运筹学与控制论

DOI:10.11721/cqnuj20230106

多目标优化问题的非单调对角最速下降算法*

杨春蓉，谭豫琳，赵克全
(重庆师范大学 数学科学学院，重庆 401331)

摘要:【目的】为了更高效的求解多目标优化问题,得到更有效的 Pareto 前沿面。【方法】通过引入非单调 Armijo 准则,得到新的步长搜索方式,进而提出了多目标优化问题的非单调对角最速下降算法。【结果】在目标函数无凸性、梯度 Lipschitz 连续性和下有界假设下,证明了算法产生序列的每个聚点均是多目标优化问题的 Pareto 弱有效解,并在适当条件下证明了算法的次线性收敛性。【结论】数值实验表明提出的算法目标函数值的平均值更小。

关键词:多目标优化;非单调线搜索;对角最速下降算法;Pareto 弱有效解

中图分类号:O221.6　　**文献标志码**:A　　**文章编号**:1672-6693(2023)01-0114-09

近些年来,多目标优化理论与方法已广泛应用于经济、工业、交通等领域[1-3],关于多目标优化研究发展十分迅速[4-7]。标量化方法常被用于求解无约束多目标优化问题,即将多目标优化问题转化成一个或者多个单目标优化问题进行处理。但从数值计算的角度来看,标量化方法也存在一些不足。首先,标量化方法中几乎所有的参数选择都可能导致无界的最优化问题[8];其次,标量化方法的参数不是先验已知的,必须预先给定[9];最后,在非凸的情况下可能无法获得 Pareto 最优解[10-11]。

求解多目标优化问题的另一类思路就是直接发展求解多目标优化问题的算法。较早期的工作可以追溯到 2000 年,Fliege 等人[12]将最速下降算法应用到直接求解多目标优化问题上,在目标函数凸性假设下,该算法的收敛性得到保证;2005 年,Drummond 等人[13]在文献[12]的基础上通过引入偏序提出了多目标优化的最速下降方法;2009 年,Fliege 等人[9]将牛顿算法应用于求解多目标优化问题,在目标函数二次可微条件下具有二阶收敛性,能够快速求出多目标优化问题的局部有效解;2011 年,Qu 等人[14]提出了多目标优化问题的拟牛顿算法,该算法近似了目标函数的 Hessian 矩阵,同时避免了凸性假设;2021 年,Mustapha 等人[15]针对无约束多目标优化问题提出了加速对角最速下降算法,该算法主要利用一个对角矩阵近似每个目标函数的 Hessian 矩阵,避免了二阶连续可微和凸性假设。这些迭代方法都是从单目标优化算法发展而来,步长均由单调线搜索方法确定,且步长的选择将使目标函数值在每一次迭代中都会减少,不会增加。

然而,在非单调的线搜索步长中可以允许函数值增加。文献[16-17]指出单目标优化的非单调线搜索可以提高找到全局最优解的可能性。当非单调的线搜索方法应用于较复杂的单目标优化时,已经有较好的数值结果[18-23]。此外,Kanako 等人[24]将非单调线搜索步长运用到多目标优化算法中;Qu 等人[25-26]提出了一些非单调的梯度方法;2020 年,Mahdavi 等人[27]提出了超线性收敛的非单调拟牛顿算法,以求解无约束的多目标优化问题。

本文在 Mustapha 等人研究工作基础上,提出了非单调对角最速下降算法,采用非单调的 Armijo 准则求解下降方向。在目标函数连续可微及非凸假设下,证明了非单调对角最速下降算法产生的序列的每个聚点是多目标优化问题的 Pareto 弱有效解。数值实验表明,本文提出的非单调对角最速下降算法在求解多目标线性优化模型上,函数的平均值更低。

* **收稿日期**:2022-08-15　**修回日期**:2022-12-05　**网络出版时间**:2023-02-22 16:31

资助项目:国家自然科学基金(No. 11991024;No. 12171063);重庆市高校创新研究群体项目(No. CXQT20014);重庆市自然科学基金面上项目(No. cstc2021jcyj-msxmX0280);重庆英才计划(No. CQYC20210302270);重庆市教育委员会科学技术研究项目(No. KJQN202100521)

第一作者简介:杨春蓉,女,研究方向为多目标优化理论与算法,E-mail:1622010843@qq.com;**通信作者**:赵克全,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.20230222.1338.018.html

1 预备知识

设 $\mathbf{R}$ 是实数集,$\mathbf{R}_+$ 是非负实数集,$\mathbf{R}_{++}$ 是正实数集。文中所使用的符号说明如下:$\langle,\rangle$ 表示 $\mathbf{R}^n$ 空间中的内积,$\|\cdot\|$ 表示范数,定义偏序 $\leqslant(<)$,$\boldsymbol{x}\leqslant\boldsymbol{y}$ 当且仅当 $\boldsymbol{y}-\boldsymbol{x}\in\mathbf{R}_+^m(\boldsymbol{y}-\boldsymbol{x}\in\mathbf{R}_{++}^m)$。函数 $\boldsymbol{F}:\mathbf{R}^n\to\mathbf{R}^m$,且 $\boldsymbol{F}(\boldsymbol{x})=(F_1(\boldsymbol{x}),F_2(\boldsymbol{x}),\cdots,F_m(\boldsymbol{x}))$,$\boldsymbol{F}$ 在 $\boldsymbol{x}$ 处的雅可比矩阵为 $\nabla\boldsymbol{F}(\boldsymbol{x})$,$\nabla F_i(\boldsymbol{x})$ 和 $\nabla^2F_i(\boldsymbol{x})$ 分别表示 $\boldsymbol{F}$ 的第 i 个分量在 x 处的梯度和 Hessian 矩阵。考虑如下多目标优化问题:

$$\begin{aligned}&\min\ \boldsymbol{F}(\boldsymbol{x}),\\&\text{s.t. }\ \boldsymbol{x}\in\mathbf{R}^n。\end{aligned}\tag{1}$$

其中:$F_i:\mathbf{R}^n\to\mathbf{R}$ 连续可微。

定义 1　对所有的 $i\in I$,如果不存在 $\boldsymbol{x}\in\mathbf{R}^n$,使得 $F_i(\boldsymbol{x})\leqslant F_i(\boldsymbol{x}^*)$ 且 $F_i(\boldsymbol{x})\neq F_i(\boldsymbol{x}^*)$,则称 $\boldsymbol{x}^*$ 是问题(1)的 Pareto 有效解。

定义 2　对所有的 $i\in I$,如果不存在 $\boldsymbol{x}\in\mathbf{R}^n$,使得 $F_i(\boldsymbol{x})<F_i(\boldsymbol{x}^*)$,则称 x^* 是问题(1)的 Pareto 弱有效解。

定义 3　对所有的 $i\in I$,如果存在 $\lambda_i\in\mathbf{R}_+^m$,使得 $\sum_{i\in I}\lambda_i\nabla F_i(\boldsymbol{x}^*)=0$,则称 $\boldsymbol{x}^*$ 为问题(1)的 Pareto 临界点。

定义 4　如果对所有的 $i\in I$,有 $\langle\nabla F_i(\boldsymbol{x}),\boldsymbol{d}\rangle<0$,$\boldsymbol{d}\in\mathbf{R}^n$,$\boldsymbol{x}\in\mathbf{R}^n$ 则称 $\boldsymbol{d}$ 为每个目标函数 $F_i(\boldsymbol{x})$ 的下降方向。

2 超线性收敛的非单调对角最速下降算法

2.1 下降方向的选择

考虑问题(1),令 $\boldsymbol{x}\in\mathbf{R}^n$ 是当前迭代点。对每一个目标函数 F_i 都给出一个线性近似,搜索方向可以通过极小化这些线性近似加上二次正则化项得到。因此,令 $\Gamma:\mathbf{R}^n\to\mathbf{R}$,具体函数表达式如下:

$$\Gamma(\boldsymbol{d})=\max_{i\in I}\ \langle\nabla F_i(\boldsymbol{x}_k),\boldsymbol{d}\rangle+\tau^k\|\boldsymbol{d}\|^2/2,$$

其中:τ^k 是正则化参数。极小化 $\Gamma(\boldsymbol{d})$,可得下降方向 $\boldsymbol{d}$。同理,本文可求解以下子问题获取下降方向 d:

$$\begin{aligned}&\min\ \max_{i\in I}\ \langle\nabla F_i(\boldsymbol{x}_k),\boldsymbol{d}\rangle+\tau^k\|\boldsymbol{d}\|^2/2,\\&\text{s.t. }\ \boldsymbol{d}\in\mathbf{R}^n。\end{aligned}\tag{2}$$

注 1　假设 F_i 为凸函数,问题(1)在 $\boldsymbol{x}^k\in\mathbf{R}^n$ 的牛顿方向[9],是每个目标函数 F_i 在 x^k 处包含二次近似的最优解:

$$\begin{aligned}&\min\ \max_{i\in I}\langle\nabla F_i(\boldsymbol{x}_k),\boldsymbol{d}\rangle+\langle\nabla^2F_i(\boldsymbol{x}_k)\boldsymbol{d},\boldsymbol{d}\rangle/2,\\&\text{s.t. }\ \boldsymbol{d}\in\mathbf{R}^n。\end{aligned}$$

本文对目标函数没有凸性假设,为了避免计算每个目标函数 F_i 的 Hessian 矩阵,在每一次迭代中采用一个公共的对角矩阵 τ^kI 去替代原来的 Hessian 矩阵。易知,问题(2)是非光滑优化问题,等价问题为:

$$\begin{aligned}&\min\ \xi,\\&\text{s.t. }\ \langle\nabla F_i(\boldsymbol{x}_k),\boldsymbol{d}\rangle+\tau^k\|\boldsymbol{d}\|^2/2\leqslant\xi,i\in I,\boldsymbol{d}\in\mathbf{R}^n。\end{aligned}\tag{3}$$

问题(3)的对偶问题为一个二次规划问题:

$$\begin{aligned}&\min\ \Big\|\sum_{i\in I}\lambda_i\nabla F_i(\boldsymbol{x}_k)\Big\|^2/(2\tau^k),\\&\text{s.t. }\ \lambda_i\geqslant0,i\in I,\sum_{i\in I}\lambda_i=1。\end{aligned}\tag{4}$$

假设问题(4)的最优解为 $\boldsymbol{\lambda}^k$,那么问题(3)的解的形式为:

$$\begin{aligned}\boldsymbol{d}^k&=-\Big[\sum_{i\in I}\lambda_i^k\nabla F_i(\boldsymbol{x}_k)\Big]/\tau^k,\\\xi^k&=-\Big[\Big\|\sum_{i\in I}\lambda_i^k\nabla F_i(\boldsymbol{x}_k)\Big\|^2\Big]/\tau^k。\end{aligned}\tag{5}$$

那么一定存在 $\lambda_i=\lambda_i(\boldsymbol{x})$,$i=1,2,\cdots,m$,因此令 $\boldsymbol{d}=\boldsymbol{d}(\boldsymbol{x})$,$\theta(\boldsymbol{x})=\xi(\boldsymbol{x})$。

引理 1　假设 $\boldsymbol{\lambda}^k$ 是问题(4)的最优解,$\boldsymbol{d}^k$ 是由定义(5)给出的,假设 $\tau^k\geqslant\tau_0>0$。那么,以下结论中一个必定成立:i) 若 $\boldsymbol{d}^k=0$,则 $\boldsymbol{x}^k$ 是 Pareto 临界点;ii) 若 $\boldsymbol{d}^k\neq0$,则 $\langle\nabla F_i(\boldsymbol{x}^k),\boldsymbol{d}^k\rangle\leqslant-\tau^k\|\boldsymbol{d}^k\|^2$,$\forall i\in I$ 成立。

2.2 **更新正则化参数 τ^k**

首先考虑以下的二次模型：

$$F_i(\boldsymbol{x}^k)+\langle\nabla F_i(\boldsymbol{x}^k),\boldsymbol{x}-\boldsymbol{x}^k\rangle+\tau^k\|\boldsymbol{x}-\boldsymbol{x}^k\|^2/2,$$
$$F_i(\boldsymbol{x}^{k+1})+\langle\nabla F_i(\boldsymbol{x}^{k+1}),\boldsymbol{x}-\boldsymbol{x}^{k+1}\rangle+\tau^{k+1}\|\boldsymbol{x}-\boldsymbol{x}^{k+1}\|^2/2。$$

由前面可知 $\sum_{i\in I}\lambda_i^k=1$，以上式子同时乘以 $\sum_{i\in I}\lambda_i^k=1$ 可以得到：

$$\Lambda^k(\boldsymbol{x}):=\sum_{i\in I}\lambda_i^k(F_i(\boldsymbol{x}^k)+\langle\nabla F_i(\boldsymbol{x}^k),\boldsymbol{x}-\boldsymbol{x}^k\rangle)+\tau^k\|\boldsymbol{x}-\boldsymbol{x}^k\|^2/2,$$
$$\Lambda^{k+1}(\boldsymbol{x}):=\sum_{i\in I}\lambda_i^k(F_i(\boldsymbol{x}^{k+1})+\langle\nabla F_i(\boldsymbol{x}^{k+1}),\boldsymbol{x}-\boldsymbol{x}^{k+1}\rangle)+\tau^{k+1}\|\boldsymbol{x}-\boldsymbol{x}^{k+1}\|^2/2。$$

如果$\nabla\Lambda^k(\boldsymbol{x})=\nabla\Lambda^{k+1}(\boldsymbol{x})$，则 $\tau^{k+1}I$ 能够很好地逼近 $\sum_{i\in I}\lambda_i^k\nabla^2F_i(\boldsymbol{x}^{k+1})$。易知$\nabla\Lambda^k(\boldsymbol{x})=\nabla\Lambda^{k+1}(\boldsymbol{x})$等价于：

$$\sum_{i\in I}\lambda_i^k\nabla F_i(\boldsymbol{x}^k)=\sum_{i\in I}\lambda_i^k\nabla F_i(\boldsymbol{x}^{k+1})+\tau^{k+1}(\boldsymbol{x}^k-\boldsymbol{x}^{k+1})。$$

更进一步，可以写成 $\tau^{k+1}(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k)=\sum_{i\in I}\lambda_i^k(\nabla F_i(\boldsymbol{x}^{k+1})-\nabla F_i(\boldsymbol{x}^k))$。令

$$\boldsymbol{s}^k:=\boldsymbol{x}^{k+1}-\boldsymbol{x}^k,\boldsymbol{\gamma}_i^k:=\nabla F_i(\boldsymbol{x}^{k+1})-\nabla F_i(\boldsymbol{x}^k),\boldsymbol{u}^k:=\sum_{i\in I}\lambda_i^k\boldsymbol{\gamma}_i^k,$$

则上式可以写成 $\tau^{k+1}\boldsymbol{s}^k=\boldsymbol{u}^k$，那么可以求解下面子问题得到 τ^{k+1}，对每一个 $\tau_0>0$，有：

$$\min\ \|\tau\boldsymbol{s}^k-\boldsymbol{u}^k\|^2,$$
$$\text{s.t.}\ \tau\geqslant\tau_0。$$

以上问题最优解的显示表达式为 $\tau^{k+1}=\max\{\tau_0,u_i^k/s_i^k\}$。

2.3 **算法 1**

2021 年，Mustaph 等人[16]提出了利用对角矩阵去逼近函数的 Hessian 矩阵，通过解决一个含有对角二次项的极大极小值问题，获得了搜索方向，并采用了类似于 Armijo 准则的方法来计算步长，虽然此方法在每一次迭代之后函数值都会下降，但是找到全局最优解的可能性降低。基于上述的考虑，本文提出了非单调的对角最速下降算法，此算法不需要函数值在每一次迭代之后都减少，在某些点函数值可以增加，提高了找到全局最优解的可能性。具体算法步骤如下：

步骤 0，选择初始点 $\boldsymbol{x}^0\in\mathbf{R}^n$，取 $\tau^0>0$，$C_j^0=F_j(\boldsymbol{x}^0)$，$j\in I$，$Q_0=1$，$\eta_{\min}$，$\eta_{\max}\in[0,1]$，令 $k=0$。

步骤 1，解决以下子问题寻找 λ_i^k，$i\in I$：

$$\min\ \Big\|\sum_{i\in I}\lambda_i\nabla F_i(\boldsymbol{x}_k)\Big\|^2/(2\tau^k),$$
$$\text{s.t.}\ \lambda_i\geqslant 0,i\in I,\sum_{i\in I}\lambda^i=1。$$

则可求得 $\boldsymbol{d}^k=-\Big[\sum_{i\in I}\lambda_i^k\nabla F_i(\boldsymbol{x}_k)\Big]/\tau^k$，$\theta(\boldsymbol{x}_k)=-\Big[\Big\|\sum_{i\in I}\lambda_i^k\nabla F_i(\boldsymbol{x}_k)\Big\|^2\Big]/\tau^k$。令

$$\boldsymbol{\gamma}_i^k:=\nabla F_i(\boldsymbol{x}_{k+1})-\nabla F_i(\boldsymbol{x}_k),\boldsymbol{s}^k:=\boldsymbol{x}_{k+1}-\boldsymbol{x}_k,\boldsymbol{u}^k:=\sum_{i\in I}\lambda_i^k\boldsymbol{\gamma}_i^k,$$

更新 $\tau^k=\max\{\tau_0,u_i^k/s_i^k\}$。

步骤 2，如果 $|\theta(\boldsymbol{x}_k)|\leqslant\varepsilon$，则算法终止；否则转步骤 3。

步骤 3，令 $\boldsymbol{x}_{k+1}:=\boldsymbol{x}_k+\alpha_k\boldsymbol{d}^k$，其中 α_k 满足非单调 Armijo 准则，h_k 是使得以下条件成立的最小非负整数：

$$F_j(\boldsymbol{x}_k+\alpha_k\boldsymbol{d}^k)\leqslant C_j^k+\sigma\alpha_k\theta(\boldsymbol{x}_k),j=1,\cdots,m。\tag{6}$$

选取 $\eta_k\in[\eta_{\min},\eta_{\max}]$，令 $Q_{k+1}=\eta_kQ_k+1$，有：

$$C_j^{k+1}=[\eta_kQ_kC_j^k+F_j(\boldsymbol{x}_{k+1})]/Q_{k+1}。\tag{7}$$

令 $k:=k+1$ 转步骤 1。

3 收敛性分析

引理 2 对于问题 1，由 τ^k 的更新迭代方式易知 $\tau^k\boldsymbol{I}$，$k=1,2,\cdots$是正定的，则有：1) 对所有的 $\boldsymbol{x}\in\mathbf{R}^n$，$\theta(\boldsymbol{x})\leqslant 0$。2) 下述 3 个表述是等价的：a) $\boldsymbol{x}$ 不是 Pareto 临界点；b) $\theta(\boldsymbol{x})<0$；c) $\boldsymbol{d}(\boldsymbol{x})\neq 0$。3) 函数 $\theta(\boldsymbol{x})$是连续的。

引理 3　点 $\boldsymbol{x}\in\mathbf{R}^n$ 是函数 $\boldsymbol{F}$ 的 Pareto 临界点当且仅当 $\theta(\boldsymbol{x})=0$,或者等价为 $\boldsymbol{d}(\boldsymbol{x})=0$。

引理 4　算法 1 的每一次迭代 k,都有 $F(\boldsymbol{x}_k)\leqslant\boldsymbol{C}^k\leqslant\boldsymbol{A}^k$。

引理 5　针对算法 1,如果 x_k 是函数 F 的非 Pareto 临界点,那么一定存在满足非单调 Armijo 条件的 α_k。

引理 6　假设采用了算法 1,$\nabla F_j(j\in I)$是 Lipschitz 连续的,即存在 L 是 Lipschitz 常数,使得对 $\boldsymbol{x}_k$ 和 $\boldsymbol{x}_k+(\alpha_k/\rho)\boldsymbol{d}(\boldsymbol{x}_k)$之间的任何数 $\boldsymbol{x}$,都有$\|\nabla F_j(\boldsymbol{x})-\nabla F_j(\boldsymbol{x}_k)\|\leqslant L\|\boldsymbol{x}-\boldsymbol{x}_k\|$,其中 $\alpha_k\leqslant\rho$。如果 α_k 满足非单调的 Armijo 准则,则以下不等式成立:

$$\alpha_k\geqslant\min\{\rho,[2\rho(1-\tau)/L][|\theta(\boldsymbol{x}_k)|/\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2]\}。\tag{8}$$

证明　1) 如果 $\alpha_k\geqslant\rho$,则(8)式显然成立。

2) 如果 $\alpha_k<\rho$,因为 h_k 是满足算法 1 中的最小非负整数,当 $\alpha_k=\rho^{h_k}$ 时,则由引理 4 有 $F(\boldsymbol{x}_k)\leqslant\boldsymbol{C}^k$,且:

$$F_j[\boldsymbol{x}_k+\alpha_k\boldsymbol{d}(\boldsymbol{x}_k)/\rho]>C_j^k+\tau\alpha_k\theta(\boldsymbol{x}_k)/\rho>F_j(\boldsymbol{x}_k)+\tau\alpha_k\theta(\boldsymbol{x}_k)/\rho,\ j=1,2,\cdots,m。\tag{9}$$

因为$\nabla F_j(j\in I)$是 Lipschitz 连续的,可以得到:

$$F_j[\boldsymbol{x}_k+(\alpha_k/\rho)\boldsymbol{d}(\boldsymbol{x}_k)]-F_j(\boldsymbol{x}_k)=(\alpha_k/\rho)\nabla F_j(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{d}(\boldsymbol{x}_k)+\int_0^{\alpha_k/\rho}[\nabla F_j(\boldsymbol{x}_k+t\boldsymbol{d}(\boldsymbol{x}_k))-\nabla F_j(\boldsymbol{x}_k)]^{\mathrm{T}}\mathrm{d}t\leqslant$$
$$(\alpha_k/\rho)\theta(\boldsymbol{x}_k)+L\alpha_k^2\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2/(2\rho^2),\ j=1,2,\cdots,m。$$

结合上式和(9)式,有 $L\alpha_k\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2/2\geqslant(1-\tau)\rho|\theta(\boldsymbol{x}_k)|$。

综上所述,不等式(8)成立。　　证毕

定理 1　(Pareto 弱有效解的收敛性)假设 $F_j(\boldsymbol{x})(j\in I)$有下界,则存在 $c_1>0$ 使得:

$$|\theta(\boldsymbol{x}_k)|\geqslant c_1\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2,\forall k=1,2,3\cdots,\tag{10}$$

由引理 6 知,算法 1 产生的任何序列$\{\boldsymbol{x}_k\}$的极限点都是 $\boldsymbol{F}$ 的 Pareto 弱有效解。

证明　首先,对所有的 $j\in I$,令 $\beta=\min\{\tau\rho,[2\tau\rho(1-\tau)c_1]/L\}$,可以有:

$$F_j(\boldsymbol{x}_{k+1})\leqslant C_j^k-\beta|\theta(\boldsymbol{x}_k)|。\tag{11}$$

1) 使用非单调的 Armijo 准则且 $\alpha_k<\rho$,有:

$$F_j(\boldsymbol{x}_{k+1})\leqslant C_j^k+\tau\alpha_k\theta(\boldsymbol{x}_k)=C_j^k-\tau\rho|\theta(\boldsymbol{x}_k)|,\ j=1,2,\cdots,m;$$

2) 使用非单调的 Armijo 准则且 $\alpha_k\geqslant\rho$,由(8)式有:

$$\alpha_k\geqslant[2\rho(1-\tau)/L]/[|\theta(\boldsymbol{x})|/\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2]。$$

根据(6)式可得:

$$F_j(\boldsymbol{x}_{k+1})\leqslant C_j^k-[2\tau\rho(1-\tau)/L][|\theta(\boldsymbol{x}_k)|^2/\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2],j\in I;$$

最后,由(10)式有:

$$F_j(\boldsymbol{x}_{k+1})\leqslant C_j^k-[2\tau\rho(1-\tau)c_1]|\theta(\boldsymbol{x}_k)|/L,\ j=1,2,\cdots,m。$$

综上所述,(11)式成立。

结合(7)式和(11)式,有:$\boldsymbol{C}^{k+1}\leqslant[\eta_k\theta_k\boldsymbol{C}^k+\boldsymbol{C}^k-\beta|\theta(\boldsymbol{x}_k)|\boldsymbol{e}]/\theta_{k+1}=\boldsymbol{C}^k-[\beta|\theta(\boldsymbol{x}_k)|\boldsymbol{e}]/\theta_{k+1}$,其中 $\boldsymbol{e}=(1,\cdots,1)^{\mathrm{T}}\in\mathbf{R}^n$,因为 F_j 是有下界的,$j=1,2,\cdots,m$,且由引理 5 有 $\boldsymbol{F}(\boldsymbol{x}_k)\leqslant\boldsymbol{C}^k$,对所有的 k,可以得出 C_j^k 是有下界的,综上可得:

$$\sum_{k=0}^{+\infty}[|\theta(\boldsymbol{x}_k)|/\theta(\boldsymbol{x}_{k+1})]<\infty。\tag{16}$$

假设由算法 1 所产生的序列$\{\boldsymbol{x}_k\}$的极限点为 $\boldsymbol{x}^*$,则假设$\{\boldsymbol{x}_k\}$的子列$\{\boldsymbol{x}_k\}_{k\in K}$收敛到 $\boldsymbol{x}^*$,可得 $\theta(\boldsymbol{x}^*)=0$。否则,假设 $\theta(\boldsymbol{x}^*)<0$,则存在 $\varepsilon>0,\delta_0>0$,对所有的 $0<\delta<\delta_0$ 和所有的 $k\in K$,当$\|\boldsymbol{x}_k-\boldsymbol{x}^*\|\leqslant\delta$ 时有$|\theta(\boldsymbol{x}_k)|\geqslant\varepsilon>0$。这意味着:

$$\sum_{k=0}^{\infty}[|\theta(\boldsymbol{x}_k)|/Q_{k+1}]\geqslant\sum_{k\in\{k\in K|\|\boldsymbol{x}_k-\boldsymbol{x}^*\|<\delta\}}(\varepsilon/Q_{k+1})。\tag{13}$$

通过(7)式,再结合 $Q_0=1$ 和 $\eta_k\in[0,1]$,对每一个 k 都有:

$$Q_{k+1}=1+\sum_{i=0}^{k}\prod_{l=0}^{i}\eta_{k-l}\leqslant k+2。\tag{14}$$

结合(14)式和 $\eta_{\max}<1$ 有：$Q_{k+1}=1+\sum_{i=0}^{k}\prod_{l=0}^{i}\eta_{k=l}\leqslant 1+\sum_{i=0}^{k}\eta_{\max}^{i+1}\leqslant\sum_{i=0}^{+\infty}\eta_{\max}^{i}=1/(1-\eta_{\max})$。所以由(13)式有：

$$\sum_{k=0}^{+\infty}[|\theta(\boldsymbol{x}_k)|/Q_{k+1}]\geqslant\sum_{k\in\{k\in K|\|\boldsymbol{x}_k-\boldsymbol{x}^*\|\leqslant\delta\}}(\varepsilon/Q_{k+1})\geqslant\sum_{k\in\{k\in K|\|\boldsymbol{x}_k-\boldsymbol{x}^*\|\leqslant\delta\}}(1-\eta_{\max})\varepsilon=+\infty,$$

与(12)式矛盾。

由引理2知 $\theta(\boldsymbol{x}^*)=0$，故 $\boldsymbol{x}^*$ 是 $\boldsymbol{F}(\boldsymbol{x})$ 的 Pareto 弱有效解。 证毕

假设1 函数 $\nabla^2F_j(j=1,2,\cdots,m)$ 是一致连续的，对任意的 $\varepsilon>0$，存在 $\delta>0$，对任意的 $\boldsymbol{x},\boldsymbol{y}\in\mathbf{R}^n$，当 $\|\boldsymbol{y}-\boldsymbol{x}\|<\delta$ 时，有 $\|\nabla f_j^2(\boldsymbol{y})-\nabla f_j^2(\boldsymbol{x})\|<\varepsilon/2,j=1,2,\cdots,m$。

假设2 对任意 $\varepsilon>0$，存在 $K_0\in\mathbf{N}$，对任意的 $k\geqslant K_0$，有：

$$[\|(\nabla^2F_j(\boldsymbol{x}_k)-\tau^k\boldsymbol{I})\boldsymbol{s}_k\|/\|\boldsymbol{s}_k\|]<\varepsilon/2,j=1,2,\cdots,m。\tag{15}$$

其中：$\{x_k\}$ 是由算法1产生的，$\{\tau^k\boldsymbol{I}\}$ 是由 τ 的更新公式产生的。

定理2 若假设1和假设2成立，$\{\boldsymbol{x}_k\}$ 是由算法1所产生的序列，收敛到 $\boldsymbol{x}^*$，$\nabla F_j(\boldsymbol{x}^*)(j=1,2,\cdots,m)$ 是正定的，假设 $\sum_{k=1}^{+\infty}\|\boldsymbol{x}_k-\boldsymbol{x}^*\|<\infty$，Hessian 矩阵 $\nabla^2F_j(\boldsymbol{x})$ 在 $\boldsymbol{x}^*$ 这点是 Lipschitz 连续的，$j=1,2,\cdots,m$，即对任意的 $\|\boldsymbol{x}-\boldsymbol{x}^*\|<\varepsilon$，$L_j$ 是正常数，有 $\|\nabla^2F_j(\boldsymbol{x})-\nabla^2F_j(\boldsymbol{x}^*)\|\leqslant L_j\|\boldsymbol{x}-\boldsymbol{x}^*\|$，则有：

$$\lim_{k\to+\infty}[\|(\nabla^2F_j(\boldsymbol{x}_k)-\tau^k\boldsymbol{I})\boldsymbol{s}_k\|/\|\boldsymbol{s}_k\|]=0,\ j=1,2,\cdots,m。$$

引理7 假设 $\{\boldsymbol{x}_k\}$ 由算法1产生，$a>0$ 且为标量，对所有的 $j=1,2,\cdots,m$ 和 k，有 $\tau^k\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}\geqslant a$，$\forall\boldsymbol{z}\in\mathbf{R}^n$，$\|\boldsymbol{z}\|=1$。或者等价地有 $\tau^k\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}\geqslant a\|\boldsymbol{z}\|^2$，$\forall\boldsymbol{z}\in\mathbf{R}^n$。则有：

1) $|\theta(\boldsymbol{x}_k)|\geqslant(a/2)\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2$；

2) $|\theta(\boldsymbol{x}_k)|\leqslant(1/2a)\left\|\sum_{j=1}^{m}\lambda_j\nabla F_j(\boldsymbol{x}_k)\right\|^2$，$\forall\lambda_j\geqslant0,j=1,2,\cdots,m,\sum_{j=1}^{m}\lambda_j=1$。

命题1 假设 $\{\boldsymbol{x}_k\}$ 是由算法1产生的有界序列，并且它的极限点是函数 $\boldsymbol{F}$ 的 Pareto 临界点，假设标量 $a>0$，对所有的 k 和 $j=1,2,\cdots,m$，有 $\tau^k\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}\geqslant a\|\boldsymbol{z}\|^2$，$\forall\boldsymbol{z}\in\mathbf{R}^n$。更进一步，有 $\lim_{k\to+\infty}\theta(\boldsymbol{x}_k)=0$，$\lim_{k\to+\infty}\|\boldsymbol{d}(\boldsymbol{x}_k)\|=0$。

定理3 (局部超线性收敛性的充分条件)设 $\{\boldsymbol{x}_k\}$ 是由算法1所产生的有界序列，$\{\boldsymbol{x}_k\}$ 的极限点 $\boldsymbol{x}^*$ 是 $\boldsymbol{F}$ 的 Pareto 临界点，$a>0$ 且 a 是一个标量，故对任意的 k 和 $j=1,2,\cdots,m$，有：

$$\tau\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}\geqslant a\|\boldsymbol{z}\|^2,\ \forall\boldsymbol{z}\in\mathbf{R}^n。$$

且假设1和假设2成立，则当 $\alpha_k=1$ 且 k 充分大时，序列 $\{\boldsymbol{x}_k\}$ 是局部超线性收敛到 Pareto 临界点 $\boldsymbol{x}^*$。

证明 首先由命题1知，存在充分大的 $k_1\in K$，对所有的 $k\geqslant K_1$，存在 $\delta>0$ 和 $\varepsilon>0$，这里取 $\varepsilon/a<1$，可得：

$$\|\boldsymbol{d}(\boldsymbol{x}_k)\|<\min\{\delta,\varepsilon\}[1-(\varepsilon/a)],\tag{16}$$

其次，存在充分大的 $K_2\in\mathbf{N}$，对任意的 $k>K_2$，令 $\alpha_k=1$，有：

$$F_j(\boldsymbol{x}_k+\boldsymbol{d}(\boldsymbol{x}_k))=F_j(\boldsymbol{x}_k)+\nabla F_j(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{d}(\boldsymbol{x}_k)+\tau\boldsymbol{d}(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{d}(\boldsymbol{x}_k)/2+o(\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2),$$

对所有的 $j=1,\cdots,m$，结合引理7和 $\theta(x)$ 的定义有：

$$F_j(\boldsymbol{x}_k+\boldsymbol{d}(\boldsymbol{x}_k))\leqslant C_j^k+\tau\theta(\boldsymbol{x}_k)+(1-\tau)\theta(\boldsymbol{x}_k)+o(\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2),$$

通过化简可以得到如下不等式：

$$F_j(\boldsymbol{x}_k+\boldsymbol{d}(\boldsymbol{x}_k))\leqslant C_j^k+\tau\theta(\boldsymbol{x}_k)+\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2[-a(1-\tau)/2+o(\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2)/\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2],\tag{17}$$

因为 $F_j(\boldsymbol{x})$ 是连续可微的，所以由泰勒定理有：

$$F_j(\boldsymbol{x}_k+\alpha\boldsymbol{d}(\boldsymbol{x}_k))=F_j(\boldsymbol{x}_k)+\alpha\nabla F_j(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{d}(\boldsymbol{x}_k)+o_j(\alpha),j=1,2,\cdots,m。\tag{18}$$

结合(18)式可得，存在充分大的 $k_2\in\mathbf{N}$，对任意的 $k\geqslant K_2$，有 $-a(1-\tau)/2+o(\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2)/\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2\leqslant0$。则由(17)式，对任意的 $k\geqslant K_2$，有 $F_j(\boldsymbol{x}_k+\boldsymbol{d}(\boldsymbol{x}_k))\leqslant C_j^k+\tau\theta(\boldsymbol{x}_k),j=1,2,\cdots,m$。

因此，当 $\alpha_k=1$ 时，非单调的 Armijo 准则是成立的。

由假设2可得，对所有的 $\varepsilon>0$，存在充分大的 $K_3\in\mathbf{N}$，对所有 $k\geqslant K_3$，有(15)式成立。令 $K_0=\max\{K_1,K_2,K_3\}$，对所有的 $k\geqslant K_0$，令 $\alpha_k=1$，对任意 $k\geqslant K_0$，有 $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\boldsymbol{d}(\boldsymbol{x}_k)$。定义 $\boldsymbol{v}_{k+1}:=\sum_{j=1}^{m}\lambda_j(\boldsymbol{x}_k)\nabla F_j(\boldsymbol{x}_{k+1})$，其

中：$\lambda_j(\boldsymbol{x}_k)(j=1,2,\cdots,m)$是拉格朗日乘子，并且$\lambda_j(\boldsymbol{x}_k)\geqslant 0$，$\sum_{j=1}^{m}\lambda_j(\boldsymbol{x}_k)=1$，$j=1,2,\cdots,m$。

因此由引理 7 的 2)，有：

$$|\theta(\boldsymbol{x}_{k+1})|\leqslant\|\boldsymbol{v}_{k+1}\|^2/(2a),\tag{19}$$

定义：

$$G_k(\boldsymbol{x}):=\sum_{j=1}^{m}\lambda_j(\boldsymbol{x}_k)F_j(\boldsymbol{x}),\boldsymbol{v}_{k+1}=\nabla G_k(\boldsymbol{x}_{k+1}),\tag{20}$$

由$\boldsymbol{d}^k$的定义有$\boldsymbol{d}(\boldsymbol{x}_k)=-\nabla G_k(\boldsymbol{x}_k)/\tau^k$。由假设 1，对任意$\varepsilon>0$，存在$\delta>0$，使得对所有的$\boldsymbol{x},\boldsymbol{y}\in\mathbf{R}^n$，当$\|\boldsymbol{y}-\boldsymbol{x}\|<\delta$时，有：

$$\|\nabla^2 G_k(\boldsymbol{y})-\nabla^2 G_k(\boldsymbol{x})\|<\varepsilon/2,\tag{21}$$

并由假设 2，对任意的$\varepsilon>0$，$k\geqslant K_0$有：

$$\|(\tau^k\boldsymbol{I}-\nabla^2 G_k(\boldsymbol{x}_k))\boldsymbol{d}(\boldsymbol{x}_k)\|<\varepsilon\|\boldsymbol{d}(\boldsymbol{x}_k)\|/2,\tag{22}$$

结合(21)、(22)式和泰勒定理，对任意的$k\geqslant K_0$有：

$$\begin{gathered}\|\nabla G(\boldsymbol{x}_k+\boldsymbol{d}(\boldsymbol{x}_k))-(\nabla G(\boldsymbol{x}_k)+\tau^k\boldsymbol{I}\boldsymbol{d}(\boldsymbol{x}_k))\|\leqslant\\ \|\nabla G(\boldsymbol{x}_k+\boldsymbol{d}(\boldsymbol{x}_k))-\nabla G(\boldsymbol{x}_k)-\nabla^2 G(\boldsymbol{x}_k)\boldsymbol{d}(\boldsymbol{x}_k)\|+\|(\nabla^2 G_k(\boldsymbol{x}_k)-\tau^k\boldsymbol{I})\boldsymbol{d}(\boldsymbol{x}_k)\|\leqslant\\ \int_0^1\|\nabla^2 G_k(\boldsymbol{x}_k+t\boldsymbol{d}(\boldsymbol{x}_k))-\nabla^2 G_k(\boldsymbol{x}_k)\|\|\boldsymbol{d}(\boldsymbol{x}_k)\|\boldsymbol{d}t+\|(\nabla^2 G(\boldsymbol{x}_k)-\tau^k\boldsymbol{I})\boldsymbol{d}(\boldsymbol{x}_k)\|<\varepsilon\|\boldsymbol{d}(\boldsymbol{x}_k)\|。\end{gathered}\tag{23}$$

当所有的$k\geqslant K_0$时，有$\|\boldsymbol{d}(\boldsymbol{x}_k)\|<\delta$。

通过$\boldsymbol{d}(\boldsymbol{x})$的定义，可以得到$\nabla G_k(\boldsymbol{x}_k)+\tau^k\boldsymbol{d}(\boldsymbol{x}_k)=0$。结合(20)和(23)式可得：

$$\|\boldsymbol{v}_{k+1}\|<\varepsilon\|\boldsymbol{d}(\boldsymbol{x}_k)\|,\tag{24}$$

结合(19)和(24)式，有$|\theta(\boldsymbol{x}_{k+1})|<\varepsilon^2\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2/(2a)$。利用引理 7 的 1)，可得$a\|\boldsymbol{d}(\boldsymbol{x}_{k+1})\|^2/2<\varepsilon^2\|\boldsymbol{d}(\boldsymbol{x}_k)\|^2/(2a)$。因此，有$\|\boldsymbol{d}(\boldsymbol{x}_{k+1})\|<\varepsilon\|\boldsymbol{d}(\boldsymbol{x}_k)\|/a$，故有：

$$\|\boldsymbol{d}(\boldsymbol{x}_{k+1})\|\leqslant(\varepsilon/a)^{k+1-K_0}\|\boldsymbol{d}(\boldsymbol{x})\|,\tag{25}$$

对所有的$k\in\mathbf{N}$和$l\in\mathbf{N}$，当$k>l\geqslant K_0$，有：

$$\|\boldsymbol{x}_k-\boldsymbol{x}_l\|\leqslant\sum_{i=1}^{k-1}\|\boldsymbol{x}_{i+1}-\boldsymbol{x}_i\|\leqslant\|\boldsymbol{d}(\boldsymbol{x}_{K_0})\|\sum_{i=0}^{+\infty}(\varepsilon/a)^i=\|\boldsymbol{d}(\boldsymbol{x}_{K_0})\|/(1-\varepsilon/a)。$$

由(16)式有$\|\boldsymbol{d}(\boldsymbol{x}_{K_0})\|/(1-\varepsilon/a)<[\varepsilon(1-\varepsilon/a)]/[(1-\varepsilon/a)]=\varepsilon$。

因此，对任意的$\varepsilon>0$，存在$K_0\in\mathbf{N}$，使得对所有的$k>l\geqslant K_0$，可以得到$\|\boldsymbol{x}_k-\boldsymbol{x}_l\|<\varepsilon$。所以$\{\boldsymbol{x}_k\}$是柯西序列，且存在$\boldsymbol{x}^*\in\mathbf{R}^n$，使得$\lim_{k\to+\infty}\boldsymbol{x}_k=\boldsymbol{x}^*$。由引理 1 的 3)且$\theta$连续可得$\lim_{k\to+\infty}\theta(\boldsymbol{x}_k)=\theta(\boldsymbol{x}^*)$，由命题 1 可知，$\theta(\boldsymbol{x}^*)=0$。因此$\boldsymbol{x}^*$是$F$的 Pareto 临界点。

选择任意的$\varphi>0$，定义$\bar{\varepsilon}:=a\varphi/(1+2\varphi)$。易知$\bar{\varepsilon}/a<1$，结合(25)式和三角不等式，存在$K_0\in\mathbf{N}$，使得对任意的$k\geqslant K_0$有$\|\boldsymbol{x}_k-\boldsymbol{x}_{K_0}\|\leqslant\{[1-(\bar{\varepsilon}/a)^{k-K_0}]/(1-\bar{\varepsilon}/a)\}\|\boldsymbol{d}(\boldsymbol{x}_{K_0})\|$。又对所有的$l\in\mathbf{N}$，当$l\geqslant K_0$，有$\|\boldsymbol{x}^*-\boldsymbol{x}^l\|\leqslant\|\boldsymbol{d}(\boldsymbol{x}_l)\|/(1-\bar{\varepsilon}/a)$。利用此式和(25)式可得：

$$\|\boldsymbol{x}^*-\boldsymbol{x}_{l+1}\|\leqslant\|\boldsymbol{d}(\boldsymbol{x}_{l+1})\|/(1-\bar{\varepsilon}/a)\leqslant[\bar{\varepsilon}\|\boldsymbol{d}(\boldsymbol{x}_l)\|/a]/(1-\bar{\varepsilon}/a)。\tag{26}$$

通过三角不等式和(26)式，有：

$$\|\boldsymbol{x}^*-\boldsymbol{x}^l\|\geqslant\|\boldsymbol{d}(\boldsymbol{x}_l)\|-[(\bar{\varepsilon}/a)/(1-\bar{\varepsilon}/a)]\|\boldsymbol{d}(\boldsymbol{x}_l)\|=[(1-2\bar{\varepsilon}/a)/(1-\bar{\varepsilon}/a)]\|\boldsymbol{d}(\boldsymbol{x}_l)\|。\tag{27}$$

结合(26)、(27)式可得$\|\boldsymbol{x}^*-\boldsymbol{x}_{l+1}\|\leqslant[(\bar{\varepsilon}/a)/(1-2\bar{\varepsilon}/a)]\|\boldsymbol{x}^*-\boldsymbol{x}_l\|$，因此，$\|\boldsymbol{x}^*-\boldsymbol{x}_{l+1}\|\leqslant\varphi\|\boldsymbol{x}^*-\boldsymbol{x}_l\|$。

综上所述，因为$\varphi>0$且是任意的，那么序列$\{\boldsymbol{x}_k\}$局部超线性收敛到$\boldsymbol{x}^*$。　　证毕

4 数值实验

本文在参数为内存 16 GB CPU i5 2.11 GHz 华为笔记本电脑上进行数值实验，使用 Matlab 2019b 中的 FMINCON 和 QUADPROG 求解器进行求解，终止准则为$\|\theta^k\|\leqslant 10^{-2}$或者最大迭代步数为 500。实验结果见表 1。$n_f$表示函数均值，$m$表示问题的维度，$n$代表变量的维度，$L^T$和$U^T$表示初始变量的上下界，本文从初始

区间中随机选择 100 个初始点进行计算。最后从函数平均值和弱有效解的分布进行比较,DSD 代表对角最速下降算法,NDSD 代表非单调对角最速下降算法。DSD、NDSD 算法计算 JOS1(c)的 Pareto 前沿面见图 1、图 2。

表 1　DSD 和 NDSD 对比情况

Tab. 1　Comparison of DSD and NDSD

名称	来源	m	n	L^T	U^T	算法	n_f	t
PNR	31	2	2	$-[2,2]$	$[2,2]$	DSD	9.94	0.022 4
						NDSD	9.91	0.036 2
SP	32	2	2	$-[2,2]$	$[2,2]$	DSD	2.04	0.025 8
						NDSD	1.95	0.018 9
SD	27	2	4	$[1,\sqrt{2},\sqrt{2},1]$	$[3,3+\sqrt{2},3+\sqrt{2},3]$	DSD	9.16	0.095 4
						NDSD	9.13	0.133 1
Binh(1)	33	2	2	$-[5,5]$	$[10,10]$	DSD	24.88	0.027 2
						NDSD	24.10	0.050 2
JOS1(a)	14	2	10	$[0,0,\cdots,0]$	$[1,1,\cdots,1]$	DSD	2.01	0.025 8
						NDSD	1.98	0.048 1
JOS1(b)	14	2	30	$[0,0,\cdots,0]$	$[1,1,\cdots,1]$	DSD	2.01	0.037 4
						NDSD	1.67	0.041 8
JOS1(c)	14	2	50	$[0,0,\cdots,0]$	$[1,1,\cdots,1]$	DSD	1.98	0.028 1
						NDSD	1.88	0.023 9
JOS1(d)	14	2	50	$-[10,\cdots,10]$	$[10,\cdots,10]$	DSD	1.85	0.064 7
						NDSD	1.77	0.049 4
JOS1(f)	14	2	100	$-[50,\cdots,50]$	$[50,\cdots,50]$	DSD	2.00	0.079 9
						NDSD	1.55	0.091 2
JOS1(g)	14	2	100	$-[100,\cdots,100]$	$[100,\cdots,100]$	DSD	2.00	0.156 5
						NDSD	1.68	0.201 8

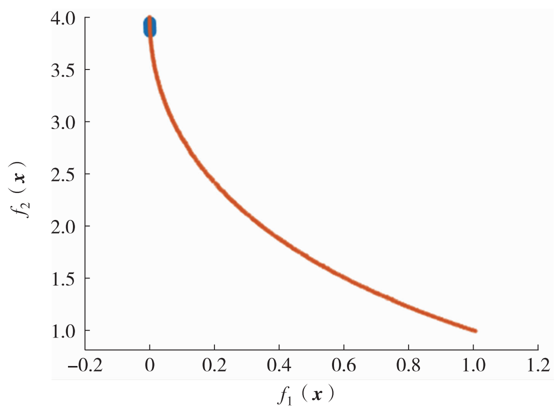

图 1　DSD 算法计算 JOS1(c)的 Pareto 前沿面

Fig. 1　The DSD algorithm calculates the Pareto front of JOS1(c)

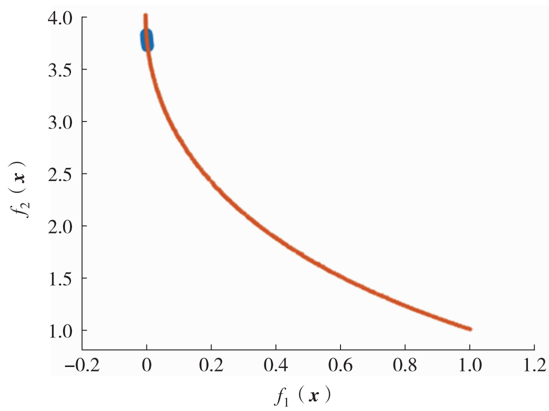

图 2　NDSD 算法计算 JOS1(c)的 Pareto 前沿面

Fig. 2　The NDSD algorithm calculates the Pareto front of JOS1(c)

由图 1、图 2 可知：两类算法产生的最优解和最优解区间不同，但是最终的函数值都在 Pareto 前沿面上，并且当目标函数的维数较低或者自变量的维度较高时，本文算法产生的函数值平均值是明显优于 DSD 算法。

参考文献：

[1] EHRGOTT M，WIECEK M. Mutiobjective programming[M]//FIGUFIRA J，GRECO S，EHROGOTT M. Multiple criteria decision analysis. New York：Springer，2005：667-722.

[2] MARLER R T，ARORA J S. Survey of multi-objective optimization methods for engineering[J]. Structural Multidiplinary Optimization，2004，26(6)：369-395.

[3] ZOPOUNIDIS C，PARDALOS P M. Handbook of multicriteria analysis[M]. Berlin：Springer，2010.

[4] CHINCHULUUN A，PARDALOS P M. A survey of recent developments in multiobjective optimization[J]. Annals of Operations Research，2007，154(1)：29-50.

[5] YU P L. Multiple-criteria decision making：concepts，techniques，and extension[M]. New York：Plenum Press，1985.

[6] ZHENG X Y，YANG X Q. The structure of weak Pareto solution sets in piecewise linear multiobjective optimization in normed spaces[J]. Science China Mathematics，2008，51(7)：1243-1256.

[7] GEOFFROY A M. Proper efficiency and the theory of vector maximization[J]. Journal of Mathematical Analysis and Applications，1968，22(3)：618-630.

[8] DRUMMOND L G，MACULAN N，SVAITER B F. On the choice of parameters for the weighting method in vector optimization [J]. Mathematical Programming，2008，111(1/2)：201-216.

[9] FLIEGE J，DRUMMOND L G，SVAITER B F. Newton's method for multiobjective optimization[J]. SIAM Journal on Optimization，2009，20(2)：602-626.

[10] MIETTINEN K. Nonlinear multiobjective optimization[M]. Berlin：Springer，2012.

[11] EHRGOTT M. Multicriteria optimization[M]. Berlin：Springer，2013.

[12] FLIEGE J，SVAITER B F. Steepest descent methods for multicriteria optimization[J]. Mathematical Methods of Operations Research，2000，51(3)：479-494.

[13] DRUMMOND L G，SVAITER B F. A steepest descent method for vector optimization[J]. Journal of Computational and Applied Mathematics，2005，175(2)：395-414.

[14] QU S J，GOH M，CHAN F T F. Quasi-newton methods for solving multiobjective optimization[J]. Operations Research Letters，2011，39(5)：397-399.

[15] MUSTAPHA E M，ABDELKRIM E M. Accelerated diagonal steepest descent method for unconstrained multiobjective optimization[J]. Journal of Optimization Theory and Applications，2021，188：220-242.

[16] DAI Y H. On the nonmonotone line search[J]. Journal of Optimization Theory and Applications，2002，112(2)：315-330.

[17] TOINT P L. A nassessment of nonmonotone line search techniques for unconstrained optimization[J]. SIAM Journal on Scientific Computing，1996，17：725-739.

[18] DAI Y. A nonmonotone conjugate gradient algorithm for unconstrained optimization[J]. Journal of Systems Science & Omplexity，2002，15：139-145.

[19] GRIPPO L，LAMPARIELLO F，LUCIDI S. A truncated Newton method with nonmonotone line search for unconstrained optimization[J]. Journal of Optimization and Applications，1989，60：401-419.

[20] LUCIDI S，ROCHETICH F，ROMA M. Curvilinear stabilization techniques for truncated Newton methods in large scale unconstrained optimization[J]. SIAM Journal on Optimization，1998，8：916-939.

[21] PANIER E R，TITS A L. Avoiding the Maratos effect by means of a nonmonotone line search[J]. SIAM Journal on Optimization，1991，28：1183-1195.

[22] RAYDAN M. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem[J]. SIAM Journal on Optimization，1997，7：26-33.

[23] ZHANG H，HAGER W W. A nonmonotone line search technique and its application to uncon-strained optimization[J]. SIAM Journal on Optimization，2004，14：1043-1056.

[24] MITA K，FUKUDA E H，YAMASHITA N. Nonmonotone line searches for unconstrained multi-objective optimization problems[J]. Journal of Global Optimization，2019，75：63-90.

[25] FAZZIO N S，SCHUVERDT M L. Convergence analysis of a nonmonotone projected gradient method for multiobjective

optimization problems[J]. Optimization Letters, 2019, 13: 1365-1379.

[26] QU S, JI Y, J JIANG J. Nonmonotone gradient methods for vector optimization with a portfolio optimization application[J]. European Journal of Operational Research, 2017, 236(2): 356-366.

[27] MAHDAVI-AMIRI N, SADAGHIANI F S. A superlinearly convergent nonmonotone quasi-Newton method for unconstrained multiobjective optimization[J]. Optimization Methods and Software, 2020, 6: 1223-1247.

Operations Research and Cybernetics

Nonmonotonic Diagonal Steepest Descent Algorithm for Multi-Objective Optimization

YANG Chunrong, TAN Yulin, ZHAO Kequan

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

Abstract: [**Purposes**]For solving multiobjective optimization problems more efficiently, a more effective Pareto frontier is obtained. [**Methods**]By introducing the nonmonotone Armijo criterion, a new step-size search method is obtained, and then a nonmonotone diagonal steepest descent algorithm for multi-objective optimization problems is proposed. [**Findings**]Under the assumptions of non-convexity of the objective function, gradient Lipschitz continuity and lower boundedness, it is proved that each accumulation point of the sequence generated by the algorithm is a Pareto weak efficient solution of the multiobjective optimization problem, and the sublinear convergence of the algorithm is proved under appropriate conditions. [**Conclusions**]Numerical experiments show that the average value of the objective function value of the proposed algorithm is smaller.

Keywords: multi-objective optimization; nonmonotone line search; diagonal steepest descent algorithm; Pareto weakly efficient solution

（责任编辑　黄　颖）