

A Smoothed Analysis of Online Lasso for the Sparse Linear Contextual Bandit Problem

Zhiyuan Liu¹, Huazheng Wang², Bo Waggoner¹, Youjian (Eugene) Liu³, Lijun Chen¹

¹ Department of Computer Science, University of Colorado, Boulder.

² Department of Computer Science, University of Virginia.

³ Department of Electrical, Computer and Energy Engineering, University of Colorado, Boulder.

Sparse Linear Contextual Bandit Problem

- **Arm Set:** each arm i is associated with a feature(context) $x_i^t \in \mathcal{R}^d$.
 - **Noisy Reward:** $r_i^t = (x_i^t)^\top \theta^* + \eta^t$, $|\theta^*|_0 \leq k$, θ^* is unknown to the learner.
- ❖ $k \ll d$ and d could be very large, e.g., $d \gg T$. Denote the set of non-zero entries of θ^* by S (effective dimension set).

Estimate θ^* (Lasso regression) reward vector $Y^t = [r_{a_1}^1, \dots, r_{a_t}^t]^\top$; Context matrix $X^t = [x_{a_t}^1, \dots, x_{a_t}^t]$.

$$\min_{\theta} \|Y^t - (X^t)^\top \theta\|_2^2 + \lambda^t \|\theta\|_1,$$

Lasso regression has a strong requirement for X^t to achieve **sampling efficiency**.

Compressed Sensing(C.S.)

Null Space Condition (Cohen et al., 2009),
Restricted Isometry Property(RIP) (Donoho, 2006),
Restricted Eigenvalue(RE) condition (Bickel et al., 2009),
Compatibility condition (Van De Geer et al., 2009),
.....



Sparse Bandit

“Online-to-Conversion” (Abbasi et al. 2012)
Bandit with C.S. (Carpentier and Munos, 2012)
Hypercube Arm Bandit (Lattimore et al. 2015)
Doubly-robust Lasso Bandit (Kim and Paik, 2019)
High Covariate Sparse Bandit (Bastani et al. 2020)

Smoothed Contexts

To improve the sampling efficiency, we consider the perturbed adversary setting (Kannan et al. 2018).

Perturbed Adversary

- 1) Arms' contexts $(\mu_1^t, \dots, \mu_m^t)$ are produced adversarially.
- 2) Perturbed adversary adds small random perturbations (e_1^t, \dots, e_m^t) (i.i.d., non-adaptive) to the context and outputs them as arm features to the learner, that is, $(x_1^t, \dots, x_m^t) = (\mu_1^t + e_1^t, \dots, \mu_m^t + e_m^t)$.

Online Lasso For Sparse Bandit Under Perturbed Adversary

Initialize θ^0, X^0 and Y^0

For $t = 1, 2, 3, \dots, T$:

Perturbed adversary produce m contexts (x_1^t, \dots, x_m^t) .

The learner **greedily** chooses arm $a_t = \arg_i \max (x_i^t)^\top \theta^t$, receives the reward $r_{a_t}^t$ and Update (X^t, Y^t) to (X^{t+1}, Y^{t+1}) . Calculate θ^{t+1} by Lasso regression:

$$\theta^{t+1} = \arg \min_{\theta} \| Y^{t+1} - (X^{t+1})^\top \theta \|_2^2 + \lambda^{t+1} \|\theta\|_1$$

Sparse Bandit: low dimensional case

When $d < T$, we prove a linearly strong convex condition which leads to the optimal sparse recovery.

$$\lambda_{\min}(\mathbb{E}_{e_i^t \sim D}[x_i^t (x_i^t)^\top]) \geq \lambda \quad \text{Perturbed Diversity} \quad + \quad \text{Random Matrix Theory} \quad \equiv \quad \text{Linearly Strong Convex}$$

Key Result

With the perturbed adversary and $t > \frac{2R^2}{g\left(\frac{2q}{\sigma_1}, 0\right)\sigma_1^2} \log dT$, the following is satisfied with probability $1 - \frac{1}{T}$:

$$\lambda_{\min}(X^t (X^t)^\top) \geq g\left(\frac{2q}{\sigma_1}, 0\right) (1 - \tau) \sigma_1^2 t,$$

$$\text{where } \tau = \sqrt{\frac{2R^2}{g\left(\frac{2q}{\sigma_1}, 0\right)\sigma_1^2} \log dT}.$$

$$\frac{2R^2}{g\left(\frac{2q}{\sigma_1}, 0\right)\sigma_1^2} \log dT$$

Number of necessary samplings (exploration length)

Linearly strong convex guarantees the optimal sparse recovery $\mathcal{O}\left(\sqrt{\frac{k \log d}{t}}\right)$.

Sparse Bandit: high dimensional case

RE with high probability

Consider perturbation $e_i^t \sim \mathcal{N}(0, \Sigma)$ where $\|\Sigma^{\frac{1}{2}}\Delta\|_2 \geq \gamma\|\Delta\|_2$ for $\Delta \in \mathcal{C}(S; 3)$. If

$$t > \max\left(\frac{4c''q(\Sigma)}{\gamma^2} k \log d, \frac{8196aR^2\lambda_{max}(\Sigma) \log t}{\gamma^4}\right)$$

Exploration length

then with the probability $1 - \left(\frac{c'}{ect} + \frac{1}{Ta}\right)$,


$$\Delta^T X^t (X^t)^T \Delta \geq ht \|\Delta\|_2^2,$$

where c, c', c'' are universal constants, $q(\Sigma) = \max_i \Sigma_{ii}$ and $h = \left(\frac{\gamma^2}{64} - R\|\Delta\|_2^2 \sqrt{\frac{2a\lambda_{max}(\Sigma) \log T}{t}}\right)$.

- The larger perturbation does not indicate the better regret. $\text{Cond}(\Sigma) \geq \frac{q(\Sigma)}{\gamma^2} \geq 1$.
- Condition number and SPR (the signal to perturbation ratio).

$$\frac{R^2 \lambda_{max}(\Sigma)}{\gamma^4} = \frac{R^2}{\gamma^2} \frac{\lambda_{max}(\Sigma)}{\gamma^2}$$

SPR Condition number

 We prove both cases will achieve the optimal regret $\mathcal{O}(\sqrt{kT \log d})$.