8.1 Problem Setup

In linear bandit, we choose a decision \( x_t \) on each round, where the action space is a compact set: \( x_t \in D \subseteq \mathbb{R}^d \). Then we obtain a reward \( r_t \in [-1, 1] \). The reward is linear + i.i.d. noise, where \( \mathbb{E}[r_t \mid x_t = x] = \mu^* \cdot x \in [-1, 1] \) and noise sequence \( \eta_t = r_t - \mu^* \cdot x_t \) is i.i.d. noise.

If \( x_0, \ldots, x_T \) are our decisions, then our cumulative regret is

\[
\text{Reg}_T = T \cdot \langle \mu^*, x^* \rangle - \sum_{t=0}^{T} \langle \mu^*, x_t \rangle
\]

where \( x^* \in D \) is an optimal decision for \( \mu^* \), i.e.

\[
x^* \in \text{argmax}_{x \in D} \mu^* \cdot x
\]

8.2 LinUCB Algorithm

\begin{algorithm}
\caption{Linear UCB}
\textbf{Input}: \( \lambda, \beta_t \)

\begin{algorithmic}
\State for \( t = 0, 1, 2, \ldots \) do
\State \hspace{1em} Execute \( x_t = \text{argmax}_{x \in D} \max_{\mu \in \text{BALL}_t} \langle x, \mu \rangle \)
\State \hspace{1em} and observe the reward \( r_t \)
\State \hspace{1em} Update \( \text{BALL}_{t+1} \).
\EndFor
\end{algorithmic}
\end{algorithm}

LinUCB is based on “optimism in the face of uncertainty,” which is described in Algorithm 1. At episode \( t \), we use all previous experience to define an uncertainty region (an ellipse) \( \text{BALL}_t \). The center of this region, \( \hat{\mu}_t \), is the solution of the following ridge regression problem:

\[
\hat{\mu}_t = \arg \min_{\theta} \sum_{i=0}^{t-1} (x_i^\top \theta - r_i)^2 + \lambda \| \theta \|_2^2
\]
If we consider the matrix form of $x_t$ that $X_t = [x_0, x_1, \ldots, x_{t-1}]^\top \in \mathbb{R}^{t \times d}$ and set $r_t = [r_0, r_1, \ldots, r_{t-1}]^\top \in \mathbb{R}^t$, the solution of the ridge regression is that:

$$
\hat{\mu}_t = \arg \min_{\theta} \|X_t^\top \theta - r_t\|^2_2 + \lambda \|\theta\|^2_2
= (X_t^\top X_t + \lambda I)^{-1} X_t^\top r_t
= \Sigma_t^{-1} \sum_{i=0}^{t-1} r_i x_i
$$

where $\lambda$ is a parameter and where

$$
\Sigma_t = \lambda I + \sum_{i=0}^{t-1} x_i x_i^\top, \text{ with } \Sigma_0 = \lambda I
$$

The shape of the region $\text{BALL}_t$ is defined through the feature covariance $\Sigma_t$. Precisely, the uncertainty region, or confidence ball, is defined as:

$$
\text{BALL}_t = \{ \mu | (\mu - \hat{\mu}_t)^\top \Sigma_t (\mu - \hat{\mu}_t) \leq \beta_t \}
$$

where $\beta_t$ is a parameter of the algorithm.

### 8.3 Regret bound of LinUCB

Our main result here is that we have **sublinear regret**: $R_T \leq O^*(d\sqrt{T})$, **poly dependence** on $d$ and **no dependence** on the cardinality $|D|$.

**Theorem 8.1.** Suppose: bounded noise $|\eta_t| \leq \sigma$, that $\|\mu^*\| \leq W$, and that $\|x\| \leq B$ for all $x \in D$. Set $\lambda = \sigma^2/W^2$ and

$$
\beta_t := \sigma^2 \left( 2 + 4d \log \left( 1 + \frac{TB^2W^2}{d} \right) + 8 \log(4/\delta) \right)
$$

With probability greater than $1 - \delta$, that 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 $c$ is an absolute constant.

To proof the Theorem 8.1, we need two key components. The first is in showing that the confidence region is appropriate.

**Proposition 8.2.** *(Uniform confidence bound)*

Let $\delta > 0$. We have that

$$
\Pr (\forall t, \mu^* \in \text{BALL}_t) \geq 1 - \delta.
$$

The second main step in analyzing LinUCB is to show that as long as the aforementioned high-probability event holds, we have some control on the growth of the regret. Let us define the instantaneous regret as $\text{regret}_t = \mu^* \cdot x^* - \mu^* \cdot x_t$, the following bounds the sum of the squares of instantaneous regret.
Proposition 8.3. (Sum of Squares Regret Bound)
Define:
\[
\text{regret}_t = \mu^* \cdot x^* - \mu^* \cdot x_t
\]
Suppose \(\|x\| \leq B\) for \(x \in D\). Suppose \(\beta_t\) is increasing and larger than 1. Suppose \(\mu^* \in \text{BALL}_t\) for all \(t\), then
\[
\sum_{t=0}^{T-1} \text{regret}_t^2 \leq 8\beta_T d \log \left(1 + \frac{TB^2}{d\lambda}\right)
\]

Using these two results we are able to prove our upper bound as follows:

Proof of Theorem 8.1. By Propositions 8.2 and 8.3 along with the Cauchy-Schwarz inequality, we have, with probability at least \(1 - \delta\),
\[
R_T = \sum_{t=0}^{T-1} \text{regret}_t \leq \sqrt{T \sum_{t=0}^{T-1} \text{regret}_t^2} \leq \sqrt{8T\beta_T d \log \left(1 + \frac{TB^2}{d\lambda}\right)}.
\]

The remainder of the proof follows from using our chosen value of \(\beta_T = \sigma^2 \left(2 + 4d \log \left(1 + \frac{TB^2W^2}{d}\right) + 8 \log(4/\delta)\right)\)
and algebraic manipulations (that \(2ab \leq a^2 + b^2\)).

8.3.1 Plan of the proof

1. First prove the Proposition that bounds the sum of square regret
   - By bounding instantaneous regret
   - And then bounding the sum of squares with “Information Gain”

2. Prove the uniform confidence bound
   - Basically show that the choice of \(\beta_t\) "works".

Lemma 8.4. ("Width" of Confidence Ball)
Let \(x \in D\). If \(\mu \in \text{BALL}_t\) and \(x \in D\). Then
\[
\left| (\mu - \hat{\mu}_t)^\top x \right| \leq \sqrt{\beta_t x^\top \Sigma_t^{-1} x}
\]

Proof. By Cauchy-Schwarz, we have:
\[
\left| (\mu - \hat{\mu}_t)^\top x \right| = \left| (\mu - \hat{\mu}_t)^\top \Sigma_t^{1/2} \Sigma_t^{-1/2} x \right| = \left| \Sigma_t^{1/2} (\mu - \hat{\mu}_t)^\top \Sigma_t^{-1/2} x \right|
\leq \left\| \Sigma_t^{1/2} (\mu - \hat{\mu}_t) \right\| \left\| \Sigma_t^{-1/2} x \right\| = \left\| \Sigma_t^{1/2} (\mu - \hat{\mu}_t) \right\| \sqrt{x^\top \Sigma_t^{-1} x}
\leq \sqrt{\beta_t x^\top \Sigma_t^{-1} x}
\]
where the last inequality holds since \(\mu \in \text{BALL}_t\).
Define
\[ w_t := \sqrt{x_t^\top \Sigma_t^{-1} x_t} \]
which is the "normalized width" at time \( t \) in the direction of the chosen decision. We now see that the width, \( 2\sqrt{\beta_t} w_t \), is an upper bound for the instantaneous regret.

**Lemma 8.5.** *(Instantaneous Regret is bounded by the width of the ellipsoid)*

Fix \( t \leq T \). If \( \mu^* \in \text{BALL}_t \), then
\[
\text{regret}_t \leq 2 \min \left( \sqrt{\beta_t} w_t, 1 \right) \leq 2 \sqrt{\beta_T} \min (w_t, 1)
\]

**Proof.** Let \( \bar{\mu} \in \text{BALL}_t \) denote the vector which minimizes the dot product \( \bar{\mu}^\top x_t \). By choice of \( x_t \), we have
\[
\bar{\mu}^\top x_t = \max_{\mu \in \text{BALL}_t} \max_{x \in D} \mu^\top x \geq (\mu^*)^\top x^*
\]
where the inequality used the hypothesis \( \mu^* \in \text{BALL}_t \). Hence,
\[
\text{regret}_t = (\mu^*)^\top x^* - (\mu^*)^\top x_t \leq (\bar{\mu} - \mu^*)^\top x_t
\]
\[
= (\bar{\mu} - \bar{\mu}_t)^\top x_t + (\bar{\mu}_t - \mu^*)^\top x_t \leq 2\sqrt{\beta_t} w_t
\]
where the last step follows from Lemma 8.4 since \( \bar{\mu} \) and \( \mu^* \) are in \( \text{BALL}_t \). Since \( r_t \in [-1, 1] \), \( \text{regret}_t \) is always at most 2 and the first inequality follows. The final inequality is due to that \( \beta_t \) is increasing and larger than 1. \( \square \)

The following two lemmas prove useful in showing that we can treat the log determinant as a potential function, where we can bound the sum of widths independently of the choices made by the algorithm.

**Lemma 8.6.** We have:
\[
\det \Sigma_T = \det \Sigma_0 \prod_{t=0}^{T-1} (1 + w_t^2)
\]

**Proof.** By the definition of \( \Sigma_{t+1} \), we have
\[
\det \Sigma_{t+1} = \det \left( \Sigma_t + x_t x_t^\top \right) = \det \left( \Sigma_t^{1/2} \left( I + \Sigma_t^{-1/2} x_t x_t^\top \Sigma_t^{-1/2} \right) \Sigma_t^{1/2} \right)
\]
\[
= \det (\Sigma_t) \det \left( I + \Sigma_t^{-1/2} x_t \left( \Sigma_t^{-1/2} x_t \right)^\top \right) = \det (\Sigma_t) \det (I + v_t v_t^\top)
\]
where \( v_t := \Sigma_t^{-1/2} x_t \). Now observe that \( v_t^\top v_t = w_t^2 \) and
\[
(I + v_t v_t^\top) v_t = v_t + v_t (v_t^\top v_t) = (1 + w_t^2) v_t
\]
Hence \( 1 + w_t^2 \) is an eigenvalue of \( I + v_t v_t^\top \). Since \( v_t v_t^\top \) is a rank one matrix, all other eigenvalues of \( I + v_t v_t^\top \) equal 1. Hence, \( \det (I + v_t v_t^\top) \) is \( (1 + w_t^2) \), implying \( \det \Sigma_{t+1} = (1 + w_t^2) \det \Sigma_t \). The result follows by induction. \( \square \)
Lemma 8.7. ("Potential Function" Bound)  
For any sequence $x_0, \ldots, x_{T-1}$ such that, for $t < T$, $\|x_t\|_2 \leq B$, we have.

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

Proof. Denote the eigenvalues of $\sum_{t=0}^{T-1} x_t x_t^\top$ as $\sigma_1, \ldots, \sigma_d$, and note:

$$
\sum_{i=1}^d \sigma_i = \text{Trace} \left( \sum_{t=0}^{T-1} x_t x_t^\top \right) = \sum_{t=0}^{T-1} \|x_t\|^2 \leq TB^2.
$$

Using the AM-GM inequality,

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

$$
\leq d \log \left( \frac{1}{d} \sum_{i=1}^d (1 + \sigma_i/\lambda) \right) \leq d \log \left( 1 + \frac{TB^2}{d\lambda} \right)
$$

which concludes the proof. \qed

Finally, we are ready to prove that if $\mu^*$ always stays within the evolving confidence region, then our regret is under control.

Proof of Proposition 8.3. Assume that $\mu^* \in \text{BALL}_t$ for all $t$. We have that:

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

$$
\leq \max \{8, \frac{4}{\log 2}\} \beta_T \sum_{t=0}^{T-1} \log (1 + w_t^2) \leq 8\beta_T \log (\det \Sigma_{T-1} / \det \Sigma_0)
$$

$$
= 8\beta_T d \log \left( 1 + \frac{TB^2}{d\lambda} \right)
$$

where the first inequality follow from Lemma 8.5, the second from that $\beta_t$ is an increasing function of $t$; the third uses that for $0 \leq y \leq 1, y \geq \log(1 + y) \geq \frac{y}{1+y} \geq \frac{y}{2}$, when $w_t^2 \leq 1$, $w_t^2 \leq 2 \log(1 + w_t^2)$, and when $w_t^2 > 1$,

$$
4\beta_t = \frac{1}{\log 2} \beta_t \log 2 \leq \frac{4}{\log 2} \beta_t \log (1 + w_t^2);
$$

the final two inequalities follow by Lemmas 8.6 and 8.7. \qed

Then we can do confidence analysis to prove the uniform confidence bound:

Lemma 8.8. (Self-Normalized Bound for Vector-Valued Martingales; [Abbasi-Yadkori et al., 2011]). Let $\{\varepsilon_i\}_{i=1}^\infty$ be a real-valued stochastic process with corresponding filtration $\{\mathcal{F}_t\}_{t=1}^\infty$ such that $\varepsilon_i$ is $\mathcal{F}_i$ measurable, $\mathbb{E}[\varepsilon_i | \mathcal{F}_{t-1}] = 0$, and $\varepsilon_i$ is conditionally $\sigma$-sub-Gaussian with $\sigma \in \mathbb{R}^+$. Let $\{X_i\}_{i=1}^\infty$ be a stochastic process with $X_i \in \mathcal{H}$ (some Hilbert space) and $X_i$ being $\mathcal{F}_i$ measurable. Assume that a linear operator $\Sigma : \mathcal{H} \to \mathcal{H}$ is positive definite, i.e., $x^\top \Sigma x > 0$ for any $x \in \mathcal{H}$. For any $t$, define the linear operator $\Sigma_t = \Sigma_0 + \sum_{i=1}^t X_i X_i^\top$ (here $xx^\top$ denotes outer-product in $\mathcal{H}$). With probability at least $1 - \delta$, we have for all $t \geq 1$:

$$
\left\| \sum_{i=1}^t X_i \varepsilon_i \right\|_{\Sigma_t^{-1}}^2 \leq \sigma^2 \log \left( \frac{\det(\Sigma_t) \det(\Sigma)^{-1}}{\delta^2} \right).
$$
Proof of Proposition 8.2. Since \( r_\tau = x_\tau \cdot \mu^* + \eta_\tau \), we have:

\[
\hat{\mu}_t - \mu^* = \Sigma_t^{-1} \sum_{\tau=0}^{t-1} r_\tau x_\tau - \mu^* = \Sigma_t^{-1} \sum_{\tau=0}^{t-1} x_\tau (x_\tau \cdot \mu^* + \eta_\tau) - \mu^* \\
= \Sigma_t^{-1} \left( \sum_{\tau=0}^{t-1} x_\tau(x_\tau)^\top \right) \mu^* - \mu^* + \Sigma_t^{-1} \sum_{\tau=0}^{t-1} \eta_\tau x_\tau \\
= \lambda \Sigma_t^{-1} \mu^* + \Sigma_t^{-1} \sum_{\tau=0}^{t-1} \eta_\tau x_\tau
\]

For any \( 0 < \delta_t < 1 \), using triangle inequality and Lemma 8.8, it holds with probability at least \( 1 - \delta_t \),

\[
\sqrt{(\hat{\mu}_t - \mu^*)^\top \Sigma_t (\hat{\mu}_t - \mu^*)} = \left\| (\Sigma_t)^{1/2} (\hat{\mu}_t - \mu^*) \right\| \\
\leq \left\| \lambda \Sigma_t^{-1/2} \mu^* \right\| + \left\| \Sigma_t^{-1/2} \sum_{\tau=0}^{t-1} \eta_\tau x_\tau \right\| \\
\leq \sqrt{\lambda} \| \mu^* \| + \sqrt{2\sigma^2 \log \left( \det (\Sigma_t) \det (\Sigma_0)^{-1} / \delta_t \right)}
\]

where we have also used the triangle inequality and that \( \| \Sigma_t^{-1} \| \leq 1/\lambda \). We seek to lower bound \( \Pr (\forall t, \mu^* \in \text{BALL}_t) \). Note that at \( t = 0 \), by our choice of \( \lambda \), we have that \( \text{BALL}_0 \) contains \( W^* \), so \( \Pr (\mu^* \notin \text{BALL}_0) = 0 \). For \( t \geq 1 \), let us assign failure probability \( \delta_t = (3/\pi^2) / t^2 \cdot 2\delta \) for the \( t \)-th event, which, using the above and union bound, gives us an upper bound on the sum failure probability as

\[
1 - \Pr (\forall t, \mu^* \in \text{BALL}_t) = \Pr (\exists t, \mu^* \notin \text{BALL}_t) \leq \sum_{t=1}^{\infty} \Pr (\mu^* \notin \text{BALL}_t) < \sum_{t=1}^{\infty} \left( 1/t^2 \right) (3/\pi^2) \cdot 2\delta = 1/2 \cdot 2\delta = \delta
\]

This along with Lemma 8.7 completes the proof.

### 8.4 Remarks

- The regret of LinUCB is optimal up to \( \tilde{O}(d\sqrt{T}) \)
- The analysis of LinUCB is based on strong assumption on realizability.
- For agnostic linear bandits, EXP4 [Auer et al., 2002] can achieve the regret of \( O(d\sqrt{T}) \), and works in the adversarial settings, but is computationally inefficient.
- In contextual version with a finite list of available actions are given at each \( t \), assuming i.i.d. setting, the “Taming the Monster” algorithm [Agarwal et al., 2014] achieves a regret bound of \( O(\sqrt{dkT}) \) where \( k \) is the number of actions with an oracle-efficient algorithm.

### References

