arXiv:2302.01310v2 [stat.ML] 9 Oct 2024

采用解耦评估的多目标贝叶斯优化知识梯度

Jack M. Buckingham[1], Sebastian Rojas Gonzalez[2], 和 Juergen Branke[1]

[1] 华威大学，考文垂，英国[2] 代理建模实验室，根特大学，比利时

摘要。 多目标贝叶斯优化旨在以尽可能少的样本，寻找一组高成本目标之间权衡关系的帕累托前沿。在某些情况下，可以分别评估各个目标，并且每个目标都可对应不同的延迟或评估成本。目标之间的这种解耦为更快地学习帕累托前沿提供了机会，因为它可以避免不必要且高成本的评估。我们提出了一种基于标量化的知识梯度采集函数，该函数考虑了不同目标的评估成本。我们证明了在任意 D 维实紧搜索空间上，最优值估计量具有渐近一致性；并通过实验证明，该算法的性能与最先进水平相当，且显著优于始终同时评估两个目标的版本。

关键词： 贝叶斯优化 · 解耦目标。

1 引言

贝叶斯优化（BO）是一种在尽可能少地获取样本的同时，寻找评估成本高昂（例如计算或财务成本高）的目标函数全局最大值的技术 [11,17]。数据可以按顺序或批量收集，并使用贝叶斯模型（通常为高斯过程（GP））根据目前收集的样本估计目标函数及其估计的不确定性。随后，使用采集函数在探索输入空间中不确定性较高的区域与利用已知包含良好取值的区域之间进行权衡。在现实问题中，通常存在多个相互竞争的目标。在这类问题中，通常不存在唯一的最佳解；相反，目标是找到一组最优权衡方案，其中改善一个目标必然会使另一个目标退化。设计空间中的这组最优权衡方案称为帕累托集，而其在目标空间中的映像则称为帕累托前沿。

随机标量化是解决多目标贝叶斯优化问题的一种常用技术 [21,27]。每一步都会随机选择一个标量化向量，将多目标问题转化为单目标问题。通过在整个

优化过程中改变标量化权重，最终可以发现帕累托前沿上的不同点[26]。与使用随机标量化不同，一种更复杂的方法是对可能的标量化取期望，这意味着每一步都会考虑所有标量化，并且不会将时间浪费在对应于帕累托前沿上已充分了解部分的标量化上 [4]。

传统上，多目标贝叶斯优化假设在每个采样的设计向量处都会评估所有目标——这被称为耦合评估。然而，在实践中这并不总是合理。例如，在涡轮机设计中，效率等目标通过计算流体力学计算，而机械应力则通过有限元分析计算 [38]。类似地，计算涡轮机设计的效率可能比计算给定配置的成本更加昂贵。在这些可以进行解耦评估的问题中，我们可能希望更频繁地评估某些目标而非其他目标，原因有二。首先，某些目标的计算成本可能更低或计算速度更快，因此更频繁地评估这些目标是合理的。其次，某些目标可能比其他目标更难学习，因此，为找到帕累托前沿，这些目标需要更多次评估。

本文研究以下多目标优化问题：(1) 目标可以分别进行评估；(2) 目标共享资源（称为竞争性解耦）；(3) 所有目标在一定程度上都具有较高成本；以及 (4) 某些目标比其他目标更难学习。我们的贡献如下：

1. 我们将自然的 多属性知识梯度采集函数 [4] 扩展到目标函数分别进行评估且具有已知恒定延迟/成本的情形。2. 我们提供了理论保证，证明对于采用任意线性效用函数的决策者，该算法都将收敛于全局最优解。3. 我们针对两类双目标问题开展了实验：一类问题的目标函数长度尺度不同，另一类问题的观测噪声方差不同。我们的算法与当前最先进的方法相比具有竞争力，并且我们证明，对于所有算法，利用目标函数之间的解耦都能提高收敛速度。4. 我们证明了对标量化取期望的重要性，而不是使用一系列随机标量化。

相关工作 对具有不同延迟的多个解耦目标（也称为异质目标）进行优化，主要受到进化算法（EAs）领域的关注 [1,10,9,25,39]。 [2] 提供了近期的综述。进化算法是高效的黑箱优化器，但往往需要大量数据。另一方面，贝叶斯优化在黑箱优化的基础上还具有数据效率。例如，Loka 等人 [24] 将对廉价目标函数的评估直接纳入用于贝叶斯优化的两个基于超体积的采集函数中。因此，在优化采集函数时，这些目标函数会被评估许多次。这种方法仅适用于廉价目标函数的评估成本极低的情况，例如目标函数具有已知的解析公式。

在基于信息的采集函数中，通常会考虑解耦目标的情况 [19,20,35]。这些算法试图最大化帕累托前沿或帕累托集与所提议观测之间的互信息。另一种称为超体积知识梯度（HVKG）的采集函数 [12]，会选择下一个采样点，以最大化与高斯过程代理模型后验均值相关联的超体积增量。知识梯度和基于信息的采集函数都可以自然地扩展到部分观测，例如解耦目标的情况。

2 问题定义

设 $\boldsymbol{f}^*\colon \mathcal{X} \to \mathbb{R}^M$ 是定义在空间 $\mathcal{X} \subset \mathbb{R}^D$ 上的未知多目标函数，其各个分量可以分别评估。[3] 设 $\{u_{\boldsymbol{\lambda}}\colon \boldsymbol{\lambda} \in \Lambda\}$ 是一组单调效用函数 $u_{\boldsymbol{\lambda}}\colon \mathbb{R}^M \to \mathbb{R}$，每个函数代表一位可能的决策者。由于效用函数是单调的，其最大化解将是帕累托最优解。我们的目标是找到帕累托最优解的一个有限子集 $S \subset \mathcal{X}$，使随机决策者从 S 中进行选择时的期望效用最大化。也就是说，我们求解

$$S^* \in \operatorname*{arg\,max}_{S\subset\mathcal{X},\,|S|\le N_S} \mathbb{E}_{\boldsymbol{\lambda}\sim p(\boldsymbol{\lambda})}\left[\max_{\boldsymbol{x}\in S} u_{\boldsymbol{\lambda}}(\boldsymbol{f}^*(\boldsymbol{x}))\right] \tag{1}$$

其中某个 $N_S \in \mathbb{N}$。

假设我们可以根据以下模型对 $\boldsymbol{f}^*$ 进行观测

$$y_m = f^*_m(\boldsymbol{x}) + \varepsilon_m, \quad \varepsilon_m \sim \mathcal{N}(0, \sigma_m^2) \tag{2}$$

其中 $m \in \{1,\dots,M\}$ 是正在评估的 $\boldsymbol{f}^*$ 分量，$\boldsymbol{x} \in \mathcal{X}$ 是输入，ε_m 表示观测噪声。进一步假设，评估每个目标函数的成本 c_1，$\dots$，c_M 是常数且已知。

该问题需要迭代选择采样位置 $\boldsymbol{x}_1$、$\boldsymbol{x}_2$、$\cdots \in \mathcal{X}$ 以及要评估的目标 m_1、m_2、$\cdots \in \{1,\dots,M\}$，并据此推导出方程 **(1)** 的近似解 S_1、S_2，**.....**，使其相对于总采样成本 $C_N = \sum_{n=1}^{N} c_{m_n}$，快速收敛至最优期望效用。

3 背景

贝叶斯优化 贝叶斯优化包含两个要素：目标函数的概率模型，以及通过最大化来确定下一个采样位置的采集函数。

对于概率模型，会在所有可能的目标函数上设定多输出高斯过程先验分布 [3]，通常还会配合一个假设观测含有加性高斯噪声的观测模型

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \Sigma). \tag{3}$$

[3] 我们将使用 $\boldsymbol{f}^*$ 表示未知的真实函数，并使用 $\boldsymbol{f}$ 表示高斯过程（GP）模型。

通常假设协方差矩阵 Σ 为对角矩阵，这意味着添加到不同目标函数中的噪声彼此独立。

多输出高斯过程 $\boldsymbol{f} \sim \mathcal{GP}(\boldsymbol{\mu}, \mathbf{K})$ 完全由其向量值均值函数 $\boldsymbol{\mu}: \mathcal{X} \to \mathbb{R}^M$ 以及半正定矩阵值协方差函数 或核函数 $\mathbf{K}: \mathcal{X} \times \mathcal{X} \to \mathbb{R}^{M\times M}$决定。本文中，我们将$\mathbf{K}$ 设为对角值函数；若再配合对角的 Σ，这就等价于使用相互独立的单输出高斯过程对各目标进行建模，$f_m \sim \mathcal{GP}(\mu_m, k_{m,\,m})$。

给定输入 $\boldsymbol{x}_1, \ldots\ldots, \boldsymbol{x}_N \in \mathcal{X}$ 处的 N 观测 $y_1, \ldots\ldots, y_N \in \mathbb{R}$，且观测具有方差为 σ^2 的加性高斯噪声时，单输出高斯过程 $f \sim \mathcal{GP}(\mu, k)$ 的后验分布仍是一个高斯过程，其均值函数$\mu_N: \mathcal{X} \to \mathbb{R}$ 和协方差函数 $k_N: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ 由 [31] 给出。

$$\mu_N(\boldsymbol{x}) = \mu(\boldsymbol{x}) + k(\boldsymbol{x}, \mathbf{X})\big(k(\mathbf{X}, \mathbf{X}) + \sigma^2\big)^{-1}\big(\boldsymbol{y} - \mu(\mathbf{X})\big), \tag{4a}$$

$$k_N(\boldsymbol{x}, \boldsymbol{x}') = k(\boldsymbol{x}, \boldsymbol{x}') - k(\boldsymbol{x}, \mathbf{X})\big(k(\mathbf{X}, \mathbf{X}) + \sigma^2\big)^{-1}k(\mathbf{X}, \boldsymbol{x}'). \tag{4b}$$

这里，我们用$\mathbf{X}$ 表示由观测位置 $\boldsymbol{x}_1, \ldots\ldots, \boldsymbol{x}_N$堆叠而成的矩阵，并用 $\mu(\mathbf{X}) \in \mathbb{R}^N$和 $k(\mathbf{X}, \mathbf{X}) \in \mathbb{R}^{N\times N}$ 表示批量评估。关于贝叶斯优化的完整介绍，请参阅 [32,14,17]。

多属性知识梯度（*maKG*）贝叶斯优化的第二个要素是采集函数。 [4]中介绍的多属性知识梯度采集函数所依据的原理是：评估预算用尽后，对于给定的 $\boldsymbol{\lambda} \in \Lambda$，我们将推荐输入 $\boldsymbol{x}^* \in \mathcal{X}$，使 GP 模型的效用后验均值最大。

更正式地说，令 $\mu_N(\boldsymbol{x}'; \boldsymbol{\lambda}) = \mathbb{E}_N[u_{\boldsymbol{\lambda}}(\boldsymbol{f}(\boldsymbol{x}'))]$ 表示在效用函数 $u_{\boldsymbol{\lambda}}$ 下，给定 $\boldsymbol{x}_1$、……、$\boldsymbol{x}_N$ 处 $\boldsymbol{f}$ 的值时， $\boldsymbol{f}(\boldsymbol{x}')$ 的效用后验均值。在优化过程中，令 $\boldsymbol{x} \in \mathcal{X}$ 表示潜在的下一个采样位置，并令 $\boldsymbol{y}^{\boldsymbol{x}} \in \mathbb{R}^M$ 表示根据模型 (3) 在该位置可能得到的观测值这一随机变量。需要注意的是，此处我们假设观测到了所有 M 个目标。应用效用函数 $u_{\boldsymbol{\lambda}}$ 后，将其他某个位置 $\boldsymbol{x}' \in \mathcal{X}$ 处的后验均值记为 $\mu_{N+}(\boldsymbol{x}'; \boldsymbol{x}, \boldsymbol{\lambda}) = \mathbb{E}_N\left[u_{\boldsymbol{\lambda}}(\boldsymbol{f}(\boldsymbol{x}')) \mid \boldsymbol{y}^{\boldsymbol{x}}\right]$。从当前 $\boldsymbol{f}$ 后验均值的最大值中减去相应数值，并对 $\boldsymbol{y}^{\boldsymbol{x}}$ 取期望，便得到效用 $u_{\boldsymbol{\lambda}}$ 的 知识梯度，

$$\alpha_{\mathrm{KG}}(\boldsymbol{x}; \boldsymbol{\lambda}) = \mathbb{E}_N\left[\max_{\boldsymbol{x}' \in \mathcal{X}} \mu_{N+}(\boldsymbol{x}'; \boldsymbol{x}, \boldsymbol{\lambda})\right] - \max_{\boldsymbol{x}' \in \mathcal{X}} \mu_N(\boldsymbol{x}'; \boldsymbol{\lambda}). \tag{5}$$

一种简单的贝叶斯优化方法是在每次迭代时，为效用函数 $u_{\boldsymbol{\lambda}}$ 随机选择一个新的参数 $\boldsymbol{\lambda} \in \Lambda$，并优化 $\alpha_{\mathrm{KG}}(\cdot, \boldsymbol{\lambda})$。另一种方法则是对应用于标量化目标的采集函数关于标量化权重取期望。这种方法用于 [4] 中的多属性知识梯度。用公式表示为，

$$\alpha_{\mathrm{maKG}}(\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{\lambda} \sim p(\boldsymbol{\lambda})}\big[\alpha_{\mathrm{KG}}(\boldsymbol{x}; \boldsymbol{\lambda})\big]. \tag{6}$$

作者将范围限定在线性效用函数上， $u_{\boldsymbol{\lambda}}(\boldsymbol{f}(\boldsymbol{x}')) = \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}')$。这样做的优点是，线性效用函数与取期望运算可交换。

无论是使用随机标量化，还是对标量化取期望，都需要选择分布 $p(\boldsymbol{\lambda})$。Paria 等人 [27] 利用它来表示决策者在了解帕累托前沿不同区域时的偏好。本文将重点讨论线性标量化的情形，其中 $p(\boldsymbol{\lambda})$ 是 $\mathbb{R}^M$ 中标准单纯形上的均匀分布。

4 成本加权多目标知识梯度

为了将 maKG 扩展到多个目标，我们只需注意到，多输出高斯过程可以根据每次仅观测 $\boldsymbol{f}$ 一个坐标的观测进行条件化。在解耦设置中，我们有观测 $y_{1、m_1}$、.....、$y_{N、m_N}$，其中 $y_{n、m_n} = f_{m_n}(\boldsymbol{x}_n) + \varepsilon_{n、m_n}$，对每个 n 均如公式（2）所示。

对于给定的 $\boldsymbol{\lambda} \in \Lambda$，记 $\mu_N(\boldsymbol{x}'; \boldsymbol{\lambda}) = \mathbb{E}_N[u_{\boldsymbol{\lambda}}(\boldsymbol{f}(\boldsymbol{x}'))]$ 为给定 N 观测条件下 $\boldsymbol{f}$ 的后验均值。令 $\boldsymbol{x} \in \mathcal{X}$ 表示潜在的下一个采样位置，令 $m \in \{1, \ldots, M\}$ 表示潜在的下一个目标函数，并令 $y_m^{\boldsymbol{x}} = f_m(\boldsymbol{x}) + \varepsilon_m$ 表示相应潜在观测的随机变量。记 $\mu_{N+}(\boldsymbol{x}'; \boldsymbol{x}、 m、 \boldsymbol{\lambda}) = \mathbb{E}_N[u_{\boldsymbol{\lambda}}(\boldsymbol{f}(\boldsymbol{x}')) \mid y_m^{\boldsymbol{x}}]$ 为给定 $y_m^{\boldsymbol{x}}$ 和 N 实际观测条件下的后验均值。将这两个函数的最大值相减，并对 $y_m^{\boldsymbol{x}}$ 取期望，即得到多目标知识梯度，

$$\alpha_{\mathrm{MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda}) = \mathbb{E}_N\left[\max_{\boldsymbol{x}' \in \mathcal{X}} \mu_{N+}(\boldsymbol{x}'; \boldsymbol{x}, m, \boldsymbol{\lambda})\right] - \max_{\boldsymbol{x}' \in \mathcal{X}} \mu_N(\boldsymbol{x}'; \boldsymbol{\lambda}). \tag{7}$$

这是单个效用 $u_{\boldsymbol{\lambda}}$ 的结果。对 $\boldsymbol{\lambda} \sim p(\boldsymbol{\lambda})$ 取期望可得

$$\overline{\alpha}_{\mathrm{MOKG}}(\boldsymbol{x}, m) = \mathbb{E}_{\boldsymbol{\lambda} \sim p(\boldsymbol{\lambda})}\left[\alpha_{\mathrm{MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda})\right]. \tag{8}$$

在目标函数的评估成本各不相同的情况下，我们将 MOKG 除以评估所提议目标函数的成本，从而得到单位成本的价值。这是 Snoek 等人 [33] 在单目标情形下采用的方法。本文的主要贡献是提出了成本加权多目标知识梯度（*C-MOKG*）采集函数，

$$\alpha_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda}, \boldsymbol{c}) = \frac{1}{c_m} \alpha_{\mathrm{MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda}), \tag{9a}$$

$$\overline{\alpha}_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{c}) = \frac{1}{c_m} \overline{\alpha}_{\mathrm{MOKG}}(\boldsymbol{x}, m), \tag{9b}$$

其中， $\boldsymbol{c}$ 是与 M 个目标函数分别对应的成本向量。

4.1 高效计算与优化

我们首先介绍一种方法，在采用离散近似的随机标量化情形下计算并优化 C-MOKG。随后，我们将结合准蒙特卡洛近似，使用该方法在对标量化取期望时计算并优化 C-MOKG。

与 [4], 一样，我们现在将范围限制为线性效用函数， $u_{\boldsymbol{\lambda}}(\boldsymbol{f}(\boldsymbol{x}')) = \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}')$，以利用期望与线性函数可交换这一事实。

离散近似 早期引入知识梯度的研究[15,16] 聚焦于离散搜索空间。这启发了一种常见的计算策略，用于在输入维度较低时计算知识梯度 [40,28,29], : 对输入空间使用离散近似 $\mathcal{X}$进行内层优化，同时保留完整的连续空间来确定建议的下一个采样位置。具体而言，设 $\mathcal{X}_{\text{disc}}$ 是近似 $\mathcal{X}$ 的有限子集。那么

$$\alpha_{\text{MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda}) \approx \hat{\alpha}_{\text{MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda}) = \mathbb{E}_N \left[\max_{\boldsymbol{x}' \in \mathcal{X}_{\text{disc}}} \mu_{N+}(\boldsymbol{x}'; \boldsymbol{x}, m, \boldsymbol{\lambda}) \right] - \max_{\boldsymbol{x}' \in \mathcal{X}_{\text{disc}}} \mu_N(\boldsymbol{x}'; \boldsymbol{\lambda}). \quad (10)$$

由于期望与线性算子可交换，我们可以利用公式 (4a) 将 $\mu_{N+}(\boldsymbol{x}'; \boldsymbol{x}, m, \boldsymbol{\lambda})$ 写成假设观测值 $y_m^{\boldsymbol{x}}$ 的仿射函数，

$$\mu_{N+}(\boldsymbol{x}'; \boldsymbol{x}, m, \boldsymbol{\lambda}) = \boldsymbol{\lambda} \cdot \tilde{\boldsymbol{\mu}}(\boldsymbol{x}') + \boldsymbol{\lambda} \cdot \tilde{\mathbf{K}}_{:,m}(\boldsymbol{x}', \boldsymbol{x}) \frac{y_m^{\boldsymbol{x}} - \tilde{\mu}_m(\boldsymbol{x})}{\tilde{k}_{m,m}(\boldsymbol{x}, \boldsymbol{x}) + \sigma_m^2}. \quad (11)$$

这里，$\tilde{\boldsymbol{\mu}}$ 和 $\mathbf{K}$ 分别是给定截至目前的 N 个观测值时 $\boldsymbol{f}$ 的后验均值与协方差函数。因此，可以使用 [16] 中的算法 2 高效地计算 $\hat{\alpha}_{\text{MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda})$。此外，所得解析表达式是确定性的，并且可以通过自动微分求出其关于候选输入 $\boldsymbol{x}$ 的导数。因此，我们可以使用多起点 L-BFGS-B 等确定性梯度优化器来寻找全局最大值。

然后，我们可以通过针对每个 m 优化结果并选取最大值，对 m 和 $\boldsymbol{x}$ 同时优化 C-MOKG。事实上，写作 $[M] = \{1, \ldots, M\}$,

$$\max_{\substack{\boldsymbol{x} \in \mathcal{X} \\ m \in [M]}} \hat{\alpha}_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda}, \boldsymbol{c}) = \max_{m \in [M]} \frac{1}{c_m} \max_{\boldsymbol{x} \in \mathcal{X}} \hat{\alpha}_{\text{MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda}). \quad (12)$$

在优化知识梯度采集函数时，通常会忽略公式（10）中的第二项，因为它相对于 $\boldsymbol{x}$ 是常数。然而，在对公式（12）进行外层最大化时，不能忽略这一项，因为因子 $1/c_m$ 意味着它不再相对于 m 为常数。

准蒙特卡洛 上一节中的离散近似足以在只有一个标量化的情况下优化 C-MOKG。我们可以通过采用（准）蒙特卡洛近似（qMC），将其扩展为一种在对标量化取期望时优化 C-MOKG 的方法。事实上，

$$\overline{\alpha}_{\text{C-MOKG}}(\boldsymbol{x}, m) \approx \frac{1}{Q} \sum_{j=1}^{Q} \hat{\alpha}_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda}^{(j)}). \quad (13)$$

这里，$\boldsymbol{\lambda}^{(1)}, \ldots, \boldsymbol{\lambda}^{(Q)}$ 是一个大小为 Q 的准蒙特卡洛样本。总和中的每一项都可以使用上一节介绍的离散化技术计算，并且可以通过多起点 L-BFGS-B 对平均值进行优化。

4.2 理论结果

令 $\Lambda \subset \mathbb{R}^M$ 表示标准单纯形。对于使用线性效用函数定义的 C- 多目标知识梯度，以下结果成立： $u_{\boldsymbol{\lambda}}(\boldsymbol{f}(\boldsymbol{x}'))=\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}')$，其中 $\boldsymbol{\lambda}\in\Lambda$。

我们的首个结果表明，成本感知多目标知识梯度处处非负。这是知识梯度采集函数的一项标准结果，也正因如此，我们无需像使用期望改进时那样，在期望内部取正部。

引理 1. 成本感知多目标知识梯度的两种形式均非负。也就是说，对于所有 $\boldsymbol{x}\in\mathcal{X}$、$m\in\{1,\dots,M\}$ 和所有 $\boldsymbol{\lambda}\in\Lambda$,

$$\alpha_{\text{C-MOKG}}(\boldsymbol{x},m;\,\boldsymbol{\lambda})\geq 0 \quad and \quad \overline{\alpha}_{\text{C-MOKG}}(\boldsymbol{x},m)\geq 0,$$

几乎必然。

这是因为随机过程的期望的最大值不大于该过程最大值的期望；证明见附录A。

我们的主要理论贡献保证，使用 C-MOKG 选择样本时，无论采用对标量化取期望还是随机标量化，算法推荐点所对应的标量化目标函数值都会收敛到最优值。我们通过不再区分 $\boldsymbol{f}$ 和 $\boldsymbol{f}^*$，假设不存在模型失配。具体而言，由于此处我们将 $\boldsymbol{f}^*$ 视为高斯过程，而非函数，因此以下结果应理解为对所有可能的 $\boldsymbol{f}^*$ 整体成立，而不是对任何单个样本成立。

对于每个 $N\in\mathbb{N}_0$ 和每个偏好向量 $\boldsymbol{\lambda}\in\Lambda$，令

$$\boldsymbol{x}^*_{N,\boldsymbol{\lambda}}\in\operatorname*{arg\,max}_{\boldsymbol{x}\in\mathcal{X}}\mathbb{E}_N[\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x})] \tag{14}$$

为一个随机变量，它使第 N 阶段标量化目标函数的后验均值达到最大。因此，$\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}^*_{1,\boldsymbol{\lambda}})$ $\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}^*_{2,\boldsymbol{\lambda}})\dots$是优化过程中每个阶段都采用推荐点时所得到的（无噪声）标量化目标函数值序列。以下定理说明，该序列收敛于标量化隐藏目标函数的真实最大值 $\boldsymbol{\lambda}\cdot\boldsymbol{f}$。

定理 2（C-MOKG 的一致性）。假设 $\mathcal{X}\subset\mathbb{R}^D$ 是紧集，并按公式 (14)定义 $\boldsymbol{x}^*_{n,\boldsymbol{\lambda}}$。使用随机标量化或对标量化取期望的 *C-MOKG* 时，我们有

$$\forall\boldsymbol{\lambda}\in\Lambda,\quad \boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}^*_{N,\boldsymbol{\lambda}})\to\max_{\boldsymbol{x}\in\mathcal{X}}\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}) \quad as \quad N\to\infty$$

几乎必然收敛且在均值意义下收敛。

这一结果的证明基于 Bect等人 [8]的工作。证明过程表明，对于 $\boldsymbol{\lambda}$ 的每一种选择，$\alpha_{\text{C-MOKG}}$ 都会收敛到零，并以此证明后验均值收敛到真实的目标函数（可能相差一个常数）。定理 2 随即得证。有关这些结果的证明，请参阅附录 A。

5 实验

更频繁地评估一个目标函数而非另一个目标函数，主要有两个原因：一是它的评估成本相对较低，二是它更难学习。因此，当昂贵的目标函数更容易学习时，改进幅度最大。本节针对合成双目标问题开展的实验通过两种情形展示了这一点：分别采用较短的长度尺度和加入观测噪声，使成本较低的目标函数更难学习。

合成问题 我们在两类测试问题上测试该算法，每类包含 100 个测试问题，每个问题都有两个输入维度和两个目标。为避免模型失配，我们使用 Matérn-$5/2$ 核函数，分别从不同的高斯过程中独立采样生成这两个目标函数。在第一类测试问题中，第一目标的长度尺度为 0.2，第二目标的长度尺度为 1.8。这一区别使第一目标更难学习。在这一类问题中，采样时不添加观测噪声。在第二类测试问题中，两个目标的长度尺度均为 0.4，并通过加入观测噪声使第一目标更难学习。所添加噪声的标准差为 1，与同样为 1 的输出尺度相比相当大。第二目标没有噪声。在这两种情况下，我们都假设第一目标的成本或延迟为 1，而第二目标为 10。用于生成这些测试问题的超参数完整说明见附录 B。

贝叶斯遗憾性能指标 根据方程（1）中的问题定义，自然的指标是多目标优化中常用的 R2 性能指标的一种变体 [18,37]。该指标假设效用函数由参数化定义，且参数的分布已知。我们假设$u_{\boldsymbol{\lambda}}(\boldsymbol{f}^*(\boldsymbol{x})) = \boldsymbol{\lambda} \cdot \boldsymbol{f}^*(\boldsymbol{x})$，其中 $\boldsymbol{\lambda}$在标准单纯形上均匀分布。解集 S 的质量即为期望效用，也就是：

$$R2(S) = \mathbb{E}_{\boldsymbol{\lambda} \sim p(\boldsymbol{\lambda})} \left[\max_{\boldsymbol{x} \in S} \boldsymbol{\lambda} \cdot \boldsymbol{f}^*(\boldsymbol{x}) \right]. \tag{15}$$

我们的算法会为每个目标函数返回一个后验均值预测， $\boldsymbol{f}$ 作为真实函数 $\boldsymbol{f}^*$的估计。基于此，使用 NSGA-II [13]从预测的帕累托集 $S_N \subset \mathcal{X}$ 中推导出一个包含 1000 个点的近似。具有特定效用函数 $\boldsymbol{\lambda}$ 的决策者会选择一个解$\boldsymbol{x}^*_{N,\ \boldsymbol{\lambda}} \in \arg\max_{\boldsymbol{x} \in S_N} \mathbb{E}_N[\boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x})]$。然而，由于得到的帕累托前沿基于预测的值，从该帕累托前沿中选出的解不一定能实现他们所期望的效用。我们可以通过仅记录所选解的真实效用来对此进行处理， $\boldsymbol{\lambda} \cdot \boldsymbol{f}^*(\boldsymbol{x}^*_{N,\ \boldsymbol{\lambda}})$。从任意解的最大可能效用 $\boldsymbol{x} \in \mathcal{X}$中减去该值，并对 $\boldsymbol{\lambda}$ 取期望，便得到贝叶斯遗憾，

$$R_N = \mathbb{E}_{\boldsymbol{\lambda} \sim p(\boldsymbol{\lambda})} \left[\max_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{\lambda} \cdot \boldsymbol{f}^*(\boldsymbol{x}) - \boldsymbol{\lambda} \cdot \boldsymbol{f}^*(\boldsymbol{x}^*_{N,\boldsymbol{\lambda}}) \right]. \tag{16}$$

这与 [27],中贝叶斯遗憾的构造类似，但允许模型与真实函数之间存在不匹配。

实践中，我们使用大小为 $N_{\boldsymbol{\lambda}} = 1024$ 的准蒙特卡洛（Sobol' [34]）样本（$\boldsymbol{\lambda}^{(j)}$）$_{j=1}^{N_{\boldsymbol{\lambda}}}$ 来估计该期望。通过以下步骤计算对 $\boldsymbol{x} \in \mathcal{X}$ 的最大化：首先使用 NSGA-II 为 $\boldsymbol{f}^*$ 在帕累托集上生成 1000 个点的近似值 $\mathcal{X}^*_{\text{帕累托}}$，然后选出最大值。总之，我们计算

$$R_N \approx \hat{R}_N = \frac{1}{N_{\boldsymbol{\lambda}}} \sum_{j=1}^{N_{\boldsymbol{\lambda}}} \left(\max_{\boldsymbol{x} \in \mathcal{X}^*_{\text{Pareto}}} \boldsymbol{\lambda}^{(j)} \cdot \boldsymbol{f}^*(\boldsymbol{x}) - \boldsymbol{\lambda}^{(j)} \cdot \boldsymbol{f}^*(\boldsymbol{x}^*_{N,\boldsymbol{\lambda}}) \right). \tag{17}$$

超体积性能指标 仅使用线性效用函数定义的贝叶斯遗憾指标，只能衡量帕累托前沿与其凸包交集上的性能。可以通过使用帕累托相容效用函数 $u_{\boldsymbol{\lambda}}$ 来解决这一问题，但一种常见的替代方法是衡量参考点与估计的帕累托集的映像之间所围成的超体积，$\boldsymbol{f}^*$（S_N）。超体积使用 BoTorch [7,22] 中的支配分区实现进行计算。对于参考点，我们取 NSGA-II 生成的帕累托前沿上 1000 个点在各维度上的最小值，并减去帕累托前沿各维度取值范围的 1%。

5.1 实验细节

对于两类测试问题中的每一类，我们运行贝叶斯优化 100 次，并报告贝叶斯遗憾的均值。每次重复实验都使用测试问题的一个不同且独立采样的实例，以及从扰乱 Sobol' 序列中生成的六个不同初始点样本 [34]。初始点均在两个目标函数上进行评估。

我们将算法与超体积知识梯度（HVKG） [12] 以及针对解耦目标的联合熵搜索下界近似（JES-LB）的一种修改版本进行比较 [36]。对于解耦目标问题，HVKG 是当前的最先进方法。尽管 JES-LB 的作者没有考虑解耦情形，但该情形历来已在基于熵的采集函数中得到考虑 [19,20,35] ，而 JES-LB 是其中的最先进方法。

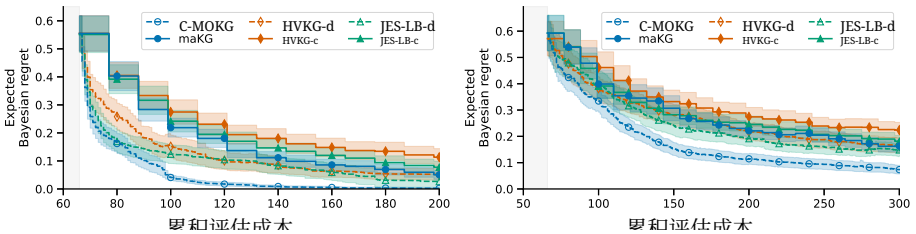
我们还与一种使用多属性知识梯度（maKG） [4]的基准算法进行比较。该算法始终评估两个目标函数，但除此之外与我们的算法完全相同。对于我们的采集函数和 maKG 基准算法，我们都使用均匀 11×11 网格进行离散化 $\mathcal{X}_{\text{disc}}$。式（13）中的对标量化取期望通过准蒙特卡洛估计计算，使用一个维度为 $D-1$、包含 $Q = 16$ 个点的扰乱 Sobol' 序列，并在每次贝叶斯优化迭代时随机重新生成。

对于代理模型，我们使用与测试问题相同的 Matérn-$5/2$ 核函数，不过其超参数是通过最大后验估计，根据观测数据拟合得到的。对于第一类测试问题，我们在长度尺度上设置了具有引导作用的先验，以提示模型第一目标具有较短的长度尺度。如前所述，当

成本较低的目标函数更难学习时，该算法的效果最佳。如果工程师知道某个目标函数的长度尺度较短或观测噪声较大，就可以通过相关超参数的先验分布将这些信息纳入模型。所用先验分布的完整细节见附录 B。所有参与比较的采集函数均使用相同的超先验。

5.2 结果

图 1 展示了在两类测试问题中，实验重复 100 次期间贝叶斯遗憾均值的变化。阴影区域表示跨测试问题族期望值的 95% 边际置信区间，按均值的两个标准误计算。在两种情况下，解耦算法都优于耦合算法，因为它们可以跳过对较慢且更易学习的目标函数进行采样，从而节省时间。C-MOKG 优于 HVKG 和 JES-LB，因为它是针对具有线性效用函数的决策者设计的。然而，第二类问题中使用的噪声水平较高，因此这两种算法的收敛速度都比第一类问题慢得多。

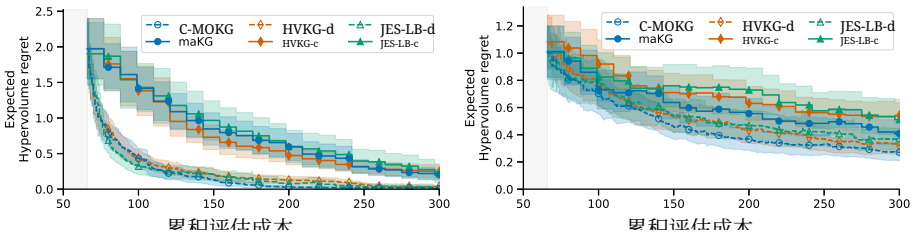


（a）不同的长度尺度　　（b）不同的观测噪声

图 1：比较 C-MOKG/maKG（蓝色圆圈）、HVKG（红色菱形）和 JES-LB（绿色三角形）的期望贝叶斯遗憾变化。对于 HVKG 和 JES-LB，使用后缀“-c”或“-d”分别区分耦合算法和解耦算法。耦合版本以实线表示，解耦版本以虚线表示。阴影区域表示跨测试问题族期望值的 95% 置信区间（均值上下两个标准误）。

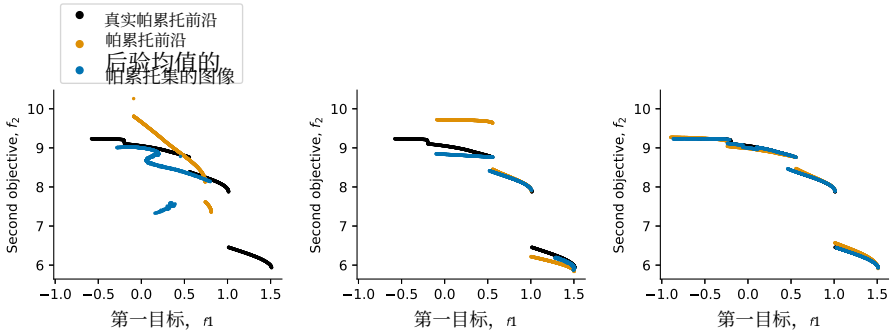
本研究全程使用线性效用函数。不过，如图2所示，从超体积遗憾度量来看，C-MOKG依然表现出色，这一点值得关注，因为这些测试问题的帕累托前沿通常并非凸的。众所周知，对于非凸帕累托前沿，线性标量化并不满足帕累托一致性。然而，利用 C-MOKG 收集的数据可以为帕累托前沿的所有部分提供信息，而不仅仅是与凸包相交的部分；我们认为，这正是该算法表现出色的原因。

图3c展示了一个 C-MOKG 算法收敛至帕累托前沿的示例。



（a）不同的长度尺度　　（b）不同的观测噪声

图 2：使用超体积遗憾得到的图 1 中实验的结果。



(a) 仅初始样本　　(b) 30 个贝叶斯优化样本　　(c) 70 个贝叶斯优化样本

图 3：随着使用 C-MOKG 收集样本，高斯过程代理模型的收敛情况。

5.3 与随机标量化的比较

我们在全文中都指出，当目标函数分别评估时，对标量化取期望尤其有益。本节将 C-MOKG 对标量化取期望的版本（如 [4]）与采用随机标量化的版本（如 ParEGO [21]）进行比较。为生成随机标量化权重，我们在每次迭代时取扰乱 Sobol' 序列中的下一个元素，从而确保权重在单纯形上分布均匀。

图 4 的结果显示，在解耦情况下，使用对标量化取期望（蓝色圆圈）与使用随机标量化（橙色三角形）之间存在显著差异

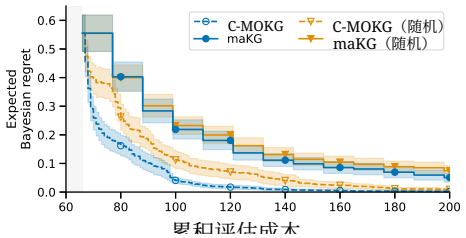


图 4：对于第一类测试问题，使用随机标量化（橙色三角形）和对标量化取期望（蓝色圆形）时的预期贝叶斯遗憾比较。解耦的 C-MOKG（虚线）显著受益于对标量化取期望。

（虚线）。而在目标函数一同评估的耦合情况下（实线），差异则没有那么大。原因可能是，某些随机标量化会过度偏向学习速度慢、容易学习的目标函数。因此，算法浪费时间获取样本，以了解帕累托前沿上已经十分熟悉的部分。相反，采用对标量化取期望的算法能够考察整个帕累托前沿上的潜在改进，并更频繁地对速度更快、难以学习的目标函数进行采样。

6 结论

本文提出了将 maKG 采集函数扩展到解耦目标情形的方法。我们为优化的收敛性提供了理论保证，并通过实验表明，该方法的表现与 HVKG 和 JES-LB 等其他最先进的采集函数一样出色。鉴于该方法依赖线性标量化，这一点令人惊讶；我们采用线性标量化是为了使数值近似在计算上可行，并简化收敛性证明。

未来，我们计划将该算法与采用其他效用函数（例如（增广）切比雪夫标量化）的版本进行比较，以量化采用帕累托相容效用函数所带来的收敛性改进。另一个值得研究的方向是将这项工作扩展到必须学习目标成本的情形，并将我们的方法与那些处理输入空间中成本变化的方法相结合 [33,5,23]。

致谢。第一作者ľúšu英国工程与自然科学研究委员会资助，通过华威大学系统数学II博士培训中心（项目编号 EP/S022244/1）获得支持。第二作者获得了 FWO（比利时）资助（项目编号 1216021N）以及比利时佛兰德斯人工智能研究计划（VCCM核心实验室）的支持。

利益披露。 作者声明，他们不存在与本文内容相关的利益冲突。

参考文献

1. Allmendinger, R., Handl, J., Knowles, J.：多目标优化：当目标函数具有非均匀延迟时。《欧洲运筹学杂志》**243**，497–513（2015）。https://doi.org/10.1016/j.ejor.2014.09.033 2. Allmendinger, R., Knowles, J.：异质目标：最新进展与未来研究。arXiv arXiv:2103.15546（2021年3月）3. Álvarez, M.A., Rosasco, L., Lawrence, N.D.：向量值函数的核：综述。《机器学习基础与趋势》**4**(3)，195–266（2012）。https://doi.org/10.1561/2200000036 4. Astudillo, R., Frazier, P.I.：效用不确定性下的多属性贝叶斯优化。载于：NIPS 贝叶斯优化研讨会。美国加利福尼亚州长滩（2017年12月），https://bayesopt.github.io/papers/2017/41.pdf 5. Astudillo, R., Jiang, D., Balandat, M., Bakshy, E., Frazier, P.：评估成本未知情况下的多步预算贝叶斯优化。《神经信息处理系统进展》**34**，20197–20209（2021年）6. Azaïs, J.M., Wschebor, M.：随机过程与随机场的水平集和极值。约翰·威立父子公司（2009年）7. Balandat, M., Karrer, B., Jiang, D., Daulton, S., Letham, B., Wilson, A.G., Bakshy, E.：BoTorch：一个用于高效蒙特卡洛贝叶斯优化的框架。载于：Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., 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），
h
t
t
p
s://
p
r
ocee
dings.neurips.cc/paper/2020/file/f5b1b89d98b7286673128a5fb112cb9a-Paper.pdf 8. Bect, J., Bachoc, F., Ginsbourger, D.：基于高斯过程的序贯实验设计的上鞅方法。《伯努利》**25**（2019年11月）。https://doi.org/10.3150/18-BEJ1074 9. Blank, J., Deb, K.：采用异质评估时间处理约束多目标优化问题：原理验证结果。《模因计算》**14**，135–150（2022年6月）。https://doi.org/10.1007/s12293-022-00362-z 10. Chugh, T., Allmendinger, R., Ojalehto, V., Miettinen, K.：针对具有非均匀延迟目标的代理辅助进化双目标优化。载于：《遗传与进化计算会议论文集》，第609–616页。GECCO '18，计算机协会，纽约，美国（2018年）。https://doi.org/10.1145/3205455.3205514 11. Couckuyt, I., Rojas Gonzalez, S., Branke, J.：贝叶斯优化：教程。载于：《遗传与进化计算会议附属会议论文集》，第843–863页（2022年）12. Daulton, S., Balandat, M., Bakshy, E.：超体积知识梯度：一种适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。《机器学习研究集刊》，第202卷，第7167–7204页。PMLR（2023年7月23–29日），https://proceedings.mlr.press/v202/daulton23a.html 13. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.：快速且精英化的多目标遗传算法：NSGA-II。《IEEE 进化计算汇刊》**6**（2002年），182–197。https://doi.org/10.1109/4235.996017 14. Frazier, P.I.：贝叶斯优化。载于：《当代问题优化与建模的最新进展》，第255–278页。INFORMS（2018年10月）。https://doi.org/10.1287/educ.2018.0188

14 J.M. Buckingham 等人。

15. Frazier, P.I., Powell, W.B., Dayanik, S.: 序贯信息收集的知识梯度策略。《SIAM 控制与优化期刊》**47**(5), 2410– 2439 070693424 16 2008 10 1137 ( )。 https://doi.org/ . / . Frazier, P.I., Powell, W.B., Dayanik, S.: 相关正态信念的知识梯度策略。《 I NFORMS 计算期刊》**21**, 599–613 (11 2009)。 https://doi.org/10 0314 17 1287 1080 . / ijoc. . . Garnett, R.: 贝叶斯优化。剑桥大学出版社 (2023), https: //bayesoptbook.co m/ 18. Hansen, M.P., Jaszkiewicz, A.: 非支配集近似解质量的评估。技术报告，数学建模研究所，丹麦技术大学 (1994) 19. Hernandez-Lobato, D., Hernandez-Lobato, J., Shah, A., Adams, R.: 面向多目标贝叶斯优化的预测熵搜索。载于：《第33届国际机器学习会议论文集》。《机器学习研究论文集》，第48卷，第1492–1501页。PMLR，纽约，纽约州，美国 (20–22 Jun 2016), https://proceedings.mlr.press/v48/hernandez-lobatoa16.html 20. Hernández-Lobato, J.M., Gelbart, M.A., Adams, R.P., Hoffman, M.W., Ghahramani, Z.: 使用基于信息的搜索进行约束贝叶斯优化的通用框架。《机器学习研究期刊》**17**(160), 1–53 (2016)，http://jmlr.org/papers/v17/15-616.html 21. Knowles, J.: ParEGO：一种用于昂贵的多目标优化问题、带有在线景观近似的混合算法。《IEEE 进化计算汇刊》**10**, 50–66 (2 2006)。 https://doi.org/10.1109/TEVC.2005. 851274 22. Lacour, R., Klamroth, K., Fonseca, C.M.: 用于计算超体积指标的盒分解算法。《计算机与运筹研究》**79**, 347 360 2017 10 1016 2016 06 021, – ( )。 https://doi.org/https://doi.org/ . /j.cor. . . https: // www.sciencedirect.com/science/article/pii/S0305054816301538 23. Lee, E.H., Eriksson, D., Perrone, V., Seeger, M.: 一种用于成本约束贝叶斯优化的非短视方法。载于：de Campos, C., Maathuis, M.H. (编)《第三十七届人工智能不确定性会议论文集》。《机器学习研究论文集》，第 161 卷，第 568–577 页。PMLR (27–30 Jul 2021) 24. Loka, N., Couckuyt, I., Garbuglia, F., Spina, D., Nieuwenhuyse, I.V., Dhaene, T.: 具有低成本与高成本目标函数的工程问题双目标贝叶斯优化。《计算机辅助工程》 (1 2022)。 https://doi.org/ 10 7 25 1007 00366 021 01573 . /s - - - . Mamun, M.M., Singh, H.K., Ray, T.: 一种针对具有可独立评估目标的计算昂贵型多目标优化问题的方法。《群体与进化计算》 **75**, 101146 (12 2022)。 https://doi. org/10 101146 26 1016 2022 . /j.swevo. . . Miettinen, K.: 非线性多目标优化，第12卷。施普林格科学与商业媒体 (1999) 27. Paria, B., Kandasamy, K., Póczos, B.: 使用随机标量化进行多目标贝叶斯优化的灵活框架。载于：Adams, R.P., Gogate, V. (编)《第35届人工智能不确定性会议论文集》。《机器学习研究论文集》，第115卷，第766–776页。PMLR (22–25 Jul 2020)，https://proceedings.mlr.press/v115/paria20a.html 28. Pearce, M., Branke, J.: 具有输入不确定性的贝叶斯仿真优化。载于：2017年冬季仿真会议 (WSC)。第2268–2278页。IEEE (12 2017)。 https://doi.org/10.1109/WSC.2017.8247958

29. Pearce, M., Branke, J.: 具有相关性的连续多任务贝叶斯优化。《欧洲运筹学杂志》**270**, 1074–1085 (11 2018)。 https://doi.org/10 017 30 1016 2018 03 . /j.ejor. . . . Pisier, G.: 巴拿赫空间中的鞅。剑桥高等数学研究丛书，剑桥大学出版社 (2016)。 https://doi.org/10.1017/ CBO9781316480588 31. Rasmussen, C.E., Williams, C.K.I.: 机器学习中的高斯过程（自适应计算与机器学习）。麻省理工学院出版社 (2006)。 https: //doi.org/10 1162254 32 5555. / . Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., de Freitas, N.: 将人类排除在循环之外：贝叶斯优化综述。《IEEE论文集》**104**2494218 33 1 148 175 2016 10 1109 2015 ( ), – ( )。 https://doi.org/ . /JPROC. . . Snoek, J., Larochelle, H., Adams, R.P.: 机器学习算法的实用贝叶斯优化。载于：Pereira, F., Burges, C., Bottou, L., Weinberger, K. (编)《神经信息处理系统进展》。第25卷。Curran Associates, Inc. (2012) 34. Sobol', I.M.: 立方体中的点分布与积分的近似求值。《计算数学与数学物理杂志》**7**(4), 784– 802 (1967) 35. Suzuki, S., Takeno, S., Tamura, T., Shitara, K., Karasuyama, M.: 使用帕累托前沿熵的多目标贝叶斯优化。载于：III, H.D., Singh, A. (编)《第37届国际机器学习会议论文集》。《机器学习研究论文集》，第119卷，第9279–9288页。PMLR (13–18 Jul 2020), https://proceedings.mlr.press/v119/suzuki20a.html 36. Tu, B., Gandy, A., Kantas, N., Shafei, B.: 面向多目标贝叶斯优化的联合熵搜索。载于：Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (编)《神经信息处理系统进展》。第35卷，第9922–9938页。Curran Asso- ciates, Inc. (2022), https://proceedings.neurips.cc/paper_files/paper/2022/file/4086fe59dc3584708468fba0e459f6a7-Paper-Conference.pdf 37. Tu, B., Kantas, N., Lee, R.M., Shafei, B.: 基于R2效用的多目标优化。arXiv (2024 11774 38 10 48550 2305)。 https://doi.org/ . /arXiv. . . Van den Braembussche, R.A.: 先进涡轮机械设计的数值优化。载于：Thévenin, D., Janiga, G. (编)《优化与计算流体力学》，第6章，第147–189页。施普林格 (2008)。 https://doi.org/10.1007/ 978-3-540-72153-6 39. Wang, X., Jin, Y., Schmitt, S., Olhofer, M.: 通过具有多个异质目标的贝叶斯进化优化缓解搜索偏差。arXiv arxiv:2208 12217 40 12217 2022 10 48550 2208 . ( )。 https://doi.org/ . /ARXIV. . . Wu, J., Frazier, P.: 用于批量贝叶斯优化的并行知识梯度法。载于：Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., Garnett, R. (编)《神经信息处理系统进展》。第29卷。Curran Associates, Inc. (2016), https://proceedings.neurips.cc/paper/2016/file/18d10dc6e666eab6de9215ae5b3d54df-Paper.pdf

A 理论结果

在本附录中，我们证明与成本加权多目标知识梯度（C-MOKG）相关的理论结果。附录最后证明，对于任意标量化权重 $\boldsymbol{\lambda}$，最优值估计量渐近一致。证明的结构大体沿袭 Bect等人 [8], 的方法，但没有使用那么复杂的工具。遗憾的是，我们无法直接应用 [8] 中的结果，因为这些结果并未涵盖多目标高斯过程模型或随机标量化的情形。

由于条件期望仅在零测事件意义下定义，因此本节的结果均注明为“几乎必然”（a.s.）。读者应将这些结果证明中的大多数等式和不等式理解为几乎必然成立。

A.1 统计模型

为方便读者，我们回顾所用的统计模型。设 $\boldsymbol{f} \sim \mathcal{GP}(\boldsymbol{\mu}, \mathbf{K})$ 是一个指标集为紧集 $\mathcal{X} \subset \mathbb{R}^D$ 的多输出高斯过程，定义在概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上。这里 $\boldsymbol{\mu}: \mathcal{X} \to \mathbb{R}^M$ 是均值函数，$\mathbf{K}: \mathcal{X} \times \mathcal{X} \to \mathbb{R}^{M \times M}$ 是协方差函数。我们假设 $\boldsymbol{\mu}$ 和 $\mathbf{K}$ 连续，并且进一步采用具有连续样本路径的 $\boldsymbol{f}$ 版本。重要的是，我们不区分高斯过程 $\boldsymbol{f}$ 与“真实”函数 $\boldsymbol{f}^*$。

假设可以根据以下模型观测 $\boldsymbol{f}$

$$y_m = f_m(\boldsymbol{x}) + \varepsilon_m \tag{18}$$

其中，$m \in [M] = \{1, \dots, M\}$表示正在评估的 $\boldsymbol{f}$分量的索引，$\boldsymbol{x} \in \mathcal{X}$是输入位置，而 $\varepsilon_m \sim \mathcal{N}(0, \sigma_m^2)$ 表示观测噪声。令 $\boldsymbol{c} \in (0, \infty)^M$ 为与每个目标函数相关联的严格正评估成本向量。我们考虑按顺序选取（依赖于数据）的设计点和索引 $\boldsymbol{x}_1, m_1, \boldsymbol{x}_2, m_2, \dots$，并用 $y_{1,m_1}, y_{2,m_2}, \dots$ 表示相应的观测。我们用 ε_{1,m_1}, $\varepsilon_{2,m_2}, \dots$ 表示添加到每次观测中的独立噪声项。

令 $\Lambda = \{\boldsymbol{\lambda} \in [0, \infty)^M : \sum_{j=1}^M \lambda_j = 1\}$ 表示 $\mathbb{R}^M$ 中的标准单纯形。我们将分别证明对标量化取期望和随机标量化两种情形下的结果。在这两种情形中，我们都假设概率密度$p(\boldsymbol{\lambda})$ 在 Λ 上处处严格为正。对于使用随机标量化的算法，我们将标量化权重序列记为 $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots$。对于每个 $n \in \mathbb{N}_0$，记 $\mathcal{F}_n = \sigma(\{\boldsymbol{x}_j, m_j, y_{j,m_j}, \varepsilon_{j,m_j}\}_{j=1}^n \cup \{\boldsymbol{\lambda}_j\}_{j=1}^\infty)$ 为由截至时间 n 可获得的所有信息以及完整的标量化权重序列生成的σ-代数。当然，虽然这样做无妨，但只有在随机标量化的情形下才需要纳入 $\boldsymbol{\lambda}_j$。与正文中的记号略有不同，我们将 $\mathbb{E}_n[\cdot]$ 记为在给定 $\mathcal{F}_n$ 条件下的期望。为了妥善处理这两种算法，当存在随机标量化权重向量 $\boldsymbol{\lambda}$ 时，我们将其视为条件变量的一部分，

$$\mathbb{E}_n[\cdot] = \mathbb{E}[\cdot \mid \sigma(\mathcal{F}_n, \boldsymbol{\lambda})].$$

那么，为了表示仅对 $\boldsymbol{\lambda}$ 取期望，我们使用 $\mathbb{E}_{\boldsymbol{\lambda}\sim p(\boldsymbol{\lambda})}[\cdot]$。

回顾正文中的两个 C-MOKG 采集函数。虽然现在 $\mathbb{E}_n$ 的定义包含了完整的随机标量化序列，而不只是截至时间 n 的标量化，但由于因果结构，公式并未因此改变，

$$\alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda}, \boldsymbol{c}) = \frac{1}{c_m}\left(\mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}} \mu^s_{n+}(\boldsymbol{x}'; \boldsymbol{x}, m, \boldsymbol{\lambda})\right] - \max_{\boldsymbol{x}'\in\mathcal{X}} \mu^s_n(\boldsymbol{x}'; \boldsymbol{\lambda})\right), \tag{19a}$$

$$\overline{\alpha}^n_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{c}) = \mathbb{E}_{\boldsymbol{\lambda}\sim p(\boldsymbol{\lambda})}\left[\alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda}, \boldsymbol{c})\right], \tag{19b}$$

其中 $\mu^s_n(\boldsymbol{x}';\ \boldsymbol{\lambda}) = \mathbb{E}_n[\boldsymbol{\lambda}\cdot \boldsymbol{f}(\boldsymbol{x})]$, $\mu^s_{n+}(\boldsymbol{x}';\ \boldsymbol{x},\ m,\ \boldsymbol{\lambda}) = \mathbb{E}_n[\boldsymbol{\lambda}\cdot \boldsymbol{f}(\boldsymbol{x}')\mid y^{\boldsymbol{x}}_m]$ 以及 $y^{\boldsymbol{x}}_m = f_m(\boldsymbol{x}) + \varepsilon_m$ 是假设的下一次观测。请注意，本附录中使用上标 s，以区别标量化高斯过程的后验均值与未标量化高斯过程的后验均值；后者将在命题 12 之前引入。

非正式地说，我们假设设计点 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，.....以及目标索引 m_1，m_2，..... 在每一步都被选为相关采集函数的极大化点。然而，知识梯度并非处处连续，因此它是否能取到最大值并不明显。此外，在实践中，我们永远无法完美地最大化采集函数。基于这些原因，我们假设设计点仅被选为采集函数的近似极大化点。也就是说，我们假设存在一个由小的非负实数组成的序列 $\eta = (\eta_n)_{n=0}^{\infty}$，满足 $\eta_n \to 0$（当 $n\to\infty$ 时），并且以下条件之一成立

$$\begin{aligned}\forall n\in\mathbb{N}_0 \quad & \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}_{n+1}, m_{n+1};\ \boldsymbol{\lambda}_{n+1}, \boldsymbol{c}) \\ & > \sup_{\substack{\boldsymbol{x}\in\mathcal{X}\\ m\in[M]}} \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\ \boldsymbol{\lambda}_{n+1}, \boldsymbol{c}) - \eta_n,\end{aligned} \tag{20a}$$

$$\begin{aligned}\forall n\in\mathbb{N}_0 \quad & \overline{\alpha}^n_{\text{C-MOKG}}(\boldsymbol{x}_{n+1}, m_{n+1}; \boldsymbol{c}) \\ & > \sup_{\substack{\boldsymbol{x}\in\mathcal{X}\\ m\in[M]}} \overline{\alpha}^n_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{c}) - \eta_n,\end{aligned} \tag{20b}$$

具体取决于我们使用随机标量化，还是对标量化权重取期望。这被称为 Bect等人[8][4]提出的 η-*quasi-SUR* 序贯设计。

A.2 $\alpha^n_{\text{C-MOKG}}$ 趋于零的收敛

定义 3. 类比于 *[8],*，我们为第 $n\in\mathbb{N}_0$ 步各采集函数所关联的残余不确定性 引入一些记号，

$$\forall \boldsymbol{\lambda}\in\Lambda \quad H_n(\boldsymbol{\lambda}) = \mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}} \boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}')\right] - \max_{\boldsymbol{x}'\in\mathcal{X}} \mathbb{E}_n[\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}')], \tag{21a}$$

$$\overline{H}_n = \mathbb{E}_{\boldsymbol{\lambda}\sim p(\boldsymbol{\lambda})}[H_n(\boldsymbol{\lambda})] \tag{21b}$$

[4] 请注意，在 [8], 中，字符 ε 用于代替 η

引理 4. 定义 *3* 中的残余不确定性定义明确。也就是说，

$$\forall n \in \mathbb{N}_0 \, \forall \boldsymbol{\lambda} \in \Lambda \quad H_n(\boldsymbol{\lambda}) < \infty \quad a.s. \qquad and \qquad \forall n \in \mathbb{N}_0 \quad \overline{H}_n < \infty \quad a.s..$$

证明。 我们先考虑第一种形式。令 $\boldsymbol{\lambda} \in \Lambda$。由于 $\boldsymbol{\lambda}$的所有分量都介于 0 和 1 之间，因此对于所有 $\boldsymbol{x}' \in \mathcal{X}$，有 $|\boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}')| \leq \|\boldsymbol{f}(\boldsymbol{x}')\|_1$。因此，

$$\begin{aligned} H_n(\boldsymbol{\lambda}) &\leq \mathbb{E}_n \left[\max_{\boldsymbol{x}' \in \mathcal{X}} |\boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}')| \right] + \max_{\boldsymbol{x}' \in \mathcal{X}} \mathbb{E}_n |\boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}')| \\ &\leq 2\mathbb{E}_n \left[\max_{\boldsymbol{x}' \in \mathcal{X}} |\boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}')| \right] \\ &\leq 2\mathbb{E}_n \left[\max_{\boldsymbol{x}' \in \mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_1 \right] \\ &\leq 2 \sum_{m=1}^{M} \mathbb{E}_n \left[\max_{\boldsymbol{x}' \in \mathcal{X}} |f_m(\boldsymbol{x}')| \right] < \infty. \end{aligned}$$

这里的最终不等式成立，是因为每个 f_m 都有连续样本路径，而$\mathcal{X}$是紧的。例如，使用 [6] 中的定理 2.9，并注意到对于任意 m和任意 $\boldsymbol{x} \in \mathcal{X}$，都有 $\mathbb{E}_n[\max_{\boldsymbol{x}' \in \mathcal{X}} |f_m(\boldsymbol{x}')|] \leq 2\mathbb{E}_n[\max_{\boldsymbol{x}' \in \mathcal{X}} f_m(\boldsymbol{x}')] + \mathbb{E}_n|f_m(\boldsymbol{x})|$。

为了对 H_n证明同样的结论，我们只需对 $\boldsymbol{\lambda} \sim p(\Lambda)$取期望。也就是说，

$$\overline{H}_n = \mathbb{E}_{\boldsymbol{\lambda} \sim p(\boldsymbol{\lambda})}[H_n(\boldsymbol{\lambda})] \leq \sum_{m=1}^{M} \mathbb{E}_n \left[\sup_{\boldsymbol{x} \in \mathcal{X}} |f_m(\boldsymbol{x})| \right] < \infty.$$

□

注 *5*。 考虑随机标量化的情形。由于在以 $\mathcal{F}_n$为条件化后， $\boldsymbol{x}_{n+1}$ 和 m_{n+1} 是确定性的，因此我们可以将它们直接代入 $\alpha^n_{\text{C-MOKG}}(\boldsymbol{x}_{n+1}, m_{n+1}; \boldsymbol{\lambda}_{n+1}, \boldsymbol{c})$中的外层期望。此外，对于所有 $\boldsymbol{x}' \in \mathcal{X}$，$\mu^s_{n+}(\boldsymbol{x}'; \boldsymbol{x}_{n+1}, m_{n+1}, \boldsymbol{\lambda}_{n+1}) = \mu^s_{n+1}(\boldsymbol{x}'; \boldsymbol{\lambda}_{n+1})$，并且我们有

$$\begin{aligned} &\alpha^n_{\text{C-MOKG}}(\boldsymbol{x}_{n+1}, m_{n+1}; \boldsymbol{\lambda}_{n+1}, \boldsymbol{c}) \\ &\quad = \frac{1}{c_{m_{n+1}}} \left(\mathbb{E}_n \left[\max_{\boldsymbol{x}' \in \mathcal{X}} \mu^s_{n+1}(\boldsymbol{x}'; \boldsymbol{\lambda}_{n+1}) \right] - \max_{\boldsymbol{x}' \in \mathcal{X}} \mu^s_n(\boldsymbol{x}'; \boldsymbol{\lambda}_{n+1}) \right) \\ &\quad = \frac{1}{c_{m_{n+1}}} \left(\mathbb{E}_n \left[\max_{\boldsymbol{x}' \in \mathcal{X}} \mathbb{E}_{n+1} \left[\boldsymbol{\lambda}_{n+1} \cdot \boldsymbol{f}(\boldsymbol{x}') \right] \right] - \max_{\boldsymbol{x}' \in \mathcal{X}} \mathbb{E}_n [\boldsymbol{\lambda}_{n+1} \cdot \boldsymbol{f}(\boldsymbol{x}')] \right) \\ &\quad = \frac{1}{c_{m_{n+1}}} \Big(H_n(\boldsymbol{\lambda}_{n+1}) - \mathbb{E}_n[H_{n+1}(\boldsymbol{\lambda}_{n+1})] \Big) \quad \text{a.s.} \end{aligned} \tag{22}$$

对于平均标量化的情况，我们也得到了类似的结果：

$$\begin{aligned} \overline{\alpha}^n_{\text{C-MOKG}}(\boldsymbol{x}_{n+1}, m_{n+1}; \boldsymbol{c}) &= \mathbb{E}_{\boldsymbol{\lambda} \sim p(\boldsymbol{\lambda})} \left[\alpha^n_{\text{C-MOKG}}(\boldsymbol{x}_{n+1}, m_{n+1}, \boldsymbol{\lambda}; \boldsymbol{c}) \right] \\ &= \frac{1}{c_{m_{n+1}}} \Big(\overline{H}_n - \mathbb{E}_n \big[\overline{H}_{n+1} \big] \Big) \quad \text{a.s.} \end{aligned} \tag{23}$$

[5] 形式上，我们称 $\boldsymbol{x}_{n+1}$ 和 m_{n+1} 是 $\mathcal{F}_n$-可测的 .

我们先介绍三个密切相关的引理。

引理 1（由正文重述）。*C-MOKG* 的两种形式均为非负。也就是说，对于所有 $n \in \mathbb{N}_0$、 $\boldsymbol{x} \in \mathcal{X}$、 $m \in \{1, \ldots, M\}$ 以及所有 $\boldsymbol{\lambda} \in \Lambda$,

$$\alpha^n_{C\text{-}MOKG}(\boldsymbol{x}, m; \boldsymbol{\lambda}, \boldsymbol{c}) \geq 0 \quad a.s. \qquad and \qquad \overline{\alpha}^n_{C\text{-}MOKG}(\boldsymbol{x}, m; \boldsymbol{c}) \geq 0 \quad a.s..$$

证明。我们首先证明 $\alpha^n_{\text{C-MOKG}}(\,\cdot\,,\,\cdot\,;\ \boldsymbol{\lambda},\ \boldsymbol{c})$ 的结论，其中标量化权重 $\boldsymbol{\lambda}$ 作为参数。令 $n \in \mathbb{N}_0$、 $\boldsymbol{x} \in \mathcal{X}$、 $m \in \{1, \ldots, M\}$，并令$\boldsymbol{\lambda} \in \Lambda$ 为随机变量。那么，对于任意 $\boldsymbol{x}'' \in \mathcal{X}$,

$$\max_{\boldsymbol{x}' \in \mathcal{X}} \mu^s_{n+}(\boldsymbol{x}'; \boldsymbol{x}, m, \boldsymbol{\lambda}) \geq \mu^s_{n+}(\boldsymbol{x}''; \boldsymbol{x}, m, \boldsymbol{\lambda}).$$

利用 $\mathbb{E}_n$ ，在给定 $\mathcal{F}_n$ 和 $\boldsymbol{\lambda}$ 的条件下取期望，得到

$$\mathbb{E}_n\left[\max_{\boldsymbol{x}' \in \mathcal{X}} \mu^s_{n+}(\boldsymbol{x}'; \boldsymbol{x}, m, \boldsymbol{\lambda})\right] \geq \mathbb{E}_n\big[\mu^s_{n+}(\boldsymbol{x}''; \boldsymbol{x}, m, \boldsymbol{\lambda})\big] = \mu^s_n(\boldsymbol{x}''; \boldsymbol{\lambda}).$$

最后，这对所有 $\boldsymbol{x}'' \in \mathcal{X}$ 都成立，因此对最大值也成立

$$\begin{aligned} & \mathbb{E}_n\left[\max_{\boldsymbol{x}' \in \mathcal{X}} \mu^s_{n+}(\boldsymbol{x}'; \boldsymbol{x}, m, \boldsymbol{\lambda})\right] \geq \max_{\boldsymbol{x}' \in \mathcal{X}} \mu^s_n(\boldsymbol{x}'; \boldsymbol{\lambda}) \\ \Rightarrow \quad & \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda}, \boldsymbol{c}) \\ & \qquad = \frac{1}{c_m}\left(\mathbb{E}_n\left[\max_{\boldsymbol{x}' \in \mathcal{X}} \mu^s_{n+}(\boldsymbol{x}'; \boldsymbol{x}, m, \boldsymbol{\lambda})\right] - \max_{\boldsymbol{x}' \in \mathcal{X}} \mu^s_n(\boldsymbol{x}'; \boldsymbol{\lambda})\right) \geq 0. \end{aligned}$$

为了证明 $\alpha_{\text{C-MOKG}}$ 的结论，我们只需取期望以消去 $\boldsymbol{\lambda}$,

$$\overline{\alpha}^n_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{c}) = \mathbb{E}_{\boldsymbol{\lambda} \sim p(\boldsymbol{\lambda})}[\alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda}, \boldsymbol{c})] \geq 0.$$

□

引理 6。 在这两种情况下，残余不确定性至少为按目标成本缩放后的 *C-MOKG*。也就是说，对于所有 $n \in \mathbb{N}_0$、 $\boldsymbol{x} \in \mathcal{X}$、 $m \in \{1, \ldots, M\}$ 以及所有 $\boldsymbol{\lambda} \in \Lambda$,

$$c_m \alpha^n_{C\text{-}MOKG}(\boldsymbol{x}, m; \boldsymbol{\lambda}, \boldsymbol{c}) \leq H_n(\boldsymbol{\lambda}) \qquad and \qquad c_m \overline{\alpha}^n_{C\text{-}MOKG}(\boldsymbol{x}, m; \boldsymbol{c}) \leq \overline{H}_n.$$

证明。 令 $\boldsymbol{\lambda} \in \Lambda$。通过与证明引理1中 C-MOKG 非负时非常相似的论证,

$$\begin{aligned} & \forall \boldsymbol{x}'' \in \mathcal{X} && \max_{\boldsymbol{x}' \in \mathcal{X}} \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}') \geq \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}'') \\ \Rightarrow \quad & \forall n \in \mathbb{N}_0\ \forall \boldsymbol{x}'' \in \mathcal{X} && \mathbb{E}_n\left[\max_{\boldsymbol{x}' \in \mathcal{X}} \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}') \,\middle|\, y^{\boldsymbol{x}}_m\right] \geq \mathbb{E}_n\left[\boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}'') \,|\, y^{\boldsymbol{x}}_m\right] \\ & && \qquad = \mu^s_{n+}(\boldsymbol{x}''; \boldsymbol{x}, m, \boldsymbol{\lambda}) \\ \Rightarrow \quad & \forall n \in \mathbb{N}_0 && \mathbb{E}_n\left[\max_{\boldsymbol{x}' \in \mathcal{X}} \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}') \,\middle|\, y^{\boldsymbol{x}}_m\right] \geq \max_{\boldsymbol{x}' \in \mathcal{X}} \mu^s_{n+}(\boldsymbol{x}'; \boldsymbol{x}, m, \boldsymbol{\lambda}) \\ \Rightarrow \quad & \forall n \in \mathbb{N}_0 && \mathbb{E}_n\left[\max_{\boldsymbol{x}' \in \mathcal{X}} \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}')\right] \geq \mathbb{E}_n\left[\max_{\boldsymbol{x}' \in \mathcal{X}} \mu^s_{n+}(\boldsymbol{x}'; \boldsymbol{x}, m, \boldsymbol{\lambda})\right] \\ \Rightarrow \quad & \forall n \in \mathbb{N}_0 && H_n(\boldsymbol{\lambda}) \geq c_m \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{\lambda}, \boldsymbol{c}). \end{aligned}$$

为了证明 $H_n \geq c_m\overline{\alpha}^n_{\text{C-MOKG}}(\boldsymbol{x},\ m;\boldsymbol{c})$，我们只需对 $\boldsymbol{\lambda} \sim p(\Lambda)$ 取期望。□

引理 7. 对于所有 $\boldsymbol{\lambda} \in \Lambda$，序列 $(H_n(\boldsymbol{\lambda}))^{n\in\mathbb{N}_0}$ 是关于滤过 $(\mathcal{F}_n)_{n=0}^{\infty}$的非负上鞅。同样，$(H_n)^{n\in\mathbb{N}_0}$ 也是关于同一滤过的非负上鞅。也就是说，

$$\begin{aligned}\forall\boldsymbol{\lambda}\in\Lambda\,\forall n\in\mathbb{N}_0 \quad H_n(\boldsymbol{\lambda})\geq 0\ a.s. \qquad and \qquad H_n(\boldsymbol{\lambda})\geq\mathbb{E}_n[H_{n+1}(\boldsymbol{\lambda})]\ a.s.\\ \forall n\in\mathbb{N}_0 \qquad \overline{H}_n\geq 0\ a.s. \qquad and \qquad \overline{H}_n\geq\mathbb{E}_n[\overline{H}_{n+1}]\ a.s.\end{aligned}$$

证明。设 $\boldsymbol{\lambda}\in\Lambda$。根据引理 1 和 6，对于所有 $\boldsymbol{x}\in\mathcal{X}$和所有 $m\in\{1,\ldots,M\}$，$H_n(\boldsymbol{\lambda}) \geq c_m\alpha^n_{\text{C-MOKG}}(\boldsymbol{x},\ m;\ \boldsymbol{\lambda},\ \boldsymbol{c}) \geq 0$以及 $H_n \geq c_m\alpha^n_{\text{C-MOKG}}(\boldsymbol{x},\ m;\ \boldsymbol{c}) \geq 0$几乎必然成立。此外，利用与证明引理 1 和 6 时类似的论证，对于所有 $n\in\mathbb{N}_0$,

$$\begin{aligned}&\forall\boldsymbol{x}''\in\mathcal{X} && \max_{\boldsymbol{x}'}\mathbb{E}_{n+1}[\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}')]\geq\mathbb{E}_{n+1}[\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}'')]\\ \Rightarrow\quad &\forall\boldsymbol{x}''\in\mathcal{X} && \mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}}\mathbb{E}_{n+1}[\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}')]\right]\geq\mathbb{E}_n[\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}'')]\\ \Rightarrow\quad & && \mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}}\mathbb{E}_{n+1}[\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}')]\right]\geq\max_{\boldsymbol{x}'\in\mathcal{X}}\mathbb{E}_n[\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}')]\\ \Rightarrow\quad & && \mathbb{E}_n[H_{n+1}(\boldsymbol{\lambda})]\leq H_n(\boldsymbol{\lambda}).\end{aligned}$$

对 $\boldsymbol{\lambda}\sim p(\boldsymbol{\lambda})$ 取期望即可得到 H_n 的结果。 □

接下来，我们证明 C-MOKG 和残余不确定性具有利普希茨性质。

引理 8. *C-MOKG* 和残余不确定性在 $\boldsymbol{\lambda}\in\Lambda$ 中具有以下利普希茨型性质，且这些性质几乎必然成立。设 $\boldsymbol{\lambda}$、$\boldsymbol{\lambda}'\in\Lambda$ 和 $n\in\mathbb{N}_0$。则

1. 对所有 $\boldsymbol{x}\in\mathcal{X}$ 和 $m\in\{1,\ldots,M\}$,

$$\begin{aligned}&\left|\alpha^n_{C\text{-}MOKG}(\boldsymbol{x},m;\boldsymbol{\lambda},\boldsymbol{c})-\alpha^n_{C\text{-}MOKG}(\boldsymbol{x},m;\boldsymbol{\lambda}',\boldsymbol{c})\right|\\ &\qquad\leq\frac{2}{c_m}\|\boldsymbol{\lambda}-\boldsymbol{\lambda}'\|_2\,\mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}}\|\boldsymbol{f}(\boldsymbol{x}')\|_2\right];\end{aligned}$$

2.

$$|H_n(\boldsymbol{\lambda})-H_n(\boldsymbol{\lambda}')|\leq 2\|\boldsymbol{\lambda}-\boldsymbol{\lambda}'\|_2\,\mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}}\|\boldsymbol{f}(\boldsymbol{x}')\|_2\right].$$

证明。我们先从 C-MOKG 开始。令 $\boldsymbol{\lambda},\ \boldsymbol{\lambda}'\in\Lambda$。那么对于所有 $\boldsymbol{x},\ \boldsymbol{x}'\in\mathcal{X}$、$m\in\{1,\ldots,M\}$ 和 $n\in\mathbb{N}_0$,

$$\begin{aligned}\mu^s_{n+}(\boldsymbol{x}';\boldsymbol{x},m,\boldsymbol{\lambda})-\mu^s_{n+}(\boldsymbol{x}';\boldsymbol{x},m,\boldsymbol{\lambda}')&=\mathbb{E}_n\left[(\boldsymbol{\lambda}-\boldsymbol{\lambda}')\cdot\boldsymbol{f}(\boldsymbol{x}')\,|\,y^{\boldsymbol{x}}_m\right]\\ &\leq\mathbb{E}_n\left[\|\boldsymbol{\lambda}-\boldsymbol{\lambda}'\|_2\,\|\boldsymbol{f}(\boldsymbol{x}')\|_2\,\middle|\,y^{\boldsymbol{x}}_m\right]\\ &\leq\|\boldsymbol{\lambda}-\boldsymbol{\lambda}'\|_2\,\mathbb{E}_n\left[\max_{\boldsymbol{x}''\in\mathcal{X}}\|\boldsymbol{f}(\boldsymbol{x}'')\|_2\,\middle|\,y^{\boldsymbol{x}}_m\right]\end{aligned}$$

其中第二行由柯西-施瓦茨不等式得出。对于 $\mu_n^s(\boldsymbol{x}';\boldsymbol{\lambda})$，我们也可以类似地得到

$$\mu_n^s(\boldsymbol{x}';\boldsymbol{\lambda}') - \mu_n^s(\boldsymbol{x}';\boldsymbol{\lambda}) \leq \|\boldsymbol{\lambda}-\boldsymbol{\lambda}'\|_2 \, \mathbb{E}_n\left[\max_{\boldsymbol{x}''\in\mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}'')\|_2\right].$$

因此，

$$\begin{aligned}
c_m\Big(&\alpha^n_{\text{C-MOKG}}(\boldsymbol{x},m;\boldsymbol{\lambda},\boldsymbol{c}) - \alpha^n_{\text{C-MOKG}}(\boldsymbol{x},m;\boldsymbol{\lambda}',\boldsymbol{c})\Big) \\
&= \mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}} \mu^s_{n+}(\boldsymbol{x}';\boldsymbol{x},m,\boldsymbol{\lambda})\right] - \max_{\boldsymbol{x}'\in\mathcal{X}} \mu_n^s(\boldsymbol{x}';\boldsymbol{\lambda}) \\
&\quad - \mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}} \mu^s_{n+}(\boldsymbol{x}';\boldsymbol{x},m,\boldsymbol{\lambda}')\right] + \max_{\boldsymbol{x}'\in\mathcal{X}} \mu_n^s(\boldsymbol{x}';\boldsymbol{\lambda}') \\
&\leq \mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}} \mu^s_{n+}(\boldsymbol{x}';\boldsymbol{x},m,\boldsymbol{\lambda}) - \mu^s_{n+}(\boldsymbol{x}';\boldsymbol{x},m,\boldsymbol{\lambda}')\right] \\
&\quad + \max_{\boldsymbol{x}'\in\mathcal{X}}\Big(\mu_n^s(\boldsymbol{x}';\boldsymbol{\lambda}') - \mu_n^s(\boldsymbol{x}';\boldsymbol{\lambda})\Big) \\
&\leq \mathbb{E}_n\left[\|\boldsymbol{\lambda}-\boldsymbol{\lambda}'\|_2 \, \mathbb{E}_n\Big[\max_{\boldsymbol{x}'\in\mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2 \,\Big|\, y_m^{\boldsymbol{x}}\Big]\right] \\
&\quad + \|\boldsymbol{\lambda}-\boldsymbol{\lambda}'\|_2 \, \mathbb{E}_n\Big[\max_{\boldsymbol{x}'\in\mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2\Big] \\
&= 2\|\boldsymbol{\lambda}-\boldsymbol{\lambda}'\|_2 \, \mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2\right].
\end{aligned}$$

由于交换 $\boldsymbol{\lambda}$ 和 $\boldsymbol{\lambda}'$ 后结论仍然成立，因此我们已证明 $\alpha^n_{\text{C-MOKG}}$ 的不等式。

残余不确定性的证明类似。再次令 $\boldsymbol{\lambda}$，$\boldsymbol{\lambda}' \in \Lambda$。那么对于所有 $n \in \mathbb{N}_0$，

$$\begin{aligned}
H_n(\boldsymbol{\lambda}) - &H_n(\boldsymbol{\lambda}') \\
&= \mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}} \boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}')\right] - \max_{\boldsymbol{x}'\in\mathcal{X}} \mathbb{E}_n[\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}')] \\
&\quad - \mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}} \boldsymbol{\lambda}'\cdot\boldsymbol{f}(\boldsymbol{x}')\right] + \max_{\boldsymbol{x}'\in\mathcal{X}} \mathbb{E}_n[\boldsymbol{\lambda}'\cdot\boldsymbol{f}(\boldsymbol{x}')] \\
&\leq \mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}} (\boldsymbol{\lambda}-\boldsymbol{\lambda}')\cdot\boldsymbol{f}(\boldsymbol{x}')\right] + \max_{\boldsymbol{x}'\in\mathcal{X}} \mathbb{E}_n\left[(\boldsymbol{\lambda}'-\boldsymbol{\lambda})\cdot\boldsymbol{f}(\boldsymbol{x}')\right] \\
&\leq \mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}} \|\boldsymbol{\lambda}-\boldsymbol{\lambda}'\|_2 \, \|\boldsymbol{f}(\boldsymbol{x}')\|_2\right] + \max_{\boldsymbol{x}'\in\mathcal{X}} \mathbb{E}_n\left[\|\boldsymbol{\lambda}-\boldsymbol{\lambda}'\|_2 \, \|\boldsymbol{f}(\boldsymbol{x}')\|_2\right] \\
&\leq 2\|\boldsymbol{\lambda}-\boldsymbol{\lambda}'\|_2 \mathbb{E}_n\left[\max_{\boldsymbol{x}'\in\mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2\right].
\end{aligned}$$

同样，将 $\boldsymbol{\lambda}$ 和 $\boldsymbol{\lambda}'$ 互换后，该结论仍然成立，由此得出 H_n 的不等式。□

现在我们可以陈述并证明前两个主要结果。这两个结果表明，无论采用随机标量化还是对标量化取期望，$\alpha^n_{\text{C-MOKG}}(\boldsymbol{x},\ m;\boldsymbol{\lambda},\ \boldsymbol{c})$ 都会在某种意义下收敛到零。

定理 9. 假设我们使用 *C-MOKG* 选择样本，并采用依据分布 $p(\boldsymbol{\lambda})$独立选取的随机标量化权重 $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots$。那么，

$$\forall \boldsymbol{\lambda} \in \Lambda \quad \mathbb{P}\left(\liminf_{n\to\infty} \sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha^n_{C\text{-}MOKG}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}, \boldsymbol{c}) = 0\right) = 1.$$

注 *10.* 事实上，对于所有 $\boldsymbol{\lambda} \in \Lambda$，以下结论都成立:

$$\sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}, \boldsymbol{c}) \to 0$$

当 $n \to \infty$ 时，几乎必然如此。不过，这一结论的证明要等到推论 15 才给出。

证明。令 $\boldsymbol{\lambda} \in \Lambda$。单目标情形下的证明见 [8]，其依据是我们在优化的每一步都使用相同的采集函数。然而，对于采用随机标量化的多目标情形 $(\boldsymbol{\lambda}_n)_{n=1}^{\infty}$，实际上每一步使用的采集函数都不同。使证明得以继续的关键观察是：尽管我们不会无限多次使用完全相同的权重 $\boldsymbol{\lambda}$，但会无限多次使用任意接近的权重，而且这些权重足够相似。形式化地说，我们注意到存在一个（随机）子序列 $(\boldsymbol{\lambda}_{1+n_j})_{j=0}^{\infty}$，使得 $\boldsymbol{\lambda}_{1+n_j} \to \boldsymbol{\lambda}$ 当 $j \to \infty$ 时几乎必然成立。不失一般性地，为了稍微简化后续记号，我们假设 $n_0 = 0$。我们首先证明 $\sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha^{n_j}_{\text{C-MOKG}}(\boldsymbol{x}, m;\ \boldsymbol{\lambda}_{1+n_j}, \boldsymbol{c}) \to 0$ 几乎必然成立，然后利用该子序列的收敛性，断言将 $\boldsymbol{\lambda}_{1+n_j}$ 替换为 $\boldsymbol{\lambda}$ 时也有相同结论。

受 [8]，中符号表示的启发，对于每个 $j \in \mathbb{N}_0$，令

$$\begin{aligned}
\Delta^{(1)}_{j+1} &= H_{n_j}(\boldsymbol{\lambda}_{1+n_j}) - H_{n_{(j+1)}}(\boldsymbol{\lambda}_{1+n_{(j+1)}}), & \overline{\Delta}^{(1)}_{j+1} &= \mathbb{E}_{n_j}[\Delta^{(1)}_{j+1}],\\
\Delta^{(2)}_{j+1} &= H_{n_{(j+1)}}(\boldsymbol{\lambda}_{1+n_{(j+1)}}) - H_{n_{(j+1)}}(\boldsymbol{\lambda}_{1+n_j}), & \overline{\Delta}^{(2)}_{j+1} &= \mathbb{E}_{n_j}[\Delta^{(2)}_{j+1}].
\end{aligned}$$

这里第一对给出了序列中相邻项之间的差值 $(H_{n_j}(\boldsymbol{\lambda}_{1+n_j}))_{j=0}^{\infty}$，而第二项则是用于“撤销” $\boldsymbol{\lambda}_{1+n_j}$变化的修正项。接着，由于 $H_n(\boldsymbol{\lambda})$是一个上鞅（引理 7），并应用公式 (22) 和 (20a)，我们有

$$\begin{aligned}
\overline{\Delta}^{(1)}_{j+1} + \overline{\Delta}^{(2)}_{j+1} &= H_{n_j}(\boldsymbol{\lambda}_{1+n_j}) - \mathbb{E}_{n_j}[H_{n_{(j+1)}}(\boldsymbol{\lambda}_{1+n_j})]\\
&\geq H_{n_j}(\boldsymbol{\lambda}_{1+n_j}) - \mathbb{E}_{n_j}[H_{1+n_j}(\boldsymbol{\lambda}_{1+n_j})]\\
&= c_{m_{1+n_j}} \alpha^{n_j}_{\text{C-MOKG}}(\boldsymbol{x}_{1+n_j}, m_{1+n_j};\, \boldsymbol{\lambda}_{1+n_j}, \boldsymbol{c})\\
&> c_{m_{1+n_j}} \left(\sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha^{n_j}_{\text{C-MOKG}}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}_{1+n_j}, \boldsymbol{c}) - \eta_{n_j}\right).
\end{aligned}$$

因此，若我们能证明几乎必然地 $\Delta_j^{(1)} \to 0$ 且当 $j \to \infty$时 $\Delta_j^{(2)} \to 0$，那么我们就有几乎必然地 $\sup_{\boldsymbol{x}\in\mathcal{X},\ m\in[M]} \alpha^{n_j}\text{C-MOKG}(\boldsymbol{x},\ m;\ \boldsymbol{\lambda}_{1+n_j},\ \boldsymbol{c}) \to 0$。

对于任意 $j \in \mathbb{N}$，我们有 $\sum_{i=1}^{j} \Delta_i^{(1)} = H_0(\boldsymbol{\lambda}_1) - H_{n_j}(\boldsymbol{\lambda}_{1+n_j})$。因此，由于每个 $H_{n_j}(\boldsymbol{\lambda}_{1+n_j})$ 几乎必然非负（引理 7），

$$\begin{aligned}\forall j \in \mathbb{N} \quad \mathbb{E}\left[\sum_{i=1}^{j} \Delta_i^{(1)}\right] &= \mathbb{E}\left[\sum_{i=1}^{j} \Delta_i^{(1)}\right] \\ &= \mathbb{E}[H_0(\boldsymbol{\lambda}_1) - H_{n_j}(\boldsymbol{\lambda}_{1+n_j})] \le \mathbb{E}[H_0(\boldsymbol{\lambda}_1)] < \infty.\end{aligned}$$

因此，$\urcorner j \to \infty$ 几乎必然成立。

为了证明 $\Delta_j^{(2)} \to 0$ 几乎必然成立，我们将使用引理 8。事实上，由此可得，对所有 j,

$$\begin{aligned}\left|\Delta_{j+1}^{(2)}\right| &= \left|H_{n_{(j+1)}}(\boldsymbol{\lambda}_{1+n_{(j+1)}}) - H_{n_{(j+1)}}(\boldsymbol{\lambda}_{1+n_j})\right| \\ &\le 2\left\|\boldsymbol{\lambda}_{1+n_{(j+1)}} - \boldsymbol{\lambda}_{1+n_j}\right\|_2 \mathbb{E}_{n_{(j+1)}}\left[\max_{\boldsymbol{x}'\in\mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2\right].\end{aligned}$$

但是，根据 Lévy 零一律，$\mathbb{E}_{n_{(j+1)}}[\max_{\boldsymbol{x}'\in\mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2] \to \mathbb{E}_{\infty}[\max_{\boldsymbol{x}'\in\mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2]$ 在 $j \to \infty$ 时几乎必然成立。此外，根据构造，$\|\boldsymbol{\lambda}_{1+n_{(j+1)}} - \boldsymbol{\lambda}_{1+n_j}\|_2 \to 0$ 在 $j \to \infty$ 时几乎必然成立。因此，$\Delta_j^{(2)} \to 0$ 几乎必然成立。

因此,

$$0 \le \sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha_{\text{C-MOKG}}^{n_j}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}_{1+n_j}, \boldsymbol{c}) < \frac{1}{c_{m_{1+n_j}}}(\Delta_{j+1}^{(1)} + \Delta_{j+1}^{(2)}) + \eta_{n_j} \to 0$$

几乎必然有 $j \to \infty$。

为说明我们可以用其极限 $\boldsymbol{\lambda}$ 替代 $\boldsymbol{\lambda}_{1+n_j}$，注意到对于所有 $j \ge 0$,

$$\begin{aligned}&\sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha_{\text{C-MOKG}}^{n_j}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}, \boldsymbol{c}) \\ &\quad\le \sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha_{\text{C-MOKG}}^{n_j}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}_{1+n_j}, \boldsymbol{c}) \\ &\qquad+ \sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \left(\alpha_{\text{C-MOKG}}^{n_j}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}, \boldsymbol{c}) - \alpha_{\text{C-MOKG}}^{n_j}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}_{1+n_j}, \boldsymbol{c})\right) \\ &\quad\le \underbrace{\sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha_{\text{C-MOKG}}^{n_j}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}_{1+n_j}, \boldsymbol{c})}_{\to 0} \\ &\qquad+ 2\,\underbrace{\|\boldsymbol{\lambda} - \boldsymbol{\lambda}_{1+n_j}\|_2}_{\to 0}\ \underbrace{\mathbb{E}_{n_j}\left[\max_{\boldsymbol{x}'\in\mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2\right]}_{\to \mathbb{E}_{\infty}[\max_{\boldsymbol{x}'\in\mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2] < \infty} \\ &\quad\to 0 \quad \text{as } j \to \infty.\end{aligned}$$

在倒数第二行中，我们使用了 $\max_{\boldsymbol{x}'\in\mathcal{X}}\|\boldsymbol{f}(\boldsymbol{x}')\|_2$ 是 L^1-可积的这一事实，因此根据 Lévy 零一律，条件期望 $\mathbb{E}_{n_j}[\max_{\boldsymbol{x}'\in\mathcal{X}}\|\boldsymbol{f}(\boldsymbol{x}')\|_2]$ 几乎必然收敛于 $\mathbb{E}_\infty[\max_{\boldsymbol{x}'\in\mathcal{X}}\|\boldsymbol{f}(\boldsymbol{x}')\|_2]<\infty$。

因此，整个上界收敛于零，我们由此得出

$$\sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha^{n_j}_{\text{C-MOKG}}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}, \boldsymbol{c}) \to 0 \quad \text{as} \quad j\to\infty \quad \text{a.s..}$$

因此，$\liminf_{n\to\infty}\sup_{\boldsymbol{x}\in\mathcal{X},\ m\in[M]}\alpha^n_{\text{C-MOKG}}(\boldsymbol{x},\ m;\ \boldsymbol{\lambda},\ \boldsymbol{c}) = 0$ 几乎必然成立，从而完成证明。□

对于对标量化取期望的情形，我们也可以作出类似的陈述，不过在这种情况下，我们可以直接将下极限替换为真正的极限。

定理 11. 假设我们使用对标量化取期望的 *C-MOKG* 来选择样本。那么，对于所有偏好向量 $\boldsymbol{\lambda}\in\Lambda$, $\alpha^n_{C\text{-}MOKG}(\boldsymbol{x}, m;\ \boldsymbol{\lambda}, \boldsymbol{c}) \to 0$，当 $n\to\infty$ 时，该结论关于 $\boldsymbol{x}\in\mathcal{X}$ 和 $m\in\{1,\dots,M\}$ 一致成立，且几乎必然成立。也就是说，

$$\forall\boldsymbol{\lambda}\in\Lambda \quad \mathbb{P}\left(\sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha^n_{C\text{-}MOKG}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}, \boldsymbol{c}) \to 0 \quad as \quad n\to\infty\right) = 1.$$

证明。 对标量化取期望这一情形的证明比定理 9 的证明更简单，因为在优化的每一步中，我们都使用了相同的采集函数。

我们首先证明，$\alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{c}) \to 0$ 关于 $\boldsymbol{x}\in\mathcal{X}$ 和 $m\in\{1,2,\dots,M\}$ 一致收敛，且几乎必然。对于每个 $n\in\mathbb{N}_0$，我们重新定义 $\Delta_{n+1} = H_n - H_{n+1}$ 和 $\Delta_{n+1} = \mathbb{E}_n[\Delta_{n+1}]$。如前文公式 (23) 中所述，$\alpha^n_{\text{C-MOKG}}(\boldsymbol{x}_{n+1}, m_{n+1};\ \boldsymbol{c}) = \frac{1}{c_{m_{n+1}}}(H_n - \mathbb{E}_n[H_{n+1}])$。因此，根据公式 (20b)，由于 $\boldsymbol{x}_{n+1}$ 和 m_{n+1} 的选取旨在最大化 $\alpha^n_{\text{C-MOKG}}$，

$$\begin{aligned}\frac{1}{c_{m_{n+1}}}\Delta_{n+1} &= \frac{1}{c_{m_{n+1}}}\left(H_n - \mathbb{E}_n\left[H_{n+1}\right]\right)\\ &= \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}_{n+1}, m_{n+1};\ \boldsymbol{c}) > \sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{c}) - \eta_n. \quad (24)\end{aligned}$$

现在，我们的问题归结为证明当 $n\to\infty$ 时 $\Delta_n\to 0$。对于任意 $n\in\mathbb{N}_0$，我们有 $\sum_{j=1}^n \Delta_j = H_0 - H_n$。因此，

$$\forall n\in\mathbb{N}_0 \quad \mathbb{E}\left[\sum_{j=1}^n \Delta_j\right] = \mathbb{E}\left[\sum_{j=1}^n \Delta_j\right] = \mathbb{E}[H_0 - H_n] \le \mathbb{E}[H_0] < \infty$$

其中，倒数第二个不等式使用了引理 7。另请注意，将公式 (24) 与引理 1 结合可得

$$\forall n\in\mathbb{N}_0 \quad \Delta_{n+1} = c_{m_{n+1}}\alpha^n_{\text{C-MOKG}}(\boldsymbol{x}_{n+1}, m_{n+1}; \boldsymbol{c}) \ge 0$$

因此，当 $\Delta_{n+1} \to 0$ 时， $n \to \infty$。于是，

$$0 \leq \sup_{\boldsymbol{x} \in \mathcal{X},\, m \in [M]} \overline{\alpha}^n_{\text{C-MOKG}}(\boldsymbol{x}, m; \boldsymbol{c}) < \frac{1}{c_{m_{n+1}}} \overline{\Delta}_{n+1} + \eta_n \to 0$$

几乎必然地，作为 $n \to \infty$。

现在我们将利用这一点证明，对于所有 $\boldsymbol{\lambda} \in \Lambda$, $\alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\ \boldsymbol{\lambda}, \boldsymbol{c})$, $\to 0$关于 $\boldsymbol{x} \in \mathcal{X}$和 $m \in \{1, \ldots, M\}$一致收敛，且几乎必然。反设此结论不成立。则存在 $\boldsymbol{\lambda} \in \Lambda$，使得以非零概率

$$\limsup_{n \to \infty} \sup_{\boldsymbol{x} \in \mathcal{X},\, m \in [M]} \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}, \boldsymbol{c}) > 0.$$

由引理 8可知，对于所有 $\boldsymbol{\lambda}' \in \Lambda$，以及所有 $\boldsymbol{x} \in \mathcal{X}$、 $m \in \{1, \ldots, M\}$和 $n \in \mathbb{N}_0$,

$$\alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}', \boldsymbol{c}) \geq \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}, \boldsymbol{c}) - \frac{2}{c_m} \|\boldsymbol{\lambda} - \boldsymbol{\lambda}'\|_2\, \mathbb{E}_n \left[\max_{\boldsymbol{x}' \in \mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2 \right].$$

现在，$\max_{\boldsymbol{x}' \in \mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2$是 L^1-可积的，因此根据 Lévy 零一律， $\mathbb{E}_n \left[\max_{\boldsymbol{x}' \in \mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2\right] \to \mathbb{E}_\infty \left[\max_{\boldsymbol{x}' \in \mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2\right]$ 几乎必然地趋于 $n \to \infty$。因此，存在一个随机的 $N \in \mathbb{N}_0$，使得几乎必然地，对于所有 $n \geq N$，$\mathbb{E}_n \left[\max_{\boldsymbol{x}' \in \mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2\right] \leq \frac{3}{2} \mathbb{E}_\infty \left[\max_{\boldsymbol{x}' \in \mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2\right]$。也就是说，我们得到了 $\mathbb{E}_n[\max_{\boldsymbol{x}' \in \mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2]$ 的一个上界，该上界与 n无关。因此，存在一个以 $\boldsymbol{\lambda}$ 为中心、半径严格为正且随机的小 L^2-球 $B_{\boldsymbol{\lambda}} \subset \Lambda$，使得对于所有 $\boldsymbol{\lambda}' \in B_{\boldsymbol{\lambda}}$ 和 $n \geq N$，我们都有

$$\frac{2}{c_m} \|\boldsymbol{\lambda} - \boldsymbol{\lambda}'\|_2\, \mathbb{E}_n \left[\max_{\boldsymbol{x}' \in \mathcal{X}} \|\boldsymbol{f}(\boldsymbol{x}')\|_2 \right] < \min(A_{\boldsymbol{\lambda}}, 1)$$

其中 $A_{\boldsymbol{\lambda}} = \frac{1}{2} \limsup_{n \to \infty} \sup_{\boldsymbol{x} \in \mathcal{X},\, m \in [M]} \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\ \boldsymbol{\lambda}, \boldsymbol{c}) > 0$。这里取与 1 的最小值，以涵盖 $A_{\boldsymbol{\lambda}} = +\infty$ 的情况。因此，对于所有$n \geq N$和 $\boldsymbol{\lambda}' \in B_{\boldsymbol{\lambda}}$，以及所有 $\boldsymbol{x} \in \mathcal{X}$ 和 $m \in \{1, \ldots, M\}$,

$$\alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}', \boldsymbol{c}) \geq \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}, \boldsymbol{c}) - \min(A_{\boldsymbol{\lambda}}, 1).$$

由于 C-MOKG 非负（引理 1），将范围限制为 $B_{\boldsymbol{\lambda}}$ 并进行积分，可得对于所有 $\boldsymbol{x} \in \mathcal{X}$、 $m \in \{1, \ldots, M\}$ 以及所有 $n \geq N$，期望的一个下界，

$$\overline{\alpha}^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\, \boldsymbol{c}) \geq \mathbb{P}(B_{\boldsymbol{\lambda}}) (\alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}, \boldsymbol{c}) - \min(A_{\boldsymbol{\lambda}}, 1)).$$

然后，取上确界并求极限，得到

$$\begin{aligned}
&\limsup_{n \to \infty} \sup_{\boldsymbol{x} \in \mathcal{X},\, m \in [M]} \overline{\alpha}^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\, \boldsymbol{c}) \\
&\qquad \geq \mathbb{P}(B_{\boldsymbol{\lambda}}) \left(\limsup_{n \to \infty} \sup_{\boldsymbol{x} \in \mathcal{X},\, m \in [M]} \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}, \boldsymbol{c}) - \min(A_{\boldsymbol{\lambda}}, 1) \right) \\
&\qquad \geq \frac{1}{2} \mathbb{P}(B_{\boldsymbol{\lambda}})\ \limsup_{n \to \infty} \sup_{\boldsymbol{x} \in \mathcal{X},\, m \in [M]} \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}, \boldsymbol{c}) \\
&\qquad > 0.
\end{aligned}$$

根据 $\boldsymbol{\lambda}$ 的选取，并且由于 Λ 上的分布处处严格为正，最终的严格不等式以非零概率成立。因此，我们得出矛盾，从而得出结论：对于所有 $\boldsymbol{\lambda} \in \Lambda$，$\sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha^n_{\text{C-MOKG}}(\boldsymbol{x},\ m;\ \boldsymbol{\lambda},\ \boldsymbol{c}) \to 0$几乎必然成立。□

A.3 收敛至最优推荐点

我们已经证明，对于这两个采集函数中的每一个，$\alpha^n_{\text{C-MOKG}}(\cdot,\ \cdot;\ \boldsymbol{\lambda},\ \boldsymbol{c})$的下极限几乎必然对每个 $\boldsymbol{\lambda} \in \Lambda$ 均为零。事实上，对于 $\alpha^n_{\text{C-MOKG}}$，我们已经证明其几乎必然一致收敛于 0。然而，我们的目标是得到一个强得多的结果。

回顾正文中的公式（14），我们在此将其重述为公式（25）。对于每个 $n \in \mathbb{N}_0$ 和每个偏好向量 $\boldsymbol{\lambda} \in \Lambda$，令

$$\boldsymbol{x}^*_{n,\boldsymbol{\lambda}} \in \arg\max_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_n[\boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x})] \tag{25}$$

它是一个随机变量，在第 n 阶段使标量化目标的后验均值达到最大。因此，假设不存在模型失配，$\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}^*_{1,\ \boldsymbol{\lambda}})$ $\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}^*_{2,\ \boldsymbol{\lambda}})\ldots$就是在优化的每个阶段都采用推荐点时，我们将获得的（无噪声）标量化目标值序列。下一定理针对单目标情形作为 [8], 中的命题 4.9 给出，它表明，该序列收敛于标量化隐藏目标函数 $\boldsymbol{\lambda} \cdot \boldsymbol{f}$ 的真实最大值。

定理 2（重述自正文）。定义 $\boldsymbol{x}^*_{n,\ \boldsymbol{\lambda}}$，如公式 (25) 所示。使用 *C-MOKG* 并采用随机标量化或对标量化取期望时，我们有

$$\forall \boldsymbol{\lambda} \in \Lambda \quad \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}}) \to \max_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}) \quad as \quad n \to \infty$$

几乎必然且在均值意义下成立。

在证明这一点之前，我们将先证明一个重要结果：无论使用何种采集函数，高斯过程代理模型的后验均值与协方差函数都几乎必然一致收敛。该结果对于证明定理 2 至关重要。

令 $\mathcal{F}_\infty = \sigma(\cup_{n=0}^\infty \mathcal{F}_n)$ 为包含所有 $\mathcal{F}_n$ 的最小 σ-代数，并令 $\mathbb{E}_\infty[\cdot] = \mathbb{E}[\cdot\,|\,\mathcal{F}_\infty]$ 表示在给定所有观测和选择条件下的期望。此外，对于每个 $n \in \mathbb{N}_0 \cup \{\infty\}$，定义 $\boldsymbol{f}$ 的条件均值函数和协方差函数，

$$\begin{aligned}
&\boldsymbol{\mu}_n : \mathcal{X} \to \mathbb{R}^M, \quad \boldsymbol{x} \mapsto \mathbb{E}_n[\boldsymbol{f}(\boldsymbol{x})],\\
&\mathbf{K}_n : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^{M\times M}, \quad (\boldsymbol{x}, \boldsymbol{x}') \mapsto \mathbb{E}_n[\boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x}')^T] - \mathbb{E}_n[\boldsymbol{f}(\boldsymbol{x})]\mathbb{E}_n[\boldsymbol{f}(\boldsymbol{x}')]^T.
\end{aligned}$$

请注意，这些是向量值和矩阵值随机过程。

命题 12. 对于查询位置的任意选择$(\boldsymbol{x}_n, m_n)_{n=1}^{\infty}$，随机过程序列 $\boldsymbol{\mu}_n \to \boldsymbol{\mu}_\infty$ 和 $\mathbf{K}_n \to \mathbf{K}_\infty$ 均匀收敛于 $n \to \infty$，且对所有 $1 \leq p < \infty$，几乎必然地并在 L^p 中收敛。此外，极限 $\boldsymbol{\mu}_\infty$ 和$\mathbf{K}_\infty$ 是连续的。

证明。 由于 $\mathcal{X}$是紧的，连续函数 $\mathcal{X} \to \mathbb{R}^M$组成的空间 $\mathcal{C}(\mathcal{X}, \mathbb{R}^M)$ 在赋予上确界范数 $\|\cdot\|_\infty$后构成一个巴拿赫空间。多输出高斯过程 $\boldsymbol{f}$ 具有连续样本路径，因此可视为该空间中的随机元素。对于任意 $1 \leq p < \infty$，它也是 L^p-可积的。事实上，$\sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} |f_m(\boldsymbol{x})|$ 的所有矩均有限[6]，因此

$$\|\boldsymbol{f}\|_{L^p(\mathcal{C}(\mathcal{X},\mathbb{R}^M))} = \mathbb{E}\left[\|\boldsymbol{f}\|_\infty^p\right]^{1/p} = \mathbb{E}\left[\left(\sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} |f_m(\boldsymbol{x})|\right)^p\right]^{1/p} < \infty.$$

注意，条件均值 $\boldsymbol{\mu}_n = \mathbb{E}_n[\boldsymbol{f}]$ 和 $\boldsymbol{\mu}_\infty = \mathbb{E}_\infty[\boldsymbol{f}]$ 是连续的，因为巴拿赫空间在取条件期望时是封闭的（例如参见 [30]中的命题1.10）。事实上，我们已证明它们是 [30], 中定理1.14和1.30所述形式的鞅，因此 $\boldsymbol{\mu}_n \to \boldsymbol{\mu}_\infty$ 几乎必然且在 L^p 中收敛。这是在函数空间 $\mathcal{C}(\mathcal{X}, \mathbb{R}^M)$ 中使用上确界范数的收敛，这等价于过程一致收敛。也就是说，对于任意$1 \leq p < \infty$， $\boldsymbol{\mu}_n \to \boldsymbol{\mu}_\infty$ 一致收敛，且这种收敛既几乎必然，也在 L^p 中成立。

我们可以对巴拿赫空间 $\mathcal{C}(\mathcal{X} \times \mathcal{X}, \mathbb{R}^{M\times M})$ 中 $\boldsymbol{f}$ 的二阶矩序列使用同样的论证。令 $\mathbf{M}_n^{(2)}(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}_n[\boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x}')^T]$ 表示 $\boldsymbol{f}$ 的二阶矩。由$\mathbf{F}^{(2)}(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x}')^T$定义的过程 $\mathbf{F}^{(2)} \in \mathcal{C}(\mathcal{X} \times \mathcal{X}, \mathbb{R}^{M\times M})$ 对所有 $1 \leq p < \infty$ 都是 L^p-可积的，因为

$$\begin{aligned}\|\mathbf{F}^{(2)}\|_{L^p(\mathcal{C}(\mathcal{X}\times\mathcal{X},\, \mathbb{R}^{M\times M}))} &= \mathbb{E}\left[\Big(\sup_{\substack{\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}\\ m,m'\in[M]}} |f_m(\boldsymbol{x}) f_{m'}(\boldsymbol{x}')|\Big)^p\right]^{1/p} \\ &= \mathbb{E}\left[\Big(\sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} |f_m(\boldsymbol{x})|\Big)^{2p}\right]^{1/p} = \|\boldsymbol{f}\|^2_{L^{2p}(\mathcal{C}(\mathcal{X},\, \mathbb{R}^M))} < \infty.\end{aligned}$$

因此，我们可以再次应用 [30] 中的鞅收敛定理1.14和1.30，得出 $\mathbf{M}_n^{(2)} \to \mathbf{M}_\infty^{(2)}$ 几乎必然且在 L^p 中成立，其中我们记 $\mathbf{M}_\infty^{(2)}(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}_\infty[\boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x}')^T]$。由于

$$\forall n \in \mathbb{N} \cup \{\infty\}\ \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X} \quad \mathbf{K}_n(\boldsymbol{x}, \boldsymbol{x}') = \mathbf{M}_n^{(2)}(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{\mu}_n(\boldsymbol{x})\boldsymbol{\mu}_n(\boldsymbol{x}')^T,$$

我们得出，对于所有 $1 \leq p < \infty$，$\mathbf{K}_n \to \mathbf{K}_\infty$ 一致收敛，且几乎必然并在 L^p 中成立。根据前面的同一论证，$\mathbf{M}_n^{(2)}$ 和 $\mathbf{M}_\infty^{(2)}$ 是连续的，因为它们是巴拿赫空间中连续函数的随机元素的条件期望。因此，$\mathbf{K}_n$ 和 $\mathbf{K}_\infty$ 也是连续的。□

[6] 例如参见 [6] 中的公式 2.34，并将 $\boldsymbol{f}$ 视为定义在输入空间$\mathcal{X} \times [M]$上的 $\mathbb{R}$ 值高斯过程

证明定理 2 的策略是，首先证明极限估计 $\boldsymbol{\mu}_\infty$ 几乎必然与目标函数 $\boldsymbol{f}$ 相差一个常数。也就是说，我们将证明 $\boldsymbol{f}-\boldsymbol{\mu}_\infty$ 几乎必然具有常值样本路径。直观而言，知道 $\boldsymbol{\mu}_\infty$ 就足以确定 $\boldsymbol{f}$的最大值所在位置。为了证明 $\boldsymbol{f}-\boldsymbol{\mu}_\infty$ 几乎必然具有常值样本路径，我们将首先证明极限协方差函数 $\mathbf{K}_\infty$ 几乎必然具有常值样本路径。

引理 13. 极限协方差函数 $\mathbf{K}_\infty\colon\ \mathcal{X}\times\mathcal{X}\to\mathbb{R}^{M\times M}$ 几乎必然具有常值样本路径。也就是说，

$$\mathbb{P}\big(\forall \boldsymbol{x},\boldsymbol{u},\boldsymbol{x}',\boldsymbol{u}'\in\mathcal{X}\quad \mathbf{K}_\infty(\boldsymbol{u},\boldsymbol{x})=\mathbf{K}_\infty(\boldsymbol{u}',\boldsymbol{x}')\big)=1.$$

证明. 此证明利用了标量化是线性的这一事实，因此它们与期望算子可交换。

设 $\boldsymbol{x}\in\mathcal{X}$、 $m\in\{1,\dots,M\}$ 和 $\boldsymbol{\lambda}\in\Lambda$。对于每个 $n\in\mathbb{N}_0$ 和 $\boldsymbol{u}\in\mathcal{X}$，令

$$\begin{aligned}W_{n,\boldsymbol{u}}&=\mathbb{E}_n[\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{u})\,|\,y_m^{\boldsymbol{x}}]-\mathbb{E}_n[\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}_{n,\boldsymbol{\lambda}}^*)\,|\,y_m^{\boldsymbol{x}}]\\&=\boldsymbol{\lambda}\cdot\mathbb{E}_n[\boldsymbol{f}(\boldsymbol{u})\,|\,y_m^{\boldsymbol{x}}]-\boldsymbol{\lambda}\cdot\mathbb{E}_n[\boldsymbol{f}(\boldsymbol{x}_{n,\boldsymbol{\lambda}}^*)\,|\,y_m^{\boldsymbol{x}}].\end{aligned}$$

注意， $\alpha^n_{\text{C-MOKG}}(\boldsymbol{x},m;\ \boldsymbol{\lambda},\boldsymbol{c})=\frac{1}{c_m}\mathbb{E}_n[\sup_{\boldsymbol{x}'\in\mathcal{X}}W_{n,\boldsymbol{x}'}]$。

设 $n\in\mathbb{N}_0$ 和 $\boldsymbol{u}\in\mathcal{X}$。由于 $\sup_{\boldsymbol{x}'\in\mathcal{X}}W_{n,\boldsymbol{x}'}\ge 0$，我们有

$$\begin{aligned}\sup_{\boldsymbol{x}'\in\mathcal{X}}W_{n,\boldsymbol{x}'}&=\sup_{\boldsymbol{x}'\in\mathcal{X}}\max(W_{n,\boldsymbol{x}'},0)\\\Rightarrow\qquad 0\le\mathbb{E}_n[\max(W_{n,\boldsymbol{u}},0)]&\le\mathbb{E}_n\left[\sup_{\boldsymbol{x}'\in\mathcal{X}}\max(W_{n,\boldsymbol{x}'},0)\right]\\&=c_m\alpha^n_{\text{C-MOKG}}(\boldsymbol{x},m;\,\boldsymbol{\lambda},\boldsymbol{c}).\end{aligned}$$

根据采集策略应用定理 9 或定理 11，我们几乎必然有 $\liminf_{n\to\infty}\sup_{\boldsymbol{x}\in\mathcal{X},\ m\in[M]}\alpha^n_{\text{C-MOKG}}(\boldsymbol{x},\ m;\ \boldsymbol{\lambda},\ \boldsymbol{c})=0$。因此，$\liminf_{n\to\infty}\mathbb{E}_n[\max(W_{n,\ \boldsymbol{u}},\ 0)]=0$ 几乎必然成立。这意味着，几乎必然，

$$\forall\delta>0\quad\liminf_{n\to\infty}\mathbb{P}_n(W_{n,\boldsymbol{u}}>\delta)=0,\tag{26}$$

这与依概率收敛的定义非常相似。这里，(随机)分布 $\mathbb{P}_n$ 定义为示性变量的条件期望。也就是说，对于每个事件 $A\in\mathcal{F}$， $\mathbb{P}_n(A)=\mathbb{E}_n[\mathbb{I}_A]$。令$\boldsymbol{k}_{\infty,\ :,\ m}$ 表示$\mathbf{K}_\infty$的 m^{th} 列，我们将使用公式 (26)来证明

$$\boldsymbol{\lambda}\cdot\boldsymbol{k}_{\infty,:,m}(\boldsymbol{u},\boldsymbol{x})=\boldsymbol{\lambda}\cdot\boldsymbol{k}_{\infty,:,m}(\boldsymbol{x},\boldsymbol{x})\tag{27}$$

几乎必然成立——注意到 $W_{n,\ \boldsymbol{u}}$ 是一个高斯变量，这使证明容易得多。

令 $k_{n,m,m}$ 表示 $\mathbf{K}_n$ 的 $(m,m)^{\text{th}}$ 元素。若 $k_{n,m,m}(\boldsymbol{x},\boldsymbol{x})\ =\ 0$，则 $k_{n',m,m}(\boldsymbol{x},\boldsymbol{x})=0$ 对所有 $n'\ge n$ 均成立；并且根据柯西-施瓦茨不等式，$k_{n',m',m}(\boldsymbol{u},\boldsymbol{x})=0$ 对所有 $m'\in\{1,\dots,M\}$ 以及所有 $n'\ge n$ 也成立。因此，$\boldsymbol{k}_{n,:,m}(\boldsymbol{u},\boldsymbol{x})-\boldsymbol{k}_{n,:,m}(\boldsymbol{x},\boldsymbol{x})=0$

对所有充分大的 n 均成立，因此应用命题 12，我们得到了公式 (27)。

因此，我们关注满足以下条件的事件：$k_{n,\,m,\,m}(\boldsymbol{x},\ \boldsymbol{x}) > 0$ 对所有 n 均成立。利用公式 (4a) 中高斯过程条件均值的公式，我们可以将$W_{n,\,\boldsymbol{u}}$ 表示为

$$W_{n,\boldsymbol{u}} = \boldsymbol{\lambda} \cdot \left(\boldsymbol{\mu}_n(\boldsymbol{u}) - \boldsymbol{\mu}_n(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}}) + \frac{\boldsymbol{k}_{n,:,m}(\boldsymbol{u},\boldsymbol{x}) - \boldsymbol{k}_{n,:,m}(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}},\,\boldsymbol{x})}{k_{n,m,m}(\boldsymbol{x},\boldsymbol{x}) + \sigma_m^2} (y_m^{\boldsymbol{x}} - \mu_{n,m}(\boldsymbol{x})) \right)$$

其中，$y_m^{\boldsymbol{x}} \sim \mathcal{N}(\mu_{n,\,m}(\boldsymbol{x})\, k_{n,\,m,\,m}(\boldsymbol{x},\ \boldsymbol{x}) + \sigma_m^2)$ 是在 $\boldsymbol{x}$处对第 m^{th} 个目标函数进行采样时观测到的（随机）值。因此，在给定 $\mathcal{F}_n$的条件下，变量$W_{n,\,\boldsymbol{u}}$ 本身服从正态分布，其均值和协方差为

$$\begin{aligned} \mathbb{E}_n[W_{n,\boldsymbol{u}}] &= \boldsymbol{\lambda} \cdot (\boldsymbol{\mu}_n(\boldsymbol{u}) - \boldsymbol{\mu}_n(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}})), \\ \operatorname{Var}_n[W_{n,\boldsymbol{u}}] &= \frac{(\boldsymbol{\lambda} \cdot [\boldsymbol{k}_{n,:,m}(\boldsymbol{u},\boldsymbol{x}) - \boldsymbol{k}_{n,:,m}(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}},\boldsymbol{x})])^2}{k_{n,m,m}(\boldsymbol{x},\boldsymbol{x}) + \sigma_m^2}. \end{aligned}$$

因此，对任意 $\delta > 0$,

$$\begin{aligned} &\mathbb{P}_n(W_{n,\boldsymbol{u}} > \delta) \\ &= 1 - \Phi\left(\frac{\sqrt{k_{n,m,m}(\boldsymbol{x},\boldsymbol{x}) + \sigma_m^2}}{|\boldsymbol{\lambda} \cdot (\boldsymbol{k}_{n,:,m}(\boldsymbol{u},\boldsymbol{x}) - \boldsymbol{k}_{n,:,m}(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}},\boldsymbol{x}))|} [\delta - \boldsymbol{\lambda} \cdot (\boldsymbol{\mu}_n(\boldsymbol{u}) - \boldsymbol{\mu}_n(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}}))] \right) \end{aligned}$$

其中，Φ 是标准正态变量的累积分布函数。

根据命题 12，$\boldsymbol{\mu}_n \to \boldsymbol{\mu}_\infty$ 关于 $n \to \infty$ 一致收敛，且几乎必然成立。此外，$\boldsymbol{\mu}_\infty$ 是连续的，且 $\mathcal{X}$ 是紧的。因此，存在一个负的随机变量 a （不依赖于 n），使得当n充分大时，$\mathbb{E}_n[W_{n,\boldsymbol{u}}] = \boldsymbol{\lambda} \cdot (\boldsymbol{\mu}_n(\boldsymbol{u}) - \boldsymbol{\mu}_n(\boldsymbol{x}^*_{n,\ \boldsymbol{\lambda}})) > a$ 几乎必然成立。例如，取任意$a < \min \boldsymbol{\lambda} \cdot \boldsymbol{\mu}_\infty - \max\ \boldsymbol{\lambda} \cdot \boldsymbol{\mu}_\infty$。于是，

$$\mathbb{P}_n(W_{n,\boldsymbol{u}} > \delta) \geq 1 - \Phi\left(\frac{\sqrt{k_{n,m,m}(\boldsymbol{x},\boldsymbol{x}) + \sigma_m^2}}{|\boldsymbol{\lambda} \cdot (\boldsymbol{k}_{n,:,m}(\boldsymbol{u},\boldsymbol{x}) - \boldsymbol{k}_{n,:,m}(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}},\boldsymbol{x}))|} (\delta - a) \right).$$

根据公式 (26)，存在一个（随机）子序列 $(n_j)_{j=0}^{\infty}$，使得概率 $\mathbb{P}_{n_j}(W_{n_j,\ \boldsymbol{u}} > \delta) \to 0$，当 $j \to \infty$ 时。因此，在这个子序列上，$\frac{\sqrt{k_{n_j\ m\ m}\ \boldsymbol{x}\ \boldsymbol{x} + \sigma_m^2}\ ,\ ,\ (\ ,\ )\ (\boldsymbol{k}_{n_j,}, (\ ,\ ) - \boldsymbol{k}_{n_j,}, (\boldsymbol{x}^*_{n_j,}\boldsymbol{\lambda},\boldsymbol{x}))|}{|\boldsymbol{\lambda} \cdot \ :\ m\ \boldsymbol{u}\ \boldsymbol{x}} \to \infty$ 作为 $j \to \infty$。由于 $(k_{n,}\boldsymbol{m},\boldsymbol{m}(\boldsymbol{x},\boldsymbol{x}))_{n=1}^{\infty}$ 是一个递减序列，因此必有 $\boldsymbol{\lambda} \cdot (\boldsymbol{k}_{n_j,}\boldsymbol{:}\ m(\boldsymbol{u},\boldsymbol{x}) - \boldsymbol{k}_{n_j,}\boldsymbol{:}\ m(\boldsymbol{x}^*_{n_j,}\boldsymbol{\lambda},\boldsymbol{x})) \to 0$

几乎必然。对于所有 $\boldsymbol{u}$，包括 $\boldsymbol{u}=\boldsymbol{x}$ 的情况，这一点都成立，因此对于所有 $\boldsymbol{u}\in\mathcal{X}$，

$$\boldsymbol{\lambda}\cdot(\boldsymbol{k}_{n_j,:,m}(\boldsymbol{u},\boldsymbol{x})-\boldsymbol{k}_{n_j,:,m}(\boldsymbol{x},\boldsymbol{x}))= \left(\boldsymbol{\lambda}\cdot(\boldsymbol{k}_{n_j,:,m}(\boldsymbol{u},\boldsymbol{x})-\boldsymbol{k}_{n_j,:,m}(\boldsymbol{x}^*_{n_j,\boldsymbol{\lambda}},\boldsymbol{x})) -\boldsymbol{\lambda}\cdot(\boldsymbol{k}_{n_j,:,m}(\boldsymbol{x},\boldsymbol{x})-\boldsymbol{k}_{n_j,:,m}(\boldsymbol{x}^*_{n_j,\boldsymbol{\lambda}},\boldsymbol{x}))\right)\to 0$$

几乎必然地满足 $j\to\infty$。根据命题 12，序列 $\boldsymbol{\lambda}\cdot(\boldsymbol{k}_{n,:,m}(\boldsymbol{u},\boldsymbol{x})-\boldsymbol{k}_{n,:,m}(\boldsymbol{x},\boldsymbol{x}))\to\boldsymbol{\lambda}\cdot(\boldsymbol{k}_{\infty,:,m}(\boldsymbol{u},\boldsymbol{x})-\boldsymbol{k}_{\infty,:,m}(\boldsymbol{x},\boldsymbol{x}))$ 收敛于 $n\to\infty$。我们已经证明该极限为零，因为一个子序列的极限为零。这就是公式 (27)。

因此，对于所有 $\boldsymbol{u}\in\mathcal{X}$, $\boldsymbol{\lambda}\cdot\boldsymbol{k}_{\infty,:,m}(\boldsymbol{u},\boldsymbol{x})=\boldsymbol{\lambda}\cdot\boldsymbol{k}_{\infty,:,m}(\boldsymbol{x},\boldsymbol{x})$，结论成立。对于所有 $\boldsymbol{\lambda}\in\Lambda$ 和 $m\in\{1,\dots,M\}$，这一结论也成立，因此 $\mathbf{K}_\infty(\boldsymbol{u},\boldsymbol{x})=\mathbf{K}_\infty(\boldsymbol{x},\boldsymbol{x})$。此外，对于所有 $\boldsymbol{x}\in\mathcal{X}$，结论也成立，因此利用 $\mathbf{K}_\infty$ 关于其参数的对称性，我们有：

$$\forall\boldsymbol{x},\boldsymbol{u},\boldsymbol{x}',\boldsymbol{u}'\in\mathcal{X}\quad\mathbb{P}\big(\mathbf{K}_\infty(\boldsymbol{u},\boldsymbol{x})=\mathbf{K}_\infty(\boldsymbol{u}',\boldsymbol{x}')\big)=1.$$

为将这一结论扩展到同时适用于所有 $\boldsymbol{x},\boldsymbol{u}\in\mathcal{X}$ 的情形，注意到 $\mathcal{X}$ 是一个紧致度量空间，因而是可分的。也就是说，它有一个可数稠密子集 $\mathcal{X}'\subset\mathcal{X}$。根据 $\mathbb{P}$ 的可数次可加性，我们有

$$\mathbb{P}\big(\forall\boldsymbol{x},\boldsymbol{u},\boldsymbol{x}',\boldsymbol{u}'\in\mathcal{X}'\quad\mathbf{K}_\infty(\boldsymbol{u},\boldsymbol{x})=\mathbf{K}_\infty(\boldsymbol{u}',\boldsymbol{x}')\big)=1.$$

于是，由$\mathbf{K}_\infty$的样本路径连续性可知，这对所有 $\mathcal{X}$都成立，

$$\mathbb{P}\big(\forall\boldsymbol{x},\boldsymbol{u},\boldsymbol{x}',\boldsymbol{u}'\in\mathcal{X}\quad\mathbf{K}_\infty(\boldsymbol{u},\boldsymbol{x})=\mathbf{K}_\infty(\boldsymbol{u}',\boldsymbol{x}')\big)=1.$$

也就是说，$\mathbf{K}_\infty$ 的样本路径几乎必然为常数。 □

引理 14. 若$\mathbf{K}_\infty$ 的样本路径几乎必然为常数，则 $\boldsymbol{f}-\boldsymbol{\mu}_\infty$ 的样本路径也几乎必然为常数。

证明。 令 $\boldsymbol{u},\boldsymbol{x}\in\mathcal{X}$。则

$$\begin{aligned}&\mathrm{Var}_\infty[(\boldsymbol{f}(\boldsymbol{u})-\boldsymbol{\mu}_\infty(\boldsymbol{u}))-(\boldsymbol{f}(\boldsymbol{x})-\boldsymbol{\mu}_\infty(\boldsymbol{x}))]\\&=\mathrm{Var}_\infty[\boldsymbol{f}(\boldsymbol{u})-\boldsymbol{f}(\boldsymbol{x})]=\underbrace{\mathbf{K}_\infty(\boldsymbol{u},\boldsymbol{u})-\mathbf{K}_\infty(\boldsymbol{x},\boldsymbol{u})}_{0\text{ a.s.}}+\underbrace{\mathbf{K}_\infty(\boldsymbol{x},\boldsymbol{x})-\mathbf{K}_\infty(\boldsymbol{u},\boldsymbol{x})}_{0\text{ a.s.}}=0.\end{aligned}$$

注意到 $\mathbb{E}_\infty[(\boldsymbol{f}(\boldsymbol{u})-\boldsymbol{\mu}_\infty(\boldsymbol{u}))-(\boldsymbol{f}(\boldsymbol{x})-\boldsymbol{\mu}_\infty(\boldsymbol{x}))]=0$ 几乎必然成立，我们可以利用全方差公式将这一结果推广到无条件方差。

$$\begin{aligned}&\mathrm{Var}[(\boldsymbol{f}(\boldsymbol{u})-\boldsymbol{\mu}_\infty(\boldsymbol{u}))-(\boldsymbol{f}(\boldsymbol{x})-\boldsymbol{\mu}_\infty(\boldsymbol{x}))]\\&\qquad=\mathbb{E}\left[\underbrace{\mathrm{Var}_\infty[(\boldsymbol{f}(\boldsymbol{u})-\boldsymbol{\mu}_\infty(\boldsymbol{u}))-(\boldsymbol{f}(\boldsymbol{x})-\boldsymbol{\mu}_\infty(\boldsymbol{x}))]}_{=0\text{ a.s.}}\right]\\&\qquad\quad+\mathrm{Var}\left[\underbrace{\mathbb{E}_\infty[(\boldsymbol{f}(\boldsymbol{u})-\boldsymbol{\mu}_\infty(\boldsymbol{u}))-(\boldsymbol{f}(\boldsymbol{x})-\boldsymbol{\mu}_\infty(\boldsymbol{x}))]}_{=0\text{ a.s.}}\right]\\&\qquad=0.\end{aligned}$$

因此，$(\boldsymbol{f}(\boldsymbol{u}) - \boldsymbol{\mu}_\infty(\boldsymbol{u})) - (\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{\mu}_\infty(\boldsymbol{x}))$ 是一个均值为零、协方差矩阵为零的随机向量，因此 $\boldsymbol{f}(\boldsymbol{u}) - \boldsymbol{\mu}_\infty(\boldsymbol{u}) = \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{\mu}_\infty(\boldsymbol{x})$ 几乎必然成立。

对于每一对 $\boldsymbol{x}$，$\boldsymbol{u} \in \mathcal{X}$，这以概率 1 成立。为了得出它对所有 $\boldsymbol{x}$，$\boldsymbol{u} \in \mathcal{X}$ 同时以概率 1 成立，我们采用引理 13 末尾的相同论证。为此，只需注意到 $\mathcal{X}$ 是可分的，且 $\boldsymbol{f}$ 和 $\boldsymbol{\mu}_\infty$ 的样本路径连续。□

既然已经证明 $\boldsymbol{\mu}_\infty$ 与 $\boldsymbol{f}$ 相差一个常数因子，我们现在可以加强定理 9。这对定理 2 的证明并非必要，但确实能增进我们对多目标知识梯度行为的理解。

推论 15. 假设我们使用 *C-MOKG* 选取样本，其中随机标量化权重 $\boldsymbol{\lambda}_1$、$\boldsymbol{\lambda}_2$、……相互独立，并按照分布 $p(\boldsymbol{\lambda})$选取。那么，

$$\forall \boldsymbol{\lambda} \in \Lambda \quad \mathbb{P}\left(\sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha^n_{C\text{-}MOKG}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}, \boldsymbol{c}) \to 0 \quad as \quad n \to \infty \right) = 1.$$

证明。令 $\boldsymbol{\lambda} \in \Lambda$。根据 Lévy 零一律，

$$\mathbb{E}_n\left[\max_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x})\right] \to \mathbb{E}_\infty\left[\max_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x})\right]$$

当 $n \to \infty$ 时，几乎必然成立。此外，我们在命题 12 中证明了，当 $n \to \infty$ 时，$\boldsymbol{\mu}_n \to \boldsymbol{\mu}_\infty$一致收敛。因此，残余不确定性也收敛，

$$\begin{aligned} H_n(\boldsymbol{\lambda}) = \mathbb{E}_n\left[\max_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x})\right] - \max_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{\lambda}\cdot\boldsymbol{\mu}_n(\boldsymbol{x}) \\ \to \mathbb{E}_\infty\left[\max_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x})\right] - \max_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{\lambda}\cdot\boldsymbol{\mu}_\infty(\boldsymbol{x}) = H_\infty(\boldsymbol{\lambda}) \end{aligned}$$

作为 $n \to \infty$ 几乎必然。然而，根据引理 14，$\boldsymbol{f} - \boldsymbol{\mu}_\infty$的样本路径几乎必然为常数，因此 $\max_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}^*_{\infty,\boldsymbol{\lambda}})$ 几乎必然。于是，

$$H_n(\boldsymbol{\lambda}) \to H_\infty(\boldsymbol{\lambda}) = \mathbb{E}_\infty[\boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}^*_{\infty,\boldsymbol{\lambda}})] - \boldsymbol{\lambda}\cdot\boldsymbol{\mu}_\infty(\boldsymbol{x}^*_{\infty,\boldsymbol{\lambda}}) = 0 \quad \text{(a.s.)}.$$

最后，我们注意到，根据引理 6，

$$0 \le \sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\, \boldsymbol{\lambda}, \boldsymbol{c}) \le H_n(\boldsymbol{\lambda}) \to 0$$

因此 $\sup_{\boldsymbol{x}\in\mathcal{X},\, m\in[M]} \alpha^n_{\text{C-MOKG}}(\boldsymbol{x}, m;\ \boldsymbol{\lambda}, \boldsymbol{c}) \to 0$几乎必然。 □

根据引理 14，我们现在也可以证明定理 2。

证明（定理 *2* 的证明）。令 $\boldsymbol{\lambda} \in \Lambda$。回顾公式（25），对于每个$n$，我们定义了 $\boldsymbol{x}^*_{n,\boldsymbol{\lambda}} \in \arg\max \mathbb{E}_n[\boldsymbol{\lambda} \cdot \boldsymbol{f}]$ 为一个随机变量，它在阶段 n使标量化目标的后验均值最大化。类似地，令 $\boldsymbol{x}^*_{\boldsymbol{\lambda}} \in \arg\max \boldsymbol{\lambda} \cdot \boldsymbol{f}$为一个使标量化目标$\boldsymbol{\lambda} \cdot \boldsymbol{f}$最大化的随机变量。注意，这些选择不一定唯一。

我们首先希望证明，$\boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}}) \to \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{\boldsymbol{\lambda}})$ 几乎必然成立。根据 $\boldsymbol{x}^*_{\boldsymbol{\lambda}}$ 的定义，对于每个 n，量 $\boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{\boldsymbol{\lambda}}) - \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}})$ 都是非负的。因此，只需证明几乎必然有 $\limsup_{n\to\infty} \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{\boldsymbol{\lambda}}) - \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}}) \le 0$。

注意，我们可以拆分上极限，得到

$$\begin{aligned}
&\limsup_{n\to\infty} \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{\boldsymbol{\lambda}}) - \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}}) \\
&\quad \le \limsup_{n\to\infty} \Big(\boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{\boldsymbol{\lambda}}) - \boldsymbol{\lambda} \cdot \boldsymbol{\mu}_\infty(\boldsymbol{x}^*_{\boldsymbol{\lambda}})\Big) - \Big(\boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}}) - \boldsymbol{\lambda} \cdot \boldsymbol{\mu}_\infty(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}})\Big) \\
&\qquad + \limsup_{n\to\infty} \Big(\boldsymbol{\lambda} \cdot \boldsymbol{\mu}_\infty(\boldsymbol{x}^*_{\boldsymbol{\lambda}}) - \boldsymbol{\lambda} \cdot \boldsymbol{\mu}_n(\boldsymbol{x}^*_{\boldsymbol{\lambda}})\Big) - \Big(\boldsymbol{\lambda} \cdot \boldsymbol{\mu}_\infty(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}}) - \boldsymbol{\lambda} \cdot \boldsymbol{\mu}_n(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}})\Big) \\
&\qquad + \limsup_{n\to\infty} \boldsymbol{\lambda} \cdot \boldsymbol{\mu}_n(\boldsymbol{x}^*_{\boldsymbol{\lambda}}) - \boldsymbol{\lambda} \cdot \boldsymbol{\mu}_n(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}}).
\end{aligned}$$

根据引理 13 和引理 14，$\boldsymbol{\lambda} \cdot \boldsymbol{f} - \boldsymbol{\lambda} \cdot \boldsymbol{\mu}_\infty$ 的样本路径几乎必然为常数，因此上面第一行几乎必然为零。由于 $\boldsymbol{\mu}_n \to \boldsymbol{\mu}_\infty$ 几乎必然一致收敛（命题 12），中间一行也几乎必然为零。最后，根据 $\boldsymbol{x}^*_{n,\boldsymbol{\lambda}}$ 的定义，最下面一行不大于零。因此

$$\limsup_{n\to\infty} \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{\boldsymbol{\lambda}}) - \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}}) \le 0$$

如所需，由此得出 $\boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{n,\boldsymbol{\lambda}}) \to \boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x}^*_{\boldsymbol{\lambda}})$ 几乎必然。

均值收敛由控制收敛定理得出，其支配变量为 $\sup_{\boldsymbol{x}} |\boldsymbol{\lambda} \cdot \boldsymbol{f}(\boldsymbol{x})|$。□

B 更多实验细节

本节包含正文中所述实验的更多技术细节。

B.1 测试问题

本文介绍了两类测试问题，$\boldsymbol{f}^*\colon\ [0,1]^2 \to \mathbb{R}^2$。它们通过为每个目标函数分别采样一个高斯过程生成（例如）。第一类中，各目标的长度尺度不同；第二类中，各目标是否存在加性高斯观测噪声有所不同。表 1 展示了两个高斯过程使用的超参数。下方的图 5 和图 6 分别展示了来自第一类和第二类的测试问题示例。

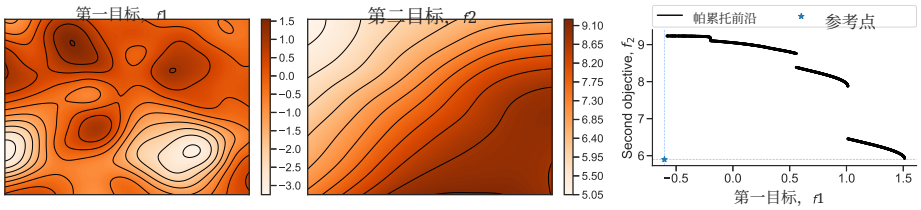
对高斯过程进行采样时，我们不知道贝叶斯优化算法之后会尝试在哪些位置进行采样。为解决这一问题，我们在 $[0,1]^2$ 上使用经过扰乱的二维 Sobol' 序列，选取 100 个点对每个高斯过程进行采样。随后，我们将高斯过程在以这些点为条件化依据后的后验均值用作测试函数。

表 1：用于生成两类测试问题的高斯过程超参数。

	系列 1		系列 2	
	目标 1	目标 2	目标 1	目标 2
核函数	Matérn-$5/2$	Matérn-$5/2$	Matérn-$5/2$	Matérn-$5/2$
各向同性长度尺度	0.2	1.8	0.4	0.4
输出尺度	1	50	1	1
常数均值	0	0	0	0
噪声标准差	0	0	1	0

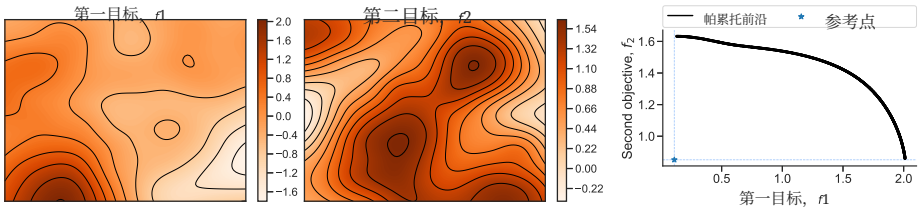
B.2 代理模型

在贝叶斯优化过程中，为了对各目标进行建模，我们为每个目标使用相互独立的高斯过程。我们使用 Matérn-$5/2$ 核函数，并在每一步中，利用最大后验估计，根据迄今为止观测到的数据拟合超参数。在拟合高斯过程之前，会对观测进行标准化，使其均值为零、方差为一。这是处理实际问题时常用的一种技巧，因为先验分布必须根据数据指定。使用高斯过程进行预测时，会应用逆变换。



（a）等高线 (b) 帕累托前沿

图5：第一类测试问题中的一个示例，由高斯过程的样本生成。面板 (a) 展示了通过为高斯过程设置更短的长度尺度，使第一目标更难学习。面板 (b) 展示了该测试问题的帕累托前沿。



(a) 等高线 (b) 帕累托前沿

图6：第二类测试问题中的一个示例，由高斯过程的样本生成。面板 (a) 显示两个目标的长度尺度相同。相反，在采样时加入观测噪声，使第一目标更难学习。面板 (b) 展示了该测试问题的帕累托前沿。

先验均值 高斯过程的先验均值被建模为一个常数函数。贝叶斯优化旨在采样目标函数值较大的点，因此，若根据这个有偏样本拟合先验均值，就会使推断值产生偏差。因此，先验均值作为最大后验估计的一部分，根据最初的六个数据点进行拟合；之后的所有迭代中，该均值都固定为此值，仅拟合其他超参数。

观测噪声 公式（2）中的观测模型包含加性高斯观测噪声。第二类测试问题中，仅对第一目标拟合该观测噪声的方差，因为只有这个目标加入了噪声。对于第一类测试问题以及第二类测试问题中的第二目标，噪声方差固定为可忽略的值 10^{-4}。没有将其设为零，因为这会导致数值不稳定。实验人员通常知道自己的问题是随机的还是确定性的，因此我们认为这是对目标函数进行建模最自然的方式。

先验分布 对于长度尺度、输出尺度和观测噪声方差（在拟合时），使用伽马分布作为先验分布。伽马分布

由形状参数 α 和率参数 β确定。其定义域为 (0， ∞)，概率密度函数为

$$p(z) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} z^{\alpha-1} e^{-\beta z} \tag{28}$$

其中 Γ 是伽马函数。

高斯过程先验均值建模为一个常数函数。对于该常数，采用定义在整个实数轴上的均匀分布作为非正规先验。具体实现时，只需在边际对数似然中不加入高斯过程先验均值的先验分布所产生的贡献。

如正文所述，第一类测试问题所使用的先验分布包含了关于哪个目标函数具有较短长度尺度的信息。相反，第二类测试问题中各目标函数的长度尺度先验相同。图 7 和图 8 展示了用于对这两类测试问题建模的长度尺度先验分布。

所有超参数的先验分布汇总于表 2 和表 3 中。

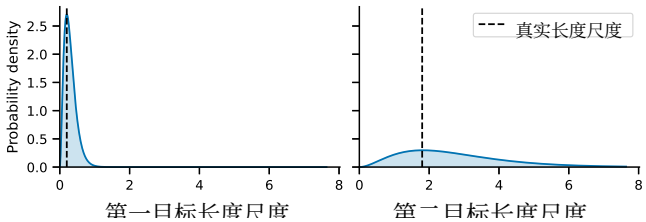


图 7：用于第一类测试问题的代理模型的长度尺度先验分布。这些分布编码了第一目标具有较短长度尺度这一先验信息。

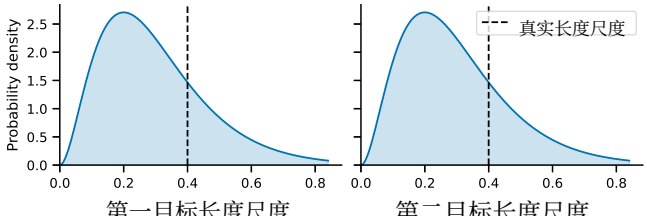


图 8：用于第二类测试问题的代理模型的长度尺度先验分布。两个目标函数使用相同的先验分布。

表2：用于第一类测试问题中标准化数据代理模型超参数的先验分布。

	目标 1	目标 2
核函数	Matérn-$5/2$	Matérn-$5/2$
各向同性长度尺度	伽马分布($\alpha = 3,\ \beta = 10$)	伽马分布($\alpha = 3,\ \beta = 1.1$)
输出尺度	伽马分布($\alpha = 2,\ \beta = 0.15$)	伽马分布($\alpha = 2,\ \beta = 0.15$)
常数均值	无先验	无先验
噪声方差	固定为 10^{-4}	固定为 10^{-4}

表 3：第二类测试问题中标准化数据的代理模型超参数所用的先验分布。

	目标 1	目标 2
核函数	Matérn-$5/2$	Matérn-$5/2$
各向同性长度尺度	伽马分布($\alpha = 3,\ \beta = 10$)	伽马分布($\alpha = 3, \beta = 10$)
输出尺度	伽马分布($\alpha = 2, \beta = 0.15$)	伽马分布($\alpha = 2, \beta = 0.15$)
常数均值	无先验	无先验
噪声方差	伽马分布($\alpha = 1.1,\ \beta = 0.05$)	固定为 10^{-4}

B.3 随机数

在将 C-MOKG 与基准算法进行比较时，我们尽可能确保整个实验过程中的各项条件保持一致。因此，

– 每个系列中的 100 个测试问题均使用相同的算法进行测试；– 为每个测试问题分配一组不同的六个初始点，供两种算法共同使用；– 两种算法均使用相同的点序列，通过蒙特卡洛近似来计算关于 $\boldsymbol{\lambda}$ 的期望。