arXiv:2602.02432v1 [cs.LG] 2 Feb 2026

使用贝叶斯优化最大化可靠性

Jack M. Buckingham [1] **Ivo Couckuyt** [2 3] **Juergen Branke** [4]

摘要

贝叶斯优化（BO）是一种广受欢迎的样本效率高的方法，适用于代价高昂的黑箱优化。制造业中就有这样一类问题：在存在随机扰动的情况下，如何最大化设计的可靠性，或者等价地最小化其失效概率；这类问题可能涉及极其罕见的失效（$P_{\text{fail}} = 10^{-6} - 10^{-8}$）。本文提出两种基于汤普森采样和知识梯度的贝叶斯优化方法，其中后者通过近似一步贝叶斯最优策略来最小化失效概率的对数。两种方法都采用重要性采样，以针对极小的失效概率。实证结果表明，在极端和非极端情形下，所提方法的表现均优于现有方法。

1. 引言

系统可靠性通过设计满足所有性能约束的概率来量化，其中性能取决于随机输入或扰动。在许多应用中，自然的目标是提高可靠性，或者等价地降低失效概率。例如，电子分量的良率优化（Wang 等，2018年；Weller 等，2022年）、机械部件的设计（Huang& Chan，2010年；Bichon等人，2012年）、广告活动的优化（Betlei等人，2024年）以及医疗场景中的治疗剂量选择（Durham等人，1998年）。

该问题的线性版本最初由Charnes & Cooper(1963作为机会约束规划中的“P模型”变体进行研究。然而，上述示例通常是非线性且非凸的，其数据获取依赖于资源密集型仿真或现实世界实验。

在某些情况下，所需的可靠性非常高，这意味着必须采用重要性采样或其他技术来估计稀有事件概率。例如，在电子学领域，由于数百万个SRAM单元堆叠在一起，单个单元的故障率必须至多为 10^{-8} – 10^{-6}，以确保整体具有合理的故障率（Sun等人，2015）。此外，在仿真和实验中，梯度通常无法获得，并且我们只能进行少量经过谨慎选择的评估。因此，贝叶斯优化（BO）是一种理想的算法（Frazier，2018年；Garnett，2023年）。

最小化失效概率，关键在于准确表征分隔可行区域与不可行域的相关边界部分。在可靠性工程中，这被称为极限状态面。考虑机械或电子元件的设计：由于无法精确制造出名义设计，最终设计会是对该名义设计的扰动。高效的优化算法会针对既接近极限状态面、又接近名义设计的设计开展实验。然而，现有方法未能利用这一点，而是均匀探索整个极限状态面。这一点可见图1，其中将知识梯度（本文提出的一种算法）与（Huang & Chan，2010年）提出的一种替代方法进行了比较。

贡献

1. 我们提出两种采集策略，将贝叶斯优化扩展到最大化可靠性问题：一种基于汤普森采样（称为 TS-MR），另一种基于知识梯度（称为 KG-MR）。

2. 我们为 TS-MR 提供了一种近似，为 KG-MR 提供了两种近似；这些近似都融入了重要性采样，从而将适用范围扩展到了最优失效概率极小的情形。

3. 我们通过多个示例证明，TS-MR 和 KG-MR 的表现均优于最先进方法，其中单次评估的 KG-MR 是表现最佳且最可靠的方法。

[1]华威大学数学系统CDT，英国考文垂 [2]根特大学工程与建筑学院，比利时根特 [3]imec，比利时鲁汶 [4]华威商学院，英国华威大学考文垂。通信作者：Jack M.Buckingham <jack.buckingham@warwick.ac.uk>、Juergen Branke<juergen.branke@wbs.ac.uk>。

预印本。*2026*年*2*月*3*日。

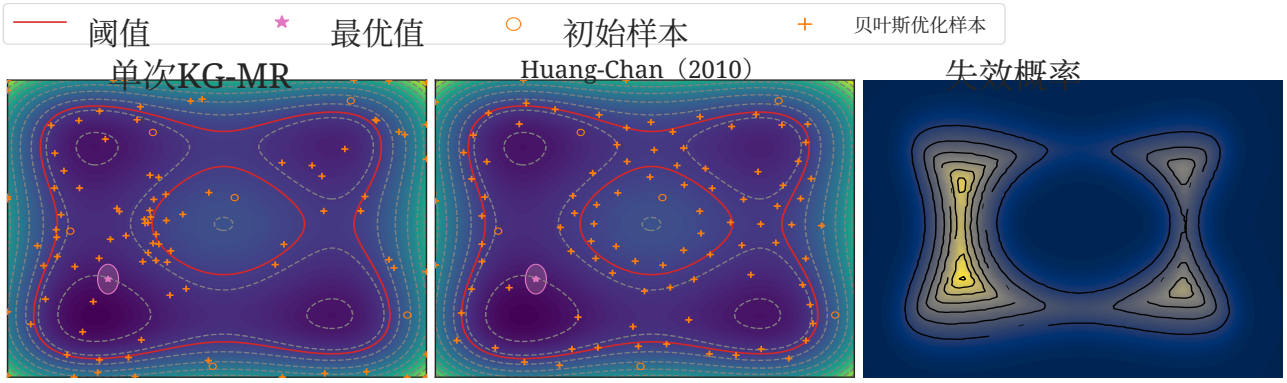

图 *1.* 等高线图展示了 Styblinski-Tang（2维）问题，其中名义设计通过添加一个正态分布随机变量进行扰动，值高于阈值的扰动设计被归类为失效。左侧和中间面板展示了黑盒函数 f 及其阈值水平集，以及用于最大化可靠性（KG-MR）的单次知识梯度和 Huang & Chan (2010年) 算法所收集的观测。最优名义设计以粉色圆点表示，周围的椭圆表示正态分布扰动的一个标准差。右侧面板展示了定义于（2）中的域上的失效概率。

2. 问题定义

令 $\mathcal{X} \subset \mathbb{R}^{d_x}$、$\mathcal{Y} \subset \mathbb{R}^{d_y}$ 和 $\mathcal{U} \subset \mathbb{R}^{d_u}$ 分别表示名义设计、扰动设计和扰动变量的空间。设 $f\colon \mathcal{Y} \to \mathbb{R}$ 为一个黑盒函数，并假设我们可以在自行选择的点 $\boldsymbol{y} \in \mathcal{Y}$ 处进行代价高昂的观测 $v = f(\boldsymbol{y})$。设 $\mathbf{u}$ 为一个取值于 $\mathcal{U}$ 的随机向量，且 $\boldsymbol{g}\colon \mathcal{X} \times \mathcal{U} \to \mathcal{Y}$ 为一个已知函数，它将名义设计 $\boldsymbol{x}$ 与扰动 $\boldsymbol{u}$ 映射为扰动设计 $\boldsymbol{y}$。设 $c \in \mathbb{R}$ 为一个阈值，使得满足 $f(\boldsymbol{y}) \geq c$ 的扰动设计 $\boldsymbol{y} \in \mathcal{Y}$ 被视为不可行。最后，设 $\mathcal{Y}_{\text{feas}} \subset \mathcal{Y}$ 为一个子空间，其中的取值 $\boldsymbol{y} \in \mathcal{Y} \setminus \mathcal{Y}_{\text{feas}}$ 无论 $f(\boldsymbol{y})$ 的值如何，都被视为不可行。这包括对 $\boldsymbol{y}$ 的一些更简单的约束（并非黑盒约束），例如对受到正态扰动的名义设计施加的边界约束。本文要解决的问题是最小化与名义设计相关的失效概率，

$$\min_{\boldsymbol{x} \in \mathcal{X}} P(\boldsymbol{x}) \tag{1a}$$

$$P(\boldsymbol{x}) \coloneqq \mathbb{P}_{\mathbf{u}}\Big(f \circ \boldsymbol{g}(\boldsymbol{x}, \mathbf{u}) \geq c \text{ or } \boldsymbol{g}(\boldsymbol{x}, \mathbf{u}) \notin \mathcal{Y}_{\text{feas}}\Big). \tag{1b}$$

这里，$f \circ \boldsymbol{g}(\boldsymbol{x}, \mathbf{u}) = f(\boldsymbol{g}(\boldsymbol{x}, \mathbf{u}))$ 表示函数复合，$\mathbb{P}_{\mathbf{u}}$ 表示关于$\mathbf{u} \sim \mathbb{P}_{\mathbf{u}}$的概率。

以下引理在附录A中得到证明，可确保问题（1）至少有一个解。

引理 2.1。假设 $f\colon \mathcal{Y} \to \mathbb{R}$连续，$\boldsymbol{g}\colon \mathcal{X} \times \mathcal{U} \to \mathcal{Y}$ 连续，且$\mathbb{P}_{\mathbf{u}}$ 的分布在极限状态面或$\mathcal{Y}_{\text{feas}}$的边界上没有质量。则失效概率 $P\colon \mathcal{X} \to [0, 1]$ 连续。若此外 $\mathcal{X}$ 是紧致的，则 P 会在某个 $\boldsymbol{x}^* \in \mathcal{X}$处取得其最小值。

在实验中，我们重点关注以下情形：$d_x = d_y = d_u = d$，$\mathcal{X} = \mathcal{Y}_{\text{feas}} \subset \mathbb{R}^d$ 是一个超矩形，$\mathcal{Y} = \mathcal{U} = \mathbb{R}^d$ 且 $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{x} + \boldsymbol{u}$，但核心概念适用于一般情形。在这一特定情形下，（1b）变为

$$P(\boldsymbol{x}) = \mathbb{P}_{\mathbf{u}}\Big(f(\boldsymbol{x} + \mathbf{u}) \geq c \text{ or } \boldsymbol{x} + \mathbf{u} \notin \mathcal{Y}_{\text{feas}}\Big). \tag{2}$$

3. 相关工作

贝叶斯优化此前已应用于基于可靠性的设计优化（Dubourg 等人，2011年；Janusevskis& Le Riche，2013；El Amri 等人，2023；Pelamatti 等人，2023），其中作者寻求在满足约束的前提下，使目标函数的期望最小化，而这些约束必须以较高的最低概率成立。Tsai 等人（2023）解决了一个类似问题，其中约束必须在期望意义上成立。相比之下，我们考虑的是将约束提升为目标函数的问题，并寻求以最大概率满足某项约束的解。

分位数优化和最坏情况优化也是与之密切相关的问题，贝叶斯优化已成功应用于这些问题（Cakmak 等人，2020年；Nguyen 等人，2021；Picheny 等人，2022年；ur Rehman 等人，2014；Han 等人，2025）。不过，此处风险水平固定，并选择名义设计 $\boldsymbol{x}$ 以获得尽可能优的阈值 c，而本文讨论的是相反的问题。

主动学习中也会出现相关问题，包括估计失效概率 $\mathbb{P}(f(\mathbf{u}) \geq c)$ (称为可靠性分析)，以及估计黑盒函数 $\{\boldsymbol{u} \in \mathcal{U}\colon f(\boldsymbol{u}) = c\}$ 的特定等值集，其中极限状态面就是这样一个集合。该领域的研究包括Picheny 等人 (2010年), Bichon等人 (2008年), Bect等人 (2011年),Knudde等人 (2019)以及Booth&Renganathan(2025)。

一阶和二阶可靠性方法（FORM / SORM）（Hasofer &Lind，1974年；Breitung，1984年）利用极限状态面的不同部分对失效概率的贡献不均这一特点，以极限状态面上概率密度最高的点为中心，用局部线性或二次近似替代该曲面$p(\boldsymbol{u} \mid f(\boldsymbol{u}) = c)$。对于许多估计问题而言，这种近似是可接受的；但在优化问题中，最优名义设计 $\boldsymbol{x}^*$ 反而会被困在$d+1$ 概率同样最大的点之间——对于一个多面体而言，这是 $\mathbb{R}^d$中的最少顶点数。虽然可以设想一种算法，在所提出的名义设计收敛过程中跟踪这些点，但本文并未采用这种方法。

如引言中所述，我们尤其关注失效概率极低的情形，这需要采用稀有事件估计技术，例如重要性采样（Kloek & Van Dijk，1978 年）或子集模拟（ Au& Beck，2001 年）。这两种方法的现代方案都采用马尔可夫链蒙特卡洛（ MCMC）（Papaioan- nou 等，2015；Papakonstantinou 等，2023年；Tong &Stadler，2023年）和密度估计（De Boer 等，2005年；Kroese 等，2013）。Tabandeh 等（2022年）综述了现代重要性抽样方法。Fonseca 等人Fonseca 等（2006年）以不依赖于是否存在代理模型的方式，将重要性采样应用于基于可靠性的设计优化问题。

在旨在最大化可靠性的研究中，Wang 等 （2018年）使用贝叶斯优化来决定在哪里进行自适应蒙特卡罗可靠性估计，而Weller 等（2022年）提出了一种嵌套方案，利用贝叶斯优化训练一个在极限状态面上精度较高的高斯过程，然后应用进化算法选取有前景的名义设计。他们分别需要数万次和数千次仿真，并假设仿真速度足够快，因此这些计算是可行的。据我们所知，只有两项研究不需要如此大量的仿真。Huang& Chan（2010年）提出了一种算法，可从四种不同的采集准则中进行选择，以探索极限状态面。Bichon 等人（ 2012 年）提出了对其早期高效全局可靠性分析（EGRA）工作的三项扩展，以解决与基于可靠性的设计优化相关的各种问题，其中一项是（1）的约束版本。这些方法用作第8节数值比较的基线算法。

4. 背景

贝叶斯优化（BO）是一种样本高效的黑盒函数全局优化算法（Frazier，2018年；Garnett，2023年）。通常会为目标函数设定一个先验信念，通常采用高斯过程（GP），并通过采集策略权衡探索不确定性较高的区域与利用预期已能取得良好值的区域。

在标准情形下，评估按顺序进行：每一步先根据现有数据拟合概率先验的超参数，然后利用采集策略确定下一个样本位置，最后在该位置评估黑盒函数。整个过程从初始的空间填充设计开始，例如 Sobol' 样本，并通常持续运行，直到评估预算耗尽。

汤普森采样（TS）是一种采集策略，其下一个采样位置是目标函数后验分布的随机样本所对应的最大值。相比之下，知识梯度（KG）（Frazier 等，2009；Scott 等，2011年）是一步前瞻的贝叶斯最优策略，适用于最终推荐值为目标函数后验均值最大值的情形。

重要性采样是一种用于稀有事件概率的蒙特卡罗（ MC）估计的方差缩减技术（Tabandeh 等，2022年）。对于适中的样本量， N_u，标准MC估计量 $\underline{P}(\boldsymbol{x}) \approx \frac{1}{N}\sum_{i=1}^{N_u} \mathbb{I}_{\Omega_i}$很可能观测不到失效；此处我们将$\Omega_i = \{f \circ \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}_i) \geq c$或 $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}_i) \notin \mathcal{Y}_{\text{feas}}\}$记作相应符号。事实上，该估计量的标准差为$\sqrt{P(\boldsymbol{x})(1-P(\boldsymbol{x}))/N_u}$。如果最优 $\boldsymbol{x}^*$ 处的真实失效概率为 $P(\boldsymbol{x}^*) = 10^{-7}$，那么若要使估计量的标准差低于失效概率的10%，我们需要 $N_u \gtrsim 10^9$ 个样本。

重要性采样通过从一种使失效更频繁发生的分布 $\mathbb{Q}_{\mathbf{u}}$ 中进行采样来降低方差，然后对样本进行重加权，以反映它们在 $\mathbb{P}_{\mathbf{u}}$ 下的真实概率。假设 $\mathbb{P}_{\mathbf{u}}$ 和 $\mathbb{Q}_{\mathbf{u}}$ 的密度分别记为 $p(\cdot)$ 和 $q(\cdot)$，则重要性采样估计为 $P(\boldsymbol{x}) \approx \frac{1}{N_u}\sum_{i=1}^{N_u} \frac{p}{q(\boldsymbol{u}_i)(\boldsymbol{u}_i)} \mathbb{I}_{\Omega_i}$。对于 $\mathbb{Q}_{\mathbf{u}}$，一种简单的选择是采用与 $\mathbb{P}_{\mathbf{u}}$ 相同的参数族，但增大方差。例如，如果 $\mathbb{P}_{\mathbf{u}} = \mathcal{N}(\mathbf{0}, \Sigma_u)$，则选择 $\mathbb{Q}_{\mathbf{u}} = \mathcal{N}(\mathbf{0}, \tau^2\Sigma_u)$。

5. 推荐设计

我们首先假设已经对黑盒函数 f 进行了若干次观测，并希望提取一个推荐的名义设计。问题 (1) 属于部分信息或间接信息问题：我们观测的不是目标函数 $P(\boldsymbol{x})$，而是观测 $f(\boldsymbol{y})$ 在有限个$\boldsymbol{y}$上的取值。因此，我们收集的信息不足以完全了解 $P(\boldsymbol{x})$，只能求助于模型。

让我们使用高斯过程先验 $f \sim \mathcal{GP}(\mu, k)$ 来对黑盒目标 f：$\mathcal{Y} \to \mathbb{R}$ 建模。为确保优化问题定义明确，我们假设 f 是样本连续的。

基于这一先验信念，我们使用高斯过程的后验均值作为 f 的代理模型。因此，在 n 次观测之后

$v_1 = f(\boldsymbol{y}_1) \ldots, v_n = f(\boldsymbol{y}_n)$，推荐的名义设计为

$$\boldsymbol{x}_n^* \in \arg\min_{x \in \mathcal{X}} P_n(\boldsymbol{x}) \coloneqq \mathbb{E}[P(\boldsymbol{x}) \,|\, \mathcal{F}_n] \tag{3}$$

其中 $\mathcal{F}_n$ 是由 n 观测生成的σ代数，且 $P_n(\boldsymbol{x})$ 是期望失效概率。引理 5.1 证明（3）中的argmin非空。

以下引理在附录A中证明。

引理 5.1. 假设 f 是样本连续的高斯过程，且 $\boldsymbol{g}: \mathcal{X} \times \mathcal{U} \to \mathcal{Y}$ 是连续的。则在对 $\mathbb{P}_{\mathbf{u}}$ 所施加的、于附录A中明确给出的温和正则性条件下，$P_n: \mathcal{X} \to [0,1]$ 是连续的。若 $\mathcal{X}$还是紧致的，则 P_n 在某个$\boldsymbol{x}_n^* \in \mathcal{X}$处取得其最小值。

将（1b）写成关于示性函数组合的期望，并借助富比尼定理交换积分次序，我们得到

$$\begin{aligned} P_n(\boldsymbol{x}) &= \mathbb{E}[P(\boldsymbol{x}) \,|\, \mathcal{F}_n] \\ &= \mathbb{E}\Big[\mathbb{E}_{\mathbf{u}}\Big[\mathbb{I}_{\{f \circ \boldsymbol{g}(\boldsymbol{x},\mathbf{u}) \geq c\}} \mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x},\mathbf{u}) \in \mathcal{Y}_{\text{feas}}\}} \\ &\qquad + \mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x},\mathbf{u}) \notin \mathcal{Y}_{\text{feas}}\}}\Big] \,\Big|\, \mathcal{F}_n\Big] \\ &= \mathbb{E}_{\mathbf{u}}\Big[\Phi_n(\boldsymbol{x}, \mathbf{u}; c)\, \mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x},\mathbf{u}) \in \mathcal{Y}_{\text{feas}}\}} + \mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x},\mathbf{u}) \notin \mathcal{Y}_{\text{feas}}\}}\Big] \end{aligned} \tag{4}$$

其中我们记

$$\Phi_n(\boldsymbol{x}, \boldsymbol{u}; c) = \Phi\left(\frac{\mu_n\big(\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u})\big) - c}{\sqrt{k_n\big(\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}), \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u})\big)}}\right) \tag{5}$$

其中，μ_n 和 k_n 分别是 f 在给定 $\mathcal{F}_n$条件下的后验均值和协方差函数。

最后，对于非常小的失效概率，我们更关注失效概率的数量级，而不是其本身。因此，我们定义

$$R_n(\boldsymbol{x}) = -\log P_n(\boldsymbol{x}). \tag{6}$$

这里的负号只是将最小化问题转换为最大化问题。虽然 R_n 和 P_n 分别在同一点处取得最大值和最小值，但取对数会更方便，因为基于梯度的优化器的默认收敛容差可能会导致 P_n 过早终止，但仍适用于 R_n。当然，只有当 $P_n(\boldsymbol{x}) > 0$时，$R_n(\boldsymbol{x})$才有定义。不过，如果 $P_n(\boldsymbol{x}) = 0$，那么 $P(\boldsymbol{x}) = 0$ 几乎必然成立，因此 $\boldsymbol{x}$是一个失效概率为零的最优点，优化问题也就解决了。因此，我们可以放心地忽略这种情况。

6. 采集策略

现在，我们介绍本文提出的两种采集策略。

6.1. 用于最大化可靠性的汤普森采样（TS-MR）

我们提出的第一个算法基于汤普森采样。在优化的每个阶段，我们从关于 f 的高斯过程后验中抽取一个随机˜样本 f，并用它计算具有最大可靠性的名义设计，$\boldsymbol{x}_{n+1}$。

$$\boldsymbol{x}_{n+1} \in \arg\min_{\boldsymbol{x} \in \mathcal{X}} \mathbb{P}_{\mathbf{u}}\Big(\tilde{f} \circ \boldsymbol{g}(\boldsymbol{x}, \mathbf{u}) \geq c \;\text{ or }\; \boldsymbol{g}(\boldsymbol{x}, \mathbf{u}) \notin \mathcal{Y}_{\text{feas}}\Big). \tag{7}$$

然后，我们选择一个扰动 $\boldsymbol{u}_{n+1}$，使该扰动的概率密度与该扰动处失效指示函数的方差之积最大化，

$$\boldsymbol{u}_{n+1} \in \arg\max_{\boldsymbol{u} \in \mathcal{U}_{\text{feas}}^{n+1}} \alpha_n^{\text{MV}}(\boldsymbol{u}; \boldsymbol{x}_{n+1}), \tag{8a}$$

$$\begin{aligned} &\alpha_n^{\text{MV}}(\boldsymbol{u}; \boldsymbol{x}_{n+1}) \\ &\quad = p(\boldsymbol{u}) \mathrm{Var}\big[\mathbb{I}_{\{f \circ \boldsymbol{g}(\boldsymbol{x}_{n+1}, \boldsymbol{u}) \geq c\}} \,\big|\, \mathcal{F}_n\big] \\ &\quad = p(\boldsymbol{u}) \Phi_n(\boldsymbol{x}_{n+1}, \boldsymbol{u}; c)\Big(1 - \Phi_n(\boldsymbol{x}_{n+1}, \boldsymbol{u}; c)\Big). \end{aligned} \tag{8b}$$

为了使该组合可行，优化被约束在 $\mathcal{U}^{n+1}$ 可行 $= \boldsymbol{g}_{\boldsymbol{x}_{n+1}}^{-1}(\mathcal{Y}_{\text{feas}}) \subset \mathcal{U}$ 中，其中我们将 $\boldsymbol{g}_{\boldsymbol{x}_{n+1}}(\cdot) = \boldsymbol{g}(\boldsymbol{x}_{n+1}, \cdot)$ 记为。α_n^{MV}的理念是鼓励探索极限状态曲面中那些接近所提出的名义设计 $\boldsymbol{x}_{n+1}$ 的部分。最后，下一个采样位置为

$$\boldsymbol{y}_{n+1} = \boldsymbol{g}(\boldsymbol{x}_{n+1}, \boldsymbol{u}_{n+1}). \tag{9}$$

6.2. 用于最大化可靠性的知识梯度（KG-MR）

如第5节中所述，式（1）中的问题属于部分或间接信息问题。知识梯度此前已成功应用于此类问题（Daulton 等人，2023年；Buathong 等人，2024；Buckingham 等人，2025a；b）。

KG应用于价值函数 $(R_n)_{n \geq 1}$（定义于（6））是失效概率负对数最大值的期望值在下一次观测前后的变化，条件为该观测在指定的 $\boldsymbol{y} \in \mathcal{Y}_{\text{feas}}$处进行。也就是说，在 n 次观测之后，它是函数 $\alpha_n^{\text{KG-MR}}: \mathcal{Y}_{\text{feas}} \to \mathbb{R}$由下式给出

$$\begin{aligned} &\alpha_n^{\text{KG-MR}}(\boldsymbol{y}) = \\ &\quad \mathbb{E}\left[\max_{\boldsymbol{x} \in \mathcal{X}} R_{n+1}(\boldsymbol{x}) \,\middle|\, \mathcal{F}_n, \boldsymbol{y}_{n+1} = \boldsymbol{y}\right] - \max_{\boldsymbol{x} \in \mathcal{X}} R_n(\boldsymbol{x}). \end{aligned} \tag{10}$$

它通过在期望效用 $\mathbb{E}[-\log P(\boldsymbol{x}) \,|\, \mathcal{F}_n] \approx -\log P_n(\boldsymbol{x}) = R_n(\boldsymbol{x})$ 中用 $P_n(\boldsymbol{x})$ 替换 $P(\boldsymbol{x})$，近似了用于最小化失效概率对数的一步前瞻贝叶斯最优策略。这

显著提高了计算效率，而类似的插件近似此前已在多目标贝叶斯优化中取得良好效果（Daulton 等人，2023年）。

注 6.1。 如前所述，如果 $\max_{\boldsymbol{x}\in\mathcal{X}} R_{n+1}(\boldsymbol{x}) = \infty$，那么我们就找到了一个几乎必然具有零失效概率的点，因此无需担心 $\alpha_n^{\text{KG-MR}}(\boldsymbol{y})$ 未定义。

注 6.2。 如果最优失效概率没有那么极端，我们完全可以去掉对数项，并直接根据 P_n 定义它。

引理 6.3。 式 (10) 中的 *KG-MR* 采集函数处处非负。也就是说，

$$\forall n \geq 0 \; \forall \boldsymbol{y} \in \mathcal{Y}_{\text{feas}} \quad \alpha_n^{\textit{KG-MR}}(\boldsymbol{y}) \geq 0.$$

证明见附录 A。

7. 近似

为了优化并找到（3）中的推荐点或（7）中的查询点，我们需要对（8）和（10）以及采集函数进行近似。对于（4）的一种简单近似，是在 N_u 个独立点 $\boldsymbol{u}_1$、...、$\boldsymbol{u}_{N_u} \sim \mathbb{P}_{\mathbf{u}}$ 上进行蒙特卡洛近似，

$$P_n(\boldsymbol{x}) \approx \frac{1}{N_u}\sum_{i=1}^{N_u} \Phi_n(\boldsymbol{x}, \boldsymbol{u}_i; c)\, \mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x},\boldsymbol{u}_i)\in\mathcal{Y}_{\text{feas}}\}} + \mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x},\boldsymbol{u}_i)\notin\mathcal{Y}_{\text{feas}}\}}. \quad (11)$$

然而，直接优化这种近似存在两个问题。首先，虽然得益于高斯过程的不确定性，我们能够通过解析方法消除指示函数 $\mathbb{I}_{\{f\circ\boldsymbol{g}(\boldsymbol{x},\mathbf{u})\geq c\}}$，但无法去掉所有指示函数；剩余的指示函数意味着（11）中的近似在 $\boldsymbol{x}$ 上不连续。其次，如果最优失效概率很小，那么所有 $\boldsymbol{u}_i$ 都很可能远低于失效阈值，这意味着 $\Phi_n(\boldsymbol{x}, \boldsymbol{u}_i;\ c) \approx 0$ 的梯度几乎为零，因此我们没有可用于优化 $P_n(\boldsymbol{x})$的梯度信息。

为了解决这些问题中的第一个，我们对指示函数进行临时平滑，详见附录 D.1。具体而言，我们近似 $\mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x},\boldsymbol{u})\in\mathcal{Y}_{\text{feas}}\}} \approx \iota(\boldsymbol{g}(\boldsymbol{x},\boldsymbol{u});\delta)$，其中 δ 决定边界附近平滑区域的宽度 $\partial\mathcal{Y}_{\text{feas}}$。实践中，我们发现只需进行极少量的平滑，因此设定 $\delta = \min(0.05\ell_{\min}, 0.1)$ 其中 $\ell_{\min}$ 是 $\mathcal{Y}_{\text{feas}}$的最短边长。

对于第二个问题，即失效概率贡献接近于零所导致的问题，我们采用抽样分布为 $\mathbb{Q}_{\mathbf{u}} = \mathcal{N}(\mathbf{0},\ \tau^2\Sigma_u)$ 的重要性采样。在实践中，我们发现 $\tau = 3$ 在最优失效概率处于 $10^{-6} - 10^{-8}$ 范围内时效果良好。

7.1 推荐点近似

将指示函数平滑近似和重要性采样代入（3）可得 $P_n(\boldsymbol{x})$，$R_n(\boldsymbol{x})$ 以及建议 $\boldsymbol{x}_n^*$，

$$\hat{\boldsymbol{x}}_n^* \in \arg\max_{\boldsymbol{x}\in\mathcal{X}} \hat{R}_n(\boldsymbol{x}) = \arg\min_{\boldsymbol{x}\in\mathcal{X}} \hat{P}_n(\boldsymbol{x}), \quad (12a)$$

$$\hat{R}_n(\boldsymbol{x}) = -\log \hat{P}_n(\boldsymbol{x}), \quad (12b)$$

$$\hat{P}_n(\boldsymbol{x}) = \frac{1}{N_u}\sum_{i=1}^{N_u} \frac{p(\boldsymbol{u}_i)}{q(\boldsymbol{u}_i)} \hat{J}_n(\boldsymbol{x}, \boldsymbol{u}_i), \quad (12c)$$

$$\hat{J}_n(\boldsymbol{x}, \boldsymbol{u}) = \Phi_n(\boldsymbol{x}, \boldsymbol{u}; c)\iota\big(\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}); \delta\big) + \big(1 - \iota\big(\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}); \delta\big)\big). \quad (12d)$$

我们通过多起点L-BFGS-B求解 (12a)，这是一种基于梯度的确定性优化器（见附录 D.3）。

此外，众所周知，准蒙特卡罗（qMC）可将蒙特卡罗估计的收敛速率从 $\mathcal{O}(1/\sqrt{N_u})$ 提升至$\mathcal{O}((\log N_u)^{d_u}/N_u)$，适用于许多被积函数（Lemieux，2009）。当分布 $\mathbb{P}_{\mathbf{u}}$ 适用时，我们使用准蒙特卡罗近似来计算 (12c)，以 Sobol' 样本代替独立样本。对于高斯分布和均匀分布的 $\mathbf{u}$，此方法可行。

7.2. 汤普森采样的近似

TS-MR 中所需的第一个近似是抽取~样本 f，条件为 $\mathcal{F}_n$。为此，我们采用Wilson等人 (2020年; 2021) 提出的路径采样方法，使用 N_{rff} =1024 个随机傅里叶特征（RFFs）。我们采用如下~近似，类似于（12），使用 $P(\boldsymbol{x})$ 表示 ~ 使用 7 ）中得到的失效概率，f，

$$\tilde{P}(\boldsymbol{x}) \approx \frac{1}{N_u}\sum_{i=1}^{N_u} \frac{p(\boldsymbol{u}_i)}{q(\boldsymbol{u}_i)} \tilde{J}(\boldsymbol{x}, \boldsymbol{u}_i) \quad (13a)$$

$$\tilde{J}(\boldsymbol{x}, \boldsymbol{u}) \approx \Phi\left(\frac{\tilde{f}\circ\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}) - c}{\rho}\right)\iota\big(\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}); \delta\big) + \big(1 - \iota\big(\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}); \delta\big)\big). \quad (13b)$$

这里，我们使用参数 $\rho > 0$ 模拟了 (12d) 中 $\mathbb{I}_{\{\tilde{f}\circ\boldsymbol{g}(\boldsymbol{x},\boldsymbol{u})\geq c\}}$ 的平滑处理。同样，我们在对数空间中使用多起点L-BFGS-B优化 $P(\boldsymbol{x})$ 的这一近似 ~。

TS-MR 的第二阶段需要优化 (8) 中的 α_n^{MV}。这既不需要重要性采样，也不需要指示函数平滑。不过，我们确实在对数空间中进行优化，并使用 PyTorch 中的 log ndtr() 函数，以数值稳定的方式计算 $\log\ \Phi_n(\boldsymbol{x}_{n+1}, \boldsymbol{u}; c)$ 和 $\log\ (1 - \Phi_n(\boldsymbol{x}_{n+1}, \boldsymbol{u}; c))$。完整算法见算法 1。

算法 1 用于最大可靠性的汤普森采样 **输入**：初始样本量 n_0、评估预算 n_{tot}、准蒙特卡洛样本量 N_u、傅里叶特征数量 N_{rff}、平滑参数 ρ, δ 1：根据在 $\mathcal{Y}_{\text{feas}}$ 上生成的扰乱 Sobol' 序列，选取 n_0 个点并评估 f 2：**对于** $n = n_0, \dots, n_{\text{tot}} - 1$，**执行** 3：在 f 上拟合高斯过程先验的最大后验超参数 $\hat{}$ 4：为 $\mathcal{GP}(\mu_n, k_n)$ 采样随机傅里叶特征及权重 f $\sim$ 5：生成准蒙特卡罗样本 $(\boldsymbol{u}_1, \dots, \boldsymbol{u}_{N_u}) \sim \mathbb{Q}_{\mathbf{u}}$ 6：使用 (13) 中的近似，优化 $\boldsymbol{x}_{n+1} \leftarrow \arg\min_{\boldsymbol{x} \in \mathcal{X}} \log P(\boldsymbol{x})$ 7：优化 (8) 中定义的 α_n^{MV}，得到 $\boldsymbol{u}_{n+1} \leftarrow \arg\max_{\boldsymbol{u} \in \mathcal{U}^{n+1}} \text{feas} \log \alpha_n^{\text{MV}}(\boldsymbol{u}; \boldsymbol{x}_{n+1})$ 8：设置 $\boldsymbol{y}_{n+1} \leftarrow \boldsymbol{g}(\boldsymbol{x}_{n+1}, \boldsymbol{u}_{n+1})$ 9：评估代价高昂的函数 $f(\boldsymbol{y}_{n+1})$ 10：**结束循环** 11：使用 (12) 计算建议值 $\hat{}\boldsymbol{x}^*_{n_{\text{tot}}}$

算法 2 最大可靠性的知识梯度 **输入**：初始样本量 n_0、评估预算 n_{tot}、准蒙特卡洛样本量 N_u, N_v；对于离散 KG-MR，离散化大小 N_x；对于单次KG-MR，平滑参数 δ1：按照在 $\mathcal{Y}_{\text{feas}}$ 上生成的扰乱 Sobol' 序列，选取 n_0 个点并在这些点处评估 f 2：**对于** $n = n_0, \dots, n_{\text{tot}} - 1$，**执行** 3：在 f 上拟合高斯过程先验的最大后验超参数 4：生成准蒙特卡洛样本 $(\boldsymbol{u}_1, \dots, \boldsymbol{u}_{N_u}) \sim \mathbb{Q}_{\mathbf{u}}$ 和 $(z_1, \dots, z_{N_v}) \sim \mathcal{N}(0, 1)$ 5：若使用离散 KG-MR，则使用扰乱的 Sobol' 序列抽取大小为 N_x 的样本 $\mathcal{X}_{\text{disc}}$ 6：使用式 (15) 或式 (14) 优化 $\boldsymbol{y}_{n+1} \leftarrow \arg\max_{\boldsymbol{x} \in \mathcal{Y}_{\text{feas}}} \hat{}\alpha_n(\boldsymbol{y})$ 7：评估昂贵函数，$f(\boldsymbol{y}_{n+1})$ 8：**结束循环** 9：使用式 (12) 计算推荐值 $\hat{}\boldsymbol{x}^*_{n_{\text{tot}}}$

7.3. 知识梯度的近似

KG 的"一次性"近似（Balandat 等人，2020年）使用 MC 或 qMC 估计，对 N_v 个虚拟观测取期望，以计算对下一次观测的期望 v_{n+1}（10）中的期望。将 MC 求和中每个元素对应的最大化变量提取到前面，从而构造出一个单一的高维优化问题，

$$\max_{\boldsymbol{y} \in \mathcal{Y}_{\text{feas}}} \hat{\alpha}_n(\boldsymbol{y}) = \max_{\substack{\boldsymbol{y} \in \mathcal{Y}_{\text{feas}} \\ \boldsymbol{x}_1, \dots, \boldsymbol{x}_{N_v} \in \mathcal{X}}} \frac{1}{N_v} \sum_{i=1}^{N_v} \hat{R}_{n+1}(\boldsymbol{x}_i; \boldsymbol{y}, z_i) - \max_{\boldsymbol{x} \in \mathcal{X}} \hat{R}_n(\boldsymbol{x}). \quad (14)$$

这里，$\hat{} R_n(\boldsymbol{x})$ 与（12b）中的定义相同，而 $\hat{} R_{n+1}(\boldsymbol{x}; \boldsymbol{y}, z)$ 也相同，只是假设 f 的第（$n+1$）个观测位于 $\boldsymbol{y}$，其值为 $v = f(\boldsymbol{y}) = \mu_n(\boldsymbol{y}) + z\sqrt{k_n(\boldsymbol{y}, \boldsymbol{y})}$。这种对重参数化技巧的运用将随机性与对 $\boldsymbol{y}$ 的依赖分离开来。样本 $z_1, \dots, z_{N_v}$ 是由 N_v 个标准高斯变量构成的样本。如果这些变量相互独立，那么这就是标准的 MC 估计；不过在实践中，我们会使用经过 Box-Muller 变换处理的 Sobol' 样本来得到 qMC 估计，这也是 BoTorch 中的标准做法（Balandat 等人，2020年）。（14）中优化问题的维度为 $d_y + N_v d_x$。

另一种避免式（14）中高维优化的方法，是将式（10）中对 $\boldsymbol{x} \in \mathcal{X}$ 的最大化，替换为在有限集合 $\mathcal{X}_{\text{disc}} \subset \mathcal{X}$ 上进行最大化。该集合从扰乱的 Sobol' 序列或其他空间填充设计中抽取（例如Buathong 等人，2024；Buckingham 等人，2025a）。将此方法与期望的（准）蒙特卡洛估计结合起来 over v_{n+1} 给出

$$\hat{\alpha}_n(\boldsymbol{y}) = \frac{1}{N_v} \sum_{i=1}^{N_v} \max_{\boldsymbol{x} \in \mathcal{X}_{\text{disc}}} \hat{R}_{n+1}(\boldsymbol{x}; \boldsymbol{y}, z_i) - \max_{\boldsymbol{x} \in \mathcal{X}_{\text{disc}}} \hat{R}_n(\boldsymbol{x}). \quad (15)$$

当 d_x 较小时，这种近似通常很准确，但对于较大的 d_x，会受到维度灾难的影响。

对于式（15）中的离散近似，集合 $\mathcal{X}_{\text{disc}}$ 是固定的$\hat{}$，因此我们无需担心 R_n 和$\hat{} R_{n+1}$ 关于 $\boldsymbol{x}$ 的可微性，并可将边界平滑因子 $\delta = 0$设为。对于式（14）中的单次近似，我们令 $\delta = \min(0.05\ell_{\text{min}}, 0.1)$。

两种近似均使用多起点 L-BFGS-B 进行优化，尤其注意单次近似所用的初始点（见附录 B）。完整算法总结于算法 2。

8. 实验

为了评估 TS-MR 和两种 KG-MR 近似方法的表现，我们将它们与文献中已有的两种算法进行比较：由 Huang &Chan (2010年) 提出的算法（记为 HC），以及由 Bichon等人 (2008年;2012年) 提出的高效全局可靠性分析（EGRA）。这些方法概述于附录 C，参数选择详见附录 D.5。此外，还与两种更简单的基线算法进行了比较。第一种是随机抽样方法，它根据扰乱后的 Sobol' 序列选择点$\boldsymbol{y}_1$，$\boldsymbol{y}_2$，.....；第二种则使用期望改进（EI）搜索$\min_{\boldsymbol{y}} f(\boldsymbol{y})$，而不是最小化失效概率。对于所有比较的方法，生成推荐点时均采用相同的流程。也就是说，我们使用多起点优化（12a）

，采用附录 D.3中所述的多起点 L-BFGS-B 方法。

我们在三个从高斯过程采样得到的随机问题和九个常用的合成测试问题上比较这些算法，这些问题的输入维度从 2 到 16 不等。所有问题均具有 $\mathcal{X} = \mathcal{Y} = \mathcal{U}$ 和 $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{x} + \boldsymbol{u}$，以表示围绕名义设计的扰动。所用扰动服从正态分布，$\mathbf{u} \sim \mathcal{N}(0, \Sigma_u)$，其对角协方差矩阵 Σ_u 在不同问题之间有所变化。

每种算法在每个测试问题上运行 30 次，每次使用不同的随机种子。每次重复运行时，都会使用扰乱 Sobol' 样本随机选择一个不同的初始设计，样本大小取决于问题维度，具体说明见附录 D.5。对于所有算法，高斯过程代理模型均采用 Mat´ern-5/2 核函数，其最大后验估计超参数按照附录D.2所述方法拟合。

8.1. 推荐结果评估

给定推荐的名义设计 $\boldsymbol{x}_n^*$，我们使用重要性采样对 $P(\boldsymbol{x}_n^*)$ 进行准蒙特卡洛近似评估，并使用真实测试函数，

$$P(\boldsymbol{x}_n^*) \approx \frac{1}{N_u} \sum_{i=1}^{N_u} \frac{p(\boldsymbol{u}_i)}{q(\boldsymbol{u}_i)} \Big(\mathbb{I}_{\{f \circ \boldsymbol{g}(\boldsymbol{x}_n^*, \boldsymbol{u}_i) \geq c\}} \mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x}_n^*, \boldsymbol{u}_i) \in \mathcal{Y}_{\text{feas}}\}} + \mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x}_n^*, \boldsymbol{u}_i) \notin \mathcal{Y}_{\text{feas}}\}} \Big). \tag{16}$$

为此，我们使用 $N_u = 2^{20} \approx 1,000,000$，这个数值大于我们用于采集函数或生成推荐点的数值。对于这一估计，重要性采样所用的膨胀方法和尺度因子，与采集函数中使用的以及生成推荐点时使用的相同。由于所有情况下 $\mathbb{Q}_{\mathbf{u}}$均服从高斯分布，准蒙特卡洛近似使用一个 d 维 Sobol' 样本，并将坐标成对地通过 Box-Muller 变换。

8.2. 结果

图1 展示了在 Styblinski-Tang（2D）测试问题上，单次 KG-MR 和 HC 算法生成的样本示例。两种算法主要都在极限状态面附近采集样本。然而，KG-MR 将样本集中在极限状态面上对最优设计处失效概率有贡献的部分，而 HC 则将样本均匀分布在整个极限状态面周围。当极限状态面与问题边界相交时，这尤其会成为一个问题（见附录 E）。

图 2展示了 BO 过程中每个时点推荐设计的失效概率中位数和四分位距。第一列包含三个由高斯过程生成的问题，其余列包含可能存在模型失配的常见合成问题。中间两列的问题经过挑选，具有有趣的极限状态面；最后一列的问题则用于展示 KG-MR 和 TS-MR 预计不会优于 HC 和 EGRA 的情况。

单次KG-MR（osKG-MR）在12个问题中的10个上表现最佳或并列最佳，而离散KG-MR（dKG-MR）在较低维度下的表现与其相当。在 10+维度下，dKG-MR中的近似过于粗糙，减缓了收敛速度。TS-MR也表现出色，但在二维高斯过程、16维高斯过程、Branin和裁剪的 10 维 Styblinski-Tang 问题上稍逊一筹。HC和EGRA在二维问题上的表现通常较差，因为它们未能聚焦于极限状态面的相关区域（例如，图1和附录E）。在更高维度下，它们在高斯过程问题上的表现较差，但在合成示例上的表现尚可。

图2 最右列中的问题经过选择，旨在展示KG-MR和TS-MR算法不应具有优势的情况。二维二次函数具有圆形极限状态面，这意味着极限状态面的各个部分同等重要。因此，HC 算法的表现与KG-MR和TS-MR相当，而EGRA也并未落后太多。同样，对于六维Hartmann问题，较高的阈值会将极限状态面推向域的边缘，使HC 算法具备竞争力；不过在此问题上，EGRA的表现与Sobol搜索相近。最后，HC和EGRA算法在未裁剪的十维Styblinski-Tang问题上的实测表现良好，EGRA略胜于HC和osKG-MR。该问题具有高度多峰性，包含1024个局部极小值，其中848个可行。我们推测，HC和EGRA具有良好的局部搜索特性，会聚焦于初始设计中发现的某个优良局部极小值附近，学习极限状态面的信息。它们的计算成本也低于osKG-MR（见附录F）。众所周知，汤普森采样在高维空间中过度探索（Papenmeier等人，2025），在这个问题上毫无进展；而dKG-MR也毫无进展，因为近似集无法捕捉问题足够多的结构。

在失效概率并非极端的情况下，也可以得出类似结论；此时无需使用重要性采样，且 KG-MR 中的对数项会被省略（见附录 G）。值得注意的差异包括：osKG-MR 在所有问题中均表现最佳或并列最佳，而 TS-MR、HC 和 EGRA 在部分问题中有所改进。总体而言，所有算法的收敛速度都更快。附录 H 中的敏感性研究表明，KG-MR 方法对参数 N_u 和 N_v的选择相当稳健，但对于 dKG-MR，离散化规模 $N_x = |\mathcal{X}_{\text{disc}}|$ 在高维情况下会产生影响。

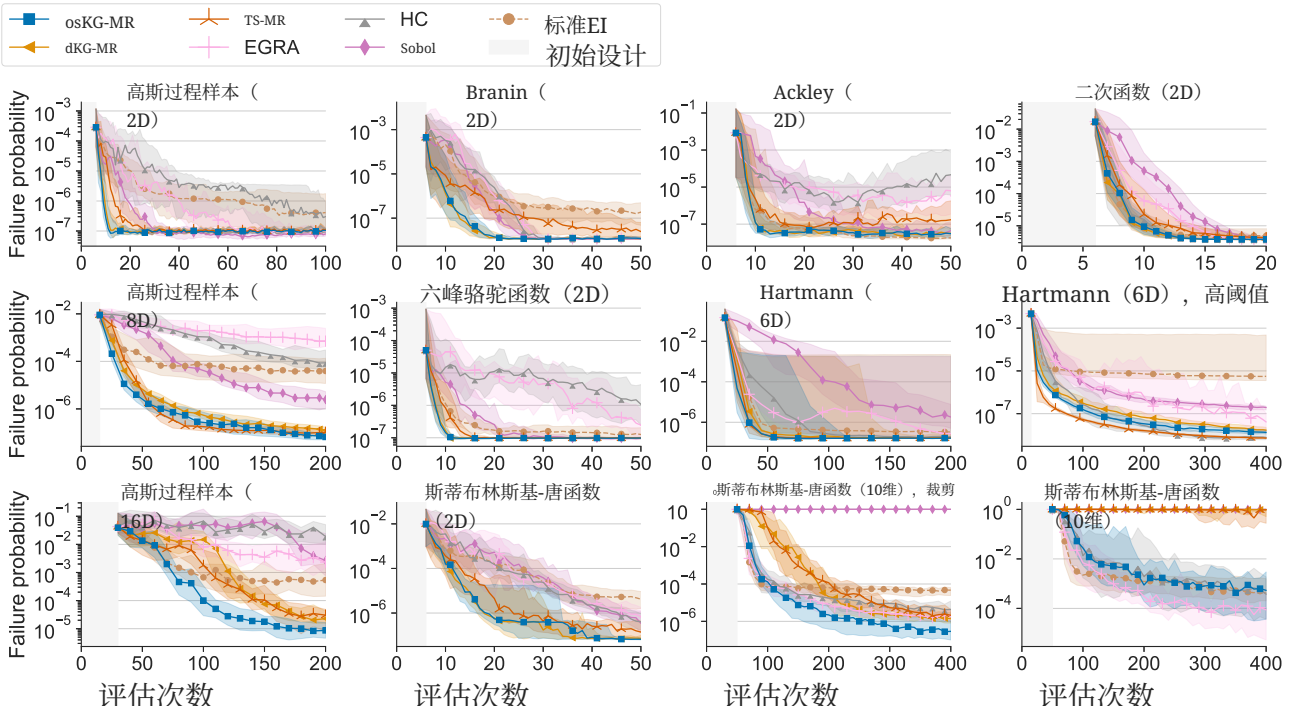

图*2.* 12个测试问题的失效概率。第一列包含GP测试问题，其余各列为存在模型失配可能的问题。最后一列包含预期KG-MR不会带来优势的问题。图中显示了推荐解的失效概率随高成本黑箱函数评估次数的变化。实线表示30次重复试验的失效概率中位数，阴影区域表示上四分位数和下四分位数。

9. 结论

本文介绍了两种不同的贝叶斯优化策略，分别针对等价的问题：最大化可靠性、最小化失效概率和最大化产率。本文提出了结合准蒙特卡罗和重要性采样的近似方法，并在一系列存在或不存在模型失配的合成问题上开展了实验。在二维测试问题中，KG-MR 方法表现同样出色，且收敛速度快于所有其他方法。在高维问题中，离散 KG-MR 采集函数中的近似过于粗糙，妨碍了收敛；但单次 KG-MR 仍表现最佳，有时与 TS-MR 并列。在所有情况下，KG-MR 的参数都很容易调整；唯一需要做出的选择，是在高维情况下如何为 osKG-MR 分配计算预算，以增加 N_u 和 N_v。KG-MR 算法通常比其他方法开销更大；在最慢的情况下，使用图形处理器进行优化需要约 30–40 秒（见附录 F）。总之，除非计算成本相对于评估 f的成本不再可以忽略，否则建议使用单次 KG-MR；在这种情况下，TS-MR 是一种快速的替代方案。

Huang & Chan 的算法（2010年）旨在处理具有多个极限状态函数的问题 f，本文尚未对此进行实证研究。不过，该框架可以十分自然地扩展到这种情形，因此这是一个值得未来研究的方向。同样，也可以处理相关问题的约束版本，例如 Bichon等人（2012年）所测试的情形。此外，该框架允许名义设计 $\boldsymbol{x}$ 与扰动 $\boldsymbol{u}$ 进行任意组合，并可通过实验进行测试，例如采用其他相关研究中使用过的乘法扰动（Cakmak 等人，2020年；Daulton 等人，2022年）。最后，采用核密度估计或 MCMC 进行自适应重要性抽样，可以通过确保用于估计 $\mathbb{E}_{\mathbf{u}}$ 的所有蒙特卡罗点都靠近极限状态面来提高效率。

致谢

作者谨此感谢 Tobias Grafke 就稀有事件建模所作的深刻讨论。第一作者获得了工程与物理科学研究委员会的资助，资助项目为华威大学系统数学 II 博士培训中心（编号 EP/S022244/1）。第二作者获得了弗拉芒政府通过弗兰德斯人工智能

智能研究计划提供的资助。计算设施由华威大学科学计算研究技术平台提供。

数据访问声明

这项研究完全是理论性的，没有数据支撑本文。

影响声明

本文介绍的研究旨在推动机器学习领域的发展。我们的工作可能会对社会产生多种影响，但我们认为没有哪一种需要在此特别指出。

参考文献

Au, S.-K. 与 Beck, J. L. 基于子集模拟估计高维小失效概率。《概率工程力学》，16(4):263–277，2001年。ISSN 0266-8920。网址 https://doi.org/10.1016/S0266-8920(01)00019-4。

Balandat, M.、Karrer, B.、Jiang, D.、Daulton, S.、Letham, B.、Wilson, A. G. 与 Bakshy, E. Botorch：一种高效蒙特卡洛贝叶斯优化框架。见 Larochelle, H.、Ranzato, M.、Hadsell, R.、Balcan, M. 与 Lin, H.（编），《神经信息处理系统进展》，第33卷，第21524–21538页。Curran Associates, Inc.，2020年。网址 https://proceedings.neurips.cc/paper/2020/file/f5b1b89d98b7286673128a5fb112cb9a-Paper.pdf。

Bect, J., Ginsbourger, D., Li, L., Picheny, V., 和 Vazquez, E. 用于估计失效概率的计算机实验序贯设计。《统计与计算》，22(3):773–793，2011年。ISSN 1573-1375。网址 https://doi.org/10.1007/s11222-011-9241-4。

Betlei, A., Vladimirova, M., Sebbar, M., Urien, N., Rahier, T., 和 Heymann, B. 提高在线系统中策略分配的成功概率。*AAAI*人工智能大会论文集，38(10): 11061–11068，2024年。ISSN 2374-3468。网址 https://doi.org/10.1609/aaai.v38i10.28982。

Bichon, B. J., Eldred, M. S., Swiler, L. P., Mahadevan, S., 和 McFarland, J. M. 面向非线性隐式性能函数的高效全局可靠性分析。*AIAA*期刊，46(10):2459–2468，2008年。ISSN 0001-1452。网址https://doi.org/10.2514/1.34321。

Bichon, B. J.、Eldred, M. S.、Mahadevan, S. 和 McFarland, J. M. 用于可靠性的高效全局代理建模优化。*Journal ofMechanicalDesign*，135(011009)，2012年。ISSN 1050-0472。网址 https://doi.org/10.1115/1.4022999。de Boer, P.-T.、Kroese, D. P.、Mannor, S. 和 Rubinstein, R. Y. 交叉熵方法教程。*Annalsof OperationsResearch*，134(1):19–67，2005年。ISSN 1572-9338。网址 https://doi.org/10.1007/s10479-005-5724-z。

Booth, A. S. 和 Renganathan, S. A. 使用高斯过程代理模型进行失效概率估计的两阶段设计。*JournalofQuality Technology*，57(5):500–516，2025年10月。网址 https://doi.org/10.1080/00224065.2025.2562868。

Breitung, K. 多元正态积分的渐近近似。*Journalof EngineeringMe-chanics*，110(3):357–366，1984年。ISSN0733-9399。网址 https://doi.org/10.1061/(ASCE)0733-9399(1984)110:3(357)。

Buathong, P.、Wan, J.、Astudillo, R.、Daulton, S.、Balandat, M. 和 Frazier, P. I. 基于部分评估的函数网络贝叶斯优化。载于 Salakhutdinov, R.、Kolter, Z.、Heller, K.、Weller, A.、Oliver, N.、Scarlett, J. 和 Berkenkamp, F.（编），国际机器学习会议，第235卷，*ProceedingsofMachineLearningResearch*，第4752–4784页。PMLR，2024年7月。网址 https://proceedings.mlr.press/v235/buathong24a.html。

Buckingham, J. M., Couckuyt, I., 和 Branke, J. 用于非凸两阶段随机优化问题的贝叶斯优化。*arXiv* 预印本，2025年2月a。网址https://arxiv.org/abs/2408.17387。

Buckingham, J. M., Rojas Gonzalez, S., 和 Branke, J. 用于解耦评估的多目标贝叶斯优化的知识梯度。载于 Singh, H., Ray, T., Knowles, J., Li, X., Branke, J., Wang, B., 和 Oyama, A.（编），进化多准则优化，第 15513 卷，《计算机科学讲义》，第 117–132 页，新加坡，2025年2月b。Springer Nature Singapore。国际标准书号 978-981-96-3538-2。网址 https://doi.org/10.1007/978-981-96-3538-2_9。

Cakmak, S.、Astudillo Marban, R.、Frazier, P. I. 和 Zhou, E. 风险度量的贝叶斯优化。载于《神经信息处理系统进展》，体积 33，第 20130–20141 页。Curran Associates, Inc.，2020年。网址 https://proceedings.neurips.cc/paper_files/paper/2020/file/e8f2779682fd11fa2067beffc27a9192-Paper.pdf。

Charnes, A. 和 Cooper, W. W. 机会约束下优化与满意化的确定性等价物。运筹学，11(1)：18–39，1963 年。ISSN 0030- 364X。网址 https://doi.org/10.1287/opre.11.1.18。

Daulton, S.、Cakmak, S.、Balandat, M.、Osborne, M. A.、Zhou, E. 和 Bakshy, E. 输入噪声下的鲁棒多目标贝叶斯优化。载于 Chaudhuri, K. 、 Jegelka, S. 、 Song, L. 、 Szepesvari, C. 、 Niu, G. 和 Sabato, S.（编），国际机器学习会议，第 162 卷，机器学习研究论文集，第 4831–4866 页。PMLR，2022年7月。网址 https://proceedings.mlr.press/v162/daulton22a.html。

Daulton, S.、Balandat, M. 和 Bakshy, E. 超体积知识梯度：一种用于部分信息下多目标贝叶斯优化的前瞻方法。载于第 *40* 届国际机器学习会议论文集，第 202 卷，机器学习研究论文集，第 7167–7204 页。PMLR，2023年7月。网址 https://proceedings.mlr.press/v202/daulton23a.html。

Dubourg, V., Sudret, B., 和 Bourinet, J.-M. 使用克里金代理模型和子集模拟进行基于可靠性的设计优化。结构与多学科优化，44(5):673–690，2011年。ISSN 1615-1488。网址 https://doi.org/10.1007/s00158-011-0653-8。

Durham, S., Flournoy, N., 和 Li, W. 一种用于最大化有利响应概率的序贯设计。加拿大统计学杂志，26(3):479–495，1998年。ISSN 1708-945X。网址 https://doi.org/10.2307/3315771。

El Amri, R., Le Riche, R., Helbert, C., Blanchet-Scalliet, C., 和 Da Veiga, S. 一种用于含不确定性的约束贝叶斯优化的采样准则。*SMAI*计算数学杂志，9:285–309，2023年。ISSN 2426-8399。网址 https://doi.org/10.5802/smai-jcm.102。

Fonseca, J. R.、Friswell, M. I. 和 Lees, A. W. 通过蒙特卡洛样本重加权实现高效鲁棒设计。《工程数值方法国际期刊》，69(11):2279–2301，2006年。国际标准书号 1097-0207。网址 https://doi.org/10.1002/nme.1850。

Frazier, P. I. 贝叶斯优化。载于《当代问题优化与建模最新进展》，第 255–278 页。INFORMS，2018年10月。网址 https://doi.org/10.1287/educ.2018.0188。

Frazier, P. I., Powell, W. B., and Dayanik, S. 相关正态信念的知识梯度策略。*INFORMS*计算杂志，21:599–613，2009年11月。ISSN 1091-9856。网址 https://doi.org/10.1287/ijoc.1080.0314。

Gablonsky, J. 和 Kelley, C. 局部偏向的 DIRECT算法形式。《全局优化期刊》，21(1):27–37，2001年。国际标准书号 1573-2916。网址 https://doi.org/10.1023/A:1017930332101。

Garnett, R. 贝叶斯优化。剑桥大学出版社，2023年。国际标准书号 9781108425780。网址 https://bayesoptbook.com/。

Han, J.、Zheng, Y.、Wang, K.、Yang, C. 和 Yuan, X. 基于自适应增量Kriging元模型的最坏情况鲁棒优化。《专家系统与应用》，260:125372，2025年1月。国际标准书号 0957-4174。网址 https://doi.org/10.1016/j.eswa.2024.125372。

Hasofer, A. M. and Lind, N. C. 精确且不变的二阶矩代码格式。《工程力学部期刊》，100(1):111–121，1974年。网址 https://doi.org/10.1061/JMCEA3.0001848。

Huang, Y.-C. and Chan, K.-Y. 用于概率约束空间中最大可靠性的改进高效全局优化算法。《机械设计期刊》，132(61002)，2010年。ISSN 1050-0472。网址 https://doi.org/10.1115/1.4001532。

Janusevskis, J.andLeRiche, R. 平均响应的基于克里金的估计与优化同步进行。《全局优化期刊》，55(2):313–336，2013。ISSN 1573-2916。网址 https://doi.org/10.1007/s10898-011-9836-5。

Jones, D. R., Perttunen, C. D., and Stuckman, B. E. 不使用利普希茨常数的利普希茨优化。《优化理论与应用杂志》，79 (1):157–181，1993年。ISSN 1573-2878。网址 https://doi.org/10.1007/BF00941892。

Jones, D. R., Schonlau, M., andWelch, W. J. 对高成本黑箱函数进行高效的全局优化。《全局优化期刊》，13:455–492，1998年12月。ISSN 09255001。网址 https://doi.org/10.1023/A:1008306431147。

Kloek, T. 和 van Dijk, H. K. 方程组参数的贝叶斯估计：蒙特卡罗积分的应用。*Econometrica*，46(1):1–19，1978年。ISSN 0012-9682。网址 https://doi.org/10.2307/1913641。

Knudde, N.、Couckuyt, I.、Shintani, K. 和 Dhaene, T. 面向可行域发现的主动学习。载于*2019*年第*18*届*IEEE*机器学习与应用国际会议（*ICMLA*），第567–572页，2019。网址 https://doi.org/10.1109/ICMLA.2019.00106。

Kroese, D. P.、Rubinstein, R. Y. 与 Glynn, P. W. 第2章——用于估计的交叉熵方法。载于 Rao, C. R. 与 Govindaraju, V.（编），统计学手册，体积31的统计学手册，第19–34页。爱思唯尔，2013。网址 https://doi.org/10.1016/B978-0-444-53859-8.00002-3。

Lemieux。蒙特卡罗与准蒙特卡洛抽样。施普林格统计学丛书。施普林格，第1版，2009。国际标准书号 978-0-387-78164-8 978-0-387-78165-5。网址 https://doi.org/10.1007/978-0-387-78165-5。

Nguyen, Q. P., Dai, Z., Low, B. K. H. 和 Jaillet, P. 基于高斯过程的风险价值优化。载于第*38*届国际机器学习会议论文集，第8063–8072页。PMLR，2021年7月。网址 https://proceedings.mlr.press/v139/nguyen21b.html。

Papaioannou, I., Betz, W., Zwirglmaier, K. 和 Straub, D. 用于子集模拟的马尔可夫链蒙特卡罗算法。概率工程力学，41:89–103，2015。ISSN 0266-8920。网址 https://doi.org/10.1016/j.probengmech.2015.06.006。

Papakonstantinou, K. G., Nikbakht, H. 和 Eshra, E. 用于估计高维问题中稀有事件概率的哈密顿马尔可夫链蒙特卡罗方法。概率工程力学，74:103485，2023年。ISSN 0266-8920。网址 https://doi.org/10.1016/j.probengmech.2023.103485。

Papenmeier, L., Cheng, N., Becker, S., 和 Nardi, L. 探索贝叶斯优化中的探索。在 第四十一届人工智能不确定性会议论文集，第 3388–3415 页。PMLR，2025年7月。网址 https://proceedings.mlr.press/v286/papenmeier25a.html。

Pelamatti, J., Le Riche, R., Helbert, C., 和 Blanchet-Scalliet, C. 不确定性下贝叶斯优化中的约束耦合与选择。《优化与工程》，25(1):373–412，2023年。ISSN 1573-2924。网址 https://doi.org/10.1007/s11081-023-09807-x。

Picheny, V.、Ginsbourger, D.、Roustant, O.、Haftka, R.T. 和 Kim, N.-H。自适应试验设计，用于目标区域的精确近似。《机械设计杂志》，132(071008)，2010年。ISSN 1050-0472。网址 https://doi.org/10.1115/1.4001873。Picheny, V.、Moss, H.、Torossian, L. 和 Durrande, N。贝叶斯分位数与期望分位数优化。载于 第三十八届人工智能不确定性会议论文集，第1623–1633页。PMLR，2022年8月。网址 https://proceedings.mlr.press/v180/picheny22a.html。

Qing, J., Couckuyt, I., 和 Dhaene, T. 一种考虑输入不确定性的稳健多目标贝叶斯优化框架。《全局优化期刊》，86(3): 693–711，2023年7月。ISSN 1573-2916。网址 https://doi.org/10.1007/s10898-022-01262-9。

Scott, W., Frazier, P. I. 和 Powell, W. B. 使用高斯过程回归，通过相关知识梯度对连续参数进行仿真优化。《*SIAM*优化学报》，21(3):996–1026，2011年。网址 https://doi.org/10.1137/100801275。

Sun, S., Li, X., Liu, H., Luo, K. 和 Gu, B. 采用缩放西格玛抽样，对高维变化空间中的罕见电路失效事件进行快速统计分析。*IEEE*集成电路与系统计算机辅助设计汇刊，34(7):1096–1109，2015。ISSN 1937-4151。网址 https://doi.org/10.1109/TCAD.2015.2404895。

Tabandeh, A., Jia, G. 和 Gardoni, P. 关于可靠性分析中重要性抽样方法的综述与评估。《结构安全》，97:102216，2022年。ISSN 0167-4730。网址 https://doi.org/10.1016/j.strusafe.2022.102216。

Tong, S. 和 Stadler, G. 基于大偏差理论的自适应重要性采样，用于高维问题中的罕见事件。《*SIAM/ASA*不确定性量化学报》，11(3):788–813，2023年。网址 https://doi.org/10.1137/22M1524758。

Tsai, S. C.、Park, C. 和 Chang, M. H. 考虑单个随机约束的基于元模型的仿真优化。《计算机与运筹学》，155: 106239，2023年。ISSN 0305-0548。网址 https://doi.org/10.1016/j.cor.2023.106239。

ur Rehman, S.、Langelaar, M. 和 van Keulen, F. 无约束问题的高效基于克里金法的鲁棒优化。《计算科学杂志》，5(6):872–881，2014年11月。ISSN 1877-7503。网址 https://doi.org/10.1016/j.jocs.2014.04.005。

Wang, M., Lv, W., Yang, F., Yan, C., Cai, W., Zhou, D. 和 Zeng, X. 模拟电路与

基于高斯过程回归和自适应良率估计的 SRAM电路良率优化。 *IEEE* 集成电路与系统计算机辅助设计汇刊，37(10)：1929–1942，2018年。ISSN 1937-4151。网址 https://doi.org/10.1109/TCAD.2017.2778061。

Weller, D. D., Hefenbrock, M., Beigl, M., and Tahoori, M. B. 使用基于贝叶斯优化失效率模型的核密度估计，快速高效地进行高西格玛良率分析与优化。 *IEEE* 计算机辅助设计集成电路与系统学报, 41(3):695–708, 2022年。ISSN 1937- 4151。网址 https://doi.org/10.1109/TCAD.2021.3064440。

Wilson, J., Borovitskiy, V., Terenin, A., Mostowsky, P., and Deisenroth, M. 高效地从高斯过程后验中采样函数。载于 第*37*届国际机器学习会议论文集，第10292–10302页。PMLR，2020年。网址 https://proceedings.mlr.press/v119/wilson20a.html。

Wilson, J., Borovitskiy, V., Terenin, A., Mostowsky, P., and Deisenroth, M. 高斯过程的路径条件化。机器学习研究杂志，22 (105):1–47，2021。ISSN 1533-7928。网址 http://jmlr.org/papers/v22/20-1260.html。

A. 理论结果的证明

在本附录中，我们给出正文中所述理论结果的证明。为方便起见，我们在证明之前重述所有结果。

引理 2.1（正式版本）。假设 $f\colon\ \mathcal{Y}\to\mathbb{R}$ 是连续的，且 $\boldsymbol{g}\colon\ \mathcal{X}\times\mathcal{U}\to\mathcal{Y}$ 是连续的。进一步假设 $\mathbb{P}_{\mathbf{u}}$ 的分布在 $\mathcal{Y}_{\text{feas}}$ 的极限状态面或边界上没有质量，因此对于任意 $\boldsymbol{x}\in\mathcal{X}$,

$$\mathbb{P}_{\mathbf{u}}\big(f\circ\boldsymbol{g}(\boldsymbol{x},\mathbf{u})=c\big)=0,$$
$$\mathbb{P}_{\mathbf{u}}\big(\boldsymbol{g}(\boldsymbol{x},\mathbf{u})\in\partial\mathcal{Y}_{\text{feas}}\big)=0.$$

则失效概率 $P\colon\ \mathcal{X}\to[0,1]$ 是连续的。如果 $\mathcal{X}$ 还是紧致的，那么对于某个$\boldsymbol{x}^*\in\mathcal{X}$， P可以取得其最小值。

证明。令 $\boldsymbol{x}\in\mathcal{X}$，并令 $\mathcal{S}_{\boldsymbol{x}}^{(1)}=\{\boldsymbol{u}\in\mathcal{U}\colon\ f\circ\boldsymbol{g}(\boldsymbol{x},\ \boldsymbol{u})=c\}$ 以及 $\mathcal{S}_{\boldsymbol{x}}^{(2)}=\{\boldsymbol{u}\in\mathcal{U}\colon\ \boldsymbol{g}(\boldsymbol{x},\ \boldsymbol{u})\in\partial\mathcal{Y}_{\text{feas}}\}$ 分别表示 $\boldsymbol{x}$ 的极限状态面的分量，并令 $\mathcal{S}_{\boldsymbol{x}}=\mathcal{S}_{\boldsymbol{x}}^{(1)}\cup\mathcal{S}_{\boldsymbol{x}}^{(2)}$。则 $\mathbb{P}(\mathbf{u}\in\mathcal{U}\setminus\mathcal{S}_{\boldsymbol{x}})=1$。注意到

$$P(\boldsymbol{x})=\mathbb{E}_{\mathbf{u}}\Big[\mathbb{I}_{\{f\circ\boldsymbol{g}(\boldsymbol{x},\mathbf{u})\geq c\}}\mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x},\mathbf{u})\in\mathcal{Y}_{\text{feas}}\}}+\mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x},\mathbf{u})\notin\mathcal{Y}_{\text{feas}}\}}\Big].$$

令 $\boldsymbol{u}\in\mathcal{U}\setminus\mathcal{S}_{\boldsymbol{x}}$，并设 $(\boldsymbol{x}_n)$ 是 $\mathcal{X}$ 中的一个序列，其中 $\boldsymbol{x}_n\to\boldsymbol{x}$ 以 $n\to\infty$ 为条件。则 $f\circ\boldsymbol{g}(\cdot,\ \boldsymbol{u})\colon\ \mathcal{X}\to\mathbb{R}$ 连续，且 $f\circ\boldsymbol{g}(\boldsymbol{x}_n,\ \boldsymbol{u})\to f\circ\boldsymbol{g}(\boldsymbol{x},\ \boldsymbol{u})$，当 $n\to\infty$ 时。此外，由于 $f\circ\boldsymbol{g}(\boldsymbol{x},\ \boldsymbol{u})\neq c$，可知 $\mathbb{I}_{\{f\circ\boldsymbol{g}(\boldsymbol{x}_n,\boldsymbol{u})\geq c\}}$ 最终为常数，因此 $\mathbb{I}_{\{f\circ\boldsymbol{g}(\boldsymbol{x}_n,\boldsymbol{u})\geq c\}}\to\mathbb{I}_{\{f\circ\boldsymbol{g}(\boldsymbol{x},\boldsymbol{u})\geq c\}}$。类似地，$\boldsymbol{g}(\boldsymbol{x},\ \boldsymbol{u})\notin\partial\mathcal{Y}_{\text{feas}}$，所以由同样的论证可得 $\mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x}_n,\mathbf{u})\in\mathcal{Y}_{\text{feas}}\}}\to\mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x},\mathbf{u})\in\mathcal{Y}_{\text{feas}}\}}$。
这对 $\mathbb{P}_{\mathbf{u}}$-几乎处处的 $\boldsymbol{u}\in\mathcal{U}$ 成立，因此由控制收敛定理， $P(\boldsymbol{x}_n)\to P(\boldsymbol{x})$，从而 P 在 $\boldsymbol{x}$ 处连续。由于 $\boldsymbol{x}$ 是任意的，P 处处连续。
最后，如果 $\mathcal{X}$ 也是紧致 ，则作为紧集上的连续函数， P在某个 $\boldsymbol{x}^*\in\mathcal{X}$处取得其最小值。 □

在证明引理5.1之前，我们正式陈述所需的正则性条件。
假设 A.1。

a) 对于任意 $\boldsymbol{y}_1'$，……， $\boldsymbol{y}_N'\in\mathcal{Y}$（不一定是优化过程中选取的点）以及任意 v_1，……， $v_N\in\mathcal{R}$，集合$\big\{\boldsymbol{y}\in\mathcal{Y}\colon\ \mathrm{Var}\big[f(\boldsymbol{y})\ \big|\ f(\boldsymbol{y}_1')=v_1,\ \ldots\ldots,\ f(\boldsymbol{y}_N')=v_N\big]=0\big\}$的勒贝格测度为零,
b) 对于任意 $\boldsymbol{x}\in\mathcal{X}$, $\boldsymbol{g}(\boldsymbol{x},\ \mathbf{u})$（其中 $\mathbf{u}\sim\mathbb{P}_{\mathbf{u}}$）相对于 $\mathcal{Y}$ 上的勒贝格测度绝对连续,
c) 对任意 $\boldsymbol{x}\in\mathcal{X}$, $\mathbb{P}_{\mathbf{u}}\big(\boldsymbol{g}(\boldsymbol{x},\ \mathbf{u})\in\partial\mathcal{Y}_{\text{feas}}\big)=0$。

第一个假设确保高斯过程不会过于退化。结合第二个假设，足以推导出 $\mathbb{P}(f\circ\boldsymbol{g}(\boldsymbol{x},\ \mathbf{u})=c\,|\,\mathcal{F}_n)=0$。也就是说，$\mathbb{P}_{\mathbf{u}}$和 f在现在具有随机性的极限状态面上联合赋予的测度为零。结合假设A.1 的最后一部分，这些条件与引理2.1中的条件类似。

引理5.1（正式版本）。假设 f 是样本连续高斯过程， $\boldsymbol{g}\colon\ \mathcal{X}\times\mathcal{U}\to\mathcal{Y}$ 是连续的，且假设A.1 成立。那么 $P_n\colon\ \mathcal{X}\to[0,1]$ 是连续的。若此外 $\mathcal{X}$ 是紧致的，则 P_n 在某个 $\boldsymbol{x}_n^*\in\mathcal{X}$处取得其最小值。

证明。将 $(\Omega\ \mathcal{F},\ \mathbb{P})$ 记为底层概率空间。

我们首先证明，对于所有 $\boldsymbol{x}\in\mathcal{X}$, $\mathbb{P}(f\circ\boldsymbol{g}(\boldsymbol{x},\mathbf{u})=c\,|\,\mathcal{F}_n)=0$ 几乎必然成立。事实上，假设 A.1 的前两部分表明，对于所有 $\boldsymbol{x}\in\mathcal{X}$, $\mathrm{Var}[f\circ\boldsymbol{g}(\boldsymbol{x},\mathbf{u})\,|\,\mathbf{u}\ \mathcal{F}_n]>0$ 几乎必然成立。因此， $f\circ\boldsymbol{g}(\boldsymbol{x},\mathbf{u})\,|\,\mathbf{u}\ \mathcal{F}_n$ 是一个几乎必然非退化的高斯随机变量，进而 $\mathbb{P}_{\mathbf{u}}(f\circ\boldsymbol{g}(\boldsymbol{x},\mathbf{u})=c\,|\,\mathbf{u}\ \mathcal{F}_n)=0$ 几乎必然成立。因此，

$$\mathbb{P}\big(f\circ\boldsymbol{g}(\boldsymbol{x},\mathbf{u})=c\,\big|\,\mathcal{F}_n\big)=\mathbb{E}\big[\mathbb{P}\big(f\circ\boldsymbol{g}(\boldsymbol{x},\mathbf{u})=c\,\big|\,\mathbf{u},\mathcal{F}_n\big)\,\big|\,\mathcal{F}_n\big]=0\quad\text{a.s.}$$

证明的其余部分与引理 2.1相同。

令 $\boldsymbol{x} \in \mathcal{X}$，并定义事件 $A = \{\omega \in \Omega\colon \ f(\boldsymbol{g}(\boldsymbol{x},\ \mathbf{u}(\omega);\ \omega)\ \geq c\}$ 以及 $B = \{\omega \in \Omega\colon \ \boldsymbol{g}(\boldsymbol{x},\ \mathbf{u}(\omega)) \notin \mathcal{Y}_{\text{feas}}\}$。这些确实是事件，因为 f 是样本连续的（因此是联合可测的），而 $\boldsymbol{g}$ 是连续的。于是 $P_n(\boldsymbol{x}) = \mathbb{E}[\mathbb{I}_A(1-\mathbb{I}_B) + \mathbb{I}_B \,|\, \mathcal{F}_n]$ 也有良好定义。

设 $(\boldsymbol{x}_m)$ 是 $\mathcal{X}$ 中的一个序列，其中 $\boldsymbol{x}_m \to \boldsymbol{x}$ 为 $m \to \infty$，并定义事件 A_m 和 B_m，其定义与 A 和 B 相同，但将 $\boldsymbol{x}_m$ 替换为 $\boldsymbol{x}$。也就是说，$A_m = \{\omega \in \Omega\colon \ f(\boldsymbol{g}(\boldsymbol{x}_m,\ \mathbf{u}(\omega);\ \omega)\ \geq c\}$，$B_m = \{\omega \in \Omega\colon \ \boldsymbol{g}(\boldsymbol{x}_m,\ \mathbf{u}(\omega)) \notin \mathcal{Y}_{\text{feas}}\}$ 且 $P_n(\boldsymbol{x}_m) = \mathbb{E}[\mathbb{I}_{A_m}(1-\mathbb{I}_{B_m}) + \mathbb{I}_{B_m} \,|\, \mathcal{F}_n]$。

由于 $\boldsymbol{g}$ 是连续的，我们得出 $f(\boldsymbol{g}(\cdot,\ \mathbf{u}))$ 几乎必然具有连续的样本路径。此外，$\mathbb{P}(f(\boldsymbol{g}(\boldsymbol{x},\ \mathbf{u})) = c \,|\, \mathcal{F}_n) = 0$，因此 $\mathbb{I}_{A_m}$ 几乎必然最终为常数，进而 $\mathbb{I}_{A_m} \to \mathbb{I}_A$ 为 $m \to \infty$。类似的论证可证明 $\mathbb{I}_{B_m} \to \mathbb{I}_B$ 几乎必然成立，因此由控制收敛定理可得 $P_n(\boldsymbol{x}_m) \to P_n(\boldsymbol{x})$，当 $m \to \infty$时。

最后，如果 $\mathcal{X}$ 也是紧致的，那么作为紧集上的连续函数，P_n 会在某个 $\boldsymbol{x}_n^* \in \mathcal{X}$处取得其最小值。 □

引理6.3. 式(10)中的*KG-MR* 采集函数处处非负。也就是说，

$$\forall n \geq 0\ \forall \boldsymbol{y} \in \mathcal{Y}_{\text{feas}} \quad \alpha_n^{\textit{KG-MR}}(\boldsymbol{y}) \geq 0.$$

证明。对所有 $\boldsymbol{x}' \in \mathcal{X}$

$$\max_{\boldsymbol{x}\in\mathcal{X}} R_{n+1}(\boldsymbol{x}) \geq R_{n+1}(\boldsymbol{x}')$$
$$\Rightarrow \quad \mathbb{E}\left[\max_{\boldsymbol{x}\in\mathcal{X}} R_{n+1}(\boldsymbol{x}) \,\middle|\, \mathcal{F}_n, \boldsymbol{y}_{n+1} = \boldsymbol{y}\right] \geq \mathbb{E}\left[R_{n+1}(\boldsymbol{x}') \,\middle|\, \mathcal{F}_n, \boldsymbol{y}_{n+1} = \boldsymbol{y}\right] \geq R_n(\boldsymbol{x}')$$

其中最后一个不等式由詹森不等式得出，因为 $z \mapsto -\log z$ 是凸的。这对所有 $\boldsymbol{x}'$ 都成立，因此

$$\mathbb{E}\left[\max_{\boldsymbol{x}\in\mathcal{X}} R_{n+1}(\boldsymbol{x}) \,\middle|\, \mathcal{F}_n, \boldsymbol{y}_{n+1} = \boldsymbol{y}\right] \geq \max_{\boldsymbol{x}\in\mathcal{X}} R_n(\boldsymbol{x})$$
$$\Rightarrow \quad \alpha_n^{\text{KG-MR}}(\boldsymbol{y}) = \mathbb{E}\left[\max_{\boldsymbol{x}\in\mathcal{X}} R_{n+1}(\boldsymbol{x}) \,\middle|\, \mathcal{F}_n, \boldsymbol{y}_{n+1} = \boldsymbol{y}\right] - \max_{\boldsymbol{x}\in\mathcal{X}} R_n(\boldsymbol{x}) \geq 0.$$

□

B. 单次知识梯度近似的优化

我们优化单次KG-MR采集函数（14），采用多起点L-BFGS-B。为多起点优化选择起始位置时必须谨慎，因为优化景观包含许多局部极小值。事实上，假设 $R_{n+1}(\boldsymbol{x};\ \boldsymbol{y}, z_j)$ 具有 m 个局部极小值，位于 $\boldsymbol{x}$。则（14）右端的被最大化函数具有 m^{N_v} 个局部极小值，分别位于 $\boldsymbol{x}_1$，……，$\boldsymbol{x}_{N_v}$，且各自具有其自身的最优 $\boldsymbol{y}$。

在我们的实验中，我们采用以下流程来选择初始点。首先，我们在大量 N_{raw} 个 $\boldsymbol{y}$ 值处评估离散KG-MR（15）。随后，我们使用玻尔兹曼采样从中随机选取 N_{restarts} 个值作为起始位置，并倾向于选择数值较高的位置。如果没有选中最佳值，我们就用最佳值替换最后一个采样值。这些 $\boldsymbol{y}$ 值及其对应的 $\boldsymbol{x}_1$、……、$\boldsymbol{x}_{N_v} \in \mathcal{X}_{\text{disc}}$ 用于初始化对（14）进行的多起点L-BFGS-B优化。BoTorch 中的这种玻尔兹曼采样实现是标准做法。

C. 基线算法

在本附录中，我们介绍文献中已有、并被用作基线的算法。

C.1. Huang &Chan（2010年）的方法

Huang & Chan （2010年）提出了一种贝叶斯优化方法，该方法会在四种采集函数之间进行选择。如果当前没有可行样本，则优化可行概率（17a）。否则，依次尝试另外三种采集函数，直到找到一个与所有现有样本的距离至少为 ε_s 的点。首先尝试（17b），它旨在最大化与现有观测的距离，同时使该点位于预测的极限状态面 Δ 范围内。如果未能找到这样的点，

（17c）会被最大化，以寻找极限状态面上彼此不相连的新区域，这称为隧穿。最后，如果这些方法均未奏效，则优化（17d），以寻找方差最大的点，从而降低模型的不确定性。列出的采集函数如下：

$$\alpha_n^{\mathrm{F}}(\boldsymbol{y}) = \mathbb{P}(f(\boldsymbol{y}) \leq c) = \Phi\left(\frac{c - \mu_n(\boldsymbol{y})}{\sqrt{k_n(\boldsymbol{y}, \boldsymbol{y})}}\right), \tag{17a}$$

$$\alpha_n^{\mathrm{LS}}(\boldsymbol{y}) = \mathbb{I}_{\{|\mu_n(\boldsymbol{y}) - c| \leq \Delta\}} \min_{i=1,\ldots,n} \frac{\|\boldsymbol{y}_i - \boldsymbol{y}\|}{\|\boldsymbol{b} - \boldsymbol{a}\|}, \tag{17b}$$

$$\alpha_n^{\mathrm{TN}}(\boldsymbol{y}) = \alpha_n^{\mathrm{F}}(\boldsymbol{y}) \min_{\substack{i=1,\ldots,n \\ \text{s.t. } f(\boldsymbol{y}_i) \leq c}} \frac{\|\boldsymbol{y}_i - \boldsymbol{y}\|}{\|\boldsymbol{b} - \boldsymbol{a}\|}, \tag{17c}$$

$$\alpha_n^{\mathrm{MV}}(\boldsymbol{y}) = \sqrt{k_n(\boldsymbol{y}, \boldsymbol{y})}. \tag{17d}$$

与之前一样，这里 $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ 是迄今为止 n 观测的位置，而 $\boldsymbol{a}, \boldsymbol{b}$ 定义了问题的边界，使得 $\mathcal{Y}_{\text{feas}} = \prod_{j=1}^{d_y} [a_j, b_j]$。请注意，在 (17) 中，我们沿用了 (Huang & Chan，2010年) 中的名称，但 α_n^{MV}是一个不同于第 6.1节中为 TS-MR 定义的采集函数。

我们使用 DIRECT（Jones等人，1993；Gablonsky& Kelley，2001年）优化 (17b) 中的极限状态探索函数 α_n^{LS}，并使用多起点L-BFGS-B优化所有其他采集函数。为确保数值稳定性，我们优化 (17a) 的对数，而不是直接优化采集函数。在 (Huang & Chan，2010年) 中，高斯过程超参数通过变差函数进行拟合；不过，为了与其他算法进行公平比较，我们对所有算法使用相同的高斯过程模型和超参数最大后验（MAP）估计方法。我们没有采用该研究中的终止准则，而是为了公平比较，让算法在与其他算法相同的固定评估预算下运行。在实验部分，我们将该算法记为 HC。

C.2. Bichon等人的方法(2008年)

Bichon等人 (2012年) 提出了一系列方法来解决基于可靠性的设计优化中的三个问题，其中包括式（1）所提出的问题。这项工作扩展了他们早期关于主动学习的研究，在该研究中，他们提出了高效全局可靠性分析（EGRA），以高效估计失效概率（Bichon等人，2008年）。他们提出了三种将贝叶斯优化与 EGRA 相结合的方法：

1. 嵌套方法：针对每个已访问的 $\boldsymbol{x} \in \mathcal{X}$，使用单独的代理模型来表示函数 $\boldsymbol{u} \mapsto f(\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}))$；2. 嵌套方法：使用单个代理模型表示 $h(\boldsymbol{x}, \boldsymbol{u}) = f(\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}))$；3. 序贯方法：交替使用贝叶斯优化寻找优质的 $\boldsymbol{x}$，并使用 EGRA 确定极限状态面。

他们对问题的一般表述并不假设存在复合函数 $f(\boldsymbol{g}(\boldsymbol{x}, \mathbf{u}))$，这意味着极限状态面可能因不同的 $\boldsymbol{x}$而异。然而，在我们的情形中，经过依赖于 $\boldsymbol{x}$ 的适当变换后，每个 $\boldsymbol{g}$对应的极限状态面实际上相同。 Bichon等人 (2012年) 在第5.3节（钢柱）中纳入了此类问题，并指出他们提出的算法在这种情况下本质上完全相同。此外，他们将问题表述为一个约束优化问题，其中确定性约束来自原始目标函数。在第一个设计点，会进行完整的 EGRA 分析；随后利用各设计点，通过约束贝叶斯优化在设计空间中搜索满足这些约束的点。然而，本文所考虑的问题中，式（1）不含来自原始目标函数的约束。因此，据我们所知，在式（1）所述且不存在黑盒约束的情形下，Bichon等人，2012年）提出的算法会简化为 EGRA。

The EGRA算法的灵感来自期望改进（Jones等人，1998年），其概括而言是最大化 g

$$\alpha_n^{\mathrm{EGRA}}(\boldsymbol{y}) = \mathbb{E}\left[\max\left\{\epsilon - \left|c - f(\boldsymbol{y})\right|, 0\right\} \,\middle|\, \mathcal{F}_n\right] \tag{18}$$

其中 $\epsilon = \kappa\sqrt{k_n(\boldsymbol{y}, \boldsymbol{y})}$ 与 f 的边缘后验标准差成正比。与 HC 一样，它会探索整个极限状态面，而不是集中于在最大可靠性点处对失效概率贡献最大的部分。

为确保比较有意义，在实验中，所有算法均使用相同的 GP 模型，并采用 MAP 估计的超参数，而不是使用 (Bichon等人，2008年) 或 (Bichon等人，2012年提出的任一模型。对于κ的取值，我们采用 $\kappa = 2$，这是 Bichon等人(2008年；2012年) 的建议。Bect等人(2011年) 发现，取值为 $\kappa = 0.5$ 时，估计结果更好

然而，对于失效概率的估计，结果更好；但在最大化可靠性方面，我们发现差异很小，如图3所示。

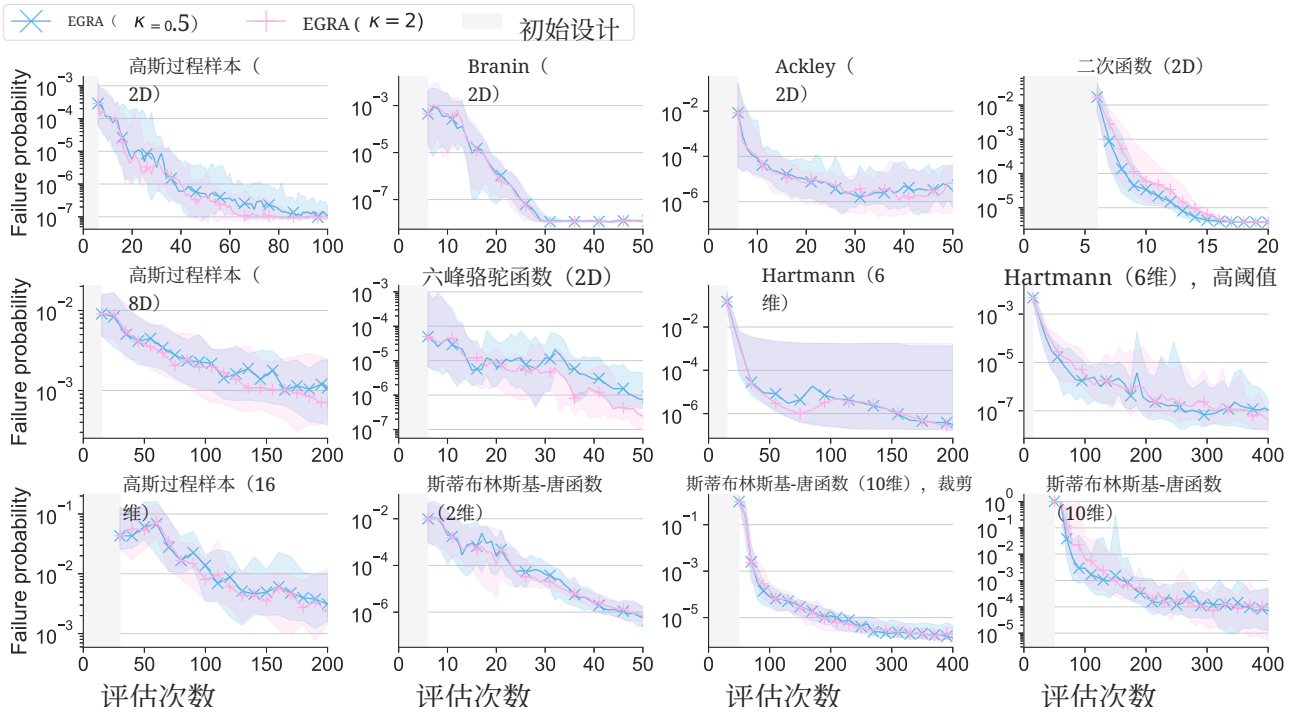

图3. 两种EGRA变体在12个测试问题上的失效概率（图2）。与推荐解相关联的失效概率显示为高成本黑箱函数评估次数的函数。实线表示30次重复的失效概率中位数，阴影区域表示上四分位数和下四分位数。

D. 进一步的实验细节

在本附录中，我们给出有关实验流程的进一步细节；这些细节对于结果的复现很重要，但在首次阅读时可以放心略过。

D.1. 指示函数平滑

在第7节中，我们简要介绍了对 $\mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x},\boldsymbol{u})\in\mathcal{Y}_{\text{feas}}\}}$ 的平滑近似，用于解决式（11）中的不可微问题。

我们希望避免在 f 之外 $\mathcal{Y}_{\text{feas}}$ 进行求值，因为例如，黑盒函数可能由仿真给出，而仿真在此区域之外会报错，或产生违反高斯过程建模假设的输出。因此，我们使用非对称平滑函数 $\iota(\boldsymbol{g}(\boldsymbol{x},\ \boldsymbol{u})\ ;\ \delta)\ \approx \mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x},\boldsymbol{u})\in\mathcal{Y}_{\text{feas}}\}}$，当 $\boldsymbol{g}(\boldsymbol{x},\ \boldsymbol{u})\ \notin \mathcal{Y}_{\text{feas}}$ 时，其值为零，并在可行区域内平滑地增大。这里 $\delta > 0$ 控制受平滑影响的边界周围区间的宽度。图4展示了这种近似对一维失效概率的贡献。

当 $\mathcal{Y}_{\text{feas}} = \prod_{j=1}^{d_y}[a_j, b_j]$ 是超矩形时，边界平滑所用的精确公式为

$$\iota(\boldsymbol{g}(\boldsymbol{x},\boldsymbol{u});\delta) = \prod_{j=1}^{d_y} G\left(\frac{g_j(\boldsymbol{x},\boldsymbol{u}) - a_j}{\delta}\right) G\left(\frac{b_j - g_j(\boldsymbol{x},\boldsymbol{u})}{\delta}\right) \tag{19}$$

其中 δ 是控制平滑区域宽度的平滑参数，g_j 是计算 $\boldsymbol{g}$第j个分量的函数。对于 $z \in (0,1)$，函数 $G(z) = F(z/1-z)$，其中 $F(\cdot)$是形状参数为0.5、率参数为1的伽马分布的累积分布函数。对于 $z \le 0$，我们定义 $G(z) = 0$；对于 $z \ge 1$，我们设定$G(z) = 1$。该公式源于将第 j个下界 a_j 视为随机变量 $A_j = a_j + \delta\frac{Z_j}{1+Z_j}$，其中$Z_j \sim$ 服从 Gamma(0.5,1)，上界同理。

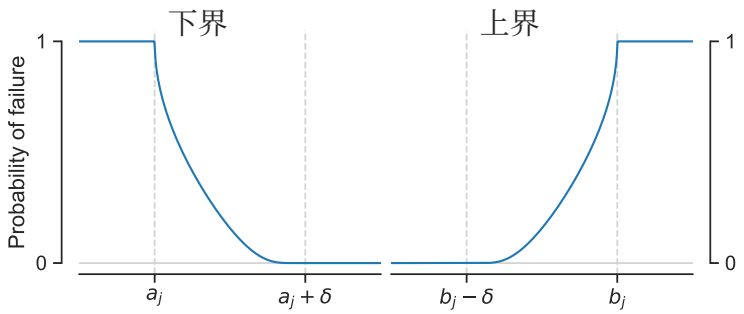

图 4。 近似 $1 - \iota(\boldsymbol{g}(\boldsymbol{x},\ \boldsymbol{u});\ \delta)$ 关于 $\mathbb{I}_{\{\boldsymbol{g}(\boldsymbol{x},\boldsymbol{u}) \notin \mathcal{Y}_{\text{feas}}\}}$ 的函数表示为 $\boldsymbol{g}(\boldsymbol{x},\ \boldsymbol{u})$。其定义见（19），并用于平滑（11）中的不连续性。

实践中，我们将 δ 保持得很小，取值为 $\delta = \min(0.05\ell_{\min}, 0.1)$，其中 $\ell_{\min}$ 是 $\mathcal{Y}_{\text{feas}}$的最小边长。

D.2. 模型参数

对于高斯过程超参数的先验，我们对输出尺度使用 $s^2 \sim \text{Gamma}(2,\ 0.15)$，并对 $\ell_1,\ \ldots,\ \ell_d \sim$ 使用 $\text{Gamma}(3,\ 10)$。这里，伽马分布采用形状参数和率参数约定进行参数化。常数均值没有先验（即非正规先验），观测噪声的方差固定为 $\sigma^2 = 0.01^2$，以确保数值稳定性。实际观测值$v_i = f(\boldsymbol{y}_i)$未添加噪声。

每收集到一条新观测，都会重新拟合高斯过程超参数。拟合前，我们会对观测进行归一化和标准化，使输入值落在 $[0, 1]^{d_y}$ 内，输出值的均值为零、方差为一。使用高斯过程进行预测时，会对输入和输出进行相应的变换与逆变换。这是 BoTorch 中的标准做法。

D.3. 优化推荐结果

推荐点 $\boldsymbol{x}_n^*(n = 1,\ 2,\ \ldots)$ 对所有算法，均使用（12）从高斯过程代理模型中生成。对于二维问题，通过在 1024 个 Sobol'点上评估后验均值并使用玻尔兹曼采样，选出 10 个重新启动点；必要时，手动确保将找到的最佳值纳入初始点。这些初始值随后使用 L-BFGS-B 进行优化。在（12c）中使用准蒙特卡洛估计 $N_u = 1024$，而（12d）中的边界平滑参数 δ 设为$\delta = \min(0.05\ell,\ 0.1)$，其中 ℓ 是 $\mathcal{Y}_{\text{feas}}$中的最小边长。为保证数值稳定性，会在高斯过程的方差中加入一个较小的微扰项。

对于高维问题（即维度大于 2），采用相同的做法，但随后使用 L-BFGS-B 对找到的最佳值进行微调，$N_u = 131,072$，因为根据经验，这些问题在 N_u 取较大值时效果更好。

D.4. 测试问题

本节总结了所使用的测试问题，以及阈值 c 和扰动分布 $\mathbb{P}_{\mathbf{u}}$。所有问题均采用加性扰动，$\boldsymbol{g}(\boldsymbol{x},\ \boldsymbol{u}) = \boldsymbol{x} + \boldsymbol{u}$。

BRANIN（2D）

可行域为 $\mathcal{Y}_{\text{feas}} = [-5, 10] \times [0, 15]$，阈值为 $c = 60$，公式为

$$f(\boldsymbol{y}) = \left(y_2 - \frac{5.1}{4\pi^2} y_1^2 + \frac{5}{\pi} y_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(y_1) + 10. \tag{20}$$

扰动分布为 $\mathbb{P}_{\mathbf{u}} = \mathcal{N}\left(0, \text{diag}(0.8, 0.8)^2\right)$。

六峰骆驼函数（2D）

可行域为 $\mathcal{Y}_{\text{feas}} = [-3,3]\times[-2,2]$，阈值为 $c = 2$，公式为

$$f(\boldsymbol{y}) = \left(4 - 2.1y_1^2 + \frac{1}{3}y_1^4\right)y_1^2 + y_1y_2 + 4\big(y_2^2 - 1\big)y_2^2. \tag{21}$$

扰动分布为 $\mathbb{P}_{\mathbf{u}} = \mathcal{N}\big(0,\ \text{diag(0.2,\ 0.1)}^2\big)$。

ACKLEY（2D）

可行域为 $\mathcal{Y}_{\text{feas}} = [-32.768, 32.768]^2$，阈值为 $c = 20.5$，公式为

$$f(\boldsymbol{y}) = -20\exp\left(-0.2\sqrt{\frac{1}{2}\big(y_1^2 + y_2^2\big)}\right) - \exp\left(\frac{1}{2}\big[\cos(2\pi y_1) + \cos(2\pi y_2)\big]\right) + 20 + e. \tag{22}$$

扰动分布为 $\mathbb{P}_{\mathbf{u}} = \mathcal{N}\big(0,\ \text{diag(3,\ 3)}^2\big)$。

二次函数（2维）

可行域为 $\mathcal{Y}_{\text{feas}} = [0,1]^2$，阈值为 $c = 0.09$，公式为

$$f(\boldsymbol{y}) = (y_1 - 0.3)^2 + (y_2 - 0.3)^2. \tag{23}$$

扰动分布为 $\mathcal{N}\big(0,\ \text{diag(0。06,\ 0。06)}^2\big)$。

STYBLINSKI-TANG（2维、10维）

用 $d = 2,\ 10$ 表示维度，可行域为 $\mathcal{Y}_{\text{feas}} = [-5,5]^d$。公式为

$$f(\boldsymbol{y}) = \frac{1}{2}\sum_{i=1}^{d}\big(y_i^4 - 16y_i^2 + 5y_i\big). \tag{24}$$

对于 2维问题，阈值为 $c = -20$，扰动分布为 $\mathbb{P}_{\mathbf{u}} = \mathcal{N}\big(0,\ \text{diag(0。25,\ 0。5)}^2\big)$。对于 10维问题，阈值为 $c = -300$，扰动分布为 $\mathbb{P}_{\mathbf{u}} = \mathcal{N}\big(0,\ \text{diag(0。4,\ 0。4,\ 0。4,\ 0。1,\ .....,\ 0。1)}^2\big)$。对于 10维问题的裁剪版本，可行域缩小为 $\mathcal{Y}_{\text{feas}} = [-5,0]\times[-5,5]^3\times[-5,0]^6$。

Hartmann（6D）

可行域为 $\mathcal{Y}_{\text{feas}} = [0,1]^6$，公式为

$$f(\boldsymbol{y}) = -\sum_{i=1}^{4}\alpha_i\exp\left(-\sum_{j=1}^{6}A_{ij}(y_j - P_{ij})^2\right) \tag{25}$$

其中

$$\alpha = \begin{pmatrix}1 & 1.2 & 3 & 3.2\end{pmatrix}^T$$

$$\boldsymbol{A} = \begin{pmatrix}10 & 3 & 17 & 3.5 & 1.7 & 8\\ 0.05 & 10 & 17 & 0.1 & 8 & 14\\ 3 & 3.5 & 1.7 & 10 & 17 & 8\\ 17 & 8 & 0.05 & 10 & 0.1 & 14\end{pmatrix}$$

$$\boldsymbol{P} = \begin{pmatrix}0.1312 & 0.1696 & 0.5569 & 0.0124 & 0.8283 & 0.5886\\ 0.2329 & 0.4135 & 0.8307 & 0.3736 & 0.1004 & 0.9991\\ 0.2348 & 0.1451 & 0.3522 & 0.2883 & 0.3047 & 0.6650\\ 0.4047 & 0.8828 & 0.8732 & 0.5743 & 0.1091 & 0.0381\end{pmatrix}.$$

对于标准版本，阈值为 $c = -1$，扰动分布为 $\mathbb{P}_{\mathbf{u}} = \mathcal{N}\big(0,\ \text{diag(0。05,\ .....,\ 0。05)}^2\big)$。对于高阈值版本，我们使用 $c = -0.05$，其中 $\mathbb{P}_{\mathbf{u}} = \mathcal{N}\big(0,\ \text{diag(0。07,\ .....,\ 0。07)}^2\big)$。扰动分布为 $\mathbb{P}_{\mathbf{u}} = \mathcal{N}\big(0,\ \text{diag(0。05,\ .....,\ 0。05)}^2\big)$。

高斯过程样本（2D、8D、16D）

高斯过程样本测试问题采用包含 1024 个傅里叶特征的随机傅里叶特征近似生成。高斯过程的均值为零，并采用 Mat´ern-5/2 核函数。输出尺度为 10，长度尺度约为 $0.2\sqrt{d}$，以便在维度增加时保持问题复杂度不变。阈值经过设定，使得 $\mathcal{Y}_{\text{feas}}$ 的体积中有 33% 不可行。扰动分布选为各向同性高斯分布，其尺度经过选择，以达到合适的最优失效概率。所用绝对值汇总于表 1。

可行域为 $\mathcal{Y}_{\text{feas}} = [0,1]^d$。

表 *1.* 用于生成GP测试问题的参数。

维度	各向同性长度尺度	阈值，c	失效体积	扰动分布，$\mathbb{P}_{\mathbf{u}}$
2	0.28	3.6	33%	$\mathcal{N}\left(0, 0.04^2 \boldsymbol{I}\right)$
8	0.57	1.2	33%	$\mathcal{N}\left(0, 0.06^2 \boldsymbol{I}\right)$
16	0.8	0.6	33%	$\mathcal{N}\left(0, 0.07^2 \boldsymbol{I}\right)$

D.5. 算法参数

在所有实验中，osKG-MR 和 dKG-MR 使用的幻想样本数量为 $N_v = 64$。在两种 KG-MR 算法和 TS-MR 中，用于计算 $\mathbf{u}$ 期望的准蒙特卡洛点数量为 $N_u = 64$。对于 dKG-MR，离散化规模$\mathcal{X}_{\text{disc}}$ 为 $N_x = 512$。在使用重要性采样的情况下（即针对极端失效概率的实验），我们采用的尺度因子为 $\tau = 3$。边界平滑参数 δ 设为$\min(0.05\ell_{\text{min}}, 0.1)$ 其中 ℓ_{min} 是可行域 $\mathcal{Y}_{\text{feas}}$的最小边长。TS-MR 中的阈值平滑参数设为 $\rho = 0.01$。对于 EGRA，大多数实验中我们采用$\kappa = 2$，但在附录C.2 中图3 的敏感性研究中采用。多起点 L-BFGS-B 在采集函数优化中使用 10 次重新启动。

（Huang & Chan，2010年）中没有说明如何设置 HC 算法中的参数 ε_s 和 Δ。我们根据 $\varepsilon_s = 0.01\ell_{\text{diagonal}}$ 选择点之间的最小距离，其中 ℓ_{diagonal} 是可行框对角线的长度 $\mathcal{Y}_{\text{feas}}$。这样一来，最小距离便可随问题维度增加而增大，因为高维问题自然包含更大的体积和更大的距离。这样选取的值也与（Huang & Chan，2010年）中报告的值相符。对于参数 Δ，它描述极限状态面周围感兴趣区域的宽度，（Huang & Chan，2010年）中没有报告任何取值，因此我们采用尽力而为的方法。由于它是在目标空间中的度量，直观上，所用的 Δ 应随极限状态面附近的梯度 $\|\nabla f\|$ 而缩放。对于二维问题，我们绘制函数景观，并根据等高线选择一个 Δ 值，使极限状态面周围形成一条窄带。对于高维问题，我们则根据公式（Hartmann 和 Styblinski-Tang）或函数值直方图（高斯过程问题）选择合适的值。当然，对于真正的黑盒问题，这种选取方法并不可行，而且会使我们的实验对 HC 算法不公平地有利。不过，由于 HC 的表现远不如 KG-MR 和 TS-MR，这并不会影响比较结论。HC 使用的参数以及所有算法采用的初始设计规模 n_0 列于表 2。

生成推荐点时使用的参数详见附录D.3。

E. 示例查询模式

图5 和6 展示了在二维高斯过程和六峰骆驼测试问题上测试的五种算法收集的前 30 个样本。osKG-MR 和 TS-MR 算法显然聚焦于极限状态面的相关部分，而 HC 和 EGRA 算法有时起初会聚焦于错误的部分。在一个案例中，HC 未能发现极限状态面的所有不连通部分。HC 和 EGRA 算法还尝试在边界上采样 $\partial\mathcal{Y}_{\text{feas}}$，在这些位置，$f$ 的值接近阈值 c。其他关于稳健贝叶斯优化的研究中也观察到了这一现象（Qing 等人，2023年），在这些情况下，算法无法有效降低边界之外高斯过程的不确定性。KG-MR 和 TS-MR 自然会纳入这样的知识：扰动解 $\boldsymbol{y} \in \mathcal{Y} \setminus \mathcal{Y}_{\text{feas}}$ 越过边界后将不可行，因此不会受到这一问题的影响。

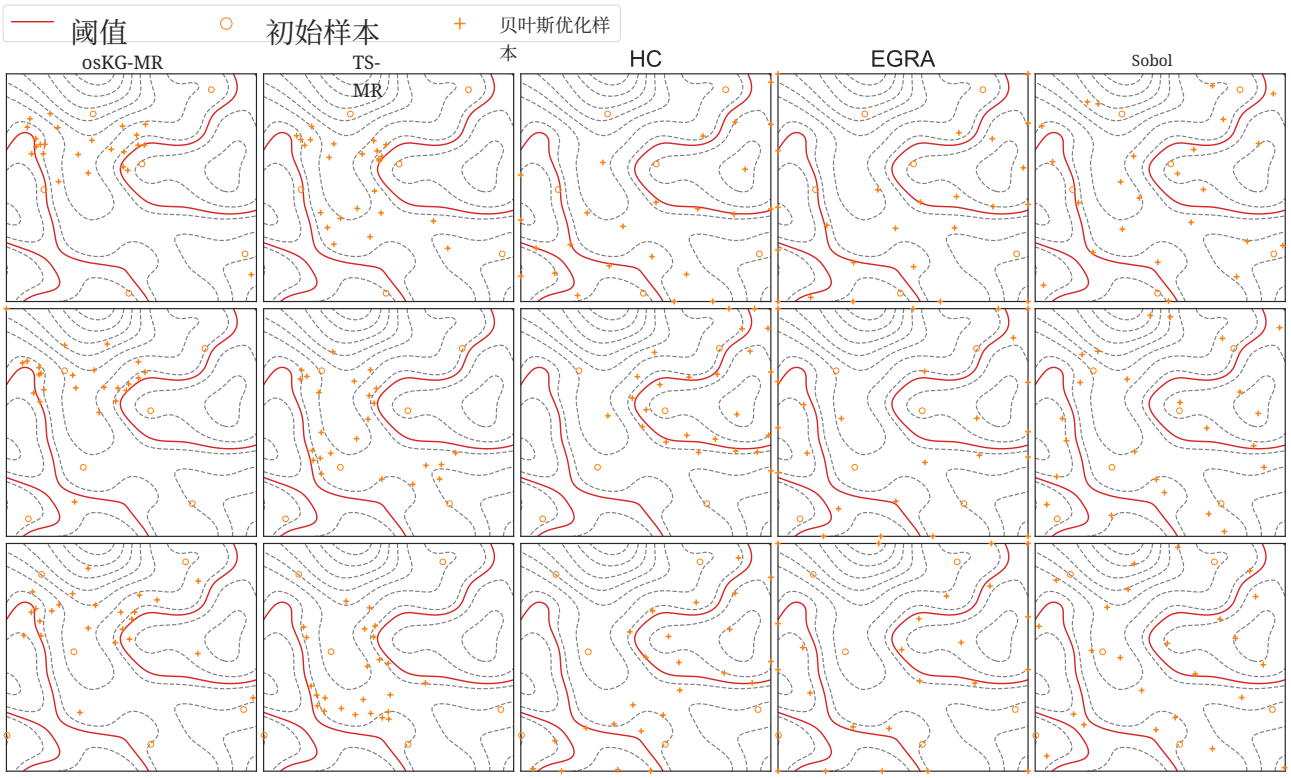

图 *5*。 在二维高斯过程测试问题上测试的五种算法的前 30 个采样位置。各行展示了三次不同的运行结果，分别对应三种不同的初始设计。

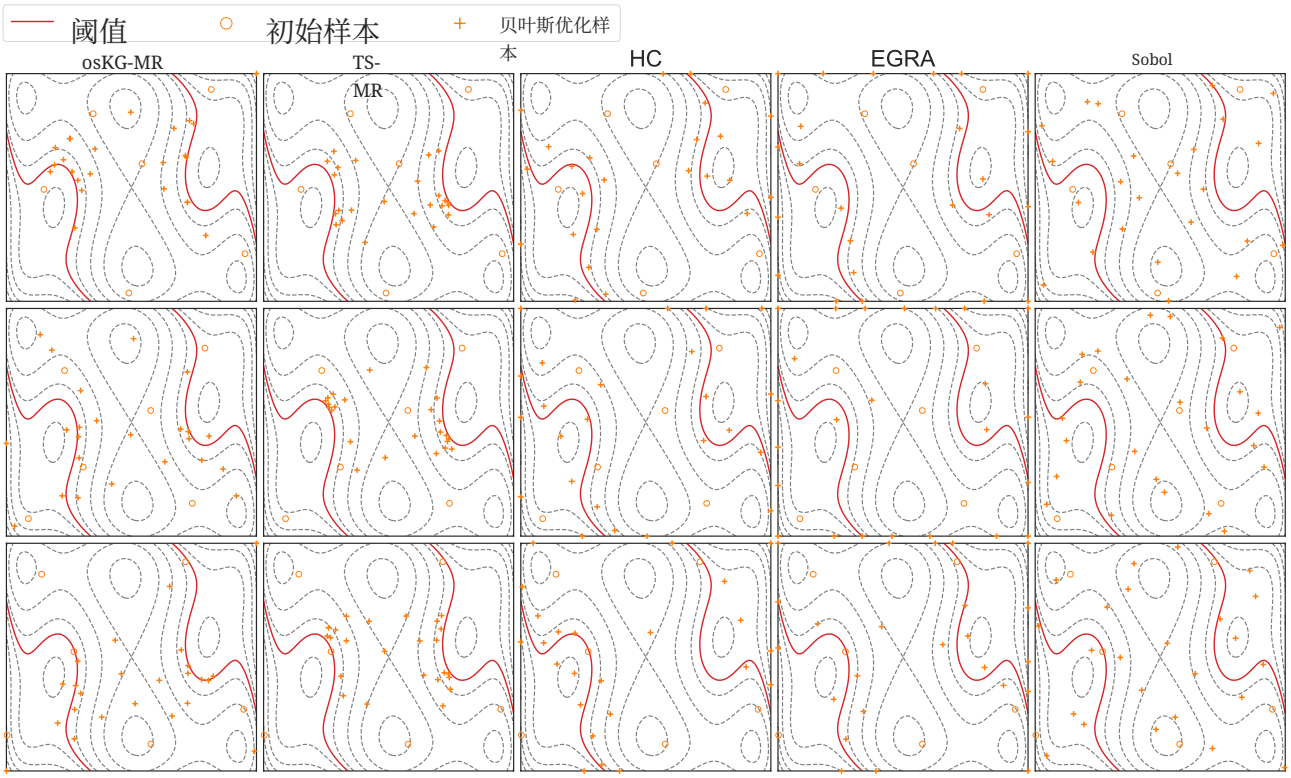

图 *6*。 在六驼峰测试问题上测试的五种算法的前 30 个采样位置。各行展示了三次不同的运行结果，分别对应三种不同的初始设计。

表2. 算法的实验特定参数。对于 HC，这些参数是样本之间的最小距离 ε_s 以及目标空间中的距离 Δ，它们共同定义了极限状态面周围的感兴趣带。此外，所有算法都使用初始设计的大小 n_0。

		HC	
实验	n_0	ε_s	Δ
高斯过程（2D）	6	0.014	0.6
高斯过程（8D）	15	0.028	0.6
高斯过程（16D）	30	0.04	0.6
Branin（2维）	6	0.21	10
六峰骆驼函数（2维）	6	0.072	0.4
Styblinski-Tang（2维）	6	0.14	10
Ackley（2维）	6	0.93	0.2
二次函数（2维）	6	0.014	0.01
Hartmann（6维）	15	0.024	0.02
Hartmann（6维），高阈值	15	0.024	0.02
Styblinski-Tang（10维）	50	0.32	10
Styblinski-Tang（10维），裁剪	50	0.22	10

F. 采集函数耗时

选择采集函数时，了解优化该采集函数所需的时间非常重要，因为这段时间应远短于评估黑盒函数所需的时间或成本 f。表3 和4 展示了图2中间两列所列问题对应的采集函数优化耗时的下四分位数和上四分位数。这些数据并非用于宣称某种算法优于另一种，而是为了说明高成本黑箱函数 f 的成本应达到什么程度。例如，如果评估 f 需要约一小时，那么花费30秒来优化裁剪后的10维 Styblinski-Tang 问题的单次KG-MR便微不足道。

KG-MR算法，尤其是单次KG-MR，计算量很大，因此对于$d > 2$的问题，它们使用图形处理器运行，而所有其他算法-问题组合仅使用中央处理器运行。所使用的图形处理器为英伟达 L40（48GB内存），搭配AMD 9354P Genoa（Zen 4）处理器的10个核心和59GB内存。对于仅使用中央处理器的实验，使用英特尔Xeon Platinum8268（CascadeLake）处理器的8个核心，并配备19.6GB内存。

表3。图2 中间两列所列二维测试问题在优化过程中的若干时点，优化采集函数所需时间的下四分位数和上四分位数。 n 表示当前正在优化的样本编号。对于二维测试问题，所有算法均使用8个中央处理器和总计29.6GB内存运行。

		Ackley（2D）	Branin（2D）	六峰 骆驼（2D）	斯蒂布林斯基 唐函数（2D）
离散KG-MR	$n = 7$	19s-23s	16s-19s	15s-19s	17s-19s
	$n = 50$	25s-33s	9秒-65秒	7秒-14秒	8秒-30秒
单次KG-MR	$n = 7$	21秒-30秒	24秒-33秒	20秒-27秒	23秒-27秒
	$n = 50$	27秒-38秒	9秒-23秒	25秒-35秒	22秒-35秒
TS-MR	$n = 7$	1.0秒–1.3秒	0.9秒–1.1秒	1.2秒–2.2秒	1.0秒–1.4秒
	$n = 50$	1.4秒–1.5秒	1.7秒-2.0秒	2.3秒-2.7秒	1.6秒-2.0秒
HC	$n = 7$	17秒-23秒	16秒-24秒	19秒-29秒	15秒–24秒
	$n = 50$	9秒–12秒	8秒–11秒	11秒–14秒	12秒–15秒
EGRA	$n = 7$	0.34秒–0.44秒	0.29秒-0.45秒	0.29秒-0.44秒	0.31秒-0.50秒
	$n = 50$	0.46秒-0.58秒	0.98秒-1.43秒	0.46秒-0.66秒	0.55秒-0.77秒

表4. 图2中间两列所示高维测试问题在优化过程多个时点优化采集函数所需时间的下四分位数和上四分位数。 2 其中， n 表示当前正在优化的样本编号。对于这些测试问题，KG-MR算法使用图形处理器（英伟达 L40，48GB内存）和10个中央处理器运行，总内存为59GB。其余测试问题使用8个中央处理器，总内存为29.6GB。

		Hartmann（6维）	斯蒂布林斯基-唐函数（10维），裁剪
离散KG-MR（图形处理器）	$n = 16$	1.6秒–2.3秒	
	$n = 100$	2.2秒-4.4秒	2.4秒-4.3秒
	$n = 200$	2.4秒-7.1秒	2.8秒-6.0秒
单次KG-MR（图形处理器）	$n = 16$	17秒-26秒	
	$n = 100$	27秒-39秒	25秒-38秒
	$n = 200$	28秒-44秒	29秒-37秒
TS-MR	$n = 16$	1.6秒-3.3秒	
	$n = 100$	4.2秒-5.9秒	1.9秒-2.8秒
	$n = 200$	4.6秒-5.8秒	2.6秒-3.5秒
HC	$n = 16$	1秒-13秒	
	$n = 100$	2.2秒-2.9秒	1秒–11秒
	$n = 200$	1.9秒–3.4秒	1.6秒–4.2秒
EGRA	$n = 16$	1.2秒–3.6秒	
	$n = 100$	1.5秒–2.0秒	1.7秒–2.5秒
	$n = 200$	1.5秒-2.2秒	2.3秒-4.3秒

G. 较不极端的最优失效概率问题

为补充主论文中的问题，我们还研究了最优失效概率不那么极端的问题上的算法。在这种情况下，我们不需要重要性采样，并且还从 KG-MR 的定义中去掉对数。也就是说，我们使用 $R_n(\boldsymbol{x}) = P_n(\boldsymbol{x})$，而不是公式（6）。

为了获得非极端的最优失效概率，我们对 GP测试问题进行了修改，使 $\mathcal{Y}_{\text{feas}}$为失效的比例随着维度增加而提高。此外，高斯扰动的标准差被增加到 0.1。用于生成问题的核函数和随机种子未作改变，因此实际样本 f 保持不变。这些变化总结于表 5中。对于非GP测试问题，阈值保持不变，但噪声方差的标准差增加到表 6所示的数值。

结果如图 7 所示。结论与图2中的极端情况类似，但也有一些例外。在高斯过程问题中，EGRA表现出更好的性能；而在（未裁剪的）Styblinski-Tang（10维）问题上，osKG-MR现在的性能与EGRA相当。在Styblinski-Tang（2维）问题上，TS-MR未能收敛到与其他算法相同的最优值。总体而言，所有算法的收敛速度都更快，因为方差更高的扰动分布意味着失效概率景观更简单，并且受极限状态面中小幅扰动的影响更小。

表5。 用于生成具有非极端失效概率的GP测试问题的参数。

维度	阈值	$, c$	失效体积	扰动分布	$, \mathbb{P}_{\mathbf{u}}$
2	3.6		33%	$\mathcal{N}(0,\ 0.1^2\boldsymbol{I})$	
8	-1.4		67%	$\mathcal{N}(0,\ 0.1^2\boldsymbol{I})$	
10	-4.5		90%	$\mathcal{N}(0,\ 0.1^2\boldsymbol{I})$	

表6. 非极端失效概率下非GP测试问题使用的扰动分布

问题	扰动分布
Branin（2D）	$\mathcal{N}(0, 2.5^2\boldsymbol{I})$
六峰骆驼函数（2D）	$\mathcal{N}(0,\ \text{diag}(0.6,\ 0.3)^2)$
Styblinski-Tang（2维）	$\mathcal{N}(0,\ \text{diag}(1,\ 2)^2)$
Ackley（2维）	$\mathcal{N}(0,\ 8^2\boldsymbol{I})$
二次函数（2维）	$\mathcal{N}(0,\ 0.12^2\boldsymbol{I})$
Hartmann（6维）	$\mathcal{N}(0, 0.1^2\boldsymbol{I})$
Hartmann（6维），高阈值	$\mathcal{N}(0, 0.18^2\boldsymbol{I})$
Styblinski-Tang（10维）	$\mathcal{N}(0, \text{对角矩阵}(0.8, 0.8, 0.8, 0.2, \ldots, 0.2)^2)$
Styblinski-Tang（10维），裁剪	$\mathcal{N}(0,\ \text{diag}(0.8,\ 0.8,\ 0.8,\ 0.2,\ \ldots,\ 0.2)^2)$

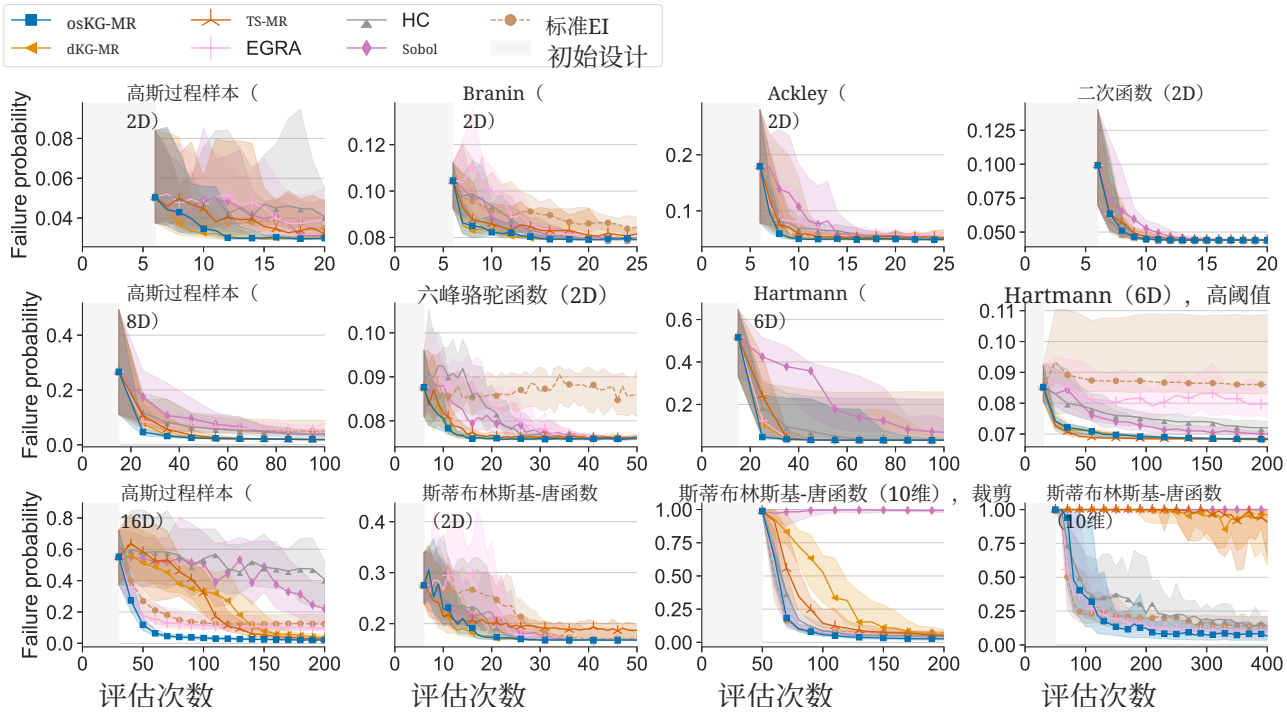

图7. 具有非极端最优失效概率的12个测试问题的失效概率。如图2所示，第一列包含高斯过程测试问题，其余各列为可能存在模型失配的问题。最后一列包含预计KG-MR不会具有优势的问题。图中显示了推荐解对应的失效概率随高成本黑箱函数评估次数的变化。实线表示30次重复试验中的失效概率中位数，阴影区域表示上四分位数和下四分位数。

H. 敏感性研究

本节研究单次和离散 KG-MR 近似对其参数的敏感性。研究的参数包括:

- 用于估计期望 $\mathbb{E}_{\mathbf{u}}$ 的准蒙特卡洛点数量 N_u ,
- 用于估计对下一个潜在样本的期望的准蒙特卡洛点数量 N_v v_{n+1},
- 离散化 $\mathcal{X}$ 中的点数 $N_x = |\mathcal{X}_{\text{disc}}|$ （仅适用于 dKG-MR）,
- 使用 L-BFGS-B 优化采集函数时的重新启动次数 N_{restarts} 。

我们在图2 中间两列的问题上测试了敏感性，结果见图 8和9。两种算法对所有参数都相当稳健。然而,

- 在裁剪的 10 维 Styblinski-Tang 问题中，dKG-MR 对 N_x 的大小很敏感,
- 较大的 N_u 对裁剪的 10 维 Styblinski-Tang 问题中的 osKG-MR 略有帮助,
- 较小的 N_v 对 Hartmann（六维）问题中的 osKG-MR 略有不利影响。

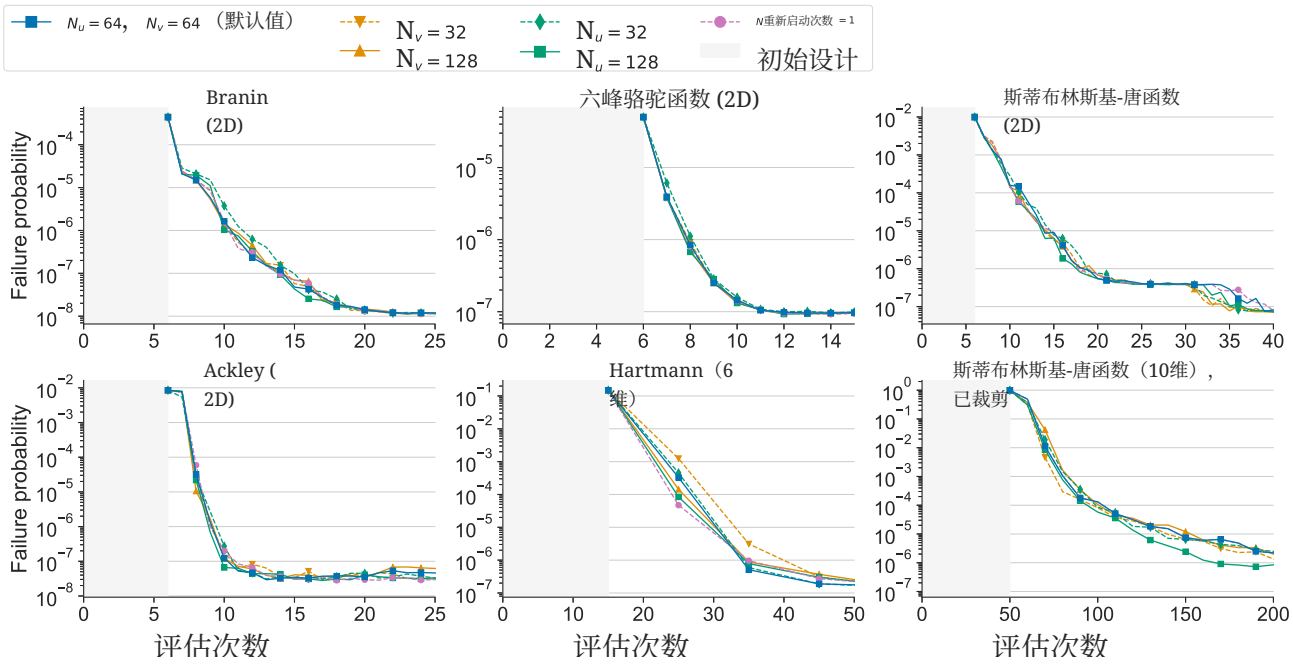

图*8*。 图2中间两列所示的最优失效概率极小的问题上，单次知识梯度-中位数回归近似（osKG-MR）的敏感性分析2。图中展示了推荐解对应的失效概率随高成本黑箱函数评估次数的变化。实线表示30次重复试验中的失效概率中位数。为清晰起见，省略了上四分位数和下四分位数。图例中仅显示与默认值不同的参数。

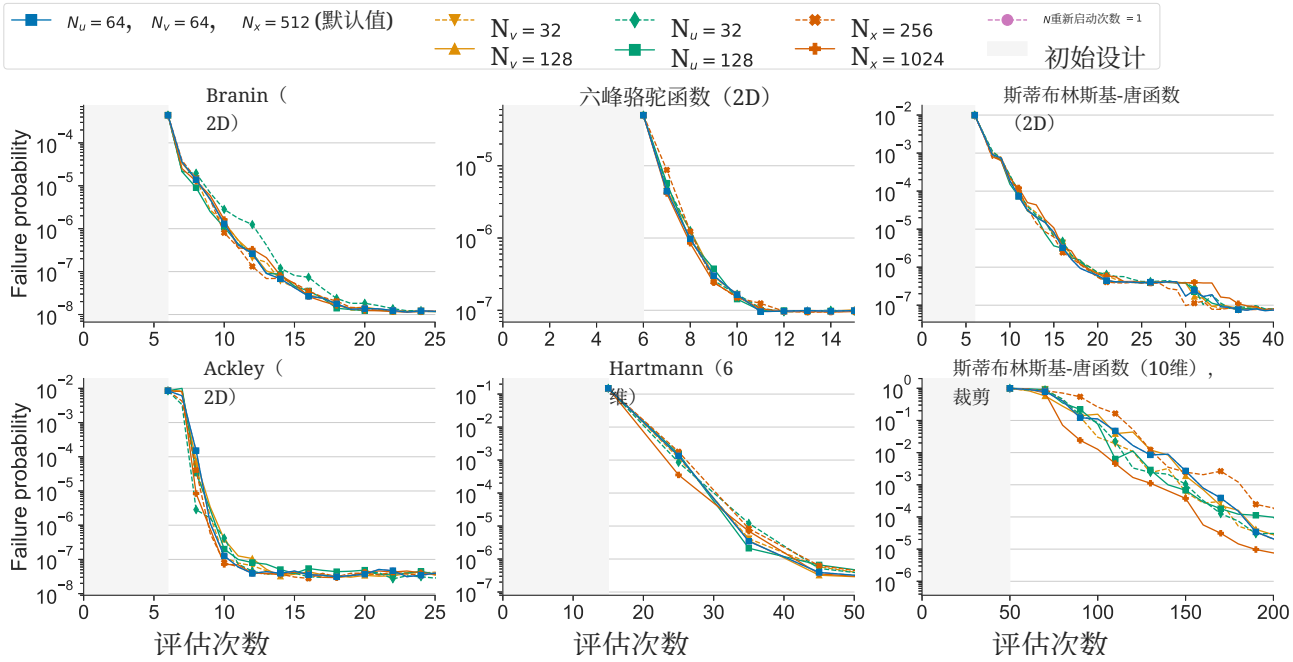

图*9*。 对图2中间两列所示的最优失效概率极小的问题进行离散 KG-MR 近似（dKG-MR）的敏感性分析。推荐解对应的失效概率表示为高成本黑箱函数评估次数的函数。实线表示30次重复试验中的失效概率中位数。为清晰起见，省略了上四分位数和下四分位数。图例中仅显示与默认值不同的参数。