

多面体约束非光滑复合函数的序列有效集方法

时闪闪, 宇振盛

(上海理工大学 理学院, 上海 200093)

摘要: 对带多面体约束的非光滑复合函数问题的求解进行了研究。针对非光滑复合函数问题, 首先, 构造光滑函数来逼近非光滑目标函数, 通过求解光滑近似问题来达到求解原问题的目的。在此基础上, 考虑多面体约束的特殊结构, 运用序列二次规划算法的思想, 利用有效集策略, 通过逐次求解一系列仅含等式约束的二次规划问题来逼近搜索方向的最优解, 再通过线搜索求得步长, 进而得到下一步的迭代点。最后, 从理论上证明了算法的全局收敛性, 并进行了初步的数值实验。将该算法与光滑序列投影收缩算法作对比, 结果表明, 该算法在迭代次数和计算时间上都有一定的优势。

关键词: 光滑近似; 多面体约束; 复合函数; 序列二次规划; 有效集

中图分类号: O 221 **文献标志码:** A

A sequential active set algorithm for nonsmooth composite functions with polyhedral constraints

SHI Shanshan, YU Zhensheng

(College of Science, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: The solution of nonsmooth composite function problem with polyhedral constraints was studied. For the problem of nonsmooth composite function, a smooth function was firstly constructed to approximate the nonsmooth objective function, and the purpose of solving the original problem was then achieved by handling the smooth approximation problem. On this basis, considering the special structure of polyhedral constraints, the idea of sequential quadratic programming algorithm was adopted. A quadratic programming subproblem was solved to obtain the search direction. The active set strategy was adopted by the solution of quadratic programming to approach the optimal solution of the search direction by solving a series of quadratic programming problems with equality constraints step by step. After that, the step size was obtained by line search. The next iteration point was then obtained. Finally, the global convergence of the algorithm was proved theoretically. A preliminary numerical experiment was carried out to compare the algorithm with the smoothed sequential projection shrinkage algorithm. The results show that the algorithm has certain advantages not only in iteration times but also in

收稿日期: 2021-09-01

基金项目: 教育部行指委教育改革创新项目 (HBKC213014)

第一作者: 时闪闪 (1996-), 女, 硕士研究生。研究方向: 最优化理论与方法。E-mail: 18438605319@163.com

通信作者: 宇振盛 (1973-), 男, 教授。研究方向: 最优化理论与方法。E-mail: zhsh-yu@163.com

computing time.

Keywords: smooth approximation; polyhedral constraint; composite function; sequential quadratic programming; active set

1 问题的提出

本文考虑如下复合函数优化问题

$$\begin{cases} \min_x & \phi(\mathbf{x}) := f(\mathbf{x}) + g(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in \Omega \end{cases} \quad (1)$$

式中： $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是连续可微函数； $g: \mathbb{R}^n \rightarrow \mathbb{R}$ 是凸的非光滑函数；可行集 $\Omega = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}_i^T \mathbf{x} = \mathbf{b}_i, i \in \mathcal{M}_E, \mathbf{a}_i^T \mathbf{x} \geq \mathbf{b}_i, i \in \mathcal{M}_I\}$, $\mathcal{M}_E = \{1, \cdots, m\}$, $\mathcal{M}_I = \{m+1, \cdots, p\}$, $\mathbf{a}_i \in \mathbb{R}^n$, $\mathbf{b}_i \in \mathbb{R}$, $i = 1, \cdots, p$ 。

从现实生活中的问题抽象出来的数学模型，可以分为光滑和非光滑两大类。对于光滑问题，由于其梯度及 Hessian 矩阵易于求出，可以使用经典的优化算法求解。而对非光滑问题，函数是不可微的，因此无法直接使用经典的优化算法求解。此外，非光滑复合函数的极小化问题在信息理论^[1]、无线通信^[2-4]、图像恢复^[5-7]、信号处理^[8-9]、变量选择^[4, 10]等领域都有广泛的应用。

当可行集 $\Omega = \mathbb{R}^n$ 时，问题(1)就是无约束优化问题。目前，已经有很多学者提出了各种非常有效的算法^[11-18]解决无约束复合函数问题。比较成熟的算法有迭代收缩阈值算法^[12-13]、临近牛顿法^[14]、分裂算法^[15-16]、邻近点算法^[17-18]等。

实际生活中的很多问题都要求满足一些特定的条件，所以研究带约束的复合函数问题十分有必要。例如，压缩感知理论中的稀疏信号重构问题就是经典的复合函数优化问题。稀疏重构主要解决这两类问题：在对非负信号处理时，有非负约束的限制；进行图像恢复时，灰度强度取值在 0~255 之间。在信号恢复、图像恢复模型里加入约束，可以提高恢复的准确率。同样地，对于实际问题，考虑约束限制，能更加精确地求出符合实际条件的解，从而节约成本。

当可行集 $\Omega = \{\mathbf{l} \leq \mathbf{x} \leq \mathbf{u}\}$ 时，问题(1)就是界约束的复合函数问题。Bian 等^[6]提出光滑二次正则化方法，用 SQR 表示解决图像恢复中出现的界约束非 Lipschitz 连续优化问题。有效集方法是求解约束优化问题的常用方法，采用有效集策略，可以加快算法的收敛速度。张涛^[19]提出了一种求解

界约束优化问题的新积极集信赖域算法。

生产生活中的大多数问题，往往具有更加复杂的约束，因此研究复杂约束的复合函数优化问题更具有现实意义。当可行集 Ω 为多面体约束时，问题(1)就是多面体约束复合函数优化问题。简单约束复合函数优化问题的理论已经发展得十分成熟，但是关于多面体约束的复合函数优化问题，研究工作仍处于基础阶段，还有很多问题尚未解决。Liu 等^[4]提出了光滑序列二次规划框架(SSQP)求解一类多面体上复合函数 L_q 极小化问题。实际上，有效集方法已成功地应用于大规模的线性约束光滑优化问题的求解。William 等^[20]提出了求解多面体约束优化问题的新的有效集算法(PASA)。Zhang 等^[1]提出了求解线性约束非凸非 Lipschitz 优化的光滑有效集算法(SASA)，且用于求解简单约束非光滑复合函数优化问题效果好，即

$$\begin{cases} \min_x & \phi(\mathbf{x}) := f(\mathbf{x}) + \|\mathbf{x}\|_q^q \\ \text{s.t.} & \mathbf{x} \in \Omega \end{cases} \quad (2)$$

也就是本文实际求解的模型：

$$\begin{cases} \min_x & \phi(\mathbf{x}) := f(\mathbf{x}) + \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_q^q \\ \text{s.t.} & \mathbf{x} \in \Omega \end{cases} \quad (3)$$

它是 Liu 等^[4]在求解一类多面体上复合函数 L_q 极小化问题中求解的模型。

这两个问题之间的一个明显区别是复合项 $\|\cdot\|_q^q$ 不同。具体来说：问题(2)试图找到一个解 $\mathbf{x}^*$ ，使 $\mathbf{x}$ 的非零项的数量尽可能得少；而问题(3)尝试找到一个解 $\mathbf{x}^*$ ，使向量 $\mathbf{b} - \mathbf{A}\mathbf{x}$ 正分量的数量尽可能少。问题(2)本质上是一般约束稀疏优化，问题(3)本质上是不等式约束稀疏优化。

以上算法在求解大多数问题时表现良好，但它们不能很好地解决一般多面体约束复合函数问题，例如，SQR 算法^[6]不能求解问题中 $f(\mathbf{x})$ 的复合项 L_q 和一般多面体约束问题。SSQP 算法^[4]的目标函数不具一般性，且由于投影收缩算法是不精确求解子问题，这会使得算法收敛速度慢。PASA 算法^[20]需要求解投影梯度信息，SASA 算法^[1]对线性约束优化器有特殊条件限制，这两种算法都是两

阶段算法, 程序实现难度较大。

序列二次规划 (sequential quadratic programming, SQP) 是求解非线性约束优化问题的主要方法之一, 它在求解中等规模和小规模的非线性问题^[21-22]中表现不俗。有效集算法 (active set algorithm, ASA) 在求解不等式约束优化问题时十分有效, 通过有效集策略, 每次只需求解仅含等式约束的优化问题, 从而可以降低维度, 达到提高算法收敛速度的目的。

本文提出了将 ASA 和 SQP 结合的新算法, 该算法主要借鉴文献 [4] 中提到的求解多面体上最小化复合函数的光滑近似技巧及 SSQP 框架, 将非光滑目标函数转化成光滑函数, 并利用有效集策略求解二次规划子问题。

2 光滑近似

由于问题 (1) 中涉及非光滑项 $\|x\|_q^q$, 导致无法直接使用经典的优化算法来求解。因此先给出近似光滑函数这一重要概念。

定义 1 ^[23] 函数 $\tilde{g}_\mu: \mathbb{R}^n \times \mathbb{R}_+ \rightarrow \mathbb{R}$ 称为非光滑函数 $g: \mathbb{R}^n \rightarrow \mathbb{R}$ 的近似光滑函数, 若 $\forall \mu > 0$, $\tilde{g}_\mu$ 是连续可微的, 且对 $\forall \mu \in \mathbb{R}^n$, 有 $\lim_{z \rightarrow x, \mu \downarrow 0} \tilde{g}(z, \mu) = g(x)$ 。

非光滑优化问题的光滑近似已经得到了广泛的应用^[23-24]。本文对绝对值函数

$$\theta(t) = |t|$$

构造光滑函数^[25]:

$$\theta_\mu(t, \mu) = \begin{cases} t, & t > \frac{\mu}{2} \\ \frac{t^2}{\mu} + \frac{\mu}{4}, & -\frac{\mu}{2} \leq t \leq \frac{\mu}{2} \\ -t, & t < -\frac{\mu}{2} \end{cases} \quad (4)$$

基于式 (4), 可以构造非光滑函数 ϕ 的近似光滑函数 $\tilde{\phi}$ 。定义

$$\begin{cases} \min_x \tilde{\phi}(x, \mu) := f(x) + \tilde{g}(x, \mu) \\ \text{s.t. } x \in \Omega \end{cases} \quad (5)$$

则式 (5) 是问题 (1) 的光滑近似函数。

不难得到 $\theta_\mu(t, \mu)$ 具有的特殊性质。显然地, 对任意固定的 $\mu > 0$, 有

$$\theta_\mu(t, \mu) = \theta(t), \quad \forall t \geq \mu$$

和

$$\theta(t, \mu) \geq \frac{\mu}{4}, \quad \forall t$$

此外, $\theta^q(t, \mu)$ 是连续可微的。除在 $t = \frac{\mu}{2}$ 和 $t = -\frac{\mu}{2}$ 外, $\theta^q(t, \mu)$ 在其他处都是二次连续可微的。

$\theta^q(t, \mu)$ 关于 t 的一阶导数和二阶导数分别为

$$[\theta^q(t, \mu)]' = \begin{cases} qt^{q-1}, & t > \frac{\mu}{2} \\ 2q\theta^{q-1}(t, \mu) \frac{t}{\mu}, & -\frac{\mu}{2} \leq t \leq \frac{\mu}{2} \\ -qt^{q-1}, & t < -\frac{\mu}{2} \end{cases}$$

$$[\theta^q(t, \mu)]'' = \begin{cases} q(q-1)t^{q-2}, & t > \frac{\mu}{2} \\ q\theta^{q-1}(t, \mu) \frac{2}{\mu} + q(q-1), & -\frac{\mu}{2} \leq t \leq \frac{\mu}{2} \\ -q(q-1)t^{q-2}, & t < -\frac{\mu}{2} \end{cases}$$

接下来, 给出一个非常重要的引理。

引理 1 对 $\forall q \in (0, 1)$, $\mu \in (0, +\infty)$,

$$\begin{aligned} \text{a. } 0 \leq \theta^q(t, \mu) - \theta^q(t) &\leq \left(\frac{\mu}{4}\right)^q, \quad \forall t \leq \frac{\mu}{2}, \text{ 且有 } [\theta^q(t, \mu)]'' \leq \\ 16q\mu^{q-2}, \quad \forall t \notin \left\{-\frac{\mu}{2}, \frac{\mu}{2}\right\}. \end{aligned}$$

b. 定义

$$\kappa(t, \mu) = \begin{cases} 16q\mu^{q-2}, & 0 \leq t \leq \mu \\ 0, & \text{其他} \end{cases}$$

对 $\forall t$ 和 $\hat{t}$, 如果 $\hat{t} > \mu$ 使得 $t - \hat{t} \geq -\frac{\hat{t}}{2}$, 或者 $0 \leq t \leq \mu$ 时, $t \in (-\infty, \infty)$, 或者 $\hat{t} < 0$ 时, $t - \hat{t} \leq 0$, 有

$$\theta^q(t, \mu) \leq \theta^q(\hat{t}, \mu) + [\theta^q(\hat{t}, \mu)]'(t - \hat{t}) + \frac{\kappa(\hat{t}, \mu)}{2}(t - \hat{t})^2$$

成立。

引理 1 的详细证明参见文献 [4] 中的引理 4.1。

3 算法

本文受文献 [4] 的启发, 采用该文献中的 SSQP 框架, 提出了序列有效集算法 (SQP-ASA), 在每一迭代步中利用有效集策略求解一个凸二次规划 (QP) 子问题。此 QP 子问题的目标函数是问题 (1) 光滑近似后式 (5) 的目标函数 $\tilde{\phi}(x, \mu)$ 的局部上界。光滑参数的更新依赖于问题 (1) 的残差。

对固定的光滑参数 $\mu > 0$, 定义非光滑函数 $g(\cdot)$ 的光滑近似函数 $\tilde{g}(\cdot, \mu)$ 在 x_k 处的二次近似为

$$Q_1(x, x_k, \mu) = \tilde{g}(x_k, \mu) + \nabla \tilde{g}(x_k, \mu)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \tilde{B}(x_k, \mu)(x - x_k)$$

式中, $\tilde{g}(x_k, \mu) = \sum_{i=1}^n \theta^q(x_k, \mu)$, $\tilde{B}(x, \mu) = \text{Diag}(\kappa(x_1, \mu), \dots, \kappa(x_n, \mu))$ 。

类似地, 定义光滑函数 $f(x)$ 在 x_k 处的二次近

似为

$$Q_2(\mathbf{x}, \mathbf{x}_k) = f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T (\mathbf{x} - \mathbf{x}_k) + \frac{1}{2} L_k \|\mathbf{x} - \mathbf{x}_k\|^2$$

假设 $f(\mathbf{x})$ 的梯度是 Lipschitz 连续的, 即

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq L \|\mathbf{x} - \mathbf{y}\|_2, \forall \mathbf{x}, \mathbf{y} \in \Omega$$

定义

$$Q(\mathbf{x}, \mathbf{x}_k, \mu) = Q_1(\mathbf{x}, \mathbf{x}_k, \mu) + Q_2(\mathbf{x}, \mathbf{x}_k) \tag{6}$$

由于 $Q_1(\mathbf{x}, \mathbf{x}_k, \mu)$ 和 $Q_2(\mathbf{x}, \mathbf{x}_k)$ 分别是 $\tilde{g}(\cdot, \mu)$ 和 $f(\mathbf{x})$ 在 $\mathbf{x}_k$ 处的二次近似, 因此, 可以把 $Q_1(\mathbf{x}, \mathbf{x}_k, \mu)$ 看作是 $\tilde{\phi}(\mathbf{x}, \mu)$ 在 $\mathbf{x}_k$ 处的二次近似。

下面的引理给出了凸二次规划(6)的目标函数 $Q(\mathbf{x}, \mathbf{x}_k, \mu)$ 是式(5)的目标函数 $\tilde{\phi}(\mathbf{x}, \mu)$ 之间的大小关系, 即 $Q(\mathbf{x}, \mathbf{x}_k, \mu)$ 是 $\tilde{\phi}(\mathbf{x}, \mu)$ 在 $\mathbf{x}_k$ 处的一个上界, 这保证了问题(6)的最小值是问题(5)最小值的上界。

引理 2 对任意的 $\mathbf{x}_k, \mathbf{x}$, 使得

$$\begin{aligned} (\mathbf{x}_k - \mathbf{x})_e &\leq 0, e \in I_{\mathbf{x}_k}^\mu \\ (\mathbf{x}_k - \mathbf{x})_e &\geq -\frac{(\mathbf{x}_k)_e}{2}, e \in \mathcal{J}_{\mathbf{x}_k}^\mu \end{aligned}$$

其中

$$\begin{aligned} I_{\mathbf{x}_k}^\mu &= \{e | (\mathbf{x}_k)_e < 0\} \\ \mathcal{J}_{\mathbf{x}_k}^\mu &= \{e | (\mathbf{x}_k)_e > 0\} \end{aligned}$$

如果 $L_k > 0$, $f(\mathbf{x}) \leq Q_2(\mathbf{x}, \mathbf{x}_k)$, 那么 $\tilde{\phi}(\mathbf{x}, \mu) \leq Q(\mathbf{x}, \mathbf{x}_k, \mu)$ 。

该引理的证明可参考文献 [4] 中的引理 4.2。

基于引理 2, 提出序列有效集二次规划算法。

本文研究的问题(2)光滑后可以表示为

$$\begin{cases} \min_{\mathbf{x}} & \tilde{\phi}(\mathbf{x}, \mu) := f(\mathbf{x}) + \tilde{g}(\mathbf{x}, \mu) \\ \text{s.t.} & \mathbf{a}_i^T \mathbf{x} = \mathbf{b}_i, i \in \mathcal{M}_E \\ & \mathbf{a}_i^T \mathbf{x} \geq \mathbf{b}_i, i \in \mathcal{M}_I \end{cases} \tag{7}$$

式中: 函数 $\tilde{\phi}$ 是连续可微的; 指标集 $\mathcal{M}_E = \{1, \dots, m\}$, $\mathcal{M}_I = \{m+1, \dots, p\}$; $\mathbf{a}_i \in \mathbb{R}^n$, $\mathbf{b}_i \in \mathbb{R}$, $i = 1, \dots, p$ 。

那么, 问题(7)在 $\mathbf{x}_k$ 处的 QP 子问题为

$$\begin{cases} \min & Q(\mathbf{x}, \mathbf{x}_k, \mu) \\ \text{s.t.} & (\mathbf{x}_k - \mathbf{x})_e \leq 0, e \in I_{\mathbf{x}_k}^\mu \\ & (\mathbf{x}_k - \mathbf{x})_e \geq -\frac{(\mathbf{x}_k)_e}{2}, e \in \mathcal{J}_{\mathbf{x}_k}^\mu \\ & \mathbf{a}_i^T \mathbf{x} = \mathbf{b}_i, i \in \mathcal{M}_E \\ & \mathbf{a}_i^T \mathbf{x} \geq \mathbf{b}_i, i \in \mathcal{M}_I \end{cases} \tag{8}$$

令 $\mathbf{d}_k = \mathbf{x} - \mathbf{x}_k$, $\mathbf{H}_k = L_k \mathbf{I}_n + \tilde{\mathbf{B}}(\mathbf{x}_k, \mu)$, 则

$$Q(\mathbf{x}, \mathbf{x}_k, \mu) = \nabla \tilde{\phi}(\mathbf{x}_k, \mu)^T \mathbf{d}_k + \frac{1}{2} \mathbf{d}_k^T \mathbf{H}_k \mathbf{d}_k$$

求解问题(7)的算法思想为: 在每一迭代步中, 首先使用有效集方法求出其二次规划子问题(8)的最优解 $\mathbf{d}^*$, 并将它作为问题(7)下一步的搜索方向 $\mathbf{d}_k$, 再利用价值函数通过 Armijo 线搜索求

出步长 α_k , 进而得到问题(7)的下一步迭代点 $\mathbf{x}_{k+1}$, 重复这一过程, 直到求得问题(7)的最优解 $\mathbf{x}^*$ 。具体求解过程如算法 1 所示。

算法 1 序列有效集算法第 k 步迭代

步骤 1 选择初始可行点 $\mathbf{x}_k$ 和参数 $\epsilon \in (0, 1]$, $0 < \sigma < 1$, $\eta > 1$ 和 $L^{\max} \geq L_0 \geq L^{\min} > 0$ 。令 $z = 0$ (外部迭代索引), $k = 0$ (内部迭代索引), $\mu_0 = \frac{\epsilon}{\sigma^{\lfloor \log_{\sigma} \epsilon \rfloor}} \in (\sigma, 1]$ 。

步骤 2 令 $\mu = \mu_z$, 如果 $\|\mathbf{d}_k\| \leq \mu$, 执行步骤 5; 否则执行步骤 3。

步骤 3 参照文献 [26] 中的有效集方法求解 QP 子问题, 得到 $\mathbf{d}_k$ 。通过价值函数 $\psi(\mathbf{x}, \mu, \vartheta)$, 根据 Armijo 线搜索求出步长 β_k , 进而得到 $\mathbf{x}_{k+1} = \mathbf{x}_k + \beta_k \mathbf{d}_k$ 和 r_k , 其中

$$r_k = \frac{f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) - \nabla f(\mathbf{x}_k)^T (\mathbf{x}_{k+1} - \mathbf{x}_k)}{\frac{1}{2} L_k \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2}$$

步骤 4 令 $\mathbf{s}_{k+1} = \mathbf{x}_{k+1} - \mathbf{x}_k$, $\mathbf{y}_{k+1} = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$, 如果 $r_k \leq 1$, $L_{k+1} = \max \left\{ \min \left\{ L^{\max}, \frac{\mathbf{s}_{k+1}^T \mathbf{y}_{k+1}}{\|\mathbf{s}_{k+1}\|^2}, L^{\min} \right\} \right\}$, $k = k + 1$, 执行步骤 2, 否则 $L_k = \eta L_k$, 执行步骤 3。

步骤 5 如果 $\mu \leq \epsilon$, 则终止该算法; 否则, 转到步骤 6。

步骤 6 令 $\mu_{z+1} = \sigma \mu_z$, $z = z + 1$, $\mathbf{x}_0 = \mathbf{x}_k$, $k = 0$, 然后转到步骤 2。

令 $\mathbf{x} - \mathbf{x}_k = \mathbf{d}$, 则可将 QP 子问题(8)转化与 $\mathbf{d}$ 有关的子问题求解。通过积极集策略, 会有越来越多的 $\mathbf{x}_k$ 逐渐满足 $\mathbf{a}_i^T \mathbf{x}_k = \mathbf{b}_i$, $i \in \mathcal{S}_k$ 。这样问题就转化为了等式约束优化问题, 再利用拉格朗日乘子算法对其进行求解。因此, 本文的算法可以在大大减少问题维数的同时提高计算的效率。

接下来, 利用文献 [26] 中的有效集方法求解 QP 子问题, 如算法 2 所示。

算法 2 有效集方法第 k 步迭代

步骤 1 选择初始可行点 $\forall \mathbf{x}_0, \mathbf{d}_0 \in \Omega$, 对称正定矩阵 $\mathbf{H}_0$, 令 $\forall \mathbf{x} \in \Omega$ 。

步骤 2 确定相应的有效集 $\mathcal{S}_k$, 求解子问题

$$\begin{aligned} \min & q_k(\mathbf{t}) = \frac{1}{2} \mathbf{t}^T \mathbf{H}_k \mathbf{t} + \mathbf{C}_k^T \mathbf{t} \\ \text{s.t.} & \mathbf{a}_i^T \mathbf{t} = 0, i \in \mathcal{S}_k \end{aligned}$$

得极小点 $\mathbf{d}_k$ 和拉格朗日乘子 λ_k , 其中 $\mathbf{C}_k = \mathbf{H}_0 \mathbf{d}_k + \mathbf{g}_0$, 如果 $t_k = 0$, 执行步骤 3; 否则执行步骤 4。

步骤 3 令 $(\lambda_k)_l = \min_{i \in \mathcal{S}_k} \{(\lambda_k)_l\}$, 如果 $(\lambda_k)_l \geq 0$, 则 $\mathbf{d}_k$ 是全局极小点, 停算。否则, 若 $(\lambda_k)_l < 0$, 则令 $\mathcal{S}_k := \mathcal{S}_k \setminus \{l\}$, 转步骤 2。

步骤4 令 $\alpha_k = \min\{1, \bar{\alpha}_k\}$, 其中

$$\bar{\alpha}_k = \min_{i \in S_k} \left\{ \frac{c_i - \mathbf{a}_i^T \mathbf{d}_k}{\mathbf{a}_i^T \mathbf{t}_k} \mid \mathbf{a}_i^T \mathbf{t}_k < 0 \right\}$$

令 $\mathbf{d}_{k+1} := \mathbf{d}_k + \alpha_k \mathbf{t}_k$, 其中 $c_i = \mathbf{b}_i - \mathbf{a}_i^T \mathbf{x}_0$ 。

步骤5 若 $\alpha_k = 1$, 则令 $S_{k+1} := S_k$; 否则, 若 $\alpha_k < 1$, 则令 $S_{k+1} := S_k \cup \{j_k\}$, 其中 $\{j_k\}$ 满足

$$\bar{\alpha}_k = \frac{c_{j_k} - \mathbf{a}_{j_k}^T \mathbf{d}_k}{\mathbf{a}_{j_k}^T \mathbf{t}_k}。$$

步骤6 $k = k + 1$, 转步骤2。

此处的价值函数 $\psi(\mathbf{x}, \mu, \vartheta) = \nabla \tilde{\phi}(\mathbf{x}, \mu)$ 。这是因为由有效集方法得到的方向 $\mathbf{d}_k$ 一定是可行方向。

4 收敛性分析

为证明算法1的全局收敛, 现给出如下基本假设:

假设1 对 $\forall \mathbf{x} \in \Omega$, 线性独立约束限定 (LICQ) 成立, 即梯度 $\mathbf{a}_i$ ($i \in S_k$) 线性无关。

由 $\mathbf{H}_k$ 的构造可知, $\mathbf{H}_k$ 一定是对称正定的。本文提出的 SQP-ASA 算法通过求解如下形式的二次规划问题

$$\begin{cases} \min & \nabla \tilde{\phi}(\mathbf{x}_k, \mu)^T \mathbf{d} + \frac{1}{2} \mathbf{d}^T \mathbf{H}_k \mathbf{d} \\ \text{s.t.} & \mathbf{a}_i^T \mathbf{d} + \mathbf{a}_i^T \mathbf{x}_k = \mathbf{b}_i, i \in M_E \\ & \mathbf{a}_i^T \mathbf{d} + \mathbf{a}_i^T \mathbf{x}_k \geq \mathbf{b}_i, i \in M_I \end{cases} \quad (9)$$

得到解 $\mathbf{d}^*$, 作为原问题变量 $\mathbf{x}$ 在第 k 次迭代过程中的搜索方向 $\mathbf{d}_k$ 。在求解过程中利用有效集策略, 通过逐次求解仅含带等式约束的相关二次规划问题 (9) 来逼近其最优解。

下面证明 SQP-ASA 算法的全局收敛性, 即证该算法产生的点列 $\{\mathbf{x}_k\}$ 的任何聚点 $\mathbf{x}^*$ 都是问题 (7) 的 KT 点。

引理3 由算法2中的有效集方法求得的 QP 子问题方向 $\mathbf{d}_k$ 一定是原问题 (7) 的可行下降方向。

证明 先证明 $\mathbf{d}_k$ 是可行方向。依据有效集的求解过程可知: $\mathbf{d}_k$ 是 QP 子问题 (9) 的可行方向, 从而有 $\mathbf{a}_i^T \mathbf{d}_k \geq c_i$, $c_i = \mathbf{b}_i - \mathbf{a}_i^T \mathbf{x}_k$, $\mathbf{a}_i^T \mathbf{d}_k = 0$, $i \in S_k$ 。又因初始点 $\mathbf{x}_k$ 是可行点, 则有 $\mathbf{a}_i^T \mathbf{x}_k \geq \mathbf{b}_i$, $\mathbf{a}_i^T \mathbf{x}_k = \mathbf{b}_i$, $i \in S_k$ 。设由序列二次规划求出的下一步迭代点是 $\mathbf{x}_{k+1}$, 则

$$\begin{aligned} \mathbf{a}_i^T \mathbf{x}_{k+1} &= \mathbf{a}_i^T (\mathbf{x}_k + \beta_k \mathbf{d}_k) = \mathbf{a}_i^T \mathbf{x}_k + \beta_k \mathbf{a}_i^T \mathbf{d}_k = \\ & \mathbf{a}_i^T \mathbf{x}_k + 0 = \mathbf{b}_i, i \in M_E \end{aligned}$$

$$\begin{aligned} \mathbf{a}_i^T \mathbf{x}_{k+1} &= \mathbf{a}_i^T (\mathbf{x}_k + \beta_k \mathbf{d}_k) = \mathbf{a}_i^T \mathbf{x}_k + \beta_k \mathbf{a}_i^T \mathbf{d}_k \geq \\ & \mathbf{a}_i^T \mathbf{x}_k + \beta_k (\mathbf{b}_i - \mathbf{a}_i^T \mathbf{x}_k) \geq (1 - \beta_k) \mathbf{a}_i^T \mathbf{x}_k + \\ & \beta_k \mathbf{b}_i \geq (1 - \beta_k) \mathbf{b}_i + \beta_k \mathbf{b}_i = \mathbf{b}_i, i \in M_I \end{aligned}$$

因此, $\mathbf{d}_k$ 一定是原问题 (7) 的可行方向。

其次, 证明方向 $\mathbf{d}_k$ 是下降的。对 $\forall \mu \in (0, 1]$, 利用 QP 子问题 (9) 的 KT 条件, 有

$$\begin{cases} \nabla \tilde{\phi}(\mathbf{x}_k, \mu) + \mathbf{H}_k \mathbf{d}_k = \sum_{i \in M_I} \lambda_i \mathbf{a}_i + \sum_{i \in M_E} \lambda_i \mathbf{a}_i \\ \mathbf{a}_i^T \mathbf{d}_k = 0, i \in M_E \\ \mathbf{a}_i^T \mathbf{d}_k \geq c_i, c_i = \mathbf{b}_i - \mathbf{a}_i^T \mathbf{x}_k, i \in M_I \\ \lambda_i \geq 0, i \in M_I; \lambda_i = 0, i \in M \setminus S_k \end{cases} \quad (10)$$

由式 (10) 的第一个式子可得

$$\begin{aligned} \nabla \tilde{\phi}(\mathbf{x}_k, \mu) \mathbf{d}_k &= -\mathbf{d}_k^T \mathbf{H}_k \mathbf{d}_k - \sum_{i \in M_I} \lambda_i (\mathbf{a}_i^T \mathbf{x}_k - \mathbf{b}_i) - \\ & \sum_{i \in M_E} \lambda_i (\mathbf{a}_i^T \mathbf{x}_k - \mathbf{b}_i) = -\mathbf{d}_k^T \mathbf{H}_k \mathbf{d}_k - \\ & \sum_{i \in S_k} \lambda_i (\mathbf{a}_i^T \mathbf{x}_k - \mathbf{b}_i) - \sum_{i \in M \setminus S_k} \lambda_i (\mathbf{a}_i^T \mathbf{x}_k - \mathbf{b}_i) \end{aligned}$$

当 $i \in S_k$ 时, $\mathbf{a}_i^T \mathbf{x}_k - \mathbf{b}_i = 0$; 当 $i \in M \setminus S_k$ 时, $\lambda_i = 0$,

从而 $\sum_{i \in M} \lambda_i (\mathbf{a}_i^T \mathbf{x}_k - \mathbf{b}_i) = 0$, 即 $\nabla \tilde{\phi}(\mathbf{x}_k, \mu) \mathbf{d}_k = -\mathbf{d}_k^T \mathbf{H}_k \mathbf{d}_k$ 。又因为 $\mathbf{H}_k$ 是对称正定矩阵, 所以对任何非负向量 $\mathbf{d}_k$ 都有 $\mathbf{d}_k^T \mathbf{H}_k \mathbf{d}_k > 0$, 故 $\nabla \tilde{\phi}(\mathbf{x}_k, \mu) \mathbf{d}_k < 0$ 。

引理4 对于上述二次规划子问题 (9), 设其 KT 点对为 $(\mathbf{d}(\mathbf{x}), \tilde{\lambda}(\mathbf{x}))$ 。如果 $\mathbf{d}(\mathbf{x}) = 0$, 则 $\mathbf{x}$ 为问题 (7) 的 KT 点。

证明 不妨设

$$\lambda(\mathbf{x}) = (\lambda_i(\mathbf{x}), i \in M), \quad \lambda_i(\mathbf{x}) = \begin{cases} \tilde{\lambda}_i(\mathbf{x}), i \in S(\mathbf{x}) \\ 0, i \in M \setminus S(\mathbf{x}) \end{cases}$$

如果 $\mathbf{d}(\mathbf{x}) = 0$, 依据有效集方法求解方向 $\mathbf{d}(\mathbf{x})$ 的过程, 结合问题 (7) 的 KT 条件可知, $\tilde{\lambda}_i(\mathbf{x}) \geq 0$, 否则, $\mathbf{d}(\mathbf{x})$ 不是二次规划问题 (9) 的最优解, 与 KT 点对这个条件相矛盾。因此

$$\begin{cases} \nabla \tilde{\phi}(\mathbf{x}, \mu) = \sum_{i \in M_I} \lambda_i \mathbf{a}_i + \sum_{i \in M_E} \lambda_i \mathbf{a}_i \\ \mathbf{a}_i^T \mathbf{x} = \mathbf{b}_i, i \in M_E \\ \mathbf{a}_i^T \mathbf{x} \geq \mathbf{b}_i, i \in M_I \\ \lambda_i \geq 0, i \in M_I; \lambda_i = 0, i \in M \setminus S(\mathbf{x}) \end{cases}$$

这说明 $\mathbf{x}$ 是问题 (7) 的一个 KT 点。

定理1 由序列有效集算法生成的序列 $\{\mathbf{x}_k\}$ 的任何聚点 $\mathbf{x}^*$ 都是原问题的 KT 点。

证明 对 $\forall \mu \in (0, 1]$, 当该算法是有限步终止时, 由引理4可知, 算法产生的点列 $\{\mathbf{x}_k\}$ 的任何聚点 $\mathbf{x}^*$ 都是问题 (7) 的 KT 点。

下设该算法产生无穷点列 $\{\mathbf{x}_k\}$, $\mathbf{x}^*$ 为其任一给

定聚点。由于 $S_k \in \mathcal{M}$ 为有限集，不妨设存在无穷子集 K ，使 $\mathbf{x}_k \rightarrow \mathbf{x}^*$ ， $S_k \equiv S$ ， $k \in K$ 。为不失一般性，不妨设由 SQP-ASA 算法产生的迭代点 $\mathbf{x}_{k+1} = \mathbf{x}_k + \beta_k \mathbf{d}_k (\forall k \in K)$ ，而由引理 3 可知， $\tilde{\phi}(\mathbf{x}_k, \mu)$ 单调下降，从而由 $\mathbf{x}_k \rightarrow \mathbf{x}^*$ ， $k \in K$ 可以得到 $\tilde{\phi}(\mathbf{x}_k, \mu) \rightarrow \tilde{\phi}(\mathbf{x}^*, \mu)$ ， $k \rightarrow \infty$ 。

一方面，由于步长 β 是由 Armijo 线搜索得到的，所以 $\beta \in (0, 1]$ 。另一方面，通过引理 3 知 $\mathbf{d}_k$ 是下降方向，不难得到

$$0 = \lim_{k \in K} (\tilde{\phi}(\mathbf{x}_{k+1}, \mu) - \tilde{\phi}(\mathbf{x}_k, \mu)) \leq \lim_{k \in K} \beta \nabla \tilde{\phi}(\mathbf{x}_k, \mu)^T \mathbf{d}_k \leq$$
$$\lim_{k \in K} \beta (-\mathbf{d}_k^T \mathbf{H}_k \mathbf{d}_k) \leq 0$$

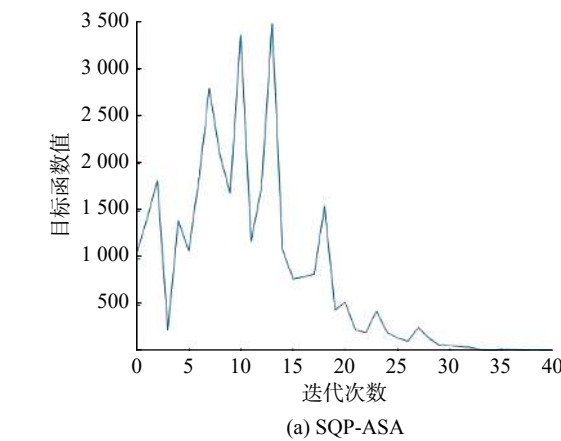
故可知 $\mathbf{d}_k \rightarrow 0$ ， $k \in K$ 。从而 $\mathbf{a}_i^T \mathbf{x}^* - b_i = 0$ ， $i \in S$ ， $S \subseteq \mathcal{M}(\mathbf{x}^*)$ 。那么

$$\begin{cases} \nabla \tilde{\phi}(\mathbf{x}^*, \mu) = \sum_{i \in \mathcal{M}_I} \lambda_i^* \mathbf{a}_i + \sum_{i \in \mathcal{M}_E} \lambda_i^* \mathbf{a}_i \\ \mathbf{a}_i^T \mathbf{x}^* = b_i, i \in \mathcal{M}_E \\ \mathbf{a}_i^T \mathbf{x}^* \geq b_i, i \in \mathcal{M}_I \\ \lambda_i^* \geq 0, i \in \mathcal{M}_I; \lambda_i^* = 0, i \in \mathcal{M} \setminus S \end{cases}$$

这说明 $\mathbf{x}^*$ 是问题(7)的一个 KT 点。

5 数值计算

利用 Matlab R2016a 实现算法 SQP-ASA。其中



试验环境为：Intel(R) Core(TM) i5-5200U CPU@ 2.20GHz 4.00GB RAM。通过求解下列问题的稀疏解来检验算法的有效性。

例 1 非负箱约束稀疏信号恢复的模型如下

$$\min_x \phi(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$$
$$\text{s.t. } 0 \leq \mathbf{x} \leq 255$$

寻求该模型的稀疏解，可以转化为求解模型：

$$\min_x \phi(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \|\mathbf{x}\|_q^q$$
$$\text{s.t. } 0 \leq \mathbf{x} \leq 255$$

分别运用本文提出的序列有效集算法 (SQP-ASA) 和文献 [4] 中的序列投影收缩算法 (SQP-PG)，求出该模型的稀疏解。相关参数取法如下： $\mu_0 = 0.1$ ， $\sigma_2 = 0.0001$ ， $\mathbf{A} = \text{randi}([-10, 10], 50, 50)$ ， $\mathbf{b} = \text{randi}([-10, 10], 50, 1)$ ， $L_0 = \|2 \times (\mathbf{A})^T \mathbf{A}\|$ ， $L^{\max} = 3000$ ， $L^{\min} = 1137.5$ ，精度参数 $\epsilon = 10^{-5}$ ，初始点 $\mathbf{d}_0 = 254 \times \text{randi}([0, 1], 50, 1)$ ，最大迭代次数为 1000。目标函数与迭代次数的关系如图 1 所示。

这两种方法的迭代次数和迭代时间的比较见表 1。从表 1 可以看出，对于同样的初始点，SQP-ASA 算法比 SQP-PG 算法的迭代次数少，运行效率也更高。

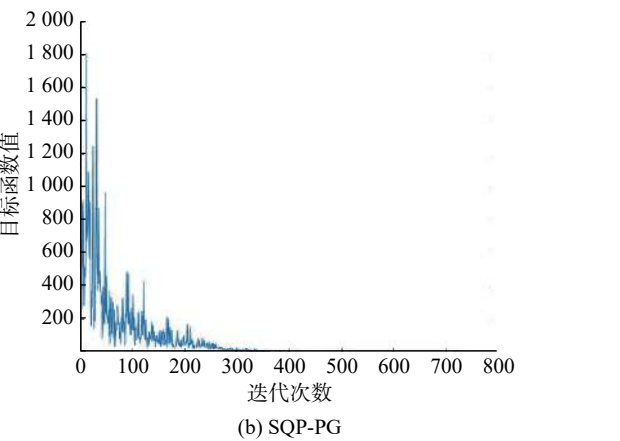


图 1 迭代次数比较

Fig.1 Comparison of iteration times

表 1 例 1 的数值结果
Tab.1 Numerical results of example 1

例1	初始值	迭代次数	CPU时间/s	最优解
SQP-ASA	(50,1)	40	0.0480	(50,1)
SQP-PG	(50,1)	715	0.3031	(50,1)

例 2 寻求如下函数

$$\min_x \phi(\mathbf{x}) = f(\mathbf{x})$$
$$\text{s.t. } \mathbf{x} \in \Omega$$

的稀疏解。可以转化为求解模型

$$\min_x \phi(\mathbf{x}) = f(\mathbf{x}) + \|\mathbf{x}\|_q^q$$
$$\text{s.t. } \mathbf{x} \in \Omega$$

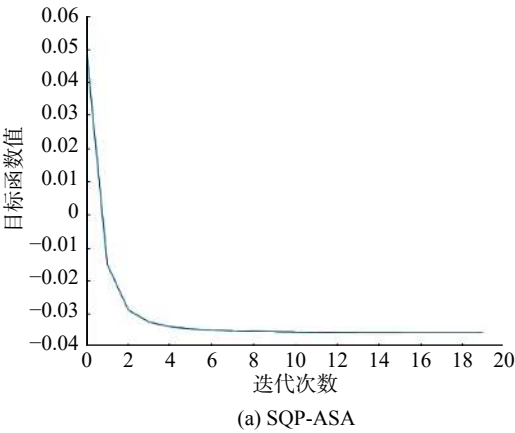
本文通过求解以下两个模型在不同约束条件下的稀疏解，来验证算法的有效性。两个模型均来自文献 [26]。

模型 1

$$f(\mathbf{x}) = \frac{1}{2}x_1^2 - x_1x_2 + x_2^2 - 6x_1 - 2x_2$$

$$\Omega = \begin{cases} -2x_1 - x_2 + 3 \geq 0 \\ x_1 - x_2 + 1 \geq 0 \\ -x_1 - 2x_2 + 2 \geq 0 \\ x_1, x_2 \geq 0 \end{cases}$$

分别使用 SQP-ASA 算法和 SQP-PG 算法, 求出该模型的稀疏解。这里的相关参数取法如下:
 $\mu_0 = 0.01, \xi = 0.5, \sigma = 0.1, \eta = 1.2, L^{\max} = 20, L^{\min} = 3,$



$L_0 = 3$, 精度参数 $\epsilon = 10^{-5}$, 初始 $d_0 = [0.5 \ -1]^T$, 最大迭代次数为 1 000。迭代过程如图 2 所示。
这两种方法的迭代次数和迭代时间的比较见表 2。从表 2 可以看出, 对于同样的初始点, SQP-ASA 算法比 SQP-PG 算法的迭代次数少, 运行效率也更高。

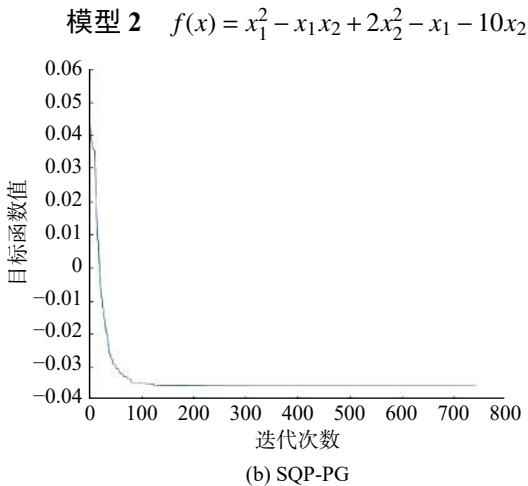


图 2 迭代次数比较
Fig.2 Comparison of iteration times

表 2 模型 1 的数值结果
Tab.2 Numerical results of model 1

算法	初始值	迭代次数	CPU时间/s	最优解
SQP-ASA	(0,1)	19	0.0341	(0.0034,0.7036)
SQP-PG	(0,1)	743	0.6895	(0.0034,0.7036)

$$\Omega = \begin{cases} -3x_1 - 2x_2 + 6 \geq 0 \\ x_1, x_2 \geq 0 \end{cases}$$

分别使用 SQP-ASA 算法和 SQP-PG 算法, 求出该模型的稀疏解。此时的相关参数取法如下:
 $\mu_0 = 0.01, \xi = 0.5, \sigma = 0.1, \eta = 1.2, L^{\max} = 20, L^{\min} = 2.181, L_0 = 2.618$, 精度参数 $\epsilon = 10^{-5}$, 初始 $d_0 = [0 \ -3]^T$, 最大迭代次数为 1 000。目标函数与迭代次数的关系如图 3 所示。

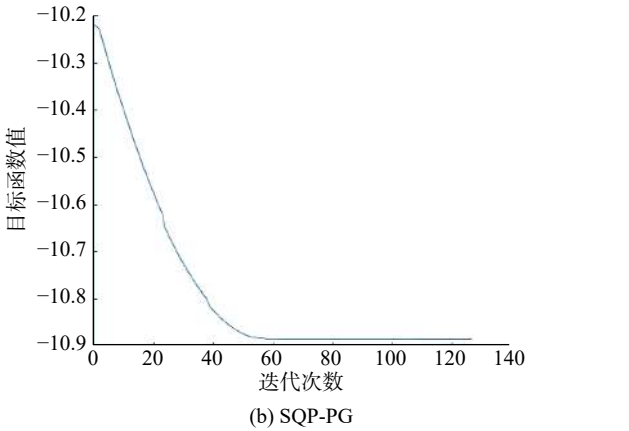
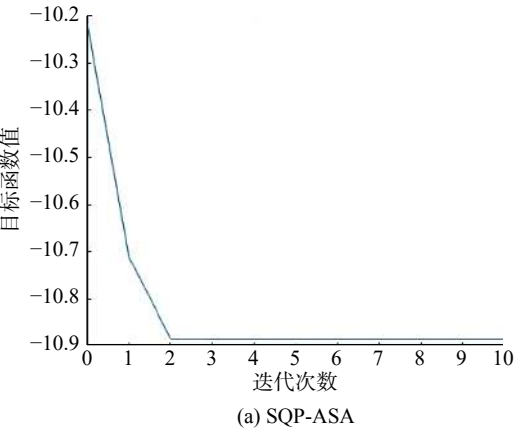


图 3 迭代次数比较
Fig.3 Comparison of iteration times

这两种方法的迭代次数和迭代时间的比较见表 3。从表 3 可以看出, 对于同样的初始点, SQP-ASA 算法比 SQP-PG 算法的迭代次数少, 运行效率也更高。

表 3 模型 2 的数值结果
Tab.3 Numerical results of model 2

算法	初始值	迭代次数	CPU时间/s	最优解
SQP-ASA	(0,3)	10	0.0677	(0.0017,2.4201)
SQP-PG	(0,3)	128	0.0800	(0.0017,2.4201)

6 结 论

提出了一种新的求解多面体约束非光滑复合函数优化问题的序列有效集算法。首先,将非光滑目标函数转化为光滑目标函数;然后,在每一迭代步中用有效集方法求解一个二次规划子问题得到方向,利用价值函数通过线搜索取得步长,重复这一过程直到求得原问题的解;最后,证明了该算法的全局收敛性。此外,通过实例进行了数值验证,结果表明,该方法较序列投影收缩方法具有一定优势。

参考文献:

- [1] ZHANG C, CHEN X J. A smoothing active set method for linearly constrained non-lipschitz nonconvex optimization[J]. *SIAM Journal on Optimization*, 2020, 13(1): 1–30.
- [2] COMBETTES P L, PESQUET J C. Proximal splitting methods in signal processing[M]//BAUSCHKE H H, BURACHIK R S, COMBETTES P L, et al. *Fixed-point Algorithms for Inverse Problems in Science and Engineering*. New York: Springer, 2011.
- [3] LIU Y F, DAI Y H, LUO Z Q. Joint power and admission control via linear programming deflation[J]. *IEEE Transactions on Signal Processing*, 2013, 61(6): 1327–1338.
- [4] LIU Y F, MA S Q, DAI Y H, et al. A smoothing SQP framework for a class of composite L_q minimization over polyhedron[J]. *Mathematical Programming*, 2016, 158(1/2): 467–500.
- [5] NESTEROV Y. Gradient methods for minimizing composite functions[J]. *Mathematical Programming*, 2013, 140(1): 125–161.
- [6] BIAN W, CHEN X J. Worst-case complexity of smoothing quadratic regularization methods for non-lipschitzian optimization[J]. *SIAM Journal on Optimization*, 2013, 23(3): 1718–1741.
- [7] WANG W N, WU C L, TAI X C. A globally convergent algorithm for a constrained non-lipschitz image restoration model[J]. *Journal of Scientific Computing*, 2020, 83(1): 14.
- [8] HUANG Y K, LIU H W. A Barzilai-borwein type method for minimizing composite functions[J]. *Numerical Algorithms*, 2015, 69(4): 819–838.
- [9] MANJU V N, FRED A L. Sparse decomposition technique for segmentation and compression of compound images[J]. *Journal of Intelligent Systems*, 2019, 29(1): 515–528.
- [10] HU J H, JIAN H, FENG Q. A group adaptive elastic-net approach for variable selection in high-dimensional linear regression[J]. *Science China Mathematics*, 2018, 61(1): 173–188.
- [11] SARABI M E. Primal superlinear convergence of SQP methods in piecewise linear-quadratic composite optimization[J]. *Set-Valued and Variational Analysis*, 2022, 30(1): 1–37.
- [12] BECH A, TEBoulLE M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. *SIAM Journal of Imaging Sciences*, 2009, 2(1): 183–202.
- [13] 李启朋. 非光滑凸优化问题的快速迭代收缩阈值算法研究[D]. 西安: 西安电子科技大学, 2018.
- [14] KANZOW C, LECHNER T. Globalized inexact proximal newton-type methods for nonconvex composite functions[J]. *Computational Optimization and Applications*, 2021, 78(2): 377–410.
- [15] CRUZ J Y B. On proximal subgradient splitting method for minimizing the sum of two nonsmooth convex functions[J]. *Set-Valued and Variational Analysis*, 2017, 25(2): 245–263.
- [16] PATRASCU A, IROFTI P. Stochastic proximal splitting algorithm for composite minimization[J]. *Optimization Letters*, 2021, 15(6): 2255–2273.
- [17] ZHANG X Y, BARRIO R, MARTÍNEZ M A, et al. Bregman proximal gradient algorithm with extrapolation for a class of nonconvex nonsmooth minimization problems[J]. *IEEE Access*, 2019, 7: 126515–126529.
- [18] NECOARA I. General convergence analysis of stochastic first-order methods for composite optimization[J]. *Journal of Optimization Theory and applications*, 2021, 189(1): 66–95.
- [19] 张涛. 一种针对盒子约束优化问题带有新积极集策略的信赖域算法[D]. 北京: 北京交通大学, 2016.
- [20] HAGER W W W, ZHANG H C. An active set algorithm for nonlinear optimization with polyhedral constraints[J]. *Science China Mathematics*, 2016, 59(8): 89–106.
- [21] FLETCHER R. Practical methods of optimization[J]. *SIAM Review*, 1984, 26(1): 143–144.
- [22] EVTUŠENKO J G. Numerical methods for nonlinear programming[J]. *Doklady Akademii Nauk SSSR*, 1975, 221(5): 1016–1019.
- [23] WEI B, CHEN X J, YE Y Y. Complexity analysis of interior point algorithms for non-Lipschitz and nonconvex minimization[J]. *Mathematical Programming*, 2014, 149(1/2): 301–327.
- [24] NESTEROV Y. Smooth minimization of non-smooth functions[J]. *Mathematical Programming*, 2005, 103(1): 127–152.
- [25] 王传芳. 解非光滑优化问题的光滑技术及理论[D]. 南京: 南京航空航天大学, 2003.
- [26] 马昌凤. 最优化方法及其 Matlab 程序设计[M]. 北京: 科学出版社, 2010.