A parametric approach to Bayesian optimization with pairwise comparisons

Marco Cox
Eindhoven University of Technology
m.g.h.cox@tue.nl

Bert de Vries
Eindhoven University of Technology
and GN Hearing
bdevries@ieee.org

Abstract

Optimizing a (preference) function through a small number of pairwise comparisons is challenging since pairwise comparisons provide limited information about the underlying function. In practice, preference functions often have a single peak, and this property could be exploited to speed up the optimization process. In this paper we describe a Bayesian optimization method aimed at achieving this.

1 Introduction

The idea behind Bayesian optimization (BO) is to use assumptions about the (expensive to evaluate) objective function in order to reduce the number of function evaluations (Brochu et al., 2010). These assumptions are defined as a prior distribution over the space of possible objective functions. The usefulness of this approach depends on both the validity and strength of the assumptions. Weak assumptions might be valid but unhelpful in reducing the number of function evaluations. Strong assumptions on the other hand might easily be violated, leading to fast convergence to the wrong solution. The Gaussian process (GP) prior is a popular choice in BO, in part because it can encode reasonable assumptions such as smoothness, and the strength of these assumptions can be tuned through (maximum likelihood) estimation of the hyperparameters.

Bayesian optimization can still be applied if the objective function cannot be evaluated directly, but only (noisy) binary pairwise comparisons are available. This is a common scenario in for example human preference learning, where users can often only judge inputs relative to each other ("I prefer `B` over `A`") but not on an absolute scale. However, such pairwise comparisons provide less information about the objective function than direct function evaluations, increasing the number of required iterations in a BO loop. On top of this, practical considerations might limit the number of available iterations. As a result, just assuming the objective function to be smooth – for example by putting a GP prior on it – might not be enough to converge to a reasonable solution within the available time budget. The problem here is that the smoothness assumption is relatively weak: it requires visiting every neighborhood of the (potentially high-dimensional) input space to obtain a good estimate of the objective function. To avoid this, optimization methods based on pairwise comparisons have been developed that avoid probabilistic inference of the objective function altogether, for example by resorting to direct stochastic gradient ascent on the objective function (Yue and Joachims, 2009).

In this work we propose a different approach: making stronger assumptions about the shape of the objective function. In a variety of practical setups it seems reasonable to assume that the objective function has a single peak and is monotonically decreasing away from this peak. Think for example about a human's comfort level as a function of the temperature setting of a climate control system. We present a BO method for setups with pairwise comparison observations that exploits this property to significantly speed up the optimization. Since a GP prior is unable to encode the "single peak assumption", we propose a different, parametric model for the objective function. This allows us to

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

retain a full Bayesian treatment of the optimization problem while reducing the number of iterations required to converge to the neighborhood of the optimum.

2 Problem definition

Consider some system or algorithm that is governed by a configuration vector $\mathbf{x} \in \mathcal{X}$, where $\mathcal{X} \subseteq \mathbb{R}^d$. We want to optimize the system's performance by tuning $\mathbf{x}$, but we can only do so by playing the following sequential game: At every step i, we propose a new configuration vector $\mathbf{x}'_i$ and observe a (noisy) binary label $y_i \in \{+1, -1\}$ indicating whether proposal $\mathbf{x}'_i$ results in better performance than the current configuration $\mathbf{x}_i$. If so, the proposal is adopted and $\mathbf{x}_{i+1}$ is set to $\mathbf{x}'_i$, otherwise $\mathbf{x}_{i+1} = \mathbf{x}_i$.

The *pairwise comparison* between two inputs $(\mathbf{x}, \mathbf{x}')$ is governed by an underlying latent objective function $f : \mathcal{X} \to \mathbb{R}$ through a response model $p(y = +1|\mathbf{x}, \mathbf{x}', f) = \Pr\{f(\mathbf{x}') + \epsilon \geq f(\mathbf{x})\}$, where ϵ is a zero-mean noise term. We assume that obtaining a comparison is expensive, for example because it involves asking a human for feedback. Moreover, we assume that f has a single peak and is monotonically decreasing away from this peak.

Our goal is to incrementally build an input sequence $[\mathbf{x}'_1, \ldots, \mathbf{x}'_N]$ that maximizes the cumulative value of the objective function for sufficiently large N: $V = \sum_{i=1}^{N} f(\mathbf{x}'_i)$. The problem of how to generate the input sequence in an optimal way is known as the (continuous) dueling bandits problem (Busa-Fekete and Hüllermeier, 2014), and it involves an exploration–exploitation trade off. On the one hand we want to select $\mathbf{x}'_i$ such that $f(\mathbf{x}'_i)$ is probably large. On the other hand we want to propose inputs that allow us to learn about f, so we can propose better inputs in the future.

3 Bayesian optimization in the continuous dueling bandit setting

We briefly introduce BO in the dueling bandit setting. BO is ususally aimed at solving the global optimization problem $\mathbf{x}_\star = \arg\max_{\mathbf{x}} f(\mathbf{x})$, where f is a black box function that is expensive to evaluate (Brochu et al., 2010; Shahriari et al., 2016). The main idea is to posit a probabilistic model for f, and to use this model to sequentially select 'good' query points, such that the number of function evaluations is kept to a minimum. The most common choice for $p(f)$ is the GP prior, which can capture assumptions about f such as smoothness and periodicity (Rasmussen and Williams, 2006).

Although f cannot be evaluated directly in the pairwise comparison setup, the BO paradigm can still be applied. This requires specifying a probabilistic generative model which factors in a response model and a prior on the latent objective function: $p(y, f|\mathbf{x}, \mathbf{x}') = p(f)p(y|\mathbf{x}, \mathbf{x}', f)$. Chu and Ghahramani (2005) worked out this model for a GP prior $p(f)$ and a probit response model $p(y|\mathbf{x}, \mathbf{x}', f) = \Phi(y \cdot (f(\mathbf{x}') - f(\mathbf{x})))$. This response model assumes that the pairwise comparison is performed under additive Gaussian noise on f. Given a set of pairwise comparisons $\mathcal{D}_i = \{\mathbf{x}_{1:i}, \mathbf{x}'_{1:i}, y_{1:i}\}$, the posterior GP $p(f|\mathcal{D}_i)$ can be obtained through approximate Bayesian inference, and can then be used to select a next input as in regular BO through some acquisition function.

Multiple methods such as BALD (Houlsby et al., 2011) and BOPPER (Gonzalez et al., 2016) have been developed for finding the next input pair $(\mathbf{x}_{i+1}, \mathbf{x}'_{i+1})$ based on the posterior $p(f|\mathcal{D}_i)$ such that y_{i+1} will provide maximum information about f. However, these methods are generally not well suited for the bandit setting since they only aim at exploration (exploring f to locate its optimum in as few steps as possible) and ignore the exploitation aspect. Thompson sampling (Russo et al., 2017) is a simple but effective method for selecting inputs that has been shown to yield good results in the bandit setting (Agrawal and Goyal, 2012; Leike et al., 2016). Under Thompson sampling, the next proposed input is chosen as $\mathbf{x}'_{i+1} = \arg\max_{\mathbf{x}} \tilde{f}(\mathbf{x})$, where $\tilde{f} \sim p(f|\mathcal{D}_i)$ is a random objective function drawn from the posterior.

4 A parametric model for pairwise comparisons

Our goal in this work is to develop a BO method that can exploit the assumption that the objective function has a single peak and is decreasing away from this peak. This can be achieved by constructing a prior $p(f)$ that enforces this property, i.e. that puts little or zero probability mass on functions that do not have this property. Unfortunately, this is not possible with a GP prior.

Bell shape, $f_{\hat{\mathbf{x}},\Lambda}(\mathbf{x}) = \exp[-(\mathbf{x}-\hat{\mathbf{x}})^T\Lambda(\mathbf{x}-\hat{\mathbf{x}})]$.

Cone, $f_{\hat{\mathbf{x}},\Lambda}(\mathbf{x}) = -\sqrt{(\mathbf{x}-\hat{\mathbf{x}})^T\Lambda(\mathbf{x}-\hat{\mathbf{x}})}$.

Parabola, $f_{\hat{\mathbf{x}},\Lambda}(\mathbf{x}) = -(\mathbf{x}-\hat{\mathbf{x}})^T\Lambda(\mathbf{x}-\hat{\mathbf{x}})$.

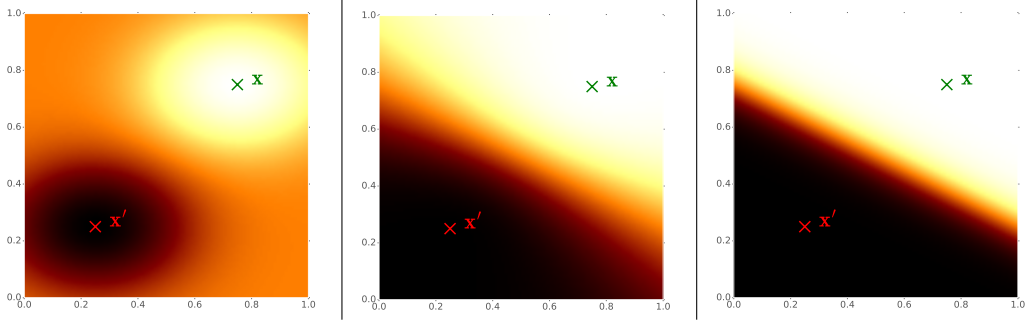


Figure 1: Analytical forms for objective function f. Parameter $\hat{\mathbf{x}}$ specifies the position of the maximum and parameter Λ is a matrix that governs the shape. The heat maps illustrate the likelihood functions for the maximizing argument, $l_{y,\mathbf{x},\mathbf{x}',\Lambda}(\hat{\mathbf{x}}) = p(y|\mathbf{x},\mathbf{x}',\Lambda,\hat{\mathbf{x}})$, that follow from the respective analytical forms for $d = 2$.

To construct a prior that encodes the single peak assumption, we propose to fix f to a specific analytical form. We consider three such forms, corresponding to the assumptions that f is either (a) bell-shaped, (b) cone-shaped or (c) parabolic. All of these forms have two parameters: parameter $\hat{\mathbf{x}} \in \mathcal{X}$ determines the position of the peak and parameter Λ is positive-definite $d \times d$ matrix that determines the shape of the function around the peak. The forms are listed in Fig. 1, together with the likelihood functions they induce for the argmax parameter $\hat{\mathbf{x}}$. Note that it is sufficient to fix the form of the objective function upto an additive constant, since the pairwise comparison response model is invariant to additive constants in the underlying objective function.

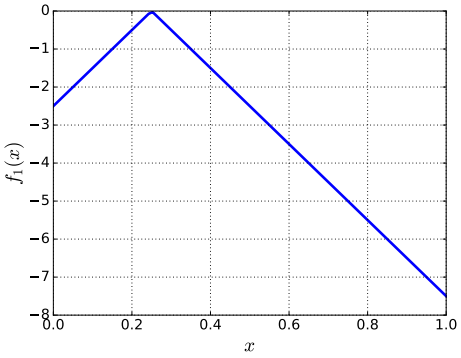
Given the analytical form of f, $p(f)$ is now specified through priors on the parameters: $p(\hat{\mathbf{x}})$ and $p(\Lambda)$. Without loss of generality, we constrain the input domain to the hypercube $\mathcal{X} = [0,1]^d$. To ensure $p(\hat{\mathbf{x}}) = 0$ for $\hat{\mathbf{x}} \notin \mathcal{X}$, we specify $p(\hat{\mathbf{x}})$ implicitly through a prior on a transformed variable $\hat{\mathbf{z}}$:

$$\begin{aligned} \hat{\mathbf{x}} &= \Phi(\hat{\mathbf{z}}), & \hat{\mathbf{z}} &\sim \mathcal{N}(\boldsymbol{\mu}, \Sigma), \\ \Lambda &= \texttt{diagm}([\lambda_1, \ldots, \lambda_D]), & \lambda_d &\sim \texttt{Gamma}(k_d, \theta_d). \end{aligned} \tag{1}$$

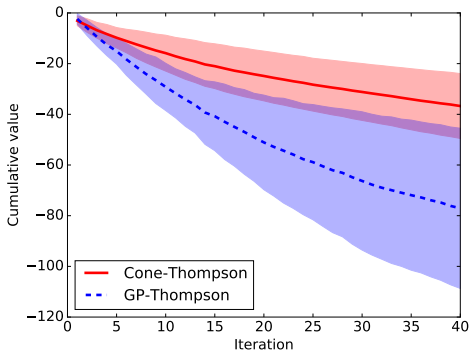
Fixing the analytical form of f makes the pairwise comparison model parametric, and posterior inference of f boils down to posterior inference of parameters $\hat{\mathbf{x}}$ and Λ. Unsurprisingly, this is not analytically tractable. However, one can apply black box variational inference methods to achieve approximate Bayesian inference. In our experiments we use the probabilistic programming language Stan and its automatic differentiation variational inference engine to automate posterior inference (Kucukelbir et al., 2015). The fact that this model leads to an explicit posterior distribution for $\hat{\mathbf{x}}$ is convenient since it makes it trivial to implement Thompson sampling: one can sample new proposals directly from $p(\hat{\mathbf{x}}|\mathcal{D})$.

Assuming that the objective function admits to a simple parametric form is clearly a strong assumption, and there will be a model mismatch if the true objective function does not have the chosen form. However, the hope is that if the model can provide a reasonable fit at the evaluation points, it can significantly speed up the localization of the peak in the objective function compared to more flexible models like the GP. This is useful in the bandit setting.

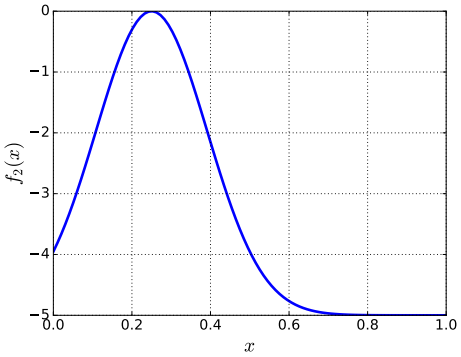
The illustrations of the likelihood functions $l_{y,\mathbf{x},\mathbf{x}',\Lambda}(\hat{\mathbf{x}})$ of the cone and parabola forms in Fig. 1 provide an interesting insight into why these assumptions may lead to faster convergence. Under these models, each pairwise comparison yields a likelihood term that suppresses a much larger part of the input space than just the direct neighborhood of one of the inputs. In contrast, the bell-shaped form leads to likelihood terms that will only affect the posterior distribution of $\hat{\mathbf{x}}$ in the direct neighborhood of the input pair. This is similar to the behavior resulting from a GP prior with a kernel that enforces smoothness. If Λ is the identity matrix, the cone form reduces to the negative Euclidean distance: $f_{\hat{\mathbf{x}}}(\mathbf{x}) = -||\mathbf{x} - \hat{\mathbf{x}}||$.



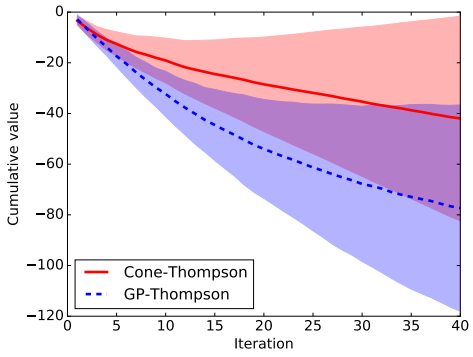
(a) $f_1(x) = -\sqrt{100(x - 0.25)^2}$



(b) Results of optimization of f_1.



(c) $f_2(x) = 5\exp(-25(x - 0.25)^2) - 5$



(d) Results of optimization of f_2.

Figure 2: Results of dueling bandits optimization of two artifical objective functions. The cumulative value curves are the averages of 50 runs and the shaded areas represent two standard deviations.

5 Experiments

We test the usefulness of the parametric model in the dueling bandit optimization setting from Section 2 by comparing its performance to that of a GP model on two artificial objective functions. The first objective function is a 1 dimensional cone, depicted in Fig. 2a. The second one is a bell-shaped function, shown in Fig. 2c. In both experiments we compare the cone variant of the parametric model (Cone-Thompson) to a GP model with a squared exponential kernel (GP-Thompson). Since the parametric model assumes the objective function to have the analytical form of a cone, there is a model mismatch in the second experiment, allowing us to test the robustness under mismatch. Priors $p(\hat{\mathbf{x}})$ and $p(\Lambda)$ are chosen to be uninformative. Inputs $[\mathbf{x}'_1, \ldots, \mathbf{x}'_{40}]$ are selected through Thompson sampling under both models. The hyperparameters of the GP model are fitted in every iteration by marginal log-likelihood optimization.

The results in Figures 2b and 2d show that Cone-Thompson consistently and significantly outperforms GP-Thompson on both acquisition functions. This is encouraging because it suggests that the parametric model might be able to outperform GP-based models on real objective functions that are peak-shaped.

6 Conclusions

If the objective function in a Bayesian optimization setup with pairwise comparisons is peak-shaped, it is possible to exploit this property to speed up the optimization. We proposed a simple parametric prior for objective functions that can be used to achieve this. In practical systems, it could be beneficial to add this model to an ensemble model that also includes more flexible priors.

References

Agrawal, S. and Goyal, N. (2012). Analysis of Thompson Sampling for the Multi-armed Bandit Problem. In *COLT*, pages 39.1–39.26.

Brochu, E., Cora, V. M., and De Freitas, N. (2010). A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*.

Busa-Fekete, R. and Hüllermeier, E. (2014). A survey of preference-based online learning with bandit algorithms. In *International Conference on Algorithmic Learning Theory*, pages 18–39. Springer.

Chu, W. and Ghahramani, Z. (2005). Preference learning with Gaussian processes. In *Proceedings of the 22nd international conference on Machine learning*, pages 137–144. ACM.

Gonzalez, J., Dai, Z., Damianou, A., and Lawrence, N. D. (2016). Bayesian optimisation with pairwise preferential returns. In *NIPS Workshop on Bayesian Optimization*.

Houlsby, N., Huszár, F., Ghahramani, Z., and Lengyel, M. (2011). Bayesian Active Learning for Classification and Preference Learning. *arXiv:1112.5745 [cs, stat]*.

Kucukelbir, A., Ranganath, R., Gelman, A., and Blei, D. (2015). Automatic variational inference in Stan. In *Advances in neural information processing systems*, pages 568–576.

Leike, J., Lattimore, T., Orseau, L., and Hutter, M. (2016). Thompson sampling is asymptotically optimal in general environments. In *Proceedings of the 2016 Conference on Uncertainty in Artificial Intelligence (UAI)*, New York.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Russo, D., Van Roy, B., Kazerouni, A., and Osband, I. (2017). A Tutorial on Thompson Sampling. *arXiv preprint arXiv:1707.02038*.

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and Freitas, N. d. (2016). Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE*, 104(1):148–175.

Yue, Y. and Joachims, T. (2009). Interactively optimizing information retrieval systems as a dueling bandits problem. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1201–1208. ACM.