Prior Diffusiveness and Regret in the Linear-Gaussian Bandit

ICML 2026 | 5-minute recording

Yifan Zhu John C. Duchi Benjamin Van Roy

Stanford University

ICML 2026

Prior Diffusiveness in Linear-Gaussian Bandits

Setting

$$R_{t+1} = \theta_\star^\top A_t + W_{t+1}, \qquad W_{t+1} \sim \mathcal{N}(0, \sigma^2)$$

$$A_t \in \mathcal{A} \subset rB_2^d, \qquad \theta_\star \sim \mathcal{N}(0, \Sigma_0).$$

- d: dimension; T: horizon.
- r: action radius; σ^2: noise variance.
- $\sqrt{\mathrm{Tr}(\Sigma_0)}$: prior scale.

Question

If the prior is diffuse, does Thompson sampling pay

a worse long-run rate?

or only

an initial burn-in cost?

Answer

Diffuse priors hurt only through additive burn-in.

Existing Picture: Scale Multiplies the Rate

High-level limitation of prior analyses

In many existing analyses, the parameter scale, reward scale, or prior scale multiplies the rest of the regret bound.

Example: Kalkanli–Özgür (2020)

Suppressing logarithmic factors,

$$\mathrm{Reg}(T) \lesssim d\sqrt{T\left(\sigma^2 + r^2\,\mathrm{Tr}(\Sigma_0)\right)}.$$

The prior scale $\sqrt{\mathrm{Tr}(\Sigma_0)}$ multiplies the $d\sqrt{T}$ term.

Suggested picture

$$\text{regret} \approx \text{scale} \times \text{long-run learning cost}$$

Our point

This picture is too pessimistic for the linear-Gaussian bandit.

Main Result: Additive Decoupling

Bayesian regret of Thompson sampling

$$\boxed{\mathrm{Reg}(T) = \widetilde{O}\left(\sigma d\sqrt{T} + dr\sqrt{\mathrm{Tr}(\Sigma_0)}\right).}$$

Long-run noisy learning

$$\sigma d\sqrt{T}$$

depends on dimension, horizon, and observation noise; not on prior scale.

Prior-localization burn-in

$$dr\sqrt{\mathrm{Tr}(\Sigma_0)}$$

depends on action radius and prior diffusiveness; not polynomially on T.

Conceptual difference

Prior diffusiveness hurts, but it does *not* multiply the long-run $\sqrt{T}$ term.

Why Standard Potential Tools Miss the Separation

Classical elliptical potential lemma

If $V_t = \lambda I + \sum_{s<t} u_s u_s^\top$, $\lambda \geq 1$, and $\|u_t\|_2 \leq 1$, then

$$\sum_{t<T} \|u_t\|^2_{V_t^{-1}} \leq 2 \log \frac{\det V_T}{\det V_0}.$$

Abeille–Lazaric (2017), Proposition 2.

Used to control

$$\sum_{t<T} \|A_t\|_{V_t^{-1}},$$

cumulative posterior uncertainty along chosen actions.

But diffuse priors mean small precision

Posterior precision is

$$V_t = \Sigma_0^{-1} + \sigma^{-2} \sum_{s<t} A_s A_s^\top.$$

Hence

$$\Sigma_0 \text{ large} \Longrightarrow V_0 = \Sigma_0^{-1} \text{ small}.$$

Need

A lemma that separates:

- long-run information accumulation;
- initial small-precision cost.

Key Tool: Generalized Elliptical Potential Lemma

Simplified lemma

If $V_{t+1} = V_t + u_t u_t^\top$, $\|u_t\|_2 \le 1$, $V_0 \succ 0$, and $p \in [0,1]$, then

$$\sum_{t<T} \|u_t\|^{2p}_{V_t^{-1}} \lesssim \underbrace{T^{1-p}\left(\log \frac{\det V_T}{\det V_0}\right)^p}_{\text{volume growth}} + \underbrace{\mathrm{Tr}(V_0^{-p}) - \mathrm{Tr}(V_T^{-p})}_{\text{small initial precision}}.$$

Log-det term

Gives the long-run contribution:

$$\sigma d \sqrt{T}.$$

Trace term

Gives the prior burn-in contribution:

$$dr\sqrt{\mathrm{Tr}(\Sigma_0)}.$$

Lower Bound: The Burn-in Term Is Real

Theorem 6: prior-sensitive cost

If $\tau_1^2 \geq \cdots \geq \tau_d^2$ are eigenvalues of Σ_0, then any policy suffers

$$\mathrm{Reg}_{\mathfrak{p}}(T) \gtrsim \frac{r}{\|\tau\|_2} \sum_{i=2}^{d} ((i-1) \wedge T)\tau_i^2.$$

Example: isotropic prior

If $\Sigma_0 = S^2 I$, then burn-in scales as

$$Sr\sqrt{d}\,\min\{T, d\}.$$

Theorem 6: noise-limited cost

For any policy, there is also a noise-limited lower bound

$$\mathrm{Reg}_{\mathfrak{p}}(T) \gtrsim \sigma d\sqrt{T} - \text{prior correction}.$$

Takeaway

The upper and lower bounds identify the same two regimes:

$$\text{prior burn-in} \quad + \quad \text{noise-limited long run}.$$

Diffuse priors hurt only through additive burn-in.

Thank you!

Questions?

Prior Diffusiveness and Regret in the Linear-Gaussian Bandit