Technical Appendix - Linear Streaming Bandit: Regret Minimization and Fixed-Budget Epsilon-Best Arm Identification

Yuming Shao, Zhixuan Fang

April 15, 2025

A Useful Facts

We list some useful facts here. They will be utilized in our proofs.

Lemma 3 (Lemma 10 in Abbasi-yadkori et al. [2011]). *Suppose $X_1, ..., X_t \in \mathbb{R}^d$ and for any $1 \le s \le t$, $||X_s||_2 \le L$. Let $V = \lambda I + \sum_{s=1}^t X_s X_s^T$ for some $\lambda > 0$. Then,*

$$\det(V) \le (\lambda + tL^2/d)^d.$$

Lemma 4 (Abbasi-yadkori et al. [2011]). *Let $\{X_t\}_{t=1}^\infty$ be an $\mathbb{R}^d$-valued stochastic process such that X_t is $\mathcal{F}_{t-1}$ measurable. For any $t \ge 0$, define*

$$V_t = \lambda I + \sum_{s=1}^t X_s X_s^T, \quad \hat{\theta}_t = V_t^{-1} \sum_{s=1}^t r_s X_s,$$

where $r_t = \langle \theta^*, X_t \rangle + \varepsilon_t$ and ε_t is σ-subGaussian. Then for any $\delta > 0$, with probability at least $1 - \delta$, for all $t \ge 0$ and $x \in \mathbb{R}^d$, we have

$$|\langle x, \theta^* - \hat{\theta}_t \rangle| \le ||x||_{V_t^{-1}} \left(\sqrt{\lambda} ||\theta^*||_2 + \sigma \sqrt{2 \ln \left(\frac{\det(V_t)^{1/2} \det(\lambda I)^{-1/2}}{\delta} \right)} \right).$$

Lemma 5 ((20.2) of Chapter 20 in Lattimore and Szepesvári [2020]). *Say $X_1, ..., X_n \in \mathbb{R}^d$ is a sequence of arms. The corresponding reward sequence is $r_1, ..., r_n$ with $r_t = \langle \theta^*, X_t \rangle + \varepsilon_t$, where $\{\varepsilon_t\}_{t \ge 1}$ is independent σ-subGaussian random noise. If $X_1, ..., X_n$ are deterministically chosen without knowing the realizations of $r_1, ..., r_n$, then for any $x \in \mathbb{R}^d$ and $\delta > 0$,*

$$\mathbb{P}\left(\langle \hat{\theta}_t - \theta^*, x \rangle \ge \sqrt{2\sigma^2 ||x||^2_{(\sum_{s=1}^t X_s X_s^T)^{-1}} \log\left(\frac{1}{\delta}\right)} \;\middle|\; X_1, ..., X_n \right) \le \delta,$$

where $\hat{\theta}_t = (\sum_{s=1}^t X_s X_s^T)^{-1} \sum_{s=1}^t r_s X_s$.

B Missing Proofs for Regret Minimization

Proof of Lemma 1. The proof is obvious by combining Lemma 3 and Lemma 4. □

Proof of Lemma 2. In the b-th pass, we let the index of $\hat{a}_b$ and a^* be i and j, respectively. That is, $a_{i,b} = \hat{a}_b$ and $a_{j,b} = a^*$. Let $t_k^{(b)} = \sum_{s=1}^k T_{s,b} + \sum_{b'=1}^{b-1} T_{b'}$.

We suppose $\mu_{\hat{a}_b} < \mu^* - 2\sqrt{\beta_T}\epsilon_b$, which implies $i \ne j$. Since the algorithm records the arm with the highest mean, we have

$$\langle \hat{a}_b, \hat{\theta}_{t_i^{(b)}} \rangle > \langle a^*, \hat{\theta}_{t_j^{(b)}} \rangle. \tag{1}$$

By Line 6 of the algorithm, in the 1-st pass, all K arms are visited, thus (1) obviously holds when $b = 1$. To show that (1) also holds when $b > 1$, we only need to show that arm a^* is visited, i.e., the condition in Line 6 is true when $\tilde{a} = a^*$. We inductively assume that a^* is visited in the 1-st, ..., $(b-1)$-th pass. Say the index of $\hat{a}_{b-1}$ in the $(b-1)$-th pass is i'. We observe that

$$\begin{aligned}
\langle a^*, \hat{\theta}_{t_{j-1}^{(b)}} \rangle \geq & \langle a^*, \theta^* \rangle - \sqrt{\beta_{t_{j-1}^{(b)}}} ||a^*||_{V^{-1}_{t_{j-1}^{(b)}}} \\
\geq & \langle \hat{a}_{b-1}, \theta^* \rangle - \sqrt{\beta_{t_{j-1}^{(b)}}} ||a^*||_{V^{-1}_{t_{j-1}^{(b)}}} \\
\geq & \langle \hat{a}_{b-1}, \hat{\theta}_{t_{i'}^{(b-1)}} \rangle - \sqrt{\beta_{t_{i'}^{(b-1)}}} ||\hat{a}_{b-1}||_{V^{-1}_{t_{i'}^{(b-1)}}} - \sqrt{\beta_{t_{j-1}^{(b)}}} ||a^*||_{V^{-1}_{t_{j-1}^{(b)}}} \\
\geq & \langle \hat{a}_{b-1}, \hat{\theta}_{t_{i'}^{(b-1)}} \rangle - 2\sqrt{\beta_T}\epsilon_{b-1},
\end{aligned}$$

where the last inequality is because $\hat{a}_{b-1}, a^*$ are both visited in the $(b-1)$-th pass. We have shown that the condition in Line 6 is surely true for $\tilde{a} = a^*$, thus arm a^* is visited in the b-th pass and (1) holds for any $b \geq 1$. For the left hand side of (1), we have

$$\langle \hat{a}_b, \hat{\theta}_{t_i^{(b)}} \rangle \leq \langle \hat{a}_b, \theta^* \rangle + \sqrt{\beta_{t_i^{(b)}}} ||\hat{a}_b||_{V^{-1}_{t_i^{(b)}}} < \langle \hat{a}_b, \theta^* \rangle + \sqrt{\beta_T}\epsilon_b < \langle a^*, \theta^* \rangle - \sqrt{\beta_T}\epsilon_b.$$

For the right hand side of (1),

$$\langle a^*, \hat{\theta}_{t_j^{(b)}} \rangle \geq \langle a^*, \theta^* \rangle - \sqrt{\beta_{t_j^{(b)}}} ||a^*||_{V^{-1}_{t_j^{(b)}}} > \langle a^*, \theta^* \rangle - \sqrt{\beta_T}\epsilon_b.$$

We observe a contradiction:

$$\langle a^*, \theta^* \rangle - \sqrt{\beta_T}\epsilon_b < \langle a^*, \hat{\theta}_{t_j^{(b)}} \rangle < \langle \hat{a}_b, \hat{\theta}_{t_i^{(b)}} \rangle < \langle a^*, \theta^* \rangle - \sqrt{\beta_T}\epsilon_b,$$

thus we have proved this lemma. □

C Missing Proofs for Best Arm Identification

C.1 Proofs for G-MP-SE algorithm

Proof of Theorem 2. The algorithm approximates the set of true arms $\mathcal{K}$ using a grid set $\mathcal{K}_1(\epsilon) = \{g \in \mathcal{G}_\epsilon \mid C_1(g) > 0\}$. To upper bound the error probability of the algorithm, it is vital to show that the optimal arm in the grid set $a^*_\epsilon := \arg\max_{a \in \mathcal{K}_1(\epsilon)} \langle a, \theta^* \rangle$ survives $B-2$ rounds of successive elimination with high probability.

Step I.

We first show that in pass b, if a^*_ϵ is still not eliminated, i.e., $C_{b-1}(a^*_\epsilon) > 0$, then the probability $\mathbb{P}(\langle \hat{\theta}_b, g \rangle > \langle \hat{\theta}_b, a^*_\epsilon \rangle \mid C_{b-1}(a^*_\epsilon) > 0)$ for some $g \in \mathcal{K}_{b-1}(\epsilon), g \neq a^*_\epsilon$, is upper-bounded using a Chernoff-type bound:

$$\begin{aligned}
& \mathbb{P}\big(\langle \hat{\theta}_b, g \rangle > \langle \hat{\theta}_b, a^*_\epsilon \rangle \bigm| C_{b-1}(a^*_\epsilon) > 0\big) \\
=& \mathbb{P}\big(\langle \hat{\theta}_b - \theta^*, g \rangle > \langle \hat{\theta}_b - \theta^*, a^*_\epsilon \rangle - \langle \theta^*, g \rangle + \langle \theta^*, a^*_\epsilon \rangle \bigm| C_{b-1}(a^*_\epsilon) > 0\big) \\
=& \mathbb{P}\big(\langle \hat{\theta}_b - \theta^*, g - a^*_\epsilon \rangle > \langle \theta^*, a^*_\epsilon - g \rangle \bigm| C_{b-1}(a^*_\epsilon) > 0\big) \\
=& \mathbb{P}\big(\langle \hat{\theta}_b - \theta^*, a^*_\epsilon - g \rangle < -\Delta_g + \Delta_{a^*_\epsilon} \bigm| C_{b-1}(a^*_\epsilon) > 0\big) \\
=& \sum_{\mathcal{A}_b} \mathbb{P}\big(\langle \hat{\theta}_b - \theta^*, a^*_\epsilon - g \rangle < -\Delta_g + \Delta_{a^*_\epsilon}, \mathcal{A}_b \bigm| C_{b-1}(a^*_\epsilon) > 0\big) \\
=& \sum_{\mathcal{A}_b} \mathbb{P}\big(\mathcal{A}_b \bigm| C_{b-1}(a^*_\epsilon) > 0\big) \mathbb{P}\big(\langle \hat{\theta}_b - \theta^*, a^*_\epsilon - g \rangle < -\Delta_g + \Delta_{a^*_\epsilon} \bigm| \mathcal{A}_b, C_{b-1}(a^*_\epsilon) > 0\big) \\
=& \sum_{\mathcal{A}_b} \mathbb{P}\big(\mathcal{A}_b \bigm| C_{b-1}(a^*_\epsilon) > 0\big) \mathbb{P}\big(\langle \hat{\theta}_b - \theta^*, a^*_\epsilon - g \rangle < -\Delta_g + \Delta_{a^*_\epsilon} \bigm| \mathcal{A}_b\big) \\
\leq& \sum_{\mathcal{A}_b} \mathbb{P}\big(\mathcal{A}_b \bigm| C_{b-1}(a^*_\epsilon) > 0\big) \exp\Big(-\frac{(\Delta_g - \Delta_{a^*_\epsilon})^2}{2\sigma^2 ||a^*_\epsilon - g||^2_{V_b^{-1}}}\Big),
\end{aligned}$$

where $\mathcal{A}_b$ denotes the actual sequence of arms pulled in pass b. The last equality is due to the fact that the distribution of $\hat{\theta}_b$ is independent of $C_{b-1}(a^*_\epsilon)$ conditioned on the realization of $\mathcal{A}_b$. The inequality is by Lemma 5. We first notice that $g(a^*) \in \mathcal{K}_1(\epsilon)$. Thus by the definition of a^*_ϵ, $\langle a^*_\epsilon, \theta^* \rangle \geq \langle g(a^*), \theta^* \rangle$, which further implies

$$\begin{aligned}\Delta_{a^*_\epsilon} = \langle \theta^*, a^* - a^*_\epsilon \rangle \leq & \langle \theta^*, a^* - g(a^*) \rangle \\ \leq & ||\theta^*||_2 ||a^* - g(a^*)||_2 \\ \leq & ||\theta^*||_2 \epsilon.\end{aligned}$$

Then we consider the term $||a^*_\epsilon - g||^2_{V_b^{-1}}, \forall \mathcal{A}_b$:

$$\begin{aligned}& ||a^*_\epsilon - g||^2_{V_b^{-1}} \\ \leq & \max_{g', g'' \in \mathcal{K}_{b-1}(\epsilon)} ||g' - g''||^2_{V_b^{-1}} \\ = & \max_{g', g'' \in \mathcal{K}_{b-1}(\epsilon)} ||g' - g''||^2_{\left(\sum_{t \text{ in pass } b} a_t a_t^T\right)^{-1}}.\end{aligned}$$

Say the sample budget is allocated with respect to a design $\lambda^{(b)}$, we have

$$\begin{aligned}& \sum_{t \text{ in pass } b} a_t a_t^T \\ = & \sum_{g \in \mathcal{G}_\epsilon} \sum_{t \text{ in pass } b: g(a_t) = g} a_t a_t^T \\ \succeq & \sum_{g \in \mathcal{G}_\epsilon} \sum_{t \text{ in pass } b: g(a_t) = g} \frac{\lambda_g^{(b)} \tilde{T}}{T_b(g)} a_t a_t^T \\ = & \sum_{g \in \mathcal{G}_\epsilon} \sum_{t \text{ in pass } b: g(a_t) = g} \frac{\lambda_g^{(b)} \tilde{T}}{T_b(g)} \big(a_t - g(a_t) + g(a_t)\big)\big(a_t - g(a_t) + g(a_t)\big)^T \\ = & \sum_{g \in \mathcal{G}_\epsilon} \sum_{t \text{ in pass } b: g(a_t) = g} \frac{\lambda_g^{(b)} \tilde{T}}{T_b(g)} \left[gg^T + g\big(a_t - g\big)^T + \big(a_t - g\big)g^T + \big(a_t - g\big)\big(a_t - g\big)^T\right] \\ \succeq & \sum_{g \in \mathcal{G}_\epsilon} \sum_{t \text{ in pass } b: g(a_t) = g} \frac{\lambda_g^{(b)} \tilde{T}}{T_b(g)} \left[gg^T + g\big(a_t - g\big)^T + \big(a_t - g\big)g^T\right],\end{aligned}$$

where $T_b(g) := \sum_{t \text{ in pass } b} \mathbb{I}(g(a_t) = g)$ and obviously $T_b = \sum_{g \in \mathcal{G}_\epsilon} T_b(g)$. We attempt to show that for any t in pass b, $g(a_t)\big[a_t - g(a_t)\big]^T$, $\big[a_t - g(a_t)\big]g(a_t)^T \succeq -\epsilon(L+\epsilon)I$. We only consider $M_t := g(a_t)\big[a_t - g(a_t)\big]^T$ since the derivation is almost the same for $\big[a_t - g(a_t)\big]g(a_t)^T$. For each $a \in \mathbb{R}^d$, $a^T M_t a \geq \lambda_{\min}(M_t) a^T a = a^T(\lambda_{\min}(M_t) I)a$. Since $\text{rank}(M_t) \leq 1$, at least $d-1$ eigenvalues of M_t are 0. There are two possibilities overall. If $\lambda_{\min}(M_t) \geq 0$, then M_t is positive semi-definite, thus our argument above trivially holds. If $\lambda_{\min}(M_t) < 0$, then we have that

$$\begin{aligned}\lambda_{\min}(M_t) = \text{tr}\Big(g(a_t)\big[a_t - g(a_t)\big]^T\Big) = & \langle g(a_t), a_t - g(a_t) \rangle \\ \geq & - ||g(a_t)||_2 ||a_t - g(a_t)||_2 \\ \geq & - \epsilon(L+\epsilon).\end{aligned}$$

As a result, we have $M_t \succeq -\epsilon(L+\epsilon)I$ and furthermore

$$\begin{aligned}& ||a^*_\epsilon - g||^2_{V_b^{-1}} \\ \leq & \max_{g', g'' \in \mathcal{K}_{b-1}(\epsilon)} ||g' - g''||^2_{\left(\sum_{g \in \mathcal{G}_\epsilon} \sum_{t \text{ in pass } b: g(a_t) = g} \frac{\lambda_g^{(b)} \tilde{T}}{T_b(g)} (gg^T - 2\epsilon(L+\epsilon)I)\right)^{-1}} \\ = & \max_{g', g'' \in \mathcal{K}_{b-1}(\epsilon)} ||g' - g''||^2_{\left(\sum_{g \in \mathcal{G}_\epsilon} \lambda_g^{(b)} \tilde{T}(gg^T - 2\epsilon(L+\epsilon)I)\right)^{-1}} \\ = & \max_{g', g'' \in \mathcal{K}_{b-1}(\epsilon)} ||g' - g''||^2_{\left(-2\epsilon(L+\epsilon)\tilde{T}I + \sum_{g \in \mathcal{G}_\epsilon} \lambda_g^{(b)} \tilde{T} gg^T\right)^{-1}} \\ = & \frac{1}{\tilde{T}} \max_{g', g'' \in \mathcal{K}_{b-1}(\epsilon)} ||g' - g''||^2_{\left(-2\epsilon(L+\epsilon)I + \sum_{g \in \mathcal{G}_\epsilon} \lambda_g^{(b)} gg^T\right)^{-1}}.\end{aligned}$$

Define

$$R^2(\lambda, \mathcal{S}, \mathcal{G}_\epsilon) := \max_{g', g'' \in \mathcal{S}} ||g' - g''||^2_{\left(-2\epsilon(L+\epsilon)I + \sum_{g \in \mathcal{G}_\epsilon} \lambda_g g g^T\right)^{-1}},$$

thus we have that $||a^*_\epsilon - g||^2_{V_b^{-1}} \le \frac{R^2(\lambda^{(b)}, \mathcal{K}_{b-1}(\epsilon), \mathcal{G}_\epsilon)}{\tilde{T}}, \forall \mathcal{A}_b$. As a consequence, we have the upper bound

$$\begin{aligned}
&\mathbb{P}\big(\langle \hat{\theta}_b, g\rangle > \langle \hat{\theta}_b, a^*_\epsilon\rangle \;\big|\; C_{b-1}(a^*_\epsilon) > 0\big) \\
\le &\sum_{\mathcal{A}_b} \mathbb{P}\big(\mathcal{A}_b \;\big|\; C_{b-1}(a^*_\epsilon) > 0\big) \exp\Big(-\frac{(\Delta_g - \Delta_{a^*_\epsilon})^2}{2\sigma^2 ||a^*_\epsilon - g||^2_{V_b^{-1}}}\Big) \\
\le &\sum_{\mathcal{A}_b} \mathbb{P}\big(\mathcal{A}_b \;\big|\; C_{b-1}(a^*_\epsilon) > 0\big) \exp\Big(-\frac{\tilde{T}(\Delta_g - \Delta_{a^*_\epsilon})^2}{2\sigma^2 R^2(\lambda^{(b)}, \mathcal{K}_{b-1}(\epsilon), \mathcal{G}_\epsilon)}\Big) \\
= &\exp\Big(-\frac{\tilde{T}(\Delta_g - \Delta_{a^*_\epsilon})^2}{2\sigma^2 R^2(\lambda^{(b)}, \mathcal{K}_{b-1}(\epsilon), \mathcal{G}_\epsilon)}\Big).
\end{aligned}$$

Step II.

Next, we show that if $C_{b-1}(a^*_\epsilon) > 0$, then the probability that arm a^*_ϵ is completely eliminated in pass b is also upper bounded. Recall that the number of active arms at the end of pass b is set to be $|\mathcal{K}_b| = \lceil K^{\frac{B-b-1}{B-2}} \rceil$. We have that

$$\begin{aligned}
&\mathbb{P}\Big(C_b(a^*_\epsilon) = 0 \;\Big|\; C_{b-1}(a^*_\epsilon) > 0\Big) \\
\le &\mathbb{P}\Big(\exists \text{ at least } |\mathcal{K}_b| \text{ arms } a : \langle g(a), \hat{\theta}_b\rangle > \langle a^*_\epsilon, \hat{\theta}_b\rangle \;\Big|\; C_{b-1}(a^*_\epsilon) > 0\Big) \\
\le &\min_{1 < i_b \le |\mathcal{K}_b|} \mathbb{P}\Big(\exists \text{ at least } |\mathcal{K}_b| - (i_b - 1) \text{ arms } a \in \mathcal{B}_{b-1}(i_b) : \langle g(a), \hat{\theta}_b\rangle > \langle a^*_\epsilon, \hat{\theta}_b\rangle \;\Big|\; C_{b-1}(a^*_\epsilon) > 0\Big),
\end{aligned}$$

where $\mathcal{B}_b(i) := \{a \in \mathcal{K}_b : \Delta_a > \Delta_{[i]}\}$. The first inequality is by the computation rule for $\boldsymbol{C}_b$ in the algorithm, while the second is by the observation that even if the set (of cardinality $i_b - 1$) containing the arms with relatively low reward gaps contributes $i_b - 1$ arms a satisfying $\langle g(a), \hat{\theta}_b\rangle > \langle a^*_\epsilon, \hat{\theta}_b\rangle$, $|\mathcal{K}_b| - (i_b - 1)$ more arms with relatively high reward gaps must also satisfy the inequality. Thus

$$\begin{aligned}
&\mathbb{P}\Big(C_b(a^*_\epsilon) = 0 \;\Big|\; C_{b-1}(a^*_\epsilon) > 0\Big) \\
\le &\min_{1 < i_b \le |\mathcal{K}_b|} \mathbb{P}\Big(\sum_{a \in \mathcal{B}_{b-1}(i_b)} \mathbb{I}\big(\langle g(a), \hat{\theta}_b\rangle > \langle a^*_\epsilon, \hat{\theta}_b\rangle\big) \ge |\mathcal{K}_b| - (i_b - 1) \;\Big|\; C_{b-1}(a^*_\epsilon) > 0\Big) \\
\le &\min_{1 < i_b \le |\mathcal{K}_b|} \frac{1}{|\mathcal{K}_b| - (i_b - 1)} \mathbb{E}\Big[\sum_{a \in \mathcal{B}_{b-1}(i_b)} \mathbb{I}\big(\langle g(a), \hat{\theta}_b\rangle > \langle a^*_\epsilon, \hat{\theta}_b\rangle\big) \;\Big|\; C_{b-1}(a^*_\epsilon) > 0\Big] \\
= &\min_{1 < i_b \le |\mathcal{K}_b|} \frac{1}{|\mathcal{K}_b| - (i_b - 1)} \sum_{a \in \mathcal{B}_{b-1}(i_b)} \mathbb{P}\big(\langle g(a), \hat{\theta}_b\rangle > \langle a^*_\epsilon, \hat{\theta}_b\rangle \;\big|\; C_{b-1}(a^*_\epsilon) > 0\big) \\
\le &\min_{1 < i_b \le |\mathcal{K}_b|} \frac{1}{|\mathcal{K}_b| - (i_b - 1)} \sum_{a \in \mathcal{B}_{b-1}(i_b)} \exp\Big(-\frac{\tilde{T}(\Delta_{g(a)} - \Delta_{a^*_\epsilon})^2}{2\sigma^2 R^2(\lambda^{(b)}, \mathcal{K}_{b-1}(\epsilon), \mathcal{G}_\epsilon)}\Big) \\
\le &\min_{1 < i_b \le |\mathcal{K}_b|} \frac{1}{|\mathcal{K}_b| - (i_b - 1)} \sum_{a \in \mathcal{B}_{b-1}(i_b)} \exp\Big(-\frac{\tilde{T}(\Delta_{[i_b]} - 2||\theta^*||_2\epsilon)^2}{2\sigma^2 R^2(\lambda^{(b)}, \mathcal{K}_{b-1}(\epsilon), \mathcal{G}_\epsilon)}\Big) \\
= &\min_{1 < i_b \le |\mathcal{K}_b|} \frac{|\mathcal{K}_{b-1}| - i_b}{|\mathcal{K}_b| - (i_b - 1)} \exp\Big(-\frac{\tilde{T}(\Delta_{[i_b]} - 2||\theta^*||_2\epsilon)^2}{2\sigma^2 R^2(\lambda^{(b)}, \mathcal{K}_{b-1}(\epsilon), \mathcal{G}_\epsilon)}\Big).
\end{aligned}$$

We give an explanation for the derivation above. The second inequality is by adopting Markov's inequality. The third inequality uses our results in Step I. As for the last inequality, note that for any $a \in \mathcal{B}_{b-1}(i_b)$, $\Delta_a \ge \Delta_{[i_b]}$ and that

$$\Delta_{g(a)} - \Delta_a = \langle \theta^*, a - g(a)\rangle \ge -||\theta^*||_2 ||a - g(a)||_2 \ge -||\theta^*||\epsilon.$$

By selecting $i_b = \lfloor |\mathcal{K}_b|/2 \rfloor + 1$, we obtain that

$$\begin{aligned}
\frac{|\mathcal{K}_{b-1}| - i_b}{|\mathcal{K}_b| - (i_b - 1)} =& \frac{|\mathcal{K}_{b-1}| - \lfloor |\mathcal{K}_b|/2 \rfloor - 1}{|\mathcal{K}_b| - (\lfloor |\mathcal{K}_b|/2 \rfloor + 1 - 1)} \\
\leq& \frac{|\mathcal{K}_{b-1}| - \lfloor |\mathcal{K}_b|/2 \rfloor - 1}{|\mathcal{K}_b| - |\mathcal{K}_b|/2} \\
\leq& \frac{|\mathcal{K}_{b-1}| - |\mathcal{K}_b|/2}{|\mathcal{K}_b| - |\mathcal{K}_b|/2} = 2\frac{|\mathcal{K}_{b-1}|}{|\mathcal{K}_b|} - 1 \\
=& 2\frac{\lceil K^{\frac{B-(b-1)-1}{B-2}} \rceil}{\lceil K^{\frac{B-b-1}{B-2}} \rceil} - 1 \leq 2\frac{K^{\frac{B-(b-1)-1}{B-2}} + 1}{K^{\frac{B-b-1}{B-2}}} - 1 \\
\leq& 2K^{\frac{B-(b-1)-1}{B-2} - \frac{B-b-1}{B-2}} + 1 = 2K^{\frac{1}{B-2}} + 1.
\end{aligned}$$

As a consequence,

$$\mathbb{P}\Big(C_b(a^*_\epsilon) = 0 \;\Big|\; C_{b-1}(a^*_\epsilon) > 0\Big) \leq \big(2K^{\frac{1}{B-2}} + 1\big) \exp\Big(-\frac{\tilde{T}(\Delta_{[\lfloor \lceil K^{\frac{B-b-1}{B-2}} \rceil /2 \rfloor + 1]} - 2||\theta^*||_2\epsilon)^2}{2\sigma^2 R^2(\lambda^{(b)}, \mathcal{K}_{b-1}(\epsilon), \mathcal{G}_\epsilon)}\Big).$$

Step III.

Finally, we control the error probability of the proposed algorithm. Recall that in the last pass, the algorithm chooses an arm a whose corresponding grid point $g(a)$ is not fully eliminated. Thus if $C_{B-1}(a^*_\epsilon) > 0$, we have that $||\hat{a} - a^*_\epsilon||_2 \leq \epsilon$, which further implies $||\hat{a} - a^*||_2 \leq 2\epsilon$. Noting that the values of $\boldsymbol{C}_b$ are non-increasing as b grows, we have

$$\begin{aligned}
&\mathbb{P}\big(\Delta_{\hat{a}} > 2||\theta^*||_2\epsilon\big) \leq \mathbb{P}\big(C_{B-1}(a^*_\epsilon) = 0\big) \\
=& \sum_{b=1}^{B-1} \mathbb{P}\big(C_b(a^*_\epsilon) = 0, C_{b-1}(a^*_\epsilon) > 0, C_{b-2}(a^*_\epsilon) > 0, ...\big) \\
\leq& \sum_{b=2}^{B-1} \mathbb{P}\big(C_b(a^*_\epsilon) = 0, C_{b-1}(a^*_\epsilon) > 0\big) \\
=& \sum_{b=2}^{B-1} \mathbb{P}\big(C_b(a^*_\epsilon) = 0 \;\big|\; C_{b-1}(a^*_\epsilon) > 0\big)\mathbb{P}\big(C_{b-1}(a^*_\epsilon) > 0\big) \\
\leq& \sum_{b=2}^{B-1} \mathbb{P}\big(C_b(a^*_\epsilon) = 0 \;\big|\; C_{b-1}(a^*_\epsilon) > 0\big) \\
\leq& \sum_{b=2}^{B-1} \big(2K^{\frac{1}{B-2}} + 1\big) \exp\Big(-\frac{\tilde{T}(\Delta - 2||\theta^*||_2\epsilon)^2}{2\sigma^2 R^2(\lambda^{(b)}, \mathcal{K}_{b-1}(\epsilon), \mathcal{G}_\epsilon)}\Big) \\
\leq& (B-2)\big(2K^{\frac{1}{B-2}} + 1\big) \exp\Big(-\frac{\tilde{T}(\Delta - 2||\theta^*||_2\epsilon)^2}{2\sigma^2 \max_{1<b<B} R^2(\lambda^{(b)}, \mathcal{K}_{b-1}(\epsilon), \mathcal{G}_\epsilon)}\Big).
\end{aligned}$$

In the derivation above, the third inequality is by our results in Step II and a simple fact that $\Delta_{[\lfloor \lceil K^{\frac{B-b-1}{B-2}} \rceil /2 \rfloor + 1]} \geq \Delta$. By Kiefer–Wolfowitz Theorem (see for example Theorem 21.1 in Lattimore and Szepesvári [2020]), for any $S \subseteq \mathcal{K}$,

$$\min_{\lambda \in \mathcal{P}(S)} \max_{g' \in S} ||g'||^2_{\left(\sum_{g \in S} \lambda_g g g^T\right)^{-1}} \leq d.$$

Thus by taking $\epsilon \downarrow 0$, we obtain that

$$\lim_{\epsilon \downarrow 0} \mathbb{P}\big(\Delta_{\hat{a}} > 2||\theta^*||_2\epsilon\big) \leq (B-2)\big(2K^{\frac{1}{B-2}} + 1\big) \exp\Big(-\frac{\Delta^2 \tilde{T}}{4\sigma^2 d}\Big). \tag{2}$$

As (2) indicates, the larger the value of $\tilde{T}$, the lower the algorithm's prediction error. However, a strict budget constraint T must be satisfied by the algorithm. Here, we will show that if the sample budget parameter $\tilde{T}$ is selected to be $\frac{T}{B-2} - \frac{d(d+1)}{2}$, the budget constraint T will not be violated. Formally, the parameter $\tilde{T}$ must satisfy that

$$\sum_{b=2}^{B-1} \sum_{g \in \mathcal{G}_\epsilon} \lceil \lambda_g^{(b)} \tilde{T} \rceil \leq T.$$

Note that by Kiefer–Wolfowitz Theorem, there exists a design λ_{b-1}^* whose support is upper bounded by $\frac{d(d+1)}{2}$, $\forall b = 2, ..., B-1$. Given this fact, we have

$$\begin{aligned}
&\sum_{b=2}^{B-1} \sum_{g \in \mathcal{G}_\epsilon} \lceil \lambda_g^{(b)} \tilde{T} \rceil \leq \sum_{b=2}^{B-1} \sum_{g \in \mathcal{G}_\epsilon : \lambda_g^{(b)} > 0} (\lambda_g^{(b)} \tilde{T} + 1) \\
&= \sum_{b=2}^{B-1} \text{supp}(\lambda^{(b)}) \quad + \sum_{b=2}^{B-1} \sum_{g \in \mathcal{G}_\epsilon : \lambda_g^{(b)} > 0} \lambda_g^{(b)} \tilde{T} \\
&\leq (B-2) \frac{d(d+1)}{2} + (B-2) \Big(\frac{T}{B-2} - \frac{d(d+1)}{2} \Big) \\
&= T.
\end{aligned}$$

Thus, our selection of design $\lambda^{(b)}$ will never cause a violation of the sample budget constraint. Finally, we conclude that

$$\lim_{\epsilon \downarrow 0} \mathbb{P}\big(\Delta_{\hat{a}} > 2||\theta^*||_2 \epsilon\big) \leq (B-2)\big(2K^{\frac{1}{B-2}} + 1\big) \exp\Big(- \frac{\Delta^2}{4\sigma^2 d} \big(\frac{T}{B-2} - \frac{d(d+1)}{2} \big) \Big).$$

□

C.2 Discussion on the choice of design

Recall that our choice of design $\lambda^{(b)}$ is λ_{b-1}^* in the theoretical analysis of the G-MP-SE algorithm, where

$$\begin{aligned}
&\lambda_b^* := \arg \min_{\lambda \in \mathcal{P}(\mathcal{K}_b(\epsilon))} \max_{g' \in \mathcal{K}_b(\epsilon)} ||g'||^2_{\left(\sum_{g \in \mathcal{K}_b(\epsilon)} \lambda_g g g^T\right)^{-1}}, \\
&\mathcal{K}_b(\epsilon) := \{ g \in \mathcal{G}_\epsilon \mid C_b(g) > 0 \}.
\end{aligned}$$

We also notice that in the proof of Theorem 2, the key to obtaining an efficient prediction error upper bound is to choose a design that minimizes $\max_{1<b<B} R^2(\lambda^{(b)}, \mathcal{K}_{b-1}(\epsilon), \mathcal{G}_\epsilon)$, where

$$R^2(\lambda, \mathcal{S}, \mathcal{G}_\epsilon) = \max_{g', g'' \in \mathcal{S}} ||g' - g''||^2_{\left(-2\epsilon(L+\epsilon)I + \sum_{g \in \mathcal{G}_\epsilon} \lambda_g g g^T\right)^{-1}}.$$

Thus, a seemingly more intuitive choice of design λ is

$$\lambda^{(b)} \leftarrow \arg \min_{\lambda \in \mathcal{L}_\epsilon} R^2(\lambda, \mathcal{K}_{b-1}, \mathcal{G}_\epsilon), \quad \forall b = 2, ..., B-1,$$

where we define

$$\mathcal{L}_\epsilon = \Big\{ \lambda' \in \mathcal{P}(\mathcal{G}_\epsilon) \, \Big| \, \sum_{g \in \mathcal{G}_\epsilon} \lambda'_g g g^T \succeq 2\epsilon(L+\epsilon) I \Big\}.$$

We can further show that the computation of $\lambda^{(b)}$ is tractable, as it suffices to solve a convex optimization problem using Frank-Wolfe-type algorithms (Jaggi [2013]).

Proposition 1. *The following optimization problem is a convex optimization for* $\forall \mathcal{S} \subseteq \mathcal{G}_\epsilon$*:*

$$\min_{\lambda \in \mathcal{L}_\epsilon} R^2(\lambda, \mathcal{S}, \mathcal{G}_\epsilon).$$

Proof of Proposition 1. We first show that the set $\mathcal{L}_\epsilon$ is convex. Consider any $\lambda, \lambda' \in \mathcal{P}(\mathcal{G}_\epsilon)$ such that $\sum_{g\in\mathcal{G}_\epsilon} \lambda_g gg^T, \sum_{g\in\mathcal{G}_\epsilon} \lambda'_g gg^T \succeq 2\epsilon(L+\epsilon)I$. Then for any $\tau \in (0,1)$, the design $\tau\lambda + (1-\tau)\lambda'$ satisfies that

$$\begin{aligned}
&\sum_{g\in\mathcal{G}_\epsilon} \big(\tau\lambda_g + (1-\tau)\lambda'_g\big)gg^T \\
=&\tau \sum_{g\in\mathcal{G}_\epsilon} \lambda_g gg^T + (1-\tau)\sum_{g\in\mathcal{G}_\epsilon} \lambda'_g gg^T \\
\succeq&2\tau\epsilon(L+\epsilon)I + 2(1-\tau)\epsilon(L+\epsilon)I \\
=&2\epsilon(L+\epsilon)I,
\end{aligned}$$

thus we also have that $\tau\lambda + (1-\tau)\lambda' \in \mathcal{L}_\epsilon$. Furthermore, we still consider any $\lambda, \lambda' \in \mathcal{L}_\epsilon$. Then for any $\tau \in (0,1)$, we have that for any $g', g'' \in \mathcal{S}$,

$$\begin{aligned}
&||g'-g''||^2_{\left(-2\epsilon(L+\epsilon)I+\sum_{g\in\mathcal{G}_\epsilon}(\tau\lambda_g+(1-\tau)\lambda'_g)gg^T\right)^{-1}} \\
=&(g'-g'')^T\Big(-2\epsilon(L+\epsilon)I + \sum_{g\in\mathcal{G}_\epsilon}(\tau\lambda_g + (1-\tau)\lambda'_g)gg^T\Big)^{-1}(g'-g'') \\
=&(g'-g'')^T\Big(\tau\big(-2\epsilon(L+\epsilon)I + \sum_{g\in\mathcal{G}_\epsilon}\lambda_g gg^T\big) \\
&+ (1-\tau)\big(-2\epsilon(L+\epsilon)I + \sum_{g\in\mathcal{G}_\epsilon}\lambda'_g gg^T\big)\Big)^{-1}(g'-g'') \\
=&(g'-g'')^T\Big(\tau P^T I P + (1-\tau)P^T\Lambda' P\Big)^{-1}(g'-g'') \\
=&(g'-g'')^T P^{-1}\Big(\tau I + (1-\tau)\Lambda'\Big)^{-1}(P^T)^{-1}(g'-g''),
\end{aligned}$$

where $P \in \mathbb{R}^{d\times d}$ is a invertible and symmetric matrix satisfying $P^TP = -2\epsilon(L+\epsilon)I + \sum_{g\in\mathcal{G}_\epsilon}\lambda_g gg^T$, $P^T\Lambda'P = -2\epsilon(L+\epsilon)I + \sum_{g\in\mathcal{G}_\epsilon}\lambda'_g gg^T$ and $\Lambda' \in \mathbb{R}^{d\times d}$ is a diagnal matrix. Let Λ'_i denote the (i,i)-th entry of Λ' and $h = (P^T)^{-1}(g'-g'')$. We continue the above derivation,

$$\begin{aligned}
&||g'-g''||^2_{\left(-2\epsilon(L+\epsilon)I+\sum_{g\in\mathcal{G}_\epsilon}(\tau\lambda_g+(1-\tau)\lambda'_g)gg^T\right)^{-1}} \\
=&h^T\Big(\tau I + (1-\tau)\Lambda'\Big)^{-1}h \\
=&\sum_{i=1}^d \frac{h_i^2}{\tau + (1-\tau)\Lambda'_i} \\
\le&\sum_{i=1}^d \tau h_i^2 + (1-\tau)\frac{h_i^2}{\Lambda'_i} \\
=&\tau\big|\big|(P^T)^{-1}(g'-g'')\big|\big|^2 + (1-\tau)\big|\big|(P^T)^{-1}(g'-g'')\big|\big|^2_{\Lambda'} \\
=&\tau||g'-g''||^2_{\left(-2\epsilon(L+\epsilon)I+\sum_{g\in\mathcal{G}_\epsilon}\lambda_g gg^T\right)^{-1}} \\
&+ (1-\tau)||g'-g''||^2_{\left(-2\epsilon(L+\epsilon)I+\sum_{g\in\mathcal{G}_\epsilon}\lambda'_g gg^T\right)^{-1}},
\end{aligned}$$

where the inequality is due to the convexity of the function $f(x) = \frac{1}{x}, \forall x > 0$. By taking maximum on both sides, we finally have that

$$\begin{aligned}
&\max_{g',g''\in\mathcal{S}} ||g'-g''||^2_{\left(-2\epsilon(L+\epsilon)I+\sum_{g\in\mathcal{G}_\epsilon}(\tau\lambda_g+(1-\tau)\lambda'_g)gg^T\right)^{-1}} \\
\le&\tau \max_{g',g''\in\mathcal{S}} ||g'-g''||^2_{\left(-2\epsilon(L+\epsilon)I+\sum_{g\in\mathcal{G}_\epsilon}\lambda_g gg^T\right)^{-1}} \\
&+(1-\tau)\max_{g',g''\in\mathcal{S}} ||g'-g''||^2_{\left(-2\epsilon(L+\epsilon)I+\sum_{g\in\mathcal{G}_\epsilon}\lambda'_g gg^T\right)^{-1}}
\end{aligned}$$

holds for any $g', g'' \in \mathcal{S}$ and $\delta \in (0,1)$, thus we complete the proof. □

However, there are some difficulties that prevent us from using this design. We are not aware of any constant upper bound on the size of $\mathrm{supp}(\lambda^{(b)})$, so the choice of $\tilde{T}$ should be overly conservative to avoid violating the sample budget constraint, which might lead to a high prediction error.

C.3 Proofs for SPC algorithm

Proof of Theorem 3. The proof involves two steps.

Step I. Verifying that the budget constraint T is never violated.

Suppose the SPC algorithm has pulled a total of t arms when it outputs its prediction $\hat{a}$. Let $V_t = \eta I + \sum_{s=1}^{t} a_s a_s^T$ be the final η-regularized design matrix. We have that

$$\begin{aligned}
\det(V_t) =& \det(V_{t-1} + a_t a_t^T) = \det(V_{t-1})\big(1 + ||a_t||^2_{V_{t-1}^{-1}}\big) \\
=& \det(\eta I) \prod_{s=1}^{t} \big(1 + ||a_s||^2_{V_{s-1}^{-1}}\big) \\
\geq& \det(\eta I) \prod_{s=1}^{t} \big(1 + \epsilon^2\big) = \det(\eta I)\big(1+\epsilon^2\big)^t \\
\geq& \det(\eta I) \exp\Big(\frac{t\epsilon^2}{2}\Big),
\end{aligned}$$

where the first inequality is by Line 5 in Algorithm 5, the second inequality is by the assumption that $\epsilon^2 \leq \frac{5}{2}$ and the fact that $e^{\frac{x}{2}} \leq 1 + x, \forall x \in [0, \frac{5}{2}]$. By adopting Lemma 3, we have that

$$\exp\Big(\frac{t\epsilon^2}{2}\Big) \leq \frac{\det(V_t)}{\det(\eta I)} \leq \frac{(\eta + tL^2/d)^d}{\eta^d}.$$

Taking logarithm on both sides, we have that t satisfies the following inequality:

$$t\epsilon^2 \leq 2d \ln\Big(1 + \frac{tL^2}{\eta d}\Big). \tag{3}$$

Recall that in the algorithm, we select $\epsilon = \sqrt{\frac{2d}{T}\ln\Big(1 + \frac{TL^2}{\eta d}\Big)}$. We define $l(t) = t\epsilon^2 - 2d\ln\Big(1 + \frac{tL^2}{\eta d}\Big)$ for any $t \geq 0$. Compute the first and second order derivatives of $l(t)$:

$$\begin{aligned}
\frac{dl(t)}{dt} &= \epsilon^2 - \frac{2L^2/\eta}{1 + tL^2/\eta d}, \\
\frac{d^2 l(t)}{dt^2} &= \frac{2L^4/\eta^2 d}{(1 + tL^2/\eta d)^2}.
\end{aligned}$$

Note that $\frac{d^2 l(t)}{dt^2} > 0$ for any $t \geq 0$, and $l(0) = l(T) = 0$. Rolle's theorem can be adopted to show that $\frac{dl(T)}{dt} > 0$. For any $t > T$, by Lagrange's mean value theorem, $\exists s \in (T, t)$ such that

$$\frac{dl(s)}{dt} = \frac{l(t) - l(T)}{t - T} \geq \frac{dl(T)}{dt} > 0,$$

which implies $l(t) > 0$. Any $t > T$ violates (3), thus we conclude that $t \leq T$.

Step II. Upper bounding the error probability.

If the prediction is incorrect, i.e., $\hat{a} \neq a^*$, there exists a sub-optimal arm whose empirical mean is higher than that of a^*. We have

$$\begin{aligned}
\mathbb{P}(\hat{a} \neq a^*) \leq& \mathbb{P}\Big(\exists a \neq a^* : \langle \hat{\theta}_{\max\{t_a, t_{a^*}\}}, a\rangle > \langle \hat{\theta}_{\max\{t_a, t_{a^*}\}}, a^*\rangle\Big), \\
\leq& \sum_{a \neq a^*} \mathbb{P}\Big(\langle \hat{\theta}_{\max\{t_a, t_{a^*}\}}, a\rangle > \langle \hat{\theta}_{\max\{t_a, t_{a^*}\}}, a^*\rangle\Big),
\end{aligned}$$

where t_a is the last time that arm a is pulled. If an arm a is never pulled, set t_a to be the time that a is read into $\tilde{a}$. Consider any arm $a \neq a^*$, we have that

$$\begin{aligned}
&\mathbb{P}\Big(\langle\hat{\theta}_{\max\{t_a,t_{a^*}\}}, a\rangle > \langle\hat{\theta}_{\max\{t_a,t_{a^*}\}}, a^*\rangle\Big)\\
=&\mathbb{P}\Big(\langle\hat{\theta}_{\max\{t_a,t_{a^*}\}} - \theta^*, a\rangle + \langle\theta^*, a\rangle > \langle\hat{\theta}_{\max\{t_a,t_{a^*}\}} - \theta^*, a^*\rangle + \langle\theta^*, a^*\rangle\Big)\\
=&\mathbb{P}\Big(\langle\hat{\theta}_{\max\{t_a,t_{a^*}\}} - \theta^*, a^* - a\rangle < -\Delta_a\Big)\\
\leq&\exp\Big(-\frac{\Delta_a^2}{2\sigma^2||a^*-a||^2_{V^{-1}_{\max\{t_a,t_{a^*}\}}}}\Big) \leq \exp\Big(-\frac{\Delta_a^2}{4\sigma^2\epsilon^2}\Big),
\end{aligned}$$

where the first inequality is by Lemma 5, the second inequality is due to Line 5 in Algorithm 5. Whenever the algorithm is comparing the empirical means of two arms a, a', we have $||a||^2_{V^{-1}_{\max\{t_a,t_{a'}\}}}$, $||a'||^2_{V^{-1}_{\max\{t_a,t_{a'}\}}} \leq \epsilon^2$. Consequently, we obtain

$$\mathbb{P}(\hat{a} \neq a^*) \leq (K-1)\exp\Big(-\frac{T\Delta^2}{8\sigma^2 d\ln\big(1+\frac{TL^2}{\eta d}\big)}\Big).$$

□

D Experiment Results

Our experiments were run on a PC with the following specifications: AMD Ryzen 9 5900X 12-Core Processor, NVIDIA GeForce RTX 3060 GPU, 32GB of memory.

D.1 Synthetic datasets

D.1.1 CR-MPS algorithm

In this experiment, we evaluate the performance of the proposed CR-MPS algorithm in the Linear Streaming Bandit setup for regret minimization, comparing it against established baseline algorithms from the Streaming Bandit literature. To the best of our knowledge, MBSE from Agarwal et al. [2022] and MPSE from Li et al. [2023] are the only existing baseline algorithms capable of operating in the regret minimization scenario.

The details of our experiment setup are as follows. We set d, the dimension of both the parameter vector θ^* and the arm profile vectors, to be 5. The algorithms are allowed to go through the arm stream 5 times, i.e., $B = 5$. The parameter vector θ^* and the arm set $\mathcal{K}$ are independently generated. Specifically, each entry of θ^* is sampled from Unif$([0, 4])$. Let the matrix $A \in \mathbb{R}^{d\times K}$ denote the arm stream. Each entry of A is sampled from Unif$([0.16994, 0.31305])$. The i.i.d arm reward noise follows a zero-mean Gaussian distribution of variance $\sigma^2 = 0.1^2$. We conducted multiple experiments under varying values of K, the size of the arm set $\mathcal{K}$. Every experiment consists of $N = 20$ independent repetitions. In each repetition, we independently generate a new pair of θ^* and $\mathcal{K}$ on which we run the proposed CR-MPS algorithm against the two aforementioned baseline algorithms. The regret curve is calculated as the average regret across all repetitions.

Figure 1 illustrates the results of this experiment. When the number of arms is small, all algorithms exhibit converging regret curves. However, the regret of CR-MPS is significantly lower than that of the baseline algorithms. As K increases, the baseline curves gradually become non-convergent, whereas CR-MPS continues to demonstrate strong performance.

D.1.2 G-MP-SE algorithm

In this experiment, we evaluate the performance of the proposed G-MP-SE algorithm for ϵ-Best Arm Identification in the Linear Streaming Bandit setup.

The details of the experimental setup are as follows. We set $d = 5$, $B = 5$, $K = 10^4$, and $\epsilon = 0.15$. The parameter vector θ^* and the arm set $\mathcal{K}$ are independently generated. Specifically, each entry of

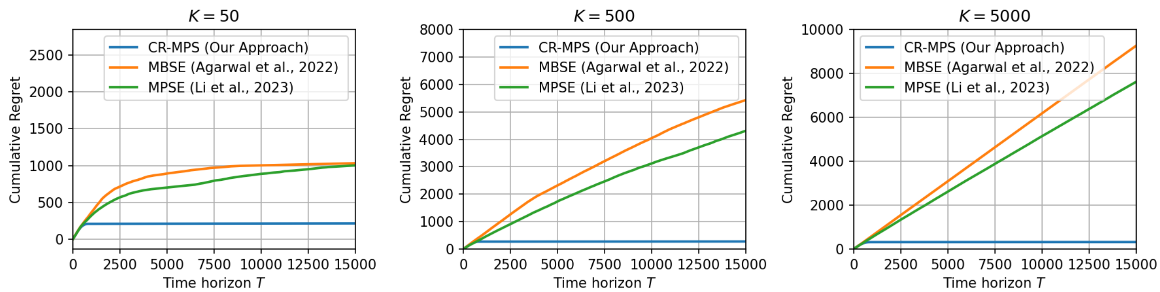

Figure 1: The regret of CR-MPS and baseline algorithms on a synthetic dataset.

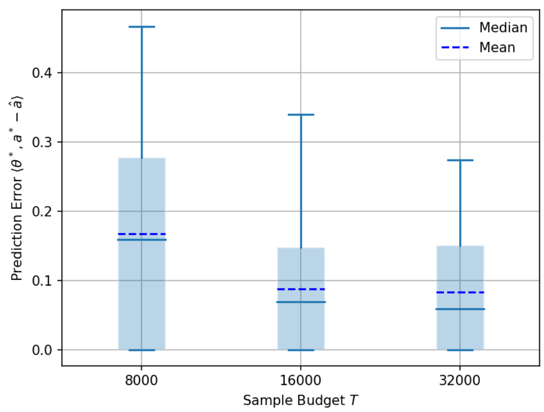

Figure 2: The prediction errors of G-MP-SE on a synthetic dataset.

θ^* is sampled from Unif($[0, 4]$). Let the matrix $A \in \mathbb{R}^{d\times K}$ denote the arm stream. Each entry of A is sampled from Unif($[0.16994, 0.31305]$). We also set the optimal arm a^* to be $0.315 * [1\ 1\ 1\ 1\ 1]^T$. The i.i.d arm reward noise follows a zero-mean Gaussian distribution of variance $\sigma^2 = 5^2$. We implement G-optimal design using the `minimize` method in the `scipy.optimize` package. We conduct several experiments with varying sample budgets T. Every experiment consists of $N = 40$ independent repetitions. In each repetition, we independently generate a new pair of θ^* and $\mathcal{K}$ on which we run the proposed G-MP-SE algorithm. Box plots are used to record the prediction errors for each experiment.

Figure 2 shows the results of this experiment. As the sample budget T increases, we can observe a decrease in the prediction error of G-MP-SE.

D.1.3 SPC algorithm

In this experiment, we evaluate the performance of the proposed SPC algorithm for Best Arm Identification in the Linear Streaming Bandit setup.

The details of the experiment setup are as follows. We set $d = 5$. The parameter vector θ^* and the arm set $\mathcal{K}$ are independently generated. Specifically, each entry of θ^* is sampled from Unif($[0, 4]$). Let the matrix $A \in \mathbb{R}^{d\times K}$ denote the arm stream. Each entry of A is sampled from Unif($[0.16994, 0.31305]$). We also set the optimal arm a^* to be $0.319 * [1\ 1\ 1\ 1\ 1]^T$. The i.i.d arm reward noise follows a zero-mean Gaussian distribution of variance $\sigma^2 = 4^2$. We conduct several experiments under different values of the sample budget T and the number of arms K. Every experiment consists of $N = 500$ independent repetitions. In each repetition, we independently generate a new pair of θ^* and $\mathcal{K}$ on which we run the proposed SPC algorithm. The error probability is estimated using these N observations.

Table 1 presents the results of this experiment. As the sample budget T increases, a clear decrease in the error probability of SPC is observed.

$\mathbb{P}(\hat{a} \neq a^*)$ \ T / K	10,000	20,000	30,000	50,000
1,000	0.120	0.074	0.056	0.022
10,000	0.186	0.102	0.070	0.036
50,000	0.206	0.142	0.094	0.058
100,000	0.266	0.148	0.096	0.076

Table 1: The error probability of SPC on a synthetic dataset.

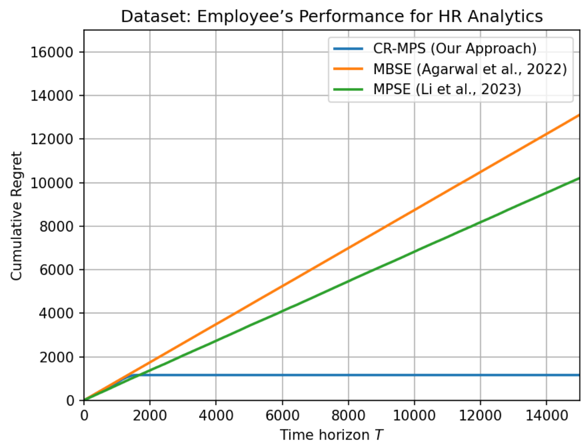

Figure 3: The regret of CR-MPS and baseline algorithms on a real-world dataset.

D.2 Real-World dataset

Since one of the motivations for our Linear Streaming Bandit model is recruitment, we use a Kaggle dataset of job applicants to evaluate the effectiveness of our proposed algorithms. The dataset we utilize is **Employee's Performance for HR Analytics**[1] (Chaudhari [2023]). This dataset contains the profile information of a large number of employees. We select $d = 8$ of the characteristics to construct the arm profile vectors. These characteristics are Education Level, Gender, Count of Training Programs Attended, Age, Length of Service, Previous KPI, Awards Won, and Average Training Score. Performance ratings in the dataset are used to represent the reward for selecting each employee. The ground truth parameter vector θ^* is estimated using linear regression on the entire dataset.

D.2.1 CR-MPS algorithm

In this experiment, we evaluate the performance of the proposed CR-MPS algorithm in the Linear Streaming Bandit setup for regret minimization, comparing it against the known baseline algorithms from the Streaming Bandit literature.

The details of the experiment setup are as follows. We set $K = 10000$. We run these algorithms on a uniformly randomly shuffled K-sized subset of the Employee's Performance for HR Analytics dataset. The algorithms are allowed to go through the arm stream 5 times, i.e., $B = 5$. The experiment is repeated $N = 30$ times. The resulting regret curve is calculated as the average regret across all repetitions.

Figure 3 presents the result for this experiment. The regret of CR-MPS is significantly lower than the regret of the baseline algorithms.

D.2.2 G-MP-SE algorithm

In this experiment, we evaluate the performance of the proposed G-MP-SE algorithm for ϵ-Best Arm Identification in the Linear Streaming Bandit setup.

[1] https://www.kaggle.com/datasets/sanjanchaudhari/employees-performance-for-hr-analytics/data

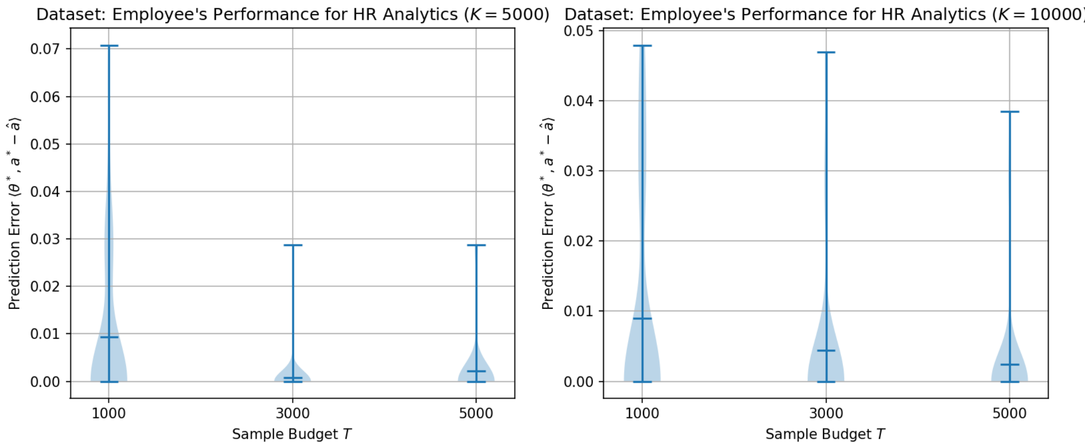

Figure 4: The prediction errors of G-MP-SE on a real-world dataset.

$\mathbb{P}(\hat{a} \neq a^*)$ K \ T	1,000	3,000	5,000	8,000
5,000	0.244	0.050	0.020	0.000
10,000	0.248	0.032	0.022	0.002
15,000	0.016	0.000	0.000	0.000

Table 2: The error probability of SPC on a real-world dataset.

The details of the experiment setup are as follows. We set $K \in \{5000, 10000\}$. We implement G-optimal design using the `minimize` method in the `scipy.optimize` package. We run the algorithm on uniformly randomly shuffled K-sized subsets of the Employee's Performance for HR Analytics dataset. Every experiment consists of $N = 40$ independent repetitions. Violin plots are used to record the prediction errors for each experiment.

Figure 4 presents the results of this experiment. As the sample budget T increases, a decrease in the prediction error of G-MP-SE is observed.

D.2.3 SPC algorithm

In this experiment, we evaluate the performance of the proposed SPC algorithm for Best Arm Identification in the Linear Streaming Bandit setup.

The details of the experiment setup are as follows. We run the algorithm on uniformly randomly shuffled K-sized subsets of the Employee's Performance for HR Analytics dataset, with various sample budget T. Every experiment consists of $N = 500$ independent repetitions. The error probability is estimated using these N observations.

Table 2 presents the results of this experiment. As the sample budget T increases, a decrease in the error probability of SPC is observed.

References

Yasin Abbasi-yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011. URL https://proceedings.neurips.cc/paper_files/paper/2011/file/e1d5be1c7f2f456670de3d53c7b54f4a-Paper.pdf.

Arpit Agarwal, Sanjeev Khanna, and Prathamesh Patil. A sharp memory-regret trade-off for multi-pass streaming bandits. In Po-Ling Loh and Maxim Raginsky, editors, *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages

1423–1462. PMLR, 02–05 Jul 2022. URL https://proceedings.mlr.press/v178/agarwal22a.html.

Sanjana Chaudhari. Employee's performance for hr analytics, 2023. URL https://www.kaggle.com/ds/3537629.

Martin Jaggi. Revisiting frank-wolfe: Projection-free sparse convex optimization. In *International conference on machine learning*, pages 427–435. PMLR, 2013.

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

Shaoang Li, Lan Zhang, Junhao Wang, and Xiang-Yang Li. Tight memory-regret lower bounds for streaming bandits. *arXiv preprint arXiv:2306.07903*, 2023.