Appendix: Multi-information Source BO

Xubo Yue
University of Michigan

Raed Al Kontar
University of Michigan

1 Formulation

The misoBO is an extension of the BO setting in our main paper. In this misoBO scenario, we have access to several sampling sources and we are interested in deciding both optimal sampling points and sampling sources.

1.1 Setting

We want to solve the unconstrained optimization problem $\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}\in\boldsymbol{\mathcal{X}}} f(\boldsymbol{x})$. Due to limited budget, sampling from the original source is expensive and incurs a cost $c(\boldsymbol{x}) : \boldsymbol{\mathcal{X}} \to \mathbb{R}^+$. Now suppose we have access to I possibly biased auxiliary sources indexed by $\mathcal{I} = \{1, \ldots, I\}$. Each source has a relatively cheap query cost $c_i(\boldsymbol{x}), i \in \mathcal{I}$. When sampling from source $i \in \mathcal{I}$ at point $\boldsymbol{x}$, we observe a noisy and biased outcome $y(i, \boldsymbol{x})$. We assume the observation $y(i, \boldsymbol{x})$ is normally distributed with mean $f(i, \boldsymbol{x})$ and variance $\sigma_i^2(\boldsymbol{x})$. Denote by $\delta_i(\boldsymbol{x}) : \boldsymbol{\mathcal{X}} \to \mathbb{R}$ the bias term and $\delta_i(\boldsymbol{x}) = f(i, \boldsymbol{x}) - f(\boldsymbol{x})$ from each auxiliary source $i \in \mathcal{I}$. We set $\delta_i \sim \mathcal{GP}(0, \Sigma_i(\boldsymbol{x}, \boldsymbol{x}'))$ and $f(\boldsymbol{x}) \sim \mathcal{GP}(\mu_0(\boldsymbol{x}), \Sigma_0(\boldsymbol{x}, \boldsymbol{x}'))$. Therefore, $f(i, \boldsymbol{x})$ is a GP with mean function $\mu(i, \boldsymbol{x})$ and covariance function $\Sigma((i, \boldsymbol{x}), (i', \boldsymbol{x}'))$. Specifically, $\mu(i, \boldsymbol{x}) = \mu_0(\boldsymbol{x}), \Sigma((i, \boldsymbol{x}), (i', \boldsymbol{x}')) = \Sigma_0(\boldsymbol{x}, \boldsymbol{x}') + \mathbb{I}(i, i')\Sigma_i(\boldsymbol{x}, \boldsymbol{x}')$, where $\mathbb{I}(i, i') = 1$ if $i = i'$. Here we note that a mean function (or a constant) can be added to model systematic discrepancy in the bias δ_i (Higdon et al., 2008).

Given data $\boldsymbol{D}_k = \{\boldsymbol{x}_1, y_1, i_1, \ldots, \boldsymbol{x}_k, y_k, i_k\}$, we would like to determine the next sampling duplet $(i_{k+1}, \boldsymbol{x}_{k+1})$ by solving the following optimization problem: $(i_{k+1}, \boldsymbol{x}_{k+1}) := (i^*, \boldsymbol{x}^*) = \arg\max_{(i,\boldsymbol{x})\in(\mathcal{I},\boldsymbol{\mathcal{X}})} Q_k(i, \boldsymbol{x}; \boldsymbol{D}_k)$. After observing the optimal sampling duplet, we augment the current training data $\boldsymbol{D}_k$ with the new observation and obtain $\boldsymbol{D}_{k+1} = \boldsymbol{D}_k \cup \{(\boldsymbol{x}_{k+1}, y_{k+1}, i_{k+1})\}$.

1.2 Dynamic Programming

Denote by $k \in \{1, ..., N\}$. At each stage k, define the state space as $\mathcal{S}_k = (\boldsymbol{\mathcal{X}} \times \mathbb{R} \times \mathcal{I})$ and denote by dataset $\boldsymbol{D}_k := s_k \in \mathcal{S}_k$ the current state, where s_k is the potential state in the state space $\mathcal{S}_k$. A policy $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_N\}$ is a sequence of rules π_k mapping the state space $\mathcal{S}_k$ to the design space $\boldsymbol{\mathcal{X}}$ and sources $\mathcal{I}$. We use $\pi_k^{\boldsymbol{\pi}}$ to emphasize the k^{th} rule under policy $\boldsymbol{\pi}$. Let $\pi_k(\boldsymbol{D}_k) = (\boldsymbol{x}_{k+1}, i_{k+1})$. Now denote by $r_k : \mathcal{S}_k \times \boldsymbol{\mathcal{X}} \times \mathcal{I} \to \mathbb{R}$ the reward function at stage k. Define the end-stage reward as $r_{N+1} : \mathcal{S}_{N+1} \to \mathbb{R}$. The discounted expected cumulative reward of a finite N-step horizon under policy $\boldsymbol{\pi}$ given initial dataset $\boldsymbol{D}_1$ can be expressed as $R^{\boldsymbol{\pi}}(\boldsymbol{D}_1) =$

$$\mathbb{E}\Big[\sum_{k=1}^{N} \alpha^{k-1} r_k(\boldsymbol{D}_k, \boldsymbol{x}_{k+1}, i_{k+1}) + \alpha^N r_{N+1}(\boldsymbol{D}_{N+1})\Big]. \tag{1}$$

In the policy space $\boldsymbol{\Pi}$, we are interested in the optimal policy $\boldsymbol{\pi}^* \in \boldsymbol{\Pi}$ which maximizes Eq. (1). Specifically,

$$R^{\boldsymbol{\pi}^*}(\boldsymbol{D}_1) := \max_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} R^{\boldsymbol{\pi}}(\boldsymbol{D}_1). \tag{2}$$

Based on the Bellman optimality equation, we can formulate (1) as a recursive DP: $R_k(\boldsymbol{D}_k) =$

$$\max_{(i_{k+1},\boldsymbol{x}_{k+1})\in(\mathcal{I},\boldsymbol{\mathcal{X}})} \mathbb{E}[r_k(\boldsymbol{D}_k, \boldsymbol{x}_{k+1}, i_{k+1}) + \alpha R_{k+1}(\boldsymbol{D}_{k+1})], \tag{3}$$

with $R_{N+1}(\boldsymbol{D}_{N+1}) = r_{N+1}(\boldsymbol{D}_{N+1})$. Therefore, the acquisition function is expressed as $Q_k(i_{k+1}, \boldsymbol{x}_{k+1}; \boldsymbol{\mathcal{D}}_k) =$

$$\mathbb{E}\big[r_k(\boldsymbol{D}_k, \boldsymbol{x}_{k+1}, i_{k+1}) + \alpha R_{k+1}(\boldsymbol{D}_{k+1})\big]. \tag{4}$$

1.3 Knowledge Gradient

The reward function at each stage k quantifies the gains of applying rule π_k given state $\boldsymbol{D}_k$. To handle multi-information source BO efficiently, we will adopt a normalized KG as our expected stage-reward function (Ryzhov et al., 2012; Poloczek et al., 2017). Specifically, $\mathbb{E}[r_k(\boldsymbol{D}_k, \boldsymbol{x}_{k+1}, i_{k+1})] =$

$$\mathbb{E}\big[\frac{1}{c_{i_{k+1}}(\boldsymbol{x}_{k+1})}\big(\max_{\boldsymbol{x}'} \mu^{k+1}(0, \boldsymbol{x}') - \max_{\boldsymbol{x}'} \mu^k(0, \boldsymbol{x}')\big)\big]. \tag{5}$$

The first part in the expected KG can be expressed as $\mathbb{E}\big[\max_{\boldsymbol{x}'} \mu^{k+1}(0, \boldsymbol{x}')\big] =$

$$\mathbb{E}\big[\max_{\boldsymbol{x}'}\{\mu^k(0, \boldsymbol{x}') + \sigma^k_{\boldsymbol{x}'}(i, \boldsymbol{x}_{k+1})Z\}\big], \tag{6}$$

where Z is a standard normal random variable and $\sigma^k_{\boldsymbol{x}'}(i_{k+1}, \boldsymbol{x}_{k+1}) =$

$$\frac{\Sigma^k((0, \boldsymbol{x}'), (i_{k+1}, \boldsymbol{x}_{k+1}))}{[\sigma^2_{i_{k+1}}(\boldsymbol{x}_{k+1}) + \Sigma^k((i_{k+1}, \boldsymbol{x}_{k+1}), (i_{k+1}, \boldsymbol{x}_{k+1}))]^{1/2}},$$

such that Σ^k is the posterior covariance function of f given current data $\boldsymbol{D}_k$. Since we are taking expectation with respect to Gaussian random variables, equations (5) and (6) are easy to compute and can be efficiently estimated by the Gauss-Hermite quadrature with n nodes.

1.4 An ADP Formulation

In our DP formulation, each reward-to-go function R_{k+1} is approximated by H_{k+1} using a heuristic (base) policy. Denote by $h (h > 1)$ the rolling horizon and $\tilde{N} = \min\{k + h, N\}$. We then can obtain the approximated reward-to-go functions:

$$\begin{aligned} H_k(\boldsymbol{D}_k) &= \mathbb{E}[r_k(\boldsymbol{D}_k, \tilde{\pi}_k) + \alpha H_{k+1}(\boldsymbol{D}_{k+1})], \\ H_{\tilde{N}}(\boldsymbol{D}_{N+1}) &= r_{\tilde{N}}(\boldsymbol{D}_{N+1}), \end{aligned} \tag{7}$$

where $\tilde{\pi}_k$ is the heuristic policy at every iteration $k \in [\tilde{N}] = \{1, \ldots, \tilde{N}\}$. We define the heuristic policy as

$$(i^*, \boldsymbol{x}^*) = \underset{(i_{k+1}, \boldsymbol{x}_{k+1}) \in \mathcal{I} \times \mathcal{X}}{\arg\max} \mathbb{E}[r_k(\boldsymbol{D}_k, \boldsymbol{x}_{k+1}, i_{k+1})]. \tag{8}$$

This is equivalent to maximizing the greedy KG in (5). At the end stage, we define policy $\tilde{\pi}_{\tilde{N}+1}$ such that

$$\boldsymbol{x}^* = \underset{\boldsymbol{x} \in \mathcal{X}}{\arg\max}\, \mu_0^{\tilde{N}}(\boldsymbol{x}), \tag{9}$$

where $\mu_0^{\tilde{N}}$ is the updated mean function from data $\boldsymbol{D}_{\tilde{N}}$. The last sampling source is chosen to exhaust the remaining budget. Approximated acquisition function can then be solved using gradient-free optimization technique such as bound optimization by quadratic approximation (BOBYQA) Powell (2009). We summarize our DP-misoKG algorithm in **Algorithm 1**.

2 Algorithm

The algorithm for the multi-information source BO is lised in **Algorithm 1**.

Algorithm 1: The Non-myopic Multi-Information Source Bayesian Optimization Algorithm

Data: Initial data $\boldsymbol{D}_1$, budget B and query cost c, c_i, number of remaining evaluations N, number of node n.
Result: Data $\boldsymbol{D}_N$, optimal value $f^{\boldsymbol{D}_N}_{max}$, Gap G.
Fit $\mathcal{GP}$ to data $\boldsymbol{D}_1$ and obtain parameters of bias terms and initial optimal value $f^{\boldsymbol{D}_1}_{max}$;
for $k = 1 : N$ **do**
  if $B - \min_i c_i < 0$ **then**
    Directly return $\boldsymbol{D}_k$ as $\boldsymbol{D}_N$;
    STOP;
  else
    Choose feasible horizon h;
    Select
     $(i_{k+1}, \boldsymbol{x}_{k+1}) = \arg\max_{i \in \mathcal{I}, \boldsymbol{x} \in \boldsymbol{\mathcal{X}}} Q_k(i, \boldsymbol{x}; \boldsymbol{D}_k)$
     s.t. $c_{i_{k+1}}(\boldsymbol{x}_{k+1}) \leq B$;
    $B \leftarrow B - c_{i_{k+1}}(\boldsymbol{x}_{k+1})$;
  end
  Evaluate $f(i_{k+1}, \cdot)$ at $\boldsymbol{x}_{k+1}$ and obtain y_{k+1};
  Augment the dataset
   $\boldsymbol{D}_{k+1} = \boldsymbol{D}_k \cup \{(\boldsymbol{x}_{k+1}, y_{k+1}, i_{k+1})\}$;
  Fit $\mathcal{GP}$ to data $\boldsymbol{D}_{k+1}$;
  $k \leftarrow k + 1$;
end
Fit $\mathcal{GP}$ to data $\boldsymbol{D}_N$;
Obtain optimal value $f^{\boldsymbol{D}_N}_{max}$;
Calculate the Gap G;
Return $\boldsymbol{D}_N$, $f^{\boldsymbol{D}_N}_{max}$ and G.

3 Performance Guarantees

Under the multi-information source setting, the heuristic KG is also sequentially consistent and sequentially improving.

Corollary 1. *The KG algorithm is sequentially consistent and sequentially improving.*

Proof. Remember that state s_k is the dataset $\boldsymbol{D}_k$. Assume KG algorithm starts at a state s_k (i.e., current dataset $\boldsymbol{D}_k$). At each iteration of KG, given a path $(\boldsymbol{D}_k, \boldsymbol{D}_{k+1}, \ldots, \boldsymbol{D}_m)$ and $\boldsymbol{D}_m$ is not the state at the end, the next state $\boldsymbol{D}_{m+1}$ is obtained by solving the acquisition function of KG and augment $\boldsymbol{D}_m$ with $(\boldsymbol{x}^*, y, i^*)$. If $\boldsymbol{D}_{m+1}$ is not the terminating state, the algorithm will start with the path $(\boldsymbol{D}_k, \boldsymbol{D}_{k+1}, \ldots, \boldsymbol{D}_m, \boldsymbol{D}_{m+1})$. Otherwise, the algorithm will terminate with state $\boldsymbol{D}_{m+1}$ and $N = m+1$. We assume that, if there is a tie in the acquisition function, the algorithm will pick up the cheapest source i or the vector $\boldsymbol{x}$ with the smallest elements. Therefore, KG is sequentially consistent.

Let $(\boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_k, \ldots, \boldsymbol{D}_N)$ be the path generated by the rollout starting from $\boldsymbol{D}_1$. Define $\sigma(s)$ as the sub σ-algebra generated by state s. Since KG is sequentially consistent, we have

$$\mathbb{E}\big[\sum_{\ell=k}^{N} r_\ell(s_\ell, \pi_\ell^{\boldsymbol{\pi}_{\mathcal{H}(s)}}(s_\ell))|\sigma(s')\big] = \mathbb{E}\big[\sum_{\ell=k}^{N} r_\ell(s_\ell, \pi_\ell^{\boldsymbol{\pi}_{\mathcal{H}(s')}}(s_\ell))|\sigma(s')\big]. \tag{10}$$

Therefore, the rollout is sequentially improving and we complete our proof. □

Bibliography

D. Higdon, J. Gattiker, B. Williams, and M. Rightley. Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, 103 (482):570–583, 2008.

M. Poloczek, J. Wang, and P. Frazier. Multi-information source optimization. In *Advances in Neural Information Processing Systems*, pages 4288–4298, 2017.

M. J. Powell. The bobyqa algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge*, pages 26–46, 2009.

I. O. Ryzhov, W. B. Powell, and P. I. Frazier. The knowledge gradient algorithm for a general class of online learning problems. *Operations Research*, 60(1):180–195, 2012.