浙江理工大学学报,第 53 卷,第 3 期,2025 年 5 月
Journal of Zhejiang Sci-Tech University
DOI:10.3969/j.issn.1673-3851(n).2025.03.015

带基数约束凸优化问题的序列凸优化算法

赖诗宇,罗和治
(浙江理工大学理学院,杭州 310018)

摘　要:针对带基数约束凸优化问题,提出了一个基于非线性 DC(Difference of two convex functions)逼近函数的序列凸优化算法,并证明了该算法收敛到 DC 逼近问题的 KKT (Karush-Kuhn-Tucker)点。数值实验结果表明:基于非线性 DC 逼近函数的序列凸优化算法能有效找到带基数约束凸优化问题的稀疏解,且得到解的质量优于已有算法。

关键词:凸优化;基数约束;序列凸优化;DC 逼近;KKT 点

中图分类号:O224　　**文献标志码**:A　　**文章编号**:1673-3851 (2025) 05-0425-07

引文格式:赖诗宇,罗和治. 带基数约束凸优化问题的序列凸优化算法[J]. 浙江理工大学学报(自然科学),2025,53(3):425-431.

Reference Format: LAI Shiyu, LUO Hezhi. Successive convex optimization algorithm for convex optimization problems with cardinality constraints[J]. Journal of Zhejiang Sci-Tech University,2025,53(3):425-431.

Successive convex optimization algorithm for convex optimization problems with cardinality constraints

LAI Shiyu, LUO Hezhi
(School of Science, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: A successive convex optimization algorithm based on a nonlinear DC (difference of two convex functions) approximation function is proposed for convex optimization problems with cardinality constraints. It is proved that the proposed algorithm converges to the KKT (Karush-Kuhn-Tucker) point of the DC approximation problem. Numerical results show that the successive convex optimization algorithm based on the DC approximation function can effectively find sparse solutions to convex optimization problems with cardinality constraints, and the quality of the obtained solutions is better than that of the existing algorithms.

Key words: convex optimization; cardinality constraint; successive convex optimization; DC approximation; KKT point

0 引　言

在现实应用的优化模型中,当决策变量要求稀疏时,人们经常会遇到基数约束的情况。本文考虑如下带基数约束的凸优化问题:

$$\begin{cases}\min & f(\boldsymbol{x});\\ \text{s.t.} & \boldsymbol{x}\in X,\\ & \|\boldsymbol{x}\|_0\leqslant K\end{cases}\tag{1}$$

收稿日期:2022—12—10　　网络出版日期:2023—01—17
基金项目:国家自然科学基金项目(12271485,11871433);浙江省自然科学基金项目(LZ21A010003)
作者简介:赖诗宇(1996—　),女,四川绵阳人,硕士研究生,主要从事最优化理论与算法方面的研究。
通信作者:罗和治,E-mail:hzluo@zstu.edu.cn

其中：$f:\mathbf{R}^n\to\mathbf{R}$ 是可微凸函数；$X=\{\boldsymbol{x}\in\mathbf{R}^n \mid g(\boldsymbol{x})\leqslant 0\}$；$g(\boldsymbol{x})=(g_1(\boldsymbol{x}),\cdots,g_m(\boldsymbol{x}))^{\mathrm{T}}$，$g_i:\mathbf{R}^n\to\mathbf{R}$，$i=1,\cdots,m$ 都是可微凸函数；$\|\boldsymbol{x}\|_0$ 是零范数，表示向量 $\boldsymbol{x}$ 的非零元个数；约束 $\|\boldsymbol{x}\|_0\leqslant K$ 称为基数约束；$K(1\leqslant K<n)$ 为给定的正整数。带基数约束凸优化问题在投资组合[1]、信号处理[2]、压缩感知[3]以及图像处理[4]等领域有着广泛的应用。由于基数约束的存在，问题(1)是 NP－难的[5]，且求解该问题的困难在于基数约束的组合性质。

近年来，带基数约束凸优化问题引起了学者们的广泛关注。根据基数约束的不同处理方式，求解问题(1)的方法大致分为混合整数规划方法、凸逼近方法和非凸逼近方法[6]。Bertsimas 等[7]引入 0－1 变量，将问题(1)等价变换为混合整数规划问题：

$$\begin{cases}\min & f(\boldsymbol{x});\\ \text{s. t.} & \boldsymbol{x}\in X,\\ & \boldsymbol{e}^{\mathrm{T}}\boldsymbol{y}\leqslant K,\\ & \boldsymbol{y}\in\{0,1\}^n,\\ & -M\boldsymbol{y}_i\leqslant \boldsymbol{x}_i\leqslant M\boldsymbol{y}_i,i=1,\cdots,n,\end{cases}$$

其中：$\boldsymbol{e}$ 是各分量均为 1 的向量；M 是一个足够大的数。求解混合整数规划问题主要使用分支定界算法。尽管该方法可以求得精确解，然而求解这个模型的连续松弛问题通常产生很弱的下界，只能求解小规模问题。在凸逼近方法中，Candès 等[8]、Bruckstein 等[9]将约束中的 $\|\boldsymbol{x}\|_0$ 松弛为 $\|\boldsymbol{x}\|_1$，在一定条件下，通过求解凸松弛问题来求解带基数约束凸优化问题。虽然 $\|\boldsymbol{x}\|_1$ 是 $\|\boldsymbol{x}\|_0$ 在单位球中的最佳凸逼近，但两者等价条件过于苛刻，在很多情形下无法满足解的稀疏度要求。

使用非凸逼近函数代替 $\|\boldsymbol{x}\|_0$ 是研究带基数约束凸优化问题的一个有效方法。Mangasarian[10]、Fung 等[11]提出了基于指数近似函数和 l_p 范数($0<p<1$)的逐次线性化算法，该算法在一定条件下能得到正则化问题的一个稳定点。Le Thi 等[12]在 DC (Difference of two convex functions)规划框架下提出了基于非凸逼近函数的 DC 算法。Zheng 等[13]提出了 $\|\boldsymbol{x}\|_0$ 的一个分段线性 DC 逼近函数和基于该函数的加强割序列凸逼近算法，并证明了该算法收敛到 DC 逼近问题的 KKT (Karush-Kuhn-Tucker)点。Gotoh 等[14]使用最大 k 范数证明，基数约束优化问题等价于一类特殊的 DC 优化问题，并提出了求解该问题的 DC 算法。Jiang 等[15]针对具有非负变量的基数约束优化问题，提出了基于该问题等价变换形式的序列凸逼近方法，并证明了该方法收敛于变换问题的 KKT 点。值得一提的是，带基数约束凸优化问题解的质量与 $\|\boldsymbol{x}\|_0$ 的非凸逼近函数的选取有着紧密的联系。

本文针对带基数约束凸优化问题，提出了一个基于非线性 DC 逼近函数的序列凸优化算法(Successive convex optimization algorithm, SCA)。首先，给出了 $\|\boldsymbol{x}\|_0$ 的一个非线性 DC 逼近函数，从而得到原问题的一个非凸逼近问题；其次，通过线性化 DC 函数凹项的方法构造一系列凸逼近子问题，进而提出求解该问题的 SCA 算法，证明了该算法收敛到 DC 逼近问题的 KKT 点；最后通过数值实验验证算法的有效性。

1 零范数的 DC 逼近函数

本节给出 $\|\boldsymbol{x}\|_0$ 的一个非线性 DC 逼近函数及问题(1)的 DC 逼近问题。

定义 1 设函数 $s:\mathbf{R}\to\mathbf{R}$，若对任意的 $t\in\mathbf{R}$，当 $t\neq 0$ 时，$s(t)=1$，否则 $s(t)=0$，则称函数 $s(t)$ 为阶梯函数。

由定义 1 有

$$\|\boldsymbol{x}\|_0=\sum_{i=1}^{n}s(\boldsymbol{x}_i)。$$

众所周知，处理 $\|\boldsymbol{x}\|_0$ 的常用方法之一是用一个连续函数 ϕ_θ 逼近不连续的阶梯函数，其中：$\theta>0$ 是控制逼近紧密度的参数。

首先，对阶梯函数 $s(t)$，提出一个非线性 DC 逼近函数，即：

$$\phi_\theta(t)=\min\left\{\left(\frac{1}{\theta}\mid t\mid+\varepsilon\right)^p,1\right\},0<p<1 \tag{2}$$

其中：ε 是一个很小的正数。特别地，当 $p=1,\varepsilon=0$ 时，$\phi_\theta(t)$ 即为文献[13]中的分段线性 DC 函数。容易验证，函数 $\phi_\theta(t)$ 具有如下性质：

a) $\forall\theta>0, t\in\mathbf{R}, \phi_\theta(t)\leqslant s(t)$；

b)对于固定的 $t\in\mathbf{R}$，当 $\theta>0$ 时，$\phi_\theta(t)$ 是递减函数；

c) $\forall t\in\mathbf{R}, \lim\limits_{\theta\to 0^+}\phi_\theta(t)=s(t)$；

d) $\forall\theta>0, \phi_\theta(t)$ 是偶函数。

当 $t=0$ 时，$\phi'_\theta(0)=\dfrac{p\varepsilon^{(p-1)}}{\theta}$。根据文献[12]中的命题 6，$\phi_\theta(t)$ 可 DC 分解为 $\phi_\theta(t)=c_\theta(t)-h_\theta(t)$，其中：

$$c_\theta(t)=\frac{p\varepsilon^{(p-1)}}{\theta}|t|, h_\theta(t)=\frac{p\varepsilon^{(p-1)}}{\theta}|t|-\min\left\{\left(\frac{1}{\theta}|t|+\varepsilon\right)^p,1\right\},$$

且 $c_\theta(t)$、$h_\theta(t)$ 是关于变量 t 的凸函数。记

$$\Phi_\theta(\boldsymbol{x})=\sum_{i=1}^n\phi_\theta(\boldsymbol{x}_i)。$$

显然，Φ_θ 是一个 DC 函数。用函数 $\Phi_\theta(\boldsymbol{x})$ 代替 $\|\boldsymbol{x}\|_0$，得到问题(1)的如下 DC 逼近问题：

$$\begin{cases}\min & f(\boldsymbol{x});\\ \text{s. t.} & \boldsymbol{x}\in X,\\ & \Phi_\theta(\boldsymbol{x})\leqslant K\end{cases}\tag{3}$$

记问题(3)的可行域为 D_θ，即：$D_\theta=\{\boldsymbol{x}\in\mathbf{R}^n\mid\boldsymbol{x}\in X,\Phi_\theta(\boldsymbol{x})\leqslant K\}$。

2　序列凸优化算法

本节给出求解问题(3)的一个序列凸优化算法及其全局收敛性。首先，记

$$C_\theta(\boldsymbol{x})=\sum_{i=1}^n c_\theta(\boldsymbol{x}_i), H_\theta(\boldsymbol{x})=\sum_{i=1}^n h_\theta(\boldsymbol{x}_i),$$

则 $\Phi_\theta(\boldsymbol{x})=C_\theta(\boldsymbol{x})-H_\theta(\boldsymbol{x})$，且 $C_\theta(\boldsymbol{x})$ 和 $H_\theta(\boldsymbol{x})$ 是凸函数。记 $\partial H_\theta(\bar{\boldsymbol{x}})$ 表示凸函数 $H_\theta(\boldsymbol{x})$ 在 $\bar{\boldsymbol{x}}$ 处的次梯度集。设 $\bar{\boldsymbol{x}}\in D_\theta$，容易验证

$$\partial H_\theta(\bar{\boldsymbol{x}})=\{(\boldsymbol{\xi}_1,\cdots,\boldsymbol{\xi}_n)^{\mathrm{T}}\mid\boldsymbol{\xi}_i\in\partial h_\theta(\bar{\boldsymbol{x}}_i), i=1,\cdots,n\},$$

其中 $\boldsymbol{\xi}_i\in\partial h_\theta(\bar{\boldsymbol{x}}_i)$ 的表达式为：

$$\boldsymbol{\xi}_i=\begin{cases}-\dfrac{p\varepsilon^{p-1}}{\theta}, & \bar{\boldsymbol{x}}_i<-\theta(1-\varepsilon);\\[2mm] \left[-\dfrac{p\varepsilon^{p-1}}{\theta},-\dfrac{p\varepsilon^{p-1}}{\theta}+\dfrac{p}{\theta}\right], & \bar{\boldsymbol{x}}_i=-\theta(1-\varepsilon);\\[2mm] -\dfrac{p\varepsilon^{p-1}}{\theta}+\dfrac{p}{\theta}\left(-\dfrac{1}{\theta}\bar{\boldsymbol{x}}_i+\varepsilon\right)^{p-1}, & -\theta(1-\varepsilon)<\bar{\boldsymbol{x}}_i<0;\\[2mm] \dfrac{p\varepsilon^{p-1}}{\theta}-\dfrac{p}{\theta}\left(\dfrac{1}{\theta}\bar{\boldsymbol{x}}_i+\varepsilon\right)^{p-1}, & 0\leqslant\bar{\boldsymbol{x}}_i<\theta(1-\varepsilon);\\[2mm] \left[\dfrac{p\varepsilon^{p-1}}{\theta}-\dfrac{p}{\theta},\dfrac{p\varepsilon^{p-1}}{\theta}\right], & \bar{\boldsymbol{x}}_i=\theta(1-\varepsilon);\\[2mm] \dfrac{p\varepsilon^{p-1}}{\theta}, & \bar{\boldsymbol{x}}_i>\theta(1-\varepsilon)。\end{cases}$$

设 $\boldsymbol{\xi}\in\partial H_\theta(\bar{\boldsymbol{x}})$，则

$$H_\theta(\boldsymbol{x})\geqslant H_\theta(\bar{\boldsymbol{x}})+\boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}}), \forall\boldsymbol{x}\in\mathbf{R}^n。$$

根据 $\Phi_\theta(\boldsymbol{x})=C_\theta(\boldsymbol{x})-H_\theta(\boldsymbol{x})$，从而有

$$\Phi_\theta(\boldsymbol{x})\leqslant C_\theta(\boldsymbol{x})-H_\theta(\bar{\boldsymbol{x}})-\boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}})\tag{4}$$

利用(4)式,得到问题(3)的凸逼近问题:

$$\begin{cases} \min & f(\boldsymbol{x}); \\ \text{s. t.} & \boldsymbol{x} \in X, \\ & C_\theta(\boldsymbol{x}) - H_\theta(\bar{\boldsymbol{x}}) - \boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{x} - \bar{\boldsymbol{x}}) \leqslant K \end{cases} \tag{5}$$

注意到

$$C_\theta(\boldsymbol{x}) = \sum_{i=1}^{n} \frac{p\varepsilon^{(p-1)}}{\theta} |\boldsymbol{x}_i|$$

是非光滑函数,因此引入等式 $\boldsymbol{z}_i = |\boldsymbol{x}_i|, i=1,\cdots,n$ 处理非光滑部分。从而得到问题(3)的如下凸逼近问题:

$$\begin{cases} \min & f(\boldsymbol{x}); \\ \text{s. t.} & \boldsymbol{x} \in X, \\ & \hat{C}_\theta(\boldsymbol{z}) - H_\theta(\bar{\boldsymbol{x}}) - \boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{x} - \bar{\boldsymbol{x}}) \leqslant K, \\ & -\boldsymbol{z}_i \leqslant \boldsymbol{x}_i \leqslant \boldsymbol{z}_i, i = 1,\cdots,n \end{cases} \tag{6}$$

其中:$\bar{\boldsymbol{x}} \in D_\theta, \boldsymbol{\xi} \in \partial H_\theta(\bar{\boldsymbol{x}})$,且 $\hat{C}_\theta(\boldsymbol{z}) = \sum_{i=1}^{n} \frac{p\varepsilon^{(p-1)}}{\theta} \boldsymbol{z}_i$。

类似于文献[13]中引理 2 和引理 3 的证明方法,可以证得如下引理:

引理 1 问题(5)与问题(6)是等价的。

证明 若 $\boldsymbol{x}^*$ 是问题(5)的最优解,令 $\boldsymbol{z}_i^* = |\boldsymbol{x}_i^*|, i=1,\cdots,n$,则$(\boldsymbol{x}^*, \boldsymbol{z}^*)$是问题(6)的最优解。反之,根据问题(6)的约束条件可知,问题(6)的最优解$(\boldsymbol{x}^*, \boldsymbol{z}^*)$一定满足 $\boldsymbol{z}_i^* = |\boldsymbol{x}_i^*|, i=1,\cdots,n$。因此,若$(\boldsymbol{x}^*, \boldsymbol{z}^*)$是问题(6)的最优解,则 $\boldsymbol{x}^*$ 是问题(5)的最优解。

证毕。

引理 2 令 $\bar{\boldsymbol{x}} \in D_\theta$,若$(\bar{\boldsymbol{x}}, \bar{\boldsymbol{z}})$是问题(6)在 $\bar{\boldsymbol{x}}$ 处的最优解且 $\bar{\boldsymbol{z}}_i = |\bar{\boldsymbol{x}}_i|, i=1,\cdots,n$,则 $\bar{\boldsymbol{x}}$ 是问题(3)的 KKT 点。

证明 根据问题(6)的 KKT 条件可知,存在乘子$(\bar{\boldsymbol{\lambda}}, \bar{\eta}, \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}) \geqslant 0$ 使得:

$$\nabla f(\bar{\boldsymbol{x}}) + \nabla g(\bar{\boldsymbol{x}})\bar{\boldsymbol{\lambda}} - \bar{\eta}\bar{\xi} + \bar{\boldsymbol{\alpha}} - \bar{\boldsymbol{\beta}} = 0 \tag{7}$$

$$\frac{p\varepsilon^{p-1}}{\theta}\bar{\eta}\boldsymbol{e} - \bar{\boldsymbol{\alpha}} - \bar{\boldsymbol{\beta}} = 0 \tag{8}$$

$$\bar{\boldsymbol{\lambda}}^{\mathrm{T}} g(\bar{\boldsymbol{x}}) = 0 \tag{9}$$

$$\bar{\eta}\left[\frac{p\varepsilon^{p-1}}{\theta}\boldsymbol{e}^{\mathrm{T}}\bar{\boldsymbol{z}} - H_\theta(\bar{\boldsymbol{x}}) - K\right] = 0 \tag{10}$$

$$\bar{\boldsymbol{\alpha}}_i(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{z}}_i) = 0, \bar{\boldsymbol{\beta}}_i(\bar{\boldsymbol{x}}_i + \bar{\boldsymbol{z}}_i) = 0, i = 1,\cdots,n \tag{11}$$

由 $\bar{\boldsymbol{z}}_i = |\bar{\boldsymbol{x}}_i|, i=1,\cdots,n$ 以及式(8)和式(11)可知,当 $\bar{\boldsymbol{x}}_i > 0$ 时,$\bar{\boldsymbol{\beta}}_i = 0, \bar{\boldsymbol{\alpha}}_i = \frac{p\varepsilon^{p-1}}{\theta}\bar{\eta}$;当 $\bar{\boldsymbol{x}}_i < 0$ 时,$\bar{\boldsymbol{\alpha}}_i = 0, \bar{\boldsymbol{\beta}}_i = \frac{p\varepsilon^{p-1}}{\theta}\bar{\eta}$。当 $\bar{\boldsymbol{x}}_i = 0$ 时,根据式(11)和乘子的非负性,有 $\bar{\boldsymbol{\alpha}}_i - \bar{\boldsymbol{\beta}}_i \in \left[-\frac{p\varepsilon^{p-1}}{\theta}\bar{\eta}, \frac{p\varepsilon^{p-1}}{\theta}\bar{\eta}\right]$。基于以上分析可以得到,$-\bar{\eta}\bar{\boldsymbol{\xi}} + \bar{\boldsymbol{\alpha}} - \bar{\boldsymbol{\beta}} \in \bar{\eta}\partial\Phi_\theta(\bar{\boldsymbol{x}})$,其中 $\partial\Phi_\theta(\bar{\boldsymbol{x}})$表示函数 $\Phi_\theta(\boldsymbol{x})$在点 $\bar{\boldsymbol{x}}$ 处的 Clarke 广义梯度集(见文献[16])。由式(7)有:

$$0 \in \nabla f(\bar{\boldsymbol{x}}) + \nabla g(\bar{\boldsymbol{x}})\bar{\boldsymbol{\lambda}} + \bar{\eta}\partial\Phi_\theta(\bar{\boldsymbol{x}}) \tag{12}$$

根据式(9)、式(10)和式(12)可以推出 $\bar{\boldsymbol{x}}$ 是问题(3)的 KKT 点。

证毕。

基于凸逼近问题(6),给出求解问题(3)的一个序列凸优化算法。

算法 1 序列凸优化算法。

步骤 1:取参数 $\theta > 0$,算法终止参数 $\varepsilon_0 > 0$。选取满足 $\boldsymbol{x}^0 \in D_\theta$ 的初始点 $\boldsymbol{x}^0$,令 $k := 0$。

步骤 2:计算 $H_\theta(\boldsymbol{x})$在 $\boldsymbol{x}^k$ 处的次梯度 $\boldsymbol{\xi}^k$。

步骤 3:求解当 $\bar{\boldsymbol{x}} = \boldsymbol{x}^k$ 和 $\boldsymbol{\xi} = \boldsymbol{\xi}^k$ 时的凸子问题(6),得到最优解 $\boldsymbol{x}^{k+1}$。

步骤 4:若 $\| \boldsymbol{x}^{k+1}-\boldsymbol{x}^k \| \leqslant \varepsilon_0$,终止算法。

步骤 5:令 $k := k+1$,转步骤 2。

类似于文献[13]中定理 2 的证明方法,可以证得算法 1 的如下收敛性结果。

定理 1　设$\{\boldsymbol{x}^k\}$为由算法 1 产生的一个无穷序列,则$\{\boldsymbol{x}^k\}$的任意一个聚点是问题(3)的 KKT 点。

证明　为便于分析,假设在算法 1 的步骤 2 中,当 $\boldsymbol{x}_i^k=\theta(1-\varepsilon)$时,$\boldsymbol{\xi}_i^k=\dfrac{p\varepsilon^{p-1}}{\theta}-\dfrac{p}{2\theta}$。

令问题(3)的最优值为 V。由 $\boldsymbol{x}^{k+1}$、$\boldsymbol{x}^k$ 分别是问题(5)在 $\boldsymbol{x}^k$ 处的最优解和可行解,可以得到 $V\leqslant f(\boldsymbol{x}^{k+1})\leqslant f(\boldsymbol{x}^k)$,即$\{f(\boldsymbol{x}^k)\}$递减,且 $\lim\limits_{k\to\infty} f(\boldsymbol{x}^k)=\inf_k f(\boldsymbol{x}^k)\geqslant V$。

设 $\boldsymbol{x}^*$ 是序列$\{\boldsymbol{x}^k\}$的一个聚点,则存在子列$\{\boldsymbol{x}^{k_j}\}\subset\{\boldsymbol{x}^k\}$使得 $\boldsymbol{x}^{k_j}\to\boldsymbol{x}^*$。由 $\boldsymbol{x}^k\in D_\theta$,有 $\boldsymbol{x}^*\in D_\theta$ 且 $f(\boldsymbol{x}^*)=\lim\limits_{k\to\infty} f(\boldsymbol{x}^k)=\inf_k f(\boldsymbol{x}^k)$。

令 $\boldsymbol{z}_i^*=|\boldsymbol{x}_i^*|$,$i=1,\cdots,n$。记 $q(\boldsymbol{x}_i,\bar{\boldsymbol{x}}_i,\boldsymbol{\xi}_i)=h_\theta(\bar{\boldsymbol{x}}_i)+\boldsymbol{\xi}_i(\boldsymbol{x}_i-\bar{\boldsymbol{x}}_i)$,$\boldsymbol{\xi}_i\in\partial h_\theta(\bar{\boldsymbol{x}}_i)$,则问题(6)中的线性约束可表示为:

$$C_\theta(\boldsymbol{z})-\sum_{i=1}^n q(\boldsymbol{x}_i,\bar{\boldsymbol{x}}_i,\boldsymbol{\xi}_i)\leqslant K \tag{13}$$

对 $\boldsymbol{x}_i^*\geqslant 0$,$i=1,\cdots,n$,考虑如下几种情形:

a)当 $\boldsymbol{x}_i^*,\boldsymbol{x}_i^{k_j}\in[0,\theta(1-\varepsilon))$时,$\boldsymbol{\xi}_i^{k_j}=\dfrac{p\varepsilon^{p-1}}{\theta}-\dfrac{p}{\theta}\left(\dfrac{1}{\theta}\boldsymbol{x}_i^{k_j}+\varepsilon\right)^{p-1}$,$\boldsymbol{\xi}_i^*=\dfrac{p\varepsilon^{p-1}}{\theta}-\dfrac{p}{\theta}\left(\dfrac{1}{\theta}\boldsymbol{x}_i^*+\varepsilon\right)^{p-1}$,

$$q(\boldsymbol{x}_i,\boldsymbol{x}_i^{k_j},\boldsymbol{\xi}_i^{k_j})=\frac{p\varepsilon^{p-1}}{\theta}\boldsymbol{x}_i^{k_j}-\left(\frac{1}{\theta}\boldsymbol{x}_i^{k_j}+\varepsilon\right)^p+\left(\frac{p\varepsilon^{p-1}}{\theta}-\frac{p}{\theta}\left(\frac{1}{\theta}\boldsymbol{x}_i^{k_j}+\varepsilon\right)^{p-1}\right)(\boldsymbol{x}_i-\boldsymbol{x}_i^{k_j}) \tag{14}$$

因为次梯度集 $\partial H_\theta(\boldsymbol{x})$是上半连续集映射(见文献[17]性质 4.2.3),故当 $\boldsymbol{x}^{k_j}\to\boldsymbol{x}^*$ 时,$\boldsymbol{\xi}^{k_j}\to\boldsymbol{\xi}^*\in\partial H_\theta(\boldsymbol{x}^*)$。对式(14)取极限可得:

$$\frac{p\varepsilon^{p-1}}{\theta}\boldsymbol{x}_i^*-\left(\frac{1}{\theta}\boldsymbol{x}_i^*+\varepsilon\right)^p+\left(\frac{p\varepsilon^{p-1}}{\theta}-\frac{p}{\theta}\left(\frac{1}{\theta}\boldsymbol{x}_i^*+\varepsilon\right)^{p-1}\right)(\boldsymbol{x}_i-\boldsymbol{x}_i^*)=q(\boldsymbol{x}_i,\boldsymbol{x}_i^*,\boldsymbol{\xi}_i^*)。$$

b)当 $\boldsymbol{x}_i^*=\theta(1-\varepsilon)$时,

$$q(\boldsymbol{x}_i,\boldsymbol{x}_i^*,\boldsymbol{\xi}_i^*)=p\varepsilon^{p-1}(1-\varepsilon)-1+\boldsymbol{\xi}_i^*(\boldsymbol{x}_i-\theta(1-\varepsilon)),$$

其中 $\boldsymbol{\xi}_i^*\in\partial h_\theta(\boldsymbol{x}_i^*)=\left[\dfrac{p\varepsilon^{p-1}}{\theta}-\dfrac{p}{\theta},\dfrac{p\varepsilon^{p-1}}{\theta}\right]$。注意到,

$$q(\boldsymbol{x}_i,\boldsymbol{x}_i^{k_j},\boldsymbol{\xi}_i^{k_j})=\begin{cases}\dfrac{p\varepsilon^{p-1}}{\theta}\boldsymbol{x}_i^{k_j}-\left(\dfrac{1}{\theta}\boldsymbol{x}_i^{k_j}+\varepsilon\right)^p+\left(\dfrac{p\varepsilon^{p-1}}{\theta}-\dfrac{p}{\theta}\left(\dfrac{1}{\theta}\boldsymbol{x}_i^{k_j}+\varepsilon\right)^{p-1}\right)(\boldsymbol{x}_i-\boldsymbol{x}_i^{k_j}), & 0\leqslant\boldsymbol{x}_i^{k_j}<\theta(1-\varepsilon);\\ p\varepsilon^{p-1}(1-\varepsilon)-1+\left(\dfrac{p\varepsilon^{p-1}}{\theta}-\dfrac{p}{2\theta}\right)(\boldsymbol{x}_i-\theta(1-\varepsilon)), & \boldsymbol{x}_i^{k_j}=\theta(1-\varepsilon);\\ \dfrac{p\varepsilon^{p-1}}{\theta}\boldsymbol{x}_i^{k_j}-1+\dfrac{p\varepsilon^{p-1}}{\theta}(\boldsymbol{x}_i-\boldsymbol{x}_i^{k_j}), & \boldsymbol{x}_i^{k_j}>\theta(1-\varepsilon)。\end{cases}$$

因此当 $\boldsymbol{x}_i^{k_j}\to\boldsymbol{x}_i^*$ 时,存在 $\boldsymbol{\xi}_i^*\in\partial h_\theta(\boldsymbol{x}_i^*)$,使得 $q(\boldsymbol{x}_i,\boldsymbol{x}_i^{k_j},\boldsymbol{\xi}_i^{k_j})\to q(\boldsymbol{x}_i,\boldsymbol{x}_i^*,\boldsymbol{\xi}_i^*)$。

c)当 $\boldsymbol{x}_i^*,\boldsymbol{x}_i^{k_j}\in(\theta(1-\varepsilon),+\infty)$时,$\boldsymbol{\xi}_i^{k_j}=\dfrac{p\varepsilon^{p-1}}{\theta}=\boldsymbol{\xi}_i^*\in\partial h_\theta(\boldsymbol{x}_i^*)$,

$$q(\boldsymbol{x}_i,\boldsymbol{x}_i^{k_j},\boldsymbol{\xi}_i^{k_j})=\frac{p\varepsilon^{p-1}}{\theta}\boldsymbol{x}_i-1=q(\boldsymbol{x}_i,\boldsymbol{x}_i^*,\boldsymbol{\xi}_i^*)。$$

同理可证,当 $\boldsymbol{x}_i^*<0$,$i=1,\cdots,n$,$j\to\infty$时,$q(\boldsymbol{x}_i,\boldsymbol{x}_i^{k_j},\boldsymbol{\xi}_i^{k_j})=q(\boldsymbol{x}_i,\boldsymbol{x}_i^*,\boldsymbol{\xi}_i^*)$成立。

综上所述,由式(13)可以推断出,当 j 充分大时,对一些 $\boldsymbol{\xi}^*\in\partial H_\theta(\boldsymbol{x}^*)$,问题(6)在 $\boldsymbol{x}^{k_j}$ 和 $\boldsymbol{x}^*$ 处有相同的可行域。一方面,$(\boldsymbol{x}^{k_j+1},\boldsymbol{z}^{k_j+1})$是问题(6)在 $\boldsymbol{x}^{k_j}$ 处的最优解,故其也是问题(6)在 $\boldsymbol{x}^*$ 处的最优解;另一方面,由$(\boldsymbol{x}^*,\boldsymbol{z}^*)$是问题(6)在 $\boldsymbol{x}^*$ 处的可行解,可以得到 $f(\boldsymbol{x}^{k_j+1})\leqslant f(\boldsymbol{x}^*)$。再根据 $f(\boldsymbol{x}^*)=\inf_k f(\boldsymbol{x}^k)$,有 $f(\boldsymbol{x}^{k_j+1})=f(\boldsymbol{x}^*)$。因此,$(\boldsymbol{x}^*,\boldsymbol{z}^*)$是问题(6)在 $\boldsymbol{x}^*$ 处的最优解,结合引理 2,可知 $\boldsymbol{x}^*$ 是问题(3)的

KKT 点。

证毕。

3 数值实验

本节通过数值实验验证算法 1 求解基数约束凸优化问题(1)的有效性。数值实验在 MATLAB R2018b 上实现,在 PC 机(1.80 GiHz ,16 GiB,ADM)上运行。算法 1 中所有凸二次子问题均由 CPLEX 中的二次规划求解器求解,该求解器采用的是凸二次规划障碍内点算法。

测试问题为有限多元化的均值—方差投资组合选择问题[18]:

$$\begin{cases}\min & \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x};\\ \text{s. t.} & \sum_{i=1}^{n}\boldsymbol{\mu}_i\boldsymbol{x}_i \geqslant \rho,\ \sum_{i=1}^{n}\boldsymbol{x}_i = 1,\\ & 0 \leqslant \boldsymbol{x}_i \leqslant \boldsymbol{u}_i,\ i=1,\cdots,n,\\ & \|\boldsymbol{x}\|_0 \leqslant K。\end{cases}$$

在数值实验中,采用 $\|\boldsymbol{x}\|_0$ 的两个不同逼近函数进行比较,即:

a)非线性 DC 函数:$\phi_\theta(t)=\min\left\{\left(\frac{1}{\theta}|t|+\varepsilon\right)^p,1\right\}$,$0<p<1$。

b)分段线性 DC 函数[13]:$\phi_\theta(t)=\min\left\{\frac{1}{\theta}|t|,1\right\}$。

测试问题中的参数 $\boldsymbol{Q}$、$\boldsymbol{\mu}_i$、ρ、$\boldsymbol{u}_i$ 由与文献[19—20]类似的方式随机生成,即:矩阵 $\boldsymbol{Q}$ 以不同的对称矩阵产生,系数 $\boldsymbol{\mu}_i$、ρ 是服从均匀分布 U[0.002,0.1]的随机数,$\boldsymbol{u}_i$ 服从 U[0.375,0.425]。在数值测试中,算法 1 中的初始解 $\boldsymbol{x}^0$ 与恢复问题(1)的可行解 $\hat{x}$,分别采用文献[21]中的启发式 1 和 2 求解;基于非线性 DC 函数和分段线性 DC 函数的逼近问题的初始点都是通过求解同一个问题得到的。表 1 给出了使用非线性 DC 函数和分段线性 DC 函数在序列凸优化算法下,计算 10 个例子的平均结果,其中:n 表示问题的维数,$p=0.5$,时间表示求解 10 个测试问题所需的平均 CPU 时间,r 是 10 个例子的平均相对改进率,计算公式为:

$$r=\frac{f(\boldsymbol{x}^0)-f(\hat{\boldsymbol{x}})}{\max(1,|f(\boldsymbol{x}^0)|)}。$$

算法终止参数 $\varepsilon_0=10^{-7}$。分段线性 DC 函数参数设置参见文献[13]。

从表 1 的数值结果可以看出,分段线性 DC 函数所花费的时间均少于非线性 DC 函数。当 $K\geqslant 10$ 时,非线性 DC 函数的相对改进率高于分段线性 DC 函数,能够得到更好的解。因此,基于非线性 DC 函数的序列凸优化算法能有效地找到带基数约束凸优化问题的稀疏解,在解的质量方面具有一定的优势。

表 1 基于不同逼近函数解质量的相对改进率

n	K	θ	非线性 DC 函数		分段线性 DC 函数	
			r/%	时间/s	r/%	时间/s
100	5	1.00	1.98	2.27	3.60	0.24
100	10	0.90	3.77	3.19	1.01	1.28
100	15	0.80	2.26	4.79	1.52	2.81
100	20	0.70	1.51	30.45	1.20	15.82
200	5	1.00	3.33	2.13	4.13	0.32
200	10	0.90	5.38	5.33	3.88	1.43
200	15	0.80	3.68	5.38	1.60	2.43
200	20	0.70	2.59	18.37	0.80	7.97
300	5	1.00	1.77	3.83	7.16	0.52
300	10	0.90	3.55	7.41	2.04	1.51
300	15	0.80	3.79	9.37	3.14	2.43
300	20	0.70	3.04	24.38	2.07	7.74

4 结　论

本文研究了带基数约束凸优化问题,提出了零范数的一个非线性 DC 逼近函数,给出了基于该非线性 DC 逼近函数的序列凸优化算法,并证明了算法的全局收敛性。数值实验表明,提出的序列凸优化算法能有效地找到问题的稀疏解,在解的质量方面具有一定的优势。本文所提出的方法是求解带基数约束凸优化问题的局部方法,后续将研究该问题的全局算法。

参考文献:

[1] Gao J J, Li D. Optimal cardinality constrained portfolio selection[J]. Operations Research, 2013, 61(3): 745-761.

[2] Chartrand R. Exact reconstruction of sparse signals via nonconvex minimization[J]. IEEE Signal Processing Letters, 2007, 14(10): 707-710.

[3] Blumensath T. Compressed sensing with nonlinear observations and related nonlinear optimization problems [J]. IEEE Transactions on Information Theory, 2013, 59(6): 3466-3474.

[4] Elad M, Figueiredo M A T, Ma Y. On the role of sparse and redundant representations in image processing [J]. Proceedings of the IEEE, 2010, 98(6): 972-982.

[5] Bienstock D. Computational study of a family of mixed-integer quadratic programming problems[J]. Mathematical Programming, 1996, 74(2): 121-140.

[6] 赵晨,罗自炎,修乃华. 稀疏优化理论与算法若干新进展[J]. 运筹学学报, 2020, 24(4): 1-24.

[7] Bertsimas D, Shioda R. Algorithm for cardinality-constrained quadratic optimization[J]. Computational Optimization and Applications, 2009, 43(1): 1-22.

[8] Candès E J, Tao T. Decoding by linear programming[J]. IEEE Transactions on Information Theory, 2005, 51(12): 4203-4215.

[9] Bruckstein A M, Donoho D L, Elad M. From sparse solutions of systems of equations to sparse modeling of signals and images[J]. SIAM Review, 2009, 51(1): 34-81.

[10] Mangasarian O L. Minimum-support solutions of polyhedral concave programs[J]. Optimization, 1999, 45(1/2/3/4): 149-162.

[11] Fung G M, Mangasarian O L. Equivalence of minimal ℓ_0- and ℓ_p-norm solutions of linear equalities, inequalities and linear programs for sufficiently small p[J]. Journal of Optimization Theory and Applications, 2011, 151(1): 1-10.

[12] Le Thi H A, Pham Dinh T, Le H M, et al. DC approximation approaches for sparse optimization[J]. European Journal of Operational Research, 2015, 244(1): 26-46.

[13] Zheng X J, Sun X L, Li D, et al. Successive convex approximations to cardinality-constrained convex programs: a piecewise-linear DC approach[J]. Computational Optimization and Applications, 2014, 59(1): 379-397.

[14] Gotoh J, Takeda A, Tono K. DC formulations and algorithms for sparse optimization problems[J]. Mathematical Programming, 2018, 169(1): 141-176.

[15] Jiang Z Y, Wu B Y, Hu Q Y, et al. Cardinality-constrained programs with nonnegative variables and an SCA method[J]. Journal of the Operational Research Society, 2022, 73(3): 634-652.

[16] Clarke F H. Optimization and Nonsmooth Analysis[M]. New York: Wiley, 1983: 27.

[17] Bertsekas D P, Nedić A, Ozdaglar A E. Convex analysis and optimization[M]. Belmont, Mass: Athena Scientific, 2003: 231-232.

[18] Bonami P, Lejeune M A. An exact solution approach for portfolio optimization problems under stochastic and integer constraints[J]. Operations Research, 2009, 57(3): 650-670.

[19] Frangioni A, Gentile C. Perspective cuts for a class of convex 0-1 mixed integer programs[J]. Mathematical Programming, 2006, 106(2): 225-236.

[20] Zheng X J, Pan Y T, Hu Z L. Perspective reformulations of semicontinuous quadratically constrained quadratic programs[J]. INFORMS Journal on Computing, 2021, 33(1): 163-179.

[21] Sun X L, Zheng X J, Li D. Recent advances in mathematical programming with semi-continuous variables and cardinality constraint[J]. Journal of the Operations Research Society of China, 2013, 1(1): 55-77.

(责任编辑:康　锋)