Supplementary Materials "Bayesian Optimization via Exact Penalty"

Jiangyan Zhao[1] and Jin Xu[1,2,*]

[1]School of Statistics, East China Normal University, Shanghai, China

[2]Key Laboratory of Advanced Theory and Application in Statistics and Data Science – MOE, East China Normal University, Shanghai, China

[*]Corresponding author: School of Statistics, East China Normal University, 3663 North Zhongshan Road, Shanghai 200062, China, email: jxu@stat.ecnu.edu.cn

Supplementary Materials

The supplement is structured as follows. Section SM§1 demonstrates the prediction performance of the proposed novel surrogate model for the equality constraint violation. Section SM§2 presents the implementation details of the seven comparators. Section SM§3 provides the visualizations and mathematical definitions of the synthetic test problems in the main text. Section SM§4 explains the distributional behaviors of design points obtained by various methods in a single run. Section SM§5 examines the discrepancy between the closed-form and the Monte Carlo ScaledEI acquisition function. Section SM§6 presents the sensitivity analysis of EPBO concerning the choice of kernel function and the size of the initial design. Section SM§7 provides additional empirical comparisons and discussions regarding various real-world challenges, including (i) disjoint feasible regions, (ii) challenges in GP modeling, and (iii) high dimensionality.

SM§1 Novel Surrogate Model for the Equality Constraint Violation

Figure SM.1 provides the prediction performance of the proposed novel surrogate model concerning the equality constraint violation $|h(x)|$, where $h(x) = \exp(-1.4x)\cos(7\pi x/2)$. Panels (b)-(d) illustrate that the predictive mean function (dashed line) approaches the true constraint violation (solid line) with the associated confidence interval shrinking to zero as the sample size increases. Panels (f)-(h) show the performance of the usual vanilla surrogate model (Gramacy et al., 2016). The latter is clearly inferior to the former, particularly in the neighborhood of the roots of $h(x) = 0$. This discrepancy is attributed to the kinks in the surface introduced by the absolute value operator. The results are consistent with what is seen in the inequality case in Figure 1.

SM§2 Implementation Details for Comparators

The `laGP` package in R (Gramacy, 2016) provides implementations for ALBO (Gramacy et al., 2016), Slack-ALBO (Picheny et al., 2016), and EIC (Schonlau et al., 1998). The `BoTorch` package in Python (Balandat et al., 2020) offers implementations for SCBO (Eriksson and Poloczek, 2021). Its official usage tutorial is available at `https://botorch.org/tutorials/scalable_constrained_bo`.

Next, we briefly review the asymmetric entropy (AE) method (Lindberg and Lee, 2015) and the barrier method (BM) (Pourmohamad and Lee, 2022). Let $p(\boldsymbol{x})$ denote the probability of point $\boldsymbol{x}$ satisfying all constraints (or being feasible).

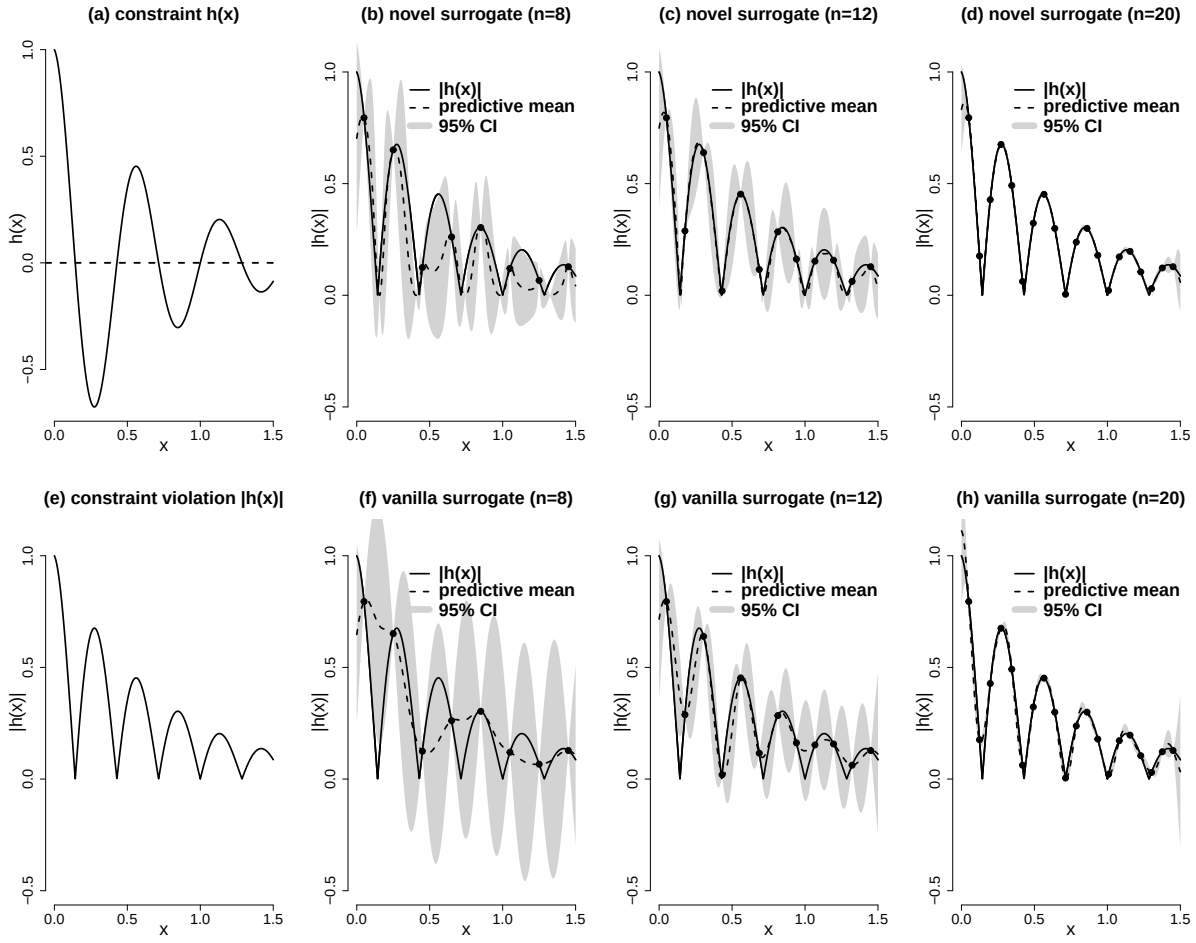

Figure SM.1: (a): True equality constraint function $h(x) = \exp(-1.4x)\cos(7\pi x/2)$; (b)–(d): the proposed novel surrogate model based on 8, 12 and 20 equally spaced samples in [0.05, 1.45], respectively, where the solid line is the true constraint violation, the dashed line is the predictive mean of the surrogate model, and the shaded areas are the corresponding 95% confidence intervals (CI) given by $\mu(x) \pm 1.96\sigma(x)$; (e): true equality constraint violation function $|h(x)|$; (f)–(h): counterparts of (b)–(d) obtained by the vanilla surrogate model.

Given the preference of finding an optimum on the constraint boundary, Lindberg and Lee (2015) proposed the asymmetric entropy (AE) acquisition function as

$$\mathrm{AE}(\boldsymbol{x}) = \mathrm{EI}(\boldsymbol{x})^{\alpha_1} \times S_a(\boldsymbol{x})^{\alpha_2},$$

where$S_a(\boldsymbol{x})$ is asymmetric entropy defined by

$$S_a(\boldsymbol{x}) = \frac{2p(\boldsymbol{x})\{1-p(\boldsymbol{x})\}}{p(\boldsymbol{x}) - \omega p(\boldsymbol{x}) + \omega^2},$$

ω is a mode location parameter, α_1 and α_2 are specified weights for EI and S_a, respectively.

Lindberg and Lee (2015) recommends $\omega = 2/3$, $\alpha_1 = 1$, and $\alpha_2 = 5$ for practical use. If no feasible solution is found in the sequential design, then $y_{\min}$ in $\mathrm{EI}(\boldsymbol{x})$ is not defined. In such a case, we replace it with the 90th percentile of the values of the objective as suggested by Gramacy (2016).

To force the sequential design to stay inside the feasible region as much as possible, Pourmohamad and Lee (2022) introduced the barrier method into the Bayesian optimization paradigm and proposed the BM acquisition function as

$$\mathrm{BM}(\boldsymbol{x}) = -\mu_f(\boldsymbol{x}) + \sigma_f^2(\boldsymbol{x}) \sum_{j=1}^{J} \left[\log\left\{-\mu_{c_j}(\boldsymbol{x})\right\} + \frac{\sigma_{c_j}^2(\boldsymbol{x})}{2\mu_{c_j}^2(\boldsymbol{x})} \right]. \tag{SM.1}$$

When the feasible space is rather small compared to the entire design space, the initial design may not include any feasible point, resulting in an undefined logarithmic function in (SM.1). In this case, we modify the BM acquisition function as

$$\mathrm{BM}^{+}(\boldsymbol{x}) = \begin{cases} \mathrm{BM}(\boldsymbol{x}), & \text{if } S \neq \varnothing, \\ p(\boldsymbol{x}), & \text{if } S = \varnothing. \end{cases}$$

where S is the set of feasible solutions.

As AE and BM do not have packages publicly available so far, we mimic the setting of `laGP` and develop a new R package to implement them, along with the proposed EPBO.

SM§3 Details of the Three Synthetic Test Problems

The first test problem is the inequality constrained problem from Gramacy (2016) with the "Herbie's tooth" objective (Lee et al., 2011), a sinusoidal inequality constraint and a quadratic inequality constraint, given by

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathcal{B}} \quad & f(\boldsymbol{x}) = -t(4x_1 - 2)t(4x_2 - 2) \quad \text{with} \\ & t(z) = \exp\{-(z-1)^2\} + \exp\{-0.8(z+1)^2\} - 0.05\sin\{8(z+0.1)\}, \\ \text{s.t.} \quad & g_1(\boldsymbol{x}) = 1.5 - x_1 - 2x_2 - 0.5\sin\{2\pi(x_1^2 - 2x_2)\} \leqslant 0, \\ & g_2(\boldsymbol{x}) = x_1^2 + x_2^2 - 1.5 \leqslant 0, \end{aligned}$$

where $\mathcal{B} = [0,1]^2$. It is referred to as the HSQ problem. The optimal solutions are $\boldsymbol{x}_1^* = (0.2397, 0.7842)^\top$ and $\boldsymbol{x}_2^* = (0.7842, 0.2397)^\top$ with $f(\boldsymbol{x}_1^*) = f(\boldsymbol{x}_2^*) = -1.0934$. See more details in the `demo("ALfhat")` of the `laGP` package. This problem is illustrated in panel

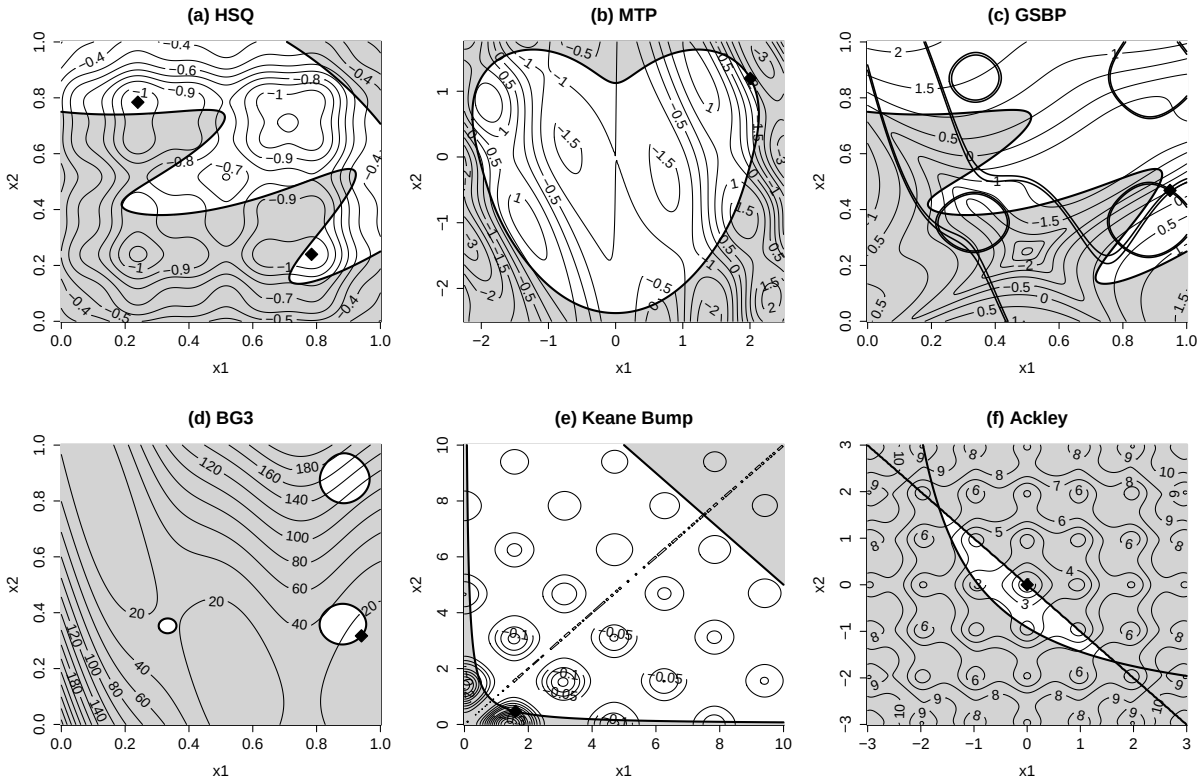

Figure SM.2: Graphical illustration of the six synthetic test problems considered in Section 4.1 and Section SM§7, where the diamonds indicate the locations of the global minimum, the contour lines show the values of the objective function, the shaded regions represent the infeasible spaces ruled out by the inequality constraints. In panel (c), the equality constraints are shown with two lines to represent the tolerance.

(a) of Figure SM.2.

The second synthetic test problem is the modified Townsend problem (MTP) from Pourmohamad and Lee (2022) given by

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathcal{B}} \quad & f(\boldsymbol{x}) = -\left[\cos\{(x_1-0.1)x_2\}\right]^2 - x_1\sin(3x_1+x_2), \\ \text{s.t.} \quad & g(\boldsymbol{x}) = x_1^2 + x_2^2 - \left\{2\cos(t) - \frac{1}{2}\cos(2t) - \frac{1}{4}\cos(3t) - \frac{1}{8}\cos(4t)\right\}^2 - 4\sin^2(t) \leqslant 0, \end{aligned}$$

where $t = \arctan(x_1/x_2)$ and $\mathcal{B} = [-2.25, 2.5]\times[-2.5, 1.75]$. The global optimum is $f(\boldsymbol{x}^*) = -2.0239884$ at $\boldsymbol{x}^* = (2.0052938, 1.1944509)^\top$. This problem is illustrated in Panel (b) of Figure SM.2.

The third synthetic test problem consists of the Goldstein-Price (rescaled and centered) objective from Picheny et al. (2013) with a sinusoidal inequality constraint from Gramacy (2016), a Branin (centered) equality constraint from Picheny et al. (2013), and an equality constraint

from Parr et al. (2012). It is referred to as the GSBP problem. The mathematical expression is given as follows.

$$
\begin{aligned}
\min_{\boldsymbol{x}\in\mathcal{B}} \quad f(\boldsymbol{x}) =& \frac{\log[\{1+a(\boldsymbol{x})\}\{30+b(\boldsymbol{x})\}-8.693]}{2.427}, \quad \text{with} \\
a(\boldsymbol{x}) =& \left(4x_1+4x_2-3\right)^2 \times \\
& \{75-56\left(x_1+x_2\right)+3\left(4x_1-2\right)^2+6\left(4x_1-2\right)\left(4x_2-2\right)+3\left(4x_2-2\right)^2\}, \\
b(\boldsymbol{x}) =& \left(8x_1-12x_2+2\right)^2 \times \\
& \{-14-128x_1+12\left(4x_1-2\right)^2+192x_2-36\left(4x_1-2\right)\left(4x_2-2\right)+27\left(4x_2-2\right)^2\}, \\
\text{s.t.} \quad g(\boldsymbol{x}) =& 1.5-x_1-2x_2-0.5\sin\{2\pi(x_1^2-2x_2)\} \leqslant 0, \\
h_1(\boldsymbol{x}) =& 15-\left\{15x_2-\frac{5}{4\pi^2}(15x_1-5)^2+\frac{5}{\pi}(15x_1-5)-6\right\}^2 \\
& -10\left(1-\frac{1}{8\pi}\right)\cos(15x_1-5)=0, \\
h_2(\boldsymbol{x}) =& 4-\left\{4-2.1(2x_1-1)^2+\frac{(2x_1-1)^4}{3}\right\}(2x_1-1)^2-(2x_1-1)(2x_2-1) \\
& -16(x_2^2-x_2)(2x_2-1)^2-3\sin\{12(1-x_1)\}-3\sin\{12(1-x_2)\}=0,
\end{aligned}
$$

where $\mathcal{B}=[0,1]^2$. The optimal solution is $\boldsymbol{x}^*=(0.9477263, 0.4685515)^\top$ with $f(\boldsymbol{x}^*)=-0.5270189$, $g(\boldsymbol{x}^*)=-0.26378$, $h_1(\boldsymbol{x}^*)=-3.831842\times10^{-7}$ and $h_2(\boldsymbol{x}^*)=5.397704\times10^{-6}$. See more details in the `demo("GSBP")` of the `laGP` package. This problem is illustrated in Panel (c) of Figure SM.2.

SM§4 Comparison of Distributional Behaviour of Design Points by Various Methods in a Single Run for the HSQ Problem

To provide additional insights into the behavior of the EPBO, AE, Slack-ALBO, and SCBO methods, Figure SM.3 graphically presents the distribution of their respective design points in a single run with the same initial design. AE primarily assigns its design points near constraint boundaries, as intended by its preference. EPBO, on the other hand, adopts a more balanced way by exploring both boundaries and interior space to identify the global optimum, reflecting its characteristic of global exploration. However, both Slack-ALBO and SCBO appear to have the risk of over-exploitation in the upper right corner of the design space. This is caused by the small difference between the local (-1.0609) and global (-1.0934) optimal values and the large distance between the local and global optimal points.

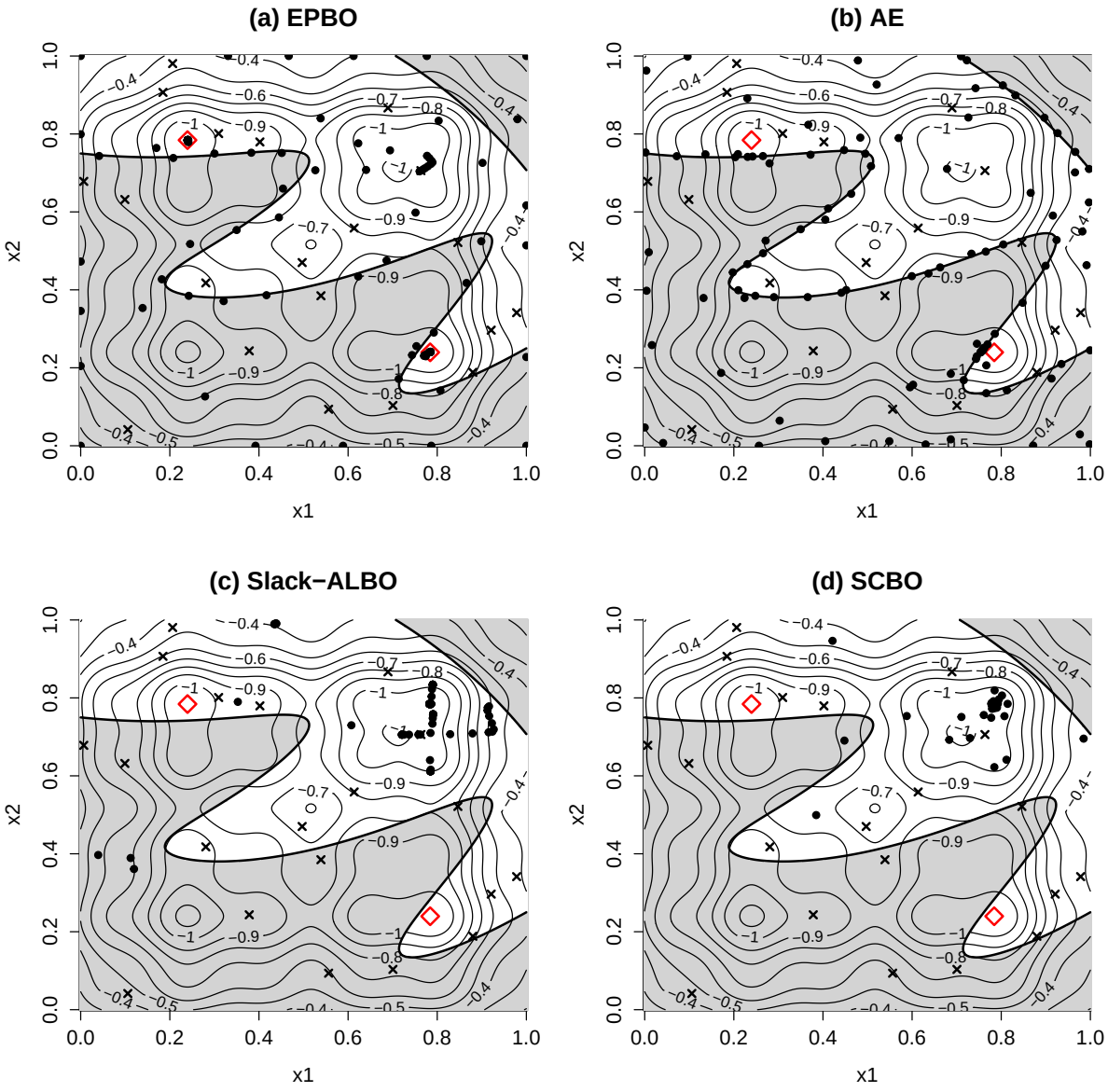

Figure SM.3: Graphical illustration of the distributions of the design points (in solid dots) obtained by EPBO, AE, Slack-ALBO, and SCBO in a single run for the HSQ problem, where the diamonds indicate the locations of the global minimum, the contour lines show the values of the objective function, and the shaded regions represent the infeasible spaces. The initial 20 LHD points are indicated with crosses.

SM§5 Monte Carlo ScaledEI Acquisition Function

As explained in Section 3, the ScaledEI acquisition function derived from model (5) lacks an analytical expression. However, a numerical approximation can be easily obtained using the Monte Carlo (MC) method (Gramacy et al., 2016).

Table SM.1: Average runtimes (in second) spent in 100 repetitions by EPBO with closed-form acquisition function and MC acquisition function using 100 or 1000 samples for the HSQ and GSBP problems

Problem	Closed Form	MC ($T = 100$)	MC ($T = 1000$)
HSQ	8.33	30.53	191.18
GSBP	23.92	31.85	163.30

Suppose $y_f^{(t)}(\boldsymbol{x}), y_{g_j}^{(t)}(\boldsymbol{x}), y_{h_\ell}^{(t)}(\boldsymbol{x}), t = 1, \dots, T, j = 1, \dots, J, \ell = 1, \dots, L$ are sampled from the GP models $Y_f(\boldsymbol{x})|\mathcal{D}_n \sim \mathcal{N}(\mu_f(\boldsymbol{x}), \sigma_f^2(\boldsymbol{x}))$, $Y_{g_j}(\boldsymbol{x})|\mathcal{D}_n \sim \mathcal{N}(\mu_{g_j}(\boldsymbol{x}), \sigma_{g_j}^2(\boldsymbol{x}))$, and $Y_{h_\ell}(\boldsymbol{x})|\mathcal{D}_n \sim \mathcal{N}(\mu_{h_\ell}(\boldsymbol{x}), \sigma_{h_\ell}^2(\boldsymbol{x}))$, respectively. Then, the penalty improvement function at the t-th sample can be expressed as

$$I_p^{(t)}(\boldsymbol{x}) = \max\left[0, y_{\min} - \left\{ y_f^{(t)}(\boldsymbol{x}) + \sum_{j=1}^{J} \rho_{g_j} \max\{0, y_{g_j}^{(t)}(\boldsymbol{x})\} + \sum_{\ell=1}^{L} \rho_{h_\ell} |y_{h_\ell}^{(t)}(\boldsymbol{x})| \right\}\right],$$

and the ScaledEI acquisition function can be approximated by

$$\widehat{\mathbb{E}}\{I_p^{(t)}(\boldsymbol{x})\}/[\widehat{\mathbb{V}}\{I_p^{(t)}(\boldsymbol{x})\}]^{1/2}, \tag{SM.2}$$

where $\widehat{\mathbb{E}}\{I_p^{(t)}(\boldsymbol{x})\} = T^{-1}\sum_{t=1}^{T} I_p^{(t)}(\boldsymbol{x})$ and $\widehat{\mathbb{V}}\{I_p^{(t)}(\boldsymbol{x})\} = (T-1)^{-1}\sum_{t=1}^{T}[I_p^{(t)}(\boldsymbol{x}) - \mathbb{E}_t\{I_p^{(t)}(\boldsymbol{x})\}]^2$ are the sample mean and the sample variance, respectively.

Figure SM.4 compares the performance of EPBO using the closed-form acquisition function ScaledEI (8) and the MC version (SM.2) for the HSQ problem. The log utility gap, defined by the logarithmic difference between the searched and theoretical BFOVs (Picheny et al., 2016), is employed instead of the BFOV to offer a more lucid visualization of the discrepancy in performance. It is seen in Panel (a) that EPBO with the closed-form acquisition function shows steady convergence toward the theoretical value, and EPBO with the MC acquisition function appears to stagnate after about 90 iterations. Panel (b) demonstrates that the former achieves a higher quality of solutions than the latter with a large number of samples. Figure SM.5 compares the two acquisition functions for the GSBP problem. The closed-form one shows clear superiority in both the average progress and the proportion of finding feasible solutions. It is noted that the MC acquisition function with 1000 samples performs comparable to that of Slack-ALBO in Figure 4.

Table SM.1 reports the average runtimes spent in 100 repetitions by EPBO with the close-form acquisition function and MC acquisition function using 100 or 1000 samples for the HSQ and GSBP problems. (The computations were performed on an Intel(R) Core(TM) i7-9750H-based laptop running R 4.3.1.) Depending on the complexity of the problem, the MC approach

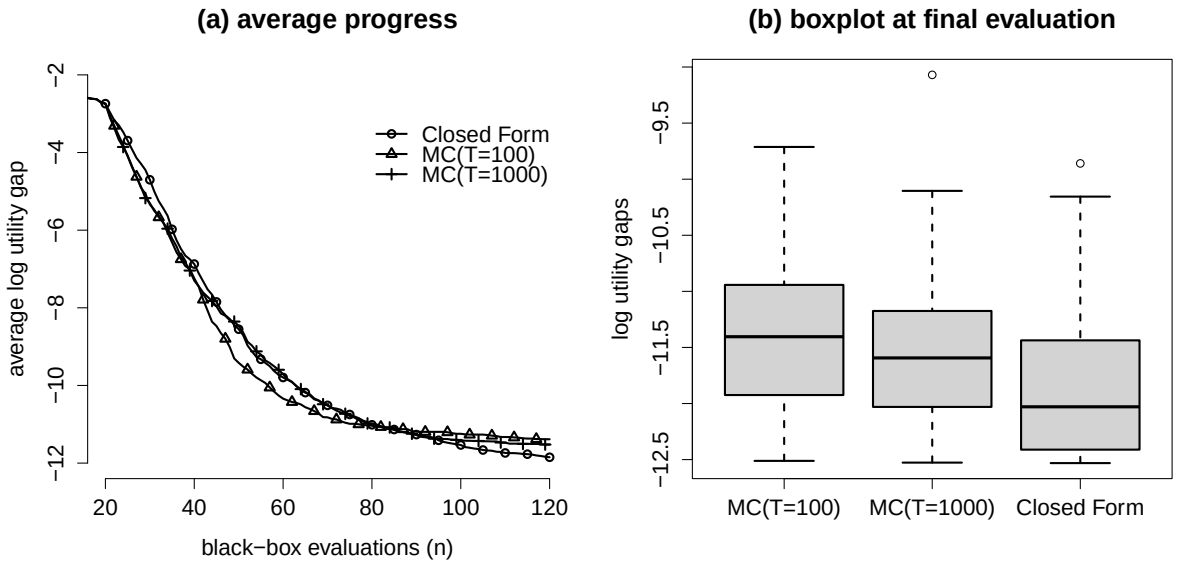

Figure SM.4: Comparison results of EPBO using different acquisition functions for the HSQ problem. (a) the average progress in the log utility gap over 120 iterations, and (b) boxplots of the log utility gaps at the final evaluation.

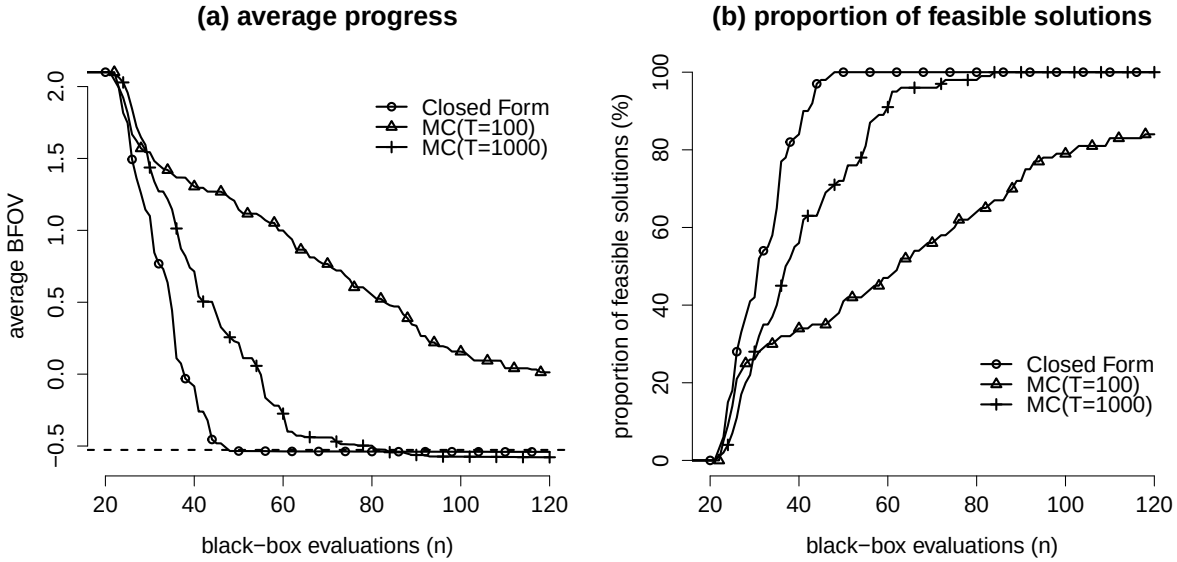

Figure SM.5: Comparison results of EPBO over iterations using different acquisition functions for the GSBP problem. (a) average progress of BFOVs, where the horizontal dashed line indicates the value of the global optimum, and (b) proportions of feasible solutions.

needs more than seven to 20 times the computation cost to obtain comparable performance to the closed-form approach.

SM§6 Sensitivity Analysis

SM§6.1 Choice of Kernel Function

We first examine the performance of EPBO with respect to different choices of kernel function. As the `laGP` package (Gramacy, 2016) used in the comparison in the main text only provides the squared exponential (SE) kernel, we will employ the `DiceKriging` package of R (Roustant et al., 2012) to assess different kernel options.

Next, we briefly review the Matérn kernel, another type of kernel function that is commonly used in BO algorithms. It is given by

$$k_\nu(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\sqrt{2\nu} r)^\nu K_\nu(\sqrt{2\nu} r),$$

where $r = \sqrt{\sum_{i=1}^d (x_i - x_i')^2 / \ell_i}$ is the weighted Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$, σ and ℓ_i are positive parameters, Γ is the gamma function, and K_ν is a modified Bessel function of the second kind with smoothness parameter ν. Here we focus on the two most used cases of matern32 with $\nu = 3/2$ and matern52 with $\nu = 5/2$, given respectively by

$$\begin{aligned} k_{3/2}(\boldsymbol{x}, \boldsymbol{x}') &= \sigma^2 (1 + \sqrt{3} r) \exp(-\sqrt{3} r), \\ k_{5/2}(\boldsymbol{x}, \boldsymbol{x}') &= \sigma^2 \left(1 + \sqrt{5} r + \frac{5}{3} r^2\right) \exp(-\sqrt{5} r). \end{aligned}$$

Note that when ν increases to infinity, the Matérn kernel coincides with the SE kernel, $k_{\mathrm{SE}}(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \exp(-r^2/2)$. For a comprehensive survey of kernel functions, please refer to Rasmussen and Williams (2006, Chapter 4), Gramacy (2020, Chapter 5), and Garnett (2023, Chapter 3).

Figures SM.6 and SM.7 show the performance of EPBO using the three kernel functions considered for the HSQ and GSBP problems. The same indices as before are used. It is seen that the results are generally comparable. The matern32 or matern52 kernel performs slightly better than the SE kernel. It is of interest to investigate further how to choose the kernel function in Bayesian optimization.

SM§6.2 Size of Initial Design

Second, we assess the sensitivity of EPBO with respect to the size of the initial design and/or the starting at infeasible points. We will evaluate the performance under the sizes of d, $2d$, and $5d$, in contrast to the $10d$ rule adopted in the main text.

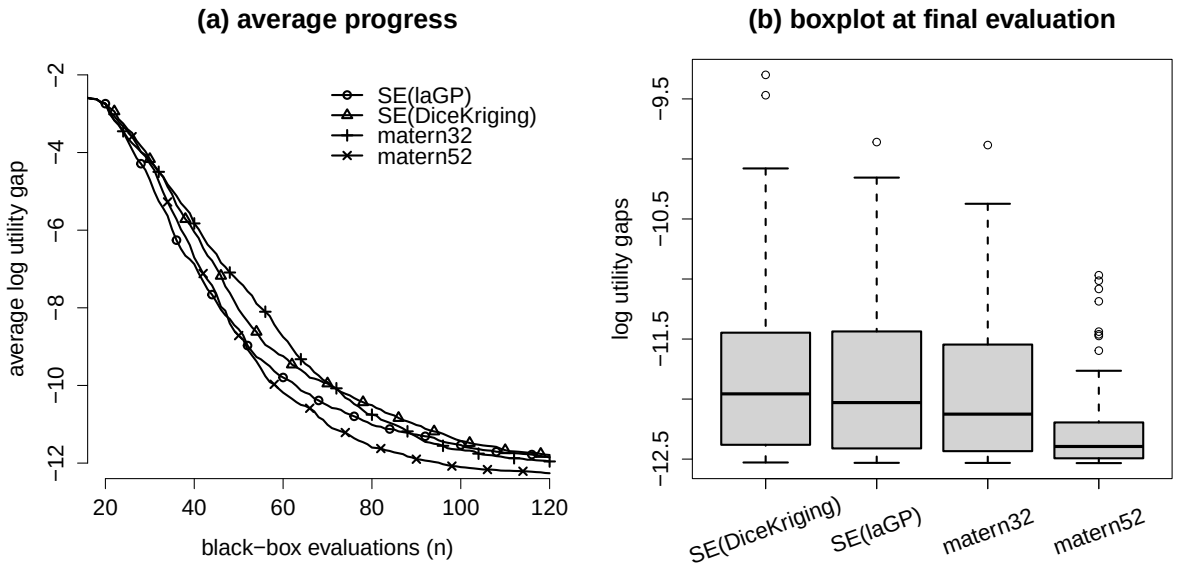

Figure SM.6: Comparison results of EPBO using different kernel functions for the HSQ problem with the same captions as in Figure SM.4.

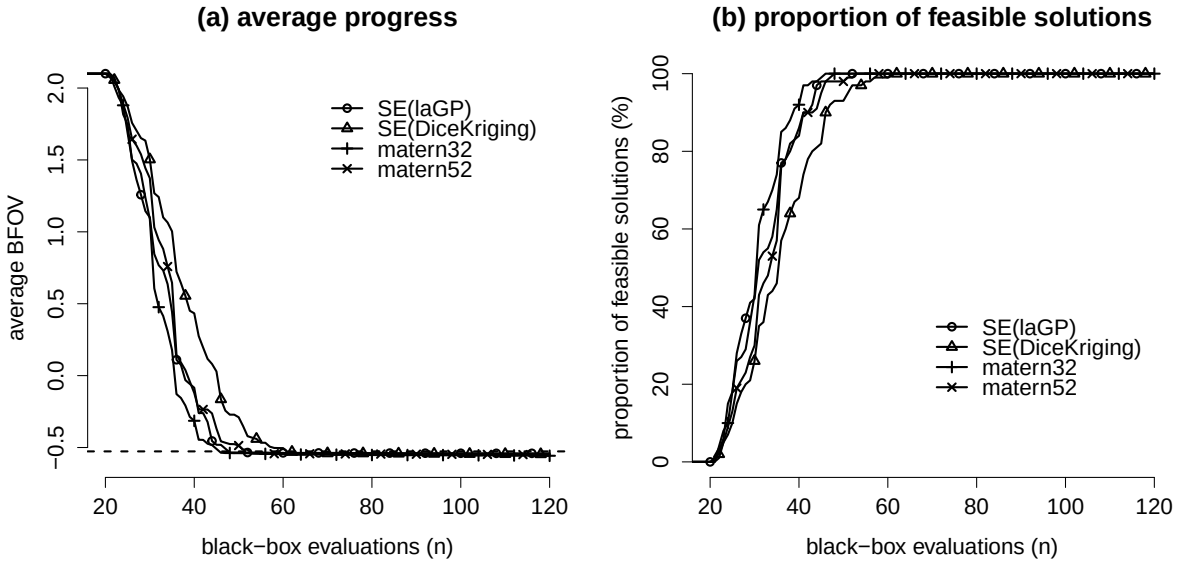

Figure SM.7: Comparison results of EPBO using different kernel functions for the GSBP problem with the same captions as in Figure SM.5.

Figures SM.8 and SM.9 show the comparison results for the HSQ and GSBP problems, respectively. For the HSQ problem, as the feasible region occupies only 45.7% of the design space, the initial designs of sizes d and $2d$ have 19% and 3% chances, respectively, of not providing any feasible points. Meanwhile, the initial designs of sizes $5d$ and $10d$ contain feasible points. For the GSBP problem, none of the initial designs considered provide any feasible

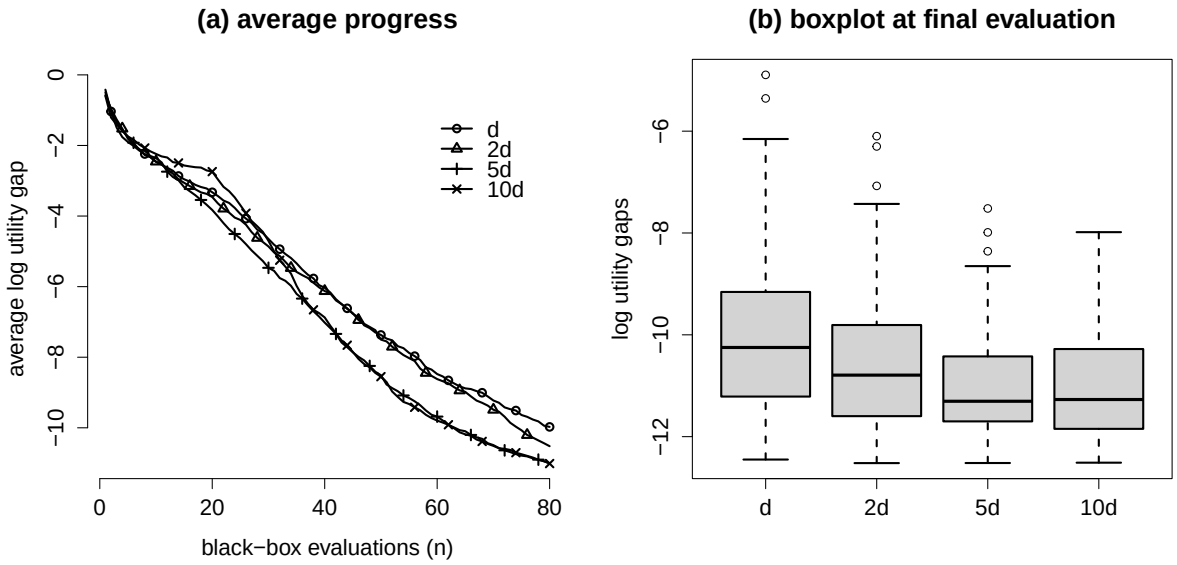

Figure SM.8: Comparison results of EPBO using different sizes of the initial design for the HSQ problem with the same captions as in Figure SM.4.

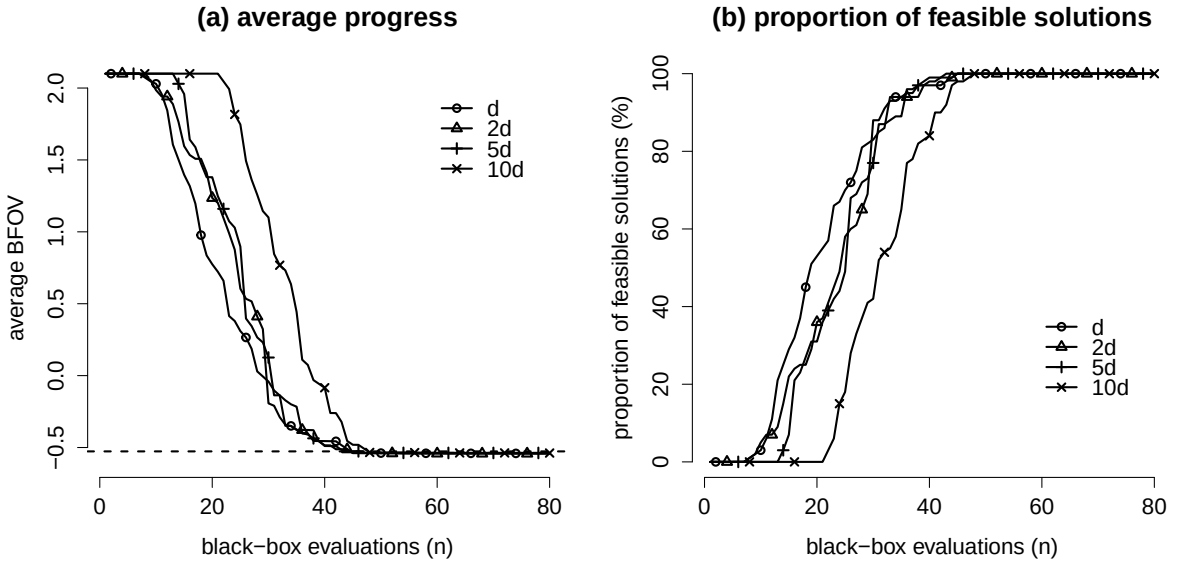

Figure SM.9: Comparison results of EPBO using different sizes of the initial design for the GSBP problem with the same captions as in Figure SM.5.

point. Given these startings from infeasibility points, EPBO still produces outstanding performance on both problems. It is of interest to note that the $5d$ rule surprisingly outperforms the $10d$ rule-of-thumb that we use in the main text.

SM§7 Additional Comparison under Various Scenarios

In this section, we provide additional comparisons of the considered methods under various scenarios of challenges that arise in reality. They include (i) disjoint feasible regions, (ii) challenges in GP modeling, and (iii) high dimensionality. We will not include the results of the reference random search method here due to its inefficiency.

SM§7.1 Problem with Disjoint Feasible Regions

The first test problem, considered by Parr et al. (2012), involves minimizing a modified version of the Branin function under the constraint posed by the Gomez#3 function. It is expressed as

$$\begin{aligned}\min_{\boldsymbol{x}\in\mathcal{B}} \quad & f(\boldsymbol{x}) = \left(a_2 - \frac{5.1}{4\pi^2} + \frac{5}{\pi}a_1 - 6\right)^2 + 10\left\{\left(1 - \frac{1}{8\pi}\right)\cos a_1 + 1\right\} + \frac{5a_1 + 25}{15} \\ \text{s.t.} \quad & g(\boldsymbol{x}) = 6 - \left[\left(4 - 2.1b_1^2 + \frac{1}{3}b_1^4\right)b_1^2 + b_1b_2 + (-4 + 4b_2^2)b_2^2 + 3\sum_{i=1}^{2}\sin\{6(1 - b_i)\}\right] \leqslant 0,\end{aligned}$$

where $a_1 = 15x_1 - 5, a_2 = 15x_2, b_1 = 2x_1 - 1, b_2 = 2x_2 - 1$, and $\mathcal{B} = [0, 1]^2$. The global optimum occurs at $\boldsymbol{x}^* = (0.9406044, 0.3171484)^\top$ with $f(\boldsymbol{x}^*) = 12.005$. The problem is referred to as the BG3 problem and is illustrated in Panel (d) of Figure SM.2. Besides the highly non-linear nature of both the objective function and the constraint, the feasible regions are disjoint, which causes additional complexity for the optimization and makes the local minimum points more deceptive.

Figure SM.10 compares the methods for the BG3 problem. Panel (a) shows that EPBO and SCBO perform the best two in the average progress of the log utility gap. Panel (b) indicates that the solution quality of EPBO is superior to ALBO, AE, and EIC. Slack-ALBO, BM, and SCBO result in the local suboptimal solutions that fall into the feasible regions in the lower left corner in panel (d) of Figure SM.2 (where the log utility gap is approximately equal to 2.16, indicated by the lower horizontal dashed line) and the upper right corner in panel (d) of Figure SM.2 (where the log utility gap is approximately equal to 4.55, corresponding to the upper horizontal dashed line). Slack-ALBO suffers the most, with more than a 50% chance of ending with some local suboptimal solution.

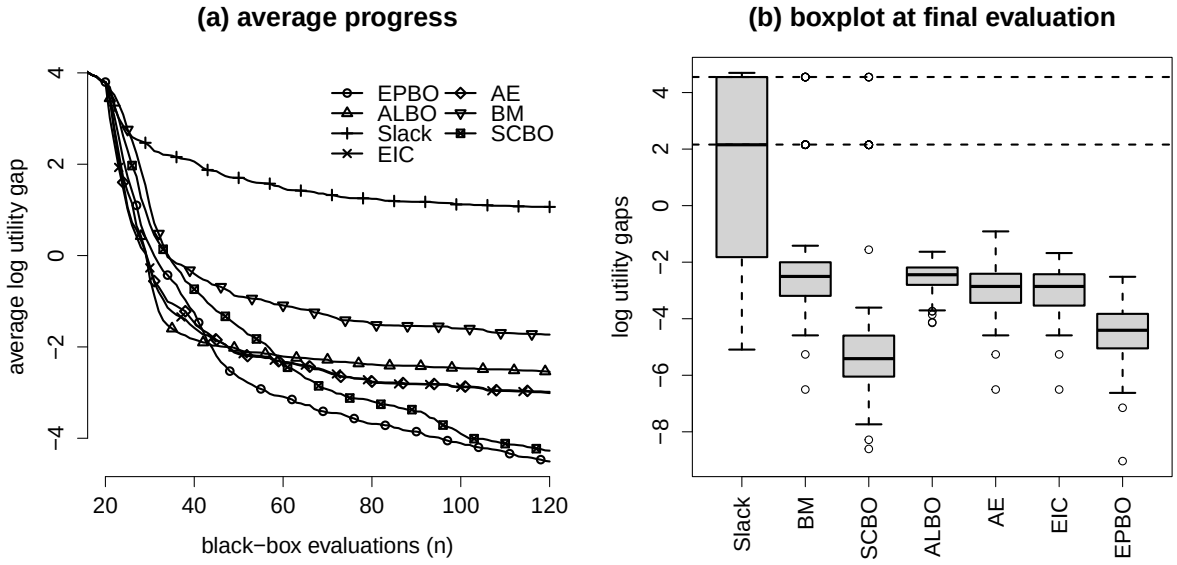

Figure SM.10: Comparison results for the BG3 problem. (a) the average progress in the log utility gap over 120 iterations and (b) boxplots of the log utility gap at the final evaluation, where the upper and lower horizontal dashed lines correspond to the suboptimal solutions located in the upper right corner and lower left corner in panel (d) of Figure SM.2, respectively.

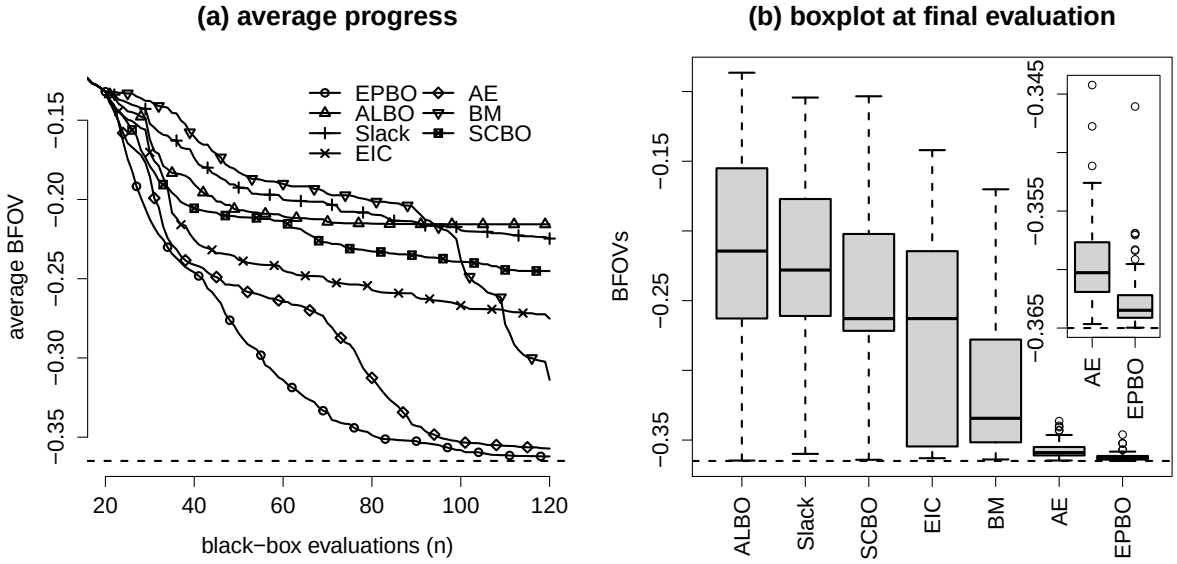

Figure SM.11: Comparison results for the Keane Bump problem. (a) average progress in BFOVs over 150 iterations, where the horizontal dashed line indicates the value of the global optimum, and (b) boxplots of the BFOVs at the final evaluation, where the sub-window displays the boxplots of AE and EPBO in a refined scale.

SM§7.2 Problems with Challenges in GP Modeling

The second test problem is the Keane Bump problem introduced by Keane (1994), which is expressed as

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathcal{B}} \quad & f(\boldsymbol{x}) = -\left|\frac{\sum_{i=1}^{d}\cos^4(x_i) - 2\prod_{i=1}^{d}\cos^2(x_i)}{\sqrt{\sum_{i=1}^{d} i x_i^2}}\right|, \\ \text{s.t.} \quad & g_1(\boldsymbol{x}) = 0.75 - \prod_{i=1}^{d} x_i \leqslant 0, \quad g_2(\boldsymbol{x}) = \sum_{i=1}^{d} x_i - 7.5d \leqslant 0, \end{aligned}$$

where $\mathcal{B} = [0, 10]^d$. In this case, we consider $d = 2$ with the global optimum taking place at $\boldsymbol{x}^* = (1.60086, 0.468498)^\top$ with $f(\boldsymbol{x}^*) = -0.3649799$. This problem is illustrated in Panel (e) of Figure SM.2 and is known for its difficulty in GP modeling (Eriksson and Poloczek, 2021).

Figure SM.11 shows the comparison of the competing methods over 120 iterations. Both indices of average progress and boxplot at the final evaluation show that EPBO and AE are the best two, with EPBO's superior performance seen in a refined scale. The IQR of BFOVs at the final evaluation by EPBO is 2.6×10^{-3}, whereas the counterpart by the runner-up AE is 6×10^{-3}, indicating the better characteristic of EPBO in robustness to the initial design.

The third test problem uses the well-known Ackley function (accessible at `https://www.sfu.ca/~ssurjano/ackley.html`) as the objective function and shares the same type of constraints as the Keane Bump problem. It is expressed as

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathcal{B}} \quad & f(\boldsymbol{x}) = -20\exp\left(-0.2\sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^{d}\cos(2\pi x_i)\right) + 20 + \exp(1), \\ \text{s.t.} \quad & g_1(\boldsymbol{x}) = 2.5^d - \prod_{i=1}^{d}(x_i + 3) \leqslant 0, \quad g_2(\boldsymbol{x}) = \sum_{i=1}^{d} x_i \leqslant 0, \end{aligned}$$

where $\mathcal{B} = [-3, 3]^d$. The global optimum takes place at $\boldsymbol{x}^* = (0, \ldots, 0)^\top$ with $f(\boldsymbol{x}^*) = 0$. The case with $d = 2$ is illustrated in Panel (f) of Figure SM.2. It is seen that the valleys on the surface of the Ackley function are dense and gradually decrease to the origin. (A 3-dimensional graph can be viewed at `https://en.wikipedia.org/wiki/File:Ackley%27s_function.pdf`.) Thus, the trust region strategy (Eriksson et al., 2019; Eriksson and Poloczek, 2021) is particularly suitable for solving this type of problem.

Panel (a) of Figure SM.12 compares the average progress of the competing method across 120 iterations. Unlike previous problems, SCBO demonstrates its superior performance over others, as this Ackley problem aligns well with the strengths of the trust region strategy em-

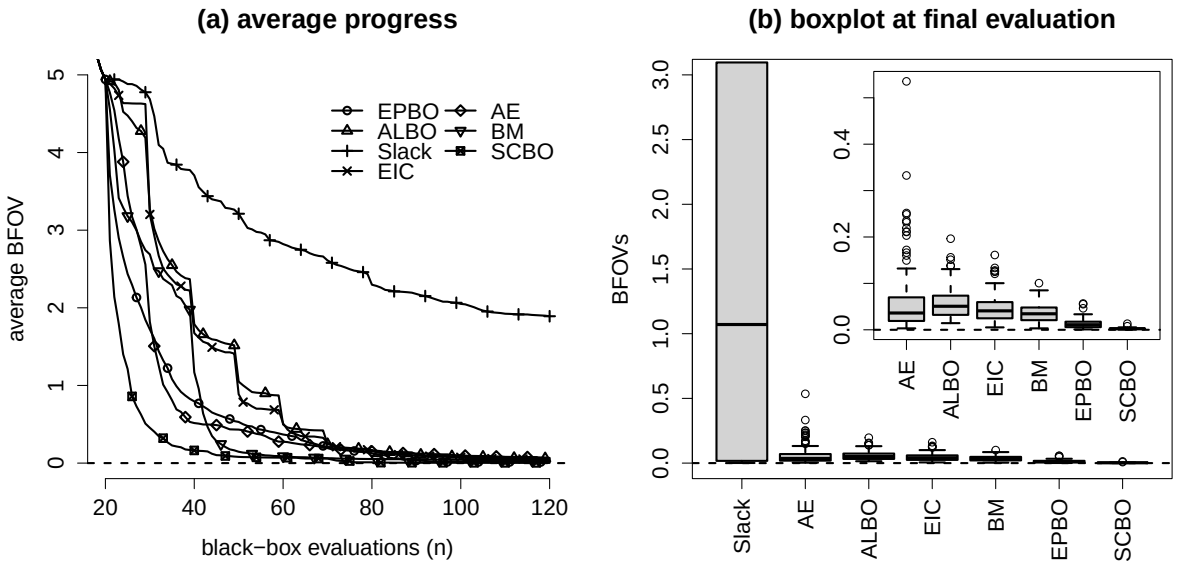

Figure SM.12: Comparison results for the two-dimensional Ackley problem. (a) the average progress in BFOVs over 120 iterations, where the horizontal dashed line indicates the value of the global optimum, and (b) boxplots of the BFOVs at the final evaluation.

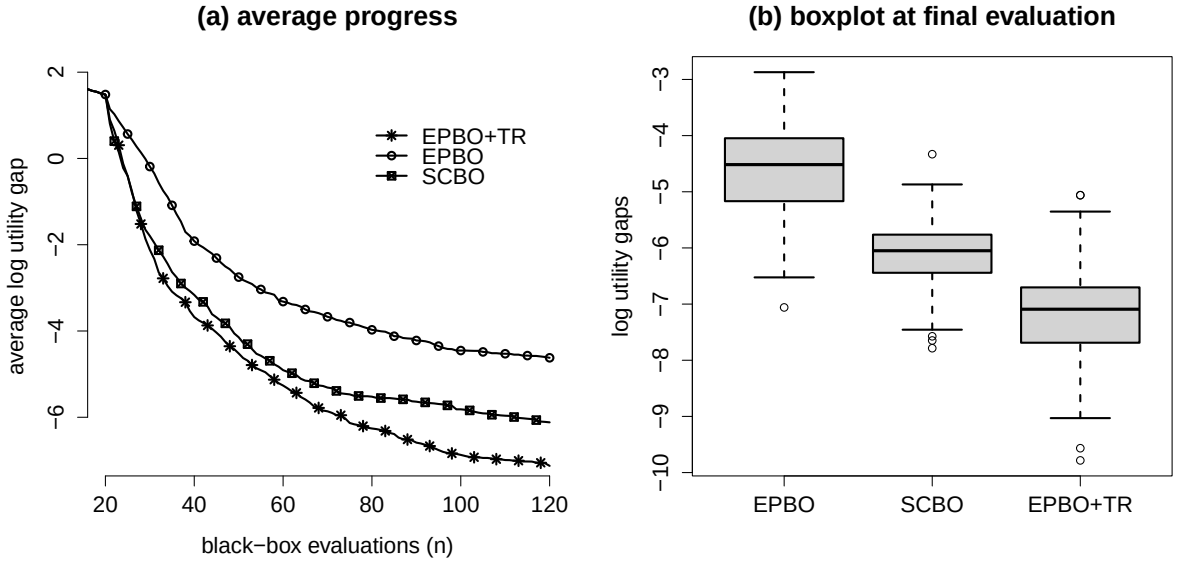

Figure SM.13: Additional comparison results of EPBO, SCBO, and EPBO+TR for the two-dimensional Ackley problem in terms of log utility gap.

ployed by SCBO. Slack-ALBO performs the worst, as it has large (greater than 50%) chance of getting stuck in local minimum points. This is caused by extra numerical integration and approximation during its GP modeling of the Ackley function. Panel (b) shows consistent performance in terms of the boxplot of the BFOVs at the final evaluation. The IQR of BFOVs

by the best-performing SCBO is 0.0015, in contrast to 0.0117 of the runner-up EPBO, which further indicates the advantage of the trust region strategy.

SM§7.3 Problems of High Dimensionality

In Bayesian optimization, problems involving a large number of variables, e.g., with a dimension exceeding 15, present particular challenges (Eriksson and Poloczek, 2021; Binois and Wycoff, 2022; Santoni et al., 2023). The complexity arises not only from the difficulty in estimating the GP surrogate models but also from the increased intricacy of optimizing the acquisition function (Binois and Wycoff, 2022). Various strategies have been proposed to tackle these challenges, including but not limited to variable selection, additive model, linear and nonlinear embedding, and trust region (Santoni et al., 2023). Santoni et al. (2023) showed empirically that the trust region strategy appears to be the most promising one. For a more comprehensive survey and detailed comparison of these strategies, we refer to the works of Binois and Wycoff (2022) and Santoni et al. (2023).

As the trust region strategy is complementary to our proposed EPBO method, it is natural to combine the two, called EPBO+TR.

Before getting to the high-dimensional Ackley problem (e.g., $d = 20$), it is instructive to assess the performance of EPBO+TR at its lower-dimensional scenario of $d = 2$. It is seen in Figure SM.13 that EPBO+TR greatly improves EPBO in convergence speed and solution quality and outperforms SCBO. However, it should be emphasized that, akin to SCBO, EPBO+TR may exhibit over-exploitation, an inherent disadvantage of the trust region strategy. In unconstrained scenarios, Diouane et al. (2023) proposed alternating between standard global steps and local steps within the trust region to avoid local optima. Employing this strategy to constrained scenarios seems a promising avenue for future research.

Now we consider the Ackley problem with dimension $d = 20$. We set the size of the initial design to be $n_0 = 50$. (The $10d$ rule-of-thumb (Loeppky et al., 2009) is typically for the low-dimensional case.) Figure SM.14 demonstrates that the performance of EPBO+TR and SCBO is consistent with their performance in the 2-dimensional case (in Figure SM.13). This sheds light on the more promising use of EPBO+TR in dealing with high-dimensional constrained problems and warrants further investigation.

On the other hand, it is common to consider only a small subset of experimental factors to be active in the objective function or the constraints (Eriksson and Jankowiak, 2021). For instance, we explore the Ackley problem of 20-dimension space, where the dimension of the objective Ackley function is set to be two while the dimension of the constraints and the design space are set to be 20. Figure SM.15 shows that the performance of EPBO+TR and SCBO on this sparse

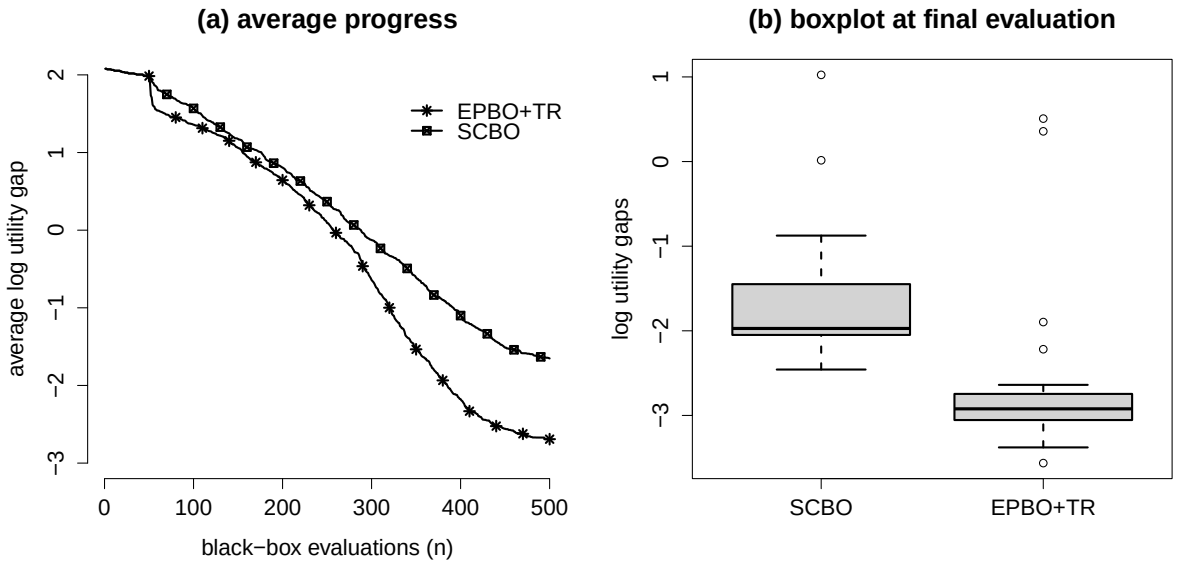

Figure SM.14: Comparison results of EPBO+TR and SCBO for the 20-dimensional Ackley problem. (a) the average progress in the log utility gap over iterations and (b) boxplots of the log utility gap at the final evaluation.

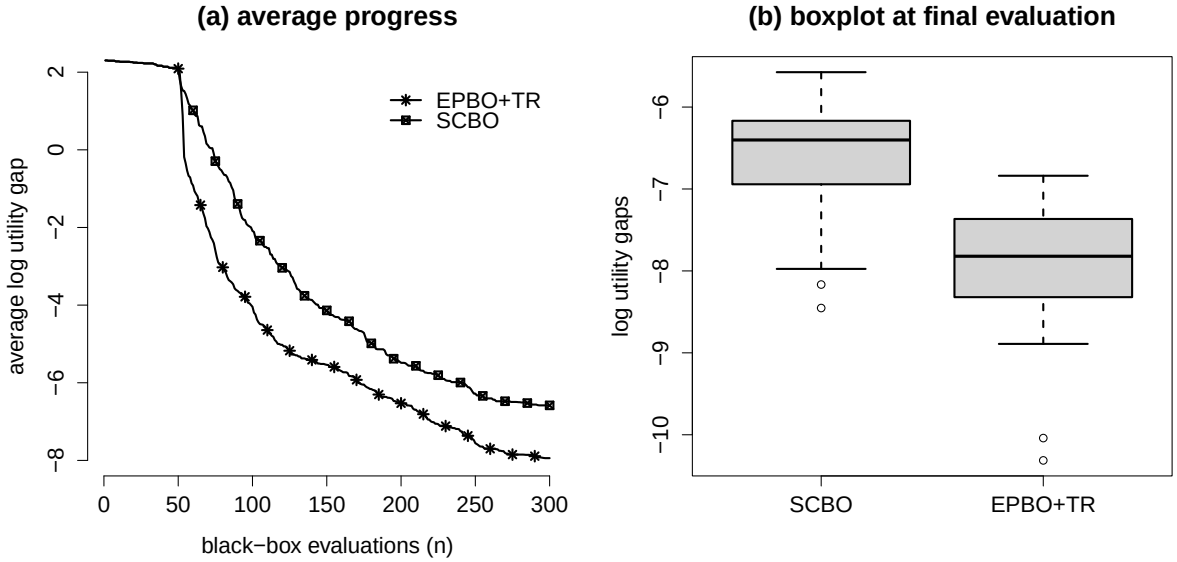

Figure SM.15: Comparison results of EPBO+TR and SCBO for the 20-dimensional sparse Ackley problem with the same captions as in Figure SM.14.

Ackley problem is consistent with their performance on all input activity (in Figure SM.14). It should be emphasized that neither EPBO+TR nor SCBO is specifically tailored for sparse constrained optimization problems. It is of interest to investigate the integration of EPBO+TR with the sparsity-inducing prior proposed by Eriksson and Jankowiak (2021).

References

Balandat, M., B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy (2020). BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (Eds.), *Advances in Neural Information Processing Systems*, Volume 33, pp. 21524–21538. Curran Associates, Inc. [SM 2]

Binois, M. and N. Wycoff (2022). A survey on high-dimensional Gaussian process modeling with application to Bayesian optimization. *ACM Transactions on Evolutionary Learning and Optimization 2*(2), 1–26. [SM 17]

Diouane, Y., V. Picheny, R. L. Riche, and A. S. D. Perrotolo (2023). TREGO: a trust-region framework for efficient global optimization. *Journal of Global Optimization 86*(1), 1–23. [SM 17]

Eriksson, D. and M. Jankowiak (2021). High-dimensional Bayesian optimization with sparse axis-aligned subspaces. In C. de Campos and M. H. Maathuis (Eds.), *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, Volume 161 of *Proceedings of Machine Learning Research*, pp. 493–503. PMLR. [SM 17, SM 18]

Eriksson, D., M. Pearce, J. Gardner, R. D. Turner, and M. Poloczek (2019). Scalable global optimization via local Bayesian optimization. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, Volume 32. Curran Associates, Inc. [SM 15]

Eriksson, D. and M. Poloczek (2021). Scalable constrained Bayesian optimization. In A. Banerjee and K. Fukumizu (Eds.), *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, Volume 130 of *Proceedings of Machine Learning Research*, pp. 730–738. PMLR. [SM 2, SM 15, SM 17]

Garnett, R. (2023). *Bayesian Optimization*. Cambridge, United Kingdom: Cambridge University Press. https://bayesoptbook.com/. [SM 10]

Gramacy, R. B. (2016). laGP: Large-scale spatial modeling via local approximate Gaussian processes in R. *Journal of Statistical Software 72*(1), 1–46. [SM 2, SM 4, SM 5, SM 10]

Gramacy, R. B. (2020). *Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences*. Boca Raton, Florida: Chapman Hall/CRC. http://bobby.gramacy.com/surrogates/. [SM 10]

Gramacy, R. B., G. A. Gray, S. L. Digabel, H. K. H. Lee, P. Ranjan, G. Wells, and S. M. Wild (2016). Modeling an augmented Lagrangian for blackbox constrained optimization. *Technometrics 58*(1), 1–11. [SM 2, SM 7]

Keane, A. (1994). Experiences with optimizers in structural design. In *Proceedings of the conference on adaptive computing in engineering design and control*, Volume 94, pp. 14–27. [SM 15]

Lee, H. H., R. B. Gramacy, C. Linkletter, and G. A. Gray (2011). Optimization subject to hidden constraints via statistical emulation. *Pacific Journal of Optimization 7*(3), 467–478. [SM 4]

Lindberg, D. V. and H. K. Lee (2015). Optimization under constraints by applying an asymmetric entropy measure. *Journal of Computational and Graphical Statistics 24*(2), 379–393. [SM 2, SM 3, SM 4]

Loeppky, J. L., J. Sacks, and W. J. Welch (2009). Choosing the sample size of a computer experiment: A practical guide. *Technometrics 51*(4), 366–376. [SM 17]

Parr, J. M., A. J. Keane, A. I. Forrester, and C. M. Holden (2012). Infill sampling criteria for surrogate-based optimization with constraint handling. *Engineering Optimization 44*(10), 1147–1166. [SM 6, SM 13]

Picheny, V., R. B. Gramacy, S. Wild, and S. Le Digabel (2016). Bayesian optimization under mixed constraints with a slack-variable augmented Lagrangian. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems 29*, pp. 1435–1443. Curran Associates, Inc. [SM 2, SM 8]

Picheny, V., T. Wagner, and D. Ginsbourger (2013). A benchmark of Kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization 48*(3), 607–626. [SM 5]

Pourmohamad, T. and H. K. H. Lee (2022). Bayesian optimization via barrier functions. *Journal of Computational and Graphical Statistics 31*(1), 74–83. [SM 2, SM 4, SM 5]

Rasmussen, C. E. and C. K. Williams (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: the MIT Press. [SM 10]

Roustant, O., D. Ginsbourger, and Y. Deville (2012). DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software 51*(1), 1–55. [SM 10]

Santoni, M. L., E. Raponi, R. De Leone, and C. Doerr (2023). Comparison of bayesian optimization algorithms for bbob problems in dimensions 10 and 60. In *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, GECCO'23 Companion, New York, NY, USA, pp. 2390–2393. Association for Computing Machinery. [SM 17]

Schonlau, M., W. J. Welch, and D. R. Jones (1998). Global versus local search in constrained optimization of computer models. *Institute of Mathematical Statistics Lecture Notes – Monograph Series 34*, 11–25. [SM 2]