\cdot)_+ = \max\{\cdot, 0\}$. *Then,* `OLS-UCBV` *(Alg. 2) satisfies the* gap-dependent *regret upper bound*

$$\mathbb{E}[R_T] = \tilde{O}\left(\log(d)^2 \sum_{i=1}^{d} \max_{a \in \mathcal{A}/i \in a, \Delta_a > 0} \frac{\sigma_{a,i}^2}{\Delta_a} \right),$$

and the distribution-free *regret upper bound*

$$\mathbb{E}[R_T] = \tilde{O}\left(\log(d) \sqrt{T \sum_{i=1}^{d} \max_{a \in \mathcal{A}/i \in a} \sigma_{a,i}^2} \right).$$

We manage to get the same gap-dependent regret upper bound as `ESCB-C` (Perrault et al., 2020), up to logarithmic factors in time but we also manage to get a new $\sqrt{T}$ gap-free bound.

Diagonal covariance. It is important to note that in the scenario of independent item rewards, where the covariance $\mathbf{\Sigma}^*$ is diagonal, then $\sigma_{a,i}^2 = \Sigma_{i,i}^*$ for all $a \in \mathcal{A}$ and $i \in a$. Our gap-dependent and gap-free upper bounds are then roughly bounded as

$$\tilde{O}\left(\sum_{i=1}^{d} \frac{\Sigma_{i,i}^*}{\min_{a \in \mathcal{A}/i \in a} \Delta_a} \right) \text{ and } \tilde{O}\left(\sqrt{\mathrm{Tr}(\mathbf{\Sigma}^*) T} \right),$$

respectively. In this particular case, our gap-free bound outperforms the standard regret upper bound for combinatorial semi-bandit problems (which does not use the independence information, see for instance Kveton et al. (2015)), typically of the order $O(\sqrt{dmT})$ where m is the maximum number of items per action.

4 Analysis of `OLS-UCBV`

In this section, we analyze `OLS-UCBV` by detailing key steps of the proof of Theorem 3.2, while certain technical arguments are deferred to Appendix C.

The proof relies on considering high-probability "favorable" events where the two following conditions are met:

- $\hat{\mu}_t$ remains within a sufficiently small ellipsoid, the size of which expands logarithmically over time with $f_{t,\delta}$ – denoted as $\mathcal{G}_t$ thereafter, see (9),
- the covariance estimators are within high confidence bounds as specified in Proposition 3.1 – denoted as $\mathcal{C}$ thereafter, see (10).

4.1 Defining "favorable" events

Let $t \in [T]$ and $a \in \mathcal{A}$. We introduce the regularized design matrix $\mathbf{Z}_t$ using the covariance matrix $\mathbf{\Sigma}^*$ (recall the empirical version defined in (5))

$$\mathbf{Z}_t = \textstyle\sum_{s=1}^{t} \mathbf{d}_{A_s} \mathbf{\Sigma}^* \mathbf{d}_{A_s} + \mathbf{d}_{\mathbf{\Sigma}^*} \mathbf{D}_t + d \mathbf{d}_B \,,$$

and the event

$$\mathcal{G}_t = \left\{ \left\| \textstyle\sum_{s=1}^{t} \mathbf{d}_{A_s} \eta_s \right\|_{\mathbf{Z}_t^{-1}} \leq f_{t,\delta} \right\}, \tag{9}$$

corresponding to the belonging of $\hat{\mu}_t$ to an ellipsoid centered in μ (since η denote rewards deviations from the mean μ), the size of which grows with $f_{t,\delta}$ given in (6). We also define the event $\mathcal{C}$ of Proposition 3.1 as

$$\mathcal{C} = \left\{ \forall t \in [T],\ (i,j) \in [d]^2 \text{ s.t. } n_{t,(i,j)} \geq 2, |\hat{\chi}_{t,(i,j)} - \mathbf{\Sigma}_{i,j}^*| \leq \mathcal{B}_{t,(i,j)}(\delta, T) \right\}, \tag{10}$$

having probability at least $1 - \delta$.

4.2 Regret template bound

The algorithm begins with an exploration phase lasting at most $d(d+1)$ rounds. Thus,

$$\mathbb{E}[R_T] \leq d(d+1)\Delta_{\max} + \textstyle\sum_{t=d(d+1)}^{T-1} \mathbb{E}[\Delta_{A_{t+1}}], \tag{11}$$

where $\Delta_{\max}$ is the largest suboptimality gap in the instance.

Using the fact that the probability of $\mathcal{C}$ is at least $1-\delta$, the formula of total probability brings

$$\mathbb{E}[R_T] \leq \Delta_{\max}\Big(d(d+1) + \underbrace{\sum_{t=d(d+1)}^{T-1} \mathbb{P}\big(\mathcal{G}_t^c\big)}_{\mathbf{A}} + T\delta\Big) + \underbrace{\sum_{t=d(d+1)}^{T-1} \mathbb{E}\Big[\Delta_{A_{t+1}} \mathbb{1}\{\mathcal{G}_t \bigcap \mathcal{C}\}\Big]}_{\mathbf{B}}. \tag{12}$$

The rest of the proof consists in upper bounding the terms **A** and **B** appearing in (12), for which we now provide respective sketches of proof.

4.3 Bounding $\mathbf{A} = \sum_{t=d(d+1)}^{T-1} \mathbb{P}\big(\mathcal{G}_t^c\big)$

The term $\mathbf{A} = \sum_{t=d(d+1)}^{T-1} \mathbb{P}\big(\mathcal{G}_t^c\big)$ upper-bounds the probability that at least one of the unfavorable events $\mathcal{G}_t^c$ is realized –i.e. that for at least one round t, $\hat{\mu}_t$ is outside of the ellipsoid defined by (9). The following result provides an upper bound for **A**.

Proposition 4.1. *For events $\{\mathcal{G}_t\}_{t\leq T}$ defined as in* (9)*, we have*

$$\textstyle\sum_{t=d(d+1)}^{T-1} \mathbb{P}\big(\mathcal{G}_t^c\big) \leq \delta T^2. \tag{13}$$

Proof. Bounding $\sum_{t=d(d+1)}^{T-1} \mathbb{P}\big(\mathcal{G}_t^c\big)$ needs a meticulous interaction of a peeling trick as used by Degenne and Perchet (2016) and martingale arguments inspired from Abbasi-Yadkori et al. (2011). In the next two paragraphs, we present the main results of each of these two elements of the proof.

Peeling trick. The peeling trick consists in separating the space of trajectories up to round t into an exponentially large number of parts, each having an exponentially small probability.

Formally, let $\epsilon > 0$. To each $p \in \mathbb{N}^d$ we associate the set

$$\mathcal{D}_p = \big\{x \in \mathbb{R}^d \text{ s.t. } \forall i \in [d],\ (1+\epsilon)^{p_i} \leq x_i < (1+\epsilon)^{p_i+1}\big\}. \tag{14}$$

As an abuse of notation, we denote by $(t \in \mathcal{D}_p)$ the event $((n_{t,(i,i)}+1)_{i\in[d]} \in \mathcal{D}_p)$.

Setting $P_{t,\epsilon} = \lfloor \frac{\log(1+t)}{\log(1+\epsilon)} \rfloor$, we define for each $p \in [P_{t,\epsilon}]^d$

$$\begin{aligned}
\mathbf{D}_p &= \operatorname{diag}\Big(\big((1+\epsilon)^{p_i}\big)_{i\in[d]}\Big) \in M_d(\mathbb{R}),\\
\mathbf{Z}_{t,p} &= \sum_{s=1}^{t} \mathbf{d}_{A_s} \boldsymbol{\Sigma}^* \mathbf{d}_{A_s} + \mathbf{d}_{\boldsymbol{\Sigma}^*} \mathbf{D}_p + d\mathbf{d}_B,\\
M_{t,p} &= \Big\| \sum_{s=1}^{t} \mathbf{d}_{A_s} \eta_s \Big\|_{\mathbf{Z}_{t,p}^{-1}}.
\end{aligned} \tag{15}$$

Proposition 4.2. *With the notations of* (14) *and* (15)*, we have*

$$\mathbb{P}(\mathcal{G}_t^c) \leq \sum_{p\in[P_{t,\epsilon}]^d} \mathbb{P}\Big\{(M_{t,p}^2 > f_{t,\delta}^2) \bigcap (t \in \mathcal{D}_p)\Big\}. \tag{16}$$

Proof. The proof is deferred to Appendix C.1. □

Martingale bounding. The following result proposes a bound for the right-hand side of (16) and is proven using martingale arguments.

Proposition 4.3. *For $p \in [P_{t,\epsilon}]^d$ and with the notations of* (14) *and* (15)*, we have*

$$\mathbb{P}\Big\{M_{t,p}^2 > f_{t,\delta}^2 \bigcap (t \in \mathcal{D}_p)\Big\} \leq 2t\delta \left(\frac{\log(1+\epsilon)}{\log(1+t)}\right)^d, \tag{17}$$

where $f_{t,\delta} = 6d\log(\log(1+t)) + 3d\log(1+e) + \log(1/\delta)$ is defined in (6)*.*

Proof. The proof is deferred to Appendix C.2. □

Armed with Propositions 4.2 and 4.3 we can finally write:

$$\mathbb{P}(\mathcal{G}_t^c) \overset{(16)}{\leq} \sum_{p\in[P_{t,\epsilon}]^d} \mathbb{P}\Big\{M_{t,p}^2 > f_{t,\delta}^2 \bigcap (t \in \mathcal{D}_p)\Big\} \overset{(17)}{\leq} (P_{t,\epsilon})^d 2t\delta \left(\frac{\log(1+\epsilon)}{\log(1+t)}\right)^d \leq 2t\delta\,,$$

where the last inequality comes from $P_{t,\epsilon} = \lfloor \frac{\log(1+t)}{\log(1+\epsilon)} \rfloor$. Summing for $t = d(d+1)$ to $T-1$ then shows (13) which is the desired result. □

4.4 Bounding $\mathbf{B} = \sum_{t=d(d+1)}^{T-1} \mathbb{E}\big[\Delta_{A_{t+1}} \mathbb{1}\{\mathcal{G}_t \bigcap \mathcal{C}\}\big]$

Leveraging the careful definition of the favorable $\{\mathcal{G}_t\}_{t\in[T]}$ and $\mathcal{C}$ from (9) and (10) enables us to provide the following result to bound term **B**, corresponding the cumulative regret incurred under $\{\mathcal{G}_t \bigcap \mathcal{C}\}$ at each round t.

Our objective is to prove that B grows at most poly-logarithmically w.r.t. T, with coefficients depending on the structure of the instance and the covariance matrix $\mathbf{\Sigma}^*$. We state our result formally in the following proposition.

Proposition 4.4. *Let $T \geq d(d+1)$ and $\delta = 1/T^2$. Then,*

$$\mathbb{E}\Big[\sum_{t=d(d+1)}^{T-1} \Delta_{A_{t+1}} \mathbb{1}\big\{\mathcal{G}_t \bigcap \mathcal{C}\big\}\Big] = O\left(\log(T)^3(\log d)^2\Big(\sum_{i=1}^{d} \max_{a\in\mathcal{A}/i\in a} \frac{\sigma_{a,i}^2}{\Delta_a}\Big)\right),$$

as $T \to \infty$, where $\sigma_{a,i}^2 = \sum_{j\in a}(\mathbf{\Sigma}_{i,j}^*)_+$.

Proof. We first prove the following result which gives an upper bound of the gap $\Delta_{A_{t+1}}$ under $\{\mathcal{G}_t \bigcap \mathcal{C}\}$.

Proposition 4.5. *Let $t \geq d(d+1)$. Then under $\{\mathcal{G}_t \bigcap \mathcal{C}\}$, $\Delta_{A_{t+1}} \leq f_{t,\delta}\big(\|\mathbf{D}_t^{-1}A_{t+1}\|_{\mathbf{Z}_t} + \|\mathbf{D}_t^{-1}A_{t+1}\|_{\hat{\mathbf{Z}}_t}\big)$.*

Proof. The proof is deferred to Appendix C.7. □

To continue with our objective to bound **B**, we take inspiration from the works of Kveton et al. (2015) and Degenne and Perchet (2016). For starters, we prove the following result (using $\bar{\sigma}_{a,i}^2$'s defined differently from the $\sigma_{a,i}$'s).

Lemma 4.6. *Let $t \geq d(d+1)$, we have that under $\{\mathcal{G}_t \bigcap \mathcal{C}\}$,*

$$\begin{aligned}\frac{\Delta_{A_{t+1}}^2}{4f_{t,\delta}^2} &\leq \frac{1}{2}\big(\|\mathbf{D}_t^{-1}A_{t+1}\|_{\mathbf{Z}_t}^2 + \|\mathbf{D}_t^{-1}A_{t+1}\|_{\hat{Z}_t}^2\big) \\ &\leq \sum_{i\in A_{t+1}} \frac{\bar{\sigma}_{A_{t+1},i}^2}{n_{t,(i,i)}} + \big(d + 3\log(T)h_{T,\delta}^2\big) \sum_{(i,j)\in A_{t+1}} \frac{B_iB_j}{n_{t,(i,j)}^2} + 3h_{T,\delta} \sum_{(i,j)\in A_{t+1}} \frac{B_iB_j}{n_{t,(i,j)}^{3/2}}\,, \end{aligned} \tag{18}$$

where

$$\bar{\sigma}_{A_{t+1},i}^2 = 2 \sum_{j\in A_{t+1}/\mathbf{\Sigma}_{j,j}^* \leq \mathbf{\Sigma}_{i,i}^*} (\mathbf{\Sigma}_{i,j}^*)_+ \leq 2\sigma_{A_{t+1},i}^2\,.$$

Proof. The proof is a consequence of Proposition 4.5 and is deferred to Appendix C.8. □

Unfortunately, only the positive coefficients of $\mathbf{\Sigma}^*$ are considered in the analysis but the inclusion of negative correlations could be advantageous to reduce the rate at which the regret increases. However, it is quite complex and thus deferred to future research.

The rest of the proof involves a decomposition of $\{\mathcal{G}_t \bigcap \mathcal{C}\}$. By considering each of the 3 sub-sum in Equation (18) and designing set of event that are implied by $\{\mathcal{G}_t \bigcap \mathcal{C}\}$ but that can happen only a finite number of times.

In practice, we introduce two well-chosen sequences $(\alpha_k)_{k\in\mathbb{N}}$ and $(\beta_k)_{k\in\mathbb{N}}$ (see Appendix C.9), and define sequences of events $\mathbb{A}^r_{t,k}$ for $k \in \mathbb{N}^*$ and $r = 1, 2$ and 3 from them (see Appendix C.10). In particular, the following result holds

Lemma 4.7. *Let $t \geq d(d+1)$ and $k_0 \in \mathbb{N}^*$ such that $0 < d\beta_{k_0} < \left(\frac{1}{d} \wedge \min_{i,a}\{\mathbf{\Sigma}^*_{i,i}\bar{\sigma}^{-2}_{a,i}\}\right)$. Then,*

$$\mathbb{1}\{\mathcal{G}_t \bigcap \mathcal{C}\} \leq \sum_{r=1}^{3}\sum_{k=1}^{k_0} \mathbb{1}\{\mathbb{A}^r_{t,k}\}\,.$$

Proof. The proof is deferred to Appendix C.11. □

The rest of the proof consists in upper bounding $\sum_{t=d(d+1)}^{T-1} \Delta_{A_{t+1}} \mathbb{1}\{\mathbb{A}^r_{t,k}\}$ for each $(r, k) \in [3] \times [k_0]$ using the definitions of the events, and is detailed in Appendix C.12.

A notable difference from previous approaches by Kveton et al. (2015); Degenne and Perchet (2016) is the use of $\mathbf{\Sigma}^*_{i,i}/\bar{\sigma}^2_{a,i}$ in $\mathbb{A}^1_{t,k}$ instead of set cardinals. This enables the explicit appearance of the $\sigma^2_{a,i}$ coefficients (as $\bar{\sigma}^2_{a,i} \leq 2\sigma^2_{a,i}$). The use of $\log(T)$ and $\frac{2\log(T)}{\log(T)-1}$ factors for Lemma 4.7 also enable to make the terms depending on $\sigma^2_{a,i}$ dominant in the final bound. □

4.5 Conclusion of the proof

Injecting results from Proposition 4.4 and Proposition 4.1 into (12) finally yields

$$\mathbb{E}[R_T] = O\left(\log(T)^3\Big(\sum_{i\in[d]} \max_{a\in\mathcal{A}/i\in a} \frac{\sigma_{a,i}}{\Delta'_a}\Big)(\log d)^2\right),$$

as $T \to \infty$. This provides the gap-dependent bound of Theorem 3.2. The gap-free bound is detailed in Appendix D. It arises from the fact that our gap-dependent bound does not incur any term in $\Delta^{-2}_{\min}$, unlike Perrault et al. (2020); Degenne and Perchet (2016).

5 Comparisons and Complexity Analysis

5.1 Regret rates

Some advantages of using variance-adaptive algorithms instead of ones for which the performance depends on potentially non-tight quantities like sub-Gaussian parameters have been outlined in Section 3.4 and discussed more in Audibert et al. (2009) and Perrault et al. (2020). They particularly yield regret rates that can be considerably improved, at the cost of only marginally more computations. Among stochastic combinatorial semi-bandit algorithms, `OLS-UCBV` thus manages to yield variance-dependant regret rates similar to those reached by Perrault et al. (2020) up to log factors, but with an additional gap-free rate evolving as $\sqrt{T}$, see Table 1. An example of experiment can be found in Figure 2 in Appendix E where we can see that `OLS-UCBV` yields similar regret as other combinatorial algorithms.

5.2 Computational complexity

Another point of comparison between algorithms is their computational complexity, outlined in Table 2. Combinatorial semi-bandit algorithms are known not to be efficiently manageable for big action sets in the absence of structure (matroid constraints for example). For our comparison, we place ourselves in settings where we can afford to iterate computationally over it, thus the P factors. In that case, for the same

theoretical guarantees, OLS-UCBV is cheaper to execute compared to ESCB-C. Indeed, the latter needs to solve a linear program in a convex set at each round for each action, and with enough precision so that it does not influence the regret substantially. Given the form of the problem, this operation is significantly more costly than computing the upper confidence bound of OLS-UCBV which has a closed form involving only a finite number of matrix operations.

5.3 Complexity / regret trade-off

We now analyze the complexity / regret trade-off between using the pure bandit feedback or using the semi-bandit one, in settings where the number of actions is reasonable. Exploiting the semi-bandit feedback may be beneficial in certain cases, as it enables to combine information collected by different actions, but this comes at the cost of an increased computational complexity which may not be worth it. Looking more precisely at the gap-free regret rates in Table 1, while the regret of UCBV is upper bounded by $\left(T \sum_a a^\top \mathbf{\Sigma}^* a\right)^{1/2}$, the upper bound for OLS-UCBV is $\left(T \sum_i \max_{a/i \in a} \sum_{j \in a} (\mathbf{\Sigma}^*{}_{i,j})_+\right)^{1/2}$. In the case where all the correlations are positive or null, OLS-UCBV performs better. But in general, it may not be the case and negative coefficients in $\mathbf{\Sigma}^*$ could make $\sum_a a^\top \mathbf{\Sigma}^* a$ smaller than $\sum_i \max_{a/i \in a} \sum_{j \in a} (\mathbf{\Sigma}^*{}_{i,j})_+$, particularly in cases where P is not significantly bigger than d. Unfortunately, leveraging negative coefficients of $\mathbf{\Sigma}^*$ to make the semi-bandit feedback "always better" than the pure bandit feedback (as intuition dictates) does not seem to be straightforward and remains an open question.

Table 2: Complexity of the following algorithms: UCB (Auer et al., 2002a), UCBV (Audibert et al., 2009), CUCB (Kveton et al., 2015), OLS-UCB (Degenne and Perchet, 2016), ESCB-C (Perrault et al., 2020), and OLS-UCBV (ours). $C^{\text{opt}}_{1/T}$ refers to the complexity of the optimisation step needed in ESCB-C.

Feed.	Algorithm	Time	Space
Bandit	UCB/UCBV	TP	P
Semi-B.	CUCB	TPm	$d + P$
	OLS-UCB	$T(m^2 + Pd^2)$	$d^2 + P$
	ESCB-C	$T(m^2 + P\, C^{\text{opt}}_{1/T})$	$d^2 + P$
	OLS-UCBV	$T(m^2 + Pd^2)$	$d^2 + P$

5.4 Empirical comparison

We compare empirically in Appendix E theoretical regret rates of UCBV and OLS-UCBV, $\left(\sum_i \max_{a/i \in a} \sigma_{a,i}\right) / \left(\sum_a a^\top \mathbf{\Sigma}^* a\right)$, for randomly generated instances with different ratios P/d. Not enforcing any particular structure, the ratio seems to remain constant for our generated instances, with a notably constant factor gain when the covariance coefficients are biased to remain positive. But when we enforce a particular structure, like actions with no overlaps, we can actually see regimes where the absence of negative coefficients in the theoretical rate of OLS-UCB is detrimental, Figure 3 in Appendix E.

6 Concluding Remarks

We propose and analyze OLS-UCBV, a covariance-adaptive algorithm for stochastic combinatorial semi-bandits. Compared to other existing approaches, ours is computationally less demanding and yields the first $\sqrt{T}$ gap-free regret rate depending explicitly on the covariance of the base items rewards. A limitation, also existing in other works, is the overlooking of negative covariance coefficients. Finding a way to take them into account might greatly improve the performances of semi-bandit algorithms and show the benefit of the additional feedback compared to a pure bandit one.

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. *Advances in Neural Information Processing Systems*.

Audibert, J.-Y., Munos, R., and Szepesvári, C. (2009). Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19):1876–1902.

Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002a). Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256.

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. (2002b). The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77.

Brégère, M., Gaillard, P., Goude, Y., and Stoltz, G. (2019). Target tracking for contextual bandits: Application to demand side management. In *International Conference on Machine Learning*.

Bubeck, S. and Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122.

Chen, W., Wang, Y., and Yuan, Y. (2013). Combinatorial multi-armed bandit: General framework and applications. In *International Conference on Machine Learning*.

Combes, R., Talebi Mazraeh Shahi, M. S., Proutiere, A., et al. (2015). Combinatorial bandits revisited. *Advances in Neural Information Processing Systems*.

Dani, V., Hayes, T. P., and Kakade, S. M. (2008). Stochastic linear optimization under bandit feedback. In *Conference on Learning Theory*.

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

Filippi, S., Cappe, O., Garivier, A., and Szepesvári, C. (2010). Parametric bandits: The generalized linear case. *Advances in Neural Information Processing Systems*.

Guo, D., Ktena, S. I., Myana, P. K., Huszar, F., Shi, W., Tejani, A., Kneier, M., and Das, S. (2020). Deep Bayesian bandits: Exploring in online personalized recommendations. In *Conference on Recommender Systems*.

Kveton, B., Wen, Z., Ashkan, A., and Szepesvári, C. (2015). Tight regret bounds for stochastic combinatorial semi-bandits. In *International Conference on Artificial Intelligence and Statistics*.

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

Perrault, P., Valko, M., and Perchet, V. (2020). Covariance-adapting algorithm for semi-bandits with application to sparse outcomes. In *Conference on Learning Theory*.

Rusmevichientong, P. and Tsitsiklis, J. N. (2010). Linearly parameterized bandits. *Mathematics of Operations Research*, 35(2):395–411.

Tabei, G., Ito, Y., Kimura, T., and Hirata, K. (2023). Design of multi-armed bandit-based routing for in-network caching. *IEEE Access*.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.

Zeng, C., Wang, Q., Mokhtari, S., and Li, T. (2016). Online context-aware recommendation with time varying multi-armed bandit. In *International Conference on Knowledge Discovery and Data Mining*.

Zhou, D., Gu, Q., and Szepesvári, C. (2021). Nearly minimax optimal reinforcement learning for linear mixture Markov decision processes. In *Conference on Learning Theory*.

A Proofs of Section 2

A.1 Proof of Theorem 2.2

Theorem 2.2. *Let $d, m \in \mathbb{N}^*$ such that $d/m \geq 2$ is an integer, $T \in \mathbb{N}^*$, and $\mathbf{\Sigma}^* \succeq 0$ a covariance matrix. Then, there exists a stochastic combinatorial semi-bandit with d base items, actions of size at most m, and a reward distribution with covariance matrix $\mathbf{\Sigma}^*$ on which for any policy π, the regret satisfies*

$$R_T \geq \frac{1}{8}\sqrt{T \sum_{i\in[d]} \max_{a\in\mathcal{A}, i\in a} \sum_{j\in a} \mathbf{\Sigma}^*_{i,j}}\,.$$

Proof. Let $d, m \in \mathbb{N}^*$ such that $d/m \geq 2$ is an integer, $T \in \mathbb{N}^*$ a horizon, and $\mathbf{\Sigma}^* \succeq 0$ a covariance matrix. We consider the structure where $\mathcal{A} = \{a_1, \dots, a_{d/m}\} \subset \{0,1\}^d$ contains d/m disjoint actions each having m base elements. We consider that for all $p \in [d/m]$, $(a_p)_{i\in[d]} = \big(\mathbb{1}\big\{(p-1)m < i \leq pm\big\}\big)_{i\in[d]}$. Let π be a policy. As all the actions are disjoints, we can reduce ourselves to a multi-armed bandit with d/m actions, where for all $p \in [d/m]$ the variance of the p-th action is $a_p^\top \mathbf{\Sigma}^* a_p$.

Let $\mathbf{\Sigma}' \in M_{d/m}(\mathbb{R})$ be the diagonal matrix where for all $p \in [d/m]$, $\mathbf{\Sigma}'_{p,p} = a_p^\top \mathbf{\Sigma}^* a_p$. Let $c > 0$, and

$$\Delta = 2c\sqrt{\mathbf{\Sigma}'_{\min} \frac{\sum_{k=1}^{d/m} \mathbf{\Sigma}'_{k,k}}{T}}\,, \tag{19}$$

where $\mathbf{\Sigma}'_{\min} = \min_{p\in[d/m]} \mathbf{\Sigma}'_{p,p}$.

We denote $G_0 \sim \mathcal{N}(0, \mathbf{\Sigma}')$ a (d/m)-dimensional centered Gaussian distribution with covariance matrix $\mathbf{\Sigma}'$. Let $p \in [d/m]$, we consider the mean vector $\mu^{(p)} \in \mathbb{R}^{d/m}$ having coordinate 0 everywhere and Δ at coordinate p, for all $i \in [d/m]$, $\mu_i^{(p)} = \Delta \mathbb{1}\{i = p\}$. We introduce the Gaussian reward distributions $G_p \sim \mathcal{N}(\mu^{(p)}, \mathbf{\Sigma}')$ and denote $T_p = \sum_{t=1}^T \mathbb{1}\{A_t = p\}$. Then, using policy π, and considering the reward distributions G_p and G_0, the average number of times action p has been chosen satisfies

$$\Big|\mathbb{E}_{\pi,G_p}[T_p] - \mathbb{E}_{\pi,G_0}[T_p]\Big| \leq T\ \mathrm{TV}\Big((\pi, G_0), (\pi, G_p)\Big) \leq T\sqrt{\frac{1}{2}\mathrm{KL}\Big((\pi, G_0), (\pi, G_p)\Big)}\,, \tag{20}$$

where TV denotes the total variation distance, KL denotes the Kullback–Leibler divergence and the last inequality uses Pinsker's inequality. Then, using the divergence decomposition between multi-armed bandits (Lemma 15.1 in Lattimore and Szepesvári, 2020),

$$\begin{aligned}
\mathrm{KL}\big((\pi, G_0), (\pi, G_p)\big) &= \sum_{k=1}^{d/m} \mathbb{E}_{\pi,G_0}\big[T_k\big]\ \mathrm{KL}\big(\mathcal{N}(0, \mathbf{\Sigma}'), \mathcal{N}(\mu^{(p)}, \mathbf{\Sigma}')\big) \\
&= \sum_{k=1}^{d/m} \mathbb{E}_{\pi,G_0}\big[T_k\big]\ \frac{\big(\mu_k^{(p)}\big)^2}{2\mathbf{\Sigma}'_{k,k}}\,.
\end{aligned}$$

Reinjecting this expression into Equation (20), we get

$$\begin{aligned}
\mathbb{E}_{\pi,G_p}[T_p] &\leq \mathbb{E}_{\pi,G_0}[T_p] + \frac{T}{2}\sqrt{\sum_{k=1}^{d/m} \frac{\big(\mu_k^{(p)}\big)^2}{\mathbf{\Sigma}'_{k,k}} \mathbb{E}_{\pi,G_0}[T_k]} \\
&= \mathbb{E}_{\pi,G_0}[T_p] + \frac{T}{2}\sqrt{\frac{1}{\mathbf{\Sigma}'_{p,p}}\Delta^2 \mathbb{E}_{\pi,G_0}[T_p]} \\
&= \mathbb{E}_{\pi,G_0}[T_p] + c\sqrt{T\mathbb{E}_{\pi,G_0}[T_p] \frac{\mathbf{\Sigma}'_{\min}}{\mathbf{\Sigma}'_{p,p}} \sum_{k=1}^{d/m} \mathbf{\Sigma}'_{k,k}} \qquad \leftarrow \text{reinjecting Equation (19)} \\
&\leq \mathbb{E}_{\pi,G_0}[T_p] + c\sqrt{T\mathbb{E}_{\pi,G_0}[T_p] \sum_{k=1}^{d/m} \mathbf{\Sigma}'_{k,k}}\,.
\end{aligned}$$

Now, summing over the actions p,

$$\begin{aligned}
\sum_{p=1}^{d/m} \mathbb{E}_{\pi,G_p}[T_p] &\leq \sum_{p=1}^{d/m} \mathbb{E}_{\pi,G_0}[T_p] + c\sqrt{T\sum_{k=1}^{d/m}\boldsymbol{\Sigma}'_{k,k}}\sum_{p=1}^{d/m}\sqrt{\mathbb{E}_{\pi,G_0}[T_p]} \\
&\leq T + c\sqrt{T\sum_{k=1}^{d/m}\boldsymbol{\Sigma}'_{k,k}}\sqrt{\frac{d}{m}}\sqrt{\sum_{p=1}^{d/m}\mathbb{E}_{\pi,G_0}[T_p]} \qquad \leftarrow \text{Cauchy–Schwarz} \\
&\leq T + cT\sqrt{\frac{d}{m}\sum_{k=1}^{d/m}\boldsymbol{\Sigma}'_{k,k}}\,. \qquad (21)
\end{aligned}$$

We denote $R_T^{(p)}$ the average cumulative regret incurred with the reward distribution G_p, then

$$\begin{aligned}
\sum_{p=1}^{d/m} R_T^{(p)} &= \Delta\sum_{p=1}^{d/m}(T - \mathbb{E}_{\pi,G_p}[T_p]) \\
&= 2c\sqrt{\boldsymbol{\Sigma}'_{\min}\frac{\sum_{k=1}^{d/m}\boldsymbol{\Sigma}'_{k,k}}{T}}\left(\frac{d}{m}T - \sum_{p=1}^{d/m}\mathbb{E}_{\pi,G_p}[T_p]\right) \qquad \leftarrow \text{reinjecting Equation (19)} \\
&\geq 2c\sqrt{\boldsymbol{\Sigma}'_{\min}\frac{\sum_{k=1}^{d/m}\boldsymbol{\Sigma}'_{k,k}}{T}}\left(\frac{d}{m}T - T - cT\sqrt{\frac{d}{m}\sum_{k=1}^{d/m}\boldsymbol{\Sigma}'_{k,k}}\right) \qquad \leftarrow \text{from Equation (21)} \\
&= 2c\sqrt{\boldsymbol{\Sigma}'_{\min}}\,\frac{d}{m}\sqrt{T\sum_{k=1}^{d/m}\boldsymbol{\Sigma}'_{k,k}}\left(1 - \frac{m}{d} - c\sqrt{\frac{1}{d/m}\sum_{k=1}^{d/m}\boldsymbol{\Sigma}'_{k,k}}\right) \\
&\geq 2c\sqrt{\boldsymbol{\Sigma}'_{\min}}\frac{d}{m}\sqrt{T\sum_{k=1}^{d/m}\boldsymbol{\Sigma}'_{k,k}}\left(1 - \frac{m}{d} - c\sqrt{\boldsymbol{\Sigma}'_{\min}}\right).
\end{aligned}$$

Taking $c = \frac{1}{2}\frac{1}{\sqrt{\boldsymbol{\Sigma}'_{\min}}}(1 - \frac{m}{d})$,

$$\begin{aligned}
\sum_{p=1}^{d/m} R_T^{(p)} &\geq \frac{d}{m}\sqrt{T\sum_{k=1}^{d/m}\boldsymbol{\Sigma}'_{k,k}}\,\frac{1}{2}\left(1 - \frac{m}{d}\right)^2 \\
&\geq \frac{1}{8}\frac{d}{m}\sqrt{T\sum_{k=1}^{d/m}\boldsymbol{\Sigma}'_{k,k}} \leftarrow \text{as } m/d \leq 1/2\,.
\end{aligned}$$

Therefore, there exists at least one instance $p^* \in [d/m]$ such that

$$R_T^{(p^*)} \geq \frac{1}{8}\sqrt{T\sum_{k=1}^{d/m}\boldsymbol{\Sigma}'_{k,k}}\,.$$

Now, decomposing

$$\sum_{k=1}^{d/m}\boldsymbol{\Sigma}'_{k,k} = \sum_{k=1}^{d/m}\Big(\sum_{i\in a_k}\sum_{j\in a_k}\boldsymbol{\Sigma}^*_{i,j}\Big) = \sum_{i\in[d]}\max_{a\in\mathcal{A}, i\in a}\sum_{j\in a}\boldsymbol{\Sigma}^*_{i,j}\,,$$

we get

$$R_T^{(p^*)} \geq \frac{1}{8}\sqrt{T\sum_{i\in[d]}\max_{a\in\mathcal{A}, i\in a}\sum_{j\in a}\boldsymbol{\Sigma}^*_{i,j}}\,.$$

□

B Proofs of Section 3

B.1 Proof of Proposition 3.1

Proposition 3.1. *Let $T \geq 3$, $\delta \in (0,1)$. Then with probability $1-\delta$, for all $t \leq T$ and $(i,j) \in [d]^2$, such that $n_{t,(i,j)} \geq 2$,*

$$|\hat{\chi}_{t,(i,j)} - \mathbf{\Sigma}^*_{i,j}| \leq \mathcal{B}_{t,(i,j)}(\delta, T)\,,$$

where $\mathcal{B}_{t,(i,j)}(\delta,T) = 3B_iB_j\left(\frac{h_{T,\delta}}{\sqrt{n_{t,(i,j)}}} + \frac{h^2_{T,\delta}}{n_{t,(i,j)}}\log(T)\right)$ with $h_{T,\delta} = \log(5d^2T^2/\delta)$.

Proof. **Diagonal coefficients.** Let $i \in [d]$, $\delta \in (0,1)$ and $t \in \mathbb{N}$. For simplicity of notation, we write $n_{t,i}$ and $\chi_{t,i}$ instead of $n_{t,(i,i)}$ and $\chi_{t,(i,i)}$ for the diagonal coefficients. Then

$$\begin{aligned}
\hat{\chi}_{t,i} &= \tfrac{1}{n_{t,i}}\textstyle\sum_{s=1}^t A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(Y_{s,i} - \hat{\mu}_{s-1,i})^2 \\
&= \frac{1}{n_{t,i}}\sum_{s=1}^t A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}\Big(Y_{s,i} - \mu_i - \frac{1}{n_{s-1,i}}\sum_{k=1}^{s-1} A_{k,i}(Y_{k,i}-\mu_i)\Big)^2 \\
&= \frac{1}{n_{t,i}}\sum_{s=1}^t A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(Y_{s,i}-\mu_i)^2 \\
&\qquad - 2\frac{1}{n_{t,i}}\sum_{s=1}^t A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(Y_{s,i}-\mu_i)\frac{1}{n_{s-1,i}}\sum_{k=1}^{s-1} A_{k,i}(Y_{k,i}-\mu_i) \\
&\qquad + \frac{1}{n_{t,i}}\sum_{s=1}^n A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(\frac{1}{n_{s-1,i}}\sum_{k=1}^{s-1} A_{k,i}(Y_{k,i}-\mu_{k,i}))^2\,.
\end{aligned}$$

Then, the triangle inequality yields

$$|\hat{\chi}_{t,i} - \mathbf{\Sigma}^*_{i,i}| \leq \underbrace{\Big|\frac{1}{n_{t,i}}\sum_{s=1}^t A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(\eta^2_{s,i} - \mathbf{\Sigma}^*_{i,i})\Big|}_{A} + \Big|\frac{1}{n_{t,i}}\mathbf{\Sigma}^*_{i,i}\Big| \tag{22}$$

$$+ \underbrace{\Big|2\frac{1}{n_{t,i}}\sum_{s=1}^t A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}\eta_{s,i}\frac{1}{n_{s-1,i}}\sum_{k=1}^{s-1} A_{k,i}\eta_{k,i}\Big|}_{B} \tag{23}$$

$$+ \underbrace{\Big|\frac{1}{n_{t,i}}\sum_{s=1}^n A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(\frac{1}{n_{s-1,i}}\sum_{k=1}^{s-1} A_{k,i}\eta_{k,i})^2\Big|}_{C}\,. \tag{24}$$

We begin with the term A. As for all $s \in [t]$, $|A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(\eta^2_{s,i} - \mathbf{\Sigma}^*_{i,i})| \leq B_i^2$, then for all $\lambda \in \mathbb{R}$, using submartingale arguments,

$$\mathbb{E}\left[\exp\Big(\lambda\sum_{s=1}^t A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(\eta^2_{s,i} - \mathbf{\Sigma}^*_{i,i}) - (n_{t,i}-1)\frac{1}{2}\lambda^2 B_i^4\Big)\right] \leq 1\,.$$

From here, we use a Laplace's method by taking $\lambda \sim \mathcal{N}(0, 1/B_i^4)$, this yields

$$\frac{B_i^2}{\sqrt{2\pi}}\mathbb{E}\Bigg[\int_{\mathbb{R}}\exp\Big(\lambda\sum_{s=1}^{t}A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(\eta_{s,i}^2-\mathbf{\Sigma}^*_{i,i})-n_{t,i}\frac{1}{2}\lambda^2B_i^4\Big)d\lambda\Bigg]\leq 1$$

$$\frac{B_i^2}{\sqrt{2\pi}}\mathbb{E}\Bigg[\int_{\mathbb{R}}\exp\Big(\frac{1}{2}\frac{1}{n_{t,i}B_i^4}\big(\sum_{s=1}^{t}A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(\eta_{s,i}^2-\mathbf{\Sigma}^*_{i,i})\big)^2 - \frac{1}{2}n_{t,i}B_i^4\big(\lambda-\frac{1}{n_{t,i}B_i^4}\sum_{s=1}^{t}A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(\eta_{s,i}^2-\mathbf{\Sigma}^*_{i,i})\big)^2\Big)d\lambda\Bigg]\leq 1$$

$$\frac{B_i^2}{\sqrt{2\pi}}\mathbb{E}\Bigg[\exp\Big(\frac{1}{2}\frac{1}{n_{t,i}B_i^4}\big(\sum_{s=1}^{t}A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(\eta_{s,i}^2-\mathbf{\Sigma}^*_{i,i})\big)^2\Big)\int_{\mathbb{R}}\exp\Big(-\frac{1}{2}n_{t,i}B_i^4\big(\lambda-\frac{1}{n_{t,i}B_i^4}\sum_{s=1}^{t}A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(\eta_{s,i}^2-\mathbf{\Sigma}^*_{i,i})\big)^2\Big)d\lambda\Bigg]\leq 1$$

$$\frac{B_i^2}{\sqrt{2\pi}}\mathbb{E}\Bigg[\exp\Big(\frac{1}{2}\frac{1}{n_{t,i}B_i^4}\big(\sum_{s=1}^{t}A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(\eta_{s,i}^2-\mathbf{\Sigma}^*_{i,i})\big)^2\Big)\frac{\sqrt{2\pi}}{\sqrt{n_{t,i}}B_i^2}\Bigg]\leq 1$$

$$\mathbb{E}\Bigg[\exp\Big(\frac{1}{2}\frac{1}{n_{t,i}B_i^4}\big(\sum_{s=1}^{t}A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(\eta_{s,i}^2-\mathbf{\Sigma}^*_{i,i})\big)^2-\frac{1}{2}\big(\log(n_{t,i})+2\log(1/\delta)\big)\Big)\Bigg]\leq\delta$$

$$\mathbb{E}\Bigg[\exp\Big(\frac{1}{2}\frac{n_{t,i}}{B_i^4}\Big(\big(\frac{1}{n_{t,i}}\sum_{s=1}^{t}A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(\eta_{s,i}^2-\mathbf{\Sigma}^*_{i,i})\big)^2-\frac{B_i^4}{n_{t,i}}\big(\log(n_{t,i})+2\log(1/\delta)\big)\Big)\Big)\Bigg]\leq\delta\,.$$

Thus a Chernoff's bounding yields

$$\mathbb{P}\Bigg(\Big|\frac{1}{n_{t,i}}\sum_{s=1}^{t}A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}(\eta_{s,i}^2-\mathbf{\Sigma}^*_{i,i})\Big|\geq\frac{B_i^2}{\sqrt{n_{t,i}}}\big(\log(T)+\log(1/\delta)\big)\Bigg)\leq\delta. \tag{25}$$

For the term B, we use the same kind of approach to get that,

$$\mathbb{P}\Bigg(2\Big|\frac{1}{n_{t,i}}\sum_{s=1}^{t}A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}\eta_{s,i}\frac{1}{n_{s-1,i}}\sum_{k=1}^{s-1}A_{k,i}\eta_{k,i}\Big|\geq\frac{2B_i^2}{\sqrt{n_{t,i}}}\big(\log(T)+\log(1/\delta)\big)\Bigg)\leq\delta\,. \tag{26}$$

The last term, C, needs a little more steps. With probability at least $1-T\delta$, for all $s\in[T]$

$$\mathbb{P}\Bigg(\Big|\frac{1}{n_{s-1,i}}\sum_{k=1}^{s-1}A_{k,i}\eta_{k,i}\Big|^2\geq\frac{B_i^2}{n_{s-1,i}}\big(\log(T)+\log(1/\delta)\big)^2\Bigg)\leq\delta\,,$$

and thus, a union bound gives that, for all $t\in[T]$,

$$\mathbb{P}\Bigg(\Big|\frac{1}{n_{t,i}}\sum_{s=1}^{n}A_{s,i}\mathbb{1}\{n_{s,i}\geq 2\}\big(\frac{1}{n_{s-1,i}}\sum_{k=1}^{s-1}A_{k,i}\eta_{k,i}\big)^2\Big|\geq\frac{B_i^2(1+\log(T))}{n_{t,i}}\big(\log(T)+\log(1/\delta)\big)^2\Bigg)\leq T\delta\,. \tag{27}$$

Reinjecting Eq. (25), (26) and (27) into Eq. (22) yields that with probability at least $1-3T\delta$, for all $t\in[T]$

$$|\hat{\chi}_{t,i}-\mathbf{\Sigma}^*_{i,i}|\leq\frac{3B_i^2}{\sqrt{n_{t,i}}}(\log(T)+\log(1/\delta))+\frac{B_i^4(1+\log(T))}{n_{t,i}}\big(\log(T)+\log(1/\delta)\big)^2+\frac{B_i^2}{n_{t,i}}\,,$$

the last term coming from the bias when we use $n_{t,i}$ instead of $n_{t,i}-1$.

Rearranging terms, for $T\geq 3$, with probability at least $1-d\delta$, for all $t\in[T]$ and $i\in[d]$,

$$|\hat{\chi}_{t,i}-\mathbf{\Sigma}^*_{i,i}|\leq\frac{3B_i^2}{\sqrt{n_{t,i}}}(\log(3T^2)+\log(1/\delta))+3\frac{B_i^2}{n_{t,i}}\log(T)\big(\log(3T^2)+\log(1/\delta)\big)^2\,.$$

Extra diagonal coefficients. Let $(i,j) \in [d]$, $\delta \in (0,1)$ and $t \in \mathbb{N}$. Then,

$$\begin{aligned}\hat{\chi}_{t,(i,j)} &= \frac{1}{n_{t,(i,j)}} \sum_{s=1}^{t} A_{s,i} A_{s,j} \mathbb{1}\{n_{s,(i,j)} \geq 2\} (Y_{s,i} - \hat{\mu}_{s-1,i})(Y_{s,j} - \hat{\mu}_{s-1,j}) \\ &= \frac{1}{n_{t,(i,j)}} \sum_{s=1}^{t} A_{s,i} A_{s,j} \mathbb{1}\{n_{s,(i,j)} \geq 2\} \Big(\eta_{s,i} - \frac{1}{n_{s-1,i}} \sum_{k=1}^{s-1} \eta_{k,i}\Big)\Big(\eta_{s,j} - \frac{1}{n_{s-1,j}} \sum_{k=1}^{s-1} \eta_{k,j}\Big) .\end{aligned}$$

Thus

$$\begin{aligned}|\hat{\chi}_{t,(i,j)} - \mathbf{\Sigma}^*_{i,j}| \leq \underbrace{\Big| \frac{1}{n_{t,(i,j)}} \sum_{s=1}^{t} A_{s,i} A_{s,j} \mathbb{1}\{n_{s,(i,j)} \geq 2\} \eta_{s,i} \eta_{s,j} - \mathbf{\Sigma}^*_{i,j} \Big|}_{D} + \frac{|\mathbf{\Sigma}^*_{i,j}|}{n_{t,(i,j)}} \\ + \underbrace{\Big| \frac{1}{n_{t,(i,j)}} \sum_{s=1}^{t} A_{s,i} A_{s,j} \mathbb{1}\{n_{s,(i,j)} \geq 2\} \eta_{s,i} \frac{1}{n_{s-1,j}} \sum_{k=1}^{s-1} A_{k,j} \eta_{k,j} \Big|}_{E} \\ + \underbrace{\Big| \frac{1}{n_{t,(i,j)}} \sum_{s=1}^{t} A_{s,i} A_{s,j} \mathbb{1}\{n_{s,(i,j)} \geq 2\} \eta_{s,j} \frac{1}{n_{s-1,i}} \sum_{k=1}^{s} A_{k,i} \eta_{k,i} \Big|}_{F} \\ + \underbrace{\Big| \frac{1}{n_{t,(i,j)}} \sum_{s=1}^{t} A_{s,i} A_{s,j} \mathbb{1}\{n_{s,(i,j)} \geq 2\} \frac{1}{n_{s-1,i}} \sum_{k=1}^{s-1} A_{k,i} \eta_{k,i} \frac{1}{n_{s-1,j}} \sum_{k=1}^{s-1} A_{k,j} \eta_{k,j} \Big|}_{G}\end{aligned}$$

The term are treated just like for the diagonal terms.

$$\begin{aligned}\mathbb{P}\left(D \geq \frac{B_i B_j}{\sqrt{n_{t,i,j}}} \big(\log(T) + \log(1/\delta) \big) \right) \leq \delta , \\ \mathbb{P}\left(E \geq \frac{B_i B_j}{\sqrt{n_{t,(i,j)}}} \big(\log(T) + \log(1/\delta) \big) \right) \leq \delta , \\ \mathbb{P}\left(F \geq \frac{B_i B_j}{\sqrt{n_{t,(i,j)}}} \big(\log(T) + \log(1/\delta) \big) \right) \leq \delta .\end{aligned}$$

And with probability at least $1 - 2T\delta$, for all $s \in [T]$,

$$\Big| \frac{1}{n_{s-1,i}} \sum_{k=1}^{s-1} A_{k,i} \eta_{k,i} \Big| \leq \frac{B_i}{\sqrt{n_{s-1,i}}} \big(\log(T) + \log(1/\delta) \big) ,$$

and

$$\Big| \frac{1}{n_{s-1,j}} \sum_{k=1}^{s-1} A_{k,j} \eta_{k,j} \Big| \leq \frac{B_j}{\sqrt{n_{s-1,j}}} \big(\log(T) + \log(1/\delta) \big) .$$

As $n_{s-1,(i,j)} \leq \sqrt{n_{s-1,i} n_{s-1,j}}$, then for all $t \in [T]$

$$G \leq \frac{B_i B_j}{n_{t,(i,j)}} (1 + \log(T)) \big(\log(T) + \log(1/\delta) \big)^2 .$$

Therefore for $T \geq 3$, with probability at least $1 - \frac{1}{2} d(d-1)\delta$, for all $t \in [T]$ and $(i,j) \in [d]^2$,

$$|\hat{\chi}_{t,(i,j)} - \mathbf{\Sigma}^*_{i,j}| \leq 3 \frac{B_i B_j}{\sqrt{n_{t,i,j}}} \big(\log(5T^2) + \log(1/\delta) \big) + 3 \frac{B_i B_j}{n_{t,(i,j)}} \log(T) \big(\log(5T^2) + \log(1/\delta) \big)^2 .$$

A final union bound for all the coefficients yield the desired result. □

C Proofs of Section 4

C.1 Proof of Proposition 4.2

Proposition 4.2. *With the notations of* (14) *and* (15)*, we have*

$$\mathbb{P}(\mathcal{G}_t^c) \leq \sum_{p\in[P_{t,\epsilon}]^d} \mathbb{P}\Big\{(M_{t,p}^2 > f_{t,\delta}^2)\bigcap(t\in\mathcal{D}_p)\Big\}\,. \tag{16}$$

Proof. Using the definition of the events $\mathcal{D}_p$ we can write the following onion bound:

$$\mathbb{P}(\mathcal{G}_t^c) \leq \textstyle\sum_{p\in[P_{t,\epsilon}]^d} \mathbb{P}\Big\{\mathcal{G}_t^c\bigcap(t\in\mathcal{D}_p)\Big\}\,. \tag{28}$$

where $P_{t,\epsilon} = \lfloor \frac{\log(1+t)}{\log(1+\epsilon)} \rfloor$. Let $p\in\mathbb{N}^d$, under the event $(t\in\mathcal{D}_p)$, we have

$$\big\|\textstyle\sum_{s=1}^t \mathbf{d}_{A_s}\eta_s\big\|_{\mathbf{Z}_t^{-1}} \leq \big\|\sum_{s=1}^t \mathbf{d}_{A_s}\eta_s\big\|_{\mathbf{Z}_{t,p}^{-1}}\,, \tag{29}$$

with $\mathbf{D}_p = \mathrm{diag}(((1+\epsilon)^{p_i})_{i\in[d]}) \in M_d(\mathbb{R})$ and $\mathbf{Z}_{t,p} = \sum_{s=1}^t \mathbf{d}_{A_s}\mathbf{\Sigma}^*\mathbf{d}_{A_s} + \mathbf{d}_{\mathbf{\Sigma}^*}\mathbf{D}_p + d\mathbf{d}_B$. Then, denoting $M_{t,p} = \|\sum_{s=1}^t \mathbf{d}_{A_s}\eta_s\|_{\mathbf{Z}_{t,p}^{-1}}$, the unfavorable event probability may be upper bounded as

$$\begin{aligned}\mathbb{P}(\mathcal{G}_t^c) \overset{(9)}{=} \mathbb{P}\Big\{\Big\|\sum_{s=1}^t \mathbf{d}_{A_s}\eta_s\Big\|_{\mathbf{Z}_t^{-1}} > f_{t,\delta}\Big\} &\overset{(28)}{\leq} \sum_{p\in[P_{t,\epsilon}]^d} \mathbb{P}\Big\{\Big(\Big\|\sum_{s=1}^t \mathbf{d}_{A_s}\eta_s\Big\|_{\mathbf{Z}_t^{-1}} > f_{t,\delta}\Big)\bigcap(t\in\mathcal{D}_p)\Big\}\\ &\overset{(29)}{\leq} \sum_{p\in[P_{t,\epsilon}]^d} \mathbb{P}\Big\{(M_{t,p}^2 > f_{t,\delta}^2)\bigcap(t\in\mathcal{D}_p)\Big\}\,. \end{aligned}\tag{30}$$

□

C.2 Proof of Proposition 4.3

Proposition 4.3. *For $p\in[P_{t,\epsilon}]^d$ and with the notations of* (14) *and* (15)*, we have*

$$\mathbb{P}\Big\{M_{t,p}^2 > f_{t,\delta}^2\bigcap(t\in\mathcal{D}_p)\Big\} \leq 2t\delta\left(\frac{\log(1+\epsilon)}{\log(1+t)}\right)^d\,, \tag{17}$$

where $f_{t,\delta} = 6d\log(\log(1+t)) + 3d\log(1+e) + \log(1/\delta)$ is defined in (6)*.*

Proof. In order to analyze $\mathbb{P}\Big\{M_{t,p}^2 > f_{t,\delta}^2\bigcap(t\in\mathcal{D}_p)\Big\}$ for each $p\in[P_{t,\epsilon}]^d$, we first upper bound $M_{t,p}^2$ by a sum of two martingales thanks to the following technical lemma.

Lemma C.1. *Let $t\in\mathbb{N}$ and $p\in\mathbb{N}^d$, then*

$$M_{t,p}^2 \leq \underbrace{2\textstyle\sum_{s=1}^t (\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1}\left(\sum_{k=1}^{s-1}\mathbf{d}_{A_k}\eta_k\right)}_{I_{t,p}} + \underbrace{\textstyle\sum_{s=1}^t \|\mathbf{d}_{A_s}\eta_s\|^2_{\mathbf{Z}_{s-1,p}^{-1}}}_{J_{t,p}}\,.$$

Proof. The complete argument is deferred to Appendix C.3. □

Conveniently, we can tie the variance of these martingales to the dimension d in each layer thanks to the the following technical lemma, inspired by Abbasi-Yadkori et al. (2011). The proof is in Appendix C.4.

Lemma C.2. *Let $t\in\mathbb{N}$, $p\in\mathbb{N}^d$, then under $(t\in\mathcal{D}_p)$, $\sum_{k=1}^s \big\|\mathbf{d}_{A_k}\mathbf{\Sigma}^{*1/2}\big\|^2_{\mathbf{Z}_{k-1,p}^{-1}} \leq 2d\log(2+\epsilon)$ for all $s\in[t]$.*

Let $(x_I, x_J) \in]0,1[^2$ such that $x_I + x_J = 1$. Then,

$$\mathbb{P}\Big\{M_{t,p}^2 > f_{t,\delta}^2 \bigcap (t \in \mathcal{D}_p)\Big\} \leq \mathbb{P}\Big\{(M_{t,p}^2 > f_{t,\delta}^2) \bigcap (\forall s \leq t, J_{s,p} \leq x_J f_{t,\delta}^2) \bigcap (t \in \mathcal{D}_p)\Big\} + \mathbb{P}\Big\{(\exists s \in [t], J_{s,p} > x_J f_{t,\delta}^2) \bigcap (t \in \mathcal{D}_p)\Big\}. \tag{31}$$

This yields the following two propositions.

Proposition C.3. *Let $t \in \mathbb{N}$, $p \in \mathbb{N}^d$, then for $x_I f_t \geq 4d \log\log(1+t) + 2d\log(2+\epsilon) + 4\log(1/\delta)$,*

$$\mathbb{P}\big\{(M_{t,p}^2 > f_{t,\delta}^2) \bigcap (\forall s \in [t], J_{s,p} \leq x_J f_{t,\delta}^2) \bigcap (t \in \mathcal{D}_p)\big\} \leq t\delta\Big(\tfrac{\log(1+\epsilon)}{\log(1+t)}\Big)^d .$$

Proof. The proof relies on bounding the probability by the sum $\sum_{k=1}^t \mathbb{P}((I_{k,p} > x_I f_t^2) \bigcap (t \in \mathcal{D}_p))$, and using a submartingale argument. See Appendix C.5. □

Proposition C.4. *Let $t \in \mathbb{N}$, $p \in \mathbb{N}^d$, then if $x_J f_t^2 \geq 2d \log\log(1+t) + 3d\log(2+\epsilon) + 2\log(1/\delta)$,*

$$\mathbb{P}\Big\{(\exists s \in [t], J_{s,p} > x_J f_t^2) \bigcap (t \in \mathcal{D}_p)\Big\} \leq t\delta\Big(\tfrac{\log(1+\epsilon)}{\log(1+t)}\Big)^d .$$

Proof. The proof is in Appendix C.6. □

Let $t \geq 5$, $d \geq 1$, the choices $x_J = 1/3$, $x_I = 2/3$, $\epsilon = e - 1$ and

$$f_t = 6d\log(\log(1+t)) + 3d\log(1+e) + \log(1/\delta) \tag{32}$$

satisfy the assumptions of Propositions C.3 and C.4. Combining (31) and Propositions C.3 and C.4 yields the wanted result. □

C.3 Proof of Lemma C.1

Lemma C.1. *Let $t \in \mathbb{N}$ and $p \in \mathbb{N}^d$, then*

$$M_{t,p}^2 \leq \underbrace{2\textstyle\sum_{s=1}^t (\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1}\Big(\sum_{k=1}^{s-1}\mathbf{d}_{A_k}\eta_k\Big)}_{I_{t,p}} + \underbrace{\textstyle\sum_{s=1}^t \|\mathbf{d}_{A_s}\eta_s\|^2_{\mathbf{Z}_{s-1,p}^{-1}}}_{J_{t,p}} .$$

Proof. Let $t \in \mathbb{N}^*$, $p \in \mathbb{N}^d$, $s \in [t]$, then $\mathbf{Z}_{0,p} = \mathbf{D}_p \mathbf{d}_{\mathbf{\Sigma}^*} + d\mathbf{d}_B$ and, as

$$\mathbf{Z}_{s,p} = \mathbf{Z}_{s-1,p} + \mathbf{d}_{A_s}\mathbf{\Sigma}^*\mathbf{d}_{A_s} ,$$

then,

$$\mathbf{Z}_{s,p}^{-1} \preccurlyeq \mathbf{Z}_{s-1,p}^{-1} . \tag{33}$$

We can now bound $M_{s,p}^2$ with a recursive expression (with respect to $M_{s-1,p}^2$), as

$$\begin{aligned}
M_{s,p}^2 &= \left\| \sum_{k=1}^{s} \mathbf{d}_{A_k} \eta_k \right\|_{\mathbf{Z}_{s,p}^{-1}}^2 \\
&\leq \left\| \sum_{k=1}^{s} \mathbf{d}_{A_k} \eta_k \right\|_{\mathbf{Z}_{s-1,p}^{-1}}^2 \quad \leftarrow \text{ Eq. (33)} \\
&= \left(\sum_{k=1}^{s-1} \mathbf{d}_{A_k} \eta_k + \mathbf{d}_{A_s} \eta_s \right)^\top \mathbf{Z}_{s-1,p}^{-1} \left(\sum_{k=1}^{s-1} \mathbf{d}_{A_k} \eta_k + \mathbf{d}_{A_s} \eta_s \right) \\
&= \left(\sum_{k=1}^{s-1} \mathbf{d}_{A_k} \eta_k \right)^\top \mathbf{Z}_{s-1,p}^{-1} \left(\sum_{k=1}^{s-1} \mathbf{d}_{A_k} \eta_k \right) + 2 \left(\mathbf{d}_{A_s} \eta_s \right)^\top \mathbf{Z}_{s-1,p}^{-1} \left(\sum_{k=1}^{s-1} \mathbf{d}_{A_k} \eta_k \right) + \left(\mathbf{d}_{A_s} \eta_s \right)^\top \mathbf{Z}_{s-1,p}^{-1} \left(\mathbf{d}_{A_s} \eta_s \right) \\
&= \left\| \sum_{k=1}^{s-1} \mathbf{d}_{A_k} \eta_k \right\|_{\mathbf{Z}_{s-1,p}^{-1}}^2 + 2 \left(\mathbf{d}_{A_s} \eta_s \right)^\top \mathbf{Z}_{s-1,p}^{-1} \left(\sum_{k=1}^{s-1} \mathbf{d}_{A_k} \eta_k \right) + \| \mathbf{d}_{A_s} \eta_t \|_{\mathbf{Z}_{s-1,p}^{-1}}^2 \\
&= M_{s-1,p}^2 + 2 \left(\mathbf{d}_{A_s} \eta_s \right)^\top \mathbf{Z}_{s-1,p}^{-1} \left(\sum_{k=1}^{s-1} \mathbf{d}_{A_k} \eta_k \right) + \| \mathbf{d}_{A_s} \eta_t \|_{\mathbf{Z}_{s-1,p}^{-1}}^2, . \qquad (34)
\end{aligned}$$

Inequality (34) being true for every $s \in [t]$, we can sum:

$$\sum_{s=1}^{t} M_{s,p}^2 \leq \sum_{s=1}^{t} M_{s-1,p}^2 + 2 \sum_{s=1}^{t} \left(\mathbf{d}_{A_s} \eta_s \right)^\top \mathbf{Z}_{s-1,p}^{-1} \left(\sum_{k=1}^{s-1} \mathbf{d}_{A_k} \eta_k \right) + \sum_{s=1}^{t} \| \mathbf{d}_{A_s} \eta_s \|_{\mathbf{Z}_{s-1,p}^{-1}}^2 .$$

Since $M_{0,p} = 0$, this finally yields

$$M_{t,p}^2 \leq 2 \sum_{s=1}^{t} \left(\mathbf{d}_{A_s} \eta_s \right)^\top \mathbf{Z}_{s-1,p}^{-1} \left(\sum_{k=1}^{s-1} \mathbf{d}_{A_k} \eta_k \right) + \sum_{s=1}^{t} \| \mathbf{d}_{A_s} \eta_s \|_{\mathbf{Z}_{s-1,p}^{-1}}^2 ,$$

which is the desired result. □

C.4 Proof of Lemma C.2

Lemma C.2. *Let $t \in \mathbb{N}$, $p \in \mathbb{N}^d$, then under $(t \in \mathcal{D}_p)$, $\sum_{k=1}^{s} \left\| \mathbf{d}_{A_k} \boldsymbol{\Sigma}^{*1/2} \right\|_{\mathbf{Z}_{k-1,p}^{-1}}^2 \leq 2d \log(2+\epsilon)$ for all $s \in [t]$.*

Proof. Let $t \in \mathbb{N}$, $p \in \mathbb{N}^d$, $s \in [t]$, $k \in [s]$.

We first observe that, under the event ($t \in \mathcal{D}_p$),

$$\begin{aligned}
\left\|\mathbf{d}_{A_k}\boldsymbol{\Sigma}^{*1/2}\right\|^2_{\mathbf{Z}^{-1}_{k-1,p}} &= \mathrm{Tr}\Big(\boldsymbol{\Sigma}^{*1/2}\mathbf{d}_{A_k}\mathbf{Z}^{-1}_{k-1,p}\mathbf{d}_{A_k}\boldsymbol{\Sigma}^{*1/2}\Big)\\
&= \mathrm{Tr}\Big(\boldsymbol{\Sigma}^{*1/2}\mathbf{d}_{A_k}(\mathbf{D}_p\mathbf{d}_{\boldsymbol{\Sigma}^*} + d\mathbf{d}_B + V_{k-1})^{-1}\mathbf{d}_{A_k}\boldsymbol{\Sigma}^{*1/2}\Big) \leftarrow \text{by Def. of } \mathbf{Z}_{k-1,p}\\
&= \mathrm{Tr}\Big(\boldsymbol{\Sigma}^{*1/2}\mathbf{d}_{A_k}\big(d\mathbf{d}_B\big)^{-1/2}\big(I + (d\mathbf{d}_B)^{1/2}\mathbf{D}_p\mathbf{d}_{\boldsymbol{\Sigma}^*}(d\mathbf{d}_B)^{1/2} + (d\mathbf{d}^*_{\boldsymbol{\Sigma}})^{1/2}V_{k-1}(d\mathbf{d}^*_{\boldsymbol{\Sigma}})^{1/2}\big)^{-1}\\
&\qquad\qquad \big(d\mathbf{d}_B\big)^{-1/2}\mathbf{d}_{A_k}\boldsymbol{\Sigma}^{*1/2}\Big)\\
&\leq \frac{1}{d}\mathrm{Tr}\Big(\boldsymbol{\Sigma}^{*1/2}\mathbf{d}_{A_k}\mathbf{d}_B^{-1}\mathbf{d}_{A_k}\boldsymbol{\Sigma}^{*1/2}\Big) \leftarrow \text{all the eigenvalues of the central matrix are} \leq 1\\
&= \frac{1}{d}\mathrm{Tr}(\boldsymbol{\Sigma}^{*1/2}\mathbf{d}_{A_k}\mathbf{d}_B^{-1}\mathbf{d}_{A_k}\mathbf{d}_{A_k}\boldsymbol{\Sigma}^{*1/2}) \leftarrow \text{ as } \mathbf{d}_{A_k} = \mathbf{d}^2_{A_k}\\
&= \frac{1}{d}\mathrm{Tr}(\mathbf{d}_{A_k}\boldsymbol{\Sigma}^*\mathbf{d}_{A_k}\mathbf{d}_B^{-1}\mathbf{d}_{A_s})\\
&= \frac{1}{d}\sum_{i\in A_k}\frac{\boldsymbol{\Sigma}^*_{i,i}}{B_i^2}\\
&\leq 1 \leftarrow \text{ as } \boldsymbol{\Sigma}^*_{i,i} \leq B_i^2\,.
\end{aligned}$$

Using that for $x \in [0,1], x \leq 2\log(1+x)$, we can derive

$$\begin{aligned}
\sum_{k=1}^{s}\left\|\mathbf{d}_{A_s}\boldsymbol{\Sigma}^{*1/2}\right\|^2_{\mathbf{Z}^{-1}_{k-1,p}} &\leq 2\sum_{k=1}^{s}\log\left(1 + \left\|\mathbf{d}_{A_k}\boldsymbol{\Sigma}^{*1/2}\right\|^2_{\mathbf{Z}^{-1}_{k-1,p}}\right)\\
&= 2\sum_{k=1}^{s}\log\left(1 + \mathrm{Tr}\Big(\boldsymbol{\Sigma}^{*1/2}\mathbf{d}_{A_k}\mathbf{Z}^{-1}_{k-1,p}\mathbf{d}_{A_k}\boldsymbol{\Sigma}^{*1/2}\Big)\right)\\
&= 2\sum_{k=1}^{s}\log\left(1 + \mathrm{Tr}\Big(\mathbf{Z}^{-1/2}_{k-1,p}\mathbf{d}_{A_k}\boldsymbol{\Sigma}^*\mathbf{d}_{A_k}\mathbf{Z}^{-1/2}_{k-1,p}\Big)\right) \leftarrow \mathrm{Tr}(AB) = \mathrm{Tr}(BA)\\
&\leq 2\sum_{k=1}^{s}\log\left(\det\Big(I + \mathbf{Z}^{-1/2}_{k-1,p}\mathbf{d}_{A_k}\boldsymbol{\Sigma}^*\mathbf{d}_{A_k}\mathbf{Z}^{-1/2}_{k-1,p}\Big)\right),
\end{aligned}$$

where the last inequality uses the fact that the eigenvalues of $\mathbf{Z}^{-1/2}_{k-1,p}\mathbf{d}_{A_k}\boldsymbol{\Sigma}^*\mathbf{d}_{A_k}\mathbf{Z}^{-1/2}_{k-1,p}$ are all non-negative. Therefore,

$$\begin{aligned}
\sum_{k=1}^{s}\left\|\mathbf{d}_{A_s}\boldsymbol{\Sigma}^{*1/2}\right\|^2_{\mathbf{Z}^{-1}_{k-1,p}} &\leq 2\sum_{k=1}^{s}\log\left(\det\Big(\mathbf{Z}^{-1/2}_{k-1,p}(\mathbf{Z}_{k-1,p} + \mathbf{d}_{A_k}\boldsymbol{\Sigma}^*\mathbf{d}_{A_k})\mathbf{Z}^{-1/2}_{k-1,p}\Big)\right) \leftarrow \det(AB) = \det(A)\det(B)\\
&= 2\sum_{k=1}^{s}\log\left(\frac{\det(\mathbf{Z}_{k-1,p} + \mathbf{d}_{A_k}\boldsymbol{\Sigma}^*\mathbf{d}_{A_k})}{\det(\mathbf{Z}_{k-1,p})}\right)\\
&= 2\sum_{k=1}^{s}\log\left(\frac{\det(\mathbf{Z}_{k,p})}{\det(\mathbf{Z}_{k-1,p})}\right)\\
&= 2\log\left(\frac{\det(\mathbf{Z}_{s,p})}{\det(\mathbf{Z}_{0,p})}\right) \leftarrow \text{ Telescoping sum.}\\
&= 2\log\left(\frac{\det(\mathbf{D}_p\mathbf{d}^*_{\boldsymbol{\Sigma}} + d\mathbf{d}_B + V_s)}{\det(\mathbf{D}_p\mathbf{d}^*_{\boldsymbol{\Sigma}} + d\mathbf{d}_B)}\right)\\
&= 2\log\Big(\det\big(I + (\mathbf{D}_p\mathbf{d}^*_{\boldsymbol{\Sigma}} + d\mathbf{d}_B)^{-1/2}V_s(\mathbf{D}_p\mathbf{d}^*_{\boldsymbol{\Sigma}} + d\mathbf{d}_B)^{-1/2}\big)\Big)\\
&\leq 2\log\left(\prod_{i=1}^{d}\left(1 + \frac{n_{s,i}\boldsymbol{\Sigma}^*_{i,i}}{(1+\epsilon)^{p_i}\boldsymbol{\Sigma}^*_{i,i} + dB_i^2}\right)\right),
\end{aligned}$$

by multiplying the diagonal elements of $\left(I + (\mathbf{D}_p\mathbf{d}_{\mathbf{\Sigma}}^* + d\mathbf{d}_B)^{-1/2}V_s(\mathbf{D}_p\mathbf{d}_{\mathbf{\Sigma}}^* + d\mathbf{d}_B)^{-1/2}\right)$, as it is a symmetric positive-definite matrix. Therefore, we get the result using the definition of $(t \in \mathcal{D}_p)$,

$$\sum_{k=1}^{s} \left\| \mathbf{d}_{A_s} \mathbf{\Sigma}^{*1/2} \right\|^2_{\mathbf{Z}^{-1}_{k-1,p}} \leq 2\log\left(\prod_{i=1}^{d} \left(1 + \frac{n_{s,i}}{(1+\epsilon)^{p_i}} \right) \right) \leq 2\log\left(\prod_{i=1}^{d} (2+\epsilon) \right) = 2d\log(2+\epsilon)\,.$$

□

C.5 Proof of Proposition C.3

Proposition C.3. *Let $t \in \mathbb{N}$, $p \in \mathbb{N}^d$, then for $x_I f_t \geq 4d\log\log(1+t) + 2d\log(2+\epsilon) + 4\log(1/\delta)$,*

$$\mathbb{P}\left\{ (M^2_{t,p} > f^2_{t,\delta}) \bigcap (\forall s \in [t], J_{s,p} \leq x_J f^2_{t,\delta}) \bigcap (t \in \mathcal{D}_p) \right\} \leq t\delta \left(\tfrac{\log(1+\epsilon)}{\log(1+t)} \right)^d .$$

Proof. Let $t \in \mathbb{N}$ and $p \in \mathbb{N}^d$. We begin by decomposing the probability into more "manageable pieces". For $s \in [t]$ we define $\mathcal{E}_{p,t,s} = \left\{ \forall k \in [s], M^2_{k,p} \leq f^2_{t,\delta} \right\}$, then

$$\begin{aligned}
&\mathbb{P}\Big\{ (M^2_{t,p} > f^2_{t,\delta}) \bigcap (\forall s \in [t], J_{s,p} \leq x_J f^2_{t,\delta}) \bigcap (t \in \mathcal{D}_p) \Big\} \\
&\qquad = \mathbb{P}\Big\{ (M^2_{t,p} > f^2_{t,\delta}) \bigcap (\mathcal{E}_{p,t,1}) \bigcap (\forall s \in [t], J_{s,p} \leq x_J f^2_{t,\delta}) \bigcap (t \in \mathcal{D}_p) \Big\} \\
&\qquad\qquad + \mathbb{P}\Big\{ (M^2_{t,p} > f^2_{t,\delta}) \bigcap (\mathcal{E}^c_{p,t,1}) \bigcap (\forall s \in [t], J_{s,p} \leq x_J f^2_{t,\delta}) \bigcap (t \in \mathcal{D}_p) \Big\} .
\end{aligned}$$

Iterating this operation on the term containing $\mathcal{E}_{p,t,k}$ for $k = 1, \ldots, t-1$

$$\begin{aligned}
&\mathbb{P}\Big\{ (M^2_{t,p} > f^2_{t,\delta}) \bigcap (\forall s \in [t], J_{s,p} \leq x_J f^2_{t,\delta}) \bigcap (t \in \mathcal{D}_p) \Big\} \\
&\quad = \mathbb{P}\Big\{ (M^2_{t,p} > f^2_{t,\delta}) \bigcap (\forall s \in [t], \mathcal{E}_{p,t,s}) \bigcap (\forall s \in [t], J_{s,p} \leq x_J f^2_{t,\delta}) \bigcap (t \in \mathcal{D}_p) \Big\} \\
&\qquad + \sum_{k=0}^{t-1} \mathbb{P}\Big\{ (M^2_{t,p} > f^2_{t,\delta}) \bigcap (\forall s \in [k], \mathcal{E}_{p,t,s}) \bigcap (\mathcal{E}^c_{p,t,k+1}) \bigcap (\forall s \in [t], J_{s,p} \leq x_J f^2_{t,\delta}) \bigcap (t \in \mathcal{D}_p) \Big\} \\
&\quad = \sum_{k=0}^{t-1} \mathbb{P}\Big\{ (M^2_{t,p} > f^2_{t,\delta}) \bigcap (\forall s \in [k], \mathcal{E}_{p,t,s}) \bigcap (\mathcal{E}^c_{p,t,k+1}) \bigcap (\forall s \in [t], J_{s,p} \leq x_J f^2_{t,\delta}) \bigcap (t \in \mathcal{D}_p) \Big\} ,
\end{aligned}$$

as $(M^2_{t,p} > f^2_{t,\delta})$ and $\mathcal{E}_{p,t,t,}$ are incompatible events. Then,

$$\begin{aligned}
&\mathbb{P}\Big\{ (M^2_{t,p} > f^2_{t,\delta}) \bigcap (\forall s \in [t], J_{s,p} \leq x_J f^2_{t,\delta}) \bigcap (t \in \mathcal{D}_p) \Big\} \\
&\qquad \leq \sum_{k=0}^{t-1} \mathbb{P}\Big\{ (\forall s \in [k], \mathcal{E}_{p,t,s}) \bigcap (\mathcal{E}^c_{p,t,k+1}) \bigcap (\forall s \in [t], J_{s,p} \leq x_J f^2_{t,\delta}) \bigcap (t \in \mathcal{D}_p) \Big\} \\
&\qquad \leq \sum_{k=0}^{t-1} \mathbb{P}\Big\{ (M^2_{k+1,p} > f^2_{t,\delta}) \bigcap (\forall s \in [k], \mathcal{E}_{p,t,s}) \bigcap (\forall s \in [t], J_{s,p} \leq x_J f^2_{t,\delta}) \bigcap (t \in \mathcal{D}_p) \Big\} ,
\end{aligned}$$

by definition of $\Big\{ (\forall s \in [k], \mathcal{E}_{p,t,s}) \bigcap (\mathcal{E}^c_{p,t,k+1}) \Big\}$. And finally, using $M_{k+1,p} \leq I_{k+1,p} + J_{k+1,p}$,

$$\begin{aligned}
&\mathbb{P}\Big\{ (M^2_{t,p} > f^2_{t,\delta}) \bigcap (\forall s \in [t], J_{s,p} \leq x_J f^2_{t,\delta}) \bigcap (t \in \mathcal{D}_p) \Big\} \\
&\qquad \leq \sum_{k=0}^{t-1} \mathbb{P}\Big\{ (I_{k+1,p} + J_{k+1,p} > f^2_{t,\delta}) \bigcap (\forall s \in [k], \mathcal{E}_{p,t,s}) \bigcap (J_{k+1,p} \leq x_J f^2_{t,\delta}) \bigcap (t \in \mathcal{D}_p) \Big\} \\
&\qquad \leq \sum_{k=0}^{t-1} \mathbb{P}\Big\{ (I_{k+1,p} > x_I f^2_{t,\delta}) \bigcap (\forall s \in [k], \mathcal{E}_{p,t,s}) \bigcap (t \in \mathcal{D}_p) \Big\} .
\end{aligned} \tag{35}$$

For $k=0$, as $x_I f_{t,\delta}^2 > 0$ and $I_{1,p}=0$,

$$\mathbb{P}\Big\{(I_{1,p} > x_I f_{t,\delta}^2) \bigcap (\forall s \in [k], \mathcal{E}_{p,t,s}) \bigcap (t \in \mathcal{D}_p)\Big\} = 0\,. \tag{36}$$

For $k \geq 1$, under the event $\Big\{(\forall s \in [k],\ \mathcal{E}_{p,t,s})\Big\}$,

$$I_{k+1,p} = 2\sum_{s=1}^{k+1}(\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1}(\sum_{l=1}^{s-1}\mathbf{d}_{A_l}\eta_l) = 2f_{t,\delta}\sum_{s=1}^{k+1}(\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1}(\sum_{l=1}^{s-1}\mathbf{d}_{A_l}\eta_l)\frac{\mathbb{1}\{\mathcal{E}_{p,t,s-1}\}}{f_{t,\delta}}\,. \tag{37}$$

We then use the following Lemma.

Lemma C.5. *(Proposition 2.10 from Wainwright, 2019) Let X be a centered random variable, bounded by $b \in \mathbb{R}^{*+}$, with variance $\sigma^2 \in \mathbb{R}^{*+}$. Then, for all $|\lambda| \leq \frac{1}{2b}$, we have $\mathbb{E}\big[\exp(\lambda X - \lambda^2\sigma^2)\big] \leq 1$.*

Let $|\lambda| \leq \frac{1}{2}$. As for all $s \in [k+1]$,

$$\left|(\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1}(\sum_{l=1}^{s-1}\mathbf{d}_{A_l}\eta_l)\frac{\mathbb{1}\{\mathcal{E}_{p,t,s-1}\}}{f_{t,\delta}}\right| \leq \|\mathbf{d}_{A_s}\eta_s\|_{\mathbf{Z}_{s-1,p}^{-1}}\frac{M_{s-1,p}}{f_{t,\delta}}\mathbb{1}\{\mathcal{E}_{p,t,s-1}\} \leq 1,$$

$$\mathbb{E}\left[(\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1}(\sum_{l=1}^{s-1}\mathbf{d}_{A_l}\eta_l)\frac{\mathbb{1}\{\mathcal{E}_{p,t,s-1}\}}{f_{t,\delta}}\middle|\mathcal{F}_{s-1}\right] = 0,$$

$$\mathbb{E}\left[\left((\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1}\Big(\sum_{l=1}^{s-1}\mathbf{d}_{A_l}\eta_l\Big)\frac{\mathbb{1}\{\mathcal{E}_{p,t,s-1}\}}{f_{t,\delta}}\right)^2\middle|\mathcal{F}_{s-1}\right] \leq \left\|\mathbf{Z}_{s-1,p}^{-1/2}\mathbf{d}_{A_s}(\mathbf{\Sigma}^*)^{1/2}\right\|^2.$$

Then, by Lemma C.5

$$\mathbb{E}\left[\exp\left(\lambda(\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1}\Big(\sum_{l=1}^{s-1}\mathbf{d}_{A_l}\eta_l\Big)\frac{\mathbb{1}\{\mathcal{E}_{p,t,s-1}\}}{f_{t,\delta}} - \lambda^2\|\mathbf{Z}_{s-1,p}^{-1/2}\mathbf{d}_{A_s}(\mathbf{\Sigma}^*)^{1/2}\|^2\right)\middle|\mathcal{F}_{s-1}\right] \leq 1\,.$$

Iterating over $s = 1, \ldots, k+1$ results in

$$\mathbb{E}\left[\exp\left(\lambda\sum_{s=1}^{k+1}(\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1}(\sum_{l=1}^{s-1}\mathbf{d}_{A_l}\eta_l)\frac{\mathbb{1}\{\mathcal{E}_{p,t,s-1}\}}{f_{t,\delta}} - \lambda^2\sum_{s=1}^{k+1}\|\mathbf{Z}_{s-1,p}^{-1/2}\mathbf{d}_{A_s}(\mathbf{\Sigma}^*)^{1/2}\|^2\right)\right] \leq 1\,. \tag{38}$$

However, this inequality cannot be exploited readily in this form as we need the event $(t \in \mathcal{D}_p)$ and Lemma C.2 to incorporate the dimension d,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda\sum_{s=1}^{k+1}(\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1}\Big(\sum_{l=1}^{s-1}\mathbf{d}_{A_l}\eta_l\Big)\frac{\mathbb{1}\{\mathcal{E}_{p,t,s-1}\}}{f_{t,\delta}} - 2d\lambda^2\log(2+\epsilon)\right)\mathbb{1}\{t \in \mathcal{D}_p\}\right] \\
&\leq \mathbb{E}\left[\exp\left(\lambda\sum_{s=1}^{k+1}(\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1}\Big(\sum_{l=1}^{s-1}\mathbf{d}_{A_l}\eta_l\Big)\frac{\mathbb{1}\{\mathcal{E}_{p,t,s-1}\}}{f_{t,\delta}} - \lambda^2\sum_{s=1}^{k+1}\|\mathbf{Z}_{s-1,p}^{-1/2}\mathbf{d}_{A_s}(\mathbf{\Sigma}^*)^{1/2}\|^2\right)\mathbb{1}\{t \in \mathcal{D}_p\}\right] \leftarrow \text{ Lemma C.2} \\
&\leq \mathbb{E}\left[\exp\left(\lambda\sum_{s=1}^{k+1}(\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1}\Big(\sum_{l=1}^{s-1}\mathbf{d}_{A_l}\eta_l\Big)\frac{\mathbb{1}\{\mathcal{E}_{p,t,s-1}\}}{f_{t,\delta}} - \lambda^2\sum_{s=1}^{k+1}\|\mathbf{Z}_{s-1,p}^{-1/2}\mathbf{d}_{A_s}(\mathbf{\Sigma}^*)^{1/2}\|^2\right)\right] \leftarrow \text{ as } \mathbb{1}\{t \in \mathcal{D}_p\} \leq 1 \\
&\leq 1\,. \leftarrow \text{ by Eq. (38)}
\end{aligned}$$

Now, taking $\lambda = \frac{1}{2}$, we can bound

$$
\begin{aligned}
&\mathbb{P}\big((I_{k+1,p} > x_I f_{t,\delta}^2) \bigcap (\forall s \in [k], \mathcal{E}_{p,t,s}) \bigcap (t \in \mathcal{D}_p)\big) \\
&\leq \mathbb{P}\bigg(\Big(2 f_{t,\delta} \sum_{s=1}^{k+1} (\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1} \Big(\sum_{l=1}^{s-1} \mathbf{d}_{A_l}\eta_l\Big) \frac{\mathbb{1}\{\mathcal{E}_{p,t,s-1}\}}{f_{t,\delta}} > x_I f_{t,\delta}^2\Big) \bigcap (t \in \mathcal{D}_p)\bigg) \leftarrow \text{ by Eq. (37)} \\
&= \mathbb{P}\bigg(\Big(\frac{1}{2} \sum_{s=1}^{k+1} (\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1} \Big(\sum_{l=1}^{s-1} \mathbf{d}_{A_l}\eta_l\Big) \frac{\mathbb{1}\{\mathcal{E}_{p,t,s-1}\}}{f_{t,\delta}} > \frac{x_I}{4} f_{t,\delta}\Big) \bigcap (t \in \mathcal{D}_p)\bigg) \\
&= \mathbb{P}\bigg(\Big(\exp\Big(\frac{1}{2} \sum_{s=1}^{k+1} (\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1} \big(\sum_{l=1}^{s-1} \mathbf{d}_{A_l}\eta_l\big) \frac{\mathbb{1}\{\mathcal{E}_{p,t,s-1}\}}{f_{t,\delta}} - 2d\frac{1}{4}\log(2+\epsilon)\Big) \\
&\qquad\qquad > \exp\Big(\frac{x_I}{4} f_{t,\delta} - 2d\frac{1}{4}\log(2+\epsilon)\Big)\Big) \bigcap (t \in \mathcal{D}_p)\bigg) \\
&\leq \mathbb{P}\bigg(\exp\Big(\frac{1}{2} \sum_{s=1}^{k+1} (\mathbf{d}_{A_s}\eta_s)^\top \mathbf{Z}_{s-1,p}^{-1} \Big(\sum_{l=1}^{s-1} \mathbf{d}_{A_l}\eta_l\Big) \frac{\mathbb{1}\{\mathcal{E}_{p,t,s-1}\}}{f_{t,\delta}} - \frac{d}{2}\log(2+\epsilon)\Big) \mathbb{1}\{t \in \mathcal{D}_p\} \\
&\qquad\qquad > \exp\Big(\frac{x_I}{4} f_{t,\delta} - \frac{d}{2}\log(2+\epsilon)\Big)\bigg) \\
&\leq \exp\Big(\frac{d}{2}\log(2+\epsilon) - \frac{x_I}{4} f_{t,\delta}\Big). \leftarrow \text{ by Markov ineq.}
\end{aligned}
$$

Taking $f_{t,\delta} \geq \frac{4}{x_I}(d \log\log(1+t) + \frac{d}{2}\log(2+\epsilon) + \log(1/\delta))$, we have

$$
\mathbb{P}\Big((I_{k+1,p} > x_I f_{t,\delta}^2) \bigcap (\forall s \in [k], \mathcal{E}_{p,t,s}) \bigcap (t \in \mathcal{D}_p)\Big) \leq \delta \Big(\frac{\log(1+\epsilon)}{\log(1+t)}\Big)^d,
$$

and the desired result,

$$
\begin{aligned}
&\mathbb{P}\big\{(M_{t,p}^2 > f_{t,\delta}^2) \bigcap (\forall s \in [t], J_{s,p} \leq x_J f_{t,\delta}^2) \bigcap (t \in \mathcal{D}_p)\big\} \\
&\leq \sum_{k=0}^{t-1} \mathbb{P}\Big\{(I_{k+1,p} > x_I f_{t,\delta}^2) \bigcap (\forall s \in [k], \mathcal{E}_{p,t,s}) \bigcap (t \in \mathcal{D}_p)\Big\} \leftarrow \text{ by Eq. (35)} \\
&\leq t\delta \Big(\frac{\log(1+\epsilon)}{\log(1+t)}\Big)^d.
\end{aligned}
$$

□

C.6 Proof of Proposition C.4

Proposition C.4. *Let $t \in \mathbb{N}$, $p \in \mathbb{N}^d$, then if $x_J f_t^2 \geq 2d \log\log(1+t) + 3d\log(2+\epsilon) + 2\log(1/\delta)$,*

$$
\mathbb{P}\Big\{(\exists s \in [t], J_{s,p} > x_J f_t^2) \bigcap (t \in \mathcal{D}_p)\Big\} \leq t\delta \Big(\tfrac{\log(1+\epsilon)}{\log(1+t)}\Big)^d.
$$

Proof. Let $t \in \mathbb{N}$, $p \in \mathbb{N}^d$. Then

$$
\mathbb{P}\Big\{(\exists s \in [t], J_{s,p} > x_J f_t^2) \bigcap (t \in \mathcal{D}_p)\Big\} \leq \sum_{s=1}^{t} \mathbb{P}\Big\{(J_{s,p} > x_J f_t^2) \bigcap (t \in \mathcal{D}_p)\Big\}. \tag{39}
$$

Let $s \in [t]$. By definition, $J_{s,t} = \sum_{k=1}^{s} \|\mathbf{d}_{A_k}\eta_k\|^2_{\mathbf{Z}_{k-1,p}^{-1}}$.

For $k \in [s]$,

$$\begin{aligned}
\|\mathbf{d}_{A_k}\eta_k\|^2_{\mathbf{Z}^{-1}_{k-1,p}} &\leq 1\,,\\
\mathbb{E}\Big[\|\mathbf{d}_{A_k}\eta_k\|^2_{\mathbf{Z}^{-1}_{k-1,p}}\Big|\mathcal{F}_{k-1}\Big] &= \mathbb{E}\Big[\eta_k^\top \mathbf{d}_{A_k}\mathbf{Z}^{-1}_{k-1,p}\mathbf{d}_{A_k}\eta_k\Big|\mathcal{F}_{k-1}\Big]\\
&= \mathbb{E}\Big[\mathrm{Tr}(\mathbf{Z}^{-1/2}_{k-1,p}\mathbf{d}_{A_k}\eta_k\eta_k^\top\mathbf{d}_{A_k}\mathbf{Z}^{-1/2}_{k-1,p})\Big|\mathcal{F}_{k-1}\Big] \leftarrow \mathrm{Tr}(AB)=\mathrm{Tr}(BA)\\
&= \mathrm{Tr}\Big(\mathbf{Z}^{-1/2}_{k-1,p}\mathbf{d}_{A_k}\mathbf{\Sigma}^*\mathbf{d}_{A_k}\mathbf{Z}^{-1/2}_{k-1,p}\Big)\\
&= \|\mathbf{d}_{A_k}(\mathbf{\Sigma}^*)^{1/2}\|^2_{\mathbf{Z}^{-1}_{k-1,p}}\,,\\
\mathbb{E}\Big[\|\mathbf{d}_{A_k}\eta_k\|^4_{\mathbf{Z}^{-1}_{k-1,p}}\Big|\mathcal{F}_{k-1}\Big] - \mathbb{E}\Big[\|\mathbf{d}_{A_k}\eta_k\|^2_{\mathbf{Z}^{-1}_{k-1,p}}\Big|\mathcal{F}_{k-1}\Big]^2 &\leq \mathbb{E}\Big[\|\mathbf{d}_{A_k}\eta_k\|^4_{\mathbf{Z}^{-1}_{k-1,p}}\Big|\mathcal{F}_{k-1}\Big]\\
&\leq \mathbb{E}\Big[\|\mathbf{d}_{A_k}\eta_k\|^2_{\mathbf{Z}^{-1}_{k-1,p}}|\mathcal{F}_{k-1}\Big] \leftarrow \text{ as } \|\mathbf{d}_{A_k}\eta_k\|^2_{\mathbf{Z}^{-1}_{k-1,p}}\leq 1\\
&= \|\mathbf{d}_{A_k}(\mathbf{\Sigma}^*)^{1/2}\|^2_{\mathbf{Z}^{-1}_{k-1,p}}\,.
\end{aligned}$$

We can then use Lemma C.5. Let $|\lambda| \leq \frac{1}{2}$, then

$$\begin{aligned}
&\mathbb{E}\Big[\exp\big(\lambda\|\mathbf{d}_{A_k}\eta_k\|^2_{\mathbf{Z}^{-1}_{k-1,p}} - \lambda(\lambda+1)\|\mathbf{Z}^{-1/2}_{k-1,p}\mathbf{d}_{A_k}(\mathbf{\Sigma}^*)^{1/2}\|^2\big)\Big|\mathcal{F}_{k-1}\Big]\\
&\quad= \mathbb{E}\Big[\exp\Big(\lambda\big(\|\mathbf{d}_{A_k}\eta_k\|^2_{\mathbf{Z}^{-1}_{k-1,p}} - \mathbb{E}\big[\|\mathbf{d}_{A_k}\eta_k\|^2_{\mathbf{Z}^{-1}_{k-1,p}}|\mathcal{F}_{k-1}\big]\big) - \lambda^2\|\mathbf{d}_{A_k}(\mathbf{\Sigma}^*)^{1/2}\|^2_{\mathbf{Z}^{-1}_{k-1,p}}\Big)\Big|\mathcal{F}_{k-1}\Big] \leq 1\,,
\end{aligned}$$

and, summing over k yields,

$$\mathbb{E}\Big[\exp\Big(\lambda\sum_{k=1}^s\|\mathbf{d}_{A_k}\eta_k\|^2_{\mathbf{Z}^{-1}_{k-1,p}} - \lambda(\lambda+1)\sum_{k=1}^s\|\mathbf{d}_{A_k}(\mathbf{\Sigma}^*)^{1/2}\|^2_{\mathbf{Z}^{-1}_{k-1,p}}\Big)\Big] \leq 1\,.$$

We can now incorporate the event $(t \in \mathcal{D}_p)$ to get the inequality

$$\begin{aligned}
\mathbb{E}\Big[\exp\Big(\lambda J_{s,p} - \lambda(\lambda+1)2d\log(2+\epsilon)\Big)\mathbb{1}\{t\in\mathcal{D}_p\}\Big] &\leftarrow \text{ by Lemma C.2}\\
&\leq \mathbb{E}\Big[\exp\Big(\lambda J_{s,p} - \lambda(\lambda+1)\sum_{k=1}^s\|\mathbf{d}_{A_k}(\mathbf{\Sigma}^*)^{1/2}\|^2_{\mathbf{Z}^{-1}_{k-1,p}}\Big)\mathbb{1}\{t\in\mathcal{D}_p\}\Big]\\
&\leq \mathbb{E}\Big[\exp\Big(\lambda J_{s,p} - \lambda(\lambda+1)\sum_{k=1}^s\|\mathbf{d}_{A_k}(\mathbf{\Sigma}^*)^{1/2}\|^2_{\mathbf{Z}^{-1}_{k-1,p}}\Big)\Big]\\
&\leq 1\,.
\end{aligned}$$

Then, using Markov inequality,

$$\begin{aligned}
\mathbb{P}\Big\{(J_{s,p} > x_J f_t^2)\bigcap(t\in\mathcal{D}_p)\Big\} &= \mathbb{P}\Big\{\Big(\exp\big(\frac{1}{2}J_{s,p} - \frac{3}{4}2d\log(2+\epsilon)\big) > \exp\big(\frac{1}{2}x_J f_t^2 - \frac{3}{4}2d\log(1+\epsilon)\big)\Big)\bigcap(t\in\mathcal{D}_p)\Big\}\\
&\leq \mathbb{P}\Big\{\exp\Big(\frac{1}{2}J_{s,p} - \frac{3}{4}2d\log(2+\epsilon)\Big)\mathbb{1}\{t\in\mathcal{D}_p\} > \exp\Big(\frac{1}{2}x_J f_t^2 - \frac{3}{4}2d\log(2+\epsilon)\Big)\Big\}\\
&\leq \exp\Big(\frac{3}{2}d\log(1+\epsilon) - \frac{1}{2}x_J f_t^2\Big)\\
&\leq \delta\Big(\frac{\log(2+\epsilon)}{\log(1+t)}\Big)^d\,,
\end{aligned} \tag{40}$$

for $f_t \geq \frac{1}{\sqrt{x_J}}\sqrt{3d\log(2+\epsilon) + 2d\log\log(1+t) + 2\log(1/\delta)}$.

Thus, we deduce the desired inequality,

$$\mathbb{P}\Big\{(\exists s\in[t], J_{s,p} > x_J f_t^2)\bigcap(t\in\mathcal{D}_p)\Big\} \le \sum_{s=1}^{t}\mathbb{P}\Big\{(J_{s,p} > x_J f_t^2)\bigcap(t\in\mathcal{D}_p)\Big\} \leftarrow \text{ by Eq. (39)}$$
$$\le t\delta\left(\frac{\log(1+\epsilon)}{\log(1+t)}\right)^d . \leftarrow \text{ by (40)}$$

□

C.7 Proof of Proposition 4.5

Proposition 4.5. *Let $t \ge d(d+1)$. Then under $\{\mathcal{G}_t \bigcap \mathcal{C}\}$, $\Delta_{A_{t+1}} \le f_{t,\delta}\big(\|\mathbf{D}_t^{-1}A_{t+1}\|_{\mathbf{Z}_t} + \|\mathbf{D}_t^{-1}A_{t+1}\|_{\hat{\mathbf{Z}}_t}\big)$.*

Proof. Let $t \ge d(d+1)$The error in estimating the mean reward for action a with $\langle a, \hat{\mu}_t\rangle$ is bounded as

$$\begin{aligned}\left|a^\top(\hat{\mu}_t - \mu)\right| &= \left|a^\top \mathbf{D}_t^{-1}\textstyle\sum_{s=1}^t \mathbf{d}_{A_s}\eta_s\right| \\ &= \left|a^\top \mathbf{D}_t^{-1}\mathbf{Z}_t^{1/2}\mathbf{Z}_t^{-1/2}\textstyle\sum_{s=1}^t \mathbf{d}_{A_s}\eta_s\right| \\ &\le \|\mathbf{D}_t^{-1}a\|_{\mathbf{Z}_t}\ \big\|\textstyle\sum_{s=1}^t \mathbf{d}_{A_s}\eta_s\big\|_{\mathbf{Z}_t^{-1}}, \end{aligned} \tag{41}$$

where the last line is by Cauchy–Schwarz' inequality.

We reminding the event where the empirical average $\hat{\mu}$ remains in our ellipsoid

$$\mathcal{G}_t = \left\{\ \big\|\textstyle\sum_{s=1}^t \mathbf{d}_{A_s}\eta_s\big\|_{\mathbf{Z}_t^{-1}} \le f_{t,\delta}\right\}.$$

On this event, injecting $a = A_{t+1}$ and $a = a^*$ into (41) yields

$$\langle A_{t+1}, \hat{\mu}_t\rangle \le \langle A_{t+1}, \mu\rangle + f_{t,\delta}\|\mathbf{D}_t^{-1}A_{t+1}\|_{\mathbf{Z}_t},$$

and

$$\langle a^*, \mu\rangle - f_{t,\delta}\|\mathbf{D}_t^{-1}a^*\|_{\mathbf{Z}_t} \le \langle a^*, \hat{\mu}_t\rangle.$$

Moreover, since by definition of A_{t+1} in (7) we have

$$\langle a^*, \hat{\mu}_t\rangle + f_{t,\delta}\|\mathbf{D}_t^{-1}a^*\|_{\hat{\mathbf{Z}}_t} \le \langle A_{t+1}, \hat{\mu}_t\rangle + f_{t,\delta}\|\mathbf{D}_t^{-1}A_{t+1}\|_{\hat{\mathbf{Z}}_t},$$

we can finally write

$$\begin{aligned}&\langle a^*, \mu\rangle + f_{t,\delta}(\|\mathbf{D}_t^{-1}a^*\|_{\hat{\mathbf{Z}}_t} - \|\mathbf{D}_t^{-1}a^*\|_{\mathbf{Z}_t}) \\ &\quad\le \langle A_{t+1}, \mu\rangle + f_{t,\delta}\|\mathbf{D}_t^{-1}A_{t+1}\|_{\mathbf{Z}_t} + f_{t,\delta}\|\mathbf{D}_t^{-1}A_{t+1}\|_{\hat{\mathbf{Z}}_t}.\end{aligned} \tag{42}$$

Besides, under $\mathcal{C}$, by the definition in 10, $\hat{Z}_t$ uses coefficient-wise upper bounds of $\mathbf{\Sigma}^*$, which yields for all $a \in \mathcal{A}$

$$\|\mathbf{D}_t^{-1}a^*\|_{\mathbf{Z}_t}^2 \le \|\mathbf{D}_t^{-1}a^*\|_{\hat{\mathbf{Z}}_t}^2.$$

Injecting this in (42) and denoting $\Delta_{A_{t+1}} = \langle a^*, \mu\rangle - \langle A_{t+1}, \mu\rangle$ concludes the proof. □

C.8 Proof of Lemma 4.6

Lemma 4.6. *Let $t \ge d(d+1)$, we have that under $\{\mathcal{G}_t \bigcap \mathcal{C}\}$,*

$$\begin{aligned}\frac{\Delta_{A_{t+1}}^2}{4f_{t,\delta}^2} &\le \frac{1}{2}\big(\|\mathbf{D}_t^{-1}A_{t+1}\|_{\mathbf{Z}_t}^2 + \|\mathbf{D}_t^{-1}A_{t+1}\|_{\hat{Z}_t}^2\big) \\ &\le \sum_{i\in A_{t+1}}\frac{\bar{\sigma}_{A_{t+1},i}^2}{n_{t,(i,i)}} + \big(d + 3\log(T)h_{T,\delta}^2\big)\sum_{(i,j)\in A_{t+1}}\frac{B_iB_j}{n_{t,(i,j)}^2} + 3h_{T,\delta}\sum_{(i,j)\in A_{t+1}}\frac{B_iB_j}{n_{t,(i,j)}^{3/2}},\end{aligned} \tag{18}$$

where

$$\bar{\sigma}_{A_{t+1},i}^2 = 2\sum_{j\in A_{t+1}/\mathbf{\Sigma}_{j,j}^*\le\mathbf{\Sigma}_{i,i}^*}(\mathbf{\Sigma}_{i,j}^*)_+ \le 2\sigma_{A_{t+1},i}^2.$$

Proof. Let $t \geq d(d+1)$, then Proposition 4.5 yields that under $\{\mathcal{G}_t \bigcap \mathcal{C}\}$, $\Delta_{A_{t+1}} \leq f_{t,\delta}\big(\|\mathbf{D}_t^{-1}A_{t+1}\|_{\mathbf{Z}_t} + \|\mathbf{D}_t^{-1}A_{t+1}\|_{\hat{\mathbf{Z}}_t}\big)$, and

$$\begin{aligned}\Delta_{A_{t+1}}^2 &\leq f_{t,\delta}^2\big(\|\mathbf{D}_t^{-1}A_{t+1}\|_{\mathbf{Z}_t} + \|\mathbf{D}_t^{-1}A_{t+1}\|_{\hat{\mathbf{Z}}_t}\big)^2 \\ &\leq 2f_{t,\delta}^2\big(\|\mathbf{D}_t^{-1}A_{t+1}\|_{\mathbf{Z}_t}^2 + \|\mathbf{D}_t^{-1}A_{t+1}\|_{\hat{\mathbf{Z}}_t}^2\big) \\ \frac{\Delta_{A_{t+1}}^2}{2f_{t\delta}^2} &\leq \|\mathbf{D}_t^{-1}A_{t+1}\|_{\mathbf{Z}_t}^2 + \|\mathbf{D}_t^{-1}A_{t+1}\|_{\hat{\mathbf{Z}}_t}^2 \,. \end{aligned} \tag{43}$$

From, here, we just need to develop the right-hand side of this inequality.

$$\begin{aligned}\|\mathbf{D}_t^{-1}A_{t+1}\|_{\mathbf{Z}_t}^2 &= A_{t+1}^\top \mathbf{D}_t^{-1}\mathbf{Z}_t\mathbf{D}_t^{-1}A_{t+1} \\ &= \sum_{(i,j)\in A_{t+1}} \frac{(\mathbf{Z}_t)_{i,j}}{n_{t,(i,i)}n_{t,(j,j)}} \,.\end{aligned}$$

As $\mathbf{Z}_t = \sum_{s=1}^t \mathbf{d}_{A_s}\mathbf{\Sigma}^*\mathbf{d}_{A_s} + \mathbf{d}_{\mathbf{\Sigma}^*}\mathbf{D}_t + d\mathbf{d}_B$, we get

$$\begin{aligned}\|\mathbf{D}_t^{-1}A_{t+1}\|_{\mathbf{Z}_t}^2 &= \sum_{(i,j)\in A_{t+1}} \frac{n_{t,(i,j)}\mathbf{\Sigma}^*_{i,j}}{n_{t,(i,i)}n_{t,(j,j)}} + \sum_{i\in A_{t+1}} \frac{n_{t(i,i)}\mathbf{\Sigma}^*_{i,i}}{n_{t,(i,i)}^2} + d\sum_{i\in A_{t+1}} \frac{B_i^2}{n_{t,(i,i)}^2} \\ &\leq \sum_{i\in A_{t+1}} \Big(2\sum_{j\in A_{t+1}/\mathbf{\Sigma}^*_{j,j}\leq\mathbf{\Sigma}^*_{i,i}} \frac{n_{t,(i,j)}\mathbf{\Sigma}^*_{i,j}}{n_{t,(i,i)}n_{t,(j,j)}}\Big) + d\sum_{i\in A_{t+1}} \frac{B_i^2}{n_{t,(i,i)}^2} \,,\end{aligned}$$

by rearranging terms. Now as for all $(i,j)\in[d]^2$, $n_{t,(i,j)} \leq \max\{n_{t,(i,i)}, n_{t,(j,j)}\}$, then

$$\|\mathbf{D}_t^{-1}A_{t+1}\|_{\mathbf{Z}_t}^2 \leq \sum_{i\in A_{t+1}} \frac{1}{n_{t,(i,i)}}\Big(2\sum_{j\in A_{t+1}/\mathbf{\Sigma}^*_{j,j}\leq\mathbf{\Sigma}^*_{i,i}} \mathbf{\Sigma}^*_{i,j}\Big) + d\sum_{i\in A_{t+1}} \frac{B_i^2}{n_{t,(i,i)}^2} \,,$$

Denoting $\bar{\sigma}^2_{A_{t+1},i} = 2\sum_{j\in A_{t+1}/\mathbf{\Sigma}^*_{j,j}\leq\mathbf{\Sigma}^*_{i,i}} (\mathbf{\Sigma}^*_{i,j})_+$ yields

$$\|\mathbf{D}_t^{-1}A_{t+1}\|_{\mathbf{Z}_t}^2 \leq \sum_{i\in A_{t+1}} \frac{\bar{\sigma}^2_{A_{t+1},i}}{n_{t,(i,i)}} + d\sum_{i\in A_{t+1}} \frac{B_i^2}{n_{t,(i,i)}^2} \,. \tag{44}$$

The second term from the right-hand side of Equation (43) is developed in the same manner.

$$\begin{aligned}\|\mathbf{D}_t^{-1}A_{t+1}\|_{\hat{\mathbf{Z}}_t}^2 &= A_{t+1}^\top \mathbf{D}_t^{-1}\hat{\mathbf{Z}}_t\mathbf{D}_t^{-1}A_{t+1} \\ &= \sum_{(i,j)\in A_{t+1}} \frac{(\hat{\mathbf{Z}}_t)_{i,j}}{n_{t,(i,i)}n_{t,(j,j)}} \,.\end{aligned}$$

We remind that $\hat{\mathbf{Z}}_t = \sum_{s=1}^t \mathbf{d}_{A_s}\hat{\mathbf{\Sigma}}_t\mathbf{d}_{A_s} + \mathbf{d}_{\hat{\mathbf{\Sigma}}_t}\mathbf{D}_t + d\mathbf{d}_B$ where for all $(i,j)\in[d]^2$, $\hat{\mathbf{\Sigma}}_{t,(i,j)} = \hat{\chi}_{t,(i,j)} + \mathcal{B}_{t,(i,j)}(\delta,T)$. Being under the event $\mathcal{C}$, Proposition 3.1 yields that $\hat{\mathbf{\Sigma}}_{t,(i,j)} \leq \mathbf{\Sigma}^*_{i,j} + 2\mathcal{B}_{t,(i,j)}(\delta,T)$, therefore,

$$\begin{aligned}\|\mathbf{D}_t^{-1}A_{t+1}\|_{\hat{\mathbf{Z}}_t}^2 &\leq \sum_{(i,j)\in A_{t+1}} \frac{n_{t,(i,j)}\mathbf{\Sigma}^*_{i,j}}{n_{t,(i,i)}n_{t,(j,j)}} + \sum_{(i,j)\in A_{t+1}} \frac{n_{t,(i,j)}2\mathcal{B}_{t,(i,j)}(\delta,T)}{n_{t,(i,i)}n_{t,(j,j)}} + \sum_{i\in A_{t+1}} \frac{n_{t(i,i)}\mathbf{\Sigma}^*_{i,i}}{n_{t,(i,i)}^2} + d\sum_{i\in A_{t+1}} \frac{B_i^2}{n_{t,(i,i)}^2} \\ &\leq \sum_{i\in A_{t+1}} \frac{\bar{\sigma}^2_{A_{t+1},i}}{n_{t,(i,i)}} + \sum_{(i,j)\in A_{t+1}} \frac{n_{t,(i,j)}2\mathcal{B}_{t,(i,j)}(\delta,T)}{n_{t,(i,i)}n_{t,(j,j)}} + d\sum_{i\in A_{t+1}} \frac{B_i^2}{n_{t,(i,i)}^2} \,.\end{aligned}$$

Reminding, $\mathcal{B}_{t,(i,j)}(\delta,T) = 3B_iB_j\left(\frac{h_{T,\delta}}{\sqrt{n_{t,(i,j)}}} + \frac{h_{T,\delta}^2}{n_{t,(i,j)}}\log(T)\right)$,

$$\begin{aligned}\|\mathbf{D}_t^{-1}A_{t+1}\|_{\hat{\mathbf{Z}}_t}^2 \leq \sum_{i\in A_{t+1}} \frac{\bar{\sigma}^2_{A_{t+1},i}}{n_{t,(i,i)}} + d\sum_{i\in A_{t+1}} \frac{B_i^2}{n_{t,(i,i)}^2} \\ + \sum_{(i,j)\in A_{t+1}} \frac{n_{t,(i,j)}6B_iB_jh_{T,\delta}}{n_{t,(i,i)}n_{t,(j,j)}\sqrt{n_{t,(i,j)}}} + \sum_{(i,j)\in A_{t+1}} \frac{n_{t,(i,j)}6B_iB_jh_{T,\delta}^2\log(T)}{n_{t,(i,i)}n_{t,(j,j)}n_{t,(i,j)}} \,.\end{aligned}$$

As $n_{t,(i,j)} \leq \sqrt{n_{t,(i,i)} n_{t,(j,j)}}$ (Cauchy–Shwarz),

$$\|\mathbf{D}_t^{-1} A_{t+1}\|_{\hat{\mathbf{Z}}_t}^2 \leq \sum_{i \in A_{t+1}} \frac{\bar{\sigma}^2_{A_{t+1},i}}{n_{t,(i,i)}} + d \sum_{i \in A_{t+1}} \frac{B_i^2}{n_{t,(i,i)}^2} + \sum_{(i,j) \in A_{t+1}} \frac{6 B_i B_j h_{T,\delta}}{n_{t,(i,j)}^{3/2}} + \sum_{(i,j) \in A_{t+1}} \frac{6 B_i B_j h_{T,\delta}^2 \log(T)}{n_{t,(i,j)}^2} \,. \tag{45}$$

Reinjecting Equation (44) and Equation (45) into Equation (43) yields

$$\begin{aligned}
\frac{\Delta^2_{A_{t+1}}}{2 f^2_{t\delta}} &\leq \|\mathbf{D}_t^{-1} A_{t+1}\|_{\mathbf{Z}_t}^2 + \|\mathbf{D}_t^{-1} A_{t+1}\|_{\hat{\mathbf{Z}}_t}^2 \\
&\leq 2 \sum_{i \in A_{t+1}} \frac{\bar{\sigma}^2_{A_{t+1},i}}{n_{t,(i,i)}} + 2d \sum_{i \in A_{t+1}} \frac{B_i^2}{n_{t,(i,i)}^2} \\
&\qquad + \sum_{(i,j) \in A_{t+1}} \frac{6 B_i B_j h_{T,\delta}}{n_{t,(i,j)}^{3/2}} + \sum_{(i,j) \in A_{t+1}} \frac{6 B_i B_j h_{T,\delta}^2 \log(T)}{n_{t,(i,j)}^2} \\
&\leq 2 \sum_{i \in A_{t+1}} \frac{\bar{\sigma}^2_{A_{t+1},i}}{n_{t,(i,i)}} + \big(2d + 6 h_{T,\delta}^2 \log(T)\big) \sum_{(i,j) \in A_{t+1}} \frac{B_i B_j}{n_{t,(i,i)}^2} + 6 h_{T,\delta} \sum_{(i,j) \in A_{t+1}} \frac{B_i B_j}{n_{t,(i,j)}^{3/2}} \,,
\end{aligned}$$

thus the desired inequality,

$$\frac{\Delta_{A^2_{t+1}}}{4 f^2_{T,\delta}} \leq \sum_{i \in A_{t+1}} \frac{\bar{\sigma}^2_{A_{t+1},i}}{n_{t,(i,i)}} + \big(d + 3 h_{T,\delta}^2 \log(T)\big) \sum_{(i,j) \in A_{t+1}} \frac{B_i B_j}{n_{t,(i,i)}^2} + 3 h_{T,\delta} \sum_{(i,j) \in A_{t+1}} \frac{B_i B_j}{n_{t,(i,j)}^{3/2}} \,.$$

□

C.9 Definition of the sequences (α_k) and (β_k)

Let $\beta = 1/5$, $x > 0$. We define $\beta_0 = \alpha_0 = 1$. For $k \geq 1$, we define

$$\beta_k = \beta^k, \qquad \alpha_k = x \beta^k \,. \tag{46}$$

Let's first look for an adequate k_0 for Proposition 4.7. As for $a \in \mathcal{A}$ and $i \in a$, $\mathbf{\Sigma}^*_{i,i} \bar{\sigma}^{-2}_{a,i} \geq \frac{1}{2d}$ (by definition of $\bar{\sigma}^2_{a,i}$), taking $k_0 = \lceil \frac{2 \log(\sqrt{2} d)}{\log(1/\beta)} + 1 \rceil$ is sufficient to have $0 < d\beta_{k_0} < \big(\frac{1}{d} \wedge \min_{i,a} \{\mathbf{\Sigma}^*_{i,i} \bar{\sigma}^{-2}_{a,i}\}\big)$. This choice yields

$$\begin{aligned}
\left(\sum_{k=1}^{k_0 - 1} \frac{\beta_{k-1} - \beta_k}{\alpha_k} + \frac{\beta_{k_0 - 1}}{\alpha_{k_0}} \right) &= \left(\sum_{k=1}^{k_0 - 1} \frac{1 - \beta}{\beta} + \frac{1}{\beta} \right) \frac{1}{x} \\
&= \left((k_0 - 1) \frac{1 - \beta}{\beta} + \frac{1}{\beta} \right) \frac{1}{x} \\
&= \left(4 k_0 + 1 \right) \frac{1}{x} \\
&\leq 1 \,,
\end{aligned} \tag{47}$$

for $x = 4k_0 + 1$.

Besides, denoting $\lesssim$ for a rough inequality up to universal multiplicative constants,

$$\begin{aligned}
\sum_{k=1}^{k_0} \frac{\alpha_k}{\beta_k} &= (4k_0 + 1) k_0 \\
&\lesssim \log(d)^2 \,,
\end{aligned} \tag{48}$$

$$\begin{aligned}
\sum_{k=1}^{k_0} \frac{\sqrt{\alpha_k}}{\beta_k} &= \sqrt{9k_0+5}\ \sum_{k=1}^{k_0} \beta^{-k/2} \\
&= \sqrt{9k_0+5}\ \sum_{k=1}^{k_0} \sqrt{5}^k \\
&= \sqrt{9k_0+5}\ \frac{\sqrt{5}^{k_0}+1}{\sqrt{5}-1} \\
&\le \sqrt{9k_0+5}\left(\exp\left(\frac{1}{2}\log(5)\Big(\frac{2\log(\sqrt{2}d)}{\log(5)}+1\Big)\right)+1\right) \\
&\le \sqrt{9k_0+5}(d\sqrt{10}+1) \\
&\lesssim d\sqrt{\log(d)}\,, \qquad (49)
\end{aligned}$$

and

$$\begin{aligned}
\sum_{k=1}^{k_0} \frac{\alpha_k^{2/3}}{\beta_k} &\le (9k_0+5)^{2/3}(d\sqrt{10}+1) \\
&\lesssim d\big(\log(d)\big)^{2/3}\,. \qquad (50)
\end{aligned}$$

C.10 Definition of the sequences $\mathbb{A}^r_{t,k}$

The two sequences $(\alpha_k)_{k\in\mathbb{N}}$ and $(\beta_k)_{k\in\mathbb{N}}$, both begin at 1 and strictly decrease to 0 (see Appendix C.9). These sequences are introduced to be able to consider the 3 terms of Equation (18) separately, and are engineered so that they only introduce a quasi-constant factor in the final regret bound.

Let $t \ge d(d+1)$, then, Lemma 4.6 states that under $\{\mathcal{G}_t \bigcap \mathcal{C}\}$,

$$\frac{\Delta^2_{A_{t+1}}}{4f^2_{t,\delta}} \le \sum_{i\in A_{t+1}} \frac{\bar\sigma^2_{A_{t+1},i}}{n_{t,(i,i)}} + \big(d+3\log(T)h^2_{T,\delta}\big) \sum_{(i,j)\in A_{t+1}} \frac{B_iB_j}{n^2_{t,(i,j)}} + 3h_{T,\delta} \sum_{(i,j)\in A_{t+1}} \frac{B_iB_j}{n^{3/2}_{t,(i,j)}}\,.$$

Let $k \in \mathbb{N}^*$, we introduce the sets

$$S^1_{t,k} = \Big\{ i\in A_{t+1},\quad n_{t,(i,i)} \le d\alpha_k \log(T)\ \frac{4f^2_{t,\delta}}{\Delta^2_{A_{t+1}}}\ \frac{\bar\sigma^4_{A_{t+1},i}}{\mathbf{\Sigma}^*_{i,i}} \Big\}\,, \qquad (51)$$

$$S^2_{t,k} = \Big\{ (i,j)\in A_{t+1},\quad n^2_{t,(i,j)} \le d^2\alpha_k\ \frac{2\log(T)}{\log(T)-1}\ \frac{4f^2_{t,\delta}}{\Delta^2_{A_{t+1}}}\ (d+3\log(T)h^2_{T,\delta})B_iB_j \Big\}\,, \qquad (52)$$

$$S^3_{t,k} = \Big\{ (i,j)\in A_{t+1},\quad n^{3/2}_{t,(i,j)} \le d^2\alpha_k \frac{2\log(T)}{\log(T)-1} \frac{4f^2_{t,\delta}}{\Delta^2_{A_{t+1}}}\ 3h_{T,\delta}B_iB_j \Big\}\,. \qquad (53)$$

They are associated to the events

$$\begin{aligned}
\mathbb{A}^1_{t,k} &= \Big\{ \sum_{i\in S^1_{t,k}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar\sigma^2_{A_{t+1},i}} \ge \beta_k d;\quad \forall l<k, \sum_{i\in S^1_{t,l}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar\sigma^2_{A_{t+1},i}} < \beta_l d \Big\}\,, \\
\mathbb{A}^2_{t,k} &= \Big\{ |S^2_{t,k}| \ge \beta_k d^2;\quad \forall l<k, |S^2_{t,l}| < \beta_l d^2 \Big\}\,, \\
\mathbb{A}^3_{t,k} &= \Big\{ |S^3_{t,k}| \ge \beta_k d^2;\quad \forall l<k, |S^3_{t,l}| < \beta_l d^2 \Big\}\,.
\end{aligned}$$

C.11 Proof of Lemma 4.7

Lemma 4.7. *Let $t \ge d(d+1)$ and $k_0 \in \mathbb{N}^*$ such that $0 < d\beta_{k_0} < \big(\frac{1}{d} \wedge \min_{i,a}\{\mathbf{\Sigma}^*_{i,i}\bar\sigma^{-2}_{a,i}\}\big)$. Then,*

$$\mathbb{1}\{\mathcal{G}_t \bigcap \mathcal{C}\} \le \textstyle\sum_{r=1}^3 \sum_{k=1}^{k_0} \mathbb{1}\{\mathbb{A}^r_{t,k}\}\,.$$

Proof. Let $t \geq d(d+1)$, $(\alpha_k)_{k\in\mathbb{N}}$ and $(\beta_k)_{k\in\mathbb{N}}$ defined in Appendix C.9 and the events $\mathbb{A}^r_{t,k}$ defined in Appendix C.10.

Let $k \geq k_0$,

$$\mathbb{A}^1_{t,k} = \left\{ \sum_{i\in S^1_{t,k}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar{\sigma}^2_{A_{t+1},i}} \geq \beta_k d; \quad \forall l < k, \sum_{i\in S^1_{t,l}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar{\sigma}^2_{A_{t+1},i}} < \beta_l d \right\}.$$

As $\beta_k d < \min_{i,a} \mathbf{\Sigma}^*_{i,i}\bar{\sigma}^{-2}_{a,i}$ and $(S^1_{t,l})_l$ is a decreasing sequence of sets, $\sum_{i\in S^1_{t,k_0}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar{\sigma}^2_{A_{t+1},i}} < \beta_{k_0} d$ imply $S^1_{t,k_0} = \emptyset$ and $\sum_{i\in S^1_{t,k}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar{\sigma}^2_{A_{t+1},i}} = 0 < \beta_k d$. Therefore, $\mathbb{A}^1_{t,k}$ cannot happen and we denote

$$\mathbb{A}^1_t = \bigcup_{k\geq 1} \mathbb{A}^1_{t,k} = \bigcup_{k\in[k_0]} \mathbb{A}^1_{t,k} = \bigcup_{k\in[k_0]} \left\{ \sum_{i\in S^1_{t,k}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar{\sigma}^2_{A_{t+1},i}} \geq \beta_k d; \quad \forall l < k, \sum_{i\in S^1_{t,l}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar{\sigma}^2_{A_{t+1},i}} < \beta_l d \right\}.$$

Likewise, for $k > k_0$ and $r = 2$ or 3,

$$\mathbb{A}^r_{t,k} = \left\{ |S^r_{t,k}| \geq \beta_k d^2; \quad \forall l < k, |S^r_{t,k}| < \beta_l d^2 \right\}.$$

As $\beta_{k_0} d^2 < 1$ and $(S^r_{t,l})_l$ is a decreasing sequence of sets, then $|S^r_{t,k_0}| < \beta_{k_0} d^2$ imply $S^r_{t,k_0} = \emptyset$ and $|S^r_{t,k}| = 0 < \beta_k d^2$. Therefore, $\mathbb{A}^r_{t,k}$ cannot happen and we denote

$$\mathbb{A}^2_t = \bigcup_{k\geq 1} \mathbb{A}^2_{t,k} = \bigcup_{k\in[k_0]} \mathbb{A}^2_{t,k} = \bigcup_{k\in[k_0]} \left\{ |S^2_{t,k}| \geq \beta_k d^2; \quad \forall l < k, |S^2_{t,k}| < \beta_l d^2 \right\},$$
$$\mathbb{A}^3_t = \bigcup_{k\geq 1} \mathbb{A}^3_{t,k} = \bigcup_{k\in[k_0]} \mathbb{A}^3_{t,k} = \bigcup_{k\in[k_0]} \left\{ |S^3_{t,k}| \geq \beta_k d^2; \quad \forall l < k, |S^3_{t,k}| < \beta_l d^2 \right\}.$$

The idea is now to prove that

$$\left(\bigcup_{r=1}^3 \mathbb{A}^r_t \right)^c = \bigcap_{r=1}^3 \left(\mathbb{A}^r_t \right)^c \subseteq \left(\mathcal{G}_t \bigcap \mathcal{C} \right)^c.$$

We begin by considering $(\mathbb{A}^1_t)^c$,

$$\begin{aligned}
(\mathbb{A}^1_t)^c &= \bigcap_{k=1}^{k_0} (\mathbb{A}^1_{t,k})^c \\
&= \bigcap_{k=1}^{k_0} \left(\left\{ \sum_{i\in S^1_{t,k}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar{\sigma}^2_{A_{t+1},i}} < \beta_k d \right\} \bigcup_{l=1}^{k-1} \left\{ \sum_{i\in S^1_{t,l}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar{\sigma}^2_{A_{t+1},i}} \geq \beta_l d \right\} \right) \\
&= \bigcap_{k=1}^{k_0} \left\{ \sum_{i\in S^1_{t,k}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar{\sigma}^2_{A_{t+1},i}} < \beta_k d \right\}.
\end{aligned} \tag{54}$$

Then, under $(\mathbb{A}_t^1)^c$, denoting $S_{t,0}^1 = A_{t+1}$, as $S_{t,k_0}^1 = \emptyset$ and the sets $S_{t,k}^1$ are decreasing with respect to k,

$$\begin{aligned}
\sum_{i\in A_{t+1}} \frac{\bar\sigma^2_{A_{t+1},i}}{n_{t,(i,i)}} &= \sum_{k=1}^{k_0} \sum_{i\in S^1_{t,k-1}\setminus S^1_{t,k}} \frac{\bar\sigma^2_{A_{t+1},i}}{n_{t,(i,i)}} \\
&\le \sum_{k=1}^{k_0} \sum_{i\in S^1_{t,k-1}\setminus S^1_{t,k}} \bar\sigma^2_{A_{t+1},i} \frac{1}{d\alpha_k} \frac{1}{\log(T)} \frac{\Delta^2_{A_{t+1}}}{4f^2_{t,\delta}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar\sigma^4_{A_{t+1},i}} \leftarrow \text{ by Eq. (51)} \\
&= \frac{\Delta^2_{A_{t+1}}}{\log(T) 4 f^2_{t,\delta} d} \sum_{k=1}^{k_0} \frac{1}{\alpha_k} \sum_{i\in S_{t,k-1}\setminus S_{t,k}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar\sigma^2_{A_{t+1},i}} \\
&= \frac{\Delta^2_{A_{t+1}}}{4\log(T) f^2_{t,\delta} d} \sum_{k=1}^{k_0} \frac{1}{\alpha_k} \Bigg(\sum_{i\in S^1_{t,k-1}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar\sigma^2_{A_{t+1},i}} - \sum_{i\in S^1_{t,k}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar\sigma^2_{A_{t+1},i}} \Bigg) \\
&= \frac{\Delta^2_{A_{t+1}}}{4\log(T) f^2_{t,\delta} d} \sum_{k=0}^{k_0-1} \frac{1}{\alpha_{k+1}} \Bigg(\sum_{i\in S^1_{t,k}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar\sigma^2_{A_{t+1},i}} - \sum_{i\in S^1_{t,k+1}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar\sigma^2_{A_{t+1},i}} \Bigg) \\
&= \frac{\Delta^2_{A_{t+1}}}{4\log(T) f^2_{t,\delta} d} \Bigg(\frac{1}{\alpha_1} \sum_{i\in S^1_{t,0}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar\sigma^2_{A_{t+1},i}} + \sum_{k=1}^{k_0-1} \Bigg(\frac{1}{\alpha_{k+1}} - \frac{1}{\alpha_k} \Bigg) \sum_{i\in S^1_{t,k}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar\sigma^2_{A_{t+1},i}} \Bigg) \\
&< \frac{\Delta^2_{A_{t+1}}}{4\log(T) f^2_{t,\delta} d} \Bigg(\frac{d}{\alpha_1} + \sum_{k=1}^{k_0-1} d\beta_k \Bigg(\frac{1}{\alpha_{k+1}} - \frac{1}{\alpha_k} \Bigg) \Bigg) \leftarrow S_{t,0} = A_{t+1} \text{ and Eq. (54)} \\
&= \frac{\Delta^2_{A_{t+1}}}{4\log(T) f^2_{t,\delta}} \Bigg(\sum_{k=1}^{k_0-1} \frac{\beta_{k-1} - \beta_k}{\alpha_k} + \frac{\beta_{k_0-1}}{\alpha_{k_0}} \Bigg) \\
&\le \frac{1}{\log(T)} \frac{\Delta^2_{A_{t+1}}}{4f^2_{t,\delta}} . \leftarrow \text{ by Eq. (47)}
\end{aligned} \tag{55}$$

Likewise for $r = 1$ and 2,

$$(\mathbb{A}_t^r)^c = \bigcap_{k=1}^{k_0} \Big\{ |S^r_{t,k}| < \beta_k d^2 \Big\} . \tag{56}$$

For $r=2$, under $(\mathbb{A}_t^2)^c$, denoting $S_{t,0}^2 = A_{t+1} \times A_{t+1}$, as $S_{t,k_0}^2 = \emptyset$,

$$\begin{aligned}
\big(d+3\log(T)h_{T,\delta}^2\big)\sum_{(i,j)\in A_{t+1}} \frac{B_iB_j}{n_{t,(i,j)}^2} &= \big(d+3\log(T)h_{T,\delta}^2\big)\sum_{k=1}^{k_0}\sum_{(i,j)\in S_{t,k-1}\setminus S_{t,k}} \frac{B_iB_j}{n_{t,i}^2} \\
&\le \sum_{k=1}^{k_0}\sum_{(i,j)\in S_{t,k-1}\setminus S_{t,k}} B_iB_j \frac{1}{\alpha_k}\frac{1}{d^2}\frac{\log(T)-1}{2\log(T)}\frac{\Delta_{A_{t+1}}^2}{4f_{t,\delta}^2}\frac{1}{B_iB_j} \leftarrow \text{ by Eq. (52)} \\
&= \frac{\log(T)-1}{2\log(T)}\frac{\Delta_{A_{t+1}}^2}{4f_{t,\delta}^2}\frac{1}{d^2}\sum_{k=1}^{k_0}\frac{1}{\alpha_k}\Big(|S_{t,k-1}| - |S_{t,k}|\Big) \\
&= \frac{\log(T)-1}{2\log(T)}\frac{\Delta_{A_{t+1}}^2}{4f_{t,\delta}^2}\frac{1}{d^2}\sum_{k=0}^{k_0-1}\frac{1}{\alpha_{k+1}}\Big(|S_{t,k}| - |S_{t,k+1}|\Big) \\
&= \frac{\log(T)-1}{2\log(T)}\frac{\Delta_{A_{t+1}}^2}{4f_{t,\delta}^2}\frac{1}{d^2}\left(\frac{|S_{t,0}|}{\alpha_1} + \sum_{k=1}^{k_0-1}|S_{t,k}|\Big(\frac{1}{\alpha_{k+1}} - \frac{1}{\alpha_k}\Big)\right) \\
&< \frac{\log(T)-1}{2\log(T)}\frac{\Delta_{A_{t+1}}^2}{4f_{t,\delta}^2}\frac{1}{d^2} \\
&\qquad \left(\frac{1}{\alpha_1}d^2 + \sum_{k=1}^{k_0-1}\beta_k d^2\Big(\frac{1}{\alpha_{k+1}} - \frac{1}{\alpha_k}\Big)\right) \leftarrow \text{ by Eq. (56)} \\
&= \frac{\log(T)-1}{2\log(T)}\frac{\Delta_{A_{t+1}}^2}{4f_{t,\delta}^2}\left(\sum_{k=1}^{k_0-1}\frac{\beta_{k-1}-\beta_k}{\alpha_k} + \frac{\beta_{k_0-1}}{\alpha_{k_0}}\right) \\
&\le \frac{\log(T)-1}{2\log(T)}\frac{\Delta_{A_{t+1}}^2}{4f_{t,\delta}^2}\,. \leftarrow \text{ by Eq. (47)}
\end{aligned} \tag{57}$$

And for $r=3$,

$$3h_{T,\delta}\sum_{(i,j)\in A_{t+1}}\frac{B_iB_j}{n_{t,(i,j)}^{3/2}} < \frac{\log(T)-1}{2\log(T)}\frac{\Delta_{A_{t+1}}^2}{4f_{t,\delta}^2}\,. \tag{58}$$

Therefore, under $\bigcap_{r=1}^3\Big(\mathbb{A}_t^r\Big)^c$, summing Equation (55), Equation (57) and Equation (58) yields

$$\begin{aligned}
\sum_{i\in A_{t+1}}\frac{\bar\sigma_{A_{t+1},i}^2}{n_{t,i}} + \big(d+3\log(T)h_{T,\delta}^2\big)\sum_{(i,j)\in A_{t+1}}\frac{B_iB_j}{n_{t,(i,j)}^2} + 3h_{T,\delta}\sum_{(i,j)\in A_{t+1}}\frac{B_iB_j}{n_{t,(i,j)}^{3/2}} &< \frac{\Delta_{A_{t+1}}^2}{4f_{t,\delta}^2}\Big(\frac{1}{\log(T)} + 2\frac{\log(T)-1}{2\log(T)}\Big) \\
&= \frac{\Delta_{A_t}^2}{4f_{t,\delta}^2}\,,
\end{aligned}$$

which contradict Equation (18) and thus imply $\{\mathcal{G}_t \bigcap \mathcal{C}\}^c$. By contraposition, we have proved that $\{\mathcal{G}_t \bigcap \mathcal{C}\}$ imply $\bigcup_{r=1}^3\big(\mathbb{A}_t^r\big)$. Therefore,

$$\mathbb{1}\{\mathcal{G}_t \bigcap \mathcal{C}\} \le \sum_{r=1}^3 \mathbb{1}\{\mathbb{A}_t^r\} \le \sum_{r=1}^3\sum_{k=1}^{k_0}\mathbb{1}\{\mathbb{A}_{t,k}^r\}\,.$$

□

C.12 Proof of Proposition 4.4

Proposition 4.4. *Let $T \ge d(d+1)$ and $\delta = 1/T^2$. Then,*

$$\mathbb{E}\Big[\sum_{t=d(d+1)}^{T-1}\Delta_{A_{t+1}}\mathbb{1}\big\{\mathcal{G}_t\bigcap\mathcal{C}\big\}\Big] = O\bigg(\log(T)^3(\log d)^2\Big(\sum_{i=1}^d\max_{a\in\mathcal{A}/i\in a}\frac{\sigma_{a,i}^2}{\Delta_a}\Big)\bigg)\,,$$

as $T \to \infty$, *where* $\sigma^2_{a,i} = \sum_{j\in a}(\mathbf{\Sigma}^*_{i,j})_+$.

Proof. Let $T \geq d(d+A)$, from Proposition 4.7 we have

$$\sum_{t=d(d+1)}^{T-1} \Delta_{A_{t+1}} \mathbb{1}\{\mathcal{G}_t \bigcap \mathcal{C}\} \leq \sum_{t=d(d+1)}^{T-1} \sum_{r=1}^{3} \sum_{k=1}^{k_0} \Delta_{A_{t+1}} \mathbb{1}\{\mathbb{A}^r_{t,k}\}$$
$$= \sum_{r=1}^{3} \sum_{t=d(d+1)}^{T-1} \Delta_{A_{t+1}} \sum_{k=1}^{k_0} \mathbb{1}\{\mathbb{A}^r_{t,k}\}\,. \tag{59}$$

We begin with $r=1$, $t \geq d(d+1)$, and $k \in [k_0]$,

$$\mathbb{A}^1_{t,k} = \left\{ \sum_{i\in S^1_{t,k}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar{\sigma}^2_{A_{t+1},i}} \geq \beta_k d; \quad \forall l < k, \sum_{i\in S^1_{t,l}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar{\sigma}^2_{A_{t+1},i}} < \beta_l d \right\} \subseteq \left\{ \frac{1}{\beta_k d} \sum_{i\in S^1_{t,k}} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar{\sigma}^2_{A_{t+1},i}} \geq 1 \right\}.$$

Therefore,

$$\mathbb{1}\{\mathbb{A}^1_{t,k}\} \leq \frac{1}{\beta_k d} \sum_{i\in[d]} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar{\sigma}^2_{A_{t+1},i}} \mathbb{1}\left\{\mathbb{A}^1_{t,k} \bigcap \{i \in S^1_{t,k}\}\right\}. \tag{60}$$

Summing over t and integrating the gaps yields

$$\sum_{t=d(d+1)}^{T-1} \Delta_{A_{t+1}} \sum_{k=1}^{k_0} \mathbb{1}\{\mathbb{A}^1_{t,k}\} \leq \sum_{t=d(d+1)}^{T} \Delta_{A_{t+1}} \sum_{k=1}^{k_0} \frac{1}{\beta_k d} \sum_{i\in[d]} \frac{\mathbf{\Sigma}^*_{i,i}}{\bar{\sigma}^2_{A_{t+1},i}} \mathbb{1}\left\{\mathbb{A}^1_{t,k} \bigcap \{i \in S^1_{t,k}\}\right\} \leftarrow \text{ by Eq. (60)}$$
$$\leq \sum_{i\in[d]} \mathbf{\Sigma}^*_{i,i} \sum_{t=d(d+1)}^{T} \sum_{k=1}^{k_0} \frac{1}{\beta_k d} \frac{\Delta_{A_{t+1}}}{\bar{\sigma}^2_{A_{t+1},i}} \mathbb{1}\{i \in S^1_{t,k}\}$$
$$= \sum_{i\in[d]} \mathbf{\Sigma}^*_{i,i} \sum_{k=1}^{k_0} \frac{1}{\beta_k d} \sum_{t=d(d+1)}^{T} \frac{\Delta_{A_{t+1}}}{\bar{\sigma}^2_{A_{t+1},i}} \mathbb{1}\left\{ n_{t,(i,i)} \leq d\alpha_k \log(T) \frac{4f^2_{t,\delta}}{\left(\frac{\Delta_{A_{t+1}}}{\bar{\sigma}^2_{A_{t+1},i}}\right)^2 \mathbf{\Sigma}^*_{i,i}} \right\}. \leftarrow \text{ by Eq. (51)} \tag{61}$$

Let $i \in [d]$, we consider all the actions associated to it. Let $q_i \in \mathbb{N}$ be the number of actions associated to item i. Let $l \in [q_i]$, we denote $e^l_i \in \mathcal{A}$ the l-th action associated to item i, sorted by decreasing $\frac{\Delta_{e^l_i}}{\bar{\sigma}^2_{e^l_i,i}}$, with $\frac{\bar{\sigma}^2_{e^0_i,i}}{\Delta_{e^0_i}} = 0$ by convention. Then

$$\sum_{t=d(d+1)}^{T-1} \frac{\Delta_{A_{t+1}}}{\bar{\sigma}^2_{A_{t+1},i}} \mathbb{1}\left\{ n_{t,(i,i)} \leq d\alpha_k \log(T) \frac{4f_{t,\delta}^2}{\left(\frac{\Delta_{A_{t+1}}}{\bar{\sigma}^2_{A_{t+1},i}}\right)^2 \mathbf{\Sigma}^*_{i,i}} \right\}$$

$$\leq \sum_{t=0}^{T-1} \frac{\Delta_{A_{t+1}}}{\bar{\sigma}^2_{A_{t+1},i}} \mathbb{1}\left\{ n_{t,(i,i)} \leq d\alpha_k \log(T) \frac{4f_{T,\delta}^2}{\left(\frac{\Delta_{A_{t+1}}}{\bar{\sigma}^2_{A_{t+1},i}}\right)^2 \mathbf{\Sigma}^*_{i,i}} \right\} \leftarrow \text{ as } f_{t,\delta} \leq f_{T,\delta} \text{ and extending the sum over all the rounds}$$

$$= \sum_{t=0}^{T-1} \sum_{l=1}^{q_i} \frac{\Delta_{e_i^l}}{\bar{\sigma}^2_{e_i^l,i}} \mathbb{1}\left\{ n_{t,(i,i)} \leq d\alpha_k \log(T) \frac{4f_{T,\delta}^2}{\left(\frac{\Delta_{e_i^l}}{\bar{\sigma}^2_{e_i^l,i}}\right)^2 \mathbf{\Sigma}^*_{i,i}}, \quad A_{t+1} = e_i^l \right\}$$

$$= \sum_{t=0}^{T-1} \sum_{l=1}^{q_i} \frac{\Delta_{e_i^l}}{\bar{\sigma}^2_{e_i^l,i}} \mathbb{1}\left\{ n_{t,(i,i)} \frac{\mathbf{\Sigma}^*_{i,i}}{d\alpha_k \log(T) 4f_{T,\delta}^2} \leq \frac{1}{\left(\frac{\Delta_{e_i^l}}{\bar{\sigma}^2_{e_i^l,i}}\right)^2}, \quad A_{t+1} = e_i^l \right\}$$

$$= \sum_{t=0}^{T-1} \sum_{l=1}^{q_i} \frac{\Delta_{e_i^l}}{\bar{\sigma}^2_{e_i^l,i}} \sum_{p=1}^{l} \mathbb{1}\left\{ \frac{1}{\left(\frac{\Delta_{e_i^{p-1}}}{\bar{\sigma}^2_{e_i^{p-1},i}}\right)^2} < n_{t,(i,i)} \frac{\mathbf{\Sigma}^*_{i,i}}{d\alpha_k \log(T) 4f_{T,\delta}^2} \leq \frac{1}{\left(\frac{\Delta_{e_i^p}}{\bar{\sigma}^2_{e_i^p,i}}\right)^2}, \quad A_{t+1} = e_i^l \right\} \leftarrow \text{decomposing the event}$$

$$\leq \sum_{t=0}^{T-1} \sum_{l=1}^{q_i} \sum_{p=1}^{l} \frac{\Delta_{e_i^p}}{\bar{\sigma}^2_{e_i^p,i}} \mathbb{1}\left\{ \frac{1}{\left(\frac{\Delta_{e_i^{p-1}}}{\bar{\sigma}^2_{e_i^{p-1},i}}\right)^2} < n_{t,(i,i)} \frac{\mathbf{\Sigma}^*_{i,i}}{d\alpha_k \log(T) 4f_{T,\delta}^2} \leq \frac{1}{\left(\frac{\Delta_{e_i^p}}{\bar{\sigma}^2_{e_i^p,i}}\right)^2}, \quad A_{t+1} = e_i^l \right\} \leftarrow \text{ as } \frac{\Delta_{e_i^l}}{\bar{\sigma}^2_{e_i^l,i}} \leq \frac{\Delta_{e_i^p}}{\bar{\sigma}^2_{e_i^p,i}}$$

$$= \sum_{p=1}^{q_i} \frac{\Delta_{e_i^p}}{\bar{\sigma}^2_{e_i^p,i}} \sum_{t=0}^{T-1} \sum_{l=p}^{q_i} \mathbb{1}\left\{ \frac{1}{\left(\frac{\Delta_{e_i^{p-1}}}{\bar{\sigma}^2_{e_i^{p-1},i}}\right)^2} < n_{t,(i,i)} \frac{\mathbf{\Sigma}^*_{i,i}}{d\alpha_k \log(T) 4f_{T,\delta}^2} \leq \frac{1}{\left(\frac{\Delta_{e_i^p}}{\bar{\sigma}^2_{e_i^p,i}}\right)^2}, \quad A_{t+1} = e_i^l \right\}$$

$$\leq \sum_{p=1}^{q_i} \frac{\Delta_{e_i^p}}{\bar{\sigma}^2_{e_i^p,i}} \sum_{t=0}^{T-1} \sum_{l=1}^{q_i} \mathbb{1}\left\{ \frac{1}{\left(\frac{\Delta_{e_i^{p-1}}}{\bar{\sigma}^2_{e_i^{p-1},i}}\right)^2} < n_{t,(i,i)} \frac{\mathbf{\Sigma}^*_{i,i}}{d\alpha_k \log(T) 4f_{T,\delta}^2} \leq \frac{1}{\left(\frac{\Delta_{e_i^p}}{\bar{\sigma}^2_{e_i^p,i}}\right)^2}, \quad A_{t+1} = e_i^l \right\} \leftarrow \text{we extend the sum over } l$$

$$= \sum_{p=1}^{q_i} \frac{\Delta_{e_i^p}}{\bar{\sigma}^2_{e_i^p,i}} \sum_{t=0}^{T-1} \mathbb{1}\left\{ \frac{1}{\left(\frac{\Delta_{e_i^{p-1}}}{\bar{\sigma}^2_{e_i^{p-1},i}}\right)^2} < n_{t,(i,i)} \frac{\mathbf{\Sigma}^*_{i,i}}{d\alpha_k \log(T) 4f_{T,\delta}^2} \leq \frac{1}{\left(\frac{\Delta_{e_i^p}}{\bar{\sigma}^2_{e_i^p,i}}\right)^2}, \quad i \in A_{t+1} \right\} \leftarrow \text{we simplify the inner sum}$$

$$\leq \sum_{p=1}^{q_i} \frac{\Delta_{e_i^p}}{\bar{\sigma}^2_{e_i^p,i}} \left(\left\lfloor \left(\frac{\bar{\sigma}^2_{e_i^p,i}}{\Delta_{e_i^p}} \right)^2 \frac{d\alpha_k \log(T) 4f_{T,\delta}^2}{\mathbf{\Sigma}^*_{i,i}} \right\rfloor - \left\lfloor \left(\frac{\bar{\sigma}^2_{e_i^{p-1},i}}{\Delta_{e_i^{p-1}}} \right)^2 \frac{d\alpha_k \log(T) 4f_{T,\delta}^2}{\mathbf{\Sigma}^*_{i,i}} \right\rfloor \right) \leftarrow \text{the event can only happen a given nbr. of ti}$$

$$= \left(\left\lfloor \left(\frac{\bar{\sigma}_{e_i^{q_i},i}}{\Delta_{e_i^{q_i}}} \right)^2 \frac{d\alpha_k \log(T) 4f_{T,\delta}^2}{\mathbf{\Sigma}^*_{i,i}} \right\rfloor \frac{\Delta_{e_i^{q_i}}}{\bar{\sigma}_{e_i^{q_i},i}} + \sum_{p=1}^{q_i-1} \left\lfloor \left(\frac{\bar{\sigma}^2_{e_i^p,i}}{\Delta_{e_i^p}} \right)^2 \frac{d\alpha_k \log(T) 4f_{T,\delta}^2}{\mathbf{\Sigma}^*_{i,i}} \right\rfloor \left(\frac{\Delta_{e_i^p}}{\bar{\sigma}^2_{e_i^p,i}} - \frac{\Delta_{e_i^{p+1}}}{\bar{\sigma}^2_{e_i^{p+1},i}} \right) \right) \leftarrow \text{ summation by parts}$$

$$\leq \frac{d\alpha_k \log(T) 4f_{T,\delta}^2}{\mathbf{\Sigma}^*_{i,i}} \left(\frac{\bar{\sigma}^2_{e_i^{q_i},i}}{\Delta_{e_i^{q_i}}} + \sum_{p=1}^{q_i-1} \left(\frac{\bar{\sigma}^2_{e_i^p,i}}{\Delta_{e_i^p}} \right)^2 \left(\frac{\Delta_{e_i^p}}{\bar{\sigma}^2_{e_i^p,i}} - \frac{\Delta_{e_i^{p+1}}}{\bar{\sigma}^2_{e_i^{p+1},i}} \right) \right) \leftarrow \text{everything is positive}$$

$$\leq \frac{4f_{T,\delta}^2 d\alpha_k \log(T)}{\mathbf{\Sigma}^*_{i,i}} \left(\frac{\bar{\sigma}^2_{e_i^{q_i},i}}{\Delta_{e_i^{q_i}}} + \int_{\left(\frac{\Delta_{e_i^{q_i}}}{\bar{\sigma}^2_{e_i^{q_i},i}}\right)}^{\left(\frac{\Delta_{e_i^1}}{\bar{\sigma}^2_{e_i^1,i}}\right)} \frac{1}{x^2} dx \right)$$

$$= \frac{4f_{T,\delta}^2 d\alpha_k \log(T)}{\mathbf{\Sigma}^*_{i,i}} \left(\frac{\bar{\sigma}^2_{e_i^{q_i},i}}{\Delta_{e_i^{q_i}}} + \frac{\bar{\sigma}^2_{e_i^{q_i}}}{\Delta_{e_i^{q_i}}} - \frac{\bar{\sigma}^2_{e_i^1}}{\Delta_{e_i^1}} \right)$$

$$\leq \frac{8f_{T,\delta}^2 d\alpha_k \log(T)}{\mathbf{\Sigma}^*_{i,i}} \frac{\bar{\sigma}^2_{e_i^{q_i},i}}{\Delta_{e_i^{q_i}}}$$

$$\leq \frac{8f_{T,\delta}^2 d\alpha_k \log(T)}{\mathbf{\Sigma}^*_{i,i}} \left(\max_{a \in \mathcal{A}/i \in a} \frac{\bar{\sigma}^2_{a,i}}{\Delta_a} \right)$$

$$\leq \frac{16f_{T,\delta}^2 d\alpha_k \log(T)}{\mathbf{\Sigma}^*_{i,i}} \left(\max_{a \in \mathcal{A}/i \in a} \frac{\sigma^2_{a,i}}{\Delta_a} \right). \leftarrow \bar{\sigma}^2_{a,i} \leq 2\sigma^2_{a,i} \tag{62}$$

Reinjecting Eq. (62) into Equation (61) yields

$$\begin{aligned}\sum_{t=d(d+1)}^{T-1}\Delta_{A_{t+1}}\sum_{k=1}^{k_0}\mathbb{1}\{\mathbb{A}_{t,k}^1\} &\leq \sum_{i\in[d]}\mathbf{\Sigma}_{i,i}^*\sum_{k=1}^{k_0}\frac{1}{\beta_k d}\frac{16 f_{T,\delta}^2 d\alpha_k\log(T)}{\mathbf{\Sigma}_{i,i}^*}\Big(\max_{a\in\mathcal{A}/i\in a}\frac{\sigma_{a,i}^2}{\Delta_a}\Big)\\ &= 16\log(T)f_{T,\delta}^2\Big(\sum_{k=1}^{k_0}\frac{\alpha_k}{\beta_k}\Big)\sum_{i\in[d]}\Big(\max_{a\in\mathcal{A}/i\in a}\frac{\sigma_{a,i}^2}{\Delta_a'}\Big)\\ &\lesssim \big(\log(T)\big)^3\log(d)^2\sum_{i\in[d]}\Big(\max_{a\in\mathcal{A}/i\in a}\frac{\sigma_{a,i}^2}{\Delta_a'}\Big), \leftarrow \text{Eq. (6) and Eq. (48)} \end{aligned} \tag{63}$$

for $d \lesssim \frac{\log(T)}{\log(\log(1+T))}$ and $\delta = 1/T^2$.

We treat the 2 other terms in a similar way. For $r=2$, let $t \geq d(d+1)$, and $k \in [k_0]$,

$$\mathbb{A}_{t,k}^2 = \Big\{|S_{t,k}^2| \geq \beta_k d^2; \quad \forall l<k, |S_{t,l}^2| < \beta_l d^2\Big\} \subseteq \Big\{\frac{1}{\beta_k d^2}|S_{t,k}^2| \geq 1\Big\}.$$

Therefore,

$$\mathbb{1}\{\mathbb{A}_{t,k}^2\} \leq \frac{1}{\beta_k d^2}\sum_{(i,j)\in[d]^2}\mathbb{1}\Big\{\mathbb{A}_{t,k}^2\bigcap\{(i,j)\in S_{t,k}^2\}\Big\}. \tag{64}$$

Summing over t and integrating the gaps yields

$$\begin{aligned}\sum_{t=d(d+1)}^{T-1}\Delta_{A_{t+1}}\sum_{k=1}^{k_0}\mathbb{1}\{\mathbb{A}_{t,k}^2\} &\leq \sum_{t=d(d+1)}^{T-1}\Delta_{A_{t+1}}\sum_{k=1}^{k_0}\frac{1}{\beta_k d^2}\sum_{(i,j)\in[d]^2}\mathbb{1}\Big\{\mathbb{A}_{t,k}^2\bigcap\{(i,j)\in S_{t,k}^2\}\Big\} \leftarrow \text{ by Eq. (64)}\\ &\leq \sum_{(i,j)\in[d]^2}\sum_{t=d(d+1)}^{T-1}\sum_{k=1}^{k_0}\frac{1}{\beta_k d^2}\Delta_{A_{t+1}}\mathbb{1}\{(i,j)\in S_{t,k}^2\}\\ &= \sum_{(i,j)\in[d]^2}\sum_{k=1}^{k_0}\frac{1}{\beta_k d^2}\\ &\qquad\sum_{t=d(d+1)}^{T-1}\Delta_{A_{t+1}}\mathbb{1}\Big\{n_{t,(i,j)} \leq d\sqrt{\alpha_k}\sqrt{\frac{2\log(T)}{\log(T)-1}}\frac{2f_{t,\delta}\sqrt{(d+3\log(T)h_{T,\delta}^2)B_iB_j}}{\Delta_{A_{t+1}}}\Big\}. \leftarrow \text{ by Eq. (52)}\end{aligned} \tag{65}$$

Let $(i,j)\in[d]^2$, we consider all the actions which are associated to it. Let $q_{(i,j)}\in\mathbb{N}$ be the number of actions associated to the tuple (i,j). Let $l\in[q_{(i,j)}]$, this time, we denote $e_{(i,j)}^l \in \mathcal{A}$ the l-th action associated to tuple (i,j), sorted by decreasing $\Delta_{e_{(i,j)}^l}$, with $\frac{1}{\Delta_{e_{(i,j)}^0}} = 0$ by convention. Then,

$$\sum_{t=d(d+1)}^{T-1} \Delta_{A_{t+1}} \mathbb{1}\left\{ n_{t,(i,j)} \leq d\sqrt{\alpha_k} \sqrt{\frac{2\log(T)}{\log(T)-1}} \frac{2f_{t,\delta}\sqrt{(d+3\log(T)h_{T,\delta}^2)B_iB_j}}{\Delta_{A_{t+1}}} \right\}$$

$$\leq \sum_{t=0}^{T-1} \Delta_{A_{t+1}} \mathbb{1}\left\{ n_{t,(i,j)} \leq d\sqrt{\alpha_k} \sqrt{\frac{2\log(T)}{\log(T)-1}} \frac{2f_{T,\delta}\sqrt{(d+3\log(T)h_{T,\delta}^2)B_iB_j}}{\Delta_{A_{t+1}}} \right\} \leftarrow \text{ as } f_{t,\delta} \leq f_{T,\delta}$$

$$= \sum_{t=0}^{T-1} \sum_{l=1}^{q_{(i,j)}} \Delta_{e^l_{(i,j)}} \mathbb{1}\left\{ n_{t,(i,j)} \leq d\sqrt{\alpha_k} \sqrt{\frac{2\log(T)}{\log(T)-1}} \frac{2f_{T,\delta}\sqrt{(d+3\log(T)h_{T,\delta}^2)B_iB_j}}{\Delta_{e^l_i}}, \quad A_{t+1} = e^l_{(i,j)} \right\}$$

$$= \sum_{t=0}^{T-1} \sum_{l=1}^{q_{(i,j)}} \Delta_{e^l_{(i,j)}} \mathbb{1}\left\{ n_{t,(i,j)} \frac{1}{d}\sqrt{\frac{1}{\alpha_k}} \sqrt{\frac{\log(T)-1}{2\log(T)}} \frac{1}{2f_{T,\delta}\sqrt{(d+3\log(T)h_{T,\delta}^2)B_iB_j}} \leq \frac{1}{\Delta_{e^l_{(i,j)}}}, \quad A_{t+1} = e^l_{(i,j)} \right\}$$

$$\leq 2f_{T,\delta}\sqrt{(d+3\log(T)h_{T,\delta}^2)B_iB_j}\sqrt{\frac{2\log(T)}{\log(T)-1}}\sqrt{\alpha_k}d\left(1 + \sum_{p=1}^{q_{(i,j)}-1} \frac{1}{\Delta_{e^p_{(i,j)}}}\left(\Delta_{e^p_{(i,j)}} - \Delta_{e^{p+1}_{(i,j)}}\right)\right) \leftarrow \text{same steps as for (62)}$$

$$\leq 2\sqrt{2}df_{T,\delta}\sqrt{(d+3\log(T)h_{T,\delta}^2)B_iB_j}\sqrt{\frac{\log(T)}{\log(T)-1}}\sqrt{\alpha_k}\left(1 + \int_{\Delta_{e^{q_{(i,j)}}_{(i,j)}}}^{\Delta_{e^1_{(i,j)}}} \frac{1}{x}dx\right)$$

$$= 2\sqrt{2}df_{T,\delta}\sqrt{(d+3\log(T)h_{T,\delta}^2)B_iB_j}\sqrt{\frac{\log(T)}{\log(T)-1}}\sqrt{\alpha_k}\left(1 + \log\left(\frac{\Delta_{e^1_{(i,j)}}}{\Delta_{e^{q_{(i,j)}}_{(i,j)}}}\right)\right)$$

$$\leq 2\sqrt{2}df_{T,\delta}\sqrt{(d+3\log(T)h_{T,\delta}^2)B_iB_j}\sqrt{\frac{\log(T)}{\log(T)-1}}\sqrt{\alpha_k}\left(1 + \log\left(\frac{\Delta_{\max}}{\Delta_{\min}}\right)\right). \tag{66}$$

Reinjecting Eq. (66) into (65), we get

$$\sum_{t=d(d+1)}^{T-1} \Delta_{A_{t+1}} \sum_{k=1}^{k_0} \mathbb{1}\{\mathbb{A}_{t,k}^2\} \leq \frac{2\sqrt{2}}{d} f_{T,\delta}\sqrt{(d+3\log(T)h_{T,\delta}^2)}\sqrt{\frac{\log(T)}{\log(T)-1}}\left(1+\log\left(\frac{\Delta_{\max}}{\Delta_{\min}}\right)\right) \sum_{(i,j)\in[d]^2} \sqrt{B_iB_j} \sum_{k=1}^{k_0} \frac{\sqrt{\alpha_k}}{\beta_k}$$

$$\lesssim \big(\log(T)\big)^{5/2} d^2 B_{max}\left(1+\log\left(\frac{\Delta_{\max}}{\Delta_{\min}}\right)\right), \tag{67}$$

for $d \lesssim \frac{\log(T)}{\log(\log(1+T))}$, $\delta = 1/T^2$ and $d \lesssim \log(T)^3$.

Likewise for $r = 3$, let $t \geq d(d+1)$, and $k \in [k_0]$,

$$\mathbb{A}_{t,k}^3 = \left\{ |S_{t,k}^3| \geq \beta_k d^2; \quad \forall l < k, |S_{t,l}^3| < \beta_l d^2 \right\} \subseteq \left\{ \frac{1}{\beta_k d^2} |S_{t,k}^3| \geq 1 \right\},$$

And

$$\mathbb{1}\{\mathbb{A}_{t,k}^3\} \leq \frac{1}{\beta_k d^2} \sum_{(i,j)\in[d]^2} \mathbb{1}\Big\{ \mathbb{A}_{t,k}^3 \bigcap \{(i,j) \in S_{t,k}^3\} \Big\}. \tag{68}$$

Summing over t and integrating the gaps yields

$$\begin{aligned}
\sum_{t=d(d+1)}^{T-1}\Delta_{A_{t+1}}\sum_{k=1}^{k_0}\mathbb{1}\{\mathbb{A}^3_{t,k}\} &\leq \sum_{t=d(d+1)}^{T-1}\Delta_{A_{t+1}}\sum_{k=1}^{k_0}\frac{1}{\beta_k d^2}\sum_{(i,j)\in[d]^2}\mathbb{1}\Big\{\mathbb{A}^3_{t,k}\bigcap\{(i,j)\in S^3_{t,k}\}\Big\} \leftarrow \text{ by Eq. (68)}\\
&\leq \sum_{(i,j)\in[d]^2}\sum_{t=d(d+1)}^{T-1}\sum_{k=1}^{k_0}\frac{1}{\beta_k d^2}\Delta_{A_{t+1}}\mathbb{1}\{(i,j)\in S^3_{t,k}\}\\
&= \sum_{(i,j)\in[d]^2}\sum_{k=1}^{k_0}\frac{1}{\beta_k d^2}\\
&\qquad\sum_{t=d(d+1)}^{T-1}\Delta_{A_{t+1}}\mathbb{1}\left\{n_{t,(i,j)}\leq d^{4/3}\left(\alpha_k\frac{2\log(T)}{\log(T)-1}3h_{T,\delta}B_iB_j\right)^{2/3}\frac{4^{2/3}f_{t,\delta}^{4/3}}{\Delta_{A_{t+1}}^{4/3}}\right\}. \leftarrow \text{ by Eq. (53)}
\end{aligned} \tag{69}$$

Let $(i,j)\in[d]^2$, we consider all the actions which are associated to it. Let $q_{(i,j)}\in\mathbb{N}$ be the number of actions associated to the tuple (i,j). Let $l\in[q_{(i,j)}]$, this time, we denote $e^l_{(i,j)}\in\mathcal{A}$ the l-th action associated to tuple (i,j), sorted by decreasing $\Delta_{e^l_{(i,j)}}$, with $\frac{1}{\Delta_{e^0_{(i,j)}}}=0$ by convention. Then, like for (62),

$$\begin{aligned}
&\sum_{t=d(d+1)}^{T-1}\Delta_{A_{t+1}}\mathbb{1}\left\{n_{t,(i,j)}\leq d^{4/3}\left(\alpha_k\frac{2\log(T)}{\log(T)-1}3h_{T,\delta}B_iB_j\right)^{2/3}\frac{4^{2/3}f_{t,\delta}^{4/3}}{\Delta_{A_{t+1}}^{4/3}}\right\}\\
&\leq \sum_{t=0}^{T-1}\sum_{l=1}^{q_{(i,j)}}\Delta_{e^l_{(i,j)}}\mathbb{1}\left\{n_{t,(i,j)}d^{-4/3}\left(\alpha_k\frac{2\log(T)}{\log(T)-1}3h_{T,\delta}B_iB_j\right)^{-2/3}4^{-2/3}f_{T,\delta}^{-4/3}\leq\frac{1}{\Delta^{4/3}_{e^l_{(i,j)}}},\quad A_{t+1}=e^l_{(i,j)}\right\}\\
&\leq d^{4/3}\left(\alpha_k\frac{2\log(T)}{\log(T)-1}3h_{T,\delta}B_iB_j\right)^{2/3}4^{2/3}f_{t,\delta}^{4/3}\left(\frac{1}{\Delta^{1/3}_{e^{q_{(i,j)}}_{(i,j)}}}+\sum_{p=1}^{q_{(i,j)}-1}\frac{1}{\Delta^{4/3}_{e^p_{(i,j)}}}\left(\Delta_{e^p_{(i,j)}}-\Delta_{e^{p+1}_{(i,j)}}\right)\right)\\
&\leq 4^{2/3}f_{t,\delta}^{4/3}d^{4/3}\left(\alpha_k\frac{2\log(T)}{\log(T)-1}3h_{T,\delta}B_iB_j\right)^{2/3}\left(\frac{1}{\Delta^{1/3}_{e^{q_{(i,j)}}_{(i,j)}}}+\int_{\Delta_{e^{q_{(i,j)}}_{(i,j)}}}^{\Delta_{e^1_{(i,j)}}}\frac{1}{x^{4/3}}dx\right)\\
&\leq 4^{2/3}f_{t,\delta}^{4/3}d^{4/3}\left(\alpha_k\frac{2\log(T)}{\log(T)-1}3h_{T,\delta}B_iB_j\right)^{2/3}\frac{4}{3}\frac{1}{\Delta^{1/3}_{e^{q_{(i,j)}}_{(i,j)}}}\\
&= \frac{4^{5/3}}{3^{1/3}}f_{t,\delta}^{4/3}h_{T,\delta}^{2/3}d^{4/3}\left(\frac{2\log(T)}{\log(T)-1}\right)^{2/3}(B_iB_j)^{2/3}\alpha_k^{1/3}\frac{1}{\Delta_{\min}^{1/3}}\,.
\end{aligned} \tag{70}$$

Reinjecting Eq. (70) into (69), we get

$$\begin{aligned}
\sum_{t=d(d+1)}^{T-1}\Delta_{A_{t+1}}\sum_{k=1}^{k_0}\mathbb{1}\{\mathbb{A}^3_{t,k}\} &\leq \frac{4^{5/3}}{3^{1/3}}f_{t,\delta}^{4/3}h_{T,\delta}^{2/3}d^{-2/3}\left(\frac{2\log(T)}{\log(T)-1}\right)^{2/3}\frac{1}{\Delta_{\min}^{1/3}}\sum_{(i,j)\in[d]^2}(B_iB_j)^{2/3}\sum_{k=1}^{k_0}\frac{\alpha_k^{1/3}}{\beta_k}\\
&\lesssim \big(\log(T)\big)^2 d^{7/3}\big(\log(d)\big)^{2/3}B_{max}^{4/3}\frac{1}{\Delta_{\min}^{1/3}}\,,
\end{aligned} \tag{71}$$

for $d\lesssim\frac{\log(T)}{\log(\log(1+T))}$ and $\delta=1/T^2$.

Summing (63), (67) and (71), the dominant term with respect to T is (63) which yields

$$\mathbb{E}\Big[\sum_{t=d(d+1)}^{T-1}\Delta_{A_{t+1}}\mathbb{1}\{\mathcal{G}_t\bigcap\mathcal{C}\}\Big]=O\left(\sum_{i\in[d]}\Big(\max_{a\in\mathcal{A}/i\in a}\frac{\sigma^2_{a,i}}{\Delta_a}\Big)\log(d)^2\log(T)^3\right). \tag{72}$$

□

D Proof of Theorem 3.2

Theorem 3.2. *Let $T \geq 5$, $B \in \mathbb{R}_+^d$, and $\delta = 1/T^2$. Let*

$$\sigma_{a,i}^2 = \sum_{j \in a} (\mathbf{\Sigma}_{i,j}^*)_+ \,, \tag{8}$$

where $i \in [d]$, $a \in \mathcal{A}$ and $(\,\cdot\,)_+ = \max\{\cdot, 0\}$. Then, `OLS-UCBV` (Alg. 2) satisfies the gap-dependent *regret upper bound*

$$\mathbb{E}[R_T] = \tilde{O}\bigg(\log(d)^2 \sum_{i=1}^{d} \max_{a \in \mathcal{A}/i \in a, \Delta_a > 0} \frac{\sigma_{a,i}^2}{\Delta_a} \bigg) \,,$$

and the distribution-free *regret upper bound*

$$\mathbb{E}[R_T] = \tilde{O}\bigg(\log(d) \sqrt{T \textstyle\sum_{i=1}^{d} \max_{a \in \mathcal{A}/i \in a} \sigma_{a,i}^2} \bigg) \,.$$

Proof. Let $T \geq 5$ and $\delta = 1/T^2$.

Injecting the result of Proposition 4.1 into (12) readily yields

$$\mathbb{E}[R_T] \leq \Big(d(d+1) + 1 + 1/T \Big) \Delta_{\max} + \mathbb{E}\bigg[\sum_{t=d(d+1)}^{T-1} \Delta_{A_{t+1}} \mathbb{1}\{\mathcal{G}_t \bigcap \mathcal{C}\} \bigg] \,. \tag{73}$$

Reinjecting the result of Proposition 4.4 into it, we get the gap-dependent bound

$$\mathbb{E}[R_T] = O\bigg(\sum_{i \in [d]} \Big(\max_{a \in \mathcal{A}/i \in a} \frac{\sigma_{a,i}^2}{\Delta_a'} \Big) \log(d)^2 \log(T)^3 \bigg) \tag{74}$$

For the gap-free bound, return to Eq. (73)

$$\mathbb{E}[R_T] \leq \Big(d(d+1) + 1 + 1/T \Big) \Delta_{\max} + \mathbb{E}\bigg[\sum_{t=d(d+1)}^{T-1} \Delta_{A_{t+1}} \mathbb{1}\{\mathcal{G}_t \bigcap \mathcal{C}\} \bigg] \,.$$

but we bound the last part differently. Let $\Delta > 0$,

$$\begin{aligned}
\sum_{t=d(d+1)}^{T-1} \Delta_{A_{t+1}} \mathbb{1}\{\mathcal{G}_t \bigcap \mathcal{C}\} &= \sum_{t=d(d+1)}^{T-1} \Delta_{A_{t+1}} \mathbb{1}\Big\{\mathcal{G}_t \bigcap \mathcal{C} \bigcap (\Delta_{A_{t+1}} \leq \Delta)\Big\} + \sum_{t=d(d+1)}^{T-1} \Delta_{A_{t+1}} \mathbb{1}\Big\{\mathcal{G}_t \bigcap \mathcal{C} \bigcap (\Delta_{A_{t+1}} > \Delta)\Big\} \\
&\leq T\Delta + \sum_{t=d(d+1)}^{T-1} \Delta_{A_{t+1}} \mathbb{1}\Big\{\mathcal{G}_t \bigcap \mathcal{C} \bigcap (\Delta_{A_{t+1}} > \Delta)\Big\} \,.
\end{aligned}$$

Adapting Proposition 4.4 for to account for $\Delta_{A_{t+1}} > \Delta$ yields

$$\begin{aligned}
&\sum_{t=0}^{T-1} \Delta_{A_{t+1}} \mathbb{1}\{\mathcal{G}_t \bigcap \mathcal{C} \bigcap (\Delta_{A_{t+1}} > \Delta)\} \\
&\qquad \lesssim \frac{1}{\Delta} \log(T)^3 \log(d)^2 \sum_{i \in [d]} \Big(\max_{a \in \mathcal{A}/i \in a} \sigma_{a,i}^2 \Big) + \big(\log(T) \big)^{5/2} d^2 B_{max} \bigg(1 + \log\bigg(\frac{\Delta_{\max}}{\Delta} \bigg) \bigg) \\
&\qquad\quad + \big(\log(T) \big)^2 d^{7/3} \big(\log(d) \big)^{2/3} B_{max}^{4/3} \frac{1}{\Delta^{1/3}} \,.
\end{aligned}$$

Balancing $T\Delta$ and $\frac{1}{\Delta} \log(T)^3 \log(d)^2 \sum_{i \in [d]} \Big(\max_{a \in \mathcal{A}/i \in a} \sigma_{a,i}^2 \Big)$ yields

$$\mathbb{E}[R_T] = \tilde{O}\bigg(\log(d) \sqrt{\Big(\sum_{i \in [d]} \max_{a \in \mathcal{A}/i \in a} \sigma_{a,i}^2 \Big) T} \bigg) \,. \tag{75}$$

□

E Comparison of algorithms on synthetic environments

This section provides plots for the empirical comparison described in Section 5.4.

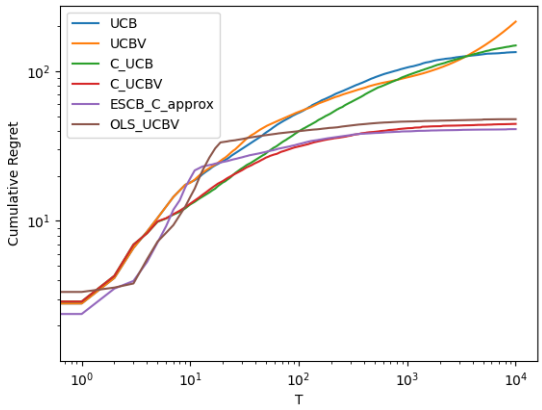


Figure 2: Average Cumulative Regret for environments generated with $d = 10$ and $P = 10$ and exploration parameters fine-tuned.

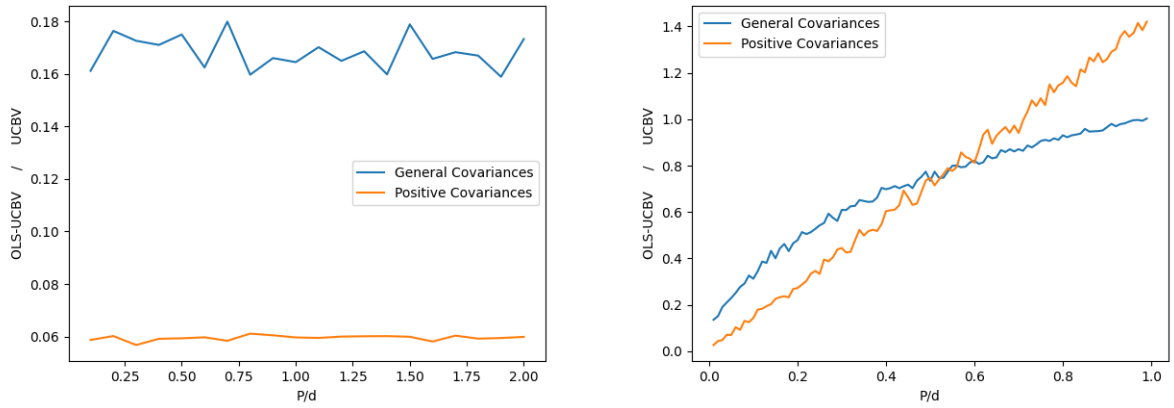


Figure 3: Comparison of the theoretical regret rate of `OLS-UCBV` and `UCBV` for a general structure on the left, and no action overlap on the right.