Structured Linear Contextual Bandits: A Sharp and Geometric Smoothed Analysis

Vidyashankar Sivakumar [1 2] **Zhiwei Steven Wu** [2] **Arindam Banerjee** [2]

Abstract

Bandit learning algorithms typically involve the balance of exploration and exploitation. However, in many practical applications, worst-case scenarios needing systematic exploration are seldom encountered. In this work, we consider a smoothed setting for structured linear contextual bandits where the adversarial contexts are perturbed by Gaussian noise and the unknown parameter θ^* has structure, e.g., sparsity, group sparsity, low rank, etc. We propose simple greedy algorithms for both the single- and multi-parameter (i.e., different parameter for each context) settings and provide a unified regret analysis for θ^* with any assumed structure. The regret bounds are expressed in terms of geometric quantities such as Gaussian widths associated with the structure of θ^*. We also obtain sharper regret bounds compared to earlier work for the unstructured θ^* setting as a consequence of our improved analysis. We show there is implicit exploration in the smoothed setting where a simple greedy algorithm works.

1. Introduction

Contextual bandits (Langford & Zhang, 2007) is a powerful framework for sequential decision-making, with many applications to clinical trials, web search, and content optimization. In a typical scenario, users arrive over time, and the algorithm chooses among various content (e.g., news articles) to present to each user and observes the outcome (e.g., clicks). A popular parametric formulation for this problem is the linear contextual bandit setting (Chu et al., 2011; Li et al., 2010): in rounds $t = 1, \ldots, T$, the algorithm selects a context $x_{i^t}^t$ from k available contexts $x_1^t, \ldots, x_k^t$ and receives a noisy reward $r^t(x_{i^t}^t) = \langle x_{i^t}^t, \theta^* \rangle + \omega^t$ where θ^* is the unknown parameter and ω^t denotes the reward noise. The goal of the algorithm is to select contexts to maximize rewards over time observing only the available contexts and the reward associated with the selected context in each round. Such algorithms typically need to balance *exploration*, making potentially sub-optimal decisions for the sake of information acquisition, and *exploitation*, selecting decisions that are optimal based on the estimate of θ^*. In particular, the greedy algorithm that myopically selects contexts maximizing rewards based on the ordinary least squares estimate $\hat{\theta}$, i.e., choosing $x_{i^t}^t = \underset{x_i^t : 1 \leq i \leq k}{\operatorname{argmax}} \langle x_i^t, \hat{\theta} \rangle$ is known to be sub-optimal in the worst case (see Mansour et al. (2018) for an example). However, the greedy algorithm offers several appealing features, including its simplicity in computation and the selection is aligned with each user's short-term interest (Bird et al., 2016; Kannan et al., 2017).

Given the advantages of the greedy algorithm, there has been recent work that investigates when the greedy algorithms perform well. On the practical side, Bietti et al. (2018) shows that there is strong empirical evidence that exploration-free algorithms perform well on real data sets. On the theoretical side, a line of work (Bastani et al., 2018; Kannan et al., 2018; Raghavan et al., 2018) analyzed conditions under which inherent diversity in the data makes explicit exploration unnecessary. In particular, the work of (Kannan et al., 2018; Raghavan et al., 2018) provide a *smoothed analysis* on the greedy algorithm under the following setting: in each round the contexts $x_i^t, 1 \leq i \leq k$ are of the form $\mu_i^t + g_i^t$, where the $\mu_i^t \in \mathbb{R}^p$'s are chosen adaptively by an adversary and $g_i^t \sim N(0, \sigma^2 \mathbb{I}_{p \times p})$ are independent Gaussian perturbations from nature. The expected reward associated with each x_i^t is then $\langle x_i^t, \theta_i^* \rangle$, where the unknown parameter θ_i^* can vary across different actions i.

Our work substantially generalizes the smoothed analysis framework for linear contextual bandits considered in Kannan et al. (2018); Raghavan et al. (2018). We enrich and refine these prior analyses by explicitly capturing the structure in the unknown parameters—specifically, θ_i^* with small atomic norms $R(\cdot)$ such as ℓ_1 norm, group-sparse norms, nuclear norms, k-support norm, etc. (Jacob et al.,

[1] Walmart Labs, Sunnyvale [2] Department of Computer Science, University of Minnesota, Twin Cities. Correspondence to: Vidyashankar Sivakumar <sivak017@umn.edu>, Zhiwei Steven Wu <steven7woo@gmail.com>, Arindam Banerjee <banerjee@cs.umn.edu>.

Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020. Copyright 2020 by the author(s).

2009; Argyriou et al., 2012; Yuan & Lin, 2006; Tibshirani, 1996; Candès & Recht, 2009). We consider two variants of the problem: the multi parameter setting when there is a separate parameter corresponding to each context, i.e., $\theta_1^*, \ldots, \theta_k^*$ and the single parameter setting when there is a single unknown parameter, i.e., $\theta^* = \theta_1^* = \theta_2^* = \ldots = \theta_k^*$. In any round t the greedy algorithm maintains estimates of the true parameters $\hat{\theta}_1^t, \ldots, \hat{\theta}_k^t$:

$$\hat{\theta}_i^t = \underset{\theta \in \mathbb{R}^p}{\operatorname{argmin}} \quad \mathcal{L}(\theta; Z_i^t, y_i^t) \quad \text{s.t.} \quad R(\theta) \leq R(\theta_i^*) \ , \tag{1}$$

where $\mathcal{L}(\theta; Z_i^t, y_i^t)$ is the least squares loss, Z_i^t is the design matrix in round t whose rows are contexts chosen in rounds prior to t and y_i^t is a vector with the corresponding rewards. The greedy algorithm then selects the context corresponding to the highest reward w.r.t. to the current parameter estimate, i.e., $x_{i^t}^t = \underset{x_i^t : 1 \leq i \leq k}{\operatorname{argmax}} \langle x_i^t, \hat{\theta}_i^t \rangle$. We analyze the regret of the greedy algorithm—the difference between the cumulative expected rewards of a clairvoyant learner with knowledge of θ_i^* and the greedy algorithm,

$$\operatorname{Reg}(T) = \sum_{t=1}^{T} \left(\max_i \langle x_i^t, \theta_i^* \rangle - \langle x_{i^t}^t, \theta_{i^t}^* \rangle \right) . \tag{2}$$

Consider first the single parameter problem setting. In any round t, denote the error vector $\Delta^t = \hat{\theta}^t - \theta^*$. It is evident from equation (1) that the error vector lies in the error set $E_c = \{\Delta \mid R(\theta^* + \Delta) \leq R(\theta^*)\}$. Now consider the error set $A = \operatorname{cone}(E_c) \cap S^{p-1}$, where S^{p-1} is the unit sphere, and its Gaussian width $w(A)$—a metric for the complexity/size of a set widely used in literature on analysis of high-dimensional statistical models (Talagrand, 2005; 2014; Gordon, 1985; Banerjee et al., 2014; Chen & Banerjee, 2015; Chandrasekaran et al., 2012; Sivakumar et al., 2015). For example, Gaussian width of the error set for $R(\cdot) = \|\cdot\|_1$ and s-sparse θ^* is $\Theta(s \log p)$. We show that the single parameter setting has a warm start phase of $t_{\min} = \tilde{\Theta}(w^2(A))$ rounds when the greedy algorithm accrues linear regret. The contexts chosen in the warm start phase are random and serve to satisfy the Restricted Eigenvalue (RE) condition, $\inf_{u \in A} \|Z^t u\|_2^2 \geq \kappa$, for some positive constant κ over all rounds. In the unconstrained setting $A = S^{p-1}$, and the RE condition essentially reduces to the condition that the empirical covariance matrix is not rank-deficient when the length of the warm start phase is $\tilde{\Theta}(p)$. After the warm start phase, the greedy algorithm has regret bounded as follows:

$$\operatorname{Reg}(T) = \tilde{O}\left(\frac{w(A)\sqrt{T}}{\sigma} \right) , \tag{3}$$

where σ^2 is the variance of the Gaussian perturbations on the contexts.

We make the following observations:

1. For the unconstrained problem, $w(A) = \Theta(\sqrt{p})$ and $\operatorname{Reg}(T) = \tilde{O}\left(\frac{\sqrt{pT}}{\sigma}\right)$. When $\sigma^2 = \Theta\left(\frac{1}{p}\right)$ as considered in (Kannan et al., 2018), ignoring logarithmic factors, the regret bounds are sharper compared to the results in (Kannan et al., 2018) by a factor $\sqrt{p}$. Moreover when $\sigma^2 = \Theta\left(\frac{1}{p}\right)$, the regret rate is of the same order as those of UCB-style algorithms and Thompson sampling in (Dani et al., 2008; Abbasi-Yadkori et al., 2011; Agarwal & Goyal, 2013) for stochastic linear bandits. With more smoothing when $\sigma^2 = \Omega\left(\frac{1}{p}\right)$ the greedy algorithm performs better with lower regret whereas less smoothing has the reverse effect.
2. For $R(\cdot) = \|\cdot\|_1$ and s-sparse θ^*, $w(A) = \Theta(\sqrt{s \log p})$ leading to the regret bounds, $\operatorname{Reg}(T) = \tilde{O}\left(\frac{\sqrt{s \log p \cdot T}}{\sigma}\right)$. Again when $\sigma^2 = \Theta\left(\frac{1}{p}\right)$, the regret rate is of the same order as that of a ℓ_1 regularized UCB method for the stochastic linear bandits problem (Abbasi-Yadkori et al., 2012). Note that the algorithm proposed in (Abbasi-Yadkori et al., 2012), in contrast to the greedy algorithm, is computationally involved.
3. Our analysis is generalized for any atomic norm $R(\cdot)$ and captures the geometry of the problem obtaining results in terms of easily computable geometric quantities like the Gaussian width (Talagrand, 2005; 2014; Gordon, 1985; Chen & Banerjee, 2015)

The multi parameter setting requires a warm start phase of $\tilde{\Theta}\left(\frac{kw^2(A)}{\sigma^4}\right)$, where k is the number of contexts, when the contexts are chosen at random or in a round robin fashion. In contrast to the single parameter setting, the warm start phase in the multi parameter setting is required to satisfy a margin condition, which we detail in Section 4. The margin condition is required for the algorithm to achieve sublinear regret. When $\sigma^2 = \Theta\left(\frac{1}{p}\right)$, in the worst case, we require the length of the warm start phase to be $\tilde{\Theta}(k \cdot p^2 \cdot w^2(A))$ when we accrue linear regret. In the unstructured setting, $w^2(A) = p$ which translates to $\tilde{\Theta}(kp^3)$ rounds in the warm start phase which improves over the $\tilde{\Theta}(kp^6)$ rounds in (Kannan et al., 2018) (see Theorem 4.2). The algorithm has regret bounded by $\tilde{O}\left(\frac{w(A)\sqrt{Tk}}{\sigma}\right)$ regret after the warm start rounds which is $\sqrt{k}$ times worse compared to the single parameter setting.

We briefly summarize the organization and notations used throughout the paper. We concisely present the main ideas and technical results in Section 2 of the paper. Results for the single parameter and multi parameter settings are presented in Section 3 and 4 respectively before concluding in Section 5.

Notation. Throughout the paper we use constants like

$c, c_1, c_2, \ldots$ whose definition may change from one line to the next. In certain places, we use the terms contexts and arms interchangeably. The notations $y = \Theta(x)$ (respectively $y = O(x)$, $y = \Omega(x)$) implies there exists absolute constants c_1, c_2, c_3, c_4 such that $c_1 \cdot x \leq y \leq c_2 \cdot x$ (respectively $y \leq c_3 \cdot x$, $y \geq c_4 \cdot x$) and $\tilde{\Theta}(\cdot)$, $\tilde{\Omega}(\cdot)$ and $\tilde{O}(\cdot)$ notations hide the dependence on polylogarithmic factors and noise variance.

2. Overview of Main Technical Results

We now summarize the major ideas and results in the paper.

Episodic algorithm. We analyze a greedy algorithm that is *episodic*—a common feature that can reduce computation and simplify statistical analyses (Javanmard & Javadi, 2018). Let T denote the total number of rounds. In the single parameter setting, denote the episode number by e and let T_e denote the total number of rounds in episode e. The number of rounds in each episode increases geometrically with time, i.e., $T_1 = 2T_0$, $T_2 = 2T_1$ and so on. The total number of rounds $T = \sum_e T_e$. The number of episodes scales as $\log T$. The regression parameter is estimated at the beginning of episode $e + 1$ using only the contexts and rewards observed in the T_e rounds in the immediately preceding episode. In the multi parameter setting, the only difference to the single parameter setting is that we maintain separate design matrices, rewards, parameter estimates and episodes for each context.

Estimation error. The regret bounds in both the single and multi parameter settings depend on the estimation error for the parameter estimated using the constrained least squares estimator at the beginning of each episode. Consider parameter estimation in episode $e + 1$. Let $Z^{(e)} \in \mathbb{R}^{T_e \times p}$ be the design matrix constructed with rows as contexts observed in episode e and $y^{(e)} \in \mathbb{R}^{T_e}$ the corresponding observed rewards. We precondition the data before parameter estimation using the Puffer transformation (Jia & Rohe, 2015). The Puffer transformation computes the SVD of the design matrix as $\frac{1}{\sqrt{T_e}} Z^{(e)} = U^{(e)} D^{(e)} (V^{(e)})^\intercal$ followed by transforming the data as $\tilde{Z}^{(e)} = F^{(e)} Z^{(e)}$, $\tilde{y}^{(e)} = F^{(e)} y^{(e)}$ where $F^{(e)} = U^{(e)} (D^{(e)})^{-1} (U^{(e)})^\intercal$. The parameter at the beginning of episode $e + 1$ is then estimated using the following least squares constrained estimator:

$$\hat{\theta}^{(e+1)} = \underset{\theta \in \mathbb{R}^p}{\text{argmin}} \frac{1}{T_e} \|\tilde{y}^{(e)} - \tilde{Z}^{(e)} \theta\|_2^2 \quad s.t. \quad R(\theta) \leq R(\theta^*) \, . \tag{4}$$

The Puffer transformation is a data preconditioning technique which makes the condition number of the design matrix unity. In the worst case over any choice of the adaptive adversaries in all rounds, Puffer transformation leads to better estimation bounds compared to the bounds obtained using raw data (Chandrasekaran et al., 2012; Negahban et al., 2012; Banerjee et al., 2014). Our analysis borrows tools and techniques from the existing vast literature on high-dimensional estimation (Wainwright, 2019; Vershynin, 2018). Specifically, following the analysis framework in (Banerjee et al., 2014), we need three main results. First, note that to satisfy the constraint in (4) the error vector Δ with $\hat{\theta}^{(e+1)} = \theta^* + \Delta$ lies in the following set,

$$E_c = \{\Delta \mid R(\theta^* + \Delta) \leq R(\theta^*)\} \, . \tag{5}$$

Second, for consistent estimation we show the design matrix satisfies the following restricted eigenvalue (RE) condition on the error set $A = \text{cone}(E_c) \cap S^{p-1}$ (Bickel et al., 2009; Negahban et al., 2012) with high probability across all episodes once $T > t_{\min} = \tilde{\Theta}(w^2(A))$,

$$\inf_{u \in A} \frac{1}{T_e} \|\tilde{Z}^{(e)} u\|_2^2 = \tilde{\Omega}(\sigma^2) \, . \tag{6}$$

Existing results on the RE condition (Mendelson et al., 2007; Banerjee et al., 2014; Negahban et al., 2012) with i.i.d. rows cannot be directly applied since the rows in the design matrix depend on previously selected contexts and rewards. We make use of recent novel results in (Banerjee et al., 2019) on bounds for sum of random quadratic quantities with dependence. Third, for rounds $T > t_{\min}$ we obtain high probability upper bounds on the estimation error with the Puffer transformed data across all episodes.

$$\max_e \|\hat{\theta}^{(e+1)} - \theta^*\|_2 \leq \tilde{O}\left(\frac{w(A)}{\sigma \sqrt{T_e}}\right) \, . \tag{7}$$

The non-asymptotic bounds on the estimation error are novel, due to new analysis techniques to handle data dependence between rounds and the Puffer transformation for which no results exist for estimation error to the best of our knowledge. The results on parameter estimation errors also holds in the multi parameter setting.

Regret. For both the single and multi parameter settings we show the regret depends on the ℓ_2 norm of the estimation error for the parameter estimated at the beginning of each episode after an initial warm start phase when the algorithm accrues linear regret. In the single parameter setting the length of the warm start phase is $t_{\min} = \tilde{\Theta}\left(w^2(A)\right)$ rounds while in the multi parameter setting it is $t_{\min} = \tilde{\Theta}\left(\frac{k w^2(A)}{\sigma^4}\right)$ rounds. The dependence of $t_{\min}$ on σ for the multi parameter setting implies a large warm start phase when σ is small. For example, if $\sigma^2 = \Theta\left(\frac{1}{p}\right)$ as assumed in (Kannan et al., 2018), then $t_{\min}$ scales as p^2 which may become prohibitive in many high-dimensional applications. After the warm start phase we show the regret in the single parameter setting is upper bounded as follows:

$$\text{Reg}(T) = \tilde{O}\left(\frac{w(A)\sqrt{T}}{\sigma}\right) \, , \tag{8}$$

The upper bound on the regret in the multi parameter setting after the warm start phase is worse compared to the single parameter setting by a factor of $\sqrt{k}$:

$$\text{Reg}(T) = \tilde{O}\left(\frac{w(A)\sqrt{kT}}{\sigma}\right) . \quad (9)$$

3. Single Parameter Regret Analysis

We present results for the single parameter setting in this section. The greedy algorithm proceeds in multiple episodes with the length of each episode increasing geometrically with time. We index episode numbers by e, time steps by t and contexts by i. We denote by T the total number of rounds and by T_e the number of rounds in episode e. In each round, the algorithm observes contexts $x_i^t, 1 \leq i \leq k$ and greedily selects the optimal context based on the current parameter estimate, i.e., $z^t = \underset{x_i^t : 1 \leq i \leq k}{\operatorname{argmax}} \langle x_i^t, \hat{\theta}^{(e)} \rangle$ and receives noisy reward $y^t = \langle z^t, \theta^* \rangle + \omega^t$ with ω^t denoting κ_ω sub-Gaussian noise at time t. The parameter is estimated at the beginning of each episode by running constrained least squares regression on the observed data from the previous epoch (with the Puffer transformation on the design matrix and response). Note that the design matrix is rank deficient in the first $e = \lceil \log t_{\min} \rceil$ rounds with $t_{\min} = \tilde{\Theta}(w^2(A))$ when the contexts will be chosen uniformly at random.

Lemma 1 gives an upper bound for the regret for Algorithm 1. The greedy algorithm accrues linear regret in the first $t_{\min}$ rounds when the design matrix is rank deficient for parameter estimation, i.e., it does not satisfy the restricted eigenvalue condition. Subsequent rounds are played in an episodic fashion with the regret in any round depending on the parameter estimation error at each episode.

Lemma 1 (Lemma 3.1 in (Kannan et al., 2018)) *Denote by $\beta = \max_{1 \leq i \leq k, 1 \leq t \leq T} \|x_i^t\|_2$. Assume $T > t_{\min}$, where $t_{\min}$ depends on properties of the true parameter θ^* and the regularizer $R(\cdot)$. Then,*

$$\mathit{Reg}(T) \leq 4\beta t_{\min} + \sum_{e=\lceil \log t_{\min} \rceil}^{\lfloor \log T \rfloor} 2\beta T_e \|\hat{\theta}^{(e)} - \theta^*\|_2 . \quad (11)$$

3.1. Gaussian Contexts

In order to build intuition, we establish results on performance of the greedy algorithm when the contexts are completely stochastic, i.e., we derive regret bounds when the contexts are sampled independently from a Gaussian distribution, , $x_i^t \sim N(0, \sigma^2 \mathbb{I}_{p \times p}), 1 \leq i \leq k, t \leq T$ in step 8 of Algorithm 1. The episodic algorithm ensures independence between data in each round of an episode. Additionally, the rows of the design matrix are sub-Gaussian and the covariance matrix satisfies the minimum eigenvalue condition.

Algorithm 1 Structured Greedy (single parameter)

```
1: Initialize empty design matrix and reward vector
   $Z^{(0)} = [], y^{(0)} = []$
2: for $e = 1, 2, 3, \ldots, \lfloor \log_2 T \rfloor$ do
3:   SVD: $\frac{1}{\sqrt{T_{e-1}}} Z^{(e-1)} = U^{(e-1)} D^{(e-1)} (V^{(e-1)})^\intercal$
4:   Puffer transformation:
     $F^{(e-1)} = U^{(e-1)} (D^{(e-1)})^{-1} (U^{(e-1)})^\intercal$
     $\tilde{Z}^{(e-1)} = F^{(e-1)} Z^{(e-1)}$
     $\tilde{y}^{(e-1)} = F^{(e-1)} y^{(e-1)}$
5:   Estimate parameter using constrained least squares
     estimator breaking ties arbitrarily when necessary

     $\hat{\theta}^{(e)} = \underset{\theta \in \mathbb{R}^p}{\operatorname{argmin}} \frac{1}{2T_{e-1}} \|\tilde{y}^{(e-1)} - \tilde{Z}^{(e-1)} \theta\|_2^2$
     $\text{s.t.} \quad R(\theta) \leq R(\theta^*) , \quad (10)$

     where $T_{e-1}$ is the number of observations in the previous episode.
6:   Initialize empty design matrix and reward vector
     $Z^{(e)} = [], y^{(e)} = []$. Set $T_e = 2^{e-1}$
7:   for $t = 2^{(e-1)} + 1$ to $2^e$ do
8:     Observe contexts $x_1^t, \ldots, x_k^t \in \mathbb{R}^p$
9:     Choose context $z^t = \underset{x_i^t : 1 \leq i \leq k}{\operatorname{argmax}} \langle x_i^t, \hat{\theta}^{(e)} \rangle$ and observe reward $y^t = \langle z^t, \theta^* \rangle + \omega^t$ where $\omega^t$ is zero
       mean $\kappa_\omega$-sub-Gaussian noise
10:    Append observations $(z^t, y^t)$ to $(Z^{(e)}, y^{(e)})$
11:  end for
12: end for
```

Lemma 2 (Single Parameter Gaussian Contexts Design Matrix Properties) *The rows of the design matrix $Z^{(e)} \in \mathbb{R}^{T_e \times p}$ in any episode e satisfy $\kappa_z = \|z^t\|_{\psi_2} \leq c_2 \sigma \sqrt{\log k}$ for c_2 some positive constant. Moreover the minimum eigenvalue of the matrix $E_{z^t}[z^t (z^T)^T]$ satisfies,*

$$\lambda_{\min}(E_{z^t}[z^t (z^t)^\intercal]) \geq c_1 \frac{\sigma^2}{\log k} , \quad (12)$$

where c_1 is some positive constant and the expectation is over the chosen contexts.

The result of Lemma 2 allows us to use existing results (Banerjee et al., 2014; Negahban et al., 2012) to establish the RE condition and estimation error bounds. The only deviation from traditional estimation is the use of the Puffer transformation (Jia & Rohe, 2015). The Puffer transformation is a data preconditioning technique which makes the condition number of the design matrix unity. We obtain the following worst case upper bound on the ℓ_2 norm of the estimation error with high probability with the Puffer transformed data:

$$\|\hat{\theta}^{(e+1)} - \theta^*\|_2 \leq \tilde{O}\left(\frac{w(A)}{\sigma\sqrt{T_e}}\right) , \tag{13}$$

where A is the error set. We provide the proof in the supplement which essentially uses the same analysis tools and techniques from (Banerjee et al., 2014). The regret bounds now follow from a straightforward application of the result of Lemma 1. When $\sigma^2 = \Theta\left(\frac{1}{p}\right)$, as assumed in (Kannan et al., 2018), the regret bound is $\tilde{O}(w(A)\sqrt{pT})$.

Theorem 1 (Gaussian Contexts Regret Bounds) *Consider Gaussian contexts. Then with probability atleast* $1-2\delta$

$$\beta = \max_{1\leq i\leq k, 1\leq t\leq T}\|x_i^t\|_2 \leq c_1\sigma(\sqrt{p} + \sqrt{\log(1/\delta)}) \tag{14}$$

Assuming $t_{\min} = O(\sqrt{T})$ *with probability atleast* $1 - 4\delta$ *the following is an upper bound on the regret for the Greedy algorithm,*

$$Reg(T) \leq O\left(\frac{\gamma \cdot \beta \cdot \log(T) \cdot \sqrt{T}}{\sigma}\right) \tag{15}$$

where $\gamma = c\kappa_\omega\sqrt{\log k}(w(A) + \sqrt{\log\log T} + \sqrt{\log(1/\delta)})$

3.2. Smoothed Perturbed Adversary

We now focus on regret bounds when the contexts are $x_i^t = \mu_i^t + g_i^t, 1 \leq i \leq k, \forall 1 \leq t \leq T$. Remember that an adaptive adversary with access to previous observed contexts and rewards can choose $\mu_i^t, \|\mu_i^t\|_2 = 1, \forall 1 \leq i \leq k$. The primary question is if an adversary can negatively influence the design matrix to affect estimation error, or in other words lower the minimum eigenvalue compared to the completely stochastic setting. The answer is in the result of Lemma 3, where we show that even in the adverserial setting the minimum eigenvalue of the covariance matrix is no worse than the completely stochastic Gaussian setting. In particular, adding small random perturbations to adverserially selected contexts leads to implicit exploration where the greedy algorithm works well.

Lemma 3 (Design matrix properties for smoothed adversary) *The rows of the design matrix* $Z^{(e)} \in \mathbb{R}^{T_e\times p}$ *in any episode* e *are* $z^t = \mu^t + g^t$ *where* $\mu^t, g^t = \underset{\mu_i^t, g_i^t : 1\leq i\leq k}{\operatorname{argmax}} \langle \mu_i^t + g_i^t, \hat{\theta}^{(e-1)}\rangle$, $g_i^t \sim N(0, \sigma^2\mathbb{I}_{p\times p})$ *with the sub-Gaussian norm of* g^t *satisfying* $\|g^t\|_{\psi_2} \leq c_2\sigma\sqrt{\log k}$ *for some constant* c_2. *Moreover we have the following lower bound on the expected minimum eigenvalue for any* μ_i^t*'s:*

$$\lambda_{\min}(E_{z^t}[z^t(z^t)^\intercal]) \geq c_1\frac{\sigma^2}{\log k} , \tag{16}$$

where c_1 *is some constant.*

Due to an adaptive adversary, the selected contexts and noise are no longer independent The dependency introduces additional complexity for analysis of the non-asymptotic estimation error. To obtain results on the RE condition, we make use of recent novel results from (Banerjee et al., 2019) on lower bounds for sum of quadratics of random variables with dependence. We also use arguments from generic chaining (Talagrand, 2005; 2014) to obtain estimation error rates with Puffer preconditioned data. The estimation error is the same as if the contexts were completely stochastic Gaussian without any adversary.

$$\|\hat{\theta}^{(e+1)} - \theta^*\|_2 \leq \tilde{O}\left(\frac{w(A)}{\sigma\sqrt{T_e}}\right) . \tag{17}$$

High probability regret bounds can now be obtained from the result of Lemma 1.

Theorem 2 (Smoothed Adversary Regret Bounds) *In the smoothed adversary setting with probability atleast* $1 - 2\delta$

$$\beta = \max_{1\leq i\leq k, 1\leq t\leq T}\|x_i^t\|_2 \leq (1 + c_1\sigma(\sqrt{p} + \sqrt{\log(1/\delta)})) . \tag{18}$$

Assuming $t_{\min} = O(\sqrt{T})$ *with probability atleast* $1 - 5\delta$ *the following is an upper bound on the regret,*

$$Reg(T) \leq O\left(\frac{\gamma \cdot \beta \cdot \log(T) \cdot \sqrt{T}}{\sigma}\right) , \tag{19}$$

where $\gamma = c\kappa_\omega\sqrt{\log k}(w(A) + \sqrt{\log\log T} + \sqrt{\log(1/\delta)})$.

3.3. Examples

We instantiate the regret bounds for a few norms with the mild assumption $\sigma^2 = \Theta\left(\frac{1}{p}\right)$. Note that for ℓ_2^2 regularization the setting is similar to (Kannan et al., 2018). The regret bounds are better than (Kannan et al., 2018) by a factor of $\sqrt{p}$. If θ^* is sparse, e.g. using the ℓ_1 norm, the regret bounds scale with $\sqrt{s\log p}$ instead of $\sqrt{p}$.

Corollary 1 *Consider the smoothed adversary setting. Let* $\sigma^2 = \Theta\left(\frac{1}{p}\right)$. *Then with probability atleast* $1 - 5\delta$*:*

1. *Let* θ^* *be* s*-sparse,* $R(\cdot)$ *the* ℓ_1 *norm.*

$$Reg(T) = \tilde{O}\left(\sqrt{s\log p}\sqrt{pT}\right) . \tag{20}$$

2. *Let* $\theta^* \in \mathbb{R}^{m\times q}$ *be a rank* r *matrix* $r \leq \min\{m, q\}$, $R(\cdot)$ *is the nuclear norm.*

$$Reg(T) = \tilde{O}\left(q\sqrt{r(m+q)}\sqrt{T}\right) . \tag{21}$$

3. *Let* $R(\cdot)$ *the* ℓ_2^2 *norm.*

$$Reg(T) = \tilde{O}\left(p\sqrt{T}\right) . \tag{22}$$

4. Multi Parameter Regret Analysis

We present results for the multi parameter setting in this section. The multi parameter setting has a separate parameter corresponding to each context. The algorithm requires a warm start phase of T_0 rounds where the contexts are chosen in a round robin fashion or randomly before employing the greedy algorithm. As we show later, the length of the warm start phase has dependence on the variance of the Gaussian perturbations and is required to obtain sublinear regret. Similar to the single parameter setting, after the warm start phase the greedy algorithm proceeds in an episodic fashion, except that we now maintain separate episodes for each context. Denote the episode numbers for context i by e_i and the maximum number of episodes for context i after T rounds as $e_{i,\max}$. In episode e_i, context i is chosen by the greedy algorithm T_{i,e_i} times. During episode e_i, before context i is chosen in T_{i,e_i} rounds by the greedy algorithm, there can also be rounds when context i was optimal but was not chosen by the algorithm, i.e., $x_i^t = \underset{x_j^t:1\leq j\leq k}{\operatorname{argmax}} \langle x_j^t, \theta_j^* \rangle$ but $x_i^t \neq \underset{x_j^t:1\leq j\leq k}{\operatorname{argmax}} \langle x_j^t, \hat{\theta}_j^{(e_j)} \rangle$. We denote the number of rounds this happens in episode e_i by T^*_{i,e_i}.

Lemma 4 below gives an upper bound for the regret for Algorithm 2.

Lemma 4 (Lemma 4.1 in (Kannan et al., 2018)) *The greedy algorithm plays the contexts in an episodic fashion with the maximum episode number for each context $e_i \leq e_{i,\max} \leq \lfloor \log T \rfloor$. Denote by $\beta = \max_{1\leq i\leq k, 1\leq t\leq T} \|x_i^t\|_2$. Let $t_{\min} < T$, where $t_{\min}$ depends on properties of the true parameters θ_i^*, the regularizer $R(\cdot)$, the noise properties, the number of contexts k and the quantity β. Then,*

$$Reg(T) \leq 2\beta t_{\min} + \\ + \beta \sum_{i=1}^{k} \sum_{e_i=1}^{e_{i,\max}} \left(T_{i,e_i} \|\theta_i^* - \hat{\theta}_i^{(e_i)}\|_2 + T^*_{i,e_i} \|\theta_i^* - \hat{\theta}_i^{(e_i)}\|_2 \right) \tag{23}$$

The regret thus depends on the following: a) the accuracy of estimating θ_i^* in each episode for all contexts; b) the number of rounds when any context i is optimal but not chosen,i.e., the quantities T^*_{i,e_i}, and c) the number of episodes per context, i.e., the quantities $e_{i,\max}$. A major difference compared to the single parameter setting is the quantity T^*_{i,e_i} and the relation of the regret with T^*_{i,e_i}. Note that the estimate of any context parameter improves with the number of times the particular context is chosen. The quantities T^*_{i,e_i}, while contributing to the regret, represent rounds when the context is not chosen and hence do not contribute to improvement of the parameter estimate. In contrast in the single parameter setting, since there is only one parameter, any chosen context contributes towards better parameter estimation rates. The larger warm start phase in the multi parameter setting is to ensure the greedy algorithm chooses contexts with constant probability when they are optimal to limit the quantities T^*_{i,e_i}.

We focus on regret bounds when the contexts are $x_i^t = \mu_i^t + g_i^t, 1 \leq i \leq k, 1 \leq t \leq T$, where μ_i^t's are chosen by an adaptive adversary and g_i^t's are the Gaussian perturbations. We begin with a characterization of the number of rounds required in the warm start phase. Remember, the goal of the warm start phase is to ensure that there is a constant probability the algorithm chooses the optimal context. This is the essence of the margin condition in Lemma 5. Propositions 1 and 2 build towards the result in Lemma 5. Proposition 1 is a straightforward observation on the relationship between the first and second optimal contexts where we introduce the quantity r. To summarize, Proposition 1 makes the observation that the dot product between the Gaussian perturbation and parameter of the optimal context exceeds r.

Proposition 1 *Consider any round t when the episode numbers of the k contexts are $e_1, \ldots, e_k$. Let i^* denote the context with the maximum reward, i.e., $i^* = \underset{l:1\leq l\leq k}{\operatorname{argmax}} \langle \mu_l^t + g_l^t, \theta_l^* \rangle$. Let j denote the context having the second largest reward, i.e., $j = \underset{l:1\leq l\leq k; l\neq i^*}{\operatorname{argmax}} \langle \mu_l^t + g_l^t, \theta_l^* \rangle$. Define $r = \langle \mu_j^t + g_j^t, \theta_j^* \rangle - \langle \mu_{i^*}^t, \theta_{i^*}^* \rangle$. Then the following condition is satisfied,*

$$\langle g_{i^*}^t, \theta_{i^*}^* \rangle \geq r \, . \tag{24}$$

Proposition 2 states conditions when the greedy algorithm chooses the optimal context. Due to parameter estimation errors, for the greedy algorithm to perceive the context to be optimal the dot product between the optimal parameter vector and Gaussian perturbation should now exceed r by a quantity which depends on the estimation error.

Proposition 2 *Assume context j' such that $j' = \underset{l:1\leq l\leq k, l\neq i^*}{\operatorname{argmax}} \langle \mu_l^t + g_l^t, \hat{\theta}_l^{(e_l)} \rangle$, i.e., the context other than i^* which has the highest estimated reward. Also assume the parameter estimate for context i^* to be $\hat{\theta}_{i^*}^{(e_{i^*})} = \theta_{i^*}^* + \Delta_{i^*}^{(e_{i^*})}$ and for context j', $\hat{\theta}_{j'}^{(e_{j'})} = \theta_{j'}^* + \Delta_{j'}^{(e_{j'})}$. Then the greedy algorithm selects context i^* if the following condition is satisfied,*

$$\langle g_{i^*}^t, \theta_{i^*}^* \rangle \geq r + \langle \mu_{j'}^t + g_{j'}^t, \Delta_{j'}^{(e_{j'})} \rangle - \langle \mu_{i^*}^t + g_{i^*}^t, \Delta_{i^*}^{(e_{i^*})} \rangle \, . \tag{27}$$

The greedy algorithm always picks the optimal context if

Algorithm 2 High-dimensional Greedy (multi parameter)

```
Set $e_1 = \ldots = e_k = 0$. Initialize empty design matrices and rewards $Z_1^{(0)}, \ldots, Z_k^{(0)} = []$,$y_1^{(0)}, \ldots, y_k^{(0)} = []$
for $t = 1$ to $T_0$ do
    Observe contexts $x_1^t, \ldots, x_k^t \in \mathbb{R}^p$
    Pick context $i^t$ from $\{1, \ldots, k\}$ in round robin fashion and observe reward $r_{i^t}^t = \langle x_{i^t}^t, \theta_{i^t}^* \rangle + \omega^t$ where $\omega^t$ is zero mean $\kappa_\omega$-sub-Gaussian noise
    Append observations $(x_{i^t}^t, r_{i^t}^t)$ to $(Z_{i^t}^{(0)}, y_{i^t}^{(0)})$
end for
SVD: $\frac{1}{\sqrt{T_{i,0}}} Z_i^{(0)} = U_i^{(0)} D_i^{(0)} (V_i^{(0)})^\intercal$
Puffer transformation: $F_i^{(0)} = U_i^{(0)} (D_i^{(0)})^{-1} (U_i^{(0)})^\intercal$
    $\tilde{y}_i^{(0)} = F_i^{(0)} y_i^{(0)}$, $\tilde{Z}_i^{(0)} = F_i^{(0)} Z_i^{(0)}$
Estimate parameters using constrained least squares estimator for each context with $T_{1,0} = \ldots = T_{i,0} = \ldots = T_{k,0} = T_0/k$
    $$\hat{\theta}_i^{(1)} = \underset{\theta \in \mathbb{R}^p}{\text{argmin}} \frac{1}{2T_{i,0}} \|\tilde{y}_i^{(0)} - \tilde{Z}_i^{(0)} \theta\|_2^2 \quad \text{s.t.} \quad R(\theta) \le R(\theta_i^*) \, , \qquad (25)$$
Increment all $e_i = e_i + 1, 1 \le i \le k$. Initialize empty design matrices and rewards $Z_1^{(e_1)}, \ldots, Z_k^{(e_k)} = []$,$y_1^{(e_1)}, \ldots, y_k^{(e_2)} = []$, and $t_1 = \ldots = t_k = 0$.
for $t = T_0$ to $T$ do
    Observe contexts $x_1^t, \ldots, x_k^t \in \mathbb{R}^p$
    Pick context $i^t$ such that $i^t = \underset{1 \le i \le k}{\text{argmax}} \langle x_i^t, \hat{\theta}_i^{(e_i)} \rangle$, receive reward $r_{i^t}^t = \langle x_{i^t}^t, \theta_{i^t}^* \rangle + \omega^t$ and increment $t_{i^t} = t_{i^t} + 1$
    Append observations $(x_{i^t}^t, r_{i^t}^t)$ to $(Z_{i^t}^{(e_{i^t})}, y_{i^t}^{(e_{i^t})})$
    if $t_{i^t} = 2T_{i^t, e_{i^t} - 1} = T_{i^t, e_{i^t}}$ then
        SVD: $\frac{1}{\sqrt{T_{i, e_{i^t}}}} Z_{i^t}^{(e_{i^t})} = U_{i^t}^{(e_{i^t})} D_{i^t}^{(e_{i^t})} (V_{i^t}^{(e_{i^t})})^\intercal$
        Puffer transformation:
        $F_{i^t}^{(e_{i^t})} = U_{i^t}^{(e_{i^t})} (D_{i^t}^{(e_{i^t})})^{-1} (U_{i^t}^{(e_{i^t})})^\intercal$,
        $\tilde{Z}_{i^t}^{(e_{i^t})} = F_{i^t}^{(e_{i^t})} Z_{i^t}^{(e_{i^t})}$ and $\tilde{y}_{i^t}^{(e_{i^t})} = F_{i^t}^{(e_{i^t})} y_{i^t}^{(e_{i^t})}$
        Estimate parameter using constrained least squares estimator
            $$\hat{\theta}_{i^t}^{(e_{i^t}+1)} = \underset{\theta \in \mathbb{R}^p}{\text{argmin}} \frac{1}{2T_{i^t, e_{i^t}}} \|\tilde{y}_{i^t}^{(e_{i^t})} - \tilde{Z}_{i^t}^{(e_{i^t})} \theta\|_2^2 \quad \text{s.t.} \quad R(\theta) \le R(\theta_{i^t}^*) \, , \qquad (26)$$
        where $T_{i^t, e_{i^t}} = 2T_{i^t, e_{i^t} - 1}$.
        Increment $e_{i^t} = e_{i^t} + 1$. Initialize empty design matrix $Z_{i^t}^{(e_{i^t})} = []$ and reward $y_{i^t}^{(e_{i^t})} = []$, $t_{i^t} = 0$.
    end if
end for
```

the condition in equation (27) is satisfied. Let us now fix the quantity r. Let the estimation errors after the warm start phase be such that $\left| \langle g_{j'}^t, \Delta_{j'}^{(e_{j'})} \rangle - \langle \mu_{i^*}^t + g_{i^*}^t, \Delta_{i^*}^{(e_{i^*})} \rangle \right| \le \frac{\sigma^2}{r}$. Then the probability that there is a match between the optimal context and the context chosen by the greedy algorithm is precisely the quantity on the l.h.s. in equation (28). Now what are values of r when equation (28) is satisfied? In the proof provided in the supplement, we will prove that the probability in equation (28) decreases with increasing r. Therefore to obtain lower bounds we assume an upper bound on r which we will show to hold with high probability over choices of contexts, μ_k^t, g_k^t, in all rounds.

Lemma 5 **(Margin Condition)** *Consider good events as when $r \le c_3 \sigma \sqrt{\log(Tk)}$ and consider errors $\Delta_{i^*}^{(e_{i^*})}$ and $\Delta_{j'}^{(e_{j'})}$ to be small enough such that $\langle \mu_{j'}^t + g_{j'}^t, \Delta_{j'}^{(e_{j'})} \rangle - \langle \mu_{i^*}^t + g_{i^*}^t, \Delta_{i^*}^{(e_{i^*})} \rangle \le \frac{\sigma^2}{r}$. Then the following holds,*

$$P\left(\langle g_{i^*}^t, \theta_{i^*}^* \rangle \ge r + \frac{\sigma^2}{r} \;\middle|\; \langle g_{i^*}^t, \theta_{i^*}^* \rangle \ge r \right) \ge \frac{1}{20} \, , \qquad (28)$$

for all $r \le c_3 \sigma \sqrt{\log(Tk)}$.

The length of the warm start phase is now influenced by the condition that $\|\Delta_{j'}^{(e_{j'})}\|_2$ and $\|\Delta_{i^*}^{(e_{i^*})}\|_2$ are small enough so that $\langle \mu_{j'}^t + g_{j'}^t, \Delta_{j'}^{(e_{j'})} \rangle - \langle \mu_{i^*}^t + g_{i^*}^t, \Delta_{i^*}^{(e_{i^*})} \rangle \le \frac{\sigma^2}{r}$ in Lemma 5 which translates to the upper bound below:

$$\|\Delta_i^{(e_i)}\|_2 = \|\hat{\theta}_i^{(e_i)} - \theta_i^*\|_2 \le \tilde{O}(\sigma) \, . \qquad (29)$$

The estimation error bounds are in turn influenced by the properties of the design matrices after the warm start phase.

Lemma 6 **(Multi parameter Design Matrix Properties)** *Consider any context i and a particular episode e_i. The rows of the design matrix $Z_i^{(e_i)} \in \mathbb{R}^{T_{i,e_i} \times p}$ are $z_i^t = \mu_i^t + g_i^t$ where in round t context i is chosen by the Greedy algorithm, i.e., $i = \underset{1 \le l \le k}{\text{argmax}} \langle x_l^t, \hat{\theta}_l^{(e_l)} \rangle$ where $x_l^t = \mu_l^t + g_l^t$, $g_l^t \sim N(0, \sigma^2 \mathbb{I}_{p \times p})$. Then under the condition $\langle g_i^t, \theta_i^* \rangle \ge r$ for some $r \le c_3 \sigma \sqrt{\log(Tk)}$,*

$$\lambda_{\min}\left(E_{z^t}\left[z_i^t (z_i^t)^\intercal \;\middle|\; z_i^t \textit{ satisfies } \zeta \right] \right) \ge c_2 \frac{\sigma^2}{\log(Tk)} \, ,$$

where ζ is the condition $z_i^t = \underset{g_l^t : 1 \le l \le k}{\text{argmax}} \langle x_l^t, \hat{\theta}_l^{(e_l)} \rangle ; \langle g_i^t, \theta_i^* \rangle \ge r ; r \le c_3 \sigma \sqrt{\log(Tk)}$.

The only difference in the properties of the design matrix compared to the single parameter setting are the sub-Gaussian norm and expected minimum eigenvalue of the covariance matrix. Using similar steps to derive estimation error as in the single parameter setting, we obtain the following upper bound on the maximum estimation error across all contexts and episodes with high probability:

$$\sup_{1\leq i\leq k}\ \sup_{e_i\leq e_{i,\max}} \|\hat{\theta}_i^{(e_i+1)} - \theta_i^*\|_2 \leq \tilde{O}\left(\frac{w(A)}{\sigma\sqrt{T_{i,e_i}}}\right) . \quad (30)$$

Comparing equations (29) and (30) it can be easily inferred that $T_{i,e_i} = \tilde{\Theta}\left(\frac{w^2(A)}{\sigma^4}\right)$ to satisfy the margin condition and since the episode length increases monotonically the length of the warm start phase $T_0 = \tilde{\Theta}\left(\frac{kw^2(A)}{\sigma^4}\right)$.

After the warm start phase, the margin condition of Lemma 5 holds and ensures that the greedy algorithm chooses the optimal context with probability atleast $1/20$. In other words, in expectation $T^*_{i,e_i} \leq 20T_{i,e_i}$, i.e., in any episode for any context the number of rounds when the context is optimal but not perceived to be optimal by the greedy algorithm in expectation is upper bounded by 20 times the length of the episode. With the result on T^*_{i,e_i}'s and the upper bound on the parameter estimation errors, the regret upper bound in the multi parameter setting can be derived from the result of Lemma 4.

Theorem 3 (Multi parameter Smoothed Adversary Regret Bounds) *Consider computation of regret for the Greedy algorithm in the multi parameter setting following Lemma 4. Let $r \leq c_3\sigma\sqrt{\log(Tk)}$, $\beta = \max_{1\leq i\leq k, 1\leq t\leq T}\|x_i^t\|_2$, and*

$$\gamma = \frac{c_{12}\kappa_\omega(w(A) + \sqrt{\log k} + \sqrt{\log(1/\delta)})\sqrt{\log(Tk)}}{\sigma}$$

. The margin condition in Lemma 5 is satisfied with probability atleast $1 - 5\delta$ when,

$$t_{\min} \geq \frac{4k\gamma^2 r^2\beta^2}{\sigma^4} + 1 + \sqrt{\frac{1}{2}\log(1/\delta)} . \quad (31)$$

Under the margin condition, the regret is maximized when in each round each context has equal probability to be selected by the Greedy algorithm. The equal probability implies that in expectation $T_1 = T_2 = \ldots = T_k = \frac{T}{k}$. Also the regret is upper bounded as follows,

$$\textit{Reg}(T) \leq 2\beta t_{\min} + 82\beta\gamma\sqrt{Tk}\log(T) . \quad (32)$$

Moreover $\beta \leq (1+c_1\sigma(\sqrt{p}+\sqrt{\log(1/\delta)}))$ with probability atleast $1 - 2\delta$. Therefore with probability atleast $1 - 7\delta$ assuming $t_{\min} = O(\sqrt{Tk})$

$$\textit{Reg}(T) \leq O\left(\gamma \cdot \beta \cdot \log(T) \cdot \sqrt{Tk}\right) \quad (33)$$

The regret is $\sqrt{k}$ times worse than that of the single parameter setting.

4.1. Examples

We instantiate the regret bounds for a few norms. When $R(\cdot)$ is $\|\cdot\|_2^2$ and $\sigma^2 = \Theta\left(\frac{1}{p}\right)$, the length of the warm start phase is $\tilde{\Theta}(kp^3)$ which improves over the $\tilde{\Theta}(kp^6)$ obtained in (Kannan et al., 2018). Ignoring logarithm terms the regret bounds are of the same order as (Agarwal & Goyal, 2013) after the warm start phase but the polynomial in p warm start rounds maybe prohibitive in many high-dimensional applications.

Corollary 2 *Let $\sigma^2 = \Theta\left(\frac{1}{p}\right)$. Then with probability atleast $1 - 7\delta$:*

1. *Let θ^* be s-sparse, $R(\cdot)$ the ℓ_1 norm.*

$$\textit{Reg}(T) = \tilde{O}\left(\sqrt{p}\sqrt{s\log p}\sqrt{Tk}\right) . \quad (34)$$

2. *Let $\theta^* \in \mathbb{R}^{m\times q}$ be a rank r matrix $r \leq \min\{m, q\}$, $R(\cdot)$ is the nuclear norm.*

$$\textit{Reg}(T) = \tilde{O}\left(q\sqrt{r(m+q)}\sqrt{Tk}\right) . \quad (35)$$

3. *Let $R(\cdot)$ the ℓ_2^2 norm.*

$$\textit{Reg}(T) = \tilde{O}\left(p\sqrt{Tk}\right) . \quad (36)$$

5. Conclusions

We analyzed the structured linear contextual bandit problem under the smoothed analysis framework. Our analysis significantly improves on the bounds obtained in (Kannan et al., 2018). While previous work have found it difficult to extend efficient exploration strategies exploiting parameter structure in the high-dimensional setting, our analysis shows that a simple greedy algorithm achieves sublinear regret under the smoothed bandits framework.

Acknowledgements: The research was supported by NSF grants OAC-1934634, IIS-1908104, IIS-1563950, IIS-1447566, IIS-1447574, IIS-1422557, CCF-1451986, FAI-1939606, a Google Faculty Research Award, a J.P. Morgan Faculty Award, and a Mozilla research grant. The authors will like to thank the anonymous reviewers for their kind and helpful feedback. This work was completed when VS was a PhD student in the Department of Computer Science at the University of Minnesota, Twin Cities. Part of this work was completed while ZSW was visiting the Simons Institute for the Theory of Computing at UC Berkeley.

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. Online Least Squares Estimation with Self-Normalized Processes: An Application to Bandit Problems. In *Conference on Learning Theory (COLT)*, 2011.

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. Online-to-Confidence-Set Conversions and Application to Sparse Stochastic Bandits. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2012.

Agarwal, S. and Goyal, N. Thompson Sampling for Contextual Bandits with Linear Payoffs. In *International Conference on Machine Learning (ICML)*, 2013.

Argyriou, A., Foygel, R., and Srebro, N. Sparse Prediction with the k-Support Norm. In *Neural Information Processing Systems (NIPS)*, 2012.

Banerjee, A., Chen, S., Fazayeli, F., and Sivakumar, V. Estimation with Norm Regularization. In *Neural Information Processing Systems (NIPS)*, 2014.

Banerjee, A., Gu, Q., Sivakumar, V., and Wu, Z. S. Random quadratic forms with dependence: Applications to restricted isometry and beyond. In *Advances in Neural Information Processing Systems (NIPS)*, 2019.

Bastani, H., Bayati, M., and Khosravi, K. Mostly exploration-free algorithms for contextual bandits. *CoRR arXiv:1704.09011*, 2018. Working paper.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009. ISSN 0090-5364.

Bietti, A., Agarwal, A., and Langford, J. Practical evaluation and optimization of contextual bandit algorithms. *CoRR arXiv:1802.04064*, 2018.

Bird, S., Barocas, S., Crawford, K., Diaz, F., and Wallach, H. Exploring or exploiting? social and ethical implications of automonous experimentation. In *Workshop on Fairness, Accountability, and Transparency in Machine Learning*, 2016.

Candès, E. J. and Recht, B. Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009. ISSN 1615-3375.

Chandrasekaran, V., Recht, B., Parrilo, P. A., and Willsky, A. S. The Convex Geometry of Linear Inverse Problems. *Foundations of Computational Mathematics*, 12(6):805–849, 2012.

Chen, S. and Banerjee, A. Structured Estimation with Atomic Norms: General Bounds and Applications. In *Neural Information Processing Systems (NIPS)*, 2015.

Chu, W., Li, L., Reyzin, L., and Schapire, R. E. Contextual bandits with linear payoff functions. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.

Dani, V., Hayes, T. P., and Kakade, S. M. Stochastic Linear Optimization Under Bandit Feedback. In *Conference on Learning Theory (COLT)*, 2008.

Gordon, Y. Some inequalities for gaussian processes and applications. *Israel Journal of Mathematics*, 50(4):265–289, 1985.

Jacob, L., Obozinski, O., and Vert, J. P. Group Lasso with Overlap and Graph Lasso. In *International Conference on Machine Learning (ICML)*, number 2009, 2009.

Javanmard, A. and Javadi, H. Dynamic Pricing in High Dimensions. *Accepted in JMLR*, 2018.

Jia, J. and Rohe, K. Preconditioning the lasso for sign consistency. *Electronic Journal of Statistics*, 9:1150–1172, 2015.

Kannan, S., Kearns, M. J., Morgenstern, J., Pai, M. M., Roth, A., Vohra, R. V., and Wu, Z. S. Fairness incentives for myopic agents. In Daskalakis, C., Babaioff, M., and Moulin, H. (eds.), *Proceedings of the 2017 ACM Conference on Economics and Computation, EC '17, Cambridge, MA, USA, June 26-30, 2017*, pp. 369–386. ACM, 2017. doi: 10.1145/3033274.3085154. URL https://doi.org/10.1145/3033274.3085154.

Kannan, S., Morgenstern, J., Roth, A., Waggoner, B., and Wu, Z. S. A smoothed analysis of the greedy algorithm for the linear contextual bandit problem. *CoRR arXiv:1801.04323*, 2018.

Langford, J. and Zhang, T. The Epoch-Greedy Algorithm for Contextual Multi-armed Bandits. In *Advances in Neural Information Processing Systems (NIPS)*, 2007.

Li, L., Chu, W., Langford, J., and Schapire, R. E. A contextual-bandit approach to personalized news article recommendation. In *International World Wide Web Conference (WWW)*, 2010.

Mansour, Y., Slivkins, A., and Wu, Z. S. Competing bandits: Learning under competition. In *Innovations in Theoretical Computer Science (ITCS)*, 2018.

Mendelson, S., Pajor, A., and Tomczak-Jaegermann, N. Reconstruction and subGaussian operators in asymptotic geometric analysis. *Geometric and Functional Analysis*, 17:1248–1282, 2007.

Negahban, S. N., Ravikumar, P., Wainwright, M. J., and Yu, B. A Unified Framework for High-Dimensional Analysis of \$M\$-Estimators with Decomposable Regularizers. *Statistical Science*, 27(4):538–557, 2012. ISSN 0883-4237.

Raghavan, M., Slivkins, A., Vaughan, J. W., and Wu, Z. S. The externalities of exploration and how data diversity helps exploitation. In *Conference on Learning Theory (COLT)*, pp. 1724–1738, 2018.

Sivakumar, V., Banerjee, A., and Ravikumar, P. Beyond subgaussian measurements: High-dimensional structured estimation with sub-exponential designs. In *Advances in Neural Information Processing Systems (NIPS)*, 2015.

Talagrand, M. *The Generic Chaining*. Springer Monographs in Mathematics. Springer Berlin, 2005.

Talagrand, M. *Upper and Lower Bounds of Stochastic Processes*. Springer, 2014.

Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*, 58(1): 267–288, 1996.

Vershynin, R. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

Wainwright, M. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press (To appear), 2019.

Yuan, M. and Lin, Y. Model Selection and Estimation in Regression With Grouped Variables. *Journal of the Royal Statistical Society*, 68(1):49–67, 2006.