

演化多目标优化中的几何热力学选择

应伟勤^{1),2)} 李元香¹⁾ SHEU Phillip C-Y^{1),2)} 吴 昱¹⁾ 余法红¹⁾

¹⁾(武汉大学软件工程国家重点实验室 武汉 430072)

²⁾(加州大学尔湾分校电气工程和计算机科学系 加利福尼亚 92697 美国)

摘 要 热力学遗传算法(Thermodynamical Genetic Algorithms, TDGAs)借鉴热力学中的自由能极小过程来统一处理多目标优化在逼近性和多样性两方面的任务. 为提高 TDGA 的运行效率和解集分布均匀性,提出了一种几何热力学选择. 在该选择中首先定义角度熵通过扇形采样来度量种群逼近方向的多样性. 然后利用距离精英定义距离能量来度量种群的逼近程度,避免了耗时的非劣分层操作. 此外,引入分量热力学替换规则以较低计算代价驱动种群的几何自由能快速下降. 在多目标 0/1 背包问题上的实验结果表明,几何热力学选择极大地提高了 TDGA 的运行效率和解集分布均匀性;采用该选择的 TDGA 算法可生成与 NSGA-II 在逼近性和分布多样性上性能相当的解,但在运行效率上明显优于 NSGA-II.

关键词 多目标优化;演化算法;热力学替换;角度熵;距离能量

中图法分类号 TP301 DOI号: 10.3724/SP.J.1016.2010.00755

Geometric Thermodynamical Selection for Evolutionary Multi-Objective Optimization

YING Wei-Qin^{1),2)} LI Yuan-Xiang¹⁾ SHEU Phillip C-Y^{1),2)} WU Yu¹⁾ YU Fa-Hong¹⁾

¹⁾(State Key Laboratory of Software Engineering, Wuhan University, Wuhan 430072)

²⁾(Department of Electrical Engineering and Computer Science, University of California, Irvine, CA 92697, USA)

Abstract Thermodynamical genetic algorithms (TDGAs) simulate the minimization of free energy in thermodynamics to deal simultaneously with both convergence and diversity in multi-objective optimization. A geometric thermodynamical selection (GTS) is proposed to improve the running efficiency and the distribution uniformity of solutions of TDGA. In GTS, an angle entropy is introduced to measure the diversity of convergent directions by sector sampling and then a distance energy is presented to measure the extent of convergence by distance elitist rather than the expensive non-dominated sorting. In addition, a component thermodynamical replacement rule is used to force the geometric free energy of population to steeply descend with low computational costs. Experimental results on multi-objective 0/1 knapsack problems show that GTS remarkably improves the running efficiency and the distribution uniformity of solutions of TDGA. At the same time, TDGA with GTS produces a perfect convergence and spread of solutions as well as NSGA-II, while its running efficiency is much higher than that of NSGA-II.

Keywords multi-objective optimization; evolutionary algorithms; thermodynamical replacement; angle entropy; distance energy

收稿日期:2008-11-13;最终修改稿收到日期:2010-03-04. 本课题得到国家自然科学基金(60773009)、国家“八六三”高技术研究发展计划项目基金(2007AA01Z290)和国家留学基金(2007101731)资助. 应伟勤,男,1979年生,博士研究生,主要研究方向为智能计算、语义计算、图像处理和模式识别. E-mail: weiqinying@yahoo.com.cn. 李元香,男,1962年生,教授,博士生导师,主要研究领域为智能计算和并行计算. SHEU Phillip C-Y,男,1956年生,教授,博士生导师,主要研究领域为语义计算、数据库、软件工程和复杂生物医学系统. 吴昱,女,1982年生,博士研究生,主要研究方向为智能计算. 余法红,男,1977年生,博士研究生,主要研究方向为智能计算和模式识别.

1 引 言

现实世界中的许多优化问题都是多属性的,通常需要对多个目标同时优化.一般来说,多目标优化问题(Multi-Objective Optimization Problems, MOPs)中各目标之间通过决策变量相互制约,对其中一个目标优化必须以其它目标作为代价,因而总体目标没有单个最优解.这使得一次能得到一组最优解的算法比那些一次只得到一个最优解的算法更具优势.基于群体搜索的演化算法可并行搜索多个目标,因此非常适合求解多目标优化问题^[1-2]. Van Veldhuizen 等^[3]从决策者的角度将多目标演化算法(Multi-Objective Evolutionary Algorithms, MOEAs)分为 3 大类:先验法、渐进法和后验法.先验法在搜索之前就需要决策者输入决策信息将多个目标合成为数量成本函数,然后通过 MOEA 产生成本函数的最优解提供给决策者.渐进法中决策者与 MOEA 搜索过程不断交互,前者为后者提供目标的优先关系,而后者为前者提供新解以产生更好的目标间优先关系.后验法则通过运行 MOEA 产生一组 Pareto 最优候选解供决策者选择.该方法使得决策者不必要事先作出各目标之间优先和偏好的决定,将决策过程从“事前”移到“事后”,这样决策者有了对各目标之间关系的总体轮廓后更容易决策^[4].

目前基于 Pareto 占优的后验法最为活跃,大多数 MOEA 采用这种方法.这些算法使用 Pareto 占优指导搜索,在进化过程中,通过 Pareto 非劣分层或 Pareto 占优计数进行适应值指派,使当前进化群体的 Pareto 非劣解集不断逼近真实 Pareto 最优解集.另外,演化算法种群大小有限,而 Pareto 最优解的数量庞大甚至无限,因而 MOEA 算法需用较少的解来代表无限多个解,从而便于决策者进行决策.因此决策者对 MOEA 所求非劣解集存在两方面的需求^[5]:(1)逼近性.即解集尽量逼近 Pareto 最优前沿;(2)多样性.即解集在最优前沿上分布均匀且覆盖宽广,从而能较好地代表整个 Pareto 最优前沿.

为达到这两个要求出现了很多种 MOEA 算法^[2],例如 Fonseca 和 Fleming 提出的 MOGA (Multi-Objective Genetic Algorithm)^[6]、Deb 提出的 NSGA (Non-dominated Sorting Genetic Algorithm)^[7]和 NSGA-II^[5]、Zitzler 和 Thiele 提出的 SPEA (Strength Pareto Evolutionary Algorithm)^[8]等.为逼近前沿,这些算法均通过基于 Pareto 占优

的适应值指派方法来指导种群向最优前沿搜索. MOGA 按 Pareto 占优计数结果为种群中每个个体分配一个等级(rank)值^[6],个体的 rank 值等于 1 加上种群中占优它的个体数. NSGA 和 NSGA-II 则首先进行非劣分层将种群划分成不同的层级,没有被任何个体占优的个体属于第 0 级,将该批个体清除出去后,再没有被任何个体占优的个体属于第 1 级,依次类推^[5,7].分层之后越低层级中的个体分配更好的适应值,使种群以层的形式向 Pareto 最优域推进. SPEA 中进化种群和精英种群的每一个体都依据 Pareto 占优计数结果被指派一个强度值,该值包含了占优和密度信息^[8].这些方法的优点是通用性较好,其不足是计算效率不高,因为这些算法在每代进化中都要进行非劣分层或占优计数操作,两者的时间复杂度均不低于 $O(mn_p^2)$,其中 m 为目标数, n_p 为种群规模.此外,由于演化算法具有随机性, SPEA 和 NSGA-II 分别采用了显式和隐式精英保留策略,保证进化过程中已发现的“好”解不至于丢失.大量实验表明精英保留策略有助于提高 MOEA 的收敛速度.

为保持分布多样性,这些算法采用了各种不同的策略. MOGA 引入基于共享函数的小生境(niching)技术来提高非劣解集的多样性^[6],首先指定一个 niche 半径 δ_{share} ,目标向量之间的距离小于 δ_{share} 的解将共享适应值. NSGA 对每个非劣层也采用适应度共享机制,但其适应度共享方式不同于 MOGA,前者是在决策向量空间共享适应度,而后者是在目标向量空间共享适应度^[7].共享函数法的缺点是计算开销较高,且进化结果对 niche 半径 δ_{share} 的选取很敏感. NSGA-II 采用一种新颖的密度估计和拥挤比较方法来取代共享函数方法^[5],先对同一层级的个体计算拥挤距离,然后将层级中个体按拥挤距离降序排序.该方法能有效保持非劣解集的多样性,不需要额外的参数,并且比共享函数法有更低的时间复杂度. SPEA 则应用聚类方法压缩精英种群^[8]来保持非劣解集的分布多样性.

而 Kita 等^[9]从热力学的自由能极小定律得到启发,在多目标优化中引入能量、熵和温度的概念,设计出一种求解多目标优化问题的热力学遗传算法(Thermodynamical Genetic Algorithm, TDGA). TDGA 最初由 Mori 在求解单目标优化问题时提出^[10],它通过模拟热力学中能量和熵的竞争模式来定量协调演化优化中提高“选择压力”和保持“种群多样性”之间的冲突.热力学系统在恒温下的任何

状态改变都遵循自由能极小定律^[11]; 对于与外界交换热量而温度保持不变的热力学封闭系统, 系统状态的自发变化总是朝着自由能减少的方向进行, 当自由能达到极小值时系统达到平衡态. 热力学中, 系统的自由能 $F = E - HT$, 其中 E 是系统的能量, H 是系统的热力学熵. 能量减少和熵增加均有利于减少自由能, 因此系统的任何状态改变都可被视为 E 和 H 竞争的结果, 温度 T 决定着两者的相对权重. 在热力学系统和演化多目标优化之间存在着诸多相似性. 例如, 种群可视为热力学系统, 个体对应粒子, 种群的平均逼近性、种群多样性和一个权重控制参数 T 则分别可担当热力学中能量、熵和温度的角色. Kita 等^[9]正是在这些相似性基础上, 将 TDGA 算法引入到多目标优化领域, 通过模拟热力学中能量与熵之间的竞争模式统一处理多目标优化在逼近性和多样性两方面的需求.

TDGA 在多目标优化时解集保持了良好的分布宽广性. 但是 TDGA 当前仍存在一些明显的不足. 首先, TDGA 采用基因熵度量种群在基因空间中的多样性^[9], 这使得解集在目标空间的分布均匀性得不到保证. 其次, TDGA 使用贪婪热力学替换规则驱动种群自由能快速下降^[9], 该规则具有较高的时间复杂度, 这导致 TDGA 的计算成本很高^[12]. 再次, TDGA 采用层级能量度量种群的逼近性^[9], 而计算层级能量时需要种群进行耗时的非劣分层操作, 这进一步降低了 TDGA 的运行效率.

本文提出了一种几何热力学选择以改进多目标优化中 TDGA 的运行效率和解集分布均匀性. 该选择充分利用多目标优化中的角度和距离等几何量定义了角度熵和距离能量. 其中角度熵通过在目标空间中扇形采样来度量种群逼近方向的多样性, 提高了解集的分布均匀性; 距离能量运用几何元素间接度量了种群的逼近程度, 巧妙地避免了耗时的非劣分层操作. 此外, 该选择引入几何分量热力学替换规则, 以较低时间复杂度驱动种群几何自由能近似最速下降, 有效降低了替换规则的计算成本. 在多目标 0/1 背包问题上的实验结果表明, 几何热力学选择在保持 TDGA 良好的覆盖宽广性的同时, 极大地提高了 TDGA 的运行效率和解集分布均匀性. 运用该选择的几何热力学遗传算法 (Geometric Thermodynamical Genetic Algorithm, GTGA) 可生成与 NSGA-II 在收敛性和分布性上相当的解, 但在运行效率上明显优于 NSGA-II.

2 多目标优化中的角度和距离

为表述方便首先给出 Pareto 占优的相关概念.

定义 1(多目标优化问题). 一个通用的多目标优化问题 (MOP) 可描述为将 n 维决策变量映射到 m 个目标的向量函数 $F: X \rightarrow Y$:

$$\begin{aligned} \min. / \max. \quad & \mathbf{y} = \mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})) \\ \text{subject to} \quad & \mathbf{x} = (x_1, x_2, \dots, x_n) \in X \subseteq \mathbb{R}^n \\ & \mathbf{y} = (y_1, y_2, \dots, y_m) \in Y \subseteq \mathbb{R}^m \end{aligned} \quad (1)$$

这里 \mathbf{x} 被称为决策向量, X 称为决策空间, \mathbf{y} 称为目标向量, Y 称为目标空间. 多目标优化问题中每一个目标函数 f_i 分别带有一个优化类型, 优化类型可以为最小化 \min 或最大化 \max .

定义 2(规范目标). 多目标优化问题 $F(\mathbf{x})$ 的规范目标为 $\mathbf{z} = \mathbf{G}(\mathbf{x}) = (g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_m(\mathbf{x})) \in Z \subseteq \mathbb{R}^m$, $\mathbf{x} \in X$, 在 f_i 的优化类型为最大化时 $g_i(\mathbf{x}) = -f_i(\mathbf{x})$, 在 f_i 的优化类型为最小化时 $g_i(\mathbf{x}) = f_i(\mathbf{x})$, $i = 1, 2, \dots, m$. 这里 Z 称为规范目标空间. 规范目标的优化类型全部为最小化. 为表述方便, 下文目标空间若无说明均指规范目标空间.

定义 3(Pareto 占优). 设 $\mathbf{z}^1 \in \mathbb{R}^m$ 和 $\mathbf{z}^2 \in \mathbb{R}^m$, 称 \mathbf{z}^1 Pareto 占优 \mathbf{z}^2 (记为 $\mathbf{z}^1 \prec \mathbf{z}^2$) 当且仅当 $(\forall i \in 1..m: z_i^1 \leq z_i^2) \wedge (\exists i \in 1..m: z_i^1 < z_i^2)$. 对于任意两个决策向量 $\mathbf{x}^1 \in X$ 和 $\mathbf{x}^2 \in X$, 称 \mathbf{x}^1 Pareto 占优 \mathbf{x}^2 (记为 $\mathbf{x}^1 \prec \mathbf{x}^2$) 当且仅当 $\mathbf{G}(\mathbf{x}^1) \prec \mathbf{G}(\mathbf{x}^2)$.

定义 4(Pareto 非劣解). 设 $S \subseteq X$, $\mathbf{a} \in S$, 若 $\neg \exists \mathbf{x} \in S: \mathbf{x} \prec \mathbf{a}$, 则称 \mathbf{a} 为 S 的 Pareto 非劣解; 若 \mathbf{a} 为 X 的非劣解, 则 \mathbf{a} 简称为 Pareto 非劣解或 Pareto 最优解.

我们在多目标优化中引入角度和距离这两个几何元素来描述解的性质. 进一步的分析可揭示这两个几何量与 Pareto 占优具有紧密联系, 这在演化多目标优化中利用几何元素度量种群自由能提供了理论依据.

定义 5(解集的基点). 设 $S \subseteq X$, $\mathbf{b} = (b_1, b_2, \dots, b_m) \in \mathbb{R}^m$, 若 $\forall \mathbf{x} \in S: g_i(\mathbf{x}) < b_i, i = 1, 2, \dots, m$, 则称 \mathbf{b} 为 S 的一个基点.

定义 6(解到基点的距离). 设 $\mathbf{x} \in S \subseteq X$, \mathbf{b} 为 S 的一个基点, 则称 $d(\mathbf{x}, \mathbf{b}) = \sqrt{\sum_{i=1}^m (g_i(\mathbf{x}) - b_i)^2}$ 为 \mathbf{x} 到基点 \mathbf{b} 的距离.

定理 1. 设 $S \subseteq X$, $\mathbf{b} \in \mathbb{R}^m$ 为 S 的基点, 若 $\mathbf{x}^* \in S$ 且 $\forall \mathbf{x} \in S: d(\mathbf{x}, \mathbf{b}) \leq d(\mathbf{x}^*, \mathbf{b})$, 则 \mathbf{x}^* 必为 S 的

Pareto非劣解.

证明. 用反证法. 假设 x^* 不是 S 的 Pareto 非劣解, 根据定义 4, 必 $\exists x' \in S: x' \prec x^*$. 再根据定义 3 可得

$$\begin{aligned} & \forall i \in 1..m: g_i(x') \leq g_i(x^*) \wedge \exists i \in 1..m: g_i(x') < g_i(x^*) \\ \Rightarrow & (\forall i \in 1..m: g_i(x') - b_i \leq g_i(x^*) - b_i) \wedge \\ & (\exists i \in 1..m: g_i(x') - b_i < g_i(x^*) - b_i) \end{aligned} \quad (2)$$

由于 b 是 S 的基点且 $x^* \in S$, 由定义 5 可知

$$\forall i \in 1..m: g_i(x^*) - b_i < 0 \quad (3)$$

结合式(2)和式(3)可得

$$\begin{aligned} & (\forall i \in 1..m: (g_i(x') - b_i)^2 \geq (g_i(x^*) - b_i)^2) \wedge \\ & (\exists i \in 1..m: (g_i(x') - b_i)^2 > (g_i(x^*) - b_i)^2) \\ \Rightarrow & \sqrt{\sum_{i=1}^m (g_i(x') - b_i)^2} > \sqrt{\sum_{i=1}^m (g_i(x^*) - b_i)^2} \\ \Rightarrow & d(x', b) > d(x^*, b) \end{aligned} \quad (4)$$

式(4)与定理中条件 $\forall x \in S: d(x, b) \leq d(x^*, b)$ 矛盾, 故假设不成立, 定理得证. 证毕.

定理 1 告诉我们集合 S 中到基点距离最大的解一定是 S 的一个 Pareto 非劣解, 这表明几何中的距离度量与多目标中的 Pareto 占优存在某种程度的关联. 但事实也清晰地表明定理 1 的逆命题并不成立, 即并不是 S 的任一个 Pareto 非劣解与基点的距离最大, 甚至 S 中的 Pareto 被占优解可以比 S 的某些 Pareto 非劣解到基点的距离更大, 这可以很容易举例说明. 为了更清晰地揭示距离度量与 Pareto 占优的关系, 需要给距离度量添加方向限定条件.

定义 7(角度). 设 $S \subseteq X, b \in \mathbb{R}^m$ 为 S 的一个基点, 则称 $\theta(x, b) = (\theta_1(x, b), \theta_2(x, b), \dots, \theta_{m-1}(x, b))$ 为 $x \in S$ 在基点 b 上的角度, 这里

$$\theta_j(x, b) = \arctan\left(\frac{g_{j+1}(x) - b_{j+1}}{g_j(x) - b_j}\right), \quad j=1, 2, \dots, m-1 \quad (5)$$

θ_j 的几何意义是 \mathbb{R}^m 空间中射线 $\overline{bG(x)}$ 在第 $j+1$ 个目标与第 1 个目标张成的坐标平面上的投影与第 1 维坐标轴负方向的夹角. 这 $m-1$ 个夹角共同确定了 \mathbb{R}^m 空间中射线 $\overline{bG(x)}$ 的方向. 最简单的情形是只有两个目标, 此时角度是一个标量. 角度和距离描述了解的很多有用信息. 在已知基点的情况下, 这两个几何量可共同确定目标空间的一个解, 其中角度描述了解向前沿逼近的方向, 距离描述了解逼近前沿的程度. 根据基点和角度的定义容易得出以下推论.

推论 1. 设 b 为 S 的一个基点, 那么 $\forall x \in S: \tan(\theta_j(x, b)) > 0, 0 < \theta_j(x, b) < \pi/2, j=1, 2, \dots,$

$m-1$.

定义 8(距离占优). 设 $S \subseteq X, b \in \mathbb{R}^m$ 为 S 的基点, 若 $x^1 \in S, x^2 \in S, d(x^1, b) > d(x^2, b)$, 则称 x^1 关于基点 b 距离占优 x^2 , 记为 $x^1 \prec_b x^2$.

定理 2. 设 $S \subseteq X, b \in \mathbb{R}^m$ 为 S 的基点, 若 $x^1 \in S, x^2 \in S, \theta(x^1, b) = \theta(x^2, b)$, 那么 $x^1 \prec_b x^2$ 的充要条件是 $x^1 \prec_b x^2$.

证明. 设 $\theta(x^1, b) = \theta(x^2, b) = (\theta_1, \theta_2, \dots, \theta_{m-1}), \theta_0 = \pi/4$, 则根据定义 7 可知

$$\begin{aligned} g_i(x^1) - b_i &= \tan(\theta_{i-1})(g_1(x^1) - b_1), \\ g_i(x^2) - b_i &= \tan(\theta_{i-1})(g_1(x^2) - b_1), \\ & \quad i=1, 2, \dots, m \end{aligned} \quad (6)$$

先证充分性. 由定义 8、式(6)和推论 1 可得

$$\begin{aligned} & x^1 \prec_b x^2 \\ \Rightarrow & \sqrt{\sum_{i=1}^m (g_i(x^1) - b_i)^2} > \sqrt{\sum_{i=1}^m (g_i(x^2) - b_i)^2} \\ \Rightarrow & (g_1(x^1) - b_1)^2 \sum_{i=1}^m \tan^2(\theta_{i-1}) > \\ & (g_1(x^2) - b_1)^2 \sum_{i=1}^m \tan^2(\theta_{i-1}) \\ \Rightarrow & (g_1(x^1) - b_1)^2 > (g_1(x^2) - b_1)^2 \\ \Rightarrow & g_1(x^1) - b_1 < g_1(x^2) - b_1 \\ \Rightarrow & \tan(\theta_{i-1})(g_1(x^1) - b_1) < \tan(\theta_{i-1})(g_1(x^2) - b_1), \\ & \quad i=1, 2, \dots, m \\ \Rightarrow & g_i(x^1) - b_i < g_i(x^2) - b_i, \quad i=1, 2, \dots, m \\ \Rightarrow & g_i(x^1) < g_i(x^2), \quad i=1, 2, \dots, m \\ \Rightarrow & x^1 \prec x^2 \end{aligned} \quad (7)$$

再证必要性. 由定义 5 和定义 6 可得

$$\begin{aligned} & x^1 \prec x^2 \\ \Rightarrow & (\forall i \in 1..m: g_i(x^1) - b_i \leq g_i(x^2) - b_i < 0) \wedge \\ & (\exists i \in 1..m: g_i(x^1) - b_i < g_i(x^2) - b_i < 0) \\ \Rightarrow & (\forall i \in 1..m: (g_i(x^1) - b_i)^2 \geq (g_i(x^2) - b_i)^2) \wedge \\ & (\exists i \in 1..m: (g_i(x^1) - b_i)^2 > (g_i(x^2) - b_i)^2) \\ \Rightarrow & \sqrt{\sum_{i=1}^m (g_i(x^1) - b_i)^2} > \sqrt{\sum_{i=1}^m (g_i(x^2) - b_i)^2} \\ \Rightarrow & x^1 \prec_b x^2 \end{aligned} \quad (8)$$

定理得证.

证毕.

定理 2 说明距离占优并不在所有情况下有实际意义, 但是在角度相等时, 距离占优总是有效的, 此时距离占优等价于 Pareto 占优. 这为本文下一步创造条件利用距离占优代替 Pareto 占优避免耗时的非劣分层操作提供了依据.

3 利用几何元素度量自由能

在演化多目标优化中引入热力学中能量与熵的竞争机制时首先要解决两个基本问题: (1) 如何定义能度量种群多样性的熵; (2) 如何定义能度量种群逼近性的能量. 为改进 TDGA 的计算成本及解集分布均匀性, 本文利用多目标优化中的角度和距离从几何观点重新定义了种群的自由能.

3.1 扇形采样与种群的角度熵

TDGA 沿用了在单目标优化中提出的基因熵, 基因熵等于种群在每一基因位上基因的信息熵的合计值^[9]. 由于基因熵是在基因空间定义的, 因而多目标优化时难以保证非劣解在目标空间中的均匀分布. 本文直接在目标空间中定义种群的熵. 在目标空间中保持非劣解均匀和宽广分布的目的是通过有限的解来代表大量甚至无穷多个最优解, 使决策者能获得各目标之间关系的总体轮廓从而方便决策. 从几何的观点来看, 这实际上是在目标空间对真实 Pareto 前沿进行空间采样的问题. Knowles^[13-14] 在 PAES 算法中提出了对目标空间进行网格采样的方法, 该方法通过均匀划分每一目标的取值范围得到矩形网格. Ding^[15] 曾把网格采样方法应用到 TDGA 中. Zheng^[4,16] 提出了对目标空间进行扇形采样的方法, 该方法具有比网格采样更低的时间和空间成本, 而且有利于进一步利用几何度量个体能量, 因此本文采用扇形采样方式. 首先利用角度定义扇形区域.

定义 9(上界和下界). 设 $P_t = (\mathbf{x}^{t,0}, \mathbf{x}^{t,1}, \dots, \mathbf{x}^{t,n_p-1})$ 为算法在第 $t(t \geq 0)$ 代的种群, $O_t = (\mathbf{x}^{t,n_p}, \mathbf{x}^{t,n_p+1}, \dots, \mathbf{x}^{t,n_p+n_o-1})$ 为算法在第 $t(t \geq 0)$ 代的子种群, 则称 $\mathbf{u}^{(t)} = (u_1^{(t)}, u_2^{(t)}, \dots, u_m^{(t)}) \in \mathbb{R}^m$ 和 $\mathbf{l}^{(t)} = (l_1^{(t)}, l_2^{(t)}, \dots, l_m^{(t)}) \in \mathbb{R}^m$ 为第 t 代的上界和下界, 这里

$$\begin{aligned} u_i^{(t)} &= \max(u_i^{(t-1)}, \max\{g_i(\mathbf{x}) \mid \mathbf{x} \in O_{t-1}\}), t > 0, \\ u_i^{(0)} &= \max\{g_i(\mathbf{x}) \mid \mathbf{x} \in P_0\}, \\ l_i^{(t)} &= \min(l_i^{(t-1)}, \min\{g_i(\mathbf{x}) \mid \mathbf{x} \in O_{t-1}\}), t > 0, \\ l_i^{(0)} &= \min\{g_i(\mathbf{x}) \mid \mathbf{x} \in P_0\}, i = 1, 2, \dots, m \end{aligned} \quad (9)$$

定义 10(第 t 代基点). 算法第 t 代基点 $\mathbf{b}^{(t)} = (b_1^{(t)}, b_2^{(t)}, \dots, b_m^{(t)}) \in \mathbb{R}^m$ 由如下两条规则生成:

(1) $t = 0$ 时, $b_i^{(0)} = u_i^{(0)} + 40(u_i^{(0)} - l_i^{(0)}), i = 1, 2, \dots, m;$

(2) $t > 0$ 时, 若 $u_i^{(t)} \geq b_i^{(t-1)}$, 则 $b_i^{(t)} = u_i^{(t)} + 40(u_i^{(t)} - l_i^{(t)})$; 若 $u_i^{(t)} < b_i^{(t-1)}$, 则 $b_i^{(t)} = b_i^{(t-1)}, i = 1,$

$2, \dots, m.$

定义 11(扇形区域). 设第 t 代基点为 $\mathbf{b}^{(t)} \in \mathbb{R}^m$, 那么在区域 $D^{(t)} = \{\mathbf{z} \in \mathbb{R}^m \mid z_i < b_i^{(t)}, i = 1, 2, \dots, m\}$ 中通过将每个角度分量的取值范围 $(0, \pi/2)$ 划分 K 等份可得到 K^{m-1} 个扇形区域 $s_v^{(t)}$, 这里

$$s_v^{(t)} = \{\mathbf{z} \in D^{(t)} \mid v_j \pi / (2K) \leq \theta_j(\mathbf{z}, \mathbf{b}^{(t)}) < (v_j + 1) \pi / (2K), j = 1, 2, \dots, m-1\} \quad (10)$$

其中整数常量 K 为每个角度分量上的等份数, $\mathbf{v} = (v_1, v_2, \dots, v_{m-1})$ 为扇形区域 $s_v^{(t)}$ 的坐标, 表示扇形 $s_v^{(t)}$ 中所有个体的第 j 个角度分量都位于第 v_j 个等份中, $v_j \in 0..K-1, j = 1, 2, \dots, m-1$. 称 $s_v^{(t)}$ 为基点 $\mathbf{b}^{(t)}$ 下坐标为 \mathbf{v} 的扇形区域. 对于任意个体 $\mathbf{x} \in P_t$, 若 $\mathbf{G}(\mathbf{x}) \in s_v^{(t)}$, 则称 \mathbf{x} 位于扇形 $s_v^{(t)}$ 中.

定义 12(角度熵). 设第 t 代种群为 $P_t \in X^{n_p}$, 第 t 代基点为 $\mathbf{b}^{(t)} \in \mathbb{R}^m$, 那么称 $H(\mathbf{b}^{(t)}, P_t)$ 为种群 P_t 在基点 $\mathbf{b}^{(t)}$ 上的角度熵, 这里

$$H(\mathbf{b}^{(t)}, P_t) = - \sum_{\mathbf{v}=(0,0,\dots,0)}^{(K-1,K-1,\dots,K-1)} \frac{n_v}{n_p} \log \frac{n_v}{n_p} \quad (11)$$

其中 n_v 为种群 P_t 中位于扇形 $s_v^{(t)}$ 中的个体数, n_p 为种群 P_t 的个体总数.

上界和下界记录了到第 t 代为止搜索过程已达个体的范围, 它们将随搜索过程的进行不断地扩张. 其中上界被任意已达个体占优, 下界占优任何已达个体. 第 t 代基点设为被当前所有已达个体强占优并且距离已达个体比较遥远的点, 这使得基点位置变化的概率非常小, 并且从基点出发对非劣前沿的采样将更加均匀. 已达个体在当前基点上的每个角度分量 $\theta_j \in (0, \pi/2), j = 1, 2, \dots, m-1$. 通过将每个角度分量的取值范围 $(0, \pi/2)$ 划分 K 个等份, 可最终将强占优基点的区域划分成 K^{m-1} 个扇形, 二目标时的扇形划分情形如图 1 所示. 网格划分法将产生 K^m 个矩形, 二目标时的网格划分情形如图 2 所示, 这明显比扇形划分法占用更多的存储空间和采样时间. 根据定义 11 可知个体 $\mathbf{x} \in P_t$ 所在扇形的坐标为 $\mathbf{v} = (v_1, v_2, \dots, v_{m-1}), v_j = \text{floor}(2K * \theta_j(\mathbf{x}, \mathbf{b}^{(t)}) / \pi), j = 1, 2, \dots, m-1$, 这里 $\text{floor}(\cdot)$ 表示下取整函数. 角度熵 $H(\mathbf{b}^{(t)}, P_t)$ 度量了种群逼近方向的分散程度. 可以证明当 P_t 中所有个体全部位于同一扇形时角度熵 $H(\mathbf{b}^{(t)}, P_t)$ 取最小值 0, 当每一扇形中均有个体出现并且数量相等即

$n_{(0,0,\dots,0)} = n_{(1,0,\dots,0)} = \dots = n_{(K-1,K-1,\dots,K-1)}$ 时 $H(\mathbf{b}^{(t)}, P_t)$ 达到最大值 $(m-1) \log K$.

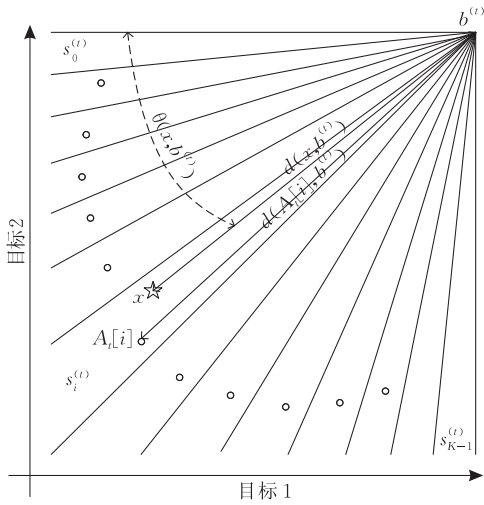


图 1 二目标时扇形划分示意图

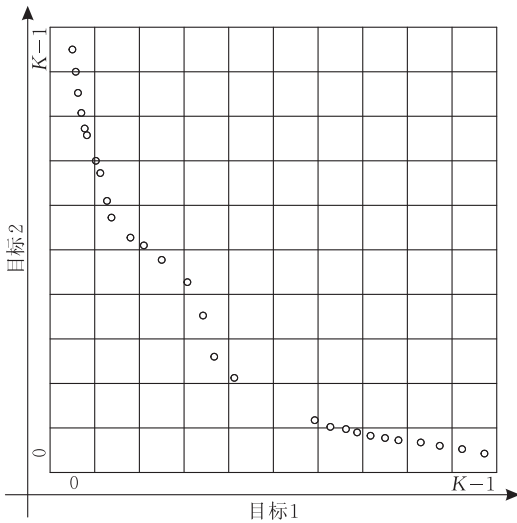


图 2 二目标时网格划分示意图

3.2 距离精英与种群的距离能量

TDGA 当前采用层级能量度量逼近性, 个体的层级能量等于该个体所在层级的编号^[9], 种群的能量等于个体能量的均值, 层级能量越低逼近程度越高. 显然计算层级能量时需要首先对种群进行耗时的非劣分层操作, 这对 TDGA 的运行效率有较大影响. 由定理 2 可知角度相等时距离占优可代替 Pareto 占优, 当扇形划分中 K 较大时每个扇形的夹角很小, 同一扇形中个体角度非常相近, 此时可用距离来近似度量同一个扇形中个体的逼近程度, 避免耗时的非劣分层操作. 为使不同扇形中的距离具可比性, 还需建立距离精英文档来保存每个扇形中的距离精英.

定义 13(距离精英文档). 距离精英文档 A_t 是一个长度为 K^{m-1} 的决策向量数组, 它保存至第 t 代每个扇形中的距离最好解, 其中坐标为 $\mathbf{v} =$

$(v_1, v_2, \dots, v_{m-1})$ 的扇形区域 $s_v^{(t)}$ 对应于文档元素 $A_t[index]$, 这里 $index = v_1 + v_2 \times K + \dots + v_{m-1} \times K^{m-2}$ 称为扇形区域 $s_v^{(t)}$ 的文档索引. 距离精英文档 A_t 由以下 3 种操作得到

(1) $t=0$ 时, $A_0 = initialize(\mathbf{b}^{(0)}, P_0)$;

(2) 若 $t>0$ 且 $\mathbf{b}^{(t)} = \mathbf{b}^{(t-1)}$, 则 $A_t = update(\mathbf{b}^{(t)}, A_{t-1}, O_{t-1})$;

(3) 若 $t>0$ 且 $\mathbf{b}^{(t)} \neq \mathbf{b}^{(t-1)}$, 则 $A_t = reset(\mathbf{b}^{(t)}, A_{t-1}, P_{t-1}, O_{t-1})$.

其中 $initialize()$, $update()$ 和 $reset()$ 3 个操作分别初始化文档, 根据子个体更新文档, 在基点改变时重置文档. 这 3 个操作都需要调用 $insert(A, \mathbf{b}, \mathbf{x})$ 函数, 函数 $insert(A, \mathbf{b}, \mathbf{x})$ 比较当前个体 \mathbf{x} 和 \mathbf{x} 所在扇形的距离精英 $A[index]$ 到基点 \mathbf{b} 的距离, 若 \mathbf{x} 距离占优 $A[index]$, 则替换距离精英 $A[index]$ 为 \mathbf{x} . 这 4 个函数可描述如下.

算法 1. 文档初始化 $A_0 = initialize(\mathbf{b}^{(0)}, P_0)$.

1. for($index=0$; $index < K^{m-1}$; $index++$)

$A_0[index] = NULL$;

2. for($i=0$; $i < n_p$; $i++$) $A_0 = insert(A_0, \mathbf{b}^{(0)}, P_0[i])$.

算法 2. 文档更新 $A_t = update(\mathbf{b}^{(t)}, A_{t-1}, O_{t-1})$.

1. $A_t = A_{t-1}$;

2. for($i=0$; $i < n_o$; $i++$) $A_t = insert(A_t, \mathbf{b}^{(t)}, O_{t-1}[i])$.

算法 3. 文档重置 $A_t = reset(\mathbf{b}^{(t)}, A_{t-1}, P_{t-1}, O_{t-1})$.

1. for($index=0$; $index < K^{m-1}$; $index++$)

$A_t[index] = NULL$;

2. for($index=0$; $index < K^{m-1}$; $index++$)

3. if($A_{t-1}[index] \neq NULL$)

$A_t = insert(A_t, \mathbf{b}^{(t)}, A_{t-1}[index])$;

4. for($i=0$; $i < n_p$; $i++$) $A_t = insert(A_t, \mathbf{b}^{(t)}, P_{t-1}[i])$;

5. for($i=0$; $i < n_o$; $i++$) $A_t = insert(A_t, \mathbf{b}^{(t)}, O_{t-1}[i])$.

算法 4. 添加个体到文档 $A = insert(A, \mathbf{b}, \mathbf{x})$.

1. $index=0$; //计算个体 \mathbf{x} 所在扇形的坐标及文档索引

2. for($j=1$; $j \leq m-1$; $j++$) {

3. $aj = \arctan((g_{j+1}(\mathbf{x}) - b_{j+1}) / (g_1(\mathbf{x}) - b_1))$;

// \mathbf{x} 的角度分量 θ_j

4. $cj = \text{floor}(2K * aj / \pi)$; // \mathbf{x} 所在扇形的坐标分量 c_j

5. $index += cj * K^{j-1}$; } // \mathbf{x} 所在扇形的的文档索引

6. if($(A[index] == NULL)$ or

$(d(\mathbf{x}, \mathbf{b}) > d(A[index], \mathbf{b}))$)

7. $A[index] = \mathbf{x}$.

定义 14(距离能量). 设 $\mathbf{b}^{(t)} \in \mathbb{R}^m$ 是第 t 代基点, $\mathbf{x} \in P_t$, 那么称 $e(\mathbf{b}^{(t)}, A_t, \mathbf{x})$ 为个体 \mathbf{x} 在基点 $\mathbf{b}^{(t)}$ 下的距离能量, $E(\mathbf{b}^{(t)}, A_t, P_t)$ 为种群 P_t 在基点 $\mathbf{b}^{(t)}$

下的距离能量, 这里

$$e(\mathbf{b}^{(t)}, A_t, \mathbf{x}) = d(A_t[\text{index}], \mathbf{b}^{(t)}) - d(\mathbf{x}, \mathbf{b}^{(t)}) \quad (12)$$

$$E(\mathbf{b}^{(t)}, A_t, P_t) = \frac{1}{n_p} \sum_{\mathbf{x} \in P^{(t)}} e(\mathbf{b}^{(t)}, A_t, \mathbf{x}) \quad (13)$$

其中 $\text{index} = v_1 + v_2 \times K + \dots + v_{m-1} \times K^{m-2}$ 为 \mathbf{x} 所在扇形的文档索引, $\mathbf{v} = (v_1, v_2, \dots, v_{m-1})$ 为 \mathbf{x} 所在扇形的坐标, n_p 为种群 P_t 的个体总数.

给定一个一般的解集 S , S 中任意两个个体 Pareto 占优比较可出现 4 种结果: 相等、占优、被占优、不可比较, 所以 Pareto 占优一般并不是 S 上的全序关系, 存在两个个体互不占优不可比较的情况. 若同一个扇形中个体的角度都相同, 则扇形成为一条射线, 根据定理 2 该扇形中距离比较可代替 Pareto 比较, 因而 Pareto 比较只可能有相等、占优、被占优 3 种结果. 此时 Pareto 占优关系是该扇形上的一个严格全序关系, 任意两个个体都是可比较的, 在每个扇形中存在一个 Pareto 最好解, 即到基点距离最大的距离精英, 故可以选取该解代表这个扇形中的所有解. 个体的距离能量等于个体所在扇形的距离精英到基点的距离与该个体到基点的距离之差. 那么距离能量小的个体到基点的距离更大, 根据定理 2 距离大的个体肯定 Pareto 占优距离小的个体, 所以距离能量小的个体更加逼近 Pareto 最优前沿. 当 K 值较大时同一扇形中个体角度非常相近, 上述扇形中个体的角度相同的假设可近似成立, 因而个体的距离能量可近似度量个体的逼近程度.

这里精英文档实际上扮演了两方面的角色. 首先为每个扇形的距离度量设立了一个标尺, 每个扇形中的精英个体距离能量最小且为 0, 不同扇形中个体比较时越靠近本扇形精英的个体距离能量越小. 其次精英文档记录了至第 t 代每个扇形中的距离最好解, 可实现精英保留策略. 根据定理 1, 无论每个扇形的角度范围有多大, 其距离精英个体一定是该扇形中个体的 Pareto 非劣解, 这使得利用几何量也可以间接构造出 Pareto 非劣前沿.

度量层级能量的主要操作在非劣分层, 需要个体两两间做 Pareto 占优比较, 时间复杂度为 $O(m(n_p + n_o)^2)$. 度量距离能量的主要操作为距离比较, 其中每个个体只需与其所在扇形中的距离精英比较, 其复杂度为 $O(m(n_p + n_o))$, 明显低于层级能量的复杂度.

3.3 种群的几何自由能

上面已经揭示了目标空间中角度、距离这两个几何元素与 Pareto 占优之间的联系, 并充分利用角

度、距离定义了种群的角度熵和距离能量. 现在可根据角度熵和距离能量从几何的观点定义种群的自由能. 种群的几何自由能将多目标优化中的逼近性和多样性在温度 T 下统一起来.

定义 15(几何自由能). 设 $\mathbf{b}^{(t)} \in \mathbb{R}^m$ 是第 t 代基点, $P_t \in X^{n_p}$ 是第 t 代种群, 那么称 $F(\mathbf{b}^{(t)}, A_t, T, P_t)$ 为种群 P_t 在基点 $\mathbf{b}^{(t)}$ 和温度 T 下的几何自由能, 这里

$$F(\mathbf{b}^{(t)}, A_t, T, P_t) = E(\mathbf{b}^{(t)}, A_t, P_t) - TH(\mathbf{b}^{(t)}, P_t) \quad (14)$$

4 几何分量热力学替换规则

热力学替换规则的设计是在演化多目标优化中引入能量和熵竞争机制的另一个关键问题. 热力学替换规则的任务是从父代种群 $P_t = (\mathbf{x}^{t,0}, \mathbf{x}^{t,1}, \dots, \mathbf{x}^{t,n_p-1})$ 的 n_p 个个体及其子种群 $O_t = (\mathbf{x}^{t,n_p}, \mathbf{x}^{t,n_p+1}, \dots, \mathbf{x}^{t,n_p+n_o-1})$ 的 n_o 个个体中挑选出 n_p 个个体组成下一代种群 $P_{t+1}^{(E)}$, 使其具有的自由能 $F(\mathbf{b}^{(t+1)}, A_{t+1}, T, P_{t+1}^{(E)})$ 最小. 然而, 使用穷举法精确地最小化每一代的自由能本身是一个颇难的组合优化问题, 基于角度熵和距离能量的穷举替换规则复杂度高达 $O(K^{m-1} C_{n_p+n_o}^{n_p})$, 实际中是不可行的. 为降低替换规则的复杂度, TDGA 当前采用贪婪热力学替换 (Greedy Thermodynamical Replacement, GTR) 规则^[9], 按贪婪策略逐个往下一代种群中填充个体. GTR 规则的计算开销较穷举规则有很大程度降低, 基于角度熵和距离能量的 GTR 规则的复杂度为 $O(m(n_p + n_o) + K^{m-1} n_p (n_p + n_o))$, 但 GTR 规则在实际应用中计算成本仍相当高, 不具实用性. 在单目标优化中我们提出了一种基于等级熵和相对能量的分量热力学替换规则 CTR^[17-18], 通过自由能分量的概念将种群的自由能分派到其各个体上, 极大地简化了热力学替换规则的设计. 本文进一步将分量替换的思想引入到多目标优化中以降低替换规则的复杂度. 首先定义多目标优化中的几何自由能分量.

定义 16(几何自由能分量). 设 P_t 为第 t 代种群, $\mathbf{b}^{(t)}$ 为第 t 代基点, $\mathbf{x} \in P_t$, 那么称 $F_c(\mathbf{b}^{(t)}, A_t, T, P_t, \mathbf{x})$ 为在基点 $\mathbf{b}^{(t)}$ 、文档 A_t 和温度 T 下个体 \mathbf{x} 在种群 P_t 中的几何自由能分量, 这里

$$F_c(\mathbf{b}^{(t)}, A_t, T, P_t, \mathbf{x}) = e(\mathbf{b}^{(t)}, A_t, \mathbf{x}) + T \log(n_v/n_p) \quad (15)$$

其中 $\mathbf{v} = (v_1, v_2, \dots, v_{m-1})$ 为 \mathbf{x} 所在扇形的坐标, $v_j = \text{floor}(2K * \theta_j(\mathbf{x}, \mathbf{b}^{(t)})/\pi)$, $j = 1, 2, \dots, m-1$, n_v 为种群 P_t 中位于坐标为 \mathbf{v} 的扇形中的个体数, n_p 为种群 P_t 的个体总数.

几何分量热力学替换规则通过计算个体在中间种群 P'_{t+1} 中的几何自由能分量可间接达到近似最小化下一代种群几何自由能的目的, 这样极大地减小了替换规则的计算开销. 如算法 5 所示, 几何 CTR 规则主要有 3 个步骤.

算法 5. 几何分量热力学替换规则 $P_{t+1} = \text{CTR}(\mathbf{b}^{(t+1)}, A_{t+1}, T, P_t, O_t)$.

1. 合并父种群 P_t 和子种群 O_t 得到中间种群 P'_{t+1} ;
2. 计算中间种群 P'_{t+1} 中所有个体的几何自由能分量 $F_c(\mathbf{b}^{(t+1)}, A_{t+1}, T, P'_{t+1}, \mathbf{x}), \mathbf{x} \in P'_{t+1}$;
3. 在 P'_{t+1} 中挑选出 n_o 个自由能分量最大的个体, 从 P'_{t+1} 中删除这 n_o 个个体即可得到下一代种群 P_{t+1} .

几何 CTR 规则中计算自由能分量的复杂度为 $O(m(n_p + n_o))$, 挑选淘汰个体的复杂度为 $O(n_o(n_p + n_o))$, 故几何 CTR 规则的复杂度为 $O((m + n_o)(n_p + n_o))$, 明显降低了 GTR 规则的计算成本. 与单目标优化中类似, 我们亦可证明多目标优化中的几何 CTR 规则具有驱动种群的几何自由能近似最速下降的能力.

定理 3. 对于父代种群 P_t 和子种群 O_t , 设 $P_{t+1}^{(E)}$ 为几何自由能最小的下一代种群, $P_{t+1}^{(C)}$ 为由几何 CTR 规则生成的下一代种群, 即 $P_{t+1}^{(C)} = \text{CTR}(\mathbf{b}^{(t+1)}, A_{t+1}, T, P_t, O_t)$, $y = n_o/(n_p + n_o)$, $D(\mathbf{b}^{(t+1)}, A_{t+1}, T, P_t, O_t) = F(\mathbf{b}^{(t+1)}, A_{t+1}, T, P_{t+1}^{(C)}) - F(\mathbf{b}^{(t+1)}, A_{t+1}, T, P_{t+1}^{(E)})$, 则

$$\lim_{y \rightarrow 0^+} (D(\mathbf{b}^{(t+1)}, A_{t+1}, T, P_t, O_t)) = 0 \quad (16)$$

运用极限夹逼准则可证明定理 3, 这里省略了定理 3 的证明, 其证明过程可参考文献[18]. 定理 3 告诉我们, 当 $n_o \ll n_p$ 时, 几何 CTR 规则具有驱动种群几何自由能至少以接近最快的进化速度下降的能力. 定理 3 保证了多目标优化中几何 CTR 规则的有效性.

5 几何热力学遗传算法

上述几何 CTR 规则亦称为基于角度熵和距离能量的 CTR 规则. 角度熵、距离能量、冷却进度表和分量热力学替换规则这 4 个要素一起构成了几何热力学选择, 它驱使种群快速向几何自由能减少的方向演化, 从而营造出热力学中能量和熵的竞争模式. 本文

称采用几何热力学选择的热力学遗传算法为几何热力学遗传算法 GTGA. 算法 6 给出了 GTGA 的一般流程.

算法 6. 几何热力学遗传算法 GTGA.

1. 随机生成 n_p 个个体组成初始种群 P_0 , 并评估 P_0 中个体;
2. 按定义 10 中规则(1)生成第 0 代基点 $\mathbf{b}^{(0)}$;
3. $A_0 = \text{initialize}(\mathbf{b}^{(0)}, P_0)$;
4. $T_0 = 10$; $t = 0$;
5. **while**($\text{terminate}(P_t, t) = \text{False}$) {
6. 从 $P_t \cup A_t$ 通过均匀选择、交叉、变异生成 n_o 个个体;
7. 将这 n_o 个个体组织成子种群 O_t , 并评估 O_t 中个体;
8. 按定义 10 中规则(2)生成第 $t+1$ 代基点 $\mathbf{b}^{(t+1)}$;
9. **if**($\mathbf{b}^{(t+1)} = \mathbf{b}^{(t)}$) $A_{t+1} = \text{update}(\mathbf{b}^{(t+1)}, A_t, O_t)$;
10. **else** $A_{t+1} = \text{reset}(\mathbf{b}^{(t+1)}, A_t, P_t, O_t)$;
11. $P_{t+1} = \text{CTR}(\mathbf{b}^{(t+1)}, A_{t+1}, T_t, P_t, O_t)$;
//调用几何 CTR 规则
12. $t = t + 1$;
13. $T_t = T_0 / (1 + t)$;
14. 对 A_t 进行非劣分层, 输出 A_t 中的 Pareto 非劣解集.

上述 GTGA 算法中初始温度 T_0 和温度衰减函数 $T_t = T_0 / (1 + t)$ 共同构成了冷却进度表. GTGA 在进化迭代结束后得到的精英文档是各扇形区域的非劣解的集合. 整个目标空间的非劣解肯定是其所在扇形的一个非劣解, 但扇形的非劣解不一定是整个目标空间的非劣解. 例如, 位于两端的扇形非劣解可能被其它解占优, 如图 3 所示. 因此算法结束时需要调用 Pareto 非劣分层操作对扇形非劣解的集合进行后处理, 将被占优的扇形非劣解清除出去即可得到最终的 Pareto 非劣前沿.

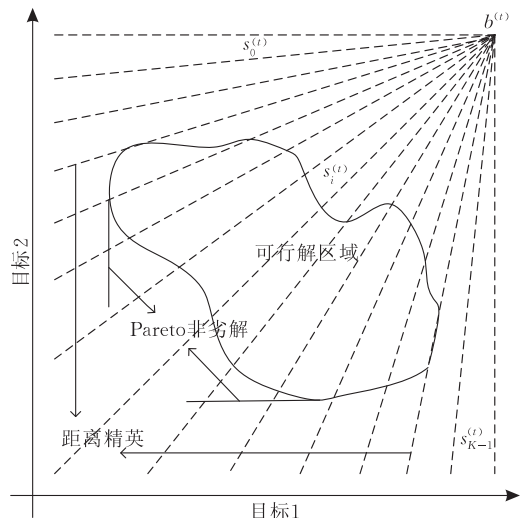


图 3 距离精英与非劣解示意图

GTGA 在一次迭代中文档更新操作 $\text{update}()$

的复杂度为 $O(n_o m)$; 由于种群的基点设为距已达个体比较遥远的点, 这使得 GTGA 中基点发生改变需要调用重置文档操作 $reset()$ 的概率非常小, 可忽略不计; 几何 CTR 规则的复杂度为 $O((m+n_o)(n_p+n_o))$. 故在 $n_o \leq n_p$ 时 GTGA 算法一次迭代的复杂度约为 $O((m+n_o)n_p)$, 因此 GTGA 算法的总复杂度约为 $O(N_e/n_o(m+n_o)n_p) = O(N_e(m+n_o)(n_p/n_o))$, 其中 N_e 为允许的最大评估次数. 而 TDGA 当前采用的基于基因熵和层级能量的 GTR 规则在 $n_o \leq n_p$ 时一次迭代的复杂度约为 $O((m+n)n_p^2)$, 故 TDGA 算法的总复杂度约为 $O(N_e/n_o(m+n)n_p^2) = O(N_e(mn_p+mn_p)(n_p/n_o))$. 为分别对比分析替换规则、熵度量方法和能量度量方法对算法运行效率的影响, 我们先将 TDGA 中 GTR 规则替换为 CTR 规则可得到算法 TDGA2. TDGA2 采用基于基因熵和层级能量的 CTR 规则, 依上面方法分析可知 TDGA2 算法的复杂度约为 $O(N_e(mn_p+n)(n_p/n_o))$. 进一步将 TDGA2 中基因熵度量方法替换为角度熵度量方法可得到算法 TDGA3. TDGA3 采用基于角度熵和层级能量的 CTR 规则, 同样可分析得到 TDGA3 算法的复杂度约为 $O(N_e(mn_p)(n_p/n_o))$. 以上分析表明分量替换规则、角度熵和距离能量均改进了 TDGA 的运行效率, 其中分量替换规则和距离能量的效果尤为明显.

6 实验与分析

为了检验上述几何热力学选择方法的性能, 本文以 C++ 实现了相关多目标演化算法, 并在多目标 0/1 背包问题上对这些算法进行了一系列对比实验. 所有实验均在 Intel(R) Core™ 2 Duo CPU T9300 /2GB 内存的计算机上进行.

6.1 性能度量

衡量多目标演化算法的性能通常需要考虑 3 方面的因素: (1) 逼近性. 所求非劣解集逼近 Pareto 最优前沿的程度; (2) 多样性. 所求非劣解集的分散程度, 这又包括均匀性和宽广性, 其中均匀性表示解集分布的均匀程度, 宽广性表示解集覆盖最优前沿的完整程度; (3) 计算成本. 本文使用 C 和 Δ 两个评价标准来比较各算法所求二目标问题非劣解集的优劣. 其中 $C(S, S')$ 用于比较两个解集逼近前沿的程度^[8,19], 这里

$$C(S, S') = \frac{|\{a' \in S' \mid \exists a \in S: a < a'\}|}{|S'|} \quad (17)$$

$C(S, S')$ 表示了 S' 被 S 占优的比率. 若 $C(S, S') >$

$C(S', S)$, 则可认为解集 S 逼近最优前沿的程度要优于解集 S' . $\Delta(S)$ 度量了二目标问题非劣解集 S 的分散程度^[20]. 设二目标问题边界点为 E_1 和 E_2 , 将边界点加入到非劣前沿 $F(S)$ 中可得到 $Q = F(S) \cup \{E_1, E_2\}$, 并按第一目标升序排列前沿 Q , 那么 $\Delta(S)$ 可如下定义:

$$\Delta(S) = \frac{1}{|Q|-1} \sum_{i=1}^{|Q|-1} |d_i - \bar{d}|, \quad \bar{d} = \frac{1}{|Q|-1} \sum_{i=1}^{|Q|-1} d_i \quad (18)$$

其中 $d_i = \|F(x^i) - F(x^{i+1})\|_2$, 这里 x^i 为序列 Q 中第 i 个解, $F(x^1) = E_1$, $F(x^{|Q|}) = E_2$. Δ 综合考虑了 S 的分布均匀性和覆盖宽广性, 分布越均匀覆盖越完整的非劣解集具有更小的 Δ 值. 另外计算成本表示算法平均搜索一个解所需要付出的计算代价, 本文使用在相同机器上算法搜索相同数目个体的运行时间度量计算成本.

6.2 测试例

为方便评估各算法的性能, 本文选用 Zitzler^[8] 提出的 3 个二目标 0/1 背包问题作为测试例, 分别记为 2-100、2-250、2-500, 其背包数目均为 2, 物品数目分别为 100、250、500, 其 Pareto 最优前沿的边界点见表 1. 实验中所有算法均采用长为 n 的二进制串编码, 其中 n 为测试例的物品数.

由于遗传操作可能产生不可行解, 本文采用加权标量 (weighted scalar) 修补算子^[21] 将不可行解转化为可行解. 该修补算子按加权标量 q_j 的升序以贪婪方式逐个清除背包中物品, 直到满足所有背包的容量限制. 这里物品 j 的加权标量 q_j 定义为

$$q_j = \sum_{i=1}^m \lambda_i p_{ij} / \sum_{i=1}^m w_{ij}, \quad j=1, 2, \dots, n \quad (19)$$

其中 p_{ij} 和 w_{ij} 分别为物品 j 在背包 i 中的收益与重量, $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$ 为关联于当前解的权重向量. 本文也进一步采用了加权标量填充算子加快局部搜索. 该填充算子只要背包容量允许, 就按加权标量 q_j 的降序以贪婪方式逐个往背包中添加剩余物品.

表 1 测试例的参数值

测试例	边界点	变异概率 P_m	Niche 半径 δ	最大评估次数 N_e
2-100	$E_1 = (4266, 3215)$ $E_2 = (3235, 4037)$	0.02	35	1×10^5
2-250	$E_1 = (9893, 7570)$ $E_2 = (7283, 10103)$	0.008	115	2×10^5
2-500	$E_1 = (20094, 16315)$ $E_2 = (15780, 20490)$	0.004	236	4×10^5

6.3 实验数据分析

本文首先选择 4 种多目标演化算法来对比分析

几何热力学选择的性能,分别为非劣分层遗传算法 NSGA^[7]、精英保留非劣分层遗传算法 NSGA-II^[5],采用基于基因熵和层级能量的分量替换规则的热力学遗传算法 TDGA2 和几何热力学遗传算法 GTGA. 所有算法均采用均匀交叉算子和翻转变异算子,变异后先对不满足约束的个体应用加权标量修补算子,之后再对该个体应用加权标量填充算子. 所有算法的种群规模 $n_p = 200$,交叉概率 $P_c = 0.8$,变异概率 $P_m = 2/n$,终止条件 *terminate* 在评估次数达到 N_e 时满足, N_e 随问题的复杂度而增加,如表 1 所示. 由于 NSGA 对 Niche 半径 δ 比较敏感,本文使用了与文献[8]中相同的 δ 值. TDGA2 和 GTGA 中子种群 O_i 的规模 $n_o = 1/10 \times n_p = 20$,初始温度 $T_0 = 10$,GTGA 中角度分量等份数 $K = 10000$.

应用 4 个算法各求解 3 个测试例 10 次,图 4~图 6 分别呈现了算法在 3 个测试例上第 1 次运行所得到的非劣前沿. 表 2~表 4 分别给出了在 3 个测试例上各算法之间利用 C 度量得到的逼近性比较结果. 其中,每一列的值皆代表一种算法与其它算法比较的 C 度量值在 10 次实验中的平均数. 例如表 3 的第 2 行第 4 列的值 $C(\text{NSGA-II}, \text{GTGA}) = 0.1858$,代表 GTGA 所得的解平均有将近二成被 NSGA-II 的解所占优了;同理,第 4 行第 2 列的值 $C(\text{GTGA}, \text{NSGA-II}) = 0.3054$,代表 NSGA-II 平均有约三成的解被 GTGA 的解所占优;从 $C(\text{NSGA-II}, \text{GTGA}) < C(\text{GTGA}, \text{NSGA-II})$ 可知 GTGA 在 2-250 上的实验中逼近效果要比 NSGA-II 略好. 从图 4~图 6 可直观看出 NSGA-II 和 GTGA 的逼近效果明显比 NSGA 和 TDGA2 好很多,进一步从表 2~表 4 可知后者的解绝大部分甚至全部被前者占优,这主要是因为 NSGA 和 TDGA2 均不带有精英保留机制. 实验中 NSGA-II 和 GTGA 的逼近效果相当,在物品数较少的 2-100 上 NSGA-II 逼近性略好于 GTGA,但在物品数较多的 2-250 和 2-500 上 GTGA 逼近性略好于 NSGA-II.

表 2 2-100 上利用 C 度量得到的逼近性比较结果

	NSGA	NSGA-II	TDGA2	GTGA
NSGA	—	0	0.4311	0.0019
NSGA-II	0.6934	—	0.7353	0.0835
TDGA2	0.0354	0.0150	—	0.0201
GTGA	0.6659	0.0256	0.7284	—

表 3 2-250 上利用 C 度量得到的逼近性比较结果

	NSGA	NSGA-II	TDGA2	GTGA
NSGA	—	0.0032	0.3406	0.0012
NSGA-II	0.9192	—	0.8561	0.1858
TDGA2	0	0.0076	—	0.0089
GTGA	0.9437	0.3054	0.8922	—

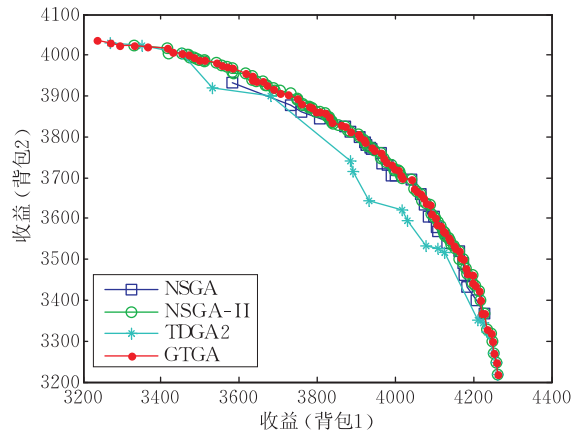


图 4 2-100 上各算法所得非劣前沿

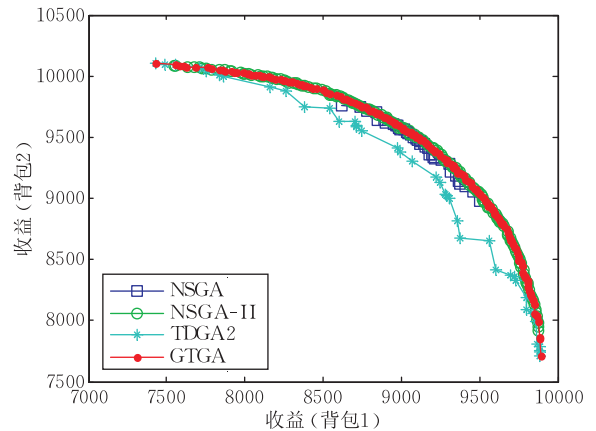


图 5 2-250 上各算法所得非劣前沿

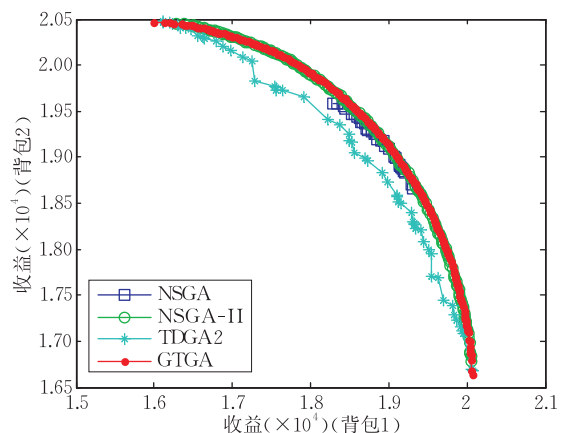


图 6 2-500 上各算法所得非劣前沿

表 4 2-500 上利用 C 度量得到的逼近性比较结果

	NSGA	NSGA-II	TDGA2	GTGA
NSGA	—	0.0052	0.2982	0
NSGA-II	0.8565	—	0.8717	0.1107
TDGA2	0	0.0006	—	0.0049
GTGA	0.9690	0.5177	0.8926	—

表 5 显示了算法在各测试例上 10 次实验中多样性度量 Δ 的均值. 除了在 2-100 上 NSGA-II 略优于 GTGA 外,在 2-250 和 2-500 上 GTGA 都具有最

小的 Δ 值,这主要得益于 GTGA 较好的分布均匀性和极好的覆盖宽广性.实验数据表明几何热力学选择在保持种群多样性方面是有效的.NSGA-II 也保持了较好的多样性,在 2-250 和 2-500 上的 Δ 值只是略差于 GTGA.从图 5 和图 6 可以看出 NSGA-II 的解分布非常均匀,只是在覆盖宽广性上 GTGA 更优.TDGA2 在保持多样性方面比 NSGA-II 和 GTGA 差,但明显好于 NSGA.从图 4~图 6 可看出 TDGA2 覆盖比较宽广,但分布不太均匀,这主要是由于 TDGA2 使用的基因熵是在决策空间而不是目标空间定义的.

表 5 3 个测试例上利用 Δ 度量得到的多样性比较结果

	2-100	2-250	2-500
NSGA	45.0861	160.4884	254.2272
NSGA-II	7.4559	12.4816	18.1551
TDGA2	39.2579	53.0492	76.5794
GTGA	9.1012	11.8158	12.6635

为更好地理解 NSGA-II 和 GTGA 中的种群进化特点,本文也呈现了 2-250 上两个算法在单次运行实验中当前评估次数 n_e 分别达到 2×10^3 、 1×10^4 、 4×10^4 、 2×10^5 时的种群快照,分别如图 7 和图 8 所示.根据算法原理,NSGA-II 对接近 Pareto 非劣前沿的个体施加大的选择压,GTGA 初始对较稀疏扇形区域中的个体施加大的选择压,其后随温度的下降对接近各扇形中距离精英的个体施加越来越大的选择压.图 7 显示 NSGA-II 锋面的前进能力非常强,种群初始迅速逼近最优前沿,但覆盖较窄,之后沿最优前沿向两端扩张.图 8 显示 GTGA 扩张分布广度的能力非常强,种群的逼近方向初始迅速多样化,其后在保持逼近方向多样性的基础上向最优前沿逼近.由此可以看出这两个算法代表了两种具有不同特点的搜索模式,两者的特点可分别形象地总结为“深度优先搜索”和“广度优先搜索”.在物品数较少时,搜索空间较小,此时增加非劣前沿选择压比增加逼近方向多样性更有效,故 NSGA-II 在 2-100 上效果略好;而物品数较多时,维持逼近方向多样性显得更为重要,因而 GTGA 在 2-250 和 2-500 上表现略好.

图 9 给出了 4 个算法在 3 个测试例上各 10 次运行中平均运行时间(以 s 为单位)的对比直方图.从图 9 可看出其中 GTGA 具有最高的运行效率,在 3 个测试例上的运行时间都明显小于其它算法.因为其它算法在每一代的适应值(个体能量)指派过程中都需要进行耗时的种群非劣分层操作,而在 GTGA 中计算个体的距离能量时个体只需与其所在扇形区域的距离精英比较,巧妙地避免了非劣分

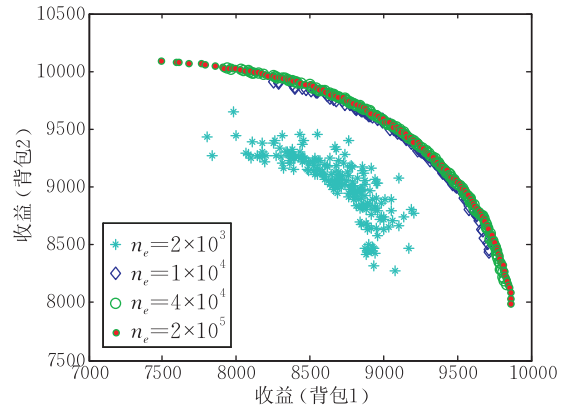


图 7 2-250 上 NSGA-II 的种群快照

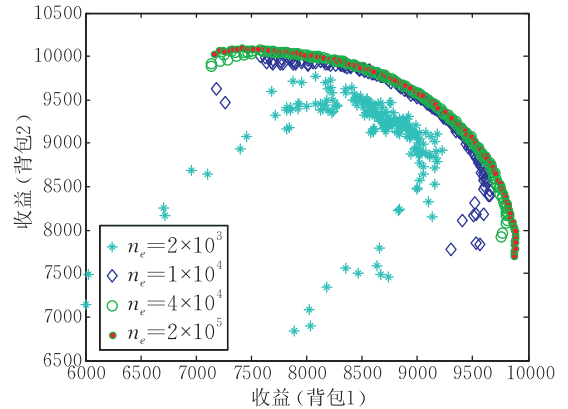


图 8 2-250 上 GTGA 的种群快照

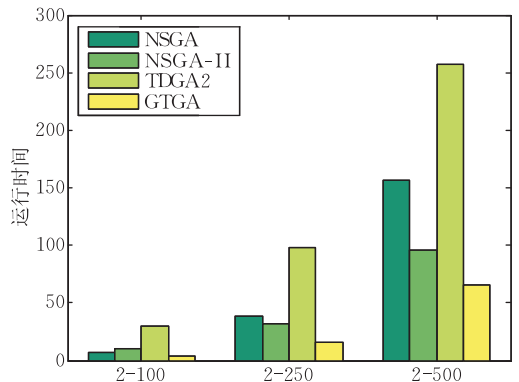


图 9 3 测试例上各算法的运行时间对比直方图

层操作.图 9 显示了 TDGA2 的计算成本最大,在每个测试例上都至少 4 倍于 GTGA.

为进一步对比分析多目标优化中熵度量、能量度量、替换规则以及种群规模对热力学遗传算法运行效率的影响,本文也在 2-250 上以 8 个种群规模运行 TDGA、TDGA2、TDGA3、GTGA 各 10 次.8 个种群规模依次为 $n_p = 100, 200, 300, 400, 500, 600, 700, 800$,TDGA 中子种群规模按文献[9]设为 $n_o = n_p$,其它 3 个算法 $n_o = 1/10 \times n_p$,所有算法终止条件 *terminate* 均在评估次数达到 $N_e = 2 \times 10^5$ 时满足.图 10 呈现了种群规模变化时 4 种热力学遗传算

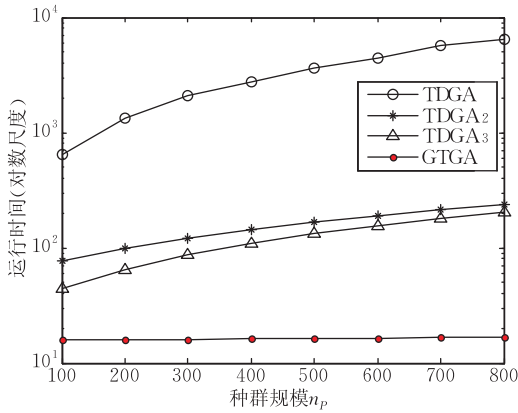


图 10 4 种热力学遗传算法的运行时间曲线

法 10 次运行的平均时间曲线. 从图中可看出 TDGA 在各种群规模下都具有最高的计算代价, 并且种群增大时其运行时间急剧增加. TDGA2 采用了分量替换规则, 其运行时间相对于 TDGA 有很大程度减少, 但其运行时间随种群规模增大有明显增长. TDGA3 采用了角度熵, 其运行时间相对于 TDGA2 有些微下降. GTGA 进一步采用距离能量避免了非劣分层操作, 从图中可知 GTGA 具有最低的计算开销, 且对种群规模较不敏感, 种群增大时计算成本只有些微增加. 这与 TDGA 极高的计算代价形成了鲜明的对比, 例如, 种群规模 $n_p = 100$ 时, TDGA 的运行时间约 8 倍于 TDGA2, 40 倍于 GTGA; 种群规模 $n_p = 800$ 时, TDGA 的运行时间约 25 倍于 TDGA2, 400 倍于 GTGA, TDGA 如此高的计算代价在实际应用中是难以承受的, 这主要是因为 TDGA 采用了耗时的非劣分层操作和贪婪替换规则. 实验结果显示分量替换规则极大地降低了多目标优化中热力学遗传算法的计算成本, 距离能量则进一步明显改进了热力学遗传算法的运行效率. 这与第 5 节中的时间复杂度分析结果是一致的.

7 总 结

针对多目标优化中 TDGA 目前计算代价太高和解集分布不够均匀的问题, 本文从几何的观点引入了几何热力学选择方法以改善 TDGA 的运行效率和解集分布. 几何热力学选择首先充分利用多目标优化中的角度和距离等几何量定义了角度熵和距离能量. 其中角度熵通过在目标空间中扇形采样来度量种群逼近方向的多样性, 提高了解集的分布均匀性; 距离能量运用几何元素间接度量了种群的逼近程度, 巧妙地避免了耗时的非劣分层操作. 此外, 引入几何分量热力学替换规则, 通过将种群的几何自由能分派到其各个体有效降低了替换规则的复杂

度. 在多目标 0/1 背包问题上的实验结果表明, 几何热力学选择在保持 TDGA 良好的覆盖宽广性的同时, 极大地改进了其运行效率和分布均匀性. 运用该选择的 GTGA 算法能够得到与 NSGA-II 性能相当的解集, 但在运行效率上明显优于 NSGA-II.

随着目标数的增多距离精英文档将急剧增长, 几何热力学选择在空间复杂度方面面临更大的挑战, 因而在 3 个及以上目标时如何存储距离精英文档值得进一步研究. 此外, 几何热力学遗传算法的收敛性分析也是我们下一步的研究方向.

参 考 文 献

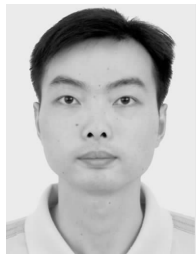
- [1] Zhou Yu-Ren, Min Hua-Qing, Xu Xiao-Yuan, Li Yuan-Xiang. A multi-objective evolutionary algorithm and its convergence. Chinese Journal of Computers, 2004, 27(10): 1415-1521(in Chinese)
(周育人, 闵华清, 许孝元, 李元香. 多目标演化算法的收敛性研究. 计算机学报, 2004, 27(10): 1415-1521)
- [2] Coello Coello C A, Lamont G B, Van Veldhuizen D A. Evolutionary Algorithms for Solving Multi-Objective Problems. Second Edition. Berlin: Springer-Verlag, 2007
- [3] Van Veldhuizen D A, Lamont G B. Multiobjective evolutionary algorithms: Analyzing the state-of-the-art. Evolutionary Computation, 2000, 8(2): 125-147
- [4] Zheng Bo-Jin. Researches on evolutionary optimization [Ph. D. dissertation]. Wuhan University, Wuhan, 2006(in Chinese)
(郑波进. 演化优化方法研究[博士学位论文]. 武汉: 武汉大学, 2006)
- [5] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 2002, 6(2): 182-197
- [6] Fonseca C M, Fleming P J. Multiobjective optimization and multiple constraint handling with evolutionary algorithms — Part I: A unified formulation. IEEE Transactions on Systems, Man, and Cybernetics — Part A: Systems and Humans, 1998, 28(1): 26-37
- [7] Srinivas N, Deb K. Multiobjective optimization using non-dominated sorting in genetic algorithms. Evolutionary Computation, 1994, 2(3): 221-248
- [8] Zitzler E, Thiele L. Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach. IEEE Transactions on Evolutionary Computation, 1999, 3(4): 257-271
- [9] Kita H, Yabumoto Y, Mori N, Nishikawa Y. Multi-objective optimization by means of the thermodynamical genetic algorithm//Proceedings of the International Conference on Parallel Problem Solving from Nature. Berlin, Germany, 1996: 504-512
- [10] Mori N, Yoshida J, Tamaki H, Kita H, Nishikawa Y. A thermodynamical selection rule for the genetic algorithm//Proceedings of the IEEE Conference on Evolutionary Computation. Perth, Australia, 1995: 188-192
- [11] Kang Li-Shan. Non-Numerical Parallel Algorithms (1):

Simulation Annealing Algorithms. Beijing: Science Press, 1997 (in Chinese)

(康立山. 非数值并行算法(第1册): 模拟退火算法. 北京: 科学出版社, 1997)

- [12] Jin Yao-Chu, Branke J. Evolutionary optimization in uncertain environments; A survey. *IEEE Transactions on Evolutionary Computation*, 2005, 9(3): 303-317
- [13] Knowles J, Corne D. Approximating the nondominated front using the Pareto archived evolution strategy. *Evolutionary Computation*, 2000, 8(2): 149-172
- [14] Knowles J, Corne D. Properties of an adaptive archiving algorithm for storing nondominated vectors. *IEEE Transactions on Evolutionary Computation*, 2003, 7(2): 100-116
- [15] Ding Wei, Hu Ting, Zhang Huan-Guo. Multi-objective optimization by a new dynamical evolutionary algorithm based on the information entropy//*Proceedings of the International Conference on Neural Networks and Brain*. Beijing, China, 2005: 50-53
- [16] Zheng Bo-Jin, Hu Ting. A novel multi-objective evolutionary algorithm//*Proceedings of the International Conference on Computational Science*. Beijing, China, 2007: 1029-1036

- [17] Ying Wei-Qin, Li Yuan-Xiang, Peng Shu-Juan, Wang Wei-Wu. A steep thermodynamical selection rule for evolutionary algorithms//*Proceedings of the International Conference on Computational Science*. Beijing, China, 2007: 997-1004
- [18] Ying Wei-Qin, Li Yuan-Xiang, Sheu P C-Y. Improving the computational efficiency of thermodynamical genetic algorithms. *Journal of Software*, 2008, 19(7): 1613-1622 (in Chinese)
(应伟勤, 李元香, Sheu P C-Y. 热力学遗传算法计算效率的改进. *软件学报*, 2008, 19(7): 1613-1622)
- [19] Zitzler E, Deb K, Thiele L. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 2000, 8(2): 173-195
- [20] Hu Xiao-Lin, Coello Coello C A, Huang Zhang-Can. A new multi-objective evolutionary algorithm: Neighborhood exploring evolution strategy. *Engineering Optimization*, 2005, 34(4): 351-379
- [21] Ishibuchi H, Kaige S. Implementation of simple multiobjective memetic algorithms and its application to knapsack problems. *International Journal of Hybrid Intelligent Systems*, 2004, 1(1-2): 22-35



YING Wei-Qin, born in 1979, Ph. D. candidate. His current research interests include intelligent computation, semantic computing, image processing and pattern recognition.

LI Yuan-Xiang, born in 1962, professor, Ph. D. supervisor. His current research interests include intelligent computation and parallel computing.

SHEU Phillip C-Y, born in 1956, professor, Ph. D. supervisor. His current research interests include semantic computing, database systems, software engineering, and complex biomedical systems.

WU Yu, born in 1982, Ph. D. candidate. Her current research interests focus on intelligent computation.

YU Fa-Hong, born in 1977, Ph. D. candidate. His current research interests include intelligent computation and pattern recognition.

Background

Multi-objective optimization problems (MOPs) are a very important research topic, not only because of the multi-objective nature of most real-world decision problems, but also because there are still many open questions in this area. The aim of multi-objective optimization is to simultaneously optimize a group of conflicting objectives. In general, there is no single optimal solution, but rather a set of Pareto-optimal solutions in MOPs as opposed to single-objective optimization problems. No other solutions in the search space are superior to any of the Pareto-optimal solutions when all objectives are simultaneously considered.

Evolutionary algorithms seem particularly suitable to solve MOPs because they could find simultaneously a set of Pareto-optimal solutions in a single "run". In order to provide flexibility for the human decision maker, evolutionary multi-objective optimization has two tasks: (i) convergence to the Pareto-optimal set, and (ii) maintenance of diversity in solutions of the Pareto-optimal set. Thermodynamical genetic algorithms (TDGAs) simulate the minimization of free ener-

gy in thermodynamics to deal simultaneously with both tasks in multi-objective optimization. However, the computational cost is very high and the distribution of solutions is not uniform enough in the current implementation of TDGA.

In order to improve the running efficiency and the distribution uniformity of solutions, the research in this paper adopts geometric quantities such as angle and distance to define the free energy and redesigns the thermodynamical replacement rule to reduce the complexity of replacement in TDGA for multi-objective optimization. The goal of the research is to incorporate biology, thermodynamics and geometry to produce a perfect convergence and spread of solutions of MOPs with low computational costs. The research is supported by the National Natural Science Foundation of China under grant No. 60773009, the National High Technology Research and Development Program (863 Program) of China under grant No. 2007AA01Z290, and the China Scholarship Council under grant No. 2007101731.