A Note on LinUCB's Smallest Eigenvalue

Alejandro Gomez-Leos

June 2023

Abstract

In this short note, we document some progress on an open question concerning the minimum eigenvalue of the action covariance matrix tracked by the celebrated LinUCB algorithm (Auer 2002, Dani et al. 2008).

1 Introduction

We consider an instance $(\mathcal{A}, \theta)$ of the stochastic linear bandit problem, where $\mathcal{A} \subseteq \mathbb{R}^d$ is the ambient action set and $\theta \in \mathbb{R}^d$ is a fixed and unknown parameter. In each round $t \in [T]$ for an unknown $T \in \mathbb{N}$, a decision maker is presented with a set of options $\mathcal{A}_t$ and chooses an action $\mathbf{a}_t \in \mathcal{A}$, receiving a scalar reward $X_t := \langle \mathbf{a}_t, \theta \rangle + \eta_t$, where η_t is independent noise. We'll assume $\mathcal{A}_t \subseteq \{x \mid ||x||_2 = 1\}$ for all $t \in [T]$ and that η_t is bounded in $[-1, 1]$, which gives $|X_t| \leq ||\theta||_2 + 1$. The decision maker's goal is to minimize their regret $R(T) = \sum_{s=1}^{T} \mathbb{E}\left[\max_{\mathbf{a} \in \mathcal{A}} \langle \mathbf{a}, \theta \rangle - \langle \mathbf{a}_t, \theta \rangle\right]$.

A classic algorithm for this problem is as follows [2, 3]. Let $\mathbf{H}_t \in \mathbb{R}^{t \times d}$ denote the matrix who's i^{th} row is the action taken at round i. Setting $\mathbf{M}_t := (\lambda \mathbf{I} + \mathbf{H}_t^\top \mathbf{H}_t)$ for some $\lambda \geq 1$, we denote the regularized ridge-regression estimator as $\hat{\theta}_t = \mathbf{M}_t^{-1} \sum_{s=1}^{t} \mathbf{a}_s X_s$. Each round t, the decision maker is to play the arm $\arg\max_{\mathbf{a} \in \mathcal{A}} \texttt{UCB}_t(\mathbf{a})$ where

$$\texttt{UCB}_t(\mathbf{a}) = \langle \mathbf{a}, \hat{\theta}_{t-1} \rangle + \sqrt{\beta_t} ||\mathbf{a}||_{\mathbf{M}_{t-1}^{-1}}.$$

Here, β_t is an appropriately chosen sequence.

Motivated by the study of a problem in multi-agent bandits[1], we ask the following. *For which instances is the smallest eigenvalue of* $\mathbf{M}_t$ *order* $\Omega(\log(t))$*?*

Considering the classic UCB algorithm for scalar bandits [1], from the structure of the reward it may seem straightforward that each arm must be pulled $\Omega(\log(t))$ times. Naturally, this observation was made precise and shown in [4]. Towards generalizing this theory to stochastic linear bandits, in the rest of this note we analyze this eigenvalue by reducing the scalar bandit problem to our setting.

2 Analysis

Preliminaries Let $\mathbf{e_1} \dots \mathbf{e_d}$ denote the standard basis. We assume that there exists $z \in (0, 1]$ fulfilling

$$\min_{i,t} \max_{\mathbf{a} \in \mathcal{A}_t} (\langle \mathbf{a}, \mathbf{e}_i \rangle)^2 \geq z$$

i.e. for every $t \in [T]$ $\mathcal{A}_t$ admits an action with some positive component towards any of the basis directions. This condition is natural—otherwise an adversary can construct a sequence of contexts whose span degenerates. We'll assume each $\mathcal{A}_t$ is the collection of the standard basis directions, so that $z = 1$ anyway. We

[1] This condition implies $\hat{\theta}_t$ is a accurate ℓ_2 estimate of θ. In the study of cooperative bandits, this could give conditions for which a selfish agent is incentivized to exploit shared information rather than exploring. In economics, such a problem is called a free-rider problem.

define the quantities

$$E_i(\ell, k) := [\ \langle \mathbf{a}_\ell, \mathbf{e}_i \rangle\ \langle \mathbf{a}_{\ell+1}, \mathbf{e}_i \rangle\ \ldots\ \langle \mathbf{a}_k, \mathbf{e}_i \rangle\ ]^\top$$

In some sense, $||E_i(\ell, k)||_2^2$ is the accrued "energy" towards direction $\mathbf{e}_i$, collected by a sequence of actions played from time ℓ to k. We consider the eigendecomposition of $\mathbf{M}_t$ as $\sum_{i\in[d]} \mu_i(t)\, \mathbf{e}_i \mathbf{e}_i^\top$, where $\mu_i(t)$ $= \lambda + ||E_i(1,t)||_2^2$ denotes the eigenvalue of $\mathbf{e}_i$ at time t (after playing t^{th} action). Note $\mu_i(0) = \lambda$.

A simple, but useful observation is that $\mu_i(t)$ monotonically increases with time

$$t_1 \le t_2 \implies \mu_i(t_1) \le \mu_i(t_2) \quad \forall i \in [d] \tag{1}$$

Yet another simple, but useful observation is that

$$t_1 \le t_2 \implies \min_{j\in[d]} \mu_j(t_1) \le \min_{j\in[d]} \mu_i(t_2) \tag{2}$$

Lemma 1. *For all $i \in [d]$ and for all time t,*

$$\langle \hat{\theta}_{t-1}, \mathbf{e}_i \rangle = \frac{1}{\lambda + ||E_i(1, t-1)||_2^2} \sum_{s=1}^{t-1} \langle \mathbf{a}_s, \mathbf{e}_i \rangle X_s$$

Proof.

$$\langle \hat{\theta}_{t-1}, \mathbf{e}_i \rangle = \mathbf{e}_i^\top \mathbf{M}_{t-1}^{-1} \sum_{s=1}^{t-1} \mathbf{a}_s X_s = \frac{1}{\mu_i(t-1)} \mathbf{e}_i^\top \sum_{s=1}^{t-1} \mathbf{a}_s X_s = \frac{1}{\mu_i(t-1)} \sum_{s=1}^{t-1} \sum_{j\in[d]} \mathbf{e}_i^\top \mathbf{e}_j \langle \mathbf{a}_s, \mathbf{e}_j \rangle X_s$$

■

Corollary 1. *For all $i \in [d]$ and for all time t, with $L := \sqrt{d}(||\theta||_2 + 1)$*

$$|\langle \hat{\theta}_{t-1}, \mathbf{e}_i \rangle| \le \frac{L}{\big(\mu_i(t-1)\big)^{0.49}}$$

Proof. This follows by

$$\begin{aligned}
|\langle \hat{\theta}_{t-1}, \mathbf{e}_i \rangle| &\le \frac{1}{\lambda + ||E_i(1, t-1)||_2^2} \sum_{s=1}^{t-1} |\langle \mathbf{a}_s, \mathbf{e}_i \rangle| |X_s| \\
&\le \frac{||\theta||_2 + 1}{\lambda + ||E_i(1, t-1)||_2^2} \sum_{s=1}^{t-1} |\langle \mathbf{a}_s, \mathbf{e}_i \rangle| \\
&= \frac{||\theta||_2 + 1}{\lambda + ||E_i(1, t-1)||_2^2} \, ||E_i(1, t-1)||_1 \\
&\le \frac{\sqrt{d}(||\theta||_2 + 1)}{\lambda + ||E_i(1, t-1)||_2^2} \, ||E_i(1, t-1)||_2 \\
&= L \bigg(\frac{||E_i(1, t-1)||_2}{\lambda + ||E_i(1, t-1)||_2^2} \bigg) \\
&\le L \, \frac{(\mu_i(t-1))^{0.51}}{\mu_i(t-1)}
\end{aligned}$$

In the last step we used $\mu_i(t-1) = \lambda + ||E_i(1, t-1)||_2^2$, as well as the inequality $x \le (\lambda + x^2)^{0.51}$ for $\lambda \ge 1$. ■

Corollary 2. *We can obtain a simple bound on $\langle \hat{\theta}_{t-1}, \mathbf{a} \rangle$ for any action $\mathbf{a}$ as*

$$|\langle \hat{\theta}_{t-1}, \mathbf{a} \rangle| \le \frac{dL}{\lambda^{0.49}} = \mathcal{O}(1)$$

Proof.

$$|\langle\hat{\theta}_{t-1},\mathbf{a}\rangle| \leq \sum_{i=1}^{d}|\langle\mathbf{a},\mathbf{e}_i\rangle\langle\hat{\theta}_{t-1},\mathbf{e}_i\rangle| \overset{(a)}{\leq} \sum_{i=1}^{d}\frac{L}{\big(\mu_i(t-1)\big)^{0.49}} \overset{(b)}{\leq} \sum_{i=1}^{d}\frac{L}{\lambda^{0.49}} \tag{3}$$

where (a) follows by $||\mathbf{a}||_2 = 1$ and Corollary 1; (b) follows by $\mu_i(t-1) \geq \lambda \quad \forall i \in [d]$ ■

We'll also apply a result from [5].

Lemma 2. *For any action* $\mathbf{a}$*, recall that* $\texttt{UCB}_s(\mathbf{a}) = \langle\mathbf{a},\hat{\theta}_{s-1}\rangle + \sqrt{\beta_s}||\mathbf{a}||_{\mathbf{M}_{s-1}^{-1}}$*. For the choice of* $\sqrt{\beta_s} := \sqrt{\lambda}||\theta||_2 + \sqrt{2\log(s) + d\log(\frac{d\lambda+s}{d\lambda})}$ *([5],* $\delta = 1/s, L = 1, m_2 = ||\theta||_2$*), it holds that*

$$\texttt{UCB}_s(\mathbf{a}) - \langle\mathbf{a},\hat{\theta}_{s-1}\rangle = \sqrt{\beta_s}\sqrt{\sum_{i\in[d]}\frac{(\mathbf{a}_i)^2}{\mu_i(s-1)}}$$

Proof. We use the following:

$$\mathbf{M}_{t-1}\mathbf{e}_i = \mu_i(t-1)\,\mathbf{e}_i \implies \frac{\mathbf{e}_i}{\mu_i(t-1)} = \mathbf{M}_{t-1}^{-1}\,\mathbf{e}_i \implies \frac{\mathbf{e}_i^\top}{\mu_i(t-1)} = \mathbf{e}_i^\top\mathbf{M}_{t-1}$$

. . . to yield

$$||\mathbf{a}||_{\mathbf{M}_{s-1}^{-1}}^2 = ||\sum_{i\in[d]}\langle\mathbf{a},\mathbf{e_i}\rangle\mathbf{e}_i||_{\mathbf{M}_{s-1}^{-1}}^2 = \sum_{i\in[d]}(\langle\mathbf{a},\mathbf{e_i}\rangle)^2||\mathbf{e}_i||_{\mathbf{M}_{s-1}^{-1}}^2 = \sum_{i\in[d]}\frac{(\langle\mathbf{a},\mathbf{e_i}\rangle)^2}{\mu_i(s-1)}$$

■

The following lemma is proved via contradiction. The contrary of the lemma allows us to find a distant time t_d such that there is a direction $\mathbf{e}_j$ deficient in energy. It follows that any action with some energy towards $\mathbf{e}_j$ remains competitive in terms of UCB score at times close to t_d ("close" is necessary because other directions may be weakly-explored at some very far time in the past from t_d—and hence more competitive with $\mathbf{e}_j$).

Examining a sequence of times in the past $t_1 < t_2 < \cdots < t_d$, we can argue the following recursively:

1. At t_1, we can guarantee some direction $\mathbf{e}_{i_1}$ gets a lot of exploration due to pigeonholing.
2. From t_1 to t_2, $\mathbf{e}_{i_1}$ cannot receive too much exploration, since we have that any action in direction of $\mathbf{e}_j$ remains competitive.
3. Thus, a distinct direction from $\mathbf{e}_{i_2} \neq \mathbf{e}_{i_1}$ gets "some" exploration by pigeonholing (this time with "less pigeonholes" by exclusion of $\mathbf{e}_{i_1}$).
4. We repeat for each subsequent time interval.

If we make sure that "some" is sufficient (i.e. about log of t_d), then by the end we reach a contradiction because we find that all directions get sufficient exploration by time t_d, directly contradicting the existance of $\mathbf{e}_j$. This is the idea of the proof.

Theorem 1. *For any sequence of actions played by* `LinUCB`*,*

$$(\exists\, c > 0)(\exists\, t^*) : t \geq t^* \implies \mu_i(t) > cf(t) \quad \forall\, i \in [d]$$

where $f(t) := \log(t)^{(1-\alpha)}$ *for any constant* $\alpha > 0$*.*

Proof. Assume not; that

$$(\forall\, c > 0)(\forall\, t \in \mathbb{N})(\exists\, \bar{t} \geq t) : \mu_i(\bar{t}) \leq cf(\bar{t}) \quad \text{for some } i \in [d] \tag{4}$$

Define $\gamma := \sqrt{1 - \frac{z}{2}}$. In particular, take $c \in \left(0, (1-\gamma)^2 \frac{z}{16d^2}\right)$. Fix a time $\tilde{t}$ such that for all $t \geq \tilde{t}$, we have

$$t - d\lceil f(t) \rceil > 0 \tag{5}$$

By (4), there exists a time, say t_d, where $t_d \geq \tilde{t}$ fulfills

$$\exists\, j \in [d] : \mu_j(t_d) \leq cf(t_d) \tag{6}$$

Applying the property (5) to the time t_d, we define the sequence of times $t_k \in \mathbb{N}$ for $k = 1 \dots d-1$ as $t_k := t_d - (d-k)\lceil f(t_d) \rceil$. To prove the theorem, it suffices to show the assumption (4) implies the following claim (justification to follow):

Claim 1. *At each time t_k for $k = 1 \dots d$, there exists a direction $\mathbf{e}_{i_k}$ distinct from $\mathbf{e}_{i_1} \dots \mathbf{e}_{i_{k-1}}$ such that*

$$\mu_{i_k}(t_k) \geq \left(1 - \gamma\right)^2 \frac{f(t_d)}{d + 1 - k} \tag{7}$$

Demonstrating the claim proves the theorem, since it shows

$$\mu_{i_k}(t_d) \geq \left(1 - \gamma\right)^2 \frac{f(t_d)}{d} \geq \left(1 - \gamma\right)^2 \frac{z f(t_d)}{16d^2} > cf(t_d) \quad \forall\, k \in [d]$$

i.e. all the directions have sufficient energy. But this contradicts the fact that we chose t_d to satisfy (6), i.e. at least one direction is deficient in energy.

We proceed to prove the claim via strong induction. The base case $k = 1$ is as follows. Since $\sum_{i=1}^{d} ||E_i(1, t_1)||_2^2 = t_1$, by pigeonholing there exists a direction $\mathbf{e}_{i_1}$ such that $\mu_{i_1}(t_1) \geq \frac{t_1}{d} = \frac{t_d - (d-1)\lceil f(t_d) \rceil}{d} \geq \frac{\lceil f(t_d) \rceil}{d}$, where the last inequality follows by (5).

Assume the claim is true up to $k-1$. Importantly, note that the induction hypothesis implies existence of distinct $\mathbf{e}_{i_1} \dots \mathbf{e}_{i_{k-1}}$ at time t_{k-1} for which, by monotonicity of $\mu_{(\cdot)}(t)$ (cf. (1)) and $t_{k-1} \geq t_1 \dots t_{k-1}$, we have

$$\mu_{i_1}(t_{k-1}) \geq \mu_{i_1}(t_1) \geq \left(1 - \gamma\right)^2 \tfrac{f(t_d)}{d}$$

$$\mu_{i_2}(t_{k-1}) \geq \mu_{i_2}(t_2) \geq \left(1 - \gamma\right)^2 \tfrac{f(t_d)}{d-1}$$

$$\mu_{i_3}(t_{k-1}) \geq \mu_{i_3}(t_3) \geq \left(1 - \gamma\right)^2 \tfrac{f(t_d)}{d-2}$$

$$\dots$$

$$\mu_{i_{k-1}}(t_{k-1}) \geq \left(1 - \gamma\right)^2 \tfrac{f(t_d)}{d-(k-1)}$$

It will later become useful to observe (from the above inequalities) a simpler uniform bound

$$\mu_{i_1}(t_{k-1}) \dots \mu_{i_{k-1}}(t_{k-1}) \geq \left(1 - \gamma\right)^2 \frac{f(t_d)}{d} \tag{8}$$

To finish the induction step, it's enough to show that not "too many" actions have "too much" energy towards the aggregate of directions $\mathbf{e}_{i_1} \dots \mathbf{e}_{i_{k-1}}$ in the window $[t_{k-1} + 1, t_k]$. Specifically,

$$\leq \gamma\, \lceil f(t_d) \rceil \text{ actions } \mathbf{a}_\ell \text{ for } \ell \in [t_{k-1} + 1, t_k] \text{ may satisfy } \sum_{i \in \{i_1 \dots i_{k-1}\}} (\mathbf{a}_\ell)_i^2 \geq \gamma \tag{9}$$

This is enough because the following sequence of arguments ensues. The window contains exactly $t_k-(t_{k-1}+1)+1=\lceil f(t_d)\rceil$ time slots, which implies that over the same window, using the fact each action has unit L_2 norm,

$$> (1-\gamma)\lceil f(t_d)\rceil \text{ actions } \mathbf{a}_\ell \text{ for } \ell\in[t_{k-1}+1,t_k] \text{ satisfy } \sum_{i\notin\{i_1\dots i_{k-1}\}}(\mathbf{a}_\ell)_i^2>1-\gamma \tag{10}$$

Hence, over the window $[t_{k-1}+1,t_k]$, (10) implies the directions excluding $\mathbf{e}_{i_1}\dots\mathbf{e}_{i_{k-1}}$ accrue energy $\geq(1-\gamma)^2\lceil f(t_d)\rceil$. Another application of pigeonholing shows there must be some direction $\mathbf{e}_{i_k}\notin\{\mathbf{e}_{i_1}\dots\mathbf{e}_{i_{k-1}}\}$ which accrues energy $\geq(1-\gamma)^2\frac{\lceil f(t_d)\rceil}{d+1-k}$, thus finishing the induction step—all of which is a consequence of (9).

We verify that (9) holds, assuming the contrary. Specifically, assume

$$> \gamma\lceil f(t_d)\rceil \text{ actions } \mathbf{a}_\ell \text{ for } \ell\in[t_{k-1}+1,t_k] \text{ may satisfy } \sum_{i\in\{i_1\dots i_{k-1}\}}(\mathbf{a}_\ell)_i^2\geq\gamma \tag{11}$$

This implies $>\gamma^2\lceil f(t_d)\rceil$ energy is accrued towards directions $\mathbf{e}_{i_1}\dots\mathbf{e}_{i_{k-1}}$ in this interval. By pigeonholing, since the interval is $\lceil f(t_d)\rceil$ wide, this implies that at least one action $\tilde{\mathbf{b}}:=\mathbf{a}_\tau$ is played at some time $\tau\in[t_{k-1}+1,t_k]$ satisfying

$$\sum_{i\in\{i_1\dots i_{k-1}\}}(\tilde{\mathbf{b}})_i^2\geq\gamma^2=1-\frac{z}{2} \tag{12}$$

In turn, this implies

$$\sum_{i\notin\{i_1\dots i_{k-1}\}}(\tilde{\mathbf{b}})_i^2\leq\frac{z}{2} \tag{13}$$

Prior to the choice of $\tilde{\mathbf{b}}$ by `LinUCB` for the round τ, each $\texttt{UCB}_\tau(\cdot)$ score for all available actions in $\mathcal{A}_\tau$ was computed (which implicitly involves computation of $\mu_i(\tau-1)\quad\forall i\in[d]$). If we let $j^*:=\arg\min\mu_j(\tau-1)$, then by the definition of z we know there exists an action $\mathbf{b}\in\mathcal{A}_\tau$ such that $(\mathbf{b})_{j^*}^2\geq z$. Furthermore, we have $\mathbf{b}\neq\tilde{\mathbf{b}}$, since otherwise $\mathbf{b}=\tilde{\mathbf{b}}$ implies $(\tilde{\mathbf{b}})_{j^*}\geq z$, hence $j^*\in\{i_1\dots i_{k-1}\}$ by (12) and (13). If $j^*\in\{i_1\dots i_{k-1}\}$, we have that (8) and the monotonicity of $\mu_{(\cdot)}(t)$ (cf. (2)) gives

$$\min_j\mu_j(t_d)\geq\mu_{j^*}(\tau-1)=\min_{j\in\{i_1\dots i_{k-1}\}}\mu_j(\tau-1)\geq\min_{j\in\{i_1\dots i_{k-1}\}}\mu_j(t_{k-1})\geq\big(1-\gamma\big)^2\frac{f(t_d)}{d}$$

But $\min_j\mu_j(t_d)\leq cf(t_d)$, so it must be that $\mathbf{b}\neq\tilde{\mathbf{b}}$.

We proceed by showing $\texttt{UCB}_\tau(\mathbf{b})-\texttt{UCB}_\tau(\tilde{\mathbf{b}})>0$, showing that $\mathbf{b}$ should have been played instead at τ, proving (9) holds. We bound the `UCB` score of $\tilde{\mathbf{b}}$ as

$$\begin{aligned}
\texttt{UCB}_\tau(\tilde{\mathbf{b}}) &\overset{(a)}{\leq} \mathcal{O}(1)+\sqrt{\beta_\tau}\bigg(\sum_{i\in\{i_1\dots i_{k-1}\}}\frac{(\tilde{\mathbf{b}})_i^2}{\mu_i(\tau-1)}+\sum_{i\notin\{i_1\dots i_{k-1}\}}\frac{(\tilde{\mathbf{b}})_i^2}{\mu_i(\tau-1)}\bigg)^{1/2}\\
&\overset{(b)}{\leq} \mathcal{O}(1)+\sqrt{\beta_\tau}\bigg(\sum_{i\in\{i_1\dots i_{k-1}\}}\frac{1}{\mu_i(\tau-1)}+\frac{1}{\min_j\mu_j(\tau-1)}\sum_{i\notin\{i_1\dots i_{k-1}\}}(\tilde{\mathbf{b}})_i^2\bigg)^{1/2}\\
&\overset{(c)}{\leq} \mathcal{O}(1)+\sqrt{\beta_\tau}\bigg(\sum_{i\in\{i_1\dots i_{k-1}\}}\frac{1}{\mu_i(\tau-1)}+\frac{(z/2)}{\min_j\mu_j(\tau-1)}\bigg)^{1/2}\\
&\overset{(d)}{\leq} \mathcal{O}(1)+\sqrt{\beta_\tau}\bigg(\frac{d^2}{(1-\gamma)^2f(t_d)}+\frac{(z/2)}{\min_j\mu_j(\tau-1)}\bigg)^{1/2}
\end{aligned}$$

where (a) follows by Lemma 2 and Corollary 2; (b) follows by unit L_2 norm assumption on actions; (c) follows by (13); and (d) follows by an application of $\mu_i(\tau-1)\geq\mu_i(t_{k-1})$ as well as the bound (8) applied d times.

On the other hand, Lemma 2 yields that $\mathbf{b}$ fulfills

$$\texttt{UCB}_\tau(\mathbf{b}) \geq -\mathcal{O}(1) + \sqrt{\beta_\tau}\left(\frac{z}{\min_j \mu_j(\tau-1)}\right)^{1/2}$$

Combining the above bounds, we have

$$\begin{aligned}\texttt{UCB}_\tau(\mathbf{b}) - \texttt{UCB}_\tau(\tilde{\mathbf{b}}) \geq -\mathcal{O}(1) \\ + \sqrt{\beta_\tau}\left(\left(\frac{z}{\min_j \mu_j(\tau-1)}\right)^{1/2} - \left(\frac{d^2}{(1-\gamma)^2 f(t_d)} + \frac{(z/2)}{\min_j \mu_j(\tau-1)}\right)^{1/2}\right) \quad (14)\end{aligned}$$

We proceed to lower bound the right hand side term as

$$\begin{aligned}&\sqrt{\beta_\tau}\left(\left(\frac{z}{\min_j \mu_j(\tau-1)}\right)^{1/2} - \left(\frac{d^2}{(1-\gamma)^2 f(t_d)} + \frac{(z/2)}{\min_j \mu_j(\tau-1)}\right)^{1/2}\right) \\ &\overset{(a)}{\geq} \sqrt{\beta_\tau}\left(\left(1-\frac{1}{\sqrt{2}}\right)\left(\frac{z}{\min_j \mu_j(\tau-1)}\right)^{1/2} - \left(\frac{d^2}{(1-\gamma)^2 f(t_d)}\right)^{1/2}\right) \\ &\overset{(b)}{\geq} \sqrt{\beta_\tau}\left(\left(1-\frac{1}{\sqrt{2}}\right)\left(\frac{z}{cf(t_d)}\right)^{1/2} - \left(\frac{d^2}{(1-\gamma)^2 f(t_d)}\right)^{1/2}\right) \\ &\overset{(c)}{\geq} \sqrt{\beta_\tau}\left(\left(1-\frac{1}{\sqrt{2}}\right)\left(\frac{16d^2}{(1-\gamma)^2 f(t_d)}\right)^{1/2} - \left(\frac{d^2}{(1-\gamma)^2 f(t_d)}\right)^{1/2}\right) \\ &= \sqrt{\beta_\tau}\left(3-\frac{4}{\sqrt{2}}\right)\left(\frac{d^2}{(1-\gamma)^2 f(t_d)}\right)^{1/2} \\ &\overset{(d)}{\geq} \Omega\big(\sqrt{\log(\tau)}\big)\left(3-\frac{4}{\sqrt{2}}\right)\left(\frac{d^2}{(1-\gamma)^2 f(t_d)}\right)^{1/2} \\ &\overset{(e)}{\geq} \Omega\big(\sqrt{\log(t_d)}\big)\left(3-\frac{4}{\sqrt{2}}\right)\left(\frac{d^2}{(1-\gamma)^2 f(t_d)}\right)^{1/2} \\ &\overset{(f)}{\geq} \Omega\big(\log(t_d)^{\alpha/2}\big)\left(3-\frac{4}{\sqrt{2}}\right)\left(\frac{d^2}{(1-\gamma)^2}\right)^{1/2}\end{aligned}$$

where (a) is an application of $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$; (b) uses the fact $\min_j \mu_j(\tau-1) \leq \min_j \mu_j(t_d) \leq cf(t_d)$; (c) follows for our choice of c; in (d) we use the fact that

$$\sqrt{\beta_\tau} := \sqrt{\lambda}||\theta||_2 + \sqrt{2\log(\tau) + d\log(\frac{d\lambda+\tau}{d\lambda})} \geq \sqrt{\lambda}||\theta||_2 + \sqrt{(2+d)\log(\tau) - d\log(d\lambda)}$$

by an application of the inequality $\log(x+1) > \log(x) + \frac{1}{x+1}$; (e) follows since $\tau \geq t_1 = t_d - (d-1)\lceil \log(t_d)^{1-\alpha} \rceil$; (f) follows by simply substituting $f(t_d) := \log(t_d)^{1-\alpha}$.

Finally, combining the above with (14) yields $\texttt{UCB}_\tau(\mathbf{b}) - \texttt{UCB}_\tau(\tilde{\mathbf{b}}) \geq -\mathcal{O}(1) + \Omega(\log(t_d)^{\alpha/2})$. Hence, for sufficiently large $\tilde{t}$, we have that $\texttt{UCB}(\mathbf{b}) - \texttt{UCB}(\tilde{\mathbf{b}}) > 0$, yielding the contradiction which implies (9) holds, which was sufficient to complete the induction, proving the claim which we showed was sufficient for the theorem.

■

Corollary 3. *In particular, for $c \in \big(0, (1-\gamma)^2 \frac{z}{16d^2}\big)$, since the theorem holds for any $\alpha > 0$, it must be that*

$$(\exists\, t_c) : t \geq t_c \implies \min_j \mu_j(t) \geq c\log(t) \quad (15)$$

3 Discussion

In the previous section we recovered the result by [4] (which also had $c = \Omega(1/d^2)$). It is possible that these techniques may be used to extend this argument to more general action sets—the main difficulty is tracking the eigenvalues as the eigenvectors of $\mathbf{M}_t$ can vary throughout.

References

[1] P Auer. Finite-time analysis of the multiarmed bandit problem, 2002.

[2] Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

[3] Varsha Dani, Thomas P Hayes, and Sham M Kakade. Stochastic linear optimization under bandit feedback. In *COLT*, volume 2, page 3, 2008.

[4] Christopher Jung, Sampath Kannan, and Neil Lutz. Quantifying the burden of exploration and the unfairness of free riding. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1892–1904. SIAM, 2020.

[5] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.