# Online Appendix

This appendix contains proofs of the theoretical results, additional simulations testing the robustness of LASSO Bandit, and new results for the OLS Bandit.

## EC.1. Proof of LASSO Tail Inequality for Adapted Observations

Recall the setup of §3.1. Let $X_t$ denote the $t^{th}$ row of $\mathbf{X}$ and $Y(t)$ denote the $t^{th}$ entry of $Y$. The sequence $\{X_t : t=1,...,n\}$ forms an adapted sequence of observations, i.e., $X_t$ may depend on past regressors and their resulting observations $\{X_{t'}, Y(t')\}_{t'=1}^{t-1}$. We also assume that all realizations of the random variable $X_t$, $t \in [n]$, satisfy $\|X_t\|_\infty \le x_{\max}$.

Before proving Proposition 1, we state and prove some lemmas, starting with the Bernstein concentration inequality for adapted sequences.

LEMMA EC.1 **(Bernstein Concentration)**. *Let $\{D_k, \mathfrak{S}_k\}_{k=1}^{\infty}$ be a martingale difference sequence, and suppose that $D_k$ is $\sigma$-subgaussian in an adapted sense, i.e., for all $\alpha \in \mathbb{R}$, $\mathbb{E}[e^{\alpha D_k}|\mathfrak{S}_{k-1}] \le e^{\alpha^2\sigma^2/2}$ almost surely. Then, for all $t \ge 0$, $\Pr[|\sum_{k=1}^{n} D_k| \ge t] \le 2\exp[-t^2/(2n\sigma^2)]$.*

Lemma EC.1 is from Theorem 2.3 of Wainwright (2016) when $\alpha_* = a_k = 0$ and $\nu_k = \sigma$ for all $k$.

LEMMA EC.2. *Define the event*

$$\mathcal{F}(\lambda_0(\gamma)) \equiv \left\{ \max_{r\in[d]} \left(2|\varepsilon^\top X^{(r)}|/n\right) \le \lambda_0(\gamma) \right\},$$

*where $X^{(r)}$ is the $r^{th}$ column of $\mathbf{X}$ and $\lambda_0(\gamma) \equiv 2\sigma x_{\max}\sqrt{(\gamma^2 + 2\log d)/n}$. Then, we have $\Pr[\mathcal{F}(\lambda_0(\gamma))] \ge 1 - 2\exp[-\gamma^2/2]$.*

*Proof of Lemma EC.2* Let $\mathfrak{S}_t$ be the sigma algebra generated by random variables $X_1, \ldots, X_{t-1}$, and $Y(1), \ldots, Y(t-1)$. First, using a union bound, we can write

$$\Pr[\mathcal{F}(\lambda_0(\gamma))] \ge 1 - \sum_{r=1}^{d} \Pr\left[|\varepsilon^\top X^{(r)}| > n\lambda_0(\gamma)/2\right].$$

Now, for each $r \in [d]$, let $D_{t,r} = \varepsilon_t X_{t,r}$ and note that $D_{1,r}, ...D_{n,r}$ is a martingale difference sequence adapted to the filtration $\mathfrak{S}_1 \subset ... \subset \mathfrak{S}_n$ since $\mathbb{E}\left[\varepsilon_t X_{t,r}|\mathfrak{S}_t\right] = 0$. On the other hand, each $D_{t,r}$ is a $(x_{\max}\sigma)$-subgaussian random variable adapted to $\{\mathfrak{S}_t\}_{t=1}^{n}$, since

$$\mathbb{E}[e^{\alpha D_{t,r}}|\mathfrak{S}_{t-1}] \;\le\; \mathbb{E}_{X_t}[e^{\alpha^2 X_{t,r}^2\sigma^2/2}|\mathfrak{S}_{t-1}] \;\le\; e^{\alpha^2(x_{max}\sigma)^2/2}.$$

Then, using Lemma EC.1, $\Pr\left[\mathcal{F}(\lambda_0(\gamma))\right] \ge 1 - 2d\exp\left[-(\gamma^2 + 2\log d)/2\right] = 1 - 2\exp[-\gamma^2/2]$. □

LEMMA EC.3 **(From page 105 of (Bühlmann and Van De Geer 2011))**. *For anly $\lambda_0 \in \mathbb{R}^+$, when $\lambda \ge 2\lambda_0$, on event $\mathcal{F}(\lambda_0)$, we have*

$$2\|\mathbf{X}(\hat\beta - \beta)\|_2^2/n + \lambda\|\hat\beta_{supp(\beta)^c}\|_1 \le 3\lambda\|\hat\beta_{supp(\beta)} - \beta_{supp(\beta)}\|_1.$$

Now we are ready to prove Proposition 1.

*Proof of Proposition 1* Let $\lambda_0(\gamma) = 2\sigma x_{\max}\sqrt{(\gamma^2 + 2\log d)/n}$ and let $\lambda$ be an arbitrary constant such that $\lambda \geq 2\lambda_0(\gamma)$. If both events $\mathcal{F}(\lambda_0(\gamma))$ and $\{\hat{\Sigma}(\mathbf{X}) \in \mathcal{C}(\text{supp}(\beta), \phi)\}$ hold, then

$$\begin{aligned}
2\|\mathbf{X}(\hat{\beta} - \beta)\|_2^2/n + \lambda\|\hat{\beta} - \beta\|_1 &= 2\|\mathbf{X}(\hat{\beta} - \beta)\|_2^2/n + \lambda\|\hat{\beta}_{\text{supp}(\beta)} - \beta_{\text{supp}(\beta)}\|_1 + \lambda\|\hat{\beta}_{\text{supp}(\beta)^c}\|_1 \\
&\leq 4\lambda\|\hat{\beta}_{\text{supp}(\beta)} - \beta_{\text{supp}(\beta)}\|_1 \\
&\leq 4\lambda\sqrt{s_0}\|\mathbf{X}(\hat{\beta} - \beta)\|_2/\sqrt{n\phi^2} \\
&\leq \|\mathbf{X}(\hat{\beta} - \beta)\|_2^2/n + 4\lambda^2 s_0/\phi^2 \,.
\end{aligned}$$

Here the three inequalities use Lemma EC.3, the definition of $\mathcal{C}(\text{supp}(\beta), \phi)$ (Definition 2), and the inequality $4uv \leq u^2 + 4v^2$, respectively. Thus, for $\lambda \geq 2\lambda_0(\gamma)$,

$$\begin{aligned}
\Pr\left\{\|\hat{\beta} - \beta\|_1 \leq \frac{4\lambda s_0}{\phi^2}\right\} &\geq \Pr\left[\mathcal{F}(\lambda_0(\gamma)) \cap \{\hat{\Sigma}(\mathbf{X}) \in \mathcal{C}(\text{supp}(\beta), \phi)\}\right] \\
&\geq \Pr\left[\mathcal{F}(\lambda_0(\gamma))\right] - \Pr\left[\hat{\Sigma}(\mathbf{X}) \notin \mathcal{C}(\text{supp}(\beta), \phi)\right] \\
&\geq 1 - 2\exp\left[-\gamma^2/2\right] - \Pr\left[\hat{\Sigma}(\mathbf{X}) \notin \mathcal{C}(\text{supp}(\beta), \phi)\right] \,.
\end{aligned}$$

Summarizing, we have shown that,

$$\lambda \geq 2\lambda_0(\gamma) \Longrightarrow \Pr\left\{\|\hat{\beta} - \beta\|_1 > \frac{4\lambda s_0}{\phi^2}\right\} \leq 2\exp\left[-\gamma^2/2\right] + \Pr\left[\hat{\Sigma}(\mathbf{X}) \notin \mathcal{C}(\text{supp}(\beta), \phi)\right] \,. \qquad \text{(EC.1)}$$

Now we choose $\gamma = \gamma(\chi) \equiv \sqrt{2n C_* \chi^2 - 2\log d}$ for a suitable constant $C_*$, to be determined. Then, the exponent of the error probability becomes $-\gamma(\chi)^2/2 = -nC_*\chi^2 + \log d$. We will now show that $C_* = C_1(\phi)$ will guarantee the condition $\lambda(\chi, \phi) \geq 2\lambda_0(\gamma(\chi))$. In particular,

$$2\lambda_0(\gamma(\chi)) \;=\; 4\sigma x_{\max}\sqrt{\left[\gamma(\chi)^2 + 2\log d\right]/n} \;=\; 4\sigma x_{\max}\chi\sqrt{2C_*} \;=\; \frac{16\sigma x_{\max} s_0}{\phi^2}\underbrace{\frac{\chi\phi^2}{4s_0}}_{\lambda(\chi,\phi)}\sqrt{2C_*}\,.$$

Therefore, for the inequality $\lambda(\chi, \phi) \geq 2\lambda_0(\gamma(\chi))$ to hold, we need $\phi^4 \geq C_*(512 s_0^2\sigma^2 x_{\max}^2)$, which is satisfied by $C_* = C_1(\phi)$. Now, we can invoke (EC.1) for $\lambda = \lambda(\chi, \phi)$, and use the inverse relation $\chi = 4\lambda(\chi, \phi)s_0/\phi^2$ to finish the proof. □

## EC.2. Proof of LASSO Tail Inequality for Non-i.i.d. Data

Recall the setup from §4.1 as well as assumptions of Lemma 1. The proof involves showing that $\|\hat{\Sigma}(\mathcal{A}') - \Sigma\|_\infty$ is small with high probability using random matrix theory. Next, we use the Azuma-Hoeffding inequality to show that $\hat{\Sigma}(\mathcal{A}) \in \mathcal{C}(\text{supp}(\beta), \phi_1\sqrt{p}/2)$ with high probability. This result provides a tail inequality for LASSO estimates $\hat{\beta}(\mathcal{A}, \lambda)$, even when part of the data is not generated i.i.d. from $\mathcal{P}_Z$.

### EC.2.1. Empirical covariance matrix via random matrix theory

LEMMA EC.4. *Given i.i.d. observations $Z_1,...,Z_n \in \mathbb{R}^d$ from the distribution $\mathcal{P}_Z$ such that all realizations of $Z_t$ satisfy $\|Z_t\|_\infty \le x_{\max}$ for all $t \in [n]$, then for all $w > 0$,*

$$\Pr\left[\|\hat{\Sigma} - \Sigma\|_\infty \ge 2x_{\max}^2 \left(w + \sqrt{2w} + \sqrt{\frac{2\log(d^2+d)}{n}} + \frac{\log(d^2+d)}{n}\right)\right] \le e^{-nw},$$

*where $\Sigma \equiv \mathbb{E}_{\mathcal{P}_Z}[ZZ^\top]$ and $\hat{\Sigma} \equiv \sum_{t=1}^n Z_t Z_t^\top / n$.*

*Proof of Lemma EC.4* First, define the family $\{\gamma_{jk}\}_{1\le j\le k\le d}$ of real-valued functions that take as input random variables $Z \sim \mathcal{P}_Z$. Precisely, for all $1 \le j \le k \le d$,

$$\gamma_{jk}(Z) \equiv \frac{Z(j)Z(k) - \mathbb{E}[Z(j)Z(k)]}{2x_{\max}^2},$$

where $Z(j)$ refers to $j^{th}$ coordinate of vector $Z$. It is easy to see that each such $\gamma_{jk}$ satisfies $\mathbb{E}[\gamma_{jk}(Z)] = 0$ and

$$\frac{1}{n}\sum_{t=1}^n \mathbb{E}|\gamma_{jk}(Z_t)|^m \le 1 \ , \quad m = 2, 3, \dots,$$

implying that the family $\{\gamma_{jk}\}_{1\le j\le k\le d}$ satisfies condition (14.5) in page 489 of Bühlmann and Van De Geer (2011) with $K = 1$. Therefore, we can apply Lemma 14.13 from page 490 of Bühlmann and Van De Geer (2011) and obtain, for all $w > 0$,

$$\Pr\left[\max_{1\le j\le k\le d}\left|\frac{1}{n}\sum_{t=1}^n \gamma_{jk}(Z_t)\right| \ge w + \sqrt{2w} + \sqrt{\frac{2\log(d^2+d)}{n}} + \frac{\log(d^2+d)}{n}\right] \le e^{-nw}.$$

Now, the result follows from the fact that $\|\hat{\Sigma} - \Sigma\|_\infty/(2x_{\max}^2) = \max_{1\le j\le k\le d}|\sum_{t=1}^n \gamma_{jk}(Z_t)|/n$. □

### EC.2.2. Compatibility condition for non-i.i.d. samples

Recall $\mathcal{A}$, $\mathcal{A}'$, $\Sigma$, $\beta$, $\mathbf{Z}$, $W$, and the corresponding assumptions on them from §4.1. We will first show that $\hat{\Sigma}(\mathcal{A}')$ satisfies the compatibility condition (with respect to supp$(\beta)$ and an appropriate constant) with high probability.

LEMMA EC.5. *If $\Sigma \in \mathcal{C}(supp(\beta), \phi_1)$ for constant $\phi_1 > 0$ and $\|\Sigma - \Sigma'\|_\infty \le \phi_1^2/(32s_0)$ holds, then $\Sigma' \in \mathcal{C}(supp(\beta), \phi_1/\sqrt{2})$.*

*Proof of Lemma EC.5* The proof follows directly from Corollary 6.8 in page 152 of Bühlmann and Van De Geer (2011). □

LEMMA EC.6. *If the assumptions of Lemma 1 hold, then*

$$\Pr\left[\hat{\Sigma}(\mathcal{A}') \in \mathcal{C}(supp(\beta), \frac{\phi_1}{\sqrt{2}})\right] \ge 1 - e^{-C_2(\phi_1)^2|\mathcal{A}'|}.$$

*Proof of Lemma EC.6* Given the assumptions of Lemma 1, we have $|\mathcal{A}'| \geq 3\log(d)/C_2(\phi_1)^2$. Together with $d > 1$, this implies that $\log(d^2+d)/|\mathcal{A}'| \leq C_2^2(\phi_1)$. Therefore, for $w = C_2^2(\phi_1)$, we have,

$$2x_{\max}^2\left(w+\sqrt{2w}+\sqrt{\frac{2\log(d^2+d)}{|\mathcal{A}'|}}+\frac{\log(d^2+d)}{|\mathcal{A}'|}\right) \leq 4x_{\max}^2\left[C_2(\phi_1)^2+\sqrt{2}C_2(\phi_1)\right]$$
$$\leq 8x_{\max}^2 C_2(\phi_1)$$
$$\leq \frac{\phi_1^2}{32s_0},$$

where the last two inequalities follow from the definition of $C_2(\phi_1) = \min[1/2, \phi_1^2/(256s_0x_{\max}^2)]$. Thus, it follows from Lemma EC.4 that

$$\Pr\left[\|\Sigma - \hat{\Sigma}(\mathcal{A}')\|_\infty \geq \frac{\phi_1^2}{32s_0}\right] \leq e^{-C_2(\phi_1)^2|\mathcal{A}'|}$$

The result then follows directly from Lemma EC.5. □

Lemma EC.7. *Given the assumptions of Lemma 1, if we have $\hat{\Sigma}(\mathcal{A}') \in \mathcal{C}(supp(\beta), \phi_1')$ for some constant $\phi_1' > 0$, then $\hat{\Sigma}(\mathcal{A}) \in \mathcal{C}(supp(\beta), \phi_1'\sqrt{|\mathcal{A}'|/|\mathcal{A}|})$.*

*Proof of Lemma EC.7* By definition, we can write

$$\hat{\Sigma}(\mathcal{A}) = \frac{|\mathcal{A}'|}{|\mathcal{A}|}\hat{\Sigma}(\mathcal{A}') + \frac{1}{|\mathcal{A}|}\sum_{t\in\mathcal{A}\setminus\mathcal{A}'} Z_t Z_t^\top = \frac{|\mathcal{A}'|}{|\mathcal{A}|}\hat{\Sigma}(\mathcal{A}') + \frac{|\mathcal{A}\setminus\mathcal{A}'|}{|\mathcal{A}|}\hat{\Sigma}(\mathcal{A}\setminus\mathcal{A}').$$

Then, for all $v$ satisfying $\|v_{\text{supp}(\beta)^c}\|_1 \leq 3\|v_{\text{supp}(\beta)}\|_1$,

$$v^\top\hat{\Sigma}(\mathcal{A})v = \frac{|\mathcal{A}'|}{|\mathcal{A}|}v^\top\hat{\Sigma}(\mathcal{A}')v + \frac{|\mathcal{A}\setminus\mathcal{A}'|}{|\mathcal{A}|}v^\top\hat{\Sigma}(\mathcal{A}\setminus\mathcal{A}')v \geq \frac{|\mathcal{A}'|}{|\mathcal{A}|}\frac{\phi_1'^2\,\|v_{\text{supp}(\beta)}\|_1^2}{s_0},$$

using the fact that $\hat{\Sigma}(\mathcal{A}\setminus\mathcal{A}')$ is positive semi-definite. □

Now we have all the ingredients to complete the proof of Lemma 1.

*Proof of Lemma 1:* Applying Lemmas EC.6 and EC.7 implies that

$$\hat{\Sigma}(\mathcal{A}) \in \mathcal{C}\left(\text{supp}(\beta), \frac{\phi_1}{\sqrt{2}}\sqrt{\frac{|\mathcal{A}'|}{|\mathcal{A}|}}\right),$$

with probability at least $1-\exp[-C_2(\phi_1)^2|\mathcal{A}'|]$. This implies that $\hat{\Sigma}(\mathcal{A}) \in \mathcal{C}(\text{supp}(\beta), \phi_1\sqrt{p}/2)$ with probability at least $1-\exp[-p\,C_2(\phi_1)^2|\mathcal{A}|/2]$. Applying Proposition 1 with compatibility constant $\phi = \phi_1\sqrt{p}/2$ yields the result. □

## EC.3. Proof of LASSO Tail Inequality for Forced-Sample Estimator

In this section, we prove a tail inequality for the forced sample estimator $\hat{\beta}(\mathcal{T}_{i,t}, \lambda_1)$ by applying Lemma 1. Recall that at each $t \in \mathcal{T}_{i,t}$, we draw a random covariate vector $X_t$, sampled i.i.d. from $\mathcal{P}_X$, and play arm $i$. Moreover, we assumed that $\Sigma_i \in \mathcal{C}(\text{supp}(\beta_i), \phi_0)$ where $\Sigma_i = \mathbb{E}_{X \sim \mathcal{P}_{X|X\in U_i}}[XX^\top]$ and also that $\Pr[X_t \in U_i] \geq p_*$.

LEMMA EC.8. *If $t \geq (Kq)^2$, then $(1/2)q\log t \leq |\mathcal{T}_{i,t}| \leq 6q\log t$.*

*Proof of Lemma EC.8* Define the $n^{th}$ round of forced sampling of all the arms

$$L_n \equiv \{(2^n - 1)Kq + 1, \ldots, (2^n)Kq\},$$

for $n \geq 0$. By construction, arm $i$ is sampled $|\mathcal{T}_i \cap L_n| = q$ times during $L_n$, so

$$\left|\mathcal{T}_i \cap \left(\bigcup_{r=0}^{n-1} L_r\right)\right| = nq\,.$$

Then for each $t \in L_n$, $nq \leq |\mathcal{T}_{i,t}| \leq (n+1)\,q$. To show the lower bound, note that for $t \in L_n$, we have $t \leq (2^n)Kq$, i.e., $\log_2[t/(Kq)] \leq n$. Therefore, using $t \geq (Kq)^2$,

$$|\mathcal{T}_{i,t}| \geq nq \geq q\,\log_2 \frac{t}{Kq} \geq q\,(\log t - \log Kq) \geq (1/2)q\log t\,.$$

To show the upper bound, note that for $t \in L_n$, $t \geq (2^n - 1)\,Kq$, i.e., $n \leq \log_2[1 + t/(Kq)]$, so

$$|\mathcal{T}_{i,t}| \leq (n+1)q \leq \left[\log_2\left(\frac{t}{Kq} + 1\right) + 1\right] q \leq \frac{\log(2t + 2\sqrt{t})}{\log 2} q \leq 6q\log t\,. \quad \square$$

LEMMA EC.9. *Let $\mathcal{T}'_{i,t} \subset \mathcal{T}_{i,t}$ be the set of all $r \in \mathcal{T}_{i,t}$ such that $X_r \in U_i$. Then for each $r \in \mathcal{T}_{i,t}$ we have $r \in \mathcal{T}'_{i,t}$ independently with probability at least $p_*$. In addition, $\{X_r\}_{r\in\mathcal{T}'_{i,t}}$ are i.i.d. from $\mathcal{P}_{X|X\in U_i}$.*

*Proof of Lemma EC.9* By construction, for each $r \in \mathcal{T}_{i,t}$, $X_r$ is drawn i.i.d. from $\mathcal{P}_X$ and therefore with probability at least $p_*$, $X_r \in U_i$, i.e., $s \in \mathcal{T}'_{i,t}$. Also, note that the events $X_r \in U_i$ are independent for different values of $r$ since the original sequence $\{X_r\}_{s\in\mathcal{T}_{i,t}}$ is i.i.d., implying that each $r \in \mathcal{T}'_{i,t}$, $X_r$ is an i.i.d. sample of $\mathcal{P}_{X|X\in U_i}$. $\square$

Using Lemma EC.9 we see that the inclusion of each member of $\mathcal{T}_{i,t}$ in $\mathcal{T}'_{i,t}$ is a Bernoulli i.i.d. random variable with mean at least $p_*$. Therefore, we get the following result using Chernoff bound.

LEMMA EC.10. *If $t \geq (Kq)^2$, for $\mathcal{T}_{i,t}$ and $\mathcal{T}'_{i,t}$ defined as in Lemma EC.9 the following holds*

$$\Pr\left[\frac{|\mathcal{T}'_{i,t}|}{|\mathcal{T}_{i,t}|} \geq \frac{p_*}{2}\right] \geq 1 - \frac{2}{t^4}\,.$$

*Proof of Lemma EC.10* We use the following version of the Chernoff inequality (Corollary A.1.14 in page 268 of Alon and Spencer 1992, for $\varepsilon = 1/2$ and $c_\varepsilon \approx 0.1082$): Letting $y$ be the sum of mutually independent indicator random variables with $\mu = \mathbb{E}[y]$, we have that

$$\Pr\left[\,|y-\mu| > \mu/2\,\right] < 2\exp\left[-0.1\mu\right]\,.$$

Applying this to indicator random variables $\mathbb{I}(r \in \mathcal{T}'_{i,i})$ for all $r \in \mathcal{T}_{i,t}$ and using

$$\mu = \mathbb{E}\left[\sum_{r\in\mathcal{T}_{i,t}} \mathbb{I}(r \in \mathcal{T}'_{i,t})\right] \geq p_*|\mathcal{T}_{i,t}|\,,$$

we can write that

$$\Pr\left[\,|\mathcal{T}'_{i,t}| < (p_*/2)|\mathcal{T}_{i,t}|\,\right] < 2e^{-\frac{p_*}{10}|\mathcal{T}_{i,t}|}\,.$$

Next, using Lemma EC.8, $t \geq (Kq)^2$, $q \geq 4q_0$, and the definition of $q_0$ from §3.3 we have

$$\Pr\left[\,|\mathcal{T}'_{i,t}| < (p_*/2)|\mathcal{T}_{i,t}|\,\right] < 2e^{-(p_*/5)q_0\log t} \leq \frac{2}{t^4}\,. \quad \square$$

Now we are ready to prove Proposition 2.

*Proof of Proposition 2* By construction, and the definition of $q_0$ from §3.3,

$$|\mathcal{T}_{i,t}| \geq (1/2)q\log t \geq 2q_0\log t \geq \frac{6\log(d)}{p_*\,C_2(\phi_0)^2}\,.$$

Then combining Lemma EC.9-EC.10 and Lemma 1, with $\mathcal{P}_Z = \mathcal{P}_{X|X\in U_i}$, $\chi = h/(4x_{\max})$, $p = p_*$, and $\lambda_1 = \lambda(h/(4x_{\max}, \phi_0\sqrt{p_*}/2)$ we obtain

$$\begin{aligned}\Pr\left[\|\hat\beta\left(\mathcal{T}_{i,t},\lambda_1\right)-\beta_i\|_1 > \frac{h}{4x_{\max}}\right] &\leq 2e^{-C_1\left(\frac{\phi_0\sqrt{p_*}}{2}\right)2q_0\log t\frac{h^2}{16x_{\max}^2}+\log d} + e^{-p_*C_2(\phi_0)^2q_0\log t} + \frac{2}{t^4}\\ &\leq 2e^{-p_*^2C_1(\phi_0)q_0\log t\frac{h^2}{128x_{\max}^2}+\log d} + \frac{1}{t^4} + \frac{2}{t^4}\\ &\leq \frac{5}{t^4}\,.\end{aligned}$$

The last two inequalities use the definition of $q_0$ and the fact that $t \geq (Kq)^2$ to show that the exponent of each term on the right hand side is at most $-4\log t$. $\square$

## EC.4. Proof of LASSO Tail Inequality for All-Sample Estimator

In this section, we prove the tail inequality for the all-sample estimator $\hat\beta(\mathcal{S}_{i,t},\lambda_{2,t})$ for arms in $\mathcal{K}_{opt}$. The approach mirrors the steps taken in Appendix EC.3. However, there is an additional complication due to the correlation between rows of $\mathbf{X}(\mathcal{S}_{i,t})$ that was discussed in §4.3. Recall the events $A_t$ defined in Eq. (3).

Lemma EC.11. *For each $i \in [K]$, if the events $X_t \in U_i$ and $A_{t-1}$ hold, and $t \notin \cup_{j\in[K]}\mathcal{T}_{j,t}$, the LASSO Bandit uses the forced-sample estimator $\hat{\beta}(\mathcal{T}_{i,t-1},\lambda_1)$ to arrive at $\hat{\mathcal{K}}=\{i\}$, implying that it plays the optimal arm at time $t$.*

*Proof of Lemma EC.11* Since $X_t \in U_i$, we know

$$X_t^\top \beta_i \geq h + \max_{j\neq i} X_t^\top \beta_j \,.$$

Then, for any $j \in [K]\setminus\{i\}$, since $A_{t-1}$ holds,

$$\begin{aligned} X_t^\top\left[\hat{\beta}(\mathcal{T}_{i,t-1})-\hat{\beta}(\mathcal{T}_{j,t-1})\right] &= X_t^\top\left[\hat{\beta}(\mathcal{T}_{i,t-1})-\beta_i\right]-X_t^\top\left[\hat{\beta}(\mathcal{T}_{j,t-1})-\beta_j\right]+X_t^\top(\beta_i-\beta_j)\\ &\geq -x_{\max}\frac{h}{4x_{\max}}-x_{\max}\frac{h}{4x_{\max}}+h\\ &\geq h/2\,. \end{aligned}$$

Thus, at time $t$, $\hat{\mathcal{K}}=\{i\}$, i.e., the LASSO Bandit will play arm $i$. □

Lemma EC.12. *For all $t$ with $t\geq (Kq)^2$ the event $A_t$ occurs with probability at least $1-5K/t^4$.*

*Proof of Lemma EC.12* For each $i\in[K]$ and all $t\geq(Kq)^2$, we have from Proposition 2,

$$\Pr\left[\|\hat{\beta}(\mathcal{T}_{i,t},\lambda_1)-\beta_i\|_1 > \frac{h}{4x_{\max}}\right] \leq \frac{5}{t^4}\,.$$

Taking a union bound over all $K$ arms gives us the result. □

Lemma EC.13. *Let $i\in[K]$. Recall from §4.3 that $\mathcal{S}'_{i,t}\subset[t]$ is the set of all time periods $r$ such the events $X_r\in U_i$ and $A_{r-1}$ hold and we are not forced-sampling any arm $j\in[K]$. Then the following properties are satisfied.*

*(1) The set of random variables $\{X_r \mid r\in\mathcal{S}'_{i,t}\}$ are i.i.d. from distribution $\mathcal{P}_{X|X\in U_i}$.*

*(2) For each $r\in[t]\setminus\cup_{j\in[K]}\mathcal{T}_{j,t}$, we have $r\in\mathcal{S}'_{i,t}$ with probability at least $p_*/2$ when $t\geq(Kq)^2$.*

*(3) $\mathcal{S}'_{i,t}\subset\mathcal{S}_{i,t}$.*

*Proof of Lemma EC.13* For (1), since $A_{r-1}$ is only a function of samples in $\mathcal{T}_{i,r-1}$, $A_{r-1}$ is independent of $X_r$. Therefore, random variables $\{X_r|A_{r-1}$ holds$\}$ are i.i.d. samples from $\mathcal{P}_X$. Now, the presence of each $X_r$ in $U_i$ is simply rejection sampling; thus, using the fact that $r\notin\cup_{j\neq i}\mathcal{T}_{j,t}$ is deterministic, for each $r\in\mathcal{S}'_{i,t}$, $X_r$ is distributed i.i.d. from $\mathcal{P}_{X|X\in U_i}$. For (2), we know that $X\in U_i$ with probability at least $p_*$ and Lemma EC.12 implies that $A_{r-1}$ holds with probability at least $1-5K(r-1)^{-4}$ when $(r-1)\geq(Kq)^2$. Note that $(r-1)\geq(Kq)^2\geq 16K^2$ (since $q\geq 4\lceil q_0\rceil\geq 4$), which implies that $A_{r-1}$ holds with probability at least $1-5K(r-1)^{-4}\geq 1/2$. Then, $r\in\mathcal{S}'_{i,t}$ with probability at least $p_*/2$. Finally, for (3), from Lemma EC.11, we know that for $X_r\sim\mathcal{P}_X$, if events $X_r\in U_i$ and $A_{r-1}$ holds and $r\notin\cup_{j\in[K]}\mathcal{T}_{j,t}$, then $r\in\mathcal{S}_{i,t}$, so $\mathcal{S}'_{i,t}\subset\mathcal{S}_{i,t}$. □

Lemma EC.14. *If $t \geq C_5$, for $\mathcal{S}'_{i,t}$ defined as in §4.3 the following holds*

$$\Pr\left[\,|\mathcal{S}'_{i,t}| \geq t\,p_*/4\,\right] \geq 1 - e^{tp_*^2/36}\,.$$

*Proof of Lemma EC.14* Note that we need to take a more refined approach than in Lemma EC.10 since the events $r \in \mathcal{S}'_{i,t}$, $r \in [t]$, are not independent. By definition of $\mathcal{S}'_{i,t}$ we have for all $r \in [t]/\mathcal{T}_{i,t}$,

$$\mathbb{I}(r \in \mathcal{S}'_{i,t}) = \mathbb{I}(A_{r-1}) \cdot \mathbb{I}(X_r \in U_i) \cdot \mathbb{I}(r \notin \cup_{j\in[K]}\mathcal{T}_{j,t})\,.$$

Note that by construction of the forced-sampling sets, for each round of forced-sampling $n \in \{0,1,2,...\}$, the time periods $t \in [2^nKq+1, (2^{n+1}-1)Kq]$ are played contiguously without any forced-sampling (or updates to the forced-sampling estimators). Then, for any time $t \notin \{\mathcal{T}_{j,t'}\}_{j\in[K],t'\in[T]}$ where we do not perform forced sampling, let $n_t$ be the largest integer satisfying $t > 2^{n_t+1}Kq$. Given arm $i$, we can define

$$M_{i,t} \equiv \sum_{r=2^{n_t}Kq+1}^{(2^{n_t+1}-1)Kq} \mathbb{I}(r \in \mathcal{S}'_{i,t}) + \sum_{r=2^{n_t+1}Kq+1}^{t} \mathbb{I}(r \in \mathcal{S}'_{i,t}) \;<\; \sum_{r=1}^{t} \mathbb{I}(r \in \mathcal{S}'_{i,t}) = |\mathcal{S}'_{i,t}|\,.$$

In other words, $M_{i,t}$ is strictly smaller than $|S_{i,t}|$, because it only considers the time interval

$$V_t = [2^{n_t}Kq+1, (2^{n_t+1}-1)Kq] \cup [2^{n_t+1}Kq+1, t] \;\subset\; [1,t]\,.$$

By construction, the time interval $[2^{n_t}Kq+1, (2^{n_t+1}-1)Kq]$ contains no forced-sampling, and thus the forced-sample estimator is never updated in this period; the same holds for $[2^{n_t+1}Kq+1, t]$. As a result, note that the event $A_r = A_{2^{n_t}Kq}$ for all times $r$ in the first interval, and $A_r = A_{2^{n_t+1}Kq}$ for all times $r$ in the second interval. Thus, we can expand

$$\begin{aligned} M_{i,t} = &\sum_{r=2^{n_t}Kq+1}^{(2^{n_t+1}-1)Kq} \mathbb{I}(A_{2^{n_t}Kq}) \cdot \mathbb{I}(X_r \in U_i) \cdot \mathbb{I}(r \notin \cup_{j\in[K]}\mathcal{T}_{j,t}) \\ &+ \sum_{r=2^{n_t+1}Kq+1}^{t} \mathbb{I}(A_{2^{n_t+1}Kq}) \cdot \mathbb{I}(X_r \in U_i) \cdot \mathbb{I}(r \notin \cup_{j\in[K]}\mathcal{T}_{j,t}) \\ \geq &\, \mathbb{I}(A_{2^{n_t}Kq}) \cdot \mathbb{I}(A_{2^{n_t+1}Kq}) \cdot \sum_{r\in V_t} \mathbb{I}(X_r \in U_i)\,, \end{aligned}$$

We are then left with a sum over independent random variables drawn from $\mathcal{P}_X$, each occurring with probability $p_*$ by Assumption 3. The definition of $n_t$ as the largest integer such that $t > 2^{n_t+1}Kq$ implies that

$$\begin{aligned} |V_t| &= (t - 2^{n_t+1}Kq) + (2^{n_t+1}Kq - Kq - 2^{n_t}Kq) \;=\; t - 2^{n_t}Kq - Kq \\ &> t - t/2 - Kq \;=\; t/2 - Kq \\ &> 3t/8\,, \end{aligned}$$

where we note that $t > 8Kq$ for $t \geq C_5$. Next, note that since $t \geq C_5 > 4(Kq)^2$ and $t < 2^{n_t+2}Kq$ by definition of $n_t$, we can write

$$2^{n_t}Kq > t/4 \geq (Kq)^2 \,,$$

allowing us to apply Lemma EC.12 for $r = 2^{n_t}Kq$ and $r = 2^{n_t+1}Kq$. A simple union bound yields

$$\begin{aligned}\Pr[A_{2^{n_t}Kq} \text{ and } A_{2^{n_t+1}Kq}] &\geq 1 - \frac{5K}{(t/4)^2} - \frac{5K}{(t/2)^2} \\ &\geq 1 - \frac{100}{K^3q^4} > 8/9\,,\end{aligned}$$

by the definition of $q$. Then,

$$\begin{aligned}\mathbb{E}[M_{i,t}] &\geq \Pr[A_{2^{n_t}Kq} \text{ and } A_{2^{n_t+1}Kq}] \cdot p_*|V_t| \\ &\geq \frac{tp_*}{3}\,,\end{aligned}$$

Applying the Hoeffding inequality, we have that

$$\begin{aligned}\Pr\left[M_{i,t} < \mathbb{E}[M_{i,t}] - \eta\right] &\leq e^{-2\eta^2/|V_t|} \\ &\leq e^{-4\eta^2/t}\,,\end{aligned}$$

Plugging in $\eta = tp_*/12$, we have

$$\Pr\left[M_{i,t} < tp_*/4\right] \leq e^{-tp_*^2/36}\,.$$

Applying the fact that $M_{i,t} \leq |S'_{i,t}|$, we have that

$$\Pr\left[|\mathcal{S}'_{i,t}| < \frac{t\,p_*}{4}\right] \leq e^{-tp_*^2/36}\,.$$

□

Now, we are ready to prove Proposition 3.

*Proof of Proposition 3:* From Lemma EC.14, for $t \geq C_5$ we have $|\mathcal{S}'_{i,t}| \geq p_*t/4$ with probability $1 - \exp[-t\,p_*^2/128]$. Therefore, using a union bound, we can apply Lemma 1 with $p = p_*/2$, $\mathcal{A} = \mathcal{S}_{i,t}$, $\mathcal{A}' = \mathcal{S}'_{i,t}$, and $\lambda = \chi\phi_0^2 p_*/(32s_0)$ to arrive at,

$$\begin{aligned}&\Pr\left[\|\hat{\beta}\left(\mathcal{S}_{i,t},\lambda\right) - \beta_i\|_1 > \chi\right] \\ &\quad\leq 2\exp\left[-C_1\left(\frac{\phi_0\sqrt{p_*/2}}{2}\right)\frac{tp_*}{4}\chi^2 + \log d\right] + \exp\left[-\frac{tp_*^2C_2(\phi_0)^2}{16}\right] + \exp\left[-\frac{tp_*^2}{36}\right] \\ &\quad\leq 2\exp\left[-\frac{tp_*^3C_1(\phi_0)}{256}\chi^2 + \log d\right] + 2\exp\left[-\frac{tp_*^2C_2(\phi_0)^2}{32}\right]\,,\end{aligned}$$

where the last inequality uses $C_2(\phi_0) \le 1/2$. Note that the condition $|\mathcal{A}'|/|\mathcal{A}| \ge p_*/4$ holds when $|\mathcal{S}'_{i,t}| \ge p_* t/4$ (since $|\mathcal{A}| \le t$). Also, the condition $|\mathcal{S}_{i,t}| \ge 6\log(d)/(p_* C_2(\phi_0)^2)$ is satisfied, using $|\mathcal{S}_{i,t}| \ge |\mathcal{S}'_{i,t}| \ge p_* t/4$, $t \ge C_5$, $q \ge 4q_0$, and the definition of $q_0$. Taking

$$\chi = 16\sqrt{\frac{\log t + \log d}{t p_*^3 C_1(\phi_0)}},$$

gives us the desired result. Note that this choice of $\chi$ implies $\lambda = \lambda_{2,t}$. □

## EC.5. Bounding the Regret in the High-Dimensional Setting

Recall from our proof strategy in §4.4, that we divide our time steps $[T]$ into three groups:

(a) Initialization ($t \le C_5$) or forced sampling ($t \in \mathcal{T}_{i,T}$ for some $i \in [K]$).

(b) Times $t > C_5$ when the event $A_{t-1}$ does not hold.

(c) Times $t > C_5$ when the event $A_{t-1}$ holds and we do not perform forced sampling.

These groups may not be disjoint but their union contains $[T]$.

We now compute an upper bound on the regret for time periods in each group (a)-(c) and sum the results. First, the following lemma gives the worst-case regret for time periods in (a).

Lemma EC.15. *The cumulative expected regret of the LASSO Bandit from initialization ($t < C_5$) and forced sampling ($t \in \mathcal{T}_{i,t}$ for some $i \in [K]$) up to time $T$ is at most*

$$2Kqbx_{\max}(6\log T + C_5).$$

*Proof of Lemma EC.15:* From Lemma EC.8, at most $6Kq\log T$ forced samples occur up to time $T$. We also have $C_5$ initialization samples. Using Cauchy-Schwarz, we can bound the worst-case regret in each time period by $\max_{i,j} X^\top(\beta_i - \beta_j) \le 2bx_{\max}$. The result follows directly. □

Before moving to time periods in (b)-(c), we state the following helpful lemma:

Lemma EC.16. *If $f$ is a monotone decreasing and integrable function on the range $[r-1, s]$, then*

$$\sum_{t=r}^{s} f(t) \le \int_{r-1}^{s} f(t)dt.$$

*Proof of Lemma EC.16:*

$$\sum_{t=r}^{s} f(t) \le \sum_{t=r}^{s} \int_{t-1}^{t} f(\tilde{t})d\tilde{t} = \int_{r-1}^{s} f(t)dt. \quad \square$$

Next, we find the worst-case regret from time periods in (b) at time $T$.

Lemma EC.17. *The cumulative expected regret of LASSO Bandit from time periods $C_5 < t \le T$ where $A_{t-1}$ does not hold is at most $2Kbx_{\max}$.*

*Proof of Lemma EC.17* From Lemma EC.12, the probability that $A_{t-1}$ does not hold is at most $5Kt^{-4} \le Kt^{-3}$ since $t \ge C_5 > 5$. Now we can sum this quantity for $t \in [C_5, T-1]$. Using Lemma EC.16,

$$\sum_{t=C_5}^{T-1} \frac{K}{t^3} \le K \int_1^T \frac{1}{t^3}\, dt \le \frac{K}{2}(1 - \frac{1}{T^2}) \le K\,.$$

As before, the worst-case regret at time $t$ is $2b\,x_{\max}$, and the result follows. □

Before analyzing the regret from group (c), we show that if the event $A_{t-1}$ holds, then the set $\hat{\mathcal{K}}$ chosen by the forced-sample estimator has two desirable properties: (i) it contains the true optimal arm, and (ii) it does not contain any sub-optimal arms. Thus, we can apply the convergence properties of the all-sample estimator (which only hold among optimal arms) to analyze the regret from choosing an arm within $\hat{\mathcal{K}}$.

Lemma EC.18. *If $A_{t-1}$ holds, then the set $\hat{\mathcal{K}}$ contains the optimal arm $i^* = \arg\max_{i\in[K]} X_t^\top \beta_i$ and no sub-optimal arms from the set $\mathcal{K}_{sub}$.*

*Proof of Lemma EC.18* To simplify notation, we call our forced-sample arm estimators $\hat{\beta}(\mathcal{T}_{i,t-1}, \lambda_1)$ at time $t$ as $\hat{\beta}_i$. Since $A_{t-1}$ holds, we have that for any pair of arms $i, j \in [K]$,

$$\begin{aligned} X_t^\top \hat{\beta}_i - X_t^\top \hat{\beta}_j &= X_t^\top(\hat{\beta}_i - \beta_i) + X_t^\top(\beta_j - \hat{\beta}_j) + X^\top(\beta_i - \beta_j) \\ &\le h/2 + X_t^\top(\beta_i - \beta_j)\,. \end{aligned}$$

Thus, if we let $i = \arg\max_{\ell\in[K]} X_t^\top \hat{\beta}_\ell$ and $j = i^*$, we see that $X_t^\top(\hat{\beta}_i - \hat{\beta}_{i^*}) \le h/2$ since $X_t^\top(\beta_i - \beta_{i^*}) < 0$ (by definition of $i^*$). Then, the optimal arm $i^* \in \hat{\mathcal{K}}$.

On the other hand, consider $i = \arg\max_{\ell\in[K]} X_t^\top \hat{\beta}_\ell$ and any sub-optimal arm $j \in \mathcal{K}_{sub}$. Then, $X_t^\top \hat{\beta}_i - X_t^\top \hat{\beta}_j \ge X_t^\top \hat{\beta}_{i^*} - X_t^\top \hat{\beta}_j$, and furthermore, since $A_{t-1}$ holds:

$$\begin{aligned} X_t^\top \hat{\beta}_{i^*} - X_t^\top \hat{\beta}_j &= X_t^\top(\hat{\beta}_{i^*} - \beta_{i^*}) + X_t^\top(\beta_j - \hat{\beta}_j) + X^\top(\beta_{i^*} - \beta_j) \\ &\ge -h/2 + X_t^\top(\beta_{i^*} - \beta_j)\,. \end{aligned}$$

Recall that for every sub-optimal arm $j \in \mathcal{K}_{sub}$, we have $X_t^\top \beta_j < X_t^\top \beta_{i^*} - h$. Then, we can write

$$\begin{aligned} X_t^\top(\hat{\beta}_i - \hat{\beta}_j) &\ge X_t^\top \hat{\beta}_{i^*} - X_t^\top \hat{\beta}_j \\ &> -h/2 + h \; = \; h/2\,. \end{aligned}$$

Thus, $j \notin \hat{\mathcal{K}}$ for every sub-optimal arm $j \in \mathcal{K}_{sub}$. □

Finally, the next two lemmas bound the regret from time periods in (c) by separately summing over expected regret when the all-sample tail inequality does and does not hold. We simplify our notation by calling our all-sample estimators $\hat{\beta}(\mathcal{S}_{i,t-1}, \lambda_{2,t-1})$ at time $t$ as $\hat{\beta}_i$, where we recall $\lambda_{2,t} = [\phi_0^2/(2s_0)]\sqrt{(\log t + \log d)/(p_* C_1 t)}$.

Lemma EC.19. *If $t > C_5$, $A_t$ holds, and we do not perform forced sampling, then the expected regret at time $t+1$ is bounded by*

$$(8Kbx_{\max})/t + 8Kbx_{\max}\exp\left[-(p_*^2 C_2(\phi_0)^2 t)/32\right] + C_3(\phi_0, p_*)\cdot(\log t + \log d)/t\,,$$

*where $C_3(\phi_0,p_*) = 1024KC_0x_{\max}^2/(p_*^3 C_1(\phi_0))$.*

*Proof of Lemma EC.19* Without loss of generality, assume that arm 1 is optimal: $\arg\max_{i\in[K]} X_{t+1}^\top \beta_i = 1$. Then, the expected regret at time $t+1$ is given by

$$r_{t+1} = \mathbb{E}\left(\sum_{i\in\hat{\mathcal{K}}} X_{t+1}^\top(\beta_1-\beta_i)\mathbb{I}[\text{choose arm } i]\right) \le \mathbb{E}\left(\sum_{i\in\hat{\mathcal{K}}} X_{t+1}^\top(\beta_1-\beta_i)\mathbb{I}\left[X_{t+1}^\top\hat\beta_i \ge X_{t+1}^\top\hat\beta_1\right]\right)\,,$$

where the last inequality uses the fact that event $\{i = \arg\max_{j\in[K]} X_{t+1}^\top\hat\beta_j\}$ is a subset of the event $\{X_{t+1}^\top\hat\beta_i \ge X_{t+1}^\top\hat\beta_1\}$, and that $X_{t+1}^\top(\beta_1-\beta_i)\ge 0$ (since we have assumed that arm 1 is optimal). Thus, we can bound $r_{t+1}$ through the regret incurred by each arm in $\hat{\mathcal{K}}$ with respect to the optimal arm independently of the other arms. We now define the event $B_i \equiv \left\{X_{t+1}^\top(\beta_1-\beta_i) > 2\delta x_{\max}\right\}$, where we take $\delta \equiv 16\sqrt{(\log t+\log d)/(p_*^3 C_1 t)}$. Then, we can write

$$\begin{aligned} r_{t+1} \le\; & \mathbb{E}\left(\sum_{i\in\hat{\mathcal{K}}} X_{t+1}^\top(\beta_1-\beta_i)\mathbb{I}\left[(X_{t+1}^\top\hat\beta_i \ge X_{t+1}^\top\hat\beta_1)\cap B_i\right]\right) \\ & + \mathbb{E}\left(\sum_{i\in\hat{\mathcal{K}}} X_{t+1}^\top(\beta_1-\beta_i)\mathbb{I}\left[(X_{t+1}^\top\hat\beta_i \ge X_{t+1}^\top\hat\beta_1))\cap B_i^c\right]\right)\,, \end{aligned}$$

which by definition of $B_i$ and using $X_{t+1}^\top(\beta_1-\beta_i)\le 2b\,x_{\max}$ gives

$$r_{t+1} \le 2b\,x_{\max}\mathbb{E}\left[\sum_{i\in\hat{\mathcal{K}}}\mathbb{I}[(X_{t+1}^\top\hat\beta_i \ge X_{t+1}^\top\hat\beta_1)\cap B_i]\right] + 2\delta\,x_{\max}\mathbb{E}\left[\sum_{i\in\hat{\mathcal{K}}}\mathbb{I}(B_i^c)\right]\,, \tag{EC.2}$$

Note that the intersection of event $B_i$ and the event of choosing arm $i\neq 1$ implies that

$$0 \ge X_{t+1}^\top\hat\beta_1 - X_{t+1}^\top\hat\beta_i > X_{t+1}^\top\left(\hat\beta_1-\beta_1\right) + X_{t+1}^\top\left(\beta_i-\hat\beta_i\right) + 2\delta\,x_{\max}\,.$$

Thus, it must be that either $X_{t+1}^\top\left(\hat\beta_1-\beta_1\right) < -\delta\,x_{\max}$ or $X_{t+1}^\top\left(\beta_i-\hat\beta_i\right) < -\delta\,x_{\max}$. Therefore,

$$\begin{aligned} \Pr\left[(X_{t+1}^\top\hat\beta_i \ge X_{t+1}^\top\hat\beta_1)\cap B_i\right] &\le \Pr\left[\|\beta_1-\hat\beta_1\|_1 > \delta\right] + \Pr\left[\|\hat\beta_i-\beta_i\|_1>\delta\right] \\ &\le \frac{4}{t} + 4\exp\left[-\frac{p_*^2C_2^2}{32}t\right]\,, \end{aligned} \tag{EC.3}$$

using a union bound and the tail inequality for the all sample estimator.

We can also bound $\Pr[B_i^c]$ using Assumption 2 on the margin condition: $\Pr[B_i^c] = \Pr\left[X_{t+1}^\top(\beta_1-\beta_i)\le 2\delta\, x_{\max}\right]\le 2C_0\delta\, x_{\max}$. Using this and Eq. (EC.3) in Eq. (EC.2) we obtain

$$r_{t+1}\le K\left\{\frac{8b\,x_{\max}}{t}+8bx_{\max}\exp\left[-\frac{p_*^2C_2^2}{32}t\right]+4C_0\delta^2x_{\max}^2\right\}$$
$$\le \frac{8Kbx_{\max}}{t}+8Kbx_{\max}\exp\left[-\frac{p_*^2C_2^2}{32}t\right]+C_3\cdot\frac{\log t+\log d}{t}. \quad \square$$

Lemma EC.20. *The cumulative expected regret from using the all-sample estimator up to time $T$ is bounded by*

$$[8Kbx_{\max}+C_3(\phi_0,p_*)\log d]\log T+C_3(\phi_0,p_*)\left(\log T\right)^2+C_4(\phi_0,p_*)\,,$$

*where* $C_4(\phi_0,p_*)=(8Kbx_{\max})/(1-\exp\left[-\frac{p_*^2C_2(\phi_0)^2}{32}\right])$.

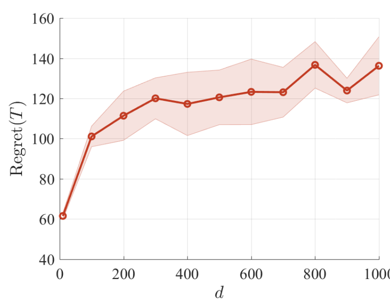
*Proof of Lemma EC.20* We sum regret from Lemma EC.19

$$\sum_{t=C_5}^{T-1}\left[(8Kbx_{\max})/t+8Kbx_{\max}\exp\left[-\frac{p_*^2C_2(\phi_0)^2}{32}t\right]+C_3(\phi_0,p_*)(\log t+\log d)/t\right]$$
$$\le[8Kbx_{\max}+C_3(\phi_0,p_*)\log d]\log T+C_3(\phi_0,p_*)\left(\log T\right)^2+C_4(\phi_0,p_*)\,. \quad \square$$

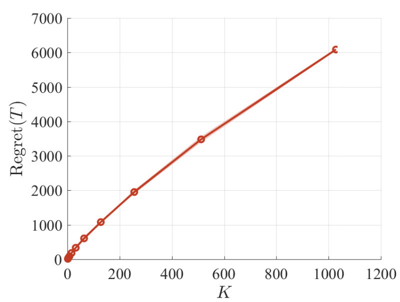
## EC.6. Additional Simulations

### EC.6.1. Dependence on $K$, $d$, and $s_0$

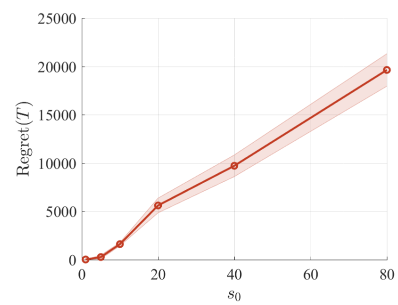
First, we study how the LASSO Bandit's cumulative expected regret scales as a function of each parameter $K$, $d$, and $s_0$. The results (see Figure EC.1) show that the regret appears to grow logarithmically with $d$, but almost linearly with $K$ and $s_0$.




(a) $K=10,\ s_0=2$




(b) $d=20,\ s_0=2$




(c) $d=500,\ K=2$

**Figure EC.1** These plots show how the regret of the LASSO Bandit scales as any single parameter $d, K, s_0$ is varied while the others are fixed.

### EC.6.2. Robustness to Algorithm Inputs

We now study the cumulative expected regret of the LASSO Bandit while varying any one of: (i) the forced sampling parameter $q \in \{1, 2, 5\}$, (ii) the localization parameter $h \in \{1, 5, 25\}$, and (iii) the regularization coefficient $c \in \{0.02, 0.05, 0.1, 0.2\}$. We only focus on scenario (a) from above.

The results are computed over $T = 10,000$ time steps and averaged over 30 trials (see Figure EC.2). We find that the cumulative regret performance is not substantially impacted despite experimenting with the parameters by up to an order of magnitude; this suggests that the LASSO Bandit is robust, which is important since input parameters are likely to be misspecified in practice.

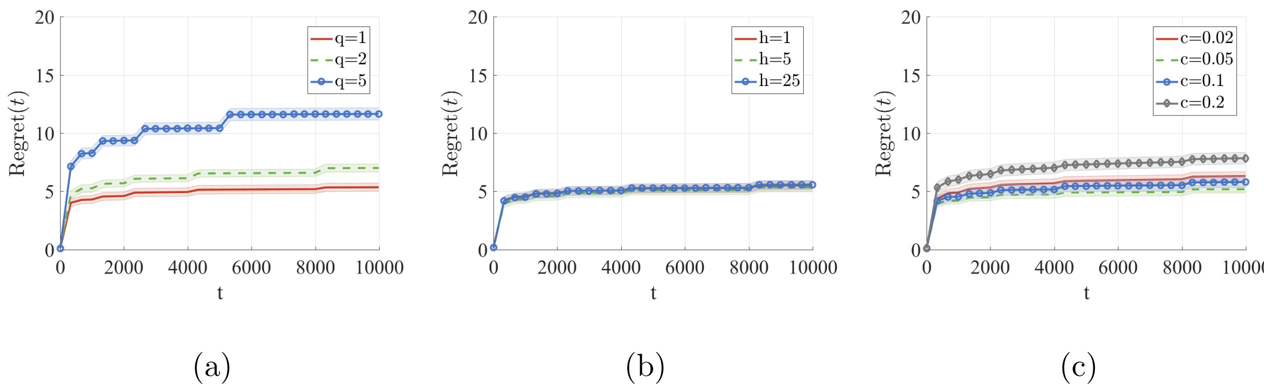



**Figure EC.2** Cumulative regret for the LASSO Bandit for varying values of (a) the forced sampling parameter $q$, (b) the localization parameter $h$, and (c) the coefficient $c$ of the regularization parameters.

### EC.6.3. Nonlinear Reward Function

Another interesting direction is considering nonlinear reward functions. The LASSO Bandit can be used in conjunction with basis expansion methods from statistical learning to approximate any nonlinear function (Hastie et al. 2001). Specifically, given a covariate vector $X = (x_1, \ldots, x_d)$, we can consider a large vector with length $O(d^n)$ consisting of all *distinct* monomials of maximum degree at most $n$, denoted by $X\otimes_n$. Then we can use a linear model with covariate vector $X\otimes_n$ to approximate a reward function that is up to a $n^{th}$-degree polynomial in $X$. Assuming that the true model is a sparse function of these monomials (i.e., the reward only depends on $s_0$ entries of $X\otimes_n$), the LASSO Bandit algorithm could be employed. From a theoretical perspective, one has to study the behavior of the constant $\phi_0$ for the compatibility condition of the covariance matrix of $X\otimes_n$ in order to prove theoretical guarantees for this approach; however, such an analysis is beyond the scope of this paper and we simply empirically test the approach. We repeat scenario (a) from above ($K = 2$, $d = 100$, $s_0 = 5$) where the true reward function of each arm is a (distinct) polynomial of degree $n = 3$. We compare two versions of the LASSO Bandit: (1) naïve-LASSO Bandit that uses only the raw covariates $X$ and does not expand them to $X \otimes_3$, and (2) NL-LASSO Bandit that uses $X \otimes_3$. For comparison, we also include a nonlinear version of the other bandit algorithms that, similar to NL-LASSO Bandit, use the expanded covariate vector and refer to them by NL-OLS-Bandit, NL-OFUL-LS, and NL-OFUL-EG respectively. Figure EC.3 shows the results. NL-LASSO Bandit outperforms all other methods. It is interesting to see that the naïve-LASSO Bandit is competitive for small $t$ since it avoids overfitting more effectively with a smaller covariate space; however, the regret is linear since its model is misspecified, and it loses to the other approaches as $T$ grows.

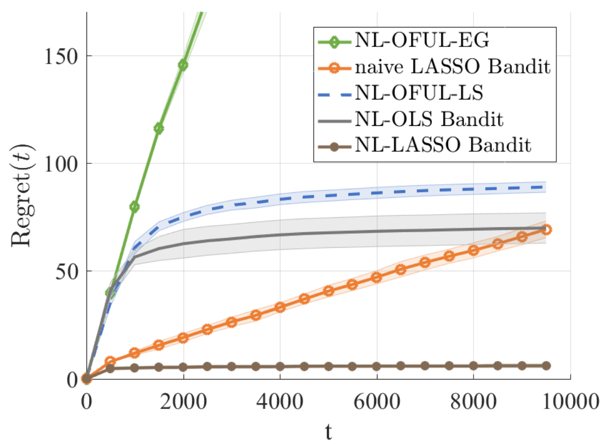



**Figure EC.3** Comparison of all methods with parameters as in Figure 1(a), but the reward functions are polynomials of degree 3. The suffix NL means that the algorithm uses the expanded version of the covariate vector that contains all monomials of degree at most 3.

### EC.6.4. When Assumptions 2 and 3 Fail

We now study two settings where some of the assumptions required by our theory fail. First, we look at Assumption 3 (arm optimality). We consider $K = 3$, $d = 100$, $s_0 = 2$, $\beta_1 = [1, 1, 0, \dots, 0]$, $\beta_2 = [0, 0, 1, 1, 0, \dots, 0]$, and $\beta_3 = r(\beta_1 + \beta_2)/2$ for $r \in \{0.9, .99, 1, 1.01, 1.1\}$. In this situation, when $r$ is close or equal to 1, the arm optimality condition fails for arm 3. Figure EC.4 shows that performance of the LASSO Bandit is robust as $r$ varies around 1. However, there is a small but noticeable loss when $r = 1.1$, which is due to the failure of the arm optimality condition. In particular, there are cases where arm 3 is the optimal arm and we incur regret if it is not played; however, the magnitude of this regret (relative to pulling the second best arm) is small, making the overall loss from the assumption's failure small. This simulation suggests that Assumption 3 could possibly be relaxed at the expense of a more cumbersome regret analysis.

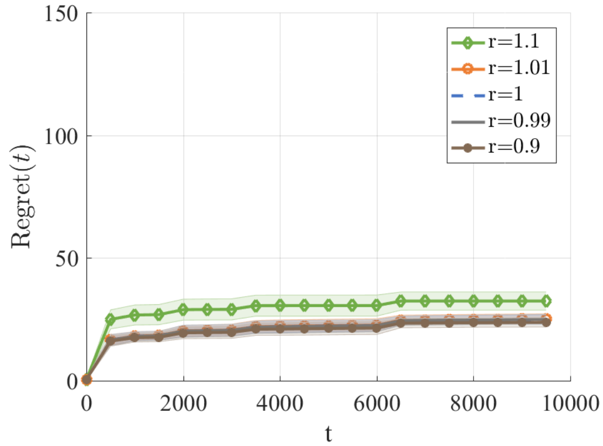



**Figure EC.4** Performance of LASSO Bandit when arm optimality condition fails (when an arm is $r$ times a convex combination of the other two arms).

Next, we study Assumption 2 (the margin condition). We consider $K=2$, $d=10$, $s_0=3$, $\beta_1=[1,0,1,0,\ldots,0]$ and $\beta_2=[1,1,1,0,\ldots,0]$. The covariates are generated according to the following procedure. First, a $(d-2)$-dimensional vector $\tilde{X}$ is sampled from the truncated normal (as above). Then, we add two coordinates at the beginning based on sampling a random variable $U$ from the uniform distribution on $[-1,1]$, independent of $\tilde{X}$. Then, our $d$-dimensional covariate vector $X$ is given by

$$X(r)=\begin{cases}1 & \text{if } r=1\\ \text{sign}(U)|U|^{1+\epsilon} & \text{if } r=2\\ \tilde{X}(r) & \text{if } r>2\,.\end{cases}$$

Now, note that

$$\begin{aligned}\Pr\left[0<|X^\top(\beta_1-\beta_2)|<\kappa\right] &= \Pr\left[0<|X(2)|<\kappa\right]\\ &= \Pr\left[0<|U|^{1+\epsilon}<\kappa\right]=2(\kappa)^{\frac{1}{1+\epsilon}}>\kappa^{\frac{1}{1+\epsilon}}\,,\end{aligned}$$

which implies that the margin condition fails for any $\epsilon>0$. We simulate the LASSO Bandit for $\epsilon\in\{0,1\}$ and the results are shown in Figure EC.5. When $\epsilon=0$ (margin condition holds) the regret grows at a slower rate than when $\epsilon=1$ (margin condition fails). In fact, when $d=1$, Goldenshluger and Zeevi (2009) prove a lower bound that the regret scales as $\mathcal{O}\left(T^{\frac{\epsilon}{2(1+\epsilon)}}\right)$ for $\epsilon>0$. Generalizing a variant of their result to $d>1$ is an open direction, but matches our simulation results.

## EC.7. OLS Bandit Algorithm and Analysis

In this section, we propose the OLS Bandit, which is a variant of the algorithm by Goldenshluger and Zeevi (2013) for the low-dimensional setting. We then apply the analytical tools we developed in the proof of the LASSO Bandit to prove an upper bound of $\mathcal{O}\left(d^2\log^{\frac{3}{2}}d\cdot\log T\right)$ on the cumulative expected regret of the OLS Bandit; this is an improvement over the existing $\mathcal{O}(d^3\log T)$ bound.

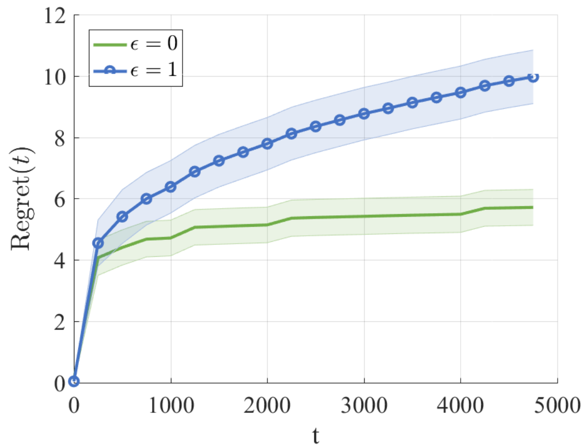



**Figure EC.5** Performance of the LASSO Bandit when the margin condition holds ($\epsilon = 0$) versus when it fails ($\epsilon = 1$). In the latter case, the probability of observing covariate vectors that lie within a margin $\kappa$ of the decision boundary is of order $\sqrt{\kappa}$.

Remark EC.1. Our analysis yields a better bound because we employ matrix martingale concentration results (Tropp 2015) to bound the difference of the true and sample covariance matrices, i.e., $\|\hat{\Sigma} - \Sigma\|_\infty$; in contrast, Goldenshluger and Zeevi (2013) rely on applying the union bound, which contributes an extra factor of $d$.

**Assumptions.** We make similar but weaker assumptions on the problem formulation as Goldenshluger and Zeevi (2013). In particular, prior work only allowed for two arms and required each arm to be optimal for some subset of users; in contrast, our formulation tackles the $K$-armed bandit and further allows for some arms $\mathcal{K}_{sub}$ to be uniformly sub-optimal.

Consequently, we make the same assumptions as that of the LASSO Bandit (including Assumptions 1-3 in §2.1) but we replace Assumption 4 on the LASSO compatibility condition with the following stronger requirement of positive-definiteness:

Assumption EC.1 **(Positive-Definiteness)**. *Define* $\Sigma_i \equiv \mathbb{E}[XX^\top \mid X \in U_i]$ *for all* $i \in [K]$. *Then, there exists a deterministic constant* $\phi_0 \in \mathbb{R}^+$ *such that for all* $i \in [K]$ *the minimum eigenvalue* $\lambda_{\min}(\Sigma_i) \geq \phi_0^2 > 0$.

**OLS Estimation.** Recall the notation we established in §3.1. Consider a linear model $Y = \mathbf{X}\beta + \varepsilon$, with design matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$, response vector $Y \in \mathbb{R}^n$, and noise vector $\varepsilon \in \mathbb{R}^n$ whose entries are independent $\sigma$-subgaussian random variables.

Definition EC.1 (OLS). If $\hat{\Sigma}(\mathbf{X}) = \mathbf{X}^\top \mathbf{X}/n$ is positive definite, the OLS estimator for the parameter $\beta$ is defined by:

$$\hat{\beta}_{\mathbf{X},Y} \equiv \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top Y . \tag{EC.4}$$

The OLS estimator converges with high probability according to the following tail inequality.

Proposition EC.1 **(OLS Tail Inequality)**. *Let $X_t$ denote the $t^{th}$ row of $\mathbf{X}$ and $Y(t)$ denote the $t^{th}$ entry of $Y$. Also, assume that $\{X_t : t = 1, ..., n\}$ forms an adapted sequence of observations, i.e., $X_t$ may depend on past regressors and their resulting observations $\{X_{t'}, Y(t')\}_{t'=1}^{t-1}$. If all realizations of random variables $X_t$ satisfy $\|X_t\|_\infty \le x_{\max}$, the following tail inequality holds for all $\chi > 0$ and all constants $\phi > 0$:*

$$\Pr\left[\|\hat{\beta} - \beta\|_1 \le \chi\right] \ge 1 - \exp\left[-\tilde{C}_1(\phi) n\chi^2 + \log 2d\right] - \Pr\left[\lambda_{min}\left(\hat{\Sigma}(\mathbf{X})\right) \le \phi^2\right],$$

*where we define $\tilde{C}_1(\phi) \equiv \phi^4 / (2d^2 x_{\max}^2 \sigma^2)$.*

**Algorithm.** We introduce the OLS Bandit algorithm below (Algorithm 2), which proceeds analogously to the LASSO Bandit (Algorithm 1). In particular, we define and use the forced-sample sets $\mathcal{T}_{i,t}$ and all-sample sets $\mathcal{S}_{i,t}$ in the same way. The key difference is that we now use OLS instead of LASSO estimation (note that we no longer require a path of regularization parameters).

---

**Algorithm** OLS Bandit

---

**Input parameters:** $q, h$
Initialize $\mathcal{T}_{i,0}$ and $\mathcal{S}_{i,0}$ by the empty set, and $\hat{\beta}(\mathcal{T}_{i,0})$ and $\hat{\beta}(\mathcal{S}_{i,0})$ by $0 \in \mathbb{R}^d$ for all $i$ in $[K]$
Use $q$ to construct force-sample sets $\mathcal{T}_i$ using Eq. (2) for all $i$ in $[K]$
**for** $t \in [T]$ **do**
  Observe $X_t \sim \mathcal{P}_X$
  **if** $t \in \mathcal{T}_i$ for any $i$ **then**
    $\pi_t \leftarrow i$
  **else**
    $\hat{\mathcal{K}} = \left\{i \in K \mid X_t^\top \hat{\beta}(\mathcal{T}_{i,t-1}) \ge \max_{j\in[K]} X_t^\top \hat{\beta}(\mathcal{T}_{j,t-1}) - h/2\right\}$
    $\pi_t \leftarrow \arg\max_{i\in\hat{\mathcal{K}}} X_t^\top \hat{\beta}(\mathcal{S}_{i,t-1})$
  **end if**
  $\mathcal{S}_{\pi_t,t} \leftarrow \mathcal{S}_{\pi_t,t-1} \cup \{t\}$
  Play arm $\pi_t$, observe $Y(t) = X_t^\top \beta_{\pi_t} + \varepsilon_{i,t}$
**end for**

---

### EC.7.1. New Upper Bound on Regret of OLS Bandit

Theorem EC.1. *When $q \ge 4\lceil \tilde{q}_0 \rceil$, $K \ge 2$, $d > 2$, and $T \ge (Kq)^2$, we have an upper bound on the expected cumulative regret at time $T$:*

$$\begin{aligned} R_T &\le 2qKbx_{\max}(6\log T + Kq) + 2Kbx_{\max} + \frac{8K\max(C_0,1)x_{\max}^2[\log(12d)]^{3/2}}{\tilde{C}_3}\log T + \tilde{C}_4 Kbx_{\max} \\ &= \mathcal{O}\left(d^2[\log d]^{3/2} \cdot \log T\right), \end{aligned}$$

*where constants $\tilde{C}_1 = \tilde{C}_1(\phi_0)$, $\tilde{C}_2 = \tilde{C}_2(\phi_0)$, $\tilde{C}_3 = \tilde{C}_3(\phi_0, p_*)$, and $\tilde{C}_4 = \tilde{C}_4(\phi_0, p_*)$ are defined by*

$$\tilde{C}_1 \equiv \frac{\phi_0^4}{2d^2 x_{\max}^2 \sigma^2},\ \tilde{C}_2 \equiv \min(\frac{1}{2}, \frac{\phi_0^2}{8x_{\max}^2}),\ \tilde{C}_3 \equiv \frac{p_*^3 \tilde{C}_1}{256},\ \text{and } \tilde{C}_4 \equiv \frac{8}{1 - \exp\left[-\tilde{C}_2 \frac{p_*^2}{64}\right]},$$

$C_0$ *is defined in Assumption 2, and we take*

$$\tilde{q}_0 \equiv \max\left\{\frac{20}{p_*}, \frac{8\log d}{p_*\tilde{C}_2}, \frac{1024x_{\max}^2 \log 2d}{h^2 p_*^2 \tilde{C}_1}\right\} = \mathcal{O}\left(d^2 \log d\right) .$$

**Key Steps.** The proof strategy is similar to that of the LASSO Bandit. First, we prove a technical lemma (analogous to Lemma 1) that shows a tail inequality holds for the OLS estimator if only a constant (but unknown) fraction of the rows of the design matrix are independent (Lemma EC.21). We use this lemma to prove analogous tail inequalities for the forced-sample estimator (Proposition EC.2) and the all-sample estimator (Proposition EC.3) in §EC.7.4. Finally, we use these tail inequalities to sum up the expected regret contributions from the three groups of time periods:

(a) Initialization ($t \leq (Kq)^2$) and forced sampling ($t \in \mathcal{T}_{i,T}$ for some $i \in [K]$).
(b) Times $t > (Kq)^2$ when the event $A_{t-1}$ does not hold.
(c) Times $t > (Kq)^2$ when the event $A_{t-1}$ holds and we do not perform forced sampling, i.e., the OLS Bandit plays the estimated best arm from $\hat{\mathcal{K}}$ using the all-sample estimator.

Summing the results concludes the proof of Theorem EC.1. The proof is given in §EC.7.5.

### EC.7.2. OLS Tail Inequality for Adapted Observations

*Proof of Proposition EC.1* For simplicity, we start with the $\ell_2$ norm. Note that, if the event $\lambda_{\min}\left(\hat{\Sigma}(\mathbf{X})\right) > \phi^2$ holds,

$$\begin{aligned}
\|\hat{\beta} - \beta\|_2 &= \|(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\varepsilon\|_2 \\
&\leq \|(\mathbf{X}^\top\mathbf{X})^{-1}\|_2 \cdot \|\mathbf{X}^\top\varepsilon\|_2 \\
&= \frac{1}{n\phi^2}\|\mathbf{X}^\top\varepsilon\|_2 .
\end{aligned}$$

Then, for any $\tilde{\chi} > 0$, we can write

$$\begin{aligned}
\Pr\left[\|\hat{\beta} - \beta\|_2 \leq \tilde{\chi}\right] &\geq \Pr\left[\left(\|\mathbf{X}^\top\varepsilon\|_2 \leq n\tilde{\chi}\phi^2\right) \cap \left(\lambda_{\min}\left(\hat{\Sigma}(\mathbf{X})\right) > \phi^2\right)\right] \\
&\geq 1 - \sum_{r=1}^{d} \Pr\left[|\varepsilon^\top X^{(r)}| > \frac{n\tilde{\chi}\phi^2}{\sqrt{d}}\right] - \Pr\left[\lambda_{\min}\left(\hat{\Sigma}(\mathbf{X})\right) \leq \phi^2\right] ,
\end{aligned}$$

where we have let $X^{(r)}$ denote the $r^{th}$ column of $\mathbf{X}$. We can expand $\varepsilon^\top X^{(r)} = \sum_{t\in[n]} \varepsilon(t) X_t(r)$, where we note that $D_{t,r} \equiv \varepsilon(j)X_j(r)$ is a $(x_{\max}\sigma)$-subgaussian random variable by Definition 1, conditioned on the sigma algebra $\mathfrak{S}_{t-1}$ that is generated by random variables $X_1, \ldots, X_{t-1}, Y(1), \ldots, Y(t-1)$. Defining $D_{0,r} = 0$, the sequence $D_{0,r}, D_{1,r}, \ldots, D_{n,r}$ is a martingale difference sequence adapted to the filtration $\mathfrak{S}_1 \subset \ldots \subset \mathfrak{S}_n$ since $\mathbb{E}[\varepsilon(t)X_t(r)|\mathfrak{S}_{t-1}] = 0$. Using Lemma EC.1,

$$\begin{aligned}
\Pr\left[\|\hat{\beta} - \beta\|_2 \leq \tilde{\chi}\right] &\geq 1 - \sum_{r=1}^{d} \Pr\left[|\varepsilon^\top X^{(r)}| > \frac{n\tilde{\chi}\phi^2}{\sqrt{d}}\right] - \Pr\left[\lambda_{\min}\left(\hat{\Sigma}(\mathbf{X})\right) \leq \phi^2\right] \\
&\geq 1 - 2d\exp\left[-\frac{n\tilde{\chi}^2\phi^4}{2dx_{\max}^2\sigma^2}\right] - \Pr\left[\lambda_{\min}\left(\hat{\Sigma}(\mathbf{X})\right) \leq \phi^2\right] .
\end{aligned}$$

Now, to bound the $\ell_1$ norm, we can use Cauchy-Schwarz to write (for $\tilde{\chi} = \chi/\sqrt{d}$)

$$\begin{aligned}
\Pr\left[\|\hat{\beta}-\beta\|_1 \le \chi\right] &\ge \Pr\left[\|\hat{\beta}-\beta\|_2 \le \tilde{\chi}\right] \\
&\ge 1 - 2d\exp\left[-\frac{n\tilde{\chi}^2\phi^4}{2dx_{\max}^2\sigma^2}\right] - \Pr\left[\lambda_{\min}\left(\hat{\Sigma}(\mathbf{X})\right) \le \phi^2\right] \\
&= 1 - \exp\left[-\underbrace{\frac{\phi^4}{2d^2x_{\max}^2\sigma^2}}_{\tilde{C}_1(\phi)} n\chi^2 + \log(2d)\right] - \Pr\left[\lambda_{\min}\left(\hat{\Sigma}(\mathbf{X})\right) \le \phi^2\right]. \quad \square
\end{aligned}$$

### EC.7.3. Positive-Definiteness for non-i.i.d. samples

In this section we prove a tail inequality for OLS with non-i.i.d. data, analogous to the result of §4.1. In particular, consider a linear model $Y = \mathbf{Z}\beta + \varepsilon$, with random design matrix $\mathbf{Z} \in \mathbb{R}^{n\times d}$ such that all realizations of $\mathbf{Z}$ satisfy $\|\mathbf{Z}\|_\infty \le x_{\max}$, response vector $Y \in \mathbb{R}^n$, and noise vector $\varepsilon \in \mathbb{R}^n$ whose entries are independent $\sigma$-subgaussian random variables. Consider a fixed subset $\mathcal{A}$ of $[n]$, and if $\mathcal{A}' \subset \mathcal{A}$ is such that $\{Z_t \mid t \in \mathcal{A}'\}$ is an i.i.d. subset of random variables with distribution $\mathcal{P}_Z$ with $\lambda_{\min}(\mathbb{E}[ZZ^T]) = \phi_1^2$ and $|\mathcal{A}'|/|\mathcal{A}| \ge p/2$ for positive constants $\phi_1$ and $p$. Similar to §3.1, we use the short notation $\hat{\beta}(\mathcal{A})$ and $\hat{\Sigma}(\mathcal{A})$ to refer to the OLS estimator and sample covariance matrix on the set $\mathcal{A}$. In this section, we will show that $\hat{\Sigma}(\mathcal{A})$ is positive-definite with minimum eigenvalue bounded below by $\phi_1^2 p/4 = \left(\phi_1\sqrt{p}/2\right)^2$ with high probability, and then apply Proposition EC.1 to obtain the following result.

Lemma EC.21. *Under the assumptions above, the following tail inequality holds for all $\chi > 0$:*

$$\Pr\left[\|\hat{\beta}(\mathcal{A})-\beta\|_1 \ge \chi\right] \le \exp\left[-\tilde{C}_1\left(\frac{\phi_1\sqrt{p}}{2}\right)|\mathcal{A}|\chi^2 + \log 2d\right] + \exp\left[-p\tilde{C}_2(\phi_1)|\mathcal{A}|/2 + \log d\right],$$

*where $\tilde{C}_1$ and $\tilde{C}_2$ are defined in §EC.7.1.*

Before formally proving Lemma EC.21, we state and prove some results.

First, we will show that $\hat{\Sigma}(\mathcal{A}')$ has minimum eigenvalue bounded below with high probability.

Lemma EC.22. *The minimum eigenvalue of $\hat{\Sigma}(\mathcal{A}')$ is bounded below by $\phi_1^2/2$ with probability* $1 - \exp\left[-\tilde{C}_2(\phi_1)|\mathcal{A}'| + \log d\right]$.

*Proof of Lemma EC.22* First, note that

$$\begin{aligned}
\lambda_{\max}\left(\hat{\Sigma}(\mathcal{A}')\right) &= \max_{\|u\|=1} u^\top \hat{\Sigma}(\mathcal{A}')u \\
&= \max_{\|u\|=1} \frac{1}{|\mathcal{A}'|}\sum_{t\in\mathcal{A}'} (Z_t^\top u)^2 \ \le\ x_{\max}^2
\end{aligned}$$

Then, it follows from the matrix Chernoff inequality, Corollary 5.2 in Tropp (2015), that

$$\Pr\left[\lambda_{\min}(\hat{\Sigma}(\mathcal{A}')) > \frac{\phi_1^2}{2}\right] \geq 1 - d \cdot \exp\left[-\frac{|\mathcal{A}'|\phi_1^2}{8x_{\max}^2}\right] \geq 1 - \exp\left[-\tilde{C}_2(\phi_1)|\mathcal{A}'| + \log d\right]$$

, if we take $\delta = 1/2$ and $R = x_{\max}^2$. □

Lemma EC.23. *If the minimum eigenvalue of $\hat{\Sigma}(\mathcal{A}')$ is bounded below by $\phi_1'^2$, then the minimum eigenvalue of $\hat{\Sigma}(\mathcal{A})$ is bounded below by $\phi_1'^2|\mathcal{A}'|/|\mathcal{A}|$.*

*Proof of Lemma EC.23* From our definition, we can write

$$\begin{aligned}\hat{\Sigma}(\mathcal{A}) &= \frac{|\mathcal{A}'|}{|\mathcal{A}|}\hat{\Sigma}(\mathcal{A}') + \frac{1}{|\mathcal{A}|}\sum_{t\in\mathcal{A}/\mathcal{A}'} Z_t Z_t^\top \\ &= \frac{|\mathcal{A}'|}{|\mathcal{A}|}\hat{\Sigma}(\mathcal{A}') + \frac{|\mathcal{A}\setminus\mathcal{A}'|}{|\mathcal{A}|}\hat{\Sigma}(\mathcal{A}\setminus\mathcal{A}') .\end{aligned}$$

Now, using the fact that the minimum eigenvalue is a concave function, it immediately follows that

$$\begin{aligned}\lambda_{\min}\left(\hat{\Sigma}(\mathcal{A})\right) &\geq \frac{|\mathcal{A}'|}{|\mathcal{A}|}\lambda_{\min}\left(\hat{\Sigma}(\mathcal{A}')\right) + \frac{|\mathcal{A}\setminus\mathcal{A}'|}{|\mathcal{A}|}\lambda_{\min}\left(\hat{\Sigma}(\mathcal{A}\setminus\mathcal{A}')\right) \\ &\geq \frac{|\mathcal{A}'|}{|\mathcal{A}|}\phi_1'^2 ,\end{aligned}$$

where the last inequality relies on the fact that $\hat{\Sigma}(\mathcal{A}\setminus\mathcal{A}')$ is always positive semi-definite. □

Now, we are ready to prove the main result of this section.

*Proof of Lemma EC.21* Combining Lemmas EC.22 and EC.23, and using $|\mathcal{A}'| \geq p|\mathcal{A}|/2$ implies that

$$\Pr\left[\lambda_{\min}\left(\hat{\Sigma}(\mathcal{A})\right) \leq \frac{\phi_1^2 p}{4}\right] \leq \exp\left[-p\tilde{C}_2(\phi_1)|\mathcal{A}|/2 + \log d\right] .$$

Now, we can apply Proposition EC.1 with $\phi = \phi_1\sqrt{p}/2$ to obtain the result. □

### EC.7.4. Proof of Tail Inequalities for OLS Force-Sample and All-Sample Estimators

Proposition EC.2. *When $t \geq (Kq)^2$, the forced sample estimator $\hat{\beta}(\mathcal{T}_{i,t})$ satisfies the tail inequality*

$$\Pr\left[\|\hat{\beta}\left(\mathcal{T}_{i,t}\right) - \beta_i\|_1 > \frac{h}{4x_{\max}}\right] \leq \frac{4}{t^4} .$$

*Proof of Proposition EC.2* Since forced-sampling schedule is the same as LASSO Bandit, using Lemma EC.8, $|\mathcal{T}_{i,t}| \geq (q/2)\log t \geq 2\tilde{q}_0$. Also, by Assumption EC.1, $\Sigma_i$ has minimum eigenvalue bounded below by $\phi_0^2$. If $|\mathcal{T}_{i,t}'|/|\mathcal{T}_{i,t}| \geq p_*/2$, Lemma EC.9 allows us to apply Lemma EC.21, with $\chi = h/(4x_{\max})$, to show that

$$\begin{aligned}&\Pr\left[\|\hat{\beta}\left(\mathcal{T}_{i,t}\right) - \beta_i\|_1 > \frac{h}{4x_{\max}}\right] \\ &\quad\leq \exp\left[-\tilde{C}_1\left(\frac{\phi_0\sqrt{p_*}}{2}\right)|\mathcal{T}_{i,t}|\frac{h^2}{16x_{\max}^2} + \log 2d\right] + \exp\left[-p_*\tilde{C}_2(\phi_0)|\mathcal{T}_{i,t}|/2 + \log d\right] \\ &\quad\leq \exp\left[-\tilde{q}_0\log t \cdot \frac{p_*^2\tilde{C}_1(\phi_0)h^2}{128x_{\max}^2} + \log 2d\right] + \exp\left[-p_*\tilde{C}_2(\phi_0)\tilde{q}_0\log t + \log d\right] .\end{aligned}$$

Combining this with the probability that $|\mathcal{T}'_{i,t}|/|\mathcal{T}_{i,t}| \geq p_*/2$ (Lemma EC.10), and using a union bound gives

$$\begin{aligned}
&\Pr\left[\|\hat{\beta}\left(\mathcal{T}_{i,t}\right)-\beta_i\|_1 > \frac{h}{4x_{\max}}\right] \\
&\qquad \leq \exp\left[-\tilde{q}_0 \log t \cdot \frac{p_*^2 \tilde{C}_1(\phi_0) h^2}{128 x_{\max}^2} + \log 2d\right] + \exp\left[-p_* \tilde{C}_2(\phi_0)\tilde{q}_0 \log t + \log d\right] + 2/t^4 \,.
\end{aligned}$$

Now, using definition of $\tilde{q}_0$, in particular

$$\tilde{q}_0 \geq \frac{8\log d}{p_* \tilde{C}_2} \quad \text{and} \quad \tilde{q}_0 \geq \frac{1024 x_{\max}^2 \log 2d}{h^2 p_*^2 \tilde{C}_1}\,,$$

and the fact that $d > 2$ and $t > (Kq)^2$, the result follows. $\square$

We again define the event $A_t$ in the same way as (3) in order to prove the tail inequality for the all-sample OLS estimator.

PROPOSITION EC.3. *When $t \geq (Kq)^2$, for $i \in \mathcal{K}_{opt}$, the all-sample estimator $\hat{\beta}(\mathcal{S}_{i,t})$ satisfies the tail inequality*

$$\Pr\left[\|\hat{\beta}\left(\mathcal{S}_{i,t}\right)-\beta_i\|_1 \leq \chi\right] \geq 1 - \exp\left[-t\chi^2 \frac{p_*^3 \tilde{C}_1(\phi_0)}{256} + \log 2d\right] - 2\exp\left[-\tilde{C}_2(\phi_0)\frac{p_*^2}{64} t + \log d\right] \,.$$

*Proof of Proposition EC.3* First, we note that Lemma EC.12 holds for the OLS estimator as well since the forced-sample tail inequality for the OLS estimator (Proposition EC.2) is slightly stronger than the forced-sample tail inequality for the LASSO estimator (Proposition 2), $5/t^4$ versus $4/t^4$ error bound.

From Lemma EC.13, we have that at time $t \geq (Kq)^2$, each of $\{1, ..., t\} \setminus \cup_{j=1}^K \mathcal{T}_{j,t}$ belongs to $\mathcal{S}'_{i,t}$ with probability at least $p_*/2$. Applying Lemma EC.14 and Lemma EC.21 with $p = p_*/2$ and $|\mathcal{A}| \geq p_* t/4$, we get, using a union bound,

$$\begin{aligned}
&\Pr\left[\|\hat{\beta}\left(\mathcal{S}_{i,t}\right)-\beta_i\|_1 > \chi\right] \\
&\quad \leq \exp\left[-\tilde{C}_1\left(\frac{\phi_0\sqrt{p_*/2}}{2}\right)\frac{tp_*}{4}\chi^2 + \log 2d\right] + \exp\left[-\frac{\tilde{C}_2(\phi_0)p_* t}{16} + \log d\right] + \exp\left[-\frac{p_*^2}{128}t\right] \\
&\qquad\qquad \leq \exp\left[-t\chi^2 \frac{p_*^3 \tilde{C}_1(\phi_0)}{256} + \log 2d\right] + 2\exp\left[-\tilde{C}_2(\phi_0)\frac{p_*^2}{64}t + \log d\right] \,,
\end{aligned}$$

where we have used $\tilde{C}_2(\phi_0) \leq 1/2$ in the last step. $\square$

### EC.7.5. Bounding the Regret in the Low-Dimensional Setting

We can now use the above tail inequalities to sum up the expected regret contributions from the three groups of time periods:

(a) Initialization ($t \le (Kq)^2$) or forced sampling ($t \in \mathcal{T}_{i,T}$ for some $i \in [K]$).

(b) Times $t > (Kq)^2$ when the event $A_{t-1}$ does not hold.

(c) Times $t > (Kq)^2$ when the event $A_{t-1}$ holds and we do not perform forced sampling, i.e., the OLS Bandit plays the estimated best arm from $\hat{\mathcal{K}}$ using the all-sample estimator.

We first note that the regret bounds of §EC.5 for groups (a) times where $t \le (Kq)^2$ or we are force-sampling, and (b) time periods where $A_{t-1}$ does not hold can be re-used. This is because the forced-sampling schedule is the same and the tail inequality we prove for the OLS forced-sample estimator is strictly stronger than the tail inequality for the LASSO forced-sample estimator. We now focus on bounding the regret from time periods (c) when $t > (Kq)^2$, we are not force-sampling, and $A_{t-1}$ holds.

In this section, we simplify our notation by letting $\hat{\beta}_i = \hat{\beta}(\mathcal{S}_{i,t})$ for all $i \in [K]$. We also define the constant $\tilde{C}_3(\phi_0, p_*) = p_*^3 \tilde{C}_1(\phi_0)/256$, but to simplify the notation, drop the references to $\phi_0$ and $p_*$ in all constants $\tilde{C}_1, \tilde{C}_2, \ldots$ since the values for $\phi_0$ and $p_*$ will be fixed in the remaining.

Lemma EC.24. *If Algorithm 2 does not use the forced-sample estimator and $A_{t-1}$ holds, then the expected regret at time $t$ is bounded by*

$$\frac{8K\max(C_0,1)x_{\max}^2[\log(12d)]^{3/2}}{t\tilde{C}_3} + 8Kbx_{\max}e^{-\tilde{C}_2\frac{p_*^2}{64}t}.$$

*Proof of Lemma EC.24* Recall from Lemma EC.18 that since $A_{t-1}$ holds, the set $\hat{\mathcal{K}}$ contains the optimal arm $i^* = \arg\max_{i\in[K]} X_t^\top \beta_i$ and no sub-optimal arms from the set $\mathcal{K}_{sub}$. Without loss of generality, assume that arm 1 is optimal, i.e., $1 = \arg\max_{i\in\{1,\ldots,K\}} X_t^\top \beta_i$. Then, the expected regret at time $t$ is given by

$$\begin{aligned}\mathbb{E}[r_t] &= \mathbb{E}\left(\sum_{i\in\hat{\mathcal{K}}, i\neq 1} X_t^\top(\beta_1-\beta_i)\cdot \mathbb{I}\left[i = \arg\max_{j\in\{1,\ldots,K\}} X_t^\top\hat{\beta}_j\right]\right)\\ &\le \mathbb{E}\left(\sum_{i\in\hat{\mathcal{K}}, i\neq 1} X_t^\top(\beta_1-\beta_i)\,\mathbb{I}\left[X_t^\top\hat{\beta}_i > X_t^\top\hat{\beta}_1\right]\right)\end{aligned}$$

where the inequality follows from the fact that the event where $i = \arg\max_{j\in\{1,\ldots,K\}} X_t^\top\hat{\beta}_j$ is a subset of the event $X_t^\top\hat{\beta}_i > X_t^\top\hat{\beta}_1$, and that $\mathbb{E}[X_t^\top(\beta_1-\beta_i)] \ge 0$ (since we have assumed that arm 1 is optimal). Thus, we can bound $r_t$ through the regret incurred by each arm with respect to the optimal arm independently of the other arms. We now define, for each $r = 0,1,2,3,\ldots$ the event

$$B_r^i = \left\{2x_{\max}r\delta \le X_t^\top(\beta_1-\beta_i) < 2x_{\max}(r+1)\delta\right\}$$

where $\delta$ is a parameter we will choose later to minimize regret. Note that, since $X_t^\top(\beta_1-\beta_i) < 2x_{\max}b$, $B_r^i$ is empty for $r+1 > b/\delta$. Then, we can write

$$\mathbb{E}[r_t] < 2x_{\max}\delta\,\mathbb{E}\left(\sum_{r=0}^{\lfloor b/\delta\rfloor-1}(r+1)\sum_{i\in\hat{\mathcal{K}}, i\neq 1}\mathbb{I}\left[(X_t^\top\hat{\beta}_i > X_t^\top\hat{\beta}_1)\cap B_r^i\right]\right) \tag{EC.5}$$

by the definition of $B_r^i$.

Note that the event $(X_t^\top \hat{\beta}_i > X_t^\top \hat{\beta}_1) \cap B_r^i$ for $i \neq 1$ implies that

$$\begin{aligned}
0 &> X_t^\top \hat{\beta}_1 - X_t^\top \hat{\beta}_i \\
&= \left[X^\top \hat{\beta}_1 - X_t^\top \beta_1\right] + \left[X_t^\top \beta_i - X_t^\top \hat{\beta}_i\right] + \left[X_t^\top \beta_1 - X_t^\top \beta_i\right] \\
&\geq \left[X_t^\top \left(\hat{\beta}_1 - \beta_1\right)\right] + \left[X_t^\top \left(\beta_i - \hat{\beta}_i\right)\right] + 2x_{\max} r\delta \,.
\end{aligned}$$

Thus, it must be that either $X_t^\top \left(\hat{\beta}_1 - \beta_1\right) > x_{\max} r\delta$ or $X_t^\top \left(\beta_i - \hat{\beta}_i\right) > x_{\max} r\delta$ which means, using a union bound,

$$\begin{aligned}
\Pr\left[(X_t^\top \hat{\beta}_i > X_t^\top \hat{\beta}_1) \cap B_r^i\right] &\leq \Pr\left[(X_t^\top \left(\hat{\beta}_1 - \beta_1\right) > x_{\max} r\delta) \cap B_r^i\right] + \Pr\left[(X_t^\top \left(\beta_i - \hat{\beta}_i\right) > x_{\max} r\delta) \cap B_r^i\right] \\
&\leq \Pr\left[(\|\beta_1 - \hat{\beta}_1\|_1 > r\delta) \cap B_r^i\right] + \Pr\left[(\|\hat{\beta}_i - \beta_i\|_1 > r\delta) \cap B_r^i\right] \\
&= \Pr\left[\|\beta_1 - \hat{\beta}_1\|_1 > r\delta\right] \Pr\left[B_r^i\right] + \Pr\left[\|\hat{\beta}_i - \beta_i\|_1 > r\delta\right] \Pr\left[B_r^i\right] \,.
\end{aligned}$$

Note that the last equality uses the fact that event $B_r^i$ depends on the randomness of $X_t$ that is completely independent of past samples that impact the randomness of $\hat{\beta}_1$ and $\hat{\beta}_i$.

Next, recall that the tail inequality (Proposition EC.3) implies that for all $j \in \hat{\mathcal{K}}$, and all $r, \delta \geq 0$,

$$\Pr\left[\|\beta_j - \hat{\beta}_j\|_1 > r\delta\right] \leq \min\left\{1,\ \exp\left[-\tilde{C}_3 r^2 \delta^2 t + \log 2d\right] + 2\exp\left[-\tilde{C}_2 \frac{p_*^2}{64} t\right]\right\} \,.$$

Combining this with the fact that, via Assumption 2 on margin condition,

$$\Pr[B_r^i] \leq \Pr\left[X_t^\top (\beta_1 - \beta_i) \leq 2x_{\max}(r+1)\delta\right] \leq 2C_0 x_{\max}(r+1)\delta \,,$$

we get,

$$\begin{aligned}
\Pr\left[(X_t^\top \hat{\beta}_i > X_t^\top \hat{\beta}_1) \cap B_r^i\right] &\leq 2\Pr[B_r^i] \min\left\{1, e^{-\tilde{C}_3 r^2 \delta^2 t + \log 2d} + 2e^{-\tilde{C}_2 \frac{p_*^2}{64} t}\right\} \\
&\leq 2\Pr[B_r^i] \min\left\{1, e^{-\tilde{C}_3 r^2 \delta^2 t + \log 2d}\right\} + 4\Pr[B_r^i] e^{-\tilde{C}_2 \frac{p_*^2}{64} t} \\
&\leq 4C_0 x_{\max}(r+1)\delta \min\left\{1, e^{-\tilde{C}_3 r^2 \delta^2 t + \log 2d}\right\} + 4\Pr[B_r^i] e^{-\tilde{C}_2 \frac{p_*^2}{64} t} \\
&\leq 4C_0 x_{\max}(r+1)\delta \min\left\{1, e^{-\tilde{C}_3 r^2 \delta^2 t + \log 2d}\right\} + 4e^{-\tilde{C}_2 \frac{p_*^2}{64} t} \,. \qquad \text{(EC.6)}
\end{aligned}$$

Note that, for a large $r$ the term $e^{-\tilde{C}_3 t r^2 \delta^2 + \log 2d}$ will be small. Therefore, we will use the term 1 for small $r$ and the second term for large $r$. Combining (EC.5) and (EC.6), setting $\delta = 1/\sqrt{\tilde{C}_3 t}$, and defining

$$R \equiv R(d, t, \delta) = \left\lfloor \sqrt{\log(12d)} \right\rfloor \,,$$

we have

$$\mathbb{E}[r_t] \le \frac{8KC_0x_{\max}^2}{t\tilde{C}_3}\left[\sum_{r=0}^{R}(r+1)^2 + 2d\sum_{r=R+1}^{\lfloor b\sqrt{\tilde{C}_3t}\rfloor-1}(r+1)^2e^{-r^2}\right] + \left[\frac{8Kbx_{\max}}{\sqrt{\tilde{C}_3t}}\sum_{r=R+1}^{\lfloor b\sqrt{\tilde{C}_3t}\rfloor-1}(r+1)e^{-\tilde{C}_2\frac{p_*^2}{64}t}\right]$$

$$\le \frac{8KC_0x_{\max}^2[\log(12d)]^{3/2}}{t\tilde{C}_3} + \frac{16Kdx_{\max}^2}{t\tilde{C}_3}\sum_{r=R+1}^{\lfloor b\sqrt{\tilde{C}_3t}\rfloor-1}(r+1)^2e^{-r^2} + 8Kbx_{\max}e^{-\tilde{C}_2\frac{p_*^2}{64}t}.$$

Now, note that

$$\begin{aligned}\sum_{r=R+1}^{\infty}(r+1)^2e^{-r^2} &\le 4\sum_{r=R+1}^{\infty}r^2e^{-r^2}\\ &\le 4\int_R^\infty u^2e^{-u^2}du\\ &= 2Re^{-R^2} + 2\int_R^\infty e^{-u^2}du\end{aligned}$$

where the second inequality follows from Lemma EC.16 and the equality is via integration by parts. Therefore,

$$\begin{aligned}\sum_{r=R+1}^{\infty}(r+1)^2e^{-r^2} &\le 2Re^{-R^2} + 2\int_R^\infty\left(\frac{u}{R}\right)e^{-u^2}du\\ &= 2Re^{-R^2} + \frac{e^{-R^2}}{R}\\ &\le \frac{\sqrt{\log(12d)}}{3d}.\end{aligned}$$

Summarizing,

$$\begin{aligned}\mathbb{E}[r_t] &\le \frac{8KC_0x_{\max}^2[\log(12d)]^{3/2}}{t\tilde{C}_3} + \frac{6Kx_{\max}^2[\log(12d)]^{1/2}}{t\tilde{C}_3} + 8Kbx_{\max}e^{-\tilde{C}_2\frac{p_*^2}{64}t}\\ &\le \frac{8K\max(C_0,1)x_{\max}^2[\log(12d)]^{3/2}}{t\tilde{C}_3} + 8Kbx_{\max}e^{-\tilde{C}_2\frac{p_*^2}{64}t}. \quad \square\end{aligned}$$

LEMMA EC.25. *The cumulative expected regret from the time periods in group (c), times $t \in [T] \setminus [(Kq)^2]$ when the event $A_{t-1}$ holds and we do not perform forced sampling, is bounded by*

$$\frac{8K\max(C_0,1)x_{\max}^2[\log(12d)]^{3/2}\log T}{\tilde{C}_3} + \tilde{C}_4Kbx_{\max},$$

*where*

$$\tilde{C}_4 = \frac{8}{1-\exp\left[-\tilde{C}_2\frac{p_*^2}{64}\right]}.$$

*Proof of Lemma EC.25* Using Lemma EC.24,

$$\sum_{t=(Kq)^2+1}^{T}\mathbb{E}[r_t] \le \frac{8K\max(C_0,1)x_{\max}^2[\log(12d)]^{3/2}\log T}{\tilde{C}_3} + \frac{8Kbx_{\max}}{1-\exp\left[-\tilde{C}_2\frac{p_*^2}{64}\right]}. \quad \square$$

Summing up the regret contributions from the previous subsection gives us our main result.

*Proof of Theorem EC.1* The total expected cumulative regret of the OLS Bandit up to time $T$ is upper-bounded by summing all the terms from Lemmas EC.15, EC.17, and EC.25):

$$R_T \leq 2qKbx_{\max}(6\log T + Kq) + 2Kbx_{\max} + \frac{8K\max(C_0,1)x_{\max}^2[\log(12d)]^{3/2}\log T}{\tilde{C}_3} + \tilde{C}_4 Kbx_{\max}. \quad \square$$