CSE 541: Interactive Learning

Lecture 13: The confidence ellipsoid, pure exploration, adversarial linear bandits

Kevin Jamieson

Paul G. Allen School of
Computer Science & Engineering
University of Washington

Monday, November 16, 2026

Recap, 1: supermartingales and Ville's inequality

$\mathcal{F}_t$ = what is known after round t. Z_t conditionally 1-sub-Gaussian given $\mathcal{F}_{t-1}$ (no independence needed), $S_t = \sum_{s \le t} Z_s$.

- **The supermartingale behind Chernoff:** $M_t(\lambda) = \exp(\lambda S_t - t\lambda^2/2)$ is nonnegative, starts at 1, and $\mathbb{E}[M_t(\lambda) \mid \mathcal{F}_{t-1}] \le M_{t-1}(\lambda)$.
- **Ville:** a nonnegative supermartingale with $M_0 \le 1$ has $\mathbb{P}(\exists t : M_t \ge 1/\delta) \le \delta$. Markov for the whole path, no union bound over t.
- **Linear boundary** (one λ): $\mathbb{P}\big(\exists t : S_t \ge \frac{t\lambda}{2} + \frac{\log(1/\delta)}{\lambda}\big) \le \delta$. Tight at one time only.
- **Curved boundary** (mix over $\lambda \sim \mathcal{N}(0, \nu^2)$; an average of supermartingales is a supermartingale, and the Gaussian integral is explicit):

$$\mathbb{P}\left(\exists t : \ |S_t| \ge \sqrt{2\big(t+\nu^{-2}\big)\left(\log\tfrac{1}{\delta} + \tfrac{1}{2}\log\big((t+\nu^{-2})\nu^2\big)\right)}\right) \le \delta.$$

Recap, 1: supermartingales and Ville's inequality

$\mathcal{F}_t$ = what is known after round t. Z_t conditionally 1-sub-Gaussian given $\mathcal{F}_{t-1}$ (no independence needed), $S_t = \sum_{s \le t} Z_s$.

- **The supermartingale behind Chernoff:** $M_t(\lambda) = \exp(\lambda S_t - t\lambda^2/2)$ is nonnegative, starts at 1, and $\mathbb{E}[M_t(\lambda) \mid \mathcal{F}_{t-1}] \le M_{t-1}(\lambda)$.
- **Ville:** a nonnegative supermartingale with $M_0 \le 1$ has $\mathbb{P}(\exists t : M_t \ge 1/\delta) \le \delta$. Markov for the whole path, no union bound over t.
- **Linear boundary** (one λ): $\mathbb{P}\big(\exists t : S_t \ge \frac{t\lambda}{2} + \frac{\log(1/\delta)}{\lambda}\big) \le \delta$. Tight at one time only.
- **Curved boundary** (mix over $\lambda \sim \mathcal{N}(0, \nu^2)$; an average of supermartingales is a supermartingale, and the Gaussian integral is explicit):

$$\mathbb{P}\left(\exists t : \ |S_t| \ge \sqrt{2\big(t+\nu^{-2}\big)\left(\log\tfrac{1}{\delta} + \tfrac{1}{2}\log\big((t+\nu^{-2})\nu^2\big)\right)}\right) \le \delta.$$

The recipe. Build a nonnegative supermartingale out of the quantity you care about; if it depends on a free parameter you would like to choose after seeing the data, *mix* over that parameter instead; apply Ville; solve for the quantity.

Recap, 2: these supermartingales are likelihood ratios

- For Gaussian data, $M_t(\lambda) = \prod_{s \le t} p_\lambda(Z_s)/p_0(Z_s)$: the likelihood ratio of "mean λ" against "mean 0", a martingale under the null. Ville: the evidence against the truth exceeds $1/\delta$ at *some* time with probability $\le \delta$.
- **SPRT and Wald's identity:** stop when the likelihood ratio hits $1/\delta$; then $\mathbb{E}_1[\tau] \approx \log(1/\delta)/\mathrm{KL}(p_1 \| p_0)$, and by the Lecture 8 template no test does better. Gaussians with gap Δ: $2\log(1/\delta)/\Delta^2$.
- **The mixture is a Bayes factor:** $\bar{M}_t = \int M_t(\lambda) h(\lambda) d\lambda$ is the likelihood ratio of "the gap is drawn from the prior h" against the null; the boundary is tightest where the prior puts its mass.

Recap, 2: these supermartingales are likelihood ratios

- For Gaussian data, $M_t(\lambda) = \prod_{s \le t} p_\lambda(Z_s)/p_0(Z_s)$: the likelihood ratio of "mean λ" against "mean 0", a martingale under the null. Ville: the evidence against the truth exceeds $1/\delta$ at *some* time with probability $\le \delta$.
- **SPRT and Wald's identity:** stop when the likelihood ratio hits $1/\delta$; then $\mathbb{E}_1[\tau] \approx \log(1/\delta)/\mathrm{KL}(p_1 \| p_0)$, and by the Lecture 8 template no test does better. Gaussians with gap Δ: $2\log(1/\delta)/\Delta^2$.
- **The mixture is a Bayes factor:** $\bar{M}_t = \int M_t(\lambda) h(\lambda) d\lambda$ is the likelihood ratio of "the gap is drawn from the prior h" against the null; the boundary is tightest where the prior puts its mass.

Why we are doing this

LinUCB needs $\theta^* \in \{\theta : \left\|\theta - \hat{\theta}_t\right\|_{V_t + \gamma I} \le \beta_t\}$ for all t at once, with adaptively chosen arms. The noise process is the *vector* $S_t = \sum_{s \le t} x_s \eta_s$. Today: the same recipe in $\mathbb{R}^d$, mixing with a Gaussian prior on θ.

Vector-valued martingales

The confidence ellipsoid for adaptive designs

Vector-valued martingales: the bandit's noise process

Now $x_t \in \mathbb{R}^d$ predictable (the arm chosen from the past), η_t conditionally 1-sub-Gaussian, and $Z_t = x_t \eta_t$, $S_t = \sum_{s \le t} x_s \eta_s$, $V_t = \sum_{s \le t} x_s x_s^\top$. For a fixed $\lambda \in \mathbb{R}^d$, $\mathbb{E}\left[e^{\langle \lambda, Z_t \rangle} \mid \mathcal{F}_{t-1}\right] = \mathbb{E}\left[e^{\langle \lambda, x_t \rangle \eta_t} \mid \mathcal{F}_{t-1}\right] \le$

Vector-valued martingales: the bandit's noise process

Now $x_t \in \mathbb{R}^d$ predictable (the arm chosen from the past), η_t conditionally 1-sub-Gaussian, and $Z_t = x_t \eta_t$, $S_t = \sum_{s \le t} x_s \eta_s$, $V_t = \sum_{s \le t} x_s x_s^\top$. For a fixed $\lambda \in \mathbb{R}^d$, $\mathbb{E}\big[e^{\langle \lambda, Z_t \rangle} \mid \mathcal{F}_{t-1}\big] = \mathbb{E}\big[e^{\langle \lambda, x_t \rangle \eta_t} \mid \mathcal{F}_{t-1}\big] \le e^{\langle \lambda, x_t \rangle^2 / 2}$, because $\langle \lambda, x_t \rangle$ is a constant given $\mathcal{F}_{t-1}$, and $\langle \lambda, x_t \rangle^2 = \|\lambda\|^2_{x_t x_t^\top}$ adds up to $\|\lambda\|^2_{V_t}$. Exactly as before,

$$M_t(\lambda) = \exp\left(\langle \lambda, S_t \rangle - \tfrac{1}{2} \|\lambda\|^2_{V_t} \right)$$

is a nonnegative supermartingale with $M_0(\lambda) = 1$, and Ville gives a *linear* boundary for each fixed λ: $\mathbb{P}\big(\exists t : \langle \lambda, S_t \rangle - \tfrac{1}{2} \|\lambda\|^2_{V_t} \ge \log \tfrac{1}{\delta}\big) \le \delta$.

Vector-valued martingales: the bandit's noise process

Now $x_t \in \mathbb{R}^d$ predictable (the arm chosen from the past), η_t conditionally 1-sub-Gaussian, and $Z_t = x_t \eta_t$, $S_t = \sum_{s \le t} x_s \eta_s$, $V_t = \sum_{s \le t} x_s x_s^\top$. For a fixed $\lambda \in \mathbb{R}^d$, $\mathbb{E}\left[e^{\langle \lambda, Z_t \rangle} \mid \mathcal{F}_{t-1}\right] = \mathbb{E}\left[e^{\langle \lambda, x_t \rangle \eta_t} \mid \mathcal{F}_{t-1}\right] \le e^{\langle \lambda, x_t \rangle^2 / 2}$, because $\langle \lambda, x_t \rangle$ is a constant given $\mathcal{F}_{t-1}$, and $\langle \lambda, x_t \rangle^2 = \|\lambda\|^2_{x_t x_t^\top}$ adds up to $\|\lambda\|^2_{V_t}$. Exactly as before,

$$M_t(\lambda) = \exp\left(\langle \lambda, S_t \rangle - \tfrac{1}{2} \|\lambda\|^2_{V_t}\right)$$

is a nonnegative supermartingale with $M_0(\lambda) = 1$, and Ville gives a *linear* boundary for each fixed λ: $\mathbb{P}\left(\exists t : \langle \lambda, S_t \rangle - \tfrac{1}{2} \|\lambda\|^2_{V_t} \ge \log \tfrac{1}{\delta}\right) \le \delta$.

The same obstacle as in the fixed design

We want $\|S_t\|_{V_t^{-1}}$, and $\max_\lambda \left(\langle \lambda, S_t \rangle - \tfrac{1}{2} \|\lambda\|^2_{V_t}\right) = \tfrac{1}{2} \|S_t\|^2_{V_t^{-1}}$ at $\lambda = V_t^{-1} S_t$. But that λ is random, and the boundary holds for fixed λ only. Fixed design: a net. Here: a better move.

Vector-valued martingales: the bandit's noise process

Now $x_t \in \mathbb{R}^d$ predictable (the arm chosen from the past), η_t conditionally 1-sub-Gaussian, and $Z_t = x_t\eta_t$, $S_t = \sum_{s\le t} x_s\eta_s$, $V_t = \sum_{s\le t} x_s x_s^\top$. For a fixed $\lambda \in \mathbb{R}^d$, $\mathbb{E}\big[e^{\langle\lambda, Z_t\rangle} \mid \mathcal{F}_{t-1}\big] = \mathbb{E}\big[e^{\langle\lambda, x_t\rangle\eta_t} \mid \mathcal{F}_{t-1}\big] \le e^{\langle\lambda, x_t\rangle^2/2}$, because $\langle\lambda, x_t\rangle$ is a constant given $\mathcal{F}_{t-1}$, and $\langle\lambda, x_t\rangle^2 = \|\lambda\|^2_{x_t x_t^\top}$ adds up to $\|\lambda\|^2_{V_t}$. Exactly as before,

$$M_t(\lambda) = \exp\Big(\langle\lambda, S_t\rangle - \tfrac{1}{2}\|\lambda\|^2_{V_t}\Big)$$

is a nonnegative supermartingale with $M_0(\lambda) = 1$, and Ville gives a *linear* boundary for each fixed λ: $\mathbb{P}\big(\exists t : \langle\lambda, S_t\rangle - \tfrac{1}{2}\|\lambda\|^2_{V_t} \ge \log\tfrac{1}{\delta}\big) \le \delta$.

The same obstacle as in the fixed design

We want $\|S_t\|_{V_t^{-1}}$, and $\max_\lambda\big(\langle\lambda, S_t\rangle - \tfrac{1}{2}\|\lambda\|^2_{V_t}\big) = \tfrac{1}{2}\|S_t\|^2_{V_t^{-1}}$ at $\lambda = V_t^{-1}S_t$. But that λ is random, and the boundary holds for fixed λ only. Fixed design: a net. Here: a better move.

Mix over λ: $h = \mathcal{N}(0, \gamma^{-1}I_d)$. Then $\bar{M}_t = \int M_t(\lambda)h(\lambda)d\lambda$ is a supermartingale with $\bar{M}_0 = 1$, and the integral is Gaussian.

The mixture, computed

With $h(\lambda) = (\gamma/2\pi)^{d/2} e^{-\gamma \|\lambda\|_2^2/2}$, write $\bar{V}_t = V_t + \gamma I$ and complete the square:

$$\begin{aligned} \bar{M}_t &= \left(\frac{\gamma}{2\pi}\right)^{d/2} \int \exp\left(\langle \lambda, S_t \rangle - \tfrac{1}{2} \|\lambda\|_{\bar{V}_t}^2\right) d\lambda \\ &= \\ &= \end{aligned}$$

The mixture, computed

With $h(\lambda) = (\gamma/2\pi)^{d/2} e^{-\gamma \|\lambda\|_2^2/2}$, write $\bar{V}_t = V_t + \gamma I$ and complete the square:

$$\begin{aligned}
\bar{M}_t &= \left(\frac{\gamma}{2\pi}\right)^{d/2} \int \exp\left(\langle \lambda, S_t \rangle - \tfrac{1}{2} \|\lambda\|_{\bar{V}_t}^2\right) d\lambda \\
&= \left(\frac{\gamma}{2\pi}\right)^{d/2} \int \exp\left(\tfrac{1}{2} \|S_t\|_{\bar{V}_t^{-1}}^2 - \tfrac{1}{2} \left\|\lambda - \bar{V}_t^{-1} S_t\right\|_{\bar{V}_t}^2\right) d\lambda \\
&= \left(\frac{\gamma}{2\pi}\right)^{d/2} (2\pi)^{d/2} \det(\bar{V}_t)^{-1/2} \exp\left(\tfrac{1}{2} \|S_t\|_{\bar{V}_t^{-1}}^2\right) = \left(\frac{\gamma^d}{\det \bar{V}_t}\right)^{1/2} e^{\frac{1}{2}\|S_t\|_{\bar{V}_t^{-1}}^2}.
\end{aligned}$$

(The λ-integral is Gaussian with covariance $\bar{V}_t^{-1}$.) Now Ville: $\bar{M}_t \geq 1/\delta$ iff

$$\|S_t\|_{\bar{V}_t^{-1}}^2 \;\geq$$

The mixture, computed

With $h(\lambda) = (\gamma/2\pi)^{d/2} e^{-\gamma\|\lambda\|_2^2/2}$, write $\bar{V}_t = V_t + \gamma I$ and complete the square:

$$\begin{aligned}
\bar{M}_t &= \left(\frac{\gamma}{2\pi}\right)^{d/2} \int \exp\left(\langle \lambda, S_t \rangle - \tfrac{1}{2}\|\lambda\|_{\bar{V}_t}^2\right) d\lambda \\
&= \left(\frac{\gamma}{2\pi}\right)^{d/2} \int \exp\left(\tfrac{1}{2}\|S_t\|_{\bar{V}_t^{-1}}^2 - \tfrac{1}{2}\left\|\lambda - \bar{V}_t^{-1} S_t\right\|_{\bar{V}_t}^2\right) d\lambda \\
&= \left(\frac{\gamma}{2\pi}\right)^{d/2} (2\pi)^{d/2} \det(\bar{V}_t)^{-1/2} \exp\left(\tfrac{1}{2}\|S_t\|_{\bar{V}_t^{-1}}^2\right) = \left(\frac{\gamma^d}{\det \bar{V}_t}\right)^{1/2} e^{\frac{1}{2}\|S_t\|_{\bar{V}_t^{-1}}^2}.
\end{aligned}$$

(The λ-integral is Gaussian with covariance $\bar{V}_t^{-1}$.) Now Ville: $\bar{M}_t \geq 1/\delta$ iff

$$\|S_t\|_{\bar{V}_t^{-1}}^2 \;\geq\; 2\log\frac{1}{\delta} + \log\frac{\det \bar{V}_t}{\gamma^d}.$$

The mixture, computed

With $h(\lambda) = (\gamma/2\pi)^{d/2} e^{-\gamma\|\lambda\|_2^2/2}$, write $\bar{V}_t = V_t + \gamma I$ and complete the square:

$$\begin{aligned}
\bar{M}_t &= \left(\frac{\gamma}{2\pi}\right)^{d/2} \int \exp\left(\langle \lambda, S_t\rangle - \tfrac{1}{2}\|\lambda\|_{\bar{V}_t}^2\right) d\lambda \\
&= \left(\frac{\gamma}{2\pi}\right)^{d/2} \int \exp\left(\tfrac{1}{2}\|S_t\|_{\bar{V}_t^{-1}}^2 - \tfrac{1}{2}\left\|\lambda - \bar{V}_t^{-1} S_t\right\|_{\bar{V}_t}^2\right) d\lambda \\
&= \left(\frac{\gamma}{2\pi}\right)^{d/2} (2\pi)^{d/2} \det(\bar{V}_t)^{-1/2} \exp\left(\tfrac{1}{2}\|S_t\|_{\bar{V}_t^{-1}}^2\right) = \left(\frac{\gamma^d}{\det \bar{V}_t}\right)^{1/2} e^{\frac{1}{2}\|S_t\|_{\bar{V}_t^{-1}}^2}.
\end{aligned}$$

(The λ-integral is Gaussian with covariance $\bar{V}_t^{-1}$.) Now Ville: $\bar{M}_t \geq 1/\delta$ iff

$$\|S_t\|_{\bar{V}_t^{-1}}^2 \;\geq\; 2\log\frac{1}{\delta} + \log\frac{\det \bar{V}_t}{\gamma^d}.$$

Self-normalized bound (Abbasi-Yadkori et al. 2011; notes eq. 1.1)

$$\mathbb{P}\left(\exists t :\; \|S_t\|_{(V_t+\gamma I)^{-1}} \geq \sqrt{2\log\tfrac{1}{\delta} + \log\tfrac{\det(V_t+\gamma I)}{\gamma^d}}\right) \leq \delta.$$

Sanity check: $d = 1$, $x_t \equiv 1$, $\gamma = \nu^{-2}$ gives $V_t = t$ and exactly the curved boundary from the recap.

Finale: the confidence ellipsoid, proved

Ridge with parameter γ (called λ last lecture): $\hat{\theta}_t = \bar{V}_t^{-1} \sum_{s \le t} x_s y_s$, $\bar{V}_t = V_t + \gamma I$. Substitute $y_s = \langle x_s, \theta^* \rangle + \eta_s$:

$$\hat{\theta}_t = \bar{V}_t^{-1} \big(V_t \theta^* + S_t \big) =$$

so $$\left\| \hat{\theta}_t - \theta^* \right\|_{\bar{V}_t} =$$

Finale: the confidence ellipsoid, proved

Ridge with parameter γ (called λ last lecture): $\hat{\theta}_t = \bar{V}_t^{-1} \sum_{s \le t} x_s y_s$, $\bar{V}_t = V_t + \gamma I$. Substitute $y_s = \langle x_s, \theta^* \rangle + \eta_s$:

$$\hat{\theta}_t = \bar{V}_t^{-1}\big(V_t\theta^* + S_t\big) = \theta^* + \bar{V}_t^{-1}\big(S_t - \gamma\theta^*\big),$$

so $$\left\|\hat{\theta}_t - \theta^*\right\|_{\bar{V}_t} = \|S_t - \gamma\theta^*\|_{\bar{V}_t^{-1}}\,.$$

Finale: the confidence ellipsoid, proved

Ridge with parameter γ (called λ last lecture): $\hat{\theta}_t = \bar{V}_t^{-1} \sum_{s \le t} x_s y_s$, $\bar{V}_t = V_t + \gamma I$. Substitute $y_s = \langle x_s, \theta^* \rangle + \eta_s$:

$$\hat{\theta}_t = \bar{V}_t^{-1}\big(V_t \theta^* + S_t\big) = \theta^* + \bar{V}_t^{-1}\big(S_t - \gamma\theta^*\big),$$

$$\text{so} \qquad \left\|\hat{\theta}_t - \theta^*\right\|_{\bar{V}_t} = \|S_t - \gamma\theta^*\|_{\bar{V}_t^{-1}}\,.$$

Triangle inequality, then $\bar{V}_t \succeq \gamma I$:
$\|S_t - \gamma\theta^*\|_{\bar{V}_t^{-1}} \le$

Finale: the confidence ellipsoid, proved

Ridge with parameter γ (called λ last lecture): $\hat{\theta}_t = \bar{V}_t^{-1} \sum_{s \le t} x_s y_s$, $\bar{V}_t = V_t + \gamma I$. Substitute $y_s = \langle x_s, \theta^* \rangle + \eta_s$:

$$\hat{\theta}_t = \bar{V}_t^{-1}\big(V_t \theta^* + S_t\big) = \theta^* + \bar{V}_t^{-1}\big(S_t - \gamma\theta^*\big),$$

$$\text{so} \qquad \left\|\hat{\theta}_t - \theta^*\right\|_{\bar{V}_t} = \|S_t - \gamma\theta^*\|_{\bar{V}_t^{-1}}\,.$$

Triangle inequality, then $\bar{V}_t \succeq \gamma I$:

$$\|S_t - \gamma\theta^*\|_{\bar{V}_t^{-1}} \le \|S_t\|_{\bar{V}_t^{-1}} + \gamma\,\|\theta^*\|_{\bar{V}_t^{-1}} \le \|S_t\|_{\bar{V}_t^{-1}} + \sqrt{\gamma}\,\|\theta^*\|_2.$$

Finale: the confidence ellipsoid, proved

Ridge with parameter γ (called λ last lecture): $\hat{\theta}_t = \bar{V}_t^{-1} \sum_{s \le t} x_s y_s$, $\bar{V}_t = V_t + \gamma I$. Substitute $y_s = \langle x_s, \theta^* \rangle + \eta_s$:

$$\hat{\theta}_t = \bar{V}_t^{-1}\big(V_t\theta^* + S_t\big) = \theta^* + \bar{V}_t^{-1}\big(S_t - \gamma\theta^*\big),$$

$$\text{so} \qquad \left\|\hat{\theta}_t - \theta^*\right\|_{\bar{V}_t} = \|S_t - \gamma\theta^*\|_{\bar{V}_t^{-1}}\,.$$

Triangle inequality, then $\bar{V}_t \succeq \gamma I$:
$\|S_t - \gamma\theta^*\|_{\bar{V}_t^{-1}} \le \|S_t\|_{\bar{V}_t^{-1}} + \gamma\,\|\theta^*\|_{\bar{V}_t^{-1}} \le \|S_t\|_{\bar{V}_t^{-1}} + \sqrt{\gamma}\;\|\theta^*\|_2$.

Theorem (the confidence ellipsoid, proved)

If $\|\theta^*\|_2 \le S$, then with probability at least $1-\delta$, simultaneously for all $t \ge 0$,

$$\begin{aligned}\left\|\hat{\theta}_t - \theta^*\right\|_{V_t+\gamma I} &\le \sqrt{\gamma}\,S + \sqrt{2\log\tfrac{1}{\delta} + \log\tfrac{\det(V_t+\gamma I)}{\gamma^d}} \\ &\le \sqrt{\gamma}\,S + \sqrt{2\log\tfrac{1}{\delta} + d\log\big(1 + \tfrac{tL^2}{d\gamma}\big)}.\end{aligned}$$

The last step: $\log\det\bar{V}_t \le d\log(\mathrm{tr}\bar{V}_t/d)$ (AM–GM on the eigenvalues) and $\mathrm{tr}\bar{V}_t \le d\gamma + tL^2$ when $\|x_s\|_2 \le L$. ■

Back to LinUCB: this is why we needed vector martingales

LinUCB plays $x_t = \arg\max_x \langle x, \hat{\theta}_{t-1} \rangle + \beta_{t-1} \|x\|_{V_{t-1}^{-1}}$, using *all* past data, with x_t chosen *because of* past noise. Its analysis needed exactly one thing we could not prove at the time:

The good event

$\theta^* \in \mathcal{C}_t = \{\theta : \left\|\theta - \hat{\theta}_t\right\|_{V_t + \gamma I} \leq \beta_t\}$ for every t simultaneously, with probability $1 - \delta$.

Back to LinUCB: this is why we needed vector martingales

LinUCB plays $x_t = \arg\max_x \langle x, \hat{\theta}_{t-1} \rangle + \beta_{t-1} \|x\|_{V_{t-1}^{-1}}$, using *all* past data, with x_t chosen *because of* past noise. Its analysis needed exactly one thing we could not prove at the time:

The good event

$\theta^* \in \mathcal{C}_t = \{\theta : \left\|\theta - \hat{\theta}_t\right\|_{V_t + \gamma I} \leq \beta_t\}$ for every t simultaneously, with probability $1 - \delta$.

- **Adaptive**, because x_t depends on $y_1, \ldots, y_{t-1}$, so the fixed-design ellipsoid does not apply. The supermartingale $M_t(\lambda)$ only needed $\langle \lambda, x_t \rangle$ known before η_t arrives: exactly what a bandit algorithm guarantees.
- **For all t at once**, because the algorithm consults $\mathcal{C}_{t-1}$ every round and each consultation must be right: Ville, not a union bound over T rounds.
- **Every direction at once**, because the arm set may be infinite or change: the mixture over λ, not a union bound over arms; the $\sqrt{d \log t}$ radius is its price.

The LinUCB proof, now complete

1. **The good event** $\theta^* \in \mathcal{C}_t$ for all t holds with probability $1 - \delta$: today's theorem, via the vector supermartingale, Ville, and the Gaussian mixture.

The LinUCB proof, now complete

1. **The good event** $\theta^* \in \mathcal{C}_t$ for all t holds with probability $1 - \delta$: today's theorem, via the vector supermartingale, Ville, and the Gaussian mixture.
2. **One round**, on the good event: optimism and Cauchy–Schwarz, exactly as in the multi-armed UCB proof,

$$r_t \;\le\; \mathrm{UCB}_t(x_t) - \langle x_t, \theta^* \rangle \;\le\; 2\beta_{t-1} \, \|x_t\|_{V_{t-1}^{-1}} \, .$$

The LinUCB proof, now complete

1. **The good event** $\theta^* \in \mathcal{C}_t$ for all t holds with probability $1-\delta$: today's theorem, via the vector supermartingale, Ville, and the Gaussian mixture.
2. **One round**, on the good event: optimism and Cauchy–Schwarz, exactly as in the multi-armed UCB proof,

$$r_t \;\le\; \mathrm{UCB}_t(x_t) - \langle x_t, \theta^* \rangle \;\le\; 2\beta_{t-1}\,\|x_t\|_{V_{t-1}^{-1}}\,.$$

3. **Summing the widths**: the elliptical potential lemma, $\sum_t \min\{1, \|x_t\|^2_{V_{t-1}^{-1}}\} \le 2d\log(1 + TL^2/(d\gamma))$, which is linear algebra about any arm sequence.

The LinUCB proof, now complete

1. **The good event** $\theta^* \in \mathcal{C}_t$ for all t holds with probability $1-\delta$: today's theorem, via the vector supermartingale, Ville, and the Gaussian mixture.
2. **One round**, on the good event: optimism and Cauchy–Schwarz, exactly as in the multi-armed UCB proof,

$$r_t \;\le\; \mathrm{UCB}_t(x_t) - \langle x_t, \theta^* \rangle \;\le\; 2\beta_{t-1}\, \|x_t\|_{V_{t-1}^{-1}} .$$

3. **Summing the widths**: the elliptical potential lemma, $\sum_t \min\{1, \|x_t\|^2_{V_{t-1}^{-1}}\} \le 2d \log(1 + TL^2/(d\gamma))$, which is linear algebra about any arm sequence.
4. **Cauchy–Schwarz** glues them, so with probability $1-\delta$,

$$R_T \;\le\; 2\beta_T \sqrt{2dT \log\left(1 + \tfrac{TL^2}{d\gamma}\right)} = \widetilde{O}\big(d\sqrt{T}\big), \qquad \beta_T \approx \sqrt{d \log T}.$$

The LinUCB proof, now complete

1. **The good event** $\theta^* \in \mathcal{C}_t$ for all t holds with probability $1-\delta$: today's theorem, via the vector supermartingale, Ville, and the Gaussian mixture.
2. **One round**, on the good event: optimism and Cauchy–Schwarz, exactly as in the multi-armed UCB proof,

$$r_t \;\le\; \mathrm{UCB}_t(x_t) - \langle x_t, \theta^* \rangle \;\le\; 2\beta_{t-1}\,\|x_t\|_{V_{t-1}^{-1}}\,.$$

3. **Summing the widths**: the elliptical potential lemma, $\sum_t \min\{1, \|x_t\|^2_{V_{t-1}^{-1}}\} \le 2d\log(1 + TL^2/(d\gamma))$, which is linear algebra about any arm sequence.
4. **Cauchy–Schwarz** glues them, so with probability $1-\delta$,

$$R_T \;\le\; 2\beta_T\sqrt{2dT\log\left(1+\tfrac{TL^2}{d\gamma}\right)} = \widetilde{O}\big(d\sqrt{T}\big), \qquad \beta_T \approx \sqrt{d\log T}.$$

Both $\sqrt{d}$'s are accounted for: one from the mixture (the radius β_T), one from the potential lemma (how many wide directions there can be). Only step 1 was ever in doubt, and it is the only step that needed a martingale.

What we did, and what it cost

- **Fixed design:** $\left\|\hat{\theta} - \theta^*\right\|_V \lesssim \sqrt{d} + \sqrt{\log(1/\delta)}$, by independence, Chernoff in each direction, and a union bound over a net of the sphere.
- **Adaptive design, all t:** $\left\|\hat{\theta}_t - \theta^*\right\|_{\bar{V}_t} \lesssim \sqrt{d \log t} + \sqrt{\log(1/\delta)}$, by a supermartingale in each direction λ, Ville's inequality in place of Chernoff, and a Gaussian mixture over λ in place of the union bound over the net.
- **The price, $\log\det \bar{V}_t \approx d \log t$,** is the continuous analog of $\log|\mathcal{C}_\epsilon| \approx d$: the mixture has to spread its mass over directions *and* scales, because it does not know in advance how large V_t will grow. It is also what makes the bound valid for all t at once.
- **One construction, three bounds.** Lines ($h = \delta_\lambda$), curves for means (h Gaussian, $d = 1$), and ellipsoids for regression (h Gaussian on $\mathbb{R}^d$). The last one is what every linear bandit algorithm with adaptive data rests on: LinUCB's index, Thompson's inflation, and the analysis of both.

Pure exploration in linear bandits

Playing well is not the same as learning well

A different goal: find the best arm

Same model, $y_t = \langle x_t, \theta^* \rangle + \eta_t$ with arms $\mathcal{X} \subset \mathbb{R}^d$, $|\mathcal{X}| = n$, but now we do not care about regret. Stop as soon as possible and output $\hat{x}$ with $\mathbb{P}(\hat{x} = x^\star) \geq 1 - \delta$, where $x^\star = \arg\max_x \langle x, \theta^* \rangle$ and $\Delta_x = \langle x^\star - x, \theta^* \rangle$.

A different goal: find the best arm

Same model, $y_t = \langle x_t, \theta^* \rangle + \eta_t$ with arms $\mathcal{X} \subset \mathbb{R}^d$, $|\mathcal{X}| = n$, but now we do not care about regret. Stop as soon as possible and output $\hat{x}$ with $\mathbb{P}(\hat{x} = x^\star) \geq 1 - \delta$, where $x^\star = \arg\max_x \langle x, \theta^* \rangle$ and $\Delta_x = \langle x^\star - x, \theta^* \rangle$.

- **What we already have.** One G-optimal design with $\tau \approx d\,\Delta_{\min}^{-2} \log(n/\delta)$ pulls makes every arm's interval narrower than $\Delta_{\min}/2$ and identifies $x^\star$; elimination gets the same. No adaptivity was needed, which should make us suspicious.

A different goal: find the best arm

Same model, $y_t = \langle x_t, \theta^* \rangle + \eta_t$ with arms $\mathcal{X} \subset \mathbb{R}^d$, $|\mathcal{X}| = n$, but now we do not care about regret. Stop as soon as possible and output $\hat{x}$ with $\mathbb{P}(\hat{x} = x^\star) \geq 1 - \delta$, where $x^\star = \arg\max_x \langle x, \theta^* \rangle$ and $\Delta_x = \langle x^\star - x, \theta^* \rangle$.

- **What we already have.** One G-optimal design with $\tau \approx d\,\Delta_{\min}^{-2} \log(n/\delta)$ pulls makes every arm's interval narrower than $\Delta_{\min}/2$ and identifies $x^\star$; elimination gets the same. No adaptivity was needed, which should make us suspicious.
- **LinUCB or Thompson sampling?** Near-optimal regret; run them and output the arm played most?

A different goal: find the best arm

Same model, $y_t = \langle x_t, \theta^* \rangle + \eta_t$ with arms $\mathcal{X} \subset \mathbb{R}^d$, $|\mathcal{X}| = n$, but now we do not care about regret. Stop as soon as possible and output $\hat{x}$ with $\mathbb{P}(\hat{x} = x^\star) \geq 1 - \delta$, where $x^\star = \arg\max_x \langle x, \theta^* \rangle$ and $\Delta_x = \langle x^\star - x, \theta^* \rangle$.

- **What we already have.** One G-optimal design with $\tau \approx d\,\Delta_{\min}^{-2} \log(n/\delta)$ pulls makes every arm's interval narrower than $\Delta_{\min}/2$ and identifies $x^\star$; elimination gets the same. No adaptivity was needed, which should make us suspicious.
- **LinUCB or Thompson sampling?** Near-optimal regret; run them and output the arm played most?
- In the multi-armed case low-regret algorithms are decent explorers: to learn that arm i is bad you must pull arm i, about Δ_i^{-2} times either way.

A different goal: find the best arm

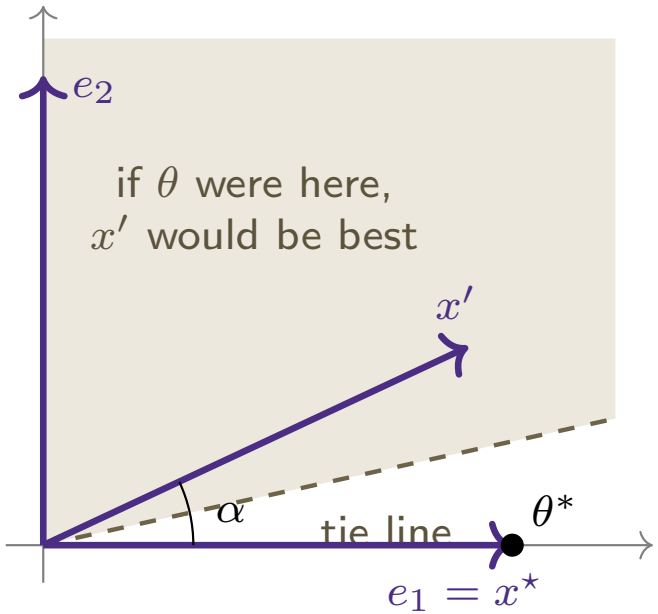
Same model, $y_t = \langle x_t, \theta^* \rangle + \eta_t$ with arms $\mathcal{X} \subset \mathbb{R}^d$, $|\mathcal{X}| = n$, but now we do not care about regret. Stop as soon as possible and output $\hat{x}$ with $\mathbb{P}(\hat{x} = x^\star) \geq 1 - \delta$, where $x^\star = \arg\max_x \langle x, \theta^* \rangle$ and $\Delta_x = \langle x^\star - x, \theta^* \rangle$.

- **What we already have.** One G-optimal design with $\tau \approx d\,\Delta_{\min}^{-2} \log(n/\delta)$ pulls makes every arm's interval narrower than $\Delta_{\min}/2$ and identifies $x^\star$; elimination gets the same. No adaptivity was needed, which should make us suspicious.
- **LinUCB or Thompson sampling?** Near-optimal regret; run them and output the arm played most?
- In the multi-armed case low-regret algorithms are decent explorers: to learn that arm i is bad you must pull arm i, about Δ_i^{-2} times either way.

With features this breaks

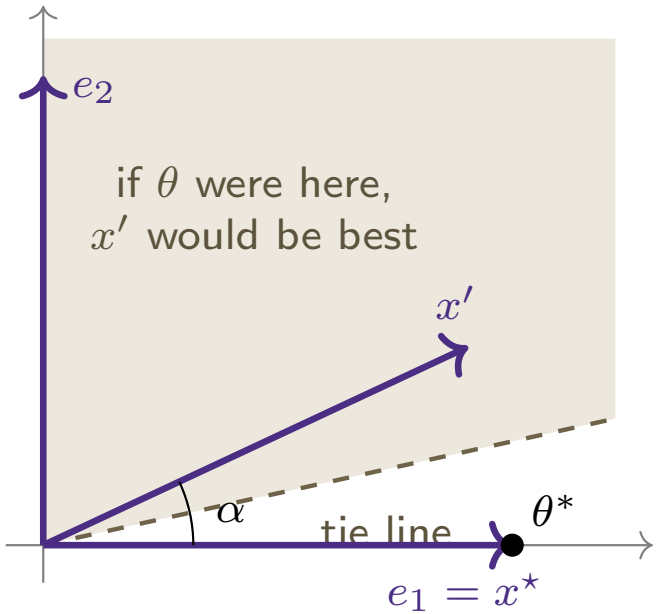
The arm that tells you the most about a comparison need not be either of the arms being compared. A regret-minimizing algorithm refuses to pull bad arms; a good explorer may *need* to. The cleanest illustration is a three-arm instance due to Soare, Lazaric & Munos (2014).

The Soare instance



$\mathcal{X} = \{e_1, e_2, x'\}$ in $\mathbb{R}^2$, with
$x' = (\cos\alpha, \sin\alpha)$ for a small angle α,
and $\theta^* = e_1$.

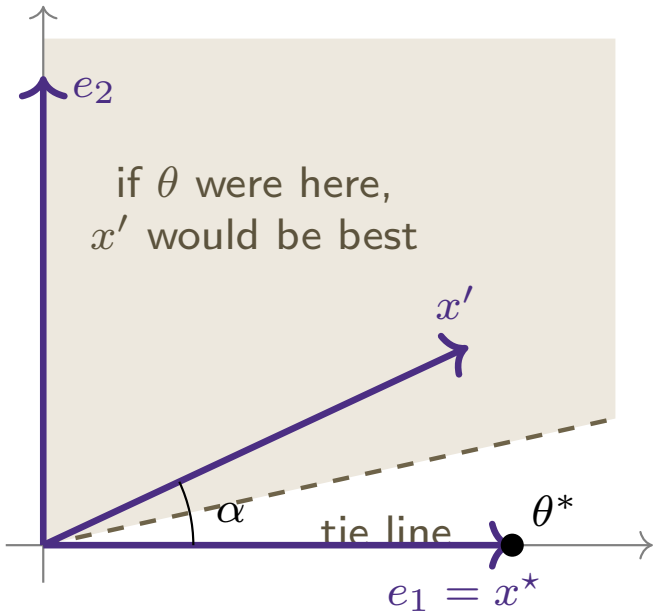
The Soare instance



$\mathcal{X} = \{e_1, e_2, x'\}$ in $\mathbb{R}^2$, with $x' = (\cos\alpha, \sin\alpha)$ for a small angle α, and $\theta^* = e_1$.

- Rewards: $e_1 \mapsto 1$, $x' \mapsto \cos\alpha \approx 1 - \alpha^2/2$, $e_2 \mapsto 0$.
- e_2 is terrible (gap 1). x' is nearly optimal (gap $\approx \alpha^2/2$) and is the only arm that is hard to rule out.

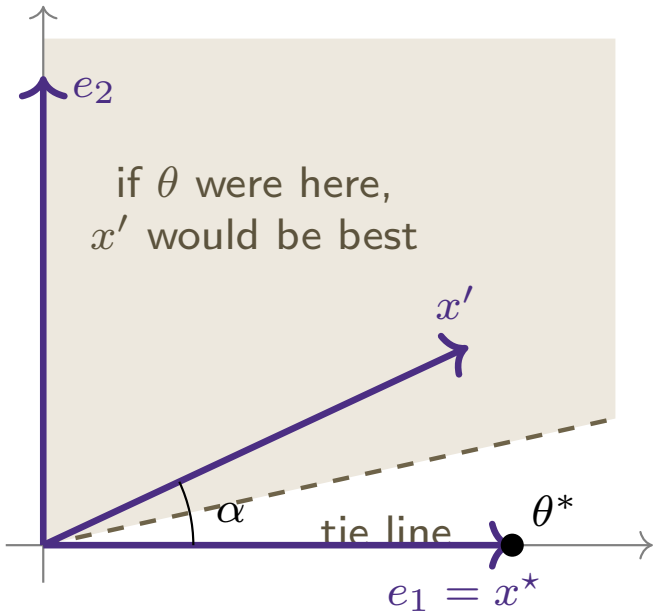
The Soare instance



$\mathcal{X} = \{e_1, e_2, x'\}$ in $\mathbb{R}^2$, with $x' = (\cos\alpha, \sin\alpha)$ for a small angle α, and $\theta^* = e_1$.

- Rewards: $e_1 \mapsto 1$, $x' \mapsto \cos\alpha \approx 1 - \alpha^2/2$, $e_2 \mapsto 0$.
- e_2 is terrible (gap 1). x' is nearly optimal (gap $\approx \alpha^2/2$) and is the only arm that is hard to rule out.
- To rule out x' we must show θ^* is below the dashed tie line $\{\theta : \langle e_1 - x', \theta\rangle = 0\}$.

The Soare instance

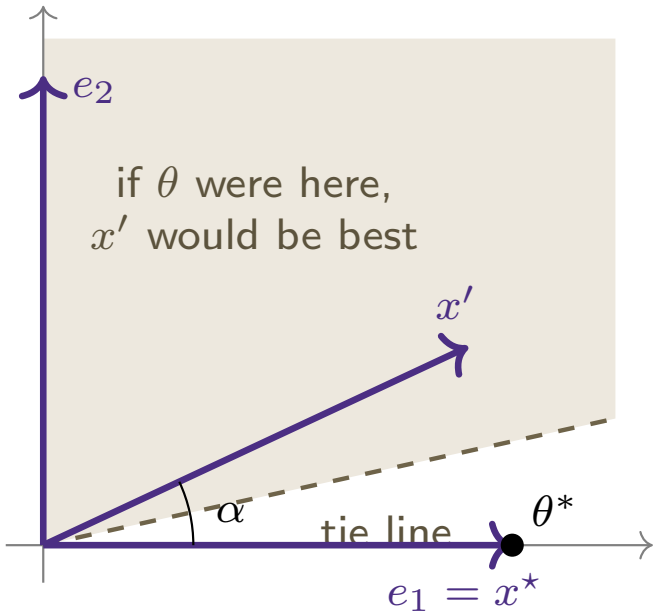


$\mathcal{X} = \{e_1, e_2, x'\}$ in $\mathbb{R}^2$, with $x' = (\cos\alpha, \sin\alpha)$ for a small angle α, and $\theta^* = e_1$.

- Rewards: $e_1 \mapsto 1$, $x' \mapsto \cos\alpha \approx 1 - \alpha^2/2$, $e_2 \mapsto 0$.
- e_2 is terrible (gap 1). x' is nearly optimal (gap $\approx \alpha^2/2$) and is the only arm that is hard to rule out.
- To rule out x' we must show θ^* is below the dashed tie line $\{\theta : \langle e_1 - x', \theta\rangle = 0\}$.

Which direction matters? $e_1 - x' = (1 - \cos\alpha,\ -\sin\alpha) \approx (\alpha^2/2,\ -\alpha)$: almost exactly e_2. Whether e_1 beats x' is a question about the *second coordinate* of θ^*, and the arm that measures it directly is e_2, the worst arm on the table.

The Soare instance



$\mathcal{X} = \{e_1, e_2, x'\}$ in $\mathbb{R}^2$, with $x' = (\cos\alpha, \sin\alpha)$ for a small angle α, and $\theta^* = e_1$.

- Rewards: $e_1 \mapsto 1$, $x' \mapsto \cos\alpha \approx 1 - \alpha^2/2$, $e_2 \mapsto 0$.
- e_2 is terrible (gap 1). x' is nearly optimal (gap $\approx \alpha^2/2$) and is the only arm that is hard to rule out.
- To rule out x' we must show θ^* is below the dashed tie line $\{\theta : \langle e_1 - x', \theta \rangle = 0\}$.

Which direction matters? $e_1 - x' = (1 - \cos\alpha,\ -\sin\alpha) \approx (\alpha^2/2,\ -\alpha)$: almost exactly e_2. Whether e_1 beats x' is a question about the *second coordinate* of θ^*, and the arm that measures it directly is e_2, the worst arm on the table. Pulling e_1 and x' measures θ^* along two nearly parallel directions: a lot about the first coordinate, almost nothing about the second.

What it takes to stop: each gap must dominate its error

After τ pulls from a design λ we have $\hat{\theta}$. To declare $x^\star$ the winner, $\langle x^\star - x, \hat{\theta} \rangle > 0$ must reflect the truth for *every* other x:

$$\langle x^\star - x, \hat{\theta} \rangle \;=\; \underbrace{\langle x^\star - x, \theta^* \rangle}_{\text{signal } \Delta_x} \;+\; \underbrace{\langle x^\star - x, \hat{\theta} - \theta^* \rangle}_{\text{error in direction } x^\star - x} .$$

What it takes to stop: each gap must dominate its error

After τ pulls from a design λ we have $\hat{\theta}$. To declare $x^\star$ the winner, $\langle x^\star - x, \hat{\theta} \rangle > 0$ must reflect the truth for *every* other x:

$$\langle x^\star - x, \hat{\theta} \rangle \;=\; \underbrace{\langle x^\star - x, \theta^* \rangle}_{\text{signal } \Delta_x} \;+\; \underbrace{\langle x^\star - x, \hat{\theta} - \theta^* \rangle}_{\text{error in direction } x^\star - x} .$$

The error in a direction z is at most $\|z\|_{A(\lambda)^{-1}} \sqrt{2\log(n/\delta)/\tau}$ (fixed design, union bound). So we can stop once, for every $x \neq x^\star$,

$$\|x^\star - x\|_{A(\lambda)^{-1}} \sqrt{\frac{2\log(n/\delta)}{\tau}} \;\leq\; \Delta_x \quad\Longleftrightarrow\quad \tau \;\geq$$

What it takes to stop: each gap must dominate its error

After τ pulls from a design λ we have $\hat{\theta}$. To declare $x^\star$ the winner, $\langle x^\star - x, \hat{\theta} \rangle > 0$ must reflect the truth for *every* other x:

$$\langle x^\star - x, \hat{\theta} \rangle \;=\; \underbrace{\langle x^\star - x, \theta^* \rangle}_{\text{signal } \Delta_x} \;+\; \underbrace{\langle x^\star - x, \hat{\theta} - \theta^* \rangle}_{\text{error in direction } x^\star - x} .$$

The error in a direction z is at most $\|z\|_{A(\lambda)^{-1}} \sqrt{2\log(n/\delta)/\tau}$ (fixed design, union bound). So we can stop once, for every $x \neq x^\star$,

$$\|x^\star - x\|_{A(\lambda)^{-1}} \sqrt{\frac{2\log(n/\delta)}{\tau}} \;\leq\; \Delta_x \quad\Longleftrightarrow\quad \tau \;\geq\; 2\log(n/\delta)\, \frac{\|x^\star - x\|^2_{A(\lambda)^{-1}}}{\Delta_x^2}.$$

What it takes to stop: each gap must dominate its error

After τ pulls from a design λ we have $\hat{\theta}$. To declare $x^\star$ the winner, $\langle x^\star - x, \hat{\theta}\rangle > 0$ must reflect the truth for *every* other x:

$$\langle x^\star - x, \hat{\theta}\rangle \;=\; \underbrace{\langle x^\star - x, \theta^*\rangle}_{\text{signal } \Delta_x} \;+\; \underbrace{\langle x^\star - x, \hat{\theta} - \theta^*\rangle}_{\text{error in direction } x^\star - x}.$$

The error in a direction z is at most $\|z\|_{A(\lambda)^{-1}} \sqrt{2\log(n/\delta)/\tau}$ (fixed design, union bound). So we can stop once, for every $x \neq x^\star$,

$$\|x^\star - x\|_{A(\lambda)^{-1}} \sqrt{\frac{2\log(n/\delta)}{\tau}} \;\le\; \Delta_x \quad\Longleftrightarrow\quad \tau \;\ge\; 2\log(n/\delta)\,\frac{\|x^\star - x\|^2_{A(\lambda)^{-1}}}{\Delta_x^2}.$$

The worst x sets the sample size, and the design is ours to choose:

The complexity of pure exploration (Soare et al. 2014)

$$\rho^\star \;:=\; \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \neq x^\star} \frac{\|x^\star - x\|^2_{A(\lambda)^{-1}}}{\langle x^\star - x, \theta^*\rangle^2}, \qquad \tau \approx \rho^\star \log(n/\delta) \text{ pulls suffice.}$$

What it takes to stop: each gap must dominate its error

After τ pulls from a design λ we have $\hat{\theta}$. To declare $x^\star$ the winner, $\langle x^\star - x, \hat{\theta}\rangle > 0$ must reflect the truth for *every* other x:

$$\langle x^\star - x, \hat{\theta}\rangle \;=\; \underbrace{\langle x^\star - x, \theta^*\rangle}_{\text{signal } \Delta_x} \;+\; \underbrace{\langle x^\star - x, \hat{\theta} - \theta^*\rangle}_{\text{error in direction } x^\star - x}.$$

The error in a direction z is at most $\|z\|_{A(\lambda)^{-1}} \sqrt{2\log(n/\delta)/\tau}$ (fixed design, union bound). So we can stop once, for every $x \neq x^\star$,

$$\|x^\star - x\|_{A(\lambda)^{-1}} \sqrt{\frac{2\log(n/\delta)}{\tau}} \;\le\; \Delta_x \quad\Longleftrightarrow\quad \tau \;\ge\; 2\log(n/\delta)\,\frac{\|x^\star - x\|^2_{A(\lambda)^{-1}}}{\Delta_x^2}.$$

The worst x sets the sample size, and the design is ours to choose:

The complexity of pure exploration (Soare et al. 2014)

$$\rho^\star \;:=\; \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \neq x^\star} \frac{\|x^\star - x\|^2_{A(\lambda)^{-1}}}{\langle x^\star - x, \theta^*\rangle^2}, \qquad \tau \approx \rho^\star \log(n/\delta) \text{ pulls suffice.}$$

G-optimal controls every *arm* equally; this controls every *difference* $x^\star - x$, as finely as its own gap requires.

Back to the Soare instance: α^{-2} versus α^{-4}

The hard comparison is $x^\star - x' \approx (\alpha^2/2, -\alpha)$ with gap $\Delta_{x'} \approx \alpha^2/2$. Three ways to spend the samples:

strategy	error^2 in direction $x^\star - x'$, per sample	samples needed
pull only the good arms e_1, x'	≈ 1	$\approx \alpha^{-4} \log(1/\delta)$
G-optimal design, once	$\approx d$ over the worst arm	$\approx d\,\alpha^{-4} \log(n/\delta)$
the $\rho^\star$ design: mostly e_2	$\approx \alpha^2$	$\approx \alpha^{-2} \log(1/\delta)$

Back to the Soare instance: α^{-2} versus α^{-4}

The hard comparison is $x^\star - x' \approx (\alpha^2/2, -\alpha)$ with gap $\Delta_{x'} \approx \alpha^2/2$. Three ways to spend the samples:

strategy	error^2 in direction $x^\star - x'$, per sample	samples needed
pull only the good arms e_1, x'	≈ 1	$\approx \alpha^{-4} \log(1/\delta)$
G-optimal design, once	$\approx d$ over the worst arm	$\approx d\,\alpha^{-4} \log(n/\delta)$
the $\rho^\star$ design: mostly e_2	$\approx \alpha^2$	$\approx \alpha^{-2} \log(1/\delta)$

- Restricted to $\{e_1, x'\}$ this is a two-armed bandit with gap $\alpha^2/2$: $\approx \alpha^{-4}$ samples. That is what LinUCB and Thompson do: after a few rounds e_2 is obviously bad and optimism never touches it again.

Back to the Soare instance: α^{-2} versus α^{-4}

The hard comparison is $x^\star - x' \approx (\alpha^2/2, -\alpha)$ with gap $\Delta_{x'} \approx \alpha^2/2$. Three ways to spend the samples:

strategy	error^2 in direction $x^\star - x'$, per sample	samples needed
pull only the good arms e_1, x'	≈ 1	$\approx \alpha^{-4} \log(1/\delta)$
G-optimal design, once	$\approx d$ over the worst arm	$\approx d\,\alpha^{-4} \log(n/\delta)$
the $\rho^\star$ design: mostly e_2	$\approx \alpha^2$	$\approx \alpha^{-2} \log(1/\delta)$

- Restricted to $\{e_1, x'\}$ this is a two-armed bandit with gap $\alpha^2/2$: $\approx \alpha^{-4}$ samples. That is what LinUCB and Thompson do: after a few rounds e_2 is obviously bad and optimism never touches it again.
- Pulling e_2 measures the second coordinate of θ^* with variance 1 per sample, and it only has to be resolved to accuracy $\approx \alpha/2$: $\approx \alpha^{-2}$ samples, by sampling the arm with the largest regret.

Back to the Soare instance: α^{-2} versus α^{-4}

The hard comparison is $x^\star - x' \approx (\alpha^2/2, -\alpha)$ with gap $\Delta_{x'} \approx \alpha^2/2$. Three ways to spend the samples:

strategy	error^2 in direction $x^\star - x'$, per sample	samples needed
pull only the good arms e_1, x'	≈ 1	$\approx \alpha^{-4} \log(1/\delta)$
G-optimal design, once	$\approx d$ over the worst arm	$\approx d\,\alpha^{-4} \log(n/\delta)$
the $\rho^\star$ design: mostly e_2	$\approx \alpha^2$	$\approx \alpha^{-2} \log(1/\delta)$

- Restricted to $\{e_1, x'\}$ this is a two-armed bandit with gap $\alpha^2/2$: $\approx \alpha^{-4}$ samples. That is what LinUCB and Thompson do: after a few rounds e_2 is obviously bad and optimism never touches it again.
- Pulling e_2 measures the second coordinate of θ^* with variance 1 per sample, and it only has to be resolved to accuracy $\approx \alpha/2$: $\approx \alpha^{-2}$ samples, by sampling the arm with the largest regret.
- Regret and pure exploration want different designs: "which arm might be best?" versus "which arm best separates the candidates?"

Is $\rho^\star$ the truth? A sketch of the lower bound

Lecture 8's template again: a δ-PAC algorithm must behave differently on any instance θ' whose best arm is not $x^\star$.

Is $\rho^\star$ the truth? A sketch of the lower bound

Lecture 8's template again: a δ-PAC algorithm must behave differently on any instance θ' whose best arm is not $x^\star$.

1. **The cheapest alternative for a given x:** move θ^* just across the tie hyperplane $\{\theta : \langle x^\star - x, \theta \rangle = 0\}$, so that x beats $x^\star$.

Is $\rho^\star$ the truth? A sketch of the lower bound

Lecture 8's template again: a δ-PAC algorithm must behave differently on any instance θ' whose best arm is not $x^\star$.

1. **The cheapest alternative for a given x:** move θ^* just across the tie hyperplane $\{\theta : \langle x^\star - x, \theta \rangle = 0\}$, so that x beats $x^\star$.
2. **How different do the transcripts look?** If the pulls are spread like λ over τ rounds, $\mathrm{KL} = \frac{\tau}{2} \|\theta^* - \theta'\|^2_{A(\lambda)}$ (information adds up over pulls), and the closest point across the hyperplane gives

$$\min_{\theta'} \ \frac{\tau}{2} \|\theta^* - \theta'\|^2_{A(\lambda)} \ = \ \frac{\tau \, \Delta_x^2}{2 \, \|x^\star - x\|^2_{A(\lambda)^{-1}}}.$$

Is $\rho^\star$ the truth? A sketch of the lower bound

Lecture 8's template again: a δ-PAC algorithm must behave differently on any instance θ' whose best arm is not $x^\star$.

1. **The cheapest alternative for a given x:** move θ^* just across the tie hyperplane $\{\theta : \langle x^\star - x, \theta \rangle = 0\}$, so that x beats $x^\star$.
2. **How different do the transcripts look?** If the pulls are spread like λ over τ rounds, $\mathrm{KL} = \frac{\tau}{2} \|\theta^* - \theta'\|^2_{A(\lambda)}$ (information adds up over pulls), and the closest point across the hyperplane gives

$$\min_{\theta'} \ \tfrac{\tau}{2} \|\theta^* - \theta'\|^2_{A(\lambda)} \ = \ \frac{\tau \, \Delta_x^2}{2 \, \|x^\star - x\|^2_{A(\lambda)^{-1}}}.$$

3. **Telling them apart with error δ needs $\mathrm{KL} \gtrsim \log(1/\delta)$,** for every $x \neq x^\star$ and whatever λ the algorithm uses:

$$\mathbb{E}[\tau] \ \gtrsim \ 2 \log \tfrac{1}{2.4\delta} \ \min_{\lambda} \max_{x \neq x^\star} \frac{\|x^\star - x\|^2_{A(\lambda)^{-1}}}{\Delta_x^2} \ = \ 2\rho^\star \log \tfrac{1}{2.4\delta}.$$

Is $\rho^\star$ the truth? A sketch of the lower bound

Lecture 8's template again: a δ-PAC algorithm must behave differently on any instance θ' whose best arm is not $x^\star$.

1. **The cheapest alternative for a given x:** move θ^* just across the tie hyperplane $\{\theta : \langle x^\star - x, \theta \rangle = 0\}$, so that x beats $x^\star$.
2. **How different do the transcripts look?** If the pulls are spread like λ over τ rounds, $\mathrm{KL} = \frac{\tau}{2} \|\theta^* - \theta'\|^2_{A(\lambda)}$ (information adds up over pulls), and the closest point across the hyperplane gives

$$\min_{\theta'} \ \tfrac{\tau}{2} \|\theta^* - \theta'\|^2_{A(\lambda)} \ = \ \frac{\tau \, \Delta_x^2}{2 \, \|x^\star - x\|^2_{A(\lambda)^{-1}}}.$$

3. **Telling them apart with error δ needs $\mathrm{KL} \gtrsim \log(1/\delta)$,** for every $x \neq x^\star$ and whatever λ the algorithm uses:

$$\mathbb{E}[\tau] \ \gtrsim \ 2 \log \tfrac{1}{2.4\delta} \ \min_{\lambda} \max_{x \neq x^\star} \frac{\|x^\star - x\|^2_{A(\lambda)^{-1}}}{\Delta_x^2} \ = \ 2\rho^\star \log \tfrac{1}{2.4\delta}.$$

The same ratio as in the upper bound, signal against error: $\rho^\star \log(1/\delta)$ is the right answer. On the Soare instance, α^{-2}.

The algorithm: elimination, with the right design

$\rho^\star$ depends on the unknown θ^* and $x^\star$. The usual fix: rounds of increasing precision, with a design that only needs the survivors.

RAGE (Fiez, Jain, Jamieson & Ratliff 2019; notes §11.5)

Initialize $\mathcal{X}_1 = \mathcal{X}$. **for** $\ell = 1, 2, \ldots$ while $|\mathcal{X}_\ell| > 1$:

$\hat{\lambda}_\ell = \arg\min_{\lambda \in \triangle_\mathcal{X}} \max_{x, x' \in \mathcal{X}_\ell} \|x - x'\|^2_{A(\lambda)^{-1}}$ (a design over *all* of $\mathcal{X}$ that best separates the survivors)

pull $\tau_\ell \approx \epsilon_\ell^{-2}$ (that min-max value) $\log(\ell^2 n/\delta)$ times from $\hat{\lambda}_\ell$; least squares $\hat{\theta}_\ell$ on this round's data

$\mathcal{X}_{\ell+1} = \mathcal{X}_\ell \setminus \{x : \max_{x' \in \mathcal{X}_\ell} \langle x' - x, \hat{\theta}_\ell \rangle > \epsilon_\ell\}, \quad \epsilon_\ell = 2^{-\ell}$

The algorithm: elimination, with the right design

$\rho^\star$ depends on the unknown θ^* and $x^\star$. The usual fix: rounds of increasing precision, with a design that only needs the survivors.

RAGE (Fiez, Jain, Jamieson & Ratliff 2019; notes §11.5)

Initialize $\mathcal{X}_1 = \mathcal{X}$. **for** $\ell = 1, 2, \dots$ while $|\mathcal{X}_\ell| > 1$:

$\hat{\lambda}_\ell = \arg\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{x, x' \in \mathcal{X}_\ell} \|x - x'\|^2_{A(\lambda)^{-1}}$ (a design over *all* of $\mathcal{X}$ that best separates the survivors)

pull $\tau_\ell \approx \epsilon_\ell^{-2}$ (that min-max value) $\log(\ell^2 n/\delta)$ times from $\hat{\lambda}_\ell$; least squares $\hat{\theta}_\ell$ on this round's data

$\mathcal{X}_{\ell+1} = \mathcal{X}_\ell \setminus \{x : \max_{x' \in \mathcal{X}_\ell} \langle x' - x, \hat{\theta}_\ell \rangle > \epsilon_\ell\}, \quad \epsilon_\ell = 2^{-\ell}$

- **Guarantee:** with probability $1 - \delta$ it returns $x^\star$ after $\lesssim \rho^\star \log(1/\Delta_{\min}) \big[\log(1/\delta) + \log n\big]$ pulls: the lower bound, up to logs.
- **Two changes from last week's elimination.** The design targets *differences* between survivors, not the survivors themselves; and it ranges over all arms, so it can put weight on arms already known to be bad. On the Soare instance it pulls e_2.

Adversarial linear bandits

EXP3 with features, and a design in place of uniform exploration

The setting, and what we already have

Adversarial linear bandit with finitely many arms (notes §7.3)

Known: arms $\mathcal{X} = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$. **Adversary** fixes $\theta_1, \dots, \theta_T \in \mathbb{R}^d$ in advance, $|\langle x_i, \theta_t \rangle| \le 1$.
for $t = 1, \dots, T$: draw $I_t \sim q_t \in \triangle_n$; suffer and observe $\ell_{t,I_t} = \langle x_{I_t}, \theta_t \rangle$ only.
Regret: $\max_i \mathbb{E}\big[\sum_t \langle x_{I_t} - x_i, \theta_t \rangle\big]$.

No noise, no distribution. Ignoring the features, EXP3 gives $\sqrt{12\, nT \log n}$ (Lecture 8). Can the linear structure make d replace n, as in the stochastic case?

The setting, and what we already have

Adversarial linear bandit with finitely many arms (notes §7.3)

Known: arms $\mathcal{X} = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$. **Adversary** fixes $\theta_1, \dots, \theta_T \in \mathbb{R}^d$ in advance, $|\langle x_i, \theta_t \rangle| \le 1$.
for $t = 1, \dots, T$: draw $I_t \sim q_t \in \triangle_n$; suffer and observe $\ell_{t,I_t} = \langle x_{I_t}, \theta_t \rangle$ only.
Regret: $\max_i \mathbb{E}\big[\sum_t \langle x_{I_t} - x_i, \theta_t \rangle\big]$.

No noise, no distribution. Ignoring the features, EXP3 gives $\sqrt{12\, nT \log n}$ (Lecture 8). Can the linear structure make d replace n, as in the stochastic case?

Recall Theorem 12 (Lecture 8)

$I_t \sim q_t = (1-\gamma)p_t + \gamma\lambda$, p_t exponential weights on estimates $\hat{\ell}_{t,i}$ with $\mathbb{E}[\hat{\ell}_{t,i} \mid \mathcal{F}_{t-1}] = \ell_{t,i}$ and $-\eta\hat{\ell}_{t,i} \le 1$:

$$\max_i \mathbb{E}\Big[\sum_{t=1}^{T} \ell_{t,I_t} - \ell_{t,i}\Big] \;\le\; 2\gamma T + \frac{\log n}{\eta} + (1-\gamma)\,\eta\,\mathbb{E}\Big[\sum_{t=1}^{T}\sum_{j=1}^{n} p_{t,j}\hat{\ell}_{t,j}^2\Big].$$

The estimator: least squares with one sample

Write $A(q) = \sum_{i=1}^{n} q_i \, x_i x_i^\top$, the design matrix of the sampling distribution. Estimate the whole loss vector θ_t from the single observation (x_{I_t}, ℓ_{t,I_t}):

$$\hat{\theta}_t \;=\; A(q_t)^{-1} \, x_{I_t} \, \ell_{t,I_t}, \qquad \hat{\ell}_{t,i} \;=\; \langle x_i, \hat{\theta}_t \rangle = x_i^\top A(q_t)^{-1} x_{I_t} \, \ell_{t,I_t}.$$

The estimator: least squares with one sample

Write $A(q) = \sum_{i=1}^{n} q_i \, x_i x_i^\top$, the design matrix of the sampling distribution. Estimate the whole loss vector θ_t from the single observation (x_{I_t}, ℓ_{t,I_t}):

$$\hat{\theta}_t \;=\; A(q_t)^{-1} \, x_{I_t} \, \ell_{t,I_t}, \qquad \hat{\ell}_{t,i} \;=\; \langle x_i, \hat{\theta}_t \rangle = x_i^\top A(q_t)^{-1} x_{I_t} \, \ell_{t,I_t}.$$

Unbiased. Since $I_t \sim q_t$ given the past and $\ell_{t,I_t} = x_{I_t}^\top \theta_t$,

$$\mathbb{E}[\hat{\theta}_t \mid \mathcal{F}_{t-1}] = A(q_t)^{-1} \, \mathbb{E}\big[x_{I_t} x_{I_t}^\top \mid \mathcal{F}_{t-1}\big] \, \theta_t =$$

The estimator: least squares with one sample

Write $A(q) = \sum_{i=1}^{n} q_i \, x_i x_i^\top$, the design matrix of the sampling distribution. Estimate the whole loss vector θ_t from the single observation (x_{I_t}, ℓ_{t,I_t}):

$$\hat{\theta}_t \;=\; A(q_t)^{-1} \, x_{I_t} \, \ell_{t,I_t}, \qquad \hat{\ell}_{t,i} \;=\; \langle x_i, \hat{\theta}_t \rangle = x_i^\top A(q_t)^{-1} x_{I_t} \, \ell_{t,I_t}.$$

Unbiased. Since $I_t \sim q_t$ given the past and $\ell_{t,I_t} = x_{I_t}^\top \theta_t$,

$$\mathbb{E}[\hat{\theta}_t \mid \mathcal{F}_{t-1}] = A(q_t)^{-1} \, \mathbb{E}\big[x_{I_t} x_{I_t}^\top \mid \mathcal{F}_{t-1}\big] \, \theta_t = A(q_t)^{-1} \Big(\sum_{j=1}^{n} q_{t,j} x_j x_j^\top \Big) \theta_t = \theta_t,$$

so $\mathbb{E}[\hat{\ell}_{t,i} \mid \mathcal{F}_{t-1}] = \langle x_i, \theta_t \rangle = \ell_{t,i}$ for *every* arm, including arms never played.

The estimator: least squares with one sample

Write $A(q) = \sum_{i=1}^{n} q_i \, x_i x_i^\top$, the design matrix of the sampling distribution. Estimate the whole loss vector θ_t from the single observation (x_{I_t}, ℓ_{t,I_t}):

$$\hat{\theta}_t \;=\; A(q_t)^{-1} \, x_{I_t} \, \ell_{t,I_t}, \qquad \hat{\ell}_{t,i} \;=\; \langle x_i, \hat{\theta}_t \rangle = x_i^\top A(q_t)^{-1} x_{I_t} \, \ell_{t,I_t}.$$

Unbiased. Since $I_t \sim q_t$ given the past and $\ell_{t,I_t} = x_{I_t}^\top \theta_t$,

$$\mathbb{E}[\hat{\theta}_t \mid \mathcal{F}_{t-1}] = A(q_t)^{-1} \, \mathbb{E}\big[x_{I_t} x_{I_t}^\top \mid \mathcal{F}_{t-1}\big] \, \theta_t = A(q_t)^{-1} \Big(\sum_{j=1}^{n} q_{t,j} x_j x_j^\top \Big) \theta_t = \theta_t,$$

so $\mathbb{E}[\hat{\ell}_{t,i} \mid \mathcal{F}_{t-1}] = \langle x_i, \theta_t \rangle = \ell_{t,i}$ for *every* arm, including arms never played.

- Compare the multi-armed estimator $\hat{\ell}_{t,i} = \mathbf{1}\{I_t = i\} \ell_{t,I_t} / q_{t,i}$: it is the special case $\mathcal{X} = \{e_i\}$, where $A(q) = \mathrm{diag}(q)$. Importance weighting *is* one-sample least squares.
- One pull informs every arm, through $A(q_t)^{-1}$. The price is variance: $\hat{\ell}_{t,i}$ can be as large as $\|x_i\|_{A(q_t)^{-1}} \, \|x_{I_t}\|_{A(q_t)^{-1}}$ (Cauchy–Schwarz), which is where the exploration distribution λ will matter.

Checking the two conditions

Boundedness. $q_t = (1-\gamma)p_t + \gamma\lambda$ gives $A(q_t) \succeq \gamma A(\lambda)$, so $\|x\|^2_{A(q_t)^{-1}} \le \gamma^{-1} \|x\|^2_{A(\lambda)^{-1}}$. Cauchy–Schwarz and $|\ell_{t,I_t}| \le 1$:

$$|\eta\hat{\ell}_{t,i}| \le \eta\,|x_i^\top A(q_t)^{-1} x_{I_t}| \;\le$$

Checking the two conditions

Boundedness. $q_t = (1-\gamma)p_t + \gamma\lambda$ gives $A(q_t) \succeq \gamma A(\lambda)$, so $\|x\|^2_{A(q_t)^{-1}} \le \gamma^{-1} \|x\|^2_{A(\lambda)^{-1}}$. Cauchy–Schwarz and $|\ell_{t,I_t}| \le 1$:

$$|\eta\hat{\ell}_{t,i}| \le \eta\, |x_i^\top A(q_t)^{-1} x_{I_t}| \;\le\; \eta\, \|x_i\|_{A(q_t)^{-1}} \|x_{I_t}\|_{A(q_t)^{-1}} \;\le\; \tfrac{\eta}{\gamma} \max_j \|x_j\|^2_{A(\lambda)^{-1}}$$

We choose λ: make the right side small, i.e. the G-optimal design, and Kiefer–Wolfowitz says $\max_j \|x_j\|^2_{A(\lambda^\star)^{-1}} = d$. With $\gamma = \eta d$, $|\eta\hat{\ell}_{t,i}| \le 1$.

Checking the two conditions

Boundedness. $q_t = (1-\gamma)p_t + \gamma\lambda$ gives $A(q_t) \succeq \gamma A(\lambda)$, so $\|x\|^2_{A(q_t)^{-1}} \le \gamma^{-1} \|x\|^2_{A(\lambda)^{-1}}$. Cauchy–Schwarz and $|\ell_{t,I_t}| \le 1$:

$$|\eta\hat{\ell}_{t,i}| \le \eta\,|x_i^\top A(q_t)^{-1} x_{I_t}| \;\le\; \eta\,\|x_i\|_{A(q_t)^{-1}}\,\|x_{I_t}\|_{A(q_t)^{-1}} \;\le\; \tfrac{\eta}{\gamma}\max_j \|x_j\|^2_{A(\lambda)^{-1}}$$

We choose λ: make the right side small, i.e. the G-optimal design, and Kiefer–Wolfowitz says $\max_j \|x_j\|^2_{A(\lambda^\star)^{-1}} = d$. With $\gamma = \eta d$, $|\eta\hat{\ell}_{t,i}| \le 1$.

Second moment. With $\ell^2 \le 1$ and $\mathbb{E}[x_{I_t} x_{I_t}^\top \mid \mathcal{F}_{t-1}] = A(q_t) \succeq (1-\gamma)A(p_t)$:

$$\begin{aligned}
\mathbb{E}\Big[\sum_j p_{t,j}\hat{\ell}^2_{t,j} \,\Big|\, \mathcal{F}_{t-1}\Big] &\le \sum_j p_{t,j}\, x_j^\top A(q_t)^{-1}\, \mathbb{E}\big[x_{I_t} x_{I_t}^\top \mid \mathcal{F}_{t-1}\big] A(q_t)^{-1} x_j \\
&= \\
&\le
\end{aligned}$$

Checking the two conditions

Boundedness. $q_t = (1-\gamma)p_t + \gamma\lambda$ gives $A(q_t) \succeq \gamma A(\lambda)$, so $\|x\|^2_{A(q_t)^{-1}} \le \gamma^{-1}\|x\|^2_{A(\lambda)^{-1}}$. Cauchy–Schwarz and $|\ell_{t,I_t}| \le 1$:

$$|\eta\hat{\ell}_{t,i}| \le \eta\,|x_i^\top A(q_t)^{-1}x_{I_t}| \;\le\; \eta\,\|x_i\|_{A(q_t)^{-1}}\,\|x_{I_t}\|_{A(q_t)^{-1}} \;\le\; \tfrac{\eta}{\gamma}\max_j \|x_j\|^2_{A(\lambda)^{-1}}$$

We choose λ: make the right side small, i.e. the G-optimal design, and Kiefer–Wolfowitz says $\max_j \|x_j\|^2_{A(\lambda^\star)^{-1}} = d$. With $\gamma = \eta d$, $|\eta\hat{\ell}_{t,i}| \le 1$.

Second moment. With $\ell^2 \le 1$ and $\mathbb{E}[x_{I_t}x_{I_t}^\top \mid \mathcal{F}_{t-1}] = A(q_t) \succeq (1-\gamma)A(p_t)$:

$$\begin{aligned}
\mathbb{E}\Big[\sum_j p_{t,j}\hat{\ell}^2_{t,j}\,\Big|\,\mathcal{F}_{t-1}\Big] &\le \sum_j p_{t,j}\,x_j^\top A(q_t)^{-1}\,\mathbb{E}\big[x_{I_t}x_{I_t}^\top \mid \mathcal{F}_{t-1}\big]A(q_t)^{-1}x_j \\
&= \sum_j p_{t,j}\,\|x_j\|^2_{A(q_t)^{-1}} \\
&\le \frac{1}{1-\gamma}\,\mathrm{tr}\Big(A(p_t)^{-1}\sum_j p_{t,j}x_jx_j^\top\Big) = \frac{d}{1-\gamma}.
\end{aligned}$$

The trace identity again: the p_t-average leverage is exactly d, whatever p_t is.

EXP3 for adversarial linear bandits: $\sqrt{12\,dT\log n}$

Corollary (notes Cor. 4)

EXP3(γ) with $\hat{\ell}_{t,i} = x_i^\top A(q_t)^{-1} x_{I_t} \ell_{t,I_t}$, λ the G-optimal design for $\mathcal{X}$, $\eta = \sqrt{\log(n)/(3dT)}$, $\gamma = \eta d$:

$$\max_i \; \mathbb{E}\Big[\sum_{t=1}^{T} \langle x_{I_t} - x_i, \theta_t \rangle\Big] \;\le\; \sqrt{12\,dT\log n}.$$

EXP3 for adversarial linear bandits: $\sqrt{12\,dT\log n}$

Corollary (notes Cor. 4)

EXP3(γ) with $\hat{\ell}_{t,i} = x_i^\top A(q_t)^{-1} x_{I_t} \ell_{t,I_t}$, λ the G-optimal design for $\mathcal{X}$, $\eta = \sqrt{\log(n)/(3dT)}$, $\gamma = \eta d$:

$$\max_i \; \mathbb{E}\Big[\sum_{t=1}^{T} \langle x_{I_t} - x_i, \theta_t \rangle\Big] \;\le\; \sqrt{12\,dT\log n}.$$

Proof. Plug the two bounds into Theorem 12:

$$\text{regret} \le 2\gamma T + \frac{\log n}{\eta} + (1-\gamma)\eta \cdot \frac{dT}{1-\gamma} =$$

EXP3 for adversarial linear bandits: $\sqrt{12\,dT\log n}$

Corollary (notes Cor. 4)

EXP3(γ) with $\hat{\ell}_{t,i} = x_i^\top A(q_t)^{-1} x_{I_t} \ell_{t,I_t}$, λ the G-optimal design for $\mathcal{X}$, $\eta = \sqrt{\log(n)/(3dT)}$, $\gamma = \eta d$:

$$\max_i \; \mathbb{E}\Big[\sum_{t=1}^{T} \langle x_{I_t} - x_i, \theta_t \rangle\Big] \;\le\; \sqrt{12\,dT\log n}.$$

Proof. Plug the two bounds into Theorem 12:

$$\text{regret} \le 2\gamma T + \frac{\log n}{\eta} + (1-\gamma)\eta \cdot \frac{dT}{1-\gamma} = 3\eta dT + \frac{\log n}{\eta},$$

balanced by $\eta = \sqrt{\log(n)/(3dT)}$. ■

EXP3 for adversarial linear bandits: $\sqrt{12\,dT\log n}$

Corollary (notes Cor. 4)

EXP3(γ) with $\hat{\ell}_{t,i} = x_i^\top A(q_t)^{-1} x_{I_t} \ell_{t,I_t}$, λ the G-optimal design for $\mathcal{X}$, $\eta = \sqrt{\log(n)/(3dT)}$, $\gamma = \eta d$:

$$\max_i \; \mathbb{E}\Big[\sum_{t=1}^{T} \langle x_{I_t} - x_i, \theta_t \rangle\Big] \;\le\; \sqrt{12\,dT\log n}.$$

Proof. Plug the two bounds into Theorem 12:

$$\text{regret} \le 2\gamma T + \frac{\log n}{\eta} + (1-\gamma)\eta \cdot \frac{dT}{1-\gamma} = 3\eta dT + \frac{\log n}{\eta},$$

balanced by $\eta = \sqrt{\log(n)/(3dT)}$. ■

- d **replaces** n **again;** the arm count survives only as $\log n$, as in the stochastic $\sqrt{dT\log(nT)}$. With $\mathcal{X} = \{e_i\}$ this is Corollary 3 verbatim.
- **No stochastic assumption anywhere,** yet the same structure is exploited: a design spreads a small exploration budget so every direction has leverage at most d.

Linear bandits: the scoreboard

Finite arm set $\mathcal{X} \subset \mathbb{R}^d$, $|\mathcal{X}| = n$, horizon T

setting	algorithm	regret
stochastic	elimination with designs	$\sqrt{dT \log(nT)}$
stochastic, any $\mathcal{X}_t$	LinUCB / linear Thompson	$d\sqrt{T}\log T$ / $d^{3/2}\sqrt{T}$
adversarial	EXP3 with a G-optimal design	$\sqrt{12\, dT \log n}$
lower bound	any algorithm, $\mathcal{X} = \{e_i\}$	$\frac{1}{27}\sqrt{(d-1)T}$

- For a finite arm set, adversarial and stochastic cost the same up to logs, both within logs of the lower bound. Optimism pays an extra $\sqrt{d}$ for arbitrary, changing arm sets.
- Two confidence tools for the stochastic side, both now proved: fixed-design intervals plus a union bound (elimination), and the martingale ellipsoid (LinUCB).
- On the adversarial side the design caps the variance of the one-sample estimator *and* is the exploration distribution: Kiefer–Wolfowitz, a third time.