LEARNING IN PROPHET INEQUALITIES WITH NOISY OBSERVATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study the prophet inequality, a fundamental problem in online decision-making and optimal stopping, in a practical setting where rewards are observed only through noisy realizations and reward distributions are unknown. At each stage, the decision-maker receives a noisy reward whose true value follows a linear model with an unknown latent parameter, and observes a feature vector drawn from a distribution. To address this challenge, we propose algorithms that integrate learning and decision-making via lower-confidence-bound (LCB) thresholding. In the i.i.d. setting, we establish that both an Explore-then-Decide strategy and an ε-Greedy variant achieve the sharp competitive ratio of $1 - 1/e$. For non-identical distributions, we show that a competitive ratio of $1/2$ can be guaranteed against a relaxed benchmark. Moreover, with window access to past rewards, the optimal ratio of $1/2$ against the optimal benchmark is achieved. Experiments on synthetic datasets confirm our theoretical results and demonstrate the efficiency of our algorithms.

1 INTRODUCTION

The prophet inequality is a fundamental problem in online decision-making and optimal stopping (Hill & Kertz, 1992). A decision-maker (or gambler) sequentially observes a stream of random variables (or rewards) revealed one by one and must decide at each stage whether to *accept the current value and stop*, or *continue to the next stage*. The benchmark is the *prophet*, an omniscient agent who knows all realizations in advance. The objective of the gambler is to design an online stopping rule whose expected payoff is competitive with that of the prophet, aiming to maximize the competitive ratio. This framework has been extensively studied, owing to its rich mathematical structure and broad applications such as posted-price mechanisms (Lucier, 2017), online ad allocation (Alaei et al., 2012), and hiring processes in labor markets (Arsenis & Kleinberg, 2022).

Classical work has established sharp guarantees when the underlying distributions are known. In particular, Samuel-Cahn (1984) showed that a single-threshold strategy achieves the optimal ratio of $1/2$ for independent but non-identical distributions, while in the i.i.d. case, $1 - 1/e$ was achieved in Hill & Kertz (1982) and later improved by Abolhassani et al. (2017); Correa et al. (2017).

Crucially, all these results rely on full knowledge of the distributions, an assumption that rarely holds in practice. More recently, attention has shifted toward the prophet inequality under *unknown* distributions (Correa et al., 2019a; 2020; Goldenshluger & Zeevi, 2022; Immorlica et al., 2023). In particular, Correa et al. (2019a) showed that, in the unknown-distribution setting, a competitive ratio of $1/e(\approx 0.368)$ can be achieved by the classical optimal algorithm for the secretary problem. To obtain the higher ratio of $1 - 1/e(\approx 0.632)$ in the i.i.d. reward distribution setting, however, $\Theta(n)$ additional offline reward samples are required. Likewise, in the non-i.i.d. setting with unknown distribution, attaining $1/2$ competitive ratio also demands $\Theta(n)$ offline samples (Rubinstein et al., 2019). Such requirements limit the applicability of these results in real-world scenarios.

In this work, we study the prophet inequality in a novel and practical setting, in which at each stage only a *noisy* realization of the random variable is observed, and reward distributions are *unknown* without available offline reward samples. Instead, the decision-maker has access to observable feature vectors drawn from distributions, and the rewards follow a linear model with an unknown latent parameter. This structural information enables estimation of the reward distribution and fundamentally distinguishes our setting from the classical unknown-distribution model (Correa et al., 2019a).

This feature-based formulation is motivated by applications such as online advertising, hiring, and recommendation systems, where contextual information (e.g., ad profiles, candidate attributes, or item descriptions) and noisy feedback are observable, while the underlying reward distributions remain unknown.

To address these challenges, we integrate learning and decision-making under noisy reward observations and feature information. Furthermore, we employ a lower-confidence-bound (LCB) thresholding strategy to handle the uncertainty in the estimator. The main contributions are as follows:

Summary of Contributions.

- Motivated by practical scenarios, we introduce a novel setting of the prophet inequality where the gambler only observes noisy rewards together with feature information and reward distributions are unknown.
- In the i.i.d. case, we propose learning-decision algorithms that integrate lower-confidence-bound (LCB) thresholding, achieving the sharp competitive ratio of $1 - 1/e$ against the optimal benchmark.
- For the non-identical case, we analyze an algorithm that attains a competitive ratio of $1/2$ against a relaxed benchmark. Furthermore, with window access to past rewards, the algorithm achieves the optimal competitive ratio of $1/2$ against the optimal benchmark.
- We validate our algorithms through experiments on synthetic datasets.

2 RELATED WORK

Prophet Inequalities under Known Reward Distributions. The study of prophet inequalities originates from Krengel & Sucheston (1977; 1978). A key milestone was established by Samuel-Cahn (1984), who showed that a single-threshold strategy achieves the optimal competitive ratio of $1/2$ in the case of independent but non-identical distributions. In the order-selection variant, where the gambler can choose the order of arrivals, Chawla et al. (2010) achieved a ratio of $1-1/e$. For the i.i.d. case, Hill & Kertz (1982) established a ratio of $1 - 1/e$, which was subsequently improved by Abolhassani et al. (2017) and Correa et al. (2017). Extending beyond exact observations, Assaf et al. (1998) demonstrated that analogous guarantees remain valid under noisy observations, though only with respect to a Bayesian version of the prophet benchmark, which is weaker than the classical one. Indeed, under noisy observations, any non-trivial guarantee with respect to the classical benchmark becomes impossible without additional structural assumptions, as we will show later. Finally, all of these results assume full knowledge of the underlying reward distributions—an assumption rarely satisfied in practical applications.

Prophet Inequalities under Unknown Reward Distributions. To address this limitation, recent work has studied prophet inequalities under unknown reward distributions (Correa et al., 2019a; 2020; Goldenshluger & Zeevi, 2022; Immorlica et al., 2023; Gatmiry et al., 2024; Li et al., 2022). For the i.i.d. setting, Correa et al. (2019a) showed that a competitive ratio of $1/e(\approx 0.368)$ can be achieved by the classical optimal algorithm for the secretary problem as the horizon grows. To obtain the higher ratio of $1-1/e(\approx 0.632)$, however, $\Theta(n)$ additional offline reward samples are required. Building on this, Goldenshluger & Zeevi (2022) showed that an asymptotic ratio approaching 1 is attainable, but only for fixed distributions whose maxima lie in the Gumbel or reverse-Weibull domains of attraction as the horizon grows.

The case of unknown non-identical distributions has been examined in Kaplan et al. (2020); Rubinstein et al. (2019); Gatmiry et al. (2024); Liu et al. (2025). However, achieving a $1/2$ competitive ratio requires $\Theta(n)$ offline samples. Although Gatmiry et al. (2024) and Liu et al. (2025) avoid offline samples, their setting involves repeated sequences of rounds rather than a single sequence. This repetition allows information to be aggregated across rounds, making the learning problem tractable under bandit feedback. In contrast, our setting involves only a single sequence, and is therefore fundamentally different. Prophet inequalities under unknown and non-independent distributions were also studied in Immorlica et al. (2023), achieving a ratio of $1/(2er)$ for r-sparse correlated structures, but their model still assumes distributional knowledge of the independent components of the rewards.

In contrast, we study a novel and practical setting that targets the optimal prophet under noisy reward observations and unknown reward distributions without available offline reward samples. Instead, we exploit observable feature vectors and their distribution, a setting motivated by real-world applications where feature information are available but the reward distribution is unknown.

3 Problem Statement

We consider n non-negative random variables (or rewards) $X_1, \dots, X_n$, where each X_i is independently drawn from an *unknown* distribution $\mathcal{D}_i$. In particular, we assume that

$$X_i = x_i^\top \theta, \quad i \in [n],$$

where $x_i \in \mathbb{R}^d$ is a feature vector drawn independently from a known distribution $\mathcal{D}_{x,i}$, and $\theta \in \mathbb{R}^d$ is an *unknown* latent parameter. Since θ is unknown, the induced distributions $\mathcal{D}_i$ of the X_i are also unknown to the gambler.

At each stage i, the gambler does not observe X_i directly. Instead, it observes a *noisy* measurement

$$y_i = X_i + \eta_i,$$

where the noise η_i is i.i.d drawn from a σ-sub-Gaussian distribution for $\sigma > 0$. The noisy observations $y_1, y_2, \dots$ are revealed sequentially.

After observing y_i and x_i at stage i, the gambler must make an irrevocable decision on whether to accept index i (and stop) or continue to the next stage. We denote by $\tau \in [n+1]$ the stopping time at which the gambler accepts an index, with $\tau = n+1$ meaning that the gambler rejects all variables. For completeness, we allow X_{n+1} to be any non-negative value, so that our analysis applies uniformly in this case.

The gambler's expected payoff is $\mathbb{E}[X_\tau]$. As a benchmark, we consider the prophet—an omniscient decision maker who knows all values $X_1, \dots, X_n$ in advance—which achieves $\mathbb{E}\left[\max_{i\in[n]} X_i\right]$. The goal of the gambler is to maximize the *asymptotic competitive ratio* against the prophet, defined:

$$\lim_{n\to\infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[n]} X_i]}.$$

Notation. For a square matrix M, $\lambda_{\min}(M)$ denotes its minimum eigenvalue.

We consider regularization conditions as follows.

Assumption 3.1. *There exists $S > 0$ such that $\|\theta\|_2^2 \le S$.*

Assumption 3.2. *There exists $L > 0$ such that, for all $i \in [n]$ and $x \sim \mathcal{D}_{x,i}$, $\|x_i\|_2^2 \le L$*

Remark 3.3. *Our regularization assumptions are standard in the online linear learning literature (Abbasi-Yadkori et al., 2011; Ruan et al., 2021; Liu et al., 2025). We emphasize that in our setting L may depend on n and can diverge as $n \to \infty$; this point will be revisited later.*

4 The i.i.d. Setting

Here, we focus on the case where all reward distributions are identical, i.e., $\mathcal{D}_i = \mathcal{D}$ for every $i \in [n]$. This holds, for instance, when the feature distributions are identical across stages, i.e., $\mathcal{D}_{x,i} = \mathcal{D}_x$ for $i \in [n]$. Under this setting, we propose algorithms and analyze their competitive ratios.

4.1 Explore-then-Decide with LCB Thresholding

We first propose an algorithm (Algorithm 1) based on Explore-then-Decide with lower confidence bound (LCB) thresholding. To address the unknown distribution $\mathcal{D}$, the algorithm begins with an exploration phase of length l_n, provided as an input. Afterward, during the decision phase, it computes an LCB for the reward and applies an LCB-based thresholding rule to decide at each stage whether to stop or continue. The details of this procedure are described below.

Algorithm 1 Explore-Then-Decide with LCB Thresholding (ETD-LCBT)

Input: Exploration length l_n; regularization parameter β
Output: Stopping time τ

for $i = 1, \dots, n$ **do**
  if $i \le l_n$ **then**
    Observe (y_i, x_i)
    if $i = l_n$ **then**
      $V \leftarrow \sum_{t=1}^{l_n} x_t x_t^\top + \beta I_d$; $\hat{\theta} \leftarrow V^{-1} \sum_{t=1}^{l_n} y_t x_t$
      Compute α from (2) (or (5) for non-i.i.d.)
  else
    Observe (y_i, x_i)
    Compute X_i^{LCB} from (1)
    if $X_i^{LCB} \ge \alpha$ **then**
      Stop and set $\tau \leftarrow i$

4.1.1 Strategy

Exploration. With setting $l_n = o(n)$, during the first l_n stages, we collect pairs of noisy rewards y_t and features x_t at each stage t. Using these observations, we estimate the unknown parameter θ as $\hat{\theta} = V^{-1} \sum_{t=1}^{l_n} y_t x_t$, where $V = \sum_{t=1}^{l_n} x_t x_t^\top + \beta I_d$ for a constant $\beta > 0$.

After this exploration phase, the algorithm enters the decision phase, where it determines at each stage whether to stop or continue. The details regarding LCB Thresholding are given below.

Lower Confidence Bound (LCB). We define the lower confidence bound for X_i as

$$X_i^{LCB} = x_i^\top \hat{\theta} - \xi(x_i), \tag{1}$$

where $\xi(x_i) := \sqrt{x_i^\top V^{-1} x_i}(\sigma \sqrt{d \log(n + n l_n L / d\beta)} + \sqrt{S\beta})$.

Decision with LCB Threshold. Using the CDF of $\mathbb{P}_{z \sim \mathcal{D}_x}(Z^{LCB} \le \alpha | \hat{\theta}, V)$ where $Z^{LCB} = z^\top \hat{\theta} - \xi(z)$, we set threshold α s.t.

$$\mathbb{P}_{z \sim \mathcal{D}_x}(Z^{LCB} \le \alpha | \hat{\theta}, V) = 1 - \frac{1}{n} \tag{2}$$

The algorithm stops at stage $i > l_n$ if $X_i^{LCB} \ge \alpha$, in which case we set $\tau = i$. By definition, if no stopping occurs throughout the horizon, we set $\tau = n + 1$.

4.1.2 Theoretical Analysis

Now we provide theoretical analyses. In this setting, a fundamental difficulty emerges due to noisy observations. In fact, it is possible to construct instances where the observation noise drives the competitive ratio to a trivial limit, as formalized below (see Appendix A.1 for the proof).

Proposition 4.1. *There exists a bounded i.i.d. distribution for $(X_i)_{i=1}^n$ together with an observation noise model such that, for any (possibly randomized) algorithm τ based on the observations,* $\lim_{n \to \infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i \in [n]} X_i]} = 0$.

The trivial outcome in Proposition 4.1 explains why Assaf et al. (1998) studied a Bayesian version of the prophet inequality rather than the classical one ($\mathbb{E}[\max_{i \in [n]} X_i]$) under the noisy observation. As Proposition 4.1 shows, even with full knowledge of the reward distribution, no algorithm can avoid this collapse to a trivial competitive ratio. To overcome this fundamental challenge—both in targeting the classical prophet under noisy observation and in the presence of an unknown latent parameter in the reward distribution—we later impose a mild non-degeneracy condition on reward scaling.

For notational convenience, let $\lambda = \lambda_{\min}\big(\mathbb{E}_{x \sim \mathcal{D}_x}[x x^\top]\big)$, the minimum eigenvalue of the covariance matrix of the feature distribution. Without loss of generality, we restrict attention to the case $\lambda > 0$, ensuring non-degeneracy of the feature covariance. Under this notation, we can now state our main guarantee on the competitive ratio (see Appendix A.2 for the proof).

Theorem 4.2. *Algorithm 1 with $l_n = o(n)$, $l_n = \omega(\frac{L \log d}{\lambda})$, and a constant $\beta > 0$, achieves an asymptotic competitive ratio of*

$$\lim_{n\to\infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[n]} X_i]} \geq 1 - \frac{1}{e} - \mathcal{O}\left(\limsup_{n\to\infty} \frac{1}{\mathbb{E}[\max_{i\in[n]} X_i]} \sqrt{\frac{Ld(\sigma^2 d + S)\log(Ln)}{\lambda l_n}} \right).$$

This result highlights the critical role of the optimal value $OPT = \mathbb{E}[\max_{i\in[n]} X_i]$ in determining the competitive ratio under noisy learning. As shown in Proposition 4.1, without further structural assumptions, the competitive ratio can collapse to zero. To circumvent this issue, we impose a non-degeneracy condition on reward scaling, specifically on the growth of OPT, which ensures learnability under noise and allows us to recover the sharp bound established in Theorem 4.2.

Corollary 4.3. *We set $l_n = \frac{Ld(\sigma^2 d+S)}{\lambda} f(n)\log(Ln)$ for some function $f(n)$ (e.g., $f(n) = \Theta(\log^p n)$ for $p > 0$, or $\Theta(n^q)$ for $0 < q < 1$) satisfying $l_n = o(n)$. If $OPT = \omega(1/\sqrt{f(n)})$, then Algorithm 1 achieves an asymptotic competitive ratio of*

$$\lim_{n\to\infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[n]} X_i]} \geq 1 - \frac{1}{e}.$$

The growth condition of $OPT = \omega(1/\sqrt{f(n)})$ in Corollary 4.3 is mild in practice. For example, by using $f(n) = n^{2/3}$ for setting l_n, the requirement is satisfied in most applications since OPT typically remains bounded away from zero. In particular, it suffices that $OPT \geq C$ for some constant $C > 0$ and all sufficiently large n.

Our competitive ratio of $1 - 1/e$ matches that of Hill & Kertz (1982) in the known i.i.d. setting and that of Correa et al. (2019a) in the unknown i.i.d. setting but with $\Theta(n)$ additional offline reward samples. Without such samples, only a $1/e$ ratio can be guaranteed (Correa et al., 2019a), which is strictly weaker than our result. Moreover, because rewards in our setting are observed only through noisy realizations, these prior guarantees no longer apply. Also note that $1 - 1/e$ is the best possible competitive ratio attainable by single quantile-threshold policies (Correa et al., 2019b; 2017), and our algorithm provides a close approximation to this class of policies.

Remark 4.4. *Importantly, while Correa et al. (2019a) show that $1/e$ is optimal for unknown distributions without sufficiently many offline reward samples of $\Omega(n)$, we demonstrate that by exploiting feature information under structural assumptions, the sharp bound of $1-1/e$ can in fact be achieved. Moreover, our analysis accommodates distributions whose support grows with the horizon n (e.g., $L = \sqrt{n}$ when setting $f(n) = \log n$ in Corollary 4.3), so that both the support and the variance of $\mathcal{D}$ may diverge as $n \to \infty$. This highlights that our framework is not restricted to the fixed distributional domains considered in Goldenshluger & Zeevi (2022), but instead applies more broadly to settings where distributions may evolve with the horizon.*

4.2 ε-GREEDY WITH LCB THRESHOLDING

While the Explore-then-Decide method achieves a sharp competitive ratio, its deterministic separation between exploration and decision phases—and the fact that exploration is confined to the early stages—limits its practicality in applications where exploration spread across time is preferable, such as online advertising or sequential recommendation systems. To address this, we propose an ε-Greedy approach (Algorithm 2) that selects decision stages uniformly at random over the time horizon. The details of the strategy are described as follows.

Randomized Exploration. At each stage $i \in [n]$, we draw a Bernoulli random variable $b_i \sim \text{Bernoulli}(\varepsilon)$, where $\varepsilon = \sqrt{l_n/n}$ with setting $l_n = o(n)$.

- If $b_i = 1$, we perform exploration by observing the noisy reward y_i and feature x_i, and update, $\hat{\theta}_i = V_i^{-1} \sum_{t\in\mathcal{I}_i} y_t x_t$, where $V_i = \sum_{t\in\mathcal{I}_i} x_t x_t^\top + \beta I_d$ for a constant $\beta > 0$.
- If $b_i = 0$, we enter the decision phase and determine whether to stop based on an dynamic threshold.

Algorithm 2 ε-Greedy with LCB Thresholding (ε-Greedy-LCBT)

Input: Bernoulli parameter ε; regularization parameter β
Output: Stopping time τ
for $i = 1, \ldots, n$ **do**
  Sample $b_i \sim \text{Bernoulli}(\varepsilon)$
  if $b_i = 1$ **then**
    $\mathcal{I}_i \leftarrow \mathcal{I}_{i-1} \cup \{i\}$
    Observe (x_i, y_i)
    $V_i \leftarrow \sum_{t \in \mathcal{I}_i} x_t x_t^\top + \beta I_d$; $\hat{\theta}_i \leftarrow V_i^{-1} \sum_{t \in \mathcal{I}_i} y_t x_t$
  else
    $\mathcal{I}_i \leftarrow \mathcal{I}_{i-1}, \hat{\theta}_i \leftarrow \hat{\theta}_{i-1}, V_i \leftarrow V_{i-1}$
    Observe (x_i, y_i)
    Compute X_i^{LCB} from (3) and α_i using (4)
    if $X_i^{LCB} \geq \alpha_i$ **then**
      Stop with $\tau \leftarrow i$

Unlike the Explore-then-Decide method, here the exploration rounds are distributed over the entire horizon. Consequently, $\hat{\theta}_i$ and V_i are updated continuously, which in turn affects both the LCB and the threshold dynamically, described below.

Lower Confidence Bound. We redefine the LCB for $X_i = x_i^\top \theta$ as

$$X_i^{LCB} = x_i^\top \hat{\theta}_i - \xi_i(x_i), \tag{3}$$

where $\xi_i(x_i) := \sqrt{x_i^\top V_i^{-1} x_i}\left(\sigma\sqrt{d \log(n + n|\mathcal{I}_i|L/d\beta)} + \sqrt{S\beta}\right)$.

Dynamic Threshold. Using a CDF of $\mathbb{P}_{z\sim\mathcal{D}_x}(Z_i^{LCB} \leq \alpha \mid \hat{\theta}_i, V_i, \mathcal{I}_i)$ where $Z_i^{LCB} = z^\top \hat{\theta}_i - \xi_i(z)$, for each $i \in [n]$, we set the dynamic threshold α_i such that

$$\mathbb{P}_{z\sim\mathcal{D}_x}(Z_i^{LCB} \leq \alpha_i \mid \hat{\theta}_i, V_i, \mathcal{I}_i) = 1 - \frac{1}{n}. \tag{4}$$

The algorithm stops at stage i if $X_i^{LCB} \geq \alpha_i$, in which case we set $\tau = i$. Unlike Explore-then-Decide, this procedure employs a dynamic threshold. By definition, if no stopping occurs over the entire horizon, we set $\tau = n + 1$. Recall $\lambda = \lambda_{\min}\big(\mathbb{E}_{x\sim\mathcal{D}_x}[xx^\top]\big)$. Then, the algorithm satisfies the following theorem (see Appendix A.3 for the proof).

Theorem 4.5. *Algorithm 2 with $\varepsilon = \sqrt{l_n/n}$, $l_n = o(n)$, $l_n = \Omega(\frac{L \log d \log n}{\lambda})$, and a constant $\beta > 0$, achieves an asymptotic competitive ratio of*

$$\lim_{n\to\infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[n]} X_i]} \geq 1 - \frac{1}{e} - \mathcal{O}\Big(\limsup_{n\to\infty} \frac{1}{\mathbb{E}[\max_{i\in[n]} X_i]}\sqrt{\frac{Ld(\sigma^2 d + S)\log(Ln)}{\lambda l_n}}\Big).$$

Furthermore, by setting $l_n = \frac{Ld(\sigma^2 d+S)}{\lambda} f(n)\log(Ln)$ for some function $f(n)$ (e.g., $f(n) = \Theta(\log^p n)$ for $p > 0$, or $\Theta(n^q)$ for $0 < q < 1$) satisfying $l_n = o(n)$, if $OPT = \omega(1/\sqrt{f(n)})$, then Algorithm 2 achieves the asymptotic ratio

$$\lim_{n\to\infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[n]} X_i]} \quad \geq \quad 1 - \frac{1}{e}.$$

Notably, the ε-Greedy approach achieves the same competitive ratio as established for the Explore-then-Decide method in Corollary 4.3, while ensuring uniformly random decision stages.

5 Non-Identical Distributions

In this section, we consider the setting where the distributions $\mathcal{D}_i$ are not identical across $i \in [n]$. In what follows, we propose algorithms and analyze their competitive ratios.

5.1 Explore-then-Decide with LCB Thresholding

We build on the Explore-then-Decide framework in Algorithm 1, adapting the thresholding policy accordingly. In the initial exploration phase of length l_n, we collect data and estimate $\hat{\theta} = V^{-1}\sum_{t=1}^{l_n} y_t x_t$, where $V = \sum_{t=1}^{l_n} x_t x_t^\top + \beta I$ for a constant $\beta > 0$. In the subsequent decision phase, we apply LCB-based thresholding for non-identical distributions, as described below.

Decision with LCB Threshold. For each time $i > l_n$, for $z_s \sim \mathcal{D}_{x,s}$ for all $s \in [l_n+1, n]$, we define the threshold:

$$\alpha = \frac{1}{2}\mathbb{E}\left[\max_{s\in[l_n+1,n]} z_s^\top \hat{\theta} \mid \hat{\theta}\right] \tag{5}$$

Recall the lower confidence bound for X_i in the Explore-then-Deicide framework: $X_i^{LCB} = x_i^\top \hat{\theta} - \xi(x_i)$, where $\xi(x_i) := \sqrt{x_i^\top V^{-1} x_i}(\sigma\sqrt{d\log(n + nl_nL/d\beta)} + \sqrt{S\beta})$. The algorithm stops at stage i if $X_i^{LCB} \geq \alpha$.

For notational convenience, let $\lambda' = \min_{i\in[n]} \lambda_{\min}\left(\mathbb{E}_{x\sim\mathcal{D}_{x,i}}[xx^\top]\right)$. Then, the algorithm satisfies with the following theorem (see Appendix A.5 for the proof).

Theorem 5.1. *Consider Algorithm 1 with $l_n = o(n)$, $l_n = \omega(\frac{L\log d}{\lambda'})$, and a constant $\beta > 0$, where the threshold value is chosen according to* (5)*. Then the algorithm achieves the following asymptotic competitive ratio:*

$$\lim_{n\to\infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[l_n+1,n]} X_i]} \geq \frac{1}{2} - \mathcal{O}\left(\limsup_{n\to\infty} \frac{1}{\mathbb{E}[\max_{i\in[l_n+1,n]} X_i]}\sqrt{\frac{(\sigma^2 d + S)\log(Ln)}{\lambda' l_n}}\right).$$

Furthermore, by setting $l_n = \frac{L(\sigma^2 d+S)}{\lambda} f(n)\log(Ln)$ for some function $f(n)$ (e.g., $f(n) = \Theta(\log^p n)$ for $p > 0$, or $\Theta(n^q)$ for $0 < q < 1$) satisfying $l_n = o(n)$, if $OPT = \omega(1/\sqrt{f(n)})$, then Algorithm 1 with threshold (5) achieves the following asymptotic competitive ratio:

$$\lim_{n\to\infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[l_n+1,n]} X_i]} \geq \frac{1}{2}.$$

In the theorem, we target the relaxed prophet of $\mathbb{E}[\max_{i\in[l_n+1,n]} X_i]$ due to the inherent difficulty of the non-i.i.d. setting against the original prophet, as shown in Proposition 5.2 (see Appendix A.4 for the proof).

Proposition 5.2. *There exist non-identical distributions $\{\mathcal{D}_{x,i}\}_{i=1}^n$ for the feature vectors x_i's, and a parameter vector θ, such that when observing noise-free rewards $X_i = x_i^\top\theta$ for $i \in [n]$, the following holds: for any stopping rule τ, $\mathbb{E}[X_\tau]/\mathbb{E}\left[\max_{i\in[n]} X_i\right] \leq \min\{\frac{1}{2}, \frac{1}{d}\}$. Furthermore, there exists $\{\mathcal{D}_{x,i}\}_{i=1}^n$ and θ such that, for any stopping rule τ, $\lim_{n\to\infty} \mathbb{E}[X_\tau]/\mathbb{E}\left[\max_{i\in[d+1,n]} X_i\right] \leq 1/2$.*

Proposition 5.2 shows that, even in the noise-free case ($\sigma = 0$), the initial stages must be sacrificed to learn θ. For the prophet of $\mathbb{E}[\max_{i\in[n]} X_i]$, the competitive ratio approaches zero with large enough d (e.g. $d = log(n)$). For the relaxed prophet of $\mathbb{E}[\max_{i\in[d+1,n]} X_i]$, the upper bound becomes non-trivially $1/2$. The noise enhances this effect. In our setting with noise, the first l_n observations are necessarily reserved for learning and are thus excluded from the stopping decision. This motivates our focus on a relaxed prophet benchmark based on $\mathbb{E}[\max_{i\in[l_n+1,n]} X_i]$, which allows for non-trivial guarantees.

Furthermore, based on Proposition 4.1, noisy observations also lead to trivial outcomes in the case of non-identical distributions without any structural assumptions. To address this, we impose a mild non-degeneracy condition on reward scaling—specifically on the growth of OPT—which allows us to recover the sharp bound stated in Theorem 5.1.

The optimal competitive ratio for non-identical distributions is known to be $1/2$ (Samuel-Cahn, 1984). In our setting with unknown distributions, Theorem 5.1 shows that attaining this ratio requires relaxing the prophet benchmark by excluding the initial exploration phase. Equivalently, if l_n offline reward samples with features were available, the original optimal prophet benchmark could be targeted while still achieving the $1/2$ ratio. In the next subsection, we present another practical condition under which the optimal prophet benchmark can be attained in our learning setting without relying on additional offline reward samples.

5.2 Explore-Then-Decide with Window Access

In the standard non-identical distribution setting, items are revealed sequentially, and the gambler must decide immediately whether to accept or reject the *current* observation. As discussed in Marshall et al. (2020); Benomar et al. (2024), this assumption, however, can be overly pessimistic: in many practical scenarios, early opportunities are not irrevocably lost but may remain available for a short period of time. For instance, in a hiring process, one may be able to interview several candidates sequentially before making a final choice among them.

Window Access. Motivated by this observation, we consider a mild relaxation of the standard setting by using window access for the previous time steps, same as Marshall et al. (2020). More specifically, for a window size of w_n, at time i, the decision-maker is allowed to choose among the first w_n values $\{X_{i-w_n+1}, \ldots, X_i\}$ before deciding whether to continue. Interestingly, for $w_n \leq n-1$, the optimal competitive ratio in the non-i.i.d. distributions is the same with the standard setting (i.e. window size 1) as shown in the following (see Appendix A.6 for the proof).

Proposition 5.3. *In the non-i.i.d. setting with window access of size $w_n \leq n-1$, for any algorithm, there always exist non-identical distributions such that the competitive ratio is bounded above by $\mathbb{E}[X_\tau]/\mathbb{E}[\max_{i\in[n]} X_i] \leq 1/2$.*

These observations raise the following question: *Can the optimal competitive ratio under window access also be achieved in the setting of unknown non-identical distributions and noisy reward observations? If so, what window size w_n is required, and how frequently is window access required?*

To handle this setting, we propose an algorithm (Algorithm 3 in Appendix A.7) adopting the Explore-then-Decide method. After the exploration phase, at time $l_n + 1$ the decision-maker, with window size $w_n = l_n + 1$, may select from the values $\{X_1, \ldots, X_{l_n+1}\}$ before deciding whether to continue. From this perspective, the early observations collected during exploration are no longer wasted but can be revisited together with the $(l_n + 1)$-st observation, thereby mitigating the inefficiency of pure exploration. Notably, our algorithm requires only a single window access at time $l_n + 1$, with window size $l_n + 1$. Due to space constraints, we defer the detailed description of the algorithm to Appendix A.7. In what follows, we provide a theorem for the competitive ratio of the method with window access (see Appendix A.8 for the proof).

Theorem 5.4. *In the non-i.i.d. setting with unknown distributions and window access of size $w_n > l_n$, Algorithm 3 with $l_n = o(n)$, $l_n = \omega(\frac{L\log d}{\lambda'})$, and a constant $\lambda > 0$ achieves the following asymptotic competitive ratio:*

$$\lim_{n\to\infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[n]} X_i]} \geq \frac{1}{2} - \mathcal{O}\left(\lim_{n\to\infty} \frac{1}{\mathbb{E}[\max_{i\in[n]} X_i]}\sqrt{\frac{L(\sigma^2 d + S)\log(Ln)}{\lambda' l_n}}\right).$$

Furthermore, by setting $l_n = (L(\sigma^2 d + S)/\lambda) f(n)\log(Ln)$ for some function $f(n)$ (e.g., $f(n) = \Theta(\log^p n)$ for $p > 0$, or $\Theta(n^q)$ for $0 < q < 1$) satisfying $l_n = o(n)$, if $OPT = \omega(1/\sqrt{f(n)})$, then Algorithm 3 achieves the following asymptotic competitive ratio:

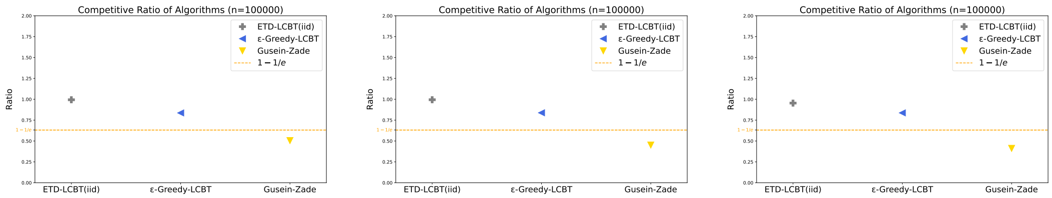
$$\lim_{n\to\infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[n]} X_i]} \geq \frac{1}{2}.$$

Notably, Algorithm 3 achieves the optimal ratio, matching the upper bound in Proposition 5.3.

Remark 5.5. *The guarantees in Theorem 5.4 can also be obtained by running LCB-Thresholding with l_n offline samples, without window access. In prior work, the non-i.i.d. setting requires $\Theta(n)$ offline samples to achieve a $1/2$ competitive ratio (Rubinstein et al., 2019). In contrast, exploiting the linear structure allows our method to achieve the optimal $1/2$ ratio using only l_n samples, which can be significantly smaller—for example, $l_n = O(\log^{p+1} n)$ when $f(n) = \Theta(\log^p n)$ for $p > 0$. Further details are provided in Appendix A.9.*

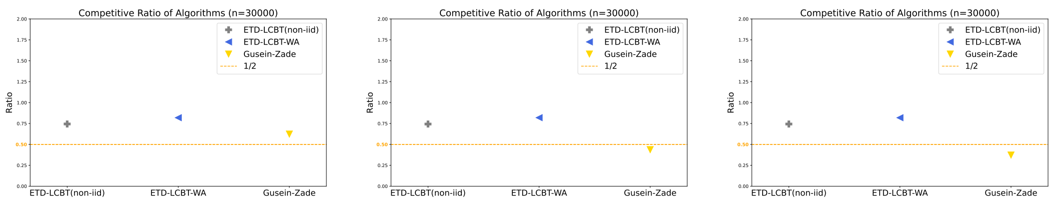
6 Experiments

In this section, we evaluate our algorithms on synthetic datasets. Gaussian noise with variance σ^2 is added to the rewards, and each experiment is repeated 10 times. We consider dimension $d = 2$ for the feature and latent parameter. For our algorithms, we set $l_n = n^{2/3}$ and $\beta = 1$. Since no existing algorithm directly applies to our setting with noisy rewards and without additional reward



(a) $\sigma = 0.1$ (b) $\sigma = 0.5$ (c) $\sigma = 0.8$

Figure 1: Competitive ratio under i.i.d. distribution setting with noise variance σ.



(a) $\sigma = 0.1$ (b) $\sigma = 0.5$ (c) $\sigma = 0.8$

Figure 2: Competitive ratios under non-identical distributions with noise variance σ.

samples under unknown distributions, we adopt the rule of `Gusein-Zade` (Gusein-Zade, 1966) as a benchmark. This rule observes the first n/e stages and then stops at the first record exceeding the maximum among these initial n/e values. Although it does not handle noisy rewards, it has been shown to extend to the prophet inequality with unknown i.i.d. distributions without additional reward samples. In particular, it guarantees a competitive ratio of $1/e$ in the worst case (Correa et al., 2019a), and can even achieve an asymptotic ratio of 1 under certain problem-specific distributions (Goldenshluger & Zeevi, 2022).

We first consider the i.i.d. setting with $n = 100000$, where θ and each x_i are drawn uniformly over each dimension and then normalized. Figure 1 shows that our algorithms of `ETD-LCBT(iid)` (Algorithm 1) and ε-`Greedy-LCBT` (Algorithm 2) achieve competitive ratios exceeding $1 - 1/e$, consistent with the theoretical guarantees in Corollary 4.3 and Theorem 4.5, and significantly outperform the benchmark of `Gusein-Zade`. Furthermore, as the noise variance increases, the performance gap between our algorithms and the benchmark becomes even larger, highlighting the robustness of our methods to noise.

Next, we consider the non-identical distribution setting with $n = 30000$, where θ is drawn uniformly over each dimension, but each x_i is drawn from a distinct distribution: the range of each dimension is randomly sampled, and each coordinate is then drawn uniformly within its range. Figure 2 demonstrates that `ETD-LCBT-WA` (Algorithm 3) achieves a competitive ratio exceeding $1/2$, consistent with the theoretical guarantee in Theorem 5.4. Even `ETD-LCBT(non-iid)` (Algorithm 1), which is guaranteed only against the relaxed benchmark (Theorem 5.1), empirically attains a ratio above $1/2$. As expected, `ETD-LCBT-WA` outperforms `ETD-LCBT(non-iid)` due to its access to the window. Notably, both algorithms outperform the benchmark of `Gusein-Zade`. Furthermore, as the noise variance increases, the performance gap between our algorithms and the benchmark becomes even larger, highlighting the robustness of our methods to noise.

7 Conclusion

We introduced a new framework for prophet inequalities under noisy observations and unknown reward distributions, motivated by real-world applications where noisy reward and contextual information are observable but reward distributions are not. By combining learning with LCB-based stopping rules, we achieved the sharp competitive ratio of $1 - 1/e$ in the i.i.d. setting. For non-identical distributions, we showed that the optimal bound of $1/2$ can be attained under window access. Our empirical results demonstrate the efficiency of our algorithms. (**Future Directions**) Several directions remain open for future work, including extensions to correlated rewards and applications to richer contextual models beyond linear structure. We believe that bridging prophet inequalities with modern online learning techniques will continue to uncover new insights at the interface of optimal stopping, learning, and decision-making.

REPRODUCIBILITY STATEMENT

All theoretical claims are stated with explicit assumptions and are accompanied by complete proofs in the appendix. Algorithmic details, including pseudocode, are provided in the main paper and supplementary materials. For the experimental results, we describe the data generation process in the main, and we attach source code for reproducing all figures and numerical results as part of the supplementary material.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Melika Abolhassani, Soheil Ehsani, Hossein Esfandiari, MohammadTaghi Hajiaghayi, Robert Kleinberg, and Brendan Lucier. Beating 1-1/e for ordered prophets. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 61–71, 2017.

Saeed Alaei, MohammadTaghi Hajiaghayi, and Vahid Liaghat. Online prophet-inequality matching with applications to ad allocation. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 18–35, 2012.

Makis Arsenis and Robert Kleinberg. Individual fairness in prophet inequalities. *arXiv preprint arXiv:2205.10302*, 2022.

David Assaf, Larry Goldstein, and Ester Samuel-Cahn. A statistical version of prophet inequalities. *The Annals of Statistics*, 26(3):1190–1197, 1998.

Ziyad Benomar, Dorian Baudry, and Vianney Perchet. Lookback prophet inequalities. *Advances in Neural Information Processing Systems*, 37:42123–42161, 2024.

Shuchi Chawla, Jason D Hartline, David L Malec, and Balasubramanian Sivan. Multi-parameter mechanism design and sequential posted pricing. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pp. 311–320, 2010.

José Correa, Patricio Foncea, Ruben Hoeksma, Tim Oosterwijk, and Tjark Vredeveld. Posted price mechanisms for a random stream of customers. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pp. 169–186, 2017.

José Correa, Paul Dütting, Felix Fischer, and Kevin Schewior. Prophet inequalities for iid random variables from an unknown distribution. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pp. 3–17, 2019a.

Jose Correa, Patricio Foncea, Ruben Hoeksma, Tim Oosterwijk, and Tjark Vredeveld. Recent developments in prophet inequalities. *ACM SIGecom Exchanges*, 17(1):61–70, 2019b.

José Correa, Paul Dütting, Felix Fischer, Kevin Schewior, and Bruno Ziliotto. Unknown iid prophets: Better bounds, streaming algorithms, and a new impossibility. *arXiv preprint arXiv:2007.06110*, 2020.

Khashayar Gatmiry, Thomas Kesselheim, Sahil Singla, and Yifan Wang. Bandit algorithms for prophet inequality and pandora's box. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 462–500. SIAM, 2024.

Alexander Goldenshluger and Assaf Zeevi. Optimal stopping of a random sequence with unknown distribution. *Mathematics of Operations Research*, 47(1):29–49, 2022.

SM Gusein-Zade. The problem of choice and the sptimal stopping rule for a sequence of independent trials. *Theory of Probability & Its Applications*, 11(3):472–476, 1966.

Theodore P Hill and Robert P Kertz. Comparisons of stop rule and supremum expectations of iid random variables. *The Annals of Probability*, pp. 336–345, 1982.

Theodore P Hill and Robert P Kertz. A survey of prophet inequalities in optimal stopping theory. *Contemporary Mathematics*, 125(1):191, 1992.

Nicole Immorlica, Sahil Singla, and Bo Waggoner. Prophet inequalities with linear correlations and augmentations. *ACM Transactions on Economics and Computation*, 11(3-4):1–29, 2023.

Haim Kaplan, David Naori, and Danny Raz. Competitive analysis with a sample and the secretary problem. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2082–2095. SIAM, 2020.

Ulrich Krengel and Louis Sucheston. Semiamarts and finite values. 1977.

Ulrich Krengel and Louis Sucheston. On semiamarts, amarts, and processes with finite value. *Probability on Banach spaces*, 4(197-266):1–2, 1978.

Bo Li, Xiaowei Wu, and Yutong Wu. Prophet inequality on iid distributions: beating 1-1/e with a single query. *arXiv preprint arXiv:2205.05519*, 2022.

Junyan Liu, Ziyun Chen, Kun Wang, Haipeng Luo, and Lillian J Ratliff. Improved regret and contextual linear extension for pandora's box and prophet inequality. *arXiv preprint arXiv:2505.18828*, 2025.

Brendan Lucier. An economic view of prophet inequalities. *ACM SIGecom Exchanges*, 16(1):24–47, 2017.

William Marshall, Nolan Miranda, and Albert Zuo. Windowed prophet inequalities. *arXiv preprint arXiv:2011.14929*, 2020.

Yufei Ruan, Jiaqi Yang, and Yuan Zhou. Linear bandits with limited adaptivity and learning distributional optimal design. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 74–87, 2021.

Aviad Rubinstein, Jack Z Wang, and S Matthew Weinberg. Optimal single-choice prophet inequalities from samples. *arXiv preprint arXiv:1911.07945*, 2019.

Ester Samuel-Cahn. Comparison of threshold stop rules and maximum for independent nonnegative random variables. *the Annals of Probability*, pp. 1213–1216, 1984.

Joel A Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

A APPENDIX

A.1 PROOF OF PROPOSITION 4.1

To show this proposition, we follow the example in Assaf et al. (1998). Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli with success probability $p_n = c/n$ for some fixed $c \in (0, \infty)$. Let the observations be obtained through a symmetric flip-noise channel:

$$Z_i = \begin{cases} X_i, & \text{with probability } 1/2, \\ 1 - X_i, & \text{with probability } 1/2, \end{cases}$$

independently across i and independently of $(X_i)_{i=1}^n$. Then $Z_i \perp X_i$ and in fact $Z_i \sim \text{Bernoulli}(1/2)$ regardless of p_n.

Let τ be any (possibly randomized) index valued in $\{1, \ldots, n\}$ that is measurable with respect to $(Z_1, \ldots, Z_n)$. Write $P_i(Z) := \Pr(\tau = i \mid Z)$ where $Z = (Z_1, \ldots, Z_n)$. By independence and $\mathbb{E}[X_i \mid Z] = \mathbb{E}[X_i] = p_n$,

$$\mathbb{E}[X_\tau \mid Z] = \mathbb{E}[\sum_{i=1}^n X_i P_i(Z) \mid Z] = \sum_{i=1}^n \mathbb{E}[X_i \mid Z]\mathbb{E}[P_i(Z) \mid Z] = \sum_{i=1}^n p_n \mathbb{E}[P_i(Z) \mid Z] = p_n = \frac{c}{n},$$

hence $\mathbb{E}[X_\tau] = c/n$ for every algorithm τ.

On the other hand, the oracle that sees the true X's obtains the maximum $\max_i X_i$, which equals 1 iff at least one success occurs. Therefore

$$\mathbb{E}\left[\max_{1 \le i \le n} X_i\right] = 1 - (1 - p_n)^n = 1 - \left(1 - \frac{c}{n}\right)^n \xrightarrow[n \to \infty]{} 1 - e^{-c} \; > \; 0.$$

Combining the two displays,

$$\frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_i X_i]} = \frac{c/n}{1 - (1 - c/n)^n} \xrightarrow[n \to \infty]{} 0.$$

Since τ was arbitrary, the conclusion holds for any algorithm.

A.2 PROOF OF THEOREM 4.2

We first provide a lemma for estimation error.

Lemma A.1 (Theorem 2 in Abbasi-Yadkori et al. (2011))**.** *For $\delta > 0$, we have*

$$\mathbb{P}\left(\|\hat{\theta} - \theta\|_V \le \sqrt{S\beta} + \sigma \sqrt{d \log\left(\frac{1 + \sum_{i=1}^{l_n} \|x_i\|_2^2 / d\beta}{\delta} \right)} \right) \ge 1 - \delta.$$

Proof. This lemma follows from Theorem 2 in Abbasi-Yadkori et al. (2011), using the inequality $\det(V) \le (\text{Tr}(V)/d)^d = (\beta + \sum_{s=1}^{l_n} \|x_s\|_2^2/d)^d$, where $\text{Tr}(V)$ denotes the trace of V. ☐

The above lemma implies that

$$\mathbb{P}\left(\left| x^\top (\hat{\theta} - \theta) \right| \le \sqrt{x^\top V^{-1} x} \left(\sigma \sqrt{d \log(n + n l_n L / d\beta)} + \sqrt{S\beta} \right), \forall x \in \mathbb{R}^d \right) \ge 1 - 1/n. \quad (6)$$

We define an event $\mathcal{E}_1 = \{|x^\top (\hat{\theta} - \theta)| \le \sqrt{x^\top V^{-1} x}(\sigma \sqrt{d \log(n + n l_n L/d\beta)} + \sqrt{S\beta}), \forall x \in \mathbb{R}^d\}$, which holds with $\mathbb{P}(\mathcal{E}_1) \ge 1 - \frac{1}{n}$.

We define $g := \sqrt{L \|V^{-1}\|_2}(\sigma \sqrt{d \log(n + n l_n L/d\beta)} + \sqrt{S\beta})$. Recall

$$\xi(x_i) = \sqrt{x_i^\top V^{-1} x_i}(\sigma \sqrt{d \log(n + n l_n L/d\beta)} + \sqrt{S\beta}).$$

Then we have $\xi(x_i) \le g$. Under $\mathcal{E}_1$, for any $i > l_n$ we have

$$X_i - 2g \le x_i^\top \hat{\theta} - g \le x_i^\top \hat{\theta} - \xi(x_i) \le X_i^{LCB} \le X_i. \quad (7)$$

We define α^* s.t. $\mathbb{P}_{X \sim \mathcal{D}}(X \le \alpha^*) = 1 - \frac{1}{n}$. Then we have the following lemma regarding the bounds for the threshold value. We denote $\mathcal{H}_{l_n} = \{\hat{\theta}, V\}$.

Lemma A.2. *Under $\mathcal{E}_1$, for any given $\mathcal{H}_{l_n}$, we have*

$$\alpha^* - 2g \le \alpha \le \alpha^*.$$

Proof. For $z \sim \mathcal{D}_x$, we define $Z = z^\top \theta$ and $\hat{Z} = z^\top \hat{\theta}$. Then, under $\mathcal{E}_1$, for any given $\mathcal{H}_{l_n}$, we have $Z - 2g \le \hat{Z} - g \le \hat{Z} - \xi(z) \le Z$ with $\xi(z) \le g$.

Since $\hat{Z} - \xi(z) \le Z$ and $\mathbb{P}(\hat{Z} - \xi(z) \ge \alpha \mid \mathcal{H}_{l_n}) = \mathbb{P}(Z \ge \alpha^* \mid \mathcal{H}_{l_n}) = 1/n$, we can easily obtain

$$\alpha \le \alpha^*.$$

Likewise, for α' s.t. $\mathbb{P}(Z - 2g \ge \alpha' \mid \mathcal{H}_{l_n}) = \frac{1}{n}$, from $Z - 2g \le \hat{Z} - \xi(z)$ and $\mathbb{P}(Z - 2g \ge \alpha' \mid \mathcal{H}_{l_n}) = \mathbb{P}(\hat{Z} - \xi(z) \ge \alpha \mid \mathcal{H}_{l_n}) = 1/n$, we have $\alpha' \le \alpha$. Therefore, with $\alpha' + 2g = \alpha^*$ from $\mathbb{P}(Z \ge \alpha^* \mid \mathcal{H}_{l_n}) = \mathbb{P}(Z - 2g \ge \alpha' \mid \mathcal{H}_{l_n}) = 1/n$, we have

$$\alpha^* - 2g \le \alpha,$$

which concludes the proof. □

Lemma A.3. *For $l \ge 1$, let $z_1, \dots, z_l \overset{i.i.d.}{\sim} \mathcal{D}_x$ satisfying Assumption 3.2. Recall $\lambda = \lambda_{\min}(\mathbb{E}_{z \sim \mathcal{D}_x}[zz^\top]) > 0$. Then*

$$\mathbb{P}\left(\frac{1}{l}\sum_{s=1}^{l} z_s z_s^\top \succeq \frac{\lambda}{2} I_d\right) \ge 1 - d\exp\left(-\frac{\lambda l}{8L}\right).$$

Proof. Let $\mu_{\min} = \lambda_{\min}(\mathbb{E}[\sum_{s=1}^{l} z_s z_s^\top])$. By the matrix Chernoff bound (Theorem 5.1.1 in Tropp et al. (2015)) for sums of independent PSD matrices with Assumption 3.2, for any $\delta \in [0,1]$,

$$\Pr\left[\lambda_{\min}\left(\sum_{s=1}^{l} z_s z_s^\top\right) \le (1-\delta)\mu_{\min}\right] \;\le\; d\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu_{\min}/L} \;\le\; d\exp\Bigl(-\frac{\delta^2}{2}\cdot\frac{\mu_{\min}}{L}\Bigr).$$

Choosing $\delta = \frac{1}{2}$ yields

$$\Pr\left[\lambda_{\min}\left(\sum_{s=1}^{l} z_s z_s^\top\right) \le \frac{\mu_{\min}}{2}\right] \;\le\; d\exp\Bigl(-\frac{\mu_{\min}}{8L}\Bigr) \le d\exp\left(-\frac{l\lambda}{8L}\right),$$

where the last inequality is obtained from Weyl's eigenvalue inequalities. Equivalently, with probability at least $1 - d\exp(-\lambda l/(8L))$,

$$\sum_{s=1}^{l} z_s z_s^\top \succeq \frac{\mu_{\min}}{2} I_d \succeq \frac{l\lambda}{2} I_d,$$

which completes the proof. □

Let $\mathcal{E}_2 = \{\sum_{s=1}^{l_n} x_s x_s^\top \succeq \frac{\lambda l_n}{2} I_d\}$, which holds with probability at least $1 - \frac{d}{e^{\lambda l_n/8L}}$ from Lemma A.3. Then under $\mathcal{E}_2$, we have $\|V^{-1}\|_2 \le \|(\sum_{s=1}^{l_n} x_s x_s^\top)^{-1}\|_2 \le 2\frac{1}{\lambda l_n}$. Then, we have

$$\begin{aligned}
\xi(x_i) &\le \sqrt{L\|V^{-1}\|_2}(\sigma\sqrt{d\log(n + nl_nL/d\beta)} + S\sqrt{\beta})(= g)\\
&\le \sqrt{L\frac{2}{\lambda l_n}}(\sigma\sqrt{d\log(n + n^2L/d\beta)} + S\sqrt{\beta}).
\end{aligned}$$

Here we define $h := \sqrt{L\frac{2}{\lambda l_n}}(\sigma\sqrt{d\log(n+n^2L/d\beta)} + S\sqrt{\beta})$ and $\mathcal{E} := \mathcal{E}_1 \cup \mathcal{E}_2$. For analyzing X_τ, we first examine the probability that the stopping time τ equals i given $\mathcal{E}$. Recall $\mathcal{H}_{l_n} = \{\hat{\theta}, \{x_s\}_{s\in[l_n]}\}$.

Lemma A.4. *For $i > l_n$, we have*

$$\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n}) = \left(1 - \frac{1}{n}\right)^{i - l_n - 1}\frac{1}{n}.$$

Proof. For $i > l_n$, we have

$$\begin{aligned}
&\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n}) \\
&= \mathbb{P}(X_{l_n+1}^{LCB} \le \alpha, \ldots, X_{i-1}^{LCB} \le \alpha, X_i^{LCB} > \alpha \mid \mathcal{H}_{l_n}) \\
&= \mathbb{P}(X_{l_n+1}^{LCB} \le \alpha, \ldots, X_{i-1}^{LCB} \le \alpha \mid \mathcal{H}_{l_n})\mathbb{P}(X_i^{LCB} > \alpha \mid \mathcal{H}_{l_n}) \\
&= \mathbb{P}(X_{l_n+1}^{LCB} \le \alpha, \ldots, X_{i-1}^{LCB} \le \alpha \mid \mathcal{H}_{l_n})\frac{1}{n},
\end{aligned}$$

where the first equality is obtained from the fact that, given $\hat{\theta}$ and $\{x_s\}_{s\in[l_n]}$, X_i^{LCB} is independent to $X_{l_n+1}^{LCB}, \ldots, X_{i-1}^{LCB}$. Similarly, for the last term above, we have

$$\begin{aligned}
&\mathbb{P}(X_{l_n+1}^{LCB} \le \alpha, \ldots, X_{i-1}^{LCB} \le \alpha \mid \mathcal{H}_{l_n})\frac{1}{n} \\
&= \mathbb{P}(X_{l_n+1}^{LCB} \le \alpha, \ldots, X_{i-2}^{LCB} \le \alpha \mid \mathcal{H}_{l_n})\mathbb{P}(X_{i-1}^{LCB} \le \alpha \mid \mathcal{H}_{l_n})\frac{1}{n} \\
&= \mathbb{P}(X_{l_n+1}^{LCB} \le \alpha, \ldots, X_{i-2}^{LCB} \le \alpha \mid \mathcal{H}_{l_n})\left(1 - \frac{1}{n}\right)\frac{1}{n} \\
&\ \ \vdots \\
&= \left(1 - \frac{1}{n}\right)^{i-l_n-1}\frac{1}{n},
\end{aligned}$$

which concludes the proof. □

From the exploration phase in the algorithm, we have $\mathbb{P}(\tau = i \mid \mathcal{E}) = 0$ for all $1 \le i \le l_n$. Therefore, given $\mathcal{E}$ we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}[X_\tau \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}]\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n}\mathbb{P}(\tau = i \mid \{x_s\}_{s\in[l_n]})\mathbb{E}[X_i \mathbb{1}(\mathcal{E}) \mid \tau = i, \mathcal{H}_{l_n}]\right] \\
&= \mathbb{E}\left[\sum_{i=l_n+1}^{n}\mathbb{P}(\tau = i \mid \{x_s\}_{s\in[l_n]})\mathbb{E}[X_i \mathbb{1}(\mathcal{E}) \mid \tau = i, \mathcal{H}_{l_n}]\right] \\
&\ge \mathbb{E}\left[\sum_{i=l_n+1}^{n}\mathbb{P}(\tau = i \mid \{x_s\}_{s\in[l_n]})\mathbb{E}[X_i^{LCB} \mathbb{1}(\mathcal{E}) \mid \tau = i, \mathcal{H}_{l_n}]\right] \\
&\ge \mathbb{E}\left[\sum_{i=l_n+1}^{n}\left(1 - \frac{1}{n}\right)^{i-l_n-1}\frac{1}{n}\mathbb{E}[X_i^{LCB} \mathbb{1}(\mathcal{E}) \mid \tau = i, \mathcal{H}_{l_n}]\right].
\end{aligned} \tag{8}$$

For the last term above, we have

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i=l_n+1}^{n}\left(1-\frac{1}{n}\right)^{i-l_n-1}\frac{1}{n}\mathbb{E}[X_i^{LCB}\mathbb{1}(\mathcal{E})\mid \tau=i,\mathcal{H}_{l_n}]\right]\\
&=\mathbb{E}\left[\sum_{i=l_n+1}^{n}\left(1-\frac{1}{n}\right)^{i-l_n-1}\frac{1}{n}\times\left(\mathbb{E}\left[\alpha\mathbb{1}(\mathcal{E})\mid\tau=i,\mathcal{H}_{l_n}\right]+\mathbb{E}[(X_i^{LCB}-\alpha)\mathbb{1}(\mathcal{E})\mid X_i^{LCB}\geq\alpha,\mathcal{H}_{l_n}]\right)\right]\\
&=\mathbb{E}\left[\sum_{i=l_n+1}^{n}\left(1-\frac{1}{n}\right)^{i-l_n-1}\frac{1}{n}\times\left(\mathbb{E}\left[\alpha\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}\right]+\mathbb{E}[(X_i^{LCB}-\alpha)^+\mathbb{1}(\mathcal{E})\mid X_i^{LCB}\geq\alpha,\mathcal{H}_{l_n}]\right)\right]\\
&=\mathbb{E}\left[\sum_{i=l_n+1}^{n}\left(1-\frac{1}{n}\right)^{i-l_n-1}\frac{1}{n}\times\left(\mathbb{E}\left[\alpha\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}\right]+\frac{\mathbb{E}[(X_i^{LCB}-\alpha)^+\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}]}{\mathbb{P}(X_i^{LCB}\geq\alpha\mid\mathcal{H}_{l_n})}\right)\right]\\
&\geq\mathbb{E}\left[\sum_{i=l_n+1}^{n}\left(1-\frac{1}{n}\right)^{i-l_n-1}\frac{1}{n}\times\left(\mathbb{E}\left[\alpha^*\mathbb{1}(\mathcal{E})-2g\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}\right]+n\mathbb{E}\left[(X_i-2g-\alpha^*)^+\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}\right]\right)\right],
\end{aligned}
\tag{9}
$$

where the second inequality is obtained from independency between $\tau=i$ for $i>l_n$ and $\mathcal{E}$ given $\hat{\theta}$ and $\{x_s\}_{s\in[l_n]}$, and the last inequality is obtained from Lemma A.2 and (7). Let $Z_s\sim\mathcal{D}$ for $s\in[n]$. Then, with $l_n=o(n)$, for the last term in (9), we have

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i=l_n+1}^{n}\left(1-\frac{1}{n}\right)^{i-l_n-1}\frac{1}{n}\times\left(\mathbb{E}\left[(\alpha^*-2g)\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}\right]+n\mathbb{E}\left[(X_i-2g-\alpha^*)^+\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}\right]\right)\right]\\
&\geq\sum_{i=l_n+1}^{n}\left(1-\frac{1}{n}\right)^{i-l_n-1}\frac{1}{n}\times\mathbb{E}\left[\mathbb{E}[\alpha^*\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}]-2\mathbb{E}[h\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}]+n\mathbb{E}[(X_i-2h-\alpha^*)^+\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}]\right]\\
&\geq\sum_{i=l_n+1}^{n}\left(1-\frac{1}{n}\right)^{i-l_n-1}\frac{1}{n}\times\mathbb{E}\left[\mathbb{E}[\alpha^*\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}]-2\mathbb{E}[h\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}]+\mathbb{E}[\sum_{s\in[n]}(Z_s-2h-\alpha^*)^+\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}]\right]\\
&\geq\sum_{i=l_n+1}^{n}\left(1-\frac{1}{n}\right)^{i-l_n-1}\frac{1}{n}\mathbb{E}\left[\mathbb{E}[\alpha^*\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}]-2\mathbb{E}[h\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}]+\mathbb{E}[\max_{s\in[n]}(Z_s-2h-\alpha^*)^+\mathbb{1}(\mathcal{E})\mid\mathcal{H}_{l_n}]\right]\\
&\geq\sum_{i=l_n+1}^{n}\left(1-\frac{1}{n}\right)^{i-l_n-1}\frac{1}{n}\times\left(\alpha^*\mathbb{P}(\mathcal{E})-2h\mathbb{P}(\mathcal{E})+\mathbb{E}\left[\max_{s\in[n]}(Z_s-2h-\alpha^*)\mathbb{1}(\mathcal{E})\right]\right)\\
&\geq\sum_{i=l_n+1}^{n}\left(1-\frac{1}{n}\right)^{i-l_n-1}\frac{1}{n}\times\left(\alpha^*\mathbb{P}(\mathcal{E})-2h\mathbb{P}(\mathcal{E})+\mathbb{E}\left[\max_{s\in[n]}Z_s\right]\mathbb{P}(\mathcal{E})-(2h+\alpha^*)\mathbb{P}(\mathcal{E})\right)\\
&\geq\sum_{i=l_n+1}^{n}\left(1-\frac{1}{n}\right)^{i-l_n-1}\frac{1}{n}\times\left(\mathbb{E}\left[\max_{s\in[n]}X_s\right]\mathbb{P}(\mathcal{E})-4h\right)\\
&\geq\sum_{i=l_n+1}^{n}\left(1-\frac{1}{n}\right)^{i-l_n-1}\frac{1}{n}\times\left(\mathbb{E}\left[\max_{s\in[n]}X_s\right]\mathbb{P}(\mathcal{E})-O\left(\sqrt{\frac{dL\log(nL)}{\lambda l_n}}(\sigma\sqrt{d}+\sqrt{S})\right)\right)\\
&\geq\frac{1-(1-\frac{1}{n})^{n-l_n}}{1/n}\frac{1}{n}\left(\mathbb{E}\left[\max_{s\in[n]}X_s\right]\left(1-\frac{1}{n}-\frac{d}{e^{\lambda l_n/8L}}\right)-O\left(\sqrt{\frac{dL\log(nL)}{\lambda l_n}}(\sigma\sqrt{d}+\sqrt{S})\right)\right),
\end{aligned}
\tag{10}
$$

where the first inequality is obtained from $g\leq h$ and the second last inequality is obtained from the definition of h.

Finally, from (8), (9), and (10), we have

$$\lim_{n\to\infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[n]} X_i]} = \lim_{n\to\infty} \frac{\mathbb{E}\left[\mathbb{E}[X_\tau \mid \mathcal{H}_{l_n}]\right]}{\mathbb{E}[\max_{i\in[n]} X_i]}$$
$$\geq \lim_{n\to\infty} \frac{\mathbb{E}\left[\mathbb{E}[X_\tau \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}]\right]}{\mathbb{E}[\max_{i\in[n]} X_i]}$$
$$\geq \lim_{n\to\infty} \frac{1-(1-\frac{1}{n})^{n-l_n}}{1/n}\frac{1}{n}\left(\left(1-\frac{1}{n}-\frac{d}{e^{\lambda l_n/8L}}\right) - \mathcal{O}\left(\frac{1}{\mathbb{E}[\max_{i\in[n]} X_i]}\sqrt{Ld\frac{\log(Ln)}{\lambda l_n}}(\sigma\sqrt{d}+\sqrt{S})\right)\right)$$
$$= \left(1-\frac{1}{e}\right) - \mathcal{O}\left(\limsup_{n\to\infty} \frac{1}{\mathbb{E}[\max_{i\in[n]} X_i]}\sqrt{\frac{Ld(\sigma^2 d+S)\log(Ln)}{\lambda l_n}}\right), \tag{11}$$

where the last inequality is obtained from limits $\lim_{n\to\infty}(1-1/n)^n = 1/e$ and $\lim_{n\to\infty}(1-1/n)^{l_n} = 1$ (since $l_n = o(n)$) and $l_n = \omega(\frac{L\log d}{\lambda})$.

A.3 Proof of Theorem 4.5

Lemma A.5 (Theorem 2 in Abbasi-Yadkori et al. (2011))**.** *We have*

$$\mathbb{P}\left(\forall i\in[n], \|\hat{\theta}_i-\theta\|_{V_i} \leq \sqrt{S\beta}+\sigma\sqrt{d\log\left(\frac{1+\sum_{s\in\mathcal{I}_i}\|x_s\|_2^2/d\beta}{\delta}\right)}\right) \geq 1-\delta$$

The above lemma implies that

$$\mathbb{P}\left(\left|x^\top(\hat{\theta}_i-\theta)\right| \leq \sqrt{x^\top V_i^{-1}x}\left(\sigma\sqrt{d\log\left(n+n|\mathcal{I}_i|L/d\beta\right)}+\sqrt{S\beta}\right), \forall x\in\mathbb{R}^d, \forall i\in[n]\right) \geq 1-1/n. \tag{12}$$

We define an event $\mathcal{E}_1 = \{|x^\top(\hat{\theta}_i-\theta)| \leq \sqrt{x^\top V_i^{-1}x}(\sigma\sqrt{d\log(n+n|\mathcal{I}_i|L/d\beta)}+\sqrt{S\beta}), \forall x\in\mathbb{R}^d, \forall i\in[a_n+1,n]\}$. From (12), we have $\mathbb{P}(\mathcal{E}_1)\geq 1-\frac{1}{n}$. Let $a_n = \lceil\sqrt{nl_n}\rceil$. Then we define $g_i := \sqrt{L\|V_{a_n}^{-1}\|_2}(\sigma\sqrt{d\log(n+n|\mathcal{I}_i|L/d\beta)}+\sqrt{S\beta})$ so that, for $i > a_n$, $\xi_i(x_i)\leq g_i$ (recall $\xi_i(x_i) = \sqrt{x_i^\top V_i^{-1}x_i}(\sigma\sqrt{d\log(n+n|\mathcal{I}_i|L/d\beta)}+\sqrt{S\beta})$). We denote $\mathcal{H}_i = \{\hat{\theta}_i, V_i, \mathcal{I}_i\}$.

Lemma A.6. *Under $\mathcal{E}_1$, for any $i > a_n$, and any given $\mathcal{H}_i$, we have*

$$\alpha^* - 2g_i \leq \alpha_i \leq \alpha^*.$$

Proof. For $z\sim\mathcal{D}$, we define $Z = z^\top\theta$ and $\hat{Z}_i = z^\top\hat{\theta}_i$. Then, under $\mathcal{E}_1$, for any given V_i and $\hat{\theta}_i$, we have $Z-2g_i \leq \hat{Z}_i - g_i \leq \hat{Z}_i - \xi_i(z) \leq Z$ with $\xi_i(z)\leq g_i$.

Let α^* be the oracle threshold satisfying $\mathbb{P}(Z\geq\alpha^*|\mathcal{H}_i)(=\mathbb{P}(Z\geq\alpha^*)) = 1/n$. From $\hat{Z}_i-\xi_i(z)\leq Z$ and $\mathbb{P}(\hat{Z}_i-\xi_i(z)\geq\alpha_i \mid \mathcal{H}_i) = \mathbb{P}(Z\geq\alpha^* \mid \mathcal{H}_i)(=1/n)$, we can easily obtain

$$\alpha_i \leq \alpha^*.$$

Likewise, for α' s.t. $\mathbb{P}(Z-2g_i\geq\alpha' \mid \mathcal{H}_i) = \frac{1}{n}$, from $Z-2g_i\leq\hat{Z}_i-g_i\leq\hat{Z}_i-\xi_i(z)$ and $\mathbb{P}(Z-2g_i\geq\alpha' \mid \mathcal{H}_i) = \mathbb{P}(\hat{Z}_i-\xi_i(z)\geq\alpha_i \mid \mathcal{H}_i)$, we have $\alpha'\leq\alpha_i$. Therefore, with $\alpha'+2g_i = \alpha^*$ from $\mathbb{P}(Z-2g_i\geq\alpha' \mid \mathcal{H}_i) = \mathbb{P}(Z\geq\alpha^* \mid \mathcal{H}_i) = 1/n$, we have

$$\alpha^* - 2g_i \leq \alpha_i,$$

which concludes the proof. □

Lemma A.7 (Multiplicative Chernoff Bound). *Let $Z_1, \dots Z_l$ be Bernoulli random variables with mean μ. Then for $0 \leq \delta \leq 1$ we have*

$$\mathbb{P}\left(\left|\sum_{s=1}^{l} Z_s - l\mu\right| \geq \delta l\mu\right) \leq 2\exp(-\delta^2 l\mu/3)$$

From the above lemma, we define $\mathcal{E}_2 = \left\{\left||\mathcal{I}_i| - i\sqrt{l_n/n}\right| \leq \frac{1}{2}i\sqrt{l_n/n}\,, i \in \{a_n, n\}\right\}$, which holds with probability at least $1 - 2\exp(\frac{-l_n}{12}) - 2\exp(\frac{-\sqrt{nl_n}}{12})$.

From Lemma A.3, for any $l \geq 1$, suppose $z_1, \dots, z_l \sim \mathcal{D}_x$ are i.i.d drawn from a distribution $\mathcal{D}_x$ satisfying Assumption 3.2. Recall $\lambda = \lambda_{\min}(\mathbb{E}_{z\sim\mathcal{D}_x}[zz^\top]) > 0$. We have that

$$\mathbb{P}\left(\frac{1}{l}\sum_{s=1}^{l} z_s z_s^\top \succeq \frac{\lambda}{2}I_d\right) \geq 1 - d\exp\left(-\frac{\lambda l}{8L}\right).$$

Then, we define $\mathcal{E}_3 = \{\sum_{s\in\mathcal{I}_{a_n}} x_s x_s^\top \succeq \frac{|\mathcal{I}_{a_n}|}{2}\lambda I_d\}$, which holds, under $\mathcal{E}_2$, with probability at least $1 - \frac{d}{e^{\lambda l_n/4L}}$. This implies $\mathbb{P}(\mathcal{E}_2 \cap \mathcal{E}_3) = \mathbb{P}(\mathcal{E}_3 \mid \mathcal{E}_2)\mathbb{P}(\mathcal{E}_2) \geq \left(1 - \frac{d}{e^{\lambda l_n/4L}}\right)\left(1 - 2\exp(\frac{-l_n}{12}) - 2\exp(\frac{-\sqrt{nl_n}}{12})\right)$.

Then under $\mathcal{E}_2 \cap \mathcal{E}_3$, we have $\|V_{a_n}^{-1}\|_2 \leq \|(\sum_{s\in I_{a_n}} x_s x_s^\top)^{-1}\|_2 \leq 2\frac{1}{\lambda|I_{a_n}|} \leq 4\frac{1}{\lambda l_n}$. Then for $i > a_n$, we have

$$\begin{aligned}\xi_i(x_i) &\leq \sqrt{L\|V_{a_n}^{-1}\|_2}(\sigma\sqrt{d\log(n + n|\mathcal{I}_i|L/d\beta)} + \sqrt{S\beta})(= g_i)\\ &\leq \sqrt{L\frac{4}{\lambda l_n}}(\sigma\sqrt{d\log(n + n^2L/d\beta)} + \sqrt{S\beta}).\end{aligned} \tag{13}$$

Here we define $h := \sqrt{L\frac{4}{\lambda l_n}}(\sigma\sqrt{d\log(n + n^2L/d\beta)} + \sqrt{S\beta})$ and $\mathcal{E} := \mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3$.

We define the set of decision stages until i as $\mathcal{J}_i := [i]\backslash\mathcal{I}_i$ so that $\mathcal{J}_i \cup \mathcal{I}_i = [i]$ and $\mathcal{J}_1 \subseteq \mathcal{J}_2, \dots, \subseteq \mathcal{J}_n$. Then, we analyze the stopping probability at i in the following lemma.

Lemma A.8. *For $i \in \mathcal{J}_n$ with any given $\mathcal{J}_i = \{j_1, j_2, \dots, j_{|\mathcal{J}_i|}\}$, we have*

$$\mathbb{P}(\tau = i \mid \mathcal{J}_i, \mathcal{E}) = \left(1 - \frac{1}{n}\right)^{|\mathcal{J}_i|-1}\frac{1}{n}.$$

Proof. For notation simplicity, we define $\mathcal{J}_i^{(k)} := \{j_1, \ldots, j_k\} \subseteq \mathcal{J}_i$ for $k \in [|\mathcal{J}_i|]$. Then, for $i \in \mathcal{J}_n$, we have

$$
\begin{aligned}
&\mathbb{P}(\tau = i \mid \mathcal{J}_i, \mathcal{E}) \\
&= \mathbb{E}[\mathbb{P}(\tau = i \mid \mathcal{H}_i, \mathcal{J}_i, \mathcal{E}) \mid \mathcal{J}_i, \mathcal{E}] \\
&= \mathbb{E}[\mathbb{P}(\{X_t^{LCB} \leq \alpha \, \forall t \in \mathcal{J}_i^{(|\mathcal{J}_i|-1)}\}, X_i^{LCB} > \alpha \mid \mathcal{H}_i, \mathcal{J}_i, \mathcal{E}) \mid \mathcal{J}_i, \mathcal{E}] \\
&= \mathbb{E}\left[\mathbb{P}(\{X_t^{LCB} \leq \alpha \, \forall t \in \mathcal{J}_i^{(|\mathcal{J}_i|-1)}\} \mid \mathcal{H}_i, \mathcal{J}_i, \mathcal{E})\mathbb{P}(X_i^{LCB} > \alpha \mid \mathcal{H}_i, \mathcal{J}_i, \mathcal{E}) \mid \mathcal{J}_i, \mathcal{E}\right] \\
&= \mathbb{E}\left[\mathbb{P}(\{X_t^{LCB} \leq \alpha \, \forall t \in \mathcal{J}_i^{(|\mathcal{J}_i|-1)}\} \mid \mathcal{H}_i, \mathcal{J}_i, \mathcal{E}) \mid \mathcal{J}_i, \mathcal{E}\right] \frac{1}{n} \\
&= \mathbb{P}(\{X_t^{LCB} \leq \alpha \, \forall t \in \mathcal{J}_i^{(|\mathcal{J}_i|-1)}\} \mid \mathcal{J}_i, \mathcal{E}) \frac{1}{n} \\
&= \mathbb{E}[\mathbb{P}(\{X_t^{LCB} \leq \alpha \, \forall t \in \mathcal{J}_i^{(|\mathcal{J}_i|-1)}\} \mid \mathcal{H}_{j_{|\mathcal{J}_i|-1}}, \mathcal{J}_i, \mathcal{E}) \mid \mathcal{J}_i, \mathcal{E}] \frac{1}{n} \\
&= \mathbb{E}\left[\mathbb{P}(\{X_t^{LCB} \leq \alpha \, \forall t \in \mathcal{J}_i^{(|\mathcal{J}_i|-2)}\} \mid \mathcal{H}_{j_{|\mathcal{J}_i|-1}}, \mathcal{J}_i, \mathcal{E}) \times \mathbb{P}(X_{j_{|\mathcal{J}_i|-1}}^{LCB} \leq \alpha \mid \mathcal{H}_{j_{|\mathcal{J}_i|-1}}, \mathcal{J}_i, \mathcal{E}) \mid \mathcal{J}_i, \mathcal{E}\right] \frac{1}{n} \\
&= \mathbb{E}\left[\mathbb{P}(\{X_t^{LCB} \leq \alpha \, \forall t \in \mathcal{J}_i^{(|\mathcal{J}_i|-2)}\} \mid \mathcal{H}_{j_{|\mathcal{J}_i|-1}}, \mathcal{J}_i, \mathcal{E}) \mid \mathcal{J}_i, \mathcal{E}\right] \left(1 - \frac{1}{n}\right) \frac{1}{n} \\
&= \mathbb{P}(\{X_t^{LCB} \leq \alpha \, \forall t \in \mathcal{J}_i^{(|\mathcal{J}_i|-2)}\} \mid \mathcal{J}_i, \mathcal{E}) \left(1 - \frac{1}{n}\right) \frac{1}{n} \\
&\;\;\vdots \\
&= \left(1 - \frac{1}{n}\right)^{|\mathcal{J}_i|-1} \frac{1}{n}
\end{aligned}
$$

□

From the decision strategy of the algorithm, we have $\mathbb{P}(\tau = i \mid \mathcal{J}_n) = 0$ for all $i \in \mathcal{I}_n$. Therefore, for analyzing X_τ, we have

$$
\begin{aligned}
&\mathbb{E}[X_\tau \mid \mathcal{E}] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{P}(\tau = i \mid \mathcal{J}_i, \mathcal{E})\mathbb{E}[X_i \mid \tau = i, \mathcal{J}_i, \mathcal{E}] \mid \mathcal{E}\right] \\
&= \mathbb{E}\left[\sum_{i \in \mathcal{J}_n} \mathbb{P}(\tau = i \mid \mathcal{J}_i, \mathcal{E})\mathbb{E}[X_i \mid \tau = i, \mathcal{J}_i, \mathcal{E}] \mid \mathcal{E}\right] \\
&\geq \mathbb{E}\left[\sum_{i \in \mathcal{J}_n \setminus [a_n]} \mathbb{P}(\tau = i \mid \mathcal{J}_i, \mathcal{E})\mathbb{E}[X_i \mid \tau = i, \mathcal{J}_i, \mathcal{E}] \mid \mathcal{E}\right] \\
&\geq \mathbb{E}\left[\sum_{i \in \mathcal{J}_n \setminus [a_n]} \mathbb{P}(\tau = i \mid \mathcal{J}_i, \mathcal{E})\mathbb{E}\left[\mathbb{E}[X_i^{LCB} \mid \tau = i, \mathcal{H}_i, \mathcal{J}_i, \mathcal{E}] \mid \tau = i, \mathcal{J}_i, \mathcal{E}\right] \mid \mathcal{E}\right] \\
&\geq \mathbb{E}\left[\sum_{i \in \mathcal{J}_n \setminus [a_n]} \left(1 - \frac{1}{n}\right)^{|\mathcal{J}_i|-1} \frac{1}{n}\mathbb{E}\left[\mathbb{E}[X_i^{LCB} \mid \tau = i, \mathcal{H}_i, \mathcal{J}_i, \mathcal{E}] \mid \tau = i, \mathcal{J}_i, \mathcal{E}\right] \mid \mathcal{E}\right] \\
&\geq \mathbb{E}\left[\sum_{i \in \mathcal{J}_n \setminus [a_n]} \left(1 - \frac{1}{n}\right)^{|\mathcal{J}_i|-1} \frac{1}{n}\mathbb{E}\left[\mathbb{E}[X_i^{LCB} \mid \tau = i, \mathcal{H}_i, \mathcal{J}_i, \mathcal{E}] \mid \tau = i, \mathcal{J}_i, \mathcal{E}\right] \mid \mathcal{E}\right]. \quad (14)
\end{aligned}
$$

For the last term above, for $i \in \mathcal{J}_n \setminus [a_n]$, we have

$$\begin{aligned}
&\mathbb{E}[X_i^{LCB} \mid \tau = i, \mathcal{H}_i, \mathcal{J}_i, \mathcal{E}] \\
&= \mathbb{E}\left[\alpha_i \mid \tau = i, \mathcal{H}_i, \mathcal{J}_i, \mathcal{E}\right] + \mathbb{E}\left[X_i^{LCB} - \alpha_i \mid X_i^{LCB} \geq \alpha_i, \mathcal{H}_i, \mathcal{J}_i, \mathcal{E}\right] \\
&= \mathbb{E}\left[\alpha_i \mid \mathcal{H}_i, \mathcal{E}\right] + \frac{\mathbb{E}[(X_i^{LCB} - \alpha_i)^+ \mid \mathcal{H}_i, \mathcal{E}]}{\mathbb{P}(X_i^{LCB} \geq \alpha_i \mid \mathcal{H}_i, \mathcal{E})} \\
&\geq \mathbb{E}\left[\alpha^* - 2g_i \mid \mathcal{H}_i, \mathcal{E}\right] + n\mathbb{E}\left[(X_i - 2g_i - \alpha^*)^+ \mid \mathcal{H}_i, \mathcal{E}\right] \qquad (15)
\end{aligned}$$

where the last term is obtained from Lemma A.6, $\xi_i(x_i) \leq g_i$, and the definition of α_i.

In what follows, we consider the case of $\mathbb{E}[\max_{i\in[n]} X_i] - \mathcal{O}\left(\sqrt{dL(\sigma^2 d + S)\frac{\log(Ln)}{\lambda l_n}}\right) > 0$, because otherwise, it is trivially holds:

$$\mathbb{E}[X_\tau \mid \mathcal{E}] \geq \left(\left(1-\frac{1}{n}\right)^{\sqrt{nl_n}} - \left(1-\frac{1}{n}\right)^{n-\frac{3}{2}\sqrt{nl_n}-1}\right)\left(\mathbb{E}[\max_{i\in[n]} X_i] - \mathcal{O}\left(\sqrt{Ld(\sigma^2 d + S)\frac{\log(Ln)}{\lambda l_n}}\right)\right).$$

Recall $h := \sqrt{L\frac{4}{\lambda l_n}}(\sigma\sqrt{d\log(n + n^2L/d\beta)} + \sqrt{S\beta})$. Let $Z_k \sim \mathcal{D}$ for $k \in [n]$. Then, for the last term above in (14) with (15), under $\mathcal{E}$, we have

$$\begin{aligned}
&\sum_{i\in\mathcal{J}_n\setminus[a_n]}\left(1-\frac{1}{n}\right)^{|\mathcal{J}_i|-1}\frac{1}{n}\left(\mathbb{E}\left[\alpha^* - 2g_i \mid \mathcal{H}_i, \mathcal{E}\right] + n\mathbb{E}\left[(X_i - 2g_i - \alpha^*)^+ \mid \mathcal{H}_i, \mathcal{E}\right]\right) \\
&= \sum_{i\in\mathcal{J}_n\setminus[a_n]}\left(1-\frac{1}{n}\right)^{|\mathcal{J}_i|-1}\frac{1}{n}\left(\mathbb{E}\left[\alpha^* - 2g_i \mid \mathcal{H}_i, \mathcal{E}\right] + n\mathbb{E}\left[(Z_1 - 2g_i - \alpha^*)^+ \mid \mathcal{H}_i, \mathcal{E}\right]\right) \\
&\geq \sum_{i\in\mathcal{J}_n\setminus[a_n]}\left(1-\frac{1}{n}\right)^{|\mathcal{J}_i|-1}\frac{1}{n}\left(\mathbb{E}\left[\alpha^* - 2g_i \mid \mathcal{H}_i, \mathcal{E}\right] + \mathbb{E}\left[\max_{k\in[n]}(Z_k - 2g_i - \alpha^*)^+ \mid \mathcal{H}_i, \mathcal{E}\right]\right) \\
&\geq \sum_{i\in\mathcal{J}_n\setminus[a_n]}\left(1-\frac{1}{n}\right)^{|\mathcal{J}_i|-1}\frac{1}{n}\left(\mathbb{E}[\max_{k\in[n]} Z_k \mid \mathcal{H}_i, \mathcal{E}] - 4h\right) \\
&\geq \sum_{i\in\mathcal{J}_n\setminus[a_n]}\left(1-\frac{1}{n}\right)^{|\mathcal{J}_i|-1}\frac{1}{n}\left(\mathbb{E}[\max_{k\in[n]} Z_k \mid \mathcal{H}_i, \mathcal{E}] - 4h\right) \\
&\geq \left(\left(1-\frac{1}{n}\right)^{\sqrt{nl_n}} - \left(1-\frac{1}{n}\right)^{|\mathcal{J}_n|-1}\right)\left(\mathbb{E}\left[\max_{k\in[n]} Z_k \mid \mathcal{H}_i, \mathcal{E}\right] - 4h\right) \\
&= \left(\left(1-\frac{1}{n}\right)^{\sqrt{nl_n}} - \left(1-\frac{1}{n}\right)^{n-|\mathcal{I}_n|-1}\right)\left(\mathbb{E}\left[\max_{k\in[n]} Z_k\right] - 4h\right) \\
&\geq \left(\left(1-\frac{1}{n}\right)^{\sqrt{nl_n}} - \left(1-\frac{1}{n}\right)^{n-\frac{3}{2}\sqrt{nl_n}-1}\right)\left(\mathbb{E}\left[\max_{k\in[n]} Z_k\right] - 4h\right), \qquad (16)
\end{aligned}$$

where the last inequality is obtained from $\mathcal{E}$

Finally, from (14), (15), and (16), we have

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[n]} X_i]} \\
&\geq \lim_{n\to\infty} \frac{\mathbb{E}[X_\tau \mid \mathcal{E}]\mathbb{P}(\mathcal{E})}{\mathbb{E}[\max_{i\in[n]} X_i]} \\
&\geq \lim_{n\to\infty} \frac{1}{\mathbb{E}[\max_{i\in[n]} X_i]} \left(\left(1-\frac{1}{n}\right)^{\sqrt{nl_n}} - \left(1-\frac{1}{n}\right)^{n-\frac{3}{2}\sqrt{nl_n}-1}\right) \left(\mathbb{E}\left[\max_{k\in[n]} X_k\right] - 4h\right) \mathbb{P}(\mathcal{E}) \\
&\geq \lim_{n\to\infty} \left(\left(1-\frac{1}{n}\right)^{\sqrt{nl_n}} - \left(1-\frac{1}{n}\right)^{n-\frac{3}{2}\sqrt{nl_n}-1}\right) \left(1 - \mathcal{O}\left(\frac{1}{\mathbb{E}[\max_{i\in[n]} X_i]}\sqrt{\frac{Ld(\sigma^2 d + S)\log(Ln)}{\lambda l_n}}\right)\right) \mathbb{P}(\mathcal{E}) \\
&\geq \left(1-\frac{1}{e}\right)\left(1 - \mathcal{O}\Big(\limsup_{n\to\infty} \frac{1}{\mathbb{E}[\max_{i\in[n]} X_i]}\sqrt{\frac{Ld(\sigma^2 d + S)\log(Ln)}{\lambda l_n}}\Big)\right), \qquad (17)
\end{aligned}
$$

where the last equality is obtained from $l_n = \Omega(\frac{L\log d \log n}{\lambda})$ and $l_n = o(n)$, and $\mathbb{P}(\mathcal{E}) \geq \Big(1 - \frac{1}{n} - \Big(1 - \big(1 - \frac{d}{e^{\lambda l_n/4L}}\big)\Big(1 - 2\exp(\frac{-l_n}{12}) - 2\exp(\frac{-\sqrt{nl_n}}{12})\Big)\Big)\Big)$.

A.4 Proof of Proposition 5.2

We first provide a proof for $\frac{\mathbb{E}[x_\tau^\top \theta]}{\mathbb{E}[\max_{i\in[n]} x_i^\top \theta]} \leq \frac{1}{d}$. Let $\theta = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d$. Consider a non-identical distribution $\mathcal{D}_{x,i}$ that generates the following deterministic points:

$$x_1 = (1,0,\ldots,0), \quad x_2 = (0,1,0,\ldots,0), \quad \ldots, \quad x_d = (0,\ldots,0,1), \quad x_i = (0,\ldots,0) \text{ for } i \in \{d{+}1,\ldots,n\}.$$

For any algorithm, let τ denote its stopping time.

Case 1. Set $\theta_1 = \epsilon$ for $\epsilon > 0$. If $\mathbb{P}(\tau = 1) \leq 1/d$, we set $\theta_2 = \cdots = \theta_d = 0$. Then

$$\mathbb{E}\left[\max_{i\in[n]} x_i^\top \theta\right] = \theta_1, \qquad \mathbb{E}[x_\tau^\top \theta] \leq \frac{\theta_1}{d},$$

so the competitive ratio satisfies $\frac{\mathbb{E}[x_\tau^\top \theta]}{\mathbb{E}[\max_{i\in[n]} x_i^\top \theta]} \leq 1/d$.

Case 2. Otherwise if $\mathbb{P}(\tau = 1) > 1/d$ we set $\theta_2 = \theta_1/\epsilon$ for some $0 < \epsilon < 1$. If $\mathbb{P}(\tau = 2) \leq 1/d$, then we set $\theta_3 = \theta_4 = \cdots = \theta_d = 0$. Then

$$\mathbb{E}\left[\max_{i\in[n]} x_i^\top \theta\right] = \theta_2, \qquad \mathbb{E}[x_\tau^\top \theta] \leq \frac{\theta_2}{d} + \epsilon = \frac{1}{d} + \epsilon,$$

again yielding, as $\epsilon \to 0$, $\frac{\mathbb{E}[x_\tau^\top \theta]}{\mathbb{E}[\max_{i\in[n]} x_i^\top \theta]} \leq 1/d$.

Case 3. Likewise, otherwise if $\mathbb{P}(\tau = 2) > 1/d$, we set $\theta_3 = \theta_2/\epsilon = 1/\epsilon$. If $\mathbb{P}(\tau = 3) \leq 1/d$, the we set $\theta_4 = \theta_5 = \cdots = \theta_d = 0$, Then

$$\mathbb{E}\left[\max_{i\in[n]} x_i^\top \theta\right] = \theta_3, \qquad \mathbb{E}[x_\tau^\top \theta] \leq \frac{\theta_3}{d} + 1 + \epsilon,$$

again yielding, as $\epsilon \to 0$, $\frac{\mathbb{E}[x_\tau^\top \theta]}{\mathbb{E}[\max_{i\in[n]} x_i^\top \theta]} \leq 1/d$.

There must exist some $i \in \{1,\ldots,d\}$ such that $\mathbb{P}(\tau = i) \leq 1/d$. Therefore, in a similar way, by choosing θ to place the largest mass of θ_{i-1}/ϵ on that coordinate, as $\epsilon \to 0$, we can easily show that

$$\frac{\mathbb{E}[x_\tau^\top \theta]}{\mathbb{E}[\max_{i\in[n]} x_i^\top \theta]} \leq \frac{1}{d}.$$

Thus, in all cases, one can construct θ such that the competitive ratio satisfies $\frac{\mathbb{E}[x_\tau^\top \theta]}{\mathbb{E}[\max_{i\in[n]} x_i^\top \theta]} \leq 1/d$.

We next provide a proof for $\frac{\mathbb{E}[x_\tau^\top \theta]}{\mathbb{E}[\max_{i\in[n]} x_i^\top \theta]} \le \frac{1}{2}$. We can construct $D_{x,1}$ such that $x_1 = (1,0,\dots,0)$ are drawn deterministically. We also consider $\theta = (1,0,\dots,0)$ such that $X_1 = 1$. We also consider $\mathcal{D}_{x,2}$ such that it generates $x_2 = (1/\epsilon, 0, \dots, 0)$ with probability ϵ and otherwise, $x_2 = (0,0,\dots,0)$ with probability $1-\epsilon$. For $i \ge 3$, we consider $x_i = (0,\dots,0)$.

Then for any algorithm τ which does know X_i for $i \in [n]$ in advance, we have $\mathbb{E}[x_\tau^\top \theta] \le 1$. On the other hands, the prophet who knows X_i in advance can stop at $\tau = 1$ with $X_1 = 1$ if $X_2 = 0$ with probability $1-\epsilon$ or stop at $d+2$ if $X_2 = 1/\epsilon$ with probability ϵ. This implies that $\frac{\mathbb{E}[x_\tau^\top \theta]}{\mathbb{E}[\max_{i\in[n]} x_i^\top \theta]} \le 1/(2-\epsilon)$. As $\epsilon \to 0$, we can conclude $\frac{\mathbb{E}[x_\tau^\top \theta]}{\mathbb{E}[\max_{i\in[n]} x_i^\top \theta]} \le 1/2$.

Lastly, we provide a proof for $\mathbb{E}[X_\tau]/\mathbb{E}\big[\max_{i\in\{d+1,\dots,n\}} X_i\big] \le \frac{1}{2}$. We can construct $D_{x,i}$ for $i \in [d+1]$ such that $x_1 = x_2 = \cdots = x_{d+1} = (1,0,\dots,0)$ are drawn deterministically. We also consider $\theta = (1,0,\dots,0)$ such that $X_1 = X_2 = \cdots = X_{d+1} = 1$. We also consider $\mathcal{D}_{x,d+2}$ such that it generates $x_{d+2} = (1/\epsilon, 0,\dots,0)$ with probability ϵ and otherwise, $x_{d+2} = (0,0,\dots,0)$ with probability $1-\epsilon$. For $i \ge d+2$, we consider $x_i = (0,\dots,0)$.

Then for any algorithm τ which does know X_i for $i \in [n]$ in advance, we have $\mathbb{E}[x_\tau^\top \theta] \le 1$. On the other hands, the prophet who knows X_i in advance can stop at $\tau = 1$ with $X_1 = 1$ if $X_{d+2} = 0$ with probability $1-\epsilon$ or stop at $d+2$ if $X_{d+2} = 1/\epsilon$ with probability ϵ. This implies that $\frac{\mathbb{E}[x_\tau^\top \theta]}{\mathbb{E}[\max_{i\in[n]} x_i^\top \theta]} \le 1/(2-\epsilon)$. As $\epsilon \to 0$, we can conclude $\frac{\mathbb{E}[x_\tau^\top \theta]}{\mathbb{E}[\max_{i\in[n]} x_i^\top \theta]} \le 1/2$.

A.5 Proof of Theorem 5.1

From Lemma A.1, we can show that

$$\mathbb{P}\left(\left|x^\top(\hat{\theta}-\theta)\right| \le \sqrt{x^\top V^{-1} x}\left(\sigma\sqrt{d\log(n + nl_n L/d\beta)} + \sqrt{S\beta}\right), \forall x \in \mathbb{R}^d\right) \ge 1 - 1/n.$$

We define an event $\mathcal{E}_1 = \{|x^\top(\hat{\theta}-\theta)| \le \sqrt{x^\top V^{-1} x}(\sigma\sqrt{d\log(n+nl_nL/d\beta)} + \sqrt{S\beta}), \forall x \in \mathbb{R}^d\}$, which holds with $\mathbb{P}(\mathcal{E}_1) \ge 1 - \frac{1}{n}$. Then under $\mathcal{E}_1$, we have

$$X_i - \xi(x_i) \le x_i^\top \hat{\theta} \le X_i + \xi(x_i). \tag{18}$$

Lemma A.9. *For $l \ge 1$, let $z_t \sim \mathcal{D}_{x,t}$ for $t \in [l]$ be independent random vectors (not necessarily i.i.d.) satisfying Assumption 3.2. Then*

$$\Pr\left(\frac{1}{l}\sum_{t=1}^{l} z_t z_t^\top \succeq \lambda' I_d\right) \ge 1 - d\exp\left(-\frac{\lambda' l}{8L}\right).$$

Proof. Let $\mu_{\min} = \lambda_{\min}(\mathbb{E}[\sum_{t=1}^{l} z_t z_t^\top])$. By the matrix Chernoff bound (Theorem 5.1.1 in Tropp et al. (2015)) for sums of independent PSD matrices with Assumption 3.2, for any $\delta \in [0,1]$,

$$\Pr\left[\lambda_{\min}\left(\sum_{t=1}^{l} z_t z_t^\top\right) \le (1-\delta)\mu_{\min}\right] \le d\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu_{\min}/L} \le d\exp\left(-\frac{\delta^2}{2}\cdot\frac{\mu_{\min}}{L}\right).$$

Choosing $\delta = \frac{1}{2}$ yields

$$\Pr\left[\lambda_{\min}\left(\sum_{t=1}^{l} z_t z_t^\top\right) \le \frac{\mu_{\min}}{2}\right] \le d\exp\left(-\frac{\mu_{\min}}{8L}\right) \le d\exp\left(-\frac{l\lambda'}{8L}\right),$$

where the last inequality is obtained from Weyl's eigenvalue inequalities. Equivalently, with probability at least $1 - d\exp(-\lambda' l/(8L))$,

$$\sum_{t=1}^{l} z_t z_t^\top \succeq \frac{\mu_{\min}}{2} I_d \succeq \frac{l\lambda'}{2} I_d,$$

which completes the proof. ∎

Let $\mathcal{E}_2 = \{\sum_{t=1}^{l_n} x_t x_t^\top \succeq \frac{\lambda' l_n}{2} I_d\}$, which holds with probability at least $1 - \frac{d}{e^{\lambda' l_n/8L}}$ from Lemma A.9. Then under $\mathcal{E}_2$, we have $\|V^{-1}\|_2 \leq \|(\sum_{t=1}^{l_n} x_t x_t^\top)^{-1}\|_2 \leq 2\frac{1}{\lambda' l_n}$. Then for $i > l_n$, we have

$$\begin{aligned}\xi(x_i) &\leq \sqrt{\|x_i\|_2^2 \|V^{-1}\|_2}(\sigma\sqrt{d\log(n + nl_nL/d\beta)} + \sqrt{S\beta})(:= g_i)\\ &\leq \sqrt{\frac{2L}{\lambda' l_n}}(\sigma\sqrt{d\log(n + nl_nL/d\beta)} + \sqrt{S\beta}). \end{aligned} \tag{19}$$

Here we define $h := \sqrt{\frac{2L}{\lambda' l_n}}(\sigma\sqrt{d\log(n + nl_nL/d\beta)} + \sqrt{S\beta})$. Let $z_i \sim \mathcal{D}_{x,i}$ and $\alpha^* = \frac{1}{2}\mathbb{E}\left[\max_{i\in[l_n+1,n]} z_i^\top \theta\right]$. Then, from (18) and (19), we have

$$\alpha^* - \frac{1}{2}h \leq \alpha \leq \alpha^* + \frac{1}{2}h. \tag{20}$$

Let $\mathcal{E} := \mathcal{E}_1 \cup \mathcal{E}_2$ and $\mathcal{H}_{l_n} = \{\hat{\theta}, V\}$. Then for $i > l_n$, we have

$$\begin{aligned}&\mathbb{E}[X_\tau^{LCB}\mathbb{1}(\mathcal{E}) \mid \tau = i, \mathcal{H}_{l_n}]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n})\\ &= \mathbb{E}[\alpha\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n}) + \mathbb{E}[X_i^{LCB}\mathbb{1}(\mathcal{E}) - \alpha\mathbb{1}(\mathcal{E}) \mid \tau = i, \mathcal{H}_{l_n}]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n})\\ &= \mathbb{E}[\alpha\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n})\\ &\quad + \mathbb{E}[(X_i^{LCB} - \alpha)_+\mathbb{1}(\mathcal{E}) \mid X_i^{LCB} \geq \alpha, \mathcal{H}_{l_n}]\mathbb{P}(X_i^{LCB} \geq \alpha \mid \mathcal{H}_{l_n}) \prod_{j\in[i-1]} \mathbb{P}(X_j^{LCB} < \alpha_j | \mathcal{H}_{l_n})\\ &\geq \mathbb{E}[\alpha\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n}) + \mathbb{E}[(X_i^{LCB} - \alpha)_+\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}]\mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n})\\ &\geq \mathbb{E}[\alpha\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n})\\ &\qquad + \mathbb{E}\left[(X_i - 2\xi(x_i) - \alpha)_+ \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right]\mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n})\\ &\geq \mathbb{E}\left[(\alpha^* - \frac{1}{2}h)\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n})\\ &\qquad + \left(\mathbb{E}\left[\left(X_i - \alpha^* - \frac{5}{2}h\right)_+ \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right]\right)\mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n}), \end{aligned} \tag{21}$$

where the last inequality is obtained from (20) and $\xi(x_i) \leq h$.

Using the above, we have

$$
\begin{aligned}
&\mathbb{E}[X_\tau \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}] \\
&\geq \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}[X_\tau^{LCB}\mathbb{1}(\mathcal{E}) \mid \tau = i, \mathcal{H}_{l_n}] \cdot \mathbb{P}(\tau = i \mid \mathcal{H}_{l_n}) \mid \mathcal{H}_{l_n}\right] \\
&\geq \sum_{i=l_n+1}^{n} \mathbb{E}\left[\mathbb{E}[X_i^{LCB}\mathbb{1}(\mathcal{E}) \mid \tau = i, \mathcal{H}_{l_n}] \cdot \mathbb{P}(\tau = i \mid \mathcal{H}_{l_n}) \mid \mathcal{H}_{l_n}\right] \\
&\geq \sum_{i=l_n+1}^{n} \mathbb{E}\Bigg[\mathbb{E}\left[(\alpha^* - \frac{1}{2}h)\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n}) \\
&\qquad + \left(\mathbb{E}\left[\left(X_i - \alpha^* - \frac{5}{2}h\right)_+ \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right]\right)\mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n}) \mid \mathcal{H}_{l_n}\Bigg] \\
&\geq \mathbb{E}\Bigg[\left(\mathbb{E}\left[\alpha^*\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right] - \frac{1}{2}h\right)\sum_{i=l_n+1}^{n}\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n}) \\
&\qquad\qquad + \max_{i\in[l_n+1,n]}\mathbb{E}\left[\left(X_i - \alpha^* - \frac{5}{2}h\right)_+ \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right]\mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n}) \mid \mathcal{H}_{l_n}\Bigg] \\
&\geq \mathbb{E}\Bigg[\left(\mathbb{E}\left[\alpha^*\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right] - \frac{1}{2}h\right)(1 - \mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n})) \\
&\quad + \left(\max_{i\in[l_n+1,n]}\mathbb{E}\left[X_i\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right] - \mathbb{E}[\alpha^*\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}] - \frac{5}{2}h\right)\mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n}) \mid \mathcal{H}_{l_n}\Bigg] \\
&\geq \mathbb{E}\Bigg[(\alpha^*\mathbb{P}(\mathcal{E} \mid \mathcal{H}_{l_n}) - \frac{1}{2}h)(1 - \mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n})) \\
&\qquad + \left(\max_{i\in[l_n+1,n]}\mathbb{E}[X_i]\mathbb{P}(\mathcal{E} \mid \mathcal{H}_{l_n}) - \alpha^*\mathbb{P}(\mathcal{E} \mid \mathcal{H}_{l_n}) - \frac{5}{2}h\right)\mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n}) \mid \mathcal{H}_{l_n}\Bigg] \\
&\geq \alpha^*\mathbb{P}(\mathcal{E} \mid \mathcal{H}_{l_n}) - \frac{5}{2}h. \qquad (22)
\end{aligned}
$$

Finally, using the above, we have

$$
\begin{aligned}
&\lim_{n\to\infty}\frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[l_n+1,n]}X_i]} \geq \lim_{n\to\infty}\frac{\mathbb{E}[\mathbb{E}[X_\tau\mathbb{1}(\mathcal{E})|\mathcal{H}_{l_n}]]}{\mathbb{E}[\max_{i\in[l_n+1,n]}X_i]} \\
&\geq \lim_{n\to\infty}\frac{1}{\mathbb{E}[\max_{i\in[l_n+1,n]}X_i]}\left(\alpha^*\mathbb{P}(\mathcal{E}) - \frac{5}{2}h\right) \\
&\geq \lim_{n\to\infty}\left(\frac{1}{2}\left(1 - \frac{1}{n} - \frac{d}{e^{\lambda' l_n/8L}}\right) - \mathcal{O}\left(\frac{1}{\mathbb{E}[\max_{i\in[l_n+1,n]}X_i]}\sqrt{\frac{Ld\log(Ln)}{\lambda' l_n}}(\sigma\sqrt{d} + \sqrt{S})\right)\right) \\
&= \frac{1}{2} - \mathcal{O}\left(\limsup_{n\to\infty}\frac{1}{\mathbb{E}[\max_{i\in[l_n+1,n]}X_i]}\sqrt{\frac{Ld(\sigma^2 d + S)\log(Ln)}{\lambda' l_n}}\right).
\end{aligned}
$$

A.6 Proof of Proposition 5.3

The argument follows the statement used in Marshall et al. (2020). For completeness, we provide the details here. Consider the instance where $X_1 = 1$ deterministically, $X_2 = X_3 = \cdots = X_{n-1} = 0$ deterministically, and X_n takes value $1/\epsilon$ with probability ϵ (for any $0 < \epsilon < 1$) and 0 otherwise. For any $w_n \leq n-1$, the gambler receives an expected payoff of 1, while the prophet receives an expected payoff of $2 - \epsilon$. Thus, the ratio satisfies $\mathbb{E}[X_\tau]/\mathbb{E}[\max_{i\in[n]} X_i] \leq 1/(2-\epsilon)$. We can conclude the proof with $\epsilon \to 0$.

Algorithm 3 Explore-Then-Decide with LCB Thresholding under Window Access (ETD-LCBT-WA)

Input: Exploration length l_n; regularization parameter β
Output: Stopping time τ

for $i = 1, \dots, n$ **do**
  if $i \le l_n$ **then**
    Observe (x_i, y_i)
  else if $i = l_n + 1$ **then**
    Compute $\hat{\theta}^{(k)}$ and $V^{(k)}$ for $k \in [l_n + 1]$ from (23)
    Compute α from (24) and X_k^{LCB} for $k \le l_n + 1$ from (25).
    if $\max_{k \in [1, l_n+1]} X_k^{LCB} \ge \alpha$ **then**
      Stop with $\tau \leftarrow \arg\max_{k \in [1, l_n+1]} X_i^{LCB}$
  else
    Observe (x_i, y_i)
    Compute X_i^{LCB} from (25).
    if $X_i^{LCB} \ge \alpha$ **then**
      Stop with $\tau \leftarrow i$

A.7 Details of an Algorithm for Non-iid Distributions under Window Access

Individual Estimators. After the l_n exploration stages, we define for each $i \in [l_n + 1]$

$$\hat{\theta}^{(i)} = \left(V^{(i)}\right)^{-1} \sum_{t \in [l_n+1] \setminus \{i\}} y_t x_t, \quad \text{where } V^{(i)} = \sum_{t \in [l_n+1] \setminus \{i\}} x_t x_t^\top + \beta I_d. \tag{23}$$

This construction ensures that the estimator $\hat{\theta}^{(i)}$ is independent of (x_i, y_i). For ease of presentation, for $i \ge l_n + 2$, we define $\hat{\theta}^{(i)} := \hat{\theta}^{(l_n+1)}$ and $V^{(i)} := V^{(l_n+1)}$.

Decision with LCB Threshold under Window Access. Let $z_k \sim \mathcal{D}_{x,k}$ for $k \in [n]$. Then the threshold value is set to

$$\alpha = \tfrac{1}{2} \mathbb{E}\left[\max_{k \in [n]} z_k^\top \hat{\theta}^{(l_n+1)} \,\middle|\, \hat{\theta}^{(l_n+1)}\right], \tag{24}$$

and we define LCBs as

$$X_i^{LCB} = x_i^\top \hat{\theta}^{(i)} - \xi_i(x_i), \tag{25}$$

where $\xi_i(x_i) := \sqrt{x_i^\top V_i^{-1} x_i}\left(\sigma\sqrt{d \log(n^2 + n^2 l_n L / d\beta)} + \sqrt{S\beta}\right)$.

At stage $l_n + 1$, the algorithm checks whether $\max_{k \in [1, l_n+1]} X_k^{LCB} \ge \alpha$. If so, it stops with $\tau = \arg\max_{k \in [1, l_n+1]} X_k^{LCB}$; otherwise, it continues. For $i > l_n + 1$, the algorithm stops at stage i if $X_i^{LCB} \ge \alpha$.

A.8 Proof of Theorem 5.4

Lemma A.10. *We have*

$$\mathbb{P}\left(\forall k \in [l_n+1], \|\hat{\theta}^{(i)} - \theta\|_{V^{(k)}} \le \sqrt{S\beta} + \sigma\sqrt{d \log\left(\frac{n(1 + \sum_{s \in [l_n+1] \setminus \{i\}} \|x_s\|_2^2 / d\beta)}{\delta}\right)}\right) \ge 1 - \delta$$

Proof. We can show this lemma easily by using Theorem 2 in Abbasi-Yadkori et al. (2011) with the union bound for each $\hat{\theta}^{(k)}$ for $k \in [l_n + 1]$. □

From Lemma A.10, we can show that
$\mathbb{P}\left(\left|x^\top(\hat{\theta}^{(i)}-\theta)\right| \le \sqrt{x^\top {V^{(i)}}^{-1} x}\left(\sigma\sqrt{d\log(n^2+n^2 l_n L/d\beta)}+\sqrt{S\beta}\right), \forall x\in\mathbb{R}^d, \forall i\in[1,l_n+1]\right)$
$\ge 1-1/n.$
We define an event
$\mathcal{E}_1=\left\{\left|x^\top(\hat{\theta}^{(i)}-\theta)\right| \le \sqrt{x^\top {V^{(i)}}^{-1} x}\left(\sigma\sqrt{d\log(n^2+n^2 l_n L/d\beta)}+\sqrt{S\beta}\right), \forall x\in\mathbb{R}^d, \forall i\in[1,l_n+1]\right\}.$
We have $\mathbb{P}(\mathcal{E}_1)>1-\frac{1}{n}$. Then under $\mathcal{E}_1$, for $i\in[n]$ we have

$$X_i-\xi_i(x_i)\le x_i^\top\hat{\theta}^{(i)}\le X_i+\xi_i(x_i).$$

Let $\mathcal{E}_2=\{\sum_{t\in[1,l_n+1]\setminus\{i\}} x_t x_t^\top \succeq \frac{\lambda' l_n}{2} I_d, \forall i\in[l_n+1]\}$, which holds with probability at least $1-\frac{d(l_n+1)}{e^{\lambda' l_n/8L}}$ from Lemma A.9. Then under $\mathcal{E}_2$, we have $\|{V^{(i)}}^{-1}\|_2\le \|(\sum_{t\in[l_n+1]\setminus\{i\}} x_t x_t^\top)^{-1}\|_2\le 2\frac{1}{\lambda' l_n}$. Then for $i\ge l_n+1$, we have

$$\begin{aligned}
\xi_i(x_i) &\le \sqrt{\|x_i\|_2^2\|{V^{(i)}}^{-1}\|_2}(\sigma\sqrt{d\log(n^2+n^2 l_n L/d\beta)}+\sqrt{S\beta})(=g_i)\\
&\le \sqrt{L\frac{2}{\lambda' l_n}}(\sigma\sqrt{d\log(n^2+n^3 L/d\beta)}+\sqrt{S\beta}). \qquad (26)
\end{aligned}$$

Here we define $h:=\sqrt{L\frac{2}{\lambda' l_n}}(\sigma\sqrt{d\log(n^2+n^3 L/d\beta)}+\sqrt{S\beta})$ and $\mathcal{E}:=\mathcal{E}_1\cup\mathcal{E}_2$.

Let $z_i\sim\mathcal{D}_{x,i}$ and $\alpha^*=\frac{1}{2}\mathbb{E}\left[\max_{i\in[1,n]} z_i^\top\theta\right]$. Then at time l_n+1, by following the step in (21), we have for $i\in[l_n+1]$,

$$\begin{aligned}
&\mathbb{E}[X_\tau^{LCB}\mathbb{1}(\mathcal{E})\mid \tau=i,\hat{\theta}^{(i)},\{x_s\}_{s\in[l_n+1]\setminus\{i\}}]\mathbb{P}(\tau=i\mid\hat{\theta}^{(i)},V^{(i)})\\
&\ge \mathbb{E}\left[(\alpha^*-\frac{1}{2}h_i)\mathbb{1}(\mathcal{E})\mid\hat{\theta}^{(i)},\{x_s\}_{s\in[l_n+1]\setminus\{i\}}\right]\mathbb{P}(\tau=i\mid\hat{\theta}^{(i)},V^{(i)})\\
&\quad+\left(\mathbb{E}\left[\left(X_i-\alpha^*-\frac{5}{2}h_i\right)_+\mathbb{1}(\mathcal{E})\mid\hat{\theta}^{(i)},V^{(i)}\right]\right)\mathbb{P}(\tau=n+1\mid\hat{\theta}^{(i)},V^{(i)}),
\end{aligned}$$

For ease of presentation, recall that we define $\hat{\theta}^{(i)}=\hat{\theta}^{(l_n+1)}$ for all $i>l_n+1$. Then we also have, for $i>l_n+1$,

$$\begin{aligned}
&\mathbb{E}[X_\tau^{LCB}\mathbb{1}(\mathcal{E})\mid \tau=i,\hat{\theta}^{(i)},V^{(i)}]\mathbb{P}(\tau=i\mid\hat{\theta}^{(i)},V^{(i)})\\
&\ge \mathbb{E}\left[(\alpha^*-\frac{1}{2}h)\mathbb{1}(\mathcal{E})\mid\hat{\theta}^{(i)},V^{(i)}\right]\mathbb{P}(\tau=i\mid\hat{\theta}^{(i)},V^{(i)})\\
&\quad+\left(\mathbb{E}\left[\left(X_i-\alpha^*-\frac{5}{2}h\right)_+\mathbb{1}(\mathcal{E})\mid\hat{\theta}^{(i)},V^{(i)}\right]\right)\mathbb{P}(\tau=n+1\mid\hat{\theta}^{(i)},V^{(i)}),
\end{aligned}$$

Combining them all, by following the steps in (22), we obtain:

$$\mathbb{E}[X_\tau\mathbb{1}(\mathcal{E})]\ge\alpha^*\mathbb{P}(\mathcal{E})-\frac{5}{2}h.$$

Finally, using the above, we have

$$\begin{aligned}
&\lim_{n\to\infty}\frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[l_n+1,n]}X_i]}\\
&\ge\lim_{n\to\infty}\frac{1}{\mathbb{E}[\max_{i\in[l_n+1,n]}X_i]}\left(\alpha^*\mathbb{P}(\mathcal{E})-\frac{5}{2}h\right)\\
&\ge\lim_{n\to\infty}\left(\frac{1}{2}\left(1-\frac{1}{n}-\frac{d(l_n+1)}{e^{\lambda' l_n/8L}}\right)-\mathcal{O}\left(\frac{1}{\mathbb{E}[\max_{i\in[l_n+1,n]}X_i]}\sqrt{\frac{Ld\log(Ln)}{\lambda' l_n}}(\sigma\sqrt{d}+\sqrt{S})\right)\right)\\
&=\frac{1}{2}-\mathcal{O}\left(\limsup_{n\to\infty}\frac{1}{\mathbb{E}[\max_{i\in[l_n+1,n]}X_i]}\sqrt{\frac{Ld(\sigma^2 d+S)\log(Ln)}{\lambda' l_n}}\right).
\end{aligned}$$

Algorithm 4 Decide with Offline Samples and LCB Thresholding (DOS-LCBT)

Input: Offline samples (x_t^o, y_t^o) for $t \in \mathcal{S}$; regularization parameter β
Output: Stopping time τ

```
V ← Σ_{t=1}^{l_n} x_t^o x_t^{o⊤} + βI_d;  θ̂ ← V^{-1} Σ_{t=1}^{l_n} y_t^o x_t^o
Compute α from (27)
for i = 1, ..., n do
    Observe (y_i, x_i)
    Compute X_i^{LCB} from (1)
    if X_i^{LCB} ≥ α then
        Stop and set τ ← i
```

A.9 Further Details on the Method Using Offline Samples under Non-i.i.d. Distributions

For any $\mathcal{S} \subseteq [n]$ such that $|S| = l_n$ for $l_n > 0$ (specified later), we assume that the gambler receives offline samples (x_t^o, y_t^o) where $x_t^o \sim \mathcal{D}_t$ for $t \in \mathcal{S}$ and $y_t^o = x_t^{o\top}\theta + \eta_t$. In Algorithm 4, using this offline samples, we obtain $V = \sum_{t=1}^{l_n} x_t^o x_t^{o\top} + \beta I_d$ and $\hat{\theta} = V^{-1}\sum_{t=1}^{l_n} y_t^o x_t^o$ for constant $\beta > 0$. Then we follow the following decision strategy.

Decision with LCB Threshold. For each time $i \geq 1$, for $z_s \sim \mathcal{D}_{x,s}$ for all $s \in [1, n]$, we define the threshold:

$$\alpha = \frac{1}{2}\mathbb{E}\left[\max_{s\in[n]} z_s^\top \hat{\theta} \mid \hat{\theta}\right] \tag{27}$$

Recall the lower confidence bound for X_i in the Explore-then-Deicide framework: $X_i^{LCB} = x_i^\top \hat{\theta} - \xi(x_i)$, where $\xi(x_i) := \sqrt{x_i^\top V^{-1} x_i}(\sigma\sqrt{d\log(n + nl_nL/d\beta)} + \sqrt{S\beta})$. The algorithm stops at stage i if $X_i^{LCB} \geq \alpha$.

Theorem A.11. *In the non-i.i.d. setting with unknown distributions and window access of size $w_n > l_n$, Algorithm 4 with $l_n = o(n)$, $l_n = \omega(\frac{L\log d}{\lambda'})$, and a constant $\lambda > 0$ achieves the following asymptotic competitive ratio:*

$$\lim_{n\to\infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[n]} X_i]} \geq \frac{1}{2} - \mathcal{O}\left(\lim_{n\to\infty} \frac{1}{\mathbb{E}[\max_{i\in[1,n]} X_i]}\sqrt{\frac{L(\sigma^2 d + S)\log(Ln)}{\lambda' l_n}}\right).$$

Furthermore, by setting $l_n = \frac{L(\sigma^2 d+S)}{\lambda} f(n)\log(Ln)$ for some function $f(n)$ (e.g., $f(n) = \Theta(\log^p n)$ for $p > 0$, or $\Theta(n^q)$ for $0 < q < 1$) satisfying $l_n = o(n)$, if $OPT = \omega(1/\sqrt{f(n)})$, then Algorithm 4 achieves the following asymptotic competitive ratio:

$$\lim_{n\to\infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i\in[n]} X_i]} \geq \frac{1}{2}.$$

Proof. From Lemma A.1, we can show that

$$\mathbb{P}\left(\left|x^\top(\hat{\theta} - \theta)\right| \leq \sqrt{x^\top V^{-1} x}\left(\sigma\sqrt{d\log(n + nl_nL/d\beta)} + \sqrt{S\beta}\right), \forall x \in \mathbb{R}^d\right) \geq 1 - 1/n.$$

We define an event $\mathcal{E}_1 = \{|x^\top(\hat{\theta} - \theta)| \leq \sqrt{x^\top V^{-1} x}(\sigma\sqrt{d\log(n + nl_nL/d\beta)} + \sqrt{S\beta}), \forall x \in \mathbb{R}^d\}$, which holds with $\mathbb{P}(\mathcal{E}_1) \geq 1 - \frac{1}{n}$. Then under $\mathcal{E}_1$, we have

$$X_i - \xi(x_i) \leq x_i^\top \hat{\theta} \leq X_i + \xi(x_i). \tag{28}$$

Lemma A.12. *For any $\mathcal{S} \subset [n]$ with $|\mathcal{S}| = l$ for $l > 0$, let $z_t \sim \mathcal{D}_{x,t}$ for $t \in \mathcal{S}$ be independent random vectors (not necessarily i.i.d.) satisfying Assumption 3.2. Then*

$$\Pr\left(\frac{1}{l}\sum_{t\in\mathcal{S}} z_t z_t^\top \succeq \lambda' I_d\right) \geq 1 - d\exp\left(-\frac{\lambda' l}{8L}\right).$$

Proof. Let $\mu_{\min} = \lambda_{\min}(\mathbb{E}[\sum_{t\in\mathcal{S}} z_t z_t^\top])$. By the matrix Chernoff bound (Theorem 5.1.1 in Tropp et al. (2015)) for sums of independent PSD matrices with Assumption 3.2, for any $\delta \in [0,1]$,

$$\Pr\left[\lambda_{\min}\left(\sum_{t\in\mathcal{S}} z_t z_t^\top\right) \le (1-\delta)\mu_{\min}\right] \le d\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu_{\min}/L} \le d\exp\left(-\frac{\delta^2}{2}\cdot\frac{\mu_{\min}}{L}\right).$$

Choosing $\delta = \frac{1}{2}$ yields

$$\Pr\left[\lambda_{\min}\left(\sum_{t\in\mathcal{S}} z_t z_t^\top\right) \le \frac{\mu_{\min}}{2}\right] \le d\exp\left(-\frac{\mu_{\min}}{8L}\right) \le d\exp\left(-\frac{l\lambda'}{8L}\right),$$

where the last inequality is obtained from Weyl's eigenvalue inequalities. Equivalently, with probability at least $1 - d\exp(-\lambda' l/(8L))$,

$$\sum_{t\in\mathcal{S}} z_t z_t^\top \succeq \frac{\mu_{\min}}{2} I_d \succeq \frac{l\lambda'}{2} I_d,$$

which completes the proof.

□

Let $\mathcal{E}_2 = \{\sum_{t\in\mathcal{S}} x_t^o x_t^{o\top} \succeq \frac{\lambda' l_n}{2} I_d\}$, which holds with probability at least $1 - \frac{d}{e^{\lambda' l_n/8L}}$ from Lemma A.12. Then under $\mathcal{E}_2$, we have $\|V^{-1}\|_2 \le \|(\sum_{t\in\mathcal{S}} x_t^o x_t^{o\top})^{-1}\|_2 \le 2\frac{1}{\lambda' l_n}$. Then for $i > l_n$, we have

$$\begin{aligned}\xi(x_i) &\le \sqrt{\|x_i\|_2^2 \|V^{-1}\|_2}(\sigma\sqrt{d\log(n + nl_nL/d\beta)} + \sqrt{S\beta})(:= g_i)\\ &\le \sqrt{\frac{2L}{\lambda' l_n}}(\sigma\sqrt{d\log(n + nl_nL/d\beta)} + \sqrt{S\beta}). \end{aligned} \tag{29}$$

Here we define $h := \sqrt{\frac{2L}{\lambda' l_n}}(\sigma\sqrt{d\log(n + nl_nL/d\beta)} + \sqrt{S\beta})$. Let $z_i \sim \mathcal{D}_{x,i}$ and $\alpha^* = \frac{1}{2}\mathbb{E}\left[\max_{i\in[n]} z_i^\top\theta\right]$. Then, from (28) and (29), we have

$$\alpha^* - \frac{1}{2}h \le \alpha \le \alpha^* + \frac{1}{2}h. \tag{30}$$

Let $\mathcal{E} := \mathcal{E}_1 \cup \mathcal{E}_2$ and $\mathcal{H}_{l_n} = \{\hat\theta, V\}$. Then for $i \ge 1$, we have

$$\begin{aligned}&\mathbb{E}[X_\tau^{LCB}\mathbb{1}(\mathcal{E}) \mid \tau = i, \mathcal{H}_{l_n}]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n})\\ &= \mathbb{E}[\alpha\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n}) + \mathbb{E}[X_i^{LCB}\mathbb{1}(\mathcal{E}) - \alpha\mathbb{1}(\mathcal{E}) \mid \tau = i, \mathcal{H}_{l_n}]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n})\\ &= \mathbb{E}[\alpha\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n})\\ &\quad + \mathbb{E}[(X_i^{LCB} - \alpha)_+\mathbb{1}(\mathcal{E}) \mid X_i^{LCB} \ge \alpha, \mathcal{H}_{l_n}]\mathbb{P}(X_i^{LCB} \ge \alpha \mid \mathcal{H}_{l_n}) \prod_{j\in[i-1]} \mathbb{P}(X_j^{LCB} < \alpha_j | \mathcal{H}_{l_n})\\ &\ge \mathbb{E}[\alpha\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n}) + \mathbb{E}[(X_i^{LCB} - \alpha)_+\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}]\mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n})\\ &\ge \mathbb{E}[\alpha\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n})\\ &\qquad + \mathbb{E}\left[(X_i - 2\xi(x_i) - \alpha)_+ \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right]\mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n})\\ &\ge \mathbb{E}\left[(\alpha^* - \frac{1}{2}h)\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right]\mathbb{P}(\tau = i \mid \mathcal{H}_{l_n})\\ &\qquad + \left(\mathbb{E}\left[\left(X_i - \alpha^* - \frac{5}{2}h\right)_+ \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right]\right)\mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n}),\end{aligned}$$

where the last inequality is obtained from (30) and $\xi(x_i) \le h$.

Using the above, we have

$$
\begin{aligned}
&\mathbb{E}[X_\tau \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}] \\
&\geq \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}[X_\tau^{LCB} \mathbb{1}(\mathcal{E}) \mid \tau = i, \mathcal{H}_{l_n}] \cdot \mathbb{P}(\tau = i \mid \mathcal{H}_{l_n}) \mid \mathcal{H}_{l_n}\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}[X_i^{LCB} \mathbb{1}(\mathcal{E}) \mid \tau = i, \mathcal{H}_{l_n}] \cdot \mathbb{P}(\tau = i \mid \mathcal{H}_{l_n}) \mid \mathcal{H}_{l_n}\right] \\
&\geq \sum_{i=1}^{n} \mathbb{E}\Bigg[\mathbb{E}\left[(\alpha^* - \frac{1}{2}h)\mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right] \mathbb{P}(\tau = i \mid \mathcal{H}_{l_n}) \\
&\qquad + \left(\mathbb{E}\left[\left(X_i - \alpha^* - \frac{5}{2}h\right)_+ \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right]\right) \mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n}) \mid \mathcal{H}_{l_n}\Bigg] \\
&\geq \mathbb{E}\Bigg[\left(\mathbb{E}\left[\alpha^* \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right] - \frac{1}{2}h\right) \sum_{i=1}^{n} \mathbb{P}(\tau = i \mid \mathcal{H}_{l_n}) \\
&\qquad\qquad + \max_{i \in [n]} \mathbb{E}\left[\left(X_i - \alpha^* - \frac{5}{2}h\right)_+ \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right] \mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n}) \mid \mathcal{H}_{l_n}\Bigg] \\
&\geq \mathbb{E}\Bigg[\left(\mathbb{E}\left[\alpha^* \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right] - \frac{1}{2}h\right)(1 - \mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n})) \\
&\quad + \left(\max_{i \in [n]} \mathbb{E}\left[X_i \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}\right] - \mathbb{E}[\alpha^* \mathbb{1}(\mathcal{E}) \mid \mathcal{H}_{l_n}] - \frac{5}{2}h\right) \mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n}) \mid \mathcal{H}_{l_n}\Bigg] \\
&\geq \mathbb{E}\Bigg[(\alpha^* \mathbb{P}(\mathcal{E} \mid \mathcal{H}_{l_n}) - \frac{1}{2}h)(1 - \mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n})) \\
&\qquad + \left(\max_{i \in [n]} \mathbb{E}[X_i] \mathbb{P}(\mathcal{E} \mid \mathcal{H}_{l_n}) - \alpha^* \mathbb{P}(\mathcal{E} \mid \mathcal{H}_{l_n}) - \frac{5}{2}h\right) \mathbb{P}(\tau = n+1 \mid \mathcal{H}_{l_n}) \mid \mathcal{H}_{l_n}\Bigg] \\
&\geq \alpha^* \mathbb{P}(\mathcal{E} \mid \mathcal{H}_{l_n}) - \frac{5}{2}h.
\end{aligned}
$$

Finally, using the above, we have

$$
\begin{aligned}
&\lim_{n \to \infty} \frac{\mathbb{E}[X_\tau]}{\mathbb{E}[\max_{i \in [n]} X_i]} \geq \lim_{n \to \infty} \frac{\mathbb{E}[\mathbb{E}[X_\tau \mathbb{1}(\mathcal{E}) | \mathcal{H}_{l_n}]]}{\mathbb{E}[\max_{i \in [n]} X_i]} \\
&\geq \lim_{n \to \infty} \frac{1}{\mathbb{E}[\max_{i \in [n]} X_i]} \left(\alpha^* \mathbb{P}(\mathcal{E}) - \frac{5}{2}h\right) \\
&\geq \lim_{n \to \infty} \left(\frac{1}{2}\left(1 - \frac{1}{n} - \frac{d}{e^{\lambda' l_n / 8L}}\right) - \mathcal{O}\left(\frac{1}{\mathbb{E}[\max_{i \in [n]} X_i]} \sqrt{\frac{Ld \log(Ln)}{\lambda' l_n}} (\sigma \sqrt{d} + \sqrt{S})\right)\right) \\
&= \frac{1}{2} - \mathcal{O}\left(\limsup_{n \to \infty} \frac{1}{\mathbb{E}[\max_{i \in [n]} X_i]} \sqrt{\frac{Ld(\sigma^2 d + S) \log(Ln)}{\lambda' l_n}}\right).
\end{aligned}
$$

□