Subspace Learning with Gaussian Processes for Sparse Contextual Bandits

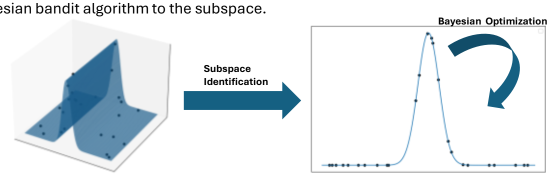
Author: Yair Chizi Y.Chizi@student.tudelft.nl
Supervisor: Julia Olkhovskaia

Thursday, June 26, 2025

1 Introduction

The Contextual Bandit Problem:
A problem in sequential learning, wherein a bandit algorithm must select an arm, or action, to obtain a reward with limited initial knowledge of its structure. In contextual bandits, the bandit algorithm is also presented with contexts for every arm that may influence the reward. The performance of bandit algorithms is measured through (cumulative) regret: the difference between the expected reward of the best constant-action algorithm and the selected bandit algorithm.

Sparsity and the SI-BO Algorithm:
In practical domains, the context vectors often lie in a high-dimensional ambient space, while the reward function often lies in a low-dimensional space. The SI-BO bandit algorithm [1] capitalizes on this distinction by employing a two-phase framework: first identifying the underlying subspace first and then applying a Bayesian bandit algorithm to the subspace.



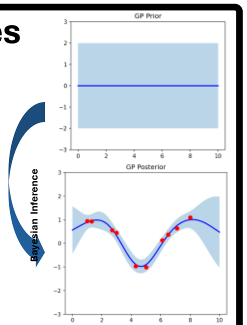
Research Sub-Questions:
- How can the framework of SI-BO be extended to create a novel algorithm SI-BKB?
- Does SI-BKB offer comparable empirical performance to SI-BO?
- What trade-offs emerge between the initial subspace identification cost and the long-term optimization benefits?

2 Gaussian Processes

UCB Algorithms:
select the arm with the largest overestimate of the reward, using uncertainty.

Overestimate computation?
Use the posterior mean and variance of the Bayesian predictive distribution using GP regression [2].

Computational Inefficiency: instead of using all past observations, select a subset of "inducing points" (the BKB bandit algorithm) [3].



4 Empirical Results

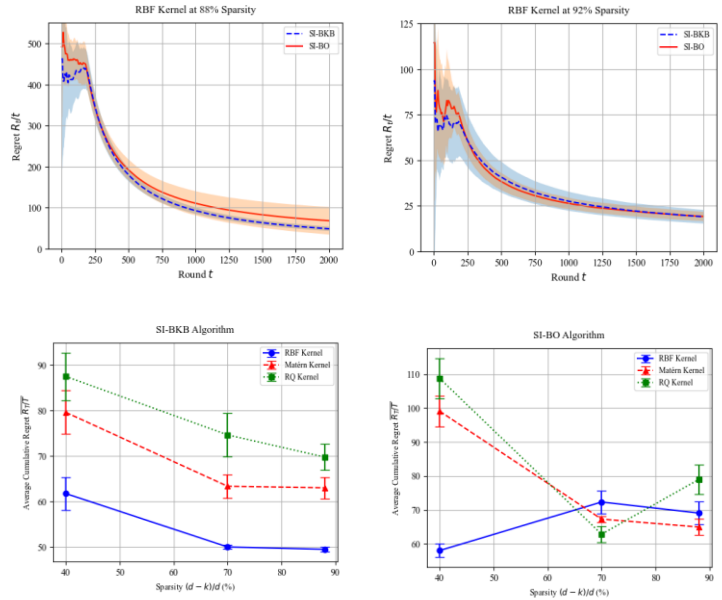
Synthetic Data Generation:
Every round, contexts are generated for each arm from a Gaussian distribution: the mean is uniformly sampled from the context space before the simulation. Additionally, before the simulation, every arm is queried 10,000 to assign the arm with the best cumulative reward as the best constant-action algorithm for the regret computation.

Experimental Conditions:
- Bandit Environments: linear, Branin (non-linear)
- Isotropic Kernels: RBF, Matern, Rational Quadratic

SI-BKB Behaves Similarly to SI-BO:
The erratic plateau of the subspace identification phase and the decay of the Bayesian optimization phase are clearly identifiable. The average regret of the SI-BKB lies within 1σ of the SI-BO algorithm. The consistency in behavior persists across bandit environments.

Violation of No-Regret:
None of the simulated conditions exhibited a regret trend that approach zero regret.

Regret Does Not Deteriorate under High Sparsity:
In some cases, the regret in the final 100 rounds improves under high sparsity. Kernels that are less suitable for the estimation of infinitely smooth functions (Matern, RQ) show substantial improvement.



3 Theoretical Analysis (SI-BKB)

SI-BKB Upper Regret Bounds
are sub-exponential with respect to the ambient dimension. Bounds are comparable to SI-BO.

Sub-exponential Bounds:
The BKB regret bound is on the true subspace and, thus, does not depend on the ambient dimension.

Misalignment Costs
are dominated by the regret bounds of the BKB on the true reward function on the horizon: $R_{\mathrm{BKB}} = \mathcal{O}(\sqrt{T}\log T)$

Brief Summary of Proof

Duration of Subspace Learning; Approx. Error equal to $\sqrt{T}$; Regret bound of BKB on the learned subspace

$$R_T \leq \underbrace{n}_{} + \underbrace{\eta T + R_{\mathrm{BKB}}(T, \hat{g}, \kappa)}_{}$$

$$\leq \mathcal{O}(k^3 d^2 \log^2(1/\delta)) + \sqrt{2}\, R_{\mathrm{BKB}}(T, g, \kappa)$$

5 Conclusion

1. **The SI-BKB algorithm achieves comparable performance** to SI-BO both theoretically and empirically.
2. **SI-BKB and SI-BO perform well** even **under high sparsity constraints** across all kernels.

Future Improvements:
- extending the subspace learning framework to other non-GP bandit algorithms
- performing subspace learning and Bayesian optimization simultaneously.

References
[1] Josip Djolonga, Andreas Krause, and Volkan Cevher. High-Dimensional Gaussian Process Bandits. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.
[2] Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias W. Seeger. Information-theoretic regret bounds for gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, 58(5):3250–3265, 2012.
[3] Joaquin Quinonero-Candela, Carl Edward Rasmussen, and Christopher K. I. Williams. Approximation Meth ods for Gaussian Process Regression. In Leon Bottou, Olivier Chapelle, Dennis DeCoste, and Jason Weston, editors, *Large-Scale Kernel Machines*, pages 203–224. The MIT Press, August 2007.