Relevant bibliography for Bayesian Optimization

Ricardo Lopez

September 1, 2022

We present here the most relevant bibliography found of Bayesian Optimization (BO).

1 Bayesian optimization methods

Frazier in [1] presents a nice tutorial on BO. He describes how Bayesian optimization works, including Gaussian process (GP) regression and three common acquisition functions: expected improvement, entropy search, and knowledge gradient. Then he discusses more techniques: multiple function evaluations in parallel, expensive-to-evaluate constraints, random environmental conditions and the inclusion of derivative information. Brochu et al. [2] is a nice complement tutorial and lecture on BO to [1]. Besides to [1], they also present two detailed extensions of Bayesian optimization, with experiments—active user modelling with preferences, and hierarchical reinforcement learning and a discussion of the pros and cons of Bayesian optimization based on their experiences. Maggi et al. in [3] provide a short and accessible introduction to Bayesian optimisation, as an addition they include a radio resource management problem of power control in 5G cellular networks with aim to utilize the limited radio-frequency spectrum resources and radio network infrastructure as efficiently as possible, for which BO is able to converge to almost optimal solutions in tens of iterations without significant performance drops during exploration

1.1 Methods for low dimensional BO

Gelbart et al. in [4] study Bayesian optimization for constrained problems in the general case that noise may be present in the constraint functions, and the objective and constraints may be evaluated independently. Besides, they develop a framework to solve such problems. Similarly to [4], Garder et al. [5] extend BO to the inequality-constrained optimization setting, particularly the setting in which evaluating feasibility is just as expensive as evaluating the objective. Likewise to [4], they present constrained Bayesian optimization which places a prior distribution on both the objective and the constraint functions. Astudillo & Frazier in [6] consider optimization of composite objective functions $f(x) = g(h(x))$, where h is a black-box derivative-free expensive to-evaluate function with vector-valued outputs, and g is a cheap-to-evaluate real-valued function. These problems arise, for example, in calibration of simulators to real-world data as in [8]. While these problems can be solved with standard Bayesian optimization, [6] propose a novel approach that exploits the composite structure of the objective function to substantially improve sampling efficiency. Their approach models h using a multi-output Gaussian process and chooses where to sample using the expected improvement evaluated on the implied non-Gaussian posterior on f. Hoffman et al. in [7] instead of using a single acquisition function, they adopt a portfolio of acquisition functions governed by an online multi-armed bandit strategy. Based in this strategy they propose several portfolio strategies, the best of which they call GP-Hedge, and show that this method outperforms the best individual acquisition function.

1.2 Methods for high dimensional BO

Solutions have been proposed to tackle high-dimensional Bayesian optimization. Wang et al. in [9] projected the high-dimensional space into a low-dimensional subspace and then optimized the acquisition function in a low-dimensional subspace. The assumption that only some dimensions are effective is often restrictive. Qian et al. in [10] studied the case when all dimensions are effective, but many of them only have a small bounded effect by using sequential random embedding to reduce the embedding gap. These methods may not work if all dimensions in the high-dimensional function are similarly effective. The additive decomposition assumption is another solution for high dimensional function analysis. Kandasamy et al. in [11] proposed a model in which the objective function is assumed to be the sum of a set of low-dimensional functions with disjoint dimensions and then BO can be performed in the low-dimensional space. However, in practice it is difficult to know the decomposition of functions in advance, especially for non-separable functions. Li et al. in [12] propose a method for high-dimensional BO, that uses a dropout strategy to optimize only a subset of variables at each iteration.

Wang et al. in [13] used a projection matrix with standard Gaussian entries and showed that the active subspace is rank-preserved but strongly dilated. Thus, they expanded the low-dimensional search space to ensure that its projection to the high-dimensional space contains an optimal solution with reasonable probability. Their algorithm fits a GP model to the low-dimensional space and projects points back to the high-dimensional space using the inverse random projection. Points whose high-dimensional image is outside the search domain are convex-projected to the boundary, which may lead to over-exploration of the boundary regions. To tackle the distortion problem in [13], Munteanu et al. in [14] present a theoretically founded approach for high-dimensional Bayesian optimization based on low-dimensional subspace embeddings. They prove that the error in the Gaussian process model is bounded tightly when going from the original high-dimensional search domain to the low dimensional embedding. This implies that the optimization process in the low-dimensional embedding proceeds essentially as if it were run directly on an unknown active subspace of low dimensionality. In likewise formulations, Jaquier et al. in [15] propose to exploit the geometry of non-Euclidean search spaces as $n-1$ sphere in an n dimensional euclidean space [16], which often arise in a variety of domains, to learn structure-preserving mappings and optimize the acquisition function of BO in low-dimensional latent spaces. Their approach, built on Riemannian manifolds theory, features geometry-aware Gaussian processes that jointly learn a nested-manifold embedding and a representation of the objective function in the latent space.

2 Bayesian optimization in budget allocation problems

Our desired problem formulation falls specifically within the budgeted bayesian optimization literature. Snoek et al. in [17] propose a BO framework by considering the expected improvement per unit of cost based on the knowledge of a modelled cost function. Lee et al. in [18] proposes a simple variation called the expected improvement per unit of cost with cost cooling, where is considered the ratio between the current remaining and initial budgets. The intuition behind of the method in [18] is that evaluating points with high cost should be discouraged early in the BO loop and accommodated as the budget is consumed. Lee et al. in [19] generalize their work in [18] and formulate BO as a constrained Markov decision process by considering a black-box cost function instead of a modeled cost function and a finite horizon. Astudillo & Frazier in [20] formulate the budgeted bayesian optimization problem in a Lee et al. [19] approach by an acquisition function that generalizes classical expected improvement to the setting of unknown evaluation costs at a random horizon.

References

[1] Frazier, P. I. (2018). 'A tutorial on Bayesian optimization'. arXiv preprint arXiv:1807.02811.

[2] Brochu, E., Cora, V. M., & 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.

[3] Maggi, L., Valcarce, A., & Hoydis, J. (2021). 'Bayesian optimization for radio resource management: Open loop power control'. IEEE Journal on Selected Areas in Communications, 39(7), 1858-1871

[4] Gelbart, M. A., Snoek, J., & Adams, R. P. (2014). 'Bayesian optimization with unknown constraints'. arXiv preprint arXiv:1403.5607.

[5] Gardner, J. R., EDU, W., Kusner, M. J., & Xu, E. Z. E. 'Supplementary Material: Bayesian Optimization with Inequality Constraints'.

[6] Astudillo, R., & Frazier, P. (2019, May). 'Bayesian optimization of composite functions'. In International Conference on Machine Learning (pp. 354-363). PMLR.

[7] Hoffman, M., Brochu, E., & De Freitas, N. (2011, July). 'Portfolio Allocation for Bayesian Optimization'. In UAI (pp. 327-336).

[8] Vrugt, J. A., Hopmans, J. W., & Šimunek, J. (2001). 'Calibration of a two-dimensional root water uptake model'. Soil Science Society of America Journal, 65(4), 1027-1037.

[9] Ziyu Wang, Masrour Zoghi, Frank Hutter, David Matheson, & Nando De Freitas. 'Bayesian optimization in high dimensions via random embeddings'. In IJCAI, pages 1778–1784, 2013.

[10] Hong Qian, Yi-Qi Hu, & Yang Yu. 'Derivative-free optimization of high-dimensional nonconvex functions by sequential random embeddings'. In Proceedings of the 25th International Joint Conference on Artificial Intelligence, 2016.

[11] Kirthevasan Kandasamy, Jeff G. Schneider, & Barnabs Pczos. 'High Dimensional Bayesian Optimisation and Bandits via Additive Models'. In ICML, volume 37, pages 295–304, 2015

[12] Li, C., Gupta, S., Rana, S., Nguyen, V., Venkatesh, S., & Shilton, A. (2018). 'High dimensional Bayesian optimization using dropout'. arXiv preprint arXiv:1802.05400.

[13] Wang, Z., Hutter, F., Zoghi, M., Matheson, D., & de Freitas, N. 'Bayesian optimization in a billion dimensions via random embeddings'. Journal of Artificial Intelligence Research, 55: 361–387

[14] Nayebi, A., Munteanu, A., & Poloczek, M. (2019, May). 'A framework for Bayesian optimization in embedded subspaces'. In International Conference on Machine Learning (pp. 4752-4761). PMLR.

[15] Jaquier, N., & Rozo, L. (2020). 'High-dimensional Bayesian optimization via nested Riemannian manifolds'. Advances in Neural Information Processing Systems, 33, 20939-20951.

[16] Daverman, R. J. (1978). 'Embeddings of $(n-1)$-spheres in Euclidean n-space'. Bulletin of the American Mathematical Society, 84(3), 377-405.

[17] Snoek, J., Larochelle, H., & Adams, R. P. (2012). 'Practical Bayesian optimization of machine learning algorithms'. Advances in neural Information Processing Systems, volume 25, pages 2951–2959. Curran Associates, Inc.

[18] Lee, E. H., Perrone, V., Archambeau, C., & Seeger, M. (2020b). 'Cost-aware bayesian optimization'. arXiv preprint arXiv:2003.10870.

[19] Lee, E. H., Eriksson, D., Perrone, V., & Seeger, M. (2021). 'A nonmyopic approach to costconstrained Bayesian optimization'. In Conference on Uncertainty in Artificial Intelligence.

[20] Astudillo, R., Jiang, D., Balandat, M., Bakshy, E., & Frazier, P. (2021). 'Multi-Step Budgeted Bayesian Optimization with Unknown Evaluation Costs'. Advances in Neural Information Processing Systems, 34.