2013 年 11 月 计 算 数 学 第 35 卷第 4 期
Nov., 2013 MATHEMATICA NUMERICA SINICA Vol.35, No.4

无约束极大极小问题的广义梯度投影算法*1)

简金宝 2)
(广西大学数学与信息科学学院, 南宁 530004;
玉林师范学院数学与信息科学学院, 玉林 537000)

唐 菲 黎健玲 唐春明
(广西大学数学与信息科学学院, 南宁 530004)

摘 要

本文讨论 $\mathbf{R}^n$ 空间上的无约束极大极小问题. 通过 $\mathbf{R}^{n+1}$ 空间上的广义梯度投影技术产生 $\mathbf{R}^n$ 上的下降搜索方向, 进而结合 Armijo 非精确线搜索建立了原问题 $\mathbf{R}^n$ 上的一个广义梯度投影型算法. 算法在仿射线性无关条件下, 具有全局收敛性和强收敛性. 文中对算法进行了初步的数值试验.

关键词: 无约束极大极小问题; 广义梯度投影算法; 全局收敛性; 强收敛性

MR (2000) 主题分类: 90C30, 90C47

1. 引 言

考虑求解如下无约束极大极小问题

$$\min_{x\in\mathbf{R}^n} F(x), \tag{1.1}$$

其中 $F(x)$=max$\{f_i(x), i\in I\}$, $I=\{1,\ 2,\ \cdots,\ m\}$, $x=(x_1,\ x_2,\ \cdots,\ x_n)^\top\in\mathbf{R}^n$, $f_i(x),\ i\in I$ 是定义在 R^n 上的光滑实值函数. 极大极小问题是一类有广泛应用背景的非光滑优化, 但其非光滑性使之理论与数值方法研究较光滑优化困难, 不能应用已知的光滑优化算法对其直接求解. 有关极大极小问题部分数值方法见文献 [1-7], 非光滑极大极小问题一种有效的处理技术是将其等价转化为约束光滑优化, 事实上, 在一定的条件下, 问题 (1.1) 与以下光滑约束优化是等价的 (就最优性条件而言)

$$\min_{(z,x)\in\mathbf{R}^{n+1}} z,\ \text{s.t. } f_i(x)-z\leqslant 0,\ i\in I. \tag{1.2}$$

而 $\mathbf{R}^{n+1}$ 空间上的光滑不等式约束优化 (1.2), 有许多有效的求解方法和技术, 其中以 Rosen 梯度投影法 [8] 为基础的 (广义) 梯度投影型算法是一类典型的方法. 其特点是搜索方向有简单的显式结构, 引起优化学者的青睐和关注 [9,10]. 但利用广义梯度投影思想设计极大极小问题的数值方法目前少见报道. 本文旨在借鉴 [11 第 2 章] 求解一般不等式约束光滑优化广义投影的基本思想, 以 $\mathbf{R}^{n+1}$ 上的问题 (1.2) 为桥梁, 构建问题 (1.1) 的一个广义投影型算法. 其

* 2013 年 2 月 14 日收到.
1) 基金项目: 国家自然科学基金 (11271086, 11171250), 广西自然科学基金 (2011GXNSFD018022, 2013GXNSFAA019013) 和广西高校人才小高地建设创新团队资助计划.
2) 通讯作者: 简金宝, E-mail: jianjb@gxu.edu.cn.

成功之处在于算法迭代只在 $\mathbf{R}^n$ 上进行, 与辅助变量 z 无直接联系, 尽管投影是在 $\mathbf{R}^{n+1}$ 上执行. 文中除对算法进行详细的理论分析外, 还进行了初步的数值试验.

2. 算法构造

对于问题 (1.1), 其稳定点 (KKT 点) 最优性条件为

$$\sum_{i\in I}\mu_i\nabla f_i(x)=0,\ \sum_{i\in I}\mu_i=1,\ \mu_i\geqslant 0,\ \mu_i(f_i(x)-F(x))=0,\ i\in I. \tag{2.1}$$

为便于讨论, 对于 $x^k\in R^n$ 及参数 $\delta_k\geqslant 0$, 基于问题 (1.1) 和 (1.2) 引入如下记号:

$$I(x^k)=\{i\in I\mid f_i(x^k)=F(x^k)\},\ I_k=\{i\in I\mid -\delta_k\leqslant f_i(x^k)-F(x^k)\}, \tag{2.2}$$

$$N_k=\begin{pmatrix}A_k\\-e_k\end{pmatrix},\ A_k=(\nabla f_i(x^k),i\in I_k),\ e_k=(1,\ 1,\ \cdots,\ 1)\in\mathbf{R}^{|I_k|}, \tag{2.3}$$

假设 A. 函数 $f_i,\ i\in I$ 均一阶连续可微, 且对于任意 $x\in R^n$, 向量组 $\{\binom{\nabla f_i(x)}{-1},\ i\in I(x)\}$ 线性无关.

假设 A 等价于: 存在 $i_x\in I(x)$, 使得 $\{\nabla f_i(x)-\nabla f_{i_x}(x),\ i\in I(x)\setminus\{i_x\}\}$ 线性无关, 即 $\{\nabla f_i(x),\ i\in I(x)\}$ 仿射线性无关.

为检验当前迭代点 x^k 是否满足最优性条件 (2.1), 对于参数 $p>0$, 定义以下各量:

$$D_{ki}=(F(x^k)-f_i(x^k))^p,\quad D_k=\mathtt{diag}(D_{ki},i\in I_k),\ Q_k=(N_k^\top N_k+D_k)^{-1}N_k^\top, \tag{2.4}$$

$$P_k=E_{n+1}-N_kQ_k,\ \mu^k=(\mu_i^k,\ i\in I_k)=-Q_ke^0,\ w_k=\sum_{i\in I_k}\mathtt{max}\{-\mu_i^k,\ \mu_i^kD_{ki}\}, \tag{2.5}$$

$$\rho_k=\frac{\|P_ke^0\|^2+w_k}{1+|e_k\mu^k|}, \tag{2.6}$$

其中 $e^0=\binom{0_{n\times 1}}{1}=\binom{0}{1}$, 有以下引理.

引理 1. 在假设 A 下, 有

(1) 矩阵 $(N_k^\top N_k+D_k)$ 正定, 且若 $x^k\to x^*,\ N_k\to N_*,\ D_k\to D_*$, 则极限矩阵 $(N_*^\top N_*+D_*)$ 也正定;

(2) $N_k^\top P_k=D_kQ_k,\quad N_k^\top Q_k^\top=E_{|I_k|}-D_k(N_k^\top N_k+D_k)^{-1}$;

(3) $e^{0\top}P(x^k)e^0=\|P_ke^0\|^2+\sum\limits_{i\in I_k}(\mu_i^k)^2D_{ki}$;

(4) $\rho_k=0$ 当且仅当 x^k 是问题 (1.1) 的稳定点.

证明. 首先, 在假设 A 下, 由 [11 定理 1.1.9] 可知 $(N_k^\top N_k+D_k)$ 正定. 其次, 参照文 [11 引理 2.2.2] 的证明可证 $(N_*^\top N_*+D_*)$ 也正定. 结论 (2) 和 (3) 则是文 [11 定理 1.1.9] 结论 (2) 的具体表现. 下面详细证明结论 (4).

设 $\rho_k=0$, 由 (2.6) 式有 $P_ke^0=0,\quad w_k=0$. 由 $w_k=0$ 有 $\max\{-\mu_i^k,\ \mu_i^kD_{ki}\}=0$, 进而 $\mu_i^k\geqslant 0,\ \mu_i^kD_{ki}=0,\ \forall\ i\in I_k$. 于是, $\mu_i^k(F(x^k)-f_i(x^k))=0,\ i\in I_k$. 又由 (2.5) 及 $P_ke^0=0$, 有

$$P_ke^0=e^0+N_k\mu^k=\begin{pmatrix}0\\1\end{pmatrix}+\begin{pmatrix}\sum\limits_{i\in I_k}\mu_i^k\nabla f_i(x^k)\\-\sum\limits_{i\in I_k}\mu_i^k\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}.$$

因此, x^k 为问题 (1.1) 的稳定点.

反之, 如果 x^k 为问题 (1.1) 的稳定点, 则由 (2.1) 易知, 存在乘子 μ 使得

$$\sum_{i\in I_k}\mu_i\begin{pmatrix}\nabla f_i(x^k)\\-1\end{pmatrix}=-\begin{pmatrix}0\\1\end{pmatrix}=-e^0,\ \mu_iD_{ki}=0,\ \mu_i\geqslant 0,\ \forall\ i\in I_k.$$

进而 $N_k\mu=-e^0,\ N_k^\top N_k\mu=-N_k^\top e^0,\ D_k\mu=0$. 于是 $\mu=-(N_k^\top N_k+D_k)^{-1}N_k^\top e^0=\mu^k$. 进一步, 由 (2.5) 有 $w_k=0,\ P_ke^0=e^0+N_k\mu=0$. 从而 $\rho_k=0$.

下面构造搜索方向

$$\tilde{d}^k=\begin{pmatrix}d^k\\s_k\end{pmatrix}=\rho_k^{\xi}(-P_ke^0+Q_k^\top v^k),\tag{2.7}$$

其中常参数 $\xi\geqslant 0$, 以及

$$v^k=(v_i^k,i\in I_k)\ ,\ v_i^k=\begin{cases}-1-\rho_k, & \text{如果 } \mu_i^k<0;\\ D_{ki}-\rho_k, & \text{如果 } \mu_i^k\geqslant 0,\end{cases}\tag{2.8}$$

关于以上构造的 $\mathbf{R}^n$ 上的搜索方向 d^k, 具有以下下降性.

引理 2. 在假设 A 下, 有 $s_k\leqslant-\rho_k^{1+\xi},\ \nabla f_i(x^k)^\top d^k\leqslant-2\rho_k^{1+\xi},\ i\in I(x^k)$. 进而 $F'(x^k;d^k)\leqslant-2\rho_k^{1+\xi}$, 故 d^k 是问题 (1.1) 在点 x^k 处的下降方向, 其中 $F'(x^k;d^k)$ 表示函数 $F(x)$ 在点 x^k 处沿方向 d^k 的方向导数.

证明. 首先, 由 (2.7) 式及引理 1 之 (3), 有

$$\begin{aligned}s_k={e^0}^\top\tilde{d}^k&=\rho_k^{\xi}(-{e^0}^\top P_ke^0+(Q_ke^0)^\top v^k)=\rho_k^{\xi}(-{e^0}^\top P_ke^0-{\mu^k}^\top v^k)\\&=\rho_k^{\xi}(-\|P_ke^0\|^2-\sum_{i\in I_k}D_{ki}(\mu_i^k)^2-\sum_{i\in I_k}\mu_i^kv_i^k)\\&=\rho_k^{\xi}[-\|P_ke^0\|^2+\rho_k\sum_{i\in I_k}\mu_i^k-\sum_{i\in I_k}D_{ki}(\mu_i^k)^2-(\sum_{\mu_i^k<0}(-\mu_i^k)+\sum_{\mu_i^k\geqslant 0}D_{ki}\mu_i^k)]\\&\leqslant\rho_k^{\xi}(-\|P_ke^0\|^2+\rho_k|e^k\mu^k|-w_k)\\&=-\rho_k^{1+\xi}.\end{aligned}$$

另一方面, 由引理 1 之 (2) 可得,

$$N_k^\top\tilde{d}^k=\rho_k^{\xi}(-N_k^\top P_ke^0+N_k^\top Q_k^\top v^k)=\rho_k^{\xi}(D_k\mu^k+v^k-D_k(N_k^\top N_k+D_k)^{-1}v^k).\tag{2.9}$$

当 $i\in I(x^k)$ 时, $D_{ki}=0$, 由上式有,

$$(\nabla f_i(x^k)^\top,-1)\tilde{d}^k=\nabla f_i(x^k)^\top d^k-s_k=\rho_k^{\xi}v_i^k\leqslant-\rho_k^{1+\xi},\ i\in I(x^k).$$

从而, $\nabla f_i(x^k)^\top d^k\leqslant-\rho_k^{1+\xi}+s_k\leqslant-2\rho_k^{1+\xi},\ i\in I(x^k)$.

最后, 注意到 $F'(x^k;d^k)=\max\{\nabla f_i(x^k)^\top d^k,\ i\in I(x^k)\}$, 立知 $F'(x^k;d^k)\leqslant-2\rho_k^{1+\xi}$. 证毕.

基于以上分析, 下面给出算法具体步骤.

算法:

步骤 0(初始化). 选取初始点 $x^0\in\mathbf{R}^n$, 参数 $\alpha,\ \beta\in(0,1),\ p>0,\ \xi\geqslant 0,\ \delta_0\geqslant 0$, 置 $k:=0$.

步骤 1(计算搜索方向). 对于点 x^k, 根据 (2.2)–(2.6) 式计算 ρ_k, 如果 $\rho_k=0$, 则 x^k 为问题 (1.1) 的稳定点, 停止; 否则, 按 (2.7)–(2.8) 计算搜索方向 $\tilde{d}^k=\binom{d^k}{s_k}$, 转入步骤 2.

步骤 2(线搜索). 计算 $\{1,\ \beta,\ \beta^2,\ \beta^3,\ \cdots\}$ 中满足不等式

$$f_i(x^k+\lambda d^k)\leqslant F(x^k)-2\alpha\lambda\rho_k^{1+\xi},\ \forall\ i\in I \tag{2.10}$$

的最大值 λ_k.

步骤 3(更新). 令 $x^{k+1}=x^k+\lambda_k d^k$, 用合适的方法产生新的非负参数 δ_{k+1}. 置 $k:=k+1$, 返回步骤 1.

3. 收敛性分析

本节将分析以上构建算法的全局收敛性和强收敛性. 一旦算法有限步终止于 x^k, 则 $\rho_k=0$, 进而由引理 1 之 (4) 可知, x^k 是问题 (1.1) 的一个稳定点. 下面假设算法产生无穷点列 $\{x^k\}$, 往证其任意一个聚点 x^* 都是问题 (1.1) 的稳定点. 注意到指标集 I_k 的有限选取性, 存在无穷子列 $\mathcal{K}$, 使得 $x^k\xrightarrow{\mathcal{K}}x^*,\ I_k\equiv\bar{I},\ \forall\ k\in\mathcal{K}$. 记

$$D_{*i}=(F(x^*)-f_i(x^*))^p,\ D_*=\mathtt{diag}(D_{*i},\ i\in\bar{I}),\ Q_*=(N_*^\top N_*+D_*)^{-1}N_*^\top,$$

$$\mu^*=(\mu_i^*,\ i\in\bar{I})=-Q_*e^0,\ P_*=E_{n+1}-N_*Q_*,\ w_*=\sum_{i\in\bar{I}}\mathtt{max}\{-\mu_i^*,\ \mu_i^*D_{*i}\},\ \rho_*=\frac{\|P_*e^0\|^2+w_*}{1+|e\mu^*|}.$$

其中 $e=(1,1,\ \cdots,1)\in\mathbf{R}^{|\bar{I}|}$. 于是 $D_k\to D_*,\ \rho_k\to\rho_*,\ F(x^k)\to F(x^*),\ k\in\mathcal{K}$. 另外, 由引理 1 可知, 矩阵 $(N_*^\top N_*+D_*)$ 正定, 故序列 $\{v^k\}_\mathcal{K}$ 及 $\{\tilde{d}^k\}_\mathcal{K}$ 有界. 为保证算法的收敛性, 需对参数序列 $\{\delta_k\}$ 作一个基本要求.

假设 B. 存在正的常数 $\bar{\delta}$, 使得当 k 充分大时, 有 $\delta_k\geqslant\bar{\delta}$.

引理 3. 在假设 A 及 B 下, 如果 x^* 不是问题 (1.1) 的稳定点, 则 $\rho_*>0$, 进而当 $k\in\mathcal{K}$ 充分大时, $\rho_k\geqslant 0.5\rho_*$, 且 $\lambda_*=\inf\{\lambda_k,\ k\in\mathcal{K}\}>0$.

证明. 首先, 类似于引理 1 之 (4) 的分析, 易证当 x^* 不是问题 (1.1) 的稳定点时, $\rho_*>0$, 进而 $\rho_k\geqslant 0.5\rho_*$. 为证 $\lambda_*>0$, 只需证明对于 $\mathcal{K}$ 中充分大的 k 以及充分小的正数 λ(与 k 无关), 不等式 (2.10) 成立. 对 i 分如下两种情况分析.

对于 $i\notin I(x^*)$, 即 $f_i(x^*)<F(x^*)$, 注意到 $f_i(x)$ 的可微性及 $\{d^k\}_\mathcal{K}$ 的有界性, 有

$$f_i(x^k+\lambda d^k)-F(x^k)+2\alpha\lambda\rho_k^{1+\xi}=f_i(x^k)-F(x^k)+O(\lambda)\leqslant 0.5(f_i(x^*)-F(x^*))+O(\lambda)\leqslant 0.$$

对于 $i\in I(x^*)$, 即 $f_i(x^*)=F(x^*)$. 此时, 由 $\{d^k\}_\mathcal{K}$ 的有界性及 Taylor 展式有

$$\begin{aligned} f_i(x^k+\lambda d^k)-F(x^k)+2\alpha\lambda\rho_k^{1+\xi} &\leqslant f_i(x^k+\lambda d^k)-f_i(x^k)+2\alpha\lambda\rho_k^{1+\xi}\\ &=\lambda\nabla f_i(x^k)^\top d^k+2\alpha\lambda\rho_k^{1+\xi}+o(\lambda), \end{aligned}$$

另外, 注意到假设 B, 易知 $i\in I(x^*)$ 时, 有 $i\in I_k$, 进而 $D_{ki}=(F(x^k)-f_i(x^k))^p\to 0$. 于是由 (2.9) 及引理 1 之 (1), 有

$$\begin{pmatrix}\nabla f_i(x^k)\\ -1\end{pmatrix}^\top\tilde{d}^k=\nabla f_i(x^k)^\top d^k-s_k=\rho_k^\xi v_i^k+O(D_{ki}).$$

进而, 结合 (2.8) 式及 $s_k \leqslant -\rho_k^{1+\xi}$, 有 $\nabla f_i(x^k)^\top d^k \leqslant -2\rho_k^{1+\xi} + O(D_{ki})$. 因此,

$$\begin{aligned} f_i(x^k+\lambda d^k) - F(x^k) + 2\alpha\lambda\rho_k^{1+\xi} &\leqslant -2\lambda\rho_k^{1+\xi} + \lambda O(D_{ki}) + 2\alpha\lambda\rho_k^{1+\xi} + o(\lambda) \\ &= 2(\alpha-1)\lambda\rho_k^{1+\xi} + \lambda O(D_{ki}) + o(\lambda) \\ &\leqslant 2(0.5^{1+\xi})(\alpha-1)\lambda\rho_*^{1+\xi} + o(\lambda) \\ &\leqslant 0. \end{aligned}$$

综上所述可知, 对充分小的 $\lambda > 0$ 及充分大的 $k \in \mathcal{K}$, 不等式 (2.10) 恒成立, 证毕.

定理 1. 在假设 A 和假设 B 下, 算法或有限步终止于问题 (1.1) 的稳定点, 或产生无穷点列 $\{x^k\}$, 使得 $\{x^k\}$ 的任何聚点 x^* 都是问题 (1.1) 的稳定点, 即算法是全局收敛的.

证明. 不妨设算法产生无穷点列 $\{x^k\}$, x^* 为其一给定的聚点. 由 (2.10) 式引理 2 易知序列 $\{F(x^k)\}$ 单调下降. 又 $\lim\limits_{k\in\mathcal{K}} F(x^k) = F(x^*)$, 因此 $\lim\limits_{k\to\infty} F(x^k) = F(x^*)$. 若 x^* 不是问题 (1.1) 的稳定点, 则由引理 3 及 (2.10) 式, 有

$$F(x^*) = \lim_{k\in\mathcal{K}} F(x^{k+1}) \leqslant \lim_{k\in\mathcal{K}} (F(x^k) - 0.5^{1+\xi}\alpha\lambda_*\rho_*^{1+\xi}) = F(x^*) - 0.5^{1+\xi}\alpha\lambda_*\rho_*^{1+\xi},$$

此与 $\lambda_* > 0$ 和 $\rho_* > 0$ 矛盾. 证毕.

为进一步分析算法的强收敛性, 需要进一步假设:

假设 C. 设算法产生的迭代点列 $\{x^k\}$ 有界.

定理 2. 设假设 A, 假设 B 和假设 C 均成立, 且参数 $\xi > 0$, 有: (1) 存在常数 $c > 0$, 使得 $\|d^k\| \leqslant c\rho_k^{\xi}$; (2) $\lim\limits_{k\to\infty} \|x^{k+1} - x^k\| = 0$; (3) 如果 $\{x^k\}$ 有孤立聚点 x^*, 则 $\lim\limits_{k\to\infty} x^k = x^*$, 即算法是强收敛的.

证明. (1) 首先利用 $\{x^k\}$ 的有界性, 引理 1 之 (1) 及 (2.7) 和 (2.8) 式, 易知方向序列 $\{\tilde{d}^k\}$ 有界, 且存在 $c > 0$, 使得 $\|\tilde{d}^k\| \leqslant c\rho_k^{\xi}$, 从而 $\|d^k\| \leqslant c\rho_k^{\xi}$.

(2) 由于序列 $\{F(x^k)\}$ 单调有界, 故整列收敛, 于是由 (2.10) 易知 $\lim\limits_{k\to\infty} \lambda_k\rho_k^{1+\xi} = 0$. 因此, 注意到 $\xi > 0$, 有

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = \lim_{k\to\infty} \lambda_k\|d^k\| \leqslant \lim_{k\to\infty} c\lambda_k\rho_k^{\xi} = c\lim_{k\to\infty} [(\lambda_k\rho_k^{1+\xi})^{\xi}\lambda_k]^{\frac{1}{1+\xi}} = 0.$$

(3) 当 $\{x^k\}$ 有孤立聚点 x^* 时, 结合 $\lim\limits_{k\to\infty} \|x^{k+1} - x^k\| = 0$, 由 [11 推论 1.1.8] 立知 $\lim\limits_{k\to\infty} x^k = x^*$. 证毕.

4. 数值试验

本节通过在 MATLAB R2007b(win32) 上编程对如下的 7 个问题进行测试, 问题取自文 [1,2,12], 运行程序的计算机配置为 Windows 7, Intel(R)Pentium(R) CPU 2.00GHz, 内存为 2.00GB.

P1. $F(x) = \max\limits_{1\leqslant i\leqslant 3}\{f_i(x)\}$, 目标函数 $F(x)$ 中:

$$f_1(x) = x_1^2 + x_2^4,\ f_2(x) = (2-x_1)^2 + (2-x_2)^2,\quad f_3(x) = 2\mathtt{exp}(-x_1+x_2).$$

P2. $F(x)=\max\limits_{1\leqslant i\leqslant 3}\{f_i(x)\}$, 目标函数 $F(x)$ 中:

$$f_1(x)=x_1^4+x_2^2,\ f_2(x)=(2-x_1)^2+(2-x_2)^2,\quad f_3(x)=2\mathtt{exp}(-x_1+x_2).$$

P3. $F(x)=\max\limits_{1\leqslant i\leqslant 4}\{f_i(x)\}$, 目标函数 $F(x)$ 中:

$$\begin{aligned}
&f_1(x)=x_1^2+x_2^2+2x_3^2+x_4^2+5x_1-5x_2-21x_3+7x_4,\\
&f_2(x)=f_1(x)-10(-x_1^2-x_2^2-x_3^2-x_4^2-x_1+x_2-x_3+x_4+8),\\
&f_3(x)=f_1(x)-10(-x_1^2-2x_2^2-x_3^2-2x_4^2+x_1+x_4+10),\\
&f_4(x)=f_1(x)-10(-2x_1^2-x_2^2-x_3^2-2x_1+x_2+x_4+5).
\end{aligned}$$

P4. $F(x)=\max\limits_{1\leqslant i\leqslant 3}\{f_i(x)\}$, 目标函数 $F(x)$ 中:

$$f_1(x)=x_1^2+x_2^2+x_1x_2,\ f_2(x)=\mathtt{sin}(x_1),\quad f_3(x)=\mathtt{cos}(x_2).$$

P5. $F(x)=\max\limits_{1\leqslant i\leqslant 6}\{f_i(x)\}$, 目标函数 $F(x)$ 中:

$$\begin{aligned}
&f_1(x)=x_1^2+x_2^2+x_3^2-1,\ f_2(x)=x_1^2+x_2^2+(x_3-2)^2,\quad f_3(x)=x_1+x_2+x_3-1,\\
&f_4(x)=x_1+x_2-x_3+1,\quad f_5(x)=2x_1^3+6x_2^2+2(5x_3-x_1+1)^2,\quad f_6(x)=x_1^2-9x_3.
\end{aligned}$$

P6. $F(x)=\max\limits_{1\leqslant i\leqslant 3}\{f_i(x)\}$, 目标函数 $F(x)$ 中:

$$f_1(x)=\tfrac{1}{2}(x_1+\tfrac{10x_1}{x_1+0.1}+2x_2^2),\ f_2(x)=\tfrac{1}{2}(-x_1+\tfrac{10x_1}{x_1+0.1}+2x_2^2),\quad f_3(x)=\tfrac{1}{2}(x_1-\tfrac{10x_1}{x_1+0.1}+2x_2^2).$$

P7. $F(x)=\max\limits_{1\leqslant i\leqslant 4}\{f_i(x)\}$, 目标函数 $F(x)$ 中:

$$f_1(x)=x_1^2,\ f_2(x)=x_2^2,\quad f_3(x)=x_3^2,\quad f_4(x)=x_4^2.$$

算法中的参数取为 $\alpha=0.1,\ \beta=0.5,\ p=1,\ \xi=0.05,\ \delta=0.001$, 终止准则为 $\|\rho_k\|\leqslant 10^{-4}$.

为了深入观测本文算法 (记为算法 A) 的有效性, 在数值报告中还将其与文 [1] 中算法 (记为算法 B) 和文 [2] 中算法 (记为算法 C) 进行比较. 数值结果列示于表 1, 各符号含义如下: n: 自变量 x 的维数; m: 目标函数中 $f_i(x)$ 的个数; x^0: 初始点; ni: 算法迭代次数; nf: 目标中各函数 $f_i(x)$ 计算的总次数; ng: 目标中需要计算的梯度 $\nabla f_i(x)$ 的总次数; x^*: 终止时获得的近似解; $F(x^*)$: 终止时获得的近似最优目标函数值.

由表 1 的数值结果可总体上看出算法 A 是有效的. 对问题 P3, P6 和 P7, 算法 A 的 nf, ng 要优于算法 B; 对问题 P2, 算法 A 与算法 B 效果相当; 对问题 P1, P4 和 P5, 算法 A 的 nf 虽不及算法 B, 但 ng 较少. 对问题 P2, P6 和 P7, 算法 A 的 nf, ng 要优于算法 C; 对问题 P1 和 P3, 算法 A 和算法 C 相当; 对问题 P4 和 P5, 算法 A 的 nf 要多于算法 C. 值得指出的是, 文献 [1,2] 属于序列二次规划方法, 每次迭代需要求解一个二次规划子问题. 而本文算法产生搜索方向的计算量主要在于计算 (2.4) 式中的 Q_k, 其相当于求解一个线性方程组. 一般认为, 求解二次规划的计算量要大于求解线性方程组, 因此, 就此而言, 本文算法每步迭代的计算量相对较小.

表 1 算法 A, 算法 B 和算法 C 的数值结果比较

算例	x^0	n,m	算法	ni	nf	ng	x^*	$F(x^*)$
P1	$(1,-0.1)^{\mathrm{T}}$	2,3	算法 A	14	141	17	$(1.139, 0.899)^{\top}$	1.9522
			算法 B	14	57	32	$(1.139, 0.899)^{\top}$	1.9522
			算法 C	12	171	20	$(1.139, 0.899)^{\top}$	1.9522
P2	$(0.1,0.1)^{\mathrm{T}}$	2,3	算法 A	9	64	13	$(1.000, 1.000)^{\top}$	2.000
			算法 B	6	33	18	$(1.000, 1.000)^{\top}$	2.000
			算法 C	7	117	14	$(1.000, 1.000)^{\top}$	2.000
P3	$(0.1,0,0.2,0)^{\mathrm{T}}$	4,4	算法 A	59	727	86	$(-0.8203, 0.9170,$ $2.1256, 0.8526)^{\top}$	-48.016
			算法 B	57	972	220	$(-0.8200, 0.9174,$ $2.1257, 0.8525)^{\top}$	-48.016
			算法 C	32	1016	72	$(-0.8200, 0.9174,$ $2.1257, 0.8525)^{\top}$	-48.016
P4	$(2,0)^{\mathrm{T}}$	2,3	算法 A	13	116	17	$(0.4624, -0.9066)^{\top}$	0.6164
			算法 B	11	48	32	$(0.4533, -0.9066)^{\top}$	0.6164
			算法 C	10	93	22	$(0.4533, -0.9066)^{\top}$	0.6164
P5	$(0.1,0.1,0.2)^{\mathrm{T}}$	3,6	算法 A	16	402	22	$(0.329, -0.000, 0.131)^{\top}$	3.5997
			算法 B	13	108	38	$(0.328, -0.000, 0.131)^{\top}$	3.5997
			算法 C	10	198	23	$(0.328, -0.000, 0.131)^{\top}$	3.5997
P6	$(0.1,0.001)^{\mathrm{T}}$	2,3	算法 A	7	86	10	$(0.000, 0.989)^{\top}$	0.000
			算法 B	9	72	30	$(0.000, -0.901)^{\top}$	0.000
			算法 C	5	111	13	$(0.000, -0.255)^{\top}$	0.000
P7	$(0.01,0.01,-1,-1)^{\mathrm{T}}$	4,4	算法 A	6	24	18	$(0.0052, 0.0052,$ $-0.0105, -0.0105)^{\top}$	0.000
			算法 B	18	76	76	$(0.000, 0.0059,$ $-0.7628, -0.7628)^{\top}$	0.000
			算法 C	11	48	46	$(0.000, 0.9766,$ $0.000, -0.000)^{\top}$	0.000

参 考 文 献

[1] Jian J B, Quan R, Zhang X L. Generalised monotone line search algorithm for degenerate nonlinear minimax problems[J]. Bulletin of the Ausralian Mathematical Society, 2006, 73: 117-127.

[2] 简金宝, 石露, 唐春明. 基于积极集技术求解无约束极大极小问题的摄动 SQP 方法 [J]. 高校应用数学学报, 2013, 28(1): 107-114.

[3] Polak E, Mayne D Q, Higgins J E. Superlinearly convergent algorithm for min-max problems[J]. Journal of Optimization Theory and Applications, 1991, 69: 407-439.

[4] Vardi A. New minimax algorithm[J]. Journal of Optimization Theory and Applications, 1992, 75(3): 613-634.

[5] Zhou J L, Tits A L. Nonmonotone line search for minimax problems[J]. Journal of Optimization Theory and Applications, 1993, 76: 455-476.

[6] Jian J B, Quan R, Hu Q J. A new superlinearly convergent SQP algorithm for nonlinear minimax problems[J]. Acta Mathematicatae Applicatae Sinica, English Series, 2007, 23(3): 395-410.

[7] 薛毅. 求解 Minimax 优化问题的 SQP 方法 [J]. 系统科学与数学, 2002, 22(3): 355-364.
[8] Zoutendijk G. Methods of Feasible Directions[M]. Amsterdam: Elsevier, 1960.
[9] 赖炎连, 简金宝. 初始点任意的一个非线性优化的广义梯度投影法 [J]. 系统科学与数学, 1995, 374-380.
[10] 赖炎连, 高自友, 贺国平. 非线性最优化的广义梯度投影法 [J]. 中国科学, 1992.
[11] 简金宝. 光滑约束优化快速算法 – 理论分析与数值试验 [M]. 北京: 科学出版社, 2010, 227-251.
[12] Ladislav L, Jan V. Test Problems for Nonsmooth Unconstrained and Linearly Constrained Optimization[R]. Institute of Computer Science, Academy of Sciences of the Czech Republic, 2000.

GENERALIZED GRADIENT PROJECTION ALGORITHM FOR UNCONSTRAINED MINIMAX PROBLEMS

Jian Jinbao
(*College of Mathematics and Information Science, Guangxi University, Nanning 530004, Chian; School of Mathematics and Information Science, Yulin Normal University, Yulin 537000, China*)
Tang Fei Li Jianling Tang Chunming
(*College of Mathematics and Information Science, Guangxi University, Nanning 530004, Chian*)

Abstract

In this paper, the unconstrained minimax problems on $\mathbf{R}^n$ are discussed. The search direction of descent in $\mathbf{R}^n$ is obtained by a generalized gradient projection on $\mathbf{R}^{n+1}$, then, with Armjio non-exact line search, a generalized gradient projection algorithm on $\mathbf{R}^n$ for the discussed minimax problems is presented. The proposed algorithm possesses global and strong convergence under affine linearly independent condition. Some preliminary numerical experiments are carried out.

Keywords: unconstrained minimax problems; generalized gradient projection algorithm; global convergence; strong convergence

2000 Mathematics Subject Classification: 90C30, 90C47