Active Learning for Accurate Estimation of Linear Models

Carlos Riquelme, Mohammad Ghavamzadeh, Alessandro Lazaric

Active Learning for Accurate Estimation of Linear Models

Carlos Riquelme [1]   **Mohammad Ghavamzadeh** [2]   **Alessandro Lazaric** [3]

Abstract

We explore the sequential decision-making problem where the goal is to estimate a number of linear models uniformly well, given a shared budget of random contexts independently sampled from a known distribution. For each incoming context, the decision-maker selects one of the linear models and receives an observation that is corrupted by the unknown noise level of that model. We present Trace-UCB, an adaptive allocation algorithm that learns the models' noise levels while balancing contexts accordingly across them, and prove bounds for its simple regret in both expectation and high-probability. We extend the algorithm and its bounds to the high dimensional setting, where the number of linear models times the dimension of the contexts is more than the total budget of samples. Simulations with real data suggest that Trace-UCB is remarkably robust, outperforming a number of baselines even when its assumptions are violated.

1. Introduction

We study the problem faced by a decision-maker whose goal is to estimate a number of regression problems equally well (i.e., with a small prediction error for each of them), and has to adaptively allocate a limited budget of samples to the problems in order to gather information and improve its estimates. Two aspects of the problem formulation are key and drive the algorithm design: **1)** The observations Y collected from each regression problem depend on side information (i.e., contexts $X \in \mathbb{R}^d$) and we model the relationship between X and Y in each problem i as a linear function with unknown parameters $\beta_i \in \mathbb{R}^d$, and **2)** The "hardness" of learning each parameter β_i is unknown in advance and may vary across the problems. In particular, we

---

[1]Stanford University, Stanford, CA, USA. [2]DeepMind, Mountain View, CA, USA (The work was done when the author was with Adobe Research). [3]Inria Lille, France. Correspondence to: Carlos Riquelme <rikel@stanford.edu>.

assume that the observations are corrupted by noise levels that are problem-dependent and must be learned as well.

This scenario may arise in a number of different domains where a fixed experimentation budget (number of samples) should be allocated to different problems. Imagine a drug company that has developed several treatments for a particular form of disease. Now it is interested in having an accurate estimate of the performance of each of these treatments for a specific population of patients (e.g., at a particular geographical location). Given the budget allocated to this experiment, a number of patients n can participate in the clinical trial. Volunteered patients arrive sequentially over time and they are represented by a context $X \in \mathbb{R}^d$ summarizing their profile. We model the health status of patient X after being assigned to treatment i by scalar $Y_i \in \mathbb{R}$, which depends on the specific drug through a linear function with parameter β_i (i.e., $Y_i \approx X^\mathsf{T} \beta_i$). The goal is to assign each incoming patient to a treatment in such a way that at the end of the trial, we have an accurate estimate for all β_i's. This will allow us to reliably predict the expected health status of each new patient X for any treatment i. Since the parameters β_i and the noise levels are initially unknown, achieving this goal requires an adaptive allocation strategy for the n patients. Note that while n may be relatively small, as the ethical and financial costs of treating a patient are high, the distribution of the contexts X (e.g., the biomarkers of cancer patients) can be precisely estimated in advance.

This setting is clearly related to the problem of pure exploration and active learning in multi-armed bandits (Antos et al., 2008), where the learner wants to estimate the mean of a finite set of arms by allocating a finite budget of n pulls. Antos et al. (2008) first introduced this setting where the objective is to minimize the largest mean square error (MSE) in estimating the value of each arm. While the optimal solution is trivially to allocate the pulls proportionally to the variance of the arms, when the variances are unknown an exploration-exploitation dilemma arises, where variance and value of the arms must be estimated at the same time in order to allocate pulls where they are more needed (i.e., arms with high variance). Antos et al. (2008) proposed a forcing algorithm where all arms are pulled at least $\sqrt{n}$ times before allocating pulls proportionally to the estimated variances. They derived bounds on the regret, measuring the difference between the MSEs of the learn-

ing algorithm and an optimal allocation showing that the regret decreases as $O(n^{-3/2})$. A similar result is obtained by Carpentier et al. (2011) that proposed two algorithms that use upper confidence bounds on the variance to estimate the MSE of each arm and select the arm with the larger MSE at each step. When the arms are embedded in $\mathbb{R}^d$ and their mean is a linear combination with an unknown parameter, then the problem becomes an optimal experimental design problem (Pukelsheim, 2006), where the objective is to estimate the linear parameter and minimize the prediction error over all arms (see e.g., Wiens & Li 2014; Sabato & Munos 2014). In this paper, we consider an orthogonal extension to the original problem where a finite number of linear regression problems is available (i.e., the arms) and random contexts are observed at each time step. Similarly to the setting of Antos et al. (2008), we assume each problem is characterized by a noise with different variance and the objective is to return regularized least-squares (RLS) estimates with small prediction error (i.e., MSE). While we leverage on the solution proposed by Carpentier et al. (2011) to deal with the unknown variances, in our setting the presence of random contexts make the estimation problem considerably more difficult. In fact, the MSE in one specific regression problem is not only determined by the variance of the noise and the number of samples used to compute the RLS estimate, but also by the contexts observed over time.

Contributions. We propose TRACE-UCB, an algorithm that simultaneously learns the "hardness" of each problem, allocates observations proportionally to these estimates, and balances contexts across problems. We derive performance bounds for TRACE-UCB in expectation and high-probability, and compare the algorithm with several baselines. TRACE-UCB performs remarkably well in scenarios where the dimension of the contexts or the number of instances is large compared to the total budget, motivating the study of the high-dimensional setting, whose analysis and performance bounds are reported in App. F of Riquelme et al. (2017a). Finally, we provide simulations with synthetic data that support our theoretical results, and with real data that demonstrate the robustness of our approach even when some of the assumptions do not hold.

2. Preliminaries

The problem. We consider m linear regression problems, where each instance $i \in [m] = \{1, \ldots, m\}$ is characterized by a parameter $\beta_i \in \mathbb{R}^d$ such that for any context $X \in \mathbb{R}^d$, a random observation $Y \in \mathbb{R}$ is obtained as

$$Y = X^\mathsf{T} \beta_i + \epsilon_i, \qquad (1)$$

where the noise ϵ_i is an i.i.d. realization of a Gaussian distribution $\mathcal{N}(0, \sigma_i^2)$. We denote by $\sigma_{\max}^2 = \max_i \sigma_i^2$ and

by $\overline{\sigma}^2 = 1/m \sum_i \sigma_i^2$, the largest and the average variance, respectively. We define a sequential decision-making problem over n rounds, where at each round $t \in [n]$, the learning algorithm $\mathcal{A}$ receives a context X_t drawn i.i.d. from $\mathcal{N}(0, \Sigma)$, selects an instance I_t, and observes a random sample $Y_{I_t,t}$ according to (1). By the end of the experiment, a training set $\mathcal{D}_n = \{X_t, I_t, Y_{I_t,t}\}_{t \in [n]}$ has been collected and all the m linear regression problems are solved, each problem $i \in [m]$ with its own training set $\mathcal{D}_{i,n}$ (i.e., a subset of $\mathcal{D}_n$ containing samples with $I_t = i$), and estimates of the parameters $\{\hat{\beta}_{i,n}\}_{i \in [m]}$ are returned. For each $\hat{\beta}_{i,n}$, we measure its accuracy by the mean-squared error (MSE)

$$L_{i,n}(\hat{\beta}_{i,n}) = \mathbb{E}_X\big[(X^\mathsf{T}\beta_i - X^\mathsf{T}\hat{\beta}_{i,n})^2\big] = \|\beta_i - \hat{\beta}_{i,n}\|_\Sigma^2. \quad (2)$$

We evaluate the overall accuracy of the estimates returned by the algorithm $\mathcal{A}$ as

$$L_n(\mathcal{A}) = \max_{i \in [m]} \mathbb{E}_{\mathcal{D}_n}\big[L_{i,n}(\hat{\beta}_{i,n})\big], \qquad (3)$$

where the expectation is w.r.t. the randomness of the contexts X_t and observations $Y_{i,t}$ used to compute $\hat{\beta}_{i,n}$. The objective is to design an algorithm $\mathcal{A}$ that minimizes the loss (3). This requires defining an allocation rule to select the instance I_t at each step t and the algorithm to compute the estimates $\hat{\beta}_{i,n}$, e.g., ordinary least-squares (OLS), regularized least-squares (RLS), or Lasso. In designing a learning algorithm, we rely on the following assumption.

Assumption 1. *The covariance matrix Σ of the Gaussian distribution generating the contexts $\{X_t\}_{t=1}^n$ is known.*

This is a standard assumption in active learning, since in this setting the learner has access to the input distribution and the main question is for which context she should ask for a label (Sabato & Munos, 2014; Riquelme et al., 2017b). Often times, companies, like the drug company considered in the introduction, own enough data to have an accurate estimate of the distribution of their customers (patients).

While in the rest of the paper we focus on $L_n(\mathcal{A})$, our algorithm and analysis can be easily extended to similar objectives such as replacing the maximum in (3) with average across all instances, i.e., $1/m \sum_{i=1}^m \mathbb{E}_{\mathcal{D}_n}\big[L_{i,n}(\hat{\beta}_{i,n})\big]$, and using weighted errors, i.e., $\max_i w_i \, \mathbb{E}_{\mathcal{D}_n}\big[L_{i,n}(\hat{\beta}_{i,n})\big]$, by updating the score to focus on the estimated standard deviation and by including the weights in the score, respectively. Later in the paper, we also consider the case where the expectation in (3) is replaced by the high-probability error (see Eq. 17).

Optimal static allocation with OLS estimates. While the distribution of the contexts is fixed and does not depend on the instance i, the errors $L_{i,n}(\hat{\beta}_{i,n})$ directly depend on the variances σ_i^2 of the noise ϵ_i. We define an optimal baseline

obtained when the noise variances $\{\sigma_i^2\}_{i=1}^m$ are known. In particular, we focus on a static allocation algorithm $\mathcal{A}_{\text{stat}}$ that selects each instance i exactly $k_{i,n}$ times, independently of the context,[1] and returns an estimate $\hat{\beta}_{i,n}$ computed by OLS as

$$\widehat{\beta}_{i,n} = \left(\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n}\right)^{-1}\mathbf{X}_{i,n}^\mathsf{T}\mathbf{Y}_{i,n}, \tag{4}$$

where $\mathbf{X}_{i,n} \in \mathbb{R}^{k_{i,n}\times d}$ is the matrix of (random) samples obtained at the end of the experiment, and $\mathbf{Y}_{i,n} \in \mathbb{R}^{k_{i,n}}$ is its corresponding vector of observations. It is simple to show that the global error corresponding to $\mathcal{A}_{\text{stat}}$ is

$$L_n(\mathcal{A}_{\text{stat}}) = \max_{i\in[m]} \frac{\sigma_i^2}{k_{i,n}}\text{Tr}\left(\Sigma\mathbb{E}_{\mathcal{D}_n}\left[\widehat{\Sigma}_{i,n}^{-1}\right]\right), \tag{5}$$

where $\widehat{\Sigma}_{i,n} = \mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n}/k_{i,n} \in \mathbb{R}^{d\times d}$ is the empirical covariance matrix of the contexts assigned to instance i. Since the algorithm does not change the allocation depending on the contexts and $X_t \sim \mathcal{N}(0,\Sigma)$, $\widehat{\Sigma}_{i,n}^{-1}$ is distributed as an inverse-Wishart and we may write (5) as

$$L_n(\mathcal{A}_{\text{stat}}) = \max_{i\in[m]} \frac{d\sigma_i^2}{k_{i,n} - d - 1}. \tag{6}$$

Thus, we derive the following proposition for the optimal static allocation algorithm $\mathcal{A}_{\text{stat}}^*$.

Proposition 1. *Given m linear regression problems, each characterized by a parameter β_i, Gaussian noise with variance σ_i^2, and Gaussian contexts with covariance Σ, let $n > m(d+1)$, then the optimal OLS static allocation algorithm $\mathcal{A}_{\text{stat}}^*$ selects each instance*

$$k_{i,n}^* = \frac{\sigma_i^2}{\sum_j \sigma_j^2}\,n + (d+1)\left(1 - \frac{\sigma_i^2}{\bar{\sigma}^2}\right), \tag{7}$$

times (up to rounding effects), and incurs the global error

$$L_n^* = L_n(\mathcal{A}_{\text{stat}}^*) = \bar{\sigma}^2\frac{md}{n} + O\left(\bar{\sigma}^2\left(\frac{md}{n}\right)^2\right). \tag{8}$$

Proof. See Appendix A.1.[2] □

Proposition 1 divides the problems into two types: those for which $\sigma_i^2 \geq \bar{\sigma}^2$ (*wild* instances) and those for which $\sigma_i^2 < \bar{\sigma}^2$ (*mild* instances). We see that for the first type, the second term in (7) is negative and the instance should be selected less frequently than in the context-free case (where the optimal allocation is given just by the first term). On the other hand, instances whose variance is below the

---

[1]This strategy can be obtained by simply selecting the first instance $k_{1,n}$ times, the second one $k_{2,n}$ times, and so on.

[2]All the proofs can be found in the appendices of the extended version of the paper (Riquelme et al., 2017a).

mean variance should be pulled more often. In any case, we see that the correction to the context-free allocation (i.e., the second term) is *constant*, as it does not depend on n. Nonetheless, it does depend on d and this suggests that in high-dimensional problems, it may significantly skew the optimal allocation.

While $\mathcal{A}_{\text{stat}}^*$ effectively minimizes the prediction loss L_n, it cannot be implemented in practice since the optimal allocation k_i^* requires the variances σ_i^2 to be known at the beginning of the experiment. As a result, we need to devise a learning algorithm $\mathcal{A}$ whose performance approaches L_n^* as n increases. More formally, we define the regret of $\mathcal{A}$ as

$$R_n(\mathcal{A}) = L_n(\mathcal{A}) - L_n(\mathcal{A}_{\text{stat}}^*) = L_n(\mathcal{A}) - L_n^*, \tag{9}$$

and we expect $R_n(\mathcal{A}) = o(1/n)$. In fact, any allocation strategy that selects each instance a linear number of times (e.g., uniform sampling) achieves a loss $L_n = O(1/n)$, and thus, a regret of order $O(1/n)$. However, we expect that the loss of an effective learning algorithm decreases not just at the same rate as L_n^* but also with the very same constant, thus implying a regret that decreases faster than $O(1/n)$.

3. The TRACE-UCB Algorithm

In this section, we present and analyze an algorithm of the form discussed at the end of Section 2, which we call TRACE-UCB, whose pseudocode is in Algorithm 1.

Algorithm 1 TRACE-UCB Algorithm
___
1: **for** $i = 1, \ldots, m$ **do**
2:     Select problem instance i exactly $d+1$ times
3:     Compute its OLS estimates $\hat{\beta}_{i,m(d+1)}$ and $\hat{\sigma}_{i,m(d+1)}^2$
4: **end for**
5: **for** steps $t = m(d+1) + 1, \ldots, n$ **do**
6:     **for** problem instance $1 \leq i \leq m$ **do**
7:        Compute score   ($\Delta_{i,t-1}$ is defined in (11))

$$s_{i,t-1} = \frac{\widehat{\sigma}_{i,t-1}^2 + \Delta_{i,t-1}}{k_{i,t-1}}\text{Tr}\left(\Sigma\hat{\Sigma}_{i,t-1}^{-1}\right)$$

8:     **end for**
9:     Select problem instance $I_t = \arg\max_{i\in[m]} s_{i,t-1}$
10:    Observe X_t and $Y_{I_t,t}$
11:    Update its OLS estimators $\hat{\beta}_{I_t,t}$ and $\hat{\sigma}_{I_t,t}^2$
12: **end for**
13: Return RLS estimates $\{\hat{\beta}_{i,n}^\lambda\}_{i=1}^m$ with regularization λ
___

The regularization parameter $\lambda = O(1/n)$ is provided to the algorithm as input, while in practice one could set λ independently for each arm using cross-validation.

Intuition. Equation (6) suggests that while the parameters of the context distribution, particularly its covariance Σ, do

not impact the prediction error, the noise variances play the most important role in the loss of each problem instance. This is in fact confirmed by the optimal allocation $k_{i,n}^*$ in (7), where only the variances σ_i^2 appear. This evidence suggests that an algorithm similar to GAFS-MAX (Antos et al., 2008) or CH-AS (Carpentier et al., 2011), which were designed for the context-free case (i.e., each instance i is associated to an expected value and not a linear function) would be effective in this setting as well. Nonetheless, (6) holds only for static allocation algorithms that completely ignore the context and the history to decide which instance I_t to choose at time t. On the other hand, adaptive learning algorithms create a strong correlation between the dataset $\mathcal{D}_{t-1}$ collected so far, the current context X_t, and the decision I_t. As a result, the sample matrix $\mathbf{X}_{i,t}$ is no longer a random variable independent of $\mathcal{A}$, and using (6) to design a learning algorithm is not convenient, since the impact of the contexts on the error is completely overlooked. Unfortunately, in general, it is very difficult to study the potential correlation between the contexts $\mathbf{X}_{i,t}$, the intermediate estimates $\hat{\beta}_{i,t}$, and the most suitable choice I_t. However, in the next lemma, we show that if at each step t, we select I_t as a function of $\mathcal{D}_{t-1}$, and *not* X_t, we may still recover an expression for the final loss that we can use as a basis for the construction of an effective learning algorithm.

Lemma 2. *Let $\mathcal{A}$ be a learning algorithm that selects the instances I_t as a function of the previous history, i.e., $\mathcal{D}_{t-1} = \{X_1, I_1, Y_{I_1,1}, \ldots, X_{t-1}, I_{t-1}, Y_{I_{t-1},t-1}\}$ and computes estimates $\hat{\beta}_{i,n}$ using OLS. Then, its loss after n steps can be expressed as*

$$L_n(\mathcal{A}) = \max_{i \in [m]} \mathbb{E}_{\mathcal{D}_n}\left[\frac{\sigma_i^2}{k_{i,n}} \mathrm{Tr}\left(\Sigma \hat{\Sigma}_{i,n}^{-1} \right) \right], \quad (10)$$

where $k_{i,n} = \sum_{t=1}^{n} \mathbb{I}\{I_t = i\}$ and $\hat{\Sigma}_{i,n} = \mathbf{X}_{i,n}^\mathsf{T} \mathbf{X}_{i,n}/k_{i,n}$.

Proof. See Appendix B. $\square$

Remark 1 (assumptions). We assume noise and contexts are Gaussian. The noise Gaussianity is crucial for the estimates of the parameter $\hat{\beta}_{i,t}$ and variance $\hat{\sigma}_{i,t}^2$ to be independent of each other, for each instance i and time t (we actually need and derive a stronger result in Lemma 9, see Appendix B). This is key in proving Lemma 2, as it allows us to derive a closed form expression for the loss function which holds under our algorithm, and is written in terms of the number of pulls and the trace of the inverse empirical covariance matrix. Note that $\hat{\beta}_{i,t}$ drives our loss, while $\hat{\sigma}_{i,t}^2$ drives our decisions. One way to remove this assumption is by defining and directly optimizing a surrogate loss equal to (10) instead of (3). On the other hand, the Gaussianity of contexts leads to the whitened inverse covariance estimate $\Sigma \hat{\Sigma}_{i,n}^{-1}$ being distributed as an inverse Wishart. As there

is a convenient closed formula for its mean, we can find the exact optimal static allocation $k_{i,n}^*$ in Proposition 1, see (7). In general, for sub-Gaussian contexts, no such closed formula for the trace is available. However, as long as the optimal allocation $k_{i,n}^*$ has no second order n^α terms for $1/2 \leq \alpha < 1$, it is possible to derive the same regret rate results that we prove later on for TRACE-UCB.

Equation (10) makes it explicit that the prediction error comes from two different sources. The first one is the noise in the measurements $\mathbf{Y}$, whose impact is controlled by the unknown variances σ_i^2's. Clearly, the larger the σ_i^2 is, the more observations are required to achieve the desired accuracy. At the same time, the *diversity* of contexts across instances also impacts the overall prediction error. This is very intuitive, since it would be a terrible idea for the research center discussed in the introduction to estimate the parameters of a drug by providing the treatment only to a hundred almost identical patients. We say contexts are balanced when $\hat{\Sigma}_{i,n}$ is well conditioned. Therefore, a good algorithm should take care of both aspects.

There are two extreme scenarios regarding the contributions of the two sources of error. **1)** If the number of contexts n is relatively large, since the context distribution is fixed, one can expect that contexts allocated to each instance eventually become balanced (i.e., TRACE-UCB does not bias the distribution of the contexts). In this case, it is the difference in σ_i^2's that drives the number of times each instance is selected. **2)** When the dimension d or the number of arms m is large w.r.t. n, balancing contexts becomes critical, and can play an important role in the final prediction error, whereas the σ_i^2's are less relevant in this scenario. While a learning algorithm cannot deliberately choose a specific context (i.e., X_t is a random variable), we may need to favor instances in which the contexts are poorly balanced and their prediction error is large, despite the fact that they might have small noise variances.

Algorithm. TRACE-UCB is designed as a combination of the upper-confidence-bound strategy used in CH-AS (Carpentier et al., 2011) and the loss in (10), so as to obtain a learning algorithm capable of allocating according to the estimated variances and at the same time balancing the error generated by context mismatch. We recall that all the quantities that are computed at every step of the algorithm are indexed at the beginning and end of a step t by $i, t-1$ (e.g., $\hat{\sigma}_{i,t-1}^2$) and i, t (e.g., $\hat{\beta}_{i,t}$), respectively. At the end of each step t, TRACE-UCB first computes an OLS estimate $\hat{\beta}_{i,t}$, and then use it to estimate the variance $\hat{\sigma}_{i,t}^2$ as

$$\hat{\sigma}_{i,t}^2 = \frac{1}{k_{i,t} - d} \big\| \mathbf{Y}_{i,t} - \mathbf{X}_{i,t}^\mathsf{T} \hat{\beta}_{i,t} \big\|^2,$$

which is the average squared deviation of the predictions based on $\hat{\beta}_{i,t}$. We rely on the following concentration in-

equality for the variance estimate of linear regression with Gaussian noise, whose proof is reported in Appendix C.1.

Proposition 3. *Let the number of pulls $k_{i,t} \geq d + 1$ and $R \geq \max_i \sigma_i^2$. If $\delta \in (0, 3/4)$, then for any instance i and step $t > m(d+1)$, with probability at least $1 - \frac{\delta}{2}$, we have*

$$|\hat{\sigma}_{i,t}^2 - \sigma_i^2| \leq \Delta_{i,t} \triangleq R\sqrt{\frac{64}{k_{i,t}-d}\left(\log\frac{2mn}{\delta}\right)^2}. \quad (11)$$

Given (11), we can construct an upper-bound on the prediction error of any instance i and time step t as

$$s_{i,t-1} = \frac{\hat{\sigma}_{i,t-1}^2 + \Delta_{i,t-1}}{k_{i,t-1}}\, \mathrm{Tr}\left(\Sigma\hat{\Sigma}_{i,t-1}^{-1}\right), \quad (12)$$

and then simply select the instance which maximizes this score, i.e., $I_t = \arg\max_i s_{i,t-1}$. Intuitively, TRACE-UCB favors problems where the prediction error is potentially large, either because of a large noise variance or because of significant unbalance in the observed contexts w.r.t. the target distribution with covariance Σ. A subtle but critical aspect of TRACE-UCB is that by ignoring the current context X_t (but using all the past samples $\mathbf{X}_{t-1}$) when choosing I_t, the distribution of the contexts allocated to each instance stays untouched and the second term in the score $s_{i,t-1}$, i.e., $\mathrm{Tr}(\Sigma\hat{\Sigma}_{i,t-1}^{-1})$, naturally tends to d as more and more (random) contexts are allocated to instance i. This is shown by Proposition 4 whose proof is in Appendix C.2.

Proposition 4. *Force the number of samples $k_{i,t} \geq d + 1$. If $\delta \in (0, 1)$, for any $i \in [m]$ and step $t > m(d+1)$ with probability at least $1 - \delta/2$, we have*

$$\left(1 - C_{\mathrm{Tr}}\sqrt{\frac{d}{k_{i,t}}}\right)^2 \leq \frac{\mathrm{Tr}\left(\Sigma\hat{\Sigma}_{i,t}^{-1}\right)}{d} \leq \left(1 + 2C_{\mathrm{Tr}}\sqrt{\frac{d}{k_{i,t}}}\right)^2,$$

with $C_{\mathrm{Tr}} = 1 + \sqrt{2\log(4nm/\delta)/d}$.

While Proposition 4 shows that the error term due to context mismatch tends to the constant d for all instances i as the number of samples tends to infinity, when t is small w.r.t. d and m, correcting for the context mismatch may significantly improve the accuracy of the estimates $\widehat{\beta}_{i,n}$ returned by the algorithm. Finally, note that while TRACE-UCB uses OLS to compute estimates $\widehat{\beta}_{i,t}$, it computes its returned parameters $\widehat{\beta}_{i,n}$ by ridge regression (RLS) with regularization parameter λ as

$$\hat{\beta}_i^\lambda = (\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n} + \lambda\mathbf{I})^{-1}\mathbf{X}_{i,n}^\mathsf{T}\mathbf{Y}_{i,n}. \quad (13)$$

As we will discuss later, using RLS makes the algorithm more robust and is crucial in obtaining regret bounds both in expectation and high probability.

Performance Analysis. Before proving a regret bound for TRACE-UCB, we report an intermediate result (proof in

App. D.1) that shows that TRACE-UCB *behaves* similarly to the optimal static allocation.

Theorem 5. *Let $\delta > 0$. With probability at least $1 - \delta$, the total number of contexts that TRACE-UCB allocates to each problem instance i after n rounds satisfies*

$$k_{i,n} \geq k_{i,n}^* - \frac{C_\Delta + 8C_{\mathrm{Tr}}}{\sigma_{\min}^2}\sqrt{\frac{nd}{\lambda_{\min}}} - \Omega(n^{1/4}) \quad (14)$$

where $R \geq \sigma_{\max}^2$ is known by the algorithm, and we defined $C_\Delta = 16R\log(2mn/\delta)$ and $\lambda_{\min} = \sigma_{\min}^2/\sum_j \sigma_j^2$.

We now report our regret bound for the TRACE-UCB algorithm. The proof of Theorem 6 is in Appendix D.2.

Theorem 6. *The regret of the Trace-UCB algorithm, i.e., the difference between its loss and the loss of optimal static allocation (see Eq. (8)), is upper-bounded by*

$$L_n(\mathcal{A}) - L_n^* \leq O\left(\frac{1}{\sigma_{\min}^2}\left(\frac{d}{\lambda_{\min}n}\right)^{3/2}\right). \quad (15)$$

Eq. (15) shows that the regret decreases as $O(n^{-3/2})$ as expected. This is consistent with the context-free results (Antos et al., 2008; Carpentier et al., 2011), where the regret decreases as $n^{-3/2}$, which is conjectured to be optimal. However, it is important to note that in the contextual case, the numerator also includes the dimensionality d. Thus, when $n \gg d$, the regret will be small, and it will be larger when $n \approx d$. This motivates studying the high-dimensional setting (App. F). Eq. (15) also indicates that the regret depends on a problem-dependent constant $1/\lambda_{\min}$, which measures the complexity of the problem. Note that when $\sigma_{\max}^2 \approx \sigma_{\min}^2$, we have $1/\lambda_{\min} \approx m$, but $1/\lambda_{\min}$ could be much larger when $\sigma_{\max}^2 \gg \sigma_{\min}^2$.

Remark 2. We introduce a baseline motivated by the context-free problem. At round t, let VAR-UCB selects the instance that maximizes the score[3]

$$s'_{i,t-1} = \frac{\hat{\sigma}_{i,t-1}^2 + \Delta_{i,t-1}}{k_{i,t-1}}. \quad (16)$$

The only difference with the score used by TRACE-UCB is the lack of the trace term in (12). Moreover, the regret of this algorithm has similar *rate* in terms of n and d as that of TRACE-UCB reported in Theorem 6. However, the simulations of Sect. 4 show that the regret of VAR-UCB is actually much higher than that of TRACE-UCB, specially when dm is close to n. Intuitively, when n is close to dm, balancing contexts becomes critical, and VAR-UCB suffers because its score does not explicitly take them into account.

Sketch of the proof of Theorem 6. The proof is divided into three parts. **1)** We show that the behavior of the ridge

---

[3]Note that VAR-UCB is similar to both the CH-AS and B-AS algorithms in Carpentier et al. (2011).

loss of TRACE-UCB is similar to that reported in Lemma 2 for algorithms that rely on OLS; see Lemma 19 in Appendix E. The independence of the $\hat{\beta}_{i,t}$ and $\hat{\sigma}_{i,t}^2$ estimates is again essential (see Remark 1). Although the loss of TRACE-UCB depends on the ridge estimate of the parameters $\hat{\beta}_{i,n}^\lambda$, the decisions made by the algorithm at each round only depend on the variance estimates $\hat{\sigma}_{i,t}^2$ and observed contexts. **2)** We follow the ideas in Carpentier et al. (2011) to lower-bound the total number of pulls $k_{i,n}$ for each $i \in [m]$ under a good event (see Theorem 5 and its proof in Appendix D.1). **3)** We finally use the ridge regularization to bound the impact of those cases *outside* the good event, and combine everything in Appendix D.2.

The regret bound of Theorem 6 shows that the largest *expected* loss across the problem instances incurred by TRACE-UCB quickly approaches the loss of the optimal static allocation algorithm (which knows the true noise variances). While $L_n(\mathcal{A})$ measures the worst *expected* loss, at any specific *realization* of the algorithm, there may be one of the instances which is very poorly estimated. As a result, it would also be desirable to obtain guarantees for the (random) maximum loss

$$\widetilde{L}_n(\mathcal{A}) = \max_{i \in [m]} \|\beta_i - \hat{\beta}_{i,n}\|_\Sigma^2. \tag{17}$$

In particular, we are able to prove the following high-probability bound on $\widetilde{L}_n(\mathcal{A})$ for TRACE-UCB.

Theorem 7. *Let $\delta > 0$, and assume $\|\beta_i\|_2 \leq Z$ for all i, for some $Z > 0$. With probability at least $1 - \delta$,*

$$\widetilde{L}_n \leq \frac{\sum_{j=1}^m \sigma_j^2}{n}\Big(d + 2\log\frac{3m}{\delta}\Big) + O\Big(\frac{1}{\sigma_{\min}^2}\Big(\frac{d}{n\lambda_{\min}}\Big)^{\frac{3}{2}}\Big). \tag{18}$$

Note that the first term in (18) corresponds to the first term of the loss for the optimal static allocation, and the second term is, again, a $n^{-3/2}$ deviation. However, in this case, the guarantees hold *simultaneously* for all the instances.

Sketch of the proof of Theorem 7. In the proof we slightly modify the confidence ellipsoids for the $\hat{\beta}_{i,t}$'s, based on self-normalized martingales, and derived in (Abbasi-Yadkori et al., 2011); see Thm. 13 in App. C. By means of the confidence ellipsoids we control the loss in (17). Their radiuses depend on the number of samples per instance, and we rely on a high-probability events to compute a lower bound on the number of samples. In addition, we need to make sure the mean norm of the contexts will not be too large (see Corollary 15 in App. C). Finally, we combine the lower bound on $k_{i,n}$ with the confidence ellipsoids to conclude the desired high-probability guarantees in Thm. 7.

High-Dimensional Setting. High-dimensional linear models are quite common in practice, motivating the study of the $n < dm$ case, where the algorithms discussed so far

break down. We propose SPARSE-TRACE-UCB in Appendix F, an extension of TRACE-UCB that assumes and takes advantage of *joint* sparsity across the linear functions. The algorithm has two-stages: first, an approximate support is recovered, and then, TRACE-UCB is applied to the induced lower dimensional space. We discuss and extend our high-probability guarantees to SPARSE-TRACE-UCB under suitable standard assumptions in Appendix F.

4. Simulations

In this section, we provide empirical evidence to support our theoretical results. We consider both synthetic and real-world problems, and compare the performance (in terms of normalized MSE) of TRACE-UCB to uniform sampling, optimal static allocation (which requires the knowledge of noise variances), and the context-free algorithm VAR-UCB (see Remark 2). We do not compare to GFSP-MAX and GAFS-MAX (Antos et al., 2008) since they are outperformed by CH-AS Carpentier et al. (2011) and VAR-UCB is the same as CH-AS, except for the fact that we use the concentration inequality in Prop. 3, since we are estimating the variance from a regression problem using OLS.

First, we use synthetic data to ensure that all the assumptions of our model are satisfied, namely we deal with linear regression models with Gaussian context and noise. We set the number of problem instances to $m = 7$ and consider two scenarios: one in which all the noise variances are equal to 1 and one where they are *not* equal, and $\sigma^2 = (0.01, 0.02, 0.75, 1, 2, 2, 3)$. In the latter case, $\sigma_{\max}^2 / \sigma_{\min}^2 = 300$. We study the impact of (independently) increasing dimension d and horizon n on the performance, while keeping all other parameters fixed. Second, we consider real-world datasets in which the underlying model is non-linear and the contexts are not Gaussian, to observe how TRACE-UCB behaves (relative to the baselines) in settings where its main underlying assumptions are violated.

Synthetic Data. In Figures 1(a,b), we display the results for fixed horizon $n = 350$ and increasing dimension d. For each value of d, we run $10,000$ simulations and report the median of the maximum error across the instances for each simulation. In Fig. 1(a), where σ_i^2's are equal, uniform sampling and optimal static allocation execute the same allocation since there is no difference in the expected losses of different instances. Nonetheless we notice that VAR-UCB suffers from poor estimation as soon as d increases, while TRACE-UCB is competitive with the optimal performance. This difference in performance can be explained by the fact that VAR-UCB does not control for contextual balance, which becomes a dominant factor in the loss of a learning strategy for problems of high dimensionality. In Fig. 1(b), in which σ_i^2's are different, uniform sampling is no longer optimal but even in this case VAR-UCB performs
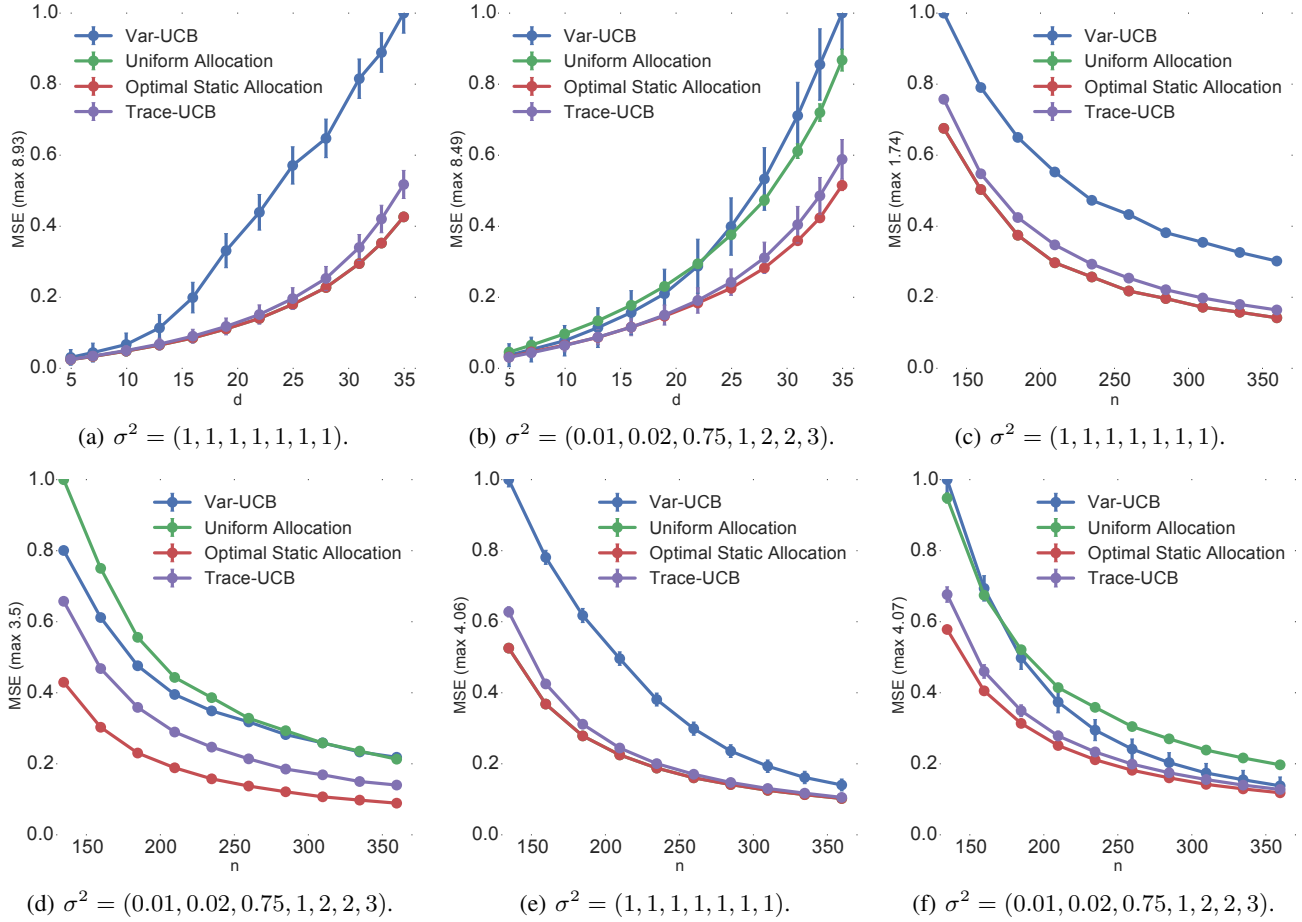
Figure 1. White Gaussian synthetic data with $m = 7$. In Figures (a,b), we set $n = 350$. In Figures (c,d,e,f), we set $d = 10$.

better than uniform sampling only for small $d < 23$, where it is more important to control for the σ_i^2's. For larger dimensions, balancing uniformly the contexts eventually becomes a better strategy, and uniform sampling outperforms VAR-UCB. In this case too, TRACE-UCB is competitive with the optimal static allocation even for large d, successfully balancing both noise variance and contextual error.

Next, we study the performance of the algorithms w.r.t. n. We report two different losses, one in expectation (3) and one in high probability (17), corresponding to the results we proved in Theorems 6 and 7, respectively. In order to approximate the loss in (3) (Figures 1(c,d)) we run $30,000$ simulations, compute the average prediction error for each instance $i \in [m]$, and finally report the maximum mean error across the instances. On the other hand, we estimate the loss in (17) (Figures 1(e,f)) by running $30,000$ simulations, taking the maximum prediction error across the instances for each simulation, and finally reporting their median.

In Figures 1(c, d), we display the loss for fixed dimension $d = 10$ and horizon from $n = 115$ to 360. In Figure 1(c), TRACE-UCB performs similarly to the optimal static allocation, whereas VAR-UCB performs significantly worse,

ranging from 25% to 50% higher errors than TRACE-UCB, due to some catastrophic errors arising from unlucky contextual realizations for an instance. In Fig. 1(d), as the number of contexts grows, uniform sampling's simple context balancing approach is enough to perform as well as VAR-UCB that again heavily suffers from large mistakes. In both figures, TRACE-UCB smoothly learns the σ_i^2's and outperforms uniform sampling and VAR-UCB. Its performance is comparable to that of the optimal static allocation, especially in the case of equal variances in Fig. 1(c).

In Figure 1(e), TRACE-UCB learns and properly balances observations extremely fast and obtains an almost optimal performance. Similarly to figures 1(a,c), VAR-UCB struggles when variances $\hat{\sigma}_i^2$ are almost equal, mainly because it gets confused by random deviations in variance estimates $\hat{\sigma}_i^2$, while overlooking potential and harmful context imbalances. Note that even when $n = 360$ (rightmost point), its median error is still 25% higher than TRACE-UCB's. In Fig. 1(f), as expected, uniform sampling performs poorly, due to mismatch in variances, and only outperforms VAR-UCB for small horizons in which uniform allocation pays off. On the other hand, TRACE-UCB is able to success-
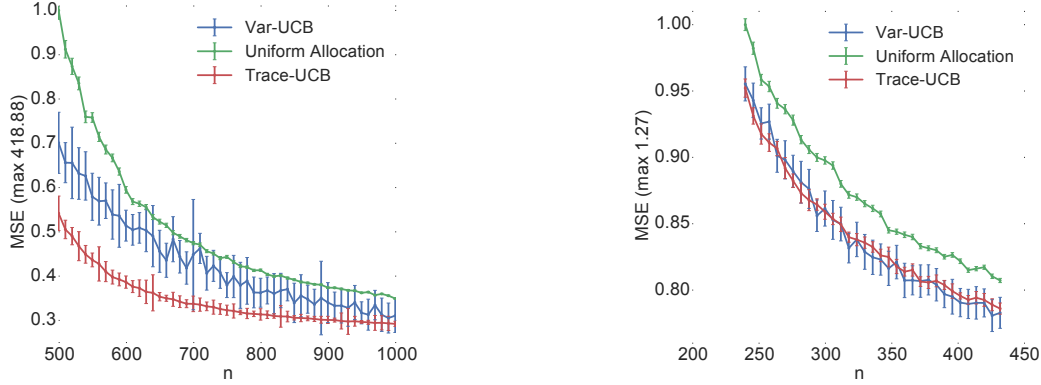
Figure 2. Results on Jester *(left)* with $d = 40, m = 10$ and MovieLens *(right)* with $d = 25, m = 5$. Median over 1000 simulations.

fully handle the tradeoff between learning and allocating according to variance estimates $\hat{\sigma}_i^2$, while accounting for the contextual trace $\widehat{\Sigma}_i$, even for very low n. We observe that for large n, VAR-UCB eventually reaches the performance of the optimal static allocation and TRACE-UCB.

In practice the loss in (17) (figures 1(e,f)) is often more relevant than (3), since it is in high probability and not in expectation, and TRACE-UCB shows excellent performance and robustness, regardless of the underlying variances σ_i^2.

Real Data. TRACE-UCB is based on assumptions such as linearity, and Gaussianity of noise and context that may not hold in practice, where data may show complex dependencies. Therefore, it is important to evaluate the algorithm with real-world data to see its robustness to the violation of its assumptions. We consider two collaborative filtering datasets in which users provide ratings for items. We choose a dense subset of k users and p items, where every user has rated every item. Thus, each user is represented by a p-dimensional vector of ratings. We define the user context by d out of her p ratings, and learn to predict her remaining $m = p - d$ ratings (each one is a problem instance). All item ratings are first centered, so each item's mean is zero. In each simulation, n out of the k users are selected at random to be fed to the algorithm, also in random order. Algorithms can select any instance as the dataset contains the ratings of every instance for all the users. At the end of each simulation, we compute the prediction error for each instance by using the $k - n$ users that did not participate in training for that simulation. Finally, we report the median error across all simulations.

Fig. 2(a) reports the results using the Jester Dataset by (Goldberg et al., 2001) that consists of joke ratings in a continuous scale from -10 to 10. We take $d = 40$ joke ratings as context and learn the ratings for another 9 jokes. In addition, we add another function that counts the total number of movies originally rated by the user. The latter is also centered, bounded to the same scale, and has higher variance (without conditioning on X). The number of to-

tal users is $k = 3811$, and $m = 10$. When the number of observations is limited, the advantage of TRACE-UCB is quite significant (the improvement w.r.t. uniform allocation goes from 45% to almost 20% for large n, while w.r.t. VAR-UCB it goes from almost 30% to roughly 5%), even though the model and context distribution are far from linear and Gaussian, respectively.

Fig. 2(b) shows the results for the MovieLens dataset (Maxwell Harper & Konstan, 2016) that consists of movie ratings between 0 and 5 with 0.5 increments. We select 30 popular movies rated by $k = 1363$ users, and randomly choose $m = 5$ of them to learn (so $d = 25$). In this case, all problems have similar variance ($\hat{\sigma}_{\max}^2 / \hat{\sigma}_{\min}^2 \approx 1.3$) so uniform allocation seems appropriate. Both TRACE-UCB and VAR-UCB modestly improve uniform allocation, while their performance is similar.

5. Conclusions

We studied the problem of adaptive allocation of n contextual samples of dimension d to estimate m linear functions equally well, under heterogenous noise levels σ_i^2 that depend on the linear instance and are unknown to the decision-maker. We proposed TRACE-UCB, an optimistic algorithm that successfully solves the exploration-exploitation dilemma by simultaneously learning the σ_i^2's, allocating samples accordingly to their estimates, and balancing the contextual information across the instances. We also provide strong theoretical guarantees for two losses of interest: in expectation and high-probability. Simulations were conducted in several settings, with both synthetic and real data. The favorable results suggest that TRACE-UCB is reliable, and remarkably robust even in settings that fall outside its assumptions, thus, a useful and simple tool to implement in practice.

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, Cs. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pp. 2312–2320, 2011.

Antos, A., Grover, V., and Szepesvári, Cs. Active learning in multi-armed bandits. In *International Conference on Algorithmic Learning Theory*, pp. 287–302, 2008.

Carpentier, A., Lazaric, A., Ghavamzadeh, M., Munos, R., and Auer, P. Upper-confidence-bound algorithms for active learning in multi-armed bandits. In *Algorithmic Learning Theory*, pp. 189–203. Springer, 2011.

Goldberg, K., Roeder, T., Gupta, D., and Perkins, C. Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4(2):133–151, 2001.

Hastie, T., Tibshirani, R., and Wainwright, M. *Statistical learning with sparsity: the lasso and generalizations*. CRC Press, 2015.

Maxwell Harper, F. and Konstan, J. The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 5(4):19, 2016.

Negahban, S. and Wainwright, M. Simultaneous support recovery in high dimensions: Benefits and perils of block-regularization. *IEEE Transactions on Information Theory*, 57(6):3841–3863, 2011.

Obozinski, G., Wainwright, M., and Jordan, M. Support union recovery in high-dimensional multivariate regression. *The Annals of Statistics*, pp. 1–47, 2011.

Pukelsheim, F. *Optimal Design of Experiments*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 2006.

Raskutti, G., Wainwright, M. J, and Yu, B. Restricted eigenvalue properties for correlated gaussian designs. *Journal of Machine Learning Research*, 11(8):2241–2259, 2010.

Riquelme, C., Ghavamzadeh, M., and Lazaric, A. Active learning for accurate estimation of linear models. *arXiv preprint arXiv:1703.00579*, 2017a.

Riquelme, C., Johari, R., and Zhang, B. Online active linear regression via thresholding. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017b.

Sabato, S. and Munos, R. Active regression by stratification. In *Advances in Neural Information Processing Systems*, pp. 469–477, 2014.

Vershynin, R. Introduction to the non-asymptotic analysis of random matrices. *arXiv:1011.3027*, 2010.

Wainwright, M. *High-dimensional statistics: A non-asymptotic viewpoint*. Draft, 2015.

Wang, W., Liang, Y., and Xing, E. Block regularized lasso for multivariate multi-response linear regression. In *AISTATS*, 2013.

Wiens, D. and Li, P. V-optimal designs for heteroscedastic regression. *Journal of Statistical Planning and Inference*, 145:125–138, 2014.

A. Optimal Static Allocation

A.1. Proof of Proposition 1

Proposition. *Given m linear regression problems, each characterized by a parameter β_i, Gaussian noise with variance σ_i^2, and Gaussian contexts with covariance Σ, let $n > m(d+1)$, then the optimal OLS static allocation algorithm $\mathcal{A}_{stat}^*$ selects each instance*

$$k_{i,n}^* = \frac{\sigma_i^2}{\sum_j \sigma_j^2}\, n + (d+1)\left(1 - \frac{\sigma_i^2}{\overline{\sigma}^2}\right), \tag{19}$$

times (up to rounding effects), and incurs the global error

$$L_n^* = L_n(\mathcal{A}_{stat}^*) = \overline{\sigma}^2 \frac{md}{n} + O\left(\overline{\sigma}^2\left(\frac{md}{n}\right)^2\right). \tag{20}$$

Proof. For the sake of readability in the following we drop the dependency on n.

We first derive the equality in Eq. 2

$$
\begin{aligned}
L_i(\hat{\beta}_i) &= \mathbb{E}_X\left[(X^\mathsf{T}\beta_i - X^\mathsf{T}\hat{\beta}_i)^2\right] \\
&= \mathbb{E}_X[(\hat{\beta}_i - \beta_i)^\mathsf{T} XX^\mathsf{T}(\hat{\beta}_i - \beta_i)] \\
&= (\hat{\beta}_i - \beta_i)^\mathsf{T}\mathbb{E}[XX^\mathsf{T}](\hat{\beta}_i - \beta_i) \\
&= (\hat{\beta}_i - \beta_i)^\mathsf{T}\Sigma(\hat{\beta}_i - \beta_i) \\
&= \|\beta_i - \hat{\beta}_i\|_\Sigma^2.
\end{aligned}
$$

As a result, we can write the global error as

$$
\begin{aligned}
L_n(\mathcal{A}_{\text{stat}}) &= \max_{i\in[m]}\ \mathbb{E}_{\mathcal{D}_{i,n}}\left[\|\beta_i - \hat{\beta}_i\|_\Sigma^2\right] \\
&= \max_{i\in[m]}\ \mathbb{E}_{\mathcal{D}_{i,n}}\left[\text{Tr}\Big((\beta_i - \hat{\beta}_i)^\mathsf{T}\Sigma(\beta_i - \hat{\beta}_i)\Big)\right] \\
&= \max_{i\in[m]}\ \mathbb{E}_{\mathcal{D}_{i,n}}\left[\text{Tr}\Big(\Sigma(\beta_i - \hat{\beta}_i)(\beta_i - \hat{\beta}_i)^\mathsf{T}\Big)\right] \\
&= \max_{i\in[m]}\ \text{Tr}\Big(\mathbb{E}_{\mathcal{D}_{i,n}}\left[\Sigma(\beta_i - \hat{\beta}_i)(\beta_i - \hat{\beta}_i)^\mathsf{T}\right]\Big),
\end{aligned}
$$

where $\mathcal{D}_{i,n}$ is the training set extracted from $\mathcal{D}_n$ containing the samples for instance i. Since contexts and noise are independent random variables, we can decompose $\mathcal{D}_{i,n}$ into the randomness related to the context matrix $\mathbf{X}_i \in \mathbb{R}^{k_i \times d}$ and the noise vector $\epsilon_i \in \mathbb{R}^{k_i}$. We recall that for any fixed realization of $\mathbf{X}_i \in \mathbb{R}^{k_i \times d}$, the OLS estimates $\hat{\beta}_i$ is distributed as

$$\hat{\beta}_i \mid \mathbf{X}_i \sim \mathcal{N}(\beta_i, \sigma_i^2(\mathbf{X}_i^\mathsf{T}\mathbf{X}_i)^{-1}), \tag{21}$$

which means that $\hat{\beta}_i$ conditioned on $\mathbf{X}_i$ is unbiased with covariance matrix given by $\sigma_i^2(\mathbf{X}_i^\mathsf{T}\mathbf{X}_i)^{-1}$. Thus, we can further develop $L_n(\mathcal{A}_{\text{stat}})$ as

$$
\begin{aligned}
L_n(\mathcal{A}_{\text{stat}}) &= \max_{i\in[m]}\ \text{Tr}\left(\mathbb{E}_{\mathbf{X}_i}\left[\mathbb{E}_{\epsilon_i}\left[\Sigma(\beta_i - \hat{\beta}_i)(\beta_i - \hat{\beta}_i)^\mathsf{T}\mid \mathbf{X}_i\right]\right]\right), \tag{22} \\
&= \max_{i\in[m]}\ \sigma_i^2\text{Tr}\left(\Sigma\mathbb{E}_{\mathbf{X}_i}\left[(\mathbf{X}_i^\mathsf{T}\mathbf{X}_i)^{-1}\right]\right) \\
&= \max_{i\in[m]}\ \sigma_i^2\text{Tr}\left(\mathbb{E}_{\mathbf{X}_i}\left[(\overline{\mathbf{X}}_i^\mathsf{T}\overline{\mathbf{X}}_i)^{-1}\right]\right),
\end{aligned}
$$

where $\overline{X} = \Sigma^{-1/2}X$ is a whitened context and $\overline{\mathbf{X}}_i$ is its corresponding whitened matrix. Since whitened contexts $\overline{X}$ are distributed as $\mathcal{N}(0, I)$, we know that $(\overline{\mathbf{X}}_i^\mathsf{T}\overline{\mathbf{X}}_i)^{-1}$ is distributed as an inverse Wishart $\mathcal{W}^{-1}(I_d, k_i)$, whose expectation is

$I_d/(k_i - d - 1)$, and thus,

$$L_n(\mathcal{A}_{\text{stat}}) = \max_{i \in [m]} \sigma_i^2 \text{Tr}\left[\frac{1}{k_i - d - 1} I_d\right] = \max_{i \in [m]} \frac{\sigma_i^2 \, d}{k_i - d - 1}. \tag{23}$$

Note that this final expression requires that $k_i > d + 1$, since it is not possible to compute an OLS estimate with less than $d + 1$ samples. Therefore, we proceed by minimizing Eq. 23, subject to $k_i > d + 1$. We write $k_i = k_i' + d + 1$ for some $k_i' > 0$. Thus, equivalently, we minimize

$$L_n(\mathcal{A}_{\text{stat}}) = \max_i \frac{\sigma_i^2 \, d}{k_i'}. \tag{24}$$

Since $\sum_i k_i' = n - m(d + 1)$, we may conclude that the optimal k_i' is given by

$$k_i' = \frac{\sigma_i^2}{\sum_j \sigma_j^2}\left(n - m(d + 1)\right),$$

so that all the terms in the RHS of Eq. 24 are equal. This gives us the optimal static allocation

$$\begin{aligned} k_i^* &= \frac{\sigma_i^2}{\sum_j \sigma_j^2}\left(n - m(d + 1)\right) + d + 1 \\ &= \frac{\sigma_i^2}{\sum_j \sigma_j^2} n + (d + 1)\left(1 - \frac{\sigma_i^2}{\bar{\sigma}^2}\right), \end{aligned} \tag{25}$$

where $\bar{\sigma}^2 = (1/m)\sum_i \sigma_i^2$ is the mean variance across the m problem instances.

Thus, for the optimal static allocation, the expected loss is given by

$$\begin{aligned} L_n^* = L_n(\mathcal{A}_{\text{stat}}^*) &= d \max_i \frac{\sigma_i^2}{\frac{\sigma_i^2}{\sum_j \sigma_j^2} n - (d + 1)\frac{\sigma_i^2}{\bar{\sigma}^2}} \\ &= \frac{\left(\sum_j \sigma_j^2\right) d}{n - m(d + 1)} \\ &= \frac{\left(\sum_j \sigma_j^2\right) d}{n} + \frac{\left(\sum_j \sigma_j^2\right) md(d + 1)}{n\left(n - m(d + 1)\right)} \\ &= \frac{\left(\sum_j \sigma_j^2\right) d}{n} + O\left(\frac{\left(\sum_j \sigma_j^2\right) md^2}{n^2}\right), \end{aligned}$$

which concludes the proof. Furthermore the following bounds trivially holds for any $n \geq 2m(d + 1)$

$$\frac{md\bar{\sigma}^2}{n} \leq L_n^* \leq 2\frac{md\bar{\sigma}^2}{n}.$$

$\square$

B. Loss of an OLS-based Learning Algorithm (Proof of Lemma 2)

Unlike in the proof of Proposition 1, when the number of pulls is random *and* depends on the value of the previous observations (through $\mathcal{D}_n$), then in general, the OLS estimates $\widehat{\beta}_{i,n}$ are no longer distributed as Eq. 21 and the derivation for $\mathcal{A}_{\text{stat}}$ no longer holds. In fact, for a learning algorithm, the value $k_{i,t}$ itself provides some information about the observations that have been obtained up until time t and were used by the algorithm to determine $k_{i,t}$. In the following, we show that by ignoring the current context X_t when choosing instance I_t, we are still able to analyze the loss of TRACE-UCB and obtain a result very similar to the static case.

We first need two auxiliary lemmas (Lemmas 8 and 9), one on the computation of an empirical estimate of the variance of the noise, and an independence result between the variance estimate and the linear regression estimate.

Lemma 8. *In any linear regression problem with noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$, after $t \geq d + 1$ samples, given an OLS estimator $\widehat{\beta}_t$, the noise variance estimator can be computed in a recurrent form as*

$$\hat{\sigma}_{t+1}^2 = \frac{t-d}{t-d+1} \hat{\sigma}_t^2 + \frac{1}{t-d+1} \frac{(X_{t+1}^{\mathsf{T}} \hat{\beta}_t - Y_{t+1})^2}{1 + X_{t+1}^{\mathsf{T}} (\mathbf{X}_t^{\mathsf{T}} \mathbf{X}_t)^{-1} X_{t+1}}, \tag{26}$$

where $\mathbf{X}_t \in \mathbb{R}^{t \times d}$ is the sample matrix.

Proof. We first recall the "batch" definition of the variance estimator

$$\widehat{\sigma}_t^2 = \frac{1}{t-d} \sum_{s=1}^{t} (Y_s - X_s^{\mathsf{T}} \widehat{\beta}_t)^2 = \frac{1}{t-d} \|\mathbf{Y}_t - \mathbf{X}_t^{\mathsf{T}} \widehat{\beta}_t\|^2$$

Since $\mathbf{Y}_t = \mathbf{X}_t \beta + \boldsymbol{\epsilon}_t$ and $\widehat{\beta}_t = \beta + (\mathbf{X}_t^{\mathsf{T}} \mathbf{X}_t)^{-1} \mathbf{X}_t^{\mathsf{T}} \boldsymbol{\epsilon}_t$, we have

$$\widehat{\sigma}_t^2 = \frac{1}{t-d} \|(\mathbf{X}_t^{\mathsf{T}} \mathbf{X}_t)^{-1} \mathbf{X}_t^{\mathsf{T}} \boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_t\|^2 = \frac{1}{t-d} \Big(\boldsymbol{\epsilon}_t^{\mathsf{T}} \boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_t^{\mathsf{T}} \mathbf{X}_t (\mathbf{X}_t^{\mathsf{T}} \mathbf{X}_t)^{-1} \mathbf{X}_t^{\mathsf{T}} \boldsymbol{\epsilon}_t \Big) = \frac{1}{t-d} (E_{t+1} - V_{t+1}).$$

We now devise a recursive formulation for the two terms in the previous expression. We have

$$E_{t+1} = \boldsymbol{\epsilon}_{t+1}^{\mathsf{T}} \boldsymbol{\epsilon}_{t+1} = \boldsymbol{\epsilon}_t^{\mathsf{T}} \boldsymbol{\epsilon}_t + \epsilon_{t+1}^2 = E_t + \epsilon_{t+1}^2.$$

In order to analyze the second term we first introduce the design matrix $\mathbf{S}_t = \mathbf{X}_t^{\mathsf{T}} \mathbf{X}_t$, which has the simple update rule $\mathbf{S}_{t+1} = \mathbf{S}_t + X_{t+1} X_{t+1}^{\mathsf{T}}$. Then we have

$$\begin{aligned}
V_{t+1} &= \boldsymbol{\epsilon}_{t+1}^{\mathsf{T}} \mathbf{X}_{t+1} (\mathbf{X}_{t+1}^{\mathsf{T}} \mathbf{X}_{t+1})^{-1} \mathbf{X}_{t+1}^{\mathsf{T}} \boldsymbol{\epsilon}_{t+1} \\
&= \big(\boldsymbol{\epsilon}_t^{\mathsf{T}} \mathbf{X}_t + \epsilon_{t+1} X_{t+1}^{\mathsf{T}} \big) \big(\mathbf{S}_t + X_{t+1} X_{t+1}^{\mathsf{T}} \big)^{-1} \big(\boldsymbol{\epsilon}_t^{\mathsf{T}} \mathbf{X}_t + \epsilon_{t+1} X_{t+1}^{\mathsf{T}} \big)^{\mathsf{T}} \\
&= \big(\boldsymbol{\epsilon}_t^{\mathsf{T}} \mathbf{X}_t + \epsilon_{t+1} X_{t+1}^{\mathsf{T}} \big) \bigg(\mathbf{S}_t^{-1} - \frac{\mathbf{S}_t^{-1} X_{t+1} X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1}}{1 + X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1} X_{t+1}} \bigg) \big(\boldsymbol{\epsilon}_t^{\mathsf{T}} \mathbf{X}_t + \epsilon_{t+1} X_{t+1}^{\mathsf{T}} \big)^{\mathsf{T}},
\end{aligned}$$

where we used the Sherman-Morrison formula in the last equality. We further develop the previous expression as

$$\begin{aligned}
V_{t+1} = &V_t + \epsilon_{t+1} X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1} X_{t+1} \epsilon_{t+1} + 2\epsilon_{t+1} X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1} \mathbf{X}_t^{\mathsf{T}} \boldsymbol{\epsilon}_t \\
&- \boldsymbol{\epsilon}_t^{\mathsf{T}} \mathbf{X}_t \frac{\mathbf{S}_t^{-1} X_{t+1} X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1}}{1 + X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1} X_{t+1}} \mathbf{X}_t^{\mathsf{T}} \boldsymbol{\epsilon}_t - \epsilon_{t+1} X_{t+1}^{\mathsf{T}} \frac{\mathbf{S}_t^{-1} X_{t+1} X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1}}{1 + X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1} X_{t+1}} X_{t+1} \epsilon_{t+1} - 2\boldsymbol{\epsilon}_t^{\mathsf{T}} \mathbf{X}_t \frac{\mathbf{S}_t^{-1} X_{t+1} X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1}}{1 + X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1} X_{t+1}} X_{t+1} \epsilon_{t+1}.
\end{aligned}$$

We define $\alpha_{t+1} = X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1} \mathbf{X}_t^{\mathsf{T}} \boldsymbol{\epsilon}_t$ and $\psi_{t+1} = X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1} X_{t+1}$, and then obtain

$$\begin{aligned}
V_{t+1} &= V_t + \epsilon_{t+1}^2 \psi_{t+1} + 2\alpha_{t+1} \epsilon_{t+1} - \frac{\alpha_{t+1}^2}{1 + \psi_{t+1}} - \epsilon_{t+1}^2 \frac{\psi_{t+1}^2}{1 - \psi_{t+1}} - 2\epsilon_{t+1} \frac{\alpha_{t+1} \psi_{t+1}}{1 + \psi_{t+1}} \\
&= V_t + \epsilon_{t+1}^2 \Big(\psi_{t+1} + \frac{\psi_{t+1}^2}{1 + \psi_{t+1}} \Big) + 2\epsilon_{t+1} \frac{\alpha_{t+1}}{1 + \psi_{t+1}} - \frac{\alpha_{t+1}^2}{1 + \psi_{t+1}}.
\end{aligned}$$

Bringing everything together we obtain

$$E_{t+1} - V_{t+1} = E_t - V_t + \epsilon_{t+1}^2 \left(1 - \psi_{t+1} + \frac{\psi_{t+1}^2}{1 + \psi_{t+1}} \right) - 2\epsilon_{t+1} \frac{\alpha_{t+1}}{1 + \psi_{t+1}} + \frac{\alpha_{t+1}^2}{1 + \psi_{t+1}}$$

$$= E_t - V_t + \frac{1}{1 + \psi_{t+1}} \left(\epsilon_{t+1}^2 - 2\epsilon_{t+1}\alpha_{t+1} + \alpha_{t+1} \right) = E_t - V_t + \frac{\left(\epsilon_{t+1} - \alpha_{t+1} \right)^2}{1 + \psi_{t+1}}.$$

Since $\epsilon_{t+1} = Y_{t+1} - X_{t+1}^\mathsf{T}\beta$, we may write

$$E_{t+1} - V_{t+1} = E_t - V_t + \frac{\left(Y_{t+1} - X_{t+1}^\mathsf{T}(\beta + \mathbf{S}_t^{-1}\mathbf{X}_t^\mathsf{T}\boldsymbol{\epsilon}_t) \right)^2}{1 + \psi_{t+1}} = E_t - V_t + \frac{\left(Y_{t+1} - X_{t+1}^\mathsf{T}\widehat{\beta}_t \right)^2}{1 + \psi_{t+1}}.$$

Recalling the definition of the variance estimate, we finally obtain

$$\hat{\sigma}_{t+1}^2 = \frac{1}{t - d + 1}(E_{t+1} - V_{t+1}) = \frac{1}{t - d + 1}(E_t - V_t) + \frac{1}{t - d + 1}\frac{\left(Y_{t+1} - X_{t+1}^\mathsf{T}\widehat{\beta}_t \right)^2}{1 + X_{t+1}^\mathsf{T}\mathbf{S}_t^{-1}X_{t+1}}$$

$$= \frac{t - d}{t - d + 1}\hat{\sigma}_t^2 + \frac{1}{t - d + 1}\frac{\left(Y_{t+1} - X_{t+1}^\mathsf{T}\widehat{\beta}_t \right)^2}{1 + X_{t+1}^\mathsf{T}\mathbf{S}_t^{-1}X_{t+1}},$$

which concludes the proof. $\qquad\square$

Lemma 9. *Let $\mathcal{F}_j$ be the σ-algebra generated by $X_1, \ldots, X_n$ and $\hat{\sigma}_1^2, \ldots, \hat{\sigma}_j^2$. Then, for any $j \geq d$,*

$$\hat{\beta}_j \mid \mathcal{F}_j \sim \mathcal{N}(\beta, \sigma^2 \, (X_{1:j}^T X_{1:j})^{-1}). \tag{27}$$

Proof. We prove the lemma by induction. The statement is true for $t = d$. We want to prove the induction, that is if $\hat{\beta}_t \mid \mathcal{F}_t \sim \mathcal{N}(\beta, \sigma^2 \, (\mathbf{X}_t^\mathsf{T}\mathbf{X}_t)^{-1})$, then

$$\hat{\beta}_{t+1} \mid \mathcal{F}_{t+1} \sim \mathcal{N}(\beta, \sigma^2 (\mathbf{X}_{t+1}^\mathsf{T}\mathbf{X}_{t+1})^{-1}). \tag{28}$$

Let us first derive a recursive expression for $\hat{\beta}_{t+1}$. Let $\mathbf{S}_t = \mathbf{X}_t^\mathsf{T}\mathbf{X}_t$, then

$$\widehat{\beta}_{t+1} = \beta + \mathbf{S}_{t+1}^{-1}\mathbf{X}_{t+1}^\mathsf{T}\boldsymbol{\epsilon}_{t+1} = \left(\mathbf{S}_t + X_{t+1}X_{t+1}^\mathsf{T} \right)^{-1}\left(\mathbf{X}_t^\mathsf{T}\boldsymbol{\epsilon}_t + \epsilon_{t+1}X_{t+1} \right)$$

$$= \left(\mathbf{S}_t^{-1} - \frac{\mathbf{S}_t^{-1}X_{t+1}X_{t+1}^\mathsf{T}\mathbf{S}_t^{-1}}{1 + X_{t+1}^\mathsf{T}\mathbf{S}_t^{-1}X_{t+1}} \right)\left(\mathbf{X}_t^\mathsf{T}\boldsymbol{\epsilon}_t + \epsilon_{t+1}X_{t+1} \right),$$

where we used Sherman-Morrison formula. By developing the previous expression we obtain

$$\widehat{\beta}_{t+1} = \left(\beta + \mathbf{S}_t^{-1}\mathbf{X}_t^\mathsf{T}\boldsymbol{\epsilon}_t \right) + \epsilon_{t+1}\mathbf{S}_t^{-1}X_{t+1}\left(1 - \frac{X_{t+1}^\mathsf{T}\mathbf{S}_t^{-1}X_{t+1}}{1 + X_{t+1}^\mathsf{T}\mathbf{S}_t^{-1}X_{t+1}} \right) - \frac{\mathbf{S}_t^{-1}X_{t+1}X_{t+1}^\mathsf{T}\mathbf{S}_t^{-1}\mathbf{X}_t^\mathsf{T}\boldsymbol{\epsilon}_t}{1 + X_{t+1}^\mathsf{T}\mathbf{S}_t^{-1}X_{t+1}}$$

$$= \widehat{\beta}_t + \frac{\epsilon_{t+1}\mathbf{S}_t^{-1}X_{t+1}}{1 + X_{t+1}^\mathsf{T}\mathbf{S}_t^{-1}X_{t+1}} - \frac{\mathbf{S}_t^{-1}X_{t+1}X_{t+1}^\mathsf{T}(\widehat{\beta}_t - \beta)}{1 + X_{t+1}^\mathsf{T}\mathbf{S}_t^{-1}X_{t+1}}.$$

We can conveniently rewrite the previous expression as

$$\widehat{\beta}_{t+1} - \beta = \left(I - \frac{\mathbf{S}_t^{-1}X_{t+1}X_{t+1}^\mathsf{T}}{1 + X_{t+1}^\mathsf{T}\mathbf{S}_t^{-1}X_{t+1}} \right)(\widehat{\beta}_t - \beta) + \frac{\epsilon_{t+1}\mathbf{S}_t^{-1}X_{t+1}}{1 + X_{t+1}^\mathsf{T}\mathbf{S}_t^{-1}X_{t+1}}$$

$$= (I - \alpha_t)(\widehat{\beta}_t - \beta) + \gamma_t\epsilon_{t+1}, \tag{29}$$

where $\alpha_t \in \mathbb{R}^{d \times d}$ and $\gamma_t \in \mathbb{R}^d$ are defined implicitly. By Lemma 8, we notice that the sequence of empirical variances in $\mathcal{F}_t$ is equivalent to the sequence of squared deviations up to t. In order to make this equivalence more apparent we define the filtration

$$\mathcal{G}_t = \left\{ \{X_s\}_{s=1}^n \cup \hat{\sigma}_2^2 \cup \{(X_{s+1}^\mathsf{T}\hat{\beta}_s - \epsilon_{s+1})^2\}_{s=2}^{t-1} \right\},$$

so that $\hat{\beta}_{t+1} \mid \mathcal{F}_{t+1} \sim \hat{\beta}_{t+1} \mid \mathcal{G}_{t+1}$. We introduce two auxiliary random vectors conditioned on $\mathcal{G}$

$$U = \epsilon_{t+1} - X_{t+1}^{\mathsf{T}}(\hat{\beta}_t - \beta) \mid \mathcal{G}_t, \qquad V = \hat{\beta}_{t+1} - \beta \mid \mathcal{G}_t.$$

We want to show that the random vectors $U \in \mathbb{R}$ and $V \in \mathbb{R}^d$ are independent. We first recall that the noise $\epsilon_{t+1} \mid \mathcal{G}_t \sim \mathcal{N}(0, \sigma^2)$, and it is independent of $\epsilon_1, \ldots, \epsilon_t$, and $\hat{\beta}_t$ under $\mathcal{G}_t$. Furthermore, by the induction assumption $\hat{\beta}_t \mid \mathcal{G}_t$ is also Gaussian, so we have that $(\hat{\beta}_t, \epsilon_{t+1})$ are jointly Gaussian given $\mathcal{G}_t$. Then we can conveniently rewrite U as

$$U = (-X_{t+1}, 1)^{\mathsf{T}}(\hat{\beta}_t, \epsilon_{t+1}) + X_{t+1}^{\mathsf{T}}\beta,$$

which shows that it is a Gaussian vector. Using the recursive formulation in Eq. 29 we can also rewrite V as

$$V = (\mathrm{Id} - \alpha_t)(\hat{\beta}_t - \beta) + \gamma_t\,\epsilon_{t+1} = \begin{bmatrix} \mathrm{I} - \alpha_t & \gamma_t \end{bmatrix} \begin{bmatrix} \hat{\beta}_t - \beta \\ \epsilon_{t+1} \end{bmatrix},$$

which is also Gaussian. Furthermore, we notice that under the induction assumption, $\mathbb{E}_{\mathcal{G}_t}[U] = 0$ and $\mathbb{E}_{\mathcal{G}_t}[V] = 0$ and thus we need to show that $\mathbb{E}[UV \mid \mathcal{G}_t] = 0$ to prove that U and V are uncorrelated

$$\begin{aligned}
\mathbb{E}[UV \mid \mathcal{G}_t] &= \mathbb{E}_{\mathcal{G}_t}\left[\left(\epsilon_{t+1} - X_{t+1}^{\mathsf{T}}(\hat{\beta}_t - \beta)\right)\left((\mathrm{Id} - \alpha_t)(\hat{\beta}_t - \beta) + \gamma_t\,\epsilon_{t+1}\right)\right] \\
&= \gamma_t\,\mathbb{E}_{\mathcal{G}_t}\left[\epsilon_{t+1}^2\right] - \mathbb{E}_{\mathcal{G}_t}\left[X_{t+1}^{\mathsf{T}}(\hat{\beta}_t - \beta)(\mathrm{I} - \alpha_t)(\hat{\beta}_t - \beta)\right] \\
&= \sigma^2\gamma_t - \mathbb{E}_{\mathcal{G}_t}\left[(\mathrm{I} - \alpha_t)(\hat{\beta}_t - \beta)(\hat{\beta}_t - \beta)^{\mathsf{T}}X_{t+1}\right] \\
&= \sigma^2\gamma_t - (\mathrm{I} - \alpha_t)\,\mathbb{E}_{\mathcal{G}_t}\left[(\hat{\beta}_t - \beta)(\hat{\beta}_t - \beta)^{\mathsf{T}}\right]X_{t+1} \\
&= \sigma^2\gamma_t - \sigma^2(\mathrm{I} - \alpha_t)(\mathbf{X}_t^{\mathsf{T}}\mathbf{X}_t)^{-1}X_{t+1} \\
&= \sigma^2 \frac{\mathbf{S}_t^{-1}X_{t+1}}{1 + X_{t+1}^{\mathsf{T}}\mathbf{S}_t^{-1}X_{t+1}} - \sigma^2\left(\mathrm{I} - \frac{\mathbf{S}_t^{-1}X_{t+1}X_{t+1}^{\mathsf{T}}}{1 + X_{t+1}^{\mathsf{T}}\mathbf{S}_t^{-1}X_{t+1}}\right)\mathbf{S}_t^{-1}X_{t+1} \\
&= \sigma^2 \frac{\mathbf{S}_t^{-1}X_{t+1} - (1 + X_{t+1}^{\mathsf{T}}\mathbf{S}_t^{-1}X_{t+1})\mathbf{S}_t^{-1}X_{t+1} + \mathbf{S}_t^{-1}X_{t+1}X_{t+1}^{\mathsf{T}}\mathbf{S}_t^{-1}X_{t+1}}{1 + X_{t+1}^{\mathsf{T}}\mathbf{S}_t^{-1}X_{t+1}} \\
&= 0.
\end{aligned}$$

It thus follows that, as U and V are uncorrelated, they are also independent. Combining the definition of $\mathcal{G}_t$, U and its independence w.r.t V, we have

$$\begin{aligned}
V \mid \mathcal{G}_{j+1} &= V \mid U, \mathcal{G}_j \\
&= V \mid \{X_1, \ldots, X_T, \hat{\sigma}_2^2, \{(X_{s+1}^{\mathsf{T}}\hat{\beta}_s - \epsilon_{s+1})^2\}_{s=2}^{t-1}\} \\
&= \begin{bmatrix} \mathrm{I} - \alpha_t & \gamma_t \end{bmatrix} \begin{bmatrix} \hat{\beta}_t - \beta \\ \epsilon_{t+1} \end{bmatrix} \mid \mathcal{G}_t.
\end{aligned}$$

By the induction hypothesis the vector in the previous expression is distributed as

$$\begin{bmatrix} \hat{\beta}_t - \beta \\ \epsilon_{t+1} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} S_t^{-1} & 0 \\ 0 & 1 \end{bmatrix}\right).$$

Therefore, we conclude that

$$V \mid \mathcal{G}_{t+1} \sim \mathcal{N}\left(0, \sigma^2 \begin{bmatrix} \mathrm{I} - \alpha_t & \gamma_t \end{bmatrix} \begin{bmatrix} S_t^{-1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathrm{I} - \alpha_t & \gamma_t \end{bmatrix}^{\mathsf{T}}\right) = \mathcal{N}(0, \sigma^2\,\Sigma'),$$

where the covariance matrix Σ' can be written as

$$\begin{aligned}
\Sigma' &= \begin{bmatrix} \mathrm{I} - \alpha_t & \gamma_t \end{bmatrix} \begin{bmatrix} S_t^{-1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathrm{I} - \alpha_t & \gamma_t \end{bmatrix}^{\mathsf{T}} \\
&= \begin{bmatrix} \mathrm{I} - \alpha_t & \gamma_t \end{bmatrix} \begin{bmatrix} S_t^{-1}(\mathrm{I} - \alpha_t)^{\mathsf{T}} \\ \gamma_t^{\mathsf{T}} \end{bmatrix} \\
&= (\mathrm{I} - \alpha_t)S_t^{-1}(\mathrm{I} - \alpha_t)^{\mathsf{T}} + \gamma_t\gamma_t^{\mathsf{T}}.
\end{aligned}$$

Recalling the definitions of α_t and γ_t, and defining $\psi_{t+1} = X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1} X_{t+1}$

$$
\begin{aligned}
\Sigma' &= \left(\mathrm{I} - \frac{\mathbf{S}_t^{-1} X_{t+1} X_{t+1}^{\mathsf{T}}}{1 + X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1} X_{t+1}} \right) \mathbf{S}_t^{-1} \left(\mathrm{I} - \frac{\mathbf{S}_t^{-1} X_{t+1} X_{t+1}^{\mathsf{T}}}{1 + X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1} X_{t+1}} \right)^{\mathsf{T}} \\
&\quad + \left(\frac{\mathbf{S}_t^{-1} X_{t+1}}{1 + X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1} X_{t+1}} \right) \left(\frac{\mathbf{S}_t^{-1} X_{t+1}}{1 + X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1} X_{t+1}} \right)^{\mathsf{T}} \\
&= \mathbf{S}_t^{-1} - 2 \frac{\mathbf{S}_t^{-1} X_{t+1} X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1}}{1 + r} + \psi_{t+1} \frac{\mathbf{S}_t^{-1} X_{t+1} X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1}}{(1 + \psi_{t+1})^2} + \frac{\mathbf{S}_t^{-1} X_{t+1} X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1}}{(1 + \psi_{t+1})^2} \\
&= \mathbf{S}_t^{-1} - \frac{\mathbf{S}_t^{-1} X_{t+1} X_{t+1}^{\mathsf{T}} \mathbf{S}_t^{-1}}{1 + \psi_{t+1}} = S_{t+1}^{-1} = (\mathbf{X}_{t+1}^{\mathsf{T}} \mathbf{X}_{t+1})^{-1},
\end{aligned}
$$

where we applied the Woodbury matrix identity in the last step. Finally, it follows that

$$
\hat{\beta}_{t+1} \mid \mathcal{F}_{t+1} \sim \mathcal{N}(\beta, \sigma^2 (\mathbf{X}_{t+1}^{\mathsf{T}} \mathbf{X}_{t+1})^{-1}),
$$

and the induction is complete. $\qquad\square$

Now we can prove Lemma 2:

Lemma. *Let $\mathcal{A}$ be a learning algorithm that selects instances I_t as a function of the previous history, that is, $\mathcal{D}_{t-1} = \{X_1, I_1, Y_{I_1, 1}, \ldots, X_{t-1}, I_{t-1}, Y_{I_{t-1}, t-1}\}$ and computes estimates $\widehat{\beta}_{i,n}$ using OLS. Then, its loss after n steps can be expressed as*

$$
L_n(\mathcal{A}) = \max_{i \in [m]} \mathbb{E}_{\mathcal{D}_t} \left[\frac{\sigma_i^2}{k_{i,n}} \mathrm{Tr}\left(\Sigma \widehat{\Sigma}_{i,n}^{-1} \right) \right], \tag{30}
$$

where $k_{i,n} = \sum_{t=1}^{n} \mathbb{I}\{I_t = i\}$ and $\widehat{\Sigma}_{i,n} = \mathbf{X}_{i,n}^{\mathsf{T}} \mathbf{X}_{i,n} / k_{i,n}$.

Proof. For any instance i, we can assume that the following random variables are sampled before TRACE-UCB starts collecting observations (we omit the i index in the table):

$t = 1$	$t = 2$	$\ldots$	$t = n$
X_1	X_2	$\ldots$	X_n
ϵ_1	ϵ_2	$\ldots$	ϵ_n
$\hat{\beta}_1$	$\hat{\beta}_2$	$\ldots$	$\hat{\beta}_n$
$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\ldots$	$\hat{\sigma}_n^2$

As a result, we can interpret TRACE-UCB as controlling the *stopping* time $t_i = k_{i,n}$ for each problem i, that is, the total number of samples $k_{i,n}$, leading to the final estimates $\hat{\beta}_{t_i}$ and $\hat{\sigma}_{t_i}^2$. In the following we introduce the notation $\mathbf{X}_{1:j}$ as the sample matrix constructed from exactly j samples, unlike $\mathbf{X}_{i,n}$ which is the sample matrix obtained with $k_{i,n}$. So we have $\mathbf{X}_{1:k_{i,n}} = \mathbf{X}_{i,n}$. Crucially, when the errors ϵ_j are Gaussian, then $\hat{\beta}_j \mid \mathbf{X}_{1:j}$ and $\hat{\sigma}_j^2 \mid \mathbf{X}_{1:j}$ are independent for any fixed j (note these random variables have nothing to do with the algorithm's decisions).

Let $\mathcal{F}_j$ be the σ-algebra generated by $X_1, \ldots, X_n$ and $\hat{\sigma}_1^2, \ldots, \hat{\sigma}_j^2$. We recall that from Lemma 9

$$
\widehat{\beta}_j \mid \mathbf{X}_{1:j} = \hat{\beta}_j \mid \mathcal{F}_j \sim \mathcal{N}(\beta_j, \sigma^2 (X_{1:j}^T X_{1:j})^{-1}). \tag{31}
$$

Intuitively, this results says that, given the data $\mathbf{X}_{1:n}$, if we are additionally given all the estimates for the variance $\{\hat{\sigma}_s^2\}_{s=2}^{j}$ —which obviously depend on $\epsilon_1, \ldots, \epsilon_j$—, then the *updated* distribution for $\hat{\beta}_j$ does not change at all. This is a crucial property since TRACE-UCB ignores the current context X_t and it makes decisions only based on previous contexts and the variance estimates $\{\hat{\sigma}_s^2\}_{s=2}^{j}$, thus allowing us to proceed and do inference on $\hat{\beta}_j$ as in the fixed allocation case.

We now need to take into consideration the filtration $\mathbf{F}_{i,j}$ for a specific instance i and the *environment* filtration $\mathcal{E}_{-i}$ containing all the contexts X and noise ϵ from all other instances (different from i). Since the environment filtration $\mathcal{E}_{-i}$ is independent from the samples from instance i, then we can still apply Lemma 9 and obtain

$$\widehat{\beta}_{i,j} \mid \mathcal{F}_{i,j}, \mathcal{E}_{-i} \sim \widehat{\beta}_{i,j} \mid \mathcal{F}_{i,j}. \tag{32}$$

Now we can finally study the expected prediction error

$$
\begin{aligned}
L_{i,n}(\widehat{\beta}_{i,n}) =& \mathbb{E}[(\hat{\beta}_i - \beta_i)(\hat{\beta}_i - \beta_i)^\mathsf{T}] \\
=& \mathbb{E}_{\mathbf{X}_{1:n}, \varepsilon_{-i}} \left[\mathbb{E}[(\hat{\beta}_i - \beta_i)(\hat{\beta}_i - \beta_i)^\mathsf{T} \mid \mathbf{X}_{1:n}, \varepsilon_{-i}] \right] \\
=& \mathbb{E}_{\mathbf{X}_{1:n}, \varepsilon_{-i}} \left[\sum_{j=1}^{n} \mathbb{E}[(\hat{\beta}_{k_i} - \beta_i)(\hat{\beta}_{k_i} - \beta_i)^\mathsf{T} \mid \mathbf{X}_{1:n}, \varepsilon_{-i}, k_i = j] \, \mathbf{P}(k_i = j) \right] \\
=& \mathbb{E}_{\mathbf{X}_{1:n}, \varepsilon_{-i}} \left[\sum_{j=1}^{n} \mathbb{E} \left[\mathbb{E}_{\mathcal{F}_j}[(\hat{\beta}_j - \beta_i)(\hat{\beta}_j - \beta_i)^\mathsf{T} \mid \mathcal{F}_j, \mathbf{X}_{1:n}, \varepsilon_{-i}, k_i = j] \mid \mathbf{X}_{1:n}, \varepsilon_{-i}, k_i = j \right] \, \mathbf{P}(k_i = j) \right] \\
=& \mathbb{E}_{\mathbf{X}_{1:n}, \varepsilon_{-i}} \left[\sum_{j=1}^{n} \mathbb{E} \left[\mathbb{E}_{\mathcal{F}_j}[(\hat{\beta}_j - \beta_i)(\hat{\beta}_j - \beta_i)^\mathsf{T} \mid \mathcal{F}_j, \mathbf{X}_{1:n}] \mid \mathbf{X}_{1:n}, \varepsilon_{-i}, k_i = j \right] \, \mathbf{P}(k_i = j) \right] \tag{33} \\
=& \mathbb{E}_{\mathbf{X}_{1:n}, \varepsilon_{-i}} \left[\sum_{j=1}^{n} \mathbb{E} \left[\sigma_i^2 (\mathbf{X}_{1:j}^\mathsf{T} \mathbf{X}_{1:j})^{-1} \mid \mathbf{X}_{1:n}, k_i = j \right] \, \mathbf{P}(k_i = j) \right] \\
=& \mathbb{E}_{\mathbf{X}_{1:n}, \varepsilon_{-i}} \left[\sum_{j=1}^{n} \sigma_i^2 (\mathbf{X}_{1:j}^\mathsf{T} \mathbf{X}_{1:j})^{-1} \, \mathbf{P}(k_i = j) \right] \\
=& \sigma_i^2 \, \mathbb{E}_{\mathbf{X}_{1:n}, \varepsilon_{-i}} \left[\mathbb{E}_{k_i}[(\mathbf{X}_{1:k_i}^\mathsf{T} \mathbf{X}_{1:k_i})^{-1}] \right] \\
=& \sigma_i^2 \, \mathbb{E} \left[(\mathbf{X}_{1:k_i}^\mathsf{T} \mathbf{X}_{1:k_i})^{-1} \right],
\end{aligned}
$$

where in Eq. 33 we applied Lemma 9. Hence, going back to the definition of loss (see e.g., Eq. 22), we obtain an expression for the loss which applies under TRACE-UCB (while not in general for other algorithms)

$$
\begin{aligned}
L_n(\mathcal{A}) &= \max_i \, \mathbb{E} \left[\sigma_i^2 \, \mathrm{Tr}(\Sigma (\mathbf{X}_{i,n}^\mathsf{T} \mathbf{X}_{i,n})^{-1}) \right] \\
&= \max_i \, \mathbb{E} \left[\frac{\sigma_i^2}{k_{i,n}} \mathrm{Tr} \left(\Sigma \hat{\Sigma}_{i,n}^{-1} \right) \right].
\end{aligned}
$$

$\square$

C. Concentration Inequalities (Proofs of Propositions 3 and 4)

In the next two subsections, we prove Propositions 3 and 4, respectively. In addition, we also show a confidence ellipsoid result for the $\widehat{\beta}$ estimates, and a concentration inequality for the norm of the observations X_t.

C.1. Concentration Inequality for the Variance (Proof of Proposition 3)

We use the following concentration inequality for sub-exponential random variables.

Proposition 10. *Let X be a mean-zero (τ^2, b)-subexponential random variable. Then, for all $\eta > 0$,*

$$\mathbb{P}(|X| \geq \eta) \leq \exp\left(-\frac{1}{2}\min\left\{\frac{\eta^2}{\tau^2}, \frac{\eta}{b}\right\}\right). \tag{34}$$

Proof. See Proposition 2.2 in (Wainwright, 2015). □

We first prove the concentration inequality for one single instance.

Proposition 11. *Let $t > d$, $\mathbf{X}_t \in \mathbb{R}^{t \times d}$ be a random matrix whose entries are independent standard normal random variables, $\mathbf{Y}_t = \mathbf{X}_t^{\mathsf{T}}\beta + \boldsymbol{\epsilon}_t$, where the noise $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \sigma_t^2\, I_d)$ is independent from $\mathbf{X}_t$, and $\delta \in (0, 3/4]$. Then, with probability at least $1 - \delta$, we have*

$$|\hat{\sigma}_t^2 - \sigma^2| \leq \sigma^2\sqrt{\frac{64}{t-d}\left(\log\frac{1}{\delta}\right)^2}, \tag{35}$$

where $\hat{\sigma}_t^2$ is the unbiased estimate $\widehat{\sigma}_t^2 = \frac{1}{t-d}\|\mathbf{Y}_t - \mathbf{X}_t\widehat{\beta}_t\|^2$ and $\widehat{\beta}_t$ is the OLS estimator of β, given $\mathbf{X}_t$ and $\mathbf{Y}_t$.

Proof. First note that the distribution of $\hat{\sigma}_t^2$ conditioned on $\mathbf{X}_t$ follows the scaled chi-squared distribution, i.e.,

$$\hat{\sigma}_t^2 \mid \mathbf{X} \sim \frac{\sigma^2}{t-d}\chi_{t-d}^2.$$

Also note that the distribution of the estimate does *not* depend on $\mathbf{X}_t$ and we can integrate out the randomness in $\mathbf{X}_t$. In order to show concentration around the mean, we directly use the sub-exponential properties of $\hat{\sigma}_t^2$. The χ_k^2 distribution is sub-exponential with parameters $(4k, 4)$.[4] Furthermore, we know that for any constant $C > 0$, $C\chi_k^2$ is $(4C^2k, 4C)$-sub-exponential. As a result, we have that $\hat{\sigma}_t^2$ is subexponential with parameters

$$(\tau^2, b) = \left(\frac{4\sigma^4}{t-d}, \frac{4\sigma^2}{t-d}\right).$$

Now we use Proposition 10 as our concentration bound. In our case, $\eta^2/\tau^2 < \eta/b$, when $\eta < \sigma^2$. In such a case, if we denote the RHS of (34) by δ, we conclude that

$$\eta = \sigma^2\sqrt{\frac{8}{t-d}\log\frac{1}{\delta}}.$$

Then, $\eta < \sigma^2$ holds when $t \geq d + 8\log(1/\delta)$. Otherwise, if $\eta^2/\tau^2 > \eta/b$, by Eq. 34, we have

$$\eta = \frac{8\sigma^2}{t-d}\log\frac{1}{\delta}.$$

In this case, when $t < d + 8\log(1/\delta)$, we have that

$$|\hat{\sigma}_t^2 - \sigma^2| \leq \sigma^2\frac{8}{t-d}\log\frac{1}{\delta}.$$

---

[4]See Example 2.5 in (Wainwright, 2015).

We would like to derive a bound that is valid in both cases. Let $x = 8\log(1/\delta)/(t-d)$, then we have

$$\mathbb{P}\left(|\hat{\sigma}_t^2 - \sigma^2| \geq \sigma^2 \max(x, \sqrt{x})\right) \leq \delta. \tag{36}$$

Suppose $x \geq \sqrt{x}$, so $t < d + \log(1/\delta)$. Then, we would like to find C, such that $x \leq C\sqrt{x}$. As $t \geq d+1$, we see that

$$\sqrt{x} = \sqrt{\frac{8\log(1/\delta)}{t-d}} \leq \sqrt{8\log(1/\delta)} \triangleq C.$$

if $C > 1$, it does follow that $\max(x, \sqrt{x}) < \max(C\sqrt{x}, \sqrt{x}) < \sqrt{8\log(1/\delta)x}$, which corresponds to $\delta < 0.88$. By (36), we now conclude that

$$\mathbb{P}\left(|\hat{\sigma}_t^2 - \sigma^2| \geq \sigma^2 \sqrt{\frac{64}{t-d}\left(\log\frac{1}{\delta}\right)^2}\right) \leq \delta,$$

and the proof is complete. $\qquad\square$

In order to prove Proposition 3, we are just left to apply a union bound over steps $t \in \{1, \ldots, n\}$ and instances $i \in \{1, \ldots, m\}$. In order to avoid confusion, let $\hat{\sigma}_{i,t}$ be the estimate obtained by the algorithm after t steps and $\hat{\sigma}_i(j)$ the estimate obtained using j samples. Let $j > d$, then

$$\mathcal{E}_i(j) = \left\{|\hat{\sigma}_i^2(j) - \sigma_i^2| \geq \sigma_i^2 \sqrt{\frac{64}{j-d}\left(\log\frac{1}{\delta}\right)^2}\right\},$$

is the high-probability event introduced in Proposition 11, which holds with probability $1 - \delta$. Then we have that the event

$$\mathcal{E} = \bigcap_{i=1}^m \bigcap_{j=1}^n \mathcal{E}_i(j),$$

holds with probability $1 - \delta'$, with $\delta' = mn\delta$. We complete the proof of Proposition 3 by properly tuning δ and taking $R \geq \max_i \sigma_i^2$. Recall that Proposition 3 is as follows.

Proposition. *Let the number of pulls $k_{i,t} \geq d+1$ and $R \geq \max_i \sigma_i^2$. If $\delta \in (0, 3/4)$, then for any instance i and step $t > m(d+1)$, with probability at least $1 - \frac{\delta}{2}$, we have*

$$|\hat{\sigma}_{i,t}^2 - \sigma_i^2| \leq \Delta_{i,t} \triangleq R\sqrt{\frac{64}{k_{i,t}-d}\left(\log\frac{2mn}{\delta}\right)^2}. \tag{37}$$

C.2. Concentration Inequality for the Trace (Proof of Proposition 4)

We first recall some basic definitions. For any matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$, the i-th singular value $s_i(\mathbf{A})$ is equivalent to $s_i(\mathbf{A})^2 = \lambda_i(\mathbf{A}^\mathsf{T}\mathbf{A})$, where λ_i is the i-th eigenvalue. The smallest and largest singular values $s_{\min}$ and $s_{\max}$ satisfy

$$s_{\min}\|x\|_2 \leq \|\mathbf{A}x\|_2 \leq s_{\max}\|x\|_2 \qquad \text{for all } x \in \mathbb{R}^d.$$

The extreme singular values measure the maximum and minimum distortion of points and their distance when going from $\mathbb{R}^d$ to $\mathbb{R}^n$ via the linear operator $\mathbf{A}$. We also recall that the spectral norm of $\mathbf{A}$ is given by

$$\|\mathbf{A}\| = \sup_{x \in \mathbb{R}^d \setminus 0} \frac{\|\mathbf{A}x\|_2}{\|x\|_2} = \sup_{x \in S^{n-1}} \|\mathbf{A}x\|_2,$$

and thus, $s_{\max}(\mathbf{A}) = \|\mathbf{A}\|$ and $s_{\min}(\mathbf{A}) = 1/\|\mathbf{A}^{-1}\|$, if $\mathbf{A}$ is invertible.

We report the following concentration inequality for the eigenvalues of random Gaussian matrices.

Proposition 12. *Let $n \geq d$, $\overline{\mathbf{X}} \in \mathbb{R}^{n \times d}$ be a random matrix whose entries are independent standard normal random variables, and $\overline{\Sigma} = \overline{\mathbf{X}}^\mathsf{T} \overline{\mathbf{X}} / n$ be the corresponding empirical covariance matrix. Let $\alpha > 0$, then with probability at least $1 - 2\exp(-\alpha^2 d/2)$, we have*

$$\mathrm{Tr}\left(\overline{\Sigma}^{-1}\right) \geq d\left(1 - \frac{2(1+\alpha)\sqrt{d} + (1+\alpha)^2 d/\sqrt{n}}{\sqrt{n} + 2(1+\alpha)\sqrt{d} + (1+\alpha)^2 d/\sqrt{n}}\right),$$

and

$$\mathrm{Tr}\left(\overline{\Sigma}^{-1}\right) \leq d\left(1 + \frac{2(1+\alpha)\sqrt{d} - (1+\alpha)^2 d/\sqrt{n}}{\sqrt{n} - 2(1+\alpha)\sqrt{d} + (1+\alpha)^2 d/\sqrt{n}}\right).$$

In particular, we have

$$d\left(1 - (1+\alpha)\sqrt{\frac{d}{n}}\right)^2 \leq \mathrm{Tr}\left(\overline{\Sigma}^{-1}\right) \leq d\left(1 + 2(1+\alpha)\sqrt{\frac{d}{n}}\right)^2.$$

Proof. We first derive the concentration inequality for the eigenvalues of the empirical covariance matrix and then we invert it to obtain the guarantee for the inverse matrix. From Corollary 5.35 in (Vershynin, 2010), we have that for any $t > 0$

$$\left(\sqrt{n} - \sqrt{d} - t\right)^2 \leq \lambda_{\min}(\overline{\mathbf{X}}^\mathsf{T} \overline{\mathbf{X}}) = s_{\min}(\overline{\mathbf{X}})^2 \leq s_{\max}(\overline{\mathbf{X}})^2 = \lambda_{\max}(\overline{\mathbf{X}}^\mathsf{T} \overline{\mathbf{X}}) \leq \left(\sqrt{n} + \sqrt{d} + t\right)^2, \tag{38}$$

with probability at least $1 - 2\exp(-t^2/2)$. Let $\alpha > 0$ and take $t = \alpha\sqrt{d}$, then with probability at least $1 - 2\exp(-\alpha^2 d/2)$, we obtain the desired statement

$$\left(1 - (1+\alpha)\sqrt{\frac{d}{n}}\right)^2 \leq \lambda_{\min}\left(\overline{\Sigma}\right) \leq \lambda_{\max}\left(\overline{\Sigma}\right) \leq \left(1 + (1+\alpha)\sqrt{\frac{d}{n}}\right)^2.$$

We now proceed by studying the eigenvalues of the inverse of the empirical covariance matrix $\lambda_{\min}(\overline{\Sigma}^{-1}) = 1/\lambda_{\max}(\overline{\Sigma})$ and $\lambda_{\max}(\overline{\Sigma}^{-1}) = 1/\lambda_{\min}(\overline{\Sigma})$. Combined with Eq. 38 we have

$$\lambda_{\min}\left(\overline{\Sigma}^{-1}\right) \geq \frac{1}{\left(1 + (1+\alpha)\sqrt{\frac{d}{n}}\right)^2}$$

$$= \frac{1}{1 + 2(1+\alpha)\sqrt{\frac{d}{n}} + (1+\alpha)^2 \frac{d}{n}}$$

$$= 1 - \frac{2(1+\alpha)\sqrt{\frac{d}{n}} + (1+\alpha)^2 \frac{d}{n}}{1 + 2(1+\alpha)\sqrt{\frac{d}{n}} + (1+\alpha)^2 \frac{d}{n}}.$$

Similarly, we have that

$$\lambda_{\max}\left(\overline{\Sigma}^{-1}\right) \leq \frac{1}{\left(1 - (1+\alpha)\sqrt{\frac{d}{n}}\right)^2}$$

$$= \frac{1}{1 - 2(1+\alpha)\sqrt{\frac{d}{n}} + (1+\alpha)^2 \frac{d}{n}}$$

$$= 1 + \frac{2(1+\alpha)\sqrt{\frac{d}{n}} - (1+\alpha)^2 \frac{d}{n}}{1 - 2(1+\alpha)\sqrt{\frac{d}{n}} + (1+\alpha)^2 \frac{d}{n}}.$$

Using the fact that for any matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$, we may write $d\,\lambda_{\min}(\mathbf{A}) \leq \mathrm{Tr}(\mathbf{A}) \leq d\,\lambda_{\max}(\mathbf{A})$, we obtain the final statement on the trace of $\overline{\Sigma}^{-1}$. The first of the two bounds can be further simplified by using $1/(1+x) \geq 1 - x$ for any

$x \geq 0$, thus obtaining

$$\lambda_{\min}\left(\overline{\Sigma}^{-1}\right) \geq \left(1 - (1+\alpha)\sqrt{\frac{d}{n}}\right)^2.$$

While under the assumption that $n \geq 4(1+\alpha)^2 d$ we can use $1/(1-x) \leq 1 + 2x$ (for any $x \geq 1/2$) and obtain

$$\lambda_{\max}\left(\overline{\Sigma}^{-1}\right) \leq \left(1 + 2(1+\alpha)\sqrt{\frac{d}{n}}\right)^2.$$

$\square$

The statement of Proposition 4 (below) is obtained by recalling that $\Sigma\widehat{\Sigma}_{i,n}^{-1}$ is the empirical covariance matrix of the whitened sample matrix $\overline{\mathbf{X}}_{i,n}$ and by a union bound over the number of samples $k_{i,n}$ and the number of instances i.

Proposition. *Force the number of samples $k_{i,t} \geq d + 1$. If $\delta \in (0,1)$, for any $i \in [m]$ and step $t > m(d+1)$ with probability at least $1 - \delta/2$, we have*

$$\left(1 - C_{\mathrm{Tr}}\sqrt{\frac{d}{n}}\right)^2 \leq \frac{\mathrm{Tr}\left(\Sigma\widehat{\Sigma}_{i,t}^{-1}\right)}{d} \leq \left(1 + 2C_{\mathrm{Tr}}\sqrt{\frac{d}{n}}\right)^2,$$

with $C_{\mathrm{Tr}} = 1 + \sqrt{2\log(4nm/\delta)/d}$.

C.3. Concentration Inequality for $\widehat{\beta}$ Estimates

We slightly modify Theorem 2 from (Abbasi-Yadkori et al., 2011) to obtain a confidence ellipsoid over the $\widehat{\beta}_i$'s.

Theorem 13. *Let $\{F_t\}_{t=0}^{\infty}$ be a filtration. Let $\{\eta_t\}_{t=1}^{\infty}$ be a real-valued stochastic process such that η_t is F_t measurable and η_t is conditionally R-subgaussian for some $R \geq 0$, i.e.*

$$\forall \lambda \in \mathbb{R} \qquad \mathbb{E}[e^{\lambda \eta_t} \mid F_{t-1}] \leq \exp\left(\frac{\lambda^2 R^2}{2}\right). \tag{39}$$

Let $\{X_t\}_{t=1}^{\infty}$ be an $\mathbb{R}^d$-valued stochastic process such that X_t is F_{t-1} measurable. Assume that V is a $d \times d$ positive definite matrix. For any $t \geq 0$, define

$$\bar{V}_t = V + \sum_{s=1}^{t} X_s X_s^T, \qquad\qquad S_t = \sum_{s=1}^{t} \eta_s X_s. \tag{40}$$

Let $V = \lambda \mathrm{Id}$, $\lambda > 0$, and define $Y_t = X_t^T \beta^* + \eta_t$. Assume that $\|\beta^*\|_2 \leq S$. Also, let $\hat{\beta}_t = \bar{V}_t^{-1} \mathbf{X}_t^T \mathbf{Y}_t$ be the ridge estimate for β after t observations $\mathbf{X}_t, \mathbf{Y}_t$. Then, for any $\delta > 0$, with probability at least $1 - \delta$, for all $t \geq 0$, β^* lies in

$$C_t = \left\{\beta \in \mathbb{R}^d : \|\hat{\beta}_t - \beta\|_{\bar{V}_t/t} \leq \frac{R}{\sqrt{t}}\sqrt{2\log\left(\frac{\det\left(\bar{V}_t\right)^{1/2}\det\left(\lambda I\right)^{-1/2}}{\delta}\right)} + \sqrt{\frac{\lambda}{t}}\,S\right\}. \tag{41}$$

Proof. Take $x = \frac{\bar{V}_t}{t}(\hat{\beta}_t - \beta^*)$ in equation 5 in the proof of Theorem 2 in (Abbasi-Yadkori et al., 2011). $\square$

We use the previous theorem by lower bounding the $\bar{V}_t/t$ norm in Σ norm.

C.4. Bounded Norm Lemma

Lemma 14. *Let $X_1, \dots, X_t \in \mathbb{R}^d$ be iid subgaussian random variables.*

If $\|X_1\|^2$ is subexponential with parameters (a^2, b), then, for $\alpha > 0$

$$\mathbf{P}\left(\frac{1}{t}\sum_{j=1}^{t}\|X_j\|^2 \leq \mathbb{E}[\|X_1\|^2] + \frac{\alpha}{t}\right) \geq \begin{cases} 1 - \exp\left(-\frac{\alpha^2}{2ta^2}\right) & \text{if } 0 \leq \alpha \leq ta^2/b, \\ 1 - \exp\left(-\frac{\alpha}{2b}\right) & \text{if } \alpha > ta^2/b. \end{cases} \tag{42}$$

Proof. The proof directly follows by Proposition 10, by defining zero-mean subexponential random variable Z with parameters $(a^2/t, b/t)$

$$Z = \frac{1}{t} \sum_{j=1}^{t} \|X_j\|^2 - \mathbb{E} \left[\frac{1}{t} \sum_{j=1}^{t} \|X_j\|^2 \right]. \tag{43}$$

$\square$

Corollary 15. *Let $X_1, \dots, X_t \in \mathbb{R}^d$ be iid gaussian variables, $X \sim \mathcal{N}(0, \mathrm{Id})$. Assume $t \geq d + 1$. Let $\delta > 0$. Then, with probability at least $1 - \delta$,*

$$\frac{1}{t} \sum_{j=1}^{t} \|X_j\|^2 \leq d + 8 \log \left(\frac{1}{\delta} \right) \sqrt{\frac{d}{t}}, \tag{44}$$

Proof. For standard Gaussian $X \sim \mathcal{N}(0, \mathrm{Id})$, $\|X\|^2 \sim \chi_d^2$, and $a^2 = 4d$ and $b = 4$. Note that $\mathbb{E}[\|X_j\|^2] = d$. By the proof of Lemma 14 and (44)

$$\mathbf{P} \left(|Z| \geq a \sqrt{\frac{2}{t} \log \left(\frac{1}{\delta} \right)} \right) \leq \delta, \qquad \text{when } t \geq 2 \left(\frac{b}{a} \right)^2 \log \left(\frac{1}{\delta} \right). \tag{45}$$

$$\mathbf{P} \left(|Z| \geq \frac{2b}{t} \log \left(\frac{1}{\delta} \right) \right) \leq \delta, \qquad \text{when } t < 2 \left(\frac{b}{a} \right)^2 \log \left(\frac{1}{\delta} \right). \tag{46}$$

Substituting $a = 2\sqrt{d}$ and $b = 4$ leads to

$$\mathbf{P} \left(|Z| \geq \sqrt{\frac{8d}{t} \log \left(\frac{1}{\delta} \right)} \right) \leq \delta, \qquad \text{when } t \geq \frac{8}{d} \log \left(\frac{1}{\delta} \right). \tag{47}$$

$$\mathbf{P} \left(|Z| \geq \frac{8}{t} \log \left(\frac{1}{\delta} \right) \right) \leq \delta, \qquad \text{when } t < \frac{8}{d} \log \left(\frac{1}{\delta} \right). \tag{48}$$

We would like to upper bound $8 \log (1/\delta) / t$ in (48). As $t > d$, we see

$$\frac{8}{t} \log \left(\frac{1}{\delta} \right) \leq \frac{8}{\sqrt{dt}} \log \left(\frac{1}{\delta} \right). \tag{49}$$

As a consequence,

$$\mathbf{P} \left(|Z| \geq \frac{8}{\sqrt{dt}} \log \left(\frac{1}{\delta} \right) \right) \leq \delta, \qquad \text{when } t < \frac{8}{d} \log \left(\frac{1}{\delta} \right). \tag{50}$$

It follows that for all $t > d$

$$\mathbf{P} \left(|Z| \geq \max \left(\frac{8}{\sqrt{dt}} \log \left(\frac{1}{\delta} \right), \sqrt{\frac{8d}{t} \log \left(\frac{1}{\delta} \right)} \right) \right) \leq \delta. \tag{51}$$

As $\delta < 1$, we finally conclude that

$$\mathbf{P} \left(|Z| \geq 8 \sqrt{\frac{d}{t}} \log \left(\frac{1}{\delta} \right) \right) \leq \delta. \tag{52}$$

Therefore, with probability at least $1 - \delta$,

$$\frac{1}{t} \sum_{j=1}^{t} \|X_j\|^2 \leq d + 8 \log \left(\frac{1}{\delta} \right) \sqrt{\frac{d}{t}}, \tag{53}$$

as stated in the corollary. $\square$

D. Performance Guarantees for TRACE-UCB

D.1. Lower Bound on Number of Samples (Proof of Theorem 5)

We derive the high-probability guarantee on the number of times each instance is selected.

Theorem. *Let $\delta > 0$. With probability at least $1 - \delta$, the total number of contexts that TRACE-UCB allocates to each problem instance i after n rounds satisfies*

$$k_{i,n} \geq k_{i,n}^* - \frac{C_\Delta + 8C_{\mathrm{Tr}}}{\sigma_{\min}^2} \sqrt{\frac{nd}{\lambda_{\min}}} - \Omega(n^{1/4}) \tag{54}$$

where $R \geq \sigma_{\max}^2$ is known by the algorithm, and we defined $C_\Delta = 16R \log(2mn/\delta)$, $C_{\mathrm{Tr}} = 1 + \sqrt{2 \log(4nm/\delta)/d}$, and $\lambda_{\min} = \sigma_{\min}^2 / \sum_j \sigma_j^2$.

Proof. We denote by $\mathcal{E}_\delta$ the joint event on which Proposition 3 and Proposition 4 hold at the same time with an overall probability $1 - \delta$. This immediately gives upper and lower confidence bounds on the score $s_{i,t}$ used in TRACE-UCB as

$$\left(1 - C_{\mathrm{Tr}}\sqrt{\frac{d}{k_{i,t}}}\right)^2 \frac{\sigma_i^2}{k_{i,t}} \leq \frac{s_{i,t}}{d} \leq \left(1 + 2C_{\mathrm{Tr}}\sqrt{\frac{d}{k_{i,t}}}\right)^2 \frac{\sigma_i^2 + 2\Delta_{i,t}}{k_{i,t}}.$$

Recalling the definition of $\Delta_{i,t}$ we can rewrite the last term as

$$\frac{\sigma_i^2 + 2\Delta_{i,t}}{k_{i,t}} = \left(1 + \frac{16R \log(2mn/\delta)}{\sigma_i^2 \sqrt{k_{i,t} - d}}\right) \frac{\sigma_i^2}{k_{i,t}} = \left(1 + \frac{C_\Delta}{\sigma_i^2 \sqrt{k_{i,t} - d}}\right) \frac{\sigma_i^2}{k_{i,t}},$$

where $C_\Delta = 16R \log(2mn/\delta)$. We consider a step $t + 1 \leq n$ at which $I_{t+1} = q$. By algorithmic construction we have that $s_{p,t} \leq s_{q,t}$ for every arm $p \in [m]$. Using the inequalities above we obtain

$$\left(1 - C_{\mathrm{Tr}}\sqrt{\frac{d}{k_{p,t}}}\right)^2 \frac{\sigma_p^2}{k_{p,t}} \leq \frac{s_{p,t}}{d} \leq \frac{s_{q,t}}{d} \leq \left(1 + 2C_{\mathrm{Tr}}\sqrt{\frac{d}{k_{q,t}}}\right)^2 \frac{\sigma_q^2 + 2\Delta_{q,t}}{k_{q,t}}$$

If $t + 1$ is the last time step at which arm q is pulled, then $k_{q,t} = k_{q,t+1} - 1 = k_{q,n} - 1$ and $k_{p,n} \geq k_{p,t}$. Then we can rewrite the previous inequality as

$$\left(1 - C_{\mathrm{Tr}}\sqrt{\frac{d}{k_{p,n}}}\right)^2 \frac{\sigma_p^2}{k_{p,n}} =: A_{p,n} \leq B_{q,n} := \left(1 + 2C_{\mathrm{Tr}}\sqrt{\frac{d}{k_{q,n} - 1}}\right)^2 \left(1 + \frac{C_\Delta}{\sigma_q^2 \sqrt{k_{q,n} - d - 1}}\right) \frac{\sigma_q^2}{k_{q,n} - 1}. \tag{55}$$

If every arm is pulled exactly the optimal number of times, then for any $i \in [m]$, $k_{i,n} = k_{i,n}^*$ and the statement of the theorem trivially holds. Otherwise, there exists at least one arm that is pulled more than $k_{i,n}^*$. Let q be this arm, then $k_{q,n} > k_{q,n}^*$. We recall that $L_n^* = d\sigma_q^2/(k_{q,n}^* - d - 1)$ and we rewrite the RHS of Eq. 55 as

$$B_{q,n} \leq \left(1 + 2C_{\mathrm{Tr}}\sqrt{\frac{d}{k_{q,n}^* - d - 1}}\right)^2 \left(1 + \frac{C_\Delta}{\sigma_q^2 \sqrt{k_{q,n}^* - d - 1}}\right) \frac{\sigma_q^2}{k_{q,n}^* - d - 1}$$

$$\leq \left(1 + 2C_{\mathrm{Tr}}\sqrt{\frac{L_n^*}{\sigma_q^2}}\right)^2 \left(1 + C_\Delta \sqrt{\frac{L_n^*}{d\sigma_q^6}}\right) \frac{L_n^*}{d}.$$

We also simplify the LHS of Eq. 55 as

$$A_{p,n} = \left(1 - 2C_{\mathrm{Tr}}\sqrt{\frac{d}{k_{p,n}}} + C_{\mathrm{Tr}}^2 \frac{d}{k_{p,n}}\right) \frac{\sigma_p^2}{k_{p,n}} \geq \left(1 - 2C_{\mathrm{Tr}}\sqrt{\frac{d}{k_{p,n}}}\right) \frac{\sigma_p^2}{k_{p,n}}.$$

At this point we can solve Eq. 55 for $k_{p,n}$ and obtain a lower bound on it. We study the inequality $1/A_{p,n} \geq 1/B_{p,n}$.

We first notice that

$$\frac{1}{A_{p,n}} \leq \frac{k_{p,n}}{\sigma_p^2}\left(1 + 4C_{\mathrm{Tr}}\sqrt{\frac{d}{k_{p,n}}}\right) \leq \frac{1}{\sigma_p^2}\left(\sqrt{k_{p,n}} + 2C_{\mathrm{Tr}}\sqrt{d}\right)^2,$$

where we used $1/(1-x) \leq 1 + 2x$ for $x \leq 1/2$ and we added a suitable positive term to obtain the final quadratic form. Similarly we have

$$\frac{1}{B_{q,n}} \geq \left(1 - 2C_{\mathrm{Tr}}\sqrt{\frac{L_n^*}{\sigma_q^2}}\right)^2\left(1 - C_\Delta\sqrt{\frac{L_n^*}{d\sigma_q^6}}\right)\frac{d}{L_n^*} = \left(1 - 2C_{\mathrm{Tr}}\sqrt{\frac{L_n^*}{\sigma_q^2}}\right)^2\left(\frac{d}{L_n^*} - C_\Delta\sqrt{\frac{d}{L_n^*\sigma_q^6}}\right),$$

where we used $1/(1+x) \geq 1 - x$ for any $x \geq 0$. In order to ease the derivation of an explicit lower-bound on $k_{p,n}$, we further simplify the previous expression by replacing higher order terms with a big-Ω notation. We first recall that $L_n^* = \widetilde{\Theta}(md\bar{\sigma}^2/n)$, then the terms of order $(1/L_n^*)$ and $(1/\sqrt{L_n^*})$ clearly dominate the expression, while all other terms are asymptotically constant or decreasing in n and thus we can rewrite the previous bound as

$$\frac{1}{B_{q,n}} \geq \frac{d}{L_n^*} - (C_\Delta + 4C_{\mathrm{Tr}}\sqrt{d})\sqrt{\frac{d}{L_n^*\sigma_q^6}} - \Omega(1).$$

By setting $C = C_\Delta + 4C_{\mathrm{Tr}}\sqrt{d}$ we can finally use the upper bound on $1/A_{p,n}$ and the lower bound on $1/B_{q,n}$ to obtain

$$\frac{1}{\sigma_p^2}\left(\sqrt{k_{p,n}} + 2C_{\mathrm{Tr}}\sqrt{d}\right)^2 \geq \frac{d}{L_n^*} - C\sqrt{\frac{d}{L_n^*\sigma_q^6}} - \Omega(1).$$

We proceed with solving the previous inequality for $k_{p,n}$ and obtain

$$k_{p,n} \geq \sigma_p^2\left(\left(\frac{d}{L_n^*} - C\sqrt{\frac{d}{L_n^*\sigma_q^6}} - \Omega(1)\right)^{1/2} - 2C_{\mathrm{Tr}}\sqrt{d}\right)^2.$$

Taking the square on RHS and adding and subtracting $d + 1$ we have

$$k_{p,n} \geq d + 1 + \sigma_p^2\left(\frac{d}{L_n^*} - C\sqrt{\frac{d}{L_n^*\sigma_q^6}} - 4C_{\mathrm{Tr}}\sqrt{d}\left(\frac{d}{L_n^*} - C\sqrt{\frac{d}{L_n^*\sigma_q^6}} - \Omega(1)\right)^{1/2} + 4C_{\mathrm{Tr}}^2 d\right) - d - 1 - \Omega(1).$$

We clearly notice that the first three terms in the RHS are dominant (they are higher order function of n through L_n^*) and thus we can isolate them and replace all other terms by their asymptotic lower bound as

$$k_{p,n} \geq d + 1 + \frac{d\sigma_p^2}{L_n^*} - \sqrt{\frac{1}{L_n^*}}\left(C\sqrt{\frac{d\sigma_p^4}{\sigma_q^6}} + 4C_{\mathrm{Tr}}d\right) - \Omega(n^{1/4}),$$

where we used the fact that $L_n^* = \widetilde{\Theta}(md\bar{\sigma}^2/n)$ to bound the higher order terms. Furthermore, we recall that $k_{p,n}^* = d\sigma_p^2/L_n^* + d + 1$ and thus we can finally write the previous bound as

$$k_{p,n} \geq k_{p,n}^* - \sqrt{\frac{1}{L_n^*}}\left(C\sqrt{\frac{d\sigma_p^4}{\sigma_q^6}} + 4C_{\mathrm{Tr}}d\right) - \Omega(n^{1/4}).$$

The final bound is obtained by using $\sigma_p^2/\sum_j \sigma_j^2 = \lambda_p \geq \lambda_{\min}$ and $\sigma_q^2 \geq \sigma_{\min}^2$ with the final expression

$$k_{p,n} \geq k_{p,n}^* - \sqrt{n}\left(\frac{C}{\sigma_{\min}^2}\sqrt{\frac{1}{\lambda_{\min}}} + 4C_{\mathrm{Tr}}\sqrt{d}\right) - \Omega(n^{1/4}).$$

A quite loose bound based on the definition of C for the previous expression gives the final more readable result

$$k_{p,n} \geq k_{p,n}^* - \frac{C_\Delta + 8C_{\mathrm{Tr}}}{\sigma_{\min}^2}\sqrt{\frac{nd}{\lambda_{\min}}} - \Omega(n^{1/4}).$$

$\square$

D.2. Regret Bound (Proof of Theorem 6)

Theorem. *The regret of the Trace-UCB algorithm, i.e., the difference between its loss and the loss of optimal static allocation (see Eq. 8), is upper-bounded by*

$$L_n(\mathcal{A}) - L_n^* \leq O\left(\frac{1}{\sigma_{\min}^2}\left(\frac{d}{\lambda_{\min}n}\right)^{3/2}\right), \tag{56}$$

where $\lambda_{\min} = \sigma_{\min}^2 / \sum_j \sigma_j^2$.

Proof. We first simplify the expression of the loss for TRACE-UCB in Lemma 19. We invert trace operator and expectation and have

$$L_{i,n}(\widehat{\beta}_i^\lambda) = \mathbb{E}\left(\text{Tr}\left[\Sigma \mathbf{W}_{i,n}\left(\sigma_i^2 \mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n} + \lambda^2 \beta_i \beta_i^\mathsf{T}\right)\mathbf{W}_{i,n}^\mathsf{T}\right]\right).$$

We notice that $\mathbf{W}_{i,n} = (\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n} + \lambda I)^{-1} \preceq (\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n})^{-1}$, where $\preceq$ is the Lower ordering between positive-definite matrices. We focus on the two additive terms in the trace separately. We have

$$
\begin{aligned}
\text{Tr}\left(\Sigma \mathbf{W}_{i,n}\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n}\mathbf{W}_{i,n}^\mathsf{T}\right) &= \text{Tr}\left(\mathbf{W}_{i,n}\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n}\mathbf{W}_{i,n}^\mathsf{T}\Sigma\right) \\
&\leq \text{Tr}\left((\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n})^{-1}\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n}\mathbf{W}_{i,n}^\mathsf{T}\Sigma\right) = \text{Tr}\left(\Sigma \mathbf{W}_{i,n}^\mathsf{T}\right) \\
&\leq \text{Tr}\left(\Sigma(\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n})^{-1}\right) = \frac{1}{k_{i,n}}\text{Tr}\left(\Sigma\widehat{\Sigma}_{i,n}^{-1}\right),
\end{aligned} \tag{57}
$$

where we used the fact that $\text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA})$, $\text{Tr}(\mathbf{AB}) \leq \text{Tr}(\mathbf{CB})$ if $\mathbf{A} \preceq \mathbf{C}$ and the definition of $\widehat{\Sigma}_{i,n}$.

Similarly, we have

$$
\begin{aligned}
\text{Tr}\left(\Sigma \mathbf{W}_{i,n}\beta_i \beta_i^\mathsf{T}\mathbf{W}_{i,n}^\mathsf{T}\right) &= \|\beta_i\|^2 \text{Tr}\left(\Sigma \mathbf{W}_{i,n}\mathbf{W}_{i,n}^\mathsf{T}\right) \\
&\leq \|\beta_i\|^2 \text{Tr}\left((\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n})^{-1}\Sigma \mathbf{W}_{i,n}\right) \leq \|\beta_i\|^2 \frac{\lambda_{\max}(\widehat{\Sigma}_{i,n}^{-1})}{k_{i,n}}\text{Tr}\left(\Sigma \mathbf{W}_{i,n}\right) \\
&\leq \|\beta_i\|^2 \frac{\lambda_{\max}(\widehat{\Sigma}_{i,n}^{-1})}{k_{i,n}}\text{Tr}\left(\Sigma(\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n})^{-1}\right) = \|\beta_i\|^2 \frac{\lambda_{\max}(\widehat{\Sigma}_{i,n}^{-1})}{k_{i,n}^2}\text{Tr}\left(\Sigma\widehat{\Sigma}_{i,n}^{-1}\right).
\end{aligned}
$$

Going back to the loss expression we have

$$L_{i,n}(\widehat{\beta}_i^\lambda) \leq \mathbb{E}\left[\frac{\text{Tr}\left(\Sigma\widehat{\Sigma}_{i,n}^{-1}\right)}{k_{i,n}}\left(\sigma_i^2 + \|\beta_i\|^2\frac{\lambda_{\max}(\widehat{\Sigma}_{i,n}^{-1})}{k_{i,n}}\right)\right].$$

We decompose the loss in two terms depending on the high-probability event $\mathcal{E}_\delta$ under which the concentration inequalities Proposition 3 and Proposition 4 hold at the same time

$$L_{i,n}(\widehat{\beta}_i^\lambda) \leq \mathbb{E}\left[\frac{\text{Tr}\left(\Sigma\widehat{\Sigma}_{i,n}^{-1}\right)}{k_{i,n}}\left(\sigma_i^2 + \|\beta_i\|^2\frac{\lambda_{\max}(\widehat{\Sigma}_{i,n}^{-1})}{k_{i,n}}\right)\Big|\mathcal{E}_\delta\right] + \delta\mathbb{E}\left(\text{Tr}\left[\Sigma \mathbf{W}_{i,n}\left(\sigma_i^2 \mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n} + \lambda^2 \beta_i \beta_i^\mathsf{T}\right)\mathbf{W}_{i,n}^\mathsf{T}\right]\Big|\mathcal{E}_\delta^\mathsf{c}\right),$$

where we used $\mathbb{P}(\mathcal{E}_\delta^\mathsf{c} \leq \delta)$. If we denote the second expectation in the previous expression by $L_{i,n}^\mathsf{c}(\widehat{\beta}_i^\lambda)$, then we can use Eq. 57 and obtain

$$L_{i,n}^\mathsf{c}(\widehat{\beta}_i^\lambda) \leq \sigma_i^2 \mathbb{E}\left(\text{Tr}\left(\Sigma \mathbf{W}_{i,n}^\mathsf{T}\right)\big|\mathcal{E}_\delta^\mathsf{c}\right) + \|\beta_i\|\lambda^2 \mathbb{E}\left(\text{Tr}\left(\Sigma \mathbf{W}_{i,n}\mathbf{W}_{i,n}^\mathsf{T}\right)\big|\mathcal{E}_\delta^\mathsf{c}\right)$$

Using the fact that $\text{Tr}(AB) \leq \lambda_{\max}(A)\text{Tr}(B)$, we can upper bound the previous equation as

$$L_{i,n}^\mathsf{c}(\widehat{\beta}_i^\lambda) \leq \sigma_i^2 \text{Tr}(\Sigma)\mathbb{E}\left(\lambda_{\max}(\mathbf{W}_{i,n})\big|\mathcal{E}_\delta^\mathsf{c}\right) + \|\beta_i\|\text{Tr}(\Sigma)\lambda^2 \mathbb{E}\left(\lambda_{\max}(\mathbf{W}_{i,n})^2\big|\mathcal{E}_\delta^\mathsf{c}\right)$$

Recalling that thanks to the regularization $\lambda_{\max}(\mathbf{W}_{i,n}) \leq 1/\lambda$, we finally obtain

$$L_{i,n}^\mathsf{c}(\widehat{\beta}_i^\lambda) \leq \text{Tr}(\Sigma)\left(\frac{\sigma_i^2}{\lambda} + \|\beta_i\|\right). \tag{58}$$

The analysis of the high-probability part of the bound relies on the concentration inequalities for the trace and $\lambda_{\max}$ and the lower bound on the number of samples $k_{i,n}$ from Thm. 5. We recall the three main inequalities we are going to use to bound the loss

$$k_{i,n} \geq k_{i,n}^* - C\sqrt{nd} - \Omega(n^{1/4}),$$

$$\mathrm{Tr}(\Sigma\widehat{\Sigma}_{i,n}^{-1}) \leq d\left(1 + 2(1+\alpha)\sqrt{\frac{d}{n}}\right)^2,$$

$$\lambda_{\max}(\widehat{\Sigma}_{i,n}^{-1}) \leq \frac{1}{\lambda_{\min}(\Sigma)}\left(1 + 2(1+\alpha)\sqrt{\frac{d}{n}}\right)^2,$$

where $C = \frac{C_\Delta + 8C_{\mathrm{Tr}}}{\sigma_{\min}^2\sqrt{\lambda_{\min}}}$ and the last inequality is obtained by multiplying by $\Sigma^{-1}\Sigma$ to whiten $\widehat{\Sigma}_{i,n}$ and using Proposition 12, and $\lambda_{\max}(AB) \leq \lambda_{\max}(A)\lambda_{\max}(B)$ and finally $\lambda_{\max}(\Sigma^{-1}) = 1/\lambda_{\min}(\Sigma)$. We can invert the first inequality as

$$\frac{1}{k_{i,n}} \leq \frac{1}{k_{i,n}^* - C\sqrt{nd} - \Omega(n^{1/4})} \leq \frac{1}{k_{i,n}^*} + O\left(\frac{2C}{k_{i,n}^*}\sqrt{\frac{d}{n}}\right) \leq \frac{1}{k_{i,n}^*} + O\left(\frac{\sqrt{d}}{\sigma_{\min}^2(\lambda_{\min}n)^{3/2}}\right), \tag{59}$$

where the last inequality is obtained by recalling that $k_{i,n}^* = \Theta(\lambda_i n)$ and using the definition of C (where we ignore C_Δ and C_{Tr}). We can then rewrite the high-probability loss as

$$\mathbb{E}\left[\frac{\mathrm{Tr}(\Sigma\widehat{\Sigma}_{i,n}^{-1})}{k_{i,n}}\left(\sigma_i^2 + \|\beta_i\|^2\frac{\lambda_{\max}(\widehat{\Sigma}_{i,n}^{-1})}{k_{i,n}}\right)\bigg|\mathcal{E}_\delta\right] \leq \frac{d\sigma_i^2}{k_{i,n}^*} + O\left(\frac{1}{\sigma_{\min}^2}\left(\frac{d}{\lambda_{\min}n}\right)^{3/2}\right) \leq L_n^* + O\left(\frac{1}{\sigma_{\min}^2}\left(\frac{d}{\lambda_{\min}n}\right)^{3/2}\right).$$

By recalling the regret $R_n = \max_i L_{i,n}(\beta_{i,n}^\lambda) - L_n^*$, bringing the bounds above together and setting $\delta = O(n^{-3/2-\epsilon})$ for any $\epsilon > 0$ and a suitable multiplicative constant, we obtain the final regret bound

$$R_n \leq O\left(\frac{1}{\sigma_{\min}^2}\left(\frac{d}{\lambda_{\min}n}\right)^{3/2}\right).$$

$\square$

D.3. High Probability Bound for Trace-UCB Loss (Proof of Theorem 7)

In this section, we start by defining a new loss function for algorithm $\mathcal{A}$:

$$\widetilde{L}_n(\mathcal{A}) = \max_{i\in[m]}\|\beta_i - \hat{\beta}_{i,n}\|_\Sigma^2. \tag{60}$$

Note that $\widetilde{L}_n(\mathcal{A})$ is a random variable as $\hat{\beta}_{i,n}$ is random, and the expectation is only taken with respect to the test point $X \sim \mathbf{F}$ (leading to the Σ-norm). We expect results of the following flavor: let $\delta > 0$, then with probability at least $1 - \delta$,

$$\widetilde{L}_n(\mathcal{A}) - \widetilde{L}_n^* \leq \tilde{O}\left(\left(\sum_j \sigma_j^2\frac{d}{n}\right)^{3/2}\right), \tag{61}$$

when $\mathcal{A}$ corresponds to TRACE-UCB, and $\widetilde{L}_n^*$ to the optimal static allocation under ordinary least squares.

We start by focusing on $\widetilde{L}_n(\mathcal{A})$, and proving Theorem 7:

Theorem. *Let $\delta > 0$, and assume $\|\beta_i\|_2 \leq Z$ for all i, for some $Z > 0$. With probability at least $1 - \delta$,*

$$\widetilde{L}_n(\mathcal{A}) \leq \frac{\sum_{j=1}^m \sigma_j^2}{n}\left(d + 2\log\frac{3m}{\delta}\right) + O\left(\frac{1}{\sigma_{\min}^2}\left(\frac{d}{n\lambda_{\min}}\right)^{3/2}\right), \tag{62}$$

where $\lambda_{\min} = \sigma_{\min}^2/\sum_j \sigma_j^2$.

Proof. We define a set of events that help us control the loss, and then we show that these events simultaneously hold with high probability. In particular, we need the following events:

1. $\mathcal{E}_G \equiv$ the good event holds (for all arms i, and all times t), which includes a confidence interval for $\hat{\sigma}_{i,t}^2$ and the trace of the empirical covariance matrix.

   Holds with probability $1 - \delta_G$. This event is described and controlled in Proposition 3 and Proposition 4.

2. $\mathcal{E}_{M,i} \equiv$ the confidence intervals $C_{i,t}$ created for arm i at time t contain the true β_i at all times t —based on the vector-valued martingale in (Abbasi-Yadkori et al., 2011).

   Holds with probability $1 - \delta_{M,i}$. This event is described and controlled in Theorem 13.

3. $\mathcal{E}_{C,i,t} \equiv$ the empirical covariance $\hat{\Sigma}_{i,t}$ for arm i at time t is close to Σ. This event is a direct consequence of event $\mathcal{E}_G$.

4. $\mathcal{E}_{B,i,t} \equiv$ the first t observations pulled at arm i have norm reasonably bounded. The empirical average norm is not too far from its mean. Holds with probability $1 - \delta_{B,i,t}$. This event is described and controlled in Corollary 15.

Let H be the set of all the previous events. Then, by the union bound

$$\mathbf{P}\left(\cap_{\epsilon \in H} \epsilon\right) \geq 1 - \sum_{\epsilon \in H} \delta_\epsilon. \tag{63}$$

Our goal is to show that if $\cap_{\epsilon \in H} \epsilon$ holds, then the loss $\widetilde{L}_n(\mathcal{A}) = \max_{i \in [m]} \|\beta_i - \hat{\beta}_{i,n}\|_\Sigma^2$ is upper bounded by a quantity that resembles the expected loss of the algorithm that knows the σ_i^2's in advance.

Fix $\delta > 0$. We want $\delta = \sum_{\epsilon \in H} \delta_\epsilon$, and we would like to assign equal weight to all the sets of events. First, $\delta_G = \delta/3$. Also, $\sum_i \delta_{M,i} = \delta/3$, implying $\delta_{M,i} = \delta/3m$ for every arm $i \in [m]$. Finally, to bound observation norms, we set $\sum_i \sum_t \delta_{B,i,t} = \delta/3$. It follows that we can take $\delta_{B,i,t} = \delta/3mT$, even though t really ranges from d to n.

Assume that $\mathcal{E}_G, \mathcal{E}_{M,i}$ and $\mathcal{E}_{B,i,t}$ hold for all arms i and times t. Then, by Theorem 5, the final number of pulls for arm i can be lower bounded by

$$k_i \geq \frac{\sigma_i^2}{\sum_j \sigma_j^2} n - c\left(\sqrt{\frac{\sigma_i^2}{\sigma_{\min}^2}} + 1\right)\sqrt{\frac{\sigma_i^2}{\sum_j \sigma_j^2} dn} + o\left(\sqrt{dn}\right), \tag{64}$$

where $c = 2\left(1 + \sqrt{2\log(12mn/\delta)/d}\right)$.

For notational simplicity, we denote by $\hat{\beta}_{i,t}$ the estimate after t pulls. Thus, with respect to our previous notation where $\hat{\beta}_{i,n}$ referred to our final estimate, we have that $\hat{\beta}_{i,k_{i,n}} = \hat{\beta}_{i,n}$ as $k_{i,n}$ is the total number of pulls for arm i.

If the $\mathcal{E}_{M,i}$ events hold, then we know that our $\hat{\beta}_{i,t}$ estimates are not very far from the true values β_i when t is large. In particular, we know that the error is controlled by the radius $R_{i,t}$ of the confidence ellipsoids. We expect these radiuses to decrease with the number of observations per arm, t. As we have a lower bound on the total number of pulls for arm i, $k_{i,n}$, if the confidence ellipsoids apply, then we can directly obtain an upper bound on the radius $R_{i,t}$ at the end of the process.

We need to do a bit of work to properly bound $\|\hat{\beta}_{i,k_{i,n}} - \beta_i\|_\Sigma^2$.

Fix arm i, and assume $\mathcal{E}_{M,i}$ holds. In addition, assume $\|\beta_i\|_2 \leq Z$ for all i. Let $\bar{V}_{i,t} = \lambda I + \mathbf{X}_{i,t}^T \mathbf{X}_{i,t}$, where $\mathbf{X}_{i,t}$ contains the first t observations pulled by arm i. We modify the proof of Theorem 2 in (Abbasi-Yadkori et al., 2011) by taking $x = (\hat{V}_t/t)(\hat{\beta}_t - \beta_*)$ in their equation 5 (we are using their notation in the latter expression). Assume the algorithm pulls arm i a total of t times —$k_{i,n}$ is a stopping time with respect to the σ-algebra that includes the environment (other arms)— then, by Theorem 13

$$\|\hat{\beta}_{i,t} - \beta_i\|_{\bar{V}_{i,t}/t} \leq \frac{\sigma_i}{\sqrt{t}}\sqrt{2\log\left(\frac{\det\left(\bar{V}_{i,t}\right)^{1/2}\det\left(\lambda I\right)^{-1/2}}{\delta_{M,i}}\right)} + \sqrt{\frac{\lambda}{t}}\,Z. \tag{65}$$

We would like to upper bound $\|\hat{\beta}_{i,k_{i,n}} - \beta_i\|_\Sigma$ by means of $\|\hat{\beta}_{i,k_{i,n}} - \beta_i\|_{\bar{V}_{i,k_{i,n}}/k_{i,n}}$. Note that when t grows, $\bar{V}_{i,t}/t \to \Sigma$ as the regularization is washed out. The distance between $\hat{\Sigma}_{i,t} = \bar{V}_{i,t}/t - (\lambda/t)I$ and Σ is captured by event $\epsilon_{C,i,t}$.

Formally, as $\mathcal{E}_G$ holds, we know that the difference between Σ and $\hat{\Sigma}_{i,t}$ is bounded in operator norm for any i and t by

$$\|\Sigma - \hat{\Sigma}_{i,t}\| \leq 2\left(1 + \sqrt{\frac{2}{d}\log\frac{2}{\delta_G}}\right)\sqrt{\frac{d}{t}}\|\Sigma\| = c\sqrt{\frac{d}{t}}\lambda_{\max}(\Sigma). \tag{66}$$

Then, as a consequence, for all $x \in \mathbb{R}^s$

$$x^T(\Sigma - \hat{\Sigma}_{i,t})x \leq c\,\lambda_{\max}(\Sigma)\sqrt{\frac{d}{t}}\|x\|_2^2. \tag{67}$$

In particular, by taking $x = \hat{\beta}_{i,t} - \beta_i$,

$$c\,\lambda_{\max}(\Sigma)\sqrt{\frac{d}{t}}\|\hat{\beta}_{i,t} - \beta_i\|_2^2 \geq (\hat{\beta}_{i,t} - \beta_i)^T(\Sigma - \hat{\Sigma}_{i,t})(\hat{\beta}_{i,t} - \beta_i) \tag{68}$$

$$= \|\hat{\beta}_{i,t} - \beta_i\|_\Sigma^2 - \|\hat{\beta}_{i,t} - \beta_i\|_{\hat{\Sigma}_{i,t}}^2. \tag{69}$$

In addition, note that $\|x\|_{\hat{\Sigma}_{i,t}}^2 = \|x\|_{\bar{V}_{i,t}/t}^2 - (\lambda/t)\|x\|_2^2$. We conclude that

$$\|\hat{\beta}_{i,t} - \beta_i\|_\Sigma^2 \leq \|\hat{\beta}_{i,t} - \beta_i\|_{\hat{\Sigma}_{i,t}}^2 + c\,\lambda_{\max}(\Sigma)\sqrt{\frac{d}{t}}\|\hat{\beta}_{i,t} - \beta_i\|_2^2 \tag{70}$$

$$= \|\hat{\beta}_{i,t} - \beta_i\|_{\bar{V}_{i,t}/t}^2 + \left(c\,\lambda_{\max}(\Sigma)\sqrt{\frac{d}{t}} - \frac{\lambda}{t}\right)\|\hat{\beta}_{i,t} - \beta_i\|_2^2. \tag{71}$$

On the other hand, we know that $\|\hat{\beta}_{i,t} - \beta_i\|_\Sigma^2 \geq \lambda_{\min}(\Sigma)\|\hat{\beta}_{i,t} - \beta_i\|_2^2$.

Therefore, by (65)

$$\|\hat{\beta}_{i,t} - \beta_i\|_\Sigma^2 \leq \frac{1}{1 - \frac{1}{\lambda_{\min}(\Sigma)}\left(c\,\lambda_{\max}(\Sigma)\sqrt{\frac{d}{t}} - \frac{\lambda}{t}\right)}\|\hat{\beta}_{i,t} - \beta_i\|_{\bar{V}_{i,t}/t}^2 \tag{72}$$

$$\leq \frac{1}{1 - \gamma_t}\left[\frac{\sigma_i}{\sqrt{t}}\sqrt{2\log\left(\frac{\det(\bar{V}_{i,t})^{1/2}\det(\lambda I)^{-1/2}}{\delta_{M,i}}\right)} + \frac{\sqrt{\lambda}Z}{\sqrt{t}}\right]^2 \tag{73}$$

$$\leq \frac{1}{1 - \gamma_t}\frac{1}{t}\left[\sigma_i\sqrt{2\left(\frac{1}{2}\log\left(\frac{\det(\bar{V}_{i,t})}{\det(\lambda I)}\right) + \log\left(\frac{1}{\delta_{M,i}}\right)\right)} + \sqrt{\lambda}Z\right]^2 \tag{74}$$

$$\leq \frac{1}{1 - \gamma_t}\frac{1}{t}\left[\sigma_i\sqrt{2\left(\frac{1}{2}\sum_{j=1}^{t}\|X_j\|_{\bar{V}_{i,t}^{-1}}^2 + \log\left(\frac{1}{\delta_{M,i}}\right)\right)} + \sqrt{\lambda}Z\right]^2, \tag{75}$$

where we defined $\gamma_t = \left(c\,\lambda_{\max}(\Sigma)\sqrt{\frac{d}{t}} - \frac{\lambda}{t}\right)/\lambda_{\min}(\Sigma)$, and we used Lemma 11 in (Abbasi-Yadkori et al., 2011) which shows that

$$\log\left(\frac{\det(\bar{V}_{i,t})}{\det(\lambda I)}\right) \leq \sum_{j=1}^{t}\|X_j\|_{\bar{V}_{i,t}^{-1}}^2. \tag{76}$$

We would like to approximate the $\bar{V}_{i,t}^{-1}$ norm, by means of the inverse covariance norm, Σ^{-1}. The whitened equation that is equivalent to (67) — see Lemma 12 — is given by $\|I - \hat{\bar{\Sigma}}_{i,t}\| \leq \epsilon$, with $\epsilon = c\sqrt{d/t}$.

It implies that for any $j = 1, \ldots, d$,

$$1 - c\sqrt{\frac{d}{t}} - O\left(\frac{d}{t}\right) \leq \lambda_j(\hat{\bar{\Sigma}}_{i,t}) \leq 1 + c\sqrt{\frac{d}{t}} + O\left(\frac{d}{t}\right). \tag{77}$$

The $\bar{V}_{i,t}^{-1}$ norm can be bounded as follows

$$\|x\|_{\bar{V}_{i,t}^{-1}}^2 = x^T \bar{V}_{i,t}^{-1} x = x^T \left(\lambda I + \mathbf{X}_{i,t}^T \mathbf{X}_{i,t}\right)^{-1} x \tag{78}$$

$$= x^T \Sigma^{-1/2} \Sigma^{1/2} \left(\lambda I + \mathbf{X}_{i,t}^T \mathbf{X}_{i,t}\right)^{-1} \Sigma^{1/2} \Sigma^{-1/2} x \tag{79}$$

$$= \bar{x}^T \left(\lambda \Sigma^{-1} + \bar{\mathbf{X}}_{i,t}^T \bar{\mathbf{X}}_{i,t}\right)^{-1} \bar{x} \tag{80}$$

$$= \frac{1}{t} \bar{x}^T \left(\frac{\lambda}{t} \Sigma^{-1} + \hat{\bar{\Sigma}}_{i,t}^{-1}\right)^{-1} \bar{x}, \tag{81}$$

where $\bar{x}$ denotes the whitened version of x. We can now apply the matrix inversion lemma to see that

$$\|x\|_{\bar{V}_{i,t}^{-1}}^2 = \frac{1}{t} \bar{x}^T \left(\frac{\lambda}{t} \Sigma^{-1} + \hat{\bar{\Sigma}}_{i,t}^{-1}\right)^{-1} \bar{x} \tag{82}$$

$$= \frac{1}{t} \bar{x}^T \left(\hat{\bar{\Sigma}}_{i,t} - \hat{\bar{\Sigma}}_{i,t} \Sigma^{-1/2} \left(\frac{t}{\lambda} I + \Sigma^{-1/2} \hat{\bar{\Sigma}}_{i,t} \Sigma^{-1/2}\right)^{-1} \Sigma^{-1/2} \hat{\bar{\Sigma}}_{i,t}\right) \bar{x} \tag{83}$$

$$= \frac{1}{t} \bar{x}^T \left(\hat{\bar{\Sigma}}_{i,t} - \hat{\bar{\Sigma}}_{i,t} \Sigma^{-1/2} R^{-1} \Sigma^{-1/2} \hat{\bar{\Sigma}}_{i,t}\right) \bar{x}, \tag{84}$$

where we implicitly defined $R = (t/\lambda)I + \Sigma^{-1/2} \hat{\bar{\Sigma}}_{i,t} \Sigma^{-1/2}$, a positive definite matrix. We upper bound the previous expression to conclude that

$$\|x\|_{\bar{V}_{i,t}^{-1}}^2 = \frac{1}{t} \bar{x}^T \left(\hat{\bar{\Sigma}}_{i,t} - \hat{\bar{\Sigma}}_{i,t} \Sigma^{-1/2} R^{-1} \Sigma^{-1/2} \hat{\bar{\Sigma}}_{i,t}\right) \bar{x} \tag{85}$$

$$\leq \frac{1}{t} \bar{x}^T \hat{\bar{\Sigma}}_{i,t} \bar{x} \leq \frac{\lambda_{\max}(\hat{\bar{\Sigma}}_{i,t})}{t} \|\bar{x}\|_2^2 \leq \frac{1 + c\sqrt{d/t} + O(d/t)}{t} \|\bar{x}\|_2^2. \tag{86}$$

If we now go back to (76), using the previous results, we see that

$$\sum_{j=1}^t \|X_j\|_{\bar{V}_{i,t}^{-1}}^2 \leq \left(1 + c\sqrt{\frac{d}{t}} + O\left(\frac{d}{t}\right)\right)\left(\frac{1}{t} \sum_{j=1}^t \|\bar{X}_j\|_2^2\right). \tag{87}$$

Substituting the upper bound in (75):

$$\|\hat{\beta}_{i,t} - \beta_i\|_{\Sigma}^2 \leq \frac{1}{1 - \gamma_t} \frac{1}{t} \left[\sigma_i \sqrt{2\left(\frac{1}{2} \sum_{j=1}^t \|X_j\|_{\bar{V}_{i,t}^{-1}}^2 + \log\left(\frac{1}{\delta_{M,i}}\right)\right)} + \sqrt{\lambda} Z\right]^2 \tag{88}$$

$$\leq \frac{1}{1 - \gamma_t} \frac{1}{t} \left[\sigma_i \sqrt{\left(1 + c\sqrt{\frac{d}{t}} + O\left(\frac{d}{t}\right)\right)\left(\frac{1}{t} \sum_{j=1}^t \|\bar{X}_j\|_2^2\right) + 2\log\frac{1}{\delta_{M,i}}} + \sqrt{\lambda} Z\right]^2.$$

By Corollary 15, with probability $1 - \delta_{B,i,t}$, the empirical average norm of the white gaussian observations is controlled by

$$\frac{1}{t} \sum_{j=1}^t \|\bar{X}_j\|^2 \leq d + 8\log\left(\frac{1}{\delta_{B,i,t}}\right)\sqrt{\frac{d}{t}}. \tag{89}$$

As $\delta_{B,i,t} = \delta/3mn$ and $\delta_{M,i} = \delta/3m$, we conclude that

$$
\|\hat{\beta}_{i,t} - \beta_i\|_\Sigma^2 \leq \frac{1}{1-\gamma_t} \frac{1}{t} \left[\sigma_i \sqrt{\left(1 + c\sqrt{\frac{d}{t}} + O\left(\frac{d}{t}\right)\right)\left(d + 8\log\left(\frac{3mn}{\delta}\right)\sqrt{\frac{d}{t}}\right) + 2\log\left(\frac{3m}{\delta}\right)} + \sqrt{\lambda}Z \right]^2
$$

$$
\leq \frac{1}{1 - \left(c\lambda_{\max}(\Sigma)\sqrt{\frac{d}{t}} - \frac{\lambda}{t}\right)/\lambda_{\min}(\Sigma)} \frac{1}{t} \left[\sigma_i \sqrt{\left(d + \left(c + 8\log\frac{3mn}{\delta}\right)\sqrt{\frac{d}{t}} + O\left(\frac{d}{t}\right)\right) + 2\log\frac{3m}{\delta}} + \sqrt{\lambda}Z \right]^2.
$$
(90)

At this point, recall that under our events

$$
k_{i,n} \geq k_{i,n}^* - C\sqrt{nd} - \Omega(n^{1/4}),
$$
(91)

where $C = \frac{C_\Delta + 8C_{\mathrm{Tr}}}{\sigma_{\min}^2 \sqrt{\lambda_{\min}}}$. As (90) decreases in t, we will bound the error $\|\hat{\beta}_{i,t} - \beta_i\|_\Sigma^2$ by taking the number of pulls $t = (\sigma_i^2/\sum_j \sigma_j^2)n + O(\sqrt{dn})$ (in particular, the RHS of (91)).

If we take $\lambda = 1/n$, we have that

$$
\|\hat{\beta}_{i,t} - \beta_i\|_\Sigma^2
$$
(92)

$$
\leq \frac{1}{1 - \left(c\lambda_{\max}(\Sigma)\sqrt{\frac{d}{t}} - \frac{\lambda}{t}\right)/\lambda_{\min}(\Sigma)} \frac{1}{t} \left[\sigma_i \sqrt{\left(d + \left(c + 8\log\frac{3mn}{\delta}\right)\sqrt{\frac{d}{t}} + O\left(\frac{d}{t}\right)\right) + 2\log\frac{3m}{\delta}} + \sqrt{\lambda}Z \right]^2
$$

$$
\leq \left(1 + c\frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Sigma)}\sqrt{\frac{d}{t}} + O\left(\frac{d}{t}\right)\right) \frac{1}{t} \left[\sigma_i \sqrt{\left(d + \left(c + 8\log\frac{3mn}{\delta}\right)\sqrt{\frac{d}{t}} + O\left(\frac{d}{t}\right)\right) + 2\log\frac{3m}{\delta}} + \sqrt{\lambda}Z \right]^2
$$

$$
\leq \left(1 + O\left(\sqrt{\frac{d}{t}}\right)\right) \frac{1}{t} \left[\sigma_i^2\left(d + 2\log\frac{3m}{\delta} + \left(c + 8\log\frac{3mn}{\delta}\right)\sqrt{\frac{d}{t}}\right) + \frac{Z^2}{n} + 2Z\sigma_i\sqrt{\frac{d + 2\log\frac{3m}{\delta}}{n}} + o\left(\sqrt{\frac{d}{n}}\right) \right].
$$

Now, by (91) and (59), and using the $\lambda_i = \sigma_i^2/\sum_j \sigma_j^2$ notation

$$
\|\hat{\beta}_{i,t} - \beta_i\|_\Sigma^2
$$
(93)

$$
\leq \left(1 + O\left(\sqrt{\frac{d}{n}}\right)\right) \frac{\left[\sigma_i^2\left(d + 2\log\frac{3m}{\delta}\right) + \sigma_i^2\left(c + 8\log\frac{3mn}{\delta}\right)\sqrt{\frac{d}{t}} + 2Z\sigma_i\sqrt{\frac{d}{n}} + o\left(\sqrt{\frac{d}{n}}\right)\right]}{k_{i,n}^* - C\sqrt{nd} - \Omega(n^{1/4})}
$$

$$
= \left(1 + O\left(\sqrt{\frac{d}{n}}\right)\right) \frac{\left[\sigma_i^2\left(d + 2\log\frac{3m}{\delta}\right) + \left(\sigma_i^2\left(c + 8\log\frac{3mn}{\delta}\right) + 2Z\sigma_i\right)\sqrt{\frac{d}{t}} + o\left(\sqrt{\frac{d}{n}}\right)\right]}{k_{i,n}^* - C\sqrt{nd} - \Omega(n^{1/4})}
$$

$$
= \left(1 + O\left(\sqrt{\frac{d}{n}}\right)\right)\left(\frac{1}{k_{i,n}^*} + O\left(\frac{\sqrt{d}}{\sigma_{\min}^2(\lambda_{\min}n)^{3/2}}\right)\right)\left[\sigma_i^2\left(d + 2\log\frac{3m}{\delta}\right) + \tilde{O}\left(\sqrt{\frac{d}{n}}\right)\right]
$$

$$
= \frac{\sigma_i^2}{k_{i,n}^*}\left(d + 2\log\frac{3m}{\delta}\right) + O\left(\frac{1}{\sigma_{\min}^2}\left(\frac{d}{\lambda_{\min}n}\right)^{3/2}\right).
$$
(94)

$\square$

E. Loss of a RLS-based Learning Algorithm

E.1. Distribution of RLS estimates

Proposition 16. *Given a linear regression problem with observations $Y = X^\mathsf{T}\beta + \epsilon$ with Gaussian noise with variance σ^2, after n contexts $\mathbf{X}$ and the corresponding observations $\mathbf{Y}$, the ridge estimate of parameter λ is obtained as*

$$\hat{\beta}^\lambda = (\mathbf{X}^\mathsf{T}\mathbf{X} + \lambda I)^{-1}\mathbf{X}^\mathsf{T}\mathbf{Y} = \mathbf{W}\mathbf{X}^\mathsf{T}\mathbf{Y},$$

with $\mathbf{W} = (\mathbf{X}^\mathsf{T}\mathbf{X} + \lambda I)^{-1}$, and its distribution conditioned on $\mathbf{X}$ is

$$\hat{\beta}^\lambda \mid \mathbf{X} \sim \mathcal{N}\left(\beta - \lambda \mathbf{W}\beta, \sigma^2 \,\mathbf{W}(\mathbf{X}^\mathsf{T}\mathbf{X})\mathbf{W}^\mathsf{T}\right). \tag{95}$$

Proof. Recalling the definition of the OLS estimator $\widehat{\beta}$ (assuming it exists), we can easily rewrite the RLS estimator as

$$\hat{\beta}^\lambda = (\mathbf{X}^\mathsf{T}\mathbf{X} + \lambda I)^{-1}(\mathbf{X}^\mathsf{T}\mathbf{X})(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{Y} = (\mathbf{X}^\mathsf{T}\mathbf{X} + \lambda I)^{-1}(\mathbf{X}^\mathsf{T}\mathbf{X})\widehat{\beta}.$$

This immediately gives that the conditional distribution of $\hat{\beta}^\lambda$ is Gaussian as for $\widehat{\beta}$. We just need to compute the corresponding mean vector and the covariance matrix. We first notice that the RLS estimator is biased as

$$\mathbb{E}[\hat{\beta}^\lambda | \mathbf{X}] = (\mathbf{X}^\mathsf{T}\mathbf{X} + \lambda I)^{-1}(\mathbf{X}^\mathsf{T}\mathbf{X})\beta.$$

Let $\mathbf{S} = \mathbf{X}^\mathsf{T}\mathbf{X}$, then we can further rewrite the bias as

$$\begin{aligned}
\mathbb{E}[\hat{\beta}^\lambda | \mathbf{X}] &= (\mathbf{S} + \lambda\mathbf{S}\mathbf{S}^{-1})^{-1}\mathbf{S}\beta = \left(\mathbf{S}(I + \lambda\mathbf{S}^{-1})\right)^{-1}\mathbf{S}\beta \\
&= (I + \lambda\mathbf{S}^{-1})^{-1}\beta = \left(I - \lambda(\mathbf{S} + \lambda I)^{-1}\right)\beta \\
&= \beta - \lambda(\mathbf{S} + \lambda I)^{-1}\beta = \beta - \lambda\mathbf{W}\beta,
\end{aligned}$$

where we used the matrix inversion lemma. Recalling that the covariance of $\widehat{\beta}$ is $\sigma^2(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}$, the covariance of $\widehat{\beta}^\lambda$ is then

$$\mathrm{Cov}\left[\hat{\beta}^\lambda | \mathbf{X}\right] = \mathbf{W}(\mathbf{X}^\mathsf{T}\mathbf{X})\mathrm{Cov}\left[\hat{\beta} | \mathbf{X}\right](\mathbf{X}^\mathsf{T}\mathbf{X})\mathbf{W}^\mathsf{T} = \sigma^2\mathbf{W}(\mathbf{X}^\mathsf{T}\mathbf{X})\mathbf{W}^\mathsf{T}.$$

$\square$

E.2. Loss Function of a RLS-based Algorithm

We start by proving the loss function in the case of a static algorithm.

Lemma 17. *Let $\mathcal{A}$ be a learning algorithm that selects instance i for $k_{i,n}$ times, where $k_{i,n}$ is a fixed quantity chosen in advance, and that returns estimates $\widehat{\beta}_i^\lambda$ obtained by RLS with regularization λ. Then its loss after n steps can be expressed as*

$$L_n(\mathcal{A}_{stat}) = \max_{i\in[m]} \mathrm{Tr}\left(\Sigma\mathbb{E}\left[\mathbf{W}_{i,n}\left(\sigma_i^2\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n} + \lambda^2\beta_i\beta_i^\mathsf{T}\right)\mathbf{W}_{i,n}^\mathsf{T}\right]\right), \tag{96}$$

where $\mathbf{W}_{i,n} = (\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n} + \lambda I)^{-1}$, and $\mathbf{X}_{n_i}$ is the matrix with the $k_{i,n}$ contexts from instance i.

Proof. The proof follows the same steps as in App. A up to Eq. 22, where we have

$$L_n(\mathcal{A}_{\mathrm{stat}}) = \max_{i\in[m]} \mathrm{Tr}\left(\mathbb{E}_{\mathbf{X}_i}\left[\mathbb{E}_{\boldsymbol{\epsilon}_i}\left[\Sigma(\beta_i - \hat{\beta}_i)(\beta_i - \hat{\beta}_i)^\mathsf{T}|\mathbf{X}_i\right]\right]\right).$$

Following Proposition 16, we can refine the inner expectation as

$$\begin{aligned}
&\mathbb{E}\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^\mathsf{T} \mid \mathbf{X}\right] \\
&= \mathbb{E}\left[(\hat{\beta} - \beta + \lambda\mathbf{W}\beta - \lambda\mathbf{W}\beta)(\hat{\beta} - \beta + \lambda\mathbf{W}\beta - \lambda\mathbf{W}\beta)^\mathsf{T} \mid \mathbf{X}\right] \\
&= \mathbb{E}\left[(\hat{\beta} - \mathbb{E}[\hat{\beta} \mid \mathbf{X}] - \lambda\mathbf{W}\beta)(\hat{\beta} - \mathbb{E}[\hat{\beta} \mid \mathbf{X}] - \lambda\mathbf{W}\beta)^\mathsf{T} \mid \mathbf{X}\right] \\
&= \mathbb{E}\left[(\hat{\beta} - \mathbb{E}[\hat{\beta} \mid \mathbf{X}])(\hat{\beta} - \mathbb{E}[\hat{\beta} \mid \mathbf{X}])^\mathsf{T} \mid \mathbf{X}\right] + \lambda^2\mathbf{W}\beta\beta^\mathsf{T}\mathbf{W}^\mathsf{T} \\
&= \sigma^2\,\mathbf{W}(\mathbf{X}^\mathsf{T}\mathbf{X})\mathbf{W}^\mathsf{T} + \lambda^2\mathbf{W}\beta\beta^\mathsf{T}\mathbf{W}^\mathsf{T} \\
&= \mathbf{W}\left[\sigma^2\mathbf{X}^\mathsf{T}\mathbf{X} + \lambda^2\beta\beta^\mathsf{T}\right]\mathbf{W}^\mathsf{T}.
\end{aligned}$$

Plugging the final expression back into $L_n(\mathcal{A}_{\text{static}})$ we obtain the desired expression. □

We notice that a result similar to Lemma 9 holds for RLS estimates as well.

Proposition 18. *Assume the noise ϵ is Gaussian. Let $\hat{\sigma}^2$ be the estimate for σ^2 computed by using the residuals of the OLS solution $\hat{\beta}$. Then, $\hat{\beta}^\lambda$ and $\hat{\sigma}^2$ are independent random variable conditionally to $\mathbf{X}$.*

Proof. As shown in the proof of Proposition 16, we have $\hat{\beta}^\lambda = (\mathbf{X}^\mathsf{T}\mathbf{X} + \lambda I)^{-1}(\mathbf{X}^\mathsf{T}\mathbf{X})\widehat{\beta}$ and we know that functions of independent random variables are themselves independent. Since the matrix mapping $\widehat{\beta}$ to $\widehat{\beta}^\lambda$ is fixed given $\mathbf{X}$, and $\hat{\beta}$ and $\hat{\sigma}^2$ are conditionally independent from Lemma 9, then the statement follows. □

We can now combine Proposition 18 and Lemma 17 to conclude that a similar expression to Eq. 97 holds for the ridge estimators also when a non-static algorithm such as TRACE-UCB is run.

Lemma 19. *Let $\mathcal{A}$ be a learning algorithm such that I_t is chosen as a function of $\mathcal{D}_{t-1} = \{X_1, I_1, Y_{I_1,1}, \ldots, X_{t-1}, I_{t-1}, Y_{I_{t-1},t-1}\}$, and that it returns estimates $\widehat{\beta}_i^\lambda$ obtained by RLS with regularization λ. Then its loss after n steps can be expressed as*

$$L_n(\mathcal{A}) = \max_{i \in [m]} \operatorname{Tr}\left(\Sigma\mathbb{E}\left[\mathbf{W}_{i,n}\left(\sigma_i^2\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n} + \lambda^2\beta_i\beta_i^\mathsf{T}\right)\mathbf{W}_{i,n}^\mathsf{T}\right]\right), \tag{97}$$

where $\mathbf{W}_{i,n} = (\mathbf{X}_{i,n}^\mathsf{T}\mathbf{X}_{i,n} + \lambda I)^{-1}$, and $\mathbf{X}_{i,n}$ is the matrix with the $k_{i,n}$ contexts from instance i.

Proof. The proof follows immediately by extending Lemma 9 to $\hat{\beta}_\lambda$ as, by Proposition 18, $\hat{\beta}_\lambda$ and $\hat{\sigma}_{\text{OLS}}^2$ are independent. Then, we proceed in a way similar to that in the proof of Lemma 2 to perform the required conditioning. □

F. Sparse Trace-UCB Algorithm

F.1. Summary

High-dimensional linear regression models are remarkably common in practice. Companies tend to record a large number of features of their customers, and feed them to their prediction models. There are also cases in which the number of problem instances under consideration m is large, e.g., too many courses in the MOOC example described in the introduction. Unless the horizon n is still proportionally large w.r.t. md, these scenarios require special attention. In particular, algorithms like TRACE-UCB that adaptively use contexts in their allocation strategy become more robust than their context-free counterparts.

A natural assumption in such scenarios is *sparsity*, i.e., only a small subset of features are relevant to the prediction problem at hand (have non-zero coefficient). In our setting of m problem instances, it is often reasonable to assume that these instances are related to each other, and thus, it makes sense to extend the concept of sparsity to *joint sparsity*, i.e., a sparsity pattern across the instances. Formally, we assume that there exists a $s \ll d$ such that

$$|S| \triangleq |\cup_{i \in [m]} \operatorname{supp}(\beta_i)| = s, \tag{98}$$

where $\operatorname{supp}(\beta_i) = \{j \in [d] : \beta_i^{(j)} \neq 0\}$ denotes the support of the i'th problem instance. A special case of joint sparsity is when $|\operatorname{supp}(\beta_i)| \approx s$, for all i, i.e., most of the relevant features are shared across the instances.

In this section, we focus on the scenario where $dm > n$. When we can only allocate a small (relative to d) number of contexts to each problem instance, proper balancing of contexts becomes extremely important, and thus, the algorithms that do not take into account context in their allocation are destined to fail. Although TRACE-UCB has the advantage of using context in its allocation strategy, it still needs to quickly discover the relevant features (those in the support) and only use those in its allocation strategy.

This motivates a two-stage algorithm, we call it SPARSE-TRACE-UCB, whose pseudocode is in Algorithm 2. In the first stage, the algorithm allocates contexts uniformly to all the instances, L contexts per instance, and then recovers the support. In the second stage, it relies on the discovered support $\widehat{S}$, and applies the standard TRACE-UCB to all the instances, but only takes into account the features in $\widehat{S}$. Note that L should be large enough that with high probability, support is exactly discovered, i.e., $\widehat{S} = S$.

There exists a large literature on how to perform *simultaneous* support discovery in jointly sparse linear regression problems (Negahban & Wainwright, 2011; Obozinski et al., 2011; Wang et al., 2013), which we discuss in detail below.

Most of these algorithms minimize the regularized empirical loss

$$\min_{M \in \mathbb{R}^{d \times m}} \frac{1}{k} \sum_{i=1}^{m} \|\mathbf{Y}_i - \mathbf{X}_i \, M[,i]\|^2 + \lambda \, \|M\|,$$

where k is the number of samples per problem, M be the matrix whose i'th column is $M[,i] = \hat{\beta}_i$, $\mathbf{X}_i \in \mathbb{R}^{k \times d}$, and $\mathbf{Y}_i = \mathbf{X}_i \beta_i + \epsilon_i$. In particular, they use l_a/l_b block regularization norm, i.e., $\|M\|_{l_a/l_b} = \|v\|_{l_a}$, where $v_i = \|M[i,]\|_{l_b}$ and $M[i,]$ is the i'th row of M. In short, the SPARSE-TRACE-UCB algorithm uses the l_1/l_2 block regularization Lasso algorithm (Wang et al., 2013), an extension of the algorithm in (Obozinski et al., 2011), for its support discovery stage.

We extend the guarantees of Theorem 7 to the high dimensional case with joint sparsity, assuming s is known.

The following is the main result of this section:

Theorem 20. *Let $\delta_1 > 0$. Assume $\|\beta_i\|_2 \leq Z$ for all i, for some $Z > 0$, and assume the parameters $(n, d, s, \beta_i, \Sigma)$ satisfy conditions **C1** to **C5** in (Wang et al., 2013). Let ψ be the sparsity overlap function defined in (Obozinski et al., 2011). If $L > 2(1 + v) \, \psi \, \log(d - s) \rho_u(\Sigma_{S_C S_C | S}^{(1:m)})/\gamma^2$ for some constant $v > 0$, and $n - Lm \geq (s + 1)m$, then, with probability at least $1 - \delta_1 - \delta_2$,*

$$\widetilde{L}_n(\mathcal{A}) \leq \frac{\sum_j \sigma_j^2}{n - Lm} \left(s + 2 \log \frac{3m}{\delta_1} \right) + \frac{2c}{\sqrt{\sigma_{\min}^2}} \left(\frac{s \sum_j \sigma_j^2}{n - Lm} \right)^{3/2} + o(z), \tag{99}$$

where $c \leq 2 \left(1 + \sqrt{2 \log(12mn/\delta_1)/s} \right)$ and we defined $\delta_2 = m \exp(-c_0 \log s) + \exp(-c_1 \log(d - s))$ for positive constants $c_0, c_1 > 0$, and $z = (s/(n - Lm))^{3/2}$.

The exact technical assumptions and the proof are given and discussed in below. We simply combine the high-probability results of Theorem 7, and the high-probability support recovery of Theorem 2 in (Wang et al., 2013).

In addition, we provide Corollary 21, where we study the regime of interest where the support overlap is complete (for simplicity), $n = C_1\, ms \log d \ll md$ for $C_1 > 0$, and $L = C_2\, s \log d$, for $C_1 - C_2 > 0$.

Corollary 21. *Under the assumptions of Theorem 20, let $\delta_1 > 0$, assume $n = C_1\, ms \log d$, the support of all arms are equal, and set $L = C_2\, s \log d$, for $\bar{C} := C_1 - C_2 > 0$. Then, with probability at least $1 - \delta_1 - \delta_2$,*

$$\widetilde{L}_n(\mathcal{A}) \leq \frac{\sum_j \sigma_j^2}{\bar{C} ms \log d}\left(s + 2\log\frac{3m}{\delta_1}\right) + \frac{2c}{\sqrt{\sigma_{\min}^2}}\left(\frac{\sum_j \sigma_j^2}{\bar{C} m \log d}\right)^{3/2} + o\,(z) \tag{100}$$

where $c \leq 2\left(1 + \sqrt{2\log(12mn/\delta_1)/s}\right)$ and we defined $\delta_2 = m\exp(-c_0 \log s) + \exp(-c_1 \log(d - s))$ for constants $c_0, c_1 > 0$, and $z = \left(\bar{C} m \log d\right)^{-3/2}$.

Algorithm 2 contains the pseudocode of our Sparse-TRACE-UCB algorithm.

Algorithm 2 Sparse-TRACE-UCB Algorithm.

1: **for** $i = 1, \dots, m$ **do**
2:     Select problem instance i exactly L times
3: **end for**
4: Run l_1/l_2 Lasso to recover support $\bar{S} = \cup_i \operatorname{supp}(\hat{\beta}_{i,L})$
5: **for** $i = 1, \dots, m$ **do**
6:     Select problem instance i exactly $s + 1$ times
7:     Compute its OLS estimates $\hat{\beta}_{i,m(L+s+1)}$ and $\hat{\sigma}^2_{i,m(L+s+1)}$ with respect to dimensions in $\bar{S}$.
8: **end for**
9: **for** steps $t = m(L + s + 1) + 1, \dots, n$ **do**
10:     **for** problem instance $1 \leq i \leq m$ **do**
11:         Compute score based on $\bar{S}$ dimensions only:

$$s_{i,t-1} = \frac{\widehat{\sigma}^2_{i,t-1} + \Delta_{i,t-1}}{k_{i,t-1}} \operatorname{Tr}\big(\Sigma \hat{\Sigma}^{-1}_{i,t-1}\big)$$

12:     **end for**
13:     Select problem instance $I_t = \arg\max_{i \in [m]} s_{i,t}$
14:     Observe X_t and $Y_{I_t,t}$
15:     Update OLS estimators $\hat{\beta}_{I_t,t}$ and $\hat{\sigma}^2_{I_t,t}$ based on $\bar{S}$
16: **end for**
17: Return RLS estimates $\{\hat{\beta}^\lambda_i\}^m_{i=1}$, with $\hat{\beta}^\lambda_{ij} = 0$ if $j \notin \bar{S}$

Given our pure exploration perspective, it is obviously more efficient to learn the true supports as soon as possible. That way we can adjust our behavior by collecting the right data based on our initial findings. Note that this is not always the case; for example, if the total number of pulls is unknown. Then it is not clear what is the right amount of budget to invest upfront to recover the supports (see tracking algorithms and doubling trick).

We briefly describe Algorithm 2 in words. First, in the *recovery* stage we sequentially pull all arms a number of times, say L times. We do not take into account the context, and just apply a round robin technique to pull each arm exactly L times. In total, there are exactly s components that are non-zero for at least one arm (out of d). After the Lm pulls, we use a block-regularized Lasso algorithm to recover the joint sparsity pattern. We discuss some of the alternatives later. The outcome of this stage is a common support $\widehat{S} := \cup_i \mathbf{supp}(\hat{\beta}_i)$. With high probability we recover the true support $\widehat{S} = S$. In the second stage, or *pure exploration* stage, the original TRACE-UCB algorithm is applied. The TRACE-UCB algorithm works by computing an estimate $\hat{\sigma}^2_i$ at each step t for each arm i. Then, it pulls the arm maximizing the score

$$s_{i,t-1} = \frac{\widehat{\sigma}^2_{i,t-1} + \Delta_{i,t-1}}{k_{i,t-1}} \operatorname{Tr}\big(\Sigma \hat{\Sigma}^{-1}_{i,t-1}\big).$$

The key observation is that in the second stage we only consider the components of each context that are in $\widehat{S}$. In particular, we start by pulling $s + 1$ times each arm so that we can compute the initial OLS estimates $\hat{\beta}_i^{\text{OLS}}$ and $\hat{\sigma}_i^2$. We keep updating those estimates when an arm is pulled, and the trace is computed with respect to the components in $\widehat{S}$ only.

Finally, we return the Ridge estimates based only on the data collected in the second stage.

F.2. A note on the Static Allocation

What is the optimal static performance in this setting if the σ^2's are known? For simplicity, suppose we pull arm i exactly $(\sigma_i^2 / \sum_j \sigma_j^2)\, n$ times. We are interested in Lasso guarantees for $\|X^T(\hat{\beta}_i - \beta_i)\|_2^2$. Note in this case we can actually set λ_i as a function of σ_i^2 as required in most Lasso analyses, because σ_i^2 is known.

A common guarantee is as follows (see (Hastie et al., 2015; Raskutti et al., 2010)). With high probability

$$\|\hat{\beta}_i - \beta_i\|_2^2 \le \frac{c^2 \sigma_i^2}{\gamma^2}\, \frac{\tau s \log d}{k},$$

where k is the number of observations, d the ambient dimension, s the efficient dimension, γ is the restricted eigenvalues constant for Σ, $\tau > 2$ is the parameter that tunes the probability bound, and c is a universal constant.

Thus, if we set $k = (\sigma_i^2 / \sum_j \sigma_j^2)\, n$, then we obtain that whp

$$\|\hat{\beta}_i - \beta_i\|_2^2 \le \frac{c^2 \tau}{\gamma^2} \left(\sum_{j=1}^m \sigma_j^2 \right) \frac{s \log d}{n}. \tag{101}$$

Note that the latter event is independent across different $i \in [m]$, so all of them simultaneously hold with high probability. The term γ^{-2} was expected as depending on the correlation levels in Σ the problem can be easier or harder. In addition, note that as $\|\hat{\beta}_i - \beta_i\|_\Sigma^2 = \text{Tr}(\Sigma(\hat{\beta}_i - \beta_i)(\hat{\beta}_i - \beta_i)^T)$, we have that

$$\lambda_{\min}(\Sigma)\, \|\hat{\beta}_i - \beta_i\|_2^2 \le \|\hat{\beta}_i - \beta_i\|_\Sigma^2 \le \lambda_{\max}(\Sigma)\|\hat{\beta}_i - \beta_i\|_2^2. \tag{102}$$

F.3. Simultaneous Support Recovery

There has been a large amount of research on how to perform *simultaneous* support recovery in sparse settings for multiple regressions. Let M be the matrix whose i-th column is $\text{M}^{(i)} = \beta_i$.

A common objective function after k observations per problem is

$$\min_{\bar{\text{M}} \in \mathbb{R}^{d \times m}} \frac{1}{k} \sum_{j=1}^m \|\mathbf{Y}_j - \mathbf{X}_j \bar{\text{M}}^{(j)}\|^2 + \lambda \|\bar{\text{M}}\|, \tag{103}$$

where we assumed $\mathbf{Y}_j = \mathbf{X}_j \beta_j + \epsilon_j$, and $\mathbf{X}_j \in \mathbb{R}^{k \times d}$, $\mathbf{Y}_j, \epsilon_j \in \mathbb{R}^k$ and $\beta_j \in \mathbb{R}^d$.

The l_a/l_b block regularization norm is

$$\|\bar{\text{M}}\|_{l_a/l_b} = \|v\|_a, \qquad \text{where } v_j = \|\bar{\text{M}}_j\|_b \quad \bar{\text{M}}_j \text{ is the j-th row of } \bar{\text{M}}. \tag{104}$$

There are a few differences among the most popular pieces of work.

Negahban and Wainwright (Negahban & Wainwright, 2011) consider random Gaussian designs $\mathbf{X}_j \sim \mathcal{N}(0, \Sigma_j)$ with random Gaussian noise (and common variance). The regularization norm is l_1/l_∞. In words, they take the sum of the absolute values of the maximum element per row in $\bar{\text{M}}$. This forces sparsity (via the l_1 norm), but once a row is selected there is no penalty in increasing the $\bar{\beta}$ components up to the current maximum of the row. They tune λ as in the standard analysis of Lasso, that is, proportionally to σ^2, which is unknown in our case. Results are non-asymptotic, and recovery happens with high probability when the number of observations is $k > Cs(m + \log d)$. They show that if the overlap is not large enough (2/3 of the support, for $m = 2$ regression problems), then running independent Lasso estimates has higher statistical efficiency. We can actually directly use the results in (Negahban & Wainwright, 2011) if we assume an upper bound $\sigma_{\max}^2 \le R$ is known.

Obozinski, Wainwright and Jordan (Obozinski et al., 2011) use l_1/l_2 block regularization (aka Multivariate Group Lasso). Their design is random Gaussian, but it is fixed across regressions: $\mathbf{X}_j = \mathbf{X}$. They provide asymptotic guarantees under the scaling $k, d, s \to \infty$, $d - s \to \infty$, and standard assumptions like bounded Σ-eigenspectrum, the irrepresentable condition, and self-incoherence. The first condition is not only required for support recovery, but also for l_2 consistency. The last two conditions are not required for risk consistency, while essential for support recovery. To capture the amount of non-zero pattern overlap among regressions, they define the sparsity overlap function ψ, and their sample requirements are a function of ψ. In particular, one needs $k > C \, \psi \, log(d - s)$, where the constant C depends on quantities related to the covariance matrix of the design matrices, and ψ can be equal to s/m, if all the patterns overlap, and at most s if they are disjoint.

Their theorems use a sequence of regularization parameters

$$\lambda_k = \sqrt{\frac{f(d) \log d}{k}}, \qquad \text{where } f(d) \to \infty \text{ as } d \to \infty,$$

in such a way that $\lambda_k \to 0$ as $k, d \to \infty$. Finally, $k > 2s$ is also required. They also provide a two-stage algorithm for efficient estimation of *individual* supports for each regression problem. All these optimization problems are convex, and can be efficiently solved in general.

To overcome the issue of common designs (we do not pull each context several times), we use the results by Wang, Liang, and Xing in (Wang et al., 2013). They extend the guarantees in (Obozinski et al., 2011) to the case where the design matrices are independently sampled for each regression problem. In order to formally present their result, we describe some assumptions. Let $\Sigma^{(i)}$ be the covariance matrix for the design observations of the i-th regression (in our case, they are all equal to Σ), and S the union of the sparse supports across regressions.

- **C1** There exists $\gamma \in (0, 1]$ such that $\||A\||_\infty \leq 1 - \gamma$, where

$$A_{js} = \max_{1 \leq i \leq m} \left| \left(\Sigma_{S^C S}^{(i)} \left(\Sigma_{SS}^{(i)} \right)^{-1} \right)_{js} \right|, \tag{105}$$

  for $j \in S^C$ and $s \in S$.

- **C2** There are constants $0 < C_{\min} \leq C_{\max} < \infty$, such that the eigenvalues of all matrices $\Sigma^{(i)}$ are in $[C_{\min}, C_{\max}]$.

- **C3** There exists a constant $D_{\max} < \infty$ such that

$$\max_{1 \leq i \leq m} \||\left(\Sigma_{SS}^{(i)} \right)^{-1}\||_\infty \leq D_{\max}. \tag{106}$$

- **C4** Define the regularization parameter

$$\lambda_k = \sqrt{\frac{f(d) \log d}{k}}, \qquad \text{where } f(d) \to \infty \text{ as } d \to \infty, \tag{107}$$

  such that $\lambda_k \to 0$ as $k \to \infty$.

- **C5** Define $\rho(k, s, \lambda_k)$ as

$$\rho(k, s, \lambda_k) := \sqrt{\frac{8\sigma_{\max}^2 s \log s}{k \, C_{\min}}} + \lambda_k \left(D_{\max} + \frac{12s}{C_{\min}\sqrt{k}} \right), \tag{108}$$

  and assume $\rho(k, s, \lambda_k)/b_{\min}^* = o(1)$, where $b_{\min}^* = \min_{j \in S} \|\mathrm{M}_j\|_2$.

We state the main theorem in (Wang et al., 2013); k is the number of observations *per* regression.

Theorem 22. *Assume the parameters $(k, d, s, \mathrm{M}, \Sigma^{(1:m)})$ satisfy conditions **C1** to **C5**. If for some small constant $v > 0$,*

$$k > 2(1 + v) \, \psi \, \log(d - s) \frac{\rho_u(\Sigma_{S_C S_C | S}^{(1:m)})}{\gamma^2}, \tag{109}$$

then the l_1/l_2 regularized problem given in (103) *has a unique solution* $\hat{M}$*, the support union* $\mathbf{supp}(\hat{M})$ *equals the true support* S*, and* $\|\hat{M} - M\|_{l_\infty/l_2} = o(b_{\min}^*)$*, with probability greater than*

$$1 - m \exp(-c_0 \log s) - \exp(-c_1 \log(d - s)), \tag{110}$$

where c_0 and c_1 are constants.

The following proposition is also derived in (Wang et al., 2013) (Proposition 1):

Proposition 23. *Assume $\Sigma^{(1:m)}$ satisfy **C2**, then ψ is bounded by*

$$\frac{s}{m\,C_{\min}} \leq \psi = \psi\left(M, \Sigma^{(1:m)}\right) \leq \frac{s}{C_{\min}}. \tag{111}$$

For our purposes, there is a single Σ, which implies that we can remove the $\max$ expressions in **C1** and **C3**. Corollary 2 in (Wang et al., 2013) establishes that when supports are equal for all arms, the number of samples required per arm is reduced by a factor of m.

F.4. High-Dimensional Trace-UCB Guarantees

If the support overlap is complete we can reduce the sampling complexity of the first stage by a factor of m; we only need

$$Lm > 2(1 + v)\, s\, \log(d - s) \frac{\rho_u(\Sigma_{S_C S_C | S}^{(1:m)})}{C_{\min}\, \gamma^2} \tag{112}$$

observations in total, for some small constant $v > 0$.

Now we show our main result for high-dimensional Trace-UCB, Theorem 20.

Theorem. *Let $\delta_1 > 0$. Assume $\|\beta_i\|_2 \leq Z$ for all i, for some $Z > 0$, and assume the parameters $(n, d, s, \beta_i, \Sigma)$ satisfy conditions **C1** to **C5** in (Wang et al., 2013). Let ψ be the sparsity overlap function defined in (Obozinski et al., 2011). If $L > 2(1 + v)\, \psi\, \log(d - s)\rho_u(\Sigma_{S_C S_C | S}^{(1:m)})/\gamma^2$ for some constant $v > 0$, and $n - Lm \geq (s + 1)m$, then, with probability at least $1 - \delta_1 - \delta_2$,*

$$\widetilde{L}_n(\mathcal{A}) \leq \frac{\sum_j \sigma_j^2}{n - Lm}\left(s + 2\log\frac{3m}{\delta_1}\right) + \frac{2c}{\sqrt{\sigma_{\min}^2}}\left(\frac{s\sum_j \sigma_j^2}{n - Lm}\right)^{3/2} + o\left(z\right), \tag{113}$$

where $c \leq 2\left(1 + \sqrt{2\log(12mn/\delta_1)/s}\right)$ and we defined $\delta_2 = m\exp(-c_0 \log s) + \exp(-c_1 \log(d - s))$ for positive constants $c_0, c_1 > 0$, and $z = (s/(n - Lm))^{3/2}$.

Proof. We start by assuming the recovered support $\hat{S}$ is equal to the true support S. This event, say $\mathcal{E}_S$, holds with probability at least $1 - \delta_2$ by Theorem 22 when L satisfies (112).

Then, we fix $\delta_1 > 0$, and run the second stage applying the Trace-UCB algorithm in the s-dimensional space given by the components in $\hat{S}$.

By Theorem 7, if $n - Lm \geq (s + 1)m$, then, with probability at least $1 - \delta_1$, the following holds:

$$\widetilde{L}_n(\mathcal{A})_S \leq \frac{\sum_j \sigma_j^2}{n - Lm}\left(s + 2\log\frac{3m}{\delta_1}\right) + \frac{2c}{\sqrt{\sigma_{\min}^2}}\left(\frac{s\sum_j \sigma_j^2}{n - Lm}\right)^{3/2} + o\left(\left(\frac{s}{n - Lm}\right)^{3/2}\right), \tag{114}$$

where $\widetilde{L}_n(\mathcal{A})_S$ denotes the loss restricted to the components in β that are in $\hat{S}$ (and Σ_S). However, under event $\mathcal{E}_S$, we recovered the true support, and our final estimates for β_{ij} for each $j \notin S$ and arm i will be equal to zero, which corresponds to their true value. Hence $\widetilde{L}_n(\mathcal{A}) = \widetilde{L}_n(\mathcal{A})_S$.

We conclude that (114) holds with probability at least $1 - \delta_1 - \delta_2$. $\qquad\square$

One regime of interest is when $n = C_1\, ms \log d \ll md$. In addition, let us assume complete support overlap across arms, so $\psi = s/Cm$. Then, we set the number of initial pulls per arm to be $L = C_2\, s \log d$, with $C_1 > C_2$.

In this case, we have that Corollary 21 holds.