Journal of Machine Learning Research () Submitted ; Published

Grouped q-Knowledge Gradient

Zachary Cosenza ZACOSENZA@UCDAVIS.EDU
Department of Chemical Engineering
University of California
Davis, CA 95616, USA

Editor: Zachary Cosenza

Abstract

This short document is mean to demonstrate the idea of an experimental design method which uses the well known multi-point knowledge gradient method combined with a simple optimization constraint meant optimally allocate laboratory resources when an experimental cost function is not learn-able. The result was that this method does not improve over previous methods which do not consider the information content of all information sources.

1. Knowledge Gradient

The goal of this work was to combine the multi-point knowledge gradient acquisition function $qKG(x)$ Frazier (2018) for Bayesian optimization (BO) to quantify the information value of a given group of experiments using Equation (1). Typically, such a group might be parameterized using a cost function $c(x)$ where x is the identifying information of an experimental campaign as seen in Equation (2).

$$qKG(x) = E_n\{\mu^*_{n+1}(x) - \mu^*_n(x)\} \tag{1}$$

$$qKG(x) = E_n\{\frac{\mu^*_{n+1}(x) - \mu^*_n(x)}{c(x)}\} \tag{2}$$

Where μ^*_{n+1} and μ^*_n are the maximums of the posterior (modeled by a Gaussian process) for the next and current set of data points respectively. $\mu^*_n = Y^*$ is the best point found so far, while μ^*_{n+1} is the maximization of a fantasy GP for the next step in the sequential design problem: the inner optimization problem. The goal of this work is to solve the following outer optimization problem for a group of optimal experiments X^*.

$$X^* = argmax\, qKG(x) \tag{3}$$

This is typically done using the one-shot approach Balandat et al. (2019). In short, rather than solving the inner problem and averaging many times to get a single stochastic approximation of $qKG(x)$, a set of N base sequences of outer "fantasy" samples x_f are used in an average over the posterior distribution of the model conditions on the design points x_d that the experimenter will actually use in lab as seen in Equation (4).

$$X(x_d, x_f) = argmax\, 1/N\, \Sigma^N_i\{\mu(x_{f,i}|x) - Y^*\} \tag{4}$$

2. Un-Learnable Design Constraints

In our experiments x there are parameters that are common across experiment types, concentration of a sample or measurement instrument for example $x = \{C_1, C_2, ..Instrument_1\}$. However, samples may be analyzed in many ways. So two samples of the same concentration but measured using two different instruments may be $x_1 = \{C_1, C_2, ..Instrument_1\}$ and $x_2 = \{C_1, C_2, ..Instrument_2\}$. Our purpose here is to determine the optimal shared and not-shared parameters given constraints on the use of our instruments.

In most laboratory settings there is no cost function $c(x)$ by which a set of experiments may be judged against one another on different instruments. Only physical constraints on the type of number of experiments is known (for example we can run N DNA experiments and R protein expression experiments). Thus, using the concept of information-value contained in the multi-point knowledge gradient, we set **two** important constraints on $\{x_d, x_f\}$.

(i) First, because x_f is meant to represent the "next" stage of experimentation after the current design is chosen, we want it to be ran on the "best" instrument $\{0\}$ which gives the greatest level of information. (ii) Next, because x_d is meant to represent the current set of experiments, we constrain them to whatever our lab can handle $\{0, 1, 2\}$. In effect, we are asking the question "what set of experiments, given our lab constraints, will result in the most information for the underlying / best measurement we can make"?

3. Results

I ran simulations of optimization loop against a published method Cosenza et al. (2022) which essentially ignores the information value of non-$\{0\}$ data points. Our toy problems labeled $\{f_1, f_2, f_3, f_4\}$ (10 dimensional) in Figure 1 have an underlying index $\{0\}$ which we want to optimize for and auxiliary / low-fidelity index values $\{1, 2\}$ which imperfectly approximate $\{0\}$. In each of the $B = 10$ batches of the optimization loop (with an initialization using Latin Hypercube of 15 data points where $q_0 = 2$ of them are of type $\{0\}$), $q_0 = 2$ type $\{0\}$ are done and $q - q_0 = 3$ type $\{1, 2\}$ are done. As a note, to mimic more closely our experimental setup, type $\{0\}$ evaluations automatically come with both $\{1, 2\}$ so each batch of experiments is comprised of $q_0 = 2$ high fidelity evaluations and $q = 5$ low fidelity evaluations.

4. Conclusions

The reasons for the lack of improvement are numerous. Perhaps (i) the toy problems were not difficult enough to allow the two BO methods to differentiate themselves. Also (ii) the size ($q = 5$ and dimensionality of 10) of the problem set may not be big enough (or too big) to distinguish the two methods. Additionally, (iii) only 20 total $\{0\}$ high fidelity evaluations were allowed to solve the four toy problems, which may not be enough to illicit improvement in either one of the algorithms. In any case, the code used to generate the images is posted as well.

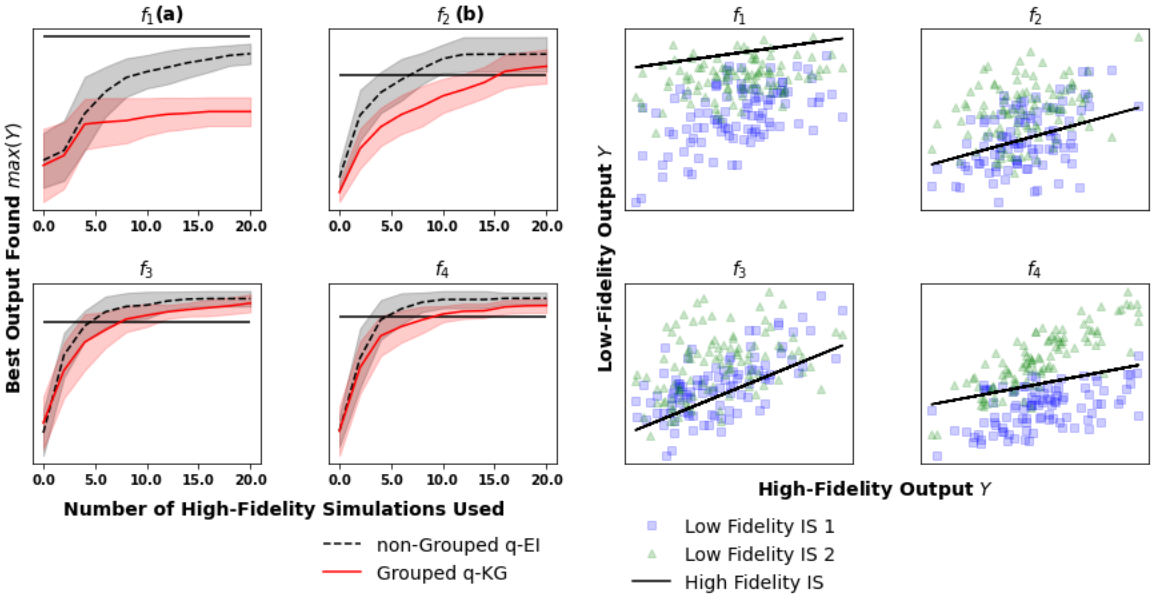

Figure 1: (a) best output for four different simulated multi-information source optimization problems. Horizontal line indicates best point found at fidelity $\{0\}$ from $M = 10^7$ MC random points (b) shows the correlation between the two lower value information sources and the desired output information source for reference.

References

Maximilian Balandat, Brian Karrer, Daniel R. Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. (MC), 2019. URL http://arxiv.org/abs/1910.06403.

Zachary Cosenza, David E Block, Peter I Frazier, and Keith Baar. Multi - information source Bayesian optimization of culture media for cellular agriculture. (April):1–12, 2022. doi: 10.1002/bit.28132.

Peter I. Frazier. A Tutorial on Bayesian Optimization. (Section 5):1–22, 2018. URL http://arxiv.org/abs/1807.02811.