

基于线性标量化方法的多目标优化牛顿算法^{*}

江木兰, 席天为, 赵克全
(重庆师范大学 数学科学学院, 重庆 401331)

摘要:【目的】为了研究更高效地求解多目标优化问题,得到更有效的 Pareto 前沿面。【方法】通过对目标函数的二次近似及近似形式的线性加权标量化构造了新的搜索方向,提出了一类新的牛顿算法。进一步考虑了 Pareto 面的均匀性的优化,利用个体聚集密度来衡量 Pareto 面的均匀程度,从而在上述新的牛顿算法基础上提出了改善 Pareto 面均匀程度的算法步骤。【结果】在目标函数二阶连续可导且局部强凸的假设条件下证明了新的牛顿算法可以超线性收敛到 Pareto 弱有效解;在目标函数具有二阶连续偏导数且 Lipschitz 连续条件下证明了该算法可以局部二次收敛到 Pareto 弱有效解。【结论】基于线性标量化方法的多目标优化牛顿算法在迭代次数以及 Pareto 前沿面均匀性具有一定优越性。

关键词:多目标优化问题;牛顿算法;线性标量化方法;二次近似;Pareto 弱有效解;收敛性

中图分类号:O221.6

文献标志码:A

文章编号:1672-6693(2022)01-0001-08

多目标优化模型与算法在实际生活中有着广泛的应用,尤其在工程优化、物流网络、路径优化、社交网络聚类、决策理论、博弈论等方面。2005 年, Kuby 等人^[1]提出了基于生态与经济发展权衡的复杂多目标模型,以便在实现经济目标的同时最大限度地减少水电和蓄水能力的损失;2006 年, Erbas 等人^[2]建立了基于处理时间、功耗和构架成本的多目标模型并用于处理软件框架,从而为异构嵌入式系统的高效设计空间探索开发出一个建模和仿真环境。在过去数十年里,对多目标优化问题(Multi-objective optimization problem, MOP)的理论与方法也已有了很深入的研究,尤其在 MOP 解的定义、标量化理论、多目标优化的最优性条件、对偶理论、广义凸性等方面。2000 年杨新民等人^[3]提出了一类不可微多目标问题的 Wolfe 型和 Mond-Weir 型对偶问题;2019 年赵克全等人^[4]基于松弛变量和广义 Tchebycheff 范数的推广形式提出一类新的标量化优化问题,并进一步通过调整几种参数范围获得了一般 MOP 弱有效解、有效解和真有效解的一些完全标量化刻画。

MOP 的求解算法大致有两种类型。一是以基因算法为代表的通过直接计算求解的进化算法,如多目标进化算法(Multi-objective evolutionary algorithm, MOEA)等。例如, Li 等人^[5]提出了基于分解的 MOEA 将 MOP 分解为多个目标子问题,并以协作的方式求出它们的解,解决了 MOEA 难以求解 MOP 的主要问题——偏差特征。二是结合 MOP 的标量化方法与传统单目标优化的经典算法^[6-7]。近些年来,第二类 MOP 的求解算法研究已取得了一些重要进展。特别地, Fliege 等人^[6]提出了采用极大极小标量化方法获得最速下降方向,该方法收敛到满足多目标优化的一阶必要性条件的点。进一步地, Fliege 等人^[7]又提出了采用对于二阶可导目标函数的二阶近似的极大极小标量化方法,以此获得牛顿方向;而该算法可以局部超线性收敛到最优点,在目标函数是 Lipschitz 连续时具有局部二阶收敛性。

文献[7]中提出的关于 MOP 一个新的求解算法,对每一次的迭代的牛顿方向都需要求解一个极大极小子问题,将它转化为约束优化问题进行计算,计算量较大。受文献[7]中研究工作的启发,本文主要研究在对目标函数进行线性加权后的牛顿方向即基于线性标量化的多目标牛顿算法。

^{*} 收稿日期:2021-04-06 修回日期:2021-06-04 网络出版时间:2022-02-09 15:49

资助项目:国家自然科学基金重大项目(No. 11991024);国家自然科学基金面上项目(No. 12171063; No. 11671062);重庆市高校创新研究群体项目(No. CXQT20014);重庆市自然科学基金(No. cstc2021jcyj-msxmX0280);重庆市教育委员会科学技术研究项目(No. KJQN202100521)

第一作者简介:江木兰,女,研究方向为多目标优化问题的理论与算法, E-mail:1821135671@qq.com;通信作者:赵克全,男,教授,博士生导师, E-mail:kequanz@163.com

网络出版地址:https://kns.cnki.net/kcms/detail/50.1165.N.20220209.1443.026.html

1 预备知识

设 $\mathbf{R}$ 是实数集, $\mathbf{R}_+$ 是非负实数集, $\mathbf{R}_{++}$ 是严格正实数集。设 $U \subset \mathbf{R}^n$ 为一个开集且 $(\text{MOP}) F: U \rightarrow \mathbf{R}^m$ 是所给的多目标函数。

假设本文中提到的多目标函数具有二阶连续偏导数, 记为 $F \in C^2(U, \mathbf{R}^m)$ 。记 $\mathbf{D}_F(\mathbf{x}) \in \mathbf{R}^{m \times n}$ 为目标函数 F 在点 $\mathbf{x}$ 的 Jacobian 矩阵, $\nabla F_j(\mathbf{x}) \in \mathbf{R}^n$ 为 F_j 在点 $\mathbf{x}$ 的梯度向量, $\nabla^2 F_j(\mathbf{x}) \in \mathbf{R}^{n \times n}$ 为 F_j 在点 $\mathbf{x}$ 的 Hessian 矩阵。对于本文的目标函数, 假定对于每一个分量函数, 都有 $\nabla^2 F_j(\mathbf{x}) > 0, \forall \mathbf{x} \in U, j=1, \dots, m$ 。

在这个假设条件下, F 在 U 的任意一个凸集上都是凸函数。在求解多目标规划的算法中, 本文还需要用到一些序关系和解的定义。

定义 1 $\mathbf{x}, \mathbf{y} \in \mathbf{R}^m$, 若 $\mathbf{x} - \mathbf{y} \in \mathbf{R}_+$, 就称 $\mathbf{x} \leq \mathbf{y}$ 。

定义 2 $\mathbf{x}, \mathbf{y} \in \mathbf{R}^m$, 若 $\mathbf{x} - \mathbf{y} \in \mathbf{R}_{++}$, 就称 $\mathbf{x} < \mathbf{y}$ 。

定义 3 若点 $\mathbf{x}^* \in U$, 使得不存在 $\mathbf{y} \in U, F(\mathbf{y}) \leq F(\mathbf{x}^*)$, 且 $F(\mathbf{y}) \neq F(\mathbf{x}^*)$, 则称 $\mathbf{x}^*$ 为 (MOP) 的 Pareto 有效点。

定义 4 若点 $\mathbf{x}^* \in U$, 使得不存在 $\mathbf{y} \in U, F(\mathbf{y}) < F(\mathbf{x}^*)$, 则称 $\mathbf{x}^*$ 为 (MOP) 的 Pareto 弱有效点。

本文是为了找到 F 的有效点或者 Pareto 前沿面, 因此本文后文会用到驻点的概念。

定义 5 如果满足 $\mathbf{R}(\mathbf{D}_F(\bar{\mathbf{x}})) \cap (-\mathbf{R}_{++}) = \emptyset$, 则称 $\bar{\mathbf{x}}$ 是 F 的一个驻点。

在多目标规划不好求解时, 考虑采用标量化方法将多目标函数转化为单目标函数, 即将本文中的问题转化为最值优化问题。下面考虑线性标量化方法, 即对于目标函数 $F(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_m(\mathbf{x}))$ 给定线性泛函 $\lambda \in \mathbf{R}_+^m$, 且为方便起见, 限定 $\|\lambda\| = |\lambda_1| + |\lambda_2| + \dots + |\lambda_m| = 1$, 将目标函数转化为 $\langle \lambda, F(\mathbf{x}) \rangle$, 即:

$$\begin{aligned} (\text{MOP}) \min F(\mathbf{x}) & \quad (\text{SOP}) \min \langle \lambda, F(\mathbf{x}) \rangle \\ \text{s. t. } \mathbf{x} \in U \subset \mathbf{R}^n & \Rightarrow \text{s. t. } \mathbf{x} \in U \subset \mathbf{R}^n \\ & F(\mathbf{x}): U \rightarrow \mathbf{R}^m \\ & \|\lambda\| = 1 \end{aligned}$$

定理 1 设 $\bar{\mathbf{x}} \in U, F(\mathbf{x})$ 为 U 上的凸函数, 则 $\bar{\mathbf{x}}$ 为 (MOP) 的弱有效解当且仅当存在 $\lambda \in \mathbf{R}_+^m$, 使得 $\bar{\mathbf{x}}$ 为 (SOP) 的最优解。

2 基于线性加权与均匀性的多目标牛顿算法

2.1 线性加权的多目标牛顿法

文献[7]提出采用对于二阶可导目标函数的二阶近似的极大极小标量化方法, 进而获得算法的牛顿方向; 假设目标函数每一个分量函数都有二阶连续偏导函数且局部强凸, 每一步的牛顿方向都是利用线性标量化目标函数的二次近似函数得到的。由此考虑以下问题:

$$\min_{\mathbf{s} \in \mathbf{R}^n} \sum_{j=1}^m \lambda_j \left(\nabla f_j(\mathbf{x}_k)^T \mathbf{s} + \frac{1}{2} \mathbf{s}^T \nabla^2 f_j(\mathbf{x}_k) \mathbf{s} \right), \quad (1)$$

其中: $\sum_{j=1}^m \lambda_j = 1, \lambda_j > 0$ 。由于 $\nabla f_j(\mathbf{x})^T \mathbf{s} + \frac{1}{2} \mathbf{s}^T \nabla^2 f_j(\mathbf{x}) \mathbf{s}, j=1, \dots, m$ 一定是凸的, 上述问题有局部最优解。故由最优性必要条件, 令 $\mathbf{s} = \mathbf{x} - \mathbf{x}_k$, 多目标牛顿法的方向为:

$$\mathbf{d}_k = \arg \min_{\mathbf{s} \in \mathbf{R}^n} \sum_{j=1}^m \lambda_j \left(\nabla f_j(\mathbf{x}_k)^T \mathbf{s} + \frac{1}{2} \mathbf{s}^T \nabla^2 f_j(\mathbf{x}_k) \mathbf{s} \right)。$$

本文在此基础上提出基于线性标量化方法的多目标牛顿算法。

算法 1 (多目标线性牛顿法)

初始化: 给定终止误差 ϵ , 初始点 $\mathbf{x}_0 \in \mathbf{R}^n$, 令 $k=0$ 。

步骤 1: 解决子问题

$$\begin{aligned} \mathbf{s}(\mathbf{x}_k) &= \arg \min_{\mathbf{s} \in \mathbf{R}^n} \sum_{j=1}^m \lambda_j \left(\nabla f_j(\mathbf{x}_k)^T \mathbf{s} + \frac{1}{2} \mathbf{s}^T \nabla^2 f_j(\mathbf{x}_k) \mathbf{s} \right), \\ \theta(\mathbf{x}_k) &= \min_{\mathbf{s} \in \mathbf{R}^n} \sum_{j=1}^m \lambda_j \left(\nabla f_j(\mathbf{x}_k)^T \mathbf{s} + \frac{1}{2} \mathbf{s}^T \nabla^2 f_j(\mathbf{x}_k) \mathbf{s} \right)。 \end{aligned}$$

步骤 2:若 $\|\theta(\mathbf{x}_k)\| \leq \epsilon$, 停算, 输出 $\mathbf{x}_k$, 否则转步骤 3。

步骤 3:采用 Armijo 线搜索, 获得步长 t_k 。

步骤 4:令 $\mathbf{x}_{k+1} = \mathbf{x}_k + t_k s(\mathbf{x}_k)$, $k = k + 1$, 转步骤 1。

注 1 上述算法的子问题为求解一个无约束优化问题, 减少了计算难度; 且由于该子问题为一个凸的函数, 因此在计算多目标规划时, 迭代次数有可能相对来说较少。同时, 可以基于个体聚集密度来让该算法获得的 Pareto 前沿面效率更高(均匀程度更好)。

2.2 基于改善 Pareto 前沿面均匀度的算法

为方便说明, 如图 1 所示, 设有两个目标函数 f_1 和 f_2 , 个体 i 的聚集距离是图中实线矩形的长与宽之和, 设 $P[i]$ 为个体 i 的聚集距离, $f_1[i]$ 为个体 i 对应于第一个目标的函数值, $f_2[i]$ 为个体 i 对应于第二个目标的函数值, 则个体 i 的聚集距离为 $P[i] = |f_1[i+1] - f_1[i-1]| + |f_2[i+1] - f_2[i-1]|$ 。

一般情况下, 当有 m 个目标函数时, 个体 i 的聚集距离为

$$P[i] = \sum_{j=1}^m |f_j[i+1] - f_j[i-1]|。$$

对于 Pareto 前沿面均匀度的评价, 用 Deb 在 2002 年提出的空间评价方法并加以改进后来评价近似解集中个体在目标空间的分布情况, 有关评价函数定义为:

$$\nu = \sum_{i=1}^{|P_F|} \left| \frac{d_i - \bar{d}}{|P_F|} \right|。$$

其中: P_F 表示已知 Pareto 前沿面, d_i 是指解集中对应于第 i 个解的最近解的距离, $\bar{d}$ 指这些距离的平均值。均匀度对于 Pareto 前沿面来说也是非常重要的参数指标, 能够很好地反映解分布状态。

算法 2(基于聚集距离的搜索算法)

步骤 1:调用算法 1, 计算出 K 个弱有效解, 获得 Pareto 前沿面(记录相对应的权值)。

步骤 2:计算出每个点的聚集距离(两个极点除外)。

步骤 3:从中选出聚集距离最大的 L 个点, 找到它们对应的权值, 且:

$$\lambda_1[i] = \lambda_1[i] + \frac{1}{K}, \lambda_2[i] = \lambda_2[i] - \frac{1}{K}。$$

步骤 4:新增的 L 个权值调用加权多目标牛顿算法获得新的 Pareto 前沿面。

步骤 5:再计算新的 Pareto 前沿面的聚集距离, 选出聚集距离最小(即相对密集的点)的 M 个点, 删除后得到均匀程度较好的 $K+L-M$ 个弱有效解。

上述算法在算法 1 的基础上增加了对于 Pareto 前沿面均匀度的调整, 将重复或者比较相似的解删除, 这样可以得到相对均匀的前沿面, 使求得的解分布更加均匀。

3 收敛性分析

引理 1 在本文的通用假设下, 有以下两个结论: 1) 对于任意的 $\mathbf{x} \in U$, $\theta(\mathbf{x}) \leq 0$ 。2) 以下 3 个结果等价: a) 点 $\mathbf{x}$ 不是驻点; b) $\theta(\mathbf{x}) < 0$; c) $s(\mathbf{x}) \neq 0$ 。

证明 由 $\theta(\mathbf{x})$ 的定义, $\theta(\mathbf{x}) \leq \sum_{j=1}^m \lambda_j \left(\nabla F_j(\mathbf{x}_k)^T \mathbf{0} + \frac{1}{2} \mathbf{0}^T \nabla^2 F_j(\mathbf{x}_k) \mathbf{0} \right) = 0$, 故结论 1) 成立。

下面证明结论 2。假设 a) 成立, $\mathbf{R}(\mathbf{D}_F(\bar{\mathbf{x}})) \cap (-\mathbf{R}_{++}^m) \neq \emptyset$, 即存在 $\tilde{\mathbf{s}} \in \mathbf{R}^n$, 使得 $\nabla F_j(\mathbf{x})^T \tilde{\mathbf{s}} < 0, \forall j = 1, \dots, m$, 故对于任意的 $t > 0$, 都有:

$$\theta(\mathbf{x}) \leq \sum_{j=1}^m \lambda_j \left(\nabla F_j(\mathbf{x}_k)^T t \tilde{\mathbf{s}} + \frac{1}{2} t \tilde{\mathbf{s}}^T \nabla^2 F_j(\mathbf{x}_k) t \tilde{\mathbf{s}} \right) = t \sum_{j=1}^m \lambda_j \left(\nabla F_j(\mathbf{x}_k)^T \tilde{\mathbf{s}} + \frac{1}{2} \tilde{\mathbf{s}}^T \nabla^2 F_j(\mathbf{x}_k) \tilde{\mathbf{s}} \right),$$

因此对于足够小的 $t > 0$, 上述公式的右边一定是负数, 从而 b) 成立。

如果 b) 成立, c) 显然成立。下面假设 c) 成立, 证明 a) 成立。由本文中对目标函数的假设, 所有的 $\nabla^2 F_j(\mathbf{x})$ 都是正定矩阵, 从而对于 $\forall j = 1, 2, \dots, m$, 都有:

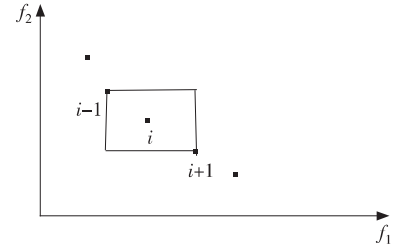


图 1 邻近个体之间的距离

Fig. 1 Distance between adjacent individuals

$$\lambda_j \nabla F_j(\mathbf{x})^T s(\mathbf{x}) < \lambda_j \left(\nabla F_j(\mathbf{x})^T s(\mathbf{x}) + \frac{1}{2} s(\mathbf{x})^T \nabla^2 F_j(\mathbf{x})^T s(\mathbf{x}) \right) = \theta(\mathbf{x}) \leq 0,$$

从而 $\mathbf{R}(\mathbf{D}_F(\bar{\mathbf{x}})) \cap (-\mathbf{R}_{++}^n) \neq \emptyset$, 则点 $\mathbf{x}$ 不是驻点, 故而 a) 成立。

证毕

推论 1 若 $\bar{\mathbf{x}}$ 不是 $F(\mathbf{x})$ 的驻点, 则 $\forall 0 < \sigma < 1, \exists r > 0, \bar{t} \in (0, 1]$, s. t. $B[\bar{\mathbf{x}}, r] \subset U$, 且对于任意的 $\mathbf{x} \in B[\bar{\mathbf{x}}, r]$, $t \in [0, \bar{t}]$, $j = 1, \dots, m$ 都有 $\mathbf{x} + ts(\mathbf{x}) \in U, F_j(\mathbf{x} + ts(\mathbf{x})) \leq F_j(\mathbf{x}) + \sigma t \theta(\mathbf{x})$ 。

推论 1 的证明与文献[7]的引理 3.3 证明类似。

注 2 推论 1 说明了步长的存在性。

引理 2 设 $\mathbf{x} \in U, a, b > 0$, 使得 $0 < a \leq b$ 。若 $a\mathbf{I} \leq \nabla^2 f_j(\mathbf{x}) \leq b\mathbf{I}, j = 1, \dots, m$, 则有:

$$\frac{a}{2} \|s(\mathbf{x})\|^2 \leq |\theta(\mathbf{x})| \leq \frac{b}{2} \|s(\mathbf{x})\|^2.$$

引理 2 的证明只需将文献[7]引理 4.2 的证明过程中的拉格朗日乘子换为线性标量化中的权重向量即可。

引理 3 设 $\mathbf{x} \in U, a > 0$, 若 $a\mathbf{I} \leq \nabla^2 f_j(\mathbf{x}), j = 1, \dots, m$, 则:

$$|\theta(\mathbf{x})| \leq \frac{1}{2a} \left\| \sum_{j=1}^m \lambda_j \nabla f_j(\mathbf{x}) \right\|^2, \forall \lambda_j \geq 0, \sum_{j=1}^m \lambda_j = 1.$$

引理 3 的证明与文献[7]的引理 4.3 证明类似。

引理 4 设 $\hat{\mathbf{x}} \in U$, 令 $\hat{\mathbf{x}}_+ = \hat{\mathbf{x}} + s(\hat{\mathbf{x}})$, 如果 $a, r, \delta, \varepsilon > 0$, 使得: a) $a\mathbf{I} \leq \nabla^2 f_j(\mathbf{x}), j = 1, \dots, m, \mathbf{x} \in B[\hat{\mathbf{x}}, r]$, b) $\|\nabla^2 f_j(\mathbf{x}) - \nabla^2 f_j(\mathbf{y})\| \leq \varepsilon, \forall \mathbf{x}, \mathbf{y} \in B[\hat{\mathbf{x}}, r], \|\mathbf{x} - \mathbf{y}\| < \delta$; c) $B[\hat{\mathbf{x}}, r] \subset U$; d) $\|s(\hat{\mathbf{x}})\| \leq \min\{\delta, r\}$, 则有 $\|s(\hat{\mathbf{x}}_+)\| \leq \frac{\varepsilon}{a} \|s(\hat{\mathbf{x}})\|$ 。若在上述条件基础上, 满足 $\|\nabla^2 f_j(\mathbf{x}) - \nabla^2 f_j(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in B[\hat{\mathbf{x}}, r], \forall \mathbf{x}, \mathbf{y} \in B[\hat{\mathbf{x}}, r]$, 则有 $\|s(\hat{\mathbf{x}}_+)\| \leq \frac{L}{a} \|s(\hat{\mathbf{x}})\|^2$ 。

引理 4 的证明只需将文献[7]的引理 4.4 的证明过程中的拉格朗日乘子看做线性标量化中的权重向量即可。

引理 5 设 $\{\mathbf{x}_k\}$ 为算法产生的数列, 则:

$$\forall k \geq 0, \exists t > 0, \text{ s. t. } t\theta(\mathbf{x}_k) \geq \max_{j=1, \dots, m} \left\{ \nabla f_j(\mathbf{x}) s(\mathbf{x}_k) + \frac{1}{2} s(\mathbf{x}_k)^T \nabla^2 f_j(\mathbf{x}_k) s(\mathbf{x}_k) \right\}.$$

证明 由 $s(\mathbf{x}_k), \theta(\mathbf{x}_k)$ 的定义以及引理 1 可得:

$$\nabla f_j(\mathbf{x}) s(\mathbf{x}_k) + \frac{1}{2} s(\mathbf{x}_k)^T \nabla^2 f_j(\mathbf{x}_k) s(\mathbf{x}_k) \leq 0, j = 1, \dots, m, \theta(\mathbf{x}_k) \leq 0,$$

从而结论显然成立。

证毕

定理 2 设 $\{\mathbf{x}_k\}$ 为算法产生的序列, 假设存在 $V \subset U, a, b, r, \delta, \varepsilon, t > 0$, 且 a) $a\mathbf{I} \leq \nabla^2 F_j(\mathbf{x}) \leq b\mathbf{I}, \forall \mathbf{x} \in V, j = 1, \dots, m$; b) $\forall \mathbf{x}, \mathbf{y} \in V$ 满足 $\|\mathbf{x} - \mathbf{y}\| < \delta$, 都有 $\|\nabla^2 F_j(\mathbf{x}) - \nabla^2 F_j(\mathbf{y})\| \leq \varepsilon$; c) $\frac{\varepsilon}{a} \leq 1 - \sigma < t$; d) $B[\mathbf{x}_0, r] \subset V$;

e) $\|s(\mathbf{x}_0)\| \leq \min\left\{\delta, r\left(1 - \frac{\varepsilon}{a}\right)\right\}$ 。从而对于所有的 k , 都有: 1) $\|\mathbf{x}_k - \mathbf{x}_0\| \leq \|s(\mathbf{x}_0)\| \frac{1 - \left(\frac{\varepsilon}{a}\right)^k}{1 - \frac{\varepsilon}{a}}$; 2) $\|s(\mathbf{x}_k)\| \leq$

$\|s(\mathbf{x}_0)\| \left(\frac{\varepsilon}{a}\right)^k$; 3) $t_k = 1$; 4) $\|s(\mathbf{x}_{k+1})\| \leq \|s(\mathbf{x}_k)\| \left(\frac{\varepsilon}{a}\right)$ 。且序列 $\{\mathbf{x}_k\}$ 会收敛到一个局部 Pareto 点 $\bar{\mathbf{x}} \in \mathbf{R}^n$ 满足 $\|\bar{\mathbf{x}} - \mathbf{x}_0\| \leq \frac{\|s(\mathbf{x}_0)\|}{1 - \frac{\varepsilon}{a}} \leq r$, 即序列 $\{\mathbf{x}_k\}$ 为超线性收敛。

证明 首先证明若对于某些 k 来说, 1) 和 2) 成立, 则对于这些 k 来说, 3) 和 4) 也成立。由三角不等式以及 1) 和 2), 有:

$$\|(\mathbf{x}_k + s(\mathbf{x}_k)) - \mathbf{x}_0\| \leq \|s(\mathbf{x}_0)\| \frac{1 - \left(\frac{\varepsilon}{a}\right)^{k+1}}{1 - \frac{\varepsilon}{a}}.$$

又由 e), 2) 及 c) 可知:

$$\mathbf{x}_k, \mathbf{x}_k + s(\mathbf{x}_k) \in B[\mathbf{x}_0, r], \|\mathbf{x}_k + s(\mathbf{x}_k) - \mathbf{x}_k\| \leq \delta.$$

利用 b)、文献[7]的引理 4.1 以及本文的引理 5 可得对 $\forall j=1, \dots, m$, 有:

$$f_j(\mathbf{x}_k + s(\mathbf{x}_k)) \leq f_j(\mathbf{x}_k) + \nabla f_j(\mathbf{x}_k)^T s(\mathbf{x}_k) + \frac{1}{2} s(\mathbf{x}_k)^T \nabla^2 f_j(\mathbf{x}_k) s(\mathbf{x}_k) + \frac{\epsilon}{2} \|s(\mathbf{x}_k)\|^2 \leq$$

$$f_j(\mathbf{x}_k) + t\theta(\mathbf{x}_k) + \frac{\epsilon}{2} \|s(\mathbf{x}_k)\|^2 = f_j(\mathbf{x}_k) + (t - (1 - \sigma))\theta(\mathbf{x}_k) + (1 - \sigma)\theta(\mathbf{x}_k) + \frac{\epsilon}{2} \|s(\mathbf{x}_k)\|^2.$$

由引理 1, $\theta(\mathbf{x}_k) \leq 0$, 且由 a) 以及引理 2 可知:

$$(1 - \sigma)\theta(\mathbf{x}_k) + \frac{\epsilon}{2} \|s(\mathbf{x}_k)\|^2 \leq (\epsilon - a(1 - \sigma)) \frac{\|s(\mathbf{x}_k)\|^2}{2} \leq 0.$$

从而 3) 成立。由 c) 可得:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + s(\mathbf{x}_k), \mathbf{x}_k, \mathbf{x}_{k+1} \in B[\mathbf{x}_0, r], \|\mathbf{x}_{k+1} - \mathbf{x}\| \leq \delta,$$

故由 a)、b) 和 d) 以及引理 4 可得 4) 成立。后续证明与文献[7]定理 5.1 后半部分的证明类似。

证毕

定理 3 在定理 1 的假设下, 若每个目标函数的二阶偏导 $\nabla^2 F_j(x)$ 在 V 上 Lipschitz 连续且 Lipschitz 常数为 L , 令 $\tau_k = \left(\frac{L}{a}\right) \|s(\mathbf{x}_k)\|$, $\zeta \in \left(0, \frac{1}{2}\right)$, 则存在 k_0 , 对于 $k \geq k_0$ 且 $\tau_k < \zeta$ 时, 有:

$$\|\bar{\mathbf{x}} - \mathbf{x}_{k+1}\| \leq \frac{L}{a} \frac{1 - \tau_k}{(1 - 2\tau_k)^2} \|\bar{\mathbf{x}} - \mathbf{x}_k\|^2 \leq \frac{L}{a} \frac{1 - \zeta}{(1 - 2\zeta)^2} \|\bar{\mathbf{x}} - \mathbf{x}_k\|^2,$$

其中 $\bar{\mathbf{x}}$ 是算法产生的序列 $\{\mathbf{x}_k\}$ 的 Pareto 最优极限, 则序列 $\{\mathbf{x}_k\}$ 二次收敛于 $\bar{\mathbf{x}}$ 。

定理 3 的证明与文献[7]的定理 2 类似。

4 数值实验

本文借助 Matlab R2016a 进行数值实验, 对每个目标函数的权重在 $[0, 1]$ 均匀取得, 停止准则为 $\|\theta(\mathbf{x}_k)\| < \delta$ 。原始目标函数一般假定具有箱型约束 $L \leq \mathbf{x} \leq U$, 且在 (1) 式中这样使用:

$$\min_{s \in \mathbf{R}^n} \sum_{j=1}^p \lambda_j \left(\nabla f_j(\mathbf{x}_k)^T s + \frac{1}{2} s^T \nabla^2 f_j(\mathbf{x}_k) s \right), L - \mathbf{x}_k \leq s \leq U - \mathbf{x}_k. \quad (2)$$

目标函数带有箱型约束会避免原问题与子问题出现无界的情况, 且对于带有箱型约束 $L \leq \mathbf{x} \leq U$ 的目标函数来说, 在每个可行局部 Pareto 有效解 $\mathbf{x}_k$, (2) 式的目标函数都为 0。此时对于原始目标函数不会存在下降方向。

数值实验的算例如表 1 所示, 其中: 算例名称的字母部分为对应文献中作者姓的首字母组合 (独立作者则采用作者姓的全称); 字母后数字表示某算例在对应文献中的排序, 如 DD1 为文献[11]中第一个算例; n 表示自变量的维数。数值实验具体结果如表 2 所示, 其中: U^T 和 L^T 分别表示目标函数自变量取值的上下界; 平均迭代次数与平均函数值分别表示在利用算法 1 且初始值在自变量上下界内随机取值 150 个并获得 150 个解的平均迭代次数和平均函数值, 均匀度分析 1 是在采用算法 2 前计算所得的 Pareto 面均匀度, 均匀度分析 2 是指采用算法 2 后所得的 Pareto 面均匀度, 均匀度分析 3 是利用文献[7]中极大极小牛顿法算出的 Pareto 面均匀度。

表 1 数值实验的算例

Tab. 1 Examples of numerical experiments

算例名称	算例来源	n	目标函数
PNR	文献[8]	2	$f_1(x) = x_1^4 + x_2^4 - x_1^2 + x_2^2 + 0.25x_1 + 20, f_2(x) = (x_1 - 1)^2 + x_2^2$
Deb	文献[9]	2	$f_1(x) = x_1, f_2(x) = \frac{2 - e^{-\left(\frac{x_2 - 0.2}{0.004}\right)^2}}{x_1} - 0.8e^{-\left(\frac{x_2 - 0.6}{0.4}\right)^2}$
DD1	文献[10]	5	$f_1(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2, f_2(x) = 3x_1 + 2x_2 - \frac{x_3}{3} + 0.01(x_4 - x_5)^2$
SD	文献[11]	4	$f_1(x) = 2x_1 + \sqrt{2}x_2 + \sqrt{2}x_3 + x_4, f_2(x) = \frac{2}{x_1} + \frac{2\sqrt{2}}{x_2} + \frac{2\sqrt{2}}{x_3} + \frac{2}{x_4}$

表 2 4 个数值实验的主要结果
Tab. 2 Main results of four numerical experiments

算例名称	n	U^T	L^T	平均迭代次数	平均函数值	均匀度分析 1	均匀度分析 2	均匀度分析 3
PNR	2	$-[2,2]$	$[2,2]$	5.860 0	$[9.417\ 3,3.882\ 3]$	0.153 5	0.087 3	0.139 4
Deb	2	$[0.1,1]$	$[1,1]$	3.620 0	$[1.512\ 0,1.749\ 2]$	0.053 7	0.026 8	0.059 6
DD1	5	$-[1,\cdots,1]$	$[1,\cdots,1]$	3.713 3	$[1.522\ 6,-3.756\ 2]$	0.040 7	0.030 5	0.042 1
SD	4	$[1,\sqrt{2},\sqrt{2},1]$	$[3,\cdots,3]$	2.980 0	$[9.823\ 3,6.357\ 6]$	0.084 2	0.042 3	0.018 8

从表 2 可以看出,对于本文选择的 4 个算例,算法 1 都可以在比较少的迭代次数达到 Pareto 弱有效解,且算法 2 在算法 1 的基础上很大程度地改善了 Pareto 前沿面的均匀性,与文献[7]中这 4 个算例的 Pareto 前沿面均匀性相比也表现出一定的优越性。

对算例 PNR 以及 Deb 采用算法 1 和算法 2 计算得到的含有 150 个点的 Pareto 前沿面的图示见图 2~5。从图 4 可以明显看出 Deb 的解相对来说均匀很多,且删掉了重复的解,更好地得到了目标函数弱有效解的分布。此外,对于 Deb 算例分别用参考文献[1]中极小极大目标牛顿法以及本文提出的线性标量化牛顿法求出含有 150 个 Pareto 解的解空间分别如图 6 和图 7 所示。

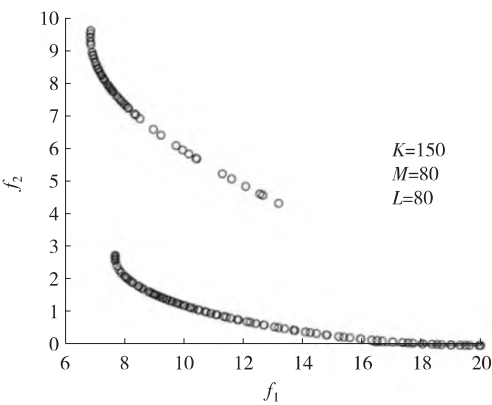


图 2 采用算法 2 计算 PNR
Fig. 2 PNR is calculated by algorithm 2

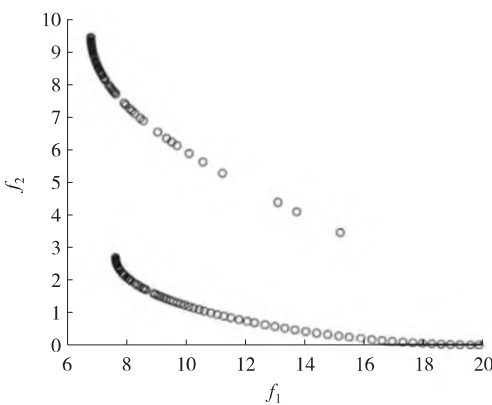


图 3 采用算法 1 计算 PNR
Fig. 3 PNR is calculated by algorithm 1

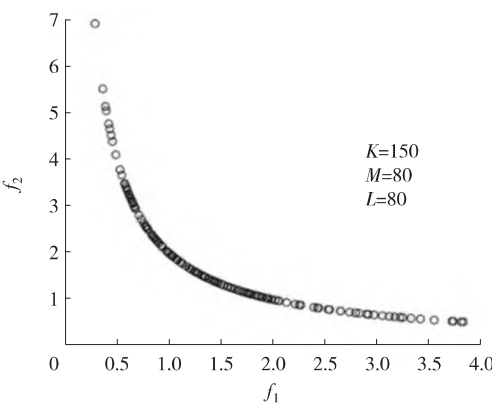


图 4 采用算法 2 计算 Deb
Fig. 4 Deb is calculated by algorithm 2

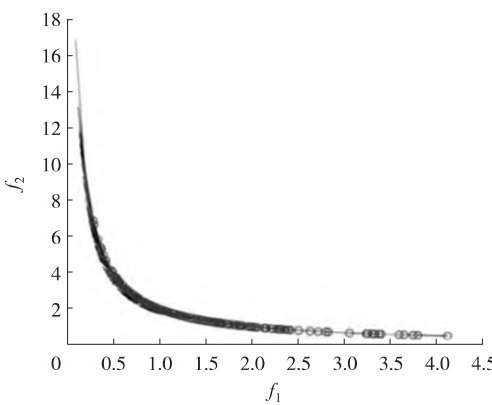
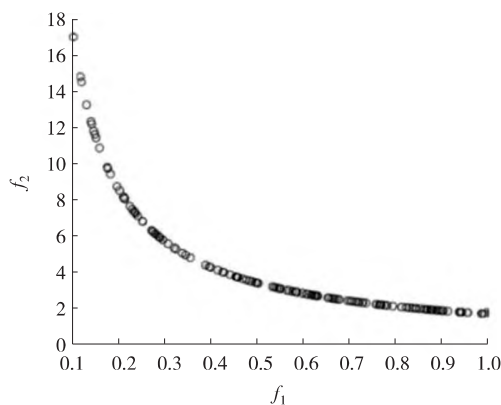


图 5 采用算法 1 计算 Deb
Fig. 5 Deb is calculated by algorithm 1

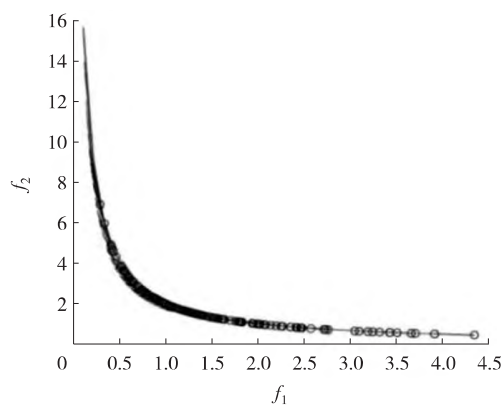
数值实验表明,对于这 4 个算例以及在上下界随机生成的初始点,由本文提出的算法所产生的序列都可以收敛到 Pareto 前沿面。因此,可以大胆推断,对于某些曲率不大的目标函数,算法 1 以及算法 2 在迭代次数、计算量以及 Pareto 前沿面均匀度上都有较好的表现。



注:函数均值为 2.394 2

图6 采用极大极小多目标牛顿法计算 Deb 的 Pareto 前沿面

Fig. 6 The Pareto front of Deb is calculated by minimax multi-objective Newton method



注:函数均值为 1.649 4

图7 采用线性标量化牛顿法计算 Deb 的 Pareto 前沿面

Fig. 7 The Pareto front of Deb is calculated by linear scalar multi-objective Newton method

5 结束语

本文研究了基于目标函数是二阶连续可导且局部强凸的多目标算法,提出了对于目标函数的二次近似进行线性加权并求它的最小值而获得的牛顿下降方向,证明了基于该牛顿方向的多目标牛顿算法超线性收敛到 Pareto 弱有效点,若目标函数的二阶偏导函数 Lipschitz 连续,可得该算法局部二次收敛。本文创新性地引进了线性标量化方法和 Pareto 前沿面的均匀程度分析,提升了解的质量,实验表明了对于局部强凸的目标函数来说,线性标量化的多目标牛顿优化算法在迭代次数以及解的均匀程度具有一定的优越性。如何进一步减弱上述算法收敛性的假设条件以及如何应用该算法处理大规模优化问题将是下进一步研究的方向。

参考文献:

- [1] KUBY M J, FAGAN W F, REVELLE C S, et al. A multiobjective optimization model for dam removal: an example trading off salmon passage with hydropower and water storage in the Willamette basin[J]. *Advances in Water Resources*, 2005, 28(8): 845-855.
- [2] ERBAS C, CERAV-ERBAS S, PIMENTEL A D. Multiobjective optimization and evolutionary algorithms for the application mapping problem in multiprocessor system-on-chip design[J]. *IEEE Transactions on Evolutionary Computation*, 2006, 10(3): 358-374.
- [3] 杨新民, 赵克全. 向量优化问题的近似解研究[J]. *运筹学学报*, 2017, 21(4): 1-18.
YANG X M, ZHAO K Q. Research on approximate solutions of vector optimization problems[J]. *Journal of Operations Research*, 2017, 21(4): 1-18.
- [4] 赵克全, 杨新民, 夏远梅. 多目标优化问题的完全标量化[J]. *中国科学: 数学*, 2021, 51(2): 411-424.
ZHAO K Q, YANG X M, XIA Y M. Complete scalarization of multiobjective optimization problems[J]. *Science China Mathematics*, 2021, 51(2): 411-424.
- [5] LI H, ZHANG Q F, DENG J D. Biased multiobjective optimization and decomposition algorithm[J]. *IEEE Transactions on Cybernetics*, 2016, 47(1): 52-66.
- [6] FLIEGE J, SVAITER B F. Steepest descent methods for multicriteria optimization[J]. *Mathematical Methods of Operations Research*, 2000, 51(3): 479-494.
- [7] FLIEGE J, GRAA L M, SVAITER D B F. Newton's method for multiobjective optimization[J]. *SIAM Journal on Optimization*, 2008, 20(2): 602-626.
- [8] PREUSS M, NAUJOKS B, RUDOLPH G. Pareto set and EMOA behavior for simple multimodal multiobjective functions [C]//RUNARSSON T P, BEYER H G, BURKE E, et al. *Parallel problem solving from nature-PPSN IX*. Berlin: Springer, 2006: 513-522.
- [9] DEB K. Multi-objective genetic algorithms: problem difficulties and construction of test problems[J]. *Evolutionary Computation*,

1999,7(3):205-230.

- [10] DAS I, DENNIS J E. Normal-boundary intersection: a new method for generating the Pareto surface in nonlinear multicriteria optimization problems[J]. SIAM Journal on Optimization, 1996, 8(3): 631-657.
- [11] STADLER W, DAUER J. Multicriteria optimization in engineering: a tutorial and survey[M]//KAMAT M P. Structural optimization: status and promise. Reston: American Institute of Aeronautics and Astronautics, 1993: 209-249.

Operations Research and Cybernetics

Newton Algorithm Based on Linear Scaling Method for Multi-Objective Optimization

JIANG Shulan, XI Tianwei, ZHAO Kequan

(School of Mathematical Science, Chongqing Normal University, Chongqing 401331, China)

Abstract: [Purposes] To solve the multi-objective optimization problem more efficiently, to get more effective Pareto front surface.

[Methods] Through the quadratic approximation of the objective function and the linear weighted scalarization of the approximate form, a new search direction is constructed, and a new Newton algorithm is proposed. The optimization of the uniformity of Pareto surface is further considered, and the individual aggregation density is used to measure the uniformity of Pareto surface. Based on the above new Newton algorithm, the algorithm steps to improve the uniformity of Pareto surface are proposed. [Findings] Under the assumption that the objective function is second-order continuous differentiable and locally strongly convex, it is proved that the new Newton algorithm can superlinearly converge to Pareto weakly efficient solution. It is proved that the algorithm can locally quadratic converge to Pareto weakly efficient solutions under the condition that the objective function has second-order continuous partial derivatives and Lipschitz continuity. [Conclusions] The multi-objective optimization Newton algorithm based on linear scalar method has certain advantages in the number of iterations and the uniformity of Pareto front.

Keywords: multi-objective optimization problem; Newton algorithm; linear scalarization method; quadratic approximation; Pareto weak efficient solution; convergence

(责任编辑 方 兴)