# Appendix

## Table of Contents



## A. Proof of Theorem 1

We state the full version of our concentration result.

**Theorem 5.** *Let $\delta \le e^{-1}$. Let $\hat{\theta}_t$ be the solution of Eq.* (1) *in the main text where, for every $s \in [t]$, $y_s$ is conditionally independent from $x_1, \ldots, x_{s-1}, x_{s+1}, \ldots, x_t$ given $x_s$ (i.e., fixed design). Let $t_{\mathsf{eff}}$ be the number of distinct vectors in $\{x_s\}_{s=1}^t$. Fix $x \in \mathbb{R}^d$ such that $\|x\| \le 1$. Define $\gamma(d) = 64(d\log(6)+\log((2+t_{\mathsf{eff}})/\delta))$. If $\xi_t^2 := \max_{s\in[t]} \|x_s\|^2_{H_t(\theta^*)^{-1}} \le \frac{1}{\gamma(d)}$,*

$$\mathbb{P}\Big(|x^\top(\hat{\theta}_t - \theta^*)| \le 2.4 \cdot \|x\|_{H_t(\theta^*)^{-1}} \sqrt{\log(2(2+t_{\mathsf{eff}})/\delta)},$$
$$\forall x' \in \mathbb{R}^d, \frac{1}{\sqrt{2.2}} \|x'\|_{(H_t(\theta^*))^{-1}} \le \|x'\|_{(H_t(\hat{\theta}_t))^{-1}} \le \sqrt{2.2}\|x'\|_{(H_t(\theta^*))^{-1}}\Big) \ge 1-\delta\ ,$$

*which implies the following empirical variance bound:*

$$\mathbb{P}\left(|x^\top(\hat{\theta}_t - \theta^*)| \le 3.6 \cdot \|x\|_{H_t(\hat{\theta}_t)^{-1}} \sqrt{\log(2(2+t_{\mathsf{eff}})/\delta)}\right) \ge 1-\delta\ .$$

To improve the concentration inequality from Li et al. (2017), we follow their analysis closely but exploit the variance term whenever possible.

We define the following:

- Let $H := H_t(\theta^*) = \sum_{s=1}^t \dot{\mu}(x_s^\top \theta^*) x_s x_s^\top$.

- Let $y_s \in \{0,1\}$ be the binary reward when arm $x_s$ is pulled at time $s$. Let $\eta_s := y_s - \mu(x_s^\top \theta^*)$ and $\sigma_s^2 := \dot\mu(x_s^\top \theta^*)$.
- Define $z_t := \sum_{s=1}^t \eta_s x_s$.
- Let $\alpha(x,\theta_1,\theta_2) = \frac{\mu(x^\top\theta_1)-\mu(x^\top\theta_2)}{x^\top(\theta_1-\theta_2)}$. We use the shorthand $\alpha_s(\hat\theta_t,\theta^*) := \alpha(x_s,\hat\theta_t,\theta^*)$.
- Let $G := \sum_{s=1}^t \alpha_s(\hat\theta_t,\theta^*) x_s x_s^\top$. Note that by the optimality condition,

$$z_t = \sum_s (\mu(x_s^\top\hat\theta_t) - \mu(x_s^\top\theta^*))x_s = \sum_s \alpha_s(\hat\theta_t,\theta^*)x_s x_s^\top(\hat\theta_t - \theta^*) = G(\hat\theta_t-\theta^*)\ . \tag{10}$$

- Define $g_t := \sum_{s=1}^t \mu(x_s^\top\theta)x_s$. The following identity is well-known (e.g., (Filippi et al., 2010, Proposition 1)):

$$\|\hat\theta_t - \theta^*\|_G = \|g_t(\hat\theta_t) - g_t(\theta^*)\|_{G^{-1}} \tag{11}$$

- Let $E := G - H$.

First, we assume the following event:

$$\mathcal{E}_0 := \left\{\forall s\in[t], \left|\frac{\alpha_s(\hat\theta_t,\theta^*) - \dot\mu(x_s^\top\theta^*)}{\dot\mu(x_s^\top\theta^*)}\right| \le Q \text{ for some } Q>0\right\}\ , \tag{12}$$

which we will show is true later under suitable stochastic events.

**The main decomposition:** We use the following decomposition based on Eq. (10) and tackle those two terms separately.

$$\begin{aligned}|x^\top(\hat\theta_t-\theta^*)| &= |x^\top G^{-1}z_t| = |x^\top(H+E)^{-1}z_t| = |x^\top H^{-1}z_t - x^\top H^{-1}E(H+E)^{-1}z_t| \\ &\le |x^\top H^{-1}z_t| + |x^\top H^{-1}E(H+E)^{-1}z_t|\ .\end{aligned}$$

We bound the two terms separately.

**Term 1:** $|x^\top H^{-1}z_t| = |\sum_s \langle x, H^{-1}x_s\rangle \eta_s|$

Note that $H^{-1}$ is deterministic (unlike $G^{-1}$) conditioning on $\{x_1,\ldots,x_t\}$, so we can apply the standard argument for the concentration inequality. With the following Bernstein's inequality in mind, we assume the event $\mathcal{E}_1(x)$ defined below.

**Lemma 1.** *Let $\delta \le e^{-1}$ and define*

$$\mathcal{E}_1(x) := \left\{|x^\top H^{-1}z_t| \le \sqrt{2}\|x\|_{H^{-1}}\sqrt{\log(2/\delta)} + \frac{2}{3}\|x\|_{H^{-1}}\xi_t\log(2/\delta)\right\}\ .$$

*Then,* $\mathbb{P}\left(\mathcal{E}_1(x)\right) \ge 1-\delta$.

*Proof.* The proof can be found in Section A.2. □

**Term 2:** $|x^\top H^{-1}E(H+E)^{-1}z_t|$

We have

$$\begin{aligned}|x^\top H^{-1}E(H+E)^{-1}z_t| = |x^\top H^{-1}EG^{-1}z_t| &\le \|x\|_{H^{-1}}\|H^{-1/2}EH^{-1/2}\|\|G^{-1}z_t\|_H \\ &= \|x\|_{H^{-1}}\|H^{-1/2}EH^{-1/2}\|\|\hat\theta_t-\theta^*\|_H\end{aligned}$$

Let us study the term $\|H^{-1/2}EH^{-1/2}\|$. For a symmetric matrix $A$, the singular values are the absolute values of the eigenvalues. Thus, we have

$$\|A\| = \max\left\{\max_{x:\|x\|\le 1} x^\top Ax, \max_{x:\|x\|\le 1} x^\top(-A)x\right\}\ .$$

With this, we need to study both $x^\top H^{-1/2}EH^{-1/2}x$ and $x^\top H^{-1/2}(-E)H^{-1/2}x$. Under the event $\mathcal{E}_0$,

$$\begin{aligned}&\max\{x^\top H^{-1/2}EH^{-1/2}x,\ \ x^\top H^{-1/2}(-E)H^{-1/2}x\} \\ &\le x^\top H^{-1/2}\left(\sum_s |\alpha_s(\hat\theta_t,\theta^*) - \dot\mu(x_s^\top\theta^*)|x_s x_s^\top\right)H^{-1/2}x\end{aligned}$$

$$\leq x^\top H^{-1/2} \left( \sum_s Q\dot{\mu}(x_s^\top \theta^*) x_s x_s^\top \right) H^{-1/2} x$$

$$= Q\|x\|^2 \leq Q \qquad (\because \|x\| \leq 1)$$

$$\implies \|H^{-1/2} E H^{-1/2}\| \leq Q \,.$$

For $\|\hat{\theta}_t - \theta^*\|_H$, we first use the lemma below to bound it by $(1+D)\|z_t\|_{H^{-1}}$. The key is to use the self-concordance control lemma (Faury et al., 2020, Lemma 9), we can relate $G$ and $H$ as a function of $D$. If $A$ and $B$ are matrices, then we use $A \succeq B$ to mean that $A - B$ is positive semi-definite.

**Lemma 2.** *Let $D = \max_{s\in[t]} |x_s^\top(\hat{\theta}_t - \theta^*)|$.*

$$G \succeq \frac{1}{1+D} \cdot H$$

*where $A \succeq B$ means that $A - B$ is positive semi-definite.*

*Proof.* We first note that, by the self-concordance control lemma (Faury et al., 2020, Lemma 9),

$$\alpha_s(\theta_1, \theta_2) \geq \frac{\dot{\mu}(x_s^\top \theta_2)}{1 + |x_s^\top(\theta_1 - \theta_2)|} \,.$$

Then,

$$G = \sum_s^t \alpha(x_s, \theta_1, \theta_2) x_s x_s^\top \succeq \frac{1}{1 + \max_{s\in[t]} |x_s^\top(\theta_1 - \theta_2)|} \sum_s^t \dot{\mu}(x_s^\top \theta_2) x_s x_s^\top \,.$$

Notice that the sum on the RHS is $H_t(\theta_2)$. □

We then bound $\|z_t\|_{H^{-1}}$ by the following concentration result via the covering argument.

**Lemma 3.** *Recall $\xi_t = \max_{s\in[t]} \|x_s\|_{H^{-1}}$. Let $\delta \leq e^{-1}$. Define $\mathcal{E}_2$ and $\beta_t$ as:*

$$\mathcal{E}_2 := \left\{ \left\| \sum_s^t \eta_s x_s \right\|_{H^{-1}} \leq 2\sqrt{2} \cdot \sqrt{d\log(6) + \log(1/\delta)} + \frac{4}{3}\xi_t(d\log(6) + \log(1/\delta)) =: \sqrt{\beta_t} \right\} \,.$$

*Then, $\mathbb{P}(\mathcal{E}_2) \geq 1 - \delta$.*

*Proof.* See Section A.2. □

Let $D := \max_{s\leq t} |x_s^\top(\hat{\theta}_t - \theta^*)|$. Therefore,

$$|x^\top H^{-1} E(H+E)^{-1} z_t| \leq \|x\|_{H^{-1}} (1+D) Q \sqrt{\beta_t} \qquad \text{(Lemma 2)}$$

To summarize, under $\mathcal{E}_0 \cap \mathcal{E}_1(x) \cap \mathcal{E}_2$, we have

$$\begin{aligned} |x^\top(\hat{\theta}_t - \theta^*)| &\leq \sqrt{2}\|x\|_{H^{-1}} \sqrt{\log(2/\delta)} + \frac{2}{3}\|x\|_{H^{-1}} \xi_t \log(2/\delta) \\ &\quad + \|x\|_{H^{-1}}(1+D)Q \left( 2\sqrt{2}\sqrt{d\log(6) + \log(1/\delta)} + \frac{4}{3}\xi_t(d\log(6) + \log(1/\delta)) \right) \,. \end{aligned} \tag{13}$$

We now aim to control $(1+D)Q$ to be small so that the entire RHS is $O(\|x\|_{H^{-1}}\sqrt{\log(1/\delta)})$.

### A.1. Controlling (1+D)Q

Assume $\mathcal{E}_2$. We first assume that for every $s \in [t]$, $\mathcal{E}_1(x_s)$ is true and then take the maximum over $s$ on the inequality implied by $\mathcal{E}_1(x_s)$ to obtain

$$D = \max_s |x_s^\top(\hat{\theta}_t - \theta^*)| \;\leq \sqrt{2}\xi_t \sqrt{\log(2/\delta)} + \frac{2}{3}\xi_t^2 \log(2/\delta) + \xi_t(1+D)Q\sqrt{\beta_t} \tag{14}$$

where $\sqrt{\beta_t}$ is defined in Lemma 3.

To control $(1+D)Q$, one can show that the self concordance control lemma (Faury et al., 2020, Lemma 9) implies the following, which we use to motivate our choice of $Q$ and satisfy $\mathcal{E}_0$:

$$\left|\frac{\alpha_s(\hat{\theta}_t, \theta^*) - \dot{\mu}(x_s^\top \theta^*)}{\dot{\mu}(x_s^\top \theta^*)}\right| \le \frac{e^D - 1 - D}{D} =: Q \;.$$

Then, one can show that

$$D \le \frac{3}{5} \implies (1+D)Q = (1+D)\frac{e^D - 1 - D}{D} \le D \;. \tag{15}$$

So, if we can ensure $D \le \frac{3}{5}$, then we have $(1+D)Q \le D$, which can be applied to Eq. (14) to arrive at

$$D \le \frac{\sqrt{2}\xi_t\sqrt{\log(2/\delta)} + \frac{2}{3}\xi_t^2\log(2/\delta)}{1 - \xi_t\sqrt{\beta_t}} \;, \tag{16}$$

assuming that $1 - \xi_t\sqrt{\beta_t} > 0$. Thus, it remains to

(A) find a sufficient condition for $D \le \frac{3}{5}$ and $1 - \xi_t\sqrt{\beta_t} > 0$,
(B) bound the RHS of Eq. (16) to obtain the bound on $D$, and
(C) use the bound on Eq. (13) to get the final bound.

For $(A)$, we prove the following lemma.

**Lemma 4.** *Under $\mathcal{E}_2$, we have*

$$\xi_t \le \frac{0.12}{\sqrt{d\log(6) + \log(1/\delta)}} \implies \xi_t\sqrt{\beta_t} \le \frac{3}{8} \implies D \le \frac{3}{5} \;. \tag{17}$$

*Proof.* Using Lemma 2,

$$\begin{aligned}
D^2 &\le \xi_t^2\|\hat{\theta}_t - \theta^*\|_H^2 \le \xi_t^2(1+D)\|\hat{\theta}_t - \theta^*\|_G^2 \\
&= \xi_t^2(1+D)\cdot\|g_t(\hat{\theta}_t) - g_t(\theta^*)\|_{G^{-1}}^2 \\
&\le \xi_t^2\cdot(1+D)^2\|g_t(\hat{\theta}_t) - g_t(\theta^*)\|_{H^{-1}}^2 \\
&\le (1+D)^2\xi_t^2\beta_t \qquad \text{(by Eq. (10) and } \mathcal{E}_2\text{)} \\
\implies D &\le (1+D)\xi_t\sqrt{\beta_t} \\
\implies D &\le \frac{\sqrt{\beta_t}\xi_t}{1 - \sqrt{\beta_t}\xi_t} \;.
\end{aligned}$$

where the last line requires an assumption that $1 - \sqrt{\beta_t}\xi_t > 0$. For this, we require that $\sqrt{\beta_t}\xi_t \le \frac{3}{8}$. Then,

$$D \le \frac{8}{5}\sqrt{\beta_t}\xi_t \;.$$

In order to control the RHS above by $3/5$, we need to satisfy $\xi_t \le \frac{5}{8\sqrt{\beta_t}} \cdot \frac{3}{5}$. However, $\beta_t$ depends on $\xi_t$, so we need to solve for $\xi_t$. Let $C := d\log(6) + \log(1/\delta)$. Then, we need to solve

$$\xi_t \le \frac{3}{8(2\sqrt{2}\cdot\sqrt{C} + \frac{4}{3}\xi_t C)} \;,$$

which is quadratic in $\xi_t$. Solving for $\xi_t$, we have $\xi \le \frac{1}{\sqrt{C}} \cdot \left(\frac{6}{16\sqrt{2} + \sqrt{16^2\cdot 2 + 12\cdot\frac{32}{3}}}\right) = \frac{0.125\ldots}{\sqrt{C}}$. Thus, it suffices to require $\xi_t \le \frac{1}{8\sqrt{C}}$. □

Hereafter, we assume that $\xi_t \le \frac{1}{8\sqrt{d\log(6) + \log(1/\delta)}}$.

For $(B)$, by Lemma 4, we can deduce from Eq. (16) that

$$D \le \frac{8}{5}\cdot\left(\sqrt{2}\xi_t\sqrt{\log(2/\delta)} + \frac{2}{3}\xi_t^2\log(2/\delta)\right) \;.$$

For $(C)$, we now turn back to the initial concentration inequality (13). We first bound the last term of Eq. (13):

$$\begin{aligned}
\|x\|_{H^{-1}}(1+D)Q\sqrt{\beta_t} &\le \|x\|_{H^{-1}} D\sqrt{\beta_t} && (\because \text{Lemma 4 \& Eq. (15)})\\
&\le \|x\|_{H^{-1}} \frac{8}{5}\cdot\left(\sqrt{2}\xi_t\sqrt{\log(2/\delta)} + \frac{2}{3}\xi_t^2\log(2/\delta)\right)\cdot\sqrt{\beta_t}\\
&\le \|x\|_{H^{-1}} \frac{8}{5}\cdot\left(\sqrt{2}\cdot\frac{3}{8}\sqrt{\log(2/\delta)} + \frac{1}{4}\xi_t\log(2/\delta)\right) && (\because \text{Lemma 4})\\
&\le \|x\|_{H^{-1}} \frac{8}{5}\cdot\left(\sqrt{2}\cdot\frac{3}{8}\sqrt{\log(2/\delta)} + \frac{1}{4}\cdot\frac{1}{8\sqrt{d\log(6)+\log(1/\delta)}}\sqrt{\log(2/\delta)}\sqrt{\log(2/\delta)}\right)\\
&\le \|x\|_{H^{-1}} \frac{8}{5}\cdot\left(\sqrt{2}\cdot\frac{3}{8}\sqrt{\log(2/\delta)} + \frac{1}{32}\sqrt{\log(2/\delta)}\right)\\
&\le 0.9\cdot\|x\|_{H^{-1}}\sqrt{\log(2/\delta)}\;.
\end{aligned}$$

Similarly, one can show that $\frac{2}{3}\|x\|_{H^{-1}}\xi_t\log(2/\delta) \le \frac{1}{12}\|x\|_{H^{-1}}\sqrt{\log(2/\delta)}$. Altogether, under our condition on $\xi_t$, $\mathcal{E}_2$, $\mathcal{E}_1(x)$, and $\cap_{s=1}^t \mathcal{E}_1(x_s)$, Eq. (13) implies that

$$|x^\top(\hat\theta_t - \theta^*)| \le 2.4\cdot\|x\|_{H^{-1}}\sqrt{\log(2/\delta)}\;.$$

Note that $\mathbb{P}(\mathcal{E}_2, \mathcal{E}_1(x), \cap_{s\in[t]}\mathcal{E}_1(x_s)) \ge 1-(t_{\mathsf{eff}}+2)\delta$. To obtain our theorem statement, we replace $\delta$ with $\delta/(t_{\mathsf{eff}}+2)$. Here, we have $t_{\mathsf{eff}}$ instead of $t$ because $\mathcal{E}(x)$ and $\mathcal{E}(x')$ are identical events when $x = x'$.

Furthermore, the following lemma shows that the empirical variance is within a constant factor of the true variance. One can easily check that the condition in the lemma is satisfied under the events we have assumed and the condition on $\xi_t$, which implies the theorem statements.

**Lemma 5.** *Suppose $D = \max_{s\in[t]}|x_s^\top(\hat\theta_t-\theta^*)| \le 1$. Then, for all $x$,*

$$\frac{1}{\sqrt{2D+1}}\|x\|_{(H_t(\theta^*))^{-1}} \le \|x\|_{(H_t(\hat\theta_t))^{-1}} \le \sqrt{2D+1}\|x\|_{(H_t(\theta^*))^{-1}}\;.$$

*Proof.* See Section A.2. □

### A.2. Proof of Auxiliary Results

**Proof of Lemma 1.** It suffices to bound $\mathbb{P}(\mathcal{E}_1(x)\mid x_1,\ldots,x_t) \ge 1-\delta$ because this implies that $\mathbb{P}(\mathcal{E}_1(x)) = \int_{x_1,\ldots,x_t}\mathbb{P}(\mathcal{E}_1(x)\mid x_1,\ldots,x_t)\,\mathrm{d}F(x_1,\ldots,x_t) \ge 1-\delta$. For brevity, we omit the conditioning on $x_1,\ldots,x_t$ from probability statements for the rest of the proof.

Define $\tau_s := \langle x, H^{-1}x_s\rangle\eta_s$. We have that $\mathbb{E}\,\tau_s = 0$, $\sigma_s^2 = \mathbb{E}\,\tau_s^2 = x^\top H^{-1}x_s x_s^\top H^{-1}x$, and $|\tau_s| \le |\langle x, H^{-1}x_s\rangle|$ for all $s\in[t]$. Let $F = \max_s f_s$ and $S = \sum_{s=1}^t \sigma_s^2$. Using the standard Bernstein inequality, we have

$$\forall\epsilon>0, \mathbb{P}(\sum_{s=1}^t \tau_s \le \epsilon) \le \exp\left(-\frac{\epsilon^2}{2S+\frac{2}{3}F\epsilon}\right)$$

This leads to, with probability at least $1-\delta$,

$$\sum_{s\in[t]}\tau_s \le \frac{2}{3}F\log(1/\delta) + \sqrt{2S\log(1\delta)}$$

By union bound, w.p. at least $1-2\delta$,

$$|\sum_{s\in[t]}\tau_s| \le \frac{2}{3}F\log(1/\delta) + \sqrt{2S\log(1\delta)}$$

Noting that $S = \sum_s \sigma_s^2 = \|x\|_{H^{-1}}^2$ and $F \le \|x\|_{H^{-1}}\max_s\|x_s\|_{H^{-1}}$ concludes the proof.

**Proof of Lemma 3.** As done in Proof of Lemma 1, we condition on $x_1,\ldots,x_t$. The proof closely follows Li et al. (2017) but we employ the Bernstein inequality. Let $\mathcal{B}(1)$ be the Euclidean ball of radius 1 and $\hat{\mathcal{B}}(1)$ be a $1/2$-cover of $\mathcal{B}(1)$. It is well-known that one can find a cover $\hat{\mathcal{B}}(1)$ of cardinality $6^d$; see Pollard (1990, Lemma 4.1). In this proof, we use the shortcut $H := H_t(\theta^*)$.

Note that $\|z_t\|_{H^{-1}} = \|H^{-1/2} z_t\|_2 = \sup_{a \in \mathcal{B}(1)} \langle a, H^{-1/2} z_t \rangle$. Fix $x \in \mathbb{R}^d$. Let $\hat{x}$ be the closes point to $x$ in the cover $\hat{\mathcal{B}}(1)$. Then,

$$\begin{aligned}\langle x, H^{-1/2} z_t \rangle &= \langle \hat{x}, H^{-1/2} z_t \rangle + \langle x - \hat{x}, H^{-1/2} z_t \rangle \\ &= \langle \hat{x}, H^{-1/2} z_t \rangle + \|x - \hat{x}\| \langle \frac{x - \hat{x}}{\|x - \hat{x}\|}, H^{-1/2} z_t \rangle \\ &\le \langle \hat{x}, H^{-1/2} z_t \rangle + \frac{1}{2} \cdot \sup_{a \in \mathcal{B}(1)} \langle a, H^{-1/2} z_t \rangle \\ &= \langle \hat{x}, H^{-1/2} z_t \rangle + \frac{1}{2} \cdot \|z_t\|_{H^{-1}} \,.\end{aligned}$$

Taking sup over $x \in \mathcal{B}(1)$ on both sides, we have

$$\|z_t\|_{H^{-1}} \le \langle \hat{x}, H^{-1/2} z_t \rangle + \frac{1}{2} \cdot \|z_t\|_{H^{-1}} \implies \|z_t\|_{H^{-1}} \le 2 \langle \hat{x}, H^{-1/2} z_t \rangle \,.$$

This implies that, for $q > 0$,

$$\mathbb{P}(\|z_t\|_{H^{-1}} > q) \le \mathbb{P}(2 \sup_{\hat{x} \in \hat{\mathcal{B}}(1)} \langle \hat{x}, H^{-1/2} z_t \rangle > q) \le \sum_{\hat{x} \in \hat{\mathcal{B}}(1)} \mathbb{P}(\langle \hat{x}, H^{-1/2} z_t \rangle > q/2) \,. \tag{18}$$

It remains to bound $\mathbb{P}(\langle \hat{x}, H^{-1/2} z_t \rangle > q/2)$ for any $\hat{x}$ and then apply the union bound. Using the standard Bernstein inequality (see the proof of Lemma 1), we have, w.p. at least $1 - \delta$,

$$\sum_s \langle \hat{x}, H^{-1/2} x_s \rangle \eta_s \le \sqrt{2 \|\hat{x}\|^2 \log(1/\delta)} + \frac{2}{3} \xi_t \log(1/\delta)$$

Let us set $q$ such that the RHS above is equal to $q/2$. Replacing $\delta$ above with $\frac{\delta}{6^d}$ and taking a union bound over all possible $\hat{x} \in \mathcal{B}(1)$ concludes the proof.

**Proof of Lemma 5.** Let $\alpha(z_1, z_2) = \frac{\mu(z_1) - \mu(z_2)}{z_1 - z_2}$. Let $s \in [t]$. Because $\alpha(z_1, z_2) = \alpha(z_2, z_1)$, Faury et al. (2020, Lemma 9) imply

$$\dot{\mu}(x_s^\top \theta^*) \frac{1 - \exp(-D)}{D} \le \alpha(x_s^\top \theta^*, x_s^\top \hat{\theta}_t) \le \dot{\mu}(x_s^\top \theta^*) \cdot \frac{\exp(D) - 1}{D}$$

and

$$\dot{\mu}(x_s^\top \hat{\theta}_t) \frac{1 - \exp(-D)}{D} \le \alpha(x_s^\top \theta^*, x_s^\top \hat{\theta}_t) \le \dot{\mu}(x_s^\top \hat{\theta}_t) \cdot \frac{\exp(D) - 1}{D} \,.$$

Then,

$$\dot{\mu}(x_s^\top \hat{\theta}_t) \ge \frac{D}{\exp(D) - 1} \cdot \alpha(x_s^\top \theta^*, x_s^\top \hat{\theta}_t) \ge \frac{D}{\exp(D) - 1} \frac{1 - \exp(-D)}{D} \cdot \dot{\mu}(x_s^\top \theta^*) \overset{(a)}{\ge} \frac{1}{2D + 1} \dot{\mu}(x_s^\top \theta^*)$$

where $(a)$ is due to the following fact: using $z \le 1 \implies e^z \le z^2 + z + 1$, we have $\frac{D}{\exp(D)-1} \frac{1-\exp(-D)}{D} = \frac{e^D - 1}{e^D(e^D - 1)} = \frac{1}{e^D} \ge \frac{1}{D^2 + D + 1} \ge \frac{1}{2D+1}$. This implies that $H_t(\hat{\theta}_t) \succeq \frac{1}{2D+1} H_t(\theta^*)$. This concludes the proof of the second inequality. One can prove the other inequality similarly.

## B. $\kappa^{-1}$-free conditioning for Theorem 5

In this section, we consider a case where the burn-in condition (i.e., the requirement on $\xi_t$) in our Theorem 5 can be satisfied without spending $\kappa^{-1} = \min_{x \colon \|x\| \le 1} \dot{\mu}(x^\top \theta^*) = \Theta(\exp(S))$ samples where $S = \|\theta^*\|$. More specifically, we show that it is possible to use a sample size that is *polynomial* rather than *exponential* in $S^*$ to satisfy our burn-in condition. The implication of this is that our improved burn-in condition $\xi_t^2 = \max_{s \in [t]} \|x_s\|^2_{H_t(\theta^*)^{-1}} \le O(\frac{1}{d + \log(t/\delta)})$ is fundamentally different from that of Li et al. (2017, Theorem 1) which requires $\frac{1}{\lambda_{\min}(V)} \le O(\frac{1}{\kappa^{-4}(d^2 + \log(1/\delta))})$. Indeed their condition can only be satisfied after $\Omega(\exp(S))$ burn-in samples at all times. The construction is based on the Gaussian measurements that are common in practice and often considered in the compressed sensing literature (Plan & Vershynin, 2012).

**Gaussian Assumption:** We consider $t$ arms sampled from the following Gaussian distribution: $x_s \sim \mathcal{N}(0, \frac{1}{d} I), 1 \le s \le t$.

We define $r := S^2/d$ and further assume $d = \Omega(S^2)$ so that $r \leq 1$.

**Note:** though this will violate the assumption that $\|x_s\| \leq 1$, needed for the theorem, one can show that for large enough $d$, the norm of $x_s$ concentrates around 1. Using this, one can find a constant $c \leq 1$ so that with high probability a sample $x \sim \mathcal{N}(0, \frac{c}{d}I)$ satisfies $\|x\| \leq 1$, and then apply our argument below.

Continuing, under our Gaussian assumption, we have $x_s^\top \theta^* \sim \mathcal{N}(0, S^2/d)$. This implies that, w.p. at least $1-\delta$, we have $\forall s \in [t], |x_s^\top \theta^*| \leq \sqrt{\frac{2S^2}{d}\log(2t/\delta)} =: W$. Let $V = \sum_{s=1}^t x_s x_s^\top$. Then with high probability,

$$H(\theta^*) \succeq \dot{\mu}(-W)V = \dot{\mu}(W)V$$

Furthermore utilizing Li et al. (2017, Proposition 1) and our Gaussian assumption on the samples $\{x_s\}_{s=1}^t$ implies that, given $B > 0$, there exists an absolute constant $C_1$ such that, w.p. at least $1-\delta$,

$$t \geq C_1 \cdot d^2(d + \log(1/\delta)) + 2dB \implies \lambda_{\min}(V) \geq B \tag{19}$$

Thus, under the condition on $t$ above, on a high probability event we have that

$$\xi_t^2 \leq \max_{x:\|x\|\leq 1} \|x\|^2_{H(\theta^*)^{-1}} \leq \frac{1}{\dot{\mu}(W)} \max_{x:\|x\|\leq 1} \|x\|^2_{V^{-1}} \leq \frac{1}{\dot{\mu}(W)} \frac{1}{\lambda_{\min}(V)} \leq \frac{1}{\dot{\mu}(W)B}$$

It remains to control the RHS above to be no larger than $\frac{1}{d+\log(6(2+t)/\delta)}$, which means that we will satisfy the burnin condition of Theorem 5. Since $6(2+t) \leq 18t$, it suffices to show that

$$\dot{\mu}\left(\sqrt{\frac{S^2}{d}\log(t/\delta)}\right) B \geq d + \log(18t/\delta)$$

Using the fact that $\dot{\mu}(z) \geq \frac{1}{4}e^{-z}$, it thus suffices to show that

$$\frac{1}{4}\exp\left(-\sqrt{2r\log(2t/\delta)}\right) \cdot B \geq d + \log(18t/\delta)$$

We will make the simple choice of

$$\begin{aligned} B &:= \frac{1}{4}\exp\left(\sqrt{2r\log(2t/\delta)}\right)\left(d + \log(18t/\delta)\right) \\ &\leq \frac{1}{4}\exp\left(1 + \log((2t/\delta)^{1/2})\right)\left(d + \log(18t/\delta)\right) && (r \leq 1 \text{ and AM-GM ineq.}) \\ &= \frac{e}{4}(\frac{2t}{\delta})^{1/2}\left(d + \log(18t/\delta)\right) \end{aligned}$$

With this choice of $B$ it suffices to compute the lower bound on the right hand side of Eq. (19). With algebra, one can show that there exists

$$t_0 = \tilde{O}(d^2\log(1/\delta) + \frac{d^4}{\delta})$$

where $\tilde{O}$ hides polylogarithmic factors. such that $t \geq t_0$ implies the condition of Eq. (19).

To summarize, we just showed that, there exists an absolute constant $C$ such that, w.p. at least $1-2\delta$,

$$t \geq C \cdot (d^2\log(1/\delta) + \frac{d^4}{\delta}) \implies \xi_t \leq \frac{1}{\gamma(d)}$$

when $r = S^2/d \leq 1$. Simply setting $r = 1$, we have $d = S^2$, so our the statement above implies that the sample size needs to be only *polynomial* in $S$ for our choice of measurements. This is in stark contrast to the result of Li et al. (2017) that requires the sample size to be *exponential* in $S$ for *any* set of measurements.

## C. Proofs for GLM-Rage

### Burn-In Results

**Lemma 6.** *For $\delta \leq 1/16$ and $d \geq 4$, with probability greater than $1-\delta$, for all $\lambda \in \Delta_{\mathcal{X}}$, we have $\frac{1}{3}H(\lambda, \theta^*) \leq H(\lambda, \hat{\theta}_0) \leq 3H(\lambda, \theta^*)$.*

*Proof.* Firstly note that,

$$H(\lambda_0, \theta^*) \geq \sum_{x \in \mathcal{X}} \lambda_{0,x} \kappa_0 x x^\top \geq \kappa_0 A(\lambda_0)$$

So for any $x \in \mathcal{X}$,

$$\|x\|^2_{H(\lambda_0,\theta^*)^{-1}} \leq \kappa_0^{-1} \|x\|^2_{A(\lambda_0)^{-1}}$$

Define $H_0(\theta^*) = \sum_{s=1}^{n_0} \dot{\mu}(x_s^\top \theta^*) x_s x_s^\top$. Thus at the end of the burn-in phase,

$$\begin{aligned}
\max_{x \in \mathcal{X}} \|x\|^2_{H_0^{-1}(\theta^*)} &\leq \frac{(1+\epsilon)}{n_0} \max_{x \in \mathcal{X}} \|x\|^2_{H^{-1}(\lambda_0,\theta^*)} && \text{(Lemma 13 rounding)} \\
&\leq \frac{3(1+\epsilon)}{n_0} \kappa_0^{-1} \max_{x \in \mathcal{X}} \|x\|^2_{A^{-1}(\lambda_0)} \\
&\leq \frac{3(1+\epsilon)\kappa_0^{-1} d}{n_0} && \text{(Kiefer-Wolfowitz)} \\
&\leq \frac{1}{\gamma(d) \log(2|\mathcal{X}|(2+|\mathcal{X}|)/\delta)}
\end{aligned}$$

where we have employed the Kiefer-Wolfowitz theorem (Lattimore & Szepesvári, 2020, Theorem 21.1), which states that $\min_{\lambda \in \Delta_\mathcal{X}} \max_{x \in \mathcal{X}} \|x\|^2_{A(\lambda)^{-1}} = d$. In particular this implies using Theorem 5,

$$\begin{aligned}
|x^\top(\theta^* - \hat{\theta}_0)| &\leq 2.4 \sqrt{\|x\|^2_{(H_0(\theta^*))^{-1}} \log(2|\mathcal{X}|(2+|\mathcal{X}|)/\delta)} \\
&\leq 2.4 \sqrt{\frac{\log(2|\mathcal{X}|(2+|\mathcal{X}|)/\delta)}{\gamma(d) \log(2|\mathcal{X}|(2+|\mathcal{X}|)/\delta)}} \\
&\leq 1
\end{aligned}$$

With this, we apply Lemma 14 to conclude the proof.

□

Define the events

$$\mathcal{R}_k = \{\frac{1}{3} H(\lambda, \theta^*) \leq H(\lambda, \hat{\theta}_k) \leq 3H(\lambda, \theta^*), \forall \lambda \in \Delta_\mathcal{X}\}, k \geq 0$$

and

$$\mathcal{E}_{2,k} = \{\forall z \in \mathcal{Z}_k, |\langle z^* - z, \hat{\theta}_k - \theta^* \rangle| \leq 2^{-k}\}, k \geq 1.$$

In addition, define $\mathcal{E}_1 = \cap_{k=0}^\infty \mathcal{R}_k$ and $\mathcal{E}_2 = \cap_{k=1}^\infty \mathcal{E}_{2,k}$.

**Lemma 7** (Closeness of $\theta_t$)**.** *We have that* $\mathbb{P}(\mathcal{R}_k | \mathcal{R}_{k-1}, \cdots, \mathcal{R}_0) \geq 1 - 2\delta$*, i.e. for all* $k \geq 1$, $\frac{1}{3} H(\lambda_k, \theta^*) \leq H(\lambda_k, \widehat{\theta}_{k-1}) \leq 3H(\lambda_k, \theta^*)$

*Proof.* We proceed by induction. The base case of $t = 0$, is handled by Lemma 6 above. Assume that the event $\mathcal{R}_{k-1}$ holds. On this event, for $k > 1$, we first verify that $\max_{x \in \mathcal{X}} \|x\|^2_{H_t(\theta^*)^{-1}} \leq 1/\gamma(d)$

$$\begin{aligned}
\max_{x \in \mathcal{X}} \|x\|^2_{H_k^{-1}(\theta^*)} &\leq \frac{(1+\epsilon)}{n_k} \max_{x \in \mathcal{X}} \|x\|^2_{H^{-1}(\lambda_k,\theta^*)} && \text{(Lemma 13 rounding)} \\
&\leq \frac{3(1+\epsilon)}{n_k} \max_{x \in \mathcal{X}} \|x\|^2_{H^{-1}(\lambda_k,\hat{\theta}_{k-1})} && \text{(On event } \mathcal{R}_{k-1}) \\
&= \frac{1}{\gamma(d) \log(2|\mathcal{X}|k^2(2+|\mathcal{X}|)/\delta)}
\end{aligned}$$

Thus with probability greater than $1 - \delta/(k^2|\mathcal{X}|)$ conditioned on $\mathcal{R}_{k-1}$

$$|x^\top(\theta^* - \hat{\theta}_k)| \leq 2.4 \sqrt{\|x\|^2_{H_k(\theta^*)^{-1}} \log(2|\mathcal{X}|k^2(2+|\mathcal{X}|)/\delta)} \qquad \text{(By Theorem 5)}$$

$$\leq 2.4\sqrt{\frac{(1+\epsilon)\|x\|^2_{H(\lambda_k,\theta^*)^{-1}}\log(2|\mathcal{X}|k^2(2+|\mathcal{X}|)/\delta)}{n_k}} \quad \text{(By Lemma 13)}$$

$$\leq 2.4\sqrt{\frac{3(1+\epsilon)\|x\|^2_{H(\lambda_k,\hat{\theta}_{k-1})^{-1}}\log(2|\mathcal{X}|k^2(2+|\mathcal{X}|)/\delta)}{n_k}} \quad \text{(By the induction hypothesis } \mathcal{R}_{k-1})$$

$$\leq 2.4\sqrt{\frac{3(1+\epsilon)\log(2|\mathcal{X}|k^2(2+|\mathcal{X}|)/\delta)}{3(1+\epsilon)\gamma(d)\log(2|\mathcal{X}|k^2(2+|\mathcal{X}|)/\delta)}}$$

$$\leq 1$$

Then, union bounding over $\mathcal{X}$ gives that conditioned on $\mathcal{R}_{k-1}$, we have the event $\cup_{x\in\mathcal{X}}\{|x^\top(\hat{\theta}_{k-1}-\theta^*)|\leq 1\}$ is true with probability greater than $1-\delta/k^2$. In particular, now applying Lemma 14 proves the claim.

□

**Lemma 8** (Concentration)**.** *In round $k$, if we take $n_k$ samples as specified in the algorithm, then*

$$|(z_*-z)^\top(\hat{\theta}_k-\theta^*)|\leq 2^{-k}$$

*for all $z\in\mathcal{Z}_k$ with probability greater than $1-\frac{\delta}{k^2}$ given $\{\mathcal{R}_s,\mathcal{E}_{2,s}\}_{s=1}^{k-1}\cap\mathcal{R}_0$, or in other words $\mathbb{P}(\mathcal{E}_{2,k}|\{\mathcal{R}_s,\mathcal{E}_{2,s}\}_{s=1}^{k-1}\cap\mathcal{R}_0)\geq 1-\frac{\delta}{k^2}$.*

*Proof.* In the previous lemma we showed that conditioned on $\{\mathcal{R}_s,\mathcal{E}_{2,s}\}_{s=1}^{k-1}$, $\max_{x\in\mathcal{X}}\|x\|^2_{H_k(\theta^*)^{-1}}\leq 1/\gamma(d)$ . Given $\mathcal{Z}_k$ (a random set), we can apply Theorem 5, to calculate for any $z\in\mathcal{Z}_k$

$$(z^*-z)^\top(\hat{\theta}_k-\theta^*)\leq 2.4\sqrt{\|z^*-z\|^2_{H_k(\theta^*)^{-1}}\log(2|\mathcal{Z}|k^2(2+|\mathcal{X}|)/\delta)} \quad \text{(Theorem 5)}$$

$$\leq 2.4\sqrt{\frac{(1+\epsilon)\|z^*-z\|^2_{H(\lambda_k,\theta^*)^{-1}}\log(2|\mathcal{Z}|k^2(2+|\mathcal{X}|)/\delta)}{n_k}} \quad \text{(Lemma 13)}$$

$$\leq 2.4\sqrt{\frac{3(1+\epsilon)\|z^*-z\|^2_{H(\lambda_k,\hat{\theta}_{k-1})^{-1}}\log(2|\mathcal{Z}|k^2(2+|\mathcal{X}|)/\delta)}{n_k}} \quad \text{(Lemma 7)}$$

$$\leq 2.4\sqrt{\frac{3(1+\epsilon)\log(2|\mathcal{Z}|k^2(2+|\mathcal{X}|)/\delta)}{2^{2k}\cdot 2.4^2\cdot 3(1+\epsilon)\log(2|\mathcal{Z}|k^2(2+|\mathcal{X}|)/\delta)}}$$

$$\leq 2^{-k}$$

Now

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_{2,k}|\{\mathcal{R}_s,\mathcal{E}_{2,s}\}_{s=1}^{k-1}\cap\mathcal{R}_0) &\leq \sum_{\mathcal{V}\subset\mathcal{Z}}\mathbb{P}(\mathcal{E}_{2,k},\mathcal{Z}_k=\mathcal{V}|\{\mathcal{R}_s,\mathcal{E}_{2,s}\}_{s=1}^{k-1}\cap\mathcal{R}_0)\\
&\leq \sum_{\mathcal{V}\subset\mathcal{Z}}\mathbb{P}(\mathcal{E}_{2,t}|\mathcal{Z}_k=\mathcal{V},\{\mathcal{R}_s,\mathcal{E}_{2,s}\}_{s=1}^{k-1}\cap\mathcal{R}_0)\mathbb{P}(\mathcal{Z}_k=\mathcal{V}|\{\mathcal{R}_s,\mathcal{E}_s\}_{s=1}^{k-1}\cap\mathcal{R}_0)\\
&\leq \frac{\delta}{k^2}\sum_{\mathcal{V}\subset\mathcal{Z}}\mathbb{P}(\mathcal{Z}_k=\mathcal{V}|\{\mathcal{R}_s,\mathcal{E}_{2,s}\}_{s=1}^{k-1}\cap\mathcal{R}_0)\\
&\leq \frac{\delta}{k^2}
\end{aligned}$$

Finally we record a consequence of the previous computation for later use, namely,

$$(z^*-z)^\top(\hat{\theta}_k-\theta^*)\leq 2.4\|z^*-z\|_{H_t(\hat{\theta}_{k-1})^{-1}}\sqrt{3\log(2|\mathcal{Z}|k^2(2+|\mathcal{X}|)/\delta)}\leq 2^{-k} \tag{20}$$

□

**Lemma 9** (Correctness.)**.** *On $\mathcal{E}_1(x)$ and $\mathcal{E}_2$, we have that $z^*\in\mathcal{Z}_k$, and $\max_{z\in\mathcal{Z}_{k+1}}\langle z^*-z,\theta^*\rangle\leq 2\cdot 2^{-k}$ for all $t$. Furthermore, we have $\mathbb{P}(\mathcal{E}_1,\mathcal{E}_2)\geq 1-\delta$.*

*Proof.* Firstly note for any set of events $\{A_k\}_{k=1}^{\infty}$,

$$\mathbb{P}(\cup_{k=1}^{\infty} A_k) = \mathbb{P}(\cup_{k=1}^{\infty} \left(A_k \setminus (\cup_{j<k} A_j)\right)) \leq \sum_{k=1}^{\infty} \mathbb{P}(A_k \setminus (\cup_{j<k} A_j)) \leq \sum_{k=1}^{\infty} \mathbb{P}(A_k \mid (\cap_{j<k} \overline{A_j}))\,.$$

Then, with $A_k = \overline{\mathcal{R}_k} \cup \overline{\mathcal{E}_{2,k}}$, we have, using Lemma 6, Lemma 8, and Lemma 9,

$$\begin{aligned}
\mathbb{P}(\overline{\mathcal{E}_2} \cup \overline{\mathcal{E}_1}) &\leq \mathbb{P}(\cup_{k=1}^{\infty} (\overline{\mathcal{R}_k} \cup \overline{\mathcal{E}_{2,k}}) \cup \overline{\mathcal{R}_0}) \\
&\leq \sum_{k=1}^{\infty} \mathbb{P}(\overline{\mathcal{R}_k} \cup \overline{\mathcal{E}_{2,k}} \mid \mathcal{R}_{k-1}, \mathcal{E}_{2,k-1}, \ldots, \mathcal{R}_1, \mathcal{E}_{2,1}) + \mathbb{P}(\overline{\mathcal{R}_0}) \\
&\leq \sum_{k=1}^{\infty} \mathbb{P}(\overline{\mathcal{E}_{2,k}} \mid \mathcal{R}_{k-1}, \mathcal{E}_{2,k-1}, \ldots, \mathcal{R}_1, \mathcal{E}_{2,1}) + \sum_{k=1}^{\infty} \mathbb{P}(\overline{\mathcal{R}_k} \mid \mathcal{R}_{k-1}, \mathcal{E}_{2,k-1}, \ldots, \mathcal{R}_1, \mathcal{E}_{2,1}) + \delta \\
&\leq \sum_{k=1}^{\infty} \frac{2\delta}{k^2} + \delta \\
&\leq 3\delta\,.
\end{aligned}$$

For the following we will assume that event $\mathcal{E}_1 \cap \mathcal{E}_2$ holds. Now we argue that $z^*$ will never be eliminated. Indeed for any $z \in \mathcal{Z}_k$, note that

$$\begin{aligned}
\langle z - z^*, \hat{\theta}_k \rangle &= \langle z - z^*, \hat{\theta}_k - \theta^* \rangle + \langle z - z^*, \theta^* \rangle \\
&\leq 2.4 \|z^* - z\|_{H_k(\hat{\theta}_{k-1})^{-1}} \sqrt{3 \log(2|\mathcal{Z}|k^2(2 + |\mathcal{X}|)/\delta)} + \langle z - z^*, \theta^* \rangle \\
&\leq 2.4 \|z^* - z\|_{H_k(\hat{\theta}_{k-1})^{-1}} \sqrt{3 \log(2|\mathcal{Z}|k^2(2 + |\mathcal{X}|)/\delta)}\,,
\end{aligned}$$

implying that $z^*$ is not kicked out. Finally, if $\langle z^* - z, \theta^* \rangle \geq 2 \times 2^{-k}$, then

$$\begin{aligned}
\langle z^* - z, \hat{\theta}_k \rangle &= \langle z^* - z, \hat{\theta}_k - \theta^* + \theta^* \rangle \\
&= \langle z^* - z, \theta^* \rangle - \|z^* - z\|_{H(\lambda_k, \hat{\theta}_{k-1})^{-1}} \sqrt{3 \log(2|\mathcal{Z}|k^2(2 + |\mathcal{X}|)/\delta)} \qquad \text{(From(20))} \\
&\geq 2 \times 2^{-k} - 2^{-k} \\
&\geq 2^{-k} \\
&\geq 2.4 \|z^* - z\|_{H_k(\hat{\theta}_{k-1})^{-1}} \sqrt{3 \log(2|\mathcal{Z}|k^2(2 + |\mathcal{X}|)/\delta)}
\end{aligned}$$

which is precisely the condition for $z$ to be removed. Finally, we have $\langle z^* - z, \theta^* \rangle = \langle z^* - z, \theta^* - \hat{\theta}_k \rangle + \langle z^* - z, \hat{\theta}_k \rangle \leq 2^{-k} + 2^{-k}$, which concludes the proof. □

**Theorem 6** (Sample Complexity)**.** *Define $\mathcal{S}_k = \{z \in \mathcal{Z} : (z^* - z)^\top \theta_* \leq 2 \cdot 2^{-(k-1)}\}$, and take $\epsilon \leq 1/2$. up constant factors, Algorithm 1 returns $z^*$ with probability greater than $1 - 3\delta$ in a number of samples no more than*

$$O\Bigg((1+\epsilon) \sum_{k=1}^{\lceil \log_2(2/\Delta_{\min}) \rceil} \min_{\lambda \in \Delta_{\mathcal{X}}} \max \left[2^{2k} \max_{z, z' \in \mathcal{S}_k} \|z - z'\|^2_{H(\lambda, \theta^*)^{-1}}, \gamma(d) \max_x \|x\|^2_{H(\lambda, \theta^*)^{-1}}\right] \log(\max(|\mathcal{X}|, |\mathcal{Z}|^2) k^2/\delta)$$
$$+ d(1+\epsilon)\kappa_0^{-1} \log(|\mathcal{X}|/\delta) + r(\epsilon) \log_2(\frac{1}{\Delta_{\min}})\Bigg)\,.$$

*Proof.* For the remainder of the proof we will assume that $\mathcal{E}_1 \cap \mathcal{E}_2$ holds.

By Lemma 9 on $\mathcal{E}_2$, we have that $\mathcal{Z}_k \subseteq \mathcal{S}_k$, in particular this implies that when $2 \times 2^{-k} \leq \Delta_{\min}$, we have $|\mathcal{Z}_k| = 1$, so this implies that the algorithm will terminate in a number of rounds not exceeding $\lceil \log_2(2/\Delta_{\min}) \rceil$

By Lemma 7 on $\mathcal{E}_1$, we have that $H(\lambda_k, \hat{\theta}_k) \geq \frac{1}{4} H(\lambda_k, \theta^*)$. Thus, in each round,

$$\min_{\lambda\in\Delta_{\mathcal{X}}}\max\left[2^{2k}2.4^2\max_{z,z'\in\mathcal{Z}_t}\|z-z'\|^2_{H(\lambda,\hat{\theta}_{k-1})^{-1}},\gamma(d)\max_{x\in\mathcal{X}}\|x\|^2_{H(\lambda,\hat{\theta}_{k-1})^{-1}}\right]$$
$$\leq O\left(\min_{\lambda\in\Delta_{\mathcal{X}}}\max\left[2^{2k}\max_{z,z'\in\mathcal{S}_k}\|z-z'\|^2_{H(\lambda,\theta^*)^{-1}},\gamma(d)\max_{x\in\mathcal{X}}\|x\|^2_{H(\lambda,\theta^*)^{-1}}\right]\right)$$

Let $c$ be an absolute constant. Thus up to doubly logarithmic factors our final sample complexity is given by

$$\begin{aligned}
&n_0+\sum_{k=1}^{\lceil\log_2(2/\Delta_{\min})\rceil}n_k\\
&\leq\frac{d(1+\epsilon)\gamma(d)\log(|\mathcal{X}|/\delta)}{\kappa_0}\\
&\quad+c_1(1+\epsilon)\sum_{k=1}^{\lceil\log_2(2/\Delta_{\min})\rceil}\min_{\lambda\in\Delta_{\mathcal{X}}}\max\left[2^{2k}\max_{z,z'\in\mathcal{S}_k}\|z-z'\|^2_{H(\lambda,\theta^*)^{-1}},\gamma(d)\max_{x}\|x\|^2_{H(\lambda,\theta^*)^{-1}}\right]\log(\max(|\mathcal{X}|,|\mathcal{Z}|)\frac{k^2}{\delta})\\
&\quad+c_2\log_2(\Delta_{\min}^{-1})r(\epsilon)\ .
\end{aligned}$$

□

**Lemma 10.** *Define,* $\mathcal{S}_k=\{z\in\mathcal{Z}:(z^*-z)^\top\theta_*\leq 2\cdot 2^{-(k-1)}\}$.

$$\begin{aligned}
\sum_{k=1}^{\log_2(2/\delta_{\min})}2^{2k}\min_{\lambda\in\Delta_{\mathcal{X}}}\max_{z,z'\in\mathcal{S}_k}\|z-z'\|^2_{H(\lambda,\theta^*)}&\leq\log\left(\frac{1}{\Delta_{\min}}\right)\min_{\lambda\in\Delta_{\mathcal{X}}}\max_{z\in\mathcal{Z}\setminus z^*}\frac{\|z^*-z\|^2_{H(\lambda,\theta^*)^{-1}}}{\langle\theta^*,z^*-z\rangle^2}\\
&=\frac{1}{4}\log\left(\frac{2}{\Delta_{\min}}\right)\left(\max_{\lambda\in\Delta_{\mathcal{X}}}\min_{\theta\in\mathcal{C}}\|\theta^*-\theta\|^2_{H(\lambda,\theta^*)}\right)^{-1}
\end{aligned}$$

*where* $\mathcal{C}=\{\theta\in\mathbb{R}^d:\exists z\in\mathcal{Z}\setminus z^*,\theta^\top(z^*-z)\leq 0\}$

*Proof.* Note that,

$$\begin{aligned}
\log\left(\frac{2}{\Delta_{\min}}\right)\min_{\lambda\in\Delta_{\mathcal{X}}}\max_{z\in\mathcal{Z}\setminus z^*}\frac{\|z^*-z\|^2_{H(\lambda,\theta^*)^{-1}}}{\langle\theta^*,z^*-z\rangle^2}&=\log\left(\frac{2}{\Delta_{\min}}\right)\min_{\lambda\in\Delta_{\mathcal{X}}}\max_{k\leq\log_2(2/\Delta_{\min})}\max_{z\in\mathcal{S}_k\setminus z^*}\frac{\|z^*-z\|^2_{H(\lambda,\theta^*)^{-1}}}{\langle\theta^*,z^*-z\rangle^2}\\
&=\log\left(\frac{2}{\Delta_{\min}}\right)\min_{\lambda\in\Delta_{\mathcal{X}}}\max_{k\leq\log_2(2/\Delta_{\min})}2^{-2k+4}\max_{z\in\mathcal{S}_k\setminus z^*}\|z^*-z\|^2_{H(\lambda,\theta^*)^{-1}}\\
&\overset{a}{\geq}16\sum_{k=1}^{\log(2/\Delta_{\min})}2^{2k}\min_{\lambda\in\Delta_{\mathcal{X}}}\max_{z\in S_k\setminus z^*}\|z^*-z\|^2_{H(\lambda,\theta^*)^{-1}}\\
&\overset{b}{\geq}4\sum_{k=1}^{\log(2/\Delta_{\min})}2^{2k}\min_{\lambda\in\Delta_{\mathcal{X}}}\max_{z,z'\in S_k}\|z'-z\|^2_{H(\lambda,\theta^*)^{-1}}
\end{aligned}$$

where $a$ is replacing a max with an average and $b$ is using $\max_{z,z'\in S_k}\|z-z'\|^2_{H(\lambda,\theta^*)^{-1}}=\max_{z,z'\in S_k}\|z-z^*\|^2_{H(\lambda,\theta^*)^{-1}}+\|z'-z^*\|^2_{H(\lambda,\theta^*)^{-1}}-2\|z'-z^*\|_{H(\lambda,\theta^*)^{-1}}\|z-z^*\|_{H(\lambda,\theta^*)^{-1}}\leq 4\max_{z\in S_k}\|z-z^*\|^2_{H(\lambda,\theta^*)^{-1}}$.

We now tackle the second equality in the theorem statement. Define $\mathcal{C}_z=\{\theta\in\mathbb{R}^d:\theta^\top(z^*-z)\leq 0\}$. Note that,

$$\max_{\lambda\in\Delta_{\mathcal{X}}}\min_{\theta\in\mathcal{C}}\|\theta^*-\theta\|^2_{H(\lambda,\theta^*)}=\max_{\lambda\in\Delta_{\mathcal{X}}}\min_{z\in\mathcal{Z}\setminus z^*}\min_{\theta\in\mathcal{C}_z}\|\theta^*-\theta\|^2_{H(\lambda,\theta^*)}$$

For a fixed $\lambda$, standard computation with Lagrange multipliers (as in Theorem 9) shows that the projection,

$$\theta_z:=\arg\min_{\theta\in\mathcal{C}_z}\|\theta^*-\theta\|^2_{H(\lambda,\theta^*)}=\theta^*-\frac{(z^*-z)^\top\theta^*H(\lambda,\theta^*)^{-1}(z^*-z)}{\|z^*-z\|^2_{H(\lambda,\theta^*)^{-1}}}$$

Thus,

$$\|\theta^* - \theta^*\|^2_{H(\lambda,\theta^*)} = \frac{(z^* - z)^\top \theta^*}{\|z^* - z\|^2_{H(\lambda,\theta^*)^{-1}}}$$

and the result follows. □

# D. RAGE-GLM-2

## D.1. Review of confidence bounds of (Faury et al., 2020)

Assume that we have observed a sequence of samples $(x_s, y_s)_{s=1}^T$, where, $\{x_s\}_{s=1}^T \in \mathcal{X}$ and the $x_s$'s are potentially chosen *adaptively*, that is $x_s, 1 \leq s \leq T$ is allowed to depend on the filtration $\mathcal{F}_{s-1} = \{(x_r, y_r)\}_{r=1}^{s-1}$.

For a regularization parameter $\eta > 0$, define

$$H_T(\eta, \theta) := \sum_{s=1}^{T} \dot{\mu}(x_s^\top \theta) x_s x_s^\top + \eta I$$

We begin by defining our estimator. Let

$$\hat{\theta}^{\mathsf{MLE}}_{\eta,T} = \arg\max_{\theta \in \mathbb{R}^d} \sum_{s=1}^{T} y_s \log \mu(x_s^\top \theta) + (1 - y_s) \log(1 - \mu(x_s^\top \theta)) - \frac{\lambda}{2}\|\theta\|_2^2. \tag{21}$$

Define,

$$\hat{\theta}_T = \arg\min_{\|\theta\|_2 \leq S_*} \|g_t(\theta) - g_t(\hat{\theta}^{\mathsf{MLE}}_{\eta,T})\|_{H_T(\eta,\theta)^{-1}} \tag{22}$$

where $g_T(\theta) = \sum_{s=1}^T \mu(x_s^\top \theta) x_s + \eta\theta$. Finally, define

$$\gamma_T(\delta) = \sqrt{\eta}(S_* + 1/2) + \frac{2}{\sqrt{\eta}} \log(1/\delta) + \frac{2d}{\sqrt{\eta}} \log(2(1 + \frac{T}{d\eta})^{1/2})$$

We recall the following lemma from (Faury et al., 2020).

**Lemma 11** (Lemma 11 of (Faury et al., 2020))**.** *On an event $\mathcal{E}$ which is true with probability greater than $1-\delta$, for all $t \geq 1$*

$$\theta^* \in \{\theta \in \mathbb{R}^d : \|\theta\| \leq S_*, \|\theta - \hat{\theta}_T\|_{H_T(\eta,\theta)} \leq (2 + 4S_*)\gamma_T(\delta)\}$$

In the following we will take $\eta = (d + \log(1/\delta))/(S_* + 1/2)$. Plugging this in to $\gamma_T(\delta)$

$$\begin{aligned}
&\sqrt{\eta}(S_* + 1/2) + \frac{2}{\sqrt{\eta}} \log(1/\delta) + \frac{2d}{\sqrt{\eta}} \log\left(2(1 + \frac{T}{d\eta}\right)^{1/2}) \\
&\qquad = \sqrt{\frac{d + \log(1/\delta)}{S_* + 1/2}}(S_* + 1/2) + \frac{2}{\sqrt{\frac{d+\log(1/\delta)}{S_*+1/2}}} \log(1/\delta) + \frac{2d}{\sqrt{\frac{d+\log(1/\delta)}{S_*+1/2}}} \log\left(2\left(1 + \frac{T}{d\frac{d+\log(1/\delta)}{S_*+1/2}}\right)^{1/2}\right) \\
&\qquad \leq \sqrt{d + \log(1/\delta)}\sqrt{S_* + 1/2} + 2\sqrt{S_* + 1/2}\sqrt{\log(1/\delta)} + \frac{2d\sqrt{S_* + 1/2}}{\sqrt{d}} \log(2(1 + \frac{T(2S_* + 1)}{2d})^{1/2}) \\
&\qquad = \sqrt{S_* + 1/2}\left(\sqrt{d + \log(1/\delta)} + 2\sqrt{\log(1/\delta)} + 2\sqrt{d} \log\left(2\left(1 + \frac{T(2S_* + 1)}{2d}\right)^{1/2}\right)\right) \\
&\qquad \leq \sqrt{S_* + 1/2}\left(\sqrt{d}\left(1 + 2\log(2) + \frac{1}{2}\log\left(1 + \frac{T(2S_* + 1)}{2d}\right)\right) + 3\sqrt{\log 1/\delta}\right) \\
&\qquad \leq 3\sqrt{S_* + 1/2}\left(\sqrt{d}\log\left(\frac{T(2S_* + 1)}{2d}\right) + \sqrt{\log 1/\delta}\right)
\end{aligned}$$

**Algorithm 4** RAGE-GLM-2

**Input:** $\epsilon$, $\delta$, $\mathcal{X}$, $\mathcal{Z}$, $\kappa_0$, $S_*$, effective rounding procedure **round**$(n, \epsilon, \lambda)$, $\eta = (d + \log(1/\delta))/(S_* + 1/2)$

**initialize** $t = 1, \mathcal{Z}_1 = \mathcal{Z}, r(\epsilon) = d^2/\epsilon, c = c(S_*, \epsilon) = 48\sqrt{(1+\epsilon)(2S_*+1)^3}$

$\theta_0 \leftarrow$ **BurnIn**$(\mathcal{X}, \kappa_0)$ ▷ Burn-in phase

**while** $|\mathcal{Z}_t| > 1$ **do** ▷ Elimination phase

  $f(\lambda) := \min_{z,z' \in \mathcal{Z}_t} \|z - z'\|^2_{H(\lambda,\theta_0)^{-1}}$

  $\lambda_t = \arg\min_{\lambda \in \Delta_{\mathcal{X}}} f(\lambda)$

  $r_t = \left\lceil 2^{2t} c^2 f(\lambda_t) \left( \sqrt{d} \log(c^2 2^{2t} (2S_* + 1) f(\lambda_t)/d) + \sqrt{\log(t^2 |\mathcal{Z}|^2/\delta)} \right)^2 \right\rceil$

  $n_t = \max\{r_t, r(\epsilon)\}$

  $x_1, \cdots, x_{n_t} \leftarrow$ **round**$(n, \epsilon, \lambda)$

  Observe rewards $y_1, \cdots, y_{n_t} \in \{0, 1\}$

  Compute $\hat{\theta}_t$ on the samples $\{(x_s, y_s)_{s=1}^{n_t}\}$ ▷ (Use Eq (22))

  $\hat{z}_t = \arg\max_{z \in \mathcal{Z}_t} \hat{\theta}_t^\top z$

  $\mathcal{Z}_{t+1} \leftarrow \mathcal{Z}_t \setminus \left\{ z \in \mathcal{Z}_t : \hat{\theta}_t^\top (\hat{z}_t - z) \geq 2^{-t} \right\}$

  $t \leftarrow t + 1$

**return** $\hat{z}_t$

where the last line uses, $(1 + 2\log(2) + 1/2\log(1+x)) \leq 3\log(x), x \geq 2$. So as long as $T \geq 4d$, we have the following bound.

$$\gamma_T(\delta) \leq 3\sqrt{2S_* + 1}\left[\sqrt{d}\log\left(\frac{T(2S_*+1)}{2d}\right) + \sqrt{\log(1/\delta)}\right] =: \Gamma_T(\delta)\ .$$

The guarantee that $T \geq 4d$ will be satisfied by the rounding procedures in the algorithm - indeed, taking $\epsilon \leq 1/2$ guarantees that the minimum number of samples we take in each round $r(\epsilon) = (d(d+1)+2)/\epsilon \geq 4d$.

### D.2. Proof of Sample Complexity

We now provide a sample complexity for Algorithm 4. In this section, we take $\theta_t$ as defined in Eq. (22) using the samples $\{(x_s, y_s)\}_{s=1}^{n_t}$ in each round $t$.

In the regularized setting, rounding implies that,

$$\begin{aligned} H_t(\eta, \theta^*) &:= H_t(\theta^*) + \eta I && (23) \\ &\geq \frac{n}{1+\epsilon} \sum_{x \in \mathcal{X}} \lambda_x \dot{\mu}(x^\top \theta^*) x x^\top + \eta I && (24) \\ &\geq \frac{n}{1+\epsilon} H(\lambda, \theta^*) + \eta I && (25) \\ &\geq \frac{n}{1+\epsilon} H(\lambda, \theta^*) && (26) \end{aligned}$$

Define

$$\mathcal{E}_1 := \left\{ \frac{1}{3} H(\lambda_0, \theta^*) \leq H(\lambda_0, \hat{\theta}_0) \leq 3H(\lambda_0, \theta^*) \right\}$$

By Lemma 6, $\mathbb{P}(\mathcal{E}_1) \geq 1 - \delta$.

Define

$$\mathcal{E}_2 = \cap_{t=1}^{\infty} \{\forall z \in \mathcal{Z}_t, |\langle z^* - z, \hat{\theta}_t - \theta^* \rangle| \leq 2^{-t}\}$$

**Lemma 12.** $\mathbb{P}(\mathcal{E}_2 \cap \mathcal{E}_1) \geq 1 - 3\delta$ *and on* $\mathcal{E}_1 \cap \mathcal{E}_2$*,* $z^* \in \mathcal{Z}_t$ *for all* $t$*.*

*Proof.* **Claim 1:** $\mathbb{P}(\mathcal{E}_2 | \mathcal{E}_1) \geq 1 - \delta$**.** Assuming $\mathcal{E}_1$, For $z \in \mathcal{Z}_t$, with probability greater than $1 - \frac{\delta}{t^2|\mathcal{Z}|}$

$$|(z^* - z)^\top (\hat{\theta}_t - \theta^*)| \leq \|z^* - z\|_{H_t(\eta,\theta^*)^{-1}} \|\theta^* - \hat{\theta}_t\|_{H_t(\eta,\theta^*)}$$

$$
\begin{aligned}
&\leq (2+4S_*)\|z^*-z\|_{H_t(\eta,\theta^*)^{-1}}\Gamma_{n_t}(\delta) && \text{(Lemma 11)}\\
&\leq 2(1+2S_*)\sqrt{\frac{1+\epsilon}{n}}\|z^*-z\|_{H(\lambda_t,\theta^*)^{-1}}\Gamma_{n_t}(\delta) && \text{(Rounding Lemma 13)}\\
&\leq 8(1+2S_*)\sqrt{\frac{1+\epsilon}{n}}\|z^*-z\|_{H(\lambda_t,\theta_0)^{-1}}\Gamma_{n_t}(\delta) && (\mathcal{E}_1)\\
&\leq 8(1+2S_*)\sqrt{\frac{(1+\epsilon)f(\lambda_t)}{n}}\Gamma_{n_t}(\delta)
\end{aligned}
$$

We wish for this quantity to be bounded above by $2^{-t}$. Plugging in $\Gamma_{n_t}(\delta)$, it suffices to take

$$\sqrt{n_t} \geq 24\cdot 2^t(2S_*+1)^{3/2}\sqrt{(1+\epsilon)f_t}\left[\sqrt{d}\log\left(\frac{n_t(2S_*+1)}{2d}\right)+\sqrt{\log(t^2|\mathcal{Z}|/\delta)}\right]$$

where for ease of notation we have denoted $f_t = f(\lambda_t)$. Using Lemma 15 below, shows that it suffices to take,

$$n_t = \left\lceil c^2 2^{2t} f_t\left(\sqrt{d}\log(c^2 2^{2t}(2S_*+1)\hat{\rho}_t/d)+\sqrt{\log(t^2|\mathcal{Z}|/\delta)}\right)^2\right\rceil$$

where $c = 2\cdot 24\sqrt{1+\epsilon}(2S_*+1)^{3/2}$, which is precisely the number of samples we take in the algorithm. Union bounding over $z\in\mathcal{Z}_t\subset\mathcal{Z}$ and $t\geq 1$ now gives the result.

**Claim 2:** $\mathbb{P}(\mathcal{E}_1\cap\mathcal{E}_2)\geq 1-3\delta$**.** Note that,

$$
\begin{aligned}
\mathbb{P}(\mathcal{E}_1^c\cup\mathcal{E}_2^c) &\leq \mathbb{P}(\mathcal{E}_2^c)+\mathbb{P}(\mathcal{E}_1^c)\\
&= \mathbb{P}(\mathcal{E}_2^c|\mathcal{E}_1^c)\mathbb{P}(\mathcal{E}_1^c)+\mathbb{P}(\mathcal{E}_2^c|\mathcal{E}_1)\mathbb{P}(\mathcal{E}_1)+\mathbb{P}(\mathcal{E}_1^c)\\
&\leq \mathbb{P}(\mathcal{E}_2^c|\mathcal{E}_1)+2\mathbb{P}(\mathcal{E}_1^c)\\
&\leq \delta+2\delta\\
&\leq 3\delta
\end{aligned}
$$

**Claim 3: On** $\mathcal{E}_1\cap\mathcal{E}_2$**,** $z^*\in\mathcal{Z}_t$ **for all** $t\geq 1$**.** Identical argument to Lemma 9

□

**Remark.** We point out that this analysis is not particularly tight, and many constants and the dependence upon $S_*$ can be improved upon in practice. In particular, we can trade off a smaller constant for a larger burn-in phase.

**Theorem 7** (Sample Complexity)**.** *Algorithm 4, returns $z^*$ with probability greater than $1-2\delta$ in a number of samples no more than*

$$O\Bigg((1+\epsilon)(2S_*+1)^3\sum_{r=1}^{\log_2(1/\Delta_{\min})}2^{2t}\rho_t\left(d\log^2\left(\frac{(2S_*+1)\rho_t}{\Delta_{\min}}\right)+\log(t^2|\mathcal{Z}|^2/\delta)\right)$$
$$+r(\epsilon)\log_2(1/\Delta_{\min})+\kappa_0^{-1}(1+\epsilon)d\gamma(d)\log(|\mathcal{X}|/\delta)\Bigg)$$

*where $\mathcal{S}_t = \{z\in\mathcal{Z}:(z^*-z)^\top\theta_*\leq 2\cdot 2^{-t}\}$ and $\rho_t = \min_{\lambda\in\Delta_\mathcal{X}}\max_{z,z'\in\mathcal{S}_t}\|z-z'\|^2_{H(\lambda,\theta^*)^{-1}}$ and we assume $\epsilon\leq 1/2$.*

*Proof.* Firstly note that $\mathbb{P}(\mathcal{E}_1^c\cup\mathcal{E}_2^c)\leq 2\delta$. For the remainder of the proof we will assume that $\mathcal{E}_1\cap\mathcal{E}_2$ holds.

By Lemma 12, we have that $\mathcal{Z}_t\subset\mathcal{S}_t$, likewise on $\mathcal{E}_1$ we have that $H(\lambda_t,\theta_0)\geq\frac{1}{4}H(\lambda_t,\theta^*)$. Thus, in each round,

$$\max_{z,z'\in\mathcal{Z}_t}\|z-z'\|^2_{H(\lambda_t,\theta_0)^{-1}}\leq 4\max_{z,z'\in\mathcal{S}_t}\|z-z'\|^2_{H(\lambda,\theta^*)^{-1}}$$

Denoting $\rho_t = \min_{\lambda\in\Delta_\mathcal{X}}\max_{z,z'\in\mathcal{S}_t}\|z-z'\|^2_{H(\lambda,\theta^*)^{-1}}$, we see that $f_t\leq\rho_t$. This implies that $n_t\leq 4c^2 2^{2t}\rho_t[\sqrt{d}\log(2c^2 2^{2t}(2S_*+1)\rho_t/d)+\sqrt{\log(t^2|\mathcal{Z}|^2/\delta)}]^2$.

Thus an upper bound on our final sample complexity is given by

$$\begin{aligned}
&\sum_{t=1}^{\log_2(1/\Delta_{\min})} n_t + r(\epsilon)\log(1/\Delta_{\min}) + n_0 \\
&\leq \sum_{t=1}^{\log_2(1/\Delta_{\min})} 4c^2 2^{2t}\rho_t[\sqrt{d}\log(c^2 2^{2t}(2S_*+1)\rho_t/d) + \sqrt{\log(t^2|\mathcal{Z}|^2/\delta)}]^2 + r(\epsilon)\log(1/\Delta_{\min}) + n_0 \\
&\leq 8c^2 \sum_{t=1}^{\log_2(1/\Delta_{\min})} 2^{2t}\rho_t[d\log^2(c^2 2^{2t}(2S_*+1)\rho_t/d) + \log(t^2|\mathcal{Z}|^2/\delta)] + r(\epsilon)\log(1/\Delta_{\min}) + n_0 \\
&\leq 8c^2 \sum_{t=1}^{\log_2(1/\Delta_{\min})} 2^{2t}\rho_t[d\log^2(c^2\frac{4}{\Delta_{\min}^2}(2S_*+1)\rho_t/d) + \log(t^2|\mathcal{Z}|^2/\delta)] + r(\epsilon)\log(1/\Delta_{\min}) + n_0 \\
&= O\Bigg((1+\epsilon)(2S_*+1)^3 \sum_{r=1}^{\log_2(1/\Delta_{\min})} 2^{2t}\rho_t[d\log^2((2S_*+1)\rho_t/\Delta_{\min}) \\
&\qquad + \log(t^2|\mathcal{Z}|^2/\Delta_{\min})] + r(\epsilon)\log(1/\delta) + \kappa_0^{-1}(1+\epsilon)d\gamma(d)\log(|\mathcal{X}|/\delta)\Bigg)
\end{aligned}$$

□

### D.3. Miscellaneous results

We let **round**$(\lambda, n)$ denote an efficient rounding procedure as explained in Chapter 12 of (Pukelsheim, 2006), or summarized in Section B of the Appendix of (Fiez et al., 2019).

**Lemma 13** (Rounding ). *Assume that $\lambda \in \Delta_{\mathcal{X}}$, and that we have sampled $x_1, \cdots, x_n \sim$ **round**$(\lambda, n, \epsilon)$ with $n \geq r(\epsilon) = (d(d+1)+2)/\epsilon$, and $\epsilon \leq 1$. Then, for any $\theta$, $\sum_{s=1}^n \dot\mu(x_s^\top\theta)x_s x_s^\top \succeq \frac{n}{1+\epsilon}\sum_{x\in\mathcal{X}}\lambda_x\dot\mu(x^\top\theta)xx^\top$. This in particular implies*

- *For any $z$,*

$$\|z\|^2_{(\sum_{s=1}^n \dot\mu(x_s^\top\theta)x_s x_s^\top)^{-1}} \leq \frac{(1+\epsilon)}{n}\|z\|^2_{(\sum_{x\in\mathcal{X}}\lambda_x\dot\mu(x^\top\theta)xx^\top)^{-1}}$$

- $\lambda_{\min}(\sum_{s=1}^n \dot\mu(x_s^\top\theta)x_s x_s^\top) \geq \frac{n}{1+\epsilon}\lambda_{\min}(\sum_{x\in\mathcal{X}}\lambda_x\dot\mu(x^\top\theta)xx^\top)$

*Proof.* Let $s = (n_x)_{x\in\mathcal{X}} \in \mathbb{N}^{\mathcal{X}}$ denote the allocation returned by the rounding procedure and let $\gamma = s/n \in \Delta_{\mathcal{X}}$ denote the associated fractional allocation. Now consider,

$$\epsilon_{\gamma/\lambda} = \min_{x\in\text{supp}(\lambda)}\frac{\gamma_x}{\lambda_x} = \max\{\kappa \geq 0 : \gamma_x \geq \kappa\lambda_x \text{ for all } x \in \mathcal{X}\}$$

By definition of $\epsilon_{\gamma/\lambda}$,

$$\sum_{x\in\mathcal{X}}\gamma_x\dot\mu(x^\top\theta)xx^\top \geq \epsilon_{\gamma/\lambda}\sum_{x\in\mathcal{X}}\lambda_x\dot\mu(x^\top\theta_*)xx^\top$$

By Theorem 12.7 of (Pukelsheim, 2006), $\epsilon_{\gamma/\lambda} \geq 1 - p/n$ where $p = |\text{supp}\lambda|$. When $\dim\operatorname{span}\mathcal{X} = d$, Caratheodory's Theorem (Vershynin, 2018), implies $p \leq d(d+1)/2 + 1$. Hence,

$$\begin{aligned}
\sum_{s=1}^n \dot\mu(x_s^\top\theta)x_s x_s^\top &= n\sum_{x\in\mathcal{X}}\gamma_x\dot\mu(x^\top\theta)xx^\top \\
&\geq n(1-\frac{p}{n})\sum_{x\in\mathcal{X}}\lambda_x\dot\mu(x^\top\theta)xx^\top \\
&\geq \frac{n}{1+\epsilon}\sum_{x\in\mathcal{X}}\lambda_x\dot\mu(x^\top\theta)xx^\top
\end{aligned}$$

as long as $n \geq (d(d+1)+2)/\epsilon$. The result now follows. □

As long as $n_t \geq r(\epsilon)$, we have a guarantee that $H_t(\theta) \geq \frac{n_t}{1+\epsilon} H(\lambda_t)$ for any $\theta$. This implies, $H_t(\theta)^{-1} \leq \frac{1+\epsilon}{n_t} H(\lambda_t)^{-1}$. This is a modification of the argument in Fiez et al. (2019).

**Lemma 14.** *Let $\theta \in \mathbb{R}^d$. Suppose $D = \max_{x\in\mathcal{X}} |x^\top(\theta - \theta^*)| \leq 1$. Then, for all $x$,*

$$\frac{1}{2D+1} H(\lambda, \theta^*) \leq H(\lambda, \theta) \leq (2D+1) H(\lambda, \theta^*)$$

*Proof.* The proof is identical to Lemma 5. □

**Lemma 15.** *Assume $a > 0, b > 2$, then for any $t \geq \max[(2a)^2(\log((2a)^2/c) + \log b + d)^2, 2c]$ we have that $\sqrt{t} \geq a[\log(b + t/c) + d]$*

*Proof.* Note that if $t > 2c$, $a\log(b + t/c) \leq a\log(b) + a\log(t/c)$, so it suffices to show $\sqrt{t} \geq a(\log(b) + \log(t/c)) + ad$ or equivalently, $\frac{1}{a}\sqrt{t} - \log(b) - d \geq \log(t/c)$, or doing the substitution $u = t/c$, $\frac{\sqrt{c}}{a}\sqrt{u} - \log(b) - d \geq \log(u)$ for $t \geq (2a)^2(\log(2a)^2/c + \log b + d)^2$. However, this follows directly from Proposition 6 of (Antos et al., 2010). □

# E. Lower Bounds

## E.1. Instance Dependent Lower Bound

The lemma shows that the main term of the sample complexity of Theorem 2 (from the main paper) given as $\rho^*$ in Section 3.1 is bounded by a natural experimental design arising from the true problem parameter.

Finally, we provide the following information theoretic lower bound for any PAC-$\delta$ algorithm. Define,

$$\beta(a, b) = \int_0^1 (1-t)\dot{\mu}(a + t(b-a))dt$$

and analogous to $H(\lambda, \theta)$ we define two additional matrix valued functions,

$$G(\lambda, \theta_1, \theta_2) = \sum_{x\in\mathcal{X}} \lambda_x \alpha(x, \theta_1, \theta_2) x x^\top \tag{27}$$

$$K(\lambda, \theta_1, \theta_2) = \sum_{x\in\mathcal{X}} \lambda_x \beta(x, \theta_1, \theta_2) x x^\top \tag{28}$$

**Theorem 8.** *Any PAC-$\delta$ algorithm for the pure exploration logistic bandits problem has a stopping time $\tau$ satisfying,*

$$\begin{aligned}\mathbb{E}[\tau] &\geq \min_{\lambda\in\Delta_Z} \max_{\substack{z\in\mathcal{Z}\setminus z^* \\ \theta\in\mathbb{R}^d \\ \theta^\top(z^*-z)\leq 0}} \frac{1}{\sum_{x\in\mathcal{X}} \lambda_x \mathsf{KL}(\nu_{x,\theta^*}|\nu_{x,\theta})} \log\left(\frac{1}{2.4\delta}\right) \\ &= c(\lambda)^{-1} \log\frac{1}{2.4\delta}, c(\lambda) = \max_{\lambda\in\Delta\mathcal{X}} \min_{\mathcal{Z}\setminus z^*} \|\theta - \theta_z\|^2_{K(\lambda,\theta^*,\theta_z)},\end{aligned}$$

*where firstly, $\nu_{x,\theta} = \mathsf{Bernoulli}(x^\top\theta)$, and secondly $\theta_z := \min_{\theta\in\mathbb{R}^d:\theta^\top(z^*-z)\leq 0} \|\theta - \theta_z\|^2_{K(\lambda,\theta^*,\theta_z)}$ and is given explicitly as the solution to the fixed-point equation*

$$\theta_z = \theta^* - \frac{(z^*-z)^\top\theta^* G(\lambda, \theta_z, \theta^*)^{-1}(z^*-z)}{\|z^*-z\|^2_{G(\lambda,\theta_z,\theta^*)^{-1}}}.$$

In general, it is not clear how to compare our upper bound from Theorem 1 to this lower bound due to the non-explicit nature of $G(\lambda, \theta_z, \theta^*)$. The quantity, $\max_{\lambda\in\Delta_\mathcal{X}} \min_{z\in\mathcal{Z}\setminus\{z^*\}} \|\theta^* - \theta_z\|^2_{H(\lambda,\theta^*)}$ can be interpreted as a lower bound arising from a quadratic approximation of the KL-divergence in the first line of the lower bound by the Fisher information matrix.

In this section, we provide an information theoretic lower bound for any PAC-$\delta$ algorithm. Define,

$$\beta(a, b) = \int_0^1 (1-t)\dot{\mu}(a + t(b-a))dt$$

and analogous to $H(\lambda, \theta)$ we define two additional matrix valued functions,

$$G(\lambda, \theta_1, \theta_2) = \sum_{x\in\mathcal{X}} \lambda_x \alpha(x, \theta_1, \theta_2) x x^\top \tag{29}$$

$$K(\lambda, \theta_1, \theta_2) = \sum_{x \in \mathcal{X}} \lambda_x \beta(x, \theta_1, \theta_2) x x^\top \tag{30}$$

**Theorem 9.** *Any PAC-$\delta$ algorithm for the pure exploration logistic bandits problem has a stopping time $\tau$ satisfying,*

$$\mathbb{E}[\tau] \geq c(\lambda)^{-1} \log \frac{1}{2.4\delta}, c(\lambda) = \max_{\lambda \in \Delta\mathcal{X}} \min_{z \neq z^* \in \mathcal{Z}} \|\theta - \theta_z\|^2_{K(\lambda, \theta^*, \theta_z)}$$

*where $\theta_z := \min_{\theta \in \mathbb{R}^d : \theta^\top (z^* - z) \leq 0} \|\theta - \theta_z\|^2_{K(\lambda, \theta^*, \theta_z)}$ and is given explicitly as the solution to the fixed-point equation*

$$\theta_z = \theta^* - \frac{(z^* - z)^\top \theta^* G(\lambda, \theta_z, \theta^*)^{-1} (z^* - z)}{\|z^* - z\|^2_{G(\lambda, \theta_z, \theta^*)^{-1}}}$$

*Proof.* Let $\mathcal{C} = \{\theta \in \Theta : \exists z \in Z, \theta^\top (z^* - z) \leq 0\}$. The transportation theorem of (Kaufmann et al., 2016) implies that any algorithm that is $\delta$-PAC, takes at least $T$ samples with

$$\begin{aligned}
\mathbb{E}[T] &\geq \log\left(\frac{1}{2.4\delta}\right) \min_{\lambda \in \Delta_Z} \max_{\theta \in \mathcal{C}} \frac{1}{\sum_{x \in \mathcal{X}} \lambda_x \mathsf{KL}(\nu_{x,\theta^*} | \nu_{x,\theta})} \\
&\geq \log\left(\frac{1}{2.4\delta}\right) \min_{\lambda \in \Delta_Z} \max_{z \in \mathcal{Z} \setminus z^*} \max_{\theta \in \mathbb{R}^d, \theta^\top (z^* - z) \leq 0} \frac{1}{\sum_{x \in \mathcal{X}} \lambda_x \mathsf{KL}(\nu_{x,\theta^*} | \nu_{x,\theta})}
\end{aligned}$$

where $\nu_{x,\theta}$ is the distribution of arm $x$ under the parameter vector $\theta$, i.e. $\nu_{x,\theta} = \mathsf{Bernoulli}(x^\top \theta)$

For a fixed $z' \in \mathcal{Z}$, s.t. $z' \neq z^*$ consider

$$\min_{\theta \in \mathbb{R}^d, \theta^\top (z^* - z') \leq 0} \sum_{x \in \mathcal{X}} \lambda_x \mathsf{KL}(\nu_{x,\theta^*} | \nu_{x,\theta}).$$

We have that

$$\begin{aligned}
\mathsf{KL}(\nu_{x,\theta^*} | \nu_{x,\theta}) &= \mu(x^\top \theta^*) \log\left(\frac{\frac{e^{x^\top \theta^*}}{1 + e^{x^\top \theta^*}}}{\frac{e^{x^\top \theta}}{1 + e^{x^\top \theta}}}\right) + (1 - \mu(x^\top \theta^*)) \log\left(\frac{\frac{1}{1 + e^{x^\top \theta^*}}}{\frac{1}{1 + e^{x^\top \theta}}}\right) \\
&= \mu(x^\top \theta^*) x^\top (\theta^* - \theta) + \log\left(\frac{\frac{1}{1 + e^{x^\top \theta^*}}}{\frac{1}{1 + e^{x^\top \theta}}}\right) \\
&= \mu(x^\top \theta^*) x^\top (\theta^* - \theta) + \log\left(\frac{1 - \mu(z^\top \theta^*)}{1 - \mu(x^\top \theta)}\right) \\
&= \mu(x^\top \theta^*) x^\top (\theta^* - \theta) + \log(1 - \mu(x^\top \theta^*)) - \log(1 - \mu(x^\top \theta))
\end{aligned}$$

Differentiating with respect to $\theta$ gives,

$$\nabla_\theta \mathsf{KL}(\nu_{x,\theta^*} | \nu_{x,\theta}) = -\mu(x^\top \theta^*) x + \frac{\dot{\mu}(x^\top \theta) x}{1 - \mu(x^\top \theta)} = (\mu(x^\top \theta) - \mu(x^\top \theta^*)) x$$

using the fact that $\dot{\mu}(a) = \mu(a)(1 - \mu(a))$ so this implies

$$\nabla_\theta \sum_{x \in \mathcal{X}} \lambda_x \mathsf{KL}(\nu_{x,\theta^*} | \nu_{x,\theta}) = \sum_{x \in \mathcal{X}} \lambda_x (\mu(x^\top \theta) - \mu(x^\top \theta^*)) x$$

Assuming $\Theta = \mathbb{R}^d$ and letting $\psi$ denote the Lagrange Multiplier corresponding to the constraint $\theta^\top (z^* - z) \leq 0$ gives that the minimal $\theta$ satisfies,

$$\sum_{x \in \mathcal{X}} \lambda_x (\mu(x^\top \theta) - \mu(x^\top \theta^*)) x = \psi \cdot (z^* - z)$$

Now by definition, $\mu(x^\top\theta) - \mu(x^\top\theta^*) = \alpha(x,\theta,\theta^*)x^\top(\theta-\theta^*)$ so, this reduces to,

$$\left(\sum_{x\in\mathcal{X}} \lambda_x \alpha(x,\theta,\theta^*)xx^\top\right)(\theta-\theta^*) = \psi(z^*-z) \Rightarrow \theta = \theta^* + \psi G(\lambda,\theta,\theta^*)^{-1}(z^*-z)$$

Let $\theta_z$ be the solution to this fixed point equation. Since we are saturating the constraint, it should be true that $\theta_z(z^*-z)=0$. With this, we take an inner product with $z^*-z$ on both sides to obtain

$$\psi = -\frac{(z^*-z)^\top\theta^*}{\|z^*-z\|^2_{G(\lambda,\theta,\theta^*)^{-1}}}.$$

So finally, we see that

$$\theta_z = \theta^* - \frac{\theta^{*\top}(z^*-z)G(\lambda,\theta_z,\theta^*)^{-1}(z^*-z)}{\|z^*-z\|^2_{G(\lambda,\theta_z,\theta^*)^{-1}}}$$

and

$$\theta_z = \arg\min_{\theta\in\mathbb{R}^d,\theta^\top(z^*-z')\le 0} \sum_{x\in\mathcal{X}} \lambda_x \mathsf{KL}(\nu_{x,\theta^*}|\nu_{x,\theta})$$

Now to finish the proof, note

$$\begin{aligned}
\mathsf{KL}(\nu_{x,\theta^*}|\nu_{x,\theta}) &= \mu(x^\top\theta^*)x^\top(\theta^*-\theta) + \log(1-\mu(x^\top\theta^*)) - \log(1-\mu(x^\top\theta)) \\
&= (\theta^*-\theta)\left[\frac{\mu(x^\top\theta^*)}{(x^\top(\theta^*-\theta))^2} + \frac{\log(1-\mu(z^\top\theta^*))}{(x^\top(\theta^*-\theta))^2} - \frac{\log(1-\mu(x^\top\theta))}{(x^\top(\theta^*-\theta))^2}\right]xx^\top(\theta^*-\theta) \\
&\overset{(a)}{=} \|\theta^*-\theta\|^2_{\beta(z^\top\theta,z^\top\theta^*)zz^\top}
\end{aligned}$$

where the last expression follows from the computation,

$$\begin{aligned}
\beta(a,b) &= \int_0^1 (1-t)\dot{\mu}(a+t(b-a))dt \\
&= \frac{\log\left(\mathrm{e}^{-a}+1\right)}{(b-a)^2} - \frac{\log\left(\mathrm{e}^{-b}+1\right)}{(b-a)^2} - \frac{1}{\left(\mathrm{e}^{b}+1\right)(b-a)}
\end{aligned}$$

□

In general, it is not clear how to compare our upper bound from Theorem 2 (in the main paper) to this lower bound due to the non-explicit nature of $G(\lambda,\theta_z,\theta^*)$. In the case of Gaussian linear bandits, previous work has shown an elimination scheme similar to Algorithm 1 is indeed near optimal.

### E.2. $1/\kappa_0$ Lower Bounds

In this section, we prove that there exist bounds where a dependence on $1/\kappa_0$ is necessary. Throughout, we take $\mathbb{E}_\theta$ and $\mathbb{P}_\theta$ to denote expectation and probability under an instance where parameter vector $\theta^*$ is equal to $\theta$.

**Theorem 10.** *Fix $\delta_1 < 1/16$, $d \ge 4$, and $\epsilon \in (0,1/2]$ such that $d\varepsilon^2 \ge 12.2$. Let $\mathcal{Z}$ denote the action set and $\Theta$ denote a family of possible parameter vectors. There exists instances satisfying the following properties simultaneously*

1. *$|\mathcal{Z}| = |\Theta| = e^{\epsilon^2 d/4}$ and $\|z\| = 1$ for all $z\in\mathcal{Z}$.*
2. *$S = \|\theta_*\| = O(\epsilon^2 d)$*
3. *Any algorithm that succeeds with probability at least $1-\delta_1$ satisfies*

$$\exists\theta\in\Theta \text{ such that } \mathbb{E}_\theta[T_{\delta_1}] > \Omega\left(e^{\epsilon^2 d/4}\right) = c\left(\frac{1}{\kappa_0}\right)^{\frac{1-\epsilon}{1+3\epsilon}}$$

*where $T_{\delta_1}$ is the random variable of the number of samples drawn by an algorithm and $c$ is an absolute constant.*

To prove this, we first state a more general theorem about lower bounds for logistic bandits.

**Theorem 11.** *Fix $\delta_1 < 1/16$, $n \in \mathbb{N} : n > 20$, and a set of arms $\mathcal{Z} = \{z_1, \cdots, z_n\} \in \mathbb{R}^d$. Consider a family of parameter vectors $\Theta = \{\theta_1, \cdots, \theta_n\} \in \mathbb{R}^d$ such that for every $i \in [n]$,*

1. $\mu(\theta_i^T z_i) \in [1-\delta_2, 1]$
2. $\mu(\theta_i^T z_j) \in [0, \delta_2]$ *for all* $j \neq i$

*If $\delta_2 \leq \frac{1}{2n}$, then any $\delta_1$-PAC algorithm satisfies*

$$\exists \theta \in \Theta \text{ such that } \mathbb{E}_\theta[T_{\delta_1}] > \frac{n}{16}$$

*where $T_{\delta_1}$ is the random variable of the number of samples drawn by an algorithm.*

***Proof of Theorem** 10.* We begin with a construction based on the technique from (Dong et al., 2019) but optimized for our setting. Then we choose $S$ to satisfy the conditions of Theorem 11 while controlling $\kappa_0$.

**Step 1: Constructing $A$ and $\Theta$**

There exists at least $n = \lfloor \exp(\epsilon^2 d/4) \rfloor > 20$ vectors on the sphere in $\mathbb{R}^{d-1}$, $a_1, \cdots, a_n$ such that $|a_i^T a_j| < 1/2$ and $\|a_i\| = 1$ for all $i$. Define arms $z_1, \cdots z_n$ such that

$$z_i = (\cos(u), \sin(u) a_i) \in \mathbb{R}^d$$

where $u = \tan^{-1}\left(\sqrt{\frac{2}{1+\epsilon}}\right)$. Similarly, define a family of $\theta$'s such that

$$\frac{\theta_i}{S} = (-\cos(u), \sin(u) a_i) \in \mathbb{R}^d$$

where $S$ is the norm of all of the $\theta_i$'s to be specified later. Since $\|a_i\| = 1$ for all $i$, we have that $\|z_i\| = 1$ and $\|\theta_i\| = S$ for all $i$. Then we have that

$$z_i^T \frac{\theta_i}{S} = -\cos(u)^2 + \sin(u)^2.$$

Plugging in our choice of $u$ and recalling that $\sin(\tan^{-1}(x)) = \frac{x}{\sqrt{1+x^2}}$ and $\cos(\tan^{-1}(x)) = \frac{1}{\sqrt{1+x^2}}$. Therefore,

$$z_i^T \frac{\theta_i}{S} = \frac{-1}{1+\frac{2}{1+\epsilon}} + \frac{\frac{2}{1+\epsilon}}{1+\frac{2}{1+\epsilon}} = \frac{1-\epsilon}{3+\epsilon} > 0$$

Furthermore, we have that

$$z_j^T \frac{\theta_i}{S} = -1 + \sin(u)^2 (1 + a_i^T a_j).$$

We have that $|a_i^T a_j| \leq \epsilon$. Therefore,

$$z_j^T \frac{\theta_i}{S} \leq -1 + \frac{\frac{2}{1+\epsilon}}{1+\frac{2}{1+\epsilon}}(1+\epsilon) = \frac{\epsilon - 1}{3+\epsilon} < 0$$

and

$$z_j^T \frac{\theta_i}{S} \geq -1 + \frac{\frac{2}{1+\epsilon}}{1+\frac{2}{1+\epsilon}}(1-\epsilon) = \frac{-1-3\epsilon}{3+\epsilon}.$$

Taken together, we have that

$$\max_{i,j} |z_i^T \theta_j| \in \left[S\frac{1-\epsilon}{3+\epsilon}, S\frac{1+3\epsilon}{3+\epsilon}\right].$$

**Step 2: Choosing $S$ to satisfy Theorem 11**

To invoke the result of Theorem 11, we require that $\mu(\theta_i^T z_i) \geq 1 - \delta_2$ and $\mu(\theta_i^T z_j) \leq \delta_2$ for $j \neq i$ for $\delta_2$ defined therein. For the above construction, we require an $S$ that satisfies: 1) $\mu\left(S\frac{1-\epsilon}{3+\epsilon}\right) \geq 1 - \delta_2$ and 2) $\mu(S\frac{\epsilon-1}{3+\epsilon}) \leq \delta_2$. Clearly this can be achieved by taking $S \to \infty$. Using the fact that $\mu(-x) = 1 - \mu(x)$ for the logistic function as well as its monotonicity, we see that both are satisfied for $S$ such that $\mu(S\frac{\epsilon-1}{3+\epsilon}) \leq \delta_2$.

Note that $\mu(x)$ is invertible with inverse $\mu^{-1}(x) = \log\left(\frac{x}{1-x}\right)$. Hence

$$S \geq \frac{3+\epsilon}{\epsilon-1}\mu^{-1}(\delta_2) = \frac{3+\epsilon}{1-\epsilon}\log\left(\frac{1-\delta_2}{\delta_2}\right).$$

implies that for any $\delta_1$-PAC algorithm there exists a $\theta \in \Theta$ such that

$$\mathbb{E}_\theta[T_{\delta_1}] > \frac{n}{16} = \frac{1}{16}\lfloor e^{d\epsilon^2/4}\rfloor.$$

**Step 3: Choosing $S$ to control $1/\kappa_0$**

Any $S$ that satisfies the constraint in step 2 satisfies the conditions of Theorem 11 and implies a sample complexity lower bounds. As $\kappa_0^{-1} = O\left(e^S\right)$, to have the tightest correspondence between $\kappa$ and $n$, we want $S$ as small as possible. Therefore, we take

$$S = \frac{3+\epsilon}{1-\epsilon}\log\left(\frac{1-\delta_2}{\delta_2}\right).$$

By construction, we have

$$\max_{i,j}|z_i^T\theta_j| \leq S\frac{1+3\epsilon}{3+\epsilon}.$$

This implies that

$$\begin{aligned}\min_{i,j}\mu\left(z_i^T\theta_j\right) \geq \mu\left(-S\frac{1+3\epsilon}{3+\epsilon}\right) &= \mu\left(-\frac{1+3\epsilon}{3+\epsilon}\times\frac{3+\epsilon}{1-\epsilon}\log\left(\frac{1-\delta_2}{\delta_2}\right)\right)\\ &= \mu\left(-\frac{1+3\epsilon}{1-\epsilon}\log\left(\frac{1-\delta_2}{\delta_2}\right)\right)\\ &= \frac{1}{1+\left(\frac{1-\delta_2}{\delta_2}\right)^{\alpha_\epsilon}} =: \delta_3\end{aligned}$$

where in the final line we have defined $\alpha_\epsilon := \frac{1+3\epsilon}{1-\epsilon}$. Then, using $1/(1-\delta_3) \leq 2$ for all $n \geq 1$,

$$\frac{1}{\kappa_0} = \frac{1}{\delta_3(1-\delta_3)} \leq \frac{2}{\delta_3} = 2\left(1+\left(\frac{1-\delta_2}{\delta_2}\right)^{\alpha_\epsilon}\right).$$

**Step 4: Putting it all together**

By Theorem 11, the above holds for $\delta_2 \leq \frac{1}{2n}$. Choose $\delta_2 = \frac{1}{2n}$. Hence,

$$\frac{1}{\kappa_0} \leq 2\left(1+\left(\frac{1-\delta_2}{\delta_2}\right)^{\alpha_\epsilon}\right) \iff n \geq \frac{1}{2}\left(\frac{1}{2}\cdot\frac{1}{\kappa_0}-1\right)^{-\alpha_\epsilon}+1 = c_1\left(\frac{1}{\kappa_0}\right)^{-\frac{1-\epsilon}{1+3\epsilon}}$$

for an absolute constant $c_1$. Furthermore, there exists a constant $c_2$ such that $S = c_2\log(n)$. Noting that $n = \Omega(e^{\epsilon^2 d})$ completes the proof.

□

***Proof of Theorem 11.*** Suppose not. Then, for every $\theta \in \Theta$, we have $\mathbb{E}_\theta[T_{\delta_1}] \leq n/16$.

**Step 1: Defining event $\mathcal{E}_i$ that leads to errors**

Let $R_t$ be the reward received at time $t$. Let $\tau$ to be the first time $t$ the algorithm receives $R_t = 1$:

$$\tau = \begin{cases}\infty & \text{if } \{t \in [1, T_{\delta_1}] : R_t = 1\} \text{ is empty}\\ \min_{t\in[1,T_{\delta_1}]:R_t=1} t & \text{otherwise}\end{cases}. \tag{31}$$

If $\tau = \infty$, then the algorithm only sees reward 0 until termination. Note that $\tau$ is a stopping time. Let $i(\theta)$ be the best arm under the parameter $\theta$. Let $T_j$ be the number of pulls of arm $j$ up to (and including) time $T_{\delta_1}$. We define the following event:

$$\mathcal{E}_{\nu,\theta} = \left\{\mathcal{A} \text{ returns } i(\nu)\right\} \cap \left\{T_{\delta_1} \le \frac{n}{4}\right\} \cap \left\{T_{i(\theta)} = 0\right\} \cap \{\tau = \infty\} \ ,$$

which is bad and should not happen frequently when $\theta$ is true but is likely to happen under $\nu$, roughly speaking. Define

$$\widehat{\mathsf{KL}}(\nu,\theta) := \sum_{t=1}^{T_{\delta_1}} \sum_{z\in\mathcal{Z}} \mathbb{1}[z = a_t] \left[ R_t \log\left(\frac{\mu(\nu^T z)}{\mu(\theta^T z)}\right) + (1-R_t)\log\left(\frac{1-\mu(\nu^T z)}{1-\mu(\theta^T z)}\right)\right] .$$

On the event $\mathcal{E}_{\nu,\theta}$ using the assumptions of the theorem, we have that

$$\widehat{\mathsf{KL}}(\nu,\theta) \le \frac{n}{4}\log\left(\frac{1}{1-\delta_2}\right)$$

Then,

$$\begin{aligned}
\mathbb{P}_\theta(\mathcal{A} \text{ returns } i(\nu)) &\ge \mathbb{P}_\theta(\mathcal{E}_{\nu,\theta}) \\
&= \mathbb{E}_\nu\left[\mathbb{1}\{\mathcal{E}_{\nu,\theta}\}\exp(-\widehat{\mathsf{KL}}(\nu,\theta))\right] \\
&\ge (1-\delta_2)^{n/4}\,\mathbb{E}_\nu\left[\mathbb{1}\{\mathcal{E}_{\nu,\theta}\}\right] \\
&= (1-\delta_2)^{n/4}\,\mathbb{P}_\nu\left(\mathcal{E}_{\nu,\theta}\right) \ .
\end{aligned} \tag{32}$$

Let us fix an arbitrary circular ordering of the members of $\Theta$.[8] Let $\mathsf{q}(\theta)$ be member of $\Theta$ that comes immediately after $\theta$ in the order, and $\mathsf{p}(\theta)$ be member of $\Theta$ that comes immediately before $\theta$. Then,

$$\begin{aligned}
\frac{1}{n}\sum_{\theta\in\Theta}\mathbb{P}_\theta(\mathcal{A} \text{ returns } i(\mathsf{q}(\theta))) &\ge (1-\delta_2)^{n/4}\,\frac{1}{n}\sum_{\theta\in\Theta}\mathbb{P}_{\mathsf{q}(\theta)}\left(\mathcal{E}_{\mathsf{q}(\theta),\theta}\right) && \text{(By (32))} \\
&\overset{(a)}{=} (1-\delta_2)^{n/4}\,\frac{1}{n}\sum_{\nu\in\Theta}\mathbb{P}_\nu\left(\mathcal{E}_{\nu,\mathsf{p}(\nu)}\right) \\
&\overset{(b)}{=} (1-\delta_2)^{n/4}\,\mathbb{P}\left(\mathcal{E}_{\theta,\mathsf{p}(\theta)}\right) \ .
\end{aligned}$$

where $(a)$ is simply a reindexing and $(b)$ is by treating the parameter $\theta$ to be drawn from $\mathsf{Uniform}(\Theta)$. Hereafter, $\mathbb{P}$ and $\mathbb{E}$ without subscripts are w.r.t. the measure on all the random variables including the prior on $\theta$. Note that the terminology 'prior' is just for brevity in this proof and does not imply that our problem setup is Bayesian.

**Step 2: Bounding $\mathbb{P}(\mathcal{E}_{\theta,\mathsf{p}(\theta)})$**

Hereafter, we shorten the notation $\mathcal{E}_{\theta,\mathsf{p}(\theta)}$ as $\mathcal{E}_\theta$. We aim to find a lower bound on $\mathbb{P}(\mathcal{E}_\theta)$. For this, we upper bound $\mathbb{P}(\overline{\mathcal{E}_\theta})$:

$$\begin{aligned}
&\mathbb{P}(\overline{\mathcal{E}_\theta}) \\
&\le \mathbb{P}\left(\overline{\mathcal{A} \text{ returns } i(\theta)}\right) + \mathbb{P}\left(T_{\delta_1} > \frac{n}{4}\right) + \mathbb{P}\left(T_{\delta_1} \le \frac{n}{4}, \tau \ne \infty\right) + \mathbb{P}\left(T_{\delta_1} \le \frac{n}{4}, \tau = \infty, T_{i(\mathsf{p}(\theta))} \ge 1\right)
\end{aligned} \tag{33}$$

Note that $\mathbb{P}\left(\overline{\mathcal{A} \text{ returns } i(\theta)}\right) = \frac{1}{n}\sum_\theta \mathbb{P}_\theta\left(\overline{\mathcal{A} \text{ returns } i(\theta)}\right) \le \delta_1$ by the assumption of the theorem. Also, by Markov's inequality,

$$\mathbb{P}_\theta\left(T_{\delta_1} > \frac{n}{4}\right) \le \frac{\mathbb{E}_\theta[T_{\delta_1}]}{\frac{n}{4}} \le \frac{\frac{n}{16}}{\frac{n}{4}} = \frac{1}{4} \implies \mathbb{P}\left(T_{\delta_1} > \frac{n}{4}\right) \le \frac{1}{4}$$

where the second inequality is by our assumption made for the sake of contradiction. For the last term in (33),

$$\begin{aligned}
\mathbb{P}\left(T_{\delta_1} \le \frac{n}{4}, \tau = \infty, T_{i(\mathsf{p}(\theta))} \ge 1\right) &= \frac{1}{n}\sum_\theta \mathbb{P}_\theta\left(T_{\delta_1} \le \frac{n}{4}, \tau = \infty, T_{i(\mathsf{p}(\theta))} \ge 1\right) \\
&= \frac{1}{n}\sum_\theta \mathbb{P}_\nu\left(T_{\delta_1} \le \frac{n}{4}, \tau = \infty, T_{i(\mathsf{p}(\theta))} \ge 1\right) \text{ for any } \nu \in \Theta \ ,
\end{aligned}$$

where the last equality uses the fact that under $\tau = \infty$ the algorithm's behavior is independent of the unknown $\theta$ because the

[8] For example, a circular orderings for $\{a, b, c\}$ is $a \prec b \prec c \prec a$. Another example is $c \prec b \prec a \prec c$.

algorithm's behavior is determined by its observed reward (and its internal randomization if any). Therefore,

$$\begin{aligned}
\mathbb{P}\left(T_{\delta_1} \le \frac{n}{4}, \tau = \infty, T_{i(\mathsf{p}(\theta))} \ge 1\right) &= \frac{1}{n}\sum_{\theta}\frac{1}{n}\sum_{\nu}\mathbb{P}_\nu\left(T_{\delta_1} \le \frac{n}{4}, \tau = \infty, T_{i(\mathsf{p}(\theta))} \ge 1\right) \\
&\overset{(a)}{=} \frac{1}{n}\sum_{j=1}^{n}\frac{1}{n}\sum_{\nu}\mathbb{P}_\nu\left(T_{\delta_1} \le \frac{n}{4}, \tau = \infty, T_j \ge 1\right) \\
&= \frac{1}{n}\sum_{j=1}^{n}\mathbb{P}\left(T_{\delta_1} \le \frac{n}{4}, \tau = \infty, T_j \ge 1\right) \\
&= \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[\mathbb{1}\left\{T_{\delta_1} \le \frac{n}{4}, \tau = \infty, T_j \ge 1\right\}\right] \\
&\le \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[\mathbb{1}\left\{T_{\delta_1} \le \frac{n}{4}, \tau = \infty, T_j \ge 1\right\}\cdot T_j\right] \\
&= \mathbb{E}\left[\mathbb{1}\left\{T_{\delta_1} \le \frac{n}{4}, \tau = \infty\right\}\cdot \frac{1}{n}\sum_{j=1}^{n} T_j\right] \\
&\le \frac{1}{4}
\end{aligned} \tag{34}$$

where $(a)$ is by reindexing.

It remains to bound the third term in Eq. (33). This case becomes a bit tricky for the following reasons. We would like to use the event $T_{\delta_1} \le n/4$ in the same way as the display (34), but, now that $\tau \ne \infty$, the random variable $T_{\delta_1}$ depends on the instance $\theta$. Thus, we cannot employ the same independence argument used above. To get around, we construct a surrogate algorithm $\mathcal{A}'$ that follows a $\delta_1$-PAC algorithm $\mathcal{A}$ but still selects arms after $\mathcal{A}$ terminates. Specifically, for $t \le T_{\delta_1}$ the algorithm $\mathcal{A}'$ follows the selection $a_t$ of $\mathcal{A}$ exactly, and for $t > T_{\delta_1}$ the algorithm $\mathcal{A}'$ employs an arbitrary policy. For example, one can choose to select arms uniformly at random for $t > T_{\delta_1}$. However, any policy works for our proof as long as the policy does not change as a function of $\theta$.

Let $\chi_j$ be the time step $t \le \frac{n}{4}$ where $i(\theta)$ is pulled for the first time:

$$\chi_j := \begin{cases} \infty & \text{if } \{t \le n/4 : a_t = i(\theta)\} \text{ is empty} \\ \min\{t \le n/4 : a_t = i(\theta)\} & \text{otherwise} \end{cases}.$$

We also define $\tau'$ as the same as (31) except that we replace $T_{\delta_1}$ therein with $n/4$. Then, by introducing the notation $\mathbb{P}^{\mathcal{A}}$ to indicate the dependency on the algorithm $\mathcal{A}$,

$$\begin{aligned}
&\mathbb{P}^{\mathcal{A}}\left(T_{\delta_1} \le \frac{n}{4}, \tau \ne \infty\right) \\
&\le \mathbb{P}^{\mathcal{A}'}\left(\tau' \ne \infty\right) \\
&= \mathbb{P}^{\mathcal{A}'}\left(\tau' \ne \infty,\ \chi_{i(\theta)} \le \tau'\right) + \mathbb{P}^{\mathcal{A}'}\left(\tau' \ne \infty,\ \chi_{i(\theta)} > \tau'\right) \\
&\overset{(a)}{\le} \mathbb{P}^{\mathcal{A}'}\left(a_{\chi_{i(\theta)}} = i(\theta),\ \chi_{i(\theta)} \ne \infty,\ R_{1:(\chi_{i(\theta)}-1)} = 0\right) + \mathbb{P}^{\mathcal{A}'}\left(\tau' \ne \infty,\ a_{\tau'} \ne i(\theta),\ R_{\tau'} = 1\right)
\end{aligned} \tag{35}$$

where $(a)$ introduces the notation $R_{1:t} = 0$ for $R_i = 0, \forall i \in [t]$.

Hereafter, we omit the dependence on $\mathcal{A}'$ for brevity. Denote by $T_j^{(n/4)}$ the number of pulls of arm $j$ up to (and including) time $n/4$. The first term above is equal to

$$\begin{aligned}
&\frac{1}{n}\sum_{\theta}\mathbb{P}_\theta\left(a_{\chi_{i(\theta)}} = i(\theta),\ \chi_{i(\theta)} \ne \infty,\ R_{1:(\chi_{i(\theta)}-1)} = 0\right) \\
&= \frac{1}{n}\sum_{\theta}\mathbb{P}\left(a_{\chi_{i(\theta)}} = i(\theta),\ \chi_{i(\theta)} \ne \infty,\ R_{1:(\chi_{i(\theta)}-1)} = 0\right) \qquad \text{(independence)}
\end{aligned}$$

$$= \frac{1}{n} \sum_{j=1}^{n} \mathbb{P}\left(a_{\chi_j} = j,\ \chi_j \neq \infty,\ R_{1:(\chi_j - 1)} = 0\right) \qquad \text{(reindexing)}$$
$$\leq \frac{1}{n} \sum_{j=1}^{n} \mathbb{P}\left(T_j^{(n/4)} \geq 1\right)$$
$$\leq \frac{1}{n} \sum_{j=1}^{n} \mathbb{E}\left[T_j^{(n/4)}\right] \qquad \text{(Markov's inequality)}$$
$$= \frac{1}{n} \cdot \frac{n}{4} = \frac{1}{4}$$

where the reasoning is mostly the same as the display (34).

The second term of the display (35) is equal to

$$\frac{1}{n} \sum_{\theta} \mathbb{P}_\theta\left(\tau' \neq \infty,\ a_{\tau'} \neq i(\theta),\ R_{\tau'} = 1\right)$$
$$= \frac{1}{n} \sum_{\theta} \mathbb{P}_\theta\left(\exists t \leq n/4 : a_t \neq i(\theta), R_t = 1, a_{1:t-1} \neq i(\theta), R_{1:t-1} = 0\right)$$
$$\leq \frac{1}{n} \sum_{\theta} \sum_{t=1}^{n/4} \mathbb{P}_\theta\left(R_t = 1 \mid a_t \neq i(\theta), a_{1:t-1} \neq i(\theta), R_{1:t-1} = 0\right) \cdot \mathbb{P}_\theta\left(a_t \neq i(\theta), a_{1:t-1} \neq i(\theta), R_{1:t-1} = 0\right)$$
$$\leq \frac{1}{n} \sum_{\theta} \sum_{t=1}^{n/4} \delta_2 \cdot 1$$
$$\leq \delta_2 \cdot \frac{n}{4}\ .$$

**Step 3: Putting it all together**

Using the results of the previous two steps, we have that

$$\frac{1}{n} \sum_{\theta} \mathbb{P}_\theta(\mathcal{A} \text{ returns } i(\mathsf{q}(\theta))) \geq (1 - \delta_2)^{n/4}\, \mathbb{P}\left(\mathcal{E}_\theta\right)$$
$$\geq (1 - \delta_2)^{n/4} \left(1 - \left(\delta_1 + \frac{1}{4} + \frac{1}{4} + \delta_2 \cdot \frac{n}{4}\right)\right)$$
$$\geq \left(1 - \frac{1}{2n}\right)^{n/4} \frac{1}{8} \qquad (\delta_1 < \tfrac{1}{16} < \tfrac{1}{4},\ \delta_2 \leq \tfrac{1}{2n})$$
$$\geq \frac{1}{16} \qquad (n \geq 20)$$
$$> \delta_1 \qquad \text{(assumption)}$$

However, we have by design that $i(\mathsf{q}(\theta)) \neq i(\theta)$. As we have assumed that $\mathcal{A}$ is $\delta_1$-PAC, we have the following contradiction, which concludes the proof:

$$\frac{1}{n} \sum_{\theta} \mathbb{P}_\theta(\mathcal{A} \text{ returns } i(\mathsf{q}(\theta))) \leq \frac{1}{n} \sum_{\theta} \delta_1 = \delta_1\ .$$

□

### E.2.1. COMPARISON TO THEOREM 9 ON THE SAME INSTANCES

The family of $\theta$'s is important to the statement of Theorems 10 and 11. It captures the complexity of *exploration* for logistic bandits and rules out pathological algorithms that have knowledge of $\theta^*$. Otherwise, if we restrict to finding the best arm and fix a set $\mathcal{Z}$ and a single $\theta^* \in \Theta$ as defined in Theorems 10 and 11, then an oracle that knows $i(\theta^*)$ can put all of its samples on it. Indeed, the constraint on $\mathcal{Z}$ and $\theta \in \Theta$ imposed by Theorem 10 implies that $\mu(i(\theta)^T\theta) \geq 1 - \delta_2$ and $\mu(z_j^T\theta) \leq \delta_2$ for

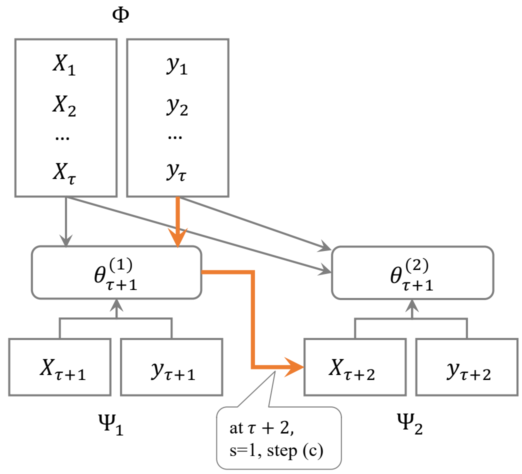


*Figure 2.* A diagram showing the dependency of the variables in SupCB-GLM of Li et al. (2017). The troublesome dependency is colored orange with thick lines. Note that we did not show all the dependencies here to avoid clutter. For example, $X_{\tau+1}$ depends on $X_1, \ldots, X_\tau$.

all $z_j \neq i(\theta)$ and all $\theta \in \Theta$. Hence, for $\theta^* \in \Theta$, the set of alternates to $\theta^*$ for the bound given in Theorem 9 is $\Theta \backslash \{\theta^*\}$. For every $\theta' \in \Theta \backslash \{\theta^*\}$, $KL(\mu(i(\theta^*)^T\theta^*)||\mu(i(\theta^*)^T\theta') \geq KL(1-\delta_2||\delta_2)) = \Omega(1)$. Therefore, the allocation

$$\lambda_a = \begin{cases} 1 & \text{if } a = i(\theta^*) \\ 0 & \text{otherwise} \end{cases}$$

which puts all of the samples on arm $i(\theta^*)$ is feasible for the optimization in Theorem 9. However, this would imply a lower bound of $KL(1-\delta_2||\delta_2)^{-1}\log(1/2.4\delta) = O(\log(1/\delta))$ independent of both the dimension and the number of arms. Naturally, this is pathological since the allocation depends on knowledge of $ß(\theta)$. In order to rule out such pathological instances, we consider a family of $\theta$'s.

## F. Comments on Li et al. (2017)

Li et al. (2017) collects the burn-in samples in a different way from our SupLogistic. They collect burn-in samples (denoted by $\Phi$ here) for the first $\tau = \sqrt{dT}$ rounds, and the buckets $\Psi_1, \ldots, \Psi_S$ are empty at the beginning of time step $\tau + 1$. Then, at time $t > \tau$, when they compute the estimate $\theta_t^{(s)}$, they use both the samples from $\Psi_s$ and $\Phi$. However, we claim that this scheme invalidates the concentration inequality. We explain below how this happens with the help of Figure 2.

- At time $\tau + 1$, we choose $X_{\tau+1}$ in $s = 1$ with step **(a)**.
- At time $\tau + 2$, we pass $s = 1$ and then choose $X_{\tau+2}$ in $s = 2$ with step **(a)**. Note that the set of arms that has survived $s = 1$ and passed onto $s = 2$ are dependent on $\theta_{\tau+1}^{(1)}$ that is a function of $y_1, \ldots, y_\tau$; see the orange thick line in Figure 2.
- At time $\tau + 3$, we pass $s = 1$, arrive at $s = 2$ step **(c)**, and we perform the arm rejections using $\theta_{\tau+2}^{(2)}$. At this point, the fixed design assumption of our concentration inequality for $\theta_{\tau+2}^{(2)}$ is violated as we describe below.

Let $\tau = 1$ for convenience. The estimator $\theta_3^{(2)}$ is computed based on $\{X_1, X_3, y_1, y_3\}$, but $X_3$ depends on $y_1$. That is, $y_1 \mid X_1$ is not conditionally independent from $X_3$. Specifically,

$$p(y_1, y_3 \mid X_1, X_3) = \frac{p(y_1, y_3, X_1, X_3)}{p(X_1, X_3)} = \frac{p(y_3 \mid X_3)p(y_1 \mid X_1, X_3)p(X_1, X_3)}{p(X_1, X_3)} = p(y_3 \mid X_3)p(y_1 \mid X_1, X_3)$$

where we use $p$ to denote both the PDF and PMF. Note that $p(y_1 \mid X_1, X_3) \neq p(y_1 \mid X_1)$ in general; the distribution of $y_1 \mid X_1, X_3$ is algorithm-dependent and thus hard to control. To be clear, note that $p(y_1 \mid X_1, X_3)$ must follow a Bernoulli distribution. It is just that we cannot guarantee that it has the mean $\mu(X_1^\top \theta^*)$.

Our algorithm SupLogistic circumvents this issue by collecting burn-in samples for each bucket $\Psi_1, \ldots, \Psi_S$, but there is another challenge in dealing with the confidence width that depends on $\theta_t^{(s)}$ due to our novel variance-dependent concentration inequality.

## G. Proofs for SupLogistic

While SupLogistic is inspired by the standard SupLinRel-type algorithms (Auer, 2002; Li et al., 2017), its design and analysis are challenging due to the fact that the confidence width scales with the unkonwn $\theta^*$ unlike standard linear bandits. To get around this issue, we design the algorithm so that the mean estimate and the variance estimate are computed from different buckets — $\{\Psi_s(t-1)\}_s$ and $\Phi$ respectively. Such a design was necessary (as far as we stick to the SupLinRel type) because using the confidence widths based on $\theta_t^{(s)}$ would introduce a dependency issue similar to what we described in Section F and invalidate our confidence bound.

While our regret bound improves upon the dependence on $\|\theta^*\|$ in the leading term, we believe it should also be possible to incorporate recent developments of SupLinRel-type algorithms by Li et al. (2019) to shave off the logarithmic factors from $\log^{3/2}(T)$ to $\log^{1/2}(T)$. The focus of our paper is, however, to show the impact of our novel confidence bounds.

We first define our notations for the proof.

- We define a shorthand $X_t := x_{t,a_t}$ for the arm chosen at time step $t$.
- Denote by $\Psi_s(t)$ the set of time steps at which the pulled arm $a_t$ was included to the bucket $s$ up to (and including) time $t$. In other words, $\Psi_s(t)$ is the variable $\Psi_s$ in the pseudocode of SupLogistic at the end of time step $t$.
- Define $H_t^{(s)}(\theta) = \sum_{u\in\Psi_s(t)} \dot\mu(X_u^\top\theta)X_uX_u^\top, \forall s\in[S]$, and $H_t^\Phi(\theta) := \sum_{u\in\Phi(\tau)}\dot\mu(X_u^\top\theta)X_uX_u^\top$. We remark that this definition depends on three variables: bucket index, time step, and the parameter $\theta$ for computing the variance $\dot\mu(X_u^\top\theta)$. Note that the bucket $\Phi$ is never updated after time step $\tau$, so we often use the notation $\Phi$ to mean $\Phi(\tau)$.

We present the proof by a bottom-up approach:

- Lemma 16 sets up the basic lemma concerned with $T_0$, the smallest budget for which we can enjoy a regret guarantee, and the key stochastic events that hold with high probability.
- Lemma 17 analyzes the instantaneous regret.
- Lemma 18 analyzes the cumulative regret per bucket $\Psi_s(T)$.
- Theorem 12 analyzes the final cumulative regret.

Note that the proof of the final regret bound becomes a matter of invoking the Cauchy-Schwarz inequality and the elliptical potential lemma (Lemma 19 below), which is standard in linear bandit analysis.

We first begin with the condition under which the algorithm collects enough burn-in samples and guarantees concentration of measure after time step $\tau$.

**Lemma 16.** *Consider* SupLogistic *with $\delta = 1/T$, $T \ge d$, $\tau = \sqrt{dT}$, and $\alpha = 2.4\sqrt{\log(2(2+K)\cdot\frac{2STK}{\delta})}$. Let $H_t^{(S+1)}(\theta^*) := H_t^\Phi(\theta^*)$. Define the following event:*

$$\mathcal{E}_{\mathsf{mean}} := \left\{ \forall t\in\{\tau+1,\dots,T\}, a\in[K], s\in[S],\ \ |x_{t,a}^\top\theta_{t-1}^{(s)} - x_{t,a}^\top\theta^*| \le \alpha\|x_{t,a}\|_{(H_{t-1}^{(s)}(\theta^*))^{-1}}, \right.$$
$$\left. \frac{1}{\sqrt{2.2}}\|x\|_{(H_{t-1}^{(s)}(\theta^*))^{-1}} \le \|x\|_{(H_{t-1}^{(s)}(\theta_\Phi))^{-1}} \le \sqrt{2.2}\|x\|_{(H_{t-1}^{(s)}(\theta^*))^{-1}} \right\} \tag{36}$$

$$\mathcal{E}_{\mathsf{diversity}} := \left\{ \forall s\in[S+1], \lambda_{\min}(H_\tau^{(s)}(\theta^*)) \ge d\log(6) + \log\left(2(2+K)\cdot\frac{2STK}{\delta}\right)\right\}.$$

*Then, there exists*

$$T_0 = \Theta(Z\log^4(Z)) \quad \text{where} \quad Z = \frac{1}{\sigma_0^4}\left(\frac{1}{\sigma_0^4} + \kappa^{-2}\right)\left(d + \frac{1}{d}\log^2(K)\right) \tag{37}$$

*such that $\forall T\ge T_0$, $\mathbb{P}(\mathcal{E}_{\mathsf{mean}}, \mathcal{E}_{\mathsf{diversity}}) \ge 1-\delta$.*

*Proof.* To avoid clutter, let us fix $s$ and drop the superscript from $H_\tau^{(s)}(\theta^*)$ and use $H_\tau(\theta^*)$. Note that each bucket has at least $\lfloor \frac{\tau}{S+1} \rfloor$ samples. Since $\lambda_{\min}(H_\tau(\theta^*)) \ge \kappa \lambda_{\min}(V_\tau)$ where $V_\tau = \sum_{u=1}^\tau X_u X_u^\top$, to ensure $\mathcal{E}_{\mathsf{diversity}}$, it suffices to show that $\lambda_{\min}(V_\tau) \ge \kappa^{-1}\left(d\log(6) + \log\left(2(2+K)\cdot\frac{2STK}{\delta}\right)\right) =: F$. Recall our stochastic assumption on the context vectors $x_{t,a}$, the definition of $\Sigma$, and our assumption $\lambda_{\min}(\Sigma) \ge \sigma_0^2$. By Li et al. (2017, Proposition 1), there exists $C_1, C_2 > 0$ such that if

$$\left\lfloor \frac{\tau}{S+1} \right\rfloor \ge \left(\frac{C_1\sqrt{d} + C_2\sqrt{\log(2STK/\delta)}}{\sigma_0^2}\right)^2 + \frac{2}{\sigma_0^2}\cdot F\ ,$$

then $\mathbb{P}(\lambda_{\min}(V) \ge F) \ge 1 - \frac{\delta}{2STK}$. Since $\tau = \sqrt{dT}$, we have $T$ in both LHS and RHS. It remains to find the smallest $T$ that satisfies the inequality above. Omitting the dependence on $\sigma_0^2$, one can show that it suffices to find a sufficient condition for $T$ such that

$$T \ge C_3 \underbrace{\frac{1}{\sigma_0^4}\left(\frac{1}{\sigma_0^4} + \kappa^{-2}\right)\left(d + \frac{1}{d}\log^2(K)\right)}_{=:Z} \log^4(T) \tag{38}$$

for some absolute constants $C_3$. One can show that $T < Z\log^4(T)$ implies $T < \Theta(Z\log^4(Z))$, whose contraposition implies that there exists $T_0 = O(Z\log^4(Z))$ such that if $T \ge T_0$, then Eq. (38) is true. Which in turn implies that $\mathbb{P}(\mathcal{E}_{\mathsf{diversity}}) \ge 1 - \frac{\delta}{2STK}\cdot(S+1)$ via union bound over $s \in [S+1]$.

When $\mathcal{E}_{\mathsf{diversity}}$ is true, it is easy to see that the condition on $\nu_t$ in Theorem 5 is satisfied if we substitute $\delta \leftarrow \delta/(2STK)$. Thus, by the union bound,

$$\mathbb{P}(\mathcal{E}_{\mathsf{mean}} \mid \mathcal{E}_{\mathsf{diversity}}) \ge 1 - \frac{\delta}{2STK}\cdot STK\ .$$

Note that $\mathbb{P}(A \cup B) = \mathbb{P}((A\cap\bar{B})\cup B) \le \mathbb{P}(A\cap\bar{B}) + \mathbb{P}(B) \le \mathbb{P}(A \mid \bar{B}) + \mathbb{P}(B)$. Setting $A = \bar{\mathcal{E}}_{\mathsf{mean}}$ and $B = \bar{\mathcal{E}}_{\mathsf{diversity}}$, we have

$$\mathbb{P}(\bar{\mathcal{E}}_{\mathsf{mean}} \cup \bar{\mathcal{E}}_{\mathsf{diversity}}) \le \frac{\delta}{2STK}\cdot STK + \frac{\delta}{2STK}(S+1) \le \delta$$

where the last inequality is by $K \ge 2$. □

**Lemma 17.** *Take $\tau$ and $\alpha$ from Lemma 16. Recall that $M = 1/4$. Suppose $\mathcal{E}_{\mathsf{mean}}$. Consider the time step $t \ge \tau + 1$. Let $s_t$ be the while loop counter $s$ at which the arm $a_t$ is chosen. Let $a_t^* = \arg\max_{a\in[K]} \mu(x_{t,a}^\top\theta^*)$ be the best arm at time $t$. Then, the best arm $a_t^*$ survives through $s_t$, i.e., $a_t^* \in A_s$ for all $s \le s_t$. Furthermore, we have*

$$\mu(x_{t,*}^\top\theta^*) - \mu(x_{t,a_t}^\top\theta^*) \le \begin{cases} \dot\mu(x_{t,a_t}^\top\theta^*)8\cdot 2^{-s_t} + M\cdot 64\cdot 2^{-2s_t} & \text{if } a_t \text{ is selected in step } \textbf{(a)} \\ \dot\mu(x_{t,a_t}^\top\theta^*)2T^{-1/2} + M\cdot 4\cdot T^{-1} & \text{if } a_t \text{ is selected in step } \textbf{(b)} \end{cases}.$$

*Proof.* This proof is adapted from Li et al. (2017, Lemma 6) while keeping the dependence on the variance $\dot\mu(x_{t,a_t}^\top\theta^*)$ to avoid introducing $\kappa^{-1}$ explicitly. Fix $t$. To avoid clutter, let us omit the subscript $t$ from $\{x_{t,a}, a_t^*\}$ and use $\{x_a, a^*\}$ instead, respectively. We also drop the subscript $t-1$ from $H_{t-1}^{(s)}(\theta)$. Let us refer to the iteration index of the while loop as *level*. We use the notation $m_a^{(s)}$ to denote $m_{t,a}$ at level $s$.

We prove the first part of the lemma by induction. For the base case, we trivially have $a^* \in A_1$. Suppose that $a^*$ has survived through the beginning of the $s$-th level (i.e., $a^* \in A_s$) where $s \le s_t - 1$. We want to prove $a^* \in A_{s+1}$. Since the algorithm proceeds to level $s+1$, we know from step **(a)** at $s$-th level that, $\forall a \in A_s$,

$$|m_a^{(s)} - x_a^\top\theta^*| \le \alpha\|x_a\|_{(H^{(s)}(\theta^*))^{-1}} \le \alpha\sqrt{2.2}\|x_a\|_{(H^{(s)}(\theta_\Phi))^{-1}} \le 2^{-s} \tag{39}$$

where both inequalities are due to $\mathcal{E}_{\mathsf{mean}}$. Specifically, it holds for $a = a^*$ because $a^* \in A_s$ by our induction step. Then, the optimality of $a^*$ implies that, $\forall a \in A_s$,

$$m_{a^*}^{(s)} \overset{(39)}{\ge} x_{a^*}^\top\theta^* - 2^{-s} \ge x_a^\top\theta^* - 2^{-s} \overset{(39)}{\ge} m_a^{(s)} - 2\cdot 2^{-s}\ .$$

Thus we have $a^* \in A_{s+1}$ according to step **(c)**.

For the second part of the lemma, suppose $a_t$ is selected at level $s_t$ in step **(a)**. If $s_t = 1$, obviously the lemma holds because

$\mu(z) \in (0,1), \forall z$. If $s_t > 1$, since we have proved $a^* \in A_{s_t}$, Eq. (39) implies that for $a \in \{a_t, a^*\}$,

$$|m_a^{(s_t-1)} - x_a^\top \theta^*| \le 2^{-s_t+1} .$$

Step **(c)** at level $s_t - 1$ implies

$$m_{a^*}^{(s_t-1)} - m_{a_t}^{(s_t-1)} \le 2 \cdot 2^{-s_t+1} .$$

Combining the two inequalities above, we get

$$x_{a_t}^\top \theta^* \ge m_{a_t}^{(s_t-1)} - 2^{-s_t+1} \ge m_{a^*}^{(s_t-1)} - 3 \cdot 2^{-s_t+1} \ge x_{a^*}^\top \theta^* - 4 \cdot 2^{-s_t+1} .$$

Recall that $M = 1/4$ is an upper bound on $\ddot{\mu}(z)$. The inequality above implies that, using Taylor's theorem,

$$\begin{aligned} \mu(x_{a^*}^\top \theta^*) - \mu(x_{a_t}, \theta^*) &= \alpha(x_{a^*}^\top \theta^*, x_{a_t}^\top \theta^*) \cdot (x_{a^*} - x_{a_t})^\top \theta^* \\ &\le \alpha(x_{a^*}^\top \theta^*, x_{a_t}^\top \theta^*) \cdot 4 \cdot 2^{-s_t+1} \\ &\le \left(\dot{\mu}(x_{a_t}^\top \theta^*) + M \cdot (x_{a^*} - x_{a_t})^\top \theta^*\right) \cdot 4 \cdot 2^{-s_t+1} \\ &= \dot{\mu}(x_{a_t}^\top \theta^*) 4 \cdot 2^{-s_t+1} + M \cdot (4 \cdot 2^{-s_t+1})^2 \\ &\le \dot{\mu}(x_{a_t}^\top \theta^*) 8 \cdot 2^{-s_t} + M \cdot 64 \cdot 2^{-2s_t} . \end{aligned} \tag{40}$$

When $a_t$ is selected in step **(b)**, since $m_{a_t}^{(s_t)} \ge m_{a^*}^{(s_t)}$, we have

$$x_{a_t}^\top \theta^* \ge m_{a_t}^{(s_t)} - 1/\sqrt{T} \ge m_{a^*}^{(s_t)} - 1/\sqrt{T} \ge x_{a^*}^\top \theta^* - 2/\sqrt{T} .$$

We now apply a similar reasoning as (40), we have

$$\mu(x_{a^*}^\top \theta^*) - \mu(x_{a_t}^\top \theta^*) \le \dot{\mu}(x_{a^*}^\top \theta^*) 2T^{-1/2} + M \cdot 4 \cdot T^{-1} .$$

□

**Lemma 18** (Regret per bucket)**.** *Assume $\mathcal{E}_{\mathsf{mean}}$ and take $\alpha$ from Lemma 16. Recall that $L = 1/4$. Then, $\forall s \in [S]$,*

$$\sum_{t \in \Psi_s(T) \setminus [\tau]} \mu(x_{t,*}^\top \theta^*) - \mu(X_t^\top \theta^*) \le 18\sqrt{L} \cdot \alpha \sqrt{|\Psi_s(T)| d \log(LT/d)} + \frac{320 M \alpha^2}{\kappa} d \log(LT/d) .$$

*Proof.* By Lemma 17 and the fact that $\mu(z) \in (0,1), \forall z$, we have

$$\begin{aligned} \sum_{t \in \Psi_s(T)} \mu(x_{t,*}^\top \theta^*) - \mu(X_t^\top \theta^*) &\le \sum_{t \in \Psi_s(T)} 1 \wedge \left(\dot{\mu}(X_t^\top \theta^*) \cdot 8 \cdot 2^{-s} + 64 M \cdot 2^{-2s}\right) \\ &\le \left(\sum_{t \in \Psi_s(T)} 1 \wedge \dot{\mu}(X_t^\top \theta^*) \cdot 8 \cdot 2^{-s}\right) + \left(\sum_{t \in \Psi_s(T)} 1 \wedge 64 M \cdot 2^{-2s}\right) \end{aligned}$$

where the last inequality is true by $1 \wedge (a+b) \le 1 \wedge a + 1 \wedge b$. For the first summation, we use $w_{t,a_t}^{(s)} > 2^{-s}$ due to step **(a)** of the algorithm:

$$\begin{aligned} \left(\sum_{t \in \Psi_s(T)} 1 \wedge \dot{\mu}(X_t^\top \theta^*) \cdot 8 \cdot 2^{-s}\right) &\le \sum_{t \in \Psi_s(T)} 1 \wedge \dot{\mu}(X_t^\top \theta^*) \cdot 8 \cdot w_{t,a_t}^{(s)} \\ &= \sum_{t \in \Psi_s(T)} 1 \wedge \dot{\mu}(X_t^\top \theta^*) \cdot 8 \cdot \alpha \sqrt{2.2} \|X_t\|_{(H_{t-1}^{(s)}(\theta_\Phi))^{-1}} && (\text{Def'n of } w_{t,a_t}^{(s)}) \\ &\le \sum_{t \in \Psi_s(T)} 1 \wedge \dot{\mu}(X_t^\top \theta^*) \cdot 18 \cdot \alpha \|X_t\|_{(H_{t-1}^{(s)}(\theta^*))^{-1}} && (\mathcal{E}_{\mathsf{mean}}) \\ &\le \sum_{t \in \Psi_s(T)} 1 \wedge 18 \alpha \sqrt{L} \|\sqrt{\dot{\mu}(X_t^\top \theta^*)} X_t\|_{(H_{t-1}^{(s)}(\theta^*))^{-1}} && (\because \sqrt{\dot{\mu}(X_t^\top \theta^*)} \le \sqrt{L}) \\ &\overset{(a)}{\le} \sqrt{|\Psi_s(T)| \sum_{t \in \Psi_s(T)} 1 \wedge (18 \alpha \sqrt{L})^2 \|\sqrt{\dot{\mu}(X_t^\top \theta^*)} X_t\|^2_{(H_{t-1}^{(s)}(\theta^*))^{-1}}} \end{aligned}$$

$$\overset{(b)}{\leq} 18\alpha\sqrt{L}\sqrt{|\Psi_s(T)| \cdot d\log\left(LT/d\right)}$$

where $(a)$ by the Cauchy-Schwarz inequality and $(b)$ by Lemma 19, with the fact that $(18\alpha\sqrt{L})^2 \geq \frac{1}{2}$, and $\lambda_{\min}(H_\tau(\theta^*)) \geq 1$ ($\because \mathcal{E}_{\mathsf{diversity}}$).

The second summation follows a similar derivation:

$$\begin{aligned}
\sum_{t\in\Psi_s(T)} 1 \wedge 64M \cdot 2^{-2s} &\leq \sum_{t\in\Psi_s(T)} 1 \wedge 64M \cdot \alpha^2 \cdot 2.2\|X_t\|^2_{(H_{t-1}(\theta_\Phi))^{-1}} \\
&\leq \sum_{t\in\Psi_s(T)} 1 \wedge 64M \cdot \alpha^2 \frac{\dot\mu(X_t^\top\theta^*)}{\dot\mu(X_t^\top\theta^*)} 5\|X_t\|^2_{(H_{t-1}(\theta^*))^{-1}} \\
&\leq \sum_{t\in\Psi_s(T)} 1 \wedge \frac{64M\alpha^2}{\kappa} 5 \cdot \left\|\sqrt{\dot\mu(X_t^\top\theta^*)}X_t\right\|^2_{(H_{t-1}(\theta^*))^{-1}} \\
&\leq \frac{320M\alpha^2}{\kappa} d\log(LT/d)
\end{aligned}$$

where $(a)$ is by Lemma 19 and $\frac{320M\alpha^2}{\kappa} \geq 320M^2\alpha^2/L \geq 1/2$, and $\lambda_{\min}(H_\tau(\theta^*)) \geq 1$. □

Finally, we prove the regret bound of SupLogistic below. Note that the statement here is slightly different from that of the main paper because we state the regret bound for large enough $T$ only. This is only an aesthetic difference. Indeed, assume that we have a regret bound $\mathsf{Reg}_T \leq A\sqrt{T} + B$ for $T \geq C$. Then, using $\mathsf{Reg}_T \leq T$, we have $\mathsf{Reg}_T \leq C$ for $T < C$. This implies that, for all $T$, we have $\mathsf{Reg}_T \leq A\sqrt{T} + B + C$.

**Theorem 12** (Regret of SupLogistic)**.** *Consider* SupLogistic *with $\delta$, $\tau$, $\alpha$, and $T_0$ from Lemma 16. Then, if $T \geq T_0$, then*

$$\mathbb{E}[\mathsf{Reg}_T] \leq 10\alpha\sqrt{dT\log(T/d)\log_2(T)} + O\left(\frac{\alpha^2}{\kappa} d \cdot (\log(T/d)) \cdot \log T\right) .$$

*Proof.* Assume $\mathcal{E}_{\mathsf{mean}}$. Recall that $\Psi_0$ contains the time step indices at which the choice $a_t$ happened in step **(b)**. Recall that we set $\tau = \sqrt{dT}$. Let $\Delta_t := \mu(x_{t,*}^\top\theta^*) - \mu(x_{t,a_t}^\top\theta^*)$. Then,

$$\begin{aligned}
R_T &= \sum_{t=1}^{\tau} \Delta_t + \sum_{t=\tau+1}^{T} \Delta_t \\
&\leq \sqrt{dT} + \sum_{t\in\Psi_0(T)} \Delta_t + \sum_{s=1}^{S} \sum_{t\in\Psi_s(T)\setminus[\tau]} \Delta_t .
\end{aligned}$$

For the first term, using Lemma 17,

$$\sum_{t\in\Psi_0(T)} \Delta_t \leq T \cdot \left(\dot\mu(X_t^\top\theta^*) \cdot \frac{2}{\sqrt{T}} + \frac{4M}{T}\right) \leq 2L\sqrt{T} + 4M .$$

For the second term, using Lemma 18, using the Cauchy-Schwarz inequality,

$$\begin{aligned}
\sum_{s=1}^{S} \sum_{t\in\Psi_s(T)\setminus[\tau]} \Delta_t &\leq \sum_{s=1}^{S} \left(18\alpha\sqrt{L|\Psi_s(T)|d\log(LT/d)} + \frac{320M\alpha^2}{\kappa} d\log(LT/d)\right) \\
&\leq 18\alpha\sqrt{Ld\log(LT/d)}\sqrt{S\sum_{s=1}^{S}|\Psi_s(T)| + S \cdot \frac{320M\alpha^2}{\kappa} d\log(LT/d)} \\
&\leq 18\alpha\sqrt{Ld\log(LT/d)} \cdot \sqrt{T\log_2(T)} + \log_2(T) \cdot \frac{320M\alpha^2}{\kappa} d\log(LT/d) .
\end{aligned}$$

Using $L = 1/4$ and $\alpha \geq 7$, the terms involving $\sqrt{T}$ is: $\sqrt{dT} + 2L\sqrt{T} + 18\alpha\sqrt{LdT\log(LT/d)\log_2(T)} \leq 10\alpha\sqrt{dT\log(T/d)\log_2(T)}$. This gives us a regret bound under the event $\mathcal{E}_{\mathsf{mean}}$. One can obtain the expected regret bound by noticing that this event $\mathcal{E}_{\mathsf{mean}}$ does not happen with probability at most $\delta = 1/T$. □

### G.1. Auxiliary Results

**Lemma 19** (Elliptical potential)**.** *Let $s \in [S]$ and $F > 0$. Then,*

$$\sum_{t \in \Psi_s(T)} \min \left\{ 1, F \|\sqrt{\dot{\mu}(X_t^\top \theta^*)} X_t\|^2_{(H^*(\Psi_s(t-1)))^{-1}} \right\} \leq (2F \vee 1) \cdot d \log \left( \frac{L|\Psi_s(T)|}{d\lambda_{\min}(H^*(\Psi_s(\tau)))} \right) .$$

*Proof.* Using Lemma 3 of Jun et al. (2017), we have that $\forall q, x > 0, \min\{q, x\} \leq \max\{2, q\} \log(1 + x)$. Thus,

$$\begin{aligned} \min \left\{ 1, F \|\sqrt{\dot{\mu}(X_t^\top \theta^*)} X_t\|^2_{(H^*(\Psi_s(t)))^{-1}} \right\} &= F \min \left\{ \frac{1}{F}, \|\sqrt{\dot{\mu}(X_t^\top \theta^*)} X_t\|^2_{(H^*(\Psi_s(t)))^{-1}} \right\} \\ &\leq F \cdot \max \left\{ 2, \frac{1}{F} \right\} \log \left( 1 + \|\sqrt{\dot{\mu}(X_t^\top \theta^*)} X_t\|^2_{(H^*(\Psi_s(t)))^{-1}} \right) \end{aligned}$$

where the last inequality is by $\max\{2, 1/F\} = 2$. Then,

$$\begin{aligned} \sum_{t \in \Psi_s(T)} \min \left\{ 1, F \|\sqrt{\dot{\mu}(X_t^\top \theta^*)} X_t\|^2_{(H^*(\Psi_s(t)))^{-1}} \right\} &\leq (2F \vee 1) \sum_{t \in \Psi_s(T)} \log \left( 1 + \|\sqrt{\dot{\mu}(X_t^\top \theta^*)} X_t\|^2_{(H^*(\Psi_s(t)))^{-1}} \right) \\ &= (2F \vee 1) \log \left( \frac{|H^*(\Psi_s(T))|}{|H^*(\Psi_s(\tau))|} \right) \\ &\leq (2F \vee 1) \cdot d \log \left( \frac{L|\Psi_s(T)|}{d \cdot \lambda_{\min}(H^*(\Psi_s(\tau)))} \right) \end{aligned}$$

where the last inequality is by the arithmetic-geometric mean inequality. □

## H. Some Additional Empirical Insights

In this section we dive a bit more into what RAGE-GLM is doing in its process of sampling. We utilized the Zappos pairwise comparison dataset (Yu & Grauman, 2014), focusing on the most "pointy" setting. As in the main text, we used the collected pairwise comparisons to learn $\theta^*$. We then sampled 10,000 shoes randomly from the set of 50,000 to be our $\mathcal{Z}$ set, and an additional randomly chosen 3,000 pairs of pairs of shoes from $\mathcal{Z}$ to be our $\mathcal{X}$ set (after being PCA-ed down to 25 dimensions). Given a query from the $x \in \mathcal{X}$ set we then hallucinate the response using $\mathbb{P}(y = 1) = \mu(x^\top \theta^*)$ according to the logistic model.

RAGE-GLM required roughly $5 \times 10^6$ samples to complete. In figure 3, the top row shows the allocation of RAGE-GLM over $\mathcal{X}$ (where we have sorted the elements of $\mathcal{X}$ by the number of times each element was pulled). As we can see only a few hundred items received any pulls.

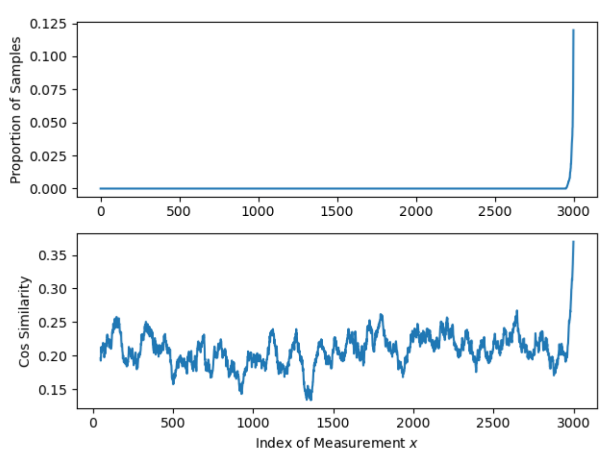


*Figure 3.* The top row is the allocation of samples for each of the 3000 pairs represented in $\mathcal{X}$. The bottom row is the cosine similarity, $\langle x, z_1 - z_2(\|x\|\|z_1 - z_2\|) \rangle$ of each $x \in \mathcal{X}$ with the difference $z_1 - z_2$ between the top two pairs of shoes.

Within a few tens of thousands of samples RAGE-GLM was able to narrow down to the top two different pair of shoes and then spent it's sampling budget differentiating between them. The second row of Figure 3 shows the absolute value of the cosine similarity between the elements of $\mathcal{X}$ and the difference between the top two pairs of shoes (we show a windowed average over 50 items to mitigate some spikes). As we can see, the most sampled items have higher average cosine similarity with the difference between the last best two items. This has an interpretation in terms of *analogies* - the algorithm focuses on two pairs of shoes in $\mathcal{X}$ whose difference is most aligned with that of the best two pairs of shoes $\mathcal{Z}$.


*Figure 4.* In the top row we present the top two shoes $z_1, z_2$. The following rows show the most asked queries to distinguish between the top two shoes along with their proportion in the allocation and their cosine similarity with $z_1 - z_2$.

In Figure 4 we show the top two pairs of shoes along with the 10 pairs that were most queried (corresponding to the allocations above). Even though the pairwise comparisons shown may not be similar to those of the pointed shoes, asking them still gives us information about the difference between the best two, illuminating the advantage of the transductive setting.