

複数のノイズ分布におけるロバスト性を考慮したベイズ最適化 Bayesian Optimization Considering Robustness under Multiple Noise Distributions

伊藤 凜[†] 増井 秀之[†]
Rin Ito Hideyuki Masui

1. はじめに

ベイズ最適化は目的関数の評価コストが高く、観測値にノイズが乗る場合にも有効なブラックボックス最適化として知られ、機械学習やシステム制御など幅広い分野でパラメータ調整に用いられている。最も一般的なベイズ最適化アルゴリズムでは、出力値の観測ノイズは考慮されるものの、入力値の不確実性については考慮されない。しかしシステム制御のパラメータ調整など、実適用先によっては入力値に無視できないノイズが乗り、その結果として出力値が変動する可能性がある。このような課題に対し、入力ノイズを考慮に入れたベイズ最適化手法が提案されている[1, 2]。これらの既存研究の多くは入力ノイズがある単一の確率分布に従うものと仮定を置いている。しかし、例えば制御パラメータの最適化において装置の仕様を絞り切れていないなど、複数の入力ノイズ分布のそれぞれに対しロバストな解を探索したいケースが考えられる。そこで本稿では、複数の入力ノイズ分布を考慮し、各ノイズ分布に対しロバストな解を探索するベイズ最適化手法を提案する。また、数値実験により提案手法の有効性を実証する。

2. 入力ノイズにロバストな最適化

2.1 ロバスト目的関数

一般的なベイズ最適化では、入力空間 \mathcal{X} における未知の目的関数 $f: \mathcal{X} \rightarrow \mathbb{R}$ について、ノイズ $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ の含まれた観測値 $y = f(x) + \epsilon$ に基づき最適解を探索する。本稿では最適化実施後、得られた最適解を用いてシステムを運用する段階において、入力 x にランダムなノイズ ξ が加わることを考える。この場合、入力ノイズによる出力値の変動は小さいほうが好ましい。次の式で表されるロバスト目的関数 $g(x)$ の最大化を行うことで、ノイズによる変動の小ささと本来の目的関数値の大きさを両立した解が得られる。

$$g(x) = \mathbb{E}_p[f(x + \xi)] = \int f(x + \xi)p(\xi)d\xi,$$

ここで、 p は入力ノイズ ξ の分布である。

2.2 入力ノイズを考慮したガウス過程回帰

ベイズ最適化では一般的に、ガウス過程回帰により目的関数 $f(x)$ をモデル化する。

$$\hat{f}(x) \sim \mathcal{GP}(m(x), v(x, x')),$$

ここで、 $m(x)$ は目的関数の予測平均、 $v(x, x')$ は予測分散である。このとき、ロバスト目的関数 $g(x)$ もまたガウス過程としてのモデル化が可能である。

$$\hat{g}(x) \sim \mathcal{GP}(m_g(x), v_g(x, x')),$$

$$m_g(x) = k_{gf}(x)^T(K + \sigma_\epsilon^2 I)^{-1}y_t,$$

$$v_g(x, x') = k_g(x, x') - k_{gf}(x)^T(K + \sigma_\epsilon^2 I)^{-1}k_{gf}(x'),$$

ここで目的関数を t 回評価した後に得られているデータセ

ットを $D_t = (\mathbf{X}_t, \mathbf{y}_t)$ とし、 $[\mathbf{X}_t]_i = x_i$ 、 $[\mathbf{y}_t]_i = y_i$ 、 $[K]_{ij} = k(x_i, x_j)$ 、 $[k_{gf}(x)]_i = k_{gf}(x, x_i)$ とした。ただし k はカーネル関数 $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ である。 k_g および k_{gf} は、入力ノイズに関してカーネル関数 k の期待値を取ったものである。

$$k_g(x, x') = \iint k(x + \xi, x' + \xi')p(\xi, \xi')d\xi d\xi',$$

$$k_{gf}(x, x') = \int k(x + \xi, x')p(\xi)d\xi.$$

カーネル関数が RBF カーネルであり、かつ入力ノイズ分布が正規分布であるとき、 k_g および k_{gf} を解析的に計算することができる[3]。

2.3 複数の入力ノイズ分布におけるロバスト性

本稿では、複数の入力ノイズ分布 $p_m \in P = \{p_1, \dots, p_M\}$ のそれぞれに対するロバスト最適解 $x_m^* = \arg \max_{x \in \mathcal{X}} g_m(x)$ を得ることを目的とする。ここで、 $g_m(x)$ は入力ノイズ分布 p_m に対するロバスト目的関数である。各 p_m に対する最適化性能を総合して評価するため、各ロバスト目的関数の累積最良値 $h_m(\mathbf{X}_t)$ の和 $h(\mathbf{X}_t)$ を評価指標として用いる。

$$h(\mathbf{X}_t) = \sum_m h_m(\mathbf{X}_t),$$

$$h_m(\mathbf{X}_t) = \max_{x \in \mathcal{X}_t} g_m(x).$$

3. 提案手法

本稿では、 $t + 1$ 回目に目的関数を評価する点 x_{t+1} を \mathbf{X}_t に加えたとき、2.3節で述べた評価指標 $h(\mathbf{X}_t \cup x_{t+1})$ がより高くなるような x_{t+1} を選ぶための獲得関数を提案する。まず、次の式で示す獲得関数を考える。

$$\begin{aligned} \alpha(x_+) &= \mathbb{E}[h(\mathbf{X}_t \cup x_+) - h(\mathbf{X}_t)] \\ &= \sum_m \mathbb{E}[h_m(\mathbf{X}_t \cup x_+) - h_m(\mathbf{X}_t)] \\ &= \sum_m \mathbb{E}[\max(g_m(x_+) - h_m(\mathbf{X}_t), 0)]. \end{aligned}$$

$g_m(x_+)$ および $h_m(\mathbf{X}_t)$ はデータセット D_t から直接値を得ることができない。そこで、これらの代わりにガウス過程回帰による予測等を用いた、次式に示す獲得関数を提案する。

$$\alpha_{\text{ours}}(x_+) = \sum_m \mathbb{E}_{y_+} [\max(m_{g_m|(x_+, y_+)}(x_+) - \hat{h}_m(\mathbf{X}_t), 0)],$$

$$\hat{h}_m(\mathbf{X}_t) = \max_{x \in \mathcal{X}_t} m_{g_m}(x).$$

まず $h_m(\mathbf{X}_t)$ については、ロバストガウス過程回帰による予測平均 $m_{g_m}(x)$ の最大値 $\hat{h}_m(\mathbf{X}_t)$ で置き換える。また $g_m(x_+)$ の代わりに用いている $m_{g_m|(x_+, y_+)}(x_+)$ は、入力 x_+ における観測値 y_+ がデータセット D_t に加わったときの、ロバストガウス過程回帰による予測平均であり、次式で与えられる。

$$\begin{aligned} m_{g_m|(x_+, y_+)}(x_+) &= m_{g_m}(x_+) \\ &\quad - A(v(x_+, x_+) + \sigma_\epsilon^2)^{-1}(m(x_+) - y_+), \\ A &= k_{gf}(x, x_+) - k_{gf}(x)^T(K + \sigma_\epsilon^2 I)^{-1}k(x_+). \end{aligned}$$

ここで y_+ がガウス過程回帰による予測分布に従うものとする。

$$y_+ = f(x_+) + \epsilon \sim \mathcal{N}(m(x_+), v(x_+, x_+) + \sigma_\epsilon^2).$$

[†]三菱電機株式会社 Mitsubishi Electric Corporation

このとき、ロバストガウス過程回帰による予測平均もまた、正規分布に従う。

$$m_{g_m|(x_+, y_+)}(x) \sim \mathcal{N}(m_{g_m}(x), A(v(x_+, x_+) + \sigma_\epsilon^2)^{-1}A^T).$$

$g_m(x_+)$ の代わりとして、新たな観測値をデータセットに追加したときの予測平均ではなく、現在のデータセットに基づく予測分布を用いることも可能である。この場合、獲得関数は一般的によく知られている Expected Improvement (EI) と似た形をとる。一般的なガウス過程回帰では、既に評価済みの入力値 $x \in \mathbf{X}_t$ において予測分散 $v(x, x)$ が小さくなり、その結果次に評価する点として選ばれにくくなる。ロバストガウス過程回帰の予測分散 $v_{g_m}(x, x)$ は、入力ノイズが加わった $(x + \xi)$ における目的関数値の不確か性が加味される。したがって既に評価済みの入力値においても予測分散が十分に小さい値とならず、次に評価する点として繰り返し採用されることで、最適化の効率が下がる可能性がある。このため本稿では前述の獲得関数の形を採用した。

4. 数値実験

提案手法の有効性を検証するため、数値実験により既存手法との比較を行う。目的関数には 2 次元空間におけるガウス混合モデル [2] を用い、入力ノイズ分布は $p_m = \mathcal{N}(\mathbf{0}, \sigma_m^2 I)$ 、分布数は $M = 5$ 、入力ノイズの分散値は $\sigma_m^2 = (0.05 * (m - 1))^2$ とした。 $m = 1$ は $\sigma_1^2 = 0$ 、すなわち入力ノイズが無い条件である。 $m = 1, 3, 5$ におけるロバスト目的関数 $g_m(x)$ を図 1 に示す。また観測ノイズは $\epsilon = 0$ とした。比較手法には EI、uGP-EI、uGP-UCB [1]、NES [2] を用いた。uGP-EI および uGP-UCB では入力ノイズを考慮したガウス過程回帰を代理モデルとして用い、獲得関数に EI または UCB (Upper Confidence Bound) を用いた。NES は入力ノイズを考慮したエントロピーベースの手法である。uGP-EI、uGP-UCB、および NES では、考慮する入力ノイズの分布は $p_3 = \mathcal{N}(\mathbf{0}, 0.1^2)$ とした。

数値実験結果を図 2 に示す。横軸は目的関数の評価回数 t 、縦軸は 2.3 節で述べた評価指標 $h(\mathbf{X}_t)$ である。評価 2 回目まではランダムに選択した入力値を用いている。それぞれの手法についてランダムシードを変えて 10 回実験を行い、その平均と標準偏差を図に示した。縦軸の値 $h(\mathbf{X}_t)$ は、各入力値 x_i とサンプルしたノイズ値 ξ に対し $f(x_i + \xi)$ を評価し、 $g(x_i)$ の近似値を得ることで算出した。これらはあくまで実験結果を評価するためのみに取得した値であり、提

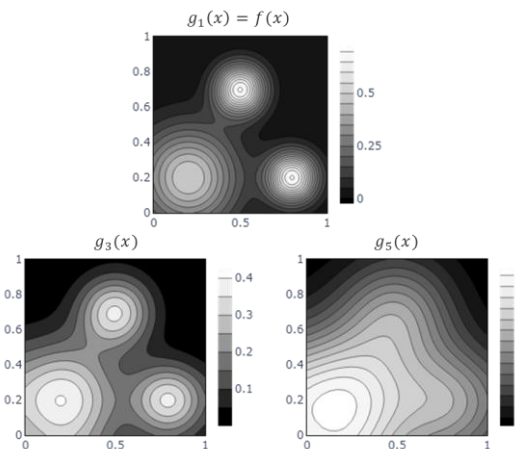


図 1 各入力ノイズ分布 p_m に関するロバスト目的関数 $g_m(x)$ 。 $g_1(x)$ は本来の目的関数 $f(x)$ にあたる。

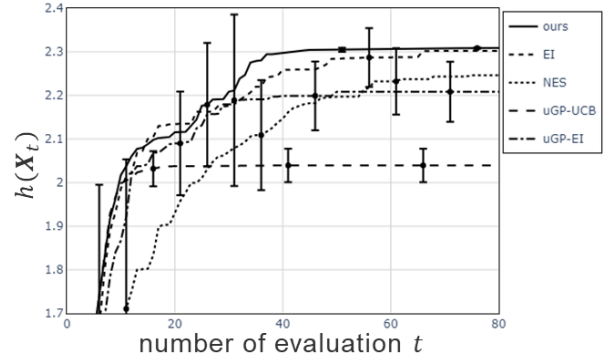


図 2 各手法の比較結果。目的関数の評価回数 t に対する評価指標 $h(\mathbf{X}_t)$ の平均と標準偏差を示す。

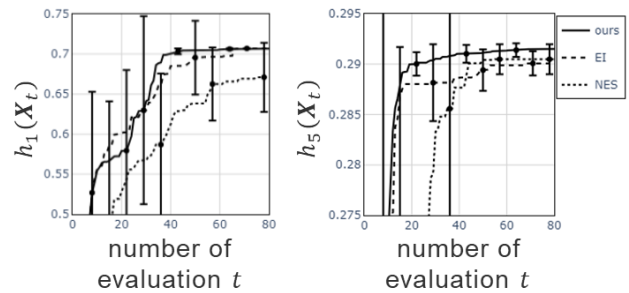


図 3 入力ノイズ無し (左) と入力ノイズ大 (右) に対するロバスト目的関数値の累積最良値 $h_m(\mathbf{X}_t)$ 。

案手法や他の手法は $y_i = f(x_i)$ のみを用いている。図 2 の結果より、提案手法が他の手法に比べ、 $h(\mathbf{X}_t)$ の値を効率的に上げていることが分かる。特に評価 50 回目以降は標準偏差も小さく、安定して良い結果が得られている。提案手法、EI、NES の三手法について、 p_1 及び p_5 に関するロバスト目的関数値の累積最良値 $h_m(\mathbf{X}_t)$ を図 3 に示す。提案手法では各入力ノイズ分布に対しロバストな入力値を効率的に探索できていることが分かる。また $h_1(\mathbf{X}_t)$ は入力ノイズを考慮しない本来の目的関数値の最良値にあたるが、その推移においても提案手法による高い効率が確認できる。これは入力ノイズを考慮したロバスト目的関数に対する最適化を並行して行うことにより、本来の目的関数の最適化において局所解に陥りにくくなる効果が得られたものと考えられる。

5. まとめ

複数の入力ノイズ分布を考慮し、各ノイズ分布のそれぞれに対しロバストな解を探索するベイズ最適化手法を提案した。また数値実験により、提案手法の有効性を示した。

参考文献

- [1] Rafael Oliveira, Lionel Ott, and Fabio Ramos. “Bayesian optimisation under uncertain inputs.” *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, Vol. 89, pp.1177-1184. (2019).
- [2] Fröhlich, Lukas, et al. “Noisy-input entropy search for efficient robust Bayesian optimization.” *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, Vol. 108, pp. 2262-2272. (2020).
- [3] Patrick Dallaire, Camille Besse, and Brahim Chaïb-Draa. “Learning Gaussian process models from uncertain data.” *Proceedings of the International Conference on Neural Information Processing (ICONIP)*, pp. 433-440. (2009).