Retrospective From the Authors Peter Frazier, Warren Powell, and Savas Dayanik

We are deeply honored by this recognition and grateful to the selection committee and editor-in-chief for this award.

This work contributes to Bayesian optimization, a powerful approach for optimizing time-consuming-to-evaluate non-convex derivative-free objective functions. Bayesian optimization builds a probabilistic surrogate (usually a Gaussian process) for the objective function $f(x)$. It chooses where to evaluate $f(x)$ next by optimizing a so-called "acquisition function" that quantifies the value of an evaluation of $f(x)$ as a function of the point evaluated, x. This work also contributes to the broader areas of optimal learning and Bayesian sequential experimental design, which use a Bayesian approach to intelligently choose which data to collect or which experiment to run in settings where data is expensive to collect.

This work played a significant role in developing the knowledge gradient (KG) acquisition function. The KG acquisition function at x is the value of evaluating the objective function at x and then, based on the $f(x)$ observed, selecting a final solution to the optimization problem. It is distinguished from the widely-used expected improvement method because the final solution selected can be a point that has not been previously evaluated. We began studying this approach as a one-step lookahead approximation to the full dynamic programming treatment of the problem (later documented by Ginsbourger and Le Riche 2010). What surprised us is that looking only one step ahead can produce an extremely effective method.

This work introduced the KG acquisition function for correlated normal prior distributions — the discrete analog of a Gaussian process. Previous work on KG, comprehensively reviewed in Garnett (2023), only looked at a restricted case of independent priors (Frazier et al. 2008, Gupta and Miescke 1996) or Gauss-Markov priors where the final solution selected is a previously-evaluated point (Mockus 1972). The more flexible correlated normal prior distributions enabled by the awarded work are critical for achieving good empirical performance in many problems.

Study of Bayesian optimization grew rapidly following the later discovery by Snoek et al. (2012) that Bayesian optimization effectively tunes hyperparameters of deep neural networks. KG became one of the main acquisition functions used alongside expected improvement, probability of improvement, entropy search, upper confidence bound, and Thompson sampling.

Papers building on the awarded work discovered that KG methods are particularly effective when we can acquire information from sources other than direct objective function evaluation. For example, in multi-information source Bayesian optimization (Poloczek et al. 2017), we can evaluate less expensive but biased surrogates for the objective (e.g., a simulation that replaces a stochastic process by a fluid approximation). Many of the most widely-used acquisition functions (including expected improvement) do not extend easily to this setting but KG methods do so naturally. This led to the development of KG methods for a wide variety of problems. Examples include active learning for allocating crowd workers to labeling tasks (Chen et al. 2013), optimization of compositions of unknown time-consuming black-box functions (Astudillo and Frazier 2021, Buathong et al. 2024), and optimization with gradient observations (Wu et al. 2017), common random numbers (Pearce et al. 2022), or across multiple tasks (Pearce and Branke 2018).

The knowledge gradient has also been adopted outside of Bayesian optimization, especially in the literature on optimal learning (Powell and Ryzhov 2012). This includes work on online learning (Ryzhov et al. 2012), preference exploration (Lin et al. 2022), hierarchical aggregation for settings where each choice is described by a vector of attributes (Mes et al. 2011), and work for parametric belief models (Negoescu et al. 2011, He and Powell 2018, He et al. 2020).

Later, KG methods were implemented in BoTorch (Balandat et al. 2020), a widely-used high-quality open-source software package for Bayesian optimization maintained by Meta. The implementation was based on a discovery that sample average approximation on top of the computing infrastructure provided by PyTorch is an effective approach for optimizing KG. KG is more complex than most of the other standard acquisition functions and this advance simplified and further accelerated the use of KG. Spurred by this award, we have migrated the original code from a zip file on a webpage to a more accessible GitHub repository (Frazier et al. 2009).

Looking forward, algorithms enabled by KG continue to be important for core applications in operations research, machine learning, and engineering. New questions and opportunities continue to emerge — for example, how to best allocate computing effort in multi-LLM routing or agentic workflows and for directing self-driving labs for materials and drug discovery. KG methods are a powerful and flexible tool in these problems and we are grateful to the selection committee for drawing attention to them.

References

Astudillo R, Frazier PI (2021) Bayesian optimization of function networks. *Advances in Neural Information Processing Systems* (Curran Associates, Inc.).

Balandat M, Karrer B, Jiang DR, Daulton S, Letham B, Wilson AG, Bakshy E (2020) BoTorch: A framework for efficient Monte-Carlo Bayesian optimization. *Advances in Neural Information Processing Systems* (Curran Associates, Inc.).

Buathong P, Wan J, Astudillo R, Daulton S, Balandat M, Frazier PI (2024) Bayesian optimization of function networks with partial evaluations. *International Conference on Machine Learning* (PMLR).

Chen X, Lin Q, Zhou D (2013) Optimistic knowledge gradient policy for optimal budget allocation in crowdsourcing. *International Conference on Machine Learning* (PMLR).

Frazier PI, Powell WB, Dayanik S (2008) A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization* 47(5):2410–2439.

Frazier PI, Powell WB, Dayanik S (2009) The knowledge-gradient policy for correlated normal beliefs. URL `http://dx.doi.org/10.1287/ijoc.1080.0314.cd`, available for download at https://github.com/INFORMSJoC/1080.0314.

Garnett R (2023) *Bayesian optimization* (Cambridge University Press).

Ginsbourger D, Le Riche R (2010) Towards Gaussian process-based optimization with finite time horizon. *International Workshop in Model-Oriented Design and Analysis* (Springer).

Gupta SS, Miescke KJ (1996) Bayesian look ahead one-stage sampling allocations for selection of the best population. *Journal of statistical planning and inference* 54(2):229–244.

He X, Powell WB (2018) Optimal learning for stochastic optimization with nonlinear parametric belief models. *SIAM Journal on Optimization* 28(3):2327–2359.

He X, Reyes KG, Powell WB (2020) Optimal learning with local nonlinear parametric models over continuous designs. *SIAM Journal on Scientific Computing* 42(4):A2134–A2157.

Lin ZJ, Astudillo R, Frazier PI, Bakshy E (2022) Preference exploration for efficient bayesian optimization with multiple outcomes. *International Conference on Artificial Intelligence and Statistics* (PMLR).

Mes MR, Powell WB, Frazier PI (2011) Hierarchical knowledge gradient for sequential sampling. *Journal of Machine Learning Research* 12(10).

Mockus J (1972) Bayesian methods of search for an extremum. *Avtomatika i Vychislitel'naya Tekhnika (Automatic Control and Computer Sciences)* 6(3):53–62.

Negoescu DM, Frazier PI, Powell WB (2011) The knowledge-gradient algorithm for sequencing experiments in drug discovery. *INFORMS Journal on Computing* 23(3):346–363.

Pearce M, Branke J (2018) Continuous multi-task bayesian optimisation with correlation. *European Journal of Operational Research* 270(3):1074–1085.

Pearce MAL, Poloczek M, Branke J (2022) Bayesian optimization allowing for common random numbers. *Operations Research* 70(6):3457–3472.

Poloczek M, Wang J, Frazier PI (2017) Multi-information source optimization. *Advances in Neural Information Processing Systems* (Curran Associates, Inc.).

Powell WB, Ryzhov IO (2012) *Optimal learning* (John Wiley & Sons).

Ryzhov IO, Powell WB, Frazier PI (2012) The knowledge gradient algorithm for a general class of online learning problems. *Operations Research* 60(1):180–195.

Snoek J, Larochelle H, Adams RP (2012) Practical bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems* (Curran Associates, Inc.).

Wu J, Poloczek M, Wilson AG, Frazier PI (2017) Bayesian optimization with gradients. *Advances in Neural Information Processing Systems* (Curran Associates, Inc.).