Online Decision Making with High-Dimensional Covariates

Siyu Xie

Department of Statistics, Northwestern University

March 8, 2024

Outline

Outline

LASSO Bandit - Motivation

- Sparsity - LASSO identifies a sparse subset of predictive covariates, which is an effective approach for treatment effect estimation in practice.
- Asymptotic performance - some techniques create substantial bias in our estimates to increase predictive accuracy for small sample sizes.
- Data-poor regimes - the performance of all existing algorithms scales polynomially in the number of covariates d, and provides no theoretical guarantees when the number of users is of order d.

Main Contributions

- Adapted LASSO to the bandit setting and tune the resulting bias-variance trade-off over time to transit from data-poor to data-rich regimes.
- Proved theoretical guarantees that the algorithm achieves good performance as soon as the number of users T is polyogarithmic in d, which is an exponential improvement over existing theory.
- Empirically demonstrated the potential benefit in a medical decision-making context with real patient data.

Outline

Notation

- $[n]$: the set $\{1, 2, \ldots, n\}$;
- β_I: for any index set $I \subset [d]$, the vector obtained by setting the elements of β that are not in I to zero, $\beta_I \in \mathrm{R}^d$.
- $\mathrm{supp}(v)$: the set of indices corresponding to nonzero entries of v.
- T: the number of (unknown) time steps.
- K: the number of arms.
- reward $X_t^\top \beta_i + \epsilon_{i,t}$, where $\epsilon_{i,t}$ are independent $\sigma-$ subgaussian random variables.
- r_t: expected regret. $r_t = \mathbb{E}\left[\max_j(X_t^\top \beta_j) - X_t^\top \beta_i\right]$
- s_0: sparsity parameter.

Assumptions

- **Assumption 1 (Parameter set).** There exist positive constants $x_{\max}$ and b such that $\|x\|_\infty \leq x_{\max}$ for all $x \in \mathcal{X}$ and $\|\beta_i\|_1 \leq b$ for all $i \in [K]$. The former implies that any realization of the random variable X_t satisfies $\|X_t\|_\infty \leq x_{\max}$ for all t.
- **Assumption 2 (Margin condition).** There exists a constant $C_0 \in \mathbb{R}^+$ such that for all i and j in $[K]$ where $i \neq j$, $\Pr\left[0 < \left|X^\top (\beta_i - \beta_j)\right| \leq \kappa\right] \leq C_0 \kappa$ for all $\kappa \in \mathbb{R}^+$.

Assumptions

- **Assumption 3 (Arm optimality).** Let $\mathcal{K}_{\text{opt}}$ and $\mathcal{K}_{\text{sub}}$ be mutually exclusive sets that include all K arms. Then there exists some $h > 0$ such that:
  (a) sub-optimal arms $i \in \mathcal{K}_{\text{sub}}$ satisfy $x^\top \beta_i < \max_{j \neq i} x^\top \beta_j - h$ for every $x \in \mathcal{X}$; and (b) for a constant $p_* > 0$, each optimal arm $i \in \mathcal{K}_{\text{opt}}$ has a corresponding set

$$U_i \equiv \left\{ x \in \mathcal{X} \mid x^\top \beta_i > \max_{j \neq i} x^\top \beta_j + h \right\}$$

  such that $\min_{i \in \mathcal{K}_{\text{opt}}} \Pr[X \in U_i] \geq p_*$.

Assumption 3: Arm Optimality

Our third assumption is a less restrictive version of an assumption introduced in Goldenshluger and Zeevi (2013). In particular, we assume that our K arms can be split into two sets:

a. Sub-optimal arms $\mathcal{K}_{\text{sub}}$ that are strictly sub-optimal for all covariate vectors in $\mathcal{X}$, i.e., there exists a constant $h_{\text{sub}} > 0$ such that for each $i \in \mathcal{K}_{sub}, x^\top \beta_i < \max_{j \neq i} x^\top \beta_j - h_{\text{sub}}$ for every $x \in \mathcal{X}$.

b. A non-empty set of optimal arms $\mathcal{K}_{\text{opt}}$ that are strictly optimal with positive probability for some covariate vectors $x \in \mathcal{X}$, i.e., there exists a constant $h_{\text{opt}} > 0$ and some region $U_i \subset \mathcal{X}$ (with $\Pr[X \in U_i] = p_i > 0$) for each $i \in \mathcal{K}_{\text{opt}}$ such that $x^\top \beta_i > \max_{j \neq i} x^\top \beta_j + h_{\text{opt}}$ for all covariate vectors x in U_i.

Assumption 4: Compatibility Condition

Definition 2 (Compatibility Condition)

For any set of indices $I \subset [d]$ and a positive and deterministic constant ϕ, define the set of matrices

$$\mathcal{C}(I, \phi) \equiv \{M \in \mathbb{R}_{\succeq 0}^{d \times d} \mid \forall v \in \mathbb{R}^d \text{ s.t. } \|v_{I^c}\|_1 \leq 3 \|v_I\|_1 , \text{ we have } \|v_I\|_1^2 \leq |I| \left(v^\top M v\right) / \phi^2\}.$$

- **Assumption 4 (Compatibility condition).** There exists a constant $\phi_0 > 0$ such that for each $i \in \mathcal{K}_{\text{opt}}$, $\Sigma_i \in \mathcal{C}\left(\text{supp}\left(\beta_i\right), \phi_0\right)$, where we define $\Sigma_i \equiv \mathbb{E}\left[XX^\top \mid X \in U_i\right]$.

Outline

Additional Notation

- design matrix $\mathbf{X}$: $T \times d$ matrix whose rows are X_t.
- Y_i: length of T vector of observations $X_t^\top \beta_i + \epsilon_{i,t}$.
- all-sample set $\mathcal{S}_i$: $\mathcal{S}_i = \{t | \pi_t = i\} \subset [T]$, set of times when arm i was played.
- $\mathbf{X}(\mathcal{S}')$: $|S'| \times d$ submatrix of $\mathbf{X}$ whose rows are X_t for each $t \in \mathcal{S}'$.
- $Y_i(\mathcal{S}')$: defined similarly, when $\mathcal{S}' \subset \mathcal{S}_i$, it is length $|\mathcal{S}'|$ vector of corresponding observed rewards $Y_i(t)$. Note that since $\pi_t = i$ for each $t \in \mathcal{S}'$, $Y_i(\mathcal{S}')$ has no missing entries.
- $\hat{\Sigma}(\mathbf{Z}) = \mathbf{Z}^\top \mathbf{Z}/n$: its sample covariance matrix.
- $\hat{\Sigma}(\mathcal{A})$ to refer to $\hat{\Sigma}(\mathbf{Z}(\mathcal{A}))$.
- $\hat{\beta}(\mathcal{S}', \lambda)$: simpler notation of $\hat{\beta}_{\mathbf{X}(\mathcal{S}'), Y(\mathcal{S}'), \lambda}$ (LASSO estimator).

LASSO Estimation

Definition 3 (LASSO).

Given a regularization parameter $\lambda \geq 0$, the LASSO estimator is

$$\hat{\beta}_{\mathbf{X},Y}(\lambda) \equiv \arg\min_{\beta'} \left\{ \frac{\| Y - \mathbf{X}\beta' \|_2^2}{n} + \lambda \left\| \beta' \right\|_1 \right\}$$

The LASSO estimator satisfies the following *tail inequality*.

LASSO Estimation

Proposition 1 (LASSO Tail Inequality for Adapted Observations).

Let X_t denote the t^{th} row of $\mathbf{X}$ and $Y(t)$ denote the t^{th} entry of Y. The sequence $\{X_t : t = 1, \ldots, n\}$ forms an adapted sequence of observations, i.e., X_t may depend on past regressors and their resulting observations $\{X_{t'}, Y(t')\}_{t'=1}^{t-1}$. Also, assume that all realizations of random vectors X_t satisfy $\|X_t\|_\infty \leq x_{\max}$. Then for any $\phi > 0$ and $\chi > 0$, if $\lambda = \lambda(\chi, \phi) \equiv \chi\phi^2/(4s_0)$, we have

$$\begin{aligned} \Pr\left[\left\|\hat{\beta}_{\mathbf{X},Y}(\lambda) - \beta\right\|_1 > \chi\right] \leq & 2\exp\left[-C_1(\phi)n\chi^2 + \log d\right] + \\ & \Pr[\hat{\Sigma}(\mathbf{X}) \notin \mathcal{C}(\text{supp}(\beta), \phi)], \end{aligned}$$

where $C_1(\phi) \equiv \phi^4/\left(512 s_0^2 \sigma^2 x_{\max}^2\right)$.

LASSO for bandit setting

We then consider estimating the parameter β_i for each arm $i \in [K]$. Using any subset of past samples $\mathcal{S}' \subset \mathcal{S}_i$ (arm i was played) and any λ, we can use the corresponding LASSO estimator $\hat{\beta}(\mathcal{S}', \lambda)$, to estimate β_i.
In order to prove regret bounds, we need to establish convergence guarantees for such estimates.
From Proposition 1 , in order to bound the error $\left\|\hat{\beta}(\mathcal{S}', \lambda) - \beta_i\right\|_1$ for each arm $i \in [K]$, we need to

- ensure with high probability $\hat{\Sigma}(\mathcal{S}') \in \mathcal{C}(\text{supp}(\beta_i), \phi)$ for some constant ϕ
- appropriately choose parameters λ over time to control the rate of convergence

Thus, the main challenge in the algorithm and analysis is constructing and maintaining sets $\mathcal{S}'$ such that with high probability $\hat{\Sigma}(\mathcal{S}') \in \mathcal{C}(\text{supp}(\beta_i), \phi)$, (although the rows of $\mathbf{X}(\mathcal{S}')$ are not i.i.d.) with sufficiently fast convergence rates.

Description of Algorithm

- The LASSO Bandit takes as input the *forced sampling parameter* $q \in \mathbb{Z}^+$(which is used to construct the forced-sample sets), a *localization parameter* $h > 0$ (defined in Assumption 3)[3], as well as initial regularization parameters $\lambda_1, \lambda_{2,0}$.
- These parameters will be specified in Theorem 1 .

We prescribe a set of times when we forced-sample arm i (regardless of the observed covariates X_t):

$$\mathcal{T}_i \equiv \{ (2^n - 1) \cdot Kq + j \mid n \in \{0, 1, 2, \ldots\} \text{ and } j \in \{q(i-1)+1, q(i-1)+2, \ldots, qi\}.$$

Thus, the set of forced samples from arm i up to time t is $\mathcal{T}_{i,t} \equiv \mathcal{T}_i \cap [t]$, with size $\mathcal{O}(q \log t)$.

As before, let $\mathcal{S}_{i,t} = \{t' \mid \pi_{t'} = i \text{ and } 1 \leq t' \leq t\}$ denote the set of times we play arm i up to time t. Note that by definition $\mathcal{T}_{i,t} \subset \mathcal{S}_{i,t}$. At any time t, the LASSO Bandit maintains two sets of parameter estimates for each β_i :

1. the forced-sample estimate $\hat{\beta}\left(\mathcal{T}_{i,t-1}, \lambda_1\right)$ based only on forced samples observed from arm i,
2. the all-sample estimate $\hat{\beta}\left(\mathcal{S}_{i,t-1}, \lambda_{2,t}\right)$ based on all samples observed from arm i.

Description of Algorithm

Execution

- If the current time t is in $\mathcal{T}_i$ for some arm i, then arm i is played.
- Otherwise, two actions are possible.
  - First, we use the forced-sample estimates to find the highest estimated reward achievable across all K arms.
  - We then select the subset of arms $\hat{\mathcal{K}} \subset [K]$ whose estimated rewards are within $h/2$ of the maximum achievable.
  - After this pre-processing step, we use the all-sample estimates to choose the arm with the highest estimated reward within the set $\hat{\mathcal{K}}$.

Algorithm LASSO Bandit

Input parameters: $q, h, \lambda_1, \lambda_{2,0}$

Initialize $\mathcal{T}_{i,0}$ and $\mathcal{S}_{i,0}$ by the empty set, and $\hat{\beta}(\mathcal{T}_{i,0}, \lambda_1)$ and $\hat{\beta}(\mathcal{S}_{i,0}, \lambda_{2,0})$ by 0 in $\mathbb{R}^d$ for all i in $[K]$

Use q to construct force-sample sets $\mathcal{T}_i$ using Eq. (2) for all i in $[K]$

for $t \in [T]$ **do**

  Observe user covariates $X_t \sim \mathcal{P}_X$

  if $t \in \mathcal{T}_i$ for any i **then**

    $\pi_t \leftarrow i$ (forced-sampling)

  else

    $\hat{\mathcal{K}} = \left\{ i \in [K] \mid X_t^\top \hat{\beta}(\mathcal{T}_{i,t-1}, \lambda_1) \geq \max_{j \in [K]} X_t^\top \hat{\beta}(\mathcal{T}_{j,t-1}, \lambda_1) - h/2 \right\}$ is the set of near-optimal arms according to the forced-sample estimators

    $\pi_t \leftarrow \arg\max_{i \in \hat{\mathcal{K}}} X_t^\top \hat{\beta}(\mathcal{S}_{i,t-1}, \lambda_{2,t-1})$ is the best arm within $\hat{\mathcal{K}}$ according to the all-sample estimators

  end if

  Update all-sample sets $\mathcal{S}_{\pi_t,t} \leftarrow \mathcal{S}_{\pi_t,t-1} \cup \{t\}$ and regularization $\lambda_{2,t} \leftarrow \lambda_{2,0}\sqrt{\frac{\log t + \log d}{t}}$

  Play arm π_t, observe $Y(t) = X_t^\top \beta_{\pi_t} + \varepsilon_{i,t}$

end for

[3] Note that if some $\bar{h}$ satisfies Assumption 3, then any $h \in (0, \bar{h}]$ also satisfies the assumption. Therefore, a conservatively small value can be chosen in practice, but this will be reflected in the constant in the regret bound.

Regret Analysis of LASSO Bandit

Theorem 1

When $q \geq 4\lceil q_0 \rceil, K \geq 2, d > 2, t \geq C_5$, and we take $\lambda_1 = \left(\phi_0^2 p_* h\right) / \left(64 s_0 x_{\max}\right)$ and $\lambda_{2,0} = \left[\phi_0^2 / \left(2 s_0\right)\right] \sqrt{1 / \left(p_* C_1\right)}$, we have the following (non-asymptotic) upper bound on the expected cumulative regret of the LASSO Bandit at time T by:

$$
\begin{aligned}
R_T \leq & \ C_3(\log T)^2 + \left[2 K b x_{\max}(6 q+4) + C_3 \log d\right] \log T \\
& + \left(2 b x_{\max} C_5 + 2 K b x_{\max} + C_4\right) \\
& = \mathcal{O}\left(s_0^2[\log T + \log d]^2\right)
\end{aligned}
$$

where the constants $C_1\left(\phi_0\right), C_2\left(\phi_0\right), C_3\left(\phi_0, p_*\right), C_4\left(\phi_0, p_*\right)$, and C_5 are given by

$$
C_1\left(\phi_0\right) \equiv \frac{\phi_0^4}{512 s_0^2 \sigma^2 x_{\max}^2}, \quad C_2\left(\phi_0\right) \equiv \min \left(\frac{1}{2}, \frac{\phi_0^2}{256 s_0 x_{\max}^2}\right), \quad C_3\left(\phi_0, p_*\right) \equiv \frac{1024 K C_0 x_{\max}^2}{p_*^3 C_1},
$$

$$
C_4\left(\phi_0, p_*\right) \equiv \frac{8 K b x_{\max}}{1-\exp \left[-\frac{p_*^2 C_2^2}{32}\right]}, \quad C_5 \equiv \min \left\{t \in \mathbb{Z}^{+} \mid t \geq 24 K q \log t + 4(K q)^2\right\},
$$

and we take $q_0 \equiv \max \left\{\frac{20}{p_*}, \frac{4}{p_* C_2^2}, \frac{12 \log d}{p_* C_2^2}, \frac{1024 x_{\max}^2 \log d}{h^2 p_*^2 C_1}\right\} = \mathcal{O}\left(s_0^2 \log d\right)$.

Outline

Key Steps of the Analysis

In this section, we outline the proof strategy of Theorem 1.

- Prove a new general LASSO tail inequality that holds even when the rows of the design matrix are not iid (Section 4.1).
- Use this result to obtain convergence guarantees for the forced-sample (Section 4.2) and all sample estimators (Section 4.3) under a fixed regularization path.
- Sum up the expected regret from the errors in the estimators.

A LASSO Tail Inequality for Non-i.i.d. Data

- Letting $\Sigma \equiv \mathbb{E}_{Z \sim \mathcal{P}_Z}\left[ZZ^\top\right]$, we further assume that $\Sigma \in \mathcal{C}\left(\text{supp}(\beta), \phi_1\right)$ for a constant $\phi_1 \in \mathbb{R}^+$.
- We will show that if the number $|\mathcal{A}'|$ of i.i.d. samples is sufficiently large, then we can prove a convergence guarantee for the LASSO estimator $\hat{\beta}(\mathcal{A}, \lambda)$ trained on samples in $\mathcal{A}$, which includes non-i.i.d. samples.

Lemma 1

For any $\chi > 0$, if $d > 1, |\mathcal{A}'| / |\mathcal{A}| \geq p/2, |\mathcal{A}| \geq 6 \log d / \left(p C_2 (\phi_1)^2 \right)$, and $\lambda = \lambda\left(\chi, \phi_1 \sqrt{p}/2\right) = \chi \phi_1^2 p / (16 s_0)$, then the following tail inequality holds:

$$
\begin{aligned}
&\Pr\left[\|\hat{\beta}(\mathcal{A}, \lambda) - \beta\|_1 > \chi \right] \\
&\leq 2 \exp\left[-C_1 \left(\frac{\phi_1 \sqrt{p}}{2} \right) |\mathcal{A}| \chi^2 + \log d \right] \\
&+ \exp\left[-p C_2 (\phi_1)^2 |\mathcal{A}|/2 \right].
\end{aligned}
$$

LASSO Tail Inequality for the Forced Sample Estimator

Section 4.2

Proposition 2

Proposition 2. For all $i \in [K]$, the forced sample estimator $\hat{\beta}\left(\mathcal{T}_{i,t}, \lambda_1\right)$ satisfies

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

when $\lambda_1 = \phi_0^2 p_* h / \left(64 s_0 x_{\max}\right)$, $t \geq (Kq)^2$, $q \geq 4\lceil q_0 \rceil$, and q_0 satisfies the definition in Section 3.3.

- The challenge is that the all-sample sets $\mathcal{S}_{i,t}$ depend on choices made online by the algorithm.
- The algorithm selects arm i at time t based both on X_t and on previous observations $\{X_{t'}\}_{1\leq t'<t}$.
- As a consequence, the variables $\{X_t \mid t \in \mathcal{S}_{i,t}\}$ may be correlated.
- We resolve this by showing that
  - (a) our algorithm uses the forced-sample estimator $\mathcal{O}(T)$ times with high probability, and
  - (b) a constant fraction of the samples where we use the forced-sample estimator are i.i.d. from the regions U_i. We then invoke Lemma 1 with a modified $\mathcal{A}'$ such that $|\mathcal{A}'| = \mathcal{O}(T)$.

LASSO Tail Inequality for the All-Sample Estimator

Section 4.3

In particular, we define the event

$$A_t \equiv \left\{ \left\| \hat{\beta}\left(\mathcal{T}_{i,t}, \lambda_1\right) - \beta_i \right\|_1 \leq \frac{h}{4x_{\max}}, \quad \forall i \in [K] \right\}.$$

Since the event A_t only depends on forced-samples, the random variables $\{X_t \mid A_{t-1} \text{ holds }\}$ are i.i.d. (with distribution $\mathcal{P}_X$). Furthermore, if we let

$$\begin{aligned} \mathcal{S}'_{i,t} \equiv & \{t' \in [t] \mid A_{t'-1} \text{ holds, } X_{t'} \in U_i \text{, and} \\ & t' \notin \cup_{j \in [K]} \mathcal{T}_{j,t}\} \end{aligned}$$

then the random variables $\left\{X_{t'} \mid t' \in \mathcal{S}'_{i,t}\right\}$ are i.i.d.

LASSO Tail Inequality for the All-Sample Estimator

Section 4.3

Proposition 3

The all-sample estimator $\hat{\beta}\left(\mathcal{S}_{i,t}, \lambda_{2,t}\right)$ for $i \in \mathcal{K}_{\text{opt}}$ satisfies the tail inequality

$$\Pr\left[\left\|\hat{\beta}\left(\mathcal{S}_{i,t}, \lambda_{2,t}\right) - \beta_i\right\|_1 > 16\sqrt{\frac{\log t + \log d}{p_*^3 C_1\left(\phi_0\right) t}}\right] < \frac{2}{t} + 2\exp\left[-\frac{p_*^2 C_2\left(\phi_0\right)^2}{32} \cdot t\right]$$

when $\lambda_{2,t} = \left[\phi_0^2/\left(2s_0\right)\right]\sqrt{\left(\log t + \log d\right)/\left(p_* C_1\left(\phi_0\right) t\right)}$ and $t \geq C_5$.

LASSO Tail Inequality for the All-Sample Estimator

- Note that the all-sample estimator tail inequality only holds for optimal arms $\mathcal{K}_{\text{opt}}$ while the forced-sample estimator tail inequality holds for all arms $[K]$.
- However, the algorithm requires a preprocessing step using the forced sample estimator to
  (a) ensure that we obtain $O(T)$ i.i.d. samples for each $i \in \mathcal{K}_{opt}$ and
  (b) to prune out suboptimal arms $\mathcal{K}_{sub}$ with high probability.

Bounding the Cumulative Expected Regret

We divide the time periods $[T]$ into three groups:

1. Initialization ($t \le C_5$), or forced sampling ($t \in \mathcal{T}_{i,T}$ for some $i \in [K]$).
2. Times $t > C_5$ when the event A_{t-1} does not hold.
3. Times $t > C_5$ when the event A_{t-1} holds and we do not perform forced sampling; that is, the LASSO Bandit plays the estimated best arm from $\hat{\mathcal{K}}$ (chosen by the forced-sampling estimator) using the all-sample estimator.

Proof of Main Result

Theorem 1

When $q \geq 4\lceil q_0 \rceil$, $K \geq 2$, $d > 2$, $t \geq C_5$, and we take $\lambda_1 = \left(\phi_0^2 p_* h\right) / (64 s_0 x_{\max})$ and $\lambda_{2,0} = \left[\phi_0^2 / (2 s_0)\right] \sqrt{1/(p_* C_1)}$, we have the following (non-asymptotic) upper bound on the expected cumulative regret of the LASSO Bandit at time T by:

$$\begin{aligned} R_T &\leq C_3 (\log T)^2 + [2Kbx_{\max}(6q+4) + C_3 \log d] \log T \\ &\quad + (2bx_{\max} C_5 + 2Kbx_{\max} + C_4) \\ &= \mathcal{O}\left(s_0^2 [\log T + \log d]^2\right) \end{aligned}$$

where the constants $C_1(\phi_0)$, $C_2(\phi_0)$, $C_3(\phi_0, p_*)$, $C_4(\phi_0, p_*)$, and C_5 are given by

$$C_1(\phi_0) \equiv \frac{\phi_0^4}{512 s_0^2 \sigma^2 x_{\max}^2}, \quad C_2(\phi_0) \equiv \min\left(\frac{1}{2}, \frac{\phi_0^2}{256 s_0 x_{\max}^2}\right), \quad C_3(\phi_0, p_*) \equiv \frac{1024 K C_0 x_{\max}^2}{p_*^3 C_1},$$

$$C_4(\phi_0, p_*) \equiv \frac{8Kbx_{\max}}{1 - \exp\left[-\frac{p_*^2 C_2^2}{32}\right]}, \quad C_5 \equiv \min\left\{t \in \mathbb{Z}^+ \mid t \geq 24Kq \log t + 4(Kq)^2\right\},$$

and we take $q_0 \equiv \max\left\{\frac{20}{p_*}, \frac{4}{p_* C_2^2}, \frac{12 \log d}{p_* C_2^2}, \frac{1024 x_{\max}^2 \log d}{h^2 p_*^2 C_1}\right\} = \mathcal{O}\left(s_0^2 \log d\right)$.

Proof of Main Result

Proof of Theorem 1

The total expected cumulative regret of the LASSO Bandit up to time T is upper-bounded by summing all the terms from Lemmas EC.15, EC.17, and EC.20:

$$
\begin{aligned}
R_T \leq\, & \overbrace{2bx_{\max}\left(6qK\log T + C_5\right)}^{\text{Regret from }(a)} + \overbrace{2Kbx_{\max}}^{\text{Regret from }(b)} \\
& + \overbrace{\left(8Kbx_{\max} + C_3\log d\right)\log T + C_3(\log T)^2 + C_4}^{\text{Regret from }(c)} \\
=\, & C_3(\log T)^2 + \left[2Kbx_{\max}(6q+4) + C_3\log d\right]\log T \\
& + \left(2bx_{\max}C_5 + 2Kbx_{\max} + C_4\right) \\
=\, & \log T\left[C_3\log T + 2Kbx_{\max}(6q+4) + C_3\log d\right] \\
& + \left(2bx_{\max}C_5 + 2Kbx_{\max} + C_4\right).
\end{aligned}
$$

Proof of Main Result

Proof of Theorem 1 (Cont'd)

Now, using $q = \mathcal{O}\left(s_0^2 \log d\right)$, and the fact that $C_0, \ldots, C_5, b, x_{\max}$, and ϕ_0 are constants,

$$R_T = \mathcal{O}\left(\log T\left[\log T + s_0^2 \log d\right]\right) = \mathcal{O}\left(s_0^2[\log T + \log d]^2\right)$$

Outline

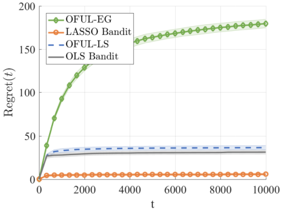
Empirical Results

- We compare the LASSO Bandit against
  a the UCB-based algorithmOFUL-LS (Abbasi-Yadkori et al.2011),which is an improved version of the algorithm sug-gested in Dani et al. (2008),
  b a sparse variant OFUL-EG for high-dimensional settings (Abbasi-Yadkori2012, Abbasi-Yadkori et al.2012), and
  c the OLSBandit by Goldenshluger and Zeevi (2013). Our re-sultsdemonstrate thattheLASSO Bandit significantlyoutperforms these benchmarks. Separately, wefindthat the LASSO Bandit is robust to changes in inputparameters by even an order of magnitude

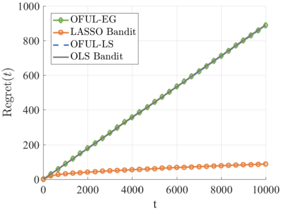
Empirical Results

- We consider three scenarios for K, d, and s_0: a) $K = 2$, $d = 100$, $s_0 = 5$; (b) $K = 10$, $d = 1000$, $s_0 = 2$; and (c) $K = 50$, $d = 20$, $s_0 = 2$.
- In each case, we consider K arms (treatments) and d user covariates, where only a randomly chosen subset of s_0 covariates are predictive of the reward for each treatment,
  - for each $i \in [K]$, the arm parameters β_i are set to zero except for s_0 randomly selected components that are drawn from a uniform distribution on $[0, 1]$.
  - Note that the OFUL-EG algorithm requires an additional technical assumption that $\Sigma_{i=1}^{K} \|\beta_i\|_1 = 1$. We scale our β_i's accordingly so that this assumption is met.
- Next, at each time t, user covariates Xt are independently sampled from a Gaussian distribution $N(\mathbf{0}_d, \mathbf{I}_d)$ and truncated so that $\|X_t\|_\infty = 1$.
- Finally, we set the noise variance to be $\sigma^2 = 0.052$.

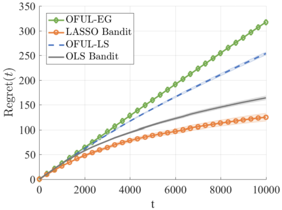
Empirical Results



(a) $K = 2,\ d = 100,\ s_0 = 5$



(b) $K = 10,\ d = 1000,\ s_0 = 2$

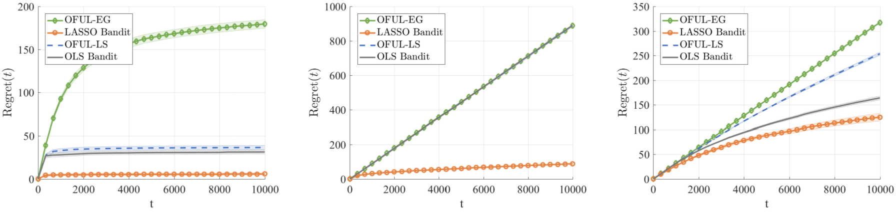


(c) $K = 50,\ d = 20,\ s_0 = 2$

(a) LASSO Bandit may be useful even in low-dimensional regime

- because other algorithms continue to overfit the arm parameters.

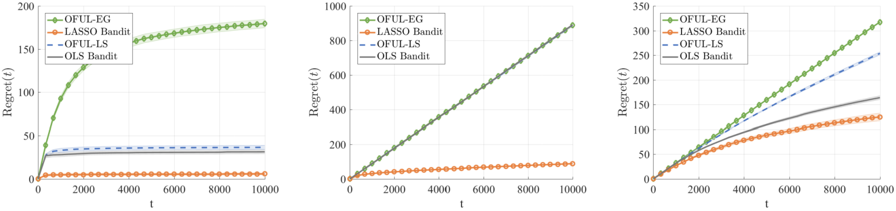
Empirical Results



(a) $K = 2,\ d = 100,\ s_0 = 5$ (b) $K = 10,\ d = 1000,\ s_0 = 2$ (c) $K = 50,\ d = 20,\ s_0 = 2$

(b) Gap between the LASSO Bandit and the other algorithm increases significantly.

- Because benchmark algorithms do not take advantage of sparsity and perform exploration for at least $O(Kd)$ samples in order to define linear regression estimates for each arm.



(a) $K=2,\ d=100,\ s_0=5$ (b) $K=10,\ d=1000,\ s_0=2$ (c) $K=50,\ d=20,\ s_0=2$

(c) Performance gap decreases.

- LASSO Bandit does not provide any improvement over existing algorithms in K, and
- provides limited improvement when the number of covariates is very small.