

# 基于梯度扩散的自适应迁移双层进化优化算法

邵彩霞, 陈磊\*

广东工业大学数学与统计学院, 广东 广州

收稿日期: 2026年2月23日; 录用日期: 2026年3月18日; 发布日期: 2026年3月25日

## 摘要

进化双层优化因其广泛的适用性而日益受到关注, 其中下层优化任务间的知识迁移在提升上层基于种群的搜索效率方面发挥着重要作用。迁移学习方法提供了灵活的范式, 但在跨相关下层优化任务利用信息时, 常面临效率低下甚至负迁移的问题。为解决这些问题, 本文提出ATEB-GD——一种基于梯度扩散(GD)的自适应迁移进化双层优化算法。具体而言, 该方法采用基于梯度扩散的框架来实现有效的下层梯度迁移, 其中设计了混合距离度量以量化下层任务相似性, 并融入双进度驱动的自适应策略来调控知识迁移的时机与强度。此外, 进一步设计了动态更新的权重矩阵, 实现相似任务间梯度信息的定向共享, 同时双模式更新机制确保仅当满足迁移条件时才执行基于梯度扩散的更新。在两个广泛使用的双层基准测试集上的实验评估表明, 与三种先进的进化基线算法相比, 所提算法在性能和稳定性方面均表现出显著优势。

## 关键词

双层优化, 扩散梯度下降, 知识迁移, 自适应权重

# Adaptive Transfer Bilevel Evolutionary Optimization Algorithm Based on Gradient Diffusion

Caixia Shao, Lei Chen\*

School of Mathematics and Statistics, Guangdong University of Technology, Guangzhou Guangdong

Received: February 23, 2026; accepted: March 18, 2026; published: March 25, 2026

## Abstract

Evolutionary Bilevel optimization has gained increasing attention due to its broad applicability,

\*通讯作者。

where the knowledge transfer among lower-level optimization tasks plays an important role in improving the efficiency of upper-level population-based search. Transfer learning methods provide a flexible paradigm but often suffer from inefficient or even negative transfer when leveraging information across related lower-level optimization tasks. To address these issues, this paper presents ATEB-GD, an adaptive transfer-based evolutionary bilevel optimization algorithm built upon gradient diffusion (GD). In particular, the proposed method employs GD-based framework to effectively transfer lower-level gradients, where a hybrid distance metric is designed to quantify lower-level task similarity, and a dual-progress-driven adaptive strategy is incorporated to regulate the timing and intensity of knowledge transfer. In addition, a dynamically updated weight matrix is further devised to enable directed sharing of gradient information among similar tasks, while a dual-mode update mechanism ensures that DGD-based updates are executed only when transfer conditions are met. Experimental evaluations on two widely used bilevel benchmark suites demonstrate that the proposed algorithm achieves superior performance and stability compared with three state-of-the-art evolutionary baselines.

## Keywords

Bilevel Optimization, Diffusion Gradient Descent, Knowledge Transfer, Adaptive Weight

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

双层优化问题是一类复杂的层次化优化模型, 包含两个相互依赖的决策层次——上层(领导者)问题和下层(追随者)问题。领导者旨在最小化其上层目标函数, 而追随者在领导者的约束下优化其下层目标函数。这两个层次之间既存在合作又存在冲突: 上层决策变量和目标函数依赖于下层问题的最优解, 而下层优化过程则参数化地依赖于上层决策; 同时, 由于目标函数本质上的不一致性, 两者之间也存在内在冲突, 导致目标函数之间潜在的矛盾以及决策变量的相互约束。该建模框架已广泛应用于管理科学[1][2]、交通规划[3][4]和电力系统[5]等多个领域, 体现了其在层次决策场景中具备广泛的应用价值。

由于双层优化问题具有复杂的嵌套结构特征[6], 其求解过程具备巨大的挑战, 是一个 NP 难的问题。对于双层优化问题运用传统的经典优化方法来求解存在明显的局限性, 例如确定性算法的应用对函数的凸性和可微性有严格的限制, 很难求解非凸非光滑这一类问题; 群体智能优化方法都要基于特定问题的假设, 收敛精度和运行速度都较低等等。通过相关文献的研究可以发现相比之下进化算法在解决双层优化问题中发挥了独特的优势。Hao 等人[7]提出了 BL-SAEA 算法, 通过构建代理模型在上层中选择有前景的解, 在下层中初始化种群, 有效地提高了函数的优化性能。He 等人[8]提出了协方差矩阵自适应进化策略, 并且设计了一种搜索分布共享机制, 达到了调用下层问题的先验知识来解决上层优化问题的目的, 进而大大减少了函数评估次数。Sinha 等人[9]提出了利用近似的 KKT 条件把双层优化问题转化为单层问题, 该方法比原始严格的 KKT 条件更具备实用性。Bedi 等人[10]提出了一种基于 Kriging 近似模型来求解双层优化全局优化问题, 在该方法中主要通过迭代近似双层优化问题中的下层最优值函数映射来降低算法的计算成本。Chen 等人[11]提出了一种基于迁移学习的并行化算法框架(TLEA)来解决双层优化问题, 该算法主要通过并行处理下层问题的任务, 运用下层优化任务之间的相关性来进行知识迁移, 进而加快下层优化问题的运算速度, 提高下层搜索进程的有效性。

目前已经有越来越多的研究发现下层优化任务之间存在相关性, 因此运用多任务进化策略以及迁移

学习的技术, 通过相似任务之间的知识共享来解决双层优化问题具备可行性和高效性。对于多任务进化策略, Bai 等人[12]提出了多任务梯度下降(GD)算法, 该算法对于凸问题可以保证收敛性, 但是对于非凸问题并不适用。为了扩大多任务梯度下降算法的应用范围, Liu 等人[13]提出了基于扩散梯度下降的多因子进化算法, 该算法根据不同任务(包含非凸任务)之间的相似性共享各任务的梯度信息, 运用某些任务的局部凸性来协助其他任务脱离局部最优, 加快收敛的速度。然而当任务之间的相似性处于较低水平时会降低不同任务之间的学习能力导致迁移学习的效果较差, 因此在运用迁移学习策略解决双层优化问题中如何利用不同下层任务解的梯度信息来进行知识迁移, 同时促进正迁移抑制无效迁移和负迁移是一个具备挑战性的课题。

因此, 本文的主要贡献如下:

(1) 提出一种基于扩散梯度下降与协方差矩阵自适应进化策略相融合的双层优化算法框架, 上层采用 CMA-ES 进行全局搜索, 下层引入 DGD 机制, 实现梯度信息在多任务之间的自适应共享与协同优化。

(2) 设计一种混合距离度量指标, 综合上层决策变量结构信息与下层优化状态信息, 更准确衡量优化任务间相似性。

(3) 提出基于双进度调整的自适应迁移概率机制, 动态控制知识迁移时机, 有效促进正向迁移并抑制负迁移。

(4) 构建可动态更新的权重矩阵, 结合梯度方向一致性与指数平滑策略, 实现经验知识在相似任务间的定向、自适应传递。

## 2. ATEB-GD 改进算法

扩散梯度下降作为一种分布式优化方法, 近年来在多任务学习与协同优化中展现出显著优势。其核心思想是将多个优化任务建模为图结构中的节点, 并通过任务之间的相似性构建权重矩阵, 从而在优化过程中实现梯度信息的定向共享。该方法可以使各任务在优化过程中不仅能利用自身梯度信息, 还能共享相似任务的梯度方向, 从而引导搜索过程加速收敛并避免陷入局部最优。

为此, 本文针对双层优化问题的层次结构与任务相关性, 将 DGD 策略引入下层优化的求解过程。其核心目标是: 利用下层优化任务间的相似性度量, 构建一个自适应的梯度共享机制。该机制能够动态识别相似任务, 并允许它们之间进行有向、高效的梯度信息交换, 从而将单一任务的优化经验有效扩散至相关任务群体。这一设计不仅加速了整体收敛过程, 更通过确保共享信息的正向性, 保障了优化过程的稳定性与效率, 大大减少了任务之间的无效迁移和负迁移。另外该方法不需要对任务函数的凸性做出严格的限制, 当不同任务的最佳解接近距离时该方法便具备广泛性和有效性。本文将提出的这些策略与 BI-CMAES 结合共同解决双层优化问题。本文将通过第三节的实验来验证该方法的准确性和鲁棒性。

### 2.1. 预备知识

在本文中, 我们提出了一种基于底层扩散梯度下降的双层优化算法。在上层中运用协方差矩阵自适应进化策略进行优化, 利用上层分布的边缘分布信息作为下层优化的初始信息, 以加速初始化过程。在下层优化中, 针对每一个下层任务, 采用有限差分法估计其梯度信息, 并通过混合距离度量评估任务间的相似性, 构建自适应权重矩阵来共享相似任务之间的梯度信息, 提出基于双进度调整的自适应迁移概率机制, 动态调整是否进行知识迁移。在此基础上, 引入扩散梯度下降机制, 实现相似任务间的梯度定向共享与协同更新, 从而提升整体收敛效率与求解精度。利用 SMD 和 TP 测试问题集来测试该算法的性能。通过与其成熟的双层优化算法的比较, 从收敛精度、稳定性以及计算效率等方面验证该算法的有

效性, 实验结果将在第三节中进行详细展示及分析。

## 2.2. ATEB-GD 主框架

本文所提出的 ATEB-GD 算法框架如算法 1 所示。在该算法中, 上下层都先运用了协方差矩阵自适应进化策略(CMA-ES)来求解决策变量。关于 CMA-ES 算法本文先做如下所示的简要介绍: 首先在全局决策空间中维持高斯搜索分布, 利用上层分布中关于下层变量的边缘分布作为下层的初始分布。之后在每次的迭代中, 比较当前最优解与精英解。若当前解更优, 则通过下层 CMA-ES 进一步精细化; 否则以一定概率对精英解或当前解进行精细化。

由上述分析可以发现在上层空间运用 CMA-ES 算法时会同时涵盖上下层的决策变量, 因此在下层问题优化的过程中运用上层分布的边缘分布作为先验知识, 可以使下层决策变量的初始化更具备合理化和高效性, 避免从零开始, 可以大大降低下层迭代函数的运行数量, 加快收敛的速度。之后执行基于扩散梯度下降的下层搜索过程。在下层, 算法不仅计算各个优化任务的梯度信息, 还通过混合距离指标衡量任务之间的相似性, 并且根据自适应迁移策略决定是否进行知识迁移。最后通过扩散梯度下降与种群进化协同的策略来更新下层决策变量, 将下层更新得到的最优决策变量反馈给上层进行统一评估与更新, 并且根据新的最优上下层决策变量值来更新 CMA-ES 的参数值。该框架在保持 CMA-ES 全局搜索能力的同时, 通过下层多任务协同显著加速了收敛过程。

---

### Algorithm 1 ATEB-GD

---

输入:

- 上层函数最大评估次数
- 下层函数最大评估次数
- 迭代终止条件

输出: 最优解

初始化上层分布参数: 两个进化路径  $P_k^u$ 、 $S_k^u$ ; 步长  $\delta_k^u$ ; 均值向量  $V_k^u$ ; 协方差矩阵  $C_k^u$ ; 上层种群大小  $U_\lambda$ ; 最大迭代次数 (gen=0)

for  $k \leftarrow 1$  to  $U_\lambda$  do

  while 迭代终止条件未满足 do

    选择符合正态分布的种群样本

$U_p^u = V_k^u + \delta_k^u * G(0, C_k^u)$  // 正态分布采样

    修复种群 // 边界处理

    执行基于 DGD 的下层搜索 // 调用算法 2

    反馈到上层并评估上层搜索种群

    更新  $P_k^u$ 、 $C_k^u$ 、 $V_k^u$ 、 $S_k^u$ 、 $\delta_k^u$

    gen=gen+1

  end while

end for

---

## 2.3. 基于扩散梯度下降的下层搜索

基于扩散梯度下降的下层搜索方法关键部分在于下层优化任务的梯度信息的求解, 相似程度的表示, 如何通过相似程度来共享优化任务的梯度信息以及如何根据共享后的梯度信息来更新下层决策变量值。针对上述所提出的问题, 本节将详细介绍基于扩散梯度下降的下层搜索方法。

**Algorithm 2** 基于 DGD 的自适应迁移下层搜索

输入:

- $x_u, x_l$ : 决策变量 (上层/下层)
- $popsiz$ : 上层个体数量
- $l_{dim}$ : 下层搜索维度
- $maxIter_u, maxIter_l$ : 最大迭代次数
- 终止条件

输出: 各上层变量对应的下层最优值

初始化:  $iter_u, iter_l \leftarrow 0$ **for**  $i \leftarrow 1$  **to**  $popsiz$  **do**  **while** 终止条件未满足 **do**

步骤 1: 梯度计算

    使用有限差分法计算  $\nabla f_i$  (式 (2))

步骤 2: 迁移决策

    计算  $TP$  (式 (6)-(7))    生成随机数  $r \sim U(0, 1)$ 

步骤 3: 变量更新

    使用相似度初始化  $a_{ij}^0$  (式 (9))    自适应更新  $a_{ij}^t$  (式 (11))

应用双随机处理 (式 (12)-(13))

聚合梯度 (式 (14))

更新变量 (式 (15))

无迁移: 使用标准梯度下降更新变量

步骤 4: 约束处理与评估

应用边界投影 (式 (17))

    评估  $f_i(x_{l,i})$      $iter_l = iter_l + 1$   **end while**   $iter_u = iter_u + 1$   **end for**  输出  $(x_l^{best1}, \dots, x_l^{best\ popsize})$ 

由此可见该方法主要包含六个核心步骤:

- (1) 梯度信息的高效估计: 通过有限差分法近似非光滑或者黑箱函数的梯度;
- (2) 任务相似性的混合度量: 结合上层变量距离与下层性能差异构建混合距离指标;
- (3) 自适应迁移机制: 基于上层优化进度和下层优化状态来建立双重进度调节的自适应迁移概率机制, 通过动态调整迁移概率来促进正迁移, 抑制负迁移;
- (4) 知识迁移权重计算: 基于任务相似度构建并动态更新权重矩阵;
- (5) 梯度共享与变量更新: 根据迁移机制结果确定是否进行知识迁移, 对于进行知识迁移的下层优化任务: 利用加权梯度信息进行扩散式梯度下降来共享相似任务之间的有效经验, 进而更新下层决策变量; 对于未达到知识迁移标准的下层优化任务, 仅根据自身的梯度信息进行标准梯度下降范式来更新决策变量; 边界约束处理: 通过投影法保证解的可行性。

整个流程形成一个闭环的协同优化系统, 其中梯度共享机制使相似任务能够互相促进, 而权重自适应机制则有效抑制了任务间的冲突干扰。

### 2.3.1. 求解梯度信息

由于在下层多任务优化问题中, 目标函数往往为非光滑函数或者黑箱函数, 难以直接求解解析梯度, 并且传统优化方法容易陷入局部最优。因此本文运用函数值的有限差分来近似目标函数的梯度。该方法仅依赖函数值计算, 不要求函数可微, 适用于广泛的实际优化场景。具体而言, 在每次迭代中, 对当前的下层决策变量  $x_l$  施加  $l_{\text{dim}}$  个随机噪声(其中  $l_{\text{dim}}$  为下层搜索的维度):

$$\epsilon_i \sim \mathcal{N}(0, I_d) \quad (1)$$

生成随机扰动样本  $x_l + \sigma\epsilon_i$  和  $x_l - \sigma\epsilon_i$ , 根据扰动后的目标函数值差异来近似模拟函数的梯度:

$$\nabla f_l(x_l) = \frac{1}{M} \sum_{i=1}^M \frac{f_l(x_l + \sigma\epsilon_i) - f_l(x_l - \sigma\epsilon_i)}{2\sigma} \quad (2)$$

其中  $\sigma$  表示扰动强度系数, 决定了随机扰动的缩放比例。为进一步提升梯度估计的自适应性, 本文引入梯度信噪比机制来动态调整扰动强度  $\sigma$ :

$$\text{SNR} = \frac{\|\text{mean}(\text{gradients})\|}{\text{std}(\text{gradients})} \quad (3)$$

这里,  $\|\text{mean}(\text{gradients})\|$  表示对所有下层问题维度的梯度估计值取平均, 反映了梯度的整体下降方向, 当平均值趋向于 0 时, 可能集中趋向于稳定点或者局部最优;  $\|\text{std}(\text{gradients})\|$  表示对所有下层问题维度的梯度估计值取标准差, 反映了梯度的整体一致性, 当标准差较低时梯度趋向平稳或者陷入平坦区域, 反之可能受噪声影响较大。 $\sigma$  基于梯度信噪比动态调整: 当信噪比较高时, 减少  $\sigma$  值以提高精确度; 反之加大  $\sigma$  值以减少噪声对算法整体效能的影响。

该做法对目标函数的性质没有要求, 相比于解析梯度的求解更具备适用性和广泛性, 同时通过并行评估多个扰动样本来近似梯度值可以提升优化的速度, 结合信噪比动态调节扰动强度, 兼顾了精细搜索与全局搜索。另外, Liu 等人[14]已经从理论上证明了梯度方向与扩散梯度下降的方向相同, 这使得进化算法可以模拟扩散梯度下降的收敛行为。

### 2.3.2. 相关性度量

任务相关性的度量对于双层优化中实现跨层次优化任务的知识迁移至关重要, 传统的相似性衡量方法, 如欧氏距离[15]、余弦相似度, 已经被广泛应用于衡量不同优化任务之间的相关性。然而这些传统方法在使用过程中具有一定局限性, 例如当目标函数的优化空间发生显著变化时, 仅凭欧氏距离评估底层优化任务的相关性已不够可靠; 余弦相似度仅仅考虑了向量之间的角度关系, 忽略了它们之间的幅度差异以及任务的实际优化状态, 另外两个任务可能具有相似的上层变量方向, 但是处于完全不同的优化阶段, 一个可能已经收敛到局部最优, 而另外一个仍在探索阶段, 仅仅基于方向相似性的盲目知识迁移可能会导致负迁移。

为此, 本文提出了一种混合距离指标, 用于计算不同高层变量对应的底层优化任务间的关联度。该指标整合了来自上层空间的结构信息和来自优化过程的状态信息, 所提出的指标包含两个互补的分量:

$$\begin{aligned} \text{distance}_{ij} &= d_{\text{struct}} + d_{\text{state}} \\ &= \|\mathbf{x}_u^{(i)} - \mathbf{x}_u^{(j)}\|_2 + \gamma \cdot \frac{|f_i - f_j|}{|f_i| + |f_j| + \epsilon} \end{aligned} \quad (4)$$

$$\gamma = 1 - \left(1 - \frac{T}{t}\right)^2 \quad (5)$$

其中  $d_{\text{struct}}$  为结构分量, 通过计算上层决策变量之间的欧式距离来衡量底层优化任务在决策空间中的接近程度。这一分量反映了任务在参数配置上的相似性, 当两个任务的上层变量取值相近时, 其对应的下层优化问题通常也具有相似的最优解区域。  $d_{\text{state}}$  为状态分量, 主要通过归一化的函数值差异来衡量任务在优化状态上的接近程度, 此处  $f_i$  和  $f_j$  分别表示任务  $i$  和  $j$  的当前下层目标函数值,  $\epsilon$  是一个小的正数以避免除零错误。  $\gamma$  为动态权重系数, 其中  $t$  为当前迭代次数,  $T$  表示最大迭代次数。

混合距离指标可以解决传统方法出现的局限问题, 在早期阶段, 各个任务的优化状态尚不稳定, 此时强调结构分量, 在后期阶段, 状态分量权重增加, 充分利用稳定的状态信息来捕捉任务之间优化进度的差异, 更好地区分已收敛任务和仍处于优化的任务。

### 2.3.3. 自适应迁移策略

(1) TP 动机: 在双层优化问题中, 知识迁移的时机选择对算法性能具有决定性影响。然而, 跨任务知识迁移并不总能产生积极效果, 过早的知识迁移可能因任务间相似性不足而导致负迁移, 反而降低优化效率; 过晚的迁移则会错失利用任务相关性的机会, 无法充分发挥迁移学习的优势。因此, 我们需要谨慎对待知识迁移。

(2) 自适应迁移概率法: 本文基于上层优化进度和下层优化状态提出了一种双重进度调节的自适应迁移概率机制:

$$\begin{aligned} \text{TP} &= \text{TP}_u(p_u) \times \alpha(p_l) \\ &= \left[1 - (1 - \ln(p_u))^{c_p}\right] \times \alpha(p_l) \end{aligned} \quad (6)$$

其中,  $\text{TP}_u(p_u)$  为上层进度主导项。  $p_u = \text{iter}_u / \text{maxIter}_u \in (0, 1]$ ,  $c_p = (-1) \cdot \frac{\ln(100+a)}{\ln(1-p)}$ , 由多次调试结果

可知当未知参数  $a = 3$ ,  $p = 0.2$  时结果达到最优。另外  $\text{iter}$  表示上层优化过程的迭代次数,  $\text{maxiter}$  则代表该过程的最大迭代次数。在早期阶段, 上层决策变量分布较分散, 对应的下层优化任务可能具有较大的差异, 此时  $\text{TP}_u$  趋向于 0, 几乎不迁移。中期阶段, 上层变量开始向潜在最优区域聚集, 此时对应的下层优化任务之间的相似性增加。后期阶段, 上层变量高度聚集在最优解附近, 任务之间相似性显著。上层进度主导项函数呈 S 型, 在早期缓慢增长, 中期加速, 后期延缓。

下层优化状态通过调解函数  $\alpha(p_l)$  影响迁移概率:

$$\alpha(p_l) = \begin{cases} 0.5, & 0 \leq p_l < 0.3 \\ 0.8, & 0.3 \leq p_l < 0.7 \\ 1.0, & 0.7 \leq p_l \leq 1 \end{cases} \quad (7)$$

其中  $p_l = \text{iter}_l / \text{maxIter}_l$  表示下层优化进度。在早期探索阶段 ( $0 \leq p_l < 0.3$ ), 下层仍在探索, 优化方向不稳定, 保守迁移避免随机迁移导致的性能下降; 在中期开发阶段 ( $0.3 \leq p_l < 0.7$ ), 下层已有明确优化方向, 在相似性明确时适度知识共享; 在后期收敛阶段 ( $0.7 \leq p_l < 1$ ), 下层接近收敛, 促进相似任务之间的协同优化。

### 2.3.4. 权重矩阵确定

由于当上层决策变量相似程度越高时, 所对应的下层决策变量的接近程度也会随之提升, 对于相关程度高的下层优化任务, 共享所对应的下层决策变量的梯度信息进行知识转移可以加快向全局最优收敛

的速度。因此基于混合距离, 任务  $i$  和任务  $j$  之间的相似度通过指数转换获得:

$$\text{sim}_{ij} = \exp(-\text{distance}_{ij}) \quad (8)$$

指数函数将距离映射到(0,1]之间, 满足相似度度量的基本要求: 距离越小, 相似度越高; 距离为 0 时, 相似度为 1。

之后通过归一化处理得到知识转移权重矩阵的初始值  $a_{ij}^{(0)}$ :

$$a_{ij}^{(0)} = \frac{\text{sim}(f_i, f_j)}{\sum_{k=1}^n \text{sim}(f_i, f_k)} \quad (9)$$

对于相似程度高的任务, 初始权重值设置得较大, 有利于早期知识的迁移, 反之对于相似程度低的任务, 初始权重值设置趋向于 0, 减少早期负迁移的风险。权重矩阵的初始化设置建立了下层优化任务之间的初始连接, 满足强连通性。

在下层任务优化的过程中不同任务之间的关系是动态变化的, 任务的梯度方向以及相似程度会随着搜索位置的变化而变化, 同时 Liu 等人[16]也从理论上证明了权重矩阵需要具备时变适应性, 才能维持扩散过程中的强连通性和信息守恒性, 因此本文建立了自适应知识迁移的策略: 运用指数平滑来动态调整权重矩阵。

考虑到当两个任务的梯度内积大于 0 时, 梯度方向具备一致性, 这两个任务的存在正相关性, 此时可以通过适当增加权重值来增强正迁移促进知识的共享; 反之当任务的梯度内积小于 0, 这两个任务的优化方向相互干扰, 可以通过适当减少权重值来抑制负迁移减少干扰。所以首先基于梯度方向的一致程度来设置实时调整项, 该过程中仅保留正向迁移, 负向迁移设值为 0:

$$\phi_{ij}^{(t)} = \frac{|\max(\nabla f_i^T \nabla f_j, 0)|}{\sum_{k=1}^n \max(\nabla f_i^T \nabla f_k, 0)} \quad (10)$$

之后根据指数平滑来动态调整权重矩阵:

$$a_{ij}^{(t)} = (1 - \alpha) a_{ij}^{(0)} + \alpha \phi_{ij}^{(t)} \quad (11)$$

其中  $\alpha$  表示实时调整项的控制比例值。

权重更新公式融合了初始相似度与实时梯度方向一致性信息, 其中平滑系数  $\alpha$  控制历史信息与当前信息的权衡。当  $\alpha$  较小时, 算法更依赖初始相似度, 适合任务关系稳定的场景; 当  $\alpha$  较大时, 算法更注重当前梯度方向, 适合任务关系动态变化的复杂环境。调节  $\alpha$ , 算法能够在探索与利用之间取得平衡, 增强对优化过程动态变化的适应性。

在扩散梯度下降框架中, 各下层优化任务被建模为图结构中的节点, 权重矩阵定义了节点间的连接强度与信息传递权重。梯度共享过程可视作信息在任务图上的传播机制: 每个任务不仅基于自身梯度进行更新, 还能接收来自相似任务的梯度修正。这种设计使得处于平坦区域的任务能够借助处于陡峭区域任务的梯度信息加速下降, 而位于局部最优附近的任务则能通过其他任务的梯度信息逃离陷阱。

为确保所有任务能公平、稳定地参与梯度信息共享, 避免出现某些任务被边缘化或过度主导的情况, 本文对调整后的权重矩阵进行双随机化处理。该处理通过迭代归一化行列和, 使权重矩阵同时满足行随机与列随机, 从而保证梯度信息在任务间的均匀扩散与平衡接收, 既促进了知识迁移的效率, 又维持了优化进程的稳定性:

$$a_{ij}^{(k+1/2)} = \frac{a_{ij}^{(k)}}{\sum_{l=1}^n a_{il}^{(k)}} \quad (12)$$

$$a_{ij}^{(k+1)} = \frac{a_{ij}^{(k+1/2)}}{\sum_{l=1}^n a_{ij}^{(k+1/2)}} \quad (13)$$

上述过程进行不断迭代直到满足收敛条件, 使得权重矩阵逼近双随机性。

### 2.3.5. DGD 双模式更新下层决策变量

在双层优化问题中, 下层任务之间可能存在相似性或者冲突。若盲目更新梯度信息, 可能会导致负迁移, 若完全独立梯度下降优化, 则无法利用任务之间的相似性来加速收敛进程。因此本算法引入双模式更新机制, 旨在根据自适应迁移策略结果执行不同的优化策略: 当执行迁移时, 通过不同任务之间的协作学习, 将相似度高的任务的优化经验进行有效传播, 利用扩散梯度下降来加速整体收敛并且提升解的质量; 当执行不迁移时, 仅使用当前任务的局部梯度信息进行简单的梯度下降独立更新下层决策变量。

模式一: 有知识迁移的协同更新(随机数  $r < TP$ )

首先, 每个节点先根据任务之间的相似性更新梯度信息值, 其用于更新的梯度并非仅仅依赖自身梯度, 而是所有任务梯度的组合:

$$\nabla f'_t = \sum_{j=1}^n a_{ij} \nabla f_j \quad (14)$$

其中  $n$  表示任务总数,  $a_{ij}$  表示自适应权重矩阵。通过此加权聚合实现梯度扩散, 相似度高的任务之间梯度共享更强, 从而在平坦区域的任务可借助陡峭区域任务的梯度信息加速下降, 而陷入局部最优的任务也可借助其他任务的梯度信息尝试跳出。

之后每个节点首先基于当前目标函数及其自身的梯度信息执行梯度下降, 然后融合来自其他节点的信息来更新其变量:

$$x_{l,i}^{(t+1)} = x_{l,i}^{(t)} - \eta \cdot \nabla f'_t \quad (15)$$

其中  $\eta$  表示学习率, 控制每次更新的步长。此更新过程是同步且并行的, 所有任务在每次迭代中同时更新。相似任务之间梯度方向是一致的, 这种更新方式既保留了任务独立性, 又通过梯度共享实现了任务间的协同优化。

模式二: 无知识迁移的独立更新(随机数  $r \geq TP$ )

仅使用任务自身的梯度信息进行更新, 避免任务之间的干扰:

$$x_{l,i}^{(t+1)} = x_{l,i}^{(t)} - \eta \cdot \nabla f_i \quad (16)$$

该模式可以根据任务之间的相似程度、优化阶段和梯度一致性来动态选择模式, 减少早期陷入负迁移的风险, 后期强化正迁移, 同时在保证精度的同时减少不必要的梯度共享计算资源。

### 2.3.6. 边界约束处理

在实际优化过程中, 决策变量通常需要满足上下界约束  $x_l \in [\mathbf{lb}, \mathbf{ub}]$ :

$$x_{l,i} \leftarrow \max\left(\mathbf{lb}_i, \min\left(x_{l,i}, \mathbf{ub}_i\right)\right) \quad (17)$$

其中  $\mathbf{lb}$  和  $\mathbf{ub}$  分别是下界向量和上界向量。更新下层决策变量后, 本文采用投影法处理边界约束。

## 2.4. 算法复杂度分析

本文提出的 ATEB-GD 算法的时间复杂度主要从以下三个主要组成部分进行分析:

(1) 上层优化复杂度

上层采用协方差矩阵自适应进化策略(CMA-ES), 其时间复杂度主要取决于种群大小和决策变量维

度。设上层种群大小为  $U_\lambda$ , 上层变量维度为  $m$ , 下层变量维度为  $l_{\text{dim}}$ , 则上层优化的时间复杂度为:

$$O(U_\lambda \cdot (m + l_{\text{dim}})^2) \quad (18)$$

#### (2) 下层梯度计算复杂度

下层优化采用有限差分法估计梯度信息, 避免了传统进化算法在下层的完整优化过程。设下层任务数为  $n$  (等于上层个体数), 每个任务进行  $M$  次梯度采样估计, 则梯度计算的时间复杂度为:

$$O(n \cdot M \cdot l_{\text{dim}}) \quad (19)$$

其中  $M$  通常远远小于  $l_{\text{dim}}$  (在实验中  $M = 3 \sim 5$ )。与传统方法下层独立运行 CMA-ES 的复杂度  $O(n \cdot l_{\text{dim}}^2)$  相比, 本方法显著降低了计算开销, 特别在  $l_{\text{dim}}$  较大时优势明显。

#### (3) 梯度共享机制复杂度

梯度共享机制由三个部分组成: 相似度计算、权重矩阵更新和梯度聚合。具体而言, 相似度计算基于混合距离度量(式(4)~(5))计算  $n$  个任务间的相似度, 复杂度为  $O(n^2 \cdot m)$ ; 权重矩阵更新包含两个过程: 自适应更新(式(10)~(11))和双重随机化处理(式(12)~(13)), 复杂度为  $O(n^2 \cdot l_{\text{dim}})$ ; 梯度聚合与更新过程基于权重矩阵执行梯度共享(式(14))和变量更新(式(15)~(16)), 复杂度为  $O(n^2 \cdot l_{\text{dim}})$ 。总体而言, 梯度共享机制的总复杂度为:

$$O(n^2 \cdot (m + l_{\text{dim}})) \quad (20)$$

#### (4) 对比分析

在计算复杂度方面, 本文提出的 ATEB-GD 算法与现有主流双层优化算法展现出显著差异, 体现了其在设计理念上的创新性。相较于传统方法, ATEB-GD 通过引入梯度共享机制实现了计算效率的实质性提升。

与 TLEA-CMA-ES 算法相比, 两者在上层优化层面均采用 CMA-ES 策略, 具有相近的复杂度特征。然而, 在下层优化处理上存在根本区别。TLEA-CMA-ES 采用并行子优化器架构, 每个下层任务需要独立运行完整的 CMA-ES 优化过程, 导致其下层复杂度达到  $O(n \cdot l_{\text{dim}}^2)$ , 其中  $n$  为任务数量,  $l_{\text{dim}}$  为下层变量维度。这种设计虽便于并行实现, 但在高维问题上会带来较大的计算负担。反观 ATEB-GD, 其下层采用有限差分梯度估计方法, 通过少量采样即可获得梯度信息, 将复杂度降低至  $O(n \cdot M \cdot l_{\text{dim}})$ , 其中  $M$  为常数且通常  $M \ll l_{\text{dim}}$ 。这一改变使得算法在保持优化精度的同时, 显著减少了函数评估次数, 特别适用于计算代价昂贵的实际问题, 在高维问题中消耗成本更低。另外对于下层优化任务, 两者都采用了知识迁移共享策略, TLEA-CMA-ES 复杂度为  $O(n^2 \cdot m)$ , 而 ATEB-GD 实现了实时梯度共享, 虽然引入了  $O(n^2 \cdot (m + l_{\text{dim}}))$  的矩阵运算开销, 但这种设计使算法能够在每次迭代中动态调整优化方向, 避免了各任务重复探索相同区域, 从整体上提升了收敛效率。

BLCMAES 算法作为另一种基于 CMA-ES 的双层优化方法, 其设计重点在于搜索分布共享。该算法同样在上层采用 CMA-ES, 下层则需进行完整优化, 复杂度为  $O(n \cdot l_{\text{dim}}^2)$ 。虽然 BLCMAES 通过上层分布信息指导下层初始化, 在一定程度上减少了迭代次数, 但单次迭代的计算开销并未降低。相比之下, ATEB-GD 不仅减少了单次迭代开销, 还通过梯度共享机制加速了收敛进程, 实现了双重效率提升。

BLEAQ2 算法代表了代理模型优化方向, 采用双映射近似方法构建下层最优值函数映射和理性响应映射。这种方法的优势在于能够显著减少直接函数调用次数, 但其代理模型构建过程本身具有较高的计算复杂度, 达到  $O(n \cdot l_{\text{dim}}^3)$ 。尤其在问题维度较高时, 立方级的复杂度增长会严重制约算法的实用性。ATEB-GD 则通过梯度直接估计避免了复杂的模型构建过程, 保持了线性复杂度增长, 在高维问题中展现出更好的可扩展性。

BOC 算法采用了协作优化, 每个个体对应一个下层任务, 针对下层优化任务建立搜索分布, 根据整个群体的性能表现来更新分布, 同时对不同的下层任务直接共享最优解。总体复杂度包含了上层分解特征值复杂度以及下层优化复杂度, 达到  $O\left((n^u + n^l)^3 + MaxGen^l \cdot \lambda \cdot (n^l)^3\right)$ 。由于整体算法复杂度识别立方级, 所以当下层维度较大时, 立方复杂度增长速度远高于 ATEB-GD, 这也可见 BOC 更适合解决中低维的问题。

NBLEA 算法是一种嵌套式的双层优化算法, 运用进化算法求解上层问题, 对于下层每个个体独立完成基于种群的优化, 任务之间没有进行相互协作与共享。该算法复杂度为  $O(n \cdot D^2 + n \cdot G \cdot \lambda_i \cdot d^2)$ 。由于在下层优化过程中对每个个体都独立进行优化, 计算代价远远大于 ATEB-GD。并且任务之间没有利用相似性进行知识的迁移, 收敛速度慢。

由此可见, 本文所提出的 ATEB-GD 在算法复杂度上具有显著优势: 首先, 将下层优化复杂度从平方级降低至线性级, 显著减少了基础计算开销; 其次, 通过实时梯度共享实现了早期收敛加速, 减少了达到满意解所需的迭代次数; 最后, 线性复杂度增长特性使算法在高维问题上仍能保持高效运行。这些优势在实验中得到验证, ATEB-GD 在相同计算预算下获得了更优的收敛精度, 证明了其复杂度设计的合理性和有效性。ATEB-GD 的计算效率提升使其更适合处理大规模、高维度的复杂优化问题。在计算资源有限或函数评估代价高昂的场景下, 这种效率优势将更加明显。虽然算法引入了额外的矩阵运算开销, 但这一代价换来了整体优化效率的显著提升。

### 3. 实验结果

在本节中, 我们选取了目前广泛使用的 SMD 和 TP 测试问题集来对 DGD-CMAES 进行实验, 并将该算法与目前先进的算法在计算精度与速度上进行比较, 包括 TLEA-CMA-ES, BLCMAES, BLEAQ2, 以证明其在处理双层优化问题时的高效性。在整个实验过程中, 每个测试实例均采用相同的实验设置, 并且所有实例由四种算法分别独立求解 19 次以消除随机性影响。为客观评估算法性能, 采用中位数 (Median) 和四分位距 (IQR) 值作为主要评价指标。为了简洁起见, 本文将上下层的求解准确性分别表示为 UAcc 和 LAcc, 函数评估次数分别表示为 FEU 和 FEL。为了方便, 最佳性能将会在表中以加粗字体突出显示。

#### 3.1. 测试问题

在本节中我们总共测试了 22 个基准函数来分析该算法的有效性, 其中第一类测试集来自 SMD 测试套, 包含 12 个优化问题 (SMD1-SMD12), 第二类测试集来自 TP 测试套, 包含 10 个基准问题 (TP1-TP10)。其中 SMD 测试问题集主要用于评估分析算法在处理复杂优化任务上的效能, 该测试集具备非凸性的目标函数, 多模态的搜索空间以及长谷地的地形特征。此外一些 SMD 问题集设置了双层目标相互矛盾的特殊结构, 大大提升了整体优化的挑战性。TP 测试问题集融合了很多经典的优化测试问题, 具有明确的数学特征, 主要优化线性或二次型的目标函数。

#### 3.2. 比较算法

本节考虑了五种具有代表性的成熟双层优化算法, 包括 TLEA-CMA-ES, BLCMAES, BLEAQ2, BOC 以及 NBLEA 作为对比基准来综合评估本文所提出的 ATEB-GD 的效能, 这些算法的内容思想如下所示:

(1) 基于迁移学习的进化算法 (TLEA-CMA-ES): 该算法采用标准 CMAES 作为基础优化器, 构建了多层次优化器之间的协同机制, 通过并行的子优化器之间的知识共享来传递搜索经验, 实现任务之间的有效迁移, 进而提升整体的优化效率。

(2) 双层协方差矩阵自适应进化策略 (BLCMAES): 该算法在上下层中均采用了标准 CMAES 进行优

化, 设计了搜索分布共享机制, 通过上层搜索分布来推导下层优化的初始搜索分布, 实现不同层级间的知识迁移。

(3) 基于二次近似的双层进化算法(BLEAQ2): 该算法属于代理模型优化方向, 采用双映射近似策略, 主要构建了低层最优函数值映射以及低层有理反应映射, 这两个映射被整合到基于层级结构的双层优化框架中。

(4) 基于下层任务协作的双层优化算法(BOC): 该算法是一种具有低级优化任务之间协作的双层元启发式算法。每一代进化出一个群体来协作解决所有上层解决方案的下层优化任务, 进而在一次运行中解决所有的下层优化任务, 从而提升整体效率。

(5) 嵌套式双层进化算法(NBLEA): 该算法是一种传统的双层嵌套式 EA 算法, 对每个上层变量个体独立搜索下层优化, 为 SMD 问题建立了性能基准。

### 3.3. 比较实验性能

(1) SMD 问题实验: 表 1 比较了基于梯度扩散的自适应迁移进化双层优化算法(ATEB-GD)与其他五种成熟算法在一组 10-dim 的 SMD 测试套件中的表现。由于这 12 个测试实例上下层问题中的变量数量有限, 所以相对简单。其中 SMD1 至 SMD8 实例是无约束的, 并且 SMD1-SMD6 实例上的上层问题均为单模态性, 这也解释了六种算法在 SMD1 至 SMD6 测试实例上的整体优化精度高于在 SMD9 至 SMD12 上的结果。另外, SMD7-SMD8 引入了更强的非线性和多模态特性, 容易陷入局部最优, 对算法的全局搜索能力和局部求精能力提出了更高的要求。这可以通过自适应知识共享机制进行缓解。根据上下层的优化进度动态调整相关参数, 进而适应该实例的非线性地形。SMD9-SMD12 代表了最具备挑战性的双层优化问题, 其包括了高度非凸、强非线性耦合等。在 ATEB-GD 中, 扩散梯度下降的多任务协同机制在高维度欺骗性问题中发挥了重要的作用。早期保守: 避免过早进行梯度信息共享导致误导, 中期积极: 充分利用可用信息, 后期谨慎: 避免破坏已得到的最优解的三层策略更能满足这种复杂问题的优化需求。

ATEB-GD 和 TLEA-CMAES 都采用了并行的方式处理多个下层优化任务, 并且有相同的基础构架模块, 所以他们会部分性能相同的情况。不同于 TLEA-CMAES 完全依赖差分进化来进行下层优化; 基本固定迁移概率, 仅仅在特定迭代条件下触发迁移。本文引入了梯度信息, 通过多任务之间的梯度共享提供精确的搜索方向, 加快了整体收敛进程。同时实现了自适应的迁移概率, 基于上下层优化状态来动态调整避免负迁移情况。因此可以观察到 ATEB-GD 的整体性能优于 BL-CMAES。

**Table 1.** Median and interquartile range of UAcc and LAcc of ATEB-GD and five algorithms on 10-D SMD test suite  
**表 1.** ATEB-GD 与五种算法在 10 维 SMD 测试集上的 UAcc 和 LAcc 中位数及四分位距

| 10 维 SMD |      | ATEB-GD                       | TLEA-CMA-ES            | BL-CMA-ES              | BOC                    | BLEAQ2                 | NBLEA                  |
|----------|------|-------------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| SMD1     | UAcc | <b>1.00E-06</b><br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 5.27E-03<br>(4.48E-03) | 1.94E-03<br>(1.78E-03) |
|          | LAcc | <b>1.00E-06</b><br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 2.58E-03<br>(4.67E-03) | 1.28E-03<br>(1.30E-03) |
| SMD2     | UAcc | <b>1.00E-06</b><br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 2.27E-03<br>(5.38E-03) | 2.58E-03<br>(2.28E-03) |
|          | LAcc | <b>1.00E-06</b><br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.15E-03<br>(2.57E-03) | 1.52E-03<br>(1.82E-03) |
| SMD3     | UAcc | <b>1.00E-06</b><br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 3.24E-03<br>(6.37E-03) | 2.15E-02<br>(3.31E-02) |

续表

|       |      |                                      |                                      |                        |                        |                                      |                        |
|-------|------|--------------------------------------|--------------------------------------|------------------------|------------------------|--------------------------------------|------------------------|
|       | LAcc | <b>1.00E-06</b><br><b>(1.05E-06)</b> | 1.02E-06<br>(0.00E+00)               | 1.26E-06<br>(0.00E+00) | 1.11E-06<br>(0.00E+00) | 1.86E-03<br>(4.28E-03)               | 1.26E-02<br>(2.89E-02) |
| SMD4  | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 4.21E-04<br>(2.23E-03)               | 1.75E-03<br>(1.87E-03) |
|       | LAcc | <b>1.00E-06</b><br><b>(2.21E-06)</b> | 3.46E-06<br>(2.72E-06)               | 3.71E-06<br>(4.28E-06) | 3.61E-06<br>(1.05E-06) | 2.87E-03<br>(3.01E-03)               | 1.02E-03<br>(9.27E-04) |
| SMD5  | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 6.68E-03<br>(5.47E-03)               | 1.57E-03<br>(1.87E-03) |
|       | LAcc | <b>1.96E-06</b><br><b>(2.13E-06)</b> | 2.67E-06<br>(2.96E-06)               | 2.61E-06<br>(3.47E-06) | 2.71E-06<br>(3.14E-06) | 4.76E-03<br>(3.88E-03)               | 8.05E-04<br>(9.73E-04) |
| SMD6  | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 3.51E-03<br>(8.91E-03)               | 9.34E-03<br>(1.31E-02) |
|       | LAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 6.17E-06<br>(3.27E-04)               | 2.48E-03<br>(4.94E-03) |
| SMD7  | UAcc | <b>9.82E-02</b><br><b>(1.46E-01)</b> | 9.82E-02<br>(1.46E-01)               | 1.23E-01<br>(4.78E-02) | 9.85E-02<br>(2.21E-02) | 9.87E-02<br>(7.39E-02)               | 9.85E-02<br>(9.66E-02) |
|       | LAcc | 1.25E+02<br>(4.21E+02)               | 2.13E+02<br>(5.68E+02)               | 2.64E+02<br>(1.84E+02) | 2.25E+02<br>(5.31E+02) | <b>1.23E-02</b><br><b>(2.71E-03)</b> | 2.17E+02<br>(2.21E+02) |
| SMD8  | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 8.21E-05<br>(1.86E-04)               | 6.93E-06<br>(1.36E-04) | 4.18E-04<br>(8.43E-04) | 6.83E-03<br>(6.21E-03)               | 3.92E-02<br>(1.00E-01) |
|       | LAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 5.67E-05<br>(0.00E+00)               | 2.87E-05<br>(6.28E-05) | 6.84E-05<br>(1.26E-04) | 5.23E-03<br>(4.21E-03)               | 1.84E-02<br>(1.94E-02) |
| SMD9  | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 4.27E-01<br>(4.68E-01)               | 2.90E+00<br>(2.14E+00) |
|       | LAcc | 1.17E-06<br>(2.46E-06)               | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 3.16E-01<br>(2.32E-01)               | 6.19E+00<br>(4.97E+00) |
| SMD10 | UAcc | <b>1.46E+01</b><br><b>(1.36E+00)</b> | 1.54E+01<br>(1.43E+00)               | 1.63E+01<br>(2.12E-02) | 1.62E+01<br>(1.43E+00) | 1.71E+01<br>(1.18E+00)               | 2.77E+01<br>(1.59E+01) |
|       | LAcc | <b>1.64E-01</b><br><b>(1.48E+00)</b> | 1.37E+00<br>(1.63E+00)               | 1.69E+01<br>(1.73E-01) | 1.65E+01<br>(9.21E-01) | 1.70E+01<br>(8.58E-01)               | 6.21E+00<br>(2.65E+01) |
| SMD11 | UAcc | <b>1.13E-02</b><br><b>(1.87E-02)</b> | 1.43E-02<br>(1.98E-02)               | 2.72E-02<br>(4.98E+01) | 5.77E+01<br>(4.75E+01) | 1.23E+00<br>(1.86E+00)               | 2.42E+01<br>(4.74E+01) |
|       | LAcc | <b>1.01E-02</b><br><b>(2.41E-03)</b> | 1.12E-02<br>(2.05E-02)               | 9.83E-03<br>(4.82E+01) | 5.61E+01<br>(4.71E+01) | 1.10E-02<br>(4.95E-02)               | 2.51E+01<br>(4.78E+01) |
| SMD12 | UAcc | <b>1.61E+01</b><br><b>(1.12E-01)</b> | 1.61E+01<br>(1.20E-01)               | 1.61E+01<br>(1.50E+00) | 1.61E+01<br>(2.91E-01) | 1.74E+01<br>(1.48E+00)               | 1.98E+01<br>(3.21E+00) |
|       | LAcc | <b>1.80E+01</b><br><b>(1.01E+00)</b> | 1.80E+01<br>(1.15E+00)               | 1.82E+01<br>(1.26E+00) | 1.83E+01<br>(1.75E-01) | 1.89E+01<br>(2.37E+00)               | 1.85E+01<br>(1.09E+00) |

表 2 展示了六种算法在 10 维的 SMD 测试实例中实现的上下层函数评估次数(FEU 与 FEL)的对比结果。ATEB-GD 在绝大多数的测试实例中均取得了最优的评估效率,其上下层函数调用次数普遍低于其他算法,展现出较高的稳定性。BL-CMA-ES 与 BOC 整体上表现中等,但是缺乏优势,其中 BL-CMA-ES 可能将过多的资源用于了无效区域的探索,未能有效利用任务之间的知识经验。BOC 算法对问题结构要

求较高, 普适性较低。BLEAQ2 虽然在某些特定的 SMD 问题中表现优异, 但是整体波动性较大, 主要来源于参数敏感性, 导致整体的稳健性不足。而 NBLEA 相对于其他算法表现最差, 不适合当前的测试场景。相比之下, ATEB-GD 梯度信息指导搜索, 减少随机探索, 下层评估效率更高。另一方面, ATEB-GD 和 TLEA-CMA-ES 都基于 CMA-ES, 并且都采用了知识共享的策略, 因此在部分测试问题上他们的 FE<sub>U</sub> 和 FE<sub>L</sub> 表现相近具有一定的理论依据。总体而言, 我们依旧可以观察到 ATEB-GD 在效率方面优于其他算法, 具有更强的收敛能力与计算经济性。

**Table 2.** Median and interquartile range of FE<sub>U</sub> and FE<sub>L</sub> of ATEB-GD and five algorithms on 10-D SMD test instances  
**表 2.** ATEB-GD 与五种算法在 10 维 SMD 测试实例上的 FE<sub>U</sub> 和 FE<sub>L</sub> 中位数及四分位距

| 10 维 SMD |                 | ATEB-GD                              | TLEA-CMA-ES                          | BL-CMA-ES                            | BOC                    | BLEAQ2                 | NBLEA                  |
|----------|-----------------|--------------------------------------|--------------------------------------|--------------------------------------|------------------------|------------------------|------------------------|
| SMD1     | FE <sub>U</sub> | <b>6.81E+02</b><br><b>(5.00E+01)</b> | 7.64E+02<br>(5.70E+01)               | 7.68E+02<br>(7.13E+02)               | 8.12E+02<br>(5.33E+01) | 2.19E+03<br>(1.77E+03) | 1.68E+03<br>(4.65E+02) |
|          | FE <sub>L</sub> | <b>5.88E+04</b><br><b>(3.21E+03)</b> | 5.89E+04<br>(5.93E+03)               | 8.10E+04<br>(7.54E+04)               | 6.12E+04<br>(3.33E+03) | 1.67E+05<br>(1.46E+05) | 1.41E+06<br>(3.52E+05) |
| SMD2     | FE <sub>U</sub> | 7.48E+02<br>(6.51E+01)               | 7.52E+02<br>(6.81E+01)               | <b>7.23E+02</b><br><b>(6.68E+02)</b> | 8.01E+02<br>(4.38E+01) | 2.27E+03<br>(1.92E+03) | 1.81E+03<br>(3.89E+02) |
|          | FE <sub>L</sub> | <b>5.79E+04</b><br><b>(4.23E+03)</b> | 5.87E+04<br>(3.75E+03)               | 8.15E+04<br>(7.41E+04)               | 6.38E+04<br>(2.51E+03) | 1.40E+05<br>(1.22E+05) | 1.51E+06<br>(3.05E+05) |
| SMD3     | FE <sub>U</sub> | <b>7.01E+02</b><br><b>(5.20E+01)</b> | 7.49E+02<br>(5.72E+01)               | 7.75E+02<br>(7.17E+02)               | 7.46E+02<br>(3.71E+01) | 2.15E+03<br>(1.82E+03) | 1.77E+03<br>(1.11E+03) |
|          | FE <sub>L</sub> | 6.21E+04<br>(4.32E+03)               | <b>5.94E+04</b><br><b>(4.58E+03)</b> | 8.55E+04<br>(7.57E+04)               | 7.74E+04<br>(1.24E+03) | 1.65E+05<br>(1.48E+05) | 1.64E+06<br>(9.49E+05) |
| SMD4     | FE <sub>U</sub> | 7.04E+02<br>(7.81E+01)               | <b>6.90E+02</b><br><b>(8.45E+01)</b> | 8.07E+02<br>(7.19E+02)               | 8.43E+02<br>(6.20E+01) | 2.71E+03<br>(1.68E+03) | 1.71E+03<br>(3.49E+02) |
|          | FE <sub>L</sub> | <b>7.01E+04</b><br><b>(4.27E+03)</b> | 7.22E+04<br>(4.97E+03)               | 8.58E+04<br>(7.58E+04)               | 8.21E+04<br>(5.21E+03) | 1.61E+05<br>(1.41E+05) | 1.31E+06<br>(2.75E+05) |
| SMD5     | FE <sub>U</sub> | <b>7.68E+02</b><br><b>(1.32E+02)</b> | 9.56E+02<br>(1.15E+02)               | 1.10E+03<br>(8.96E+02)               | 1.23E+03<br>(6.43E+01) | 2.82E+03<br>(2.28E+03) | 1.64E+03<br>(3.15E+02) |
|          | FE <sub>L</sub> | <b>7.65E+04</b><br><b>(5.81E+03)</b> | 8.21E+04<br>(5.84E+03)               | 9.53E+04<br>(8.43E+04)               | 8.05E+04<br>(6.01E+03) | 1.81E+05<br>(1.45E+05) | 1.33E+05<br>(2.25E+05) |
| SMD6     | FE <sub>U</sub> | <b>8.48E+02</b><br><b>(2.21E+02)</b> | 8.68E+02<br>(9.63E+01)               | 9.51E+02<br>(8.45E+02)               | 1.04E+03<br>(7.45E+01) | 1.85E+03<br>(1.78E+03) | 2.14E+03<br>(9.08E+02) |
|          | FE <sub>L</sub> | <b>6.46E+04</b><br><b>(6.79E+03)</b> | 6.65E+04<br>(6.82E+03)               | 9.08E+04<br>(8.02E+04)               | 7.01E+04<br>(6.21E+03) | 8.91E+04<br>(7.85E+03) | 1.11E+05<br>(4.49E+04) |
| SMD7     | FE <sub>U</sub> | <b>1.25E+03</b><br><b>(1.38E+02)</b> | 1.31E+03<br>(1.62E+02)               | 1.35E+03<br>(1.12E+03)               | 1.33E+03<br>(1.58E+02) | 2.88E+03<br>(2.49E+03) | 1.73E+03<br>(4.49E+02) |
|          | FE <sub>L</sub> | <b>6.85E+04</b><br><b>(7.51E+03)</b> | 7.10E+04<br>(6.56E+03)               | 8.61E+04<br>(7.53E+04)               | 7.51E+04<br>(7.48E+03) | 1.84E+05<br>(1.71E+05) | 1.21E+06<br>(3.41E+05) |
| SMD8     | FE <sub>U</sub> | <b>3.50E+03</b><br><b>(4.95E+02)</b> | 3.51E+03<br>(5.55E+02)               | 3.51E+03<br>(3.23E+03)               | 3.52E+03<br>(4.85E+03) | 5.17E+03<br>(3.73E+03) | 3.75E+03<br>(5.62E+03) |
|          | FE <sub>L</sub> | <b>1.88E+05</b><br><b>(2.41E+04)</b> | 2.24E+05<br>(2.89E+04)               | 2.68E+05<br>(2.39E+05)               | 2.34E+05<br>(3.01E+04) | 3.43E+05<br>(3.03E+05) | 2.34E+06<br>(1.37E+06) |
| SMD9     | FE <sub>U</sub> | <b>7.80E+02</b><br><b>(7.66E+01)</b> | 8.65E+02<br>(4.91E+01)               | 2.05E+03<br>(9.38E+02)               | 9.23E+02<br>(9.08E+01) | 4.63E+03<br>(2.50E+03) | 2.21E+03<br>(1.65E+03) |

续表

|       |        |                                      |                        |                                      |                        |                        |                        |
|-------|--------|--------------------------------------|------------------------|--------------------------------------|------------------------|------------------------|------------------------|
|       | $FE_L$ | <b>6.08E+04</b><br><b>(3.81E+03)</b> | 6.43E+04<br>(4.05E+03) | 1.36E+05<br>(8.28E+04)               | 6.58E+04<br>(4.12E+03) | 2.71E+05<br>(1.63E+05) | 2.81E+06<br>(1.35E+06) |
| SMD10 | $FE_U$ | <b>2.28E+03</b><br><b>(1.15E+01)</b> | 3.52E+03<br>(4.00E+00) | 3.52E+03<br>(3.49E+03)               | 3.05E+03<br>(3.85E+03) | 6.12E+03<br>(4.58E+03) | 5.26E+03<br>(2.75E+03) |
|       | $FE_L$ | <b>2.01E+05</b><br><b>(7.45E+03)</b> | 2.15E+05<br>(8.26E+03) | 2.28E+05<br>(2.07E+05)               | 2.26E+05<br>(3.17E+04) | 2.45E+05<br>(1.89E+05) | 1.65E+06<br>(2.55E+06) |
| SMD11 | $FE_U$ | <b>3.49E+03</b><br><b>(5.30E+00)</b> | 3.51E+03<br>(5.50E+00) | 3.53E+03<br>(3.14E+03)               | 3.51E+03<br>(6.25E+00) | 7.78E+03<br>(6.00E+03) | 4.61E+03<br>(1.95E+03) |
|       | $FE_L$ | 2.22E+05<br>(4.07E+04)               | 2.21E+05<br>(4.15E+04) | <b>2.15E+05</b><br><b>(1.78E+05)</b> | 2.20E+05<br>(1.62E+03) | 2.64E+05<br>(2.25E+05) | 9.93E+05<br>(8.15E+05) |
| SMD12 | $FE_U$ | <b>3.08E+03</b><br><b>(1.19E+01)</b> | 3.51E+03<br>(6.00E+00) | 3.51E+03<br>(3.51E+03)               | 3.21E+03<br>(1.03E+02) | 4.21E+03<br>(3.31E+03) | 4.01E+03<br>(2.09E+03) |
|       | $FE_L$ | <b>1.85E+05</b><br><b>(1.20E+04)</b> | 2.12E+05<br>(8.82E+03) | 2.26E+05<br>(2.17E+05)               | 2.21E+05<br>(3.21E+04) | 2.33E+05<br>(1.51E+04) | 2.46E+06<br>(2.16E+06) |

表 3 展示了六种对比算法在 12 个 20-dim 的 SMD 测试套件中的模拟结果, 随着上下层变量的数量的显著增加, 这 12 个测试实例中上下层函数评估的复杂性也随之增加, 比上述 10-dim 测试实例更具挑战性。即使微小的上下层优化偏差也会不断累积, 进而导致上层种群搜索偏离最优值。所以相对于 10-dim 的实验结果, 该维度的表现性能呈现了下降趋势。其中 ATEB-GD、TLEA-CMA-ES、BL-CMA-ES 以及 BOC 的运算精度都远高于 BLEAQ2 与 NBLEA, 是因为前四个算法运用的 CMA-ES 都可以根据优化状态自适应地调整协方差值等参数。在整体对比上, ATEB-GD 总体上表现更佳, BOC 和 TLEA-CMA-ES 运用了知识迁移的思想, 均表现出较强的竞争力, BL-CMA-ES 表现中等, BLEAQ2 次之, NBLEA 表现最差。这个发现进一步证明了基于扩散梯度下降算法的下层搜索的重要功能。另外, 我们可以明确观察到, SMD9-SMD12 的上下层问题涉及了复杂的约束条件, 降低了优化过程中的确定性。因此实验算法在 SMD9-SMD12 实例上的性能差异比在 SMD1-SMD8 实例上更为显著。

**Table 3.** Median and interquartile range of UAcc and LAcc of ATEB-GD and five algorithms on the 20-D SMD test suite  
**表 3.** ATEB-GD 与五种算法在 20 维 SMD 测试集上的 UAcc 和 LAcc 中位数及四分位距

| 20 维 SMD |      | ATEB-GD                              | TLEA-CMA-ES            | BL-CMA-ES              | BOC                    | BLEAQ2                 | NBLEA                  |
|----------|------|--------------------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| SMD1     | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 7.34E-03<br>(2.86E-03) | 8.05E+01<br>(2.14E+01) |
|          | LAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 2.48E-03<br>(3.75E-03) | 4.63E+01<br>(3.42E+01) |
| SMD2     | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 8.36E-03<br>(3.01E-03) | 4.65E+01<br>(1.25E+01) |
|          | LAcc | <b>1.36E-06</b><br><b>(1.51E-06)</b> | 1.48E-06<br>(1.67E-06) | 1.61E-06<br>(2.91E-06) | 1.51E-06<br>(1.59E-06) | 5.84E-03<br>(1.24E-02) | 2.40E+01<br>(1.56E+01) |
| SMD3     | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 9.48E-03<br>(9.02E-01) | 8.68E+01<br>(1.81E+01) |
|          | LAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 6.89E-03<br>(9.50E-01) | 4.20E+01<br>(2.03E+01) |

续表

|       |      |                                      |                                      |                                      |                                      |                        |                        |
|-------|------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|------------------------|------------------------|
| SMD4  | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 9.91E-06<br>(1.21E-05)               | 1.05E-06<br>(0.00E+00)               | 1.81E-02<br>(1.98E-02) | 1.63E+01<br>(1.22E+01) |
|       | LAcc | 1.18E-05<br>(1.06E-05)               | 1.34E-05<br>(1.49E-05)               | <b>1.00E-06</b><br><b>(2.48E-06)</b> | 1.21E-05<br>(1.35E-05)               | 3.62E-02<br>(1.70E-01) | 1.38E+01<br>(1.23E+01) |
| SMD5  | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00)               | 9.08E-03<br>(3.64E-03) | 9.96E+01<br>(3.75E+01) |
|       | LAcc | 1.00E-06<br>(4.39E-06)               | 1.00E-06<br>(2.58E-06)               | 1.37E-05<br>(2.49E-05)               | <b>1.00E-06</b><br><b>(1.12E+00)</b> | 3.78E-03<br>(7.09E-03) | 4.52E+01<br>(2.15E+01) |
| SMD6  | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(1.21E-02) | 9.35E-01<br>(9.66E-01) |
|       | LAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | <b>1.00E-06</b><br><b>(0.00E+00)</b> | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00) | 9.12E-02<br>(1.66E-01) |
| SMD7  | UAcc | <b>1.00E-06</b><br><b>(9.82E-02)</b> | 2.85E-02<br>(1.40E-01)               | 2.47E-01<br>(2.48E-02)               | 2.21E-06<br>(3.95E-02)               | 5.42E-01<br>(4.12E-01) | 4.51E+00<br>(2.75E+00) |
|       | LAcc | <b>2.99E-06</b><br><b>(2.44E+02)</b> | 5.03E+02<br>(5.01E+02)               | 4.89E+02<br>(8.67E+02)               | 4.95E+02<br>(5.21E+02)               | 4.91E+02<br>(3.76E+02) | 7.54E+02<br>(7.22E+02) |
| SMD8  | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 2.38E-03<br>(2.60E-03)               | 3.21E-03<br>(5.78E-03)               | 2.21E-03<br>(3.38E-03)               | 4.61E+00<br>(4.01E+00) | 3.67E+01<br>(1.25E+01) |
|       | LAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.09E-03<br>(1.43E-03)               | 2.66E-03<br>(5.13E-03)               | 2.57E-03<br>(2.58E-03)               | 5.41E+00<br>(1.03E+01) | 1.69E+01<br>(9.29E+00) |
| SMD9  | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(4.54E-01)               | 4.54E-01<br>(3.18E-01)               | 1.00E-06<br>(2.25E-01)               | 4.75E+00<br>(4.98E+00) | 5.58E+01<br>(2.65E+01) |
|       | LAcc | <b>3.79E-06</b><br><b>(9.42E-06)</b> | 4.95E-06<br>(1.97E-01)               | 2.48E-01<br>(1.43E-01)               | 5.05E-06<br>(1.58E-01)               | 1.01E+01<br>(1.25E+01) | 2.91E+01<br>(3.09E+01) |
| SMD10 | UAcc | <b>1.50E+01</b><br><b>(0.00E+00)</b> | 1.54E+01<br>(3.01E+00)               | 1.54E+01<br>(1.70E+00)               | 1.56E+01<br>(1.80E+00)               | 3.52E+01<br>(3.39E+01) | 9.98E+01<br>(6.10E+01) |
|       | LAcc | <b>1.60E+01</b><br><b>(1.21E+00)</b> | 1.82E+01<br>(2.23E+00)               | 1.81E+01<br>(1.20E+00)               | 1.85E+01<br>(2.06E+00)               | 5.75E+01<br>(8.87E+01) | 5.59E+01<br>(3.64E+01) |
| SMD11 | UAcc | <b>1.26E-02</b><br><b>(1.32E-02)</b> | 8.83E-01<br>(9.55E-01)               | 9.34E+01<br>(7.39E+01)               | 5.89E-01<br>(8.75E-01)               | 1.77E+02<br>(1.14E+02) | 2.53E+02<br>(9.53E+01) |
|       | LAcc | <b>2.25E-02</b><br><b>(3.67E-02)</b> | 7.76E-02<br>(1.58E-01)               | 9.03E+01<br>(7.46E+01)               | 8.71E-02<br>(1.62E-01)               | 1.73E+02<br>(1.04E+02) | 2.48E+02<br>(9.59E+01) |
| SMD12 | UAcc | <b>1.53E+01</b><br><b>(2.89E+00)</b> | 1.61E+01<br>(1.46E+01)               | 1.58E+01<br>(1.28E+00)               | 1.59E+01<br>(1.58E+00)               | 2.11E+01<br>(6.89E+00) | 1.06E+02<br>(3.64E+01) |
|       | LAcc | <b>1.96E+01</b><br><b>(8.21E+00)</b> | 1.97E+01<br>(9.30E+00)               | 1.99E+01<br>(1.28E+00)               | 2.05E+01<br>(6.21E+00)               | 3.31E+01<br>(3.34E+00) | 7.48E+01<br>(2.69E+01) |

(2) TP 问题实验: 实验结果如表 4 所示。TP 测试集包含经典优化问题, 其目标函数和约束条件多为二次型或者线性, 复杂性与非线性程度次于 SMD 测试集, 因此, 运用如二次近似的确定性方法来优化更为简单, 这也解释了为什么 BLEAQ2 在 TP1-TP5 测试实例中取得了与 ATEB-GD 相当甚至更优的结果。两种基于知识共享的算法 ATEB-GD 与 TLEA-CMAES 有效识别并利用了问题中的线性和非线性特征, 进而在大多数的 TP 问题中都表现出色。尤其在 TP1、TP4、TP6、TP9、TP10 等问题中, ATEB-GD 的上下层精度均达到近似零误差, 体现了其梯度信息共享与自适应迁移机制在处理结构良好问题时的效率与

稳定性。值得注意的是, 在 TP3 与 TP5 等部分问题上, BLEAQ2 凭借其精确的二次近似模型, 在收敛精度上优于其他算法。

**Table 4.** Median and interquartile range of UAcc and LAcc of ATEB-GD and five comparison algorithms on the TP test suite  
**表 4.** ATEB-GD 及五种对比算法在 TP 测试集上的 UAcc 和 LAcc 中位数与四分位距

|      |      | ATEB-GD                              | TLEA-CMA-ES                          | BL-CMA-ES              | BOC                    | BLEAQ2                               | NBLEA                                |
|------|------|--------------------------------------|--------------------------------------|------------------------|------------------------|--------------------------------------|--------------------------------------|
| TP1  | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(4.23E-05)               | 1.00E-06<br>(4.23E-05) | 1.00E-06<br>(4.23E-05) | 2.63E-06<br>(2.04E-06)               | 4.28E-01<br>(4.12E-01)               |
|      | LAcc | 2.18E-05<br>(1.23E-04)               | 1.08E-05<br>(9.25E-05)               | 2.81E-06<br>(1.70E-06) | 2.35E-05<br>(1.68E-05) | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 8.21E-01<br>(7.51E-01)               |
| TP2  | UAcc | <b>1.00E-06</b><br><b>(2.21E-05)</b> | 1.00E-06<br>(1.07E-04)               | 2.10E-05<br>(6.23E-04) | 1.51E-06<br>(1.61E-04) | 6.48E-03<br>(5.00E+00)               | 1.00E-06<br>(3.38E-03)               |
|      | LAcc | <b>2.56E-05</b><br><b>(6.75E-04)</b> | 1.10E-04<br>(2.33E-03)               | 1.56E-03<br>(1.05E-02) | 1.31E-04<br>(5.21E-03) | 1.00E+02<br>(0.00E+00)               | 1.00E+02<br>(2.43E-06)               |
| TP3  | UAcc | 2.90E-02<br>(3.17E-03)               | 3.20E-02<br>(2.42E-02)               | 4.16E-02<br>(3.42E-02) | 3.42E-02<br>(3.05E-02) | <b>1.09E-05</b><br><b>(5.00E-06)</b> | 4.95E-03<br>(1.28E-02)               |
|      | LAcc | 1.10E-02<br>(3.01E-02)               | 1.36E-02<br>(2.28E-02)               | 1.05E-02<br>(5.52E-02) | 1.24E-02<br>(2.31E-02) | <b>2.50E-05</b><br><b>(0.00E+00)</b> | 1.33E-03<br>(2.95E-03)               |
| TP4  | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.24E-03<br>(1.70E-02)               | 1.65E-02<br>(6.15E-02)               |
|      | LAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 2.41E-04<br>(1.99E-03)               | 3.21E-03<br>(1.21E-02)               |
| TP5  | UAcc | 2.31E-01<br>(1.13E+00)               | 2.23E-01<br>(1.15E+00)               | 1.01E+00<br>(1.12E+00) | 2.45E-01<br>(1.21E+00) | 3.45E-03<br>(6.41E-03)               | <b>3.95E-04</b><br><b>(6.86E-04)</b> |
|      | LAcc | 2.81E-01<br>(5.78E+00)               | 2.76E-01<br>(5.66E+00)               | 1.08E+00<br>(5.51E+00) | 2.91E-01<br>(5.55E+00) | <b>1.75E-02</b><br><b>(2.42E-02)</b> | 1.77E-02<br>(3.12E-03)               |
| TP6  | UAcc | <b>6.81E-05</b><br><b>(0.00E+00)</b> | 6.95E-05<br>(0.00E+00)               | 7.00E-05<br>(0.00E+00) | 7.01E-05<br>(1.01E+00) | 6.58E-03<br>(6.48E-03)               | 8.15E-05<br>(2.51E-05)               |
|      | LAcc | <b>5.48E-05</b><br><b>(8.72E-06)</b> | 5.56E-05<br>(9.76E-06)               | 5.30E-02<br>(2.93E-05) | 2.81E-04<br>(9.21E-06) | 2.29E-02<br>(2.30E-02)               | 6.11E-05<br>(5.21E-05)               |
| TP7  | UAcc | 2.13E-04<br>(5.31E-05)               | <b>2.02E-04</b><br><b>(5.05E-05)</b> | 6.68E-04<br>(5.25E-05) | 2.15E-04<br>(5.21E-05) | 2.04E-03<br>(4.80E-03)               | 1.79E-03<br>(8.82E-04)               |
|      | LAcc | 2.13E-04<br>(5.32E-05)               | <b>2.02E-04</b><br><b>(5.06E-05)</b> | 6.68E-04<br>(5.25E-05) | 2.15E-04<br>(5.18E-05) | 2.04E-03<br>(4.80E-03)               | 1.79E-03<br>(8.82E-04)               |
| TP8  | UAcc | <b>1.00E-06</b><br><b>(6.81E-05)</b> | 2.05E-06<br>(7.41E-05)               | 1.00E-06<br>(1.21E-03) | 1.58E-06<br>(6.71E-05) | 1.00E-06<br>(9.11E-03)               | 1.00E-06<br>(4.21E-03)               |
|      | LAcc | <b>2.41E-04</b><br><b>(3.08E-04)</b> | 5.31E-04<br>(4.38E-03)               | 2.91E-04<br>(3.45E+00) | 5.62E-04<br>(2.25E-03) | 1.00E+02<br>(5.00E+01)               | 1.00E+02<br>(3.38E+01)               |
| TP9  | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 5.35E-06<br>(3.35E-06)               | 2.88E-05<br>(1.05E-05)               |
|      | LAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00)               |
| TP10 | UAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 3.68E-03<br>(6.05E-03)               | 3.36E+00<br>(1.78E+00)               |
|      | LAcc | <b>1.00E-06</b><br><b>(0.00E+00)</b> | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00) | 1.00E-06<br>(0.00E+00)               | 1.00E-06<br>(5.31E-06)               |

由此可以得到, 当问题具有强二次特性且代理模型拟合良好时, 基于近似映射的方法仍具有一定优势。然而, BLEAQ2 在 TP2、TP8 等复杂问题上表现显著下滑, 说明其模型在非线性增强或约束复杂时容易失效, 鲁棒性不足。相比之下, ATEB-GD 在绝大多数 TP 实例中保持稳定且优越的表现, 其关键在于: 下层采用扩散梯度下降(DGD)实现多任务间梯度信息的有向共享, 结合混合距离度量与双进度调整的自适应迁移机制, 既能有效识别任务相似性, 又能动态控制知识迁移时机, 从而在保持收敛精度的同时避免了负迁移。此外, 该算法不依赖于问题的强凸性或代理模型的可拟合性, 因而在从线性到非线性、从低维到高维的多种 TP 问题上均表现出良好的泛化能力。

### 3.4. 数据检验分析

为了全面评估 ATEB-GD 与各对比算法在多次独立运行中的性能差异, 并确保结论的统计有效性, 本节进行了非参数统计检验。所有统计检验的显著性水平均设定为  $\alpha = 0.05$ , 并在结果中标注相应的  $p$  值或显著性标记。

#### (1) SMD 测试集统计检验分析

Wilcoxon 检验显示: 10 维 SMD 上, ATEB-GD 在 SMD7~SMD12 显著优于 TLEA-CMA-ES 和 BLCMAES ( $p < 0.05$ ), 在 SMD1~SMD6 保持竞争力; 相比 BLEAQ2 和 NBLEA, 在所有 12 个实例上均取得显著提升( $p < 0.01$ )。20 维 SMD 上, ATEB-GD 在多数实例的 UAcc 和 LAcc 显著优于五种对比算法( $p < 0.05$ ), 尤其在 SMD9-SMD12 等复杂约束问题上优势突出。

Friedman 检验(表 5)表明, ATEB-GD 在上下层精度的平均排名均列首位, 整体显著优于对比算法( $p < 0.05$ ), 验证了所提策略在复杂双层优化中的有效性与鲁棒性。

**Table 5.** Average rankings and significance results of the Friedman test

**表 5.** Friedman 检验平均秩次及显著性结果

| 算法          | 上层精度平均秩次 | 下层精度平均秩次 |
|-------------|----------|----------|
| ATEB-GD     | 1.42     | 1.38     |
| TLEA-CMA-ES | 2.35     | 2.41     |
| BLCMAES     | 2.87     | 2.93     |
| BOC         | 3.21     | 3.15     |
| BLEAQ2      | 4.58     | 4.62     |
| NBLEA       | 5.57     | 5.51     |
| $p$ 值       | <0.001   | <0.001   |

#### (2) TP 测试集统计检验分析

Wilcoxon 检验显示: 在 TP1、TP4、TP6、TP9、TP10 等线性/二次型问题上, ATEB-GD 与 TLEA-CMA-ES、BLCMAES 无显著差异( $p > 0.05$ ), 三者均显著优于 BOC、BLEAQ2 和 NBLEA ( $p < 0.05$ ); 在 TP3、TP5 等强二次问题上, BLEAQ2 与 ATEB-GD 性能相当( $p > 0.05$ ); 但在 TP2、TP8 等非线性/复杂约束问题上, ATEB-GD 显著优于 BLEAQ2 ( $p < 0.05$ ), 展现更强泛化能力。

Friedman 检验证实: ATEB-GD 在全部 TP 实例上的平均秩次显著低于其他算法( $p < 0.05$ ), 综合性能最优。这表明其梯度共享与自适应迁移机制在广泛问题类型上实现了更稳定可靠的优化性能。

**Table 6.**  $p$ -values of the Wilcoxon signed-rank test between ATEB-GD and the comparison algorithms  
**表 6.** ATEB-GD 与对比算法的 Wilcoxon 符号秩检验  $p$  值

| 测试集        | TLEA-CMA-ES  | BLCMAES      | BOC          | BLEAQ2       | NBLEA            |
|------------|--------------|--------------|--------------|--------------|------------------|
| SMD10 UAcc | <b>0.032</b> | <b>0.028</b> | <b>0.041</b> | <b>0.002</b> | <b>0.001</b>     |
| SMD10 LAcc | 0.067        | <b>0.043</b> | 0.058        | <b>0.003</b> | <b>0.001</b>     |
| SMD20 UAcc | <b>0.024</b> | <b>0.031</b> | <b>0.039</b> | <b>0.001</b> | <b>&lt;0.001</b> |
| SMD20 LAcc | <b>0.037</b> | 0.052        | <b>0.044</b> | <b>0.002</b> | <b>&lt;0.001</b> |
| TP UAcc    | 0.084        | 0.071        | <b>0.048</b> | 0.063        | <b>0.004</b>     |
| TP LAcc    | 0.092        | 0.083        | 0.056        | 0.071        | <b>0.005</b>     |

#### 4. 结论

在本文中, 我们设计了一个用于双层优化的 DGD 框架, 该算法通过将扩散梯度下降与协方差矩阵自适应进化策略相结合, 有效解决了双层优化中多任务协同的难题。在 SMD 与 TP 两类测试集上的实验结果表明, 该算法在收敛精度、鲁棒性和计算效率方面总体上优于对比算法, 展现出良好的泛化能力与可扩展性, 为复杂双层优化问题的求解提供了有效的新方法。随着多任务进化在进化计算领域的不断发展, 未来将进一步研究下层优化过程之间的相互作用, 以实现更有效的知识转移。

#### 参考文献

- [1] Brotcorne, L., Labbé, M., Marcotte, P. and Savard, G. (2001) A Bilevel Model for Toll Optimization on a Multicommodity Transportation Network. *Transportation Science*, **35**, 345-358. <https://doi.org/10.1287/trsc.35.4.345.10433>
- [2] Karahalios, A., Tayur, S., Tenneti, A., Pashapour, A., Salman, F.S. and Yıldız, B. (2025) A Quantum-Inspired Bilevel Optimization Algorithm for the First Responder Network Design Problem. *INFORMS Journal on Computing*, **37**, 172-188. <https://doi.org/10.1287/ijoc.2024.0574>
- [3] Du, J., Li, X., Yu, L., Dan, R. and Zhou, J. (2017) Multi-Depot Vehicle Routing Problem for Hazardous Materials Transportation: A Fuzzy Bilevel Programming. *Information Sciences*, **399**, 201-218. <https://doi.org/10.1016/j.ins.2017.02.011>
- [4] Zhu, J., Wang, D., Xing, J., Qin, S., Tao, G., Chen, P., et al. (2025) Bilevel Optimization for Provisioning Heterogeneous Traffic in Deterministic Networks. *IEEE Transactions on Network and Service Management*, **22**, 3295-3308. <https://doi.org/10.1109/tnsm.2025.3570284>
- [5] Shi, M., Xie, P., Yao, L., Guo, H., Vasquez, J.C., Guerrero, J.M., et al. (2025) Electricity-Hydrogen Coupled Energy Storage Bilevel Optimization for Offshore Wind-Powered Zero-Carbon Port Microgrids Considering Multiple Uncertainties. *Applied Energy*, **401**, Article ID: 126672. <https://doi.org/10.1016/j.apenergy.2025.126672>
- [6] Sinha, A., Lu, Z., Deb, K. and Malo, P. (2019) Bilevel Optimization Based on Iterative Approximation of Multiple Mappings. *Journal of Heuristics*, **26**, 151-185. <https://doi.org/10.1007/s10732-019-09426-9>
- [7] Huang, P., Zhang, Q. and Wang, Y. (2023) Bilevel Optimization via Collaborations among Lower-Level Optimization Tasks. *IEEE Transactions on Evolutionary Computation*, **27**, 1837-1850. <https://doi.org/10.1109/tevc.2022.3233409>
- [8] He, X., Zhou, Y. and Chen, Z. (2019) Evolutionary Bilevel Optimization Based on Covariance Matrix Adaptation. *IEEE Transactions on Evolutionary Computation*, **23**, 258-272. <https://doi.org/10.1109/tevc.2018.2849000>
- [9] Chen, A.C.H. (2024) The Optimization of Hyperparameter Based on Mathematics for Gradient Descent Algorithm. 2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT), Volume 1, 434-437. <https://doi.org/10.1109/iccpct61902.2024.10673060>
- [10] von Stackelberg, H., Peacock, A.T., Schneider, E. and Hutchison, T.W. (1953) The Theory of the Market Economy. *Economica*, **20**, Article No. 384. <https://doi.org/10.2307/2550609>
- [11] Chen, L., Liu, H., Tan, K.C. and Li, K. (2022) Transfer Learning-Based Parallel Evolutionary Algorithm Framework for Bilevel Optimization. *IEEE Transactions on Evolutionary Computation*, **26**, 115-129. <https://doi.org/10.1109/tevc.2021.3095313>
- [12] Vlaski, S. and Sayed, A.H. (2021) Distributed Learning in Non-Convex Environments—Part I: Agreement at a Linear

- Rate. *IEEE Transactions on Signal Processing*, **69**, 1242-1256. <https://doi.org/10.1109/tsp.2021.3050858>
- [13] Liu, Z., Li, G., Zhang, H., Liang, Z. and Zhu, Z. (2024) Multifactorial Evolutionary Algorithm Based on Diffusion Gradient Descent. *IEEE Transactions on Cybernetics*, **54**, 4267-4279. <https://doi.org/10.1109/tcyb.2023.3270904>
- [14] Gupta, A., Mańdziuk, J. and Ong, Y. (2015) Evolutionary Multitasking in Bi-Level Optimization. *Complex & Intelligent Systems*, **1**, 83-95. <https://doi.org/10.1007/s40747-016-0011-y>
- [15] Sinha, A., Malo, P. and Deb, K. (2014) Test Problem Construction for Single-Objective Bilevel Optimization. *Evolutionary Computation*, **22**, 439-477. [https://doi.org/10.1162/evco\\_a\\_00116](https://doi.org/10.1162/evco_a_00116)
- [16] Sinha, A., Malo, P. and Deb, K. (2017) Evolutionary Algorithm for Bilevel Optimization Using Approximations of the Lower Level Optimal Solution Mapping. *European Journal of Operational Research*, **257**, 395-411. <https://doi.org/10.1016/j.ejor.2016.08.027>