Linear Bandits

Kianté Brantley & Sham Kakade

Outline

Recap: The "Assumption Ladder" for Linear RL

- **Linear Feature Assumptions:** We explored a hierarchy of conditions for RL:
  - Bellman completeness
  - All-policy realizability
  - Linear $Q^\star$ realizability

- **The Goal:** Can we achieve sample complexities that are $\text{poly}(d, H)$?
  Crucially, we sought no dependence on the sizes of the state/action spaces ($|S|$ or $|A|$).

- **Today:** We strip away the horizon complexity and focus on the $H = 1$ case.
  - **Linear Bandits:** How do we explore efficiently in large/infinite action spaces?

Handling Large Action Spaces

- On each round, we must choose a decision $x_t \in D \subset R^d$.
- Obtain a reward $r_t \in [-1, 1]$, where

$$r_t = \mu^\star \cdot x_t + \eta_t$$

  with i.i.d noise η_t.
- the conditional expectation of r_t is linear,

$$\mathbb{E}[r_t | x_t = x] = \mu^\star \cdot x$$

Our Objective

If $x_0, \dots x_{T-1}$ are our decisions, then our cumulative regret is

$$R_T = T\mu^\star \cdot x^\star - \sum_{t=0}^{T-1} \mu^\star \cdot x_t$$

where $x^\star \in D$ is an optimal decision for $\mu^\star$, i.e.

$$x^\star \in \underset{x \in D}{\operatorname{argmax}}\, \mu^\star \cdot x$$

The "Confidence Ball"

After t rounds, define our uncertainty region $\textsc{Ball}_t$: with center, $\widehat{\mu}_t$, and shape, Σ_t, using the λ-regularized least squares solution:

$$\begin{aligned}\widehat{\mu}_t &= \operatorname*{argmin}_{\mu} \sum_{\tau=0}^{t-1} (\mu \cdot x_\tau - r_\tau)^2 + \lambda \|\mu\|_2^2 \\ &= \Sigma_t^{-1} \sum_{\tau=0}^{t-1} r_\tau x_\tau, \\ \Sigma_t &= \lambda I + \sum_{\tau=0}^{t-1} x_\tau x_\tau^\top, \text{ with } \Sigma_0 = \lambda I.\end{aligned}$$

Define the uncertainty region:

$$\textsc{Ball}_t = \left\{ \mu \mid (\widehat{\mu}_t - \mu)^\top \Sigma_t (\widehat{\mu}_t - \mu) \leq \beta_t \right\},$$

where β_t is a parameter of the algorithm.

LinUCB (the algo)

1. Input: λ, β_t
2. Initialize: $\Sigma_0 = \lambda I$, $b_0 = 0$
3. For $t = 0, 1, \ldots$
   1. Compute $\widehat{\mu}_t = \Sigma_t^{-1} b_t$
   2. Execute

$$x_t = \operatorname*{argmax}_{x \in D} \max_{\mu \in \text{BALL}_t} \mu \cdot x$$

   and observe the reward r_t. Equivalently,

$$x_t \in \operatorname*{argmax}_{x \in D} \left(\widehat{\mu}_t \cdot x \; + \; \sqrt{\beta_t} \sqrt{x^\top \Sigma_t^{-1} x} \right)$$

   3. Update: $\Sigma_{t+1} = \Sigma_t + x_t x_t^\top$ and $b_{t+1} = b_t + r_t x_t$.

LinUCB Regret Bound

Sublinear regret: $R_T \leq O^\star(d\sqrt{T})$
poly dependence on d , no dependence on the cardinality $|D|$.

Theorem

Suppose: $|\mu^\star \cdot x| \leq 1$ *and* $\|x\| \leq B$ *for all* $x \in D$*; that the noise is* σ^2 *sub-Gaussian; and that* $\|\mu^\star\| \leq W$*. Set* $\lambda = \sigma^2/W^2$ *and*

$$\beta_t := c\ \left(\sigma^2 d \log\left(1 + \frac{tB^2W^2}{d\sigma^2}\right) + \sigma^2 \log(1/\delta)\right)$$

With probability greater than $1-\delta$*, for all* $T \geq 0$*,*

$$R_T \leq c'\ \sigma\sqrt{T}\left(d \log\left(1 + \frac{TB^2W^2}{d\sigma^2}\right) + \log(4/\delta)\right)$$

where c, c' *are absolute constants.*

Outline

Confidence

In establishing the upper bounds there are two main propositions from which the upper bounds follow. The first is in showing that the confidence region is valid.

Proposition

(Confidence) Let $\delta > 0$. We have that for our choice of β_t,

$$\Pr(\forall t,\, \mu^\star \in \textsc{Ball}_t) \geq 1 - \delta.$$

Equivalently, for all $t \geq 0$:

$$(\widehat{\mu}_t - \mu^\star)^\top \Sigma_t (\widehat{\mu}_t - \mu^\star) \leq \beta_t$$

Sum of Squares Regret Bound

Assuming the confidence event holds, the following controls on the growth of the regret.

Proposition

(Sum of Squares Regret Bound) Define:

$$\text{regret}_t = \mu^\star \cdot x^\star - \mu^\star \cdot x_t$$

Suppose $\|x\| \leq B$ *for* $x \in D$. *Suppose* β_t *is increasing and larger than* 1. *Suppose* $\mu^\star \in \text{BALL}_t$ *for all* t, *then*

$$\sum_{t=0}^{T-1} \text{regret}_t^2 \leq 4\beta_T d \log\left(1 + \frac{TB^2}{d\lambda}\right)$$

Completing the Proof

Proof of Theorem 1: With the two previous Propositions, along with the Cauchy-Schwarz inequality, we have, with probability at least $1 - \delta$,

$$R_T = \sum_{t=0}^{T-1} \text{regret}_t \leq \sqrt{T \sum_{t=0}^{T-1} \text{regret}_t^2} \leq \sqrt{4T\beta_T d \log\left(1 + \frac{TB^2}{d\lambda}\right)}.$$

The remainder of the proof follows from using our chosen value of β_T and algebraic manipulations.

Outline

"Width" of Confidence Ball

Lemma

If $\mu \in \textsc{Ball}_t$ *and* $x \in D$, *then*

$$|(\mu - \widehat{\mu}_t)^\top x| \leq \sqrt{\beta_t x^\top \Sigma_t^{-1} x}$$

"Width" of Confidence Ball

Lemma

If $\mu \in \text{BALL}_t$ and $x \in D$, then

$$|(\mu - \widehat{\mu}_t)^\top x| \leq \sqrt{\beta_t x^\top \Sigma_t^{-1} x}$$

Proof: By Cauchy-Schwarz, we have:

$$\begin{aligned}
&|(\mu - \widehat{\mu}_t)^\top x| = |(\mu - \widehat{\mu}_t)^\top \Sigma_t^{1/2} \Sigma_t^{-1/2} x| \\
&\leq \|\Sigma_t^{1/2}(\mu - \widehat{\mu}_t)\| \|\Sigma_t^{-1/2} x\| = \|\Sigma_t^{1/2}(\mu - \widehat{\mu}_t)\| \sqrt{x^\top \Sigma_t^{-1} x} \leq \sqrt{\beta_t x^\top \Sigma_t^{-1} x}
\end{aligned}$$

where the last inequality holds since $\mu \in \text{BALL}_t$.

Instantaneous Regret Lemma

Normalized width at time t for decision x_t: $\quad w_t := \sqrt{x_t^\top \Sigma_t^{-1} x_t}$

Lemma

Fix $t \leq T$. If $\mu^\star \in \mathrm{BALL}_t$, then

$$\mathrm{regret}_t \leq 2 \min(\sqrt{\beta_t} w_t, 1) \leq 2\sqrt{\beta_T} \min(w_t, 1)$$

Instantaneous Regret Lemma

Normalized width at time t for decision x_t: $\quad w_t := \sqrt{x_t^\top \Sigma_t^{-1} x_t}$

Lemma

Fix $t \leq T$. If $\mu^\star \in \mathrm{BALL}_t$, then

$$\mathrm{regret}_t \leq 2 \min(\sqrt{\beta_t} w_t, 1) \leq 2\sqrt{\beta_T} \min(w_t, 1)$$

Proof: Let $\tilde{\mu} \in \mathrm{argmax}_{\mu \in \mathrm{BALL}_t} \mu^\top x_t$. By the choice of x_t and the assumption $\mu^\star \in \mathrm{BALL}_t$, we have **optimism**:

$$\tilde{\mu}^\top x_t = \max_{x \in D} \max_{\mu \in \mathrm{BALL}_t} \mu^\top x \geq (\mu^\star)^\top x^\star$$

Instantaneous Regret Lemma

Normalized width at time t for decision x_t: $\quad w_t := \sqrt{x_t^\top \Sigma_t^{-1} x_t}$

Lemma

Fix $t \leq T$. If $\mu^\star \in \mathrm{BALL}_t$, then

$$\mathrm{regret}_t \leq 2 \min(\sqrt{\beta_t} w_t, 1) \leq 2\sqrt{\beta_T} \min(w_t, 1)$$

Proof: Let $\tilde{\mu} \in \mathrm{argmax}_{\mu \in \mathrm{BALL}_t} \mu^\top x_t$. By the choice of x_t and the assumption $\mu^\star \in \mathrm{BALL}_t$, we have **optimism**:

$$\tilde{\mu}^\top x_t = \max_{x \in D} \max_{\mu \in \mathrm{BALL}_t} \mu^\top x \geq (\mu^\star)^\top x^\star$$

We can decompose the regret by adding and subtracting $\hat{\mu}_t$:

$$\begin{aligned} \mathrm{regret}_t &= (\mu^\star)^\top x^\star - (\mu^\star)^\top x_t \leq (\tilde{\mu} - \mu^\star)^\top x_t \\ &= (\tilde{\mu} - \hat{\mu}_t)^\top x_t + (\hat{\mu}_t - \mu^\star)^\top x_t \leq 2\sqrt{\beta_t} w_t \end{aligned}$$

Instantaneous Regret Lemma

Normalized width at time t for decision x_t: $\quad w_t := \sqrt{x_t^\top \Sigma_t^{-1} x_t}$

Lemma

Fix $t \leq T$. If $\mu^\star \in \mathrm{BALL}_t$, then

$$\mathrm{regret}_t \leq 2 \min(\sqrt{\beta_t} w_t, 1) \leq 2\sqrt{\beta_T} \min(w_t, 1)$$

Proof: Let $\tilde{\mu} \in \mathrm{argmax}_{\mu \in \mathrm{BALL}_t} \mu^\top x_t$. By the choice of x_t and the assumption $\mu^\star \in \mathrm{BALL}_t$, we have **optimism**:

$$\tilde{\mu}^\top x_t = \max_{x \in D} \max_{\mu \in \mathrm{BALL}_t} \mu^\top x \geq (\mu^\star)^\top x^\star$$

We can decompose the regret by adding and subtracting $\hat{\mu}_t$:

$$\begin{aligned} \mathrm{regret}_t &= (\mu^\star)^\top x^\star - (\mu^\star)^\top x_t \leq (\tilde{\mu} - \mu^\star)^\top x_t \\ &= (\tilde{\mu} - \hat{\mu}_t)^\top x_t + (\hat{\mu}_t - \mu^\star)^\top x_t \leq 2\sqrt{\beta_t} w_t \end{aligned}$$

Note: Since rewards $r_t \in [-1, 1]$, regret_t is trivially bounded by 2, yielding the "min".

Geometric Argument: Part 1

The next two lemmas give us a 'geometric' potential function argument, where we can bound the sum of widths independently of the choices made by the algorithm.

Geometric Argument: Part 1

The next two lemmas give us a 'geometric' potential function argument, where we can bound the sum of widths independently of the choices made by the algorithm.

Lemma

We have:

$$\det \Sigma_T = \det \Sigma_0 \prod_{t=0}^{T-1} (1 + w_t^2).$$

Geometric Argument: Part 1

The next two lemmas give us a 'geometric' potential function argument, where we can bound the sum of widths independently of the choices made by the algorithm.

Lemma

We have:

$$\det \Sigma_T = \det \Sigma_0 \prod_{t=0}^{T-1} (1 + w_t^2).$$

Proof: By the definition of Σ_{t+1}, we have

$$\begin{aligned} \det \Sigma_{t+1} &= \det(\Sigma_t + x_t x_t^\top) = \det(\Sigma_t^{1/2}(I + \Sigma_t^{-1/2} x_t x_t^\top \Sigma_t^{-1/2})\Sigma_t^{1/2}) \\ &= \det(\Sigma_t) \det(I + \Sigma_t^{-1/2} x_t (\Sigma_t^{-1/2} x_t)^\top) = \det(\Sigma_t) \det(I + v_t v_t^\top), \end{aligned}$$

where $v_t := \Sigma_t^{-1/2} x_t$. Now observe that $v_t^\top v_t = w_t^2$.

Geometric Argument: Part 2

Lemma

For any sequence where $\|x_t\|_2 \leq B$, we have:

$$\log\left(\frac{\det \Sigma_{T-1}}{\det \Sigma_0}\right) = \log\det\left(I + \frac{1}{\lambda}\sum_{t=0}^{T-1} x_t x_t^\top\right) \leq d\log\left(1 + \frac{TB^2}{d\lambda}\right)$$

Geometric Argument: Part 2

Lemma

For any sequence where $\|x_t\|_2 \leq B$, we have:

$$\log\left(\frac{\det \Sigma_{T-1}}{\det \Sigma_0}\right) = \log\det\left(I + \frac{1}{\lambda}\sum_{t=0}^{T-1} x_t x_t^\top\right) \leq d\log\left(1 + \frac{TB^2}{d\lambda}\right)$$

Proof: Let $\sigma_1, \ldots, \sigma_d$ be the eigenvalues of $\sum_{t=0}^{T-1} x_t x_t^\top$. We have:

$$\sum_{i=1}^{d} \sigma_i = \mathrm{Tr}\Big(\sum_{t=0}^{T-1} x_t x_t^\top\Big) = \sum_{t=0}^{T-1} \|x_t\|_2^2 \leq TB^2$$

Geometric Argument: Part 2

Lemma

For any sequence where $\|x_t\|_2 \leq B$, we have:

$$\log\left(\frac{\det \Sigma_{T-1}}{\det \Sigma_0}\right) = \log\det\left(I + \frac{1}{\lambda}\sum_{t=0}^{T-1} x_t x_t^\top\right) \leq d\log\left(1 + \frac{TB^2}{d\lambda}\right)$$

Proof: Let $\sigma_1, \ldots, \sigma_d$ be the eigenvalues of $\sum_{t=0}^{T-1} x_t x_t^\top$. We have:

$$\sum_{i=1}^{d} \sigma_i = \mathrm{Tr}\Big(\sum_{t=0}^{T-1} x_t x_t^\top\Big) = \sum_{t=0}^{T-1} \|x_t\|_2^2 \leq TB^2$$

Next, apply the AM-GM inequality to the log-determinant:

$$\log\det\left(I + \frac{1}{\lambda}\sum_{t=0}^{T-1} x_t x_t^\top\right) = \sum_{i=1}^{d} \log\left(1 + \frac{\sigma_i}{\lambda}\right) \leq d\log\left(1 + \frac{1}{d}\sum_{i=1}^{d}\frac{\sigma_i}{\lambda}\right) \leq d\log\left(1 + \frac{TB^2}{d\lambda}\right)$$

Proving "Sum of Squares Regret"

Proof of Proposition 3:
Assume $\mu^\star \in \textsc{Ball}_t$ for all t. We have:

$$\sum_{t=0}^{T-1} \text{regret}_t^2 \leq \sum_{t=0}^{T-1} 4\beta_t \min(w_t^2, 1) \leq 4\beta_T \sum_{t=0}^{T-1} \min(w_t^2, 1)$$

Proving "Sum of Squares Regret"

Proof of Proposition 3:
Assume $\mu^\star \in \textsc{Ball}_t$ for all t. We have:

$$\sum_{t=0}^{T-1} \text{regret}_t^2 \leq \sum_{t=0}^{T-1} 4\beta_t \min(w_t^2, 1) \leq 4\beta_T \sum_{t=0}^{T-1} \min(w_t^2, 1)$$
$$\leq 8\beta_T \sum_{t=0}^{T-1} \ln(1 + w_t^2) \leq 8\beta_T \log\left(\frac{\det \Sigma_{T-1}}{\det \Sigma_0}\right)$$

Proving "Sum of Squares Regret"

Proof of Proposition 3:
Assume $\mu^\star \in \textsc{Ball}_t$ for all t. We have:

$$\begin{aligned}\sum_{t=0}^{T-1} \text{regret}_t^2 &\leq \sum_{t=0}^{T-1} 4\beta_t \min(w_t^2, 1) \leq 4\beta_T \sum_{t=0}^{T-1} \min(w_t^2, 1) \\ &\leq 8\beta_T \sum_{t=0}^{T-1} \ln(1 + w_t^2) \leq 8\beta_T \log\left(\frac{\det \Sigma_{T-1}}{\det \Sigma_0}\right) \\ &= 8\beta_T d \log\left(1 + \frac{TB^2}{d\lambda}\right)\end{aligned}$$

Proving "Sum of Squares Regret"

Proof of Proposition 3:
Assume $\mu^\star \in \textsc{Ball}_t$ for all t. We have:

$$\begin{aligned}\sum_{t=0}^{T-1} \text{regret}_t^2 &\leq \sum_{t=0}^{T-1} 4\beta_t \min(w_t^2, 1) \leq 4\beta_T \sum_{t=0}^{T-1} \min(w_t^2, 1) \\ &\leq 8\beta_T \sum_{t=0}^{T-1} \ln(1 + w_t^2) \leq 8\beta_T \log\left(\frac{\det \Sigma_{T-1}}{\det \Sigma_0}\right) \\ &= 8\beta_T d \log\left(1 + \frac{TB^2}{d\lambda}\right)\end{aligned}$$

where the first inequality follows from Lemma 5; the second from that β_t is an increasing function of t; the third uses that for $0 \leq y \leq 1$, $\ln(1 + y) \geq y/2$; and the final two follow from Lemmas 6 and 7.

Outline

Confidence [Proof of Proposition 2]

Since $r_\tau = x_\tau \cdot \mu^\star + \eta_\tau$, we decompose the error:

$$\begin{aligned}
\widehat{\mu}_t - \mu^\star &= \Sigma_t^{-1} \sum_{\tau=0}^{t-1} x_\tau (x_\tau^\top \mu^\star + \eta_\tau) - \mu^\star \\
&= \Sigma_t^{-1} \left(\Sigma_t - \lambda I\right) \mu^\star - \mu^\star + \Sigma_t^{-1} \sum_{\tau=0}^{t-1} \eta_\tau x_\tau \\
&= -\lambda \Sigma_t^{-1} \mu^\star + \Sigma_t^{-1} \sum_{\tau=0}^{t-1} \eta_\tau x_\tau
\end{aligned}$$

Confidence [Proof of Proposition 2]

Since $r_\tau = x_\tau \cdot \mu^\star + \eta_\tau$, we decompose the error:

$$\begin{aligned}\widehat{\mu}_t - \mu^\star &= \Sigma_t^{-1} \sum_{\tau=0}^{t-1} x_\tau (x_\tau^\top \mu^\star + \eta_\tau) - \mu^\star \\ &= \Sigma_t^{-1} \left(\Sigma_t - \lambda I\right) \mu^\star - \mu^\star + \Sigma_t^{-1} \sum_{\tau=0}^{t-1} \eta_\tau x_\tau \\ &= -\lambda \Sigma_t^{-1} \mu^\star + \Sigma_t^{-1} \sum_{\tau=0}^{t-1} \eta_\tau x_\tau\end{aligned}$$

Using the norm $\|z\|_{\Sigma_t} := \sqrt{z^\top \Sigma_t z}$, the triangle inequality gives:

$$\begin{aligned}\|\widehat{\mu}_t - \mu^\star\|_{\Sigma_t} &\leq \left\|\lambda \Sigma_t^{-1} \mu^\star\right\|_{\Sigma_t} + \left\|\Sigma_t^{-1} \sum_{\tau=0}^{t-1} \eta_\tau x_\tau\right\|_{\Sigma_t} \\ &\leq \sqrt{\lambda} \|\mu^\star\|_2 + \underbrace{\left\|\sum_{\tau=0}^{t-1} \eta_\tau x_\tau\right\|_{\Sigma_t^{-1}}}_{??}\end{aligned}$$

Self-Normalizing Sum

Lemma (Self-Normalized Bound for Vector-Valued Martingales)

Suppose $\{\varepsilon_i\}_{i=1}^{\infty}$ *is a real-valued, mean-zero,* σ*-sub-Gaussian stochastic process. Let* $\{X_i\}_{i=1}^{\infty}$ *be an* $\mathbb{R}^d$*-valued stochastic process.*

Define $\Sigma_t = \Sigma_0 + \sum_{i=1}^{t} X_i X_i^\top$*. With probability at least* $1-\delta$*, we have for all* $t \geq 1$*:*

$$\left\| \sum_{i=1}^{t} X_i \varepsilon_i \right\|_{\Sigma_t^{-1}}^2 \leq \sigma^2 \log\left(\frac{\det(\Sigma_t)\det(\Sigma_0)^{-1}}{\delta^2} \right).$$

Completing the Confidence Bound

For all t, with probability at least $1-\delta$, we plug the lemma back into our triangle inequality bound:

$$\begin{aligned}\|\widehat{\mu}_t - \mu^\star\|_{\Sigma_t} &\leq \sqrt{\lambda}\|\mu^\star\|_2 + \left\|\sum_{\tau=0}^{t-1}\eta_\tau x_\tau\right\|_{\Sigma_t^{-1}} \\ &\leq \sqrt{\lambda}\|\mu^\star\|_2 + \sqrt{\sigma^2 \log\left(\frac{\det(\Sigma_t)}{\det(\Sigma_0)\delta^2}\right)}.\end{aligned}$$

Completing the Confidence Bound

For all t, with probability at least $1 - \delta$, we plug the lemma back into our triangle inequality bound:

$$\begin{aligned}\|\widehat{\mu}_t - \mu^\star\|_{\Sigma_t} &\leq \sqrt{\lambda}\|\mu^\star\|_2 + \left\|\sum_{\tau=0}^{t-1} \eta_\tau x_\tau\right\|_{\Sigma_t^{-1}} \\ &\leq \sqrt{\lambda}\|\mu^\star\|_2 + \sqrt{\sigma^2 \log\left(\frac{\det(\Sigma_t)}{\det(\Sigma_0)\delta^2}\right)}.\end{aligned}$$

Substitute our log-determinant bound (Lemma 7) into the term above:

$$\log\left(\frac{\det(\Sigma_t)}{\det(\Sigma_0)}\right) \leq d \log\left(1 + \frac{tB^2}{d\lambda}\right)$$

This matches our definition of $\sqrt{\beta_t}$, concluding that $\mu^\star \in \textsc{Ball}_t$.

Beyond Linear: Neural Network Reward Models

What if our expected reward is a non-linear neural network $f(x; \theta)$?

$$\mathbb{E}[r_t \mid x_t = x] \;=\; f(x; \theta^\star)$$

Optimizing UCB exploration directly in deep neural networks is notoriously hard. But in the **overparameterized regime**, neural networks behave like linear models!

Beyond Linear: Neural Network Reward Models

What if our expected reward is a non-linear neural network $f(x;\theta)$?

$$\mathbb{E}[r_t \mid x_t = x] \;=\; f(x;\theta^\star)$$

Optimizing UCB exploration directly in deep neural networks is notoriously hard. But in the **overparameterized regime**, neural networks behave like linear models!

The Local Linear Approximation: Consider a highly overparameterized network initialized at θ_0. We take a first-order Taylor expansion around the initialization:

$$f(x;\theta) \;\approx\; f(x;\theta_0) \;+\; \langle \nabla_\theta f(x;\theta_0),\, \theta - \theta_0 \rangle$$

Beyond Linear: Neural Network Reward Models

What if our expected reward is a non-linear neural network $f(x;\theta)$?

$$\mathbb{E}[r_t \mid x_t = x] \;=\; f(x;\theta^\star)$$

Optimizing UCB exploration directly in deep neural networks is notoriously hard. But in the **overparameterized regime**, neural networks behave like linear models!

The Local Linear Approximation: Consider a highly overparameterized network initialized at θ_0. We take a first-order Taylor expansion around the initialization:

$$f(x;\theta) \;\approx\; f(x;\theta_0) \;+\; \langle \nabla_\theta f(x;\theta_0),\, \theta - \theta_0 \rangle$$

This allows us to extract a fixed feature map driven by the network architecture:

$$\phi(x) \;:=\; \nabla_\theta f(x;\theta_0) \;\in\; \mathbb{R}^p$$

where p is the number of parameters.

Neural UCB & The Neural Tangent Kernel

Using the gradient feature map $\phi(x) = \nabla_\theta f(x; \theta_0)$, we can define a kernel:

$$k_{\text{NTK}}(x, x') \;:=\; \phi(x)^\top \phi(x') \;=\; \langle \nabla_\theta f(x; \theta_0), \nabla_\theta f(x'; \theta_0) \rangle$$

This is the **Neural Tangent Kernel (NTK)**.

Neural UCB & The Neural Tangent Kernel

Using the gradient feature map $\phi(x) = \nabla_\theta f(x; \theta_0)$, we can define a kernel:

$$k_{\text{NTK}}(x, x') := \phi(x)^\top \phi(x') = \langle \nabla_\theta f(x; \theta_0), \nabla_\theta f(x'; \theta_0) \rangle$$

This is the **Neural Tangent Kernel (NTK)**.

Neural UCB (NTK Bandits): We can simply run LinUCB (or GP-UCB) using these extracted features!

$$\Lambda_t = \lambda I + \sum_{\tau=0}^{t-1} \phi(x_\tau)\phi(x_\tau)^\top$$

Neural UCB & The Neural Tangent Kernel

Using the gradient feature map $\phi(x) = \nabla_\theta f(x; \theta_0)$, we can define a kernel:

$$k_{\text{NTK}}(x, x') := \phi(x)^\top \phi(x') = \langle \nabla_\theta f(x; \theta_0), \nabla_\theta f(x'; \theta_0) \rangle$$

This is the **Neural Tangent Kernel (NTK)**.

Neural UCB (NTK Bandits): We can simply run LinUCB (or GP-UCB) using these extracted features!

$$\Lambda_t = \lambda I + \sum_{\tau=0}^{t-1} \phi(x_\tau)\phi(x_\tau)^\top$$

$$b_t(x) = \beta \sqrt{\phi(x)^\top \Lambda_t^{-1} \phi(x)} \qquad \text{(or use the kernel trick)}$$

Neural UCB & The Neural Tangent Kernel

Using the gradient feature map $\phi(x) = \nabla_\theta f(x; \theta_0)$, we can define a kernel:

$$k_{\text{NTK}}(x, x') \;:=\; \phi(x)^\top \phi(x') \;=\; \langle \nabla_\theta f(x; \theta_0), \nabla_\theta f(x'; \theta_0) \rangle$$

This is the **Neural Tangent Kernel (NTK)**.

Neural UCB (NTK Bandits): We can simply run LinUCB (or GP-UCB) using these extracted features!

$$\Lambda_t = \lambda I + \sum_{\tau=0}^{t-1} \phi(x_\tau)\phi(x_\tau)^\top$$

$$b_t(x) = \beta \sqrt{\phi(x)^\top \Lambda_t^{-1} \phi(x)} \qquad \text{(or use the kernel trick)}$$

$$x_t = \underset{x \in D}{\text{argmax}} \Big(f(x; \theta_{t-1}) \;+\; b_t(x) \Big)$$

Neural UCB & The Neural Tangent Kernel

Using the gradient feature map $\phi(x) = \nabla_\theta f(x; \theta_0)$, we can define a kernel:

$$k_{\text{NTK}}(x, x') \;:=\; \phi(x)^\top \phi(x') \;=\; \langle \nabla_\theta f(x; \theta_0), \nabla_\theta f(x'; \theta_0) \rangle$$

This is the **Neural Tangent Kernel (NTK)**.

Neural UCB (NTK Bandits): We can simply run LinUCB (or GP-UCB) using these extracted features!

$$\Lambda_t = \lambda I + \sum_{\tau=0}^{t-1} \phi(x_\tau)\phi(x_\tau)^\top$$

$$b_t(x) = \beta \sqrt{\phi(x)^\top \Lambda_t^{-1} \phi(x)} \qquad \text{(or use the kernel trick)}$$

$$x_t = \underset{x \in D}{\text{argmax}} \Big(f(x; \theta_{t-1}) \;+\; b_t(x) \Big)$$

Takeaway: In the infinite-width limit, training a neural network with gradient descent is equivalent to kernel regression with the NTK. By using NTK features for our bonuses, we achieve theoretically