

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

Zhiyuan Liu

ZHIYUAN.LIU@COLORADO.EDU

Department of Computer Science, University of Colorado, Boulder

Huazheng Wang

HW7WW@VIRGINIA.EDU

Department of Computer Science, University of Virginia

Bo Waggoner

BO.WAGGNOR@COLORADO.EDU

Department of Computer Science, University of Colorado, Boulder

Youjian (Eugene) Liu

YOUJIAN.LIU@COLORADO.EDU

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

Lijun Chen

LIJUN.CHEN@COLORADO.EDU

Department of Computer Science, University of Colorado, Boulder

Abstract

We investigate the sparse linear contextual bandit problem where the parameter θ is sparse. To relieve the sampling inefficiency, we utilize the “perturbed adversary” where the context is generated adversarially but with small random non-adaptive perturbations. We prove that the simple online Lasso supports sparse linear contextual bandit with regret bound $\mathcal{O}(\sqrt{kT \log d})$ even when $d \gg T$ where k and d are the number of effective and ambient dimension, respectively. Compared to the recent work from Sivakumar et al. (2020), our analysis does not rely on the precondition processing, adaptive perturbation (the adaptive perturbation violates the i.i.d perturbation setting) or truncation on the error set. Moreover, the special structures in our results explicitly characterize how the perturbation affects exploration length, guide the design of perturbation together with the fundamental performance limit of perturbation method. Numerical experiments are provided to complement the theoretical analysis.

1. Introduction

Contextual bandit algorithms have become a referenced solution for sequential decision-making problems such as online recommendations (Li et al., 2010), clinical trials (Durand et al., 2018), dialogue systems (Upadhyay et al., 2019) and anomaly detection (Ding et al., 2019). It adaptively learns the personalized mapping between the observed contextual features and unknown parameters such as user preferences, and addresses the trade-off between exploration and exploitation (Auer, 2002; Li et al., 2010; Abbasi-Yadkori et al., 2011; Agrawal and Goyal, 2013; Abeille et al., 2017).

We consider the sparse linear contextual bandit problem where the context is high-dimensional with sparse unknown parameter θ (Abbasi-Yadkori et al., 2012; Hastie et al., 2015; Dash and Liu, 1997), i.e., most entries in θ are zero and thus only a few dimensions of the context feature are relevant to the reward. Due to insufficient data samples, the learning algorithm has to be sampling efficiency to support the sequential decision-making. However,

the data from bandit model usually does not satisfy the requirements for sparse recovery such as 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) and so on. To achieve the desired performance, current works has to consider the restricted problem settings, e.g., the unit-ball, hypercube or i.i.d. arm set (Carpentier and Munos, 2012; Lattimore et al., 2015; Kim and Paik, 2019; Bastani and Bayati, 2020), the parameter with Gaussian prior (Gilton and Willett, 2017). One exception is the Online-to-Confidence-Set Conversions (Abbasi-Yadkori et al., 2012) which considers the general setting but suffers from computation inefficiency.

In this paper, we tackle the sparse linear bandit problem using *smoothed analysis* technique (Spielman and Teng, 2004; Kannan et al., 2018), which enjoys efficient implementation and mild assumptions. Specifically, we consider the perturbed adversary setting where the context is generated adversarially but perturbed by small random noise. This setting interpolates between an i.i.d. distributional assumption on the input, and the worst-case of fully adversarial contexts. Our results show that with a high probability, the perturbed adversary inherently guarantees the (linearly) strong convex condition for the low dimensional case and the restricted eigenvalue (RE) condition for the high dimensional case, which is a key property required by the standard Lasso regression. We prove that the simple online Lasso supports sparse linear contextual bandits with regret bound $\mathcal{O}(\sqrt{kt \log d})$. We also provide numerical experiments to complement the theoretical analysis.

We also notice the recent work from Sivakumar et al. (2020) using smoothed analysis for structured linear contextual bandits. Compared to their work, our proposed method has the following advantages: (1) Our analysis only relies on the simple online Lasso instead of precondition processing and truncation on the error set. Although preconditioning transfers the non-zero singular value to 1, this could amplify the noise, and the preconditioned noises are no longer i.i.d., which makes concentration analysis difficult and the estimation unstable (Jia et al., 2015). We also observe this effect in the numeric experiments. (2) Their proof relies on the assumption that perturbations that need to be adaptively generated based on the observed history of the chosen contexts. Instead, our analysis is based on the milder assumption that the perturbation is i.i.d. and non-adaptive. (3) Their regret does not describe the full picture of the effect of variance of the perturbation. Our analysis explicitly show how the perturbation affects the exploration length, guide the design of perturbation together with the fundamental performance limit of perturbation method.

2. Model and Methodology

In the bandit problem, at each round t , the learner pulls an arm a_t among m arms (we denote the arm sets by $[m]^1$, that is, $a_t \in [m]$) and receives the corresponding noisy reward $r_{a_t}^t$. The performance of the learner is evaluated by the regret \mathcal{R} which quantifies the total loss because of not choosing the best arm a_t^* during T rounds:

$$\mathcal{R}(T) = \sum_{t=1}^T (r_{a_t^*}^t - r_{a_t}^t). \quad (1)$$

1. In this paper, we denote by $[n]$ the set $[1, \dots, n]$ for positive integer n .

In this paper, we focus on the sparse linear contextual bandit problem. Specially, each arm i at round t is associated with a feature (context) vector $\mu_i^t \in \mathbb{R}^d$. The reward of that arm is assumed to be generated by the noisy linear model which is the inner product of arm feature and an unknown S -sparse parameter θ^* where S denotes the set of effective (non-zero) entries and $|S| = k$. That is,

$$r_i^t = \langle \mu_i^t, \theta^* \rangle + \eta^t, |\theta^*|_0 = k, \quad (2)$$

where η^t follows Gaussian distribution $\mathcal{N}(0, \sigma^2)$. To handle the non-convex L_0 norm, Lasso regression is the natural way to learn the sparse θ^* with the relaxation from L_0 to L_1 norm. To achieve the desired performance, the algorithm has to rely on well designed contexts which guarantee sampling efficiency requirements such as 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) and so on. However, the data from bandit problems usually does not satisfy these conditions since the contexts could be generated adversarially. Up to now, deciding on the proper assumptions for sparsity bandit problems is still a challenge (Lattimore and Szepesvári, 2018).

Inspired by the smoothed analysis for greedy algorithm of linear bandit problem (Kannan et al., 2018), we consider the *Perturbed Adversary* defined below for the sparse linear contextual bandit problem.

Definition 1 *Perturbed Adversary* (Kannan et al., 2018). *The perturbed adversary acts as the following at round t .*

- *Given the current context μ_1^t, \dots, μ_m^t which could be chosen adversarially, the perturbation e_1^t, \dots, e_m^t are drawn independently from certain distribution. Also, each e_i^t is independently (non-adaptively) produced of the context.*
- *The perturbed adversary outputs the contexts $(x_1^t, \dots, x_m^t) = (\mu_1^t + e_1^t, \dots, \mu_m^t + e_m^t)$ as the arm features to the learner.*

Let $X \in \mathbb{R}^{d \times t}$ be the context matrix where each column contains one context vector and Y the column vector that contains the corresponding rewards. Based on perturbed adversary setting, we analyze the online Lasso in Algorithm 1 for sparse linear contextual bandit. Generally speaking, our analysis considers two cases using different techniques, one for the low dimensional case when $d < T$ and the other for the high dimensional case when $d \gg T$. For the low dimensional case, the analysis utilizes random matrix theory (Tropp, 2012) to prove that with a high probability, the minimum eigenvalue of scaled sample covariance matrix is increasing linearly with round t ; for the high dimensional case, the RE condition is guaranteed with the help of Gaussian perturbation’s property (Raskutti et al., 2010) that the nullspace of context matrix under Gaussian perturbation cannot contain any vectors that are “overly” sparse when t is larger than some threshold. The properties of both cases support $\mathcal{O}(\sqrt{\frac{k \log d}{t}})$ parameter recovery of Lasso regression under noisy environment which leads to $\mathcal{O}(\sqrt{kT \log d})$ regret.

Algorithm 1: Online Lasso For Sparse Linear Contextual Bandit Under Perturbed Adversary

- 1 Initialize θ^0 , X and Y .
 - 2 **for** $t = 1, 2, 3, \dots, T$ **do**
 - 3 The perturbed adversary produces m context $[x_1^t, \dots, x_m^t]$.
 - 4 The learner greedily chooses the arm $i = \arg \max_{j \in [m]} \langle x_j^t, \theta^t \rangle$, observes the reward r_i^t , appends the new observation (x_i^t, r_i^t) to (X, Y) , and updates θ^{t+1} by the Lasso regression:

$$\theta^{t+1} = \arg \min_{\theta} G(\theta; \lambda^t) := \|Y - X^\top \theta\|_2^2 + \lambda^t \|\theta\|_1. \quad (3)$$
 - 5 **end**
-

2.1 Low Dimensional Case

We first consider the low dimensional case when $d < T$. Under the perturbed adversary setting, we then define the property named perturbed diversity which is adopted from Bastani et al. (2017).

Definition 2 Perturbed Diversity. Let $e_i^t \sim D$ on \mathbb{R}^d . Given any context vector μ_i^t , we call x_i^t perturbed diversity if for $x_i^t = \mu_i^t + e_i^t$, the minimum eigenvalue of sample covariance matrix under perturbations satisfies

$$\lambda_{\min} \left(\mathbb{E}_{e_i^t \sim D} \left[x_i^t (x_i^t)^\top \right] \right) \geq \lambda_0,$$

where λ_0 is a positive constant.

Intuitively speaking, perturbed diversity guarantees that each context provides at least certain information about all coordinates of θ^* from the expectation which is helpful to recover the support of the parameter via regularized method. We can find several distributions D that could make the perturbed diversity happen, e.g., the Gaussian distribution. However, without any restriction, x_i^t could be very large and out of the realistic domain. Instead, the value of each dimension (we denote by $x_i^t(j)$ the j -th dimension of x_i^t) should lie in a bounded interval, in the meanwhile, the total energy of context vector is bounded by certain constant, i.e., $\|x_i^t\|_2^2 \leq R^2$. This motivates us to consider the perturbed diversity under censored perturbed adversary.

Lemma 1 Given the context vector $\mu_i^t \in \mathbb{R}^d$ and $|\mu_i^t(j)| \leq q_j$ for each $j \in [d]$, we define the censored perturbed context x_i^t under $e_i^t \sim \mathcal{N}(\mathbf{0}, \sigma_1^2 I)$ as follows:

$$x_i^t(j) = \begin{cases} \mu_i^t(j) + e_i^t(j), & \text{if } |\mu_i^t(j) + e_i^t(j)| \leq q_j, \\ q_j, & \text{if } \mu_i^t(j) + e_i^t(j) > q_j, \\ -q_j, & \text{if } \mu_i^t(j) + e_i^t(j) < -q_j. \end{cases} \quad (4)$$

Then x_i^t has the perturbed diversity with $\lambda_0 = g(\frac{2q}{\sigma_1}, 0)\sigma^2$, where $q = \min_j q_j$ and $g(\cdot, \cdot)$ is a composite function of the probability density function $\phi(\cdot)$ and the cumulative distribution function $\Phi(\cdot)$ of the normal distribution. Please refer to equation (14) for more details.

The proof is provided in the appendix and one can easily extend it to the case where $e_i^t \sim \mathcal{N}(\mathbf{0}, \Sigma)$. Based on Lemma 1, we can derive that with a high probability, $\lambda_{\min}(XX^\top)$ grows at least with a linear rate t .

Lemma 2 *With the censored perturbed diversity, when $t > \frac{2R^2}{g(\frac{2q}{\sigma_1}, 0)\sigma_1^2} \log(dT)$, the following is satisfied with probability $1 - \frac{1}{T}$: $\lambda_{\min}(XX^\top) \geq g\left(\frac{2q}{\sigma_1}, 0\right)(1 - \tau)\sigma_1^2 t$, where $\tau = \sqrt{\frac{2R^2}{g(\frac{2q}{\sigma_1}, 0)\sigma_1^2 t} \log(dT)}$.*

As one can see from Lemma 2, after certain number of (implicit) exploration rounds, i.e., $\frac{2R^2}{g(\frac{2q}{\sigma_1}, 0)\sigma_1^2} \log(dT)$, we will have enough information to support the $\mathcal{O}(\sqrt{\frac{k \log d}{t}})$ parameter recovery by Lasso regression. The regret analysis together with the high dimensional case is deferred to the next section.

2.2 High Dimensional Case

Now we turn to the high dimensional case when $d \gg T$. During the learning process, the scaled sample covariance matrix XX^\top is always rank deficiency which means $\lambda_{\min}(XX^\top) = 0$ and Lemma 2 based on random matrix theory can not be applied here any more. We then consider the restricted eigenvalue (RE) condition instead. Here the ‘‘restricted’’ means that the error $\Delta^t := \theta^t - \theta^*$ incurred by Lasso regression is restricted to a set with special structure. That is, $\Delta^t \in \mathcal{C}(S; \alpha)$ where

$$\mathcal{C}(S; \alpha) := \{\theta \in \mathbb{R}^d \mid \|\theta_{S^c}\|_1 \leq \alpha \|\theta_S\|_1\},$$

and $\alpha \geq 1$ is determined by the regularized parameter λ^t . In the following section, we focus on $\mathcal{C}(S; 3)$ which could be achieved by setting $\lambda^t = \Theta(2\sigma R \sqrt{2t \log(2d)})$.

The key is to prove that Null space of X^\top has no overlapping with $\mathcal{C}(S; 3)$. It has been proved that special cases in which contexts are purely sampled from special distributions such as Gaussian and Bernoulli distributions, satisfy this property (Zhou, 2009; Raskutti et al., 2010; Haupt et al., 2010). We make a further step to show that nullspace of context matrix under Gaussian perturbation cannot contain any vectors that are ‘‘overly’’ sparse when t is larger than some threshold.

Theorem 1 *Consider perturbation $e_i^t \sim \mathcal{N}(\mathbf{0}, \Sigma)$ where $\|\Sigma^{1/2}\Delta\|_2 \geq \gamma\|\Delta\|_2$ for $\Delta \in \mathcal{C}(S; 3)$. If $t > \max(\underbrace{\frac{4c''q(\Sigma)}{\gamma^2} k \log d}_d, \underbrace{\frac{8196aR^2\lambda_{\max}(\Sigma) \log T}{\gamma^4}}_e)$, then with probability $1 - (\frac{c'}{c^t} + \frac{1}{T^a})$, we have $\Delta^\top XX^\top \Delta \geq ht\|\Delta\|_2^2$, where c, c', c'' are universal constants, $q(\Sigma) = \max_i \Sigma_{ii}$ and $h = (\frac{\gamma^2}{64} - R\|\Delta\|_2^2 \sqrt{\frac{2a\lambda_{\max}(\Sigma) \log T}{t}})$.*

Moreover, we can design $\gamma^2 = \lambda_{\min}(\Sigma)$. By Rayleigh quotient, one can obtain $\lambda_{\max}(\Sigma) \geq q(\Sigma) = \max_i \Sigma_{ii} \geq \min_i \Sigma_{ii} \geq \lambda_{\min}(\Sigma) = \gamma^2$.

We then discuss how perturbations will affect the exploration length. First, *the larger perturbation does not indicate the less regret*. Results of Sivakumar et al. (2020) show that

the regret is $\mathcal{O}(\frac{\log T \sqrt{T}}{\sigma_1})$ where σ_1 is the perturbation's variance, and suggests choosing larger σ_1 leads to smaller regret bounds. However this is not the full picture that shows the effect of the perturbation's variance. Our results show that increasing the variance of the perturbation has limited effect over the necessary exploration and regret, which reveals theoretical limit of the perturbation method. Specifically, in the term **(d)** of Theorem 1, no matter how large the variance is, the term $\frac{q(\Sigma)}{\gamma^2} \geq 1$. So $4c''k \log d$ is the necessary exploration length and cannot be improved. Second, *Condition Number and the SPR (the signal to perturbation ratio)* are important factors. The condition number $\mathbf{Cond}(\Sigma)$ controls both term **(d)** and **(e)**, e.g., $\frac{q(\Sigma)}{\gamma^2} \leq \frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Sigma)} = \mathbf{Cond}(\Sigma)$. This also shows that the optimal perturbation design will choose $\Sigma = \sigma_1 I$. In **(e)** of Theorem 1, $\frac{R^2}{\gamma^2}$ can be regarded as the ratio between the energy of the unperturbed context and the perturbation energy. This ratio shows the trade-off between *exploration* and *fidelity*. That is, a large variance not only reduces the exploration (meanwhile, the lower bound is guaranteed by **(e)**) but also reduces the fidelity of original context.

3. Regret Analysis

Based on the properties we have proved for the low and high dimensional cases, we can obtain the following recovery guarantee by the techniques from the standard Lasso regression (Hastie et al., 2015).

Lemma 3 *If $t > T_e$ and $\lambda^t = 2\sigma R \sqrt{2t \log \frac{2d}{\delta}}$, the Lasso regression under perturbed adversary has the recovery guarantee $\|\theta^t - \theta^*\|_2 \leq \frac{3\sigma R}{C} \sqrt{\frac{2k \log 2d/\delta}{t}}$ with probability $1 - \delta$, where $T_e = \frac{2R^2}{g(\frac{2q}{\sigma_1}, 0)\sigma_1^2} \log(dT)$, $C = g(\frac{2q}{\sigma_1}, 0)(1 - \tau)\sigma_1^2$ for the low dimensional case and $T_e = \max(\frac{4c''q(\Sigma)}{\gamma^2}k \log d, \frac{8196aR^2\lambda_{\max}(\Sigma) \log T}{\gamma^4})$, $C = \frac{\gamma^2}{64} - R\sqrt{\frac{2a\lambda_{\max}(\Sigma) \log T}{t}}$ for the high dimensional case.*

We then get the final result in Theorem 2 based on all the analysis above.

Theorem 2 *The online Lasso for sparse linear contextual bandit under perturbed adversary admits the following regret with probability $1 - \delta$.*

$$\text{Regret} \leq 2R \left(T_e + \frac{6\sigma R}{C} \sqrt{2kT \frac{\log 2d}{\delta}} \right) = \mathcal{O}(\sqrt{kT \log d}). \quad (5)$$

4. Conclusion

This paper utilizes the ‘‘perturbed adversary’’ where the context is generated adversarially but with small random non-adaptive perturbations to tackle sparse linear contextual bandit problem. We prove that the simple online Lasso supports sparse linear contextual bandit with regret bound $\mathcal{O}(\sqrt{kT \log d})$ for both low and high dimensional cases and show how the perturbation affects the exploration length and the trade-off between exploration and fidelity. Future work will focus on extending our analysis to more challenge setting, i.e., defending against adversarial attack for contextual bandit model.

References

- Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 2312–2320, 2011.
- Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. Online-to-confidence-set conversions and application to sparse stochastic bandits. In *Artificial Intelligence and Statistics*, pages 1–9, 2012.
- Marc Abeille, Alessandro Lazaric, et al. Linear thompson sampling revisited. *Electronic Journal of Statistics*, 11(2):5165–5197, 2017.
- Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*, pages 127–135, 2013.
- Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.
- Hamsa Bastani and Mohsen Bayati. Online decision making with high-dimensional covariates. *Operations Research*, 68(1):276–294, 2020.
- Hamsa Bastani, Mohsen Bayati, and Khashayar Khosravi. Mostly exploration-free algorithms for contextual bandits. *arXiv preprint arXiv:1704.09011*, 2017.
- Peter J Bickel, Ya’acov Ritov, Alexandre B Tsybakov, et al. Simultaneous analysis of lasso and dantzig selector. *The Annals of statistics*, 37(4):1705–1732, 2009.
- Alexandra Carpentier and Rémi Munos. Bandit theory meets compressed sensing for high dimensional stochastic linear bandit. In *Artificial Intelligence and Statistics*, pages 190–198, 2012.
- Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. Compressed sensing and best k-term approximation. *Journal of the American mathematical society*, 22(1):211–231, 2009.
- Manoranjan Dash and Huan Liu. Feature selection for classification. *Intelligent data analysis*, 1(3):131–156, 1997.
- Kaize Ding, Jundong Li, and Huan Liu. Interactive anomaly detection on attributed networks. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 357–365, 2019.
- David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.
- Audrey Durand, Charis Achilleos, Demetris Iacovides, Katerina Strati, Georgios D Mitsis, and Joelle Pineau. Contextual bandits for adapting treatment in a mouse model of de novo carcinogenesis. In *Machine Learning for Healthcare Conference*, pages 67–82, 2018.

- Davis Gilton and Rebecca Willett. Sparse linear contextual bandits via relevance vector machines. In *2017 International Conference on Sampling Theory and Applications (SampTA)*, pages 518–522. IEEE, 2017.
- Trevor Hastie, Robert Tibshirani, and Martin Wainwright. *Statistical learning with sparsity: the lasso and generalizations*. Chapman and Hall/CRC, 2015.
- Jarvis Haupt, Waheed U Bajwa, Gil Raz, and Robert Nowak. Toeplitz compressed sensing matrices with applications to sparse channel estimation. *IEEE transactions on information theory*, 56(11):5862–5875, 2010.
- Jinzhu Jia, Karl Rohe, et al. Preconditioning the lasso for sign consistency. *Electronic Journal of Statistics*, 9(1):1150–1172, 2015.
- Sampath Kannan, Jamie H Morgenstern, Aaron Roth, Bo Waggoner, and Zhiwei Steven Wu. A smoothed analysis of the greedy algorithm for the linear contextual bandit problem. In *Advances in Neural Information Processing Systems*, pages 2227–2236, 2018.
- Gi-Soo Kim and Myunghee Cho Paik. Doubly-robust lasso bandit. In *Advances in Neural Information Processing Systems*, pages 5869–5879, 2019.
- Tor Lattimore and Csaba Szepesvári. Bandit algorithms. *preprint*, page 28, 2018.
- Tor Lattimore, Koby Crammer, and Csaba Szepesvári. Linear multi-resource allocation with semi-bandit feedback. In *Advances in Neural Information Processing Systems*, pages 964–972, 2015.
- Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670. ACM, 2010.
- Garvesh Raskutti, Martin J Wainwright, and Bin Yu. Restricted eigenvalue properties for correlated gaussian designs. *Journal of Machine Learning Research*, 11(Aug):2241–2259, 2010.
- Vidyashankar Sivakumar, Zhiwei Steven Wu, and Arindam Banerjee. Structured linear contextual bandits: A sharp and geometric smoothed analysis. *arXiv preprint arXiv:2002.11332*, 2020.
- Daniel A Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004.
- Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.
- Sohini Upadhyay, Mayank Agarwal, Djallel Bounneffouf, and Yasaman Khazaeni. A bandit approach to posterior dialog orchestration under a budget. *arXiv preprint arXiv:1906.09384*, 2019.

Sara A Van De Geer, Peter Bühlmann, et al. On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics*, 3:1360–1392, 2009.

Shuheng Zhou. Restricted eigenvalue conditions on subgaussian random matrices. *arXiv preprint arXiv:0912.4045*, 2009.

Appendix

Lemma 4 (*A variant of Matrix Chernoff Tropp (2012)*) Consider a finite sequence z_t of independent, random, self-adjoint matrices satisfy

$$z_t \succeq 0 \text{ and } \lambda_{\max}(z_t) \leq Q \text{ almost surely.}$$

Compute the minimum eigenvalue of the sum of expectations, $\psi_{\min} := \lambda_{\min}(\sum_t \mathbb{E}(z_t))$. Then for $\delta \in [0, 1]$, we have

$$\mathbb{P} \left\{ \lambda_{\min} \left(\sum_t z_t \right) \leq (1 - \delta) \psi_{\min} \right\} \leq d \left[\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right]^{\psi_{\min}/Q}. \quad (6)$$

Moreover, for any $\psi \leq \psi_{\min}$, we can get

$$\mathbb{P} \left\{ \lambda_{\min} \left(\sum_t z_t \right) \leq (1 - \delta) \psi \right\} \leq d \left[\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right]^{\psi/Q}. \quad (7)$$

Proof Since $\psi \leq \psi_{\min}$, there exists $\delta_1 \in [0, 1]$ such that $\psi = \delta_1 \psi_{\min}$. We have

$$(1 - \delta) \psi = (1 - \delta) \delta_1 \psi_{\min} = (1 - \underbrace{(1 - \delta_1 + \delta \delta_1)}_{\delta_2}) \psi_{\min}.$$

Plugging this into (6) leads to

$$\mathbb{P} \left\{ \lambda_{\min} \left(\sum_t z_t \right) \leq (1 - \delta) \psi \right\} \leq d \left[\frac{e^{-\delta_2}}{(1 - \delta_2)^{1-\delta_2}} \right]^{\psi_{\min}/Q}.$$

One can easily verify that $\delta_2 \geq \delta$. So

$$\left[\frac{e^{-\delta_2}}{(1 - \delta_2)^{1-\delta_2}} \right]^{\frac{\psi_{\min}}{Q}} \leq \left[\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right]^{\frac{\psi_{\min}}{Q}} \leq \left[\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right]^{\frac{\psi}{Q}}.$$

Then we obtain

$$\mathbb{P} \left\{ \lambda_{\min} \left(\sum_t z_t \right) \leq (1 - \delta) \psi \right\} \leq d \left[\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right]^{\psi/Q}. \quad (8)$$

Since $\frac{e^{-\delta}}{(1-\delta)^{1-\delta}} \leq e^{-\delta^2/2}$, so we have the following when $\delta \in [0, 1]$:

$$\mathbb{P} \left\{ \lambda_{\min} \left(\sum_t z_t \right) \leq (1 - \delta) \psi \right\} \leq d \left[e^{-\delta^2/2} \right]^{\psi/Q}. \quad (9)$$

■

Fact 1 Let $\eta = [\eta_1, \dots, \eta_t]^\top$ where each η_i i.i.d. from $\mathcal{N}(0, \sigma^2)$. Let $X \in \mathbb{R}^{d \times t}$ where each $|X_{ij}| \leq R$. Then with a high probability $1 - \delta$, we have

$$\|X\eta\|_\infty \leq \sigma R \sqrt{2t \log \frac{2d}{\delta}}.$$

Fact 2 (Chernoff Bound for Sum of Sub-Gaussian random variables) Let X_1, \dots, X_n be n independent random variables such that $X_i \sim \text{subG}(\sigma^2)$. Then for any $a \in \mathbb{R}^n$ and $c \geq 0$, we have

$$\Pr \left(\sum_{i=1}^n a_i X_i < -c \right) \leq \exp \left(-\frac{c^2}{2\sigma^2 \|a\|_2^2} \right). \quad (10)$$

That is, with a high probability at least $1 - \delta$, we have

$$\sum_{i=1}^n a_i X_i > -\sqrt{2\sigma^2 \|a\|_2^2 \log \frac{1}{\delta}}. \quad (11)$$

Lemma 5 (Restricted Eigenvalue Property (Corollary 1 of Raskutti et al. (2010))) Suppose that Σ satisfies the RE condition of order k with parameters $(1, \gamma)$ and denote $q(\Sigma) = \max_i \Sigma_{ii}$. Then for universal positive constants c, c', c'' , if the sample size satisfies

$$t > \frac{4c'' q(\Sigma)}{\gamma^2} k \log d, \quad (12)$$

then the matrix $\frac{\Phi \Phi^\top}{t}$ satisfies the RE condition with parameters $(1, \frac{\gamma}{8})$ with probability at least $1 - \frac{c'}{e^{ct}}$ where $\Phi \in \mathbb{R}^{d \times t}$ and each column is i.i.d. $\mathcal{N}(\mathbf{0}, \Sigma)$.

Proof of Lemma 1

Proof Since $e_i^t(j)$ is independent of each other, we can analyze it by coordinates. To simplify the analysis, we slightly abuse the notations and remove subscript i and superscript t (only within this proof), that is, $x(j) := x_i^t(j)$ and $e(j) := e_i^t(j)$.

$$\begin{aligned} \lambda_{\min} \left(\mathbb{E} \left[x x^\top \right] \right) &= \min_{\|w\|=1} w^\top \mathbb{E} [x x^\top] w \\ &= \min_{\|w\|=1} \mathbb{E} (w^\top x x^\top w) \\ &= \min_{\|w\|=1} \mathbb{E} (\langle w, x \rangle^2) \\ &\geq \min_{\|w\|=1} \text{Var}(\langle w, x \rangle) \\ &\geq \min_{\|w\|=1} \text{Var}(\langle w, e \rangle) \\ &= \min_{\|w\|=1} \sum_{i=1}^d (w(i))^2 \text{Var}(e(i) | \text{censored in } [-q_i, q_i]) \\ &\geq \min_{\|w\|=1} g(2q/\sigma, 0) \sigma_1^2 \sum_{i=1}^d (w(i))^2 \\ &= g(2q/\sigma_1, 0) \sigma_1^2, \end{aligned}$$

where $g(2q/\sigma_1, 0)$ is according to Lemma 6. ■

Lemma 6 *Let $e \sim \mathcal{N}(0, \sigma_1^2)$. For any interval $[a, b]$ which contains 0 and fixed length $2q$, e.g., $b - a = 2q$, and $q \geq \sigma_1$, we have the following result:*

$$\text{Var}(e|\text{censored in } [a, b]) \geq g(2q/\sigma_1, 0)\sigma_1^2. \quad (13)$$

Proof We first derive the variance for two sided censored Gaussian Distribution. Denote $\alpha = a/\sigma_1$ and $\beta = b/\sigma_1$. For the truncated Gaussian distribution, we have

$$\begin{aligned} \mathbb{E}(e|e \in [a, b]) &= \sigma_1 \frac{\phi(\alpha) - \phi(\beta)}{\Phi(\alpha) - \Phi(\beta)} = \sigma_1 \rho. \\ \text{Var}(e|e \in [a, b]) &= \sigma_1^2 \left(1 + \underbrace{\frac{\alpha\phi(\alpha) - \beta\phi(\beta)}{\Phi(\alpha) - \Phi(\beta)}}_{\Lambda} - \rho^2 \right). \end{aligned}$$

Then we calculate the variance of two sided censored Gaussian distribution by

$$\text{Var}(e|\text{censored in } [a, b]) = \mathbb{E}_y[\text{Var}(e|y)] + \text{Var}_y[\mathbb{E}(e|y)],$$

where y denotes the event $e \in [a, b]$. After some basic calculations, we can get the following result:

$$\begin{aligned} \text{Var}(e|\text{censored in } [a, b]) &= \sigma_1^2 (\Phi(\beta) - \Phi(\alpha))(1 + \Lambda) \\ &\quad + \sigma_1^2 [(\rho - \beta)^2 (\Phi(\beta) - \Phi(\alpha))(1 - \Phi(\beta) + \Phi(\alpha)) \\ &\quad + 2(\beta - \alpha)(\rho - \beta)(\Phi(\beta) - \Phi(\alpha))\Phi(\alpha) \\ &\quad + (\beta - \alpha)^2 (1 - \Phi(\alpha))\Phi(\alpha)] \\ &= g(\beta, \alpha)\sigma_1^2. \end{aligned} \quad (14)$$

One can show that (1) $\text{Var}(e|\text{censored in } [a, b])$ achieves minimum when $a = 0$ or $b = 0$ by the first order optimality condition. (2) $\text{Var}(e|\text{censored in } [0, b])$ is an increasing function w.r.t b . Based on (1) and (2), we obtain

$$\text{Var}(e|\text{censored in } [a, b]) \geq \text{Var}(e|\text{censored in } [0, 2q]) = g(2q/\sigma_1, 0)\sigma_1^2, \quad \blacksquare$$

Proof of Lemma 2

Proof At round t , we have

$$\begin{aligned} &\lambda_{\min}(\mathbb{E}(XX^\top)) \\ &= \lambda_{\min}(\mathbb{E}(\sum_{i=1}^t (x_{a_i}^i (x_{a_i}^i)^\top))) = \lambda_{\min}(\sum_{i=1}^t \mathbb{E}(x_{a_i}^i (x_{a_i}^i)^\top)) \\ &\geq \sum_{i=1}^t \lambda_{\min}(\mathbb{E}(x_{a_i}^i (x_{a_i}^i)^\top)), \end{aligned}$$

where the second equality is due to the independence of each round's perturbation and the inequality comes from the fact that minimum eigenvalue is an super-additive operator.

For the censored Gaussian perturbation, $\lambda_{\min}(\mathbb{E}(x_{a_i}^i (x_{a_i}^i)^\top)) \geq g(\frac{2q}{\sigma_1}, 0)\sigma_1^2$ based on Lemma 1. So $\lambda_{\min}(\mathbb{E}(XX^\top)) \geq g(\frac{2q}{\sigma_1}, 0)\sigma_1^2 t$.

Based on (9) of Lemma 4 and $\lambda_{\max}(x_{a_i}^i (x_{a_i}^i)^\top) \leq \|x_{a_i}^i\|_2^2 \leq R^2$, one can obtain

$$\mathbb{P}\left\{\lambda_{\min}(XX^\top) \leq g\left(\frac{2q}{\sigma_1}, 0\right)(1-\tau)\sigma_1^2 t\right\} \leq d \left[e^{-\tau^2/2}\right]^{\frac{g(2R/\sigma_1, 0)\sigma_1^2 t}{R^2}}.$$

Let $\frac{1}{T} = d \left[e^{-\tau^2/2}\right]^{\frac{g(2R/\sigma_1, 0)\sigma_1^2 t}{R^2}}$ and one can get the final result. \blacksquare

Proof of Theorem 1

Proof To simplify the analysis, we slightly abuse the notation and denote the unperturbed context matrix by μ where each column μ_i is one context vector. Similarly, denote e to be the perturbation matrix and e_i to be the column vector. We first decompose the $\Delta^\top XX^\top \Delta$ as follows:

$$\Delta^\top XX^\top \Delta = \underbrace{\Delta^\top \mu \mu^\top \Delta}_{\text{(a)}} + 2 \underbrace{\Delta^\top e \mu^\top \Delta}_{\text{(b)}} + \underbrace{\Delta^\top e e^\top \Delta}_{\text{(c)}}. \quad (15)$$

For the term **(a)** in equation (15), one can only show **(a)** since Δ could lie in $\mathbf{Null}(\mu^\top)$. For term **(b)** and **(c)**, we find both terms high probability lower bounds respectively.

Now consider a positive definite matrix Σ and we can design that Σ such that it satisfies the RE, that is, $\|\Sigma^{1/2}\Delta\|_2 \geq \gamma\|\Delta\|_2$. Based on Lemma 4, we can derive the following for term **(c)**. For universal positive constants c, c', c'' , if the sample size satisfies

$$t > \frac{4c''q(\Sigma)}{\gamma^2} k \log d, \quad (16)$$

where $q(\Sigma) = \max_i \Sigma_{ii}$, then with probability at least $1 - \frac{c'}{e^{ct}}$

$$\Delta^\top e e^\top \Delta \geq \frac{\gamma^2}{64} t \|\Delta\|_2^2. \quad (17)$$

We then derive a high probability bound for **(b)**. First, we decompose **(b)** into a weighted sum of i.i.d. Gaussian variable. That is,

$$\Delta^\top e \mu^\top \Delta = \sum_{i=1}^t (\mu_i^\top \Delta) (\Delta^\top e_i), \quad (18)$$

where $\mu_i^T \Delta$ is the weight and each $\Delta^\top e_i \sim \mathcal{N}(0, \Delta^\top \Sigma \Delta)$. Based on the Chernoff Bound of weighted sum of sub-Gaussian random variables in Fact 2, we have

$$\sum_{i=1}^t (\mu_i^T \Delta) (\Delta^\top e_i) \geq - \sqrt{2a \Delta^\top \Sigma \Delta \sum_{i=1}^t (\mu_i^T \Delta)^2 \log t} \quad (19)$$

$$\geq - \sqrt{2a \lambda_{\max}(\Sigma) \|\Delta\|_2^2 \sum_{i=1}^t R^2 \|\Delta\|_2^2 \log t} \quad (20)$$

$$= -Rt \|\Delta\|_2^2 \sqrt{\frac{2a \lambda_{\max}(\Sigma) \log t}{t}}. \quad (21)$$

with probability at least $1 - \frac{1}{t^a}$. We can conclude with probability at least $1 - (\frac{c'}{e^{ct}} + \frac{1}{t^a})$, both inequality (17) and (21) hold. If the round t satisfies

$$t > \max \left(\underbrace{\frac{4c'' q(\Sigma)}{\gamma^2} k \log d}_{\mathbf{d}}, \underbrace{\frac{8196aR^2 \lambda_{\max}(\Sigma) \log t}{\gamma^4}}_{\mathbf{e}} \right). \quad (22)$$

, we have $(\mathbf{b}) + (\mathbf{c}) \geq ht \|\Delta\|_2^2$, where $h = \left(\frac{\gamma^2}{64} - R \sqrt{\frac{2a \lambda_{\max}(\Sigma) \log t}{t}} \right)$. \blacksquare

Proof of Lemma 3

Proof Our proof combines the techniques from smoothed analysis and Lasso regression. Since θ^t minimizes $G(\theta)$, we have $G(\theta^t) \leq G(\theta^*)$. This yields the following inequality

$$\|X^\top \Delta^t\|_2^2 \leq \Delta^t X \eta + \lambda^t (\|\theta^*\|_1 - \|\theta^* + \Delta^t\|_1),$$

where η denotes the noise vector. Note that $\|\theta^*\|_1 = \|\theta_S^*\|_1$. Furthermore, one can verify that $\|\theta^*\|_1 - \|\theta^* + \Delta^t\|_1 \leq \|\Delta_S^t\|_1 - \|\Delta_{S^c}^t\|_1$. For $\Delta^t X \eta$, applying Hölder's inequality yields

$$\Delta^t X \eta \leq \|\Delta^t\|_1 \|X \eta\|_\infty \leq \sigma R \sqrt{2t \log \frac{2d}{\delta}} \|\Delta^t\|_1 = \frac{\lambda^t}{2} \|\Delta^t\|_1,$$

where the second inequality is due to the fact 1. Combine all above and we obtain

$$\|X^\top \Delta^t\|_2^2 \leq \frac{\lambda^t}{2} \|\Delta^t\|_1 + \lambda^t (\|\Delta_S^t\|_1 - \|\Delta_{S^c}^t\|_1) \quad (23)$$

$$\leq \frac{3}{2} \lambda^t \|\Delta^t\|_1 \leq \frac{3}{2} \lambda^t \sqrt{k} \|\Delta^t\|_2. \quad (24)$$

First from inequality (23), we can obtain $\Delta^t \in \mathcal{C}(S; 3)$. For low dimensional case, we have $\|X^\top \Delta^t\|_2^2 \geq \lambda_{\min}(XX^\top) \|\Delta^t\|_2^2 \geq Ct \|\Delta^t\|_2^2$ by Lemma 2, where $C = g\left(\frac{2q}{\sigma_1}, 0\right) (1 - \tau) \sigma_1^2$. For high dimensional case, we apply Theorem 1 since $\Delta^t \in \mathcal{C}(S; 3)$ and get $\|X^\top \Delta^t\|_2^2 \geq Ct \|\Delta^t\|_2^2$

where $C = \frac{\gamma^2}{64} - R\sqrt{\frac{2a\lambda_{\max}(\Sigma)\log T}{t}}$. Combine these with inequality (24) and we get the final result

$$\|\Delta^t\|_2 \leq \frac{3\sigma R}{C} \sqrt{\frac{2k \log 2d/\delta}{t}}.$$

■

Proof of Theorem 2

Proof As for the regret in round t , we have

$$\begin{aligned} & \langle x_{i_t^*}^t, \theta^* \rangle - \langle x_{i_t}^t, \theta^* \rangle \\ &= \langle x_{i_t^*}^t, \theta^* - \theta^t \rangle - \langle x_{i_t}^t, \theta^* - \theta^t \rangle + \langle x_{i_t^*}^t, \theta^t \rangle - \langle x_{i_t}^t, \theta^t \rangle \\ &\leq \langle x_{i_t^*}^t, \theta^* - \theta^t \rangle - \langle x_{i_t}^t, \theta^* - \theta^t \rangle \\ &\leq \|\langle x_{i_t^*}^t, \theta^* - \theta^t \rangle\|_2 + \|\langle x_{i_t}^t, \theta^* - \theta^t \rangle\|_2 \\ &\leq 2R\|\theta^* - \theta^t\|_2, \end{aligned}$$

where the first inequality comes from the greedy choice since $i_t = \arg \max_i \langle x_i^t, \theta^t \rangle$ and the last inequality is due to the censored perturbations. Based on the analysis of low and high dimensional cases, we denote the exploration length as T_e . During the exploration, we can bound the regret by $2RT_e$. So we can derive that

$$\begin{aligned} \text{Regret} &= \sum_{t=1}^{T_e} \langle x_{i_t^*}^t, \theta^* \rangle - \langle x_{i_t}^t, \theta^* \rangle + \sum_{t=T_e+1}^T \langle x_{i_t^*}^t, \theta^* \rangle - \langle x_{i_t}^t, \theta^* \rangle \\ &\leq 2RT_e + 2R \sum_{t=T_e+1}^T \|\theta^* - \theta^t\|_2 \\ &\leq 2RT_e + 2R \sum_{t=T_e+1}^T \frac{3\sigma R}{C} \sqrt{\frac{2k \log 2d/\delta}{t}} \\ &\leq 2R \left(T_e + \frac{6\sigma R}{C} \sqrt{2kT \frac{\log 2d}{\delta}} \right) \end{aligned}$$

■

Numeric Simulations

This section shows the result of numeric simulations. We choose the context's dimension $d = 2000$ where effective dimension $k = 20$ and 5 arms for each round. Our sparse bandit learning process only contains 150 rounds with each context vector are randomly generated from the uniform distribution $[0, 1]$. Each experiment are repeated 10 times to reduce the

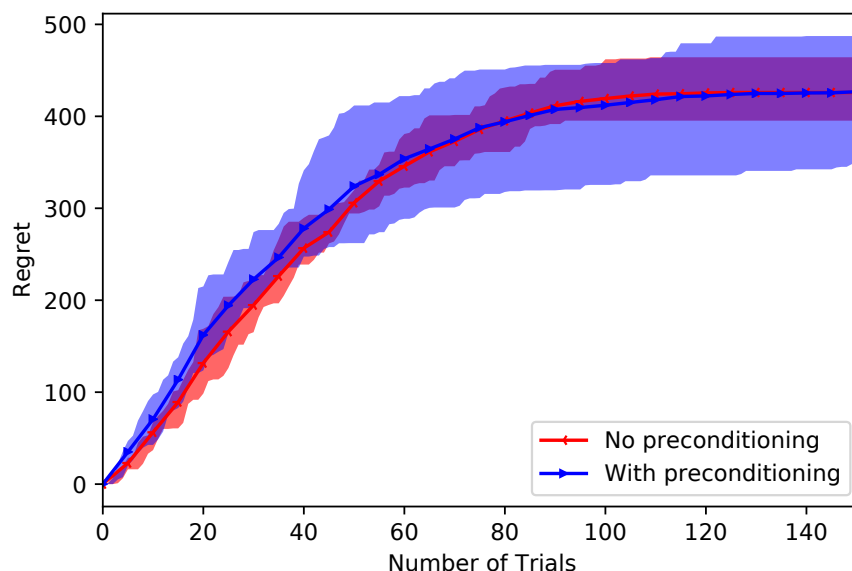


Figure 1: Regret of preconditioning and no preconditioning setting with perturbation variance 0.1.

effect of the other unnecessary factors. Solid line denotes the average performance and the shadow area contains best and worst performance during repeated running. We first compare the preconditioning via SVD from the algorithm of Sivakumar et al. (2020) which transfers all non-zero singular eigenvalues to 1. Figure 1 shows the regret results with and without preconditioning. In our experiments, we find that the average performance of preconditioning is almost the same as the one without preconditioning (see solid line). Moreover, the performance of preconditioning shows more unstable (see the shadow area). Also, for $d = 2000$, preconditioning heavily slows down the learning process. The reason could be the noise amplification incurred by preconditioning where Jia et al. (2015) shows (1) the preconditioned noise are no longer i.i.d. (2) preconditioning can amplify the noise.

We then investigate the performance under different perturbation variance. The result in Figure 2 shows the regret will first decrease then increase which is expected by our analysis. The first decreasing phase is because that the perturbation brings good property to the context matrix. When the perturbation variance becomes large, the context’s variance also becomes large which leads to more explorations under the uncertainty environment.

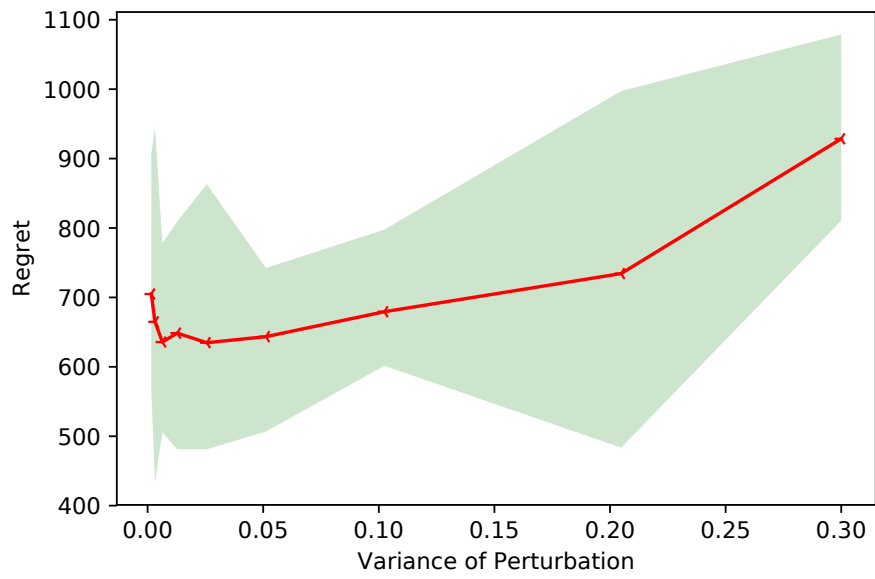


Figure 2: Regret under different perturbation variance.