A Novel Self-Normalized Bernstein-Like Dimension-Free Inequality and Regret Bounds for Generalized Kernelized Bandits

Alberto Maria Metelli, Simone Drago, and Marco Mussi
{albertomaria.metelli, simone.drago, marco.mussi}@polimi.it

Motivation

Linear Bandits (LBs, Abbasi-Yadkori et al., 2011): expected reward linear in an unknown d-dimensional parameter vector $\boldsymbol{\theta}^*$:

$$\mathbb{E}[y_t|\mathbf{x}_t;\boldsymbol{\theta}^*] = f^*(\mathbf{x}_t) = \langle \mathbf{x}_t, \boldsymbol{\theta}^* \rangle$$

Generalized Linear Bandits (GLBs, Filippi et al., 2010): apply a known *inverse link function* μ:

$$\mathbb{E}[y_t|\mathbf{x}_t;\boldsymbol{\theta}^*] = \mu(f^*(\mathbf{x}_t)) = \mu(\langle \mathbf{x}_t, \boldsymbol{\theta}^* \rangle)$$

✓ Different noise models (exponential family)
✗ Finite-dimensional $d < \infty$ **only**

Kernelized Bandits (KBs, Chowdhury and Gopalan, 2017): unknown function f^* in *reproducing kernel Hilbert space* (RKHS) with known kernel k:

$$\mathbb{E}[y_t|\mathbf{x}_t;\boldsymbol{\theta}^*] = f^*(\mathbf{x}_t), \quad f^*(\mathbf{x}) = \langle f^*, k(\mathbf{x},\cdot) \rangle$$

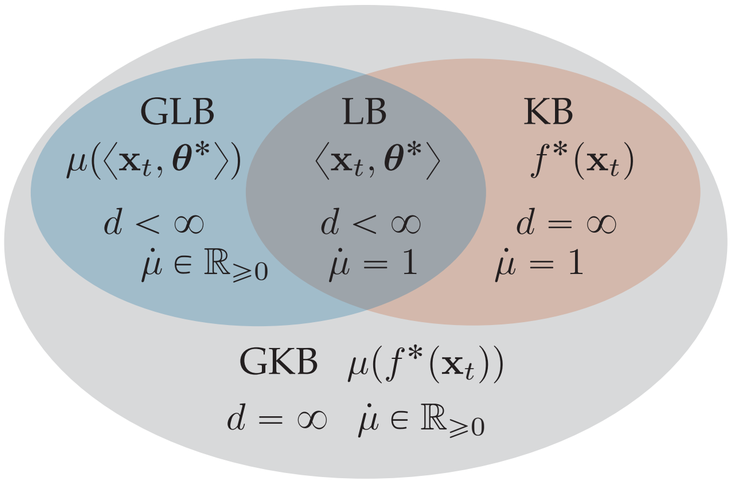
✗ Subgaussian noise model **only**
✓ Infinite-dimensional $d = \infty$

Setting

Generalized Kernelized Bandits (GKBs, **ours**): unknown function f^* in *RKHS* with known *kernel* k + apply a known *inverse link function* μ:

$$\mathbb{E}[y_t|\mathbf{x}_t;\boldsymbol{\theta}^*] = \mu(f^*(\mathbf{x}_t)), \quad f^*(\mathbf{x}) = \langle f^*, k(\mathbf{x},\cdot) \rangle$$

GKB := **GLB** + **KB**



Preliminaries

Canonical Exponential Family

$$p(y|\mathbf{x};f) = \exp\left(\frac{yf(\mathbf{x}) - m(f(\mathbf{x}))}{\mathfrak{g}(\tau)} + h(y,\tau)\right)$$

- m is 3 times differentiable and convex
- **Inverse link function**: $\mu = m'$

$$\mathbb{E}[y|\mathbf{x};f] = \mu(f(\mathbf{x})), \quad \mathbb{V}\mathrm{ar}[y|\mathbf{x};f] = \frac{\dot{\mu}(f(\mathbf{x}))}{\mathfrak{g}(\tau)}$$

RKHS and Information Gain

- Kernel k induces a countable-dimensional feature mapping ϕ s.t. $f(\mathbf{x}) = \langle \alpha, \phi(\mathbf{x}) \rangle$
- **Weighted kernel and feature**:

$$\widetilde{\phi}(\mathbf{x};f) = \sqrt{\mathfrak{g}(\tau)^{-1}\dot{\mu}(f(\mathbf{x}))}\phi(\mathbf{x})$$

$$\widetilde{k}(\mathbf{x},\mathbf{x}';f) = \mathfrak{g}(\tau)^{-1}\sqrt{\dot{\mu}(f(\mathbf{x}))}k(\mathbf{x},\mathbf{x}')\sqrt{\dot{\mu}(f(\mathbf{x}'))}$$

- **Weighted info gain**: $\frac{1}{2}\log\det(\lambda^{-1}\widetilde{\mathbf{K}}_t(\lambda;f))$

$$\widetilde{\mathbf{K}}_t(\lambda;f) = \lambda\mathbf{I} + \left(\widetilde{k}(\mathbf{x}_i,\mathbf{x}_j;f)\right)_{i,j\in[\![t-1]\!]}$$

- **Weighted covariance operator**:

$$\widetilde{V}_t(\lambda;f) = \sum_{s=1}^{t-1}\widetilde{\phi}(\mathbf{x}_s;f)\widetilde{\phi}(\mathbf{x}_s;f)^\top + \lambda I$$

$$\det(\lambda^{-1}\widetilde{V}_t(\lambda;f)) = \det(\lambda^{-1}\widetilde{\mathbf{K}}_t(\lambda;f))$$

Assumptions

Assumptions on RKHS

f^* belongs to *RKHS* $\mathcal{H}$ with *known* kernel k:
- Asm 3.1 (**Bounded norm**): $\|f^*\| \leqslant B$
- Asm 3.2 (**Bounded kernel**): $\sup_{\mathbf{x}\in\mathcal{X}} k(\mathbf{x},\mathbf{x}) \leqslant K$

Assumptions on the Reward

Reward y_t sampled as $y_t \sim p(\cdot|\mathbf{x};f)$:
- Asm 3.3 (**Bounded noise**): $|\epsilon_t| := |y_t - \mu(f^*(\mathbf{x}))| \leqslant R$
- Asm 3.4 (**Self-concordance,** Russac et al., 2021): $|\ddot{\mu}(f(\mathbf{x}))| \leqslant R_s\dot{\mu}(f(x))$

Self-Normalized Bernstein-Like Dimension-Free Concentration Inequality

We need a concentration bound:
- **Bernstein-like**: handle effectively the **heteroskedasticity** of the noise ϵ_t
- **Dimension-free**: no dependence on **dimensionality** of features ϕ (can be infinite)
- **Empirical**: dependence on the **weighted covariance operator** $\widetilde{V}_t(\lambda;f^*)$ (is random)

Self-normalized Concentrations	Condition	Properties: Dim-free	Empirical	Heterosc.	Technique
Dani et al. (2008)	Hoeffding	✗	✗	✗	Freedman
Abbasi-Yadkori et al. (2011)	Hoeffding	✓	✓	✗	Pseudo-Max
Chowdhury and Gopalan (2017)	Hoeffding	✓	✓	✗	Pseudo-Max
Faury et al. (2020)	Bernstein	✗	✓	✓	Pseudo-Max
Zhou et al. (2021)	Bernstein	✗	✗	✗	Freedman
Ziemann (2025)	Bernstein	✗	✓	✗	PAC-Bayes
Our work	**Bernstein**	✓	✓	✓	**Freedman**

$$\forall t \geqslant 1: \ \|S_t\|_{\widetilde{V}_t^{-1}(\lambda)} \leqslant \widehat{B}_t(\delta;f) := \left(\sqrt{73\log\det(\lambda^{-1}\widetilde{V}_t(\lambda)) + \sqrt{3}}\right)\sqrt{\log\frac{\pi^2(\rho+1)^2}{3\delta} + \frac{3RK}{\sqrt{\lambda}}\log\frac{\pi^2(\rho+1)^2}{3\delta}} \ \text{w.p.}\ 1-\delta$$

where: $S_t := \sum_{s=1}^{t-1}\epsilon_s\phi(\mathbf{x}_s)$, $\widetilde{V}_t(\lambda) := \sum_{s=1}^{t-1}\sigma_s^2\phi(\mathbf{x}_s)\phi(\mathbf{x}_s)^\top + \lambda I$, $\rho = \max\left\{0, \left\lceil \log\left(\frac{8R^2K^2(t-1)^3}{\lambda}\log\left(1+\frac{K^2R^2}{\lambda}\right)\right)\right\rceil\right\}$

Algorithm: GKB-UCB

Algorithm

For every round $t \in [\![T]\!]$:

1. **Maximum Likelihood Estimate**:

$$\widehat{f}_t \in \arg\min_{f\in\mathcal{H}} \mathcal{L}_t(f) := \sum_{s=1}^{t-1}\frac{-y_sf(\mathbf{x}_s) + m(f(\mathbf{x}_s))}{\mathfrak{g}(\tau)} + \frac{\lambda}{2}\|f\|^2$$

2. **Optimistic Decision Selection**: $(\widetilde{f}_t, \mathbf{x}_t) \in \arg\max_{f\in\mathcal{C}_t(\delta),\,\mathbf{x}\in\mathcal{X}} \mu(f(\mathbf{x}))$

where: $g_t(f) := \sum_{s=1}^{t-1}\frac{\mu(f(\mathbf{x}_s))}{\mathfrak{g}(\tau)}\phi(\mathbf{x}_s) + \lambda\alpha$

$$\mathcal{C}_t(\delta) = \left\{f\in\mathcal{H} : \left\|g_t(f) - g_t(\widehat{f}_t)\right\|_{\widetilde{V}_t^{-1}(\lambda;f)} \leqslant \sqrt{\lambda}B + \mathfrak{g}(\tau)^{-1}\widehat{B}_t(\delta;f)\right\}$$

Regret Analysis

If $\lambda \geqslant \Omega(K^2)$, w.p. $1-\delta$: $R(\texttt{GKB-UCB}, T) := \sum_{t\in[\![T]\!]}\left(\mu(f^*(\mathbf{x}^*)) - \mu(f^*(\mathbf{x}_t))\right) \leqslant$

$$\widetilde{O}\left((1+R_sBK)\left(\sqrt{\lambda}B + \sqrt{\widetilde{\gamma}_T(\mathcal{H})\log\frac{1}{\delta}}\right) + RK\log\frac{1}{\delta}\right)\sqrt{\widetilde{\gamma}_T(f^*)}\sqrt{\frac{T}{\kappa_*}}\right)$$

Maximal weighted information gains:

$$\widetilde{\gamma}_t(f) := \max_{\mathbf{x}_1,\dots,\mathbf{x}_{t-1}\in\mathcal{X}}\frac{1}{2}\log\det(\lambda^{-1}\widetilde{\mathbf{K}}_t(\lambda;f))$$

$$\widetilde{\gamma}_t(\mathcal{H}) := \sup_{f\in\mathcal{H}}\widetilde{\gamma}_t(f)$$

Complexity index:

$$\kappa_* := \frac{1}{\dot{\mu}(f^*(\mathbf{x}^*))}$$

References

Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, 2011.
S. R. Chowdhury and A. Gopalan. On kernelized multi-armed bandits. In *International Conference on Machine Learning*, 2017.
V. Dani, T. P. Hayes, and S. M. Kakade. Stochastic linear optimization under bandit feedback. In *Conference on Learning Theory*, 2008.
L. Faury, M. Abeille, C. Calauzènes, and O. Fercoq. Improved optimistic algorithms for logistic bandits. In *International Conference on Machine Learning*, 2020.
S. Filippi, O. Cappé, A. Garivier, and C. Szepesvári. Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems*, 2010.
Y. Russac, L. Faury, O. Cappé, and A. Garivier. Self-concordant analysis of generalized linear bandits with forgetting. In *International Conference on Artificial Intelligence and Statistics*, 2021.
D. Zhou, Q. Gu, and C. Szepesvári. Nearly minimax optimal reinforcement learning for linear mixture markov decision processes. In *Conference on Learning Theory*, 2021.
I. Ziemann. A vector bernstein inequality for self-normalized martingales. *Transactions on Machine Learning Research*, 2025.

Paper

Take a look at our work!