

基于一般二阶混合矩的高斯分布估计算法

任志刚¹ 梁永胜¹ 张爱民¹ 庞蓓¹

摘要 针对传统高斯分布估计算法 (Gaussian estimation of distribution algorithms, GEDAs) 中变量方差减小速度快、概率密度椭球体 (Probability density ellipsoid, PDE) 的长轴与目标函数的改进方向相垂直, 从而导致算法搜索效率低、容易早熟收敛这一问题, 提出一种基于一般二阶混合矩的高斯分布估计算法. 该算法利用加权的优秀样本预估高斯均值, 并根据沿目标函数的改进方向偏移后的均值来估计协方差矩阵. 理论和数值分析表明, 这一简单操作可以在不增大算法计算量的前提下自适应地调整概率密度椭球体的位置、大小和长轴方向, 提高算法的搜索效率. 在 14 个标准函数上对所提算法进行了测试, 由统计出的 Cohen's d 效应量指标可知该算法的全局寻优能力强于传统高斯分布估计算法; 与当前先进的粒子群算法、差分进化算法相比, 所提算法可以在相同的函数评价次数内获得 9 个函数的显著优解.

关键词 高斯分布估计算法, 概率密度椭球体, 早熟收敛, 协方差矩阵

引用格式 任志刚, 梁永胜, 张爱民, 庞蓓. 基于一般二阶混合矩的高斯分布估计算法. 自动化学报, 2018, 44(4): 635–645

DOI 10.16383/j.aas.2017.c160739

A Gaussian Estimation of Distribution Algorithm Using General Second-order Mixed Moment

REN Zhi-Gang¹ LIANG Yong-Sheng¹ ZHANG Ai-Min¹ PANG Bei¹

Abstract Traditional Gaussian estimation of distribution algorithms (GEDAs) are confronted with issues that variable variances decrease fast and the long axis of the probability density ellipsoid (PDE) tends to be perpendicular to the improvement direction of the objective function, leading to reduction of search efficiency of GEDA and premature convergence. To alleviate these issues, this paper proposes a novel GEDA based on the general second-order mixed moment. In each iteration, this algorithm first estimates an initial mean using weighted excellent samples, then shifts the mean along the improvement direction of the objective function and estimates the covariance matrix with the new mean as the center. Theoretical and numerical analysis results both show that this simple operation can adaptively adjust the location, size and long axis direction of PDE without increasing computation. As a consequence, the search efficiency of the algorithm is improved. Experiments are conducted on 14 benchmark functions. The resultant Cohen's d effect size demonstrates that the proposed algorithm possesses stronger global optimization ability than traditional GEDA. And compared with some state-of-the-art particle swarm optimization and differential evolution algorithms, it can produce significantly superior solutions for 9 functions within the same function evaluation times.

Key words Gaussian estimation of distribution algorithm (GEDA), probability density ellipsoid (PDE), premature convergence, covariance matrix

Citation Ren Zhi-Gang, Liang Yong-Sheng, Zhang Ai-Min, Pang Bei. A Gaussian estimation of distribution algorithm using general second-order mixed moment. *Acta Automatica Sinica*, 2018, 44(4): 635–645

分布估计算法 (Estimation of distribution algorithm, EDA) 是一类典型的基于模型的进化算法. 与基于交叉和变异等遗传操作的其他进化算法相比, EDA 具有较强的理论基础, 而且可以成功求解大量不同类型的连续和离散优化问题, 近年来一直是进化计算领域的研究热点^[1–3].

本文重点研究面向连续问题的 EDA, 这类算法通常采用高斯概率模型 (Gaussian probability model, GPM) 描述优解的分布. 根据对各变量之间相关关系的处理方式, 高斯 EDA (Gaussian EDA, GEDA) 可以分为变量无关的 EDA、部分变量相关的 EDA、全变量相关的 EDA, 其代表性算法分别为 PBIL_c (Population based incremental learning)^[1] 和 UMDA_c (Univariate marginal distribution algorithm)^[1, 4]、EGNA (Estimation of Gaussian networks algorithm)^[4–5]、EMNA_g (Estimation of multivariate normal density algorithm)^[1].

GPM 为 EDA 的理论分析提供了诸多方便, 但同时也有一些不足, 其中最显著的一点是, 直接由常用的极大似然估计 (Maximum likelihood estimation, MLE) 计算出的变量方差会快速减小, 导致算

收稿日期 2016-10-27 录用日期 2017-03-30
Manuscript received October 27, 2016; accepted March 30, 2017
国家自然科学基金 (61105126), 中国博士后科学基金 (2014M560784, 2016T90922) 资助
Supported by National Natural Science Foundation of China (61105126), Postdoctoral Science Foundation of China (2014M560784, 2016T90922)

本文责任编辑 刘艳军
Recommended by Associate Editor LIU Yan-Jun
1. 西安交通大学电子与信息工程学院自动化科学与技术系 西安 710049

1. Department of Automation Science and Technology, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049

法的探索能力急剧下降. 已有研究通过保持各方差值至少为 1^[6]、将估计出的方差固定地放大一定倍数^[7]、利用协方差矩阵 (Covariance matrix, CM) 的特征值修改方差^[8] 以及根据采样解的改进情况自适应地调整方差^[9-10] 等策略来弥补这一不足. 其中, Bosman 等学者提出的自适应方差缩放策略影响最为广泛^[10]. 他们建议当算法在远离均值的位置找到更优解时增大方差, 而当算法在连续多次迭代内找不到更优解时减小方差. 除方差快速减小这一不利因素之外, 与 CM 相对应的概率密度椭球体 (Probability density ellipsoid, PDE) 的长轴还倾向于与目标函数的改进方向相垂直, 这会大大降低 GEDA 的搜索效率^[11-12]. Cai 等学者首次发现了这一现象, 并将其与方差收缩问题合并处理, 提出了基于概率分布交叉熵的自适应方差缩放策略^[11]. Bosman 等学者则提出了预期均值偏移策略, 该策略尝试采用两组分别以预估均值和偏移均值为中心的较优解来估计 CM, 从而改变 PDE 长轴的方向^[12]. 将该策略与前期提出的自适应方差缩放策略相结合后, Bosman 等学者构建了一种称为 AMaLGaM 的有效 EDA^[13].

除上述研究之外, 文献 [14-16] 分别通过引入混沌变异算子、正则化技术、多群体-多模型方法来增强 GPM 对不同问题的适应性; 文献 [17-19] 则放弃 GPM, 分别采用直方图模型、粒子滤波、Copula 函数来估计优解的概率分布; 文献 [20] 在直方图模型的基础上, 进一步利用局部搜索技术提升 EDA 的优化性能; 文献 [21] 则尝试采用有监督学习方法估计优解的条件概率分布, 并借助 Gibbs 采样技术提高搜索效率. 这些研究工作在改进 EDA 性能的同时, 也将算法模型复杂化, 并引入了较多难以设置的自由参数.

本文在分析传统 GEDA 性能弱化原因的基础上, 提出一种简单高效的 GEDA. 该算法在每次迭代中首先根据选择出的优秀样本预估出一加权均值, 然后显式地利用预估均值的目标函数值将其偏移至一个更有希望的解区域中, 最后根据所选优秀样本关于偏移后均值的二阶混合矩来估计 CM. 这一简单操作可以在不增大算法计算量的前提下, 自适应地调整 PDE 的位置、大小和长轴方向, 使之尽可能与当前解区域的结构特征相契合, 从而提高算法的搜索效率. 根据 CM 的计算方式, 本文将所提算法命名为基于一般二阶混合矩的高斯分布估计算法 (General-second-order-mixed-moment based GEDA, GSM-GEDA).

1 基本分布估计算法

作为一种基于模型的进化算法, EDA 假设待解决问题的优解服从某种概率分布, 并利用根据当前群体中的较优解估计出的概率分布来产生下一代群体, 从而驱动算法进化. 基本 EDA 的步骤如下:

- 步骤 1. 设置算法参数, 初始化群体.
 - 步骤 2. 根据目标函数评价当前群体中各个解的质量.
 - 步骤 3. 根据选择规则选出优秀样本集合.
 - 步骤 4. 根据优秀样本集估计概率分布模型.
 - 步骤 5. 根据概率分布模型进行采样, 构建新群体.
 - 步骤 6. 更新获得的最优解并判断是否满足终止条件. 若满足, 则输出最优结果; 否则, 转至步骤 2.
- 连续型 EDA 通常采用 GPM 描述优解的分布. 对于 n 维的随机列向量 \boldsymbol{x} , GPM 的联合概率密度函数可以表示为

$$g(\boldsymbol{x}|\boldsymbol{\mu}, C) = \frac{(2\pi)^{-\frac{n}{2}}}{(\det C)^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T C^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} \quad (1)$$

其中, $\boldsymbol{\mu}$ 和 C 分别表示 \boldsymbol{x} 的均值和 CM. 在每次迭代中, GEDA 通常根据截断选择规则选出优秀样本集 S , 并采用 MLE 来估计 $\boldsymbol{\mu}$ 和 C :

$$\bar{\boldsymbol{\mu}} = \frac{1}{|S|} \sum_{i=1}^{|S|} S_i \quad (2)$$

$$\bar{C} = \frac{1}{|S|} \sum_{i=1}^{|S|} (S_i - \bar{\boldsymbol{\mu}})(S_i - \bar{\boldsymbol{\mu}})^T \quad (3)$$

EMNA_g 算法采用式 (1)~(3) 所示的 GPM 及参数估计方法来综合描述所有变量之间的相关关系. 文献 [22] 的研究表明, 诸如 EGNA 等基于高斯图的、部分变量相关的 EDA 可以纳入到全变量相关的 EDA 框架之下. UMDA_c 和 PBIL_c 则通过忽略各变量间的相关关系, 换取了具有较少参数的对角型 CM.

上述三类 GEDA 针对各类型问题的求解性能虽有不同, 但它们都采用 MLE 估计概率分布参数, 由此具有两个共性特点: 1) 各变量的方差会随着算法迭代而快速减小^[6, 7, 9-11]; 2) 与 CM 相对应的 PDE 的长轴倾向于与目标函数的改进方向相垂直^[11-13]. 图 1 给出了该现象的示意图, 造成该现象的主要原因是, GEDA 在每次迭代中选出的优秀样本主要分布在由目标函数等值线切割原 PDE 所形成的半椭球体内. 该半椭球体的长轴平行于目标函数等值线; 相应地, 根据半椭球体内的优秀样本以及 MLE 新估计出的 PDE 的长轴也平行于目标函数等值线, 即倾向于与目标函数的改进方向相垂直. 另一方面, 该半椭球体内靠近原 PDE 中心的样本较多, 而远离原 PDE 中心的样本较少, 那么根据这些样本新估计出的 PDE 自然会发生收缩现象. 传统 GEDA 的这种特性大大降低了算法的搜索效率, 导致算法即使在斜坡型解区域中也可能早熟收敛^[9].

2 基于一般二阶混合矩的高斯分布估计算法 (GSM-GEDA)

在 GEDA 中, PDE 的位置、大小和长轴方向分别决定了算法的搜索中心、范围和主要搜索方向. 我们期望 PDE 位于易于发现更优解的解区域中, 其大小能够根据当前解区域的结构特征自适应地变化, 其长轴方向与目标函数的改进方向相一致. GSM-GEDA 通过修改均值和 CM 的估计方法来实现这一目的.

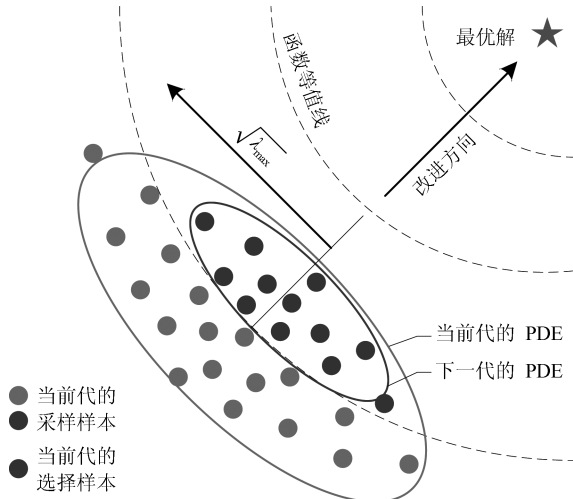


图1 传统 GEDA 中 PDE 的变化示意图
Fig.1 Schematic for the change of PDE of traditional GEDA

2.1 均值估计方法

在每次迭代中, 为了尽可能获得一个有前途的搜索中心, GSM-GEDA 根据如下两个步骤估计均值:

1) 采用加权样本预估均值. 由式 (2) 估计出的均值 $\tilde{\mu}$ 实际上是优秀样本的算术平均, 如果对较优样本赋予较大权重则有利于改进所估计均值的质量. 具体地, GSM-GEDA 首先根据下式为 GPM 预估出一均值:

$$\tilde{\mu} = \frac{\sum_{i=1}^{|S|} [\ln(|S| + 1) - \ln(i)] S_{(i)}}{\sum_{i=1}^{|S|} [\ln(|S| + 1) - \ln(i)]} \quad (4)$$

其中, $S_{(i)}$ 表示集合 S 中第 i 个最优解. 由式 (4) 可知, $S_{(i)}$ 的权重与其排序的对数值成反比; 排序越靠前, 其权重越大. 数值测试表明, $\tilde{\mu}$ 在大多数情况下都优于 $\bar{\mu}$.

2) 沿目标函数的改进方向偏移预估均值. 具体为:

$$\tilde{\delta}^t = \tilde{\mu}^t - \hat{\mu}^{t-1} \quad (5)$$

$$\hat{\mu}^t = \begin{cases} \tilde{\mu}^t + \eta_f \tilde{\delta}^t, & f(\tilde{\mu}^t + \eta_f \tilde{\delta}^t) < f(\tilde{\mu}^t) < f(\hat{\mu}^{t-1}) \\ \tilde{\mu}^t - \eta_b \tilde{\delta}^t, & \max \{ f(\tilde{\mu}^t - \eta_b \tilde{\delta}^t), f(\hat{\mu}^{t-1}) \} < f(\tilde{\mu}^t) \\ \tilde{\mu}^t, & \text{其他} \end{cases} \quad (6)$$

其中, $\hat{\mu}^t$ 表示第 t 次迭代中偏移后的均值; $\tilde{\delta}^t$ 表示当前预估均值 $\tilde{\mu}^t$ 与上次迭代中偏移后均值 $\hat{\mu}^{t-1}$ 之间的差异, 它反映了算法的进化方向; $f(\cdot)$ 表示需要极小化的目标函数. 通过显式地比较 $f(\tilde{\mu}^t)$ 与 $f(\hat{\mu}^{t-1})$, 可以获得 $f(\cdot)$ 的一个改进方向. 若 $\tilde{\mu}^t$ 优于 $\hat{\mu}^{t-1}$, 式 (6) 尝试沿 $\tilde{\delta}^t$ 的方向将 $\tilde{\mu}^t$ 偏移至 $\tilde{\mu}^t + \eta_f \tilde{\delta}^t$, 其目的是发挥算法的搜索惯性, 提高搜索效率, 其中的 η_f 称为前向偏移系数; $\tilde{\mu}^t + \eta_f \tilde{\delta}^t$ 最终能被接受的前提条件是 $f(\tilde{\mu}^t + \eta_f \tilde{\delta}^t) < f(\tilde{\mu}^t)$. 相反地, 若 $\tilde{\mu}^t$ 差于 $\hat{\mu}^{t-1}$, 式 (6) 则尝试沿 $\tilde{\delta}^t$ 的反方向将 $\tilde{\mu}^t$ 偏移至 $\tilde{\mu}^t - \eta_b \tilde{\delta}^t$, 其目的是及时修正算法的进化方向, 使其与目标函数的改进方向相一致, 其中 η_b 称为反向偏移系数; $\tilde{\mu}^t - \eta_b \tilde{\delta}^t$ 也仅在优于 $\tilde{\mu}^t$ 的情况下才会被最终接受. 当 $\tilde{\mu}^t$ 进行反向偏移时, 一个合理的设置是允许 $\tilde{\mu}^t$ 最远偏移至 $\hat{\mu}^{t-1}$, 因此 $0 < \eta_b \leq 1$; 另一方面, 大量测试表明, 当 $\eta_f \geq 1$ 时, $\tilde{\mu}^t$ 可以前向偏移至一理想位置. 综合以上两个因素, 一般可以设置 $\eta_b = 1/\eta_f$. 第 2.2 节将理论分析 η_f 对算法性能的影响; 第 3.1 节将实验测试 η_f 对算法最终求解质量的影响, 并给出取值方案.

2.2 协方差矩阵估计方法

GEDA 根据优秀样本估计 CM 的本质目的并不是准确获得样本自身的分布特征, 而是希望利用这些样本为下一次迭代确定一个合理的搜索范围和方向. 由第 1 节的分析可知, 与式 (3) 所示估计 \hat{C} 相对应的 PDE 的大小和长轴方向是不理想的. 另一方面, 由概率论中关于矩的知识可知, 随机向量的 CM 即为它的二阶混合中心矩. 那么在样本不变的前提下, 改变矩中心是调整 PDE 大小和长轴方向的一个简单方法. 对于 GSM-GEDA 来说, 一个自然选择是采用偏移后的均值 $\hat{\mu}$ 代替式 (3) 中的 MLE 均值 $\bar{\mu}$, 即:

$$\hat{C} = \frac{1}{|S|} \sum_{i=1}^{|S|} (S_i - \hat{\mu})(S_i - \hat{\mu})^T \quad (7)$$

上式表示对以偏移后的均值 $\hat{\mu}$ 这一一般位置为中心的三阶混合矩的估计.

由式 (5) 和 (6) 可知, $\hat{\mu}$ 不差于 $\tilde{\mu}$; 由式 (4) 可知, $\tilde{\mu}$ 通常优于 $\bar{\mu}$; 相应地, $\hat{\mu}$ 一般会优于 $\bar{\mu}$. 那么, $\hat{\mu} - \bar{\mu} = \tilde{\delta}$ 代表了目标函数的一个改进方向. 与式 (3) 所示估计 \hat{C} 相比, 式 (7) 给出的估计 \hat{C} 可以随

$\hat{\boldsymbol{\mu}}$ 的改变而自适应地调整 PDE 的大小 (体积), 并且可以使 PDE 的长轴方向趋近于目标函数的改进方向 ($\hat{\boldsymbol{\delta}}$). 为此给出以下定理和推论:

定理 1. 对于选定的优秀样本, 若 $\hat{\boldsymbol{\mu}} \neq \bar{\boldsymbol{\mu}}$, 那么由 \hat{C} 所确定的 PDE 的体积不小于由 \bar{C} 确定的 PDE 的体积.

证明. 由高斯分布的性质可知, PDE 的各个轴的方向分别与 CM 的各特征向量相一致, 而各半轴长度 a_d ($d = 1, 2, \dots, n$) 由 CM 的相应特征值 λ_d 确定: $a_d = \sqrt{\lambda_d}$. 另一方面, 由文献 [23] 可知, PDE 的体积与各半轴长度的乘积成正比. 记 \hat{C} 、 \bar{C} 的特征值分别为 $\hat{\lambda}_d$ 、 $\bar{\lambda}_d$ ($d = 1, 2, \dots, n$), 那么定理 1 成立的充要条件是, 当 $\hat{\boldsymbol{\mu}} \neq \bar{\boldsymbol{\mu}}$ 时, $\prod_{d=1}^n \hat{\lambda}_d \geq \prod_{d=1}^n \bar{\lambda}_d$. 为证明这一点, 对 \hat{C} 做如下变化:

$$\begin{aligned} \hat{C} &= \frac{1}{|S|} \sum_{i=1}^{|S|} (S_i - \hat{\boldsymbol{\mu}})(S_i - \hat{\boldsymbol{\mu}})^T = \\ &= \frac{1}{|S|} \sum_{i=1}^{|S|} (S_i - \bar{\boldsymbol{\mu}} - \hat{\boldsymbol{\delta}})(S_i - \bar{\boldsymbol{\mu}} - \hat{\boldsymbol{\delta}})^T = \\ &= \frac{1}{|S|} \sum_{i=1}^{|S|} (S_i - \bar{\boldsymbol{\mu}})(S_i - \bar{\boldsymbol{\mu}})^T - \\ &= \frac{1}{|S|} \sum_{i=1}^{|S|} (S_i - \bar{\boldsymbol{\mu}})\hat{\boldsymbol{\delta}}^T - \frac{1}{|S|} \hat{\boldsymbol{\delta}} \sum_{i=1}^{|S|} (S_i - \bar{\boldsymbol{\mu}})^T + \hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}^T = \\ &= \bar{C} + \hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}^T \end{aligned} \quad (8)$$

由上式可知, 矩阵 \hat{C} 是 \bar{C} 的秩 1 修正^[24]. 由于 \bar{C} 至少是半正定的, 对其进行特征值分解: $\bar{C} = \bar{Q}\bar{\Lambda}\bar{Q}^T$. 其中, $\bar{\Lambda} = \text{diag}\{\bar{\lambda}_1, \bar{\lambda}_2, \dots, \bar{\lambda}_n\}$, $0 \leq \bar{\lambda}_1 \leq \bar{\lambda}_2 \leq \dots \leq \bar{\lambda}_n$; \bar{Q} 表示由 \bar{C} 的特征向量构成的正交矩阵. 将 $\bar{Q}^T\hat{\boldsymbol{\delta}}$ 所对应的列向量简记为 \boldsymbol{p} , 并对 \hat{C} 进行如下正交相似变换:

$$\begin{aligned} \bar{Q}^T\hat{C}\bar{Q} &= \bar{Q}^T\bar{C}\bar{Q} + (\bar{Q}^T\hat{\boldsymbol{\delta}})(\bar{Q}^T\hat{\boldsymbol{\delta}})^T = \\ &= \bar{\Lambda} + \boldsymbol{p}\boldsymbol{p}^T \end{aligned} \quad (9)$$

很明显, 矩阵 $\bar{\Lambda} + \boldsymbol{p}\boldsymbol{p}^T$ 是 $\bar{\Lambda}$ 的秩 1 修正^[24]. 文献 [25] 中的引理 1.4 对秩 1 修正后对角型矩阵的特征多项式进行了讨论, 根据该引理可以推得 $\bar{\Lambda} + \boldsymbol{p}\boldsymbol{p}^T$ 也即 \hat{C} 的特征多项式:

$$\begin{aligned} F(\lambda) &= \prod_{d=1}^n (\lambda - \hat{\lambda}_d) = \\ &= \det(\lambda I - \bar{\Lambda} - \boldsymbol{p}\boldsymbol{p}^T) = \\ &= \det(\lambda I - \bar{\Lambda}) \det(I - (\lambda I - \bar{\Lambda})^{-1} \boldsymbol{p}\boldsymbol{p}^T) = \\ &= \prod_{d=1}^n (\lambda - \bar{\lambda}_d) \left(1 - \sum_{k=1}^n \frac{\boldsymbol{p}_k^2}{\lambda - \bar{\lambda}_k} \right) = \end{aligned}$$

$$\prod_{d=1}^n (\lambda - \bar{\lambda}_d) - \sum_{k=1}^n \left(\boldsymbol{p}_k^2 \prod_{d=1, d \neq k}^n (\lambda - \bar{\lambda}_d) \right) \quad (10)$$

其中, I 表示相应阶数的单位阵. 令式 (10) 中的 $\lambda = 0$, 可得:

$$\prod_{d=1}^n \hat{\lambda}_d = \prod_{d=1}^n \bar{\lambda}_d + \sum_{k=1}^n \left(\boldsymbol{p}_k^2 \prod_{d=1, d \neq k}^n \bar{\lambda}_d \right) \quad (11)$$

由上式可知, $\prod_{d=1}^n \hat{\lambda}_d \geq \prod_{d=1}^n \bar{\lambda}_d$. 由此得证. \square

推论 1. 对于选定的优秀样本, 若 $\hat{\boldsymbol{\mu}} \neq \bar{\boldsymbol{\mu}}$ 且样本数量大于问题维数 (即 $|S| > n$), 那么由 \hat{C} 所确定的 PDE 的体积将大于由 \bar{C} 所确定的 PDE 的体积.

证明. 由于式 (9) 中的 \bar{Q}^T 是正交矩阵, 那么当 $\hat{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}} = \hat{\boldsymbol{\delta}} \neq 0$ 时, $\boldsymbol{p} = \bar{Q}^T\hat{\boldsymbol{\delta}} \neq 0$. 即 $\exists k = 1, 2, \dots, n$, 满足 $\boldsymbol{p}_k \neq 0$. 另一方面, 集合 S 中各样本是相互独立的, 那么当 $|S| > n$ 时, 矩阵 \bar{C} 以概率 1 正定^[26]. 即对于 $\forall d = 1, 2, \dots, n, \bar{\lambda}_d > 0$. 综合上述两个方面, 可知 $\sum_{k=1}^n (\boldsymbol{p}_k^2 \prod_{d=1, d \neq k}^n \bar{\lambda}_d) > 0$, 进一步由式 (11) 可知 $\prod_{d=1}^n \hat{\lambda}_d > \prod_{d=1}^n \bar{\lambda}_d$. 由此得证.

推论 2. 对于选定的优秀样本, $\hat{\boldsymbol{\mu}}$ 到 $\bar{\boldsymbol{\mu}}$ 的马氏距离 $\hat{\boldsymbol{\delta}}^T \bar{C}^{-1} \hat{\boldsymbol{\delta}}$ 越大, 那么由 \hat{C} 所确定的 PDE 的体积越大.

证明. 重新考虑式 (11):

$$\begin{aligned} \prod_{d=1}^n \hat{\lambda}_d &= \prod_{d=1}^n \bar{\lambda}_d + \sum_{k=1}^n \left(\boldsymbol{p}_k^2 \prod_{d=1, d \neq k}^n \bar{\lambda}_d \right) = \\ &= \prod_{d=1}^n \bar{\lambda}_d \left(1 + \sum_{k=1}^n (\boldsymbol{p}_k^2 / \bar{\lambda}_k) \right) = \\ &= \prod_{d=1}^n \bar{\lambda}_d (1 + \boldsymbol{p}^T \bar{\Lambda}^{-1} \boldsymbol{p}) = \\ &= \prod_{d=1}^n \bar{\lambda}_d (1 + \hat{\boldsymbol{\delta}}^T \bar{Q} \bar{\Lambda}^{-1} \bar{Q}^T \hat{\boldsymbol{\delta}}) = \\ &= \prod_{d=1}^n \bar{\lambda}_d (1 + \hat{\boldsymbol{\delta}}^T \bar{C}^{-1} \hat{\boldsymbol{\delta}}) \end{aligned} \quad (12)$$

由上式可知, $\hat{\boldsymbol{\delta}}^T \bar{C}^{-1} \hat{\boldsymbol{\delta}}$ 越大, $\prod_{d=1}^n \hat{\lambda}_d$ 越大. 由此得证.

注 1. 推论 1 从理论上给出了保证由 \hat{C} 所确定的 PDE 的体积大于由 \bar{C} 所确定的 PDE 的体积的充要条件. 由上文分析可知, GSM-GEDA 在每次迭代中采用的 $\hat{\boldsymbol{\mu}}$ 通常不等于 $\bar{\boldsymbol{\mu}}$ 且优于 $\bar{\boldsymbol{\mu}}$; 另一方面, GSM-GEDA 及传统 GEDA 通常都设置群体规模远大于问题维数, 那么从中选出的优秀样本的数量一般都满足 $|S| > n$. 由此可知, GSM-GEDA 能

够使推论 1 成立. 这意味着, 对于相同的选定样本, GSM-GEDA 的探索能力强于传统 GEDA.

注 2. 推论 2 告诉我们, 若要增强 GSM-GEDA 的探索能力, 只需增大从 $\hat{\boldsymbol{\mu}}$ 到 $\bar{\boldsymbol{\mu}}$ 的马氏距离 $\hat{\boldsymbol{\delta}}^T \bar{C}^{-1} \hat{\boldsymbol{\delta}}$, 也即增大 $\mathbf{p}^T \bar{\Lambda}^{-1} \mathbf{p}$. 对于选定的优秀样本, \bar{C} 和 $\bar{\Lambda}$ 是固定的. 若要增大 $\mathbf{p}^T \bar{\Lambda}^{-1} \mathbf{p}$, 要求增大 \mathbf{p} 的各分量 p_d ($d = 1, 2, \dots, n$), 即 $\hat{\boldsymbol{\delta}}$ 在矩阵 \bar{C} 的各特征向量上的投影. 根据式 (6) 可知, 在优秀样本选定的情况下, 通过调整偏移系数 η_f 可以改变 $\hat{\boldsymbol{\mu}}$ 位置, 进而调整 $\hat{\boldsymbol{\delta}}(\mathbf{p})$ 的大小. η_f 越大, $\hat{\boldsymbol{\delta}}(\mathbf{p})$ 倾向于增大, 那么 GSM-GEDA 的探索能力会相应增强.

定理 2. 对于选定的优秀样本, 若 $\hat{\boldsymbol{\mu}} \neq \bar{\boldsymbol{\mu}}$, 那么 $\hat{\boldsymbol{\delta}}$ 与由 \hat{C} 所确定的 PDE 长轴之间的夹角不大于其与由 \bar{C} 所确定的 PDE 长轴之间的夹角.

证明. 不失一般性, 假设 $\bar{\Lambda} + \mathbf{p}\mathbf{p}^T$ 的特征值满足 $0 \leq \hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \dots \leq \hat{\lambda}_n$, 并记 $\bar{\Lambda} + \mathbf{p}\mathbf{p}^T$ 和 $\bar{\Lambda}$ 的主特征向量分别为 $\hat{\mathbf{q}}$ 、 $\bar{\mathbf{q}}$. 由式 (9) 可知, $\bar{\Lambda} + \mathbf{p}\mathbf{p}^T$ 、 $\bar{\Lambda}$ 分别是 \hat{C} 、 \bar{C} 以 \bar{Q}^T 为变换矩阵的正交变换, 那么定理 2 等价于证明: 若 $\bar{Q}^T \hat{\boldsymbol{\delta}} = \mathbf{p} \neq 0$, 那么 $\angle(\mathbf{p}, \hat{\mathbf{q}}) \leq \angle(\mathbf{p}, \bar{\mathbf{q}})$. 由于 \mathbf{p} 会与 PDE 的长轴构成两个夹角, 而这里只关注较小夹角, 所以不妨假设 $\mathbf{p}^T \hat{\mathbf{q}} \geq 0$, $\mathbf{p}^T \bar{\mathbf{q}} \geq 0$.

情况 1. 若 $\mathbf{p}^T \bar{\mathbf{q}} = 0$, 那么 $\angle(\mathbf{p}, \bar{\mathbf{q}}) = 90^\circ$; 另一方面, $\angle(\mathbf{p}, \hat{\mathbf{q}}) \leq 90^\circ$. 由此可知, 定理 2 在该情况下成立.

情况 2. 若 $\mathbf{p}^T \bar{\mathbf{q}} \neq 0$, 那么 $\hat{\lambda}_n > \bar{\lambda}_n$, $\mathbf{p}^T \hat{\mathbf{q}} \neq 0$. 这是因为 $\bar{\Lambda} + \mathbf{p}\mathbf{p}^T$ 是 $\bar{\Lambda}$ 的秩 1 修正, 满足 $\hat{\lambda}_d \geq \bar{\lambda}_d$ ($d = 1, 2, \dots, n$); 又由于 \mathbf{p} 在 $\bar{\mathbf{q}}$ 上有投影 ($\mathbf{p}^T \bar{\mathbf{q}} \neq 0$), 那么相应地, $\hat{\lambda}_n > \bar{\lambda}_n$ ^[24]. 另一方面, 由特征值定义可知:

$$(\bar{\Lambda} + \mathbf{p}\mathbf{p}^T)\hat{\mathbf{q}} = \hat{\lambda}_n \hat{\mathbf{q}} \quad (13)$$

若 $\mathbf{p}^T \hat{\mathbf{q}} = 0$, 那么 $\bar{\Lambda} \hat{\mathbf{q}} = \hat{\lambda}_n \hat{\mathbf{q}}$. 即 $\hat{\lambda}_n$ 也是 $\bar{\Lambda}$ 的特征值, 这与 $\hat{\lambda}_n > \bar{\lambda}_n$ 矛盾, 因此 $\mathbf{p}^T \hat{\mathbf{q}} \neq 0$. 由式 (13) 还可以进一步推得 $\hat{\mathbf{q}} = (\hat{\lambda}_n I - \bar{\Lambda})^{-1} \mathbf{p}\mathbf{p}^T \hat{\mathbf{q}}$.

将 $\hat{\lambda}_n$ 代入式 (10) 可知:

$$\begin{aligned} F(\hat{\lambda}_n) &= \left(1 - \sum_{d=1}^n \frac{\mathbf{p}_d^2}{\hat{\lambda}_n - \bar{\lambda}_d}\right) \prod_{d=1}^n (\hat{\lambda}_n - \bar{\lambda}_d) = 0 \\ &\Rightarrow \sum_{d=1}^n \frac{\mathbf{p}_d^2}{\hat{\lambda}_n - \bar{\lambda}_d} = 1 \\ &\Rightarrow 1 \geq \frac{\mathbf{p}_n^2}{\hat{\lambda}_n - \bar{\lambda}_n} > 0 \end{aligned} \quad (14)$$

进一步有

$$\begin{aligned} \cos \angle(\mathbf{p}, \hat{\mathbf{q}}) &= \frac{\mathbf{p}^T \hat{\mathbf{q}}}{\|\mathbf{p}\| \cdot \|\hat{\mathbf{q}}\|} = \\ &= \frac{\mathbf{p}^T \hat{\mathbf{q}}}{\|\mathbf{p}\| \cdot \left\| (\hat{\lambda}_n I - \bar{\Lambda})^{-1} \mathbf{p} \right\|} = \end{aligned}$$

$$\begin{aligned} &= \frac{1}{\|\mathbf{p}\| \sqrt{\sum_{d=1}^n \frac{\mathbf{p}_d^2}{(\hat{\lambda}_n - \bar{\lambda}_d)^2}}} \geq \\ &= \frac{|\mathbf{p}_n|}{\|\mathbf{p}\| \sqrt{(\hat{\lambda}_n - \bar{\lambda}_n) \sum_{d=1}^n \frac{\mathbf{p}_d^2}{(\hat{\lambda}_n - \bar{\lambda}_d)^2}}} \geq \\ &= \frac{|\mathbf{p}_n|}{\|\mathbf{p}\| \sqrt{(\hat{\lambda}_n - \bar{\lambda}_n) \sum_{d=1}^n \frac{\mathbf{p}_d^2}{(\hat{\lambda}_n - \bar{\lambda}_n)(\hat{\lambda}_n - \bar{\lambda}_d)}}} = \\ &= \frac{|\mathbf{p}_n|}{\|\mathbf{p}\| \sqrt{\sum_{d=1}^n \frac{\mathbf{p}_d^2}{\hat{\lambda}_n - \bar{\lambda}_d}}} = \\ &= \frac{|\mathbf{p}_n|}{\|\mathbf{p}\|} = \frac{\mathbf{p}^T \bar{\mathbf{q}}}{\|\mathbf{p}\| \cdot \|\bar{\mathbf{q}}\|} = \cos \angle(\mathbf{p}, \bar{\mathbf{q}}) \quad (15) \end{aligned}$$

即 $\angle(\mathbf{p}, \hat{\mathbf{q}}) \leq \angle(\mathbf{p}, \bar{\mathbf{q}})$. 由此可知, 若 $\mathbf{p}^T \bar{\mathbf{q}} \neq 0$, 定理 2 成立.

注 3. 在上述证明过程中, 对于情况 1 ($\mathbf{p}^T \bar{\mathbf{q}} = 0$), 可以推得 $\bar{\lambda}_n$ 、 $\bar{\mathbf{q}}$ 依然分别是 $\bar{\Lambda} + \mathbf{p}\mathbf{p}^T$ 的特征值、特征向量; 仅当 $\|\mathbf{p}\|$ 较小, 对其他特征值的调整幅度较小, 使得 $\bar{\lambda}_n$ 仍为 $\bar{\Lambda} + \mathbf{p}\mathbf{p}^T$ 的主特征值 ($\hat{\lambda}_n = \bar{\lambda}_n$) 时, 才有 $\angle(\mathbf{p}, \hat{\mathbf{q}}) = \angle(\mathbf{p}, \bar{\mathbf{q}}) = 90^\circ$. 对于情况 2, 仅当 $\mathbf{p}_n^2 = \hat{\lambda}_n - \bar{\lambda}_n$ 、 $\mathbf{p}_d = 0$ ($d = 1, 2, \dots, n-1$), 即 $\mathbf{p} // \bar{\mathbf{q}}$ 时, 才有 $\angle(\mathbf{p}, \hat{\mathbf{q}}) = \angle(\mathbf{p}, \bar{\mathbf{q}}) = 0^\circ$. 因此, 在绝大部分情况下, $\angle(\mathbf{p}, \hat{\mathbf{q}}) < \angle(\mathbf{p}, \bar{\mathbf{q}})$. 由于 PDE 的长轴, 即 CM 的主特征向量决定了 GEDA 的主要搜索方向, 那么定理 2 意味着, 在采用 \hat{C} 作为 CM 的估计之后, GEDA 的主要搜索方向更靠近目标函数的改进方向, 因此更容易找到较优解.

2.3 GSM-GEDA 的步骤

GSM-GEDA 在基本 EDA 的框架之下, 采用 GPM 描述优解的分布, 并采用了一种新的 GPM 参数估计方法, 其具体步骤如下:

步骤 1. 设置算法参数, 包括群体规模 m 、优秀样本比例系数 τ 以及前向偏移系数 η_f ; 记迭代次数 $t = 1$, 并采用伪均匀分布随机生成 m 个解, 初始化群体 M^t .

步骤 2. 根据目标函数 $f(\cdot)$ 评价 M^t 中各个解的质量.

步骤 3. 根据截断选择规则, 从 M^t 中选出前 $\lceil \tau m \rceil$ 个优解, 赋予优秀样本集 S^t .

步骤 4. 根据优秀样本集 S^t 估计概率分布参数: 1) 根据式 (4) 计算加权的预估均值 $\tilde{\boldsymbol{\mu}}^t$; 2) 根据式 (5) 和 (6) 对 $\tilde{\boldsymbol{\mu}}^t$ 进行偏移, 获得最终的估计均值 $\hat{\boldsymbol{\mu}}^t$; 3) 根据式 (7) 估计协方差矩阵 \hat{C}^t .

步骤 5. 根据高斯概率模型 $g(\mathbf{x} | \hat{\boldsymbol{\mu}}^t, \hat{C}^t)$ 新生成 $m - 2$ 个解, 并记由这些解构成的集合为 M' ; 采用精英策略更新群体 $M^{t+1} = M' \cup \{S_{(1)}^t, \hat{\boldsymbol{\mu}}^t\}$.

步骤 6. 更新 $t = t + 1$, 判断是否满足终止条件. 若满足, 则输出最优结果; 否则, 转至步骤 2.

其中, 步骤 5 采用常用的精英策略将当前群体中的最优个体 $S_{(1)}^t$ 保留到下一代群体中. 此外, 经过显式评价后的 $\hat{\mu}^t$ 也可以看作一个完整的解样本, 并且其解质量通常较高, 因此也将它保留到下一代群体中.

GSM-GEDA 与传统 GEDA 的区别主要体现在步骤 4 中的概率分布参数估计方法. 从式 (4)~(7) 可以看出, 新提出的均值、CM 估计方法的计算时间复杂度分别为 $O(\tau mn)$ 、 $O(mn)$ 、 $O(\tau mn^2)$ 、 $O(mn^2)$, 这与常用的 MLE 的计算复杂度相同. 这意味着 GSM-GEDA 可以以相同的计算复杂度获得具有更好理论性质的 GPM, 即可以避免 PDE 快速收缩, 从而保证算法具有适度的探索能力; 并且可以使 PDE 的长轴方向趋近于目标函数的改进方向, 进而提高算法的搜索效率.

3 实验与分析

为了评估 GSM-GEDA 的性能, 采用 IEEE CEC 2005 标准函数库^[27] 中的前 14 个函数对其进行了测试. 其中, $f_1 \sim f_5$ 是单模函数, $f_6 \sim f_{12}$ 是基本的多模函数, $f_{13} \sim f_{14}$ 是扩展的多模函数. 这些函数大多都经过了旋转、偏移操作, 分别使得各变量之间相互关联、最优解偏离搜索中心, 从而保证函数的优化难度. 在实验中, 所有函数的维数都设置为 30, 每个函数在每种算法上独立测试 25 次, 每次测试均以完成 300 000 次目标函数评价作为终止条件, 并采用所获优解与实际最优解之间的差值 (函数误差值) 在 25 次测试中的均值和标准差来衡量算法性能.

3.1 参数影响及其设置

GSM-GEDA 一共包括 3 个参数, 即群体规模 m 、优秀样本比例系数 τ 以及前向偏移系数 η_f . 对于 m 和 τ 这两个常规参数, 文献中已进行了较多研究^[1, 13], 本文将它们设置为常规值 $m = 1200$ 、 $\tau = 0.35$. 本节重点考察新参数 η_f 对 GSM-GEDA 的性能的影响. 由第 2 节可知, η_f 决定了预估均值 $\tilde{\mu}$ 的偏移程度, 从而改变 PDE 的中心位置、大小和长轴方向, 最终影响 GSM-GEDA 的性能.

图 2 以函数 f_2 、 f_{10} 为例, 给出了当 η_f 以 0.25 为间隔从 0.5 增大到 3.5 时, GSM-GEDA 的性能变化情况. 从图中可以看出, 当 η_f 在 0.5 ~ 3.5 这么一个较大的范围内变化时, GSM-GEDA 的性能变化并不剧烈. 特别地, 对于 f_2 , GSM-GEDA 求得的解总是非常接近最优解. 这说明 GSM-GEDA 的优化性能关于 η_f 的鲁棒性较强. 另一方面, 在不考虑随机因素的情况下, GSM-GEDA 的性能总是随着 η_f 的增大先变好后变差. 对于 f_2 这一单模函数, 当 $\eta_f \approx 1.0$ 时, GSM-GEDA 取得较优结果; 对于多模

函数 f_{10} , 当 $\eta_f \approx 2.75$ 时, GSM-GEDA 的表现较好. 实验表明, 上述结论基本上也适用于其他测试函数. 对于待优化的一般黑箱函数, 并没有先验信息确定其为单模还是多模. 为兼顾两类函数的求解质量, 我们建议在 1.0 ~ 2.75 之间为 η_f 取值; 本文选取 $\eta_f = 2.0$.

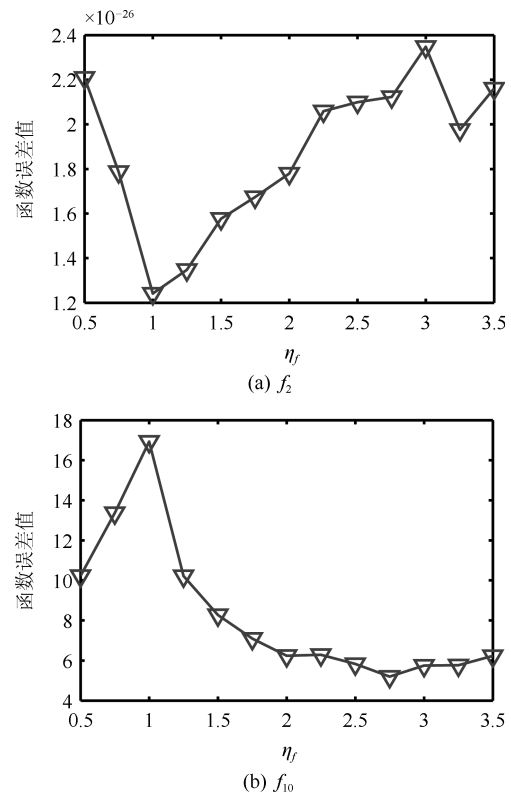


图 2 GSM-GEDA 的性能随 η_f 的变化情况

Fig. 2 The performance variation of GSM-GEDA with regard to η_f

3.2 均值-协方差矩阵估计方法的有效性分析

第 2 节理论证明了新提出的均值-协方差矩阵估计方法可以增大 PDE 的体积, 使 PDE 的长轴方向趋近于目标函数的改进方向, 本节将通过数值实验验证这一点. 实验设置如下: 在 GSM-GEDA 的每次迭代中, 除了根据式 (4)~(7) 所示的新方法估计 CM 之外, 还利用相同的优秀样本, 根据式 (2)~(3) 所示的传统 MLE 方法估计 CM, 从而满足定理 1 和 2 的限定条件. 为描述方便, 将这种依附于 GSM-GEDA 但采用传统估计方法的算法称为 quasi-GSM-GEDA. 此外, 还将 EMNA_g^[1] 作为对比算法. 除参数估计方法之外, EMNA_g 的其他设置与 GSM-GEDA 完全相同.

图 3 和 4 以函数 f_2 、 f_{10} 为例, 分别给出了各算法的 PDE 的长轴长度、长轴与目标函数改进方向 $\hat{\delta}$ 之间的夹角余弦 (简记为 $\cos(A_1)$) 随迭代次数的变化情况. 从图 3 中可以看出, GSM-GEDA 和 quasi-GSM-GEDA 的 PDE 的长轴长度都随迭代次

数的增大而逐渐减小, 但前者的长度始终不小于后者, 该现象在前期迭代过程中尤为明显. 实际上, 上述结论对于 PDE 的其他轴也是成立的, 这与定理 1 所述内容一致. 从图 4 中可以看出, GSM-GEDA 的 $\cos(A_1)$ 值始终不小于 quasi-GSM-GEDA 的相应值, 这说明 GSM-GEDA 的 PDE 的长轴与 $\hat{\delta}$ 之间的夹角不大于 quasi-GSM-GEDA 的相应夹角, 这与定理 2 所述内容一致.

从图 3 中还可以看出, 与 $EMNA_g$ 相比, GSM-GEDA 的 PDE 的长轴长度在前期迭代过程中保持了较大值, 这有利于算法进行探索性搜索; 在中后期迭代过程中, GSM-GEDA 的 PDE 的长轴收缩到一个较小的尺度范围内, 这有利于算法进行精细搜索. 为了进一步说明 GSM-GEDA 与 $EMNA_g$ 的性能差异, 图 5 和 6 分别给出了两算法的 PDE 的长轴与目标函数的最速下降方向 (即从 PDE 的中心指向

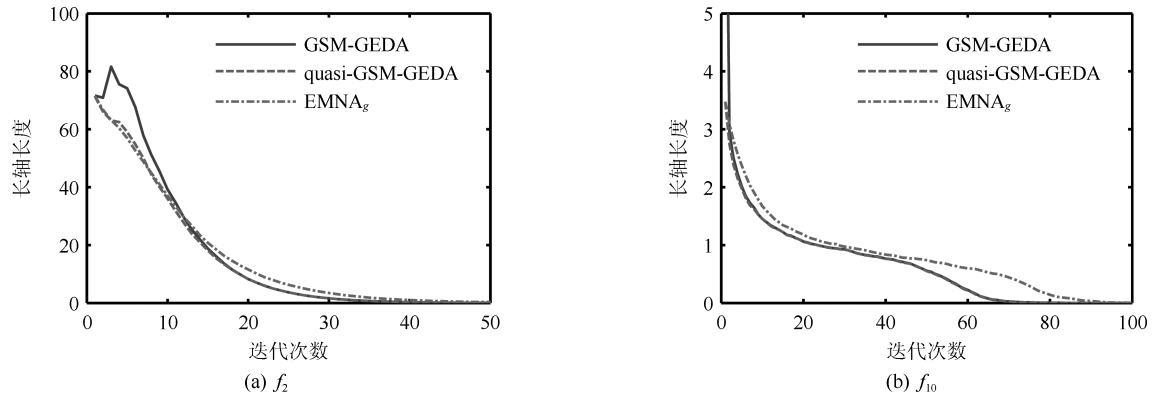


图 3 长轴长度随迭代次数的变化情况

Fig. 3 The variation of long axis length with regard to iteration times

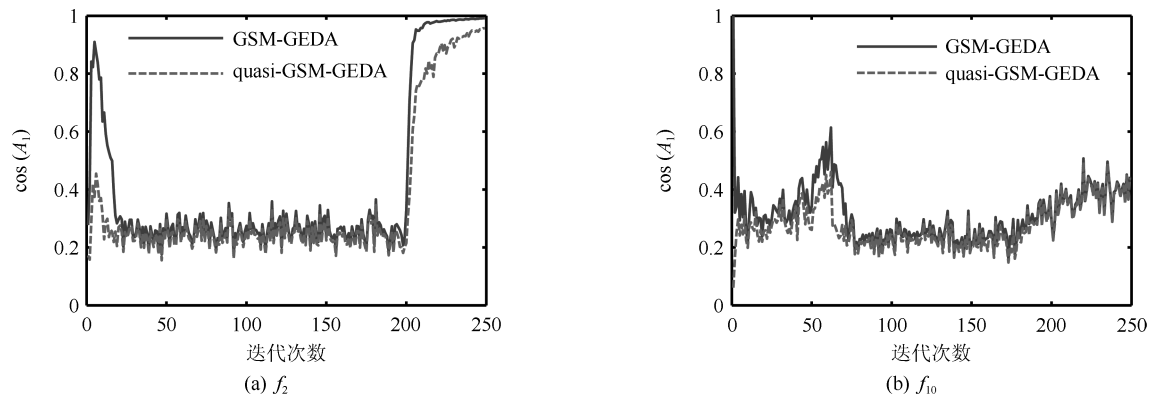


图 4 长轴与目标函数改进方向之间的夹角余弦随迭代次数的变化情况

Fig. 4 The variation of the cosine value of the angle between long axis and $\hat{\delta}$ with regard to iteration times

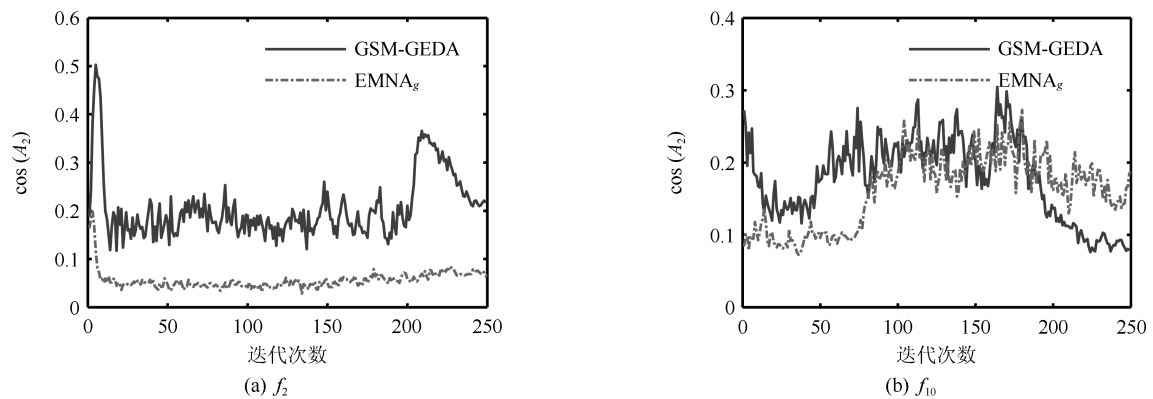


图 5 长轴与最速下降方向之间的夹角余弦随迭代次数的变化情况

Fig. 5 The variation of the cosine value of the angle between long axis and the steepest descent direction with regard to iteration times

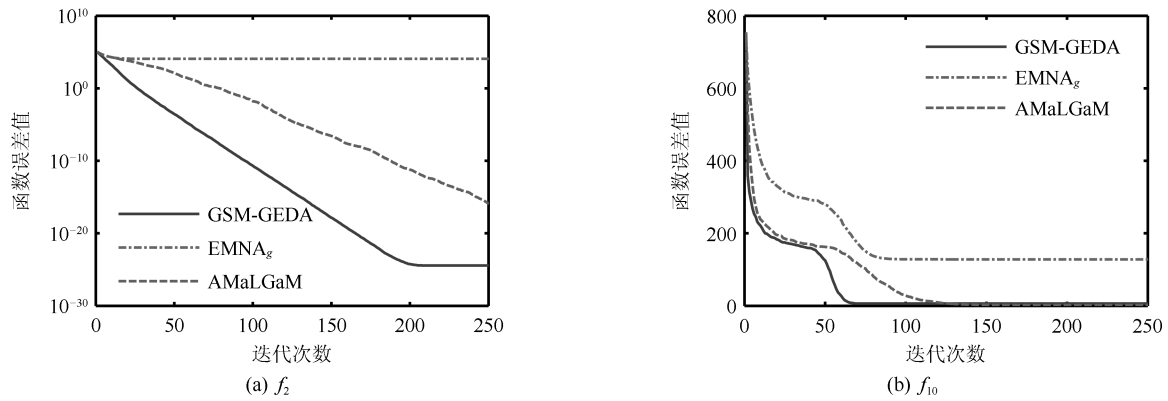


图6 函数误差值随迭代次数的变化情况

Fig. 6 The variation of the function error value with regard to iteration times

最优解的方向) 之间的夹角余弦 (简记为 $\cos(A_2)$)、所求得的最小函数误差值随迭代次数的变化情况. 从图 5 中可以看出, 对于 f_2 这一单模函数, GSM-GEDA 的 $\cos(A_2)$ 值始终大于 EMNA_g 的相应值, 说明 GSM-GEDA 能够探测到较好的搜索方向. 从图 6 中可以看出, GSM-GEDA 在经过大概 200 次迭代后基本上获得了 f_2 的最优解; EMNA_g 在经过大概 15 次迭代后便早熟收敛, 无法获得理想解. 对于多模函数 f_{10} , GSM-GEDA 在中前期迭代过程中的 $\cos(A_2)$ 值也大于 EMNA_g 的相应值, 尽管两算法都陷入了局部优解, 但 GSM-GEDA 获得的终解远优于 EMNA_g.

图 6 还给出了 AMaLGaM^[13] 获得的最小函数误差值的进化情况. AMaLGaM 利用预期均值偏移策略和自适应方差缩放策略改进了传统 GEDA, 是当前性能最为优越的 EDA 之一. 从图 6 中可以看出, AMaLGaM 和 GSM-GEDA 在不同程度上削弱了 EMNA_g 的早熟收敛问题. 在给定的目标函数评价次数内, GSM-GEDA 求得的 f_2 、 f_{10} 的终解分别优于、接近于 AMaLGaM 所得结果, 其收敛速度也快于 AMaLGaM.

3.3 与其他算法的比较

为了综合衡量 GSM-GEDA 的性能, 将其与 6 种已有算法进行了比较, 包括 EMNA_g^[1]、AMaLGaM^[13]、CMA-ES^[28]、CLPSO^[29]、CoBiDE^[30] 以及 MPEDE^[31]. 选择这 6 种算法作为对比算法的原因在于: EMNA_g 和 AMaLGaM 分别是传统 GEDA 和先进 GEDA 的典型代表, 可以为衡量 GSM-GEDA 的性能提供基准; CMA-ES、CLPSO、CoBiDE 分别是进化策略算法、粒子群算法、差分进化算法的代表算法, 并被多次用作对比算法, 而且 CoBiDE 中的交叉算子也涉及了 CM 的估计问题; MPEDE 是一种新近提出的差分进化算法, 一定程度上反映了当前函数智能优化方面的发展水平.

为了保证比较的公平性, EMNA_g、AMaLGaM 以及 GSM-GEDA 的参数 m 、 τ 取为相同值,

AMaLGaM 也选用全变量相关的 GPM 模型; AMaLGaM 的其余参数以及其他 4 种算法的参数均按照相应的原文献进行取值. 表 1 给出了各算法最终所获优解的函数误差值在 25 次独立测试中的均值和标准差, 表中还采用了下式所示的 Cohen's d 效应量^[32] 衡量 GSM-GEDA 与其他 6 种算法之间的性能差异:

$$d = \frac{\bar{x}_a - \bar{x}_b}{\sigma_{ab}}$$

$$\sigma_{ab} = \sqrt{\frac{(n_a - 1)\sigma_a^2 + (n_b - 1)\sigma_b^2}{n_a + n_b}} \quad (16)$$

其中, \bar{x}_a 、 σ_a 、 n_a 分别表示样本集 a 的均值、标准差、样本数量. 通常当 $|d| < 0.2$ 时, 认为两样本集的均值差异较小^[32], 这里意味着两个被测算法的性能无显著差异, 在表 1 中用符号“ \approx ”标识. 反过来, 若根据效应量判断出某算法显著优于或差于 GSM-GEDA, 则在表 1 中分别用符号“+”、“-”来标识.

从表 1 中可以看出, 对于单模函数 $f_1 \sim f_5$, GSM-GEDA 的性能比较理想. 尽管针对 f_1 、 f_2 的优化结果差于差分进化算法, 但其获得的解也已充分接近最优解. 对于基本的多模函数 $f_6 \sim f_{12}$, GSM-GEDA 在 3 个函数上的性能优于其他所有算法; 特别地, 它可以获得 f_7 、 f_{11} 的最优解. 对于 f_6 、 f_9 , 3 种 EDA 的表现都不太理想; 对于 f_8 , 所有被测算法的性能差异不大; 对于 f_{10} , GSM-GEDA 的表现仅次于 AMaLGaM, 优于其他 5 种算法. 对于扩展的多模函数, GSM-GEDA 在 f_{13} 上的表现与 CoBiDE 接近, 仅次于 CLPSO; 在 f_{14} 上的表现优于其他所有算法. 根据表 1 最后一行的统计结果可知, GSM-GEDA 分别在 14、9、12、9、9、9 个函数上的性能显著优于 EMNA_g、AMaLGaM、CMA-ES、CLPSO、CoBiDE、MPEDE, 表现出良好的全局寻优能力.

表 1 7 种算法最终求得的函数误差值的均值和标准差
Table 1 The mean and standard deviation of the function error values obtained by 7 algorithms

函数	EMNA _g	AMaLGaM	CMA-ES	CLPSO	CoBiDE	MPEDE	GSM-GEDA
f_1	2.21E+04 ⁻	1.63E-14 ⁻	1.58E-25 ⁻	0.00E+00 ⁺	0.00E+00 ⁺	0.00E+00 ⁺	3.98E-27
	1.65E+03	8.07E-14	3.35E-26	0.00E+00	0.00E+00	0.00E+00	7.85E-28
f_2	1.10E+04 ⁻	2.06E-16 ⁻	1.12E-24 ⁻	8.40E+02 ⁻	1.60E-12 ⁻	1.01E-26 ⁺	1.78E-26
	1.15E+03	6.64E-16	2.93E-25	1.90E+02	2.90E-12	2.05E-26	4.77E-27
f_3	1.97E+08 ⁻	3.37E-09 ⁻	5.54E-21 ⁻	1.42E+07 ⁻	7.26E+04 ⁻	1.01E+01 ⁻	1.12E-22
	3.18E+07	1.19E-08	1.69E-21	4.19E+06	5.64E+04	8.32E+00	2.85E-23
f_4	7.17E+03 ⁻	2.17E-11 ⁻	9.15E+05 ⁻	6.99E+03 ⁻	1.16E-03 ⁻	6.61E-16 ⁻	1.89E-25
	1.39E+03	1.05E-10	2.16E+06	1.73E+03	2.74E-03	5.68E-16	3.89E-26
f_5	2.81E+04 ⁻	2.46E+01 ⁻	2.77E-10 ⁻	3.86E+03 ⁻	8.03E+01 ⁻	7.21E-06 ⁻	8.90E-11
	2.22E+03	1.05E+01	5.04E-11	4.35E+02	1.51E+02	5.12E-06	3.54E-11
f_6	1.79E+09 ⁻	1.05E+01 ⁺	4.78E-01 ⁺	4.16E+00 ⁺	4.13E-02 ⁺	9.65E+00 ⁺	1.75E+01
	3.70E+08	1.49E+00	1.32E+00	3.48E+00	9.21E-02	4.65E+00	6.09E-01
f_7	1.08E+04 ⁻	1.79E-15 ⁻	1.82E-03 ⁻	4.51E-01 ⁻	1.77E-03 ⁻	2.36E-03 ⁻	0.00E+00
	2.88E+02	6.60E-15	4.33E-03	8.47E-02	3.73E-03	1.15E-03	0.00E+00
f_8	2.10E+01 ⁻	2.09E+01 [≈]	2.03E+01 ⁺	2.09E+01 [≈]	2.07E+01 ⁺	2.09E+01 [≈]	2.09E+01
	3.79E-02	5.43E-02	5.72E-01	4.41E-02	3.75E-01	5.87E-01	5.79E-02
f_9	5.46E+01 ⁻	3.66E+00 ⁺	4.45E+02 ⁻	0.00E+00 ⁺	0.00E+00 ⁺	0.00E+00 ⁺	7.48E+00
	8.62E+00	1.48E+00	7.12E+01	0.00E+00	0.00E+00	0.00E+00	1.93E+00
f_{10}	1.26E+02 ⁻	1.91E+00 ⁺	4.63E+01 ⁻	1.04E+02 ⁻	4.41E+01 ⁻	1.52E+01 ⁻	6.61E+00
	1.23E+01	1.52E+00	1.16E+01	1.53E+01	1.29E+01	2.98E+00	2.37E+00
f_{11}	6.25E+00 ⁻	3.58E-02 ⁻	7.11E+00 ⁻	2.60E+01 ⁻	5.62E+00 ⁻	2.58E+01 ⁻	0.00E+00
	1.37E+00	1.29E-01	2.14E+00	1.63E+00	2.19E+00	3.11E+00	0.00E+00
f_{12}	4.38E+04 ⁻	1.06E+03 [≈]	1.26E+04 ⁻	1.79E+04 ⁻	2.94E+03 ⁻	1.17E+03 ⁻	9.45E+02
	1.54E+04	1.87E+03	1.74E+04	5.24E+03	3.93E+03	8.66E+02	1.27E+03
f_{13}	3.85E+00 ⁻	2.85E+00 ⁻	3.43E+00 ⁻	2.06E+00 ⁺	2.64E+00 [≈]	2.92E+00 ⁻	2.77E+00
	6.32E-01	4.61E-01	7.60E-01	2.15E-01	1.13E+00	6.33E-01	3.05E-01
f_{14}	1.15E+01 ⁻	1.13E+01 ⁻	1.47E+01 ⁻	1.28E+01 ⁻	1.23E+01 ⁻	1.23E+01 ⁻	1.05E+01
	3.53E-01	3.23E-01	3.31E-01	2.48E-01	4.90E-01	4.22E-01	2.85E-01
+ / ≈ / - (个数)	0 / 0 / 14	3 / 2 / 9	2 / 0 / 12	4 / 1 / 9	4 / 1 / 9	4 / 1 / 9	—

4 结论

概率密度椭球体 (PDE) 的位置、大小和长轴方向是影响高斯分布估计算法 (GEDA) 性能的关键. 本文在分析传统 GEDA 性能弱化原因的基础上, 提出了一种基于一般二阶混合矩的高斯分布估计算法 (GSM-GEDA). 该算法以沿目标函数改进方向偏移后的加权均值作为协方差矩阵 (二阶混合中心矩) 的矩中心, 一方面将 PDE 引导至一个更有希望的解区域; 另一方面可以自适应地增大 PDE 的大小, 并使 PDE 的长轴方向趋近于目标函数的改进方向. 理论分析和数值实验均证明了上述结论的正确性. 此外, 在 14 个标准函数上的数值测试还表明, GSM-GEDA 的求解质量优于现有的 GEDA 以及

其他一些先进的进化算法. 下一步工作将研究设计更为系统化的均值偏移策略, 并在实际的复杂优化问题中检验 GSM-GEDA 的有效性.

References

- 1 Larrañaga P, Lozano J A. *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*. New York, USA: Springer, 2001. 57-100
- 2 Hauschild M, Pelikan M. An introduction and survey of estimation of distribution algorithms. *Swarm and Evolutionary Computation*, 2011, 1(3): 111-128
- 3 Wang Sheng-Yao, Wang Ling, Fang Chen, Xu Ye. Advances in estimation of distribution algorithms. *Control and Deci-*

- tion, 2012, **27**(7): 961–966, 974
(王圣尧, 王凌, 方晨, 许焯. 分布估计算法研究进展. 控制与决策, 2012, **27**(7): 961–966, 974)
- 4 Larrañaga P, Etxeberria R, Lozano J A, Peña J M. Optimization in continuous domains by learning and simulation of Gaussian networks. In: Proceedings of 2000 Genetic and Evolutionary Computation Conference. Las Vegas, USA: Morgan Kaufmann, 2000. 201–204
 - 5 Lu Q, Yao X. Clustering and learning Gaussian distribution for continuous optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: (Applications and Reviews)*, 2005, **35**(2): 195–204
 - 6 Yuan B, Gallagher M. On the importance of diversity maintenance in estimation of distribution algorithms. In: Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation. New York, USA: ACM, 2005. 719–726
 - 7 Pošík P. Preventing premature convergence in a simple EDA via global step size setting. *Parallel Problem Solving from Nature — PPSN X. Lecture Notes in Computer Science*. Berlin, Heidelberg, Germany: Springer-Verlag, 2008: 549–558
 - 8 Dong W S, Yao X. Unified Eigen analysis on multivariate Gaussian based estimation of distribution algorithms. *Information Sciences*, 2008, **178**(15): 3000–3023
 - 9 Ocenasek J, Kern S, Hansen N, Koumoutsakos P. A mixed bayesian optimization algorithm with variance adaptation. *Parallel Problem Solving from Nature — PPSN VIII. Lecture Notes in Computer Science*, Berlin, Heidelberg, Germany: Springer-Verlag, 2004. 352–361
 - 10 Bosman P A N, Grahl J, Rothlauf F. SDR: a better trigger for adaptive variance scaling in normal EDAs. In: Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation. New York, USA: ACM, 2007. 492–499
 - 11 Cai Y P, Sun X M, Xu H, Jia P F. Cross entropy and adaptive variance scaling in continuous EDA. In: Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation. New York, USA: ACM, 2007. 609–616
 - 12 Bosman P A N, Grahl J, Thierens D. Enhancing the performance of maximum-likelihood Gaussian EDAs using anticipated mean shift. *Parallel Problem Solving from Nature — PPSN X. Lecture Notes in Computer Science*. Berlin, Heidelberg, Germany: Springer-Verlag, 2008: 133–143
 - 13 Bosman P A N, Grahl J, Thierens D. Benchmarking parameter-free AMaLGaM on functions with and without noise. *Evolutionary Computation*, 2013, **21**(3): 445–469
 - 14 Cheng Yu-Hu, Wang Xue-Song, Hao Ming-Lin. An estimation of distribution algorithm with diversity preservation. *Acta Electronica Sinica*, 2010, **38**(3): 591–597
(程玉虎, 王雪松, 郝名林. 一种多样性保持的分布估计算法. 电子学报, 2010, **38**(3): 591–597)
 - 15 Karshenas H, Santana R, Bielza C, Larrañaga P. Regularized continuous estimation of distribution algorithms. *Applied Soft Computing*, 2013, **13**(5): 2412–2432
 - 16 Yang P, Tang K, Lu X F. Improving estimation of distribution algorithm on multi-modal problems by detecting promising areas. *IEEE Transactions on Cybernetics*, 2015, **45**(8): 1438–1449
 - 17 Ding N, Zhou S D, Sun Z Q. Histogram-based estimation of distribution algorithm: a competent method for continuous optimization. *Journal of Computer Science and Technology*, 2008, **23**(1): 35–43
 - 18 Zhang Jian-Hua, Zeng Jian-Chao. Estimation of distribution algorithm based on sequential importance sampling particle filters. *Acta Electronica Sinica*, 2010, **38**(12): 2929–2932
(张建华, 曾建潮. 基于序贯重点采样粒子滤波的分布估计算法. 电子学报, 2010, **38**(12): 2929–2932)
 - 19 Wang Li-Fang, Zeng Jian-Chao, Hong Yi. Estimation of distribution algorithm modeling and sampling by means of Copula. *Control and Decision*, 2011, **26**(9): 1333–1337, 1342
(王丽芳, 曾建潮, 洪毅. 利用 Copula 函数估计概率模型并采样的分布估计算法. 控制与决策, 2011, **26**(9): 1333–1337, 1342)
 - 20 Zhou A M, Sun J Y, Zhang Q F. An estimation of distribution algorithm with cheap and expensive local search. *IEEE Transactions on Evolutionary Computation*, 2015, **19**(6): 807–822
 - 21 Zhang Fang, Lu Hua-Xiang. General stochastic model for algorithm of distribution estimation with conditional probabilities and Gibbs sampling. *Control Theory & Applications*, 2013, **30**(3): 307–315
(张放, 鲁华祥. 利用条件概率和 Gibbs 抽样技术为分布估计算法构造通用概率模型. 控制理论与应用, 2013, **30**(3): 307–315)
 - 22 Zhong Wei-Cai, Liu Jing, Liu Fang, Jiao Li-Cheng. Estimation of distribution algorithm based on generic Gaussian networks. *Journal of Electronics and Information Technology*, 2005, **27**(3): 467–470
(钟伟才, 刘静, 刘芳, 焦李成. 建立在一般结构 Gauss 网络上的分布估计算法. 电子与信息学报, 2005, **27**(3): 467–470)
 - 23 Zhao Hong. Areas (Volumes) of n -dimension ellipsoid by quadratic curve (surface) enclosed. *Mathematics in Practice and Theory*, 2013, **43**(10): 279–282
(赵虹. n 元椭球面的判定及所围球体的体积. 数学的实践与认识, 2013, **43**(10): 279–282)
 - 24 Bunch J R, Nielsen C P, Sorensen D C. Rank-one modification of the symmetric eigenproblem. *Numerische Mathematik*, 1978, **31**(1): 31–48
 - 25 Yin Qing-Xiang. The inverse problem of rank-1 modification of real symmetric matrices. *Journal of Nantong Institute of Technology (Natural Science)*, 2003, **2**(4): 11–14
(殷庆祥. 实对称矩阵的秩 1 修正的特征反问题. 南通工学院学报(自然科学版), 2003, **2**(4): 11–14)

- 26 Xie Ping-Min, Chen Tu-Hao. The positive definiteness of the covariance matrix of continuous sample. *Acta Scientiarum Naturalium Universitatis Sunyatseni*, 1990, **29**(1): 102–104 (谢平民, 陈图豪. 连续型样本协差阵的正定性. 中山大学学报(自然科学版), 1990, **29**(1): 102–104)
- 27 Suganthan P N, Hansen N, Liang J J, Deb K, Chen Y P, Auger A, Tiwari S. Problem Definitions and Evaluation Criteria for the CEC 2005 Special Session on Real Parameter Optimization, Technical Report, Nanyang Technological University, Singapore, 2005.
- 28 Hansen N, Ostermeier A. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 2001, **9**(2): 159–195
- 29 Liang J J, Qin A K, Suganthan P N, Baskar S. Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. *IEEE Transactions on Evolutionary Computation*, 2006, **10**(3): 281–295
- 30 Wang Y, Li H X, Huang T W, Li L. Differential evolution based on covariance matrix learning and bimodal distribution parameter setting. *Applied Soft Computing*, 2014, **18**: 232–247
- 31 Wu G H, Mallipeddi R, Suganthan P N, Wang R, Chen H K. Differential evolution with multi-population based ensemble of mutation strategies. *Information Sciences*, 2016, **329**: 329–345
- 32 Cohen J. *Statistical Power Analysis for the Behavioral Sciences (Second Edition)*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

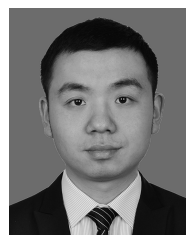


任志刚 西安交通大学电信学院自动化系副教授. 2010 年获得西安交通大学工学博士学位. 主要研究方向为机器学习与智能计算, 复杂系统的建模、优化与控制. 本文通信作者.

E-mail: renzg@mail.xjtu.edu.cn

(REN Zhi-Gang Associate professor in the Department of Automation Science and Technology, School of Electronic and Information Engineering, Xi'an Jiaotong University. He received

his Ph.D. degree from Xi'an Jiaotong University in 2010. His research interest covers machine learning and intelligent computing, modeling, optimization and control of complex systems. Corresponding author of this paper.)



梁永胜 西安交通大学电信学院自动化系硕士研究生. 2015 年获得西安交通大学电信学院自动化系学士学位. 主要研究方向为机器学习与智能计算. E-mail: liangyongsheng@stu.xjtu.edu.cn

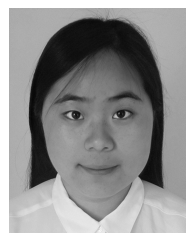
(LIANG Yong-Sheng Master student in the Department of Automation Science and Technology, School of Electronic and Information Engineering, Xi'an Jiaotong University. He received his bachelor degree from Xi'an Jiaotong University in 2015. His research interest covers machine learning and intelligent computing.)



张爱民 西安交通大学电信学院自动化系教授. 主要研究方向为控制理论与控制工程, 复杂系统的建模、优化与控制.

E-mail: zhangam@mail.xjtu.edu.cn

(ZHANG Ai-Min Professor in the Department of Automation Science and Technology, School of Electronic and Information Engineering, Xi'an Jiaotong University. Her research interest covers control theory and control engineering, modeling, optimization and control of complex systems.)



庞蓓 西安交通大学电信学院自动化系硕士研究生. 2016 年获得西安交通大学电信学院自动化系学士学位. 主要研究方向为机器学习与智能计算.

E-mail: beibei@stu.xjtu.edu.cn

(PANG Bei Master student in the Department of Automation Science and Technology, School of Electronic and Information Engineering, Xi'an Jiaotong University. She received her bachelor degree from Xi'an Jiaotong University in 2016. Her research interest covers machine learning and intelligent computing.)