

遗传算法求解多模态优化问题的动力性

李航^{1,2} 李敏强¹ 寇纪淞¹

摘要 多模态函数一般存在多个局部极值解, 局部极值解处适应值的大小很大程度上影响了它们被遗传算法搜索到的概率. 为了弄清楚这种影响机制, 通过分析基因池遗传算法的无限种群动力系统, 刻画了双峰函数局部极值解的适值差与系统不动点之间的解析关系, 进一步分析推广了理论结果的适用范围. 最后, 提出针对多模态优化问题的两阶段遗传算法, 给出了应用理论结果改善遗传搜索性能的范例, 实验结果表明该算法对多模态函数的搜索性能有明显改善, 从侧面证明了理论结果在实际应用中的正确性.

关键词 多模态优化问题, 双峰函数, Walsh 变换, 基因池遗传算法
中图分类号 TP301.6

Dynamical Behavior of Genetic Algorithms on Multi-modal Optimization

LI Hang^{1,2} LI Min-Qiang¹ Kou Ji-Song¹

Abstract There is more than one local optimum in multi-modal landscapes. The values of local optima can influence their opportunities to be exploited by GAs. By analyzing the infinite population dynamical system using the gene pool GA and the BINEEDLE fitness function, we characterize the analytic relation between the fall of local optima and the fixed points in the infinite population dynamical system of the gene pool GA. Further analysis shows that the relation is still held in more common cases. Enlightened by the result, we derive a method to improve GAs for the multi-modal landscapes and get satisfying effect. The experiment results also prove the correctness of the theory in this paper.

Key words Multi-modal optimization, the BINEEDLE fitness function, Walsh transform, the gene pool GA

很多工程问题都可以转化为多模态优化问题^[1], 这类问题往往存在多个局部极值解, 有时不仅需要找到全局最优解, 还需要找到尽可能多的局部极值解. 对那些局部极值解的适应值相差很小并且分布较均匀的多模态函数, 用遗传算法反复搜索往往能找到多个局部极值解. 观察发现, 局部极值解的适应值及其吸引域相对越大, 则算法收敛至该局部极值解的可能性越大. 各局部极值解的适值差是可以调节的, 研究它对遗传算法收敛结果的影响有助于具体地改善遗传算法对多模态优化问题的搜索性能.

Vose 和 Wright 等人引入 Walsh 变换^[2-3], 简化了遗传算法的种群动力方程, 并给出了基因池遗传算法的种群动力方程及其在分析一类特殊函数中的应用^[4-6]. 基因池遗传算法因采用基因池重组而得名^[4]. 本文在前人工作基础上, 分析了遗传算法求解多模态优化问题的动力性, 并给出了该理论的

应用范例. 本文内容具体安排如下: 第 1 节证明了 Walsh 变换前后种群动力系统的拓扑共轭性, 这是本文研究工作的基础, 得出了基因池遗传算法处理双峰函数^[5] 的动力方程, 刻画了双峰函数局部极值解的适值差与无限种群动力系统不动点间的解析关系, 分析了各不动点的稳定性, 分析了局部极值解的适值差对遗传算法性能的影响, 得出了本文的主要理论结果 (结论 1 和结论 2); 第 2 节利用 Geiringer 和 Vose 的工作^[2,7] 推广了结论 1 和结论 2 的适用范围; 第 3 节给出了应用本文理论结果改善遗传算法处理多模态优化问题性能的具体方法和实例, 从侧面说明了所得理论结果在实际应用中的有效性.

1 基因池遗传算法搜索性能与局部极值解适值差间的关系

本节通过分析变换后的无限种群动力系统 \hat{G} , 找出了局部极值解的适值差与基因池遗传算法无限种群动力系统不动点之间的解析关系, 并且分析了不动点的稳定性. 通过进一步分析, 刻画了局部极值解的适值差对遗传算法最终收敛结果的影响.

1.1 基因池遗传算法的无限种群动力系统 G 与其经 Walsh 变换后的系统 \hat{G} 的拓扑等价性

$\Omega = \{0, 1\}^l$ 表示二进制串组成的解空间, $\mathcal{L} = \{j \in \Omega : \#j = 1\}$, $\#$ 表示二进制串 j 中含 1 的个数. $x_v = P\{x = v : v \in \Omega\}$ 表示个体 v 在整个种群中

收稿日期 2006-10-24 收修改稿日期 2007-04-24
Received October 24, 2006; in revised form April 24, 2007
国家自然科学基金 (70171002, 70571057), 新世纪优秀人才支持计划 (NECT-05-0253) 资助
Supported by National Natural Science Foundation of China (70171002, 70571057), the Program for New Century Excellent Talents in University of China (NECT-05-0253)
1. 天津大学管理学院 天津 300072 2. 中国民航大学管理学院 天津 300300
1. School of Management, Tianjin University, Tianjin 300072
2. School of Management, Civil Aviation University of China, Tianjin 300300
DOI: 10.3724/SP.J.1004.2008.00180

占的比例. $\mathbf{x} = (x_1, \dots, x_{2^l})$ 表示 Ω 上随机变量的分布向量. $\Lambda = \{\mathbf{x} : \mathbf{x} = (x_1, \dots, x_{2^l}), \sum_{j \in \Omega} x_j = 1\}$ 表示所有种群分布向量形成的无限种群空间. $W = (W_{i,j})_{2^l \times 2^l}$ 表示 Walsh 矩阵, 其中 $W_{i,j} = (-1)^{\#(i \otimes j)}$. 对向量作 Walsh 变换就是将 Walsh 矩阵 W 和该向量作乘积, 如 $\hat{\mathbf{x}} = W\mathbf{x}$, $\hat{\mathbf{f}} = W\mathbf{f}$ 和 $\hat{A} = WAW$, 这里 $\mathbf{f} = (f_1, \dots, f_{2^l})$, A 表示矩阵. $\hat{\Lambda} = \{\hat{\mathbf{x}} : \hat{\mathbf{x}} = (\hat{x}_1, \dots, \hat{x}_{2^l})\}$.

定义 1^[2]. 种群动力系统模型中, U 表示变异函数, F 表示选择函数, M 表示重组函数, 记 $G(\mathbf{x}) = U \circ F \circ M(\mathbf{x})$ ($\mathbf{x} \in \Lambda$), G 就是表示整个遗传操作的函数. G 是 Λ 到 Λ 的映射, 且符合离散动力系统的定义^[8], 本文称 G 为无限种群动力系统.

引理 1. 映射 $W : \Lambda \rightarrow \hat{\Lambda}$ 是 Λ 到 $\hat{\Lambda}$ 的同胚.

证明. 任取 $\mathbf{x} \in \Lambda$, 存在映射 $W : \Lambda \rightarrow \hat{\Lambda}$, $\hat{\mathbf{x}} = W(\mathbf{x}) = W \cdot \mathbf{x}$ 且 $\hat{\mathbf{x}} \in \hat{\Lambda}$. 根据 $\hat{\Lambda}$ 的定义, $\hat{\Lambda}$ 是 Λ 在映射 W 下的值域, 因而映射 W 是满射; 又由 Walsh 矩阵 W 存在逆矩阵可知存在 W 的逆映射; 因而 W 是一一映射. 由此可知: W 是 Λ 到 $\hat{\Lambda}$ 的同胚. \square

引理 2^[3]. \mathbf{x} 是一个种群向量, 那么对 $\forall k \in \mathcal{L}$ 有

$$\widehat{G(\mathbf{x})}_k = (1 - 2\mu) \times$$

$$\frac{2^l \hat{x}_k + \sum_{i \in \Omega} (-1)^{\#(i \otimes k)} (f_i - 1) \prod_{j \in \mathcal{L}} (1 + (-1)^{\#(i \otimes j)} \hat{x}_j)}{2^l + \sum_{j \in \Omega} (f_j - 1) \prod_{j \in \mathcal{L}} (1 + (-1)^{\#(i \otimes j)} \hat{x}_j)}$$

成立.

这是整个遗传映射在 Walsh 变换下的表示. 公式中和本文中特殊符号的说明参见文献 [9].

引理 3. 令 $\hat{G}(\hat{\mathbf{x}}) = \widehat{G(\mathbf{x})}$, 则 \hat{G} 是 $\hat{\Lambda}$ 上的离散动力系统.

证明. 因为 G 是 Λ 上的离散动力系统, 由 $\mathbf{x} \in \Lambda$ 可以推出 $G(\mathbf{x}) \in \Lambda$, 因而 $\widehat{G(\mathbf{x})} \in \hat{\Lambda}$; 因为 $\hat{\mathbf{x}} \in \hat{\Lambda}$ 且 $\hat{G}(\hat{\mathbf{x}}) = \widehat{G(\mathbf{x})} \in \hat{\Lambda}$, 所以 \hat{G} 是 $\hat{\Lambda}$ 到 $\hat{\Lambda}$ 的映射. 又因为 $\hat{G}^0(\mathbf{x}) = \mathbf{x}$ 且 $\hat{G}^m(\hat{G}^n(\mathbf{x})) = \hat{G}^{m+n}(\mathbf{x})$ 对任意 $m, n \in \mathbf{Z}$ 成立 (见 \hat{G} 的定义), 由动力系统的定义^[8] 可知: \hat{G} 是离散动力系统. \square

定理 1. 动力系统 G 和 \hat{G} 是拓扑同胚的, 也是拓扑共轭的.

证明. 由定义 1 和引理 3 知, G 和 \hat{G} 分别是 Λ 和 $\hat{\Lambda}$ 上的离散动力系统. 由引理 1, 存在同胚映射 $W : \Lambda \rightarrow \hat{\Lambda}$; 又由引理 3, $\hat{G}(\hat{\mathbf{x}}) = \widehat{G(\mathbf{x})}$ 即 $\hat{G}(W(\mathbf{x})) = W(G(\mathbf{x}))$ 成立; 此外映射 W 保持时间定向性和不变性. 由文献 [8] 中的定义可以判断 G 和 \hat{G} 是拓扑同胚的, 也是拓扑共轭的. \square

由于 G 和 \hat{G} 是拓扑共轭的, 可知 G 和 \hat{G} 有拓扑等价的结构和轨迹, 且在相同的时刻达到不动点,

不动点的个数和稳定性也相同, 因此可通过研究 \hat{G} 来研究 G .

1.2 决定 \hat{G} 的不动点情况的动力方程

定理 2^[4]. 如果种群分布向量 \mathbf{x} 是联结平衡的, 那么对任意 $k \in \Omega$, $\hat{x}_k = \prod_{j \in \mathcal{L}_k} \hat{x}_j$.

定义 2^[5]. 双峰函数 (the BINEEDLE fitness function) 以一定长度的二进制串为自变量. 当 $\mathbf{x} = (0, 0, \dots, 0)$ 时, $f(\mathbf{x}) = 1 + a$; 当 $\mathbf{x} = (1, 1, \dots, 1)$ 时, $f(\mathbf{x}) = 1 + b$; 当 \mathbf{x} 等于其他值时, $f(\mathbf{x}) = 1$. 其中 $a, b > 0$, $d = b - a$, d 就是局部极值解 $(0, 0, \dots, 0)$ 和 $(1, 1, \dots, 1)$ 的适值差; 设 $r = \min\{a, b\}$ (本文假设 $b > a$, 所以 $r = a$), r 也被称为低峰值.

定理 3. 用无限种群基因池遗传算法求解双峰函数时, 若 $\hat{\mathbf{x}}$ 是 \hat{G} 的不动点, 当且仅当对 $\forall k \in \mathcal{L}$, 有

$$\hat{x}_k = \frac{(1 - 2\mu)[a \prod_{j \in \mathcal{L}} (1 + \hat{x}_j) - b \prod_{j \in \mathcal{L}} (1 - \hat{x}_j)]}{2^{l+1}\mu + [a \prod_{j \in \mathcal{L}} (1 + \hat{x}_j) + b \prod_{j \in \mathcal{L}} (1 - \hat{x}_j)]} \quad (1)$$

证明. $\hat{\mathbf{x}}$ 是 \hat{G} 的不动点, 当且仅当 $\hat{G}(\hat{\mathbf{x}}) = \hat{\mathbf{x}}$, 即 $\hat{G}(\hat{\mathbf{x}})_k = \hat{x}_k$ 对任意 $k \in \Omega$ 成立. 基因池遗传算法总是保持种群分布向量 \mathbf{x} 是联结平衡的^[4]. 根据定理 2, $\hat{G}(\hat{\mathbf{x}})_i$ ($i \in \Omega \setminus \mathcal{L}$) 的值完全可由 $\hat{G}(\hat{\mathbf{x}})_k$ ($k \in \mathcal{L}$) 决定, 因此 $\hat{\mathbf{x}}$ 是 \hat{G} 的不动点, 只需 $\hat{G}(\hat{\mathbf{x}})_k = \hat{x}_k$ 对 $\forall k \in \mathcal{L}$ 成立. 将双峰函数代入到引理 2, 化简后就得到如下对 $\forall k \in \mathcal{L}$ 成立的公式

$$\hat{G}(\hat{\mathbf{x}})_k = (1 - 2\mu) \times$$

$$\frac{2^l \hat{x}_k + a \prod_{j \in \mathcal{L}} (1 + \hat{x}_j) - b \prod_{j \in \mathcal{L}} (1 - \hat{x}_j)}{2^l + [a \prod_{j \in \mathcal{L}} (1 + \hat{x}_j) + b \prod_{j \in \mathcal{L}} (1 - \hat{x}_j)]} \quad (2)$$

$\hat{\mathbf{x}}$ 是 $\hat{G}(\hat{\mathbf{x}})$ 的不动点等价于 $\hat{G}(\hat{\mathbf{x}}) = \hat{\mathbf{x}}$, 由式 (2), 这等价于 $\hat{G}(\hat{\mathbf{x}})_k = \hat{x}_k$ 对任意 $k \in \mathcal{L}$ 成立, 将以上结果带入式 (2), 进一步整理即得到式 (1). \square

推论 1. 若 $\hat{\mathbf{x}}$ 是动力系统 \hat{G} 中的不动点, 则其分量 \hat{x}_k ($k \in \mathcal{L}$) 是等值的.

证明. 由定理 3 可知, 若 $\hat{\mathbf{x}}$ 是 \hat{G} 的不动点, 当且仅当式 (1) 对任意 $k \in \mathcal{L}$ 成立. 观察式 (1) 发现等式右边与 k 无关, 因此 $\hat{\mathbf{x}}$ 的各分量 \hat{x}_k ($k \in \mathcal{L}$) 是等值的. \square

定理 4. 令 $w = \hat{x}_k$ ($k \in \mathcal{L}$), 则双峰函数局部极值解的适值差与动力系统 \hat{G} 的不动点 $\hat{\mathbf{x}}$ 的关系可

用下式刻画

$$d = \frac{r[(1 - 2\mu - w)(1 + w)^l - (1 - 2\mu + w)(1 - w)^l] - 2^{l+1}\mu w}{(1 - 2\mu + w)(1 - w)^l} \quad (3)$$

证明. 由推论 1 可知 $\hat{\mathbf{x}}$ 的各分量 $\hat{x}_k (k \in \mathcal{L})$ 具有等值性, 可以统一用变量 w 代替. 由定理 2, $\hat{\mathbf{x}}$ 的其余分量的值都可以表示成 $\hat{x}_k (k \in \mathcal{L})$ 的函数, 进而表示成 w 的函数. 因此, w 决定着 $\hat{\mathbf{x}}$ 所有分量的值, 所以 w 和 $\hat{\mathbf{x}}$ 一一对应. 综上, 可以用 w 与 l 的关系式刻画 $\hat{\mathbf{x}}$ 与 l 的关系. 用 w 代替 $\hat{x}_k (k \in \mathcal{L})$ 代入到式 (1) 中, 由于 $r = \min\{a, b\} = a$ 且 $d = b - a$, 整理式 (1) 可得

$$2^{l+1}\mu w + d(1 - 2\mu + w)(1 - w)^l + r[(1 - 2\mu + w)(1 - w)^l - (1 - 2\mu - w)(1 + w)^l] = 0 \quad (4)$$

整理方程为以 d 为因变量并以 w 为自变量的函数, 即可得式 (3). \square

定义 3. 由于 w 与 $\hat{\mathbf{x}}$ 存在一一对应关系, 我们也把 w 称为 \hat{G} 的不动点, 并将方程 (4) 称为 \hat{G} 的不动点的等价方程.

由于动力系统 \hat{G} 与 G 是拓扑共轭的, 其不动点 $\hat{\mathbf{x}}$ 与 \mathbf{x} 也是一一对应的, 所以刻画 d 与 \hat{G} 的不动点间关系的式 (3), 也同等地刻画了 d 与 G 的不动点之间的关系. 由于动力系统 G 描述的是基因池遗传算法的动力学特性, 因此从式 (3) 可以看出双峰函数局部极值解的适值差的变化对基因池遗传算法稳定性产生的影响.

引理 4^[4]. 如果 $\mathbf{x} \in \Lambda$, 则有 $|\hat{x}_k| < 1$ 对所有 $k \in \Omega$ 成立.

由引理 4, 并且考虑到本文假设, 可推知如下定理.

定理 5. 不动点 w 满足: $w \in [-1, 1]$; 有意义的局部极值解的适值差 d 满足: $d > 0$.

1.3 动力系统 \hat{G} 的不动点的稳定性分析

如果函数 $g : \mathbf{R}^N \mapsto \mathbf{R}^N$ 是可微的, 那么它在不动点 \mathbf{x} 处的偏导数 $dg_{\mathbf{x}}$ 是一个 $N \times N$ 阶的雅克比矩阵 J , 其元素定义为: $J_{i,j} = \partial g_i(\mathbf{x}) / \partial x_j$. 众所周知, 当 J 的特征根模的最大值小于 1 时, \mathbf{x} 就是 g 的一个渐近稳定的不动点^[8]. 下面通过讨论映射 \hat{G} 在 $\hat{\mathbf{x}}$ 处的雅克比矩阵 J 的特征根模的最大值, 来说明不动点的稳定性. 由于 $\hat{\mathbf{x}}$ 的值完全取决于分量 $\hat{x}_k (k \in \mathcal{L})$, 所以只对这些分量求导, 此时雅克比矩阵 J 是 $|\mathcal{L}| \times |\mathcal{L}|$ 维的.

引理 5. 不动点 $\hat{\mathbf{x}}$ 处, 对任意 $j \neq k$ 且 $j, k \in \mathcal{L}$, 则 \hat{G} 在 $\hat{\mathbf{x}}$ 处的雅克比矩阵 J 的元素满足下式

$$\frac{\partial \hat{G}(\hat{\mathbf{x}})_k}{\partial \hat{x}_j} = (1 - 2\mu) \times \frac{2^l [a(1-w)(1+w)^{l-1} + b(1+w)(1-w)^{l-1}] + 4ab(1-w)^{l-1}(1+w)^{l-1}}{[2^l + a(1+w)^l + b(1-w)^l]^2}$$

$$\frac{\partial \hat{G}(\hat{\mathbf{x}})_k}{\partial \hat{x}_k} = (1 - 2\mu) \times \frac{2^l [2a(1+w)^{l-1} + 2b(1-w)^{l-1} + 2^l] + 4ab(1-w)^{l-1}(1+w)^{l-1}}{[2^l + a(1+w)^l + b(1-w)^l]^2}$$

当 $0 \leq \mu \leq 1/2$ 时, 上述偏导数非负, 即 J 的元素非负.

证明. 令 $P = \prod_{i \in \mathcal{L}} (1 + \hat{x}_i)$, $P_j = \prod_{i \in \mathcal{L} \setminus \{j\}} (1 + \hat{x}_i)$, $Q = \prod_{i \in \mathcal{L}} (1 - \hat{x}_i)$, $Q_j = \prod_{i \in \mathcal{L} \setminus \{j\}} (1 - \hat{x}_i)$. 对式 (2) 求导并化简, 并将其中相应部分用上述符号代替, 可得下式

$$\frac{\partial \hat{G}(\hat{\mathbf{x}})_k}{\partial \hat{x}_j} = (1 - 2\mu) \times \frac{2ab(P_j Q + P Q_j) + 2^l a(1 - \hat{x}_k) P_j + 2^l b(1 + \hat{x}_k) Q_j}{(2^l + aP + bQ)^2}$$

$$\frac{\partial \hat{G}(\hat{\mathbf{x}})_k}{\partial \hat{x}_k} = (1 - 2\mu) \times \frac{2ab(P_k Q + P Q_k) + 2^l [a(1 - \hat{x}_k) P_k + b(1 + \hat{x}_k) Q_k] + 2^l + aP + bQ}{(2^l + aP + bQ)^2}$$

由推论 1, 将上式中的 \hat{x}_k 全部替换为 w , 经化简后即可得本引理中的公式. 由于 $-1 \leq w = \hat{x}_k \leq 1$, 易证: 当 $0 \leq \mu \leq 1/2$ 时, 引理公式中的偏导数非负. \square

引理 6^[4]. 设 $l \times l$ 阶方阵 A 的对角线元素等于 s , 其余元素均等于 e , 则方阵 A 有 1 个特征值等于 $s + (l - 1)e$, 其余 $l - 1$ 个特征值均等于 $s - e$.

定义 4. 定义 $\bar{G}(w) = \frac{2^l w + a(1+w)^l - b(1-w)^l}{2^l + a(1+w)^l + b(1-w)^l} \times (1 - 2\mu)$. 设 $p = (1 + w)^l$, $p' = l(1 + w)^{l-1}$, $q = (1 - w)^l$, $q' = l(1 - w)^{l-1}$, 则可推出其导数

$$\frac{d\bar{G}}{dw} = (1 - 2\mu) \times \frac{2^l [2^l + ap + bq + ap'(1 - w) + bq'(1 + w)] + 2ab(p'q + pq')}{(2^l + ap + bq)^2} \quad (5)$$

由引理 5、引理 6 和定义 4, 并注意到 $\mu \in [0, 1/2]$ 时 $\partial \hat{G}(\hat{\mathbf{x}})_k / \partial \hat{x}_k$ 和 $\partial \hat{G}(\hat{\mathbf{x}})_k / \partial \hat{x}_j$ 均大于 0, 经推导后可得如下定理.

定理 6. $d\bar{G}/dw = \partial \hat{G}(\hat{\mathbf{x}})_k / \partial \hat{x}_k + (l - 1) \times \partial \hat{G}(\hat{\mathbf{x}})_k / \partial \hat{x}_j$; 并且当 $\mu \in [0, 1/2]$ 时, J 的特征根模的最大值就是 $d\bar{G}/dw$.

由定理 6 和雅克比矩阵特征根模的最大值对不动点稳定性的影响^[8] 可得如下定理.

定理 7. 当 $0 \leq \mu \leq 1/2$ 时, 动力系统 \hat{G} 不动点的稳定性由 $d\bar{G}/dw$ 决定: 当 $d\bar{G}/dw < 1$ 时, 不

动点是稳定的; 当 $d\bar{G}/dw > 1$ 时, 不动点是不稳定的; 当 $d\bar{G}/dw = 1$ 时, 情况不定.

设定参数 μ 、 l 和 r 的值, 并将式 (3) 代入式 (5) 即得各不动点的稳定性判别式.

1.4 局部极值解的适值差 d 对遗传算法最终收敛结果影响的图示分析

假设 $\mu = 0.02$, $l = 10$ 且 $r = 100$, 根据由不动点等价方程导出的式 (3), 可得双峰函数局部极值解的适值差 d 与不动点 w 的关系图. 再将式 (3) 带入到式 (5) 右边, 消去参数 d , 可得不动点处的雅克比矩阵特征根模的最大值 $d\bar{G}/dw$ 与不动点 w 的关系图, 根据图中 $d\bar{G}/dw$ 的值可判断不动点 w 的稳定性. 任意给定 d 值, 可从 d, w 关系图中找到与其对应的不动点 $w(d)$, 再根据 $w(d)$ 的值从 $d\bar{G}/dw, w$ 关系图中就可以找到相应的 $d\bar{G}/dw(d)$, 由此判断 d 对应的不动点 $w(d)$ 的稳定性.

由图 1 可以看出, 当 $w \in (-0.9780, -0.7326)$ 或 $w \in (0.9551, 0.9635)$ 时, 可得 $|d\bar{G}/dw| < 1$, 即不动点处雅克比矩阵特征根模的最大值小于 1, 于是上述区间中的不动点是渐近稳定的. 当 $w \in [-1, -0.9780)$, 或 $w \in (-0.7326, 0.9551)$, 或 $w \in (0.9635, 1]$ 时, 可得 $|d\bar{G}/dw| > 1$, 即不动点处雅克比矩阵特征根模的最大值大于 1, 于是这些区间中的不动点均是不稳定的.

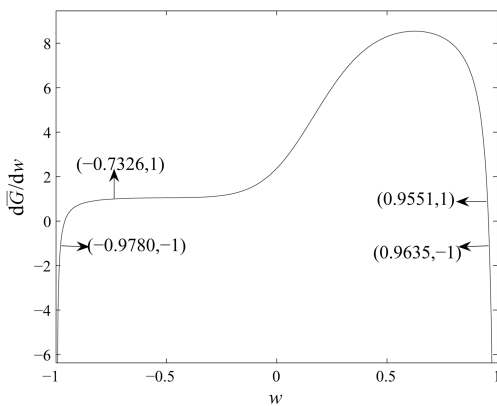


图 1 $\mu = 0.02, l = 10, r = 100$ 时 w 和 $d\bar{G}/dw$ 的关系图
Fig. 1 Relation between w and $d\bar{G}/dw$ when $\mu = 0.02, l = 10$, and $r = 100$

由图 2~4 (见下页), 可得如下结论. 首先, 由式 (3) 可知, 当分母中 $1 - 2\mu + w' = 0$ 时: $\lim_{w \rightarrow w'^-} d(w) = -\infty$ 且 $\lim_{w \rightarrow w'^+} d(w) = +\infty$; 由于 $\mu = 0.02$, 这里 $w' = -0.96$. 由图 2 知: 当 $w \in [-1, -0.96)$ 时, $d \leq -101$ 超出了 d 的取值范围, 该区域内的不动点没有意义. 当 $w \in (-0.96, -0.7326)$

时, d 由 $+\infty$ 变到 -99.5 , 随 w 的增加单调递减; 对照图 1 知, 此区间内的不动点均是稳定的. 由图 3, 4 知: 当 $w \in (-0.7326, 0.9551)$ 时, d 由 -99.5 变到 5.656×10^{15} , 随 w 的增加而递增; 由图 1 知, 此区间内的不动点均是不稳定的. 由图 4 知: 当 $w \in (0.9551, 0.9595)$ 时, d 由 5.656×10^{15} 减小到 -100 , 随 w 单调递减; 对照图 1 知, 此区间内的不动点均是稳定的. 当 $w \in (0.9595, 1]$ 时, $d \leq -100$ 超出了 d 的取值范围, 该区域内的不动点没有意义.

由观察结果可知: 当 $d \in (0, 5.656 \times 10^{15})$ 时, 有两个稳定的不动点 (分别靠近 -1 和 1) 和一个不稳定的不动点; 当 d 超过临界值 5.656×10^{15} 时, 只剩下一个稳定的不动点 (靠近 -1). 实验证明, 当参数 $0 < \mu < 0.5$ 、 r 为小于 10^{10} 的整数并且 l 为小于 500 的整数时, 上述结果都是成立的, 而这几乎涵盖了这些参数通常的取值范围.

下面分析各不动点 w 所对应的实际解的优劣.

$w = \hat{x}_k (k \in \mathcal{L})$ 是不动点的决定性等价分量 $\hat{x}_k (k \in \mathcal{L})$ 的值. 由 Walsh 变换可推知下式: $\hat{x}_k = x_{(0, \dots, 0)} + \sum_{i \in \Omega \setminus \{(0, \dots, 0), (1, \dots, 1)\}} (-1)^{i \otimes k} x_i - x_{(1, \dots, 1)}$. 从中可以看出: 对任意 k 值, \hat{x}_k 随 $x_{(0, \dots, 0)}$ 的增大而增大, 随 $x_{(1, \dots, 1)}$ 的增大而减小. 当 $i \in \Omega \setminus \{(0, \dots, 0), (1, \dots, 1)\}$ 时, 随着 $k \otimes j$ 值奇偶性的不同, 总是有一部分 x_i 导致 \hat{x}_k 的增大, 而另一部分 x_i 导致 \hat{x}_k 的减小. 由于 $k \in \mathcal{L}$ 且 \mathcal{L} 均匀覆盖所有基因位, 所以对每个固定的 x_i , 其增大总会导致一部分 \hat{x}_k 增加而另一部分 \hat{x}_k 减小. 也就是说, 指望 $i \in \Omega \setminus \{(0, \dots, 0), (1, \dots, 1)\}$ 时的 x_i 来影响 \hat{x}_k 的大小, 必然导致各 $\hat{x}_k (k \in \mathcal{L})$ 值大小的失衡. 因而根据各 \hat{x}_k 的等值性可知, 影响 \hat{x}_k 大小的主导分量只能是 $x_{(0, \dots, 0)}$ 和 $x_{(1, \dots, 1)}$. 换句话说, $w = \hat{x}_k (k \in \mathcal{L})$ 较大 (接近 1) 则反映出 $x_{(0, \dots, 0)}$ 较大, $w = \hat{x}_k (k \in \mathcal{L})$ 较小 (接近 -1) 则反映出 $x_{(1, \dots, 1)}$ 较大. 由于 $f_{(0, \dots, 0)} = 1 + a, f_{(1, \dots, 1)} = 1 + b$ 并且 $b > a$, 所以 $x_{(1, \dots, 1)}$ 较大说明最优个体 $(1, \dots, 1)$ 在最终种群中占绝大多数, $x_{(0, \dots, 0)}$ 较大表示次优个体 $(0, \dots, 0)$ 在最终种群中占绝大多数 (参见 x_v 的定义).

综上所述, 不动点 w 接近 -1 对应着实际所得的最终种群中最优个体占绝大多数, 不动点 w 接近 1 对应着实际所得的最终种群中次优个体占绝大多数, 这两类不动点都是稳定的, 所以最优和次优种群占绝大多数的情况会经常发生. 不动点 w 靠近 -1 和 1 的中间, 对应着实际所得最终种群中以其他类个体为主, 这种不动点是不稳定的, 所以这种情况不易发生. 以上讨论中我们假设 $b > a$, 但由于问题具有对称性, 所以当 $b < a$ 时以上结论依然成立.

综上所述, 可得如下结论:

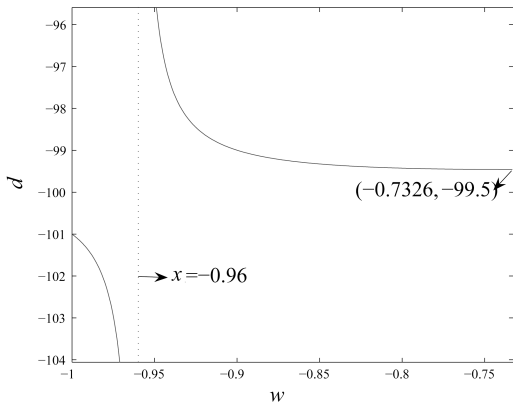


图 2 $w \in [-1, -0.76]$ 时 w 和 d 的关系图

Fig. 2 Relations between w and d when $w \in [-1, -0.76]$

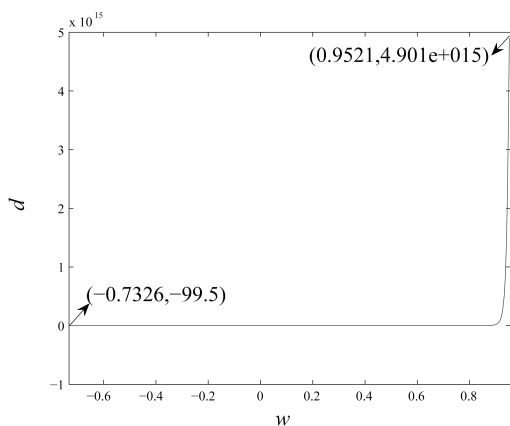


图 3 $w \in [-0.7, 0.96]$ 时 w 和 d 的关系图

Fig. 3 Relations between w and d when $w \in [-0.7, 0.96]$

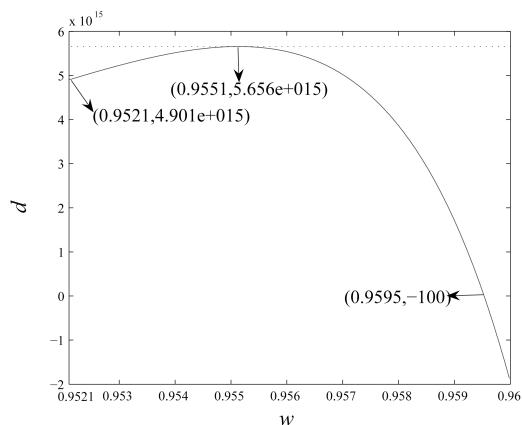


图 4 $w \in [0.95, 0.98]$ 时 w 和 d 的关系图

Fig. 4 Relations between w and d when $w \in [0.95, 0.98]$

结论 1. 用基因池遗传算法处理双峰函数, 当局部极值解的适值差小于或等于某个临界值时, 进行多次遗传搜索, 最终收敛种群会出现两种情况: 1) 以最优个体为主, 即收敛到最优个体; 2) 以次优个体

为主, 即收敛到次优个体. 当适值差超过临界值时, 进行多次遗传搜索后, 最终收敛种群只会收敛到最优个体. 最终种群收敛到最优个体的概率不是一下子就变为 1 的, 在适值差不断增大直至超过临界值的过程中, 最终种群收敛到最优个体的概率会随着适值差向临界值的逼近而逐渐增大, 直至为 1.

定义 5. 结论 1 中所提到的临界值称为绝对优势差, 记作 D . 例如, 当 $\mu = 0.02$, $l = 10$ 且 $r = 100$ 时, $D = 5.656 \times 10^{15}$.

我们希望分析出参数 r 和 l 对 D 的影响, 因而固定变异率 $\mu = 0.02$, 用相同的图示分析方法得出如下参数设置下的 D 值 (见表 1).

表 1 不同参数设置下的绝对优势差

Table 1 The critical gaps on different parameters

组别: 参数	参数 r 的值	参数 l 的值	对应 D 的值
1	100	10	5.656×10^{15}
2	100	8	3.069×10^{12}
3	100	6	1.790×10^9
4	100	2	1113
5	50	10	2.536×10^{15}
6	10	10	2.181×10^{14}
7	1	10	8.885×10^9

通过对表 1 中数据的分析并且参照大量其他参数设置下的实验结果, 可得如下结论:

结论 2. r 是双峰函数中较小的那个局部极值解与非局部极值解的适值差; l 则表征编码空间的维度, 并由参数搜索空间的规模所决定. r 或 l 的减小会引起绝对优势差 D 的减小, 相比而言 D 值对 l 的变化更为敏感. 其直观意义就是: 降低较小的那个局部极值解的适应值或缩小搜索空间都会引起绝对优势差 D 的减小. 在适值差 d 不变的情况下, D 的减小使 d 与 D 之间的差距相对缩小; 根据结论 1, d 向 D 的逼近会使遗传算法最终收敛到全局最优解的概率增加. 因此可以说降低较小的那个局部极值解的适应值或缩小参数搜索空间会使遗传算法更容易收敛到全局最优解, 而且相比而言, 后者的效果更为明显.

简而言之, 结论 1 表明: 在一定条件下, 增大局部极值解的适值差能够提高全局最优解的被搜索概率, 而减小局部极值解的适值差则会降低全局最优解的被搜索概率, 从而增加非全局最优的局部极值解的被搜索概率. 结论 2 表明: 通过调节适值差以外的其他参数也能间接达到同样的效果. 这里需要说明的是, 因为适值差的变化并未影响个体间适应值大小的序关系, 所以结论 1 和结论 2 对于采用排序选择的遗传算法并不适用.

2 理论分析结果的推广

结论 1 和结论 2 是基于基因池遗传算法得到的. 根据 Geiringer 的理论^[7], 基因池重组可以看作是两父代重组的多次重复, 即执行单次基因池遗传操作等价于执行多次简单遗传算法 (Simple genetic algorithm, SGA) 的遗传操作, 因此结论 1 和结论 2 对 SGA 也成立.

结论 1 和结论 2 是在无限种群模型下得到的. 根据 Vose 的理论^[2], 在有限种群下执行遗传算法等同于在无限种群模型下的遗传算法搜索过程中进行取样. 因此反映无限种群特性的结论 1 和结论 2 在执行有限种群遗传算法的过程中也会得以体现.

结论 1 和结论 2 是针对双峰函数得到的. 我们知道在遗传搜索过程中, 当搜索种群包含某个局部极值解吸引域内的个体, 并且此类个体在种群中达到一定规模时, 搜索过程就不可逆转地收敛到这个局部极值解, 因此可以将吸引域内的个体与局部极值解共同视为一个整体点. 这样一个只有两个局部极值解的多模态函数就可以看作是一个类双峰函数. 而多个局部极值解的多模态函数, 各局部极值解间的相互影响可以看作是两两局部极值解相互影响的叠加. 因此, 可以将结论 1 和结论 2 应用到针对一般多模态问题的遗传算法的设计中去.

3 对传统遗传算法的改进

由于结论 1 和结论 2 也适用于 SGA 和一般多模态优化问题, 所以可以用于改进针对多模态优化问题的遗传算法. 根据结论 1 和结论 2 可知: 在应用基因池遗传算法处理双峰函数的优化过程中, 保持低峰值不变的同时增大全局极值解的适值差, 就能增大全局最优解的被搜索概率; 而保持低峰值不变的同时减小局部极值解的适值差, 就能减小全局最优解的被搜索概率, 相应地增加次优个体的被搜索概率. 受此启发, 对于一般多模态函数, 保持函数值在某个阈值以下的各点的函数值不变, 通过有拉伸效果的函数标度变换增大函数值在这个阈值以上的各点的适值差, 就可以增加那些函数值在阈值以上的局部极值解 (包含全局最优解) 的被搜索概率. 相反, 保持阈值以下的各点的函数值不变, 通过有压缩效果的函数标度变换减小阈值以上的各点的适值差, 就可以减小那些阈值以上的局部极值解 (包含全局最优解) 的被搜索概率, 相应地增加各阈值以下的局部极值解的被搜索概率. 阈值小于全局最优解的函数值, 并且阈值越大, 其上的局部极值解越少, 但始终包含全局最优解. 因此可以通过调高阈值, 增强对全局最优解的搜索能力. 实际中, 阈值的大小可由单次遗传搜索后的当前最优值确定, 也可由决策者

根据最低满意值经验地确定.

下面, 介绍两种函数标度变换 (参见式 (6) 和式 (7)) 来实现对阈值以上的函数值的拉伸与压缩. 这类变换都不改变原函数值的大小序关系, 因而能够保证变换前后的局部极值解不变. 假设变换前的原函数 $f(\mathbf{x})$ 都是正函数. 如果 $f(\mathbf{x})$ 在定义域内的取值不全大于 0, 先作如下处理: 令 $f'(\mathbf{x}) = f(\mathbf{x}) + N, N \gg \max\{0 - f(\mathbf{x})\}$, 然后对正函数 $f'(\mathbf{x})$ 进行变换. 这样的处理同样保持了函数值的序关系, 保证了变换前后的局部极值解不变.

$$F_1(f(\mathbf{x})) = \begin{cases} f(\mathbf{x}), & f(\mathbf{x}) \leq f(\mathbf{x}') \\ f(\mathbf{x}) - 1 + T|f(\mathbf{x}) - f(\mathbf{x}')|, & f(\mathbf{x}) > f(\mathbf{x}') \end{cases} \quad (6)$$

$$F_2(f(\mathbf{x})) = \begin{cases} f(\mathbf{x}), & f(\mathbf{x}) \leq f(\mathbf{x}') \\ f(\mathbf{x}') + \log(1 + |f(\mathbf{x}) - f(\mathbf{x}')|), & f(\mathbf{x}) > f(\mathbf{x}') \end{cases} \quad (7)$$

变换 F_1 能在保持阈值以下函数值对应各点的函数值不变的同时, 增大阈值以上各点的适值差. 参数 $T > 0$, 且 T 越大, 适值差增大的幅度越大, 一般取 $T = 10$. 变换 F_2 能在保持阈值以下各点的函数值不变的同时, 减小阈值以上各点的适值差. \log 函数的底数越大, 适应值之差减小的幅度越大, \log 函数的底数一般取 10.

如图 5 所示, $f(x) = e^{-2 \ln 2 \times ((x-0.1)/0.8)^2} \times \sin^6(5\pi x)$ 是一个定义在 $[0, 1]$ 区间的多极值点等距非等高的函数. $x' = 0.9$ 是函数值最小的局部极值解, 我们将 $f(x')$ 设为阈值. $F_1(f(x))$ 能明显增大各局部极值解的适值差, $F_2(f(x))$ 能明显减小各局部极值解的适值差.

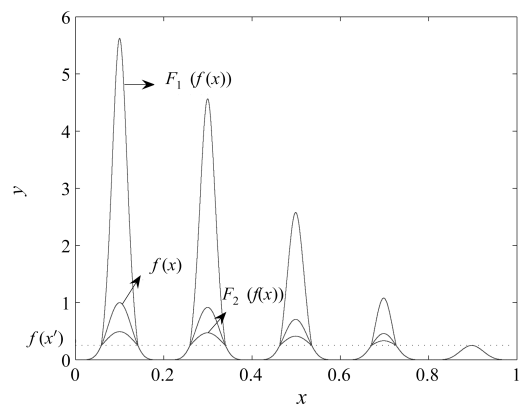


图 5 F_1, F_2 变换前后的函数 $f(x)$ 的形状变化图
Fig. 5 Different shapes of $f(x)$ under transforms F_1 and F_2

压缩阈值以上各点的适值差, 会减小阈值以上的局部极值解的被搜索概率, 相应增加阈值以下的局部极值解的被搜索概率. 若进行多次遗传搜索, 其效果参见如下定理.

定理 8. 压缩阈值以上的函数值对应各点的适值差, SGA 能在有限的搜索次数内搜索到更多的局部极值解.

证明. 假设多模态优化问题共有 M 个局部极值解, 共进行 K 次遗传搜索, 单次搜索收敛到第 i 个局部极值解的概率为 $P(i)$ ($1 \geq P(i) \geq 0$), 这里 $P(i)$ 是平均搜索概率. K 次遗传搜索后, 局部极值解 i 至少被搜索到一次的概率为: $P_K(i) = 1 - [1 - P(i)]^K$. 易证: 只要 $P(i) > 0$, 则有 $\lim_{K \rightarrow \infty} P_K(i) = 1$, 而且 $P_K(i)$ 随 K 的增加成指数型增长的, 增速很快.

由于 $\sum_{i=1}^M P(i) = 1$, 所以压缩阈值以上各点的适值差后, 在减小阈值以上的局部极值解的被搜索概率的同时会相应增加阈值以下的局部极值解的被搜索概率. 某些局部极值解的被搜索概率原来为 0, 压缩后会变得大于 0. 这样在多次遗传搜索后, 它们就会以很高的概率被发现. 而原来被搜索概率大于 0 的局部极值解, 由于各局部极值解的大小序关系未变, 所以压缩后其被搜索概率仍大于上述那些小局部极值解, 仍然大于 0.

综上, 压缩后被搜索概率大于 0 的局部极值解的个数只增不减, 因而使得有限搜索次数内搜索到的局部极值解的个数增多. \square

综上所述, 可以先对多模态函数进行 F_1 变换, 以达到更准确地收敛至全局最优解的目的; 然后再对原函数进行 F_2 变换, 以达到收敛到更多局部极值解的目的. 通过调节阈值, 可以控制所得局部极值解的质量. 这就是本文将要介绍的基于局部极值解适值差调节的两阶段遗传算法 (简称 FAGA) 的设计思想, 其主要步骤如下:

步骤 1. 根据对实际问题的最低满意度或是进行单次遗传搜索, 将当前最优解设为阈值.

步骤 2. 将适应函数 $f(\mathbf{x})$ 用 F_1 进行变换, 然后对 $F_1(f(\mathbf{x}))$ 应用 SGA 求解全局最优解.

步骤 3. 将适应函数 $f(\mathbf{x})$ 用 F_2 进行变换, 然后对 $F_2(f(\mathbf{x}))$ 反复应用 SGA 求解局部极值解, 直到达到预先设置的次数为止.

下面对 FAGA 算法的优缺点进行理论分析.

FAGA 对遗传算法的改进建立在对其动力特性细致分析的基础之上. 相比以往使用的标度变换函数 $F(f(\mathbf{x})) = e^{f(\mathbf{x})}$, F_1 和 F_2 更科学更可靠. 就双峰函数而言, 标度函数 F 不仅增大了局部极值解的适值差, 也增大了低峰值. 根据结论 2, 低峰值的增大会引起绝对优势差的成倍增大; 又根据结论 1, 只有适值差与绝对优势差间距减小才会改善算法对全局最优解的搜索效果. 因此加速函数未必能使算法

更好地收敛到全局最优解. F_1 则强调保持阈值以下函数值不变, 相当于在保持绝对优势差不变的同时增大适值差. 其次, FAGA 具有很好的可移植性. 由于算法主要是对所优化的函数进行相应变换, 所以可以将函数变换与任意采用适应值比例选择的改进型遗传算法进行结合, 以提升原算法的搜索性能.

由结论 1 和结论 2 后的注解, 本方法对采用排序选择的遗传算法无效. 另外, 由于改进算法中并未考虑局部极值解的吸引域大小对算法性能的影响, 因而当各局部极值解的吸引域十分不均匀时, 其影响超过适值差的影响, FAGA 中函数变换的调节作用就比较弱.

下面, 用 F_1 和 F_2 分别对 f_1 和 f_2 进行函数变换, 比较变换前后 SGA 的搜索精度和广度, 以此来验证 FAGA 的效果. f_1 和 f_2 都是经典的多模态测试函数. 我们设置交叉概率为 0.99, 变异概率为 0.02, 种群规模为 1024, 搜索步长为 10^{-8} , 搜索代数 500 代.

以下两个测试函数引自文献 [1], 其中 $x \in [0, 1]$.

$$\max f_1(x) = e^{-\ln 4 \times [\frac{10x-1}{8}]^2} \times \sin^6(5\pi x) \quad (8)$$

$$\max f_2(x) = e^{-\ln 4 \times [\frac{10x-1}{8}]^2} \times \sin^6[5\pi(x^{\frac{3}{4}} - \frac{1}{20})] \quad (9)$$

通过比较表 2 和表 3 (见下页上方) 中的实验结果发现, 采用两阶段遗传算法 (FAGA), 第一阶段对 f 进行 F_1 变换, 能有效提高算法对全局最优解的搜索精度; 第二阶段对 f 进行 F_2 变换, 能使算法在有限的搜索次数内 (本实验设为 100 次) 搜索到更多的局部极值解, 而搜索精度略有下降. 两阶段相结合, 既能保证全局最优解的精度, 又能搜索到更多理想的局部极值解.

经典的 PSO 本质上也只能搜索到单一的全局最优解^[10-11], 而基于小生境的适应值共享技术则强烈地依赖于生境半径的选取且需要大规模的种群^[12]. FAGA 平均分配各局部极值解被搜索的概率, 不依赖于参数设置, 另外单次搜索的时间也比较少.

4 结论

本文分析证明了在一定情况下, 增大局部极值解的适值差能够提高全局最优解的被搜索概率, 而减小局部极值解的适值差则会提高那些非全局最优极值解的被搜索概率, 进一步的工作推广了该理论的适用范围. 接下来利用该结论对遗传算法进行了改进, 提出了两阶段遗传算法. 最后, 用经典算例对算法进行了检测, 实验结果表明新算法在搜索精度和广度上较传统遗传算法有明显提高, 从侧面证明了本文理论结果在实际应用中的有效性.

表 2 SGA 和 FAGA 对函数 f_1 优化效果的对比
Table 2 Comparison between SGA and FAGA for f_1

f_1 的极值解 算法	发现概率	$f_1(0.1) = 1$		$f_1(0.3) = 0.9170$			$f_1(0.5) = 0.7071$	
		最佳值	平均值	发现概率	最佳值	平均值	发现概率	最佳值
SGA	100%	1	0.9980	0	-	-	0	-
FAGA(I)	100%	1	0.9996	0	-	-	0	-
FAGA(II)	90%	1	0.9979	10%	0.9142	0.9142	0	-

表 3 SGA 和 FAGA 对函数 f_2 优化效果的对比
Table 3 Comparison between SGA and FAGA for f_2

f_2 的极值解 算法	发现概率	$f_2(0.0797) = 0.9991$		$f_2(0.2467) = 0.9546$			$f_2(0.4506) = 0.7602$	
		最佳值	平均值	发现概率	最佳值	平均值	发现概率	最佳值
SGA	80%	0.9991	0.9981	20%	0.9546	0.9490	0	-
FAGA(I)	100%	0.9991	0.9991	0	-	-	0	-
FAGA(II)	65%	0.9991	0.9980	30%	0.9544	0.9472	5%	0.7632

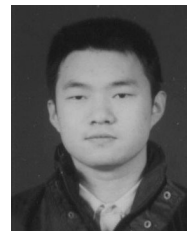
References

- Li Min-Qiang, Kou Ji-Song, Lin Dan, Li Shu-Quan. *The Basic Theories and Applications of Genetic Algorithms*. Beijing: Science Press, 2002. 216–401
(李敏强, 寇纪淞, 林丹, 李书全. 遗传算法的基本理论与应用. 北京: 科学出版社, 2002. 216–401)
- Vose M D. *The Simple Genetic Algorithm*. Cambridge, Massachusetts: The MIT Press, 1999. 21–35
- Vose M D, Wright A H. The simple genetic algorithm and the Walsh transform: Part I: theory. *Evolutionary Computation*, 1998, **6**(3): 253–273
- Wright A H, Rowe J E, Stephens C R, Poli R. *Bistability in a Gene Pool GA with Mutation*. Morgan Kaufmann, 2003. 63–80
- Wright A H, Rowe J E, Neil J R. Analysis of the simple genetic algorithm on the single-peak and double-peak landscapes. In: *Proceedings of the 2002 Congress on Evolutionary Computation*. Hawaii, USA: IEEE, 2002. 214–219
- Wright A H, Cripe G. *Bistability of the Needle Function in the Presence of Truncation Selection*. Berlin: Springer-Verlag, 2004. 330–342
- Geiringer H. On the probability theory of linkage in Mendelian heredity. *Annals of Mathematical Statistics*, 1944, **15**(1): 25–57
- Yao Miao-Xin, Chen Fang-Qi. *Mathematical Foundation of Non-linear Dynamical System*. Tianjin: Tianjin University Press, 2005. 71–92
(姚妙新, 陈芳启. 非线性理论数学基础. 天津: 天津大学出版社, 2005. 71–92)
- Zhang Wen-Xiu, Liang Yi. *Mathematical Foundation of Genetic Algorithms*. Xi'an: Xi'an Jiaotong University Press, 2000, 33–80
(张文修, 梁怡. 遗传算法的数学基础. 西安: 西安交通大学出版社, 2000, 33–80)
- Iwamatsu M. Multi-specie particle swarm optimizer for multimodal function optimization. *IEICE Transactions on Information and Systems*, 2006, **89**(3): 1181–1187
- Wang Li-Fang, Zeng Jian-Chao. A cooperative evolutionary algorithm based on particle swarm optimization and simulated annealing algorithm. *Acta Automatica Sinica*, 2006,

32(4): 630–635

(王丽芳, 曾建潮. 基于微粒群算法与模拟退火算法的协同进化方法. 自动化学报, 2006, **32**(4): 630–635)

- Li Min-Qiang, Kou Ji-Song. Dynamical behavior of genetic algorithms with finite population. *Acta Automatica Sinica*, 2002, **28**(4): 497–504
(李敏强, 寇纪淞. 多模态函数优化的协同多群体遗传算法. 自动化学报, 2002, **28**(4): 497–504)



李 航 天津大学管理学院博士研究生. 主要研究方向为人工智能与进化计算. 本文通信作者.

E-mail: lh197861@163.com

(**LI Hang** Ph. D. candidate at School of Management, Tianjin University. His research interest covers evolutionary computation and artificial intelligence. Corresponding author of this paper.)



李敏强 天津大学管理学院教授. 主要研究方向为人工智能与进化计算.

E-mail: mqli@tju.edu.cn

(**LI Min-Qiang** Professor at School of Management, Tianjin University. His research interest covers evolutionary computation and artificial intelligence.)



寇纪淞 天津大学管理学院教授. 主要研究方向为人工智能与进化计算.

E-mail: jskou@tju.edu.cn

(**KOU Ji-Song** Professor at School of Management, Tianjin University. His research interest covers evolutionary computation and artificial intelligence.)