

关于凸极小化的 Douglas-Rachford 分裂方法的一个注

康倍倍, 董云达, 王亚丽

(郑州大学 数学与统计学院, 河南 郑州 450001)

摘 要: 在一个实的无穷维 Hilbert 空间中, 研究关于凸极小化的 Douglas-Rachford 分裂方法. 假设目标函数中的 f 和 g 均为闭的真凸函数, 并且 f 的梯度是 Lipschitz 连续的. 分析了 Douglas-Rachford 分裂方法的弱收敛性, 其中邻近参数可以变化并且上界与 f 的梯度的 Lipschitz 常数有关.

关键词: 凸极小化; Douglas-Rachford 分裂方法; 邻近参数; 弱收敛性

中图分类号: O221.2 文献标志码: A doi:10.13705/j.issn.1671-6833.2017.01.023

0 引言

设 H 为一个实的无穷维 Hilbert 空间, 假设 $f, g: H \rightarrow \mathbf{R} \cup \{+\infty\}$ 是闭的真凸函数, 并且 f 还是连续可微的. 考虑下面的凸极小化问题:

$$\min_{x \in H} f(x) + g(x).$$

该问题在优化、控制、管理以及信息科学中有极为广泛的应用. 在适当的条件下, 相应的最优性条件:

$$0 \in \nabla f(x) + \partial g(x). \tag{1}$$

Douglas-Rachford 分裂方法^[1-2] (以后简称 DR 方法) 是解决此问题的一个经典方法. 任取 $z^0 \in H, \lambda > 0$, 相应的迭代格式为:

$$\begin{aligned} x^k &= (I + \lambda \nabla f)^{-1}(z^k); \\ y^k &= (I + \lambda \partial g)^{-1}(2x^k - z^k); \\ z^{k+1} &= z^k + y^k - x^k, \end{aligned}$$

此时, z^k 弱收敛于 $\bar{z} \in H$, 且 $\bar{x} = (I + \lambda \nabla f)^{-1}(\bar{z})$ 是问题(1)的解.

最近, 在 f 的梯度是 L-Lipschitz 连续的条件上, Davis^[3] 证明了当 $\lambda < \kappa/L (\kappa \approx 1.246\ 98)$ 时,

$$f(y^k) + g(y^k) - f(x^*) - g(x^*) = o(1/k),$$

成立, 其中 x^* 是问题(1)的一个解点.

在本文中, 笔者讨论 DR 方法的等价形式^[2,4]: 任取 $x^0 \in H$, 相应的迭代格式:

$$x^k - \lambda_k \nabla f(x^k) \in (I + \lambda_k \partial g)(\bar{x}^k); \tag{2}$$

$$x^{k+1} + \lambda_k \nabla f(x^{k+1}) = \bar{x}^k + \lambda_k \nabla f(x^k). \tag{3}$$

其中, $\lambda_k > 0$ 被称为邻近参数, 它在每次迭代中可

以有所变化.

笔者主要证明了若邻近参数序列同时满足:

$$\sum_{k=0}^{+\infty} \lambda_k^2 = +\infty, 0 < \lambda_k \leq \sqrt{2}/L,$$

则 $\{x^k\}$ 弱收敛于问题(1)的一个解点.

对比于 Davis 的工作, 在每次迭代中, 允许邻近参数有所变化, 并且其上界更优.

1 预备知识

在这一节中, 给出几个重要的引理.

引理 1 若 ∇f 是 L-Lipschitz 连续的, 即存在某个常数 $L > 0$, 使得

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|, \forall x, y \in H,$$

则有

$$f(x) \leq f(y) + \langle x - y, \nabla f(y) \rangle + \frac{L}{2} \|x - y\|^2, \tag{4}$$

$$\frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2 \leq \langle x - y, \nabla f(x) - \nabla f(y) \rangle, \tag{5}$$

其中性质(5)由 Baillon 等^[5] 最早提出.

引理 2^[4] 考虑任一极大单调算子 $T: H \rightarrow H$. 假设 $\text{dom} T$ 中的序列 $\{x^k\}$ 弱收敛于某一点 x , 并且 H 中的序列 $\{s^k\}$ 强收敛于某一点 s , 若 $s^k \in T(x^k)$ 对于每一个都成立, 则 $s \in T(x)$.

引理 3^[6] 若 $\{a_k\}, \{b_k\}, \{c_k\}$ 均为正数序列, 并且满足:

$$b_{k+1} \leq b_k;$$

收稿日期:2016-10-25; 修订日期:2016-11-18
通信作者:董云达(1969—),男,河南原阳人,郑州大学副教授,博士,主要从事最优化方面的研究, E-mail: ydong@zzu.edu.cn.

$$\sum c_k = +\infty;$$

$$a_{k+1}^2 \leq a_k^2 - c_k b_k,$$

则有 $\lim_{k \rightarrow \infty} b_k = 0$.

2 主要结果

在这一节中,主要研究 DR 方法的弱收敛性.

定理 4 设 \mathbf{x}^* 为问题(1)的任一解点, $\{\mathbf{x}^k\}$ 和 $\{\bar{\mathbf{x}}^k\}$ 是由 DR 方法所产生的序列,则有

$$\|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 \leq \|\mathbf{x}^k - \mathbf{x}^*\|^2 - 2\lambda_k(f(\bar{\mathbf{x}}^k) + g(\bar{\mathbf{x}}^k) - f(\mathbf{x}^*) - g(\mathbf{x}^*)) - \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 - \left(\frac{2}{\lambda_k L} - \lambda_k L\right) \cdot \|\lambda_k \nabla f(\mathbf{x}^{k+1}) - \lambda_k \nabla f(\mathbf{x}^k)\|^2. \quad (6)$$

证明:由式(2)可得

$$\lambda_k^{-1}(\mathbf{x}^k - \bar{\mathbf{x}}^k) - \nabla f(\mathbf{x}^k) \in \partial g(\bar{\mathbf{x}}^k).$$

则由次微分的定义可得,对 $\forall \mathbf{x}$,有

$$g(\mathbf{x}) \geq g(\bar{\mathbf{x}}^k) + \langle \mathbf{x} - \bar{\mathbf{x}}^k, \lambda_k^{-1}(\mathbf{x}^k - \bar{\mathbf{x}}^k) - \nabla f(\mathbf{x}^k) \rangle.$$

$$g(\bar{\mathbf{x}}^k) \leq g(\mathbf{x}) + \langle \bar{\mathbf{x}}^k - \mathbf{x}, \lambda_k^{-1}(\mathbf{x}^k - \bar{\mathbf{x}}^k) - \nabla f(\mathbf{x}^k) \rangle. \quad (7)$$

因 ∇f 是 L-Lipschitz 连续的,故由引理 1 中式(4)可得

$$f(\bar{\mathbf{x}}^k) \leq f(\mathbf{x}^{k+1}) + \nabla f(\mathbf{x}^{k+1})^T(\bar{\mathbf{x}}^k - \mathbf{x}^{k+1}) + \frac{L}{2} \|\bar{\mathbf{x}}^k - \mathbf{x}^{k+1}\|^2. \quad (8)$$

又由

$$f(\mathbf{x}) \geq f(\mathbf{x}^{k+1}) + \nabla f(\mathbf{x}^{k+1})^T(\mathbf{x} - \mathbf{x}^{k+1}),$$

即

$$f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x}^{k+1})^T(\mathbf{x}^{k+1} - \mathbf{x}), \quad (9)$$

则式(8)与式(9)相加可得

$$f(\bar{\mathbf{x}}^k) \leq f(\mathbf{x}) + \nabla f(\mathbf{x}^{k+1})^T(\bar{\mathbf{x}}^k - \mathbf{x}) + \frac{L}{2} \|\bar{\mathbf{x}}^k - \mathbf{x}^{k+1}\|^2. \quad (10)$$

再由式(7)与式(10)相加可得

$$f(\bar{\mathbf{x}}^k) + g(\bar{\mathbf{x}}^k) - f(\mathbf{x}) - g(\mathbf{x}) \leq \frac{L}{2} \|\bar{\mathbf{x}}^k - \mathbf{x}^{k+1}\|^2 + \langle \bar{\mathbf{x}}^k - \mathbf{x}, \lambda_k^{-1}(\mathbf{x}^k - \bar{\mathbf{x}}^k) - \nabla f(\mathbf{x}^k) + \nabla f(\mathbf{x}^{k+1}) \rangle,$$

从而

$$\begin{aligned} 2\lambda_k(f(\bar{\mathbf{x}}^k) + g(\bar{\mathbf{x}}^k) - f(\mathbf{x}) - g(\mathbf{x})) &\leq \\ 2\langle \bar{\mathbf{x}}^k - \mathbf{x}, \mathbf{x}^k - \bar{\mathbf{x}}^k - \lambda_k \nabla f(\mathbf{x}^k) + \lambda_k \nabla f(\mathbf{x}^{k+1}) \rangle &+ \\ \lambda_k L \|\bar{\mathbf{x}}^k - \mathbf{x}^{k+1}\|^2 &= 2\langle \bar{\mathbf{x}}^k - \mathbf{x}, \mathbf{x}^k - \mathbf{x}^{k+1} \rangle + \lambda_k L; \\ \|\lambda_k \nabla f(\mathbf{x}^{k+1}) - \lambda_k \nabla f(\mathbf{x}^k)\|^2 &= 2\langle \mathbf{x}^{k+1} - \mathbf{x}, \mathbf{x}^k - \mathbf{x}^{k+1} \rangle + \\ 2\langle \bar{\mathbf{x}}^k - \mathbf{x}^{k+1}, \mathbf{x}^k - \mathbf{x}^{k+1} \rangle &+ \\ \lambda_k L \|\lambda_k \nabla f(\mathbf{x}^{k+1}) - \lambda_k \nabla f(\mathbf{x}^k)\|^2 &= \end{aligned}$$

$$\begin{aligned} \|\mathbf{x}^k - \mathbf{x}\|^2 - \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 - \|\mathbf{x}^{k+1} - \mathbf{x}\|^2 - \\ 2\lambda_k \langle \nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x}^{k+1}), \mathbf{x}^k - \mathbf{x}^{k+1} \rangle + \\ \lambda_k L \|\lambda_k \nabla f(\mathbf{x}^{k+1}) - \lambda_k \nabla f(\mathbf{x}^k)\|^2 \leq \\ \|\mathbf{x}^k - \mathbf{x}\|^2 - \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 - \|\mathbf{x}^{k+1} - \mathbf{x}\|^2 - \\ 2\lambda_k \frac{1}{L} \|\nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x}^{k+1})\|^2 + \\ \lambda_k L \|\lambda_k \nabla f(\mathbf{x}^{k+1}) - \lambda_k \nabla f(\mathbf{x}^k)\|^2 = \\ \|\mathbf{x}^k - \mathbf{x}\|^2 - \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 - \|\mathbf{x}^{k+1} - \mathbf{x}\|^2 - \\ \left(\frac{2}{\lambda_k L} - \lambda_k L\right) \|\lambda_k \nabla f(\mathbf{x}^{k+1}) - \lambda_k \nabla f(\mathbf{x}^k)\|^2. \end{aligned}$$

故,令 $\mathbf{x} = \mathbf{x}^*$ 即可得式(6).

定理 5 若邻近参数序列同时满足

$$\sum_{k=0}^{+\infty} \lambda_k^2 = +\infty, 0 < \lambda_k \leq \sqrt{2}/L, \quad (11)$$

则由 DR 方法所产生的迭代点序列 $\{\mathbf{x}^k\}$ 弱收敛于问题(1)的一个解点.

证明:设 \mathbf{x}^* 为问题(1)的任一解点,则由命题 4 知式(6)成立,即

$$\begin{aligned} \|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 &\leq \|\mathbf{x}^k - \mathbf{x}^*\|^2 - 2\lambda_k(f(\bar{\mathbf{x}}^k) + g(\bar{\mathbf{x}}^k) - \\ f(\mathbf{x}^*) - g(\mathbf{x}^*)) - \lambda_k^2 \|\lambda_k^{-1}(\mathbf{x}^k - \mathbf{x}^{k+1})\|^2 - \\ \left(\frac{2}{\lambda_k L} - \lambda_k L\right) \lambda_k^2 \|\nabla f(\mathbf{x}^{k+1}) - \nabla f(\mathbf{x}^k)\|^2. \end{aligned}$$

由假设条件(11)知 $\frac{2}{\lambda_k L} - \lambda_k L \geq 0$,从而

$$\begin{aligned} \|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 &\leq \|\mathbf{x}^k - \mathbf{x}^*\|^2 - \\ \lambda_k^2 \|\lambda_k^{-1}(\mathbf{x}^k - \mathbf{x}^{k+1})\|^2. \end{aligned} \quad (12)$$

令 $b_k = \min_{i=1,2,\dots,k} \{\|\lambda_i^{-1}(\mathbf{x}^i - \mathbf{x}^{i+1})\|^2\}$; $c_k = \lambda_k^2$,再由引理 3 和假设条件(11)可知 $b_k \rightarrow 0$. 这意味着序列 $\{\|\lambda_k^{-1}(\mathbf{x}^k - \mathbf{x}^{k+1})\|\}$ 中存在一个子列 $\{\|\lambda_{k_j}^{-1}(\mathbf{x}^{k_j} - \mathbf{x}^{k_j+1})\|\}$ 收敛于 0.

由式(6)可得 $\|\mathbf{x}^k - \mathbf{x}^{k+1}\| \rightarrow 0$. 因为 ∇f 是 L-Lipschitz 连续的,所以

$\|\nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x}^{k+1})\| \leq L \|\mathbf{x}^k - \mathbf{x}^{k+1}\| \rightarrow 0$, 从而整个序列 $\{\nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x}^{k+1})\}$ 以范数收敛于 0. 结合式(3)知,

$$\lambda_k^{-1}(\mathbf{x}^k - \bar{\mathbf{x}}^k) = \lambda_k^{-1}(\mathbf{x}^k - \mathbf{x}^{k+1}) + \nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x}^{k+1}),$$

$$\|\lambda_{k_j}^{-1}(\mathbf{x}^{k_j} - \bar{\mathbf{x}}^{k_j})\| \rightarrow 0. \quad (13)$$

另一方面,由式(3)和假设条件(11)知,

$$\begin{aligned} \|\mathbf{x}^k - \bar{\mathbf{x}}^k\| &= \|\mathbf{x}^k - \mathbf{x}^{k+1} + \lambda_k(\nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x}^{k+1}))\| \leq \\ \|\mathbf{x}^k - \mathbf{x}^{k+1}\| &+ \lambda_k \|\nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x}^{k+1})\| \leq \\ \|\mathbf{x}^k - \mathbf{x}^{k+1}\| &+ \frac{\sqrt{2}}{L} \|\nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x}^{k+1})\| \leq \\ (1 + \sqrt{2}) \|\mathbf{x}^k - \mathbf{x}^{k+1}\| &\rightarrow 0. \end{aligned} \quad (14)$$

接下来,我们来证明迭代点序列 $\{\mathbf{x}^k\}$ 的弱收敛性. 首先,由式(12)知,序列 $\{\|\mathbf{x}^k - \mathbf{x}^*\|\}$ 是非负递减的,从而极限存在. 这意味着序列 $\{\mathbf{x}^k\}$ 是有界的. 显然,序列 $\{\mathbf{x}^{k_j}\}$ 也是有界的,从而存在子列 $\{\mathbf{x}^{k_{j_l}}\}$,使得 $\{\mathbf{x}^{k_{j_l}}\}$ 弱收敛于 \mathbf{x}^∞ . 结合式(14)知, $\{\bar{\mathbf{x}}^{k_{j_l}}\}$ 也弱收敛于 \mathbf{x}^∞ . 其次,由式(2)知,

$$\lambda_k^{-1}(\mathbf{x}^k - \bar{\mathbf{x}}^k) - (\nabla f(\mathbf{x}^k) - \nabla f(\bar{\mathbf{x}}^k)) \in (\nabla f + \partial g)(\bar{\mathbf{x}}^k).$$

沿着 k_{j_l} 取极限,并结合 $\nabla f + \partial g$ 的极大单调性、式(13)、式(14)和引理 2 可得:

$$0 \in (\nabla f + \partial g)(\mathbf{x}^\infty),$$

即 \mathbf{x}^∞ 为问题(1)的一个解点. 至于迭代序列弱聚点的唯一性的证明,读者可参考文献[4]. 证毕.

3 结论

笔者对于凸极小化式(1)的 DR 方法,研究了邻近参数可以变化的情形,并且在保证全局弱收敛性的前提下,揭示了邻近参数的上界与式(1)中相应梯度的 Lipschitz 常数的内在关系. 至于邻近点算法的 $o(1/k)$ -收敛率,可以参考文献[6]. 作为其特例的 DR 方法的情形,可以参考文献[3, 7]. 如何将笔者工作运用于压缩感知^[8],则是一个值得探讨的研究课题.

参考文献:

[1] LIONS P L, MERCIER B. Splitting algorithms for the

sum of two nonlinear operators[J]. SIAM journal on numerical analysis, 1979, 16: 964 - 979.

[2] ECKSTEIN J, BERTSEKAS D P. On douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators [J]. Mathematical programming, 1992, 55(3): 293 - 318.

[3] DAVIS D. On the design and analysis of operator splitting schemes[D]. Ph. D dissertation, UCLA, 2015.

[4] DONG Y D, FISCHER A. A family of operator splitting methods revisited[J]. Nonlinear analysis, 2010, 72: 4307 - 4315.

[5] BAILLON J B, HADDAD G. Quelques propriétés des opérateurs anglebornés et n-cycliquement monotones[J]. Israel journal of mathematics, 1977, 26: 137 - 150.

[6] DONG Y D. The proximal point algorithm revisited[J]. Journal of optimization theory and applications, 2014, 161: 478 - 489.

[7] DONG Y D. Douglas-Rachford splitting method for semidefinite programming[J]. Journal of applied mathematics and computing, 2016, 51(1): 569 - 591.

[8] 张红梅,温荟然,张向利,等. 基于压缩特征的稀疏表示运动目标跟踪[J]. 郑州大学学报(工学版), 2016, 37(3): 21 - 26.

A Note on Douglas-Rachford Splitting Method for Convex Minimization

KANG Beibei, DONG Yunda, WANG Yali

(School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, China)

Abstract: In a real infinite-dimensional Hilbert space, Douglas-Rachford splitting method for convex minimization was studied. If f and g in the objective function were closed, proper convex, and the f 's gradient was Lipschitz continuous, then the method's weak convergence was analyzed. Our analysis allowed the corresponding proximal parameters to vary from iteration to iteration and their upper bound relied on Lipschitz constant of f 's gradient.

Key words: convex minimization; Douglas-Rachford splitting method; proximal parameter; weak convergence