广西科学 Guangxi Sciences 2016,23(5):404～408

网络优先数字出版时间:2016-11-21 【DOI】10.13656/j.cnki.gxkx.20161121.012

网络优先数字出版地址:http://www.cnki.net/kcms/detail/45.1206.G3.20161121.1546.024.html

基于非精确数据的非光滑优化强次可行方向法*

Strongly Sub-feasible Direction Method with Inexact Data for Nonsmooth Optimization

唐春明,律金曼

TANG Chunming,LV Jinman

(广西大学数学与信息科学学院,广西南宁 530004)

(College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, 530004,China)

摘要:本研究针对一类目标函数非光滑优化问题,提出一个基于非精确数据的强次可行方向法.通过构造新的寻找搜索方向子问题和新型线搜索,该算法能够保证迭代点的强次可行性,且具备全局收敛性.

关键词:非光滑优化 强次可行方向法 非精确数据

中图分类号:C934 **文献标识码**:A **文章编号**:1005-9164(2016)05-0404-05

Abstract:In this paper, a strongly sub-feasible direction method with inexact data is proposed for solving a class of optimization problems with nonsmooth objectives. By constructing a new search direction finding subproblem and a new line search, the strongly sub-feasibility of the iteration points is guaranteed, and the global convergence of the algorithm is proved.

Key words:nonsmooth optimization, strongly sub-feasible direction method, inexact data

0 引言

考虑如下非线性不等式约束优化问题

$$\min_{x\in\mathbb{R}^n} \quad f(x) \qquad (0.1)$$

$$\text{s.t.} \quad c_i(x)\leqslant 0, i\in I \triangleq \{1,\cdots,m\},$$

其中 $f:\mathbb{R}^n\to\mathbb{R}$ 是凸函数但不一定光滑,$c_i(i\in I):\mathbb{R}^n\to\mathbb{R}$ 是连续可微的凸函数.

在一些实际问题中,有时很难精确计算 f 的函数值.例如,f 是如下 max-型函数

$$f(x)=\max\{F_u(x):u\in U\}, \qquad (0.2)$$

其中对任意给定的 $u\in U$,$F_u:\mathbb{R}^n\to\mathbb{R}$ 是凸函数,U 是一个无限集,此时无法计算 f 的精确值.然而,对于任意正数 ε,可以在有限的时间内找出(0.2)的一个 ε-解,即找出一个 $u_\varepsilon\in U$ 满足 $F_{u_\varepsilon}(x)\geqslant f(x)-\varepsilon$,从而得到 $f(x)$ 的近似值.因此,研究基于非精确数据的优化方法具有重要的意义[1-3].

文献[4]基于精确数据,提出一个求解问题(0.1)的强次可行方向法,其优点在于能接受不可行的初始点,且一旦产生一个可行迭代点,即自动变为可行下降算法.此外,算法可保证迭代点的强次可行性,同时能防止目标函数过度增大.文献[2]中提出一种非精确数据的思想,即假设对于给定的点 x 和误差限 $\varepsilon\geqslant 0$,能够计算得到近似的函数值 $f_\varepsilon(x)\approx f(x)$ 和一个近似的次梯度 $g_\varepsilon\approx g\in\partial f(x)$ 满足:

$$f_\varepsilon(x)\in[f(x)-\varepsilon, f(x)+\varepsilon],$$

$$g_\varepsilon\in\partial_\varepsilon f(x)=\{g:f(y)\geqslant f(x)+\langle g,y-x\rangle-\varepsilon, \forall y\in\mathbb{R}^n\}.$$

本研究旨在对文献[4]的方法进行改进,结合文献[2]的思想,提出一个基于非精确数据的非光滑优化强次可行方向法.

收稿日期:2016-08-05

修回日期:2016-09-20

作者简介:唐春明(1979-),男,博士,教授,主要从事最优化理论、方法及其应用研究,E-mail:cmtang@gxu.edu.cn。

*国家自然科学基金项目(11301095,11271086)和广西自然科学基金项目(2013GXNSFAA019013,2014GXNSFFA118001)资助。

1 算法

假设 $y^j, j=1,\cdots,k$ 为试探点列，$\varepsilon^j \geqslant 0$ 为相应的误差限，$f_{\varepsilon^j}(y^j)$ 和 $g^j_{\varepsilon^j}$ 分别为 y^j 处函数值和次梯度的近似，满足 $f_{\varepsilon^j}(y^j) \in [f(y^j)-\varepsilon^j, f(y^j)+\varepsilon^j]$ 及 $g^j_{\varepsilon^j} \in \partial_{\varepsilon^j} f(y^j)$. 记 $g^j = g^j_{\varepsilon^j}$，因此 f 在 y^j 处的近似线性化可定义为[2]

$$f_j(x) = f_{\varepsilon^j}(y^j) + \langle g^j, x-y^j \rangle - 2\varepsilon^j. \tag{1.1}$$

由 $g^j \in \partial_{\varepsilon^j} f(y^j)$ 和 $f(y^j) \geqslant f_{\varepsilon^j}(y^j) - \varepsilon^j$ 可知

$$f_j(x) \leqslant f(x), \forall x \in \mathbb{R}^n. \tag{1.2}$$

进而可定义 f 的近似割平面模型

$$\hat{f}^k(x) = \max\{f_j(x), j=1,\cdots,k\}.$$

记问题(0.1)的可行集 $F=\{x \in \mathbb{R}^n : c_i(x) \leqslant 0, i \in I\}$. 定义指标集 $I_-(x) = \{i \in I : c_i(x) \leqslant 0\}$，$I_+(x) = \{i \in I : c_i(x) > 0\}$，约束违反函数 $\varphi(x) = \max\{0, c_i(x), i \in I\}$. 引入改进函数[4]：

$$H(y;x) = \max\{f(y)-f(x)-\delta(x); c_i(y), i \in I_-(x); c_i(y)-\varphi(x), i \in I_+(x)\},$$

其中 $\delta(x): \mathbb{R}^n \to [0,+\infty)$ 连续，并满足 $\delta(x)=0$ 当且仅当 $x \in F$. 对于第 k 次迭代，缩写 $I^k_- = I_-(x^k)$，$I^k_+ = I_+(x^k)$，$\varphi^k = \varphi(x^k)$，$\delta^k = \delta(x^k)$.

下面给出改进函数的性质.

引理 1.1[4] 若问题(0.1)满足 Slater 约束规格，即存在一个向量 $x' \in \mathbb{R}^n$ 满足 $c_i(x') < 0, \forall i \in I$，则以下 3 个命题等价：(i) $\bar{x}$ 是问题(0.1)的最优解；(ii) $\min\{H(y;\bar{x}) : y \in \mathbb{R}^n\} = H(\bar{x};\bar{x}) = 0$；(iii) $0 \in \partial H(\bar{x};\bar{x})$.

基于引理 1.1，并结合邻近点方法思想[5]，选取新的试探点如下：

$$y^{k+1} = \arg\min\{H(y;x^k) + \frac{1}{2}\rho^k \| y-x^k \|^2 : y \in \mathbb{R}^n\}, \tag{1.3}$$

其中 $\rho^k > 0$ 是邻近参数，$x^k \in \{y^j\}^k_{j=1}$ 为邻近中心点. 然而，问题(1.3)的求解较困难. 因此，基于非精确数据，并结合文献[4]的思想，本文提出如下寻找搜索方向子问题：

$$\begin{aligned} \min_{d \in \mathbb{R}^n, z \in \mathbb{R}} \quad & z + \frac{1}{2}\rho^k \| d \|^2, \\ \text{s.t.} \quad & -\alpha^k_j + \langle g^j, d \rangle \leqslant z + \delta^k, j \in J^k, \\ & -\alpha^k_s + \langle s^{k-1}, d \rangle \leqslant z + \delta^k, \\ & c_i(x^k) + \langle \nabla c_i(x^k), d \rangle \leqslant z, i \in I^k_-, \\ & c_i(x^k) + \langle \nabla c_i(x^k), d \rangle \leqslant z + \varphi^k, i \in I^k_+, \end{aligned} \tag{1.4}$$

其中 $z \in \mathbb{R}$ 是一个辅助变量，$J^k \subseteq \{1,\cdots,k\}$，$\alpha^k_j = f_\varepsilon(x^k) - f^k_j + \varepsilon$，$\alpha^k_s = f_\varepsilon(x^k) - f^k_s + \varepsilon$，$f^k_j = f_j(x^k)$；$s^{k-1}$ 和 f^k_s 分别为过往的近似函数值和次梯度的聚集(具体将在算法中明确)，文献[6]中证明存在 $\bar{\lambda}_j \geqslant 0, j=1,\cdots,k-1$，使得

$$(s^{k-1}, f^k_s) = \sum_{j=1}^{k-1} \bar{\lambda}_j (g^j, f^k_j), \sum_{j=1}^{k-1} \bar{\lambda}_j = 1. \tag{1.5}$$

设 (d^k, z^k) 是(1.4)式的最优解. 由于(1.4)式是一个带线性约束的凸规划，故 (d^k, z^k) 亦为(1.4)式的 KKT 点，即存在乘子 $\lambda^k_j, j \in J^k, \lambda^k_s, \mu^k_i, i \in I$，使得

$$\begin{aligned} & \rho^k d^k + \sum_{j \in J^k} \lambda^k_j g^j + \lambda^k_s s^{k-1} + \sum_{i \in I} \mu^k_i \nabla c_i(x^k) = 0, \\ & \sum_{j \in J^k} \lambda^k_j + \lambda^k_s + \sum_{i \in I} \mu^k_i = 1, \\ & 0 \leqslant \lambda^k_j \perp (-\alpha^k_j + \langle g^j, d^k \rangle - z^k - \delta^k) \leqslant 0, j \in J^k, \\ & 0 \leqslant \lambda^k_s \perp (-\alpha^k_s + \langle s^{k-1}, d^k \rangle - z^k - \delta^k) \leqslant 0, \\ & 0 \leqslant \mu^k_i \perp (c_i(x^k) + \langle \nabla c_i(x^k), d^k \rangle - z^k) \leqslant 0, i \in I^k_-, \\ & 0 \leqslant \mu^k_i \perp (c_i(x^k) + \langle \nabla c_i(x^k), d^k \rangle - z^k - \varphi^k) \leqslant 0, i \in I^k_+. \end{aligned} \tag{1.6}$$

更新聚集次梯度如下：

$$(s^k, \check{f}^k_s) = \frac{1}{\theta^k} \Big(\sum_{j \in J^k} \lambda^k_j (g^j, f^k_j) + \lambda^k_s (s^{k-1}, f^k_s) \Big), \tag{1.7}$$

其中 $\theta^k = \sum_{j \in J^k} \lambda^k_j + \lambda^k_s$(不失一般性假设 $\theta^k > 0$).

记 $\check{\alpha}^k_s = f_\varepsilon(x^k) - \check{f}^k_s + \varepsilon$，有 $\sum_{j \in J^k} \lambda^k_j \alpha^k_j + \lambda^k_s \alpha^k_s = \theta^k \check{\alpha}^k_s$.

以下引理给出子问题(1.4)的解的性质.

引理 1.2 设 (d^k, z^k) 是问题(1.4)的最优解，则

(i) $z^k = -(\rho^k \| d^k \|^2 + \check{\alpha}^k)$, $\quad$ (1.8)

其中，

$$\check{\alpha}^k = \theta^k(\check{\alpha}^k_s + \delta^k) - \sum_{i \in I^k_-} \mu^k_i c_i(x^k) - \sum_{i \in I^k_+} \mu^k_i (c_i(x^k) - \varphi^k). \tag{1.9}$$

(ii) $g^j \in \partial_\epsilon f(x^k), \epsilon = \alpha^k_j, j=1,\cdots,k$,

$$\begin{aligned} & s^k \in \partial_\epsilon f(x^k), \epsilon = \check{\alpha}^k_s, \\ & -\rho^k d^k \in \partial_\epsilon H(x^k; x^k), \epsilon = \check{\alpha}^k. \end{aligned} \tag{1.10}$$

(iii) 如果 $z^k = 0$，则 $d^k = 0$，且 x^k 是问题(0.1)的一个最优解.

证明 (i) 由 KKT 条件(1.6)中的互补关系和(1.7)式有

$$z^k = -\rho^k \| d^k \|^2 + \sum_{j \in J^k} \lambda^k_j (-\alpha^k_j - \delta^k) + \lambda^k_s (-\alpha^k_s - \delta^k) + \sum_{i \in I^k_-} \mu^k_i c_i(x^k) + \sum_{i \in I^k_+} \mu^k_i (c_i(x^k) -$$

$\varphi^k) = -\rho^k \| d^k \|^2 - \theta^k(\check{\alpha}_s^k + \delta^k) + \sum_{i\in I_-^k} \mu_i^k c_i(x^k) + \sum_{i\in I_+^k} \mu_i^k (c_i(x^k) - \varphi^k) = -(\rho^k \| d^k \|^2 + \check{\alpha}^k)$.

故(1.8)式成立.

(ii) 由(1.2)式知,

$f(x) \geqslant f_j(x) = f_j^k + \langle g^j, x - x^k \rangle = f(x^k) - f(x^k) + f_j^k + \langle g^j, x - x^k \rangle \geqslant f(x^k) + \langle g^j, x - x^k \rangle + f_j^k - (f_\varepsilon(x^k) + \varepsilon) = f(x^k) + \langle g^j, x - x^k \rangle - \alpha_j^k$, (1.11)

从而(1.10)式的第一个式子成立. 类似地,根据(1.5)式知

$$f(x) \geqslant f(x^k) + \langle s^{k-1}, x - x^k \rangle - \alpha_s^k. \quad (1.12)$$

在(1.11)式和(1.12)式两边分别乘以 $\lambda_j^k, j \in J^k$ 和 λ_s^k 再进行求和,有

$$f(x) \geqslant f(x^k) + \langle s^k, x - x^k \rangle - \check{\alpha}_s^k, \quad (1.13)$$

故(1.10)式的第二个式子成立.

根据 c_i 的凸性,有

$$c_i(x) \geqslant c_i(x^k) + \langle \nabla c_i(x^k), x - x^k \rangle.$$

因此,

$$\sum_{i\in I} \mu_i^k c_i(x) \geqslant \sum_{i\in I} \mu_i^k (c_i(x^k) + \langle \nabla c_i(x^k), x - x^k \rangle).$$

此式结合(1.6)式,(1.13)式,$\theta^k \leqslant 1$ 及 $H(x^k; x^k) = 0$ 可得

$H(x; x^k) = \max\{f(x) - f(x^k) - \delta^k; c_i(x), i \in I_-^k; c_i(x) - \varphi^k, i \in I_+^k\} \geqslant \theta^k(f(x) - f(x^k) - \delta^k) + \sum_{i\in I} \mu_i^k c_i(x) - \sum_{i\in I_+^k} \mu_i^k \varphi^k \geqslant \theta^k(\langle s^k, x - x^k \rangle - \check{\alpha}_s^k - \delta^k) + \sum_{i\in I} \mu_i^k c_i(x^k) + \sum_{i\in I} \mu_i^k \langle \nabla c_i(x^k), x - x^k \rangle - \sum_{i\in I_+^k} \mu_i^k \varphi^k = H(x^k; x^k) + \langle -\rho^k d^k, x - x^k \rangle - (\theta^k(\check{\alpha}_s^k + \delta^k) - \sum_{i\in I_-^k} \mu_i^k c_i(x^k) - \sum_{i\in I_+^k} \mu_i^k (c_i(x^k) - \varphi^k)) = H(x^k; x^k) + \langle -\rho^k d^k, x - x^k \rangle - \check{\alpha}^k$.

由此证明(1.10)式的第三个式子.

(iii) 由于 $\check{\alpha}^k \geqslant 0$,故由(1.8)式知,当 $z^k = 0$ 时有 $\rho^k d^k = 0$ 和 $\check{\alpha}^k = 0$,从而 $0 \in \partial H(x^k; x^k)$. 故根据引理1.1(iii) 知 x^k 是问题(0.1)的一个最优解.

算法 1.1

步骤 0(初始化) 选取初始点 $x^1 \in \mathbf{R}^n$,初始误差限 $\varepsilon^1 \geqslant 0$,终止参数 $\varepsilon_{TOL} \geqslant 0$,线搜索参数 $\beta \in (0, 1)$,邻近参数 $\rho_{\min} > 0$,$\rho^1 > \rho_{\min}$,步长下界 $\underline{t} \in (0, 0.1]$. 选择两个线搜索参数 m_L 和 m_R 满足 $0 < m_L < m_R < 1$. 令 $y^1 = x^1$,$s^0 = g^1 \in \partial_{\varepsilon^1} f(x^1)$,$f_s^1 = f_1^1 = f_{\varepsilon^1}(x^1)$,$J^1 = \{1\}$. 置 $k = 1$,$l = 0$ 及 $k(0) = 1$.

步骤 1(寻找搜索方向) 令 $\varepsilon = \varepsilon^k$,求解子问题(1.4)得到最优解$(d^k, z^k)$和乘子 $\lambda_j^k, j \in J^k, \lambda_s^k, \mu_i^k, i \in I$. 根据(1.7)式计算 s^k,$\check{f}_s^k$和α_s^k.

步骤 2(误差限更新) 如果 $\varepsilon^k > -(m_R - m_L)\underline{t}z^k/5$,置 $\varepsilon^k = \varepsilon^k/2$ 并返回步骤 1;否则,转步骤 3.

步骤 3(终止准则) 如果 $z^k \geqslant -\varepsilon_{TOL}$,算法终止;否则,转步骤 4.

步骤 4(线搜索) 计算试探步长 t^k,它是序列 $\{1, \beta, \beta^2, \cdots\}$ 中第一个满足下列不等式组的 t 值:

$$c_i(x^k + t d^k) \leqslant \varphi^k + m_L t z^k, i \in I_+^k, \quad (1.14)$$

$$c_i(x^k + t d^k) \leqslant 0, i \in I_-^k. \quad (1.15)$$

如果

$$f_{\varepsilon^k}(x^k + t^k d^k) \leqslant f_{\varepsilon^k}(x^k) + m_L t^k z^k + t^k \delta^k - 2\varepsilon^k,$$

则令 $t_L^k = t^k$(有效步)及辅助步长 $t_R^k = t^k$,置 $k(l+1) = k+1$,$l := l+1$;否则令 $t_L^k = 0$(无效步)及 $t_R^k = \max\{\underline{t}, t^k\}$.

步骤 5(更新) 令 $\varepsilon^{k+1} = \varepsilon^k$,$x^{k+1} = x^k + t_L^k d^k$,$y^{k+1} = x^k + t_R^k d^k$. 选取 $\hat{J}^k \subseteq J^k$ 并令 $J^{k+1} = \hat{J}^k \cup \{k+1\}$,计算 $g^{k+1} \in \partial_{\varepsilon^{k+1}} f(y^{k+1})$ 及 $f_{k+1}^{k+1} = f_{\varepsilon^{k+1}}(y^{k+1}) + \langle g^{k+1}, x^{k+1} - y^{k+1} \rangle - 2\varepsilon^{k+1}$,

$$f_j^{k+1} = f_j^k + \langle g^j, x^{k+1} - x^k \rangle, j \in \hat{J}^k,$$

$$f_s^{k+1} = \check{f}_s^k + \langle s^k, x^{k+1} - x^k \rangle. \quad (1.16)$$

步骤 6(邻近参数选择) 如果 $x^{k+1} \neq x^k$,取 $\rho^{k+1} \in [\rho_{\min}, \rho^k]$;否则,$\rho^{k+1} = \rho^k$.

步骤 7 令 $k := k+1$,返回步骤 1.

引理 1.3[4] 算法 1.1 是适定的,即线搜索(1.14)和(1.15)能在有限次计算后终止.

引理 1.4 算法 1.1 必定出现以下两种情形之一.

(i) 存在一个指标 k_0 使得 $\varphi^{k_0} = 0$,从而 $\varphi^k \equiv 0$,$\delta^k \equiv 0$ 和 $f(x^{k+1}) \leqslant f(x^k)$,对于所有的 $k \geqslant k_0$ 成立;

(ii) 对任意的 k,有 $\varphi^k > 0$,$\varphi^{k+1} \leqslant \varphi^k$,$I_-^k \subseteq I_-^{k+1}$.

证明 (i) 由步骤 4 可知,$\varphi^k \equiv 0$ 及 $\delta^k \equiv 0$ 对 $k \geqslant k_0$ 成立. 现证明 $f(x^{k+1}) \leqslant f(x^k)$. 根据步骤 4,如果是一个有效步,由 $z^k < 0$ 可得

$$f(x^{k+1}) \leqslant f_{\varepsilon^k}(x^{k+1}) + \varepsilon^k \leqslant f_{\varepsilon^k}(x^k) + m_L t_L^k z^k - \varepsilon^k \leqslant f(x^k) + m_L t_L^k z^k \leqslant f(x^k).$$

若是一个无效步,则有 $f(x^{k+1}) = f(x^k)$.

(ii) 根据线搜索(1.14)和(1.15)易证.

2 算法的收敛性

在以下收敛分析中取 $\varepsilon_{TOL} = 0$. 若算法有限步终止于 x^k 点,则由步骤 3 可知 $z^k = 0$,因此由引理 1.2(iii) 得 x^k 是问题(0.1)的一个最优解. 假设算法

产生一个无限迭代点列，以下将证明该序列的任一聚点都是问题(0.1)的一个最优解. 根据引理1.4，在下面的分析中不妨假设 $I_-^k \equiv I_-, I_+^k \equiv I_+$.

引理2.1 邻近参数序列 $\{\rho^k\}$ 单调不增，且有正的下界.

证明 根据步骤6，显然 $\{\rho^k\}$ 单调不增，且 $\rho^k \geqslant \rho_{\min} > 0$.

引理2.2 假设存在无限指标集 $K \subseteq \{1,2,\cdots\}$ 和一个点 $\bar{x} \in \mathbf{R}^n$ 满足 $x^k \xrightarrow{K} \bar{x}$ 及 $z^k \xrightarrow{K} 0$，则 $\bar{x}$ 是问题(0.1)的一个最优解.

证明 由 $z^k \xrightarrow{K} 0$ 和(1.8)式知 $\rho^k d^k \to 0, \check{\alpha}^k \to 0, k \in K$. 故由 $x^k \xrightarrow{K} \bar{x}$ 及引理1.2知 $0 \in \partial H(\bar{x};\bar{x})$，因此 $\bar{x}$ 是问题(0.1)的一个最优解.

分两种情形证明. 首先考虑有无限个有效步的情形. 类似文献[7]，有如下引理.

引理2.3 假设存在一个无限指标集 $L \subseteq \{1,2,\cdots\}$ 和一个点 $\bar{x} \in \mathbf{R}^n$ 使得 $x^{k(l)} \to \bar{x}, l \to \infty, l \in L$，则 $\bar{x}$ 是问题(0.1)的一个最优解.

接下来考虑有效步有限的情形，即存在一个指标 $\bar{k}$ 使得 $x^k = x^{\bar{k}}, k \geqslant \bar{k}$. 下面证明 $x^{\bar{k}}$ 是问题(0.1)的一个最优解.

引理2.4 令 $\omega^k = \frac{1}{2}\|\rho^k d^k\|^2 + \rho^k \check{\alpha}^k$，则 ω^k 是以下问题的最优值

$$\begin{aligned} \min \quad & \frac{1}{2}\Big\|\sum_{j\in J^k}\lambda_j g^j + \lambda_s s^{k-1} + \sum_{i\in I}\mu_i \nabla c_i(x^k)\Big\|^2 - \rho^k\Big[\sum_{j\in J^k}\lambda_j(-\alpha_j^k - \delta^k) - \lambda_s(-\alpha_s^k - \delta^k) - \sum_{i\in I_-}\mu_i c_i(x^k) - \sum_{i\in I_+}\mu_i(c_i(x^k) - \varphi^k)\Big] \\ \text{s.t.} \quad & \lambda_j \geqslant 0, j \in J^k, \lambda_s \geqslant 0, \mu_i \geqslant 0, i \in I, \\ & \sum_{j\in J^k}\lambda_j + \lambda_s + \sum_{i\in I}\mu_i = 1. \end{aligned} \tag{2.1}$$

证明 由于问题(2.1)是问题(1.4)的对偶问题，故问题(2.1)的最优解即为问题(1.4)的KKT乘子. 因此，由(1.6)式，(1.9)式及 ω^k 的定义可得结论成立.

引理2.5[6] 设 n 维向量 d, g 及数 $\eta \in (0,1)$，$\rho > 0, C, \omega, z, \check{\alpha}$ 和 α 满足

$$\omega = \frac{1}{2}\|\rho d\|^2 + \rho\check{\alpha}, z = -(\rho\|d\|^2 + \check{\alpha}),$$

$$-\alpha + \langle g, d\rangle \geqslant \eta z, C \geqslant \max\{\|\rho d\|, \|g\|, \check{\alpha}, 1\}.$$

令 $\bar{\omega} = \min\{Q(\upsilon): \upsilon \in [0,1]\}$，其中 $Q(\upsilon) = \frac{1}{2}\|(1-\upsilon)(-\rho d) + \upsilon g\|^2 + \rho[(1-\upsilon)\check{\alpha} + \upsilon\alpha]$，则 $\bar{\omega} \leqslant \omega - \rho^2(1-\eta)^2\omega^2/8C^2$.

基于引理2.5，得到以下一个重要的结论.

引理2.6 假设对于某个 $k > 1$ 有 $t_L^k = 0$ 及 $\varepsilon^{k-1} = \varepsilon^k$，则

(i) $$-(\alpha_k^k + \delta^{k-1}) + \langle g^k, d^{k-1}\rangle \geqslant m_R z^{k-1}, \tag{2.2}$$

其中 $\varepsilon_k^k = f_{\varepsilon^k}(x^k) - f_k^k + \varepsilon^k$；

(ii) $$\omega^k \leqslant \omega^{k-1} - (\rho^{k-1})^2(1-m_R)^2(\omega^{k-1})^2/8(C^k)^2, \tag{2.3}$$

其中 C^k 满足 $C^k \geqslant \max\{\|\rho^{k-1} d^{k-1}\|, \|g^k\|, \check{\alpha}^{k-1}, 1\}$.

证明 (i) 由算法1.1中步骤4知当 $t_L^{k-1} = 0$，$\varepsilon^{k-1} = \varepsilon^k$ 时，$x^k = x^{k-1}, y^k = x^{k-1} + t_R^{k-1}d^{k-1}$，且有

$$f_{\varepsilon^k}(y^k) > f_{\varepsilon^k}(x^{k-1}) + m_L t_R^{k-1}z^{k-1} + t_R^{k-1}\delta^{k-1} - 2\varepsilon^k. \tag{2.4}$$

因此，由(1.1)式和(2.4)式有

$$\begin{aligned} &-(\alpha_k^k + \delta^{k-1}) = -(f_{\varepsilon^k}(x^k) - f_k^k + \varepsilon^k) - \delta^{k-1} = -(f_{\varepsilon^k}(x^k) - f_{\varepsilon^k}(y^k) - \langle g^k, x^k - y^k\rangle + 3\varepsilon^k) - \delta^{k-1} = -(f_{\varepsilon^k}(x^{k-1}) - f_{\varepsilon^k}(y^k) - \langle g^k, x^{k-1} - y^k\rangle + \\ &3\varepsilon^k) - \delta^{k-1} = f_{\varepsilon^k}(y^k) - t_R^{k-1}\langle g^k, d^{k-1}\rangle - f_{\varepsilon^k}(x^{k-1}) - \delta^{k-1} - 3\varepsilon^k \geqslant m_L t_R^{k-1}z^{k-1} - t_R^{k-1}\langle g^k, d^{k-1}\rangle + (1 - t_R^{k-1})\delta^{k-1} - 5\varepsilon^k. \end{aligned} \tag{2.5}$$

由 $\varepsilon_k^k = f_{\varepsilon^k}(x^k) - f_k^k + \varepsilon^k \geqslant f(x^k) - f_k^k \geqslant 0$ 知 $\varepsilon_k^k = f_{\varepsilon^k}(x^{k-1}) - f_{\varepsilon^k}(y^k) + t_R^{k-1}\langle g^k, d^{k-1}\rangle + 3\varepsilon^k \geqslant 0$，因此根据(2.4)式可得

$$\langle g^k, d^{k-1}\rangle \geqslant m_L z^{k-1} + \delta^{k-1} - 5\varepsilon^k/t_R^{k-1}. \tag{2.6}$$

结合(2.5)式，(2.6)式和 $t_R^{k-1} \in [\underline{t}, 1]$ 得

$$\begin{aligned} &-(\alpha_k^k + \delta^{k-1}) + \langle g^k, d^{k-1}\rangle \geqslant m_L t_R^{k-1}z^{k-1} + (1 - t_R^{k-1})(\langle g^k, d^{k-1}\rangle - \delta^{k-1}) - 5\varepsilon^k \geqslant m_L t_R^{k-1}z^{k-1} + (1 - \\ &t_R^{k-1})(m_L z^{k-1} - 3\varepsilon^k/t_R^{k-1}) - 5\varepsilon^k = m_L z^{k-1} - 5\varepsilon^k/t_R^{k-1} \geqslant m_L z^{k-1} - 5\varepsilon^k/\underline{t}. \end{aligned}$$

由步骤2知 $5\varepsilon^k/\underline{t} = 5\varepsilon^{k-1}/\underline{t} > -(m_R - m_L)z^{k-1}$，从而得到

$$-(\alpha_k^k + \delta^{k-1}) + \langle g^k, d^{k-1}\rangle \geqslant m_R z^{k-1}.$$

(ii) 对任意的 $\upsilon \in [0,1]$，定义问题(2.1)的可行解

$$\lambda_k(\upsilon) = \upsilon, \lambda_j(\upsilon) = 0, j \in J^k \backslash \{k\},$$

$$\lambda_s(\upsilon) = (1-\upsilon)\theta^{k-1}, \mu_i(\upsilon) = (1-\upsilon)\mu_i^{k-1}, i \in I.$$

由(1.16)式和 $x^{k-1} = x^k$ 可得

$$\alpha_s^k = f_{\varepsilon^k}(x^k) - f_s^k + \varepsilon^k = f_{\varepsilon^k}(x^{k-1}) - \check{f}_s^{k-1} + \varepsilon^k = \check{\alpha}_s^{k-1}, \tag{2.7}$$

因此

$$\sum_{j\in J^k}\lambda_j(\upsilon) g^j + \lambda_s(\upsilon)s^{k-1} + \sum_{i\in I}\mu_i(\upsilon)\nabla c_i(x^k) = \upsilon g^k + (1-\upsilon)\theta^{k-1}s^{k-1} + (1-\upsilon)\sum_{i\in I}\mu_i^{k-1}\nabla c_i(x^{k-1}) =$$

$\upsilon g^k + (1-\upsilon)(-\rho^{k-1} d^{k-1})$.

此外，根据(2.7)式有

$$-\sum_{j\in J^k}\lambda_j(-\alpha_j^k-\delta^k)-\lambda_s(-\alpha_s^k-\delta^k)-\sum_{i\in I_-}\mu_i c_i(x^k)-\sum_{i\in I_+}\mu_i(c_i(x^k)-\varphi^k)=-\upsilon(-\alpha_k^k-\delta^k)-(1-\upsilon)\theta^{k-1}(-\alpha_s^k-\delta^k)-(1-\upsilon)\sum_{i\in I_-}\mu_i^{k-1}c_i(x^k)-(1-\upsilon)\sum_{i\in I_+}\mu_i^{k-1}(c_i(x^k)-\varphi^k)=\upsilon(\alpha_k^k+\delta^{k-1})+(1-\upsilon)(\theta^{k-1}(\dot\alpha_s^{k-1}+\delta^{k-1})-\sum_{i\in I_-}\mu_i^{k-1}c_i(x^{k-1})-\sum_{i\in I_+}\mu_i^{k-1}(c_i(x^{k-1})-\varphi^{k-1}))=\upsilon(\alpha_k^k+\delta^{k-1})+(1-\upsilon)\dot\alpha^{k-1}.$$

令$\bar\omega$是如下问题的最优值

$$\min\quad \frac{1}{2}\|(1-\upsilon)(-\rho^{k-1}d^{k-1})+\upsilon g^k\|^2+\rho^{k-1}[(1-\upsilon)\dot\alpha^{k-1}+\upsilon(\alpha_k^k+\delta^{k-1})],$$
$$\text{s.t.}\quad \upsilon\in[0,1].$$

则由引理2.4知 $\omega^k\leqslant\bar\omega$. 在定理2.5中令 $d=d^{k-1}$，$g=g^k$，$\eta=m_R$，$\rho=\rho^{k-1}$，$\omega=\omega^{k-1}$，$z=z^{k-1}$，$\dot\alpha=\dot\alpha^{k-1}$，$\alpha=\alpha_k^k+\delta^{k-1}$，并结合(2.2)式，我们有 $\omega^k\leqslant\bar\omega\leqslant\omega^{k-1}-(\rho^{k-1})^2(1-m_R)^2(\omega^{k-1})^2/8(C^k)^2$.

定理2.1 (i)如果算法1.1有限步终止于第k次迭代，则x^k是问题(0.1)的一个最优解；(ii)如果算法1.1在第k次迭代时无限次在步骤1与步骤2之间循环，则x^k是问题(0.1)的一个最优解；(iii)如果算法1.1产生一个无限迭代序列$\{x^k\}$，则其任一聚点x^*都是问题(0.1)的一个最优解.

证明 (i)如果算法1.1有限次终止于点x^k，则$z^k=0$. 根据引理1.2知x^k是问题(0.1)的一个最优解.

(ii)如果算法在第k次迭代时于步骤1与步骤2之间无限次循环，此时因为每次在步骤2中ε^k减半，$(m_R-m_L)\underline{t}>0$，$-z^k=\rho^k\|d^k\|^2+\dot\varepsilon^k\geqslant0$，所以$\varepsilon^k\to0$，$z^k\to0$. 由引理2.2知$x^k$是问题(0.1)的一个最优解.

(iii)现在假设算法1.1产生一个无限迭代序列$\{x^k\}$，且x^*是其任一聚点. 则分两种情况证明x^*是问题(0.1)的一个最优解.

情形1 有无限多个有效步. 此时，必然存在无限指标集$L\subseteq\{1,2,\cdots\}$使得$x^{k(l)}\to x^*$，$l\to\infty$，$l\in L$. 因此，根据引理2.3知x^*是问题(0.1)的一个最优解.

情形2 有效步有限. 则存在一个指标$\bar k$使得$x^k\equiv x^{\bar k}=x^*$且有$\rho^k\equiv\rho$. 如果$\varepsilon^k\to0$，类似与(ii)的证明，知x^*是问题(0.1)的一个最优解. 设对于任意的$k\geqslant\bar k$有$\varepsilon^k\equiv\varepsilon>0$，此时存在一个常数$C>0$使得$C\geqslant\max\{\|\rho d^{k-1}\|,\|g^k\|,\dot\alpha^{k-1},1\}$，$\forall k\geqslant\bar k$. 结合(2.3)式，有

$$\omega^k\leqslant\omega^{k-1}-\rho^2(1-m_R)^2(\omega^{k-1})^2/8C^2,$$

由此可得$\omega^k\to0$，$k\to\infty$，从而$z^k\to0$，$k\to\infty$. 因此，由引理2.2可知x^*是问题(0.1)的一个最优解.

参考文献：

[1] KIWIEL K C. An alternating linearization bundle method for convex optimization and nonlinear multicommodity flow problems[J]. Mathematical Programming，2011，130(1)：59-84.

[2] KIWIEL K C. An algorithm for nonsmooth convex minimization with errors[J]. Mathematics of Computation，1985，171(45)：173-180.

[3] KIWIEL K C. A method of centers with approximate subgradient linearizations for nonsmooth convex optimization[J]. SIAM Journal on Optimization，2007，18(4)：1467-1489.

[4] TANG C M，JIAN J B. Strongly sub-feasible direction method for constrained optimization problems with nonsmooth objective functions[J]. European Journal of Operational Research，2012，218(1)：28-37.

[5] KIWIEL K C. Proximity control in bundle methods for convex nondifferentiable minimization[J]. Mathematical Programming，1990，46(1/2/3)：105-122.

[6] KIWIEL K C. Methods of Descent for Nondifferentiable Optimization[M]. Berlin Heidelberg：Spring-Verlag，1985.

[7] 唐春明，简金宝. 非光滑优化的强次可行方向邻近点束求解方法[J]. 广西科学，2014，21(3)：283-286.
TANG C M，JIAN J B. A proximal bundle method of strongly sub-feasible directions for nonsmooth optimization[J]. Guangxi Sciences，2014，21(3)：283-286.

(责任编辑：米慧芝)