

# 一种基于自适应代理模型的并行贝叶斯优化方法

吕志明, 王霖青, 赵 珺<sup>†</sup>, 刘 颖

(大连理工大学控制科学与工程学院, 辽宁 大连 116024)

**摘要:** 提出一种基于自适应代理模型的并行贝叶斯优化方法,用于求解计算成本高的复杂优化问题.该方法基于多点期望改进判据,通过批次采样实现并行优化.针对并行优化产生的大量历史数据会导致全局代理模型建模成本高的问题,提出一种改进的基于数据并行的高斯过程建模方法,在线构造局部代理模型.此外,针对多点期望改进判据计算成本高的问题,提出一种启发式的分层优化策略,通过序贯优化基于自适应代理模型的单点期望改进判据,近似计算多点期望改进判据.最后通过 5 个测试问题验证所提出方法的有效性.

**关键词:** 自适应; 代理; 并行; 贝叶斯优化方法; 期望改进

**中图分类号:** TP206

**文献标志码:** A

## A parallel Bayesian optimization method based on adaptive surrogate model

LYU Zhi-ming, WANG Lin-qing, ZHAO Jun<sup>†</sup>, LIU Ying

(School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China)

**Abstract:** A parallel Bayesian optimization method based on the adaptive surrogate model is proposed to solve the computationally expensive problems. Based on the multi-points expected improvement criterion, one can evaluate the batch of points simultaneously. For the problem that the global surrogate model is difficult to be constructed with a large number of historical data from the parallel optimization, an improved data parallel local learning methodology for Gaussian Process modeling is proposed to construct the local surrogate models online. Furthermore, in order to reduce the computational cost of the multi-points expected improvement criterion, a heuristic hierarchical optimization strategy is proposed to calculate the single point expected improvement criterion based on the adaptive surrogate model sequentially. Finally, the effectiveness of the proposed method is verified by 5 test problems.

**Keywords:** adaptive; surrogate; parallel; Bayesian optimization; expectation improvement

## 0 引言

在求解计算成本高的工程优化问题中,贝叶斯优化方法得到了广泛的关注<sup>[1]</sup>.然而,贝叶斯优化方法是一种基于高斯过程的序贯优化方法,其通过优化采样判据进行单点序贯采样.当给定输入,目标函数的输出响应时间很长时,单点序贯优化将不能满足实时优化的要求.因此,基于分布式并行优化的多点采样方法已成为研究热点<sup>[2-3]</sup>.

诸多文献已将贝叶斯优化方法从单点优化扩展到多点并行优化<sup>[4]</sup>.文献[5]提出一种多点期望改进判据用于序贯批次优化;文献[6]推导出多点期望改进判据的解析表达式,但是该表达式求解耗时.因此,文献[7]提出了两种启发式的多点期望改进判据.此

外,其他的多点采样判据也相继提出,如:文献[8]提出了一种多点概率改进判据;文献[9]提出了一种基于高斯过程的上置信边界采样判据;文献[10]提出了一种基于预测熵的采样判据.然而,上述采样判据采用的概率模型均服从高斯分布,当训练样本很多时,存在高斯过程回归模型建模和预测成本高的问题.

为了缓解上述问题,诸多文献采用局部代理模型代替全局代理模型.文献[11]采用基于数据并行的局部学习方法构造局部代理模型;文献[12]通过构造局部代理模型,提出一种混合专家方法;文献[13]通过对搜索空间的自适应分割,提出一种基于自适应分区的多代理模型方法.但是,上述文献只考虑了连续的单点优化方法.此外,文献[14]提出一种基于多代

收稿日期: 2017-10-30; 修回日期: 2017-12-11.

基金项目: 国家自然科学基金项目(61603069, 61473056, 61533005, 61522304, U1560102); 中央高校基本科研业务费专项资金项目(DUT17ZD231).

作者简介: 吕志明(1988—),男,博士生,从事流程工业过程建模与优化、智能优化算法设计的研究;赵珺(1981—),男,教授,博士生导师,从事生产计划与调度、现代集成制造系统、工业生产一体化、优化技术等研究.

<sup>†</sup>通讯作者. E-mail: zhaoj@dlut.edu.cn.

理的Memetic算法用于并行的局部搜索;文献[15]提出一种基于多点采样判据和空间划分策略的并行的局部优化方法.但是在所有的子空间中执行局部优化会降低算法的优化效率.

针对计算成本高的复杂优化问题,本文提出一种基于自适应代理模型的并行贝叶斯优化方法.该方法基于改进的基于数据并行的局部学习方法,在线构造局部代理模型;并根据局部代理模型的局部预测特点,构造一种基于自适应代理模型的单点期望改进判据.此外,提出一种启发式的分层优化策略,通过序贯优化基于自适应代理模型的单点期望改进判据,实现多点期望改进判据的近似优化.最后,通过不同的测试问题对所提出方法进行参数分析和有效性验证.

## 1 基于自适应代理模型的并行贝叶斯优化方法

本文考虑如下的全局优化问题:

$$\min_{\mathbf{x} \in \mathbf{H}} y(\mathbf{x}). \quad (1)$$

其中:  $y(\mathbf{x})$  为目标函数的响应值;  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbf{H} \subset \mathbf{R}^d$  为输入矢量;  $\mathbf{H} = [\mathbf{a}, \mathbf{b}] = \prod_{i=1}^d [a_i, b_i]$  为输入矢量的约束,  $a_i, b_i \in \mathbf{R} : a_i < b_i$ .

假设目标函数  $y$  为计算时间成本高的黑箱函数,为了在有限的计算成本下获得近优解,本文提出一种基于自适应代理模型的并行贝叶斯优化方法,该算法的结构如图1所示.

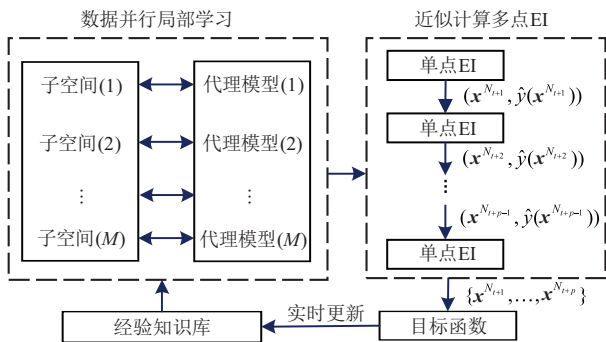


图1 基于自适应代理模型的并行贝叶斯优化方法的结构

### 1.1 基于数据并行的局部学习方法

为了缓解训练样本动态增加导致代理模型构造成本高的问题,本文提出一种基于数据并行的高斯过程回归模型在线建模方法.

#### 1.1.1 改进的GeoCluster聚类算法

为了对输入空间进行分区,首先提出一种改进的GeoCluster(Geometry motivated clustering)聚类方法.

在迭代时刻  $t$ , 根据输入空间中的历史寻优点,构造经验知识库如下:

$$D_t = \{\mathbf{x}_i, y_i | j = 1 : N_t\}, \quad (2)$$

其中  $N_t$  为  $D_t$  中的样本个数.

为了将经验知识库中的样本聚类,且保证每一类具有相等数量的样本点,需要求解如下的非线性规划方程<sup>[16]</sup>:

$$\min_{\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_m} \sum_{i=1}^m (m_i - m_d)^2. \quad (3)$$

其中:  $\mathbf{o}_i$  和  $m_i$  分别为第  $i$  类的聚类中心和样本量,  $m$  为聚类的个数,  $m_d$  为每一类中期望的样本量.

在序贯优化过程中,经验知识库中的样本量是动态增加的.由于  $m_d$  是常数,聚类个数自适应地更新如下:

$$M_t = \text{fix} \left( \frac{N_t}{m_d} \right), \quad (4)$$

其中  $\text{fix}(A)$  表示将  $A$  中的元素向零点方向取整.

基于式(4),改进的聚类中心更新公式描述如下:

$$\mathbf{o}_p(k) = \mathbf{o}_p(k-1) + \alpha \delta \mathbf{v}_p, \quad (5)$$

$$\delta \mathbf{v}_p = \sum_{q=1, q \neq p}^{M_t} \left( \frac{m_q}{m_p} \right) [\mathbf{o}_q(k-1) - \mathbf{o}_p(k-1)], \quad (6)$$

其中  $\alpha$  为学习因子.

#### 1.1.2 构造局部代理模型

已知数据集  $D_t$ , 通过改进的GeoCluster聚类方法,将输入空间划分为  $M_t$  个子空间,其中第  $s$  个子空间中的样本为  $D_s = \{\mathbf{x}_j, y_j | j = 1 : N_s\}$ ,  $\mathbf{x}_j$  为输入变量,  $y_j$  为目标函数响应值,  $N_s$  为第  $s$  个子空间中样本的个数.假设存在含噪声的响应函数  $y(\mathbf{x}_{1:N_s}) = f(\mathbf{x}_{1:N_s}) + \varepsilon$ . 其中:  $f(\mathbf{x}_{1:N_s})$  表示输入变量  $\mathbf{x}_{1:N_s}$  与响应值  $y(\mathbf{x}_{1:N_s})$  之间的映射函数,且服从高斯分布;  $\varepsilon \sim \mathcal{N}(0, \delta_y^2)$  表示零均值高斯噪声.则对应的响应函数可描述为如下的高斯分布:

$$y(\mathbf{x}_{1:N_s}) \sim \mathcal{N}(f(\mathbf{x}_{1:N_s}), \mathbf{K}_s + \delta_y^2 \mathbf{I}), \quad (7)$$

其中  $\mathbf{K}_s = \mathbf{k}(\mathbf{x}_{1:N_s}, \mathbf{x}_{1:N_s})$  为协方差矩阵.基于零均值高斯噪声假设,训练输出  $y_{1:N_s}$  和测试输出  $f(\mathbf{x}_*)$  的联合概率分布<sup>[17]</sup>如下式所示:

$$\begin{bmatrix} y_{1:N_s} \\ f(\mathbf{x}_*) \end{bmatrix} \sim \mathcal{N} \left( 0, \begin{bmatrix} \mathbf{K}_s + \delta_y^2 \mathbf{I} & \mathbf{k}(\mathbf{x}_{1:N_s}, \mathbf{x}_*) \\ \mathbf{k}(\mathbf{x}_*, \mathbf{x}_{1:N_s}) & k(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix} \right). \quad (8)$$

由此可得第  $s$  个基于GP的代理模型的后验预测分布为

$$f(\mathbf{x}_*) | D_s, \mathbf{x}_* \sim \mathcal{N}(\mu_s(\mathbf{x}_* | D_s), \delta_s^2(\mathbf{x}_* | D_s)). \quad (9)$$

其中均值和方差描述如下:

$$\mu_s(\mathbf{x}_* | D_s) = \mathbf{k}(\mathbf{x}_{1:N_s}, \mathbf{x}_*)^T (\mathbf{K}_s + \delta_y^2 \mathbf{I})^{-1} y_{1:N_s}, \quad (10)$$

$$\delta_s^2(\mathbf{x}_* | D_s) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}(\mathbf{x}_{1:N_s}, \mathbf{x}_*)^T (\mathbf{K}_s + \delta_y^2 \mathbf{I})^{-1} \mathbf{k}(\mathbf{x}_{1:N_s}, \mathbf{x}_*). \quad (11)$$

#### 1.2 近似最大化多点期望改进判据

针对多点期望改进判据优化成本高的问题,本文提出一种启发式分层优化策略,近似地最大化多点期望改进判据.

1.2.1 基于自适应代理模型的单点期望改进判据

由式(9)可知,真实函数的局部特性可以通过局部代理模型近似描述,因此,本文提出一种基于自适应代理模型的期望改进判据.

考虑到局部代理模型的选择是关于输入变量的函数,因此,当已知输入变量  $\mathbf{x}'$  时,对应的代理模型选择如下:

$$\gamma = \{\gamma | \gamma = \|\mathbf{x}' - \mathbf{x}\|, \mathbf{x} \in D_t\}; \quad (12)$$

$$\mathbf{x}_{\text{neighbor}} = \operatorname{argmin}(\gamma); \quad (13)$$

$$s = c(\mathbf{x}_{\text{neighbor}}), c(\cdot) \in \{1, 2, \dots, M_t\}; \quad (14)$$

$$\hat{f}(\mathbf{x}') = \mu(\mathbf{x}' | D_s). \quad (15)$$

其中:  $\gamma$  表示输入变量  $\mathbf{x}'$  与所有样本点的欧氏距离;  $\mathbf{x}_{\text{neighbor}}$  表示与输入变量  $\mathbf{x}'$  最近的样本点;  $c(\cdot)$  表示以  $\mathbf{x}'$  为自变量的选择函数;  $s$  表示代理模型的索引号;  $\hat{f}(\mathbf{x}')$  表示输入变量  $\mathbf{x}'$  的预测值. 因此,基于自适应代理模型的期望改进判据描述如下:

$$\text{EI}(\mathbf{x}) = \begin{cases} [y_{\min} - \mu_s(\mathbf{x})]\Phi(Z) + \delta_s(\mathbf{x})\Psi(Z), \\ \delta_s(\mathbf{x}) > 0; \\ 0, \delta_s(\mathbf{x}) = 0. \end{cases} \quad (16)$$

$$Z = \frac{y_{\min} - \mu_s(\mathbf{x})}{\delta_s(\mathbf{x})}. \quad (17)$$

$$y_{\min} = \min\{y_1, \dots, y_{N_t}\}. \quad (18)$$

其中:  $\mu_s(\mathbf{x})$  和  $\delta_s(\mathbf{x})$  分别为第  $s$  个代理模型的预测均值和方差,  $\Phi$  和  $\Psi$  分别为累积密度函数和概率密度函数.

1.2.2 启发式分层优化策略

为了近似计算多点期望改进判据,本文提出一种启发式分层优化策略.

由文献[7]可知,多点期望改进判据为单批次采样  $p$  个点所获得回报的条件期望,具体描述如下:

$$\begin{aligned} \overline{\text{EI}}(\mathbf{X}^{\text{new}}) &= \text{E}[\max\{(y_{\min} - y(\mathbf{x}_{N_{t+1}}))^+, \\ &\quad \dots, (y_{\min} - y(\mathbf{x}_{N_{t+p}}))^+\} | D_t] = \\ &= \text{E}[(y_{\min} - \min(y(\mathbf{x}_{N_{t+1}}), \\ &\quad \dots, y(\mathbf{x}_{N_{t+p}})))^+ | D_t], \end{aligned} \quad (19)$$

$$\mathbf{X}^{\text{new}} = \{\mathbf{x}_{N_{t+1}}, \dots, \mathbf{x}_{N_{t+p}}\}, \quad (20)$$

其中  $\mathbf{X}^{\text{new}}$  为单批次采样得到的  $p$  个查询点.

为获取  $\mathbf{X}^{\text{new}}$  个查询点,需要求解如下优化问题:

$$\mathbf{X}^{\text{new}} = \operatorname{argmax}_{\mathbf{x} \in D_t^p} \overline{\text{EI}}(\mathbf{x}), \quad (21)$$

其中  $D_t^p$  表示在  $t$  次迭代时获取  $p$  个查询点后的经验知识库.

然而,式(21)的计算复杂度随着  $p$  的增加而增加. 为了避免直接计算式(21),采用启发式策略,则改进的多点期望改进判据表示如下:

$$\begin{aligned} \widetilde{\text{EI}}(\mathbf{X}^{\text{new}}) &= \\ &= \text{E}[(y_{\min} - \min(\hat{f}(\mathbf{x}_{N_{t+1}}), \dots, \end{aligned}$$

$$\begin{aligned} &\hat{f}(\mathbf{x}_{N_{t+p}}))]^+ | D_t, y(\mathbf{x}_{N_{t+1}}) = \\ &y(\mathbf{x}_{N_{t+1}}), \dots, y(\mathbf{x}_{N_{t+p}}) = \hat{f}(\mathbf{x}_{N_{t+p}})], \end{aligned} \quad (22)$$

其中  $\hat{f}(\mathbf{x}_{N_{t+1}}), \dots, \hat{f}(\mathbf{x}_{N_{t+p}})$  分别表示  $\mathbf{X}^{\text{new}}$  的预测值.

基于式(22),通过序贯最大化式(16),可以实现多点期望改进判据的近似计算,具体描述如下:

$$\begin{aligned} \mathbf{x}_{N_{t+j}} &= \operatorname{argmax}_{\mathbf{x}} \widetilde{\text{EI}}[D_t, y(\mathbf{x}_{N_{t+1}}) = \hat{f}(\mathbf{x}_{N_{t+1}}), \\ &\quad \dots, y(\mathbf{x}_{N_{t+j-1}}) = \hat{f}(\mathbf{x}_{N_{t+j-1}})](\mathbf{x}), \end{aligned} \quad (23)$$

其中  $j \in [1, p]$ .

在序贯优化过程中,为了得到准确的查询点,本文提出一种分层优化策略. 首先,在整个输入空间构造如下的全局优化问题:

$$\mathbf{x}_{N_{t+j}}^{\text{global}} = \operatorname{argmax}_{\mathbf{x} \in H} \widetilde{\text{EI}}(\mathbf{x}), \quad (24)$$

其中  $\mathbf{x}_{N_{t+j}}^{\text{global}}$  为全局最优查询点. 然后,在第  $s$  个子空间中,构造如下局部优化问题:

$$\mathbf{x}_{N_{t+j}}^{\text{local}} = \operatorname{argmax}_{\mathbf{x} \in \text{SH}} \text{EI}(\mathbf{x}), \quad (25)$$

$$\text{SH} = [\min \text{SH}^{1:d}, \max \text{SH}^{1:d}], \quad (26)$$

$$\min \text{SH}^d = \min\{x_i^d | i = 1, \dots, N_s, x_i^d \in D_s\}, \quad (27)$$

$$\max \text{SH}^d = \max\{x_i^d | i = 1, \dots, N_s, x_i^d \in D_s\}. \quad (28)$$

其中:  $\mathbf{x}_{N_{t+j}}^{\text{local}}$  为局部最优查询点;  $\text{SH}$  为第  $s$  个子空间的边界;  $\min \text{SH}^d$  和  $\max \text{SH}^d$  分别为子空间中输入变量第  $d$  维的最小值和最大值.

基于分层优化策略,得到的第  $j$  个查询点为

$$\mathbf{x}_{N_{t+j}} = \begin{cases} \mathbf{x}_{N_{t+j}}^{\text{global}}, & \text{EI}(\mathbf{x}_{N_{t+j}}^{\text{global}}) \geq \text{EI}(\mathbf{x}_{N_{t+j}}^{\text{local}}); \\ \mathbf{x}_{N_{t+j}}^{\text{local}}, & \text{EI}(\mathbf{x}_{N_{t+j}}^{\text{global}}) < \text{EI}(\mathbf{x}_{N_{t+j}}^{\text{local}}). \end{cases} \quad (29)$$

综上所述,通过对多点期望改进判据的近似计算,可以得到一组新的查询点  $\mathbf{X}^{\text{new}}$ ; 然后,将该组查询点分发给不同的计算单元,通过并行计算得到真实函数值  $y(\mathbf{X}^{\text{new}})$ ; 最后,得到全局最优解

$$\mathbf{x}^{\text{best}} = \operatorname{argmin}([y_{1:N_t}; y(\mathbf{X}^{\text{new}})]), j \in [1, p]. \quad (30)$$

**算法1** 基于自适应代理模型的并行贝叶斯优化方法.

Step 1: 初始化最大迭代次数  $T$ , 子空间中期望的数据点个数, 初始化设计点的个数和经验知识库. 令  $t = 1$ .

Step 2: 通过拉丁立方采样产生初始设计点, 并计算对应的函数值.

Step 3: 根据式(2)构建经验知识库  $D_t$ .

Step 4: 根据式(4)更新聚类个数.

Step 5: 根据式(5)和(6)对经验知识库中的样本聚类.

Step 6: 根据式(10)和(11)构造局部代理模型.

Step 7: 根据式(16)~(18)构造基于自适应代理

模型的单点期望改进判据.

Step 8: 近似计算多点期望改进判据. 令  $j = 1$ .

Step 8.1: 根据式(24)~(29)最大化单点期望改进判据.

Step 8.2: 计算第  $j$  个查询点对应的预测值, 并更新代理模型.

Step 8.3: 判断  $j \geq p$  是否为假. 如果为假, 则返回 Step 1; 否则转到 Step 7.

Step 9: 得到查询点  $\mathbf{X}^{\text{new}}$ , 并计算其对应的真实函数值  $y(\mathbf{X}^{\text{new}})$ .

Step 10: 判断  $t \geq T$  是否为真. 如果为真, 则算法停止; 否则返回 Step 3.

## 2 实验分析

为了验证本文方法的有效性, 下面将在 2.1 节中研究单批次采样次数选择对本文方法性能的影响; 在 2.2 节中研究子空间大小选择对本文方法性能的影响; 在 2.3 节中将本文方法分别与 KB(Kriging believer) 和 KCL(Kriging constant liar) 两种启发式方法<sup>[7]</sup>进行实验比较.

在分析本文方法的性能过程中, 采用 5 个标准测试问题<sup>[5]</sup>, 如表 1 所示. 并由拉丁立方采样方法产生代理模型的初始设计点. 为了实验对比的公平性, 针对同一测试问题, 均采用相同的初始设计点且独立运行 30 次. 此外, 实验设计需要满足如下终止条件:

1) 当算法寻优结果与全局最优解满足如下条件

表 1 测试问题

问题	维数	搜索空间	全局最优解
Branin	2	$[-5.10, 0.15]^2$	0.3979
Goldstein-Price	2	$[-2, 2]^2$	-3.0
Hartman3	3	$[-2.048, 2.048]^3$	0
Hartman6	6	$[0, 1]^6$	-3.135
Shekel	4	$[0, 1]^4$	-3.322

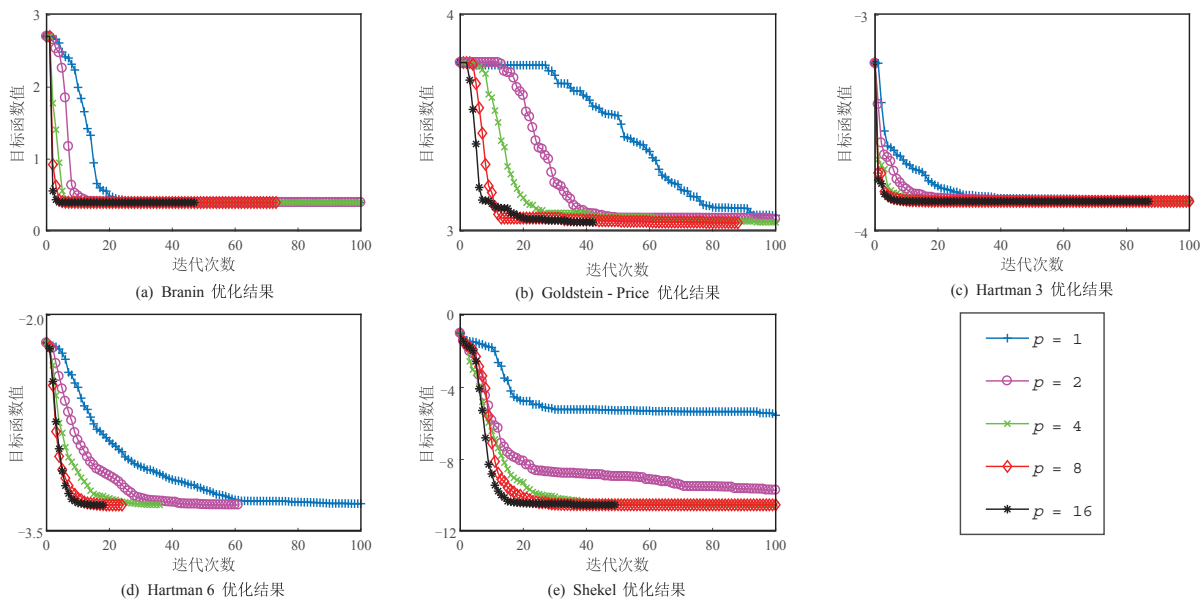


图 2 单批次采样不同次数时, 针对不同测试问题的优化结果

时, 算法终止:

$$\min(\|\mathbf{x}^{\text{best}} - \mathbf{x}^{\text{global}}\|) \leq 0.01\sqrt{d}. \quad (31)$$

其中:  $\mathbf{x}^{\text{best}}$  为算法搜索到的最优解,  $\mathbf{x}^{\text{global}}$  为测试问题真实的全局最优解.

2) 当算法迭代次数超过其最大值时, 算法终止.

### 2.1 单批次采样次数选择

本小节将研究单批次采样次数选择对并行优化收敛速度和优化效率的影响.

首先, 并行优化的收敛速度和优化效率评价指标定义如下:

$$A = \text{fix}\left(\frac{1}{\text{EN}} \sum_{i=1}^{\text{EN}} t_i\right), t_i < T; \quad (32)$$

$$B = \frac{\text{SEN}}{\text{EN}} \times 100\%. \quad (33)$$

其中:  $A$  为并行优化的收敛速度,  $\text{EN}$  为总的实验次数,  $T$  为最大迭代次数,  $t_i$  表示在满足终止条件时第  $i$  次实验的迭代次数,  $B$  为并行优化的效率,  $\text{SEN}$  表示同时满足两个终止条件的实验次数.  $A$  越小, 算法的平均迭代次数越少, 算法的收敛速度越快;  $B$  越大, 算法的成功率越高, 说明算法稳定性越好, 效率越高.

实验参数设置如下: 初始设计点的个数为  $N_0 = 10 \times d$ , 子空间中数据点的个数为  $N_s = 60$ , 单批次采样次数为  $p = 2^\eta, \eta = 0, 1, 2, 3, 4$ .

由表 2 可知: 对于 5 个测试问题, 随着  $p$  值增大, 对应的  $A$  减小, 说明  $p$  值增大可加快并行优化收敛的速度;  $p$  值越大, 对应的  $B$  越大, 说明  $p$  值增大有利于算法收敛到全局最优解. 为了更加直观地分析对比结果, 图 2 给出了针对 5 个测试问题, 单批次采样不同次数下的算法收敛曲线. 从图 2 中可以看出,  $p$  值越大, 在满足终止条件的情况下, 算法的迭代次数越少, 而且算法的稳定性越好.

表2 单批次采样不同次数的统计结果

问题	$p = 1$		$p = 2$		$p = 4$		$p = 8$		$p = 16$	
	A	B	A	B	A	B	A	B	A	B
Branin	80	50	60	66	49	83	21	100	16	100
Goldstein-Price	79	63	53	73	37	93	26	100	17	100
Hartman3	72	80	41	100	25	100	16	100	12	100
Hartman6	91	33	70	76	40	90	28	96	19	100
Shekel	100	3	74	66	44	93	31	96	21	100

2.2 子空间大小选择

本小节针对Hartman6和Shekel测试问题,研究不同子空间大小选择对并行优化速度和优化精度的影响。

实验参数设置如下:初始设计点的个数为 $N_0 = 10 \times d$ ;子空间中样本点的个数为 $N_s = 20, 40, 60, 80, 100$ ;单批次采样次数为 $p = 2^\eta; \eta = 2, 3, 4$ 。

由表3可知:对于Hartman6问题,当SN=20时,平均迭代次数最多,优化成功率最小,说明SN越小,代理模型丢失的信息越多,算法的性能越差;当SN逐渐增加时,优化成功率保持在100%,平均迭代次数逐渐减少并保持不变,原因是算法在满足终止条件后,产生的历史样本点不足以构造多个局部代理模型,因此,子空间大小的变化不会改变算法的收敛速度和效率。

表3 Hartman6问题的子空间选择不同大小时的实验结果

$p$	SN=20		SN=40		SN=60		SN=80		SN=100	
	A	B	A	B	A	B	A	B	A	B
4	39	90	27	100	25	100	19	100	16	100
8	41	80	23	100	14	100	12	100	12	100
16	41	73	11	100	12	100	10	100	10	100

由表4可知:对于Shekel问题,当 $p = 4$ 时,算法的并行优化收敛速度最慢且优化效率最低;当SN=20时,由于代理模型丢失的信息最多,算法的性能最差;当SN=60时,算法的收敛速度和优化效率基本保持不变;虽然SN=80,100时,算法的收敛速度和优化效率基本保持不变,但是SN越大,代理模型的构造成本将越高。

表4 Shekel问题的子空间选择不同大小的实验结果

$p$	SN=20		SN=40		SN=60		SN=80		SN=100	
	A	B	A	B	A	B	A	B	A	B
4	77	50	57	93	58	93	60	90	60	90
8	68	73	52	83	33	100	33	100	30	100
16	54	86	43	93	21	100	21	100	20	100

综上所述,由表3和表4中子空间选择不同大小的实验结果可知:SN越小,代理模型丢失的信息越多;随着SN的增加,平均迭代次数和成功率也增加,

说明随着SN的增加,算法的收敛速度逐渐提高,优化效率也逐渐提升.SN的建议取值为60。

2.3 与其他方法比较

为了验证本文方法的有效性,本小节针对5个测试问题,将本文方法分别与KB、KCL两种方法进行比较。

本文方法与KB和KCL均采用启发式策略,但是KB和KCL采用全局代理模型,而本文方法采用自适应局部代理模型和分层优化策略优化多点期望改进判断.实验参数设置如下:初始设计点的个数为 $N_0 = 10 \times d$ ,子空间中数据点的个数为 $N_s = 60$ ,单批次采样次数为 $p = 2^\eta, \eta = 1, 2, 3, 4$ 。

由表5和图3可知,对于Branin问题,本文方法的成功率高于KB,但是小于KCL,而且平均迭代次数要少于KB,但是与KCL相接近.说明本文方法的优化效率和收敛速度优于KB,且与KCL相接近。

表5 不同采样条件下对Branin问题的优化结果

方法	$p = 2$		$p = 4$		$p = 8$		$p = 16$	
	A	B	A	B	A	B	A	B
KB	69	56	39	83	34	86	31	83
KCL	60	76	27	100	22	100	12	100
本文方法	59	66	40	83	21	100	16	100

由表6和图4可知,对于Goldstein-Price问题,本文方法的成功率低于KB和KCL,且平均迭代次数均多于KB和KCL.说明本文方法的优化效率和收敛速度差于KB和KCL。

表6 不同采样条件下对Goldstein-Price问题的优化结果

方法	$p = 2$		$p = 4$		$p = 8$		$p = 16$	
	A	B	A	B	A	B	A	B
KB	56	86	32	96	17	100	11	100
KCL	59	90	29	100	21	100	16	100
本文方法	53	83	37	93	26	100	17	100

由表7和图5可知:对于Hartman3问题,当 $p = 2$ 时,本文方法的性能优于KB和KCL;当 $p = 4, 8, 16$ 时,本文方法的性能与KCL相近,但要优于KB。

表7 不同采样条件下对Hartman3问题的优化结果

方法	$p = 2$		$p = 4$		$p = 8$		$p = 16$	
	A	B	A	B	A	B	A	B
KB	88	30	63	70	43	80	31	93
KCL	85	40	62	90	29	96	16	100
本文方法	70	76	40	90	28	96	19	100

由表8和图6可知,对于Hartman6问题,本文方法的性能与KB和KCL相近。

由表9和图7可知,对于Shekel问题,本文方法的

成功率高于KB和KCL,而且平均迭代次数少于KB和KCL.说明本文方法的优化效率和收敛速度优于KB和KCL.

表8 不同采样条件下对Hartman6问题的优化结果

方法	$p=2$		$p=4$		$p=8$		$p=16$	
	A	B	A	B	A	B	A	B
KB	41	100	37	100	14	100	10	100
KCL	43	100	23	100	14	100	9	100
本文方法	41	100	25	100	16	100	12	100

表9 不同采样条件下对Shekel问题的优化结果

方法	$p=2$		$p=4$		$p=8$		$p=16$	
	A	B	A	B	A	B	A	B
KB	98	10	86	33	70	50	55	63
KCL	92	30	83	46	63	76	35	93
本文方法	74	66	58	93	31	96	21	100

总之,本文方法采用自适应局部代理模型代替全局代理模型减少了计算成本,同时采用分层优化策略获得了与KCL相近且优于KB的性能.

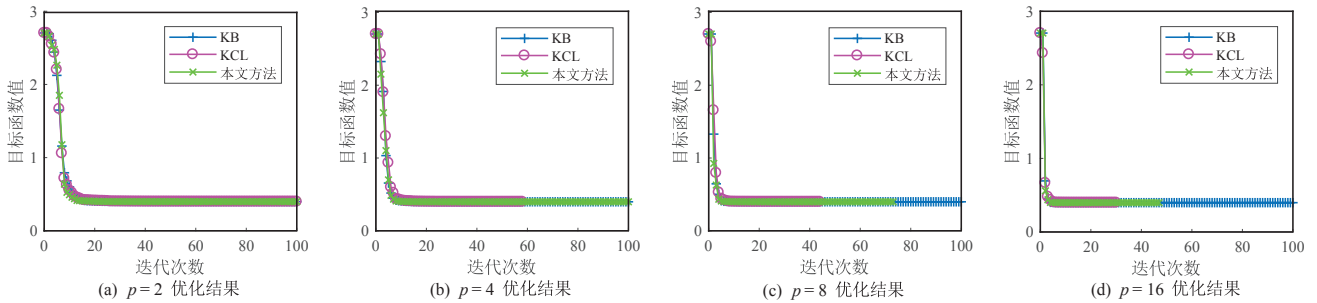


图3 针对Branin问题,当单批次采样不同次数时3种方法的对比结果

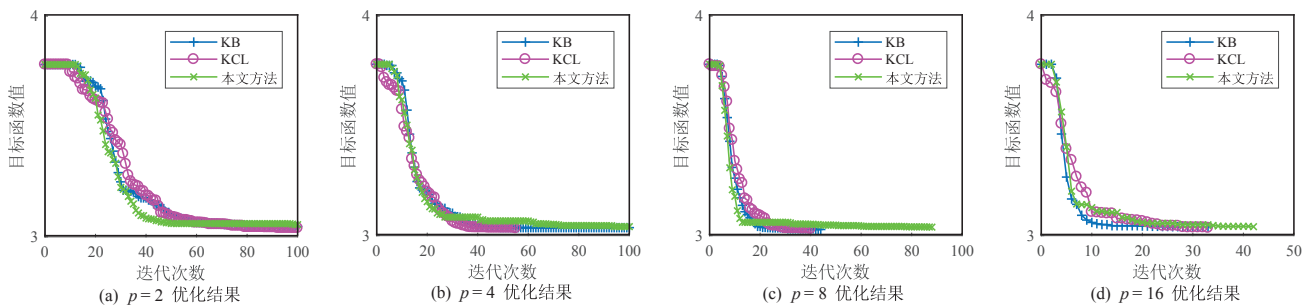


图4 针对Goldstein-Price问题,当单批次采样不同次数时3种方法的对比结果

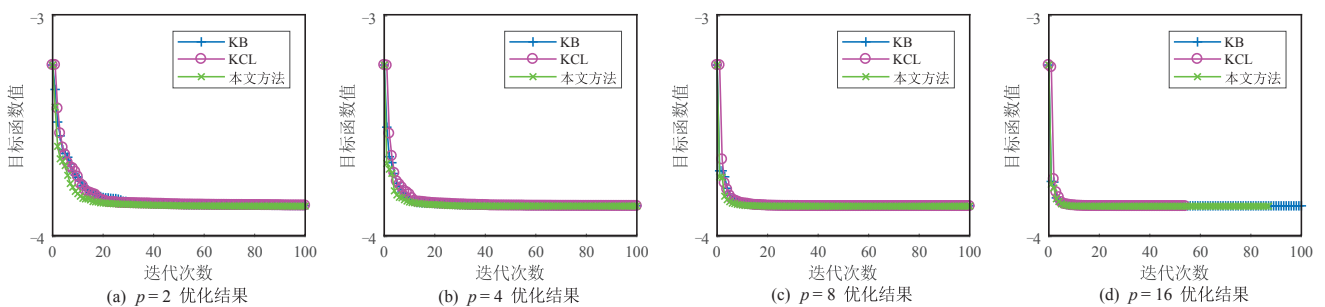


图5 针对Hartman3问题,当单批次采样不同次数时3种方法的对比结果

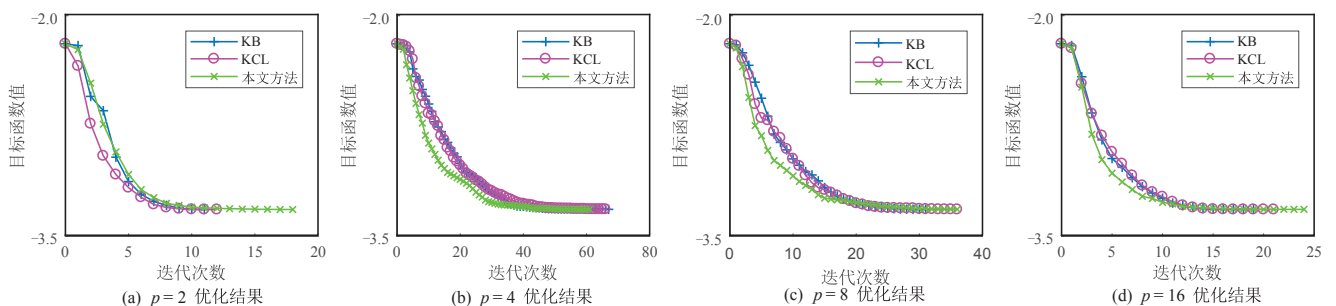


图6 针对Hartman6问题,当单批次采样不同次数时3种方法的对比结果

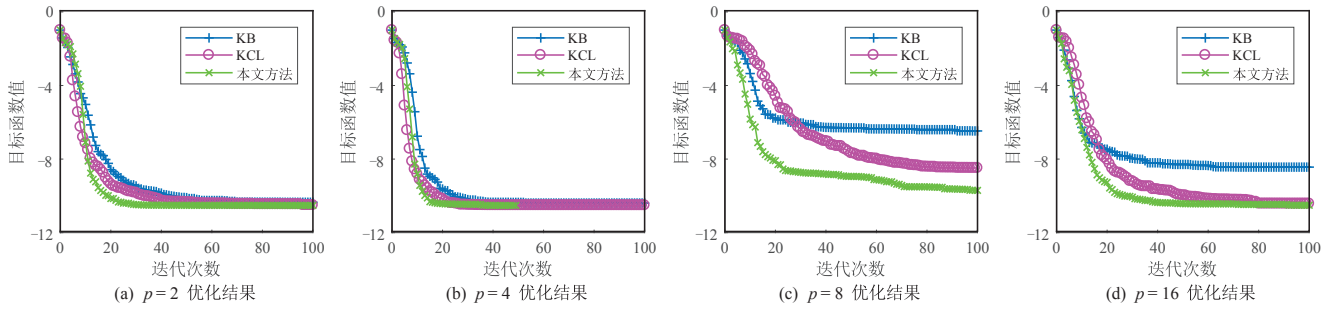


图7 针对Shekel问题,当单批次采样不同次数时3种方法的对比结果

### 3 结论

为了求解计算成本高的复杂优化问题,本文提出了一种基于自适应代理模型的并行贝叶斯优化方法.通过采用改进的数据并行局部学习方法,实现了代理模型个数的动态调整,减少了代理模型的建模成本.此外,通过基于启发式的分层优化策略对多点期望改进判据的近似计算,实现了批次采样和多点并行优化.最后,针对5个测试问题进行了对比实验,实验结果表明,本文方法的性能与KCL相近但优于KB,而且本文方法的计算复杂度要低于KCL和KB.

#### 参考文献(References)

- [1] Shahriari B, Swersky K, Wang Z, et al. Taking the human out of the loop: A review of bayesian optimization[J]. Proc of the IEEE, 2016, 104(1): 148-175.
- [2] Marmin S, Chevalier C, Ginsbourger D. Differentiating the multipoint expected improvement for optimal batch design[C]. Int Workshop on Machine Learning, Optimization and Big Data. Italy: Taormina, 2015: 37-48.
- [3] González J, Dai Z, Hennig P, et al. Batch bayesian optimization via local penalization[C]. Artificial Intelligence and Statistics. Cadiz, 2016: 648-657.
- [4] Haftka R T, Villanueva D, Chaudhuri A. Parallel surrogate-assisted global optimization with expensive functions — A survey[J]. Structural and Multidisciplinary Optimization, 2016, 54(1): 3-13.
- [5] Schonlau M. Computer experiments and global optimization[D]. Waterloo: Department of Statistics and Actuarial Science, University of Waterloo, 1997: 67-79.
- [6] Chevalier C, Ginsbourger D. Fast computation of the multi-points expected improvement with applications in batch selection[C]. Int Conf on Learning and Intelligent Optimization. Heideberg: Springer, 2013: 59-69.
- [7] Ginsbourger D, Riche R L, Carraro L. Kriging is well-suited to parallelize optimization[C]. Computational Intelligence in Expensive Optimization Problems. Berlin: Springer, 2010: 131-162.
- [8] Viana F A C, Haftka R T. Surrogate-based optimization with parallel simulations using the probability of improvement[C]. The 13th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conf. USA: Fort Worth, 2010: 13-15.
- [9] Desautels T, Krause A, Burdick J W. Parallelizing exploration-exploitation tradeoffs in gaussian process bandit optimization[J]. J of Machine Learning Research, 2014, 15(1): 3873-3923.
- [10] Shah A, Ghahramani Z. Parallel predictive entropy search for batch global optimization of expensive objective functions[C]. Int Conf on Neural Information Processing Systems. Montréal, 2015, 118(19/20): 3330-3338.
- [11] Zhou Z, Ong Y S, Nair P B, et al. Combining global and local surrogate models to accelerate evolutionary optimization[J]. IEEE Trans on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2007, 37(1): 66-76.
- [12] Liem R P, Mader C A, Martins J R R A. Surrogate models and mixtures of experts in aerodynamic performance prediction for aircraft mission analysis[J]. Aerospace Science & Technology, 2015, 43(8): 126-151.
- [13] 孙晓燕, 巩敦卫. 自适应分区多代理模型交互式遗传算法[J]. 控制与决策, 2009, 24(2): 170-175. (Sun X Y, Gong D W. Interactive genetic algorithm with multiple surrogate models based on adaptive space division[J]. Control and Decision, 2009, 24(2): 170-175.)
- [14] Zhou Z, Ong Y S, Lim M H, et al. Memetic algorithm using multi-surrogates for computationally expensive optimization problems[J]. Soft Computing — A Fusion of Foundations, Methodologies and Applications, 2007, 11(10): 957-971.
- [15] Li Z, Ruan S, Gu J, et al. Investigation on parallel algorithms in efficient global optimization based on multiple points infill criterion and domain decomposition[J]. Structural & Multidisciplinary Optimization, 2016, 54(4): 1-27.
- [16] Choudhury A, Nair P B, Keane A J. A data parallel approach for large-scale gaussian process modeling[C]. Siam Int Conf on Data Mining. Arlington, 2002: 95-111.
- [17] Petelin D. Gaussian processes for machine learning[J]. Int J of Neural Systems, 2006, 14(6): 3011-3015.

(责任编辑:李君玲)