

## 一种评估近似 Pareto 前沿多样性的方法<sup>\*</sup>

曾三友<sup>1+</sup>, 蔡振华<sup>1</sup>, 张青<sup>1</sup>, 康立山<sup>1,2</sup>

<sup>1</sup>(中国地质大学(武汉)计算机学院 空间科学技术研究中心 地质过程与矿产资源国家重点实验室,湖北 武汉 430074)

<sup>2</sup>(武汉大学 软件工程国家重点实验室,湖北 武汉 430072)

### A Novel Method for Assessing the Diversity of Approximation Pareto Fronts

ZENG San-You<sup>1+</sup>, CAI Zhen-Hua<sup>1</sup>, ZHANG Qing<sup>1</sup>, KANG Li-Shan<sup>1,2</sup>

<sup>1</sup>(State Key Laboratory of Geological Processes and Mineral Resources, School of Computer Science & Research Center for Space Science and Technology, China University of GeoSciences (Wuhan), Wuhan 430074, China)

<sup>2</sup>(State Key Laboratory of Software Engineering, Wuhan University, Wuhan 430072, China)

+ Corresponding author: E-mail: sanyou-zeng@263.net

**Zeng SY, Cai ZH, Zhang Q, Kang LS. A novel method for assessing the diversity of approximation Pareto fronts. Journal of Software, 2008,19(6):1301–1308. <http://www.jos.org.cn/1000-9825/19/1301.htm>**

**Abstract:** This paper introduces a new technique for assessing the diversity of approximation of exact set. The diversity is assessed by the “exposure degree” of the exact set against the approximation set in the new technique where the projection of the approximation set on the exact set, not the approximation set, is used to cover the exact set. Therefore, the new technique still works well for the problems with 3 and more objectives, significantly improving the existing techniques. Computational results show that it assesses the diversity well. Rigorous analysis is made for the new method.

**Key words:** set assessment; multiobjective optimization; Pareto optimal set; Pareto optimal front; approximation set

**摘要:** 提出了一种评估近似 Pareto 前沿多样性的方法.它用精确 Pareto 前沿相对于其近似集合的“暴露程度”来评估多样性,其计算基于精确 Pareto 前沿,不像其他多样性评估技术基于近似集合进行计算.对于多于两个目标(分量)的近似集合,其他技术不能或者不能很好地评估,计算结果表明,该技术能够很好地评估.并且,在理论上对该方法进行了严格的论证.

**关键词:** 集合评估;多目标优化;Pareto 解集;Pareto 前沿;近似集合

**中图法分类号:** TP301      **文献标识码:** A

与单目标问题不同,多目标优化问题的解是一个集合,称为 Pareto 解集.Pareto 解集在目标空间的映像称为

---

\* Supported by the National Natural Science Foundation of China under Grant Nos.60473037, 60483081, 40275034, 60204001, 60133010 (国家自然科学基金); the China Postdoctoral Science Foundation under Grant No.2003034505 (中国博士后科学基金) the open research program of the Geological Processes and Mineral Resources (GPMR), China University of Geosciences under Grant No.GPMR200618 (中国地质大学地质过程与矿产资源国家重点实验室开放课题)

Received 2006-12-03; Accepted 2007-05-10

Pareto 最优前沿.由于经典优化方法很难解决非线性的带约束条件的多目标优化问题,新近形成和发展起来的演化计算方法被引入,用来求解此类复杂问题.演化计算(EC)模拟自然和物理原则,并且使用群体搜索策略,使该方法具备求解约束非线性多目标优化问题的潜在优势.自 20 世纪 80 年代后期以来,掀起了演化多目标优化研究热潮,至今经久不息.不仅在演化计算国际会议(Congress on Evolutionary Computation)和遗传与演化计算国际会议(Conference on Genetic and Evolutionary Computation)两大国际会议中,演化多目标优化是最热门的话题,而且自 2001 年开始,隔年还单独召开演化多目标演化国际会议(International Conference on Evolutionary Multi-Criterion Optimization).

算法性能的比较是一个重要的话题,由于多目标优化算法输出的是集合,因而评估多目标优化算法的一个重要方面是近似前沿逼近精确 Pareto 前沿的逼近程度.在多目标演化优化领域里,大多是研究目标空间中近似前沿的评估问题,对于本文来说,近似集合既可以看成是决策空间的近似 Pareto 解集,又可以看成是目标空间的近似 Pareto 前沿.近似集合的评估问题比较困难,一般的观点是,一个好的近似解集应该满足以下条件:① 最短逼近(minimum distance)精确解集;② 在精确解集上分布均匀和分布最广(uniform distribution and maximum spread).最初只是定性评估,正如 Van Veldhuizen 所说<sup>[1]</sup>,不同演化多目标优化算法的解,仅仅通过图示近似 Pareto 前沿来进行直观比较.后来出现了比较多的定量评估方法:第 1 类方法是从不同的方面量化一个近似集合的质量,然后将各方面的量化指标值组合起来作为近似集合逼近程度评估值,例如文献[2-5];第 2 类方法是二元质量评估方法,定义一个量化函数,比较两个近似集合的质量,例如文献[6,7];第 3 类方法定义一个到达函数,通过到达函数评估近似集合到任何一个目标的可能性,例如文献[8].第 1 类方法用得最多,一般是用两个指标评估:一个指标是近似集合到精确集合的距离,例如 Van Veldhuizen 的文献[1],计算近似集合每一个点到精确集合最近的点距离,然后对所有这些距离取平均;另一个指标是近似集合在精确集合上的分布多样性,例如 Deb 的文献[2],用类似于卡方偏差估计.每一个评估方法都存在一些缺陷,并且已有的评估方法大多没有从理论上论证其定义合理性和计算可行性,特别是多样性评估非常困难,大多数人用 Deb 的文献[2]提供的方法评估.但是,此方法评估两个目标问题比较有效,对于 3 个或 3 个以上目标的问题,它的评估结果可信度令人怀疑,作者尚未看到使用 Deb 提出的多样性评估方法评估两个以上目标的近似 Pareto 前沿.

本文研究上述第 1 类评估方法中的多样性评估技术,提出了一项分布多样性评估技术,它能够评估超过两个目标的近似集合的多样性,并进行严格理论分析和例子验证.

下面首先通过例子说明目前用得较多的一种多样性评估方法中存在的问题;然后提出一种新的多样性评估方法,并通过理论分析和例子验证两种手段论证其定义的合理性和计算可行性;最后总结全文.

## 1 基本思路

前面提到,评估近似集合逼近精确集合的质量一般是用两个指标去评估:一是近似集合到精确集合的逼近距离,二是近似集合在精确集合上的分布多样性.作者认为,逼近距离和分布多样性两个指标应该尽可能地相互独立,逼近距离指标中要避免包含分布多样性信息,分布多样性指标中要避免包含逼近距离信息,即两个指标要尽可能不相关.下面我们举一个极端情形的简单例子,用两个常用的方法评估近似集合的这两个指标,并分析这两个指标之间的相关性.如图 1 所示,假设 Pareto 前沿  $Q$  为线段  $ab$ ,多目标优化算法(例如多目标演化算法)求得的近似前沿为集合  $P=\{p_1, p_2, \dots, p_{10}\}$ ,  $p'_1, p'_2, \dots, p'_{10}$  分别是  $p_1, p_2, \dots, p_{10}$  到线段  $ab$  上的垂足,或者叫垂直投影.用 Van Veldhuizen 的方法<sup>[1]</sup>计算近似集合的逼近距离就是所有近似前沿的点到精确 Pareto 前沿的平均距离.本例为  $p_1, p_2, \dots, p_{10}$  到线段  $ab$  的垂线段长度的平均值,即

$$T = \frac{1}{10} \sum_{i=1}^{10} p_i p'_i \quad (1)$$

$T$  越小,近似前沿的精度越高.式(1)中计算的逼近距离  $T$  中不包含近似集合在精确集合上的分布多样性信息.对于评估近似前沿在 Pareto 前沿上的分布多样性,研究人员大多使用 Deb 的评估方法<sup>[2]</sup>,图 1 的情形计算如下:

$$\Delta = \frac{1}{2+9}(\overline{ap_1} + (\overline{p_1p_2} - m) + (\overline{p_2p_3} - m) + \dots + (\overline{p_8p_9} - m) + (\overline{p_9p_{10}} - m) + \overline{p_{10}b}) \quad (2)$$

$$m = \frac{1}{9}(\overline{p_1p_2} + \overline{p_2p_3} + \dots + \overline{p_8p_9} + \overline{p_9p_{10}}) \quad (3)$$

$(\overline{p_1p_2} - m) + (\overline{p_2p_3} - m) + \dots + (\overline{p_8p_9} - m) + (\overline{p_9p_{10}} - m)$  越小,近似前沿分布越均匀; $\overline{ap_1} + \overline{p_{10}b}$  越小,近似前沿分布越广.因此在式(2)中, $\Delta$ 越小,近似前沿分布越均匀越广,从而多样性越好.

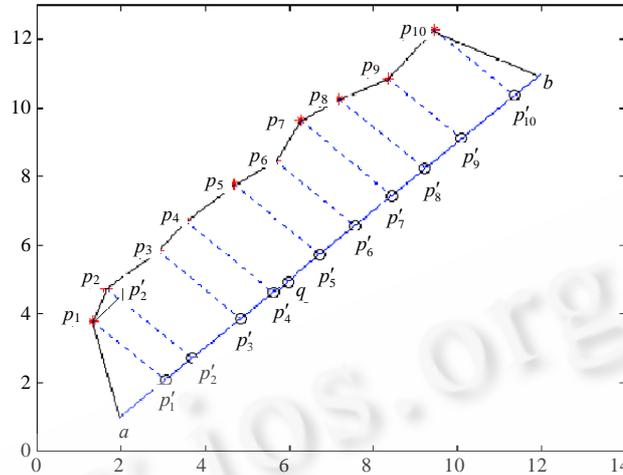


Fig.1 An example where the exact Pareto front  $Q$  is the line segment  $ab$  and the approximation Pareto fronts is  $P=\{p_1,p_2,\dots,p_{10}\}$

图 1 一个近似前沿和真实 Pareto 前沿例子:近似前沿  $P=\{p_1,p_2,\dots,p_{10}\}$ ,Pareto 前沿  $Q$ =线段  $ab$

但是,式(2)分布多样性指标值 $\Delta$ 中含有逼近距离信息.例如, $\overline{ap_1}$  和  $\overline{p_1p_2}$  出现在式(2)中,他们分别对应向量  $\overline{ap_1}$  和  $\overline{p_1p_2}$ .把它们沿精确 Pareto 前沿的垂直方向(距离逼近方向)和切线(如果是 3 个目标,则是平面;如果多于 3 个目标,则是超平面)分解,有

$$\overline{ap_1} = \overline{ap'_1} + \overline{p'_1p_1}, \quad \overline{p_1p_2} = \overline{p_1p'_2} + \overline{p'_2p_2}.$$

显然,  $\overline{p'_1p_1}$ ,  $\overline{p'_2p_2}$  属于距离逼近信息,应该要剔除.因此,可改进式(2)得到计算公式:

$$\Delta' = \frac{1}{2+9}(\overline{ap'_1} + (\overline{p'_1p'_2} - m') + (\overline{p'_2p'_3} - m') + \dots + (\overline{p'_8p'_9} - m') + (\overline{p'_9p'_{10}} - m') + \overline{p'_{10}b}),$$

$$m' = \frac{1}{9}(\overline{p'_1p'_2} + \overline{p'_2p'_3} + \dots + \overline{p'_8p'_9} + \overline{p'_9p'_{10}}).$$

但是,此计算方法适用于两个分量(目标)的问题,对于 3 个或 3 个分量以上的问题不适用.对应 3 个分量的问题,研究者往往通过三维示图来直观化多样性,这在一定程度上勉强地、定性地刻画多样性;对于 3 个分量以上的问题,人们就无能为力了.本文的主要工作是提出一种多样性评估方法,它能够评估任意多个分量的近似集合的多样性.

本文评估近似集合逼近精确集合的质量的基本思想是:第 1 步,将近似集合  $P$  投影到精确集合  $Q$ ,得到投影集合  $P'$ ,用 Van Veldhuizen 的方法<sup>[1]</sup>计算近似集合的逼近距离;第 2 步评估投影集合  $P'$  在  $Q$  上的分布性能,并作为近似集合  $P$  在  $Q$  上的分布性能.这样一来,就最大限度地避免了两个指标信息互相包含.如何评估投影集合  $P'$  在  $Q$  上的分布性能呢?我们用图 1 所给的例子来说明:近似前沿  $P=\{p_1,p_2,\dots,p_{10}\}$  在 Pareto 前沿线段  $ab$  上的投影集合为  $P'=\{p'_1,p'_2,\dots,p'_{10}\}$ ,可以解释成线段  $ab$  上的点  $p'_1,p'_2,\dots,p'_{10}$  被近似前沿覆盖,线段  $ab$  上的其他点不同程度地暴露.每个暴露点  $q$  的暴露程度(记为  $D(q,P')$ )可以用与它最近的投影点的距离来表示.

$$D(q, P') = \inf\{\overline{qp'_1}, \overline{qp'_2}, \overline{qp'_3}, \dots, \overline{qp'_{10}}\} \quad (4)$$

对于图 1 中给出的点  $q$  位置,直接观察可知  $D(q, P') = \overline{qp'_4}$ . Pareto 前沿线段  $ab$  上所有点的暴露程度的平均值定义为相对于近似前沿  $P$  而言 Pareto 前沿线段  $ab$  的暴露程度.

$$A = \frac{\int_{ab} D(q, P') dq}{ab} \quad (5)$$

相对于精确 Pareto 前沿而言,给定规模的近似前沿  $P$  的分布越均匀、越广,它对精确 Pareto 前沿的覆盖性越好,也就是精确 Pareto 前沿暴露程度越小,即  $A$  越小.因此,  $A$  可以用来评估演化算法的近似前沿  $P$  的分布多样性性能.  $A$  越小,近似前沿的分布越均匀、越广.

把研究的问题稍微抽象,下面主要研究欧氏空间中集合  $P$  逼近集合  $Q$  的精确程度评估问题.把集合  $P$  看成算法输出的近似解集(前沿),集合  $Q$  看成 Pareto 解集(前沿),研究的结果就可以用来评估近似解集(前沿)逼近 Pareto 解集(前沿)的距离逼近程度和近似解集(前沿)在 Pareto 解集(前沿)上的分布多样性性能.

## 2 相关定义

**定义 1.** 设集合  $Q \subset R^n$ , 点  $p \in R^n$ ,  $R^n$  是  $n$  维欧氏空间,点  $p$  到集合  $Q$  的距离为

$$D(p, Q) = \inf\{D(p, q) | q \in Q\} \quad (6)$$

这里,  $D(p, q)$  或者  $\overline{pq}$  表示点  $p$  与点  $q$  之间的欧氏距离.

集合  $\inf\{D(p, q) | q \in Q\}$  有下界,因此必有下确界.因此,定义 1 是有意义的.在  $D(p, q)$  有意义的前提下,才能够计算  $D(p, q)$  的值,通常  $D(p, q)$  是一个极限值,并且一般只能通过有限计算,得到  $D(p, q)$  的近似值;否则,如果  $D(p, q)$  没有意义,则计算  $D(p, q)$  的值就失去意义,计算  $D(p, q)$  的方法也就不稳定.

**定义 2.** 对于给定点  $p \in R^n$ ,若存在  $p' \in Q, Q \subset R^n$  且使  $D(p, p') = D(p, Q)$ ,则称点  $p'$  是点  $p$  在集合  $Q$  上的一个投影.

一般来说,点的投影可能不存在,也可能存在;如果投影存在,它还可能不唯一.例如,在图 2(a)中,  $Q$  为除掉  $b$  点的线段  $ab$ ,则点  $p$  在  $Q$  上没有投影;在图 2(b)中,  $Q$  为圆周,则  $Q$  中任何点都是圆心  $p$  的投影.

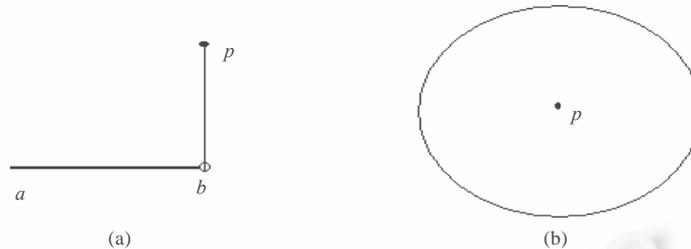


Fig.2 Illustration for the existence of projection of a point

图 2 点的投影存在性示意图

下面我们定义集合到集合的投影.

**定义 3.** 设集合  $P \subset R^n$ , 集合  $Q \subset R^n$ ,  $P$  到  $Q$  的投影定义为  $P$  中任何点在  $Q$  上投影的全体,即

$$P' = \{p' | p' \text{ 是 } p \text{ 在 } Q \text{ 上的投影}, p \in P\} \quad (7)$$

由定义 3 可知,  $P'$  是  $Q$  的子集,即  $P' \subset Q$ .

下面定义  $Q$  对  $P$  的暴露程度.

**定义 4.** 设  $P, Q$  是  $R^n$  中两个集合,  $P$  的每一点在  $Q$  上有投影,  $P'$  是  $P$  在  $Q$  上的投影集合,则  $Q$  对  $P$  的暴露程度定义为:

(1) 当  $Q$  的元素个数有限时,  $Q = \{q_1, q_2, \dots, q_N\}$ ,

$$A = \frac{\sum_{i=1}^N D(q_i, P')}{N} \quad (8)$$

(2) 当  $Q$  的 Lebesgue 测度  $|Q| \neq 0$  时, 假设  $D(q, P')$  在  $Q$  上的积分  $\int_Q D(q, P') dq$  存在

$$A = \frac{\int_Q D(q, P') dq}{|Q|} \quad (9)$$

$Q$  对  $P$  的暴露程度  $A$  越小,  $P$  在  $Q$  上的分布性能越好. 一般情形下有  $|Q| \neq 0$ , 即定义 4 中情形(2). 在这种情形下, 计算  $A$  的精确值(式(9))比较困难或者不可能, 通常利用积分定义和性质, 将积分用有限和来近似表示, 转而计算  $A$  的近似值. 易知,  $A$  近似值的计算公式与式(8)有相同的形式.

定义 4 没有考虑  $Q$  的第 3 种情形:  $Q$  是无穷集合, 但  $Q$  的 Lebesgue 测度  $|Q|=0$ , 例如,  $Q = \{q_1(1,0), q_2(1/2,0), q_3(1/3,0), \dots, q_n(1/n,0), \dots, q_0(0,0)\}$ . 这里有一个挑战性的问题, 即在第 3 种情形下,  $Q$  对  $P$  的暴露程度能不能定义.

$P$  在  $Q$  上的投影集合  $P'$  的存在性以及  $\int_Q D(q, P') dq$  的存在性是定义 4 依赖的前提条件. 如果  $P'$  不存在, 或者  $\int_Q D(q, P') dq$  不存在, 则定义 4 中计算  $A$  就失去意义, 相应地, 计算方法也就不稳定. 存在性的相关讨论见附件, 其中一些相关定理的论证请参考文献[9].

$P$  当成算法输出的近似前沿,  $Q$  当成真实 Pareto 前沿, 定义 4 就可以评估近似前沿在真实 Pareto 前沿上的分布多样性能.

### 3 例子

利用  $Q$  对  $P$  的“暴露程度”来评估  $P$  在  $Q$  上的分布多样性能(即定义 4)具有一定的创新性. 下面通过一些例子验证新的分布多样性能评估技术的有效性.

#### 3.1 示例

例 1: 设  $Q = \{q_1, q_2, \dots, q_{10}\} = \{(1,0), (2,0), (3,0), (4,0), (5,0), (6,0), (7,0), (8,0), (9,0), (10,0)\}$ ,

(1)  $P = \{p_1, p_2, p_3\} = \{(2,2), (5,1), (8,3)\}$ ,

(2)  $P = \{p_1, p_2, p_3\} = \{(2,2), (4,1), (8,3)\}$ ,

(3)  $P = \{p_1, p_2, p_3\} = \{(4,2), (5,2), (6,2)\}$ ,

(4)  $P = \{p_1, p_2, p_3\} = \{(2,3), (5,1), (8,4)\}$ ,

分别计算  $Q$  对  $P$  的暴露程度.

解:(1)  $P$  在  $Q$  上的投影集合  $P' = \{(2,0), (5,0), (8,0)\}$ ,

$$D(q_1, P') = 1, D(q_2, P') = 0, D(q_3, P') = 1, D(q_4, P') = 1, D(q_5, P') = 0, \\ D(q_6, P') = 1, D(q_7, P') = 1, D(q_8, P') = 0, D(q_9, P') = 1, D(q_{10}, P') = 2,$$

$$A = \frac{\sum_{i=1}^N D(q_i, P')}{N} = \frac{\sum_{i=1}^{10} D(q_i, P')}{10} = 8/10 = 0.8.$$

(2)  $P$  在  $Q$  上的投影集合  $P' = \{(2,0), (4,0), (8,0)\}$ ,

$$D(q_1, P') = 1, D(q_2, P') = 0, D(q_3, P') = 1, D(q_4, P') = 0, D(q_5, P') = 1, \\ D(q_6, P') = 2, D(q_7, P') = 1, D(q_8, P') = 0, D(q_9, P') = 1, D(q_{10}, P') = 2,$$

$$A = 0.9.$$

(3)  $P$  在  $Q$  上的投影集合  $P' = \{(4,0), (5,0), (6,0)\}$ ,

$$D(q_1, P') = 3, D(q_2, P') = 2, D(q_3, P') = 1, D(q_4, P') = 0, D(q_5, P') = 0, \\ D(q_6, P') = 0, D(q_7, P') = 1, D(q_8, P') = 2, D(q_9, P') = 3, D(q_{10}, P') = 4,$$

$$A = 1.6.$$

(4)  $P$  在  $Q$  上的投影集合  $P' = \{(2,0), (5,0), (8,0)\}$ , 与情形(1)有相同的投影集合, 因此与情形(1)有相同暴露程度,

$$A = 0.8.$$

在情形(1)、情形(2)中,相对于集合  $Q$  来说,投影集合  $P'$  分布广度相同,情形(1)中, $P'$  本身分布均匀,情形(2)中, $P'$  本身分布不均匀,计算结果:情形(1)的暴露程度小,情形(2)的暴露程度大.

在情形(1)、情形(3)中,投影集合  $P'$  本身分布均匀,相对于集合  $Q$  来说,情形(1)中  $P'$  分布范围广,情形(3)中  $P'$  分布范围窄,计算结果:情形(1)的暴露程度小,情形(3)的暴露程度大.

例 2: 设  $Q$  是  $x$  轴上区间  $[1, 10]$ ,

(1)  $P = \{p_1, p_2, p_3\} = \{(2, 2), (5, 1), (8, 3)\}$ ,

(2)  $P = \{p_1, p_2, p_3\} = \{(2, 2), (4, 1), (8, 3)\}$ ,

分别计算  $Q$  对  $P$  的暴露程度.

解: 在本例中,集合  $Q$  是测度非零集合.

(1)  $P$  在  $Q$  上的投影集合  $P' = \{(2, 0), (5, 0), (8, 0)\}$ ,

$$\begin{aligned} A &= \frac{\int_Q D(x, P') dx}{|Q|} = \int_1^2 (2-x) dx + \int_2^{3.5} (x-2) dx + \\ &\int_{3.5}^5 (5-x) dx + \int_5^{6.5} (x-5) dx + \int_{6.5}^8 (8-x) dx + \int_8^{10} (x-8) dx \\ &= 0.778. \end{aligned}$$

(2)  $P$  在  $Q$  上的投影集合  $P' = \{(2, 0), (4, 0), (8, 0)\}$ ,

$$A = 0.833.$$

在两种情形中,投影集合  $P'$  分布广度相同,情形(1)中  $P'$  本身分布均匀,情形(2)中  $P'$  本身分布不均匀,计算结果:情形(1)的暴露程度小,情形(2)的暴露程度大.

例 1、例 2 是两个简单的例子,从直观上能够比较得出不同的情形中近似集合分布多样性的优劣.可以看出,用新的多样性评估方法计算出来的结果与直观比较的结果一致.

### 3.2 实验比较

下面用两种多样性评估方法(Deb 的  $\Delta$ <sup>[2]</sup>和本文的“暴露程度” $A$ )来评估同一个多目标演化算法计算 10 个测试问题所得的近似 Pareto 前沿的分布多样性.

例 3: 我们使用文献[10]提出的正交多目标演化算法(OMOE A-II)来计算 10 个测试问题  $ZDT_1, ZDT_2, ZDT_3, ZDT_4, DTLZ_1, DTLZ_2, DTLZ_3, DTLZ_4, DTLZ_5, DTLZ_6$ . 前 4 个问题来源于文献[11], 都是两个目标的问题; 后 6 个问题来源于文献[12], 都是 3 个目标的问题. OMOEA-II 计算 10 个测试问题各 50 次, 分别用 Deb 的  $\Delta$  和本文的“暴露程度” $A$  来评估 OMOEA-II 计算出来的 Pareto 前沿.  $\Delta$  和  $A$  越小, 近似 Pareto 前沿的多样性越好. 表 1(上行为均值, 下行为均方差)给出 OMOEA-II 50 次计算结果的  $\Delta$  和  $A$  的均值和均方差. 从前 4 个问题  $ZDT_1, ZDT_2, ZDT_3, ZDT_4$  的计算结果看出,  $\Delta$  和  $A$  的值具有一致性, 这表明“暴露程度” $A$  能够评估近似 Pareto 前沿的多样性. 由于  $\Delta$  不适合评估具有两个目标的近似 Pareto 前沿的多样性, 后 6 个测试问题没有结果. 而本文的“暴露程度” $A$  仍然能够正确地评估近似 Pareto 前沿的多样性.

**Table 1** Means (upper rows) and variances of the diversity metric  $\Delta$  and  $A$  over 50 runs

表 1 OMOEA-II 计算 10 个测试问题各 50 次, 用两种方法量化所得的近似 Pareto 前沿的多样性

Assessing methods	$ZDT_1$	$ZDT_2$	$ZDT_3$	$ZDT_4$	$DTLZ_1$	$DTLZ_2$	$DTLZ_3$	$DTLZ_4$	$DTLZ_5$	$DTLZ_6$
$\Delta$	0.265 0.007 3	0.286 0.007 1	0.425 0.015	0.232 0.0064	-	-	-	-	-	-
$A$	0.004 8 0.000 2	0.005 1 0.000 2	0.005 8 0.000 4	0.004 4 0.000 2	0.022 5 0.003 8	0.061 1 0.001 1	0.062 4 0.001 3	0.063 5 0.002 3	0.005 1 0.000 2	0.005 1 0.000 3

总之, 3 个例子表明: “暴露程度” $A$  越小, 则相对于集合  $Q$  来说, 近似集合  $P$  分布范围就越广、越均匀,  $P$  的分布多样性能也就越好.

## 4 结束语

本文的主要创新是提出了一种集合多样性评估技术并作了严格的理论分析.研究结果表明,新方法能够有效地评估近似集合的分布多样性.新多样性评估技术有如下特点:

- 1) 本文评估近似集合逼近精确集合的质量的基本思想是:第 1 步,将近似集合  $P$  投影到精确集合  $Q$ ,得到投影集合  $P'$ ,用 Van Veldhuizen 的方法<sup>[1]</sup>计算近似集合的逼近距离;第 2 步评估投影集合  $P'$  在  $Q$  上的分布性能,并作为近似集合  $P$  在  $Q$  上的分布性能.这样一来,最大限度地避免了距离逼近指标信息和多样性分布指标信息互相包含.
- 2) Deb 的分布多样性评估技术是:首先,分别近似集合  $P$  在精确集合  $Q$  上的分布均匀性和广度,然后相加平均得到多样性评估值.本文的技术通过  $Q$  相对于  $P$  的“暴露程度”用来评估多样性,“暴露程度”既具有物理直观性,又能将分布均匀性和分布广度合二为一.
- 3) Deb 的多样性评估技术适用于两个分量(或目标)的情况,本文的技术对于任意多分量的问题都适用.

对“暴露程度”定义的前提条件在附件中进行了严格的理论论证.论证的结果是:当集合  $P, Q$  是有界闭集时,“暴露程度”定义的前提条件得到保证.

当集合  $P$  与集合  $Q$  有界但不是闭集时,我们建议以  $P, Q$  的闭包  $\bar{P}, \bar{Q}$  ( $\bar{P}, \bar{Q}$  一定是闭集)替代  $P, Q$ .也就是说,  $\bar{P}$  在  $\bar{Q}$  上的分布多样性能估计代表  $P$  在  $Q$  上的分布多样性能估计.

但当  $|Q| \neq 0$  时,计算 Lebesgue 积分的近似值运算量比较大.

值得进一步研究的问题包括:

- 1) 当  $Q$  是无穷集合,但  $Q$  的 Lebesgue 测度  $|Q|=0$  时,  $Q$  对  $P$  的暴露程度能不能定义.
- 2) 当  $P, Q$  是无界集合时,如何评估  $P$  逼近  $Q$ .

## References:

- [1] Van Veldhuizen DA. Multiobjective evolutionary algorithms: Classifications, analyses, and new innovations [Ph.D. Thesis]. Graduate School of Engineering of the Air Force Institute of Technology, Air University, 1999.
- [2] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. on Evolutionary Computation*, 2001,6(2):182–197.
- [3] Van Veldhuizen DA, Lamont GB. On measuring multiobjective evolutionary algorithm performance. In: Zalalza A, Eberhart R, eds. *Proc. of the Congress on Evolutionary Computation (CEC 2000)*, Vol.1. Piscataway: IEEE Press, 2000. 204–211.
- [4] Mostaghim S, Teich J. A new approach on many objective diversity measurement. In: Branke J, Deb K, Miettinen K, Steuer R, eds. *Proc. of the Practical Approaches to Multi-Objective Optimization*. 2005. 1862–4405. <http://drops.dagstuhl.de/opus/volltexte/2005/254>
- [5] Khare V, Yao X, Deb K. Performance scaling of multi-objective evolutionary algorithms. In: *Proc. of the 2nd Int'l Conf. on Evolutionary Multi-Criterion Optimization*. 2003. 376–390.
- [6] Zitzler E, Thiele L. Multiobjective optimization using evolutionary algorithms—A comparative case study. In: Eiben AE, Bäck T, Schoenauer M, Schwefel HP, eds. *Proc. of the 5th Int'l Conf. on Parallel Problem Solving from Nature (PPSN-V)*. Berlin: Springer-Verlag, 1998. 292–301.
- [7] Hansen MP, Jaskiewicz A. Evaluating the quality of approximations of the non-dominated set. Technical Report, 1998-7, Institute of Mathematical Modeling, Technical University of Denmark, 1998.
- [8] Da Fonseca VG, Fonseca CM, Hall AO. Inferential performance assessment of stochastic optimisers and the attainment function. In: Zitzler E, Deb K, Thiele L, Coello Coello CA, Corne D, eds. *Proc. of the 1st Int'l Conf. on Evolutionary Multi-Criterion Optimization (EMO 2001)*. LNCS 1993, Berlin: Springer-Verlag, 2001. 213–225.
- [9] Xia DX, Wu ZR, *et al.* *Functions of Real Variable & Functional Analysis (Volume 2)*. Beijing: Higher Education Press, 1979 (in Chinese).
- [10] Zeng SY, Yao SZ, Kang LS, Liu Y. An efficient multi-objective evolutionary algorithm: OMOEA-II. In: Coello Coello CA, *et al.*, eds. *Proc. of the 3rd Int'l Conf. on Evolutionary Multi-Criterion Optimization (EMO 2005)*. LNCS 3410, Mexico: Springer-Verlag, 2005. 108–119.
- [11] Zitzler E, Deb K, Thiele L. Comparison of multi-objective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 2000,8(2):173–195.

- [12] Deb K, Thiele L, Laumanns M, Zitzler E. Scalable test problems for evolutionary multi-objective optimization. KanGAL Report, No.2001001, 2001.

#### 附中文参考文献:

- [9] 夏道行,吴卓人,等.实变函数论与泛函分析(下册).北京:高等教育出版社,1979.

### 附录. 相关存在性的讨论

#### 1. 相关投影集合存在性讨论

一般来说,点的投影可能不存在.但是,当  $Q$  是有界闭集(紧集)时,点在  $Q$  上的投影存在.以下的定理或引理没有给出证明,读者可以利用文献[12]第 4 章度量空间的相关知识,很容易给出证明.

**定理 1.** 若  $Q$  是有界闭集,则对于给定的  $p, p$  在  $Q$  上有投影,即存在  $p' \in Q$ , 使

$$D(p, p') = D(p, Q).$$

为了讨论投影集合的特性,我们有如下引理:

**引理 1.** 设  $Q$  是有界闭集,  $\{p_n\}$  到  $Q$  的投影序列为  $\{p'_n\}$ , 即  $p_n$  到  $Q$  的投影为  $p'_n, n=1, 2, \dots$ , 如果  $\lim_{n \rightarrow \infty} p_n = p$ ,  $\lim_{n \rightarrow \infty} p'_n = p'$ , 则  $p$  到  $Q$  的投影为  $p'$ .

根据定理 1, 当  $Q$  是有界闭集时, 任意给定的点  $p$  在  $Q$  上有投影. 因此, 任意给定的集合  $P$  在  $Q$  上存在投影集合  $P'$ , 并且集合  $P$  中任意点在  $Q$  上有投影. 当集合  $P$  是有界闭集时, 我们还有如下定理:

**定理 2.** 设集合  $P, Q$  是有界闭集, 则  $P$  到  $Q$  的投影  $P'$  也是有界闭集.

#### 2. 相关 Lebesgue 积分存在性的讨论

**引理 2.** 设  $B$  是有界闭集,  $\{q_n\}$  到  $B$  的一投影序列为  $\{q'_n\}$ , 即  $q_n$  到  $B$  的一投影为  $q'_n, n=1, 2, \dots$ , 如果  $\lim_{n \rightarrow \infty} q_n = q$ ,  $\lim_{n \rightarrow \infty} q'_n = q'$ , 则  $\lim_{n \rightarrow \infty} D(q_n, B) = D(q, B)$ .

**定理 3.** 设集合  $B, Q$  是有界闭集,  $q \in Q$ , 则  $D(q, B)$  作为以  $q$  为变量的函数在  $Q$  上是连续的. 从而,  $D(q, B)$  在  $Q$  上 Lebesgue 可积.

下面讨论定义 4 的前提条件.

(1) 当集合  $P, Q$  是有界闭集时, 根据定理 2,  $P$  到  $Q$  的投影  $P'$  存在, 而且  $P'$  是有界闭集. 根据定理 3 可知,  $D(q, P')$  是  $Q$  上的连续函数, 从而  $D(q, P')$  在  $Q$  上 Lebesgue 可积, 保证了定义 4 的前提条件. 于是,  $Q$  对  $P$  的暴露程度可以计算或者说  $P$  在  $Q$  上的分布多样性能可以评估.

(2) 当集合  $P$  与集合  $Q$  有界但不是闭集时, 作者建议以  $P, Q$  的闭包  $\bar{P}, \bar{Q}$  ( $\bar{P}, \bar{Q}$  一定是闭集) 替代  $P, Q$ . 也就是说,  $\bar{P}$  在  $\bar{Q}$  上的分布多样性能估计代表  $P$  在  $Q$  上的分布多样性能估计.



曾三友(1964—),男,湖南攸县人,博士,教授,主要研究领域为演化计算,图像处理.



张青(1965—),女,副教授,主要研究领域为演化计算.



蔡振华(1984—),男,主要研究领域为演化计算.



康立山(1934—),男,教授,博士生导师,主要研究领域为演化计算,并行计算.