Practical Two-Step Lookahead Bayesian Optimization

Supplementary Material

A Proofs Details

Here we prove the theoretical results in the paper. We first prove Theorem 1.

Proof. We need to prove the interchange of the expectation and the gradient operators are valid. Without loss of generality, we assume $\mathcal{A} = [0,1]^d$. (If this is not true, then we can translate and rescale the domain and corresponding GP.)

We fix X_1 in the interior of $\mathcal{A}^q$. We then choose $i \in [q] = \{1, \ldots, q\}$ representing a point within the first stage of points X_1 and a component $j \in [d] = \{1, \ldots, d\}$ of that point. For $x \in [0,1]$, we then let $X_1(x)$ be X_1, but with componetn j of point i replaced by x.

We also define $\widehat{\text{2-OPT}}(X_1(x), Z)$ to be equal to $\widehat{\text{2-OPT}}(Z)$, but with X_1 replaced by $X_1(x)$, so

$$\widehat{\text{2-OPT}}(X_1(x), Z) = \max(f_0^* - \mu_0(X_1(x)) - C_0(X_1(x))Z))^+ + \Gamma(X_1(x), x_2^*, Z,).$$

With this notation, we re-state the validity of this interchange as the following proposition.

Proposition 1. *Under the conditions of Theorem 1,*

$$\frac{\partial}{\partial x}\textit{2-OPT}(X_1(x)) = \mathbb{E}_0\left[\frac{\partial}{\partial x}\widehat{\textit{2-OPT}}(X_1(x), Z)\right) \tag{6}$$

To prove the result, we use Theorem 1 in L'Ecuyer [1990]. This theorem requires three sufficient conditions be met to ensure (6) is valid: there exists an open neighborhood $\Theta \subset [0,1]$ of x such that

- (i) $\widehat{\text{2-OPT}}(X_1(x), Z)$ is continuous in x over Θ for any fixed Z;
- (ii) $\widehat{\text{2-OPT}}(X_1(x), Z)$ is differentiable in x except on a denumerable set in Θ for any given Z;
- (iii) the derivative of $\widehat{\text{2-OPT}}(X_1(x), Z)$ (when it exists) is uniformly bounded by a random variable $M(Z)$ for all $x \in \Theta$ and the expectation of $M(Z)$ is finite.

A.1 Proof of condition (i)

Because the the mean function μ and the kernel function K are assumed continuous, we see that for any given x, $\mu_0(X_1)$ and $C_0(X_1)$ are continuous in x.

Since the maximum of several continuous functions is continuous, $\max(f_0^* - \mu_0(X_1(x)) - C_0(X_1(x))Z))^+$ is continuous in x.

Since $\Gamma(X_1(x), x_2^*, Z,)$ is continuous in both $X_1(x)$ and x_2 and x_2^* is unique a.s., then $\Gamma(X_1(x), x_2^*, Z)$ is continuous in $X_1(x)$, also x. By definition, $\widehat{\text{2-OPT}}(X_1(x), Z)$ is also continuous in x.

A.2 Proof of condition (ii)

Since $\Gamma(X_1(x), x_2^*, Z,)$ is differentiable in z by the envelope theorem (see Corollary 4 of Milgrom and Segal 2002), then we need to prove $\max(f_0^* - \mu_0(X_1(x)) - C_0(X_1(x))Z))^+ + \Gamma(X_1(x), x_2^*, Z)$ is differentiable except on a denumerable set in Θ for any given $\mathbb{A}$ and Z. By definition, if $\text{argmax}(f_0^* - \mu_0(X_1(x)) - C_0(X_1(x))Z))^+ + \Gamma(X_1(x), x_2^*, Z))$ is unique, then $\widehat{\text{2-OPT}}(X_1(x), Z))$

is differentiable at x. We define $D(\mathbb{A}) \subset \Theta$ to be the set that $\max(f_0^* - \mu_0(X_1(x)) - C_0(X_1(x))Z))^+ + \Gamma(X_1(x), x_2^*, Z)$ is not differentiable, then we see that

$$D(\mathbb{A}) \subset \quad \cup_{i,j\in 1:q} \{x \in \Theta : h_i(x) = h_j(x), \\ \tfrac{dh_x(i)}{dx} \neq \tfrac{dh_j(x)}{dx}\Big\}$$

where $h_i(x) := (f_0^* - \mu_0(X_1(x)) - C_0(X_1(x))Z))_i^+$. Now we only need to show that

$$\left\{x \in \Theta : h_i(x) = h_j(x), \frac{dh_x(i)}{dx} \neq \frac{dh_j(x)}{dx}\right\}$$

is denumerable.

Defining $\eta(x) := h_i(x) - h_j(x)$ on Θ, one can see that $\eta(x)$ is continuous differentiable on Θ. We would like to show that $E := \left\{x \in \Theta : \eta(x) = 0, \frac{d\eta(x)}{dx} \neq 0\right\}$ is denumerable. To prove it, we will show that E contains only isolated points. Then one can use a theorem in real analysis: any set of isolated points in $\mathbb{R}$ is denumerable (see the proof of statement 4.2.25 on page 165 in Thomson et al. [2008]). To prove that E only contains isolated points, we use the definition of an isolated point: $y \in E$ is an isolated point of E if and only if $x \in E$ is not a limit point of E. We will prove by contradiction, suppose that $y \in E$ is a limit point of E, then it means that there exists a sequence of points $y_1, y_2, \cdots$ all belong to E such that $\lim_{n\to\infty} y_n = x$. However, by the definition of derivative and

$$\eta(y_n) = \eta(x) = 0$$
$$0 \neq \tfrac{d\eta(y)}{dy}\Big|_{y=x} = \lim_{n\to\infty} \tfrac{\eta(y_n)-\eta(x)}{y_n - x} = \lim_{n\to\infty} 0 = 0,$$

a contradiction. So we conclude that E only contains isolated points, so is denumerable.

A.3 Proof of condition (iii)

We first prove that $\frac{\partial}{\partial x}\max(f_0^* - \mu_0(X_1(x)) - C_0(X_1(x))Z))^+$ is bounded as below

$$\begin{aligned} & \frac{\partial}{\partial x}\max(f_0^* - \mu_0(X_1(x)) - C_0(X_1(x))Z))^+ \\ \leq \quad & \left|\frac{\partial}{\partial x}\mu_0(X_1(x))\right| \\ & + \left|\frac{\partial}{\partial x}C_0(X_1(x))\}\right| |Z| \end{aligned}$$

Given that $\left|\frac{\partial}{\partial x}\mu_0(X_1(x))\right|$ and $\left|\frac{\partial}{\partial x}C_0(X_1(x))\}\right|$ is bounded and $\mathbb{E}_n(|Z|)$ is finite, we get that $\mathbb{E}_n\Big(\frac{\partial}{\partial x}\max(f_0^* - \mu_0(X_1(x)) - C_0(X_1(x))Z))^+\Big)$ is bounded.

Now we proceed to prove that $\frac{\partial}{\partial x}\Gamma(X_1(x), x_2^*, Z,)$ is bounded as below

$$\begin{aligned} & \frac{\partial}{\partial x}\Gamma(X_1(x), x_2^*, Z,) \\ = \quad & \frac{\partial}{\partial x}\Gamma_n(z^*, Z, z^{1:q}) \end{aligned}$$

by the envelope theorem. Given that $\frac{\partial}{\partial x}\Gamma_n(z^*, Z, z^{1:q})$ is continuous in Z and $z^{1:q}$, so it is bounded. □

We now prove Theorem 2. We denote the gradient estimator as $G(Z_t)$, so

$$G(Z_t) = \nabla(y_n^* - y_{n+q}^*) + \nabla\Gamma_n(z^*, Z, z^{1:q})$$

Proof. We prove this theorem using Theorem 2.3 of Section 5 of Kushner and Yin [2003], which depends on the structure of the stochastic gradient G of the objective function.

The theorem from Kushner and Yin [2003], requires the following hypotheses:

1. $\epsilon_t \to 0$, $\sum_{t=1}^{\infty} \epsilon_t = \infty$, and $\sum_t \epsilon_t^2 < \infty$.
2. $\sup_t E\left[|G(Z_t)|^2\right] < \infty$
3. There exist uniformly continuous functions $\{\lambda_t\}_{t\geq 0}$ of Z, and random vectors $\{\beta_t\}_{t\geq 0}$, such that $\beta_t \to 0$ almost surely and

$$E_n\left[G(Z_t)\right] = \lambda_t(Z_t) + \beta_t.$$

Furthermore, there exists a continuous function $\bar{\lambda}$, such that for each $Z \in A^q$,

$$\lim_n \left| \sum_{i=1}^{m(r_m+s)} \epsilon_i \left[\lambda_i(Z) - \bar{\lambda}(Z)\right] \right| = 0$$

for each $s \geq 0$, where $m(r)$ is the unique value of k such that $t_k \leq t < t_{k+1}$, where $t_0 = 0$,$t_k = \sum_{i=0}^{k-1} \epsilon_i$.

4. There exists a continuously differentiable real-valued function ϕ, such that $\bar{\lambda} = -\nabla\phi$ and it is constant on each connected subset of stationary points.
5. The constraint functions defining $\mathbb{A}$ are continuously differentiable.

We now prove that our problem satisfy these conditions.

(1) is true by hypothesis of the theorem.

Let's prove (2). We have shown above that

$$E\left[|G(Z_t)|^2\right] \leq c \times \mathbb{E}_n(Z^2)$$

then is bounded.

We now prove (3). For each t, define

$$\lambda_t(Z) := \bar{\lambda} := E_n\left[G(Z_t)\right]$$

Let's prove that λ_t is continuous by noting that $\mathbb{E}_n(\nabla(y_n^* - y_{n+q}^*))$ and $\mathbb{E}_n(\nabla\Gamma_n(z^*, Z, z^{1:q}))$ are both continuous. By defining $\beta_t = 0$ for all t, and $\bar{\lambda} = \lambda_1$, we conclude the proof of (3).

Finally, define $\phi(Z) = -E\left[Q_n(Z)\right]$. Observe that in Theorem 1, we show that we can interchange the expectation and the gradient in $E\left[\nabla Q_n(Z)\right]$, and so $\lambda_m(Z) = -\nabla\phi(Z)$. In a connected subset of stationary points, we have that $\lambda_m(Z) = 0$, and so $\phi(Z)$ is constant. This ends the proof of the theorem. $\square$