doi:10.3969/j.issn.1672-4348.2016.01.019

一类复变量优化问题的次梯度投影算法

张宋传

(闽江学院 数学系,福建 福州 350121)

摘要:利用 CR 微分理论,提出求解一类线性等式约束的复变量非光滑凸优化问题的复值次梯度投影算法(CSPM),该算法能完全基于复域上运行。在较弱的条件下证明了算法的全局收敛性,数值实验进一步表明了 CSPM 的可行性和有效性,该算法尤其适合大规模优化问题的求解。

关键词:次梯度;CR 微分;复变量优化问题;非光滑

中图分类号:O221.2　　**文献标志码**:A　　**文章编号**:1672-4348(2016)01-0086-04

A complex-valued subgradient projection method (CSPM) for a class of complex variables non-smooth convex optimization problems

Zhang Songchuan

(Mathematics Department, Minjiang University, Fuzhou 350121, China)

Abstract: A complex-valued subgradient projection method (CSPM) based on CR calculus theory is presented to solve a class of complex variables non-smooth convex optimization problems with linear equality constraints, which can be completely implemented in the complex domain. The proposed method is proved to be globally convergent under mild conditions. Numerical experiments show that CSPM is feasible and effective and suitable for solving large-scale optimization problems.

Keywords: subgradient; CR calculus theory; complex variables optimization problem; non-smoothness

考虑线性等式约束的非光滑复变量凸优化问题:

$$\begin{aligned}&\min_{\boldsymbol{z}\in\mathbf{C}^n} \quad h(z) + \|\boldsymbol{Q}\boldsymbol{z}-\boldsymbol{d}\|_1\\ &s.t. \quad \boldsymbol{A}\boldsymbol{z}=\boldsymbol{b}\end{aligned} \tag{1}$$

其中.优化变量 $z\in\mathbf{C}^n$,$h(z):\mathbf{C}^n\mapsto\mathbf{R}$ 是 R 可微的凸函数,$\boldsymbol{Q}\in\mathbf{C}^{l\times n}$,$\boldsymbol{A}\in\mathbf{C}^{m\times n}$,$\boldsymbol{b}\in\mathbf{C}^m$,$\boldsymbol{d}\in\mathbf{C}^l$,另假设问题(1)有全局最优解。

1 预备知识

问题(1)的目标函数为复变量实值函数,不满足柯西-黎曼条件[1],这意味着经典的复变函数理论无法应用于问题(1)优化算法的设计上。通过拆分复值数据,将问题转化为等价的实变量优化问题求解是较为普遍的做法。不足在于转化过程繁琐,数据结构被破坏,当复变量优化问题的规模较大时,转化得到的实变量优化问题成倍地扩大了原问题的规模,大大增加了问题求解的空间复杂性。CR 微分[2]是一种适用范围更广的复变函数微分理论,近年来被广泛应用于复变量优化领域[3-4],并逐渐成为设计和分析复值优化算法,尤其是使用导数的复值优化算法的重要理论基础。

设 $g(z):\mathbf{C}^n\mapsto\mathbf{R}$,CR 微分中复变量 z 与它的共轭 $\bar{z}$ 在形式上认为是两个独立的变量,$g(z)$ 在

收稿日期:2016-01-03
基金项目:福建省中青年教师教育科研项目(JA15436)
作者简介:张宋传(1977-),男,福建光泽人,博士研究生,硕士,讲师,研究方向:复值优化算法。

形式上也可写作 $g(z,\bar{z})$ 或 $g(\boldsymbol{v})$ ，其中 $\boldsymbol{v} = \begin{pmatrix} z \\ \bar{z} \end{pmatrix}$

定义 1[2]　$g(z)$ 关于 z 与 $\bar{z}$ 的偏导数定义为：

$$\mathrm{R}\text{ 导数} \triangleq \left.\frac{\partial g}{\partial z}\right|_{\bar{z}=\text{常值}},\ \bar{\mathrm{R}}\text{ 导数} \triangleq \left.\frac{\partial g}{\partial \bar{z}}\right|_{z=\text{常值}}。$$

出于简化的考虑，R 导数与 $\bar{\mathrm{R}}$ 导数分别记为 $\frac{\partial g}{\partial z}$ 和 $\frac{\partial g}{\partial \bar{z}}$ 。如果 $g(z,\bar{z})$ 关于 z 与 $\bar{z}$ 的偏导数存在，则 $g(z,\bar{z})$ 称之为 R 可微的。令 $\boldsymbol{z} = \boldsymbol{x} + \mathrm{j}\boldsymbol{y}$ ，j 为虚数单位，$\boldsymbol{x},\boldsymbol{y} \in \mathbf{R}^n$ 分别记为 z 的实部和虚部，则

$$g(z) = u(\boldsymbol{x},\boldsymbol{y}) + \mathrm{j}v(\boldsymbol{x},\boldsymbol{y}): \mathbf{R}^{2n} \mapsto \mathbf{C}$$

其中，$u(\boldsymbol{x},\boldsymbol{y}),v(\boldsymbol{x},\boldsymbol{y}): \mathbf{R}^{2n} \mapsto \mathbf{R}$ 。可以证明 R 导数与 $\bar{\mathrm{R}}$ 导数等价于[5]

$$\frac{\partial g}{\partial z} = \frac{1}{2}\left(\frac{\partial g}{\partial \boldsymbol{x}} - \mathrm{j}\frac{\partial g}{\partial \boldsymbol{y}}\right),\quad \frac{\partial g}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial g}{\partial \boldsymbol{x}} + \mathrm{j}\frac{\partial g}{\partial \boldsymbol{y}}\right)。$$

定义 2[2]　如果 $g(z)$ 是 R 可微的，则该函数的复梯度定义为：$\nabla g = \frac{\partial g}{\partial \bar{z}}$ 。

CR 微分其它概念可参考文献[6,7]。

定义 3　若 $g(z)$ 是 $\mathbf{C}^n$ 上实值连续凸函数，如果向量 $\boldsymbol{\eta}$ 满足

$$g(\boldsymbol{\xi}) \geqslant g(z) + \mathrm{Re}((\boldsymbol{\xi} - z)^{\mathrm{H}}\boldsymbol{\eta})\quad \forall \boldsymbol{\xi} \in \mathbf{C}^n,$$

则 $\boldsymbol{\eta}$ 称为 g 在 z 上的次梯度，记为 $\tilde{\nabla} g(z)$ 。g 在 z 上的次梯度全体称为 g 在 z 上的次微分，记为 $\partial g(z)$ 。

命题 1　令 $g(z) = \|\boldsymbol{Q}z - d\|_1$ ，其中 $\boldsymbol{Q} \in \mathbf{C}^{l\times n}$ ，$\boldsymbol{d} \in \mathbf{C}^l$ ，令 $\theta_i(z) = q_i^{\mathrm{T}}z - \boldsymbol{d}_i$ 及

$$\varphi_i(z) = \begin{cases} p_i, & |\theta_i(z)| = 0 \\ \theta_i(z)/|\theta_i(z)|, & |\theta_i(z)| \neq 0 \end{cases}$$

其中，q_i 是 $\boldsymbol{Q}$ 的第 i 行，$\boldsymbol{d}_i$ 是 $\boldsymbol{d}$ 的第 i 个分量，$p_i \in \mathbf{C}$ 且 $|p_i| \leqslant 1$ ，有

$$\tilde{\nabla} g(z) = \boldsymbol{Q}^{\mathrm{H}}\phi(z),$$

其中，$\phi(z) = (\phi_1(z),\cdots,\phi_l(z))^{\mathrm{T}}$ 。

2　算法的导出

次梯度算法是用于优化非可微凸函数的一种简单的方法，其主要思想是通过次梯度反方向（负次梯度方向）搜索以找到目标函数的极小值。由于算法简单，次梯度法能快速地实施，较之牛顿法和内点法，次梯度法只需要更小的运行空间，这使得次梯度法很适合处理数据规模大的问题。相关内容参考文献[8－10]。本文首次将次梯度投影算法从解决实变量约束优化问题推广到解复变量约束优化问题，提出一类复值次梯度投影算法（complex-valued subgradient projection method，CSPM）求解问题（1）。

令 $g(z) = h(z) + \|\boldsymbol{Q}z - \boldsymbol{d}\|_1$ ，其次梯度为

$$\tilde{\nabla} g(z) = \nabla h(z) + \boldsymbol{Q}^{\mathrm{H}}\phi(z)$$

其中，$\nabla h(z)$ 为 $h(z)$ 复梯度，$\phi(z)$ 的定义见命题 1，$\tilde{\nabla} g^k$ 表示目标函数 g 在点 z^k 的次梯度。CSPM 求解问题（1）的框架如下：

算法：CSPM

输　入：初始可行点 z^1，如 $z^1 = \boldsymbol{A}^+\boldsymbol{b}$.

初始化：最大迭代数 $K_{\max}$，最大历史长度 $N > 0$，容忍度 $\varepsilon > 0$，令 $g_{\mathrm{best}}^1 = g(z^1)$，$z_{\mathrm{best}} = z^1$ 及 $k = 1$.

1：　for $i = 1$ to $K_{\max}$ do
2：　$z^{k+1} = P_\Omega(z^k - \alpha_k \tilde{\nabla} g^k)$；　$k = k + 1$
3：　if　$g_{\mathrm{best}}^{k-1} \geqslant g(z^k)$ then
4：　$g_{\mathrm{best}}^k = g(z^k)$，$z_{\mathrm{best}} = z^k$
5：　else
6：　$g_{\mathrm{best}}^k = g_{\mathrm{best}}^{k-1}$
7：　end if
8：　if　$k > N$ and $|g_{\mathrm{best}}^{k-N} - g_{\mathrm{best}}^k| / |g_{\mathrm{best}}^k| \leqslant \varepsilon$ then
9：　break
10：　end if
11：end for

输　出：z_{best}.

因为次梯度方向不一定是目标函数的下降方向，因此需要追踪每次迭代后目标函数的最优值，该算法中，g_{best}^k 记录第 k 次迭代后所达到的最优目标函数值，z_{best}^k 表示相应的最优变量。$\Omega = \{z \in \mathbf{C}^n \mid \boldsymbol{A}z = \boldsymbol{b}\}$ 表示可行域，Ω 上正交投影算子记为 $P_\Omega(\cdot)$，$\forall \boldsymbol{a} \in \mathbf{C}^n$ ，

$$P_\Omega(\boldsymbol{a}) = (\boldsymbol{I}_n - \boldsymbol{A}^+\boldsymbol{A})\boldsymbol{a} + \boldsymbol{A}^+\boldsymbol{b}$$

其中，$\boldsymbol{A}^+$ 表示复矩阵 $\boldsymbol{A}$ 的 Moore-Penrose 逆。α_k 记为第 k 次迭代的步长，次梯度法的搜索步长是预先设置的，α_k 的选取策略对于迭代序列的收敛行为起着关键作用。实验表明，有限平方和准则是较好的选择[11]，即

$$\alpha_k \geqslant 0,\quad \sum_{k=1}^{\infty}\alpha_k^2 < \infty,\quad \sum_{k=1}^{\infty}\alpha_k = \infty \tag{2}$$

例如，$\alpha_k = \frac{a}{b+k}$，其中 $a > 0$，$b \geqslant 0$，CSPM 中 α_k 选取有限平方和准则。

3 收敛性分析

令 g^* 表示问题(1)最优目标函数值，即 $g^* = \inf_{z \in \Omega} g(z)$，记最优解的集合为 Z^*，另假设：

1)算法中的次梯度有界，即存在 $G > 0$，任一 k，有 $\| \tilde{\nabla} g^k \|_2 \leqslant G$；

2)初始点与 Z^* 的距离有界，即任一 $z^* \in Z^*$，有 $\| z^1 - z^* \|_2 \leqslant R$。

定理1 设 $\{g_{\text{best}}^k\}$ 为 CSPM 求解问题(1)生成的序列，上述假设成立，则

$$\lim_{k \to \infty} g_{\text{best}}^k = g^*$$

证明： 令 $\tilde{z}^{k+1} = z^k - \alpha_k \tilde{\nabla} g^k$，设 z^* 是任一最优解，依次梯度定义，

$$g(z^*) \geqslant g(z^k) + \mathrm{Re}((z^* - z^k)^{\mathrm{H}} \tilde{\nabla} g^k),$$

因此

$$\| \tilde{z}^{k+1} - z^* \|_2^2 = \| z^k - \alpha_k \tilde{\nabla} g^k - z^* \|_2^2 = \| z^k - z^* \|_2^2 - 2\alpha_k \mathrm{Re}((z^k - z^*)^{\mathrm{H}} \tilde{\nabla} g^k) + \alpha_k^2 \| \tilde{\nabla} g^k \|_2^2 \leqslant$$

$$\| z^k - z^* \|_2^2 - 2\alpha_k (g(z^k) - g^*) + \alpha_k^2 \| \tilde{\nabla} g^k \|_2^2,$$

考虑到 $P_\Omega(\cdot)$ 是非扩张的，即

$$\| z^{k+1} - z^* \|_2 = \| P_\Omega(\tilde{z}^{k+1}) - z^* \|_2 \leqslant \| \tilde{z}^{k+1} - z^* \|_2,$$

进而

$$\| z^{k+1} - z^* \|_2^2 \leqslant \| z^k - z^* \|_2^2 - 2\alpha_k (g(z^k) - g^*) + \alpha_k^2 \| \tilde{\nabla} g^k \|_2^2$$

上式递归得

$$\| z^{k+1} - z^* \|_2^2 \leqslant \| z^1 - z^* \|_2^2 - 2\sum_{i=1}^{k} \alpha_i (g(z^i) - g^*) + \sum_{i=1}^{k} \alpha_i^2 \| \tilde{\nabla} g^i \|_2^2。$$

注意到 $\| z^{k+1} - z^* \|_2^2 \geqslant 0$ 及 $\| z^1 - z^* \| \leqslant R$，则

$$2\sum_{i=1}^{k} \alpha_i (g(z^i) - g^*) \leqslant R^2 + \sum_{i=1}^{k} \alpha_i^2 \| \tilde{\nabla} g^i \|_2^2,$$

又因为

$$\sum_{i=1}^{k} \alpha_i (g(z^i) - g^*) \geqslant \left(\sum_{i=1}^{k} \alpha_i\right) \min_{i=1,\cdots,k} (g(z^i) - g^*) = \left(\sum_{i=1}^{k} \alpha_i\right)(g_{\text{best}}^k - g^*),$$

所以

$$g_{\text{best}}^k - g^* \leqslant \left(R^2 + \sum_{i=1}^{k} \alpha_i^2 \| \tilde{\nabla} g^i \|_2^2\right) \Big/ \left(2\sum_{i=1}^{k} \alpha_i\right),$$

考虑到假设 $\| \tilde{\nabla} g^k \|_2 \leqslant G \quad (i = 1, \cdots, k)$，故

$$g_{\text{best}}^k - g^* \leqslant \left(R^2 + G^2 \sum_{i=1}^{k} \alpha_i^2\right) \Big/ \left(2\sum_{i=1}^{k} \alpha_i\right)$$

算法中，步长 α_k 选取有限平方和准则(2)，综合上式，即得 $\lim_{k \to \infty} g_{\text{best}}^k - g^* = 0$，命题得证。

4 数值实验

考虑如下问题：

$$\begin{aligned} \min_{z \in \mathbf{C}^n} \quad & \| \boldsymbol{F} z - \boldsymbol{d} \|_2^2 + \| \boldsymbol{Q} z \|_1 \\ \text{s.t.} \quad & \boldsymbol{A} z = \boldsymbol{b} \end{aligned} \tag{3}$$

实验中矩阵 $\boldsymbol{F} \in \mathbf{C}^{l \times n}$，$\boldsymbol{Q} \in \mathbf{C}^{n \times n}$，$\boldsymbol{A} \in \mathbf{C}^{m \times n}$ 及向量 $\boldsymbol{b} \in \mathbf{C}^m$，$\boldsymbol{d} \in \mathbf{C}^l$ 均随机生成，且各元素独立同分布于均值为0，标准差为1的正态分布。CSPM 的初始点选取 $\boldsymbol{A}^+ \boldsymbol{b}$，$\boldsymbol{A}$ 的伪逆算法采用 qrgin$\boldsymbol{v}$ 函数[12]。步长 α_k 选取有限平方和准则(2)，其中 $a = 1$，$b = 0$，其他参数设置为：最大迭代数 $K_{\max} = 5\,000$，最大历史长度 $N = 20$ 及容忍度 $\varepsilon = 10^{-5}$。

表1 两类算法运行时间(s)的比较

Tab.1 Comparison of running times between two algorithms

n	m/n	l/n	t(CSPM)/s	t/(CVX)/s
200	0.2	1.2	2.469	9.719
	0.2	1.6	0.469	11.5
	0.8	1.2	0.11	11.579
	0.8	1.6	0.109	12.906
300	0.2	1.2	7.734	29.5
	0.2	1.6	3	34.812
	0.8	1.2	0.313	35.578
	0.8	1.6	0.297	38.734
500	0.2	1.2	24.36	136.83
	0.2	1.6	8.656	162.45
	0.8	1.2	0.844	167.75
	0.8	1.6	0.672	162.7

实验在联想(CPU 2。1 GHz，内存2 GB)个人计算机上进行，程序采用 MATLAB 语言编写，并

在 MATLAB7.6.0 的环境下执行，每种规模下分别运行 10 次，并保证 CSPM 的相对误差小于 10^{-4}，相对误差定义为

$$RelErr = \frac{g_{\text{best}}^{k} - g^{*}}{g^{*}}$$

其中，g^{*} 事先通过凸规划软件包 CVX[13] 算出。CSPM 与 CVX 在每种规模下平均耗时（单位：s）列在表 1 中。从表 1 可以看出，CSPM 在收敛速度上优于传统凸优化方法，尤其是当 m/n 逼近 1 时，两者在平均耗时上差异悬殊。

参考文献：

[1] 钟玉泉. 复变函数论[M]. 3 版. 北京：高等教育出版社，2006：54－57.
[2] Kreutz D K. The complex gradient operator and the CR-calculus[EB/OL]. 2016－01－10. http://dsp.ucsd.edu/~kreutz/PEI－05%20Support%20Files/complex_derivatives.pdf.
[3] Wen J Z, Wen X. Fast estimation of sparse doubly spread acoustic channels[J]. Journal of the Acoustical Society of America, 2012, 131(1): 303－317.
[4] Xia Y L, Took C C, Mandic D P. An augmented affine projection algorithm for the filtering of noncircular complex signals[J]. Signal Processing, 2010, 90(6): 1788－1799.
[5] Brandwood D H. A complex gradient operator and its applica－tion in adaptive array theory[J]. IEE Proceedings of Communications Radar & Signal Processing, 1983, 130(1): 11－16.
[6] Mandic D P, Goh S L. Complex valued nonlinear adaptive filters: noncircularity, widely linear and neural models[M]. New York: Wiley, 2009: 55－68.
[7] 张贤达. 矩阵分析与应用[M]. 北京：清华大学出版社，2004：285－296.
[8] Nesterov Y, Shikkman V. Quasi-monotone subgradient methods for nonsmooth convex minimization[J]. Journal of Optimization Theory and Applications, 2015, 165(3): 917－940.
[9] Nesterov Y. Subgradient methods for huge-scale optimization problems[J]. Mathematical Programming, 2012, 146(1/2): 275－297.
[10] Nesterov Y. Primal-dual subgradient methods for convex problems[J]. Mathematical Programming, 2005, 120(1): 221－259.
[11] 龙强，李觉友. 次梯度法在求解非光滑最优化问题时的计算效果研究[J]. 重庆师范大学学报（自然科学版），2013，30(6)：25－30.
[12] Katsikis V N, Pappas D, Petralias A. An improved method for the computation of the Moore-Penrose inverse matrix[J]. Applied Mathematics & Computation, 2011, 217(23): 9828－9834.
[13] Grant M, Boyd S. CVX: MATLAB software for disciplined convex programming[EB/OL]. 2015－12－10, http://cvxr.com/cvx/.

（责任编辑：陈雯）