Regret Bounds for Safe Gaussian Process Bandit Optimization

Sanae Amani SAMANIGESHNIGANI@UCSB.EDU

Mahnoosh Alizadeh ALIZADEH@UCSB.EDU

Christos Thrampoulidis CTHRAMPO@UCSB.EDU

Department of Electrical and Computer Engineering, University of California, Santa Barbara

Editors: A. Bayen, A. Jadbabaie, G. J. Pappas, P. Parrilo, B. Recht, C. Tomlin, M.Zeilinger

Abstract

Stochastic bandit optimization has received significant attention in applications where a learner must repeatedly deal with an unknown random environment and observations are costly to obtain. The goal is to minimize the so-called cumulative pseudo-regret, i.e., the difference between the expected T-period reward generated by the algorithm and the optimal expected reward if the reward function was known to the learner. Many applications require a learner to make sequential decisions given uncertainty regarding both the system's reward function and certain safety constraints. In such safety-critical systems, it is paramount that the learner's actions do not violate the safety constraints at any stage of the learning process. In this paper, we study a stochastic bandit optimization problem where the unknown reward f and constraint function g are modeled via Gaussian Processes (GPs) with known kernels and the learner has access to both f and g through bandit feedback measurments. For this problem, we develop a safe variant of GP-UCB Srinivas et al. (2010) called SGP-UCB, with necessary modifications to respect safety constraints at every round. In fact, our algorithm can be seen as an extension of Safe-LUCB proposed by Amani et al. (2019) to safe GPs. Specifically, we show that our algorithm and guarantees are similar to those in Amani et al. (2019) for linear kernels. The algorithm proceeds in two phases to balance the goal of expanding the safe set and controlling the regret. Prior to designing the decision rule, the algorithm requires a proper exploration of $\mathcal{D}_0^{\text{S}}$, the unknown safe part of the given safe set $\mathcal{D}_0$. Hence, in the first phase, it takes actions at random from a given safe seed set $\mathcal{D}^w$ until the approximated safe set has sufficiently expanded. In the second phase, the algorithm exploits GP properties to construct confidence intervals $Q_{f,t}(\mathbf{x})$ and $Q_{g,t}(\mathbf{x})$ such that $f(\mathbf{x}) \in Q_{f,t}(\mathbf{x})$ and $g(\mathbf{x}) \in Q_{g,t}(\mathbf{x})$ with high probability. It applies $Q_{g,t}(\mathbf{x})$ to design an inner approximation $\mathcal{D}_t^{\text{S}}$ of the safe set $\mathcal{D}_0^{\text{S}}$. The chosen actions belong to $\mathcal{D}_t^{\text{S}}$ which guarantees that the safety constraints are met with high probability. We balanced the two-fold challenge of minimizing regret and expanding safe set by properly choosing the duration of the first phase T'. Our analysis suggests that the type of kernels associated with the constraint function plays a critical role in tuning the T' as well as size of $\mathcal{D}^w$, and consequently affects the regret bounds. We used *Random Fourier Feature* (RFF), a uniform kernel approximation, as a tool to facilitate our analysis when the constraint functions are associated with infinite-dimensional RKHS and derive the first sub-linear regret bounds for finite and infinite dimensional RKHS. Specifically, for a general reward function and simple constraint with linear kernel, we prove a regret bound of the same order as the standard GP-UCB, provided that the safe seed set is of size at least d. However, for more complex constraint functions, our analysis and numerical experiments suggest that guaranteeing a good regret over the entire safe set might require a very large safe seed set. We evaluate the performance of our algorithm with numerical simulations and compare it to that of existing algorithms in the literature Sui et al. (2015, 2018).

References

Sanae Amani, Mahnoosh Alizadeh, and Christos Thrampoulidis. Linear stochastic bandits under safety constraints. In *Advances in Neural Information Processing Systems*, pages 9252–9262, 2019.

Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: no regret and experimental design. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, pages 1015–1022. Omnipress, 2010.

Yanan Sui, Alkis Gotovos, Joel W. Burdick, and Andreas Krause. Safe exploration for optimization with gaussian processes. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 997–1005. JMLR.org, 2015. URL http://dl.acm.org/citation.cfm?id=3045118.3045225.

Yanan Sui, Joel Burdick, Yisong Yue, et al. Stagewise safe bayesian optimization with gaussian processes. In *International Conference on Machine Learning*, pages 4788–4796, 2018.