On Worst-Case Regret of Linear Thompson Sampling

Nima Hamidi

Stanford University

Collaborator: Mohsen Bayati

Preprint: arXiv 2006.06790
Overview

1. Problem Definition

2. Confidence-based Policies

3. Failure of LinTS 😞

4. Positive Results 😊
Stochastic Linear Bandit Problem

- Let $\Theta^* \in \mathbb{R}^d$ be fixed (and unknown).
- At time $t$, the action set $\mathcal{A}_t \subseteq \mathbb{R}^d$ is revealed to a policy $\pi$.
- The policy chooses $\tilde{\mathcal{A}}_t \in \mathcal{A}_t$.
- It observes a reward $r_t = \langle \Theta^*, \tilde{\mathcal{A}}_t \rangle + \epsilon_t$.
- Conditional on the history, $\epsilon_t$ has zero mean.
Evaluation Metric

- The objective is to improve using past experiences.
- The cumulative regret is defined as

\[
\text{Regret}(T, \Theta^*, \pi) := \mathbb{E} \left[ \sum_{t=1}^{T} \sup_{A \in \mathcal{A}_t} \langle \Theta^*, A \rangle - \langle \Theta^*, \tilde{A}_t \rangle \bigg| \Theta^* \right].
\]
Evaluation Metric

- The objective is to **improve using past experiences**.

- The **cumulative regret** is defined as

\[
\text{Regret}(T, \Theta^*, \pi) := \mathbb{E} \left[ \sum_{t=1}^{T} \sup_{A \in A_t} \langle \Theta^*, A \rangle - \langle \Theta^*, \tilde{A}_t \rangle \right| \Theta^*].
\]
Evaluation Metric

- The objective is to **improve using past experiences**.

- The **cumulative regret** is defined as

  \[
  \text{Regret}(T, \Theta^*, \pi) := \mathbb{E} \left[ \sum_{t=1}^{T} \sup_{A \in \mathcal{A}_t} \langle \Theta^*, A \rangle - \langle \Theta^*, \tilde{A}_t \rangle \middle| \Theta^* \right].
  \]

- In the Bayesian setting, the **Bayesian regret** is given by

  \[
  \text{BayesRegret}(T, \pi) := \mathbb{E}_{\Theta^* \sim P} [\text{Regret}(T, \Theta^*, \pi)].
  \]
Algorithms
Greedy

At time $t = 1, 2, \cdots, T$:

- Using the set of observations
  \[ \mathcal{H}_{t-1} := \{ (\tilde{A}_1, r_1), \cdots, (\tilde{A}_{t-1}, r_{t-1}) \}, \]

- Construct an estimate $\hat{\Theta}_{t-1}$ for $\Theta^*$,

- Choose the action $A \in \mathcal{A}_t$ with largest $\langle A, \hat{\Theta}_{t-1} \rangle$.
Greedy

The **ridge estimator** is used to obtain $\hat{\Theta}_t$ (for a fixed $\lambda$):

\[ V_t := \lambda I + \sum_{i=1}^{t} \tilde{A}_i \tilde{A}_i^\top \in \mathbb{R}^{d \times d}, \]  

(1)

and

\[ \hat{\Theta}_t := V_t^{-1} \left( \sum_{i=1}^{t} \tilde{A}_i r_i \right) \in \mathbb{R}^{d}. \]  

(2)
Greedy

Algorithm 1 Greedy algorithm

1: for $t = 1$ to $T$ do
2: \hspace{1em} Pull $\tilde{A}_t := \arg \max_{A \in \mathcal{A}_t} \langle A, \hat{\Theta}_{t-1} \rangle$
3: \hspace{1em} Observe the reward $r_t$
4: \hspace{1em} Compute $V_t = \lambda I + \sum_{i=1}^{t} \tilde{A}_i \tilde{A}_i^\top$
5: \hspace{1em} Compute $\hat{\Theta}_t = V_t^{-1} \left( \sum_{i=1}^{t} \tilde{A}_i r_i \right)$
6: end for

Greedy makes wrong decisions due to over- or under-estimating the true rewards. The over-estimation is automatically corrected. The under-estimation can cause linear regret.
Greedy

Algorithm 1 Greedy algorithm

1: for $t = 1$ to $T$ do
2: Pull $\tilde{A}_t := \arg \max_{A \in A_t} \langle A, \hat{\Theta}_{t-1} \rangle$
3: Observe the reward $r_t$
4: Compute $V_t = \lambda I + \sum_{i=1}^{t} \tilde{A}_i \tilde{A}_i^\top$
5: Compute $\hat{\Theta}_t = V_{t}^{-1} \left( \sum_{i=1}^{t} \tilde{A}_i r_i \right)$
6: end for

Greedy makes wrong decisions due to over- or under-estimating the true rewards.

- The over-estimation is automatically corrected.
- The under-estimation can cause linear regret.
Greedy

$A_1$ $A_2$ $A_3$ $A_4$ $A_5$
Greedy

A_1  A_2  A_3  A_4  A_5
Optimism in Face of Uncertainty (OFU) Algorithm

- Key idea: be optimistic when estimating the reward of actions.
Optimism in Face of Uncertainty (OFU) Algorithm

- Key idea: **be optimistic** when estimating the reward of actions.
- For $\rho > 0$, define the confidence set $C_t(\rho)$ to be

$$C_t(\rho) := \{\Theta | \|\Theta - \hat{\Theta}_t\|_{V_t} \leq \rho\},$$

where

$$\|X\|_{V_t}^2 = X^\top V_t X \in \mathbb{R}^+. $$

Theorem (Informal, Abbasi-Yadkori, Pál, and Szepesvári 2011) Letting $\rho := \tilde{O}(\sqrt{d})$, we have $\Theta^* \in C_t(\rho)$ with high probability.
Optimism in Face of Uncertainty (OFU) Algorithm

- Key idea: be optimistic when estimating the reward of actions.
- For $\rho > 0$, define the confidence set $C_t(\rho)$ to be

$$C_t(\rho) := \{ \Theta \mid \| \Theta - \hat{\Theta}_t \|_{V_t} \leq \rho \},$$

where

$$\|X\|_{V_t}^2 = X^T V_t X \in \mathbb{R}^+.$$

Theorem (Informal, Abbasi-Yadkori, Pál, and Szepesvári 2011)

Letting $\rho := \tilde{O}(\sqrt{d})$, we have $\Theta^* \in C_t(\rho)$ with high probability.
## Algorithm 2 OFUL algorithm

1: for $t = 1$ to $T$ do  
2: Pull $\tilde{A}_t := \arg \max_{A \in \mathcal{A}_t} \sup_{\Theta \in \mathcal{C}_{t-1}(\rho)} \langle A, \Theta \rangle$  
3: Observe the reward $r_t$  
4: Compute $V_t = \lambda I + \sum_{i=1}^{t} \tilde{A}_i \tilde{A}_i^T$  
5: Compute $\hat{\Theta}_t = V_t^{-1} \left( \sum_{i=1}^{t} \tilde{A}_i r_i \right)$  
6: end for
Optimism in Face of Uncertainty (OFU) Algorithm

Algorithm 2 OFUL algorithm

1: for $t = 1$ to $T$ do
2:     Pull $\tilde{A}_t := \text{arg max}_{A \in A_t} \sup_{\Theta \in C_{t-1}(\rho)} \langle A, \Theta \rangle$
3:     Observe the reward $r_t$
4:     Compute $V_t = \lambda I + \sum_{i=1}^{t} \tilde{A}_i \tilde{A}_i^T$
5:     Compute $\hat{\Theta}_t = V_t^{-1} \left( \sum_{i=1}^{t} \tilde{A}_i r_i \right)$
6: end for

It can be shown that

$$\sup_{\Theta \in C_t(\rho)} \langle A, \Theta \rangle = \langle A, \hat{\Theta}_t \rangle + \rho \| A \| V_{t-1}^{-1}.$$
Optimism in Face of Uncertainty (OFU) Algorithm

\[ A_1 \quad A_2 \quad A_3 \quad A_4 \quad A_5 \]

Greedy
Optimism in Face of Uncertainty (OFU) Algorithm

OFUL

Greedy

A₁ A₂ A₃ A₄ A₅
Linear Thompson Sampling (LinTS) Algorithm

- Key idea: use **randomization** to address under-estimation.
Linear Thompson Sampling (LinTS) Algorithm

- Key idea: use randomization to address under-estimation.
- LinTS samples from the posterior distribution of $\Theta^*$.

**Algorithm 3 LinTS algorithm**

1. for $t = 1$ to $T$ do
2. Sample $\tilde{\Theta}_t \sim P(\Theta^* | \mathcal{H}_{t-1})$
3. Pull $A_t := \arg \max_{A \in \mathcal{A}_t} \langle A, \tilde{\Theta}_t \rangle$
4. Observe the reward $r_t$
5. Update $\mathcal{H}_t \leftarrow \mathcal{H}_{t-1} \cup \{(A_t, r_t)\}$
6. end for
Linear Thompson Sampling (LinTS) Algorithm

- Under **normality**, LinTS becomes:

```
Algorithm 4 LinTS algorithm under normality
1:   for $t = 1$ to $T$ do
2:       Sample $\tilde{\Theta}_t \sim \mathcal{N}(\hat{\Theta}_{t-1}, V_{t-1})$
3:       Pull $A_t := \arg \max_{A \in A_t} \langle A, \tilde{\Theta}_t \rangle$
4:       Observe the reward $r_t$
5:       Compute $V_t = \lambda I + \sum_{i=1}^{t} \tilde{A}_i \tilde{A}_i^\top$
6:       Compute $\hat{\Theta}_t = V_t^{-1} \left( \sum_{i=1}^{t} \tilde{A}_i r_i \right)$
7:   end for
```
Linear Thompson Sampling (LinTS) Algorithm

$A_1$ $A_2$ $A_3$ $A_4$ $A_5$

OFUL
Greedy
Linear Thompson Sampling (LinTS) Algorithm

$A_1$  $A_2$  $A_3$  $A_4$  $A_5$

OFUL

Greedy
Linear Thompson Sampling (LinTS) Algorithm

\[ \text{LinTS} \rightarrow \text{OFUL} \]

\[ \text{Greedy} \]

\[ A_1 \quad A_2 \quad A_3 \quad A_4 \quad A_5 \]
Why Is LinTS Popular?

- **Empirical superiority:**
  - \( d = 120, \Theta^* \sim \mathcal{N}(0, I_d), \)
  - \( k = 10, X \sim \mathcal{N}(0, I_{12}), \)
  - Each \( A_t \) contains \( X \) as a block\(^1\).

---

\(^1\) This is the 10-armed contextual bandit with 12 dimensional covariates.
Why is LinTS Popular?

- **Computation efficiency**: when $A_t$ is a polytope \ldots
  - LinTS solves an LP problem,

- OFUL becomes an NP-hard problem!

Photo credit: Russo and Van Roy 2014
Comparison of Regret Bounds

Theorem (Abbasi-Yadkori, Pál, and Szepesvári 2011)

Under some conditions, the regret of OFUL is bounded by

\[ \text{Regret}(T, \Theta^*, \pi^{\text{OFUL}}) \leq \tilde{O}(d\sqrt{T}). \]
Comparison of Regret Bounds

**Theorem (Abbasi-Yadkori, Pál, and Szepesvári 2011)**

*Under some conditions, the regret of OFUL is bounded by*

\[
\text{Regret}(T, \Theta^*, \pi^{OFUL}) \leq \tilde{O}(d\sqrt{T}).
\]

**Theorem (Russo and Van Roy 2014)**

*Under minor assumptions, the Bayesian regret of LinTS is bounded by*

\[
\text{BayesRegret}(T, \pi^{LinTS}) \leq \tilde{O}(d\sqrt{T}).
\]
Comparison of Regret Bounds

Theorem (Abbasi-Yadkori, Pál, and Szepesvári 2011)
Under some conditions, the regret of OFUL is bounded by

\[ \text{Regret}(T, \Theta^*, \pi^{OFUL}) \leq \tilde{O}(d\sqrt{T}). \]

Theorem (Russo and Van Roy 2014)
Under minor assumptions, the Bayesian regret of LinTS is bounded by

\[ \text{BayesRegret}(T, \pi^{LinTS}) \leq \tilde{O}(d\sqrt{T}). \]

Theorem (Dani, Hayes, and Kakade 2008)
There is a Bayesian linear bandit problem that satisfies

\[ \inf_{\pi} \text{BayesRegret}(T, \pi) \geq \Omega(d\sqrt{T}). \]
A Worst-Case Regret Bound for LinTS

- Question: can one prove a similar worst-case regret bound for LinTS?
- The only known results require **inflating** the posterior variance.

**Algorithm 5** LinTS algorithm under normality

1: for $t = 1$ to $T$ do
2: Sample $\tilde{\Theta}_t \sim \mathcal{N}(\hat{\Theta}_{t-1}, \beta^2 V_{t-1}^{-1})$
3: Pull $A_t := \arg \max_{A \in A_t} \langle A, \tilde{\Theta}_t \rangle$
4: Observe the reward $r_t$
5: Compute $V_t = \lambda I + \sum_{i=1}^{t} \tilde{A}_i \tilde{A}_i^\top$
6: Compute $\hat{\Theta}_t = V_t^{-1} \left( \sum_{i=1}^{t} \tilde{A}_i r_i \right)$
7: end for
Theorem (Abeille and Lazaric 2017; Agrawal and Goyal 2013)

If $\beta \propto \sqrt{d}$, then

$$\text{Regret}(T, \Theta^*, \pi^{LinTS}) \leq \tilde{O}(d\sqrt{dT}).$$

This result is far from optimal by a $\sqrt{d}$ factor.
Empirical Performance of Inflated LinTS

- Unfortunately, the inflated variant of LinTS performs poorly...
A General Regret Bound
Randomized OFUL

- By a **worth function**, we mean a function $\tilde{M}_t$ that maps each $A \in \mathcal{A}_t$ to $\mathbb{R}$ such that

  $$|\tilde{M}_t(A) - \langle A, \hat{\Theta}_{t-1} \rangle| \leq \rho \|A\|_{V_{t-1}^{-1}}$$

  with probability at least $1 - \frac{1}{T^2}$. 

Next, define Randomized OFUL (ROFUL) to be:

Algorithm 6

1: for $t = 1$ to $T$
2: Pull $\tilde{A}_t := \arg \max_{A \in \mathcal{A}_t} \tilde{M}_t(A)$
3: Observe the reward $r_t$
4: Compute $V_t = \lambda I + \sum_{t=1}^{t-1} \tilde{A}_i \tilde{A}_i^\top$
5: Compute $\hat{\Theta}_t = V_{t-1}^{-1}(\sum_{t=1}^{t-1} \tilde{A}_i r_i)$
6: end for
Randomized OFUL

- By a **worth function**, we mean a function $\tilde{M}_t$ that maps each $A \in \mathcal{A}_t$ to $\mathbb{R}$ such that

$$|\tilde{M}_t(A) - \langle A, \hat{\Theta}_{t-1} \rangle| \leq \rho \|A\|_V^{-1}$$

with probability at least $1 - \frac{1}{T^2}$.

- Next, define **Randomized OFUL (ROFUL)** to be:

```
Algorithm 6 ROFUL algorithm
1: for $t = 1$ to $T$ do
2:   Pull $\tilde{A}_t := \arg \max_{A \in \mathcal{A}_t} \tilde{M}_t(A)$
3:   Observe the reward $r_t$
4:   Compute $V_t = \lambda I + \sum_{i=1}^{t} \tilde{A}_i \tilde{A}_i^\top$
5:   Compute $\hat{\Theta}_t = V_t^{-1} \left( \sum_{i=1}^{t} \tilde{A}_i r_i \right)$
6: end for
```
Examples of worth functions:

- **Greedy:** \( \tilde{M}_t(A) = \langle A, \hat{\Theta}_{t-1} \rangle \)

- **OFUL:** \( \tilde{M}_t(A) = \langle A, \hat{\Theta}_{t-1} \rangle + \rho \|A\| \sqrt{V_{t-1}} \)

- **LinTS:** \( \tilde{M}_t(A) = \langle A, \tilde{\Theta}_{t-1} \rangle \)
Definition

We say a worth function $\tilde{M}_t$ is **optimistic** if

$$\sup_{A \in \mathcal{A}_t} \tilde{M}_t(A) \geq \sup_{A \in \mathcal{A}_t} \langle A, \Theta^* \rangle$$

with probability at least $p$. 

Theorem

Let $(\tilde{M}_t)$ be a sequence of optimistic worth functions. Then, the regret of ROFUL with this worth function is bounded by

$$\text{Regret}(T, \pi_{\text{ROFUL}}) \leq \tilde{O}(\rho \sqrt{dT}p)$$
A General Regret Bound

Definition

We say a worth function \( \tilde{M}_t \) is optimistic if

\[
\sup_{A \in A_t} \tilde{M}_t(A) \geq \sup_{A \in A_t} \langle A, \Theta^* \rangle \tag{3}
\]

with probability at least \( p \).

Theorem

Let \( (\tilde{M}_t)_{t=1}^T \) be a sequence of optimistic worth functions. Then, the regret of ROFUL with this worth function is bounded by

\[
\text{Regret}(T, \pi^{\text{ROFUL}}) \leq \tilde{O}\left( \rho \sqrt{\frac{dT}{p}} \right).
\]
A Sufficient Condition for Optimism

Recall that the worth function for LinTS is given by

\[ \tilde{M}_t(A) = \langle A, \tilde{\Theta}_t \rangle. \]
A Sufficient Condition for Optimism

- Recall that the worth function for LinTS is given by
  \[ \tilde{M}_t(A) = \langle A, \tilde{\Theta}_t \rangle. \]

- We can decompose it as
  \[ \tilde{M}_t(A) = \langle A, \tilde{\Theta}_t - \hat{\Theta}_{t-1} \rangle + \langle A, \hat{\Theta}_{t-1} - \Theta^* \rangle + \langle A, \Theta^* \rangle. \]
A Sufficient Condition for Optimism

- Recall that the worth function for LinTS is given by
  \[ \tilde{M}_t(A) = \langle A, \tilde{\Theta}_t \rangle. \]

- We can decompose it as
  \[ \tilde{M}_t(A) = \langle A, \tilde{\Theta}_t - \hat{\Theta}_t \rangle + \langle A, \hat{\Theta}_t - \Theta^* \rangle + \langle A, \Theta^* \rangle. \]

- Hence, we have
  \[
  \sup_{A \in A_t} \tilde{M}_t(A) - \sup_{A \in A_t} \langle A, \Theta^* \rangle \geq \tilde{M}_t(A_t^*) - \langle A_t^*, \Theta^* \rangle.
  \]
A Sufficient Condition for Optimism

- Recall that the worth function for LinTS is given by
  \[ \tilde{M}_t(A) = \langle A, \tilde{\Theta}_t \rangle. \]

- We can decompose it as
  \[ \tilde{M}_t(A) = \langle A, \tilde{\Theta}_t - \hat{\Theta}_{t-1} \rangle + \langle A, \hat{\Theta}_{t-1} - \Theta^* \rangle + \langle A, \Theta^* \rangle. \]

- Hence, we have
  \[
  \sup_{A \in \mathcal{A}_t} \tilde{M}_t(A) - \sup_{A \in \mathcal{A}_t} \langle A, \Theta^* \rangle \geq \tilde{M}_t(A^*_t) - \langle A^*_t, \Theta^* \rangle \]
  \[
  = \langle A^*_t, \tilde{\Theta}_t - \hat{\Theta}_{t-1} \rangle + \langle A^*_t, \hat{\Theta}_{t-1} - \Theta^* \rangle. \]
A Sufficient Condition for Optimism

- Recall that the worth function for LinTS is given by
  \[ \tilde{M}_t(A) = \langle A, \tilde{\Theta}_t \rangle. \]

- We can decompose it as
  \[ \tilde{M}_t(A) = \langle A, \tilde{\Theta}_t - \hat{\Theta}_{t-1} \rangle + \langle A, \hat{\Theta}_{t-1} - \Theta^* \rangle + \langle A, \Theta^* \rangle. \]

- Hence, we have
  \[ \sup_{A \in \mathcal{A}_t} \tilde{M}_t(A) - \sup_{A \in \mathcal{A}_t} \langle A, \Theta^* \rangle \geq \tilde{M}_t(A^*_t) - \langle A^*_t, \Theta^* \rangle \]
  \[ = \langle A^*_t, \tilde{\Theta}_t - \hat{\Theta}_{t-1} \rangle + \langle A^*_t, \hat{\Theta}_{t-1} - \Theta^* \rangle. \]
A Sufficient Condition for Optimism

- Recall that the worth function for LinTS is given by
  \[ \tilde{M}_t(A) = \langle A, \tilde{\Theta}_t \rangle. \]

- We can decompose it as
  \[ \tilde{M}_t(A) = \langle A, \tilde{\Theta}_t - \hat{\Theta}_{t-1} \rangle + \langle A, \hat{\Theta}_{t-1} - \Theta^* \rangle + \langle A, \Theta^* \rangle. \]

- Hence, we have
  \[
  \sup_{A \in \mathcal{A}_t} \tilde{M}_t(A) - \sup_{A \in \mathcal{A}_t} \langle A, \Theta^* \rangle \\
  \geq \tilde{M}_t(A^*_t) - \langle A^*_t, \Theta^* \rangle \\
  = \langle A^*_t, \tilde{\Theta}_t - \hat{\Theta}_{t-1} \rangle + \langle A^*_t, \hat{\Theta}_{t-1} - \Theta^* \rangle.
  \]

  - Compensation term
  - Error term
A Sufficient Condition for Optimism

Define

- **Error vector** $E := \Theta^* - \hat{\Theta}_{t-1}$
- **Compensator vector** $C := \Theta_t - \hat{\Theta}_{t-1}$

The optimism assumption holds if, with probability $p$, the following holds

$$ \langle A_t^*, C \rangle \geq \langle A_t^*, E \rangle.$$

N. Hamidi, M. Bayati
On Worst-Case Regret of LinTS
Stanford University
Omniscient Adversary and LinTS

- An adversary chooses $A_t$ at time $t$.
- The adversary is omniscient if he knows $\hat{\Theta}_{t-1}$ and $\Theta^*$.
An adversary chooses $A_t$ at time $t$.

The adversary is omniscient if he knows $\hat{\Theta}_{t-1}$ and $\Theta^*$.

The adversary sets $A_t := \{0, A\}$ where $A = -c\hat{\Theta}_{t-1} + E$. 
Omniscient Adversary and LinTS

- An adversary chooses $A_t$ at time $t$.
- The adversary is omniscient if he knows $\hat{\Theta}_{t-1}$ and $\Theta^*$.
- The adversary sets $A_t := \{0, A\}$ where $A = -c\hat{\Theta}_{t-1} + E$.
- For simplicity, assume that $\|\Theta^*\|_2 = \|E\|_2 = \sqrt{d}$, and $V_{t-1} = I$. 
Omniscient Adversary and LinTS

- An **adversary** chooses $A_t$ at time $t$.
- The adversary is **omniscient** if he knows $\hat{\Theta}_{t-1}$ and $\Theta^*$.
- The adversary sets $A_t := \{0, A\}$ where $A = -c\hat{\Theta}_{t-1} + E$.
- For simplicity, assume that $\|\Theta^*\|_2 = \|E\|_2 = \sqrt{d}$, and $V_{t-1} = I$.
- Then $c > 0$ is chosen so that

$$\langle A, \Theta^* \rangle > 0$$
Omniscient Adversary and LinTS

- An **adversary** chooses $A_t$ at time $t$.
- The adversary is **omniscient** if he knows $\hat{\Theta}_{t-1}$ and $\Theta^*$.
- The adversary sets $A_t := \{0, A\}$ where $A = -c\hat{\Theta}_{t-1} + E$.
- For simplicity, assume that $\|\Theta^*\|_2 = \|E\|_2 = \sqrt{d}$, and $V_{t-1} = I$.
- Then $c > 0$ is chosen so that

$$\langle A, \Theta^* \rangle > 0 \quad \text{and} \quad \langle A, \hat{\Theta}_{t-1} \rangle < -\frac{1}{2} \cdot \|A\|_{V_{t-1}} \cdot \|E\|_{V_{t-1}} \ll 0.$$
Omniscient Adversary and LinTS

- The adversary sets $A_t = \{0, A\}$.

- LinTS chooses $A$ if and only if
  \[
  \langle A, \tilde{\Theta}_t \rangle = \langle A, \tilde{\Theta}_t - \hat{\Theta}_{t-1} \rangle + \langle A, \hat{\Theta}_{t-1} \rangle > 0.
  \]

- This requires
  \[
  \langle A, C \rangle \sim \mathcal{N}(0, V_{t-1}^{-1}) > \frac{1}{2} \cdot \|A\|_{V_{t-1}^{-1}} \cdot \|E\|_{V_{t-1}^{-1}} \approx \sqrt{d}.
  \]
Omniscient Adversary and LinTS

Next, we have

\[ \mathbb{P}(\langle A, \tilde{\Theta}_t \rangle > 0) \leq \exp(-\Omega(d))! \]

• LinTS chooses the optimal arm \( A \) w.p. \textit{exponentially small in} \( \Omega(d) \).

• When \( \tilde{A}_t = 0 \), the reward contains \textit{no new information} about \( \Theta^* \).

• The adversary reveals the same action set in the next rounds.

• The regret will grow \textit{linearly}. 
Bayesian Analyses are Brittle

- The key point was the **adversary’s knowledge of** $E$.

- This can be relaxed by **slightly modifying** the noise distribution.

- In this case, we can set up a problem so that $\mathbb{E}[E] \neq 0$.

- **Reducing the noise variance** reveals information about $E$. 
Bayesian Analyses are Brittle

We prove that the inflation is necessary for LinTS to work.

**Theorem**

*There exists a linear bandit problem such that for $T \leq \exp(\Omega(d))$, we have*

$$\text{BayesRegret}(T, \pi^{\text{LinTS}}) = \Omega(T).$$
Bayesian Analyses are Brittle

We prove that the inflation is necessary for LinTS to work.

Theorem

There exists a linear bandit problem such that for $T \leq \exp(\Omega(d))$, we have

$$\text{BayesRegret}(T, \pi^{\text{LinTS}}) = \Omega(T).$$

The counter-example satisfies the following properties:

- $\Theta^* \sim \mathcal{N}(0, I_d)$,
- LinTS uses the right prior,
- LinTS assumes noises are standard normal,
- $r_t = \langle \Theta^*, A_t \rangle$. (i.e., noiseless data!)
Reducing the Inflation Parameter
Reduction the Inflation Parameter

- Recall that a sufficient condition for optimism is that

\[ \langle A^*_t, C \rangle \geq \langle A^*_t, E \rangle \]

with probability \( p > 0 \).
Reducing the Inflation Parameter

- Recall that a sufficient condition for optimism is that
  \[ \langle A_t^*, C \rangle \geq \langle A_t^*, E \rangle \]
  with probability \( p > 0 \).

- Also, we have that
  \[ \langle A_t^*, C \rangle \sim \mathcal{N}(0, \beta^2 \| A_t^* \|^2 V_{t-1}) \].
Reducing the Inflation Parameter

- Recall that a sufficient condition for optimism is that
  \[ \langle A_t^*, C \rangle \geq \langle A_t^*, E \rangle \]
  with probability \( p > 0 \).

- Also, we have that
  \[ \langle A_t^*, C \rangle \sim \mathcal{N}(0, \beta^2 \| A_t^* \|^2_{\mathbf{V}_{t-1}}) \].

- And, in the **worst-case**, we have
  \[ \langle A_t^*, E \rangle \geq \rho \| A_t^* \|_{\mathbf{V}_{t-1}}. \]
Reducing the Inflation Parameter

- Recall that a sufficient condition for optimism is that

\[ \langle A_t^*, C \rangle \geq \langle A_t^*, E \rangle \]

with probability \( p > 0 \).

- Also, we have that

\[ \langle A_t^*, C \rangle \sim \mathcal{N}(0, \beta^2 \|A_t^*\|_V^2 V_t^{-1}). \]

- And, in the worst-case, we have

\[ \langle A_t^*, E \rangle \geq \rho \|A_t^*\|_V v_{t-1}. \]

- What if we assume that \( A_t^* \) is in a random direction?
Diversity Assumption

Assumption (Optimal arm diversity)

Assume that for any $V \in \mathbb{R}^d$ with $\|V\|_2 = 1$, we have

$$
\mathbb{P}\left( \langle A^*_t, V \rangle > \frac{\nu}{\sqrt{d}} \|A^*_t\|_2 \right) \leq \frac{1}{t^3},
$$

for some fixed $\nu \in [1, \sqrt{d}]$. 
Diversity is not Sufficient
Define **thinness** of a matrix $\Sigma$ to be

$$
\psi(\Sigma) := \sqrt{d \cdot \frac{\| \Sigma \|_{\text{op}}}{\| \Sigma \|_{\ast}}}
$$
Improved Worst-Case Regret Bound for LinTS

Define \textit{thinness} of a matrix \( \Sigma \) to be

\[
\psi(\Sigma) := \sqrt{\frac{d \cdot \| \Sigma \|_{op}}{\| \Sigma \|_*}}.
\]

Assumption
For \( \psi, \omega > 0 \), we have

\[
\mathbb{P}\left( \| A^* \|_{V_t^{-1}} < \omega \sqrt{\frac{\| V_t^{-1} \|_*}{d} \cdot \| A^* \|_2} \right) \leq \frac{1}{t^3}
\]

for any positive definite \( V_t^{-1} \) with \( \psi(V_t^{-1}) \leq \psi \).
Main Results

For $\beta := \frac{\nu \Psi}{\omega} \cdot \frac{\rho}{\sqrt{d}}$, optimism holds. So, we have the following result:

**Theorem**

If $\sum_{t=1}^{T} \mathbb{P}(\psi(V_t^{-1}) > \Psi) \leq C$, we have

$$\text{Regret}(T, \Theta^*, \pi^{TS}) \leq O\left(\rho \beta \sqrt{dT \log(T)} + C\right).$$
Empirical Scrutiny on Thinness

Thinness in the simulations in Russo and Van Roy (2014):
Empirical Scrutiny on Thinness

Thinness in the simulations in Russo and Van Roy (2014):

![Graph showing thinness over time for different methods: TS-Bayes, TS-Freq, TS-Improved.](image-url)
Conclusion

- Proved that LinTS without inflation can incur linear regret.
- Provided a general regret bound for confidence-based policies.
- Introduced sufficient conditions for reducing the inflation parameter.
Thank you!

Any questions?
Failure of LinTS: Example 1

| Environment          | LinTS          |
|----------------------|----------------|
| Prior \( \mathcal{N}(0, I_d) \) | \( \mathcal{N}(0, I_d) \) |
| Noise \( \mathcal{N}(0, 0) \)    | \( \mathcal{N}(0, 1) \) |

![Graph showing the expected number of failures for LinTS](image-url)
### Failure of LinTS: Example 2

| Environment | LinTS |
|-------------|-------|
| Prior       | $\mathcal{N}(0.1 \cdot 1_d, I_d)$ | $\mathcal{N}(0, I_d)$ |
| Noise       | $\mathcal{N}(0, 1)$            | $\mathcal{N}(0, 1)$            |

The expected number of failures for LinTS shows a significant increase with the dimension $d$. The box plot illustrates this trend, with the expected number of failures for LinTS increasing exponentially as $d$ increases from $2^0$ to $2^{17}$.
## Failure of LinTS: Example 2

| Environment | LinTS |
|-------------|-------|
| Prior       | $\mathcal{N}(\mu \cdot 1_{2000}, I_{2000})$ | $\mathcal{N}(0, I_{2000})$ |
| Noise       | $\mathcal{N}(0, 1)$ | $\mathcal{N}(0, 1)$ |

The graph shows the expected number of failures for LinTS as a function of $\mu$, with error bars indicating variability.
Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. “Improved algorithms for linear stochastic bandits”. In: Advances in Neural Information Processing Systems. 2011, pp. 2312–2320.

Marc Abeille, Alessandro Lazaric, et al. “Linear Thompson sampling revisited”. In: Electronic Journal of Statistics 11.2 (2017), pp. 5165–5197.

Shipra Agrawal and Navin Goyal. “Thompson Sampling for Contextual Bandits with Linear Payoffs.”. In: ICML (3). 2013, pp. 127–135.

Varsha Dani, Thomas P. Hayes, and Sham M. Kakade. “Stochastic Linear Optimization under Bandit Feedback”. In: COLT. 2008.

Daniel Russo and Benjamin Van Roy. “Learning to Optimize via Posterior Sampling”. In: Mathematics of Operations Research 39.4 (2014), pp. 1221–1243. DOI: 10.1287/moor.2014.0650.