MF-EDA：通过连接隶属函数调整高斯分布估计算法中的分布数量

Jun Zha ng，*IEEE* 会士，钱通、丁华、余伟杰、刘晓芳和 Jian-Yu L i

arXiv:2611.30436v1 [cs.NE] 23 Nov 2026

摘要—**分布估计算法（EDA）通常从单个高斯分布中抽样，其峰值位于大多数被选个体所在的位置。在具有两个漏斗的景观上，最佳解并不位于此处：较宽的漏斗提供了大多数被选个体，而包含全局最优解的更深漏斗则会在早期消失。本文借鉴神经模糊系统构建模型的方式，利用高斯隶属函数及其隶属度连接构建搜索分布。每一代中，所提出的 MF-EDA 会在被选个体的主方向中寻找双峰分布的方向，根据贝叶斯信息准则和分离检验，在该方向上用一个或两个隶属函数描述个体；随后将每个隶属函数与其他方向上的局部高斯分布相连接，并从归一化并集中抽样。无论其下包含多少个体，隶属函数的高度始终为 1，因此检测到的漏斗会根据其宽度获得样本。双区域选择模型表明，按个体数量比例分配采样份额（如高斯混合分布）会加剧更深漏斗在早期的不利处境，而并集则不会。使用一个隶属函数时，该算法就是其单高斯基线；这些步骤不会增加适应度评估次数，但会使每代的计算速度慢 8 至 17 倍。在 36 种双漏斗设置中，该成分使秩加权高斯估计算法的成功运行次数从 1800 次中的 805 次提升至 1712 次。在 30 维 Lunacek 双漏斗上，该算法的成功次数从 50 次运行中的 4 次提升至 49 次，而 AMaLGaM 从 0 次提升至 47 次；但在单漏斗函数和十个多峰 BBOB 函数上的结果没有变化。在标准双漏斗上，重启 CMA-ES 的评估次数仍然更少；当全局漏斗远窄于另一个漏斗时，该成分不再有帮助。**

索引词—**估计分布算法、隶属函数、神经模糊系统、高斯混合模型、漏斗景观、多峰优化、模型选择。**

I. 引言

AN 分布估计算法（EDA）通过概率模型取代遗传算法中的交叉和变异：它从当前种群中选出较优个体，为这些个体拟合一个分布，再从该分布中采样生成下一代种群 [1]–[3]。对于连续变量，该分布几乎总是单个高斯分布，从 EMNA 的最大似然高斯分布 [2] 到 IDEA [4], AMaLGaM [5], [6] 的改进模型 [7]，以及近期的后续研究 [7]，近期关于高斯 EDA 的研究仍在对单个高斯分布进行改进 [8]。单个高斯分布只有一个峰值，而峰值位于大多数被选个体所在的位置。大多数被选个体所在的位置，并不总是最佳解所在的位置。

双漏斗景观使这种差异变得十分明显。它包含两个相距甚远的宽广吸引盆地（漏斗）；其中一个包含全局最优解，另一个则更浅但更宽。在运行初期，更宽的漏斗提供了大多数被选个体，拟合得到的高斯分布向其移动，而更深的漏斗在每一代中接收到的样本越来越少，直至一个也不剩。Lunacek 等 [9] 构造了这类函数，并表明在困难变体上，采用默认种群的 CMA-ES [10]在少于 10% 的运行次数中找到全局最优解。双漏斗并非测试函数的人为产物。38 原子 Lennard-Jones团簇的能量景观包含一个通向全局最小值的狭窄深漏斗，以及一个通向次低最小值的更宽漏斗；而弛豫过程通常会进入更宽的漏斗 [11]。近期的景观研究证实，小型全局漏斗是最能预测 CMA-ES 和差分进化失败的特征之一 [12]，，并且具有快速全局信息流的算法会困陷于某个漏斗中 [13]。

高斯混合模型这一自然补救方法，已在估计分布算法中应用了二十五年 [14]–[16]，，而小生境估计分布算法则为每个簇运行一个模型 [17]–[20]。混合分布虽能减轻这一缺陷，却保留了其成因。它的各个成分是通过在所有 D 个维度上进行聚类得到的；随着D 增大，这需要每个簇包含许多个体，并且每个成分获得的样本数与其混合权重成正比，也就是与它所能解释的被选个体数成正比。因此，个体数量在早期较少的更深漏斗仍然只能获得少量样本，而它能否存续取决于其早期劣势持续多久。第 III 节表明，在采用这种个体数量 权重时，更深漏斗的优势比会在每一代乘以其当前的选择劣势。

因此，对于这个问题，现有补救方法有三个共同缺点：它们在所有 D 个维度上构建成分，这需要每个成分包含许多个体；它们按各成分的个体数量分配样本，导致在运行早期，个体数量较少但深度更深的漏斗得不到足够样本；此外，它们还会预先固定成分数量、通过距离阈值设定成分数量，或为此付出额外适应度评估的代价。

本文采用了不同于神经模糊

本研究得到韩国国家研究基金会项目 RS-2025-00555463 和 RS-2025-25456394 的资助。（通信作者：张军。）张军隶属于大韩民国的汉阳大学 ERICA 校区；以及台湾的南开大学、浙江师范大学和朝阳科技大学（电子邮箱：junzhang@ieee.org）。钱通、丁华、刘晓芳和李建宇就职于中国天津的南开大学。余伟杰就职于中国广州的中山大学。

系统 [21], [22]的构造方式。自适应网络模糊推理系统在输入端使用少量高斯隶属函数来表示复杂的输入–输出关系，并通过归一化触发强度将每个隶属函数分别连接到其自身的后件。隶属函数与混合成分有一个在此处至关重要的区别：它是正态的，也就是说，无论其下方有多少数据点，其高度始终为 1。我们以这种方式构建估计分布算法的搜索分布。在每一代中，算法（i）在被选个体的主方向中搜索其呈现最明显双峰特征的方向；（ii）在该方向上用一个或两个高斯隶属函数描述这些个体，数量由贝叶斯信息准则（BIC）[23]结合分离检验决定；（iii）根据个体的隶属度，将每个隶属函数与其他所有方向上的一个局部高斯分布连接起来；（iv）从这些隶属函数的归一化并集中进行抽样。使用一个隶属函数时，该模型就是普通的全协方差高斯分布，而该算法就是其单高斯基线。使用两个隶属函数时，数据揭示的每个漏斗都会以完整高度表示，其获得的样本份额取决于其宽度，而不是其中包含的个体数量。因此，这种构造解决了三个缺点：分量是在一个方向上拟合的，因此只需几十个被选个体即可；它们的采样份额不取决于个体数量；其数量仅根据被选个体决定，无需用户设定阈值，也无需进行适应度评估。在运行初期，当种群覆盖两个漏斗时，分布包含两个部分；之后，当其中一个漏斗变为空时，分布又恢复为单个高斯分布。我们将该算法称为 MF-EDA（隶属函数连接高斯 EDA）。

贡献如下。

1) 一种具有自适应分布数量的估计分布算法。搜索分布由一个或两个作用于数据选择方向的高斯隶属函数组成，并通过隶属度与局部高斯分布连接，再以归一化并集的形式进行抽样。分布数量在每一代根据被选个体重新确定，无需进行适应度评估，并保持全协方差高斯估计算法分布每代成本的 $O(nD^2 + D^3)$ 阶数量级不变；其常数更大，第 V-G 节对其进行了测量。

2) 解释通过隶属函数连接为何不同于混合。双区域选择模型表明，在若干世代中，人数权重会累加更深漏斗在早期所处的不利地位，而正态隶属函数的并集则不会如此（命题 1）；并集的采样密度以及每个隶属函数的样本份额均以闭式形式给出（命题2）。运行轨迹与该模型一致。

3）一个成分，而非新的优化器。我们将相同的步骤加入两种单高斯 EDA——一种秩加权高斯 EDA 和 AMaLGaM——并分别报告结果。4）支持双向主张的证据：在双漏斗景观上，更多运行次数达到了全局最优解；而在相同的适应度评估次数下，单漏斗景观上的结果没有变化。在 36 种双漏斗设置中，该成分使基线的成功运行次数从 1800 次中的 805 次提升至 1712 次；在 $D = 30$ 的 Lunacek 双漏斗上，它使秩加权 EDA 的成功次数从 50 次运行中的 4 次提升至 49 次，并使 AMaLGaM 从 0 次提升至 47 次；在单漏斗函数上，成功次数没有变化，评估次数则有几个百分点的增减；而在十个多峰 BBOB 函数 f_{15}–f_{24} 上，没有显著差异（ 0 次胜出、 160 次平局、 0 次落败）。

5) 对局限性的简要说明：一个比另一个漏斗窄得多的全局漏斗、三个或更多不在同一直线上的漏斗，以及与重启 CMA-ES 的比较；在标准双漏斗上，后者所需的评估次数更少。

第二节陈述问题并回顾相关工作。第三节分析人数计数模型为何会丢失更深漏斗。第四节介绍该方法。第五节报告实验结果，第六节讨论局限性、计算时间和未来工作，第七节作出结论。

II. 问题与相关工作

A. 问题

我们在适应度评估预算内最小化 $f\colon\ [l,u]^D \to \mathbb{R}$，其中 D 是变量数量；在所有实验中，$[l,u] = [-5,5]$。算法的成本是其使用的适应度评估次数。当算法找到 $\mathbf{x}$ 且 $f(\mathbf{x}) - f^* \le 10^{-8}$ 时，一次运行成功，其中 f^* 是全局最小值。漏斗是这样一个区域：其中的局部最小值按照一定方式排列，使得较优的局部最小值更靠近漏斗底部[9]；双漏斗则包含两个这样的区域。Lunacek 等人 [9]将最简单的一个写作

$$g_2(\mathbf{x}) = \min\Big\{\sum_{i=1}^{D}(x_i-\mu_1)^2,\ w_2 D + s_2\sum_{i=1}^{D}(x_i-\mu_2)^2\Big\}, \quad (1)$$

其中，第一个漏斗在 $\mu_1\mathbf{1}$ 处包含全局最优解，第二个漏斗抬高了 $w_2 D$，并具有宽度因子 s_2。当 $s_2 < 1$ 时，第二个漏斗比第一个更平坦、更宽，并且更大比例的搜索空间会流入其中。我们将第一个漏斗称为更深 漏斗，将第二个称为更宽 漏斗。

B. 高斯 EDA

EMNA [2]通过最大似然法为被选个体拟合单一高斯分布。Kern等。 [24] 将此类模型与进化策略进行了比较，发现最大似然高斯分布的方差未受到控制，并且逐代缩小；在他们的实验中，将被选个体聚类为高斯混合模型并未带来明显优势。IDEA [4]及其无参数后继算法 AMaLGaM [5], [6] 加入了自适应方差缩放（AVS） [25]、标准差比率触发器（SDR） [26] 和预期均值偏移（AMS），以阻止最大似然高斯分布过早收敛；AMaLGaM仅根据 D 确定其种群规模，其作者指出，该算法旨在用于单峰景观 [5]。其高斯抽样是 RV-GOMEA 的基础，后者通过适应度差异学习哪些

变量彼此依赖 [27] ，如今还可针对重叠依赖关系使用条件模型 [28],；Scholman等。 [7] 最近引入了分布的增量估计。这些发展改进了单一高斯分布的依赖结构。自然进化策略 [29]沿着期望适应度的自然梯度更新搜索分布的参数；其作者指出，原则上可以使用高斯混合模型，并且默认设置在某些高度多峰函数上表现不佳，因此建议采用更大的种群和重启。Wang等。 [8]通过使用早期世代的档案库、将采样均值移向多个领先解以及触发式收缩来修改单一高斯分布；他们指出，单峰高斯分布难以应对多峰问题，并在不使用混合分布的情况下加以解决。Soloviev等。 [30] 完全摒弃高斯假设，并让结构搜索在贝叶斯网络的每个节点上决定使用线性高斯模型还是核密度估计；他们报告称，与高斯网络相比，该模型的学习速度更慢。Larra˜naga 和 Bielza [3] 的综述将“让模型复杂度在运行过程中随景观变化”列为连续分布估计算法的发展方向之一。MF-EDA 以最小的步长实现了这一点：从一个分布变为两个分布。

CMA-ES [10], [31] 从单个高斯分布中采样，其协方差和步长沿进化路径进行调整，是连续优化中常用的基准方法。在多峰函数上，它依靠更大的种群，或通过增大种群规模进行重启（IPOP） [32],，或采用大、小种群交替进行的两种重启策略（BIPOP） [33]。在具有 Rastrigin 型局部最优的 BBOB 双漏斗问题上， f_{24}BIPOP 在 $D = 20$ [33] 的 15 次试验中有 3 次达到目标；Loshchilov 等人 [34] 还通过在每次重启时减小初始步长，将速度提高了两到三倍。在运行过程中调整种群规模，是重启策略的一种替代方案；对于结构较弱的多峰函数，重启策略中的步长修正可能会失效 [35]。当前的一个库文档指出，使用默认种群规模的 CMA-ES 往往会陷入 40 维 Rastrigin 函数的局部最优 [36]；重启策略仍是当前模块化框架的组成模块 [37],，而 Nomura 等人 [38] 表明，降低学习率的效果与扩大种群规模相似。He等人 [39] 的进化策略中的混合分布有着另一种用途：其所有分量都以同一点为中心，共同以较低成本近似高维空间中的单个高斯分布。

C. 混合分布与小生境模型

Bosman 和 Thierens [14] 通过聚类和期望最大化（EM），将混合分布引入连续IDEA； [15] 提出的混合IDEA采用随机化的领导者聚类，因此簇的数量由用户设定的距离阈值决定。Lu 和 Yao [16] 采用竞争惩罚竞争学习对被选个体进行聚类，无需预先设定簇的数量。CAMaLGaM [17]是现有方法中与我们最接近的一种：每一代都会沿最近优树构建簇层次结构，并根据密度与适应度之间的秩相关为每一层评分，然后针对得分最高层级中的每个簇运行一个 AMaLGaM；若所有得分均为负，则退回到单个簇，因此分量数量可自适应调整，也可以为一个。HillVallEA [18]以山谷检验取代层次结构；该检验需要对两个个体之间线段上的点进行适应度评估，并为每个生态位运行一个带重启的局部优化器，目标是找到所有全局最优。Yang 等人[19]提出的多峰EDA将种群划分为若干生态位，每代随机抽取生态位大小，并从一元高斯分布或柯西模型中为每个生态位采样，同时进行局部搜索。Liang等人 [20]根据最近几代被选集合估计每个生态位的协方差，并依据适应度和相对距离进行聚类。MAPS [40] 检测有前景区域并保留多个子模型。在进化策略中，RS-CMSA [41] 运行多个子种群，每个子种群都有自己的协方差矩阵；它会拒绝落入存档最小值周围禁忌区域，以及适应度更高的子种群中心周围禁忌区域的后代。子种群数量按期望找到的最小值数量成比例设定，而其后继算法仍是近期竞赛参赛方法的基础优化器 [42]。关于小生境处理的两篇近期综述 [43], [44] 对尚待解决的问题意见一致：大多数方法需要设定难以确定的小生境半径或簇的数量；较新的方法尝试从数据中估计这些参数，而在约十个以上维度中的表现基本未经测试。在估计分布算法之外，CMA-ES的小生境处理变体仍不断出现 [45], ；Le Minh 等人 [46] 通过带退火的自然梯度步优化高斯混合搜索分布。在他们的方法中，分量数量预先设定，且高于预期的模式数量；混合权重通过学习得到，评估次数随分量数量线性增长。Landais 等人 [47] 针对组合优化问题运行多个在参数空间中彼此排斥的EDA智能体，并指出智能体过多会稀释预算。Zhang 等人 [48] 从扩散模型的去噪步骤推导出一种进化算法，其中每个个体都会朝其邻居的适应度加权平均值移动，从而使子种群持续存在于不同峰值附近；该算法主要在二维函数上进行了测试。

这些方法与 MF-EDA 有三方面的不同。它们在完整的 D 维空间中构建成分；根据从分配给该成分的个体中学习到的权重进行成分采样，或为每个簇分配各自的子种群；并且是在寻找多个最优解的任务上进行评估的，且大多未报告单成分情形下的结果。

D. 隶属函数与估计分布算法

模糊集由隶属函数 [49]定义。在自适应网络模糊推理系统中， [21], [22] 输入上的高斯隶属函数产生激活强度，归一化触发强度为规则的后件赋予权重，而隶属函数和后件的参数则通过数据进行学习。此类系统的近期研究仍然手动固定规则数量，并研究相邻隶属函数应具有多强的可区分性 [50]；第四节-D中的分离检验在这里发挥着类似作用。我们团队早期的一种估计分布算法 [51]使用了模糊直方图，

将每个变量的模糊划分作为其单变量模型；它不会将隶属函数合并为一个分布，也不会选择其数量。在我们全文阅读的研究中，没有发现任何一种估计分布算法的搜索分布是由隶属函数的并集构成，并通过隶属度与局部模型相连接；我们的检索并不全面。

E. 景观分析中的漏斗

近期的景观研究直接考察漏斗。Blackwell[13] 根据障碍树生成双漏斗景观，并发现下坡行走者进入某个漏斗的概率与其表面积成正比；同时，全局最优粒子群和差分进化到达最优盆地的频率，远低于那些减缓信息流动的方法。Mizuta *et al.* [12] 生成具有多样局部最优网络的景观，并报告称，全局漏斗的大小是与 CMA-ES 和差分进化成功率相关性最强的特征之一。Thomson *et al.* [52] 记录 CMA-ES 和差分进化的运行次数在何处停滞。如今，基准测试生成器会有意控制此类全局结构：向 BBOB 函数中加入球面分量，可以逐步降低其难度 [53], ；还可以自动生成具有预设全局最优与局部最优之间对比度，或预设盆地大小的问题 [54]。根据点样本可以预测某个函数是否具有漏斗结构；在 BBOB 函数上，宏平均F1分数高于 0.99[55], ，这与本文的前提一致，即个体样本携带了足够的信息，能够区分一个漏斗和两个漏斗。这些研究描述了单模型算法何时会失败；MF-EDA 则改变了模型。

III. 人数计数模型为何会失去更深漏斗

本节通过一个包含两个区域的模型，说明当搜索分布遵循被选个体数量时会出现什么问题，以及正态隶属函数的并集会带来哪些变化。该模型有意设计得较为简单；第五节-D 将其与真实运行次数的轨迹进行比较。

A. 双区域选择模型

设景观中有两个漏斗， A（更深，包含全局最优解）和 B （更宽）。在第 t 代中，搜索分布将 a_t 比例的 N 新个体置于 A中，并将 $1-a_t$ 置于 B中。设 $q_{A,t}$ 表示在 A中采样的一个个体属于 $n = \lfloor \tau N \rfloor$ 被选个体的概率，且$q_{B,t}$ 表示在 B中的相同概率。它们的比率 $r_t = q_{A,t}/q_{B,t}$ 衡量选择过程中更深漏斗当前的优势（$r_t > 1$）或劣势（$r_t < 1$）。对于较大的 N ， A在被选个体中的比例为

$$\sigma_{t+1} = \frac{a_t q_{A,t}}{a_t q_{A,t} + (1-a_t) q_{B,t}}, \qquad \frac{\sigma_{t+1}}{1-\sigma_{t+1}} = \frac{a_t}{1-a_t}\, r_t. \tag{2}$$

比率 r_t 并非在整个运行过程中保持不变。当样本分布较为分散时，较平坦的漏斗具有更好的适应度值；一旦样本收缩至漏斗底部附近，较深的漏斗则更优。对于（1），如果两个漏斗的样本都以各自的中心为中心分布，且逐坐标方差为 v，那么在 A 中平均适应度为 Dv ，在 B 中为 $w_2 D + s_2 D v$ ，因此只要 $s_2 < 1$，当 $v > w_2/(1-s_2)$ 时， B 的平均表现更好。当 $w_2 = 0.5$ 和 $s_2 = 0.87$ 时，阈值为标准差 1。96，而 $[-5, 5]$ 上的均匀初始种群每个坐标的标准差为 2。89。因此，一次运行可能会以这样一个阶段开始：$r_t < 1$，而较深的漏斗必须挺过这一阶段；第五节-D 中的轨迹展示了这样的阶段。

B. 人数权重会加剧劣势

通过最大似然拟合的高斯混合模型，会为每个分量赋予一个混合权重，该权重等于该分量所解释的被选个体占比，并按此权重抽取分量： $a_{t+1} = \sigma_{t+1}$。单一高斯分布没有显式权重，但其均值和协方差是对被选个体的平均，因此它在某个区域内分配的概率质量也取决于该区域的人数。

命题 1. 在双区域模型中，令 $o_t = a_t/(1-a_t)$表示在 A 中抽样的优势比。
(a) 采用人数权重时， $a_{t+1} = \sigma_{t+1}$，优势比遵循 $o_T = o_0 \prod_{t=0}^{T-1} r_t$。如果在前 T_0 代中 $r_t \le r < 1$，那么 $o_{T_0} \le o_0\, r^{T_0}$。(b) 如果采样份额不取决于人数，$a_t = b_t$ ，且 b_t 仅为两个分量形状的函数，那么在每一代中，被选个体中 A 的优势比为 $\sigma_{t+1}/(1-\sigma_{t+1}) = \frac{b_t}{1-b_t}\, r_t$ ，不涉及对先前各代的连乘。——

证明。(a) 将 $a_{t+1} = \sigma_{t+1}$ 代入 (2)： $o_{t+1} = o_t r_t$，并迭代。(b) 将 $a_t = b_t$ 代入 (2)。 □

(a) 部分表明，人数计数模型会将持续数代的不利影响转化为几何衰减。在有限总体中，这种衰减最终会导致灭绝：一旦落入 A 的样本少于约一个 N，漏斗就会消失，之后无论 r_t 增长到多大，它都不会再回来。阶段 $r_t < 1$ 会持续多久取决于模型。单一高斯分布不会使其样本集中在 A 上；落入 A 的样本来自一个中心位于两个漏斗之间或 B 上的高斯分布的尾部，它们不会向 A 的底部收缩，而且在 A 灭绝之前， r_t 始终小于一。一个在 A 上含有一个成分的混合分布，会让 A 的样本收缩，因此这一阶段较短；若阶段在灭绝前结束，混合分布就会保留更深漏斗，否则便会失去它。一旦 $r_t > 1$，同样的累积效应便会有利于 A，因此在人数组权重成功的运行中，它们是速度更快的模型。

（b）部分说明，由成分形状决定权重的模型，每代只承担一次劣势 r_t ，而不是累积劣势。只要模型能检测到更深的漏斗，无论该阶段持续多久，它都会持续从中抽样；当r_t 升至1以上时，更深的漏斗便会占据被选集合。隶属函数就是这样一种模型：它呈正态分布，高度为1，其下方面积决定了它在样本中所占的比率，并且仅与其宽度成正比（命题2）。

IV. 提出的方法

A. 概述

MF-EDA 是一种高斯 EDA，它在高斯分布估计与抽样之间插入了三个步骤：搜索被选个体呈双峰分布的方向；决定该方向上使用一个还是两个隶属函数；以及将隶属函数与局部高斯分布结合。图1展示了在二维双漏斗上的构建过程。这些步骤遵循自适应网络模糊推理系统的层级 [21]：输入上的隶属函数（第1层）、激活强度及其归一化（第2层和第3层）、每个隶属函数对应一个后件（第4层），以及后件的组合（第5层）。这里，输入是所选方向上的坐标，隶属函数的后件是在其他方向上的局部高斯分布，而后件的组合就是对新个体进行抽样。

B. 选择与单一高斯分布

每一代都会评估一个包含 N 个个体的种群。选择其中最优的 $n = \lfloor \tau N \rfloor$ 个，其中 $\tau \in (0,1)$ 为选择比例，$\lfloor \cdot \rfloor$ 向下取整。按适应度排序后的被选个体 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 获得如下对数秩权重

$$w_i = \frac{\ln(n+\tfrac{1}{2}) - \ln i}{\sum_{k=1}^{n} \big(\ln(n+\tfrac{1}{2}) - \ln k\big)}, \qquad i = 1, \dots, n. \tag{3}$$

加权均值和协方差分别为 $\mathbf{m} = \sum_i w_i \mathbf{x}_i$，以及 $\mathbf{C} = \sum_i w_i (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^{\top}$，而 $\mathbf{C} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\top}$ 是特征分解，其特征向量为 $\mathbf{v}_1, \dots, \mathbf{v}_D$，特征值为 $\lambda_1, \dots, \lambda_D$。个体 i 的特征坐标为 $\mathbf{z}_i = \mathbf{V}^{\top}(\mathbf{x}_i - \mathbf{m})$，其分量为 z_{il}。对各分量相互独立的 $\mathbf{z}$ 进行抽样，令其服从 $z_l \sim \mathcal{N}(0, \alpha \lambda_l)$，并设置 $\mathbf{x} = \mathbf{m} + \mathbf{V}\mathbf{z}$，即得到单高斯基线，我们称之为 CUR-G；$\alpha \in [1, 10]$ 是自适应方差缩放：若上一代最优适应度有所提升，则除以 0.9，否则乘以 0.9；此外，将迄今为止找到的最佳个体复制到下一代种群中。

C. 双峰方向搜索

如果被选个体来自两个漏斗，那么沿连接这两个漏斗的直线方向呈双峰分布。当两个群组相距较远时，该直线接近某个主方向，但不一定是第一个主方向；而当其中一个群组较小时，这条直线甚至完全不会反映在方差中。因此，MF-EDA 会测试每个主方向。在方向 l 上，使用 $y_i = z_{il}$，它拟合 J 个高斯分布，

$$p_J(y) = \sum_{j=1}^{J} \hat{\pi}_j \, \mathcal{N}(y; \hat{c}_j, \hat{s}_j^2), \tag{4}$$

对于 $J = 1$ 和 $J = 2$，其中 $\hat{\pi}_j$、$\hat{c}_j$ 和 $\hat{s}_j^2$ 分别是成分 j 的权重、中心和方差，而 $\mathcal{N}(y; c, s^2)$ 是均值为 c、方差为 s^2 的正态密度。参数通过加权期望最大化（期望最大化）拟合。 [56]，其中每个个体按其权重 w_i 计入。期望最大化迭代 40 次；中心从 $y_1, \dots, y_n$ 的阶数为 $(j - \tfrac{1}{2})/J$ 的加权分位数开始，并将方差维持在 $10^{-6}\lambda_l$ 以上，以免任何成分塌缩。然后，算法计算

$$\mathrm{BIC}_J(l) = -2 n_{\mathrm{eff}} \sum_{i=1}^{n} w_i \ln p_J(y_i) + (3J - 1) \ln n_{\mathrm{eff}}, \tag{5}$$

使用有效样本量 $n_{\mathrm{eff}} = 1/\sum_i w_i^2$ 替代样本量，并采用加权版 BIC [23]。施瓦茨的推导假设观测值相互独立且具有相同权重；对于不等权重观测，以 n_{eff} 替代样本量是一种近似。在此处，该得分作为模型选择规则，其阈值将在第 V-E 节中进行检验。双峰方向是增益最大的方向，

$$l^{*} = \arg\max_{l} G_l, \qquad G_l = \mathrm{BIC}_1(l) - \mathrm{BIC}_2(l). \tag{6}$$

该搜索仅使用被选个体，不会产生任何适应度评估开销。

D. 一个或两个隶属函数

当方向 l^{*} 满足以下两个条件时，模型使用两个隶属函数；否则使用一个：

$$G_{l^{*}} > 2 \ln D, \tag{7}$$

$$|\hat{c}_1 - \hat{c}_2| > 2\sqrt{(\hat{s}_1^2 + \hat{s}_2^2)/2}. \tag{8}$$

条件 (7) 是考虑搜索范围后的贝叶斯信息准则判定：l^{*} 是 D 个方向中表现最好的方向，而 D 个相互独立、服从卡方分布的增益的最大值约按 $2 \ln D$ 增长；各方向并非相互独立，因此这一修正是启发式方法，第 V-E 节将检验其敏感性。条件 (8) 要求两个拟合高斯分布之间的距离大于其均方根标准差的两倍。对于方差和权重相等的情况，这是双高斯混合具有两个峰值的经典条件 [57]；它会排除这样的拟合：两个中心相同、宽度不同的高斯分布仅仅描述重尾，而贝叶斯信息准则本身会接受这种拟合。每一代都会重新作出判定，不会保留任何状态：只要其中一个漏斗变空，模型就会立即恢复为一个隶属函数。至多使用两个隶属函数是有意的限制；本文讨论的是单个与双重的情况。

E. 按隶属度合并

当两个隶属函数被打开时，个体 i 对函数 j 的隶属度，就是在上一次 EM 迭代的期望步（E-step）中计算得到的成分 j 的后验概率，

$$\rho_{ij} = \frac{\hat{\pi}_j \, \mathcal{N}(y_i; \hat{c}_j, \hat{s}_j^2)}{\sum_{k=1}^{2} \hat{\pi}_k \, \mathcal{N}(y_i; \hat{c}_k, \hat{s}_k^2)}, \qquad y_i = z_{il^{*}}. \tag{9}$$

它的形式与自适应网络模糊推理系统第 3 层的归一化触发强度相同，不同之处在于每个高斯分布都带有其归一化因子 $1/\hat{s}_j$；因此，它就是拟合混合模型的责任度。每个隶属函数

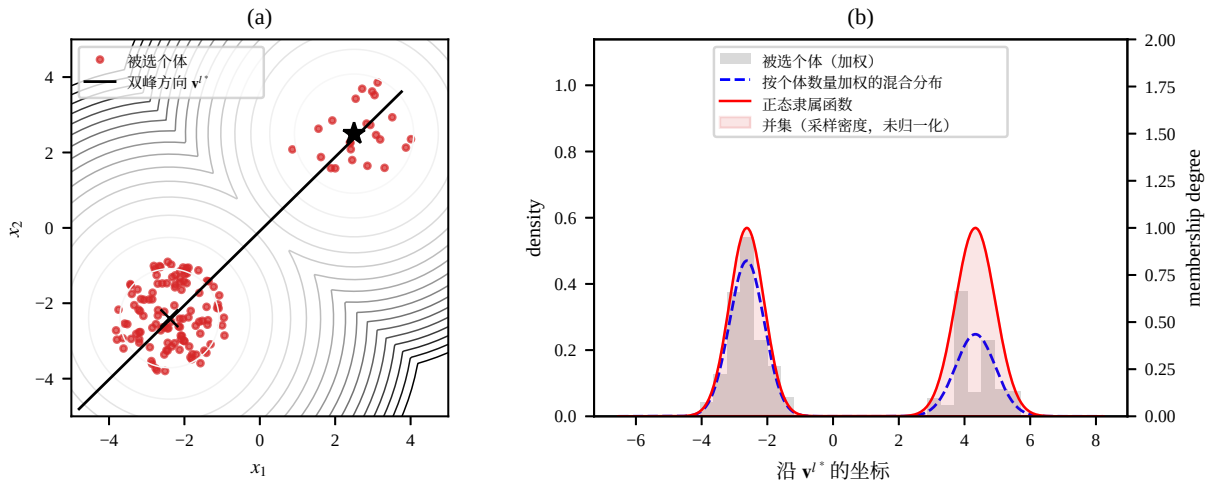

图1。二维实例（1）上搜索分布的构造（示意图）。（a）早期代中的被选个体、更深漏斗（⋆）、更宽漏斗（×）以及通过第四节-C的搜索发现的双峰方向 $\mathbf{v}_{l^*}$。（b）沿着$\mathbf{v}_{l^*}$ 更宽漏斗包含了大多数被选个体。最大似然混合分布（虚线）为每个成分分配的质量等于其头数。两个正态隶属函数（实线）的高度均为一；它们的归一化并集即为采样密度，因此，更深漏斗获得的样本份额取决于其宽度，而非其头数。

都通过隶属度加权矩，在每个特征方向 l 上与一个局部高斯分布相结合，

$$m_{jl} = \frac{\sum_i w_i \rho_{ij} z_{il}}{\sum_i w_i \rho_{ij}}, \qquad \sigma^2_{jl} = \frac{\sum_i w_i \rho_{ij} z^2_{il}}{\sum_i w_i \rho_{ij}} - m^2_{jl}, \quad (10)$$

其中 σ^2_{jl} 的下限设为 $10^{-6}\lambda_l$。没有任何个体被分配给某个簇；每个个体都根据其隶属度按比例对两个后件作出贡献。在双峰方向本身，矩给出了隶属函数的中心和宽度，即 $c_j = m_{jl^*}$ 和 $s_j^2 = \sigma^2_{jl^*}$，并将该函数取为正态函数：

$$\mu_j(y) = \exp\Big(-\frac{(y-c_j)^2}{2\alpha s_j^2}\Big), \qquad j = 1, 2. \quad (11)$$

拟合得到的权重 $\hat{\pi}_j$ 用于计算隶属度，随后将其丢弃。丢弃这些权重正是区分模型与混合分布的步骤。

F. 从并集采样

生成一个新个体分为三个步骤。(i) 以概率 $s_k/(s_1+s_2)$ 选择 $k \in \{1,2\}$，并抽样 $y \sim \mathcal{N}(c_k, \alpha s_k^2)$。(ii) 以 $\max_j \mu_j(y) / \sum_j \mu_j(y)$ 的接受概率接受 y；否则返回 (i)。(iii) 对于接受的 y，设置 $z_{l^*} = y$，对于 $l \neq l^*$抽样 $z_l \sim \mathcal{N}(m_{kl}, \alpha\sigma^2_{kl})$，并设置 $\mathbf{x} = \mathbf{m} + \mathbf{V}\mathbf{z}$，将其裁剪到该框内。

命题2。令 $Z = \int \max_j \mu_j(y)\, dy$。

(a) 被接受的 y 具有密度$\max_j \mu_j(y)/Z$，即两个隶属函数的归一化模糊并集。(b) 给定 y，隶属函数 k 的后件以概率 $\mu_k(y)/\sum_j \mu_j(y)$使用，即归一化触发强度。(c) 提议样本至少以$_1$ $_2$的概率被接受，并且$\sqrt{2\pi\alpha}\max(s_1, s_2) \le Z \le \sqrt{2\pi\alpha}\,(s_1+s_2)$。

(d) y 更接近 c_k （按 $\mu_k(y) \ge \mu_j(y)$ 的意义）的样本所占比例趋于 $s_k/(s_1+s_2)$，其中 $|c_1 - c_2|/(s_1+s_2) \to \infty$。它不依赖于 $\hat{\pi}_k$。

证明。(i) 步骤的提议密度为 $\sum_k \frac{s_k}{s_1+s_2}\mathcal{N}(y; c_k, \alpha s_k^2) = \sum_k \mu_k(y)/(\sqrt{2\pi\alpha}(s_1+s_2))$。乘以接受概率后，得到一个与 $\max_j \mu_j(y)$ 成正比的密度，这就是 (a)。给定其取值 y，提议样本来自 k 的概率为 $\mu_k(y)/\sum_j \mu_j(y)$，且接受仅取决于 y，这就是 (b)。对于 (c)，$\max_j \mu_j \ge \frac{1}{2}\sum_j \mu_j$，而 $\max_j \mu_j$介于每个 μ_j 与它们的和之间；它们的积分分别为 $\sqrt{2\pi\alpha}\,s_j$ 和 $\sqrt{2\pi\alpha}(s_1+s_2)$。对于 (d)，当重叠消失时，$\max_j \mu_j$ 在 μ_k 较大的集合上的质量趋于 $\sqrt{2\pi\alpha}\,s_k$。 □

命题2说明了搜索分布为何是并集而非总和。沿双峰方向，每个检测到的漏斗都由一个高度为 1 的隶属函数覆盖；漏斗中被选个体的数量会影响其隶属函数的位置和宽度，而不会影响其高度。因此，命题1(b)中的采样份额为 $b_t \approx s_1/(s_1+s_2)$。

算法 1列出了一个世代。使用一个隶属函数时，后件为 $m_{1l} = 0$、$\sigma^2_{1l} = \lambda_l$，步骤 8–9 简化为 CUR-G 的抽样；MF-EDA 与 CUR-G 的区别仅在于满足式 (7) 和式 (8) 的世代不同。

G. 成本

沿主方向的投影耗时 $O(nD^2)$，$2D$ 一维期望最大化拟合耗时 $O(nD)$，迭代次数固定为 40 次；获得投影后，计算后件耗时 $O(nD)$。特征分解 $O(D^3)$ 已包含在基线中。与任何全协方差高斯 EDA 一样，每一代耗时 $O(nD^2 + D^3)$。根据命题 2(c)

算法 1 MF-EDA 的一代 1：评估个体；保留迄今为止最优的 $\mathbf{x}^{\text{best}}$；更新 α2：选择最优的 $n = \lfloor \tau N \rfloor$；权重 w_i；均值 m；协方差 C =V$\mathbf{\Lambda}V^\top$3：对于每个主方向 $l = 1$，... ...， D 执行 4：通过加权期望最大化对 z_{1l}，.....， z_{nl} 拟合一个和两个高斯分布； $G_l \leftarrow$ 贝叶斯信息准则$_1(l)$ −贝叶斯信息准则$_2(l)$ 5：结束循环 6： $l^* \leftarrow \arg\max_l G_l$7：如果 BIC增益超过阈值 (7)，且两个中心之间的距离符合 (8)，则 $J \leftarrow 2$：隶属度ρ_{ij}，后件 (10)，正态隶属函数 (11)；否则 $J \leftarrow 1$8：如果 $J = 2$：采样 $N - 1$ 个体，其沿 $\mathbf{v}_{l^*}$ 的坐标从隶属函数的归一化并集中采样，其余坐标则从所选函数的后件中采样（第四节-F） 9：如果 $J = 1$：从 $\mathcal{N}$(m αC) 中采样 $N - 1$ 个体 10：添加 $\mathbf{x}^{\text{best}}$

平均而言，抽样所需的提议样本少于每个个体两个。没有任何步骤会评估适应度函数。这种复杂度表示掩盖了一个很大的常数：每代中， $2D$ 次迭代的期望最大化会评估 $120\,nD$ 个高斯密度，并对投影排序 $2D$ 次；对于所有 $D < 120$，这都超过了基线的 nD^2 次运算。第五节 G 报告了实测时间。

H. 作为其他估计分布算法成分的这些步骤

算法 1 的步骤 3–8 仅使用被选个体、其权重以及其协方差的特征分解。它们可以插入任何包含这些量的高斯 EDA 中。我们将这些步骤插入 AMaLGaM [5]：在 AMaLGaM 完成选择和最大似然估计后，方向搜索、判定和后件按上述方法以相等权重计算，与 AMaLGaM 中的做法相同；当两个隶属函数均开启时，从并集中抽取样本；AMaLGaM 的 AVS、SDR 和 AMS 则保持不变。我们将所得算法称为 AMaLGaM-MF。MF-EDA 本身是构建于 CUR-G 之上的成分。

I. 与现有方法的差异

1) 成分的加权方式。混合模型和小生境估计分布算法使用学习得到的权重对成分进行采样，或为每个簇分配一个子种群。MF-EDA 从正态隶属函数的并集中进行采样，因此某个成分所占的比例取决于其宽度（命题2）。第五节-E 用拟合混合模型的混合权重替代并集，并测量二者之间的差异。

2）成分形成的位置。CAMaLGaM、Hill-ValleA、混合分布估计算法和小生境估计分布算法在完整的 D 维空间中进行聚类。MF-EDA 沿一个方向拟合，该方向通过搜索主方向找到，并为每个隶属函数保留一个局部高斯分布 elsewhere. A one-dimensiona[1拟合] needs few individual 每个成分的样本数，无需计算成对距离。

3) 数量的选择方式。贝叶斯信息准则结合搜索范围和分离检验，在每一代中于1和2之间进行选择，且无需进行适应度评估。CAMaLGaM根据密度–适应度相关性对聚类层次结构进行评分；HillVallEA为山谷检验支付适应度评估代价； [19] 的多峰EDA随机确定小生境大小； [46] 则预先固定数量。

4) 所提出的主张。该主张包含两个方面：在适应度评估次数相同的情况下，在双漏斗景观上取得更多成功，而在单漏斗景观上不发生变化。

V. 实验

这些实验通过比较每种算法在有无该成分时的表现，检验这一双重主张（MF-EDA 对比加入 CUR-G 的 MF-EDA，AMaLGaM-MF 对比 AMaLGaM）：在双漏斗景观上，该成分会增加达到全局最优解的运行次数（第五节B和第五节C）；而在只有一个漏斗或一种全局结构的景观上，该成分既不会改变成功运行的次数，也不会使所需评估次数出现超过几个百分点的变化（第五节C和第五节F）。第五节D将运行结果与第三节中的模型进行比较；第五节E逐一移除该方法的各个要素并改变阈值；第五节G测量计算时间；第五节H报告该方法从何处开始不再发挥作用。

A. 设置

程序。除 HillVallEA 外，所有算法均以 C99 实现，并在 Linux 上使用 gcc 以优化级别 2 编译。对于 HillVallEA，我们使用作者公开发布的 C++代码，未经任何修改（ GitHub 上的 SCMaree/HillVallEA 仓库，提交 1a68ab73），并使用一个自行统计评估次数的驱动程序。双漏斗族、采用两种方案的 Lunacek 函数以及消融实验（192 个单元，每个单元运行 50 次）均使用相同源代码，在 Windows 机器上独立重新运行；在 185 个单元中，成功运行的次数完全相同，其余 7 个单元的差异最多为 5 次，这与浮点库不同所导致的情况相符。成本以适应度评估次数衡量；模型的计算时间在第五节G中单独报告。

方案。当迄今最优误差不超过 10^{-8} 时，运行即告成功。协议 A 将算法运行一次，直至预算耗尽或误差低于 10^{-10}。协议 B 会在种群坍缩（所有适应度值均在 10^{-12}范围内）时对种群进行均匀重新初始化，并持续运行直至预算耗尽；这使所有算法都能与采用重启机制的 HillVallEA 在同等条件下进行比较。两种方案的结果均予以报告。

参数。 MF-EDA 和 CUR-G 使用 N =max $\{115, \lfloor 17 + 3D^{1.5} + \frac{1}{2} \rfloor\}$ 个体,–即在 $D = 10$ 时为 115 个、在 $D = 20$ 时为 285 个、在 $D = 30$ 时为 510 个、在$D = 50$时为 1078 个，选择比例为 $\tau = 0.35$。下限确保至少有 40 个被选个体，因此，对于包含两个成分的一维拟合，每个成分约有 20 个个体。该方法没有其他需要设置的参数：允许量 2ln D、分离因子 2 以及 40 次期望最大化迭代在所有实验中都是固定的；

只有第五节的阈值测试会改变允许量。MF-EDA 和 CUR-G 每次运行都从相同的初始种群开始；此后它们的随机流不同，因此它们的比较是按初始种群配对进行的。

开发与测试种子。该方法基于一组种子开发（ Lunacek 函数的运行次数 0–49）。本节中的每个数值均来自开发过程中未使用的种子（Lunacek 函数和多漏斗函数的运行编号从 1000 开始，双漏斗族的运行编号从 2000 开始，BBOB 的实例为 16–30）。

比较算法。 使用其种群规模为 $17+3D^{1.5}$ $\tau=0.35$ 的 AMaLGaM、AVS、SDR 和 AMS，这些算法重新实现自 [5]; AMaLGaM-MF；（$\mu/\mu_w,\lambda$）-CMA-ES，采用默认的 $\lambda=4+\lfloor 3\ln\ D\rfloor\sigma_0=2$，初始均值在该框内均匀采样，重新实现自 [31]; CAMaLGaM 的重新实现版本（我们未找到公开代码），该版本来自 [17], ，对于论文中留待决定的选项，我们自行作出选择（合并边长度等于簇均值之间的距离、从均值最近的簇继承、簇权重相等、全局选择），种群规模采用 MF-EDA 的种群规模；以及作者的 HillVallEA，使用山谷聚类，并采用其默认的局部优化器、单变量 AMaLGaM 或全协方差 AMaLGaM。我们在单漏斗上检查了 AMaLGaM 和 CMA-ES 的重新实现版本的收敛性及评估次数，但未将其逐行与作者的代码进行比对；CAMaLGaM 的重新实现代表了我们对该论文的理解。

统计。 MF-EDA 与 CUR-G 的比较，以及 AMaLGaM-MF 与 AMaLGaM 的比较，均采用 Wilcoxon 符号秩检验 [58] ，以运行编号配对，比较达到目标所需的评估次数；未达到目标的情况按完整预算计。对于第一组比较，每对运行共享相同的初始种群；对于第二组比较，各次运行相互独立，而该检验允许这种情况。在评估次数相同的情况下，则在检查点处对迄今最优误差进行 Wilcoxon 秩和检验。每组检验均采用 Holm 方法 [59] 在 0.05 水平上进行校正；“胜/平/负”指校正后的检验结果。

B. 深度差和位置各异的双漏斗

第一组为 BBOB f_{24} [60] 中采用的形式 (1)，去掉其中的 Rastrigin 项和旋转：$\mu_2=-\sqrt{(\mu_1^2-\delta)/s}$，其中 $s=1-1/(2\sqrt{D+20}-8.2)$，第二个漏斗抬高 δD。深度差取值为 $\delta\in\{0.25,0.5,0.75,1\}$，全局漏斗的位置为 $\mu_1\in\{2,2.5,3\}$ 和 $D\in\{10,20,30\}$：共 36 种设置，每种运行 50 次，预算为 10^4D，采用协议 A。表 I 给出了成功次数。

CUR-G 在 1800 次运行中有 805 次达到全局最优解，而 MF-EDA 达到 1712 次。在全部 36 种设置下，MF-EDA 达到目标所需的评估次数都更少；其中 33 种设置在 Holm 校正后具有显著性，且没有任何一种设置显示其劣于 MF-EDA。单一高斯分布越弱，MF-EDA 的优势越明显。当两个漏斗的深度几乎相同时（$\delta=0.25$），CUR-G 在 450 次运行中有 28 次成功，而 MF-EDA 有 420 次；当深度差距较大时（$\delta=1$），两者分别为 348 次和 437 次。单一高斯分布的成功次数随 D 略有下降（在 $\delta=0.25$ 下 600 次运行中分别为 280、267 和 258 次，20 和 30），而 MF-EDA 的成功次数则有所增加（521、593 和 598 次）；我们将这一增长归因于较大的 D 所对应的更大种群，因为这为一维拟合提供了更多个体。

表 I 双漏斗族，深度差 δ 和全局漏斗位置 μ_1：在 10^4D 次评估内达到误差 10^{-8} 的 50 次运行中的次数，CUR-G /MF-EDA。 †：在 36 种设置下，针对目标的评估次数的配对符号秩检验经 Holm 校正后具有显著性。

δ	μ_1	$D=10$	$D=20$	$D=30$
0.25	2.0	7 / 36†	1 / 49†	2 / 50†
0.25	2.5	6 / 43†	0 / 50†	1 / 49†
0.25	3.0	5 / 44†	5 / 49†	1 / 50†
0.5	2.0	20 / 39†	16 / 48†	10 / 49†
0.5	2.5	17 / 45†	18 / 50†	16 / 50†
0.5	3.0	19 / 45†	18 / 49†	17 / 50†
0.75	2.0	31 / 42†	31 / 49†	32 / 50†
0.75	2.5	32 / 44†	29 / 50†	30 / 50†
0.75	3.0	30 / 45†	35 / 50†	28 / 50†
1	2.0	37 / 44†	38 / 49†	45 / 50
1	2.5	38 / 48†	37 / 50†	40 / 50†
1	3.0	38 / 46	39 / 50†	36 / 50
总计		280 / 521	267 / 593	258 / 598

C. Lunacek 漏斗与比较算法

第二组采用 [9]: μ_1 =2.5、 $\mu_2=-2.398179$、 $w_2=0.5$ 和 $s_2=0.87$ 的原始参数（G2-S，全局漏斗较窄）或 $s_2=1.15$（G2-L，全局漏斗较宽），并结合单漏斗 Gl $=\sum_i(x_i-2.5)^2$; 运行 50 次，预算为 $15\,000D$。表 II 和表 III 给出了两种方案下的成功次数和成功运行的中位评估次数，图 2 则绘制了 G2-S 上历史最优误差中位数随已用预算的变化。

双漏斗，不重启。 在 G2-S 上，随着 D 增大，单高斯 EDA 更频繁地错失全局漏斗：CUR- G 在 $D=10$、20 和 30 维时，50 次运行中分别成功 41、22 和 4 次；AMaLGaM 分别成功 29、3 和 0 次；采用默认种群的 CMA-ES 分别成功 23、20 和 16 次。加入隶属函数成分后，MF- EDA 的成功次数为 49、49 和 49，AMaLGaM-MF 则为 48、49 和 47。经 Holm 校正后，MF-EDA 与 CUR-G 的配对检验在 $D=20$ 和 $D=30$ （$p=3.6\times10^{-6}$ 和 1.2×10^{-9}）时显著，在 $D=10$ （$p=0.15$）时不显著；AMaLGaM-MF 与 AMaLGaM 的检验在三个维度上均显著（$p\le7.0\times10^{-7}$）。重新实现的 CAMaLGaM 成功 11、3 和 1 次；此结果仅适用于我们的重新实现及其种群规模。在 $D=50$ 维时，CUR-G 和 AMaLGaM 在 50 次运行中均未成功，CMA-ES 成功 10 次，而 MF-EDA 和 AMaLGaM-MF 分别成功 40 次和 43 次；该成分的成功率从这一维度开始下降。

单漏斗和易双漏斗：没有变化。在 G1 和 G2-L 上，无论是否包含该成分，每种估计分布算法在所有运行中都取得成功。MF-EDA 与 CUR- G 的中位评估次数差异最多为 5.5%，而 AMaLGaM-MF 与

表II LUNACEK G 型漏斗，协议 A（无重启）：达到误差 10^{-8} 的 50 次运行中运行次数，以及括号内这些运行的评估次数中位数（以千为单位）。预算 $15\,000D$。短横线表示该算法未在该单元格中运行。

函数	D	CUR-G	MF-EDA	AMaLGaM	AMaLGaM-MF	CMA-ES	CAMaLGaM-re
G1（单漏斗）	10	50（12.7）	50（12.7）	50（6.8）	50（6.7）	50（1.5）	50（7.7）
	20	50（53.4）	50（51.3）	50（26.9）	50（28.7）	50（2.8）	50（31.3）
	30	50（125）	50（118）	50 (64.0)	50 (60.8)	50 (4.1)	50 (64.9)
	50	50 (343)	50 (345)	50（171）	50（180）	50（6.4）	–
G2-L（双重的；全局漏斗更宽）	10	50（11.6）	50（11.7）	50（6.2）	50 (6.7)	33 (1.5)	50 (7.8)
	20	50（50.2）	50（49.0）	50 (27.0)	50 (27.7)	34 (2.8)	50 (28.4)
	30	50 (118)	50 (115)	50 (61.4)	50 (63.5)	33 (4.1)	50 (63.0)
	50	50 (333)	50 (331)	50 (173)	50 (172)	32 (6.3)	–
G2-S（双重的；全局漏斗更窄）	10	41 (12.2)	49 (12.3)	29 (7.7)	48 (6.7)	23 (1.5)	11 (8.6)
	20	22 (50.5)	49 (51.6)	3 (29.6)	49 (27.4)	20 (2.8)	3 (38.1)
	30	4 (126)	49 (117)	0 (–)	47 (65.1)	16 (4.2)	1（63.5）
	50	0 (–)	40（335）	0 (–)	43（169）	10（6.4）	–

表 III Lunacek G 型漏斗，协议 B（坍缩后重启）：在 50 次运行中达到 10^{-8} 误差的运行次数，以及括号中这些运行的评估次数中位数（以千计）。预算 $15\,000D$。短横线表示该算法未在该单元格中运行。HillVallEA 会自行重启，因此仅在此处列出；-u 和 -f 分别表示其单变量（默认）和全协方差局部优化器。

函数	D	CUR-G	MF-EDA	AMaLGaM	AMaLGaM-MF	CMA-ES	CAMaLGaM-re	HillVallEA-u	HillVallEA-f
G1（单漏斗）	10	50（12.7）	50（12.7）	50（6.8）	50（6.7）	50（1.5）	50（7.7）	50（2.3）	50 (7.7)
	20	50 (53.4)	50 (51.3)	50 (26.9)	50 (28.7)	50 (2.8)	50 (31.3)	50 (5.6)	50 (30.4)
	30	50 (125)	50 (118)	50 (64.0)	50（60.8）	50（4.1）	50（64.9）	50（9.1）	50（61.0）
	50	50（343）	50 (345)	50 (171)	50 (180)	50 (6.4)	–	–	–
G2-L（双重的；全局漏斗更宽）	10	50（11.6）	50（11.7）	50 (6.2)	50 (6.7)	50 (1.5)	50 (7.8)	50 (2.5)	50 (8.4)
	20	50 (50.2)	50 (49.0)	50 (27.0)	50 (27.7)	50 (2.9)	50 (28.4)	50 (5.8)	50 (28.9)
	30	50 (118)	50 (115)	50 (61.4)	50 (63.5)	50 (4.3)	50 (63.0)	50 (9.2)	50 (61.8)
	50	50 (333)	50 (331)	50 (173)	50（172）	50（6.5）	–	–	–
G2-S（双重的；全局漏斗更窄）	10	50（12.6）	50（12.3）	50（9.1）	50（6.8）	50 (3.6)	50 (19.5)	50 (4.5)	50 (17.1)
	20	47 (122)	50 (51.6)	16 (102)	50 (27.5)	50 (6.9)	44 (157)	50 (12.3)	50 (56.6)
	30	15 (285)	50 (117)	0 (–)	50 (67.1)	50 (10.4)	4 (154)	50（20.5）	50（138）
	50	0 (–)	40（335）	0 (–)	50（172）	50（29.3）	–	–	–

AMaLGaM 在四种维度下各自最多相差 8.2%，且方向不一。在这些单元格上的 16 项测试中（每对算法各进行八项，$D = 10$至 50），校正后有一项具有显著性：在 $D = 30$ 的 G1 上，MF-EDA 所需的评估次数比 CUR-G 少 5.5%。我们不认为这是该成分造成的影响，因为在那里，该成分仅在 0.4% 的世代中开启两个隶属函数（在 $D = 10$ 时为 3.9%，在$D = 20$ 时为 1.1%）；两个程序在初始种群之后使用了不同的随机流。在 G2-L 上，运行开始时有 5% 至 8% 的世代开启两个隶属函数，而模型随后恢复为一个，成本没有变化。CMA-ES 在 G2-L 上约三分之一的运行中未能成功，因为其较小的种群坍缩到次优漏斗中。

采用重启。在协议 B 下，所有采用重启的算法都能在 $D = 10$ 时快速求解 G2-S。在 $D = 20$ 和 30 时，单高斯 EDA 在预算内仍有许多次运行未能成功（CUR-G 分别有 47 次和 15 次，AMaLGaM 有 16 次和 50 次中的 0 次），因为它们每次重启都要耗费一次完整运行，且成功概率较低，与表 II 中所见相同；而 MF-EDA 和 AMaLGaM-MF 几乎只用一次运行的评估次数，就达到了 50 次全数成功。在 $D = 50$ 时，AMaLGaM-MF 达到了 50 次全数成功，MF-EDA 则在预算内保持 40 次成功。表 III 中有两项结果不支持所提出的方法，必须如实看待。首先，在单漏斗问题上，秩加权 EDA（MF-EDA、CUR-G）所需的评估次数约为 AMaLGaM 的两倍，且远多于 CMA-ES；该成分并未改变其基线算法的速度。其次，重启 CMA-ES 求解 G2-S 所需的评估次数少于任何 EDA：在 $D = 20$、30 和 50 时，中位数分别为 6.9、10.4 和 29.3 千次；相比之下，AMaLGaM-MF 分别需要 27.5、67.1 和 172 千次，MF-EDA 则分别需要 51.6、117 和 335 千次；采用单变量优化器的 HillVallEA 在 $D = 20$ 和 30 时分别需要 12.3 和 20.5 千次。CMA-ES 的每次运行在 G2-S 上的成功概率约为五分之一至二分之一，且只需数千次评估，因此少量重启比运行一次几乎必定成功的基于种群的 EDA 更省成本。证据支持这样的说法：该成分消除了单高斯 EDA 在双漏斗问题上的失效，且在单漏斗问题上不增加成本；但证据并不支持这样的说法：由此得到的 EDA 是求解这些函数成本最低的优化器。

D. 与双区域模型的比较

为了与第三节的运行结果进行比较，我们记录了在 $D = 20$ 上针对 G2-S 进行 50 次运行时，每一代中

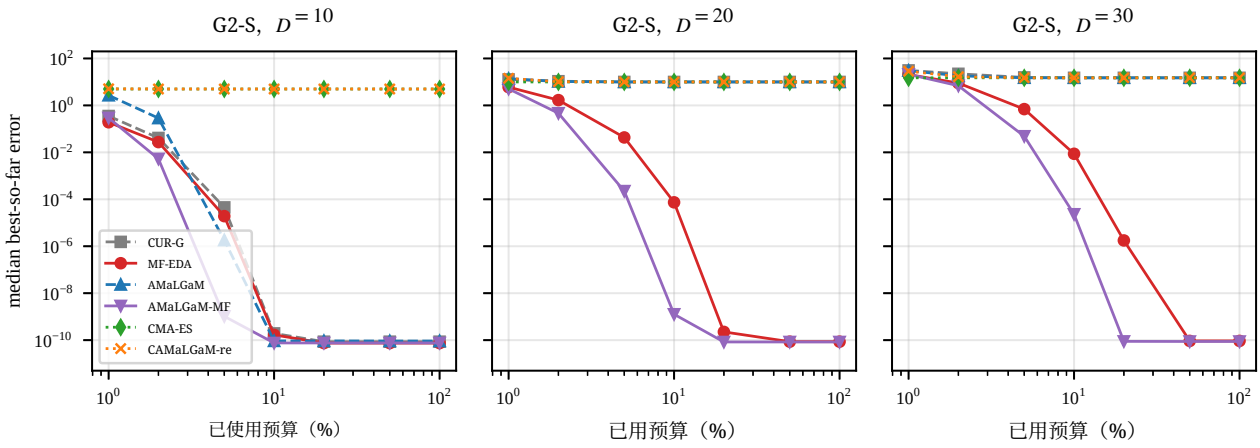

图 2. 在协议 A 下，G2-S 上 50 次运行的历史最优误差中位数与已使用预算份额的关系。曲线在 $D/2$ 处结束，表示中位数运行在次优漏斗中结束。

更靠近更深漏斗中心的被选个体所占的份额；当两个成分均开放时，还记录了更深漏斗上成分的采样份额。我们从相同的初始种群出发，跟踪了三种模型：单个高斯分布（第二个隶属函数始终未开放的 MF-EDA）；按 MF-EDA 的方式找到两个成分，但按照拟合混合模型的混合权重 $\hat{\pi}_j$（个体数量）进行抽样；以及 MF-EDA。图3显示了中位数和四分位数。

三个模型的初始深漏斗选定份额均为 0.35。采用单个高斯分布时，中位数份额随后降至 0.29、0.14、0.07 和 0.02，并从第五代开始降为零；除第一代外，之后大致每代减半，符合命题1(a)所述的几何衰减。到第 9 代时，深漏斗在 50 次运行中的 27 次里已经消失，23 次运行成功（CUR-G 是一个独立程序，其随机流在初始种群之后有所不同；表II中它有 22 次运行成功）。采用人数权重时，采样份额随选定份额变化（图 b）：中位数份额在第一代降至 0.31，随后回升；45 次运行成功，而在成功的运行中，深漏斗约在六代内接管整个种群。采用并集时，无论选定份额如何，采样份额都保持在约一半（前八代介于 0.47 和 0.56 之间），中位数选定份额从第一代起便有所上升，49 次运行成功。并集使更宽漏斗在更长时间内保持采样，因此它比人数计数模型更晚作出决策；它减少了运行失败的次数。在 MF-EDA 的前五代中，有 92% 的代数会开启两个隶属函数；之后模型会恢复为一个隶属函数。

E. 消融

表 IV 在 G2-S 上逐一移除 MF-EDA 的各个组成元素。变体 (a) 使用拟合混合模型的两个成分及其混合权重 $\hat{\pi}_j$ 进行采样，而不是使用并集份额。变体 (b) 用位置与适应度排名之间具有最大相关比的方向替代第四节-C中的搜索，且不采用 (7) 中的允许项；该规则曾用于该方法的第一个版本。下一行同时应用 (a) 和 (b)

表IV消融实验：G2-S，协议 A：达到误差 10^{-8} 的运行次数（共 50 次；括号内为这些运行的中位评估次数，单位为千次）。每个变体仅改变 MF-EDA 的一个元素。

变体	$D = 10$	$D = 20$	$D = 30$
MF-EDA（完整）	49（12.3）	49（51.6）	49（117）
（a）采样份额 $\hat{\pi}_j$（头数）	47（11.8）	45（49.6）	47（116）
(b)基于相关比率的方向	45(12.0)	45(49.2)	48 (118)
(a)和(b)，无分离检验	41(12.4)	35(50.5)	35 (117)
(c) 相等的采样份额	50(12.0)	49(48.5)	49 (118)
(d) 硬分配	50 (12.0)	49 (49.6)	49 (114)
(e) 历史变体，$h = 5$	50(12.4)	49(51.0)	49 (116)
(e) 历史变体，$h = 1$	50(13.0)	49(48.5)	49 (118)
(e) 历史变体，$h = 20$	50 (12.0)	49 (50.1)	49 (117)
单个高斯分布 (CUR-G)	41 (12.2)	22 (50.5)	4 (126)

，并省略分离检验 (8)，这就是该方法的第一个版本。变体 (c) 为两个隶属函数分别赋予二分之一的份额。变体 (d) 将每个个体分配给其隶属度较大的隶属函数。变体 (e) 使第二个隶属函数在各代中保持活跃，在每个隶属函数内部进行选择，并在某个函数连续被支配 h 代后将其关闭。

在这三个维度上，MF-EDA 在 150 次运行中成功了 147 次，而单一高斯分布成功了 67 次。头数份额 (a) 又减少了 8 次运行（139 次），相关比方向 (b) 减少了 9 次（138 次）；两者同时使用时，成功次数为 111 次。因此，这两个要素各自发挥作用；一旦其中一个到位，另一个单独就能弥补单个高斯分布与 MF-EDA 之间的大部分差距。均等份额 (c) 的成功次数与并集份额 (148 次) 相同，这与命题1(b) 一致：关键在于份额不随头数变化。硬分配 (d) 和历史变体 (e) 的成功次数也相同（各为 148 次）；按隶属度进行合并是我们所描述的构造，而数据并未显示它在这些函数上的表现优于硬分配。结果不依赖于在不同世代之间保留状态，这也正是 MF-EDA 不保留状态的原因。

对阈值的敏感性。表V将（7）中的 $2 \ln D$ 替换为 $c \ln D$，其中 c 从 0 变化到 4。在 G2-S 上，

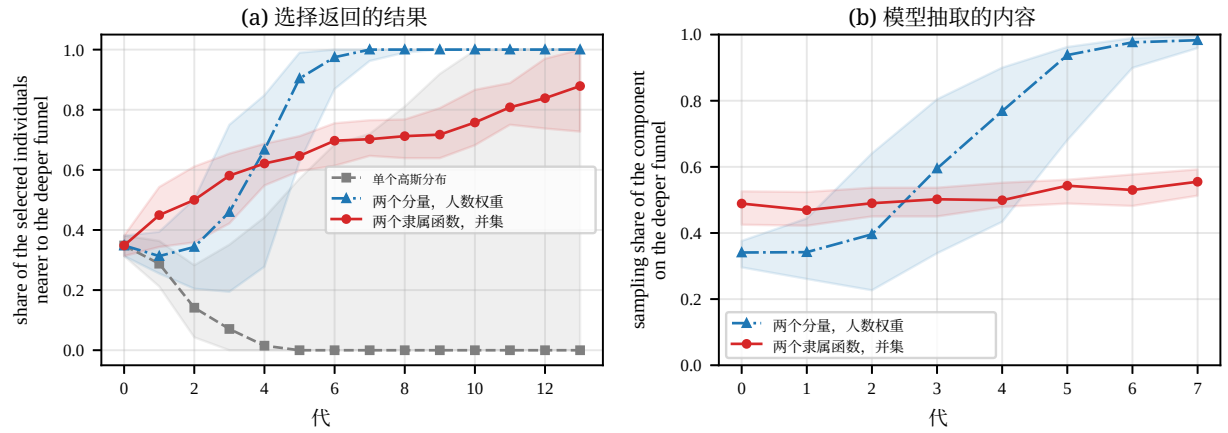

图3. G2-S 上 50 次运行的轨迹， $D = 20$ （中位数和四分位数）。(a) 被选个体中距离更深漏斗较近的个体所占比率。(b) 在两个分量均开启的运行和世代中，深漏斗上分量的采样份额。

表V BIC增益的阈值 $c \ln D$ 敏感性（协议 A，50 次运行，开发过程中未使用的随机种子； $c = 2$ 为所有其他实验的设置）。G2-S：成功运行次数。G1：成功运行次数、成功运行的中位评估次数（以千计），以及使用两个隶属函数的世代百分比。

	G2-S 运行次数			G1 运行次数			G1评估次数			G1两个函数 (%)		
c	10	20	30	10	20	30	10	20	30	10	20	30
0	48	49	49	50	50	50	13.0	51.9	121	21.1	11.5	7.1
1	49	49	49	50	50	50	12.5	52.7	121	9.2	3.5	1.9
2	49	49	49	50	50	50	12.7	51.3	118	3.9	1.1	0.4
3	50	49	49	50	50	50	12.5	53.3	121	1.9	0.3	0.1
4	49	47	48	50	50	50	12.4	51.6	121	0.7	0.1	0.0

对于每个 c和 D，成功运行次数都保持在 50 次中的 47 至 50 次。在单漏斗 G1 上，每次运行都成功，中位评估次数的变化不到 5%；发生变化的是开启两个隶属函数的频率：在 $c = 0$和 $D = 10$时，占世代的 21%，而在 $c \geq 3$时则降至 2% 或更低。因此，在只有一个漏斗的情况下，较小的阈值会开启更多第二个函数，而在 G1 上不会造成可测量的评估次数成本；较大的阈值则会减少开启次数，同时在这些运行中不会丢失双漏斗。在此范围内，结果对阈值并不敏感。

F. 单结构景观：无损失

为检验在不需要该成分时它是否不会造成负面影响，我们在十个多峰 BBOB 函数 f_{15}–f_{24} [60] at $D \in \{2, 5, 10, 20\}$ 上运行 MF-EDA 和 CUR-G，测试实例 16 至 30，预算为 $10^4 D$，并比较预算达到 1%、5%、20% 和 100% 时的迄今最优误差。表 VI 汇总了 160 次比较：经过 Holm 校正后，0 次胜出、160 次持平、0 次落败；未经校正的比较结果则几乎势均力敌。在这些函数上，两个隶属函数在 3% 至 64% 的世代中处于开启状态，具体比例取决于函数和维度；它们既不会加快也不会减慢搜索。这些函数包括 f_{24}，它是一个布满 Rastrigin 型局部最优的双漏斗函数；在该函数上，两个算法都未能达到

表 VI MF-EDA 与 CUR-G 在 BBOB f_{15}–f_{24} （各 15 个实例）上的比较：对单元格中十个函数的 Holm 校正后，基于迄今最优误差的秩和检验结果（胜/平/负），以及未经校正的秩和检验方向（+：MF-EDA 的误差平均秩较小； −：CUR-G 的误差平均秩较小）。

预算	$D = 2$	$D = 5$	$D = 10$	$D = 20$
1%	0/10/0 (+3 −7)	0/10/0 (+7 −3)	0/10/0 (+7 −3)	0/10/0 (+7 −3)
5%	0/10/0 (+5 −5)	0/10/0 (+5 −5)	0/10/0 (+7 −3)	0/10/0 (+7 −3)
20%	0/10/0 (+7 −3)	0/10/0 (+4 −6)	0/10/0 (+6 −4)	0/10/0 (+7 −3)
100%	0/10/0 (+6 −4)	0/10/0 (+5 −5)	0/10/0 (+4 −6)	0/10/0 (+6 −4)

目标值（两者的最终误差中位数均在 2 到 24 之间），而该成分并未改变这一结果。

G. 计算时间

到目前为止的比较统计的是适应度评估次数。表VII报告了在 G1 和 G2-S 上进行 10 次运行时每代的用户 CPU 时间，所用随机种子与上文相同。时间包括适应度评估；这些测试函数上的评估开销很低，因此时间主要花在模型上。每个单元格是一项测量；在 $D = 10$ 进行了三次重复，结果变化幅度最多约为 10%。MF-EDA 的一代耗时约为 CUR-G 一代的 8 到 17 倍；随着 D 增大，这一比率会下降，这与 EM 拟合的 $120\,nD$ 密度评估次数相对于基线的 nD^2 操作相符（第四节）。就绝对时间而言，MF-EDA 的一代包含 1078 个个体，在 $D = 50$ 耗时 44 毫秒，即每个采样个体约 0.04 毫秒。因此，当适应度评估的耗时低于这一水平时，开销就很重要，例如本文的测试函数；而当一次评估耗时达到数毫秒或更久时，开销便不再重要。EM 拟合采用固定的 40 次迭代；在收敛时停止拟合是降低开销的直接方法，但我们尚未测试这种做法。

H. 局限性

窄得多的全局漏斗。图4展示了协议 A 下，次优漏斗在 $D = 20$ 处的宽度因子 s_2 如何变化； s_2越小，次优漏斗相对于

表VII 计算时间：每代用户 CPU 时间（毫秒）（10 次运行，在一台 2.1 GHz 英特尔至强处理器上使用一个核心，GCC -O2，每个单元测量一次），包括适应度评估，以及 MF-EDA /CUR-G 的比率。

	G1			G2-S		
D	CUR-G	MF-EDA	比率	CUR-G	MF-EDA	比率
10	0.069	0.943	13.6	0.056	0.947	17.0
20	0.365	4.51	12.4	0.302	4.24	14.0
30	1.19	12.2	10.2	0.874	11.7	13.4
50	5.32	43.5	8.2	4.47	43.6	9.7

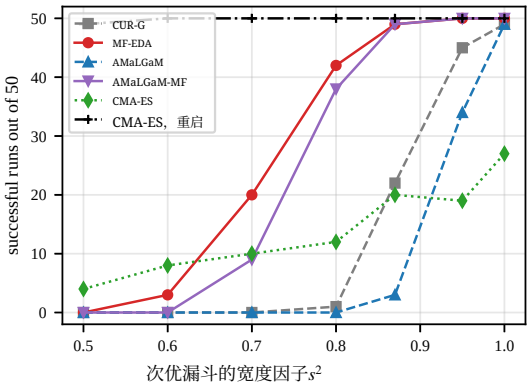

图4. 子最优漏斗的宽度因子 s_2 下，50 次运行中的成功次数； $D = 20$，协议 A（CMA-ES 也采用重启）。 $s_2 = 0.87$ 为 G2-S。

全局漏斗就越宽。该成分将两种基线的运行失败次数达到一半时对应的点，从 $s_2 \approx 0.9$ 移至 0.7 至 0.8 之间：在 $s_2 = 0.8$，CUR-G 在 50 次运行中成功 1 次，MF-EDA 成功 42 次；AMaLGaM 成功 0 次，AMaLGaM-MF 成功 38 次。在 $s_2 = 0.7$，MF-EDA 和 AMaLGaM-MF 的成功次数分别为 20 和 9；在 $s_2 = 0.6$，分别为 3 和 0；在 $s_2 = 0.5$，没有任何估计分布算法成功。一个可能的原因是，初始种群中落在全局漏斗内的被选个体太少，导致一维拟合无法检测到第二个模态；不过，我们并未逐次运行验证这一点。重启 CMA-ES 在每个 s_2下仍能成功 49 或 50 次，其中位评估次数从 $s_2 = 1$ 时的 3.0 千次上升至 $s_2 = 0.5$时的 43.5 千次。

一条直线上有两个以上的漏斗。对于漏斗中心随机抽取、因而不在同一直线上的$K \in \{3, 5\}$ 个漏斗景观，MF-EDA 在 900 次运行中成功 571 次，CUR-G 成功 544 次（ 36 种设置，每种设置运行 25 次）；校正后没有任何设置达到显著性（未校正的方向中，22 个有利，11 个不利）。沿一个方向的两个隶属函数无法表示分布在多个方向上的漏斗。

适用范围。本节中的漏斗均为球形，而 MF-EDA 在每个漏斗内沿用其基线模型。

VI. 讨论

该成分的变化。 该成分改变搜索分布中的一个方面：当被选个体沿某一主方向呈双峰分布时，两组个体将保留为两个高斯分布，其采样份额依据各自的宽度而非个体数量确定。第三节的双区域模型说明了这一点为何在双漏斗问题上很重要，而实验表明它在其他情况下影响不大：在单漏斗问题上，几乎每一代都只选择一个隶属函数，因此该算法就是其基线版本；在多峰 BBOB 函数上，第二个隶属函数会更频繁地开启（占 3% 至 64% 的世代），但对结果没有显著影响。

局限性。 第五节H部分展示了该成分不再有帮助的情况：全局漏斗窄得多（$s_2 \leq 0.6$），以及多个漏斗不在同一直线上。此外，重启 CMA-ES 求解标准双漏斗所需的评估次数少于所有受测估计分布算法，因为它的单次运行开销低；受测漏斗均为球形；而与 CAMaLGaM 的比较则基于我们的重新实现。

计算时间。 该成分不会增加适应度评估次数，但会增加计算时间：当 $D = 10$ 为 50 时，每一代所需时间约为基线方法一代的 8 至 17 倍（第 V-G 节）。因此，它适用于评估成本高于每个个体约 0.04 毫秒的问题，而不适用于廉价测试函数，因为在这类问题中，模型会占据主要运行时间。

未来工作。 未来有三个研究方向。首先，EM 拟合可以在收敛时停止，而不必固定运行 40 次迭代，这应该能减少开销；但我们尚未测量具体能减少多少。其次，在两个或三个方向上使用隶属函数，并沿用相同的决策规则，可以描述两个以上的漏斗；不过，每个成分需要更多个体。第三，可以将该成分嵌入更快速的单高斯方法，例如带重启的 CMA-ES，从而结合低成本运行与基于宽度的份额。

VII. 结论

本文提出了 MF-EDA，这是一种高斯分布估计分布算法，通过沿双峰主方向将正态隶属函数与局部高斯分布合并，并从它们的并集中抽样，从而自适应地调整分布数量。随后，采样份额依据各分量的宽度而非个体数量确定，消除了更深漏斗所带来的累积劣势。在未用于开发的随机种子上，该分量使基线算法在 36 种双漏斗设置中的成功运行次数从 805 次提高到 1712 次（共 1800 次）；在 $D = 50$ 处的 Lunacek G2-S 漏斗上，秩加权 EDA 的成功运行次数从 0 次提高到 40 次（共 50 次），AMaLGaM 则从 0 次提高到 43 次（共 50 次）。在单漏斗以及 BBOB f_{15}–f_{24} 上，结果没有显著变化，但第五节 V-C 节讨论的一个单元格除外。代价是每代的计算时间更长，而前文所述的未来工作旨在缩短这一时间。

致谢

本工作的最初构想由张俊教授提出。在张教授的指导下，研究团队使用 AI 系统 Claude（Anthropic）实现算法并开展实验，同时协助完成稿件。其他作者和学生审查并校对了代码、结果和文本。作者对本文内容负全部责任。

参考文献

[1] H. M¨uhlenbein 和 G. Paass，“从基因重组到分布估计 I：二进制参数”，载于《自然启发并行问题求解——PPSN IV》，LNCS 系列，第 1141 卷。德国柏林：施普林格，1996 年，第 178–187 页。 [2] P. Larra˜naga 和 J. A. Lozano 编，《分布估计算法：进化计算的新工具》。美国马萨诸塞州波士顿：克吕维尔，2002 年。 [3] P. Larra˜naga 和 C. Bielza，“机器学习中的分布估计算法：综述”，《IEEE 进化计算汇刊》，第 28 卷，第 5 期，第 1301–1321 页，2024 年 10 月。 [4] P. A. N. Bosman 和 D. Thierens，“从离散分布估计算法扩展到连续分布估计算法：IDEA”，载于《自然启发并行问题求解——PPSN VI》，LNCS 系列，第 1917 卷。德国柏林：施普林格，2000 年，第 767–776 页。 [5] P. A. N. Bosman、J. Grahl 和 D. Thierens，“用于连续分布估计算法数值优化的改进最大似然高斯模型”，数学与计算机科学研究中心，荷兰阿姆斯特丹，技术报告 SEN-E0704，2007 年。 [6] P. A. N. Bosman、J. Grahl 和 D. Thierens，“对无参数 AMaLGaM 在有噪声和无噪声函数上的基准测试”，《进化计算》，第 21 卷，第 3 期，第 445–469 页，2013 年。 [7] R. J. Scholman、T. Alderliesten 和 P. A. N. Bosman，“通过增量分布估计提高 RV-GOMEA 中实值灰盒优化的效率”，载于遗传与进化计算大会（GECCO）论文集，2025 年，第 755–763 页。 [8] X. Wang、Y. Li、Y. Liang、B. Wu 和 Y. Xuan，“一种采用分布修改策略的新型集成分布估计算法”，《复杂智能系统》，第 9 卷，第 5 期，第 5377–5416 页，2023 年。 [9] M. Lunacek、D. Whitley 和 A. Sutton，“全局结构对搜索的影响”，载于《自然启发并行问题求解——PPSN X》，LNCS 系列，第 5199 卷。德国柏林：施普林格，2008 年，第 498–507 页。 [10] N. Hansen 和 A. Ostermeier，“进化策略中的完全去随机化自适应”，《进化计算》，第 9 卷，第 2 期，第 159–195 页，2001 年。 [11] J. P. K. Doye、M. A. Miller 和 D. J. Wales，“38 原子 Lennard-Jones 团簇的双漏斗能量景观”，《化学物理杂志》，第 110 卷，第 6896–6906 页，1999 年。 [12] K. Mizuta、S. Tanaka、S. Tanaka 和 T. Hatanaka，“基于新颖性的多样化局部最优网络连续景观生成”，2026 年，arXiv:2604.21468。 [13] T. Blackwell，“能垒树基准测试：众多吸引盆地与双漏斗”，载于遗传与进化计算大会（GECCO）论文集，2023 年，第 13–20 页。[14] P. A. N. Bosman 和 D. Thierens，“采用混合分布和因子分解选择指标推进连续 IDEA”，载于《通过构建和使用概率模型进行优化研讨会》，GECCO，2001 年，第 208–212 页。 [15] D. Thierens 和 P. A. N. Bosman，“基于混合分布的多目标迭代密度估计进化算法”，乌得勒支大学信息与计算科学研究所，荷兰乌得勒支，技术报告 UU-CS-2001-59，2001 年。 [16] Q. Lu 和 X. Yao，“连续优化中的高斯分布聚类与学习”，《IEEE 系统、人及控制论汇刊 C 辑》，第 35 卷，第 2 期，第 195–204 页，2005 年。 [17] S. C. Maree、T. Alderliesten、D. Thierens 和 P. A. N. Bosman，“通过分层高斯混合学习对分布估计算法进行小生境处理”，载于遗传与进化计算大会（GECCO）论文集，2017 年，第 713–720 页。 [18] S. C. Maree、T. Alderliesten、D. Thierens 和 P. A. N. Bosman，“由山谷聚类驱动的实值进化多模态优化”，载于遗传与进化计算大会（GECCO）论文集，2018 年，第 857–864 页。[19] Q. Yang、W.-N. Chen、Y. Li、C. L. P. Chen、X.-M. Xu 和 J. Zhang，“多模态分布估计算法”，《IEEE 控制论汇刊》，第 47 卷，第 3 期，第 636–650 页，2017 年。 [20] Y. Liang、Z. Ren、B. Pang 和 A. Chen，“通过自适应聚类对基于档案的高斯分布估计算法进行小生境处理”，2018 年，arXiv:1803.00986。 [21] J.-S. R. Jang，“ANFIS：自适应神经模糊推理系统”，《IEEE 系统、人及控制论汇刊》，第 23 卷，第 3 期，第 665–685 页，1993 年。 [22] J.-S. R. Jang、C.-T. Sun 和 E. Mizutani，《神经模糊与软计算：学习与机器智能的计算方法》。美国新泽西州上萨德尔河：普伦蒂斯·霍尔，1997 年。 [23] G. Schwarz，“ 估计模型维数 ”，《统计年鉴》，第 6 卷

，第2期，第461–464页，1978。

[24] S. Kern, S. D. Mc, and P. Koumoutsakos, ¨ 《连续进化算法中的概率分布学习——比较综述》，《自然计算》，第3卷，第1期，第77–112页，2004年。[25] J. Grahl、P. A. N. Bosman和F. Rothlauf，《相关性触发的自适应方差缩放IDEA》，载于遗传与进化计算大会（GECCO）论文集，2006年，第397–404页。[26] P. A. N. Bosman、J. Grahl和F. Rothlauf，《SDR：正态估计分布算法中自适应方差缩放的更佳触发机制》，载于遗传与进化计算大会（GECCO）论文集，2007年，第492–499页。[27] C. Olieman、A. Bouter和P. A. N. Bosman，《实值基因池最优混合进化算法中的适应度驱动连接学习》，《IEEE进化计算汇刊》，第25卷，第2期，第358–370页，2021年。[28] G. Andreadis、T. Alderliesten和P. A. N. Bosman，《采用RV-GOMEA进行灰箱优化的适应度驱动连接学习与最大团条件连接建模》，载于遗传与进化计算大会（GECCO）论文集，2024年，第647–655页。[29] D. Wierstra、T. Schaul、T. Glasmachers、Y. Sun、J. Peters和J. Schmidhuber，《自然进化策略》，《机器学习研究杂志》，第15卷，第27期，第949–980页，2014年。[30] V. P. Soloviev、C. Bielza和P. Larra˜naga，《连续优化的半参数估计分布算法》，《IEEE进化计算汇刊》，第28卷，第4期，第1069–1083页，2024年8月。[31] N. Hansen，《CMA进化策略：教程》，2016年，arXiv:1604.00772。[32] A. Auger和N. Hansen，《一种种群规模递增的重启CMA进化策略》，载于IEEE进化计算大会（CEC）论文集，第2卷，2005年，第1769–1776页。[33] N. Hansen，《在BBOB-2009函数测试平台上对双种群CMA-ES进行基准测试》，载于遗传与进化计算大会（GECCO）伴随论文集，2009年，第2389–2396页。[34] I. Loshchilov、M. Schoenauer和M. Sebag，《CMA-ES的替代重启策略》，载于《自然界并行问题求解——PPSN XII》，LNCS丛书，第7491卷。柏林，德国：施普林格，2012年，第296–305页。[35] C. Fernando和K. De Silva，《改进种群规模自适应CMA-ES算法，以应对弱结构多峰函数中的步长爆炸》，2025年，arXiv:2506.00825。[36] M. Nomura、M. Shibata和R. Ha mano，《cmaes：一个简洁实用的CMA-ES Python库》，2026年，arXiv:2402.01373。[37] J. de Nobel、D. Vermetten、A. V. Kononova、C. Doerr和T. B¨ack，《模块化CMA-ES：现代进化策略框架》，2026年，arXiv:2610.09699。[38] M. Nomura、Y. Akimoto和I. Ono，《采用学习率自适应的CMA-ES》，《ACM进化学习与优化汇刊》，2024年，doi: 10.1145/3698203（arXiv:2401.15876）。[39] X. He、Z. Zheng和Y. Zhou，《MMES：面向大规模优化的基于混合模型的进化策略》，《IEEE进化计算汇刊》，第25卷，第2期，第320–333页，2021年4月。[40] P. Yang、K. Tang和X. Lu，《通过检测有前景区域改进多峰问题的估计分布算法》，《IEEE控制论汇刊》，第45卷，第8期，第1438–1449页，2015年。[41] A. Ahrari、K. Deb和M. Preuss，《通过具有排斥子种群的协方差矩阵自适应进化策略实现多模态优化》，《进化计算》，第25卷，第3期，第439–471页，2017年。[42] D. Chauhan，《S-CARD-CMSA：面向多模态优化的分数感知候选档案与密度过滤报告》，2026年，arXiv:2607.13764。[43] Y. Matanga、P. Owolawi、C. Du和E. van Wyk，《小生境全局优化：系统性文献综述》，《算法》，第17卷，第10期，文章编号448，2024年。[44] D. Chauhan、Shivani、D. Jung和A. Yadav，《多模态差分进化的进展：全面综述与未来展望》，《人工智能评论》，第58卷，第11期，2025年，doi: 10.1007/s10462-025-11314-7。[45] W. Karunarathne、I. Bala、D. Chauhan、M. Roughan和L. Mitchell，《用于多模态优化的改进CMA-ES算法：融入小生境策略与动态自适应机制》，2024年，arXiv:2407.00939。[46] T. Le Minh、J. Arbel、T. M¨ollenhoff、M. E. Khan和F. Forbes，《用于多模态优化的自然变分退火》，2025年，arXiv:2501.04667。[47] T. Landais、O. Goudet、A. Go¨effon、F. Saubion和S. Lam-prier，《Stein变分黑箱组合优化》，2026年，arXiv:2604.15837。[48] Y. Zhang、B. Hartl、H. Hazan和M. Levin，《扩散模型是进化算法》，2024年，arXiv:2410.02543。[49] L. A. Zadeh，《模糊集》，《信息与控制》，第8卷，第3期，第338–353页，1965年。

[50] Q. Khaled、U. Kaymak和L. Genga，《面向可解释神经模糊系统的交替双目标优化》，2026年，arXiv:2602.19253。[51] W. Liu、J. Zhong、W.-G. Wu、J. Xiao和J. Zhang，《一种用于全局数值优化的新型模糊直方图估计分布算法》，载于国际软计算与模式识别会议（*SoCPaR*）论文集，2009年，第94–99页。[52] S. L. Thomson、Q. Renau、D. Vermetten、E. Hart、N. van Stein和A. V. Kononova，《空间停滞：适用于任意算法的吸引子分析》，2024年，arXiv:2412.15848。[53] D. Vermetten、F. Ye、T. B¨ack和C. Doerr，《MA-BBOB：一种利用仿射组合与平移生成黑箱优化问题的问题生成器》，2023年，arXiv:2312.11083。[54] U. ˇSkvorc、N. van Stein、M. Seiler、B. Grimme、T. B¨ack和H. Trautmann，《利用大语言模型设计具有可控高层属性的连续优化问题》，2026年，arXiv:2601.18846。[55] M. V. Seiler、P. Kerschke和H. Trautmann，《Deep-ELA：利用自监督预训练变换器对单目标和多目标连续优化问题进行深度探索性景观分析》，2024年，arXiv:2401.01192。[56] A. P. Dempster、N. M. Laird和D. B. Rubin，《通过期望最大化算法从不完整数据中进行最大似然估计》，皇家统计学会*B*辑杂志，第39卷，第1期，第1–38页，1977年。[57] K. M. Ashman、C. M. Bird和S. E. Zepf，《检测天文数据集中的双峰性》，天文学杂志，第108卷，第2348–2361页，1994年。[58] F. Wilcoxon，《基于秩方法的个体比较》，生物计量学通报，第1卷，第6期，第80–83页，1945年。[59] S. Holm，《一种简单的序贯拒绝多重检验程序》，斯堪的纳维亚统计学杂志，第6卷，第2期，第65–70页，1979年。[60] N. Hansen、S. Finck、R. Ros和A. Auger，《实参数黑箱优化基准测试2009：无噪声函数定义》，法国国家信息与自动化研究所，法国，技术报告RR-6829，2009年。