Cost-Aware Bayesian Optimization via Information Directed Sampling

Biswajit Paria bparia@cs.cmu.edu
Willie Neiswanger wdn@cs.cmu.edu
Ramina Ghods rghods@cs.cmu.edu
Jeff Schneider jeff4@cs.cmu.edu
Barnabás Póczos bapoczos@cs.cmu.edu

School of Computer Science
Carnegie Mellon University, USA

Abstract

In this paper we propose an efficient Bayesian Optimization (BO) algorithm for expensive black-box optimization based on Information Direction Sampling (Russo and Van Roy, 2014a). We consider the setting where evaluations have varying but known costs. Our proposed approach, known as CostIDS, is cost aware and balances evaluation cost vs information gain, leading to a cost efficient algorithm. In contrast to other competing approaches, our approach does not require a set predefined budget, does not require expensive entropy computations, enjoys sub-linear regret bounds, and has the potential for much faster regret rates in the presence of informative low cost evaluations. We empirically compare our approach to various other cost aware BO baselines and show improved performance on a synthetic function.

Keywords: Bayesian Optimization, Information Gain, Regret Bounds

1. Introduction

Black box optimization problems appear frequently in a wide variety of disciplines, from designing molecules (Griffiths and Hernández-Lobato, 2017) and materials (Frazier and Wang, 2016) to optimizing hyperparameters for any model Snoek et al. (2012). Black box functions are usually expensive to evaluate and provide only noisy zeroth order function evaluations. As an example, discovering an optimal proportion of solvents to design an efficient battery requires extensive experimentation. Each experiment is usually noisy, and expensive in terms of time and resources. Thus one must follow an *exploration-exploitation* approach to identify the best proportion in a small number of experiments. Bayesian optimization is a popular approach for such noisy black box optimization problems. Typically, BO approaches focus on achieving a small regret in a small number of experiments, while ignoring the cost of each individual experiment. In practice however, each experiment may utilize a different amount of resources. Furthermore, often cheap (or low fidelity) experiments are available that provide a significant amount of information about the location of the optimum. Multi-Fidelity (Forrester et al., 2007; Kandasamy et al., 2017) and cost-aware (Lee et al., 2020) BO approaches address this by aiming to minimize the total experimental cost rather than the number of experiments. The key idea behind such approaches is to utilize low cost experiments for efficient *exploration* of the search space. In this paper, we paper we propose a method for cost-aware BO using information directed sampling (IDS) (Russo and Van Roy, 2014a).

There are various notions of regret in the literature (Bubeck and Cesa-Bianchi, 2012). In this paper we aim to minimize the *simple regret* defined as the smallest regret among the individual regrets of all the sampled points. This notion of simple regret however does fit well in the multi-fidelity setting. In the multi-fidelity we are often provided with low fidelity *approximations* of the true function. However, when measuring the simple regret, one must ignore the low fidelity evaluations, as they provide an inaccurate value of the function at that point, and hence cannot be used to make any practical decisions. Such a formulation for the simple regret in the multi-fidelity setting is also followed by Song et al. (2018). The main utility of the low-fidelity evaluations is exploration rather than exploitation.

The usual notion of simple regret holds in the cost-aware setting since there is no concept of fidelity and approximate evaluations. Hence all evaluations can be considered in the simple regret. This does not eliminate the need for cheap evaluations as they may still be used to gain information about the unknown function. In this paper, we focus on the cost-aware setting with the goal of minimizing the simple regret given some cost budget.

As a concrete example, consider the battery optimization problem described earlier. One might have access to a simulator providing approximate evaluations. This is a multi-fidelity setting, as the approximate results from the simulator cannot be relied on for making decisions. One has to perform the highest fidelity evaluations eventually, in order to verify the best found points. On the other hand, consider the setting where each experiment has a different cost – the cost-aware setting. This is a realistic setting since different configurations can utilize different amount of resources (battery chemicals in this example). All evaluations being exact in this scenario, they can be considered as reliable estimates of function values and hence included in the regret. While our proposed method can also be extended to the multi-fidelity setting, we focus on the cost-aware setting for simplicity of analysis.

Balancing *exploration* and *exploitation* is the main idea behind BO and bandit algorithms. In our setting, the cost must be factored in as well. At a high level, at each step, the algorithm should incur a small regret (exploitation), gain information (exploration) while using less resources (cost). Information Directed Sampling (IDS) (Russo and Van Roy, 2014a) provides a principled strategy for balancing regret and information gain. In this work, we follow a similar approach and propose CostIDS, a cost-aware acquisition function that balances cost, regret, and information gain.

Our Contributions. In this paper we propose a principled cost-aware acquisition function, which balances exploration, exploitation, and experimental cost. We show sub-linear regret bounds on the cumulative regret, thus resulting in zero simple regret as the budget is increased (no-regret property). Finally, we perform some preliminary experiments on a synthetic function and show promising results.

Advantages. Compared to prior work, CostIDS enjoys a number of advantages. Our approach is a theoretically motivated no-regret algorithm, while being much simpler conceptually. Furthermore, our approach also does not require expensive entropy computations as PES (Hernández-Lobato et al., 2014), MES (Wang and Jegelka, 2017), and Multi-Fidelity

MES (Takeno et al., 2019). Consequently, CostIDS is applicable to models beyond Gaussian processes where such entropy computations are prohibitive. Additionally, CostIDS does not require a pre-defined budget as assumed by Song et al. (2018).

2. Related Work

Bayesian optimization of black-box functions is a well explored topic. A number of acquisition functions have been proposed in the literature including GP-UCB (Srinivas et al., 2009), Thompson sampling (Russo and Van Roy, 2014b), expected improvement (EI) (Jones et al., 1998) and entropy based methods (Hernández-Lobato et al., 2014; Wang and Jegelka, 2017). Multi-objective BO has also been well explored. We refer to (Paria et al., 2019), for a discussion on multi-objective BO.

Experimental costs have been previously considered in BO in a variety of contexts. Kandasamy et al. (2017); Song et al. (2018) consider the multi-fidelity setting, where approximations of the true function are available as cheap low fidelity evaluations. On the other hand, Lee et al. (2020); Takeno et al. (2019) consider the cost-aware setting.

Information directed sampling (IDS) is a cumulative regret minimization approach that aims to balance regret incurred and information gained. IDS has been shown to perform well in cases where an action can provide information about other actions, also known as the complex information scenario (Russo and Van Roy, 2016). IDS has been used for reinforcement learning (Nikolov et al., 2018), bandits with heteroscedastic noise (Kirschner and Krause, 2018), and linear partial monitoring (Kirschner et al., 2020).

3. Background

Gaussian Processes. Gaussian processes (GP) (Rasmussen and Williams, 2006) are used to model the unknown function f in many BO algorithms. GPs define a prior distribution over functions defined on some input space $\mathcal{X}$. For any function f drawn from a GP, $f \sim \mathcal{GP}(\cdot)$, and some finite set of points $\boldsymbol{x}_i \in \mathcal{X}$ $(1 \leq i \leq n)$, the function values $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)$ follow a multivariate Gaussian distribution. The posterior distribution conditioned on observations can be efficiently computed due to the tractability of the Gaussian distribution. Further details can be found in Appendix A.

Information Directed Sampling. Information Directed Sampling (IDS) is an information theoretic approach for cumulative regret minimization (Russo and Van Roy, 2014a). IDS uses the concept of information ratio which emerged in a related paper (Russo and Van Roy, 2016) on an information theoretic analysis of Thompson sampling. For a finite domain $\mathcal{X}$, a policy π is computed from which the next action $\boldsymbol{x}$ with will be sampled. IDS stipulates choosing the policy that minimizes the information ratio as defined below.

$$\pi_t = \operatorname*{argmin}_{\pi} \underbrace{\frac{\Big(\mathbb{E}_{\boldsymbol{x}\sim\pi}\left[f(\boldsymbol{x}^*) - f(\boldsymbol{x}) \mid \mathcal{D}_{t-1}\right]\Big)^2}{\mathbb{E}_{\boldsymbol{x}\sim\pi}\mathrm{IG}\left(\boldsymbol{x}^*, \boldsymbol{x} \mid \mathcal{D}_{t-1}\right)}}_{\text{Information Ratio}}, \tag{1}$$

where $\boldsymbol{x}^*$ denotes the optimal action, IG denotes the information gain, and $\mathcal{D}_t = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^t$ denotes the set of observations till step t. The next candidate is chosen as $\boldsymbol{x}_t \sim \pi_t$. The optimal policy π_t can potentially be a randomized policy, that is, π_t can be supported on more than one point in $\mathcal{X}$. At a high level, IDS minimizes the regret per information-gain. The key to bounding the regret of IDS is to bound the information ratio for the chosen policy π_t. Denote such an upper bound at step t by Γ_t. The regret at step t is upper bounded an increasing function of Γ_t. Further details and derivations can be found in Russo and Van Roy (2014a).

4. Cost-aware Information Directed Sampling

We define a modified version of the information ratio for Gaussian processes.

$$\boldsymbol{x}_t = \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}} \frac{\mathbb{E}\left[f^* - f(\boldsymbol{x}) \mid \mathcal{D}_{t-1}\right]^2}{\mathrm{IG}\left(f, \boldsymbol{x} \mid \mathcal{D}_{t-1}\right)} \tag{2}$$

This formulation differs from the original in quite a few aspects. The next candidate $\boldsymbol{x}_t$ is no longer randomized. While Russo and Van Roy (2014a) argue for randomized policies as necessary for certain problems, GPs are nice in that randomized policies are not necessary for optimization. GP-UCB (Srinivas et al., 2009) is an example of such a non-randomized policy. The other difference is that information gain on the maximizer $\mathrm{IG}\left(\boldsymbol{x}^*, \boldsymbol{x} \mid \mathcal{D}_{t-1}\right)$ is replaced by an upper bound $\mathrm{IG}\left(f, \boldsymbol{x} \mid \mathcal{D}_{t-1}\right)$, and consequently a smaller information ratio. The resulting regret bound is in terms of the maximum information gain of the function f rather than the maximizer $\boldsymbol{x}^*$.

The information ratio is dependent on the expected maximum of the posterior $\mathbb{E}[f^* \mid \mathcal{D}_{t-1}]$. We bound it using a discretization technique similar to Kandasamy et al. (2018, lemma 12), as summarized in the following lemma.

Proposition 1 *At any iteration t,*

$$\mathbb{E}[f^* \mid \mathcal{D}_{t-1}] \leq \frac{1}{t^2} + \underbrace{\max_{\boldsymbol{x}\in\mathcal{X}} \ \mu_t(\boldsymbol{x}) + \sqrt{\beta}\sigma_t(\boldsymbol{x})}_{\mathcal{U}_t\text{: max UCB at iteration } t}, \tag{3}$$

where $\beta_t = C_1 d \log t + C_2$, where C_1, C_2 are constants depending on the kernel κ of the GP.

Note that the upper bound is approximately the maximum UCB at iteration t, as $1/t^2$ tends to zero as $t \to \infty$. We denote the maximum UCB by $\mathcal{U}_t$ and the maximizer by $\boldsymbol{x}_t^{\mathrm{UCB}}$. The remaining term $\mathbb{E}[f(\boldsymbol{x}) \mid \mathcal{D}_{t-1}]$ is a simply the posterior mean $\mu_t(\boldsymbol{x})$. Finally, for GPs the information gain can be expressed in closed form as $\mathrm{IG}\left(f, \boldsymbol{x} \mid \mathcal{D}_{t-1}\right) = \frac{1}{2}\log(1 + \sigma^{-2}\sigma_t(\boldsymbol{x})^2)$. However, in order to bound the information ratio we use $\sigma_t(\boldsymbol{x})^2$ instead, which is an increasing function of $\mathrm{IG}\left(f, \boldsymbol{x} \mid \mathcal{D}_{t-1}\right)$. As discussed earlier in Section 3, bounding the information ratio is essential to bound the regret.

CostIDS. With all the above substitutions and modifications, the information ratio, and the cost-aware information ratio are defined as,

$$\mathcal{R}_t(\boldsymbol{x}) = \frac{\left(\mathcal{U}_t - \mu_t(\boldsymbol{x})\right)^2}{\sigma_t(\boldsymbol{x})^2}, \qquad \mathcal{R}_t^{\mathrm{cost}}(\boldsymbol{x}) = \lambda(\boldsymbol{x})\frac{\left(\mathcal{U}_t - \mu_t(\boldsymbol{x})\right)^2}{\sigma_t(\boldsymbol{x})^2}. \tag{4}$$

The cost $\lambda(\boldsymbol{x})$ is integrated as a multiplicative factor in $\mathcal{R}_t(\boldsymbol{x})$. The cost integrated acquisition function is then optimized to yield the next candidate. In practice, optimizing this directly can present some degenerate cases where extremely cheap points are chosen repeatedly which provide little information. To avoid such a pathology, we propose the following strategy.

$$\boldsymbol{x}_t = \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}} \mathcal{R}_t^{\text{cost}}(\boldsymbol{x}) \quad \text{s.t.} \quad \mathcal{R}_t(\boldsymbol{x}) \leq \rho \mathcal{R}_t^* \quad \text{where } \mathcal{R}_t^* = \min_{\boldsymbol{x}\in\mathcal{X}} \mathcal{R}_t(\boldsymbol{x}). \tag{5}$$

The constant $\rho > 1$ is a user-defined tolerance factor. The constraint that $\mathcal{R}_t(\boldsymbol{x})$ is not too far from the optimal $\mathcal{R}_t^*$ ensures that there is some progress in each iteration.

5. Regret Bounds

A key step in bounding the regret for IDS based algorithms is bounding the information ratio. We first assume the existence of upper bounds Γ_t and Γ_t^{cost} on $\mathcal{R}_t(\boldsymbol{x}_t)$ and $\mathcal{R}_t^{\text{cost}}(\boldsymbol{x}_t)$ respectively. Thereafter, we define the simple and cumulative regrets both wrt the step t and budget used Λ. Finally, we will also provide exact expressions for the upper bounds on the information ratio.

Assumption 1 *Assume that for all $t > 0$, the information ratios can be upper bounded as*

$$\mathcal{R}_t(\boldsymbol{x}_t) \leq \Gamma_t, \quad \mathcal{R}_t^{\text{cost}}(\boldsymbol{x}_t) \leq \Gamma_t^{\text{cost}} \quad \textit{almost surely}, \tag{6}$$

where Γ_t and Γ_t^{cost} are independent of the observations $\mathcal{D}_{t-1}$ and non-decreasing in t.

Definition 2 (Regrets wrt. to t) *Define the instantaneous, cumulative, and simple regrets respectively, at step t as,*

$$r_t = f^* - f(\boldsymbol{x}_t), \quad R_t = \sum_{i=1}^{t} r_i, \quad s_t = \min_{i=1}^{t} r_t. \tag{7}$$

Next, we will propose a notion of the simple regret wrt. the total budget Λ.

Definition 3 (Regrets wrt. to Λ) *Denote by Λ the total budget used. We define t_Λ as a random variable denoting largest time step such that the budget does exceed Λ.*

$$t_\Lambda = \max \left\{ t \;\middle|\; \sum_{i=1}^{t} \lambda(\boldsymbol{x}_i) \leq \Lambda \right\} \tag{8}$$

Define the simple and (cost-weighted) cumulative regret random variables for budget Λ as,

$$s_\Lambda = \min_{i=1}^{t_\Lambda} r_i, \quad R_\Lambda = \sum_{i=1}^{t_\Lambda} \lambda(\boldsymbol{x}_i) r_i \tag{9}$$

Theorem 4 *The cumulative and simple regrets wrt. the budget is upper bounded as,*

$$\mathbb{E}[R_\Lambda] \leq \sqrt{\Lambda C_1 \mathbb{E}[\Gamma_{t_\Lambda}^{\text{cost}} \gamma_{t_\Lambda}]} + C_2, \quad \mathbb{E}[s_\Lambda] \leq C_3 \sqrt{\mathbb{E}[\Gamma_{t_\Lambda}^{\text{cost}} \gamma_{t_\Lambda}]/\Lambda} + C_4, \tag{10}$$

for all $\Lambda > \lambda_{\max}$, where $C_1 = 2(1+\sigma^{-2})^{-1}$, $C_2 = \lambda_{\max}\pi^2/6$, $C_3 = \sqrt{C_1}(1-\lambda_{\max}/\Lambda)^{-1}$, $C_4 = \pi^2/6(\Lambda/\lambda_{\max} - 1)^{-1}$, and γ_t denotes the maximum information gain.

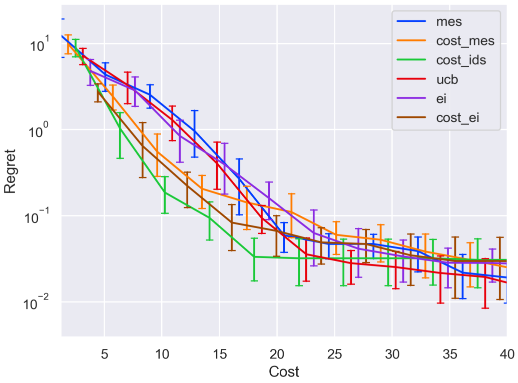

Figure 1: Cost vs. Simple regret plot for the modified Branin function

Next we provide explicit expressions for the information ratio upper bounds, Γ_t^{cost} and Γ_t. Consider the information ratio for $\boldsymbol{x}_t^{\text{UCB}}$, the maximizer of the UCB. This leads to $\mathcal{R}_t^* \leq \mathcal{R}_t(\boldsymbol{x}_t^{\text{UCB}}) = \beta_t$. Therefore, $\Gamma_t = \beta_t$ is a valid upper bound. Consequently we have $\mathcal{R}_t^{\text{cost}}(\boldsymbol{x}_{t+1}) \leq \lambda_{\max}\rho\beta_t$, leading to $\Gamma_t^{\text{cost}} = \lambda_{\max}\rho\beta_t$. Substituting them above, we get a simple regret bound of $\mathbb{E}[s_\Lambda] \leq \mathcal{O}(\sqrt{\mathbb{E}[\beta_{t_\Lambda}\gamma_{t_\Lambda}]\lambda_{\max}/\Lambda})$.

6. Experiments

We perform experiments on the Branin (2-dim) function denoted by $b(x_1, x_2)$. We modify it to simulate a hyper-parameter optimization problem for a ML model by adding a 3rd dimension as $B(x_1, x_2, l)$. The modified function denotes the validation error of hypothetical ML model, l denotes the log of number of iterations the model is trained and x_1, x_2 denote the hyper-parameters of the model. We define the modified function as $B(x_1, x_2, l) = b(x_1, x_2) - l$ to model the decrease in error with more training iterations. Furthermore, the decrease in error diminishes with more training iterations, which is a standard phenomenon in practice. The training cost for the tuple (x_1, x_2, l) is $\exp(l)$. We compare our algorithm with MES, Multi-Fidelity MES, EI, and CostEI. Figure 1 shows the simple regret vs. the cost incurred. We observe that CostIDS achieves a better cost vs. simple regret tradeoff compared to the other baselines, early in the optimization. CostIDS is out-performed when the optimum has been reasonably located and remains to be fine tuned.

7. Conclusion

In this paper we proposed a cost aware BO approach based on Information Directed Sampling. We showed that our algorithm is provably no-regret, while being conceptually simple. We also showed promising results on a synthetic function. We leave further experiments on real functions to future work.

References

Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv preprint arXiv:1204.5721*, 2012.

Alexander IJ Forrester, András Sóbester, and Andy J Keane. Multi-fidelity optimization via surrogate modelling. *Proceedings of the royal society a: mathematical, physical and engineering sciences*, 463(2088):3251–3269, 2007.

Peter I Frazier and Jialei Wang. Bayesian optimization for materials design. In *Information Science for Materials Discovery and Design*, pages 45–75. Springer, 2016.

Ryan-Rhys Griffiths and José Miguel Hernández-Lobato. Constrained bayesian optimization for automatic chemical design. *arXiv preprint arXiv:1709.05501*, 2017.

José Miguel Hernández-Lobato, Matthew W Hoffman, and Zoubin Ghahramani. Predictive entropy search for efficient global optimization of black-box functions. In *Advances in neural information processing systems*, pages 918–926, 2014.

Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

Kirthevasan Kandasamy, Gautam Dasarathy, Jeff Schneider, and Barnabás Póczos. Multi-fidelity bayesian optimisation with continuous approximations. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1799–1808. JMLR. org, 2017.

Kirthevasan Kandasamy, Akshay Krishnamurthy, Jeff Schneider, and Barnabás Póczos. Parallelised bayesian optimisation via thompson sampling. In *International Conference on Artificial Intelligence and Statistics*, pages 133–142, 2018.

Johannes Kirschner and Andreas Krause. Information directed sampling and bandits with heteroscedastic noise. *arXiv preprint arXiv:1801.09667*, 2018.

Johannes Kirschner, Tor Lattimore, and Andreas Krause. Information directed sampling for linear partial monitoring. *arXiv preprint arXiv:2002.11182*, 2020.

Eric Hans Lee, Valerio Perrone, Cedric Archambeau, and Matthias Seeger. Cost-aware bayesian optimization. *arXiv preprint arXiv:2003.10870*, 2020.

Nikolay Nikolov, Johannes Kirschner, Felix Berkenkamp, and Andreas Krause. Information-directed exploration for deep reinforcement learning. *arXiv preprint arXiv:1812.07544*, 2018.

Biswajit Paria, Kirthevasan Kandasamy, and Barnabás Póczos. A flexible framework for multi-objective bayesian optimization using random scalarizations. In *UAI*, 2019.

Carl Edward Rasmussen and Christopher KI Williams. Gaussian processes for machine learning. 2006. *The MIT Press, Cambridge, MA, USA*, 38:715–719, 2006.

Daniel Russo and Benjamin Van Roy. Learning to optimize via information-directed sampling. In *Advances in Neural Information Processing Systems*, pages 1583–1591, 2014a.

Daniel Russo and Benjamin Van Roy. Learning to optimize via posterior sampling. *Mathematics of Operations Research*, 39(4):1221–1243, 2014b.

Daniel Russo and Benjamin Van Roy. An information-theoretic analysis of thompson sampling. *The Journal of Machine Learning Research*, 17(1):2442–2471, 2016.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.

Jialin Song, Yuxin Chen, and Yisong Yue. A general framework for multi-fidelity bayesian optimization with gaussian processes. *arXiv preprint arXiv:1811.00755*, 2018.

Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*, 2009.

Shion Takeno, Hitoshi Fukuoka, Yuhki Tsukada, Toshiyuki Koyama, Motoki Shiga, Ichiro Takeuchi, and Masayuki Karasuyama. Multi-fidelity bayesian optimization with max-value entropy search. *arXiv preprint arXiv:1901.08275*, 2019.

Zi Wang and Stefanie Jegelka. Max-value entropy search for efficient bayesian optimization. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 3627–3635. JMLR. org, 2017.

Appendix A. Gaussian Processes

Gaussian processes (GP) (Rasmussen and Williams, 2006) are used to model the unknown function f in many BO algorithms due to their ability to provided well calibrated uncertainty estimates, which are also straightforward to compute.

GPs are used to define a prior distribution over functions defined on some input space $\mathcal{X}$. GPs are characterized by a mean function $\mu : \mathcal{X} \to \mathbb{R}$ and a covariance (kernel) function $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. For any function f drawn from a GP, $f \sim \mathcal{GP}(\mu, \kappa)$, and some finite set of points $\boldsymbol{x}_i \in \mathcal{X}$ $(1 \leq i \leq n)$, the function values $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)$ follow a multivariate Gaussian distribution with mean μ and covariance K given by $\mu_i = \mu(\boldsymbol{x}_i),\ K_{ij} = \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)\ \ \forall 1 \leq i, j \leq n$. Examples of popular kernel functions κ include the squared exponential and Matern kernels.

Given observations $\mathcal{D}_{t-1} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{t-1}$, and assuming the generative process $y_i = f(\boldsymbol{x}_i) + \epsilon_i \in \mathbb{R}$, $\epsilon_i \sim \mathcal{N}(\mu, \sigma^2)$ (for some specified noise variance σ^2), the posterior process is also a GP with the mean and kernel function given by

$$\begin{aligned} \mu_t(\boldsymbol{x}) &= k^T (K + \sigma^2 I)^{-1} Y, \\ \kappa_t(\boldsymbol{x}, \boldsymbol{x}') &= \kappa(\boldsymbol{x}, \boldsymbol{x}') - k^T (K + \sigma^2 I)^{-1} k'. \end{aligned} \tag{11}$$

where $Y = \{y_i\}_{i=1}^{t}$ is the vector of observed values, $K = \{\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)\}_{i,j=1}^{t}$ is the Gram matrix, with $k = \{\kappa(\boldsymbol{x}, \boldsymbol{x}_i)\}_{i=1}^{t}$, and $k' = \{\kappa(\boldsymbol{x}', \boldsymbol{x}_i)\}_{i=1}^{t}$. The posterior variance at $\boldsymbol{x}$ is given by $\sigma_t(\boldsymbol{x})^2 = \kappa_t(\boldsymbol{x}, \boldsymbol{x})$. Additional details on GPs can be found in (Rasmussen and Williams, 2006).

Appendix B. Missing Proofs

Here we provide a brief outline of the proofs.

Proof (Theorem 4) We first show that $\mathbb{E}[s_\Lambda] \leq \frac{1}{\Lambda - \lambda_{\max}} \mathbb{E}[R_\Lambda]$.

$$\mathbb{E}[s_\Lambda] \leq \mathbb{E}\left[\frac{\sum_{i=1}^{t_\Lambda} \lambda(\boldsymbol{x}_i) r_i}{\sum_{i=1}^{t_\Lambda} \lambda(\boldsymbol{x}_i)}\right] \leq \mathbb{E}\left[\frac{R_\Lambda}{\Lambda - \lambda_{\max}}\right]$$

It remains to prove the upper bound on the cumulative regret.

$$\begin{aligned}
\mathbb{E}\left[R_\Lambda\right] = \mathbb{E}\left[\sum_{i=1}^{t_\Lambda}\lambda(\boldsymbol{x}_i)r_i\right] &= \mathbb{E}\left[\sum_{i=1}^{t_\Lambda}\lambda(\boldsymbol{x}_i)(f^* - f(\boldsymbol{x}_i))\right] \\
&\leq \mathbb{E}\left[\sum_{i=1}^{t_\Lambda}\lambda(\boldsymbol{x}_i)(\mathcal{U}_i - f(\boldsymbol{x}_i))\right] + \mathbb{E}\left[\sum_{i=1}^{t_\Lambda}\frac{\lambda_{\max}}{t^2}\right] \\
&\quad \text{(by definition of } \mathcal{U}_t\text{)} \\
&\leq \mathbb{E}\sqrt{\left(\sum_{i=1}^{t_\Lambda}\lambda(\boldsymbol{x}_i)\right)\left(\sum_{i=1}^{t_\Lambda}\lambda(\boldsymbol{x}_i)(\mathcal{U}_i - f(\boldsymbol{x}_i))^2\right)} + \lambda_{\max}\frac{\pi^2}{6} \\
&\quad \text{(by Cauchy Schwartz inequality)} \\
&\leq \sqrt{\Lambda\mathbb{E}\left[\sum_{i=1}^{t_\Lambda}\Gamma_i^{\text{cost}}\sigma_i(\boldsymbol{x}_i)^2\right]} + \lambda_{\max}\frac{\pi^2}{6} \\
&\quad \text{(by concavity of the square root function)} \\
&\leq \sqrt{\Lambda C_1\mathbb{E}\left[\Gamma_{t_\Lambda}^{\text{cost}}\gamma_{t_\Lambda}\right]} + \lambda_{\max}\frac{\pi^2}{6} \\
&\quad \text{(follows from Srinivas et al. (2009))}
\end{aligned}$$

■