Generalized Linear Bandits revisited

Julian Zimmert
Humboldt University of Berlin

Csaba Szepesvri
University of Alberta

Abstract

Previous work on generalized linear bandits (GLB) suggested that algorithms based on optimism can achieve a $\sqrt{T}$-regret analogous to the linear bandit setting. We argue however, that these results should be treated as asymptotic results, as they include generally huge constants that might even be a priori unknown. Our novel lower bound for GLB proves that these constants are not mere artifacts of the analysis, but that the regret might be as large as $\Omega(\frac{d+1}{d+3})$. We prove that this lower bound has a matching upper bound up to log factors at least on a slightly restricted setting.

1 Introduction

We are concerned about the problem of general linear bandits (GLB) which are an extension of linear bandits that generalize the classical K-armed bandit problem. In the K-armed bandit problem, an agent selects one out of K possible arms at each time and immediately observes a reward. The reward depends only on the selected arm and is assumed to be drawn out of a probability distribution. The means of the distributions are unknown and independent of each other, but the tails are all assumed to be sub-Gaussian. In these problems, the agent is faced with a fundamental trade-off between playing arms to gain information about the reward structure (Exploration) and chooses the arm that is most promising given the available information (Exploitation). For linear bandits, we model an an underlying structure that links the mean rewards for different arms. In this setting, each arm is attached to a d-dimensional feature vector x and the mean reward is given by $x^T\theta_*$, for an unknown parametrization θ_*. Because this allows to infer information about arms that were never played, meaningful results exist even for infinitely many arms. GLB also assumes a feature structure, but the mean reward is $\mu(x^T\theta_*)$ for a non-decreasing link function μ.

For both the K-armed bandit as well as the linear bandit problems, algorithms with favourable theoretical properties exist. The regret, that is the accumulated difference of rewards between playing the optimal arm at each time-step and following the algorithm, is bounded by $\tilde{\mathcal{O}}(\sqrt{T})$ with high probability. Existing proofs for lower bounds show that improvement is impossible for all except constants and log factors. An analogous algorithm for GLB also guarantees $\tilde{\mathcal{O}}(\sqrt{T})$ regret, but we see two downsides of this result. First of all, it needs to assume that the derivative of μ is lower bounded away from 0. $\dot{\mu} \geq c_\mu > 0$. Secondly, the bound requires two constants that might get extremely large, making this result more asymptotic in nature. We are showing in this work that these constants are not mere artifacts of the analysis. Worst still, we show that the T dependency is greater than $\sqrt{T}$ once we do not lower bound $\dot{\mu}$. Additionally we are going to prove that the estimator of θ_* ,used in the work of [], is fundamentally sub-optimal.

2 Problem setting and existing results

The agent plays for a total of T rounds and chooses an arm X_t at each round. We assume that the arms come from a closed subset S of the d-dimensional euclidean space $x_i \in S \subset \mathbb{R}^d$, are bounded in length by L and that the features are known to the agent. There exists a bounded unknown parameter $\theta_* \in \mathbb{R}$, $||\theta_*||_2 \leq R$. When the agent plays an arm x_i at time t, he will observe the reward $R_t = \mu(x_i^T\theta_*) + \eta_t$. μ is hereby a non-decreasing link function that is k_μ-Lipschitz and η_t is a random noise that is σ-subgaussian. The optimal arm is denoted x_*: $\max_x \mu(x^T\theta_*) = \mu(x_*^T\theta_*)$. For a given strategy, we denote with x_t the arm that is chosen at time t.
The immediate regret at time t is defined as

$$\mathrm{Reg}_t := \mathbb{E}\left[R_t | X_t = x_*\right] - \mathbb{E}\left[R_t | X_t = x_t\right]$$

Preliminary work. Under review by AISTATS 2017. Do not distribute.

, and the total regret or simply regret is the cumulative sum of immediate regrets

$$\mathrm{Reg}(T) := \sum_{i=1}^{T} \mathrm{Reg}_i \,.$$

We define two different estimators in this section. The second is used in the work of **??**, the other, we believe, is more favorable in this setting. Given observations $(R_1, X_1), (R_2, X_2), \ldots, (R_{t-1}, X_{t-1})$, the least square estimator (LSE) $\hat{\theta}$ is defined as the minimizer of

$$\sum_{k=1}^{t-1} \left(R_k - \mu(X_k^T \theta_*)\right)^2 .$$

Consistency of the LSE in our setting is a standard result **??**.

The pseudo maximum likelihood estimator (pseudo-MLE), is defined as the solution $\hat{\theta}_P$ of the equation

$$\sum_{k=1}^{t-1} (R_k - \mu(X_k^T \hat{\theta}_P)) X_k = 0.$$

The estimator is consistent under very general assumptions, however it requires again that $\dot{\mu}$ is bounded away from zero, which is why this estimator is unfavourable in our setting. For self consistency, we briefly explain the rationale behind this estimator. If the rewards are actually drawn from a *canonical exponential family*, that means the density for a reference measure is

$$p_\beta(r) = \exp(r\beta b(\beta) + c(r)),$$

then the pseudo-MLE is simply the regular maximum likelihood estimator. β is a real number, c is a real-valued function and b is twice continuously differentiable. The mean is given by $\mathbb{R}[R] = \dot{b}(\beta)$ ($= \mu(\beta)$) and the tails are indeed subgaussian.

An algorithm based on the *optimism in the face of uncertainty* principles is proven to satisfy a bound on the regret of the shape $\tilde{\mathcal{O}}(\frac{k_\mu}{c_\mu}\sqrt{T})$. The motivating function for this setting is traditionally the sigmoid function $f(x) = (1 + \exp(-x))^{-1}$. W.l.o.g. we can reformulate the problem such that $||x||, ||\theta_*|| \leq 1$ and $\mu : [-1, 1] \to \mathbb{R}$, $\mu(x) = (1 + \exp(-LR_{\max} x))^{-1}$. In this case $k_\mu = \frac{LR}{4}$, $c_\mu^{-1} = \left(\frac{LR}{(e^{0.5LR} + e^{-0.5LR})^2}\right)^{-1} > \frac{\exp(LR)}{LR}$. The constant in front of T grows exponentially with the possible length of θ_* and the arms, rendering this bound meaningless for most applications.

From now on, we will replace the restriction $\dot{\mu} \geq c_\mu > 0$ by simply requiring $\dot{\mu} \geq 0$. This is taking the closure of the open cone $\dot{\mu} > 0$. Therefore any lower bound on finite time, can be matched arbitrarily close for the open cone, as long as we choose c_μ sufficiently small. Additionally, we relax the constant k_μ by reducing the Lipschitz-property to a single point, i.e.

$$\mu(1) - \mu(1 - x) \geq kx$$

For short notation, we will define the set of valid link functions as

$$M_k := \{\mu \in C_1([-1, 1]) | \dot{\mu} \geq 0, \mu(1) - \mu(1 - x) \geq kx\}$$

3 Lower bounds

. We now present the main results of the paper. First a algorithm independent lower bounds for non-decreasing link functions proves that the constants are not mere artifacts of the analysis. Second an algorithm dependent bound proofs that there is an insuperable gap when using the pseudo MLE for sub-gaussian noise.

3.1 Algorithm independent bound

Theorem 1. *For any finite time horizon T there exists a generalized linear bandit problem with finitely many arms $\mathcal{X} \subset \mathbb{R}^d$, $\mu \in M_k$, such that the regret of any algorithm will be at least of the order*

$$\Omega\left(T \wedge \left(\frac{d+1}{d+3}\right)\right)$$

.

Corollary 1.1. *It is impossible to derive a c_μ independent upper bound for the generalized linear bandit of the order $\tilde{\mathcal{O}}(\sqrt{T})$.*

Proof. Define the GLB Problem $P(T)$ as follows:
We chose the link function

$$\mu_\epsilon(x) := \max\{0, (x + \epsilon - 1)k\} \tag{1}$$

with ϵ a small number to be chosen later. For two arms x_i, x_j, that have an euclidean distance of at least $||x_i - x_j|| \geq \sqrt{2\epsilon}$, simple algebra shows that the scalar product is bounded by $x_i^T x_j \leq 1 - \epsilon$. We want to select the maximum amount K of arms on the unit sphere $\mathcal{S}^{d-1}$, such that each two arms have at least the distance $\sqrt{2\epsilon}$. Results on packing unit spheres [1] [2] show that we can find such a set $\mathcal{X}$ with cardinality at least

$$\begin{aligned} K := |\mathcal{X}| &\geq \left(\frac{1}{\sin(\sqrt{2\epsilon})} + o(1)\right)^{d-1} \\ &\geq (2\epsilon) \wedge \left(-\frac{d-1}{2}\right). \end{aligned}$$

Choosing $\theta_* \in \mathcal{X}$ at random, this problem is equivalent to the K-armed bandit problem where one arm give ϵ average reward and all other arms 0.

The lower bound proof of [**?**] implies that as long as $k\epsilon \leq \sqrt{\frac{K}{T}}$, the total regret is lower bounded by $\Omega(k\epsilon T)$. Setting

$$\epsilon = \frac{1}{2}k \wedge \left(-\frac{4}{d+3}\right) T \wedge \left(-\frac{2}{d+3}\right)$$

satisfies this inequality. So the total in regret will be at least of the order

$$\Omega\left(k \wedge \left(\frac{d-1}{d+3}\right) T \wedge \left(\frac{d+1}{d+3}\right)\right)$$

□

The exponent $\frac{d+1}{d+3}$ converges to 1 as the dimension increases, that means for a finite time horizon and large dimensions, the regret will almost be undistinguishable from linear regret. In real world application, we typically deal with relatively large dimensions. Unless we are very careful about the chosen link function, it might be impossible to learn anything meaningfull in a given time. That is why we distinguish in section 4 between different families of link functions, to provide safe-to-use applications.

3.2 Pseudo Maximum-Likelihood based lower bound

For simplicity, we assume a simple Explore then Commit algorithm in two dimensions. We show that for any time horizon T, we can construct a problem where with constant probability the pseudo MLE choses an arm not better than random. During the exploration, we suffer an regret even larger than what the general lower bound predicts.

Theorem 2. *For any finite time horizon T, there exists a generalized linear bandit problem such that after T uniform exploration steps, the regret for playing the optimal arm for $\hat{\theta}$ will be $T^{-\frac{1}{4}}$ with a T independent constant probability.*

Corollary 2.1. *The worst case regret for this algorithm is at least $T^{\frac{3}{4}}$.*

Proof. Define the GLB Problem as follows:

- $\mu(x) := \max\{0, x + \epsilon - 1\}$
- $\mathcal{X} = \mathcal{S}^1$, the 2-d unit circle
- $\eta_i \sim \mathcal{N}(0, 1)$
- $\theta_* = \binom{1}{0}$

After T steps of uniform exploration, the psuedo-MLE is the solution to

$$\sum_{k=1}^{T} \left(\mu(x_k^T\theta_*) - \mu(x_k^T\hat{\theta}_P) + \eta_k\right) x_k = 0.$$

$$\sum_{k=1}^{T} \mu(x_k^T\hat{\theta}_P)x_k = \sum_{k=1}^{T} \mu(x_k^T\theta_*)x_k + \boldsymbol{\eta},$$

with $\boldsymbol{\eta} \sim \mathcal{N}\left(0, \sum_{k=1}^{T} x_k x_k^T)\right)$. We argue that this estimator can in average not be better, than if we would perform a continuous exploration. By choosing an arbitrary uniform distribution of points on the circle, we are adding additional noise over continuous exploration. We are looking at the estimator given by

$$\frac{T}{2\pi}\int_{\mathcal{S}^1} \mu(x^T\hat{\theta}_C)x\, dx = \frac{T}{2\pi}\int_{\mathcal{S}^1} \mu(x^T\theta_*)x\, dx + \boldsymbol{\eta},$$

with $\boldsymbol{\eta} \sim \mathcal{N}(0, \frac{T}{2\pi}\int_{\mathcal{S}^1} xx^T\, dx) = \mathcal{N}(0, T\mathbf{I})$. The left hand side is co-linear to $\hat{\theta}_C$ due to symmetry, the integral on the right hand side can be explicitly solved.

$$\begin{aligned}
&\frac{T}{2\pi}\int_{\mathcal{S}^1} \mu(x^T\theta_*)x\, dx = \frac{T}{2\pi}\int_0^{2\pi} \mu(\cos(x))\cos(x)\theta_*\, dx \\
=&\frac{T}{\pi}\theta_*\int_0^{\pi} \mu(\cos(x))\cos(x)\, dx = \frac{T}{\pi}\theta_*\int_0^{1} \frac{\mu(x)x}{\sqrt{1-x^2}}\, dx \\
=&\frac{T}{\pi}\theta_*\int_{1-\epsilon}^{1} \frac{x-x^2}{\sqrt{1-x^2}}\, dx \\
=&\frac{T}{2\pi}\left(\sin^{-1}(1-\epsilon) - \frac{\pi}{2} + (1+\epsilon)\sqrt{2\epsilon-\epsilon^2}\right)\theta_* \\
=&c_\epsilon T\theta_*
\end{aligned}$$

The arm chosen after the Exploration phase is therefore:

$$\hat{x} = \frac{\hat{\theta}_C}{||\hat{\theta}_C||} = \frac{(c_\epsilon T + \eta_1)\binom{1}{0} + \eta_2\binom{0}{1}}{\sqrt{(c_\epsilon T + \eta_1)^2 + \eta_2^2}}.$$

We are interested for a bound on the immediate regret for choosing $\hat{x}$. Given that $\eta_1, \eta_2 \sim \mathcal{N}(0, T)$ i.i.d, with probability at least $\frac{1}{4}$, we have $\eta_1 \leq 0$ and $\eta_2 \geq T$.

$$\hat{x}^T\theta_* = \frac{c_\epsilon T + \eta_1}{\sqrt{(c_\epsilon T + \eta_1)^2 + \eta_2^2}} \leq \frac{c_\epsilon T}{\sqrt{c_\epsilon^2 T^2 + T}}$$

Series expansion of c_ϵ, shows that $c_\epsilon \leq \epsilon^{\frac{3}{2}}$.

$$\hat{x}^T\theta_* \leq \frac{\epsilon^{\frac{3}{2}}T}{\sqrt{\epsilon^3 T^2 + T}} \leq \frac{1}{\sqrt{2}}$$

For $T \geq$, it holdts that $\frac{1}{\sqrt{2}} < 1-\epsilon$. With probability at least $\frac{1}{4}$, the estimator will choose an arm that suffers ϵ immediate regret. Therefore even if Exploit-then-Commit would spend less than the full for exploration, the total regret will still be of order $\Omega(\epsilon T) = \Omega(T^{\frac{2}{3}})$.

There exists a gap to the general lower bound of order $\Omega(T^{\frac{2}{3}})$. We will show in the next section, that employing LSE performs better.

□

The pseudo-MLE based approach will suffer a regret that is higher than the general lower bound. In the following section, we show that the LSE does not show this behaviour, proving a true gap between the two estimators.

Remark 1. *In section 4, we will look at point symmetric functions μ for simpler analysis. We want to add that all lower bounds can be equivalently proven for a symmetric $\mu(x) = \mu_\epsilon(x) - \mu_\epsilon(-x)$.*

4 Upper bounds

The lower bound given in the previous section is of a relatively odd nature. To our best knowledge, the hasn't been a similar exponent derived for any bandit related lower bound. This naturally raises the question, whether this is even a tight lower bound. While for practical purposes, this is of little relevance. Even matching algorithms would be prohibitive with an almost linear regret. The question is however of academic interest. Secondly, it is essential for provide well behaving function that are guaranteed to suffer $\sqrt{T}$ regret instead.

We are not yet able to provide a complete answer to these question. Based on a slightly simplified setting, we can show that

- the lower bound is indeed tight up to log factors
- symmetric convex-concave link functions are safe-to-use

Restrictions We believe the results hold even without further restrictions, but removing them would complicate the proofs to a point that would burst all limits of our appendix. In this section, we assume that

- the arms consists of the d-dimensional unit ball
- $||\theta_*||$ is known.
- The derivative $\dot{\mu}$ is axis symmetric and monotonous

Formally define the sets

$$\begin{aligned} M_k &:= \{\mu \in C_1([-1,1]) | \forall x\, \dot{\mu} \geq 0, \\ &\qquad \dot{\mu}(x) = \dot{\mu}(-x), \mu(1) - \mu(1-x) \geq kx\} \\ M_k^1 &:= \{\mu \in M_k | \mu|_{[0,1]} \text{ convex})\} \\ M_k^2 &:= \{\mu \in M_k | \mu|_{[0,1]} \text{ concave})\} \end{aligned}$$

Results

Theorem 3. *For any GLB problem with $\mathcal{X} = \mathcal{B}_1^d, \mu \in M_k^1$ and $||\theta_*|| = 1$, an Exploit-then-Commit algorithm for a fixed time T will with probability $1-\delta$ suffer a regret that is not larger than*

$$\text{Reg}(T) \leq cT^{\left(\frac{d+1}{d+3}\right)} \cdot (\kappa\sigma^2 \left(\log(T)^2 + \log(\delta^{-1})\right))^p$$

While this is not an upper bound for any link function, it is wide enough to proof that a true gap exists between pseudo-MLE and LSE in this setting. The second theorem answers the question about safe-to-use function and is relevant for practical applications

Theorem 4. *For any GLB problem with $\mathcal{X} = \mathcal{B}_1^d, \mu \in M_k^2$ and $||\theta_*|| = 1$, an Exploit-then-Commit algorithm for a fixed time T will with probability $1-\delta$ suffer a regret that is not larger than*

$$\text{Reg}(T) \leq c\sqrt{T} \cdot (\kappa\sigma^2 \left(\log(T)^2 + \log(\delta^{-1})\right))^p$$

Proofs We will require the following definitions and lemmas for both proofs.

$$\begin{aligned} L(\theta) &:= \mathbb{E}\left[(\mu(x^T\theta) - R)^2\right] \\ &= \mathbb{E}\left[(\mu(x^T\theta) - \mu(x^T\theta_*) - \eta)^2\right] \\ &= \mathbb{E}\left[(\mu(x^T\theta) - \mu(x^T\theta_*))^2\right] + \sigma^2 \\ L_n(\theta) &= \frac{1}{n}\sum_{k=1}^n (\mu(x_k^T\theta) - R_k)^2 \\ &= \frac{1}{n}\sum_{k=1}^n (\mu(x_k^T\theta) - \mu(x_k^T\theta_*) - \eta_k)^2 \end{aligned}$$

Obviously θ_* is the minimum of L.
The least square estimator is

$$\hat{\theta} := \arg\min_\theta L_n(\theta) \tag{2}$$

The proofs for the main theorems will rely on an upper and lower bound of $L(\hat{\theta}) - L(\theta_*)$ that provides with high probability an upper bound of the immediate regret of the LSE.

Step 1: upper bounds on the difference $L(\hat{\theta}) - L(\theta_*)$

Theorem 5. *For any $\mu \in M_k$ it holds With probability at least $1-\delta$, that*

$$L(\hat{\theta}) - L(\theta_*) \leq \frac{8}{n}\sigma^2(c_1 \log(2n) + c_2 \log \delta^{-1}) + C_?$$

Proof.

Lemma 6.

$$L(\hat{\theta}) - L(\theta_*) \leq C_? + \frac{1}{n}\sum_{k=1}^{n}(\mu(x_k^T\hat{\theta}) - \mu(x_k^T\theta_*))^2$$

Using this lemma, it is sufficient to bound the second term. As $\hat{\theta}$ is the minimizer of L_n, it holds that

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n}(\mu(x_k^T\hat{\theta}) - \mu(x_k^T\theta_*))^2 \\
\leq &\frac{1}{n}\sum_{k=1}^{n}(\mu(x_k^T\hat{\theta}) - \mu(x_k^T\theta_*))^2 + L_n(\theta_*) - L_n(\hat{\theta}) \\
= &\frac{2}{n}\sum_{k=1}^{n}(\mu(x_k^T\hat{\theta}) - \mu(x_k^T\theta_*))\eta_k \\
= &\frac{2}{\sqrt{n}}\sqrt{\frac{1}{n}\sum_{k=1}^{n}(\mu(x_k^T\hat{\theta}) - \mu(x_k^T\theta_*))^2}\alpha_\mu^T\eta
\end{aligned}$$

With $\eta = (\eta_1, \eta_2, ...)^T$ and $\alpha_\mu \in \mathbb{R}^n$ a unit vector with components

$$\alpha_k = \frac{(\mu(x_k^T\hat{\theta}) - \mu(x_k^T\theta_*))}{\sqrt{\sum_{l=1}^{n}(\mu(x_l^T\hat{\theta}) - \mu(x_l^T\theta_*))^2}}.$$

α_μ is determined by the exploration schedule, which only depends on the starting point $x_1 \in \mathcal{S}^{d-1}$. Therefore α lays on the $d-1$-dimensional manifold $\alpha_\mu(\mathcal{S}^{d-1}) \subset \mathbb{R}^n$.

Lemma 7. *With probability at least $1-\delta$, we can bound*

$$\sup_{\alpha\in\alpha_\mu(\mathcal{S}^{d-1})} \alpha^T\eta \leq C_?\sqrt{(d-1)\log(\frac{2n}{d-1})} + C_?\sqrt{\log(\delta^{-1})}$$

Using this lemma, we obtain

$$\begin{aligned}
&\sqrt{\frac{1}{n}\sum_{k=1}^{n}(\mu(x_k^T\hat{\theta}) - \mu(x_k^T\theta_*))^2} \leq \frac{2}{\sqrt{n}}\alpha_\mu^T\eta \\
&\frac{1}{n}\sum_{k=1}^{n}(\mu(x_k^T\hat{\theta}) - \mu(x_k^T\theta_*))^2 \leq \frac{4}{n}(\alpha_\mu^T\eta)^2 \\
&\frac{1}{n}\sum_{k=1}^{n}(\mu(x_k^T\hat{\theta}) - \mu(x_k^T\theta_*))^2 \\
&\overset{\text{w.h.p.}}{\leq} \frac{4}{n}\left(C_?\sqrt{(d-1)\log(\frac{2n}{d-1})} + C_?\sqrt{\log(\delta^{-1})}\right)^2 \\
&\leq \frac{8}{n}\left(C_?(d-1)\log(\frac{2n}{d-1}) + C_?\log(\delta^{-1})\right)^2
\end{aligned}$$

Combining this with Lemma 6 finalizes the proof. □

Step 2: lower bounds on the difference $L(\hat{\theta}) - L(\theta_*)$ First we need a lemma to ensure, the length of $\hat{\theta}$ doesn't matter

Lemma 8. *Let $\tilde{\theta}$ be the vector $\hat{\theta}$ rescaled such that it matches the length of θ_*. Then for any $\mu \in M_k$*

$$L(\hat{\theta}) - L(\theta_*) \geq \frac{1}{2}(L(\tilde{\theta}) - L(\theta_*)).$$

We give a different lower bound theorem for the concave-convex functions ($\mu \in M_k^1$) and convex-concave ($\mu \in M_k^2$) respectively.

Theorem 9. *For any $\mu \in M_k^1$, $||\tilde{\theta}|| = ||\theta_*|| = 1$ it holds that*

$$L(\tilde{\theta}) - L(\theta_*) \geq (\mu(1) - \mu(\tilde{\theta}^T\theta_*))(\mu(1) - \mu(0))^{\frac{d+1}{2}}$$

Proof. We can write the difference in terms of integrals

$$\begin{aligned}
L(\tilde{\theta}) - L(\theta_*) &= \mathbb{E}\left[(\mu(x^T\tilde{\theta}) - \mu(x^T\theta_*))^2\right] \\
&= \frac{1}{|\mathcal{S}^{d-1}|}\int_{\mathcal{S}^{d-1}}\left(\mu(x^T\tilde{\theta}) - \mu(x^T\theta_*)\right)^2 dx
\end{aligned}$$

For the value inside the integral, it only matters where x intersects the span of $\tilde{\theta}$ and θ_*. Assume $(x^T\tilde{\theta})^2 + (x^T\theta_*)^2 = \sin(\alpha)^2$, then there is a $d-2$ dimensional sphere of volume $\mathcal{S}^{d-3}\cos(\alpha)^{d-3}$ satisfying the equation. We can write the integral in the following way, considering by abuse of notation $\tilde{\theta}, \theta_*$ as the 2-dimensional projections of the θ's into their common plane.

$$\begin{aligned}
&= 2\frac{|\mathcal{S}^{d-3}|}{|\mathcal{S}^{d-1}|}\int_0^{\frac{\pi}{2}}\cos(\alpha)^{d-3} \\
&\qquad \int_{\mathcal{S}^1_{\sin(\alpha)}}\left(\mu(x^T\tilde{\theta}) - \mu(x^T\theta_*)\right)^2 dx\, d\alpha \\
&= 2\frac{|\mathcal{S}^{d-3}|}{|\mathcal{S}^{d-1}|}\int_0^{\frac{\pi}{2}}\cos(\alpha)^{d-3}\sin(\alpha) \\
&\qquad \int_{\mathcal{S}^1}\left(\mu(\sin(\alpha)x^T\tilde{\theta}) - \mu(\sin(\alpha)x^T\theta_*)\right)^2 dx\, d\alpha
\end{aligned}$$

Assume the angle between $\tilde{\theta}$ and θ_* is β, then with substitution we get

$$\begin{aligned}
&= 2\frac{|\mathcal{S}^{d-3}|}{|\mathcal{S}^{d-1}|}\int_0^{\frac{\pi}{2}} d\alpha \left[\cos(\alpha)^{d-3}\sin(\alpha)\int_0^{2\pi} dx\right. \\
&\left.\left(\mu(\sin(\alpha)\cos(x - \frac{\beta}{2})) - \mu(\sin(\alpha)\cos(x + \frac{\beta}{2}))\right)^2\right] \\
&= 2\frac{|\mathcal{S}^{d-3}|}{|\mathcal{S}^{d-1}|}\int_0^{1} d\alpha \left[(1-\alpha^2)^{\frac{d-4}{2}}\alpha\int_0^{2\pi} dx\right. \\
&\left.\left(\mu(\alpha\cos(x - \frac{\beta}{2})) - \mu(\alpha\cos(x + \frac{\beta}{2}))\right)^2\right] \qquad (3)
\end{aligned}$$

Lemma 10. *For any $\mu \in M_k^1$, it holds*

$$\int_0^{2\pi} \left(\mu(\cos(x-\frac{\beta}{2})) - \mu(\cos(x+\frac{\beta}{2}))\right)^2 dx \geq (1-\cos\beta)(\mu(1)-\mu(0))^{\frac{3}{2}}$$

Using this lemma, we can upper bound (3)

$$\geq 2\frac{|\mathcal{S}^{d-3}|}{|\mathcal{S}^{d-1}|}\int_0^1 d\alpha \left[(1-\alpha^2)^{\frac{d-4}{2}}\alpha (1-\cos\beta)(\mu(\alpha)-\mu(0))^{\frac{3}{2}}\right]$$

Definition of M_k, implies that $\mu(\alpha) \geq \mu(1) - (1-\alpha)k$

$$\geq 2\frac{|\mathcal{S}^{d-3}|}{|\mathcal{S}^{d-1}|}\int_0^1 d\alpha \left[(1-\alpha^2)^{\frac{d-4}{2}}\alpha (1-\cos\beta)\max\{0,\mu(1)-\mu(0)-(1-\alpha)k\}^{\frac{3}{2}}\right]$$

Let $\mu(1)-\mu(0) = k\epsilon$, then

$$\begin{aligned}
&= 2\frac{|\mathcal{S}^{d-3}|}{|\mathcal{S}^{d-1}|}\int_{1-\epsilon}^1 d\alpha \left[(1-\alpha^2)^{\frac{d-4}{2}}\alpha (1-\cos\beta)((\epsilon-1+\alpha)k)^{\frac{3}{2}}\right] \\
&= 2\frac{|\mathcal{S}^{d-3}|}{|\mathcal{S}^{d-1}|}\int_0^\epsilon d\alpha \left[(2\alpha-\alpha^2)^{\frac{d-4}{2}}(1-\alpha) (1-\cos\beta)((\epsilon-\alpha)k)^{\frac{3}{2}}\right] \\
&= 2\frac{|\mathcal{S}^{d-3}|}{|\mathcal{S}^{d-1}|}\int_0^1 d\alpha \left[(2\alpha\epsilon-\alpha^2\epsilon^2)^{\frac{d-4}{2}}(1-\alpha\epsilon) (1-\cos\beta)((\epsilon-\alpha\epsilon)k)^{\frac{3}{2}}\epsilon\right] \\
&= 2\frac{|\mathcal{S}^{d-3}|}{|\mathcal{S}^{d-1}|}\epsilon^{\frac{d+1}{2}}\int_0^1 d\alpha \left[(2\alpha-\alpha^2\epsilon)^{\frac{d-4}{2}}(1-\alpha\epsilon) (1-\cos\beta)((1-\alpha)k)^{\frac{3}{2}}\right]
\end{aligned}$$

That concludes the proof. □

Theorem 11. *For any $\mu \in M_k^2$, $||\tilde{\theta}|| = ||\theta_*|| = 1$ it holds that*

$$L(\tilde{\theta}) - L(\theta_*) \geq (\mu(1)-\mu(\tilde{\theta}^T\theta_*))(\mu(1)-\mu(0))^{\frac{1}{2}}$$

Proof. The proof follows the argument of Theorem 8 unit 3. Lemma 8 can be bounded tighter for concave μ

Lemma 12. *For any $\mu \in M_k^2$ it holds*

$$\int_0^{2\pi} \left(\mu(\cos(x-\frac{\beta}{2})) - \mu(\cos(x+\frac{\beta}{2}))\right)^2 dx \geq (1-\cos\beta)(\mu(1)-\mu(0))$$

Using this lemma, we observe

$$\begin{aligned}
&L(\tilde{\theta}) - L(\theta_*) \\
\geq& 2\frac{|\mathcal{S}^{d-3}|}{|\mathcal{S}^{d-1}|}\int_0^1 d\alpha \left[(1-\alpha^2)^{\frac{d-4}{2}}\alpha (1-\cos\beta)(\mu(\alpha)-\mu(0))\right]
\end{aligned}$$

Concavity of $\mu|_{[0,1]}$ implies $\mu(\alpha) - \mu(0) \geq \alpha(\mu(1) - \mu(0))$.

$$\geq 2\frac{|\mathcal{S}^{d-3}|}{|\mathcal{S}^{d-1}|}\int_0^1 d\alpha \left[(1-\alpha^2)^{\frac{d-4}{2}}\alpha^2 (1-\cos\beta)(\mu(1)-\mu(0))\right]$$

□

Proof of theorem 3. We combine Lemma 6, theorem 7 and Lemma 5. Setting $\mu(1)-\mu(0) = \epsilon$, $\mu(1) - \mu(\tilde{\theta}^T\theta_*) = \lambda$ and using $\epsilon^{-1}T \wedge \left(\frac{d+1}{d+3}\right)$ many exploration steps, it holds

$$\begin{aligned}
C_?\epsilon T \wedge \left(-\frac{1}{2} - \frac{d-1}{2d+6}\right) &\geq \lambda\epsilon \wedge \left(\frac{d+1}{2}\right) \\
\lambda &\leq C_?\epsilon \wedge \left(-\frac{d-1}{2}\right) T \wedge \left(-\frac{1}{2} - \frac{d-1}{2d+6}\right)
\end{aligned}$$

Additionally λ is bounded by $\mu(1)-\mu(0) = 2\epsilon$

$$\begin{aligned}
&\min\{C_?\epsilon \wedge \left(-\frac{d-1}{2}\right) T \wedge \left(\frac{d+1}{d+3}\right), 2\epsilon\} \\
\leq& C_?T \wedge \left(-\frac{2}{d+3}\right)
\end{aligned}$$

The first term in the minimum is decreasing in ϵ, while the second term is increasing. Therefore the minimum is obtained if both terms are equal. Playing an arm with immediate regret smaller then $T \wedge \left(-\frac{2}{d+3}\right)$ for T steps results in a regret of

$$\mathrm{Reg}(T) \leq C_?T \wedge \left(\frac{d+1}{d+3}\right)$$

We set the number of exploration steps, such that with an average immediate regret of ϵ, we suffer exactly the same total regret during exploration, which concludes the proof. □

Proof of theorem 4. The proof goes analogous to theorem 3. We combine Lemma 6, theorem 8 and Lemma 5. Setting $\mu(1)-\mu(0) = \epsilon$, $\mu(1) - \mu(\tilde{\theta}^T\theta_*) = \lambda$ and using $\epsilon^{-1}T \wedge \left(\frac{1}{2}\right)$ many exploration steps, it holds

$$\begin{aligned}
C_?\epsilon T \wedge \left(-\frac{1}{2}\right) &\geq \lambda\epsilon \\
\lambda &\leq C_?T \wedge \left(-\frac{1}{2}\right)
\end{aligned}$$

Continuing for T timesteps with this immediate regret bounds the total regret to

$$\text{Reg}(T) \leq \sqrt{T} + C_? \sqrt{T} \leq C_? \sqrt{T}.$$

□

5 Discussion

We have proven that the generalized linear bandit problem, how it was originally proposed, can make it next to impossible to learn a good arm in a finite time. This should be a reminder to take constants seriously that tend to grow to infinity for interesting applications. We have good reason to belief that our lower bounds are tight, even though we assumed a restricted problem. In order to get a general lower bound, one would need to take the following steps.

- Proof lemma **??**or all μ. This is likely to hold true as convex function appear to be the worst case.
- Replace the fixed exploration time by a stopping time criteria if only an upper bound of $||\theta_*||$ is available instead of the exact value.

However this direction of research cannot lead to applicable algorithms, as we have shown in our lower bound. Therefore we will pursue the restriction to convex-concave link functions such as the sigmoid. Our results give hope that those functions might always perform better than the linear function, including all constants in the bound. We will work on ... Tompson Sampling ...

Acknowledgements

Use unnumbered third level headings for the acknowledgements. All acknowledgements go at the end of the paper. Be sure to omit any identifying information in the initial double-blind submission!

References

[1] C. E. Shannon, "Probability of error for optimal codes in a gaussian channel," *Bell System Technical Journal*, vol. 38, no. 3, pp. 611–656, 1959.

[2] A. Wyner, *Capabilities of Bounded Discrepancy Decoding*. Bell Telephone System; Technical publications; monograph, American Telephone and Telegraph Company, 1965.

6 Appendix

Lemma (6)**.**

$$L(\hat{\theta}) - L(\theta_*) \leq C_? + \frac{1}{n}\sum_{k=1}^{n}(\mu(x_k^T\hat{\theta}) - \mu(x_k^T\theta_*))^2$$

Proof. Per definition of L:

$$L(\hat{\theta}) - L(\theta_*) = \mathbb{E}\left[(\mu(\hat{\theta}^T x) - \mu(\theta_*^T x))^2\right].$$

We need to bound the discrete integral approximation

$$\frac{1}{|\mathcal{S}^{d-1}|}\int_{|\mathcal{S}^{d-1}|}(\mu(\hat{\theta}^T x) - \mu(\theta_*^T x))^2\,dx - \frac{1}{n}\sum_{k=1}^{n}(\mu(x_k^T\hat{\theta}) - \mu(x_k^T\theta_*))^2$$

... □

Lemma (7)**.** *With probability at least $1-\delta$, we can bound*

$$\sup_{\alpha\in\alpha_\mu(\mathcal{S}^{d-1})} \alpha^T\eta \leq C_?\sqrt{(d-1)\log(\frac{2n}{d-1})} + C_?\sqrt{\log(\delta^{-1})}$$

Proof. transfer from Problem 1.6 on MIT18_S997S15_Chapter1.pdf
Todo: find an ϵ net that covers $\alpha(\mathcal{S}^{d-1})$

□

Lemma (8)**.** *Let $\tilde{\theta}$ be the vector $\hat{\theta}$ rescaled such that it matches the length of θ_*. Then for any $\mu \in M_k$*

$$L(\hat{\theta}) - L(\theta_*) \geq \frac{1}{2}(L(\tilde{\theta}) - L(\theta_*)).$$

Proof. Let θ_1 and θ_2 be the two orthogonal vectors such that $\theta_* = \theta_1 - \theta_2$ and $\tilde{\theta} = \theta_1 + \theta_2$.
Define the two sets

$$\mathcal{B}_1 = \{x|\,\mathrm{sign}(x^T\theta_1) = \mathrm{sign}(x^T\theta_2)\}$$
$$\mathcal{B}_2 = \{x|\,\mathrm{sign}(x^T\theta_1) = -\mathrm{sign}(x^T\theta_2)\}.$$

Due to Symmetry it holds that

$$\mathbb{E}\left[(\mu(x^T\tilde{\theta}) - \mu(x^T\theta_*))^2|x\in\mathcal{B}_1\right] = \mathbb{E}\left[(\mu(x^T\tilde{\theta}) - \mu(x^T\theta_*))^2|x\in\mathcal{B}_2\right]$$

If $||\hat{\theta}||_2 > ||\tilde{\theta}||_2$, then for all $x \in \mathcal{B}_1$

$$(\mu(x^T\hat{\theta}) - \mu(x^T\theta_*))^2 \geq (\mu(x^T\tilde{\theta}) - \mu(x^T\theta_*))^2.$$

If $||\hat{\theta}||_2 < ||\tilde{\theta}||_2$, then for all $x \in \mathcal{B}_2$

$$(\mu(x^T\hat{\theta}) - \mu(x^T\theta_*))^2 \geq (\mu(x^T\tilde{\theta}) - \mu(x^T\theta_*))^2.$$

Finally we get

$$\begin{aligned}L(\hat{\theta}) - L(\theta_*) &= \mathbb{E}\left[(\mu(x^T\hat{\theta}) - \mu(x^T\theta_*))^2\right]\\ &= \frac{1}{2}\left(\mathbb{E}\left[(\mu(x^T\hat{\theta}) - \mu(x^T\theta_*))^2|x\in\mathcal{B}_1\right]\right.\\ &\quad\left.+\mathbb{E}\left[(\mu(x^T\hat{\theta}) - \mu(x^T\theta_*))^2\right]|x\in\mathcal{B}_2\right)\\ &\geq \frac{1}{2}\mathbb{E}\left[(\mu(x^T\tilde{\theta}) - \mu(x^T\theta_*))^2\right] = \frac{1}{2}(L(\tilde{\theta}) - L(\theta_*))\end{aligned}$$

□

Lemma (10)**.** *For any $\mu \in M_k^1$, it holds*

$$\int_0^{2\pi}\left(\mu(\cos(x-\frac{\beta}{2})) - \mu(\cos(x+\frac{\beta}{2}))\right)^2 dx \geq (1-\cos\beta)(\mu(1)-\mu(0))^{\frac{3}{2}}$$

Proof. □

Lemma (12)**.** *For any $\mu \in M_k^2$ it holds*

$$\int_0^{2\pi}\left(\mu(\cos(x-\frac{\beta}{2})) - \mu(\cos(x+\frac{\beta}{2}))\right)^2 dx \geq (1-\cos\beta)(\mu(1)-\mu(0))$$

Proof. □