arXiv:2601.11499v1 [cs.NE] 16 Jan 2026

差分进化中首次成功的概率：危险恒等式与尾界

预印本

Dimitar Nedanovski
索非亚圣克莱门特·奥赫里德斯基大学数学与信息学院，詹姆斯·鲍彻大道，1164索非亚，保加利亚 dnedanovski@gmail.com

Svetoslav Nenov
化学技术与冶金大学数学系，索非亚，保加利亚 nenov@uctm.edu

Dimitar Pilev 保加利亚索非亚化学技术与冶金大学信息学系 pilev@uctm.edu

2025年6月

摘要

我们通过条件风险框架研究差分进化（DE）中的首次命中时间。我们不通过马尔可夫链转移核或漂移论证分析收敛，而是将可测目标集 A 的生存概率表示为条件首次命中概率（风险率）$p_t = \mathbb{P}(E_t \mid \mathcal{F}_{t-1})$ 的乘积。只要在存活事件上存在确定性的风险率下界，就能由此得到生存概率的无分布假设恒等式和明确的尾界。

针对采用 current-to-$p$best/1 变异的 L-SHADE算法，我们构造了一个可检验的算法见证事件 $\mathcal{L}_t$，在该事件下，条件风险率具有一个明确的下界，其取决因素仅包括采样规则、种群规模和交叉统计量。这样便将理论常数与经验事件频率区分开来，并解释了为什么最坏情况常数风险率界通常较为保守。

我们还通过对 CEC2017基准测试套件进行 Kaplan–Meier生存分析来补充理论结果。在不同函数和预算下，我们识别出三种截然不同的经验机制：（i）强聚类成功，命中时间集中在短促的突发阶段；（ii）近似几何尾部，此时常数风险模型较为准确；（iii）难以处理的情形，在评估时限内未观察到任何命中。结果表明，尽管常数风险率界能提供有效的尾部包络，L-SHADE 的实际表现仍由突发式转变主导，而非由均匀的逐代成功概率决定。

关键词：差分进化、首次命中时间、条件风险率、成功概率、尾界、平台函数、随机优化

1 引言

差分进化（DE）是一种广泛应用于连续优化的基于种群的算法。L-SHADE 等自适应变体在各类问题上都展现出强劲的实证性能，但对其收敛行为的理论理解仍然有限。一个主要难点在于，这类算法本质上是非平稳的：在优化过程中，种群规模、参数分布和档案内容都会动态变化，因此难以应用标准分析工具。

本文从首次命中时间的角度研究 L-SHADE。设

$$A_\varepsilon = \{x \in [l,u]^d : f(x) \le f^\star + \varepsilon\}, \qquad l < u,\ \varepsilon > 0$$

表示全局最小值 $f^\star$ 的 ε-子水平集。

令 τ_A 表示种群首次包含一个属于 A 的点时的代数。当运行在函数评估次数（NFEVs）达到预算 B 时终止时，我们还考虑按评估次数索引的命中时间 τ_A^{eval}，以及由（可能随时间变化的）种群规模决定的最大代数 $n(B)$。

我们的出发点是一个关于生存概率 $\mathbb{P}(\tau_A > n)$ 的精确恒等式，该恒等式用条件首次命中风险率表示。令 E_t 表示 A 首次命中发生在第 t 代的事件，并令 $\mathcal{F}_{t-1}$ 表示代后滤过，它编码截至时间 $t-1$ 的完整算法历史。定义条件首次命中概率（风险率）

$$p_t = \mathbb{P}(E_t \mid \mathcal{F}_{t-1}), \qquad t \ge 1,$$

以及存活事件 $H_{t-1} = \{\tau_A > t-1\}$。无需任何独立性或平稳性假设，生存函数可表示为仅基于存活的风险率 $h_t = \mathbb{P}(E_t \mid H_{t-1})$ 的乘积。只要能在存活条件下为 h_t 给出下界，即可得到分布无关尾界。

为了将这一视角转化为适用于 L-SHADE 的可用界，我们构造一个可测见证事件 L_t（仅取决于时间 $t-1$ 的状态），以此确认 current-to-$p$best/1 配合二项式交叉时存在有利配置。在 $H_{t-1} \cap L_t$ 上，我们证明条件风险率具有如下形式的确定性下界

$$p_t \ge a_t > 0,$$

其中 a_t 是显式的，且仅取决于：(i) 离散抽样规则（索引选择和记忆槽选择），(ii) 种群/档案大小，以及 (iii) 交叉尾概率。关键在于，这将单步成功概率分解为：

1. 一个显式的理论因子 $a_t > 0$（通常较小且较保守）；以及 2. 一个经验频率因子，用于描述运行过程中见证配置出现的频率。

记

$$\gamma_t = \mathbb{P}(L_t \mid H_{t-1}),$$

我们得到如下形式的仅基于存活的风险率下界

$$h_t \;=\; \mathbb{P}(E_t \mid H_{t-1}) \;\ge\; a_t\,\gamma_t,$$

因此可得一个显式的生存包络

$$\mathbb{P}(\tau_A > n) \;\le\; \mathbb{P}(E_0^c)\prod_{t=1}^{n}(1 - a_t\gamma_t) \;\le\; \mathbb{P}(E_0^c)\exp\Big(-\sum_{t=1}^{n} a_t\gamma_t\Big).$$

这种“常数 $\times$ 频率”分解清楚地说明了为什么最坏情况常数风险率界可能有效却偏保守：即使 a_t 很小，当 L_t 频繁发生（通常聚集在短簇中）时，累积强度 $\sum_{t\le n} a_t\gamma_t$ 仍可能变得很大。

在有限预算内理解成功的三步方案。本文围绕以下方案展开，该方案直接着眼于在有限评估预算内取得成功：

1. 预算内经验成功率（良好函数）。对于固定预算 B 函数评估次数，我们研究成功概率 $\mathbb{P}(\tau_{A_\varepsilon}^{\mathrm{eval}} \le B)$。对于该概率接近 1 的函数，在预算 B 下可称为“良好”函数，因为从实际意义上说，有限的函数评估次数预算便足以实现高概率成功。2. 存活条件下的见证频率。对于每一代 t，我们量化（原则上可通过重复运行进行估计）存活条件频率 $\gamma_t = \mathbb{P}(L_t \mid H_{t-1})$，即 L-SHADE 进入某种配置的频率；在该配置下，单步风险率可得到明确的可认证下界。3. 通过风险率乘积进行认证。将 $h_t \ge a_t\gamma_t$ 与精确的存活乘积相结合，可得出 $\mathbb{P}(\tau_{A_\varepsilon} > n)$ 的可认证尾界，并通过 $n = n(B)$，得出 $\mathbb{P}(\tau_{A_\varepsilon}^{\mathrm{eval}} > B)$ 的可认证尾界。因此，对于“良好”函数，在有限预算内取得高成功概率，可由累积见证强度的增长来解释$\sum_{t\le n\,(B)}\ a_t\gamma_t$。

经验生存机制。为将理论与实践联系起来，我们对 CEC2017 基准测试套件在维度 $d = 10$ 下、采用标准 L-SHADE 参数并使用两种评估预算的情形，进行 Kaplan–Meier 生存分析。生存曲线揭示了三种截然不同的经验机制：(i) 强聚类成功，生存函数在短时间内骤降；(ii) 近似几何尾部，此时常数风险模型较为准确；(iii) 不可处理情形，在评估时限内未观察到任何命中。这些结果支持了核心观点：数学上有效的常风险率包络往往较为保守，因为 L-SHADE 在实际中的成功取决于依赖配置的转变，而非同质的逐代进展。

贡献。我们的贡献包括：(i) 为差分进化（DE）的首次命中时间提出一种基于风险率的表述，并由此得到精确的生存恒等式和分布无关尾界；(ii) 为 L-SHADE/current-to-$p$best/1 构造一个具体的见证事件 L_t，从而得到一个明确的生存风险率下界；以及 (iii) 开展一项实证生存分析（Kaplan–Meier 估计和几何包络），刻画聚类型与近几何型成功机制，并解释最坏情况常数风险率界为何在观察中显得过于保守。

补充材料。补充材料可在 https://github.com/snenovgmailcom/lshade_hazard_project 在线公开获取，其中包含更多实证分析，包括命中时间处的簇几何结构，以及对风险界的 Kaplan–Meier 验证。

2 成功事件与概率空间

本节中的构造引入了相关记号。例如， [16, 第 2.1.4 节、第 4 节、2], [37], [14, 第 13 节、1], [2], [3, 第 33] 节中可以找到此类构造的完整标准版本。

令 $[l,u]^d$ 为搜索长方体，并令 $\left\{P^{(t)}\right\}_{t\geq 0}$ 表示所研究的差分进化变体所产生的种群过程，其中

$$P^{(t)} = \left\{\boldsymbol{x}_1^{(t)}, \ldots, \boldsymbol{x}_{\kappa_t}^{(t)}\right\} \subset [l,u]^{\kappa_t}$$

且 κ_t 为（可能随时间变化的）种群规模。

我们在滤过概率空间 $\left(\Omega\ \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\right)$ 上对算法建模，其中 $\mathcal{F}_t$ 表示完成第 t 代后可获得的信息（包括该代中使用的所有随机抽样，以及种群、档案和参数记忆等所有派生状态变量）。附录 A 给出了支持完全自适应性的具体概率乘积空间构造。

令 S_t 表示第 t 代结束后的完整算法状态（例如种群、目标函数值、档案、参数记忆）。我们采用自然的（代后）滤过

$$\mathcal{F}_t = \sigma\left(S_0, \ldots, S_t\right), \qquad t \geq 0.$$

固定一个勒贝格测度为正的可测目标集 $A \subset [l,u]^d$（通常是目标函数 f 的全局最小值集合的 ε 邻域，或相应的水平集定义）。

定义两两互斥的首次命中事件族

$$E_0 = \left\{(\exists i)\, x_i^{(0)} \in A\right\} \in \mathcal{F}_0,$$
$$E_t = \left\{(\forall s \leq t-1)(\forall i)\, x_i^{(s)} \notin A \text{ and } (\exists i)\, x_i^{(t)} \in A\right\} \in \mathcal{F}_t, \qquad t \geq 1,$$

以及首次命中时间

$$\tau_A = \inf\{t \geq 0 :\ E_t \text{ occurs}\} = \inf\{t \geq 0 :\ (\exists i)\, x_i^{(t)} \in A\}.$$

令存活事件为

$$H_t = \{\tau_A > t\} = \bigcap_{s=0}^{t} E_s^c \in \mathcal{F}_t, \qquad t \geq 0.$$

我们将条件首次命中概率（风险率）定义为

$$p_0 = \mathbb{P}(E_0), \qquad p_t = \mathbb{P}(E_t \mid \mathcal{F}_{t-1}), \quad t \geq 1$$

以及仅基于存活的风险率 $h_t = \mathbb{P}(E_t \mid H_{t-1}),\ t \geq 1$。显然，$H_{t-1} \in \mathcal{F}_{t-1}$，并且

$$h_t = \mathbb{E}[p_t \mid H_{t-1}] = \mathbb{E}[\mathbb{P}(E_t \mid \mathcal{F}_{t-1}) \mid H_{t-1}] = \mathbb{E}[p_t \mid H_{t-1}].$$

此外，由于 $E_t = H_{t-1} \cap \left\{(\exists i)\, x_i^{(t)} \in A\right\}$，因
此

$$p_t = \mathbf{1}_{H_{t-1}} \mathbb{P}\left((\exists i)\, x_i^{(t)} \in A \,\middle|\, \mathcal{F}_{t-1}\right), \qquad t \geq 1, \tag{1}$$

因此 $p_t = 0$ 自动不属于存活事件。

3 条件成功

本节中的所有恒等式和界都适用于任意可测目标集 A。引理 1。（见 *[6]*）对于每
个 $n \geq 0$,

$$\mathbf{1}_{\{\tau_A > n\}} = \prod_{t=0}^{n} \mathbf{1}_{E_t^c}, \quad \mathbb{P}(\tau_A > n) = \mathbb{E}\left[\prod_{t=0}^{n} \mathbf{1}_{E_t^c}\right]. \tag{2}$$

此外，对于每个 $n \geq 1$

$$\mathbb{P}(\tau_A > n \mid \mathcal{F}_{n-1}) = \mathbf{1}_{H_{n-1}}(1 - p_n), \tag{3}$$

$$\mathbb{P}(\tau_A > n) = \mathbb{E}\left[\mathbf{1}_{H_{n-1}}(1 - p_n)\right], \quad n \geq 1. \tag{4}$$

引理 2。（见 *[6]*）对于任意 $n \geq 0$,

$$\mathbb{P}(\tau_A > n) \leq \mathbb{E}\Big[\prod_{t=0}^{n}(1 - p_t)\Big].$$

定理 1 中的指数包络是经典的离散时间生存界：如果给定条件危险率 γ_t，那么根据标准的生存分析论证，
可得 $\mathbb{P}(\tau > n) \leq \prod_{t=0}^{n-1}(1 - \gamma_t) \leq \exp(-\sum_{t=0}^{n-1} \gamma_t)$。
定理 1。（i）假设存在一个实数序列 $\{a_t\}_{t \geq 1} \subset [0, 1]$，使得对于所有 $t \geq 1$,

$$p_t \geq a_t \quad \textit{on } H_{t-1} \textit{ almost surely}.$$

那么对于任意 $n \geq 1$,

$$\mathbb{P}(\tau_A > n) \leq \mathbb{P}(E_0^c) \prod_{t=1}^{n}(1 - a_t).$$

特别地，如果 $\sum_{t=1}^{\infty} a_t = \infty$，那么 $\mathbb{P}(\tau_A < \infty) = 1$。

（ii）假设存在一个实数序列 $\{b_t\}_{t \geq 1} \subset [0, 1]$，使得对于所有 $t \geq 1$,

$$p_t \leq b_t \quad \textit{on } H_{t-1} \textit{ almost surely}.$$

那么，对每个 $n \geq 1$,

$$\mathbb{P}(\tau_A > n) \geq \mathbb{P}(E_0^c) \prod_{t=1}^{n}(1 - b_t).$$

此外，如果对所有 t 和 $\sum_{t=1}^{\infty} b_t < \infty$ 都有 $b_t \in [0, 1)$，那么

$$\mathbb{P}(\tau_A = \infty) = \lim_{n \to \infty} \mathbb{P}(\tau_A > n) \;\geq\; \mathbb{P}(E_0^c) \prod_{t=1}^{\infty}(1 - b_t) \;>\; 0.$$

推论1. 如果 $a_t \equiv a \in (0, 1]$, ，那么

$$\mathbb{P}(\tau_A > n) \leq \mathbb{P}(E_0^c)(1 - a)^n, \qquad \mathbb{E}[\tau_A] \leq \sum_{n=0}^{\infty} \mathbb{P}(\tau_A > n) \leq \frac{\mathbb{P}(E_0^c)}{a}.$$

特别地，$\mathbb{E}[\tau_A] \leq 1/a$，因为 $\mathbb{P}(E_0^c) \leq 1$。

如果 $a_t = Ct^{-\alpha}$ 满足 C 且 $\alpha \in (0, 1)$，那么对所有 $n \geq 1$,

$$\mathbb{P}(\tau_A > n) \leq \mathbb{P}(E_0^c) \exp\left(-\frac{C}{1-\alpha}\left((n+1)^{1-\alpha} - 1\right)\right) = \mathbb{P}(E_0^c) \exp\left(\frac{C}{1-\alpha}\right) \exp\left(-\frac{C}{1-\alpha}(n+1)^{1-\alpha}\right).$$

如果 $a_t = Ct^{-1}$ 满足 $C > 0$，那么对于所有 $n \geq 1$,

$$\mathbb{P}(\tau_A > n) \leq \mathbb{P}(E_0^c)(n+1)^{-C}.$$

引理 3. 设 $\{L_t\}_{t\geq 1}$ 为满足 $L_t \in \mathcal{F}_{t-1}$的事件。假设对于所有 $t \geq 1$,

$$p_t \geq a_t \quad on\ H_{t-1} \cap L_t,$$

其中 $\{a_t\}$ 是数值的确定性序列。于是

$$h_t = \mathbb{P}(E_t \mid H_{t-1}) \geq a_t \mathbb{P}(L_t \mid H_{t-1}).$$

证明。由于 $H_{t-1} \in \mathcal{F}_{t-1}$，我们有

$$h_t = \mathbb{P}(E_t \mid H_{t-1}) = \mathbb{E}[p_t \mid H_{t-1}]\,.$$

在事件 $H_{t-1} \cap L_t$上，假设给出 $p_t \geq a_t$；而在 $H_{t-1} \cap L_t^c$ 上，我们只使用平凡界 $p_t \geq 0$。因此,

$$\mathbb{E}[p_t \mid H_{t-1}] \geq \mathbb{E}[a_t\, \mathbf{1}_{L_t} \mid H_{t-1}] = a_t\, \mathbb{P}(L_t \mid H_{t-1}),$$

这就证明了该结论。 □

这种分解将算法常数 $\{a_t\}$ 与经验频率 $\mathbb{P}(L_t|H_{t-1})$区分开来。本文余下部分将为 L-SHADE 构造具体的 L_t。

4 针对 current-to-*p*best/1 的 L-SHADE 事件

本节始终关注采用 current-to-$p$best/1 变异和二项式交叉的 L-SHADE，其目标是子水平集

$$A_\varepsilon = \{x \in [l,u]^d : f(x) \leq f^\star + \varepsilon\}.$$

我们在固定的第 $t \geq 1$ 代进行分析。为简化记号，我们省略时间下标，写作 x_i、 N、 P、 $\mathcal{A}$ 等，而不写作 $x_i^{(t-1)}$、 $N^{(t-1)}$、 $P_{(t-1)}$， $\mathcal{A}^{(t-1)}$ 表示第 t 代开始时的状态（等价地，即第 $t-1$ 代结束后的状态）。类似地， v_i、 u_i 表示在第 t 代期间生成的变异个体和试验向量 $v_i^{(t)}$ 、 $u_i^{(t)}$ 。存活事件为 $H_{t-1} = \{\tau_{A_\varepsilon} > t-1\}$。

4.1 记号与设置

种群与档案。设 N 表示当前种群大小。种群为

$$P = \{x_1, \ldots, x_N\} \subset [l,u]^d,$$

其索引集为 $\mathcal{P} = \{1,\ldots,N\}$。设 $\mathcal{A}$ 表示档案索引集，与 $\mathcal{P}$ 不相交，并对应向量 $\{x_j :\ j \in \mathcal{A}\}$。档案存储在前几代中被更优子代替换的父代向量；其最大大小为 $\lfloor \text{arc_rate} \times N_{\text{init}} \rfloor$（默认 arc_rate $= 2.6$）。完整的 L-SHADE 规范请参见 [40]。

p-最优集合。设 $\mathcal{B} \subseteq \mathcal{P}$ 表示目标值最小的 $m = \max\{1, \lceil pN \rceil\}$ 个个体的索引，其中 $p \in (0,1]$ 是固定参数（默认值为 $p = 0.11$）。

供体池。对于目标个体索引 $i \in \mathcal{P}$ 和所选的 p-最优索引 $b \in \mathcal{B}$，可用的供体集合为:

$$S_i^{(1)} = \mathcal{P} \setminus \{i,b\}, \qquad s_i^{(1)} = |S_i^{(1)}| = N-2,$$
$$S_i^{(2)} = (\mathcal{P} \cup \mathcal{A}) \setminus \{i,b\}, \qquad s_i^{(2)} = |S_i^{(2)}| = N + |\mathcal{A}| - 2.$$

第一个供体 r_1 从 $S_i^{(1)}$ （仅种群）中抽取，而第二个供体 r_2 从 $S_i^{(2)}$ （种群 $\cup$ 档案）中抽取。

历史记忆。L-SHADE 维护 H 个记忆槽（默认 $H = 6$），用于存储 $k \in \{1,\ldots,H\}$ 的位置参数（$M_F[k]$, $M_{CR}[k]$），初始值均为 0.5。为单个 i 生成参数时，会从 $\{1,\ldots,H\}$ 中均匀采样一个记忆槽 K_i，然后:

- $F_i \sim$ 柯西分布($M_F[K_i]$, 0.1)，若 ≤ 0 则重新生成，若 > 1 则截断为 1;
- $CR_i \sim \mathcal{N}$($M_{CR}[K_i]$, 0.1)，裁剪至 $[0,1]$。

变异（current-to-$p$best/1）。给定目标索引 i、 p-best 索引 b、供体索引（r_1， r_2）以及缩放因子 $F \in (0, 1]$,，边界处理前的变异向量为

$$\tilde{v}_i(F; b, r_1, r_2) = x_i + F\Delta_i(b, r_1, r_2), \tag{5}$$

其中差分向量为

$$\Delta_i(b, r_1, r_2) = (x_b - x_i) + (x_{r_1} - x_{r_2}).$$

因此，f 对于固定索引，变异个体沿着一条从父代 x_i 发出的仿射射线 $F \mapsto x_i + F\Delta_i$.

边界处理（中点修复）。令 BH : $\mathbb{R}^d \times [l, u]^d \to [l, u]^d$ 表示中点修复算子。对于候选解 $\tilde{v}$ 和父代 x，每个坐标都按如下方式修复:

$$\mathsf{BH}(\tilde{v}, x)_j = \begin{cases} (l_j + x_j)/2, & \text{if } \tilde{v}_j < l_j, \\ (u_j + x_j)/2, & \text{if } \tilde{v}_j > u_j, \\ \tilde{v}_j, & \text{otherwise.} \end{cases} \tag{6}$$

修复后的变异向量为

$$v_i(F; b, r_1, r_2) = \mathsf{BH}\big(\tilde{v}_i(F; b, r_1, r_2),\, x_i\big) \in [l, u]^d. \tag{7}$$

算子 BH 是博雷尔可测的（它在每个坐标上都是分段仿射的）。

二项式交叉。给定修复后的变异向量 v_i、父代 x_i、交叉率 $CR_i \in [0, 1]$ 和强制索引$J_i \sim$均匀分布$\{1, \ldots, d\}$，试验向量 u_i 按坐标定义如下:

$$(u_i)_j = \begin{cases} (v_i)_j, & \text{if } U_j \le CR_i \text{ or } j = J_i, \\ (x_i)_j, & \text{otherwise,} \end{cases}$$

其中 U_1、 . . . 、 U_d i id∼ 均匀分布(0, 1)。强制索引 J_i 确保至少有一个坐标取自变异个体。

选择。如果 $f(u_i) \le f(x_i)$，则试验个体会在下一代中替代父代；否则父代保留。当试验个体严格改进时（$f(u_i) < f(x_i)$），被替代的父代会被加入档案 $\mathcal{A}$。

种群规模缩减（LPSR）。选择后，种群规模按以下方式缩减

$$N_{\text{new}} = \text{round}\left(N_{\text{init}} + (N_{\min} - N_{\text{init}}) \frac{\text{NFE}}{\text{MaxNFE}} \right),$$

其中 $N_{\min} = 4$。若 $N_{\text{new}} < N$，则移除最差的个体。

引理 4。在存活事件 H_{t-1}上，我们有

$$E_t = H_{t-1} \cap \left\{ \exists i \in \{1, \ldots, N_{t-1}\} :\, u_i^{(t)} \in A_\varepsilon \right\}. \tag{8}$$

证明。根据存活的定义，对于当前种群中的每个父代，我们都有 $x_i^{(t-1)} \notin A_\varepsilon$，因此 $f(x_i^{(t-1)}) > f^\star + \varepsilon$对所有 i均成立。现在固定任意索引 i，并假设试验个体落入成功集: $u_{t,\ i} \in A_\varepsilon$。则 $f(u_{t,\ i}) \le f^\star + \varepsilon < f($ $x_i^{(t-1)})$，因此选择必定会接受该试验个体，即 $x_i^{(t)} = u_{t,\ i} \in A_\varepsilon$。

因此，在 H_{t-1} 上，某个试验个体进入 A_ε 的事件等价于下一代种群中包含一个属于 A_ε 的个体；这恰好就是首次命中事件 E_t （因为 H_{t-1} 意味着之前没有发生过命中）。这就证明了（8）。 □

备注1。如果实现在选择后应用种群缩减步骤（如 L-SHADE 的 LPSR），那么只要缩减操作仅移除最差的个体（目标值最大的个体），(8) 仍然成立。事实上，任何被接受的$u_{t,\ i} \in A_\varepsilon$ 均满足 $f(u_{t,\ i}) \le f^\star + \varepsilon$，因此属于最优个体之列，不会被移除最差个体的操作删除。

前述描述在给定随机输入的情况下确定性地规定了 L-SHADE 的算子。现在，我们将概率采样规则表述为形式化假设。

假设1（索引采样）。固定目标索引 $i \in \mathcal{P}$。在给定 $\mathcal{F}_{t-1}$ 的条件下:

1。 p最佳索引 b 从 $\mathcal{B}$中均匀采样，因此对于每个 $b^\star \in$， $\mathbb{P}(b = b^\star \mid \mathcal{F}_{t-1}) = 1/m$ $\mathcal{B}$.

2. 供体索引（r_1, r_2）通过双池选择进行采样：

$$r_1 \sim \mathrm{Unif}(S_i^{(1)}), \qquad r_2 \mid r_1 \sim \mathrm{Unif}(S_i^{(2)} \setminus \{r_1\}).$$

由于 $S_i^{(1)} \subseteq S_i^{(2)}$，对于任何满足 $r_1^\star \in S_i^{(1)}$ 和 $r_2^\star \in S_i^{(2)} \setminus \{r_1^\star\}$的可行对 $(r_1^\star, r_2^\star)$，

$$\mathbb{P}\big((r_1, r_2) = (r_1^\star, r_2^\star) \mid \mathcal{F}_{t-1}\big) = \frac{1}{s_i^{(1)}(s_i^{(2)} - 1)}.$$

请注意：假设1(2)中的概率非常小（例如，对于具有完整档案的 $d = 10$，$\approx 10^{-6}$）。在第 5 节中，我们将证明，最优点附近的种群集中能为莫尔斯函数带来紧得多的界。

4.2 一个 L-SHADE 事件与危险下界

现在我们定义一个可测事件（仅取决于时间 $t-1$ 时的状态），在该事件下，我们将为条件首次命中概率 $p_t = \mathbb{P}(E_t \mid \mathcal{F}_{t-1})$ 找到一个确定性的下界。

固定下来

$$0 < F^- < F^+ \le 1, \qquad \Delta_F > 0, \qquad c_{\mathrm{cr}} \in (0,1), \qquad g^- > 0, \qquad q^- \in (0,1].$$

这里，$[F^-, F^+] \subset (0,1]$ 是我们要求其具有密度下界的 F 工作范围；$\Delta_F > 0$ 是成功-F 窗口的最小勒贝格测度；$c_{\mathrm{cr}} \in (0,1)$ 是交叉率阈值（我们仅考虑满足 $CR_i \in [c_{\mathrm{cr}}, 1]$ 的非退化交叉事件）；$g^- > 0$ 是 F 密度的下界；$q^- \in (0,1]$ 是概率 $\mathbb{P}(CR_i \ge c_{\mathrm{cr}})$ 的一致下界。

对于每个 (i, b, r_1, r_2)，定义成功缩放因子的集合

$$\mathcal{I}_t(i, b, r_1, r_2) \;=\; \Big\{ F \in [F^-, F^+] :\; v_i^{(t)}(F; b, r_1, r_2) \in A_{\varepsilon/2} \Big\}. \tag{9}$$

我们强调，$\mathcal{I}_t(i, b, r_1, r_2)$ 是 $[F^-, F^+]$ 的子集，并且可以是区间的并集；$\lambda(\mathcal{I}_t)$ 表示其一维勒贝格测度（总长度）。

我们将其表述为以下假设。

假设2（参数采样）。固定一个目标索引 $i \in \mathcal{P}$。在给定 $\mathcal{F}_{t-1}$ 的条件下：

1. 记忆　　槽位 K_i 从 $\{1, \dots, H\}$ 中均匀采样，因此对于每个 k，$\mathbb{P}(K_i = k \mid \mathcal{F}_{t-1}) = 1/H$　.

2. 在给定 $(\mathcal{F}_{t-1}, K_i = k)$ 的条件下，随机变量 F_i 和 CR_i 相互独立，其中 F_i 在 $(0,1]$ 上具有密度 $g_k^F()$。

备注2。假设2与 L-SHADE 的实现相符：选取记忆索引 k，然后从以 $M_F[k]$ 为中心的截尾柯西分布中抽取 F，并从以 $M_{CR}[k]$ 为中心的截尾正态分布中抽取 CR。给定索引 k 后，数值 F 和 CR 条件独立。

假设3（二项式交叉）。在给定 $(\mathcal{F}_{t-1}, CR_i)$ 的条件下，试验个体 u_i 按以下方式生成：*:*

1. 从 $\{1, \dots, d\}$ 中均匀采样 J_i；2. 对每个 $j \ne J_i$，以概率 CR_i 从变异个体 v_i 继承坐标 j（各坐标之间相互独立）；3. 确定性地从变异个体继承坐标 J_i。

备注3。在采用二项式交叉的 L-SHADE 中，试验向量 u_i 可能除强制坐标 j_{rand} 外与变异向量 v_i 完全相同，因此采样得到的交叉率 CR_i 不会实际影响子代。在这种情况下，原始 L-SHADE算法会记录 $CR_i = \perp$，并在更新交叉率记忆时排除该值。本文中，每当我们以 CR_i 或涉及交叉的事件为条件时，这一条件化均默认限制在非退化交叉事件上，此时 CR_i 有定义。

现在，我们可以定义当前状态中存在见证配置这一事件。

定义 1（L-SHADE 见证事件 $\mathcal{L}_t$）。令 $\mathcal{L}_t \in \mathcal{F}_{t-1}$ 为以下事件：存在一个目标索引 $i \in \mathcal{P}$、一个 p最佳索引 $b \in \mathcal{B}$、供体索引 $r_1 \in S_i^{(1)}$ 和 $r_2 \in S_i^{(2)} \setminus \{r_1\}$，以及一个记忆槽 $k \in \{1, \dots, H\}$，使得：

$$\lambda\big(\mathcal{I}(i, b, r_1, r_2)\big) \ge \Delta_F, \tag{10}$$

$$\inf_{F \in [F^-, F^+]} g_k^F(F) \ge g^-, \tag{11}$$

$$\mathbb{P}(CR_i \ge c_{\mathrm{cr}} \mid \mathcal{F}_{t-1}, K_i = k) \ge q^-. \tag{12}$$

注 4。根据构造， $\mathcal{L}_t$ 仅取决于时间 $t-1$的状态：种群与档案决定索引集合以及差分向量 $\Delta_i(b, r_1, r_2)$，而 L-SHADE 记忆决定密度 g_k^F和 CR的分布。因此 $\mathcal{L}_t \in \mathcal{F}_{t-1}$。

然而，条件 (10) 涉及满足 $f(v_i(F)) \le f^\star + \varepsilon/2$ 的 F值集合，这取决于目标函数几何。在一般的黑盒情形下，若没有函数知识，就无法直接观测这一点；可以 (i) 对结构化的 f 进行解析验证，或 (ii) 在 F 的网格上数值评估 $f(v_i(F))$，以此进行近似。

4.3 $H_{t-1} \cap \mathcal{L}_t$ 上 p_t 的下界

对于整数 $r \in \{0, \dots, d-1\}$，定义事件

$$C_r(i) = \{\text{the trial } u_i \text{ inherits at least } d-r \text{ coordinates from } v_i\}. \tag{13}$$

等价地， $C_r(i)$ 表示“至多有 r 个坐标来自父代 x_i”。

引理5. 在假设3下，给定 $CR_i = c$，继承到 u_i 中的变异体坐标数为

$$1 + \mathrm{Bin}(d-1, c),$$

因此，对于任意 $r \in \{0, \dots, d-1\}$，

$$\mathbb{P}\big(C_r(i) \mid CR_i = c\big) = \mathbb{P}(\mathrm{Bin}(d-1, c) \ge d-r-1)\,.$$

特别地，定义确定性常数

$$\eta_r(d, c_{\mathrm{cr}}) = \mathbb{P}(\mathrm{Bin}(d-1, c_{\mathrm{cr}}) \ge d-r-1)\,. \tag{14}$$

那么在事件 $\{CR_i \ge c_{\mathrm{cr}}\}$上，

$$\mathbb{P}\big(C_r(i) \mid \mathcal{F}_{t-1}, CR_i\big) \ge \eta_r(d, c_{\mathrm{cr}}). \tag{15}$$

证明. 假设3表明：一个坐标被强制从变异个体中选取，其余 $d-1$ 个坐标中的每一个都以概率 c独立地从变异个体中选取。因此，变异体坐标的数量为 $1 + \mathrm{Bin}(d-1, c)$。事件 $C_r(i)$恰好是该数量至少为 $d-r$的事件，即$\mathrm{Bin}(d-1, c) \ge d-r-1$。这就得到了第一个恒等式。

对于单调性界（15），注意到映射 $c \mapsto \mathbb{P}(\mathrm{Bin}(d-1, c) \ge d-r-1)$关于 $c \in [0,1]$非递减，因此在 $\{CR_i \ge c_{\mathrm{cr}}\}$上， $C_r(i)$ 的条件概率至少为 $\eta_r(d, c_{\mathrm{cr}})$。 □

注记 5. 对于任意 $c_{\mathrm{cr}} \in (0,1)$，取 $r = d - 1 - \lfloor (d-1)c_{\mathrm{cr}} \rfloor$ 可得

$$\eta_r(d, c_{\mathrm{cr}}) \ge \frac{1}{2} \quad \text{uniformly in } d,$$

因为 $\mathrm{Bin}(n, p)$ 的中位数属于 $\{\lfloor np \rfloor, \lceil np \rceil\}$。对于 L-SHADE 初始化 $c_{\mathrm{cr}} = 1/2$，这给出 $r = \lfloor (d-1)/2 \rfloor$。

为了使用 $C_r(i)$，我们需要如下形式的蕴含关系（见引理5和（10））：

$$v_i \in A_{\varepsilon/2} \text{ and } \|x_i - v_i\|_\infty \le \delta \text{ and } C_r(i) \implies u_i \in A_\varepsilon.$$

定义 2. 固定 $\varepsilon > 0$ 和一个整数 $r \in \{0, \dots, d-1\}$。如果 f在 $[l,u]^d$上是（ε， r， δ）-交叉稳定的，就称其满足以下条件：对于每个 $v \in A_{\varepsilon/2}$、每个满足 $\|x - v\|_\infty \le \delta$的 $x \in [l,u]^d$， 以及每个满足 $|J| \ge d-r$的索引集 $J \subseteq \{1, \dots, d\}$，由下式定义的向量 u

$$u_j = \begin{cases} v_j & \text{if } j \in J, \\ x_j & \text{if } j \notin J, \end{cases}$$

满足 $u \in A_\varepsilon$。

命题 2 证明，每个莫尔斯函数都是 (ε, r, δ)-交叉稳定的，只要 ε 和 $\delta \le \delta_{\max}(\varepsilon, r)$ 足够小且取值适当。

$\gamma_{t,\ i}$ 分解为条件子事件，遵循适应过程的标准概率论论证；例如参见 [24]。

引理 6（ $H_{t-1} \cap \mathcal{L}_t$ 上的风险界）。假设：

1. 假设 1–3 成立。

2. 存在 $r \in \{0, \dots, d-1\}$，使得在事件 $\mathcal{L}_t$ 上，对于定义 1 中的见证目标索引 i 和见证索引 (b, r_1, r_2)，

$$v_i(F; b, r_1, r_2) \in A_{\varepsilon/2} \text{ and } \|x_i - v_i\|_\infty \le \delta \text{ and } C_r(i) \quad \Longrightarrow \quad u_i \in A_\varepsilon. \tag{16}$$

那么，在 $H_{t-1} \cap \mathcal{L}_t$ 上，

$$p_t = \mathbb{P}(E_t \mid \mathcal{F}_{t-1}) \ge a_t, \tag{17}$$

其中

$$a_t = \frac{1}{H}\frac{1}{m}\frac{1}{s_i^{(1)}(s_i^{(2)}-1)}(g^- \Delta_F)(q^- \eta_r(d, c_{\mathrm{cr}})). \tag{18}$$

这里，i 是来自 $\mathcal{L}_t$ 的见证目标索引。

证明。考虑事件 $H_{t-1} \cap \mathcal{L}_t$。由于 $\mathcal{L}_t \in \mathcal{F}_{t-1}$，对 $\mathcal{F}_{t-1}$ 进行条件化会将见证对象（i、集合 $\mathcal{B}$、$S_i^{(1)}$、$S_i^{(2)}$、窗口 $\mathcal{I}()$ 以及依赖时隙的分布）视为固定。

步骤 1：将 E_t 化简为单个个体的试验个体。根据引理 4，在 H_{t-1} 上有

$$E_t = H_{t-1} \cap \{\exists j : u_j \in A_\varepsilon\}.$$

特别地，

$$\{u_i \in A_\varepsilon\} \subseteq \{\exists j : u_j \in A_\varepsilon\},$$

因此在 H_{t-1} 上有 $\{u_i \in A_\varepsilon\} \subseteq E_t$。在给定 $\mathcal{F}_{t-1}$ 的条件下取条件概率，得到

$$p_t = \mathbb{P}(E_t \mid \mathcal{F}_{t-1}) \ge \mathbb{P}(u_i \in A_\varepsilon \mid \mathcal{F}_{t-1}). \tag{19}$$

步骤 2：为目标个体 i 定义一个具体的成功事件。令（$b^\star$、$r_1^\star$、$r_2^\star$、$k^\star$）为由 $\mathcal{L}_t$ 保证存在的见证四元组，使得（10）–（12）成立。考虑事件 $\mathcal{A}$：在为目标个体 i 构造试验个体时，算法恰好选取以下离散选择：

$$\mathcal{A} = \{b = b^\star\} \cap \{(r_1, r_2) = (r_1^\star, r_2^\star)\} \cap \{K_i = k^\star\}.$$

根据假设 1 和 2，

$$\mathbb{P}(\mathcal{A} \mid \mathcal{F}_{t-1}) = \frac{1}{m}\frac{1}{s_i^{(1)}(s_i^{(2)}-1)}\frac{1}{H}. \tag{20}$$

步骤 3：界定命中成功-F 窗口的概率。在 $\mathcal{A}$ 上，变异个体具有固定的仿射形式 $F \mapsto v_i$（F；$b^\star$、$r_1^\star$、$r_2^\star$）。根据见证条件（10），集合 $\mathcal{I}$（i、$b^\star$、$r_1^\star$、$r_2^\star$）的勒贝格测度至少为 Δ_F。根据（11），槽位 $k^\star$ 下的 F-密度对所有 $F \in [F^-, F^+]$ 均满足 $g_{k^\star}^F(F) \ge g^-$。因此，

$$\begin{aligned}
\mathbb{P}\big(F_i \in \mathcal{I}(i, b^\star, r_1^\star, r_2^\star) \mid \mathcal{F}_{t-1}, \mathcal{A}\big) &= \int_{\mathcal{I}(i,b^\star,r_1^\star,r_2^\star)} g_{k^\star}^F(F)\, dF \\
&\ge \int_{\mathcal{I}(i,b^\star,r_1^\star,r_2^\star)} g^-\, dF = g^- \lambda\big(\mathcal{I}(i, b^\star, r_1^\star, r_2^\star)\big) \\
&\ge g^- \Delta_F.
\end{aligned} \tag{21}$$

步骤 4：界定交叉概率。根据（12），

$$\mathbb{P}(CR_i \ge c_{\mathrm{cr}} \mid \mathcal{F}_{t-1}, \mathcal{A}) \ge q^-.$$

在 $(\mathcal{F}_{t-1}, \mathcal{A}, CR_i)$ 的条件下应用引理5：在 $\{CR_i \ge c_{\mathrm{cr}}\}$ 上，$C_r(i)$ 的条件概率至少为 $\eta_r(d, c_{\mathrm{cr}})$。因此，

$$\begin{aligned}
\mathbb{P}\big(C_r(i) \cap \{CR_i \ge c_{\mathrm{cr}}\} \mid \mathcal{F}_{t-1}, \mathcal{A}\big) &= \mathbb{E}\big[\mathbf{1}_{\{CR_i \ge c_{\mathrm{cr}}\}} \mathbb{P}(C_r(i) \mid \mathcal{F}_{t-1}, \mathcal{A}, CR_i) \,\big|\, \mathcal{F}_{t-1}, \mathcal{A}\big] \\
&\ge \mathbb{E}\big[\mathbf{1}_{\{CR_i \ge c_{\mathrm{cr}}\}} \eta_r(d, c_{\mathrm{cr}}) \,\big|\, \mathcal{F}_{t-1}, \mathcal{A}\big] \\
&= \eta_r(d, c_{\mathrm{cr}}) \mathbb{P}(CR_i \ge c_{\mathrm{cr}} \mid \mathcal{F}_{t-1}, \mathcal{A}) \\
&\ge q^- \eta_r(d, c_{\mathrm{cr}}).
\end{aligned} \tag{22}$$

步骤5：合并界。考虑事件

$$\mathcal{S} = \mathcal{A} \cap \{F_i \in \mathcal{I}(i, b^\star, r_1^\star, r_2^\star)\} \cap \{CR_i \geq c_{\mathrm{cr}}\} \cap C_r(i).$$

在 $\mathcal{S}$ 上，根据 $\mathcal{I}$ 的定义，我们有 $v_i(F_i;\ b^\star, r_1^\star, r_2^\star) \in A_{\varepsilon/2}$。结合 $C_r(i)$，稳定性蕴含式（16）给出 $u_i \in A_\varepsilon$。因此 $\mathcal{S} \subseteq \{u_i \in A_\varepsilon\}$，从而

$$\mathbb{P}(u_i \in A_\varepsilon \mid \mathcal{F}_{t-1}) \geq \mathbb{P}(\mathcal{S} \mid \mathcal{F}_{t-1}).$$

使用链式法则（以 $\mathcal{A}$ 为条件化对象）以及假设2（给定槽位时 F_i 与 CR_i 的条件独立性），我们将界（20）、（21）和（22）相乘：

$$\begin{aligned}\mathbb{P}(\mathcal{S} \mid \mathcal{F}_{t-1}) &= \mathbb{P}(\mathcal{A} \mid \mathcal{F}_{t-1})\mathbb{P}\big(F_i \in \mathcal{I}() \mid \mathcal{F}_{t-1}, \mathcal{A}\big)\mathbb{P}\big(C_r(i) \cap \{CR_i \geq c_{\mathrm{cr}}\} \mid \mathcal{F}_{t-1}, \mathcal{A}\big) \\ &\geq \frac{1}{H}\frac{1}{m}\frac{1}{s_i^{(1)}(s_i^{(2)}-1)}(g^- \Delta_F)(q^- \eta_r(d, c_{\mathrm{cr}})).\end{aligned}$$

结合（19）可得（17），其中 a_t 如（18）所示。 □

注释 6（界（18）的结构）。固定阈值（$F^-, F^+, \Delta_F, c_{\mathrm{cr}}, g^-, q^-$）和整数 r。下界（18）分解为三类因子。

1. 算法/组合因素（可计算）。根据离散抽样规则：

$$\frac{1}{H}, \qquad \frac{1}{m}, \qquad \frac{1}{s_i^{(1)}(s_i^{(2)}-1)}.$$

对于初始种群规模为 $N = N_{\mathrm{init}} = 18d$ 时的 L-SHADE 默认设置（$H = 6$, $p = 0.11$），我们有$m = \max\{1, \lceil pN \rceil\}$（例如，对于 $d = 10, m = \lceil 0.11180 \rceil = 20$，因此 $1/m = 0.05$）。
对于双池采样（假设1），

$$s_i^{(1)} = N - 2, \qquad s_i^{(2)} = N + |\mathcal{A}| - 2.$$

当档案容量为 arc_、比率为 $= 2.6$ 且 $d = 10$ （因此 $N = 180$ 且 $|\mathcal{A}| \approx 2.6N_{\mathrm{init}} = 468$）时，我们有$s_i^{(1)} = 178$ 且 $s_i^{(2)} = 646$，因此

$$\frac{1}{s_i^{(1)}(s_i^{(2)}-1)} = \frac{1}{178645} \approx 8.7 \times 10^{-6}.$$

因此，

$$\frac{1}{H}\frac{1}{m}\frac{1}{s_i^{(1)}(s_i^{(2)}-1)} \approx \frac{1}{6}\frac{1}{20}8.7 \times 10^{-6} \approx 7.3 \times 10^{-8}.$$

请注意，由于种群规模缩减（LPSR），m、$s_i^{(1)}$ 和 $s_i^{(2)}$ 通常会随时间推移而减小。
交叉因子 $\eta_r(d, c_{\mathrm{cr}}) = \mathbb{P}(\mathrm{Bin}(d-1, c_{\mathrm{cr}}) \geq d-r-1)$ 是一个可通过数值计算求得的二项式尾概率。

2. 可由状态检查的因子。见证条件 (11)–(12) 仅取决于 L-SHADE 的记忆状态，因此

$$\inf_{F \in [F^-, F^+]} g_k^F(F), \qquad \mathbb{P}(CR_i \geq c_{\mathrm{cr}} \mid \mathcal{F}_{t-1}, K_i = k)$$

可根据算法内部状态进行检查。

3. 依赖问题的因子。量 $\lambda(\mathcal{I}(i, b, r_1, r_2))$ 取决于目标函数几何，若没有函数知识便无法计算（不过，对于结构化的 f，可以通过解析方法给出界）。

4. 经验频率。存活条件频率 $\gamma_t = \mathbb{P}(\mathcal{L}_t \mid H_{t-1})$ 反映了算法与问题之间的相互作用，必须通过经验估计。

4.4 基于事件频率的尾界

定义确定性的仅基于存活的风险率

$$h_t = \mathbb{P}(E_t \mid H_{t-1}), \qquad t \geq 1.$$

于是，存活事件满足精确的递推关系

$$\mathbb{P}(H_t) = \mathbb{P}(H_{t-1})\mathbb{P}(H_t \mid H_{t-1}) = \mathbb{P}(H_{t-1})\big(1 - h_t\big),$$

因此，迭代可得，

$$\mathbb{P}(\tau_{A_\varepsilon} > n) = \mathbb{P}(E_0^c) \prod_{t=1}^{n} (1 - h_t). \tag{23}$$

现在定义存活条件下的 L-SHADE 事件频率为

$$\gamma_t = \mathbb{P}(\mathcal{L}_t \mid H_{t-1}) \in [0, 1].$$

命题1. 假设引理 6 的条件成立，并假设 (18) 中的 a_t 是确定性的。那么对每个 $t \geq 1$,

$$h_t = \mathbb{P}(E_t \mid H_{t-1}) = \mathbb{E}[p_t \mid H_{t-1}] \;\geq\; a_t\, \mathbb{P}(\mathcal{L}_t \mid H_{t-1}) = a_t\, \gamma_t,$$

因此，对所有 $n \geq 1$,

$$\mathbb{P}(\tau_{A_\varepsilon} > n) \leq \mathbb{P}(E_0^c) \prod_{t=1}^{n} (1 - a_t\gamma_t) \leq \mathbb{P}(E_0^c) \exp\left(- \sum_{t=1}^{n} a_t \gamma_t \right). \tag{24}$$

特别地，如果 $\sum_{t\geq 1} a_t\gamma_t = \infty$，那么 $\mathbb{P}(\tau_{A_\varepsilon} < \infty) = 1$。

证明. 根据定义， $p_t = \mathbb{P}(E_t \mid \mathcal{F}_{t-1})$ 且 $H_{t-1} \in \mathcal{F}_{t-1}$，所以

$$h_t = \mathbb{P}(E_t \mid H_{t-1}) = \mathbb{E}[\mathbf{1}_{E_t} \mid H_{t-1}] = \mathbb{E}[\mathbb{E}[\mathbf{1}_{E_t} \mid \mathcal{F}_{t-1}] \mid H_{t-1}] = \mathbb{E}[p_t \mid H_{t-1}],$$

这正是塔式性质。

在事件 $H_{t-1} \cap \mathcal{L}_t$ 上，引理 6 给出 $p_t \geq a_t$。因此，

$$\mathbb{E}[p_t \mid H_{t-1}] \geq \mathbb{E}[a_t\, \mathbf{1}_{\mathcal{L}_t} \mid H_{t-1}] = a_t\, \mathbb{P}(\mathcal{L}_t \mid H_{t-1}) = a_t\gamma_t,$$

这里用到了 a_t 是确定性的。将 $h_t \geq a_t\gamma_t$ 代入精确的乘积恒等式 (23)，得到乘积界；再利用 $1 - x \leq e^{-x}$ 得到指数界。最后的结论由以下事实得出：若 $\sum_t a_t\gamma_t = \infty$，则 $\exp(-\sum_{t=1}^{n} a_t\gamma_t) \to 0$，因此 $\mathbb{P}(\tau_{A_\varepsilon} = \infty) = 0$。 □

注 7. 界 (24) 可分解为以下两部分：

- 一个理论因子 a_t （明确给出，但通常较小），以及
- 一个经验频率 $\gamma_t = \mathbb{P}(\mathcal{L}_t \mid H_{t-1})$，用于衡量见证配置出现的频率。

即使 a_t 很小，当 $\mathcal{L}_t$ 频繁出现时，累积求和 $\sum_{t\leq n} a_t\gamma_t$ 仍可能变得很大。这解释了为什么最坏情况界较为保守，而几乎必然收敛仍然成立。

5 莫尔斯函数

设 $D \subset \mathbb{R}^d$ 为一个有界域，使得 $[l, u]^d \subset D$，并令 $f \in C^2(D, \mathbb{R})$。回顾一下，若 f 是 D 上的莫尔斯函数，则其所有临界点均为非退化的，即 $\nabla f(x) = 0$ 蕴含 $\det\left(\nabla^2 f(x)\right) \neq 0, x \in \text{int}(D)$。

众所周知，定义在 $[l, u]^d$ 上的莫尔斯函数只有有限个临界点，因此也只有有限个局部极小值。

设 $x^\star \in \text{Int}\,([l, u]^d)$ 是 f 的一个局部极小值。根据泰勒定理，存在 $x^\star$ 的一个邻域，使得

$$f(x^\star + y) = f(x^\star) + \frac{1}{2}\, y^\top H(x^\star)\, y + o(\|y\|^2), \tag{25}$$

其中 $H(x^\star) = \nabla^2 f(x^\star)$ 是 $x^\star$ 处的海森矩阵。全文中，对于对称矩阵 A, B，我们用 $A \preceq B$ 表示 $B - A$ 是半正定的。

由于 $x^\star$ 是局部极小值且 f 是莫尔斯函数，海森矩阵 $H(x^\star)$ 是正定的。令

$$\mu_0 = \lambda_{\min}\big(H(x^\star)\big) > 0, \qquad L_0 = \lambda_{\max}\big(H(x^\star)\big) < \infty.$$

注释 8. 存在常数 $0 < \mu \le L < \infty$ 和半径 $r_0 > 0$，使得

$$\mu I \preceq \nabla^2 f(x) \preceq LI, \qquad \forall x \in B(x^\star, r_0).$$

这由 $\nabla^2 f$ 和 $H(x^\star) \succ 0$的连续性得出；例如，当 r_0 足够小时，可以取 $\mu = \mu_0/2$ 和 $L = 2L_0$。

比率 $\kappa = \frac{L}{\mu} \ge 1$ 是极小值 $x^\star$邻域内海森矩阵的条件数 .

函数 f 具有 L-利普希茨连续梯度，并且在 $B(x^\star, r_0)$上是 μ-强凸的。对于 $\varepsilon > 0$，定义局部次水平集

$$A_\varepsilon = \{x \in B(x^\star, r_0) : f(x) \le f(x^\star) + \varepsilon\}.$$

若 $2\varepsilon \le \mu r_0^2$，则 A_ε 是凸的，并且满足

$$B\left(x^\star, \sqrt{\tfrac{2\varepsilon}{L}}\right) \subseteq A_\varepsilon \subseteq B\left(x^\star, \sqrt{\tfrac{2\varepsilon}{\mu}}\right). \tag{26}$$

在本节中，我们始终假设 $2\varepsilon \le \mu r_0^2$。

注释 9（以二次增长替代莫尔斯假设）。第 5 节中的论证不依赖全局莫尔斯理论（鞍点/极大值点的非退化性）。对于相关的极小化点集 $X^\star$，只需存在常数 μ、$L > 0$和 $X^\star$的一个邻域 U，使得对所有 $x \in U$，

$$f^\star + \frac{\mu}{2}\operatorname{dist}(x, X^\star)^2 \le f(x) \le f^\star + \frac{L}{2}\operatorname{dist}(x, X^\star)^2.$$

下界是一致二次增长条件；对于非光滑目标函数，可以通过极限次微分 [12, 35]的倾斜稳定性/度量正则性来表述。在光滑情形下，这些局部界可由二阶充分条件 $\nabla f(x^\star) = 0$和 $\nabla^2 f(x^\star) \succ 0$推出，同时还需 $\nabla^2 f$ [32]的连续性；在非退化的莫尔斯极小点 [28]处，这一连续性也能得到保证。

5.1 一些预备结果

引理 7. 设 f 是一个 莫尔斯函数，并令数值 ε 的选取满足 $A_{\varepsilon/2} \subset B(x^\star, r_0)$。若 $x \in A_{\varepsilon/4}$，则

$$B\left(x,\, r_{\mathrm{safe}}\right) \subseteq A_{\varepsilon/2}, \qquad r_{\mathrm{safe}} = \left(1 - \frac{1}{\sqrt{2}}\right)\sqrt{\frac{\varepsilon}{L}}. \tag{27}$$

证明。令 $y \in B(x, r_{\mathrm{safe}})$，并设 $t = \|y - x\|$。由于 $x \in A_{\varepsilon/4} \subset B(x^\star, r_0)$，海森矩阵界 $\nabla^2 f \preceq LI$ 适用，且由泰勒定理可得

$$f(y) \le f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}t^2.$$

此外，由于 f 在 $B(x^\star, r_0)$ 上是 L-光滑的，且最小化点为 $x^\star$，我们有 $\|\nabla f(x)\|^2 \le 2L(f(x) - f^\star)$，对所有 $x \in B(x^\star, r_0)$ 均成立。因此，对于 $x \in A_{\varepsilon/4}$，

$$\|\nabla f(x)\| \le \sqrt{2L\frac{\varepsilon}{4}} = \sqrt{\frac{L\varepsilon}{2}}.$$

因此，

$$f(y) - f(x) \le t\sqrt{\frac{L\varepsilon}{2}} + \frac{L}{2}t^2.$$

对于 $t \le r_{\mathrm{safe}}$，右侧至多为 $\varepsilon/4$，因此

$$f(y) \le f(x) + \frac{\varepsilon}{4} \le f^\star + \frac{\varepsilon}{2}. \quad \square$$

命题 2. 设 f 为一个莫尔斯函数，并设数值 ε 的选取满足 $A_{\varepsilon/2} \subset B(x^\star, r_0)$。定义 $\delta_{\max}(\varepsilon, r) = (\sqrt{2} - 1)\sqrt{\frac{\varepsilon}{Lr}}$。

那么，对于每个 $0 \le \delta \le \delta_{\max}(\varepsilon, r)$，$f$ 在 $[l, u]^d$ 上是 (ε, r, δ)-交叉稳定的（见定义 2） .

证明。由于 u 与 v 至多在 r 个坐标上不同，且满足 $|u_j - v_j| \le \delta$，因此我们有 $\|u - v\| \le \sqrt{r}\,\delta$。
对于 $v \in A_{\varepsilon/2} \subset B(x^\star, r_0)$，强凸性给出 $\|v - x^\star\| \le \sqrt{\varepsilon/\mu}$。对于 $\delta \le \delta_{\max}(\varepsilon, r)$，

$$\|u - x^\star\| \le \|u - v\| + \|v - x^\star\| \le (\sqrt{2} - 1)\sqrt{\frac{\varepsilon}{L}} + \sqrt{\frac{\varepsilon}{\mu}} \le \sqrt{\frac{2\varepsilon}{\mu}} \le r_0,$$

因此 $u \in B(x^\star, r_0)$。
由于 f 在 $B(x^\star, r_0)$ 上是凸且 L-光滑的，标准梯度界 $\|\nabla f(v)\|^2 \le 2L(f(v) - f^\star)$ 成立，因此有$\|\nabla f(v)\| \le \sqrt{L\varepsilon}$。泰勒展开得

$$f(u) \le f(v) + \sqrt{L\varepsilon}\sqrt{r}\,\delta + \frac{L}{2} r\delta^2.$$

由于 $f(v) \le f^\star + \varepsilon/2$，条件 $f(u) \le f^\star + \varepsilon$可化简为

$$\sqrt{L\varepsilon}\sqrt{r}\,\delta + \frac{Lr\delta^2}{2} \le \frac{\varepsilon}{2}.$$

代入 $t = \sqrt{Lr/\varepsilon}\,\delta$ 可得 $t + t^2/2 \le 1/2$，这对 $t \le \sqrt{2} - 1$成立。 □

在陈述下一个结果之前，我们先单独讨论一个特殊的几何情形，它揭示了成功的供体步背后的核心机制。我们首先分析情形 $i = b$，此时 current-to-$p$best/1变异简化为一条从 x_b出发的纯供体射线。

只有当运行过程中产生了一个位于更深子水平集 $A_{\varepsilon/4}$ 中的 p-best 供体后，命题 3 才适用。根据经验，在稳定的开发利用阶段，这并非罕见事件：在补充材料 SM–II.2 中，我们展示了（CEC2017，$D = 10$ 和 $D = 30$），当最优个体首次进入 $A_{\varepsilon/4}$ 时，种群中已有很大一部分位于周围的吸引域次水平集中。因此，p-best 候选池包含这个深层点，并且在 L-SHADE 所采用的种群规模下，选择它作为 x_b的概率有一个严格大于零的下界。
命题 3. 设 f 是一个莫尔斯函数，并假设 $\varepsilon > 0$ 满足 $A_{\varepsilon/2} \subset B(x^\star, r_0)$。固定指标 $b \in \mathcal{P}$ 和 $(r_1, r_2) \in S_t^{(1)}{}_{,b} \times S_{t\,b}^{(2)}$,且 $r_2 \ne r_1$，并假设 $x_b \in A_{\varepsilon/4}$。定义（未经边界修复的）变异向量射线

$$\tilde{v}_b(F) = x_b + F\,(x_{r_1} - x_{r_2}), \qquad F \in [0, 1].$$

任取 $0 < F^- < F^+ \le 1$ 和任意 $\Delta_F \in (0, F^+ - F^-]$。如果

$$\|x_{r_1} - x_{r_2}\| \le \frac{r_{\mathrm{safe}}}{F^- + \Delta_F}, \tag{28}$$

then

$$[F^-,\, F^- + \Delta_F] \subseteq \mathcal{I}_t(b, b, r_1, r_2), \quad \textit{i.e.} \quad \lambda\big(\mathcal{I}_t(b, b, r_1, r_2)\big) \ge \Delta_F.$$

证明。设 $F \in [F^-, F^- + \Delta_F]$。那么

$$\|\tilde{v}_b(F) - x_b\| = F\,\|x_{r_1} - x_{r_2}\| \le (F^- + \Delta_F)\,\frac{r_{\mathrm{safe}}}{F^- + \Delta_F} = r_{\mathrm{safe}}.$$

因此 $\tilde{v}_b(F) \in B(x_b, r_{\mathrm{safe}})$。由于 $x_b \in A_{\varepsilon/4}$，由引理 7 可得 $B(x_b, r_{\mathrm{safe}}) \subseteq A_{\varepsilon/2}$，因此 $\tilde{v}_b(F) \in A_{\varepsilon/2}$。因此，对每个 $F \in [F^-, F^- + \Delta_F]$，都有 $F \in \mathcal{I}_t(b, b, r_1, r_2)$，这证明了该包含关系。测度界立即可得。 □

引理 8. 假设 f 在 $\{x_b + \tau(v_b(F_0) - x_b) : \tau \in [0, 1]\}$ 上是 μ–强凸的。设

$$x_b \in A_\varepsilon \qquad \textit{and} \qquad v_b(F_0) \in A_{\varepsilon/4}$$

对于某个 $F_0 \in (0, 1]$。定义

$$v_b(F) = x_b + \frac{F}{F_0}\big(v_b(F_0) - x_b\big), \qquad c = \frac{\mu\|v_b(F_0) - x_b\|^2}{8\varepsilon}.$$

那么，成功–F区间 $\mathcal{I}_t(b, b, r_1, r_2)$ 满足

$$\lambda(\mathcal{I}_t) \;\ge\; \frac{F_0}{3}.$$

此外，如果 $\theta_-(c) \in (0, 1)$ 是 $16c\theta^2 - (3 + 16c)\theta + 2 = 0$ 的较小根，那么

$$\lambda(\mathcal{I}_t) \;\ge\; F_0\,(1 - \theta_-(c)), \qquad \lim_{c \downarrow 0} \theta_-(c) = 2/3.$$

证明。设 $z = v_b(F_0) - x_b$，并记

$$v_b(\theta F_0) = (1-\theta)x_b + \theta v_b(F_0), \qquad \theta \in [0,1].$$

由 f 在线段 $[x_b, v_b(F_0)]$, 上的强凸性

$$f(v_b(\theta F_0)) \le (1-\theta)f(x_b) + \theta f(v_b(F_0)) - \frac{\mu}{2}\theta(1-\theta)\|z\|^2.$$

利用 $f(x_b) \le f^\star + \varepsilon$ 和 $f(v_b(F_0)) \le f^\star + \varepsilon/4$，我们得到

$$f(v_b(\theta F_0)) - f^\star \le \varepsilon\Big(1 - \frac{3}{4}\theta\Big) - 4\varepsilon c\,\theta(1-\theta).$$

因此， $v_b(\theta F_0) \in A_{\varepsilon/2}$ 的一个充分条件是

$$\frac{3}{4}\theta + 4c\,\theta(1-\theta) \;\ge\; \frac{1}{2}. \tag{29}$$

由于对所有 $\theta \in [0,1]$， $\theta(1-\theta) \ge 0$，因此 (29) 左侧有下界 $\frac{3}{4}\theta$。因此，只要$\frac{3}{4}\theta \ge \frac{1}{2} \iff \theta \ge \frac{2}{3}$.，(29) 就成立。因此，所有 $\theta \in [2/3, 1]$ 都满足 $v_b(\theta F_0) \in A_{\varepsilon/2}$，且

$$\lambda(\mathcal{I}_t) \;\ge (1 - 2/3)F_0 = \frac{F_0}{3}.$$

将式 (29) 改写为二次不等式

$$16c\,\theta^2 - (3 + 16c)\theta + 2 \;\le\; 0.$$

对于 $c > 0$，当 $\theta \in [\theta_-(c)\ \theta_+(c)]$,时，此式成立，其中 $\theta_-(c) < \theta_+(c)$ 是两个根。由于 $\theta \in [0,1]$，只需取$\theta \ge \theta_-(c)$，即可得到

$$\lambda(\mathcal{I}_t) \;\ge\; F_0\,(1 - \theta_-(c)). \quad \square$$

注释10. 命题 3 和引理 8 是互为补充的工具，而非用于证明成功–F 区间长度为正的替代方案：

$$\text{(donor concentration)} \implies \text{Prop. 3 :}\ \lambda(\mathcal{I}_t) \ge \Delta_F > 0 \implies \text{Lemma 8 (deep hit) :}\ \lambda(\mathcal{I}_t) \ge \frac{F_0}{3} \gg \Delta_F.$$

注11. 式 (6) 中定义的中点修复算子 BH 关于父代是非扩张的：对于所有 $x \in [l,u]^d$ 和所有 $\tilde{v} \in \mathbb{R}^d$,

$$\|\mathsf{BH}(\tilde{v}, x) - x\| \le \|\tilde{v} - x\|.$$

因此，如果 $\|\tilde{v}_b(F) - x_b\| \le r_{\text{safe}}$，那么修复后的变异向量

$$v_b(F) = \mathsf{BH}\big(\tilde{v}_b(F), x_b\big)$$

也满足 $\|v_b(F) - x_b\| \le r_{\text{safe}}$，且属于 $[l,u]^d$。因此，根据引理 7， $v_b(F) \in A_{\varepsilon/2}$，命题 3 的结论在中点修复下仍然成立。

定义情形 $i = b$ 下的 Δ_F-良好配对集合：

$$\mathcal{R}_{t,\Delta_F}(b) = \big\{(r_1, r_2) \in S_{t,b}^{(1)} \times S_{t,b}^{(2)} : \lambda(\mathcal{I}_t(b, b, r_1, r_2)) \ge \Delta_F\big\}.$$

根据命题 3，任何满足 $\|x_{r_1} - x_{r_2}\| \le r_{\text{safe}}/(F^- + \Delta_F)$ 的配对都属于 $\mathcal{R}_{t,\ \Delta_F}(b)$。

假设 4。存在子集 $C_1 \subseteq S_{t\ b}^{(1)}$, 和 $C_2 \subseteq S_{t\ b}^{(2)}$, , 满足 $|C_1| \ge \beta_1 s_{t\ b}^{(1)}$, 且 $|C_2| \ge \beta_2 s_{t\ b}^{(2)}$, , 使得

$$\max_{r \in C_1 \cup C_2} \|x_r - x_b\| \le \frac{r_{\text{safe}}}{2(F^- + \Delta_F)}.$$

命题 4。在假设 4 下， Δ_F-良好供体对所占的比例满足

$$c_{\text{pair}}(t) = \frac{|\mathcal{R}_{t,\Delta_F}(b)|}{s_{t,b}^{(1)}(s_{t,b}^{(2)} - 1)} \;\ge\; \beta_1\beta_2\left(1 - \frac{1}{\beta_2 s_{t,b}^{(2)}}\right). \tag{30}$$

特别是,

$$c_{\text{pair}}(t) \ge \beta_1\beta_2\,(1 + o(1)) \quad \textit{as}\ s_{t,b}^{(2)} \to \infty.$$

证明。设 $r_1 \in C_1$ 和 $r_2 \in C_2$ 满足 $r_2 \neq r_1$。根据三角不等式,

$$\|x_{r_1} - x_{r_2}\| \leq \|x_{r_1} - x_b\| + \|x_{r_2} - x_b\| \leq 2\frac{r_{\text{safe}}}{2(F^- + \Delta_F)} = \frac{r_{\text{safe}}}{F^- + \Delta_F}.$$

根据命题 3，这样的一对属于 $\mathcal{R}_{t,\ \Delta_F}(b)$。

对于每个 $r_1 \in C_1$，至少有 $|C_2| - 1$ 个 $r_2 \in C_2$ 可行选择（排除条件 $r_2 = r_1$ 至多会移除一个元素）。因此,

$$|\mathcal{R}_{t,\Delta_F}(b)| \ \geq \ |C_1|\,(|C_2| - 1) \ \geq \ \beta_1 s_{t,b}^{(1)}\big(\beta_2 s_{t,b}^{(2)} - 1\big).$$

除以 $s_{t\,b}^{(1)}$, $(s_t^{(2)}{}_{,\ b} - 1)$，得到（30）。 □

5.2 改进的风险界

定理 2. 设 f为莫尔斯函数，并假设 $\varepsilon > 0$ 满足 $A_{\varepsilon/2} \subset B(x^\star, r_0)$。固定 $0 < F^- < F^+ \leq 1$和 $\Delta_F \in (0,$ $F^+ - F^-]$。

假设在第 t 代:

(C1) 存在一个 p-最佳索引 b，使得 $x_b \in A_{\varepsilon/4}$。(C2) 假设 4 对此 b 成立，参数为 $\beta_1, \beta_2 > 0$。

(C3) 存在一个记忆槽 k，使得 $\inf_{F\in[F^-,F^+]} g_k^F(F) \ \geq g^-$，且 $\mathbb{P}(CR_b \geq c_{\text{cr}} \mid \mathcal{F}_{t-1}, K_b = k) \ \geq q^-$。

令 $r \in \{0, \ldots, d-1\}$，并令 η_r 如 (14) 中所述。则

$$\mathbb{P}(u_{t,b} \in A_\varepsilon \mid \mathcal{F}_{t-1}) \ \geq \ \tilde{a}_t(b) = \frac{c_{\text{pair}}(t)}{H}(g^-\Delta_F)(q^-\eta_r), \tag{31}$$

其中 $c_{\text{pair}}(t)$ 定义见 (30)。

界限 (31) 描述的是开发利用可靠性，而非探索成功。它保证一旦种群集中在 $A_{\varepsilon/4}$，通过 current-to-p best/1 和自适应记忆参数生成的试验向量以至少 $\tilde{a}_t$ 的概率仍处于 A_ε。$\varepsilon/4$ 与 ε 之间的差距为变异中固有的随机扰动提供了容忍空间。

证明。在 $\mathcal{F}_{t-1}$ 上进行分析，并固定 (C1) 中的情形 $i = b$。以事件 $\{K_b = k\}$ 为条件，其中 k 是 (C3) 中的槽位。这贡献了一个因子 $1/H$。

步骤 1. 按照双池规则，为目标个体 b 抽取 (r_1, r_2)。根据命题 4,

$$\mathbb{P}\big((r_1, r_2) \in \mathcal{R}_{t,\Delta_F}(b) \mid \mathcal{F}_{t-1}\big) = c_{\text{pair}}(t).$$

在事件 $\{(r_1, r_2) \in \mathcal{R}_{t,\Delta_F}(b)\}$ 上，我们根据定义有 $\lambda(\mathcal{I}_t(b, b, r_1, r_2)) \ \geq \Delta_F$。

步骤 2. 根据 (C3)，在 $\{K_b = k\}$ 上，密度在 $[F^-, F^+]$ 上满足 $g_k^F(F) \ \geq g^-$，因此

$$\mathbb{P}\big(F_b \in \mathcal{I}_t(b, b, r_1, r_2) \mid \mathcal{F}_{t-1}, K_b = k, (r_1, r_2)\big) \geq g^-\lambda(\mathcal{I}_t(b, b, r_1, r_2)) \geq g^-\Delta_F.$$

若 $F_b \in \mathcal{I}_t(b, b, r_1, r_2)$，则根据定义有 $v_b(F_b) \ \in A_{\varepsilon/2}$（采用中点修复时，根据注11，同样的包含关系仍成立。）

步骤3。由 (C3) 和引理5,

$$\mathbb{P}\big(C_r(b) \mid \mathcal{F}_{t-1}, K_b = k\big) \geq \mathbb{P}(CR_b \geq c_{\text{cr}} \mid \mathcal{F}_{t-1}, K_b = k)\eta_r \geq q^-\eta_r.$$

根据命题 2，在事件 $\{v_b(F_b) \ \in A_{\varepsilon/2}\} \cap C_r(b)$ 上，我们有 $u_{t,\ b} \in A_\varepsilon$。

利用给定槽位（假设2）时 (F_b, CR_b) 的条件独立性，并结合步骤1–3与 $\mathbb{P}(K_b = k \mid \mathcal{F}_{t-1}) \ = 1/H$，我们得到

$$\mathbb{P}(u_{t,b} \in A_\varepsilon \mid \mathcal{F}_{t-1}) \geq \frac{1}{H}c_{\text{pair}}(t)(g^-\Delta_F)(q^-\eta_r),$$

即为（31）。 □

备注12。对于莫尔斯函数，一旦种群集中在某个局部极小值点附近，局部强凸性便意味着一个足够大的缩放因子区间 F 会产生位于 $A_{\varepsilon/2}$ 中的变异个体（命题3、引理8）。此外，供体集中性意味着有正比例 $c_{\mathrm{pair}}(t)$ 的供体对在几何上是“良好”的（命题4）。

因此，个体风险界不再包含灾难性因子 $1/(s_i^{(1)}(s_i^{(2)}-1))$，而转而依赖于经验集中水平 $c_{\mathrm{pair}}(t)$。

6 见证者状态

前几节的结果表明，当种群在局部极小值点附近变得足够集中时，L-SHADE 的行为会发生定性变化。根据经验观察，并如备注12所形式化的那样，算法的演化可以自然地分为两个阶段。

在初始阶段，种群可能分散得很广。供体差分不受控制，成功 – F值可能微不足道，而个体成功风险可能受到组合因子的抑制。在这种状态下，无法期望得到关于命中时刻的统一尾界。

第二阶段始于算法进入我们所称的见证者状态。一部分不可忽略的个体位于有利吸引域中某个代表点 x_b 的邻域内。这些个体充当见证者，体现该吸引域的局部几何。事实上，强凸性保证了一个成功–F 区间。同时，供体集中性意味着有正比例 $c_{\mathrm{pair}}(t)$ 的供体对在几何上可行。

在本小节中，固定容差 $0<\varepsilon_{\mathrm{in}}<\varepsilon_{\mathrm{out}}$：目标容差和见证者（第二）状态容差。对于每个局部极小值点 x_j，$j=1,\dots,K$，定义局部次水平集

$$A_\varepsilon^{(j)}=\left\{x\in[l,u]^d:\ f(x)\le f(x_j^\star)+\varepsilon\right\}\cap B(x_j^\star,r_0),\qquad A_\varepsilon^{\mathrm{loc}}=\bigcup_{j=1}^{K}A_\varepsilon^{(j)}.$$

局部命中时刻和存活事件为

$$\tau_{\varepsilon_{\mathrm{in}}}^{\mathrm{loc}}=\inf\left\{t\ge 0:\ (\exists i)\ x_i^{(t)}\in A_{\varepsilon_{\mathrm{in}}}^{\mathrm{loc}}\right\},\qquad H_t=\left\{\tau_{\varepsilon_{\mathrm{in}}}^{\mathrm{loc}}>t\right\}.$$

6.1 见证稳定状态 G_t 与稳定化时间 T_{wit}

固定 L-SHADE 参数：p-best 比率 $p\in(0,1]$（因此，在时间 t 有 pN_t 个体组成 p-best 集合）；界限 $0<F^-<F^+\le 1$ 和交叉阈值 $c_{\mathrm{cr}}\in(0,1]$；记忆大小 $H\in\mathbb{N}$。令 $N_t=|P_{(t)}|$ 和 $A_t=|\mathcal{A}^{(t)}|$ 分别表示种群大小和档案大小。我们还固定供体集中半径 $r_{\mathrm{conc}}>0$（例如，$r_{\mathrm{conc}}=r_{\mathrm{safe}}/(2(F^-+\Delta_F))$，如命题3 中所述）以及最小簇大小 $m\ge 4$（因此，除一个 p-best 向量外，至少还有两个不同的供体）。

定义3。在第 t 代，令 $g_{t,k}^F$ 表示记忆槽 $k\in\{1,\dots,H\}$对应的柯西分布 $\mathcal{C}(M_F^{(t)}[k],0.1)$ 的密度，并令

$$q_{t,k}=\mathbb{P}\left(CR\ge c_{\mathrm{cr}}\mid M_{CR}^{(t)}[k]\right).$$

若存在一个槽位 k 使得以下条件成立，则称时间 t 的记忆是良好的：

$$\inf_{F\in[F^-,F^+]}g_{t,k}^F(F)\ \ge\ g^-\quad\text{and}\quad q_{t,k}\ \ge\ q^-,\tag{32}$$

对于某些固定常数 $g^->0$ 和 $q^->0$。

定义 4。令 $x^{(t)}_{\text{最优个体}}\in P_{(t)}$ 为时间 t 的一个最优个体（即 $P^{(t)}$ 上 f 的任意最小化点）。若存在索引 $j\in\{1,\dots,K\}$ 和子集（簇）$\mathcal{C}_t\subseteq P^{(t)}$，使得：则称见证稳定状态 G_t 成立。

(G1) 局部盆地和强凸区域：$\mathcal{C}_t\subseteq B(x_j^\star,r_0)$。

(G2) 簇大小：$|\mathcal{C}_t|\ge m$ 和 $x^{(t)}_{\text{最佳}}\in\mathcal{C}_t$。

(G3) 簇集中度：$\mathrm{diam}(\mathcal{C}_t)\le r_{\mathrm{conc}}$。

(G4) 外子水平集进展： $x^{(t)}$ 最佳$\in A^{(j)}_{\varepsilon_{\text{out}}}$。

(G_5) 非退化参数记忆：根据定义3，时间为 t 时记忆良好 .

定义5. 定义见证稳定化时间

$$T_{\text{wit}} = \inf\Big\{t \geq 0 : \forall s \geq t,\ G_s \text{ holds}\Big\} \in \mathbb{N} \cup \{\infty\}.$$

（如果集合为空，则设为 $T_{\text{wit}} = \infty$。）

6.2 见证概率

现在，我们对在时间 $t-1$ 处于见证稳定状态的条件下，第 t 代发生见证事件的概率给出下界。

定义 6. 固定 $t \geq 1$，并假设 G_{t-1} 成立，其中簇为 $\mathcal{C}_{t-1}$，最优个体为 $b = \text{best}(t-1) \in \mathcal{C}_{t-1}$。任取一个 $\mathcal{F}_{t-1}$-可测索引 $i = i(t-1) \in \mathcal{C}_{t-1} \setminus \{b\}$（例如，$\mathcal{C}_{t-1} \setminus \{b\}$ 中最小的索引）。令 $\mathcal{L}_t$ 为以下事件：在生成 i 的试验个体时所进行的变异/交叉步骤中：

(L1) p-最优选择选中 b（即 $x_{p\text{-best}} = x_b^{(t-1)}$），

(L2) 供体 r_1、r_2 被选为 $\mathcal{C}_{t-1} \setminus \{b, i\}$ 中互不相同的索引（因此 $\|x_{r_1}^{(t-1)} - x_{r_2}^{(t-1)}\| \leq r_{\text{conc}}$），

(L3) 抽样参数满足 $F \in [F^-, F^+]$ 和 $CR \geq c_{\text{cr}}$。

命题 5. 设 $\mathcal{L}_t$ 如定义 6 所述。若 G_{t-1} 成立，则

$$\mathbb{P}(\mathcal{L}_t \mid \mathcal{F}_{t-1}) \ \geq\ \gamma_0(t-1), \qquad \text{on } G_{t-1}, \tag{33}$$

其中

$$\gamma_0(t-1) = \frac{1}{H}\, g^-(F^+ - F^-)\, q^- \frac{1}{\lceil pN_{t-1} \rceil} \frac{m-2}{N_{t-1}-2} \frac{m-3}{N_{t-1}+A_{t-1}-3} \tag{34}$$

特别地，若对所有 t，$N_t \leq N_{\max}$ 和 $A_t \leq A_{\max}$ 均成立，则在 G_{t-1} 上有如下时间一致界

$$\mathbb{P}(\mathcal{L}_t \mid \mathcal{F}_{t-1}) \ \geq\ \gamma_0, \qquad \gamma_0 = \gamma_0^{\min} = \frac{1}{H}\, g^-(F^+ - F^-)\, q^- \frac{1}{\lceil pN_{\max} \rceil} \frac{m-2}{N_{\max}-2} \frac{m-3}{N_{\max}+A_{\max}-3} \ > 0.$$

证明. 在 G_{t-1} 上，根据定义，最优个体 b 位于前 $\lceil pN_{t-1} \rceil$ 名之中，因此 p-最优选择以至少 $1/\lceil pN_{t-1} \rceil$ 的概率选中 b。

由于 $|\mathcal{C}_{t-1}| \geq m$ 和 $i \neq b$，集合 $\mathcal{C}_{t-1} \setminus \{b, i\}$ 的大小至少为 $m-2$。因此，从大小为 $N_{t-1}-2$ 的 $P^{(t-1)} \setminus \{i, b\}$ 中均匀抽样得到 r_1 时，我们有

$$\mathbb{P}(r_1 \in \mathcal{C}_{t-1} \setminus \{b, i\} \mid \mathcal{F}_{t-1}) = \frac{m-2}{N_{t-1}-2}.$$

在条件化于如此选择的 r_1 后，集合 $\mathcal{C}_{t-1} \setminus \{b, i, r_1\}$ 的大小至少为 $m-3$，且 r_2 是从大小为 $N_{t-1}+A_{t-1}-3$ 的集合中均匀抽样得到的，因此

$$\mathbb{P}(r_2 \in \mathcal{C}_{t-1} \setminus \{b, i, r_1\} \mid \mathcal{F}_{t-1}, r_1 \in \mathcal{C}_{t-1} \setminus \{b, i\}) \geq \frac{m-3}{N_{t-1}+A_{t-1}-3}.$$

最后，在 G_{t-1} 处存在一个满足 (32) 的记忆槽。均匀选择槽位至少有 $1/H$ 的概率选中这样的槽位，接下来

$$\mathbb{P}(F \in [F^-, F^+] \mid \text{good slot}) \geq \int_{F^-}^{F^+} g^-\, dF = g^-(F^+ - F^-), \qquad \mathbb{P}(CR \geq c_{\text{cr}} \mid \text{good slot}) \geq q^-.$$

将各个独立的条件下界相乘，即得（34）。 □

6.3 一个局部改进引理

下一个引理表明，朝更优的参考点移动所获得的下降，在扰动足够小时仍能保持。

Lemma 9. 固定一个局部极小值 $x^\star$，并假设 f 在 $B(x^\star, r_0)$ 上是 μ-强凸且 L-光滑的。令 x、z、x_{r_1}、$x_{r_2} \in B(x^\star, r_0)$ 满足 $f(z) \le f(x)$。固定 $F \in (0,1)$，并定义扰动后的变异个体为

$$y = (1-F)x + Fz, \qquad v = y + e, \qquad e = F(x_{r_1} - x_{r_2}).$$

另外假设 $v \in B(x^\star, r_0)$（例如，通过选择“安全半径”来确保这一点）。那么

$$f(v) \;\le\; f(x) - \frac{\mu}{2}F(1-F)\|x-z\|^2 \;+\; Lr_0\|e\| + \frac{L}{2}\|e\|^2. \tag{35}$$

特别地，如果 $\|x_{r_1} - x_{r_2}\| \le r_{\mathrm{conc}}$ 且

$$Lr_0(F^+ r_{\mathrm{conc}}) + \frac{L}{2}(F^+ r_{\mathrm{conc}})^2 \;\le\; \frac{\mu}{4}F^-(1-F^+)\|x-z\|^2, \tag{36}$$

那么 $f(v) \le f(x) - c\|x-z\|^2$ 且 $c = \frac{\mu}{4}F^-(1-F^+) > 0$；因此，在选择过程中，变异个体（或任何目标函数值不大于 $f(v)$ 的试验个体）都会优先于 x 被接受。

证明。将 μ-强凸性应用于凸组合 $y = (1-F)x + Fz$，可得

$$f(y) \le (1-F)f(x) + Ff(z) - \frac{\mu}{2}F(1-F)\|x-z\|^2 \le f(x) - \frac{\mu}{2}F(1-F)\|x-z\|^2,$$

由于 $f(z) \le f(x)$。由 L-光滑性可得，

$$f(v) = f(y+e) \le f(y) + \langle \nabla f(y), e\rangle + \frac{L}{2}\|e\|^2 \le f(y) + \|\nabla f(y)\|\|e\| + \frac{L}{2}\|e\|^2.$$

最后，由于 $y \in B(x^\star, r_0)$ 且 $\nabla f(x^\star) = 0$，L-光滑性意味着 $\|\nabla f(y)\| \le L\|y - x^\star\| \le Lr_0$。合并最后三个式子可得（35）。如果 $\|x_{r_1} - x_{r_2}\| \le r_{\mathrm{conc}}$，那么 $\|e\| \le F^+ r_{\mathrm{conc}}$，并且（36）保证噪声项被控制在下降项的一半以内，从而得到 $f(v) \le f(x) - c\|x-z\|^2$ 且 $c = \frac{\mu}{4}F^-(1-F^+)$。 □

6.4 无间隙的两阶段定理：post-T_{wit} 尾部

现在，我们结合 (i) 在见证事件 $\mathcal{L}_t$ 条件下的危险下界（来自前文的危险率分析），以及 (ii) 在 G_{t-1} 上的一致见证概率，从而得到一个简洁的两阶段结果。

假设 5（$\mathcal{L}_t$ 上的单步成功危险率）。令 E_t 表示“成功在第 t 代发生”这一事件，即第 t 代新生成的试验个体中至少有一个点落在 $A^{\mathrm{loc}}_{\varepsilon_{\mathrm{in}}}$ 内。假设存在一个常数 $a_{\min} > 0$，使得对所有 $t \ge 1$，

$$\mathbb{P}(E_t \mid \mathcal{F}_{t-1}, \mathcal{L}_t) \;\ge\; a_{\min} \quad \text{on } H_{t-1}.$$

（这正是前几节的危险率定理所确立的陈述。）

定理 3（从稳定见证机制出发的两阶段收敛）。假设假设 5 成立，并令 $\gamma_0 > 0$ 为命题 5 中的一致常数。令 T_{wit} 为定义 5 中的稳定化时间，并设 $\tau = \tau^{\mathrm{loc}}_{\varepsilon_{\mathrm{in}}}$。则：

（i）（条件几乎必然成功）

$$\mathbb{P}(\tau < \infty \mid T_{\mathrm{wit}} < \infty) = 1.$$

(ii)（T_{wit} 后几何尾部）在给定 $T_{\mathrm{wit}} = t_0 < \infty$ 的条件下，对所有 $n > t_0$，

$$\mathbb{P}(\tau > n \mid T_{\mathrm{wit}} = t_0) \;\le\; \mathbb{P}(E_0^c)\exp\big(-a_{\min}\gamma_0\,(n - t_0)\big),$$

其中 $E_0 = \{\exists i : \;\; x_i^{(0)} \in A^{\mathrm{loc}}_{\varepsilon_{\mathrm{in}}}\}$。

（iii）（无条件两阶段界）对于所有 $n \ge 0$，

$$\mathbb{P}(\tau > n) \;\le\; \mathbb{P}(T_{\mathrm{wit}} > n) \;+\; \mathbb{P}(E_0^c)\,\mathbb{E}\big[\exp\big(-a_{\min}\gamma_0\,(n - T_{\mathrm{wit}})^+\big)\big].$$

证明。固定 $t \geq 1$。根据假设 5,

$$\mathbb{P}(E_t \mid \mathcal{F}_{t-1}) \geq \mathbb{P}(E_t \cap \mathcal{L}_t \mid \mathcal{F}_{t-1}) = \mathbb{P}(\mathcal{L}_t \mid \mathcal{F}_{t-1})\mathbb{P}(E_t \mid \mathcal{F}_{t-1}, \mathcal{L}_t) \geq a_{\min}\mathbb{P}(\mathcal{L}_t \mid \mathcal{F}_{t-1}),$$

在 H_{t-1} 上。在事件 $\{T_{\text{wit}} \leq t-1\}$ 中，根据稳定化的定义，我们有 G_{t-1}，而命题 5 给出 $\mathbb{P}(\mathcal{L}_t \mid \mathcal{F}_{t-1}) \geq \gamma_0$。因此，在 $H_{t-1} \cap \{T_{\text{wit}} \leq t-1\}$ 上,

$$\mathbb{P}(E_t \mid \mathcal{F}_{t-1}) \;\geq\; a_{\min}\gamma_0.$$

从时间 $t_0 + 1$ 开始，在$\{T_{\text{wit}} = t_0\}$上迭代标准的生存递推 $\mathbb{P}(\tau > n) = \mathbb{P}(E_0^c)\prod_{t=1}^{n}\mathbb{P}(E_t^c \mid H_{t-1})$，得到

$$\mathbb{P}(\tau > n \mid T_{\text{wit}} = t_0) \leq \mathbb{P}(E_0^c) \prod_{t=t_0+1}^{n} (1 - a_{\min}\gamma_0) \leq \mathbb{P}(E_0^c)\exp\big(-a_{\min}\gamma_0\,(n - t_0)\big),$$

从而证明 (ii)。令 $n \to \infty$，即可得到 (i)。对于 (iii)，分解

$$\mathbb{P}(\tau > n) \leq \mathbb{P}(T_{\text{wit}} > n) + \mathbb{E}\big[\mathbb{P}(\tau > n \mid T_{\text{wit}})\mathbf{1}_{\{T_{\text{wit}} \leq n\}}\big],$$

并在期望中应用 (ii)。 □

7 实证验证

我们在 CEC2017 基准测试套件上验证理论界，维度为 $d = 10$，每个函数进行 51 次独立运行，预算为 $10000d$ 次评估。完整的方法细节和扩展结果见补充材料：补充材料A（莫尔斯界验证）和补充材料B（Kaplan–Meier生存分析）。

7.1 莫尔斯界验证

我们对满足 (C1)–(C3) 的条件样本进行汇总，以验证逐代莫尔斯界（定理 2）。表1列出了代表性函数的实证成功率 $\hat{}p$ 与理论界均值 $\bar{a}^{\text{sum}}$。

表1：CEC2017 上的莫尔斯界验证（$d = 10$，乘积和模式）。所有函数均满足 $\hat{}p \geq \bar{a}^{\text{sum}}$（界成立）。

函数	ε	条件样本	$\hat{p}$	$\bar{a}^{\text{sum}}$	比率
F1（平移球函数）	10	1071	0.632	0.072	8.8
F4（平移Rosenbrock函数）	30	1071	0.944	0.069	13.7
F11（平移混合函数）	10	395	0.691	0.001	599
F22（组合函数）	30	42	0.643	0.053	12.1

该界在所有测试函数上均得到验证，比率范围为 8.8（F1，最紧）至 599（F11，最保守）。保守性主要源于对集中因子 $c_{\text{pair}}(t)$ 的最坏情况估计；分解分析见补充材料 A。

7.2 Kaplan–Meier生存分析

我们对整个基准测试套件中的首次命中时间 τ_{A_ε} 应用 Kaplan–Meier估计，以刻画其分布。表 2 汇总了代表性函数在 $\varepsilon = 10$下的生存分析结果。

分析揭示了三种截然不同的经验机制：(i) 聚集成功（F1、F11），命中时间集中在狭窄的突发区间内（$\mathcal{C} < 0.2$）；(ii) 近几何尾部（F5），常数风险模型近似成立（$\mathcal{C} \approx 0.4$）；以及 (iii) 不可处理情形（F22），预算时限处的生存概率为 > 0.9。

7.3 失效模式识别

该诊断框架区分了两种性质截然不同的失效模式（表3）：

利用失败（F11）：停滞期间，CR记忆不断累积较低的值，导致见证频率下降（$\hat{}\gamma \approx 0.16$），这表明算法已找到正确的吸引域，但无法有效地开发利用它。

表2：CEC2017上的Kaplan–Meier生存分析（$d = 10, \varepsilon = 10$）。$\hat{S}(n)$：预算下的生存概率； $\bar{\tau}$：平均命中时间（在成功者中）； $\mathcal{C}$：聚类系数（< 1 表示时间聚类）。

函数	**Hits**	$\hat{S}(n_{\max})$	$\bar{\tau}$	σ_τ	$\mathcal{C}$	状态
F1	51/51	0.00	186	16	0.15	聚类
F5	51/51	0.00	241	89	0.42	近几何
F11	51/51	0.00	312	178	0.08	强聚类
F22	2/51	0.96	892	241	0.004	难以处理

表3： $\varepsilon = 1$ ($d = 10$)处的失效模式比较。$\hat{\gamma}$：见证频率；L3：CR-尾部条件满足率。

函数	成功	$\hat{\gamma}$	L3满足率	失效模式
F1	100%	0.994	1.00	—
F11	73%	0.161	0.20	开发利用（L3 崩塌）
F22	4%	0.977	1.00	探索（$T_{\text{wit}} = \infty$）

探索失败（F22）：见证条件已满足（$\hat{\gamma} \approx 1$），但算法未能在预算内找到全局吸引域。理论界限虽然在空泛意义上成立，却没有参考价值，因为对于大多数运行而言， $T_{\text{wit}} = \infty$ 。
这些结果证实，风险率框架为 L-SHADE 收敛提供了有意义但较为保守的界，而条件（L2）–（L3）可作为评估算法性能的实用诊断标准。

8 结论

我们提出了一个基于风险率的框架，用于分析差分进化中的首次命中时间，尤其关注 L-SHADE 等自适应种群算法。通过将生存概率表示为条件首次命中风险率的乘积，我们得到了无需独立性或平稳性假设即可成立的分布无关恒等式和尾界。

对于采用 current-to-$p$best/1 变异的 L-SHADE，我们引入了一个明确的算法见证事件；在该事件下，条件风险率具有一个确定性下界，该下界仅取决于采样规则和种群结构。由此可将生存风险率自然分解为理论常数和经验事件频率，从而阐明：当成功以短簇的形式发生，而非以均匀速率发生时，常数风险率界为何必然较为保守。
对 CEC2017基准测试套件进行的 Kaplan–Meier分析从经验上证实了这一图景。在不同函数和评估预算下，我们观察到三种截然不同的状态：(i) 强聚类成功，即命中时间集中在短促的时间段内；(ii) 近似几何尾部，即常数风险模型能够准确描述；(iii) 在评估时限内未观察到任何命中的情况。在大多数成功案例中，最紧的有效几何包络率比经验命中后风险率小一至两个数量级，这表明最坏情况尾界会大幅低估典型表现。

结果表明，L-SHADE 的实际行为受罕见的依赖配置的转变所支配，而非由稳定的逐代进展决定。从理论角度看，风险率视角提供了一套统一的表述方式，将严格的尾界与经验生存分析联系起来。从实践角度看，它提供了诊断工具，可用于区分聚集、几何和重尾状态。

未来仍有若干研究方向有待探索。首先，通过刻画见证事件在典型轨迹中出现的频率，可以得到更精确的界。其次，将分析扩展至其他变异策略和自适应差分进化变体，或可揭示哪些设计选择会促成聚集行为，而哪些会促成几何行为。最后，将风险率框架与明确的全局探索机制相结合，或可在保持实际效率的同时，获得可证明的几乎必然成功保证。

9 致谢

感谢索非亚科技园的Discoverer超级计算机提供高性能计算资源。我们还使用了项目 BG-RRP-2.004-0002 提供的高性能计算资源。

本研究由欧盟“下一代欧盟”通过保加利亚共和国的国家复苏与韧性计划资助，项目编号为 BG-RRP-2.004-0002，“BiOrgaMCT”。

一个具体的乘积空间构造

本附录给出一个概率空间的完全严格构造，该空间支持诸如 L-SHADE 这样的自适应种群算法，其中第 t 代随机选择的分布可能取决于过去的状态。

Let

$$\Omega = [0,1]^{\mathbb{N}}, \qquad \mathcal{F} = \mathcal{B}([0,1])^{\otimes \mathbb{N}},$$

其中，$\mathcal{B}([0,1])$ 是博雷尔 σ-代数，$\mathcal{B}([0,1])^{\otimes \mathbb{N}}$ 是由柱集生成的乘积 σ-代数。令 $\mathbb{P}$ 为（可数）乘积测度 $\lambda^{\otimes \mathbb{N}}$，其中 λ 是 $[0,1]$ 上的勒贝格测度。根据柯尔莫哥洛夫扩张定理／标准博雷尔空间上的可数乘积测度的存在性，$\mathbb{P}$ 的存在性和唯一性得以保证 [4, 24]。

将坐标映射记为

$$U_n(\omega) = \omega_n, \qquad n \geq 1.$$

那么，在 $\mathbb{P}$ 下，$(U_n)_{n\geq 1}$ 独立同分布，服从 Unif(0, 1)。

自适应差分进化方法在每一代 $t \geq 0$ 中都会使用有限个随机选择：索引选择、连续随机抽样（用于参数和交叉）、强制维度索引等。令 R_t 表示第 t 代所消耗的汇总随机输入。我们不预设 (R_t) 服从固定的乘积分布，而是将每个 R_t 表示为以下内容的可测函数：(i) 过去的算法状态；(ii) 一段新的有限均匀随机数块。

形式化地，将均匀随机数流划分为有限块：选取整数 $1 = n_0 < n_1 < n_2 < s$，并定义第 t 块为

$$U^{(t)} = (U_{n_t}, U_{n_t+1}, \ldots, U_{n_{t+1}-1}) \in [0,1]^{m_t}, \quad m_t = n_{t+1} - n_t < \infty.$$

（只要 $m_t < \infty$，任意选择均可；例如，可以令 m_t 足够大，以涵盖第 t 代中的所有随机选择。）

令 S_t 表示第 t 代结束后的完整算法状态（种群、档案、记忆等）。我们假设该算法具有如下形式的可测更新

$$S_t = \Phi_t(S_{t-1}, U^{(t)}), \qquad t \geq 1, \tag{37}$$

具有可测的初始化 $S_0 = \Phi_0(U^{(0)})$。标准差分进化组件具有可测性，原因如下：(i) 离散索引采样可以通过对均匀随机数进行阈值化来实现；(ii) 从一维分布连续采样可以通过逆累积分布函数或接受-拒绝映射来实现；(iii) 变异／交叉／边界处理／选择都是欧几里得空间上的确定性博雷尔映射。更一般地，标准博雷尔空间上的任意可测随机核都可以表示为均匀随机变量的可测函数（有时称为随机化／转移引理）；参见 [24, 15]。

定义代后滤过如下：

$$\mathcal{F}_t = \sigma(S_0, \ldots, S_t), \qquad t \geq 0.$$

根据递推式 (37)，(S_t) 适应于 $(\mathcal{F}_t)$。特别地，对于任意可测目标集 $A \subset D$，事件

$$\{(\exists i)\, x_i^{(t)} \in A\} \quad \text{and} \quad E_t = \left\{(\forall s \leq t-1)(\forall i)\, x_i^{(s)} \notin A \text{ and } (\exists i)\, x_i^{(t)} \in A\right\}$$

是 $\mathcal{F}_t$-可测的，因为它们仅取决于 $(S_0, \ldots, S_t)$。因此，命中时刻 $\tau_A = \inf\{t \geq 0 \colon E_t \text{发生}\}$ 是关于 $(\mathcal{F}_t)$ 的停时。

如果希望直接通过给定过去条件下 R_t 的条件分布来指定算法，可以在路径空间 $\Omega' = \Xi_0 \times \Xi_1 \times \Xi_2 \times s$ 上配备 σ-代数 $\mathcal{F}' = \mathcal{E}_0 \otimes \mathcal{E}_1 \otimes s$，并通过指定 Ξ_0 上的初始分布，以及一系列给定过去条件下 R_t 的随机核，来定义一个概率测度。根据 Ionescu–Tulcea扩张定理 [41, 24]，诱导测度在 Ω' 上的存在性和唯一性随即成立。从思想上说，这种表述等价于上述通用均匀随机变量构造，但更直接地体现了自适应条件采样。

参考文献

[1] N. H. Awad、M. Z. Ali、J. J. Liang、B. Y. Qu 和 P. N. Suganthan。《CEC 2017 单目标实参数数值优化特别会议与竞赛的问题定义和评估标准》。技术报告，南洋理工大学，2016 年。[2] C. Bayer。《高级概率论》。讲义，2011 年。参见定理 5.14（Ionescu–Tulcea）和定理 5.16（柯尔莫哥洛夫扩张定理）。[3] P. Billingsley。《概率与测度》。Wiley，第 2 版，1986 年。[4] P. Billingsley。《概率与测度》。Wiley 概率与统计丛书。Wiley，第 3 版，1995 年。[5] J. Brest、S. Greiner、B. Boškovi´c、M. Mernik 和 V. Žumer。《差分进化中的自适应控制参数：数值基准问题的比较研究》。IEEE《进化计算汇刊》，10(6):646–657，2006 年。doi:10.1109/TEVC.2006.872133。[6] D. R. Cox 和 D. Oakes。《生存数据分析》。Chapman and Hall，1984 年。[7] S. Das 和 P. N. Suganthan。《差分进化：最新进展综述》。IEEE《进化计算汇刊》，15(1):4–31，2011 年。doi:10.1109/TEVC.2010.2059031。[8] S. Das、S. S. Mullick 和 P. N. Suganthan。《差分进化的最新进展——更新版综述》。《群体与进化计算》，27:1–30，2016 年。doi:10.1016/j.swevo.2016.01.004。[9]B. Doerr。《随机化优化启发式算法分析中的概率工具》。载于《进化计算理论：离散优化的最新进展》，第 1–87 页。Springer，2020 年。doi:10.1007/978-3-030-29414-4 1。[10] B. Doerr、D. Johannsen 和 C. Winzen。《乘法漂移分析》。《算法》，64(4):673–697，2012 年。doi:10.1007/s00453-012-9eli-4。[11] S. Droste、T. Jansen 和 I. Wegener。《关于 (1+1) 进化算法的分析》。《理论计算机科学》，276(1-2):51–81，2002 年。doi:10.1016/S0304-3975(01)00182-7。[12] D. Drusvyatskiy 和 A. S. Lewis。《次微分的倾斜稳定性、一致二次增长与强度量正则性》。SIAM《优化杂志》，23(1):256–267，2013 年。doi:10.1137/120876551。[13] D. Drusvyatskiy 和 A. S. Lewis。《近端方法的误差界、二次增长与线性收敛》。《运筹学数学》，43(3):919–948，2018 年。doi:10.1287/moor.2017.0889。[14] R. M. Dudley。《实分析与概率》。Wadsworth & Brooks/Cole，1989 年。[15] R. M. Dudley。《实分析与概率》。剑桥高等数学研究丛书。剑桥大学出版社，第 2 版，2002 年。doi:10.1017/CBO9780511755347。[16]R. Durrett。《概率：理论与例子》。剑桥大学出版社，第 5 版，2019 年。doi:10.1017/9781108591034。[17] P. Fusek、D. Klatte 和 W. Li。《线性半无限优化中 argmin 映射的 Lipschitz 模数》。SIAM《优化杂志》，23(1):17–37，2013 年。doi:10.1137/110852917。[18] J. He 和 X. Yao。《进化算法的漂移分析与平均时间复杂度》。《人工智能》，127(1):57–85，2001 年。doi:10.1016/S0004-3702(01)00058-3。[19] J. He 和 X. Yao。《迈向分析进化算法计算时间的解析框架》。《人工智能》，145(1-2):59–97，2003 年。doi:10.1016/S0004-3702(02)00381-8。[20] J. He 和 X. Yao。《用于估算进化算法计算时间的漂移分析研究》。《自然计算》，3(1):21–35，2004 年。doi:10.1023/B:NACO.0000023416.59689.4e。[21] Z. Hu、S. Xiong、Q. Su 和 X. Zhang。《差分进化算法全局收敛的充分条件》。《应用数学杂志》，2013:193196，2013 年。doi:10.1155/2013/193196。[22]T. Jansen。《进化算法分析：计算机科学视角》。Springer，2013 年。doi:10.1007/978-3-642-17339-4。[23] D. Johannsen。《随机组合结构与随机化搜索启发式算法》。博士论文，萨尔兰大学，2010 年。[24] O. Kallenberg。《现代概率论基础》，第 99 卷，《概率论与随机建模》。Springer，Cham，第 3 版，2021 年。doi:10.1007/978-3-030-61871-1。[25] A. N. Kolmogorov。《概率论基础概念》。Springer-Verlag，1933 年。英文译本：《概率论基础》，Chelsea，1956 年。

[26] D. A. Levin 和 Y. Peres。《马尔可夫链与混合时间》。美国数学会，第 2 版，2017 年。[27] S. Meyn 和 R. L. Tweedie。《马尔可夫链与随机稳定性》。剑桥大学出版社，第 2 版，2009 年。doi:10.1017/CBO9780511626630。[28] J. W. Milnor。《莫尔斯理论》。数学年刊研究丛书第 51 卷。普林斯顿大学出版社，普林斯顿，新泽西州，1963 年。根据 M. Spivak 和 R. Wells 的讲义编写。[29] B. Mitavskiy、J. Rowe 和 C. Cannings。在保持链路总数不变的条件下，优化网络通信的局部搜索策略理论分析。《智能计算与控制论国际期刊》，2(2):243–284，2009 年。doi:10.1108/17563780910959875。[30] M. Molga 和 C. Smutnicki。用于优化的测试函数。2005 年。可在线获取。[31] F. Neumann 和 C. Witt。《组合优化中的仿生计算：算法及其计算复杂性》。Springer，2010 年。doi:10.1007/978-3-642-16544-3。[32] J. Nocedal 和 S. J. Wright。《数值优化》。施普林格运筹学与金融工程丛书。Springer，纽约，纽约州，第 2 版，2006 年。doi:10.1007/978-0-387-40065-5。[33] K. Opara 和 J. Arabas。差分进化：理论分析综述。《群体与进化计算》，44:546–558，2019 年。doi:10.1016/j.swevo.2018.06.010。[34]A. K. Qin 和 P. N. Suganthan。用于数值优化的自适应差分进化算法。载于 IEEE进化计算大会（CEC）论文集，第 1785–1791 页，2005 年。doi:10.1109/CEC.2005.1554904。[35] R. T. Rockafellar 和 R. J.-B. Wets。《变分分析》，数学科学基础丛书第 317 卷。Springer，1998 年。doi:10.1007/978-3-642-02431-3。[36] G. Rudolph。《进化算法的收敛性质》。Kovač，1997 年。[37] C. R. Shalizi。由正则条件概率分布构建无限过程。讲义，36-754（卡内基梅隆大学），第 3 讲，2006 年。参见第 3.1 节（定义 30：概率核）和第 3.2 节（定理 33：Ionescu–Tulcea扩张定理）。[38] R. Storn 和 K. Price。差分进化——一种用于连续空间全局优化的简单高效启发式算法。《全局优化杂志》，11(4):341–359，1997 年。doi:10.1023/A:1008202821328。[39]R. Tanabe 和 A. Fukunaga。差分进化的基于成功历史的参数自适应。载于 IEEE进化计算大会（CEC）论文集，第 71–78 页，2013 年。doi:10.1109/CEC.2013.6557555。[40] R. Tanabe 和 A. S. Fukunaga。利用线性种群规模缩减改进 SHADE 的搜索性能。载于 IEEE进化计算大会（CEC）论文集，第 1658–1665 页，2014 年。doi:10.1109/CEC.2014.6900380。[41] C. T. Ionescu Tulcea。《乘积空间中的测度》。《林琴国家科学院物理、数学与自然科学类报告》（8），7:208–211，1949 年。[42] J. Zhang 和 A. C. Sanderson。JADE：带可选外部档案库的自适应差分进化。IEEE《进化计算汇刊》，13(5):945–958，2009 年。doi:10.1109/TEVC.2009.2014613。[43] M. Žáková。关于差分进化的收敛性。载于 MENDEL 论文集，第 318–323 页，1999 年。