网络优先数字出版时间:2016-11-21　　【DOI】10.13656/j.cnki.gxkx.20161121.005
网络优先数字出版地址:http://www.cnki.net/kcms/detail/45.1206.G3.20161121.1520.010.html

非凸两分块问题乘子交替方向法的收敛性分析*

Convergence Analysis of Alternating Direction Method of Multipliers for Two Block Nonconvex Problems

邓　钊[1],晁绵涛[1],简金宝[2]**
DENG Zhao[1],CHAO Miantao[1],JIAN Jinbao[2]

(1.广西大学数学与信息科学学院,广西南宁　530004;2.玉林师范学院数学与统计学院,广西玉林　537000)
(College of Mathematics and Information Science,Guangxi University,Nanning,Guangxi 530004,China;School of Mathematics and Statistics,Yulin Normal University,Yulin,Guangxi,537000,China)

摘要:乘子交替方向法(ADMM)求解大规模问题十分有效.ADMM 在凸情形下的收敛性已被清晰认识,但非凸问题 ADMM 的收敛性结果还很少.本文针对非凸两分块优化问题,在增广拉格朗日函数满足 Kurdyka-Lojasiewicz 不等式性质且罚参数大于某个常数的条件下,证明了 ADMM 的收敛性.

关键词:乘子交替方向法　Kurdyka-Lojasiewicz 不等式　非凸优化　收敛性
中图分类号:O221.2　　**文献标识码**:A　　**文章编号**:1005-9164(2016)05-0422-06

Abstract:The Alternating Direction Method of Multipliers(ADMM) is an effective method for large scale optimization problems. While the convergence of ADMM has been clearly recognized in the case of convex,the convergence result of ADMM in the case of nonconvex is still an open problem. In this paper,under the assumption that the augmented Lagrangian function satisfies the Kurdyka-Lojasiewicz inequality and the penalty parameter is greater than a constant,we analyze the convergence of ADMM for a class of nonconvex optimization problems whose objective function is the sum of two block nonconvex functions.

Key words:Alternating Direction Method of Multipliers,Kurdyka-Lojasiewicz inequality,nonconvex optimization,convergence

0　引言

【研究意义】乘子交替方向法(ADMM)求解大规模分布式计算问题十分有效.ADMM 既能分散地收集和存储这些数据集,又能在并行和分布式的环境下求解这些问题.ADMM 适合求解分布式凸优化问题,尤其适用于出现在统计学、机器学习和相关领域中的大规模问题,其重要性日益凸显.【前人研究进展】ADMM 的思想最早起源于 20 世纪 50 年代,算法在 20 世纪 70 年代中期由 Glowinski 和 Marrocco[1],以及 Gabay 和 Mercier[2] 首次提出.20 世纪 80 年代,乘子交替方向法的研究和应用非常广泛,直到 20 世纪 90 年代中期,乘子交替方向法求解凸优化问题的很多理论结果,已经得到很好的证明.传统 ADMM 是求解凸两分块问题十分有效的方法[1-2],其直接应

收稿日期:2016-07-01
修回日期:2016-10-27
作者简介:邓　钊(1991－),男,硕士研究生,主要从事最优化理论与方法研究。
*国家自然科学基金项目(11601095),广西自然科学基金项目(2014GXNSFFA118001,2016GXNSFDA380019)和广西高校科研项目(ZD201407)资助。
**通信作者:简金宝(1964－),男,教授,博士,博士生导师,主要从事最优化理论与算法及应用研究,E-mail:jianjb@gxu.edu.cn。

用到问题(0.5)的迭代框架如下

$$\begin{cases} x^{k+1} \in \arg\min\{L_\beta(x, y^k, \lambda^k)\}, \\ y^{k+1} \in \arg\min\{L_\beta(x^{k+1}, y, \lambda^k)\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1}), \end{cases} \tag{0.1}$$

其中 $L_\beta(x,y,\lambda)$ 为问题(0.5)的增广拉格朗日函数，定义如下

$$L_\beta(x,y,\lambda) = f(x) + g(y) - \lambda^{\mathrm{T}}(Ax + By) + \frac{\beta}{2}\|Ax + By\|^2. \tag{0.2}$$

在凸情形下，ADMM 的收敛性已被充分认识[3]. 非凸问题 ADMM 的收敛性分析仅有初步的结果，其研究是当前的热点问题[4-6]. 文献[7]考虑如下非凸问题

$$\begin{aligned} \min \quad & f(x) + g(y) \\ \text{s.t.} \quad & Ax = y, \end{aligned} \tag{0.3}$$

并提出如下改进的 ADMM

$$\begin{cases} x^{k+1} \in \arg\min\{L_\beta(x, y^k, \lambda^k)\}, \\ y^{k+1} \in \arg\min\{L_\beta(x^{k+1}, y, \lambda^k)\} + \Delta_\phi(y, y^k), \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - y^{k+1}), \end{cases} \tag{0.4}$$

其中 Δ_ϕ 是关于强凸函数 ϕ 的 Bregman 距离. 当 $\phi = 0$ 时，算法(0.4)为传统 ADMM. 在罚参数充分大且目标函数二阶连续可微的条件下，文献[7]证明算法产生点列的聚点是问题(0.3)的稳定点.

文献[8]分析如下 Bregman ADMM 算法的收敛性

$$\begin{cases} x^{k+1} \in \arg\min\{L_\beta(x, y^k, \lambda^k)\} + \Delta_\varphi(x, x^k), \\ y^{k+1} \in \arg\min\{L_\beta(x^{k+1}, y, \lambda^k)\} + \Delta_\phi(y, y^k), \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - y^{k+1}). \end{cases}$$

【本研究切入点】最近，文献[9]在矩阵 A 列满秩，增广拉格朗日函数满足 KL 性质(参见定义 1.1)且罚参数大于某个常数的条件下，分析了传统 ADMM 算法(0.1)求解非凸问题(0.3)的全局收敛性.

我们考虑如下两分块极小化问题

$$\begin{aligned} \min \quad & f(x) + g(y) \\ \text{s.t.} \quad & Ax + By = 0, \end{aligned} \tag{0.5}$$

其中函数 $f:\mathbb{R}^n \to \mathbb{R}^n \cup \{+\infty\}$ 是一个非凸正常下半连续函数，函数 $g:\mathbb{R}^m \to \mathbb{R}$ L-Lipschitz 可微，即存在 $L > 0$ 使得 $\|\nabla g(x) - \nabla g(y)\| \leqslant L\|x - y\|$，$\forall x, y \in \mathbb{R}^n$)，矩阵 $A \in \mathbb{R}^{m\times n}$，$B \in \mathbb{R}^{m\times n}$. 这类问题广泛出现在实际应用中，如矩阵分解、图像处理、信号处理等[4-5].

【拟解决的关键问题】本文在不要求矩阵 A 列满秩，B 不一定是单位阵，在 $L_\beta(w)$ 满足 KL 性质且罚参数 β 大于某个常数的条件下分析了 ADMM 算法(0.1)求解问题(0.5)的收敛性.

1 预备知识

下面，给出本文理论分析所需的一些概念与性质.

$\lambda_{++}(B^{\mathrm{T}}B)$ 表示矩阵 $B^{\mathrm{T}}B$ 最小正特征值. $\partial f(x)$ 表示函数 f 在点 x 处的极限次微分[5]，对于任意 $x \in \mathbb{R}^n$ 是函数 f 的极小值点的必要条件是：$0 \in \partial f(x)$，满足这个条件的点称为关键点或稳定点，函数 f 关键点集记为 crit f.

定义 1.1[10] (Kurdyka-Lojasiewicz 性质) 设函数 $f:\mathbb{R}^n \to \mathbb{R}^n \cup \{+\infty\}$ 是正常下半连续函数，对于任意实数 $\eta_1, \eta_2 (\eta_1 < \eta_2)$，令 $[\eta_1 < f < \eta_2] = x \in \{\mathbb{R}^n : \eta_1 < f(x) < \eta_2$，设 $x^* \in \mathrm{dom}\partial f$，若存在 $\eta \in (0, +\infty]$，x^* 的邻域 U，以及一个连续的凹函数 $\varphi:[0,\eta) \to \mathbb{R}_+$，满足如下条件

(i) $\varphi(0) = 0$;

(ii) φ 在 0 处连续，在区间 $(0,\eta)$ 上一阶连续可微;

(iii) $\varphi'(s) > 0, \forall s \in (0,\eta)$;

(iv) 对于任意的 $x \in U \cap [f(x^*) < f < f(x^*) + \eta]$，有如下 Kurdyka-Lojasiewicz 不等式成立

$$\varphi'(f(x) - f(x^*))d(0, \partial f(x)) \geqslant 1.$$

则称函数 f 在点 x^* 处满足 Kurdyka-Lojasiewicz 性质(简称 KL 性质).

满足上述性质(i)、(ii)、(iii)的函数全体记为 Φ_η.

引理 1.1[11] (uniformized KL property) 设 Ω 是一个紧集，函数 $f:\mathbb{R}^n \to \mathbb{R}^n \cup \{+\infty\}$ 是正常下半连续函数. 设函数 f 在集合 Ω 上取常数，并在 Ω 中任一点处满足 KL 性质，则存在 $\epsilon > 0$，$\eta > 0$，$\varphi \in \Phi_\eta$，对于任意的 $\bar{x} \in \Omega$ 和 $x \in \{x \in \mathbb{R}^n : d(x,\omega) < \epsilon\} \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]$，有

$$\varphi'(f(x) - f(\bar{x}))d(0, \partial f(x)) \geqslant 1.$$

引理 1.2[12] 若 $h:\mathbb{R}^n \to \mathbb{R}$ 为 L-Lipschitz 可微函数，则对任意的 $x, y \in \mathbb{R}^n$，有

$$|h(y) - h(x) - \langle \nabla h(x), y - x \rangle| \leqslant \frac{L}{2}\|y - x\|^2.$$

2 收敛性分析

假设 $\{w^k = (x^k, y^k, \lambda^k)\}$ 是算法(0.1)产生的有界序列，算法(0.1)的收敛性分析框架如下：第一步，

证明$\{L_\beta(w^k)\}$单调递减;第二步,证明$\sum_{k=0}^{+\infty}\|w^{k+1}-w^k\|^2<+\infty$;第三步,利用$L_\beta(\cdot)$的KL性质,得出序列$\{w^k\}$的任一聚点都是问题(0.5)的一个稳定点.

由算法(0.1)中每个子问题的最优性条件,有

$$\begin{cases}0\in\partial f(x^{k+1})-A^{\mathrm T}\lambda^k+\beta A^{\mathrm T}(Ax^{k+1}+By^k),\\0=\nabla g(y^{k+1})-B^{\mathrm T}\lambda^k+\beta B^{\mathrm T}(Ax^{k+1}+By^{k+1}),\\\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}).\end{cases}\tag{2.1}$$

进一步,可得

$$\begin{cases}A^{\mathrm T}\lambda^{k+1}+\beta A^{\mathrm T}B(y^{k+1}-y^k)\in\partial f(x^{k+1}),\\\nabla g(y^{k+1})=B^{\mathrm T}\lambda^{k+1},\\\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}).\end{cases}\tag{2.2}$$

本文的收敛性建立在如下假设条件下.

假设 2.1 假设以下条件成立

(i) $\mathrm{Im}(A)\subseteq\mathrm{Im}(B)$;

(ii) 存在利普希茨系数分别为$M>0,P>0$的利普希茨连续函数$H:\mathrm{Im}(B)\to\mathbb{R}^n,F:\mathrm{Im}(A)\to\mathbb{R}^n$,使得$H(u)=\arg\min_y\{g(y):By=u\},F(u)=\arg\min_x\{f(x):Ax=u\}$;

(iii) $\delta:=\frac{\beta}{2M^2}-\frac{\mu^2L^2}{\beta}-\frac{L}{2}>0$,其中$\mu:=\lambda_{++}^{-\frac12}(B^{\mathrm T}B)$.

首先,证明$\{L_\beta(w^k)\}_{k\in N}$是递减序列.

引理 2.1 $L_\beta(w^{k+1})\leqslant L_\beta(w^k)-\delta\|y^{k+1}-y^k\|^2$,其中$\delta:=\frac{\beta}{2M^2}-\frac{\mu^2L^2}{\beta}-\frac{L}{2}>0$.

证明 由增广拉格朗日函数的定义,可得

$$L_\beta(x^{k+1},y^{k+1},\lambda^{k+1})=L_\beta(x^{k+1},y^{k+1},\lambda^k)+<\lambda^k-\lambda^{k+1},Ax^{k+1}+By^{k+1}>=L_\beta(x^{k+1},y^{k+1},\lambda^k)+\frac{1}{\beta}\|\lambda^k-\lambda^{k+1}\|^2.\tag{2.3}$$

利用y^{k+1}的最优性可得

$$L_\beta(x^{k+1},y^k,\lambda^k)-L_\beta(x^{k+1},y^{k+1},\lambda^k)=f(x^{k+1})+g(y^k)-<\lambda^k,Ax^{k+1}+By^k>+\frac{\beta}{2}\|Ax^{k+1}+By^k\|^2-\{f(x^{k+1})+g(y^{k+1})-<\lambda^k,Ax^{k+1}+By^{k+1}>+\frac{\beta}{2}\|Ax^{k+1}+By^{k+1}\|^2\}=g(y^k)-g(y^{k+1})+<\lambda^k,B(y^{k+1}-y^k)>+\frac{\beta}{2}\|Ax^{k+1}+By^k\|^2-\frac{\beta}{2}\|Ax^{k+1}+By^{k+1}\|^2.\tag{2.4}$$

由引理1.2和式(2.2)中第二式可得

$$g(y^k)-g(y^{k+1})\geqslant<\lambda^{k+1},B(y^k-y^{k+1})>-\frac{L}{2}\|y^k-y^{k+1}\|^2.$$

把上式代入式(2.4)可得

$$L_\beta(x^{k+1},y^k,\lambda^k)-L_\beta(x^{k+1},y^{k+1},\lambda^k)\geqslant<\lambda^{k+1}-\lambda^k,B(y^k-y^{k+1})>+\frac{\beta}{2}\|Ax^{k+1}+By^k\|^2-\frac{\beta}{2}\|Ax^{k+1}+By^{k+1}\|^2-\frac{L}{2}\|y^k-y^{k+1}\|^2,\tag{2.5}$$

由$\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1})$可得

$$Ax^{k+1}+By^k=\frac{1}{\beta}(\lambda^k-\lambda^{k+1})+B(y^k-y^{k+1}).$$

故

$$<\lambda^{k+1}-\lambda^k,B(y^k-y^{k+1})>+\frac{\beta}{2}\|Ax^{k+1}+By^k\|^2=<\lambda^{k+1}-\lambda^k,B(y^k-y^{k+1})>+\frac{\beta}{2}\|\frac{1}{\beta}(\lambda^k-\lambda^{k+1})+B(y^k-y^{k+1})\|^2=\frac{\beta}{2}\|B(y^k-y^{k+1})\|^2+\frac{1}{2\beta}\|\lambda^{k+1}-\lambda^k\|^2.$$

把上式代入式(2.5)右端可得

$$L_\beta(x^{k+1},y^{k+1},\lambda^k)-L_\beta(x^{k+1},y^k,\lambda^k)\leqslant\frac{L}{2}\|y^k-y^{k+1}\|^2-\frac{\beta}{2}\|B(y^k-y^{k+1})\|^2.\tag{2.6}$$

由$\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1})$可得

$$\lambda^{k+1}-\lambda^k=-\beta(Ax^{k+1}+By^{k+1})\in\mathrm{Im}(B).$$

$\mu:=\lambda_{++}^{-\frac12}(B^{\mathrm T}B)>0$,则

$$\|\lambda^{k+1}-\lambda^k\|\leqslant\mu\|B(\lambda^{k+1}-\lambda^k)\|=\mu\|\nabla g(y^{k+1})-\nabla g(y^k)\|\leqslant\mu L\|y^{k+1}-y^k\|.\tag{2.7}$$

由H的性质可得$y^k=H(By^k)$,从而

$$\|y^{k+1}-y^k\|=\|H(By^{k+1})-H(By^k)\|\leqslant M\|B(y^{k+1}-y^k)\|.\tag{2.8}$$

结合式(2.6)和引理2.1可得

$$L_\beta(x^{k+1},y^{k+1},\lambda^{k+1})\leqslant L_\beta(x^{k+1},y^k,\lambda^k)+\frac{1}{\beta}\|\lambda^k-\lambda^{k+1}\|^2+\frac{L}{2}\|y^k-y^{k+1}\|^2-\frac{\beta}{2}\|B(y^k-y^{k+1})\|^2.$$

结合式(2.7)、式(2.8)有

$$L_\beta(x^{k+1},y^{k+1},\lambda^{k+1})\leqslant L_\beta(x^{k+1},y^k,\lambda^k)-(\frac{\beta}{2M^2}-\frac{\mu^2L^2}{\beta}-\frac{L}{2})\|y^{k+1}-y^k\|^2.$$

利用x^{k+1}的最优性可得

$$L_\beta(x^{k+1},y^{k+1},\lambda^{k+1})\leqslant L_\beta(x^k,y^k,\lambda^k)-(\frac{\beta}{2M^2}-\frac{\mu^2L^2}{\beta}-\frac{L}{2})\|y^{k+1}-y^k\|^2.$$

引理 2.2 $\sum_{k=0}^{+\infty}\|w^{k+1}-w^k\|^2<+\infty$.

证明　由于序列$\{w^k=(x^k,y^k,\lambda^k)\}$有界，则存在收敛子列$\{w^{k_j}\}$，设$\lim\limits_{k_j=+\infty} w^{k_j}\to w^*$. 由$f$下半连续及$g$连续，可知$L_\beta(w)$下半连续. 从而

$$L_\beta(w^*)\leqslant \liminf_{j\to+\infty} L_\beta(w^{k_j}).$$

故序列$\{L_\beta(w^{k_j})\}$有下界，又序列$\{L_\beta(w^k)\}_{k\in N}$单调递减，所以$L_\beta(w^{k_j})$收敛并且$L_\beta(w^k)\geqslant L_\beta(w^*)$. 由引理2.1可得

$$\sum_{k=0}^{n}\delta\|y^{k+1}-y^k\|^2\leqslant L_\beta(w^0)-L_\beta(w^{n+1})\leqslant L_\beta(w^0)-L_\beta(w^*)<+\infty,$$

由$\delta>0$及n的任意性可得

$$\sum_{k=0}^{+\infty}\|y^{k+1}-y^k\|^2<+\infty.$$

上式结合式(2.7)可得

$$\sum_{k=0}^{+\infty}\|\lambda^{k+1}-\lambda^k\|^2<+\infty.$$

接下来只需证明：$\sum\limits_{k=0}^{+\infty}\|x^{k+1}-x^k\|^2<+\infty$. 由ADMM算法(0.1)第三式有

$$\begin{cases}\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}),\\ \lambda^k=\lambda^{k-1}-\beta(Ax^k+By^k).\end{cases}$$

两式相减，可得

$$\lambda^{k+1}-\lambda^k=(\lambda^k-\lambda^{k-1})+\beta(Ax^k-Ax^{k+1})+\beta(By^k-By^{k+1}).$$

利用不等式$(a+b+c)^2\leqslant 3(a^2+b^2+c^2)$可得

$$\|\beta(Ax^k-Ax^{k+1})\|^2\leqslant 3(\|\lambda^{k+1}-\lambda^k\|^2+\|\lambda^k-\lambda^{k-1}\|^2+\beta^2\|B(y^{k+1}-y^k)\|^2). \tag{2.9}$$

由F的性质可得$x^k=F(Ax^k)$，进而

$$\|\beta(Ax^k-Ax^{k+1})\|^2\geqslant\frac{\beta^2}{P}\|x^k-x^{k+1}\|^2. \tag{2.10}$$

由式(2.9)和式(2.10)可得

$$\sum_{k=0}^{+\infty}\|x^{k+1}-x^k\|^2<+\infty.$$

故$\sum\limits_{k=0}^{+\infty}\|w^{k+1}-w^k\|^2<+\infty$.

引理2.3　存在$\zeta>0$，对于任意k有

$$d(0,\partial L_\beta(w^{k+1}))\leqslant\zeta\|y^{k+1}-y^k\|.$$

证明　由增广拉格朗日函数定义，可得

$$\begin{cases}\partial_x L_\beta(w^{k+1})=\partial f(x^{k+1})-A^{\mathrm T}\lambda^{k+1}+\beta A^{\mathrm T}(Ax^{k+1}+By^{k+1}),\\ \partial_y L_\beta(w^{k+1})=\nabla g(y^{k+1})-B^{\mathrm T}\lambda^{k+1}+\beta B^{\mathrm T}(Ax^{k+1}+By^{k+1}),\\ \partial_\lambda L_\beta(w^{k+1})=-(Ax^{k+1}+By^{k+1}).\end{cases} \tag{2.11}$$

进一步结合式(2.2)可得

$$\begin{cases}\beta A^{\mathrm T}B(y^{k+1}-y^k)+A^{\mathrm T}(\lambda^k-\lambda^{k+1})\in\partial_x L_\beta(w^{k+1}),\\ B^{\mathrm T}(\lambda^k-\lambda^{k+1})\in\partial_y L_\beta(w^{k+1}),\\ \dfrac{1}{\beta}(\lambda^{k+1}-\lambda^k)\in\partial_\lambda L_\beta(w^{k+1}).\end{cases} \tag{2.12}$$

令$x_{k+1}^*=\beta A^{\mathrm T}B(y^{k+1}-y^k)+A^{\mathrm T}(\lambda^k-\lambda^{k+1})$，$y_{k+1}^*=B^{\mathrm T}(\lambda^k-\lambda^{k+1})$，$\lambda_{k+1}^*=\dfrac{1}{\beta}(\lambda^{k+1}-\lambda^k)$，则

$(x_{k+1}^*,y_{k+1}^*,\lambda_{k+1}^*)\in\partial L_\beta(w^{k+1})$且存在$\zeta_1,\zeta_2>0$，使得

$$\|(x_{k+1}^*,y_{k+1}^*,\lambda_{k+1}^*)\|\leqslant\zeta_1\|y^{k+1}-y^k\|+\zeta_2\|\lambda^{k+1}-\lambda^k\|.$$

令$\zeta:=\zeta_1+\mu L\zeta_2$，结合式(2.7)可得

$$d(0,\partial L_\beta(w^{k+1}))\leqslant\zeta\|y^{k+1}-y^k\|.$$

引理2.4　设序列$\{w^k\}$的全体极限点记为$S(w^0)$，则

(i) $S(w^0)$是一个非空紧集，并且

$d(w^k,S(w^0))\to 0, k\to+\infty$；

(ii) $S(w^0)\subset \mathrm{crit}\,L_\beta$；

(iii) $L_\beta(\cdot)$在$S(w^0)$上取有限值且为常数，且

$\inf\limits_{k\in N}L_\beta(w^k)=\lim\limits_{k\to+\infty}L_\beta(w^k)$.

证明　(i)式由$S(w^0)$的定义直接可得.

(ii) 设$(x^*,y^*,\lambda^*)\in S(w^0)$，则存在子列$\{(x^{k_j},y^{k_j},\lambda^{k_j})\}$使得$(x^{k_j},y^{k_j},\lambda^{k_j})\to(x^*,y^*,\lambda^*)$，$(k_j\to+\infty)$. 由$x^{k+1}$的最优性有

$$L_\beta(x^{k+1},y^k,\lambda^k)\leqslant L_\beta(x^*,y^k,\lambda^k).$$

由引理2.2可知$\|w^{k+1}-w^k\|\to 0$，从而$(x^{k_j+1},y^{k_j+1},\lambda^{k_j+1})\to(x^*,y^*,\lambda^*)$. 又$L_\beta(\cdot)$关于$y,\lambda$连续，则有

$$\lim_{j\to+\infty}\sup L_\beta(x^{k_j+1},y^{k_j},\lambda^{k_j})=\lim_{j\to+\infty}\sup L_\beta(x^{k_j+1},y^{k_j+1},\lambda^{k_j+1})\leqslant L_\beta(x^*,y^*,\lambda^*).$$

由函数$L_\beta(\cdot)$的下半连续性，有

$$\liminf_{j\to+\infty} L_\beta(x^{k_j+1},y^{k_j+1},\lambda^{k_j+1})\geqslant L_\beta(x^*,y^*,\lambda^*).$$

由上式及式(2.13)可得$\lim\limits_{j\to+\infty}f(x^{k_j+1})=f(x^*)$. 结合$\nabla g$的连续性和$\partial f$的闭性，在式(2.2)中对序列$(x^{k_j+1},y^{k_j+1},\lambda^{k_j+1})$取极限，可得

$$\begin{cases}A^{\mathrm T}\lambda^*\in\partial f(x^*),\\ \nabla g(y^*)=B^{\mathrm T}\lambda^*,\\ Ax^*+By^*=0.\end{cases}$$

即$w^*\subset\mathrm{crit}L_\beta$.

(iii) 对于任意点$(x^*,y^*,\lambda^*)\in S(w^0)$，存在子列$(x^{k_j},y^{k_j},\lambda^{k_j})\to(x^*,y^*,\lambda^*)$. 结合$L_\beta(x^{k_j+1})$收敛，以及$\{L_\beta(w^k)\}_{k\in N}$单调递减，可得

$$\lim_{k\to+\infty} L_\beta(x^k,y^k,\lambda^k)=L_\beta(x^*,y^*,\lambda^*).$$

因此,函数$L_\beta(\cdot)$在集合$S(w^0)$上是常数,并且有$\inf_{k\in N} L_\beta(w^k)=\lim_{k\to+\infty} L_\beta(w^k)$.

最后,给出非凸问题(0.5)的ADMM算法(0.1)的收敛性分析.

定理 2.1 若$L_\beta(w)$满足KL性质,则

(i) $\sum_{k=0}^{+\infty}\|w^{k+1}-w^k\|<+\infty$;

(ii) 序列$\{w^k\}$收敛到函数$L_\beta(\cdot)$的一个关键点.

证明 由引理2.4知$\lim_{k\to+\infty} L_\beta(w^k)=L_\beta(w^*)(\forall w^*\in S(w^0))$成立.考虑如下两种情形:

(i) 存在整数k_0使得$L_\beta(w^{k_0})=L_\beta(w^*)$成立,由引理2.1可知,对于任意的$k>k_0$,有

$$\|y^{k+1}-y^k\|^2\leqslant L_\beta(w^k)-L_\beta(w^{k+1})\leqslant L_\beta(w^{k_0})-L_\beta(w^*)=0,$$

因此,对于任意的$k>k_0$,有$y^{k+1}=y^k$.结合式(2.7)、式(2.9)、式(2.10)可知,对于任意的$k>k_0+1$,有$w^{k+1}=w^k$成立,结论成立.

(ii) 对任意的k均有$L_\beta(w^k)>L_\beta(w^*)$成立.由$d(w^k,S(w^0))\to 0$可知,对于任意给定的$\varepsilon>0$,存在$k_1>0$,当$k>k_1$时,有$d(w^k,S(w^0))<\varepsilon$.又根据$L_\beta(w^k)\to L_\beta(w^*)$可知,对于任意给定的$\eta>0$,存在$k_2>0$,当$k>k_2$时,有$L_\beta(w^k)<L_\beta(w^*)+\eta$.从而当$k>\tilde{k}=\max\{k_1,k_2\}$时,有

$$d(w^k,S(w^0))<\varepsilon,$$

$$L_\beta(w^*)<L_\beta(w^k)<L_\beta(w^*)+\eta.$$

由于$S(w^0)$是非空紧集,函数$L_\beta(\cdot)$在集合$S(w^0)$上是常数,由引理1.1得

$$\varphi'(L_\beta(w^k)-L_\beta(w^*))d(0,\partial L_\beta(w^k))\geqslant 1(\forall k>\tilde{k}).$$

由函数φ的凹性,可得

$$\varphi(L_\beta(w^k)-L_\beta(w^*))-\varphi(L_\beta(w^{k+1})-L_\beta(w^*))\geqslant\varphi'(L_\beta(w^k)-L_\beta(w^*))(L_\beta(w^k)-L_\beta(w^{k+1})).$$

由引理2.3及$\varphi'(L_\beta(w^k)-L_\beta(w^*))>0$,可得

$$L_\beta(w^k)-L_\beta(w^{k+1})\leqslant\frac{\varphi(L_\beta(w^k)-L_\beta(w^*))-\varphi(L_\beta(w^{k+1})-L_\beta(w^*))}{\varphi'(L_\beta(w^k)-L_\beta(w^*))}\leqslant\zeta\|y^k-y^{k-1}\|[\varphi(L_\beta(w^k)-L_\beta(w^*))-\varphi(L_\beta(w^{k+1})-L_\beta(w^*))].$$

令$\Delta_{p,q}=\varphi(L_\beta(w^p)-L_\beta(w^*))-\varphi(L_\beta(w^q)-L_\beta(w^*))$.由引理2.1可得,对于任意的$k>\tilde{k}$有

$$\delta\|y^{k+1}-y^k\|^2\leqslant\zeta\|y^k-y^{k-1}\|\Delta_{k,k+1},$$

即

$$\|y^{k+1}-y^k\|\leqslant\sqrt{\frac{\zeta}{\delta}\Delta_{k,k+1}}\|y^k-y^{k-1}\|^{\frac{1}{2}}.$$

利用不等式$2\sqrt{\alpha\beta}\leqslant\alpha+\beta(\alpha,\beta>0)$可得

$$2\|y^{k+1}-y^k\|\leqslant\|y^k-y^{k-1}\|+\frac{\zeta}{\delta}\Delta_{k,k+1},$$

由上式可得

$$2\sum_{k=\tilde{k}+1}^{m}\|y^{k+1}-y^k\|\leqslant\sum_{k=\tilde{k}+1}^{m}\|y^k-y^{k-1}\|+\frac{\zeta}{\delta}\Delta_{\tilde{k}+1,m+1}.$$

注意到$\varphi(L_\beta(w^{m+1})-L_\beta(w^*))>0$,移项并且令$m\to+\infty$,可得

$$\sum_{k=\tilde{k}+1}^{+\infty}\|y^{k+1}-y^k\|\leqslant\|y^{\tilde{k}+1}-y^{\tilde{k}}\|+\frac{\zeta}{\delta}\varphi(L_\beta(w^{\tilde{k}+1})-L_\beta(w^*)),$$

所以

$$\sum_{k=0}^{+\infty}\|y^{k+1}-y^k\|<+\infty.$$

由式(2.7)可得

$$\sum_{k=0}^{+\infty}\|\lambda^{k+1}-\lambda^k\|<+\infty.$$

另一方面,由式(2.9)、式(2.10)可得

$$\|x^{k+1}-x^k\|\leqslant\sqrt{\frac{3P}{\beta^2}}(\|\lambda^{k+1}-\lambda^k\|^2+\|\lambda^k-\lambda^{k-1}\|^2+\beta^2\|B(y^{k+1}-y^k)\|^2)^{\frac{1}{2}}\leqslant\sqrt{\frac{3P}{\beta^2}}(\|\lambda^{k+1}-\lambda^k\|+\|\lambda^k-\lambda^{k-1}\|+\beta\|B(y^{k+1}-y^k)\|),$$

故$\sum_{k=1}^{+\infty}\|x^{k+1}-x^k\|<+\infty$.又因为

$$\|w^{k+1}-w^k\|=(\|x^{k+1}-x^k\|^2+\|y^{k+1}-y^k\|^2+\|\lambda^{k+1}-\lambda^k\|^2)^{\frac{1}{2}}\leqslant\|x^{k+1}-x^k\|+\|y^{k+1}-y^k\|+\|\lambda^{k+1}-\lambda^k\|,$$

所以

$$\sum_{k=0}^{+\infty}\|w^{k+1}-w^k\|<+\infty.$$

进一步可知序列w^k是Cauchy序列(参见文献[11]),所以序列$\{w^k\}$收敛,定理得证.

3 结论

本文针对非凸两分块优化问题,在不要求矩阵A列满秩,B不一定是单位阵,在$L_\beta(w)$满足Kurdyka-Lojasiewicz性质且罚参数大于某个常数的条件下,证明了非凸问题ADMM的收敛性.

参考文献:

[1] GLOWINSKI R,MARROCO A. Sur l'approximation,

par elements finis d'ordre un, et la resolution, par penalisation-dualité, d'une classe de problèms de Dirichlet nonlineares[J]. Revue Francaise d'Automatique, Informatique et Recherche Opérationelle, Series R, 1975, 31 (5/6): 41-76.
[2] GABAY D, MERCIER B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation[J]. Computers & Mathematics with Applications, 1976, 2(1): 17-40.
[3] BOYD S, PARIKH N, CHU E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. Foundations and Trends in Machine Learning, 2011, 3(1): 1-122.
[4] YANG L, PONG T K, CHEN X J. Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction[J]. Mathematics, 2016.
[5] HONG M Y, LUO Z Q, RAZAVIYAYN M. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems[J]. SIAM Journal on Optimization, 2014, 26(1): 337-364.
[6] WANG Y, YIN W T, ZENG J S. Global Convergence of ADMM in Nonconvex Nonsmooth Optimization[R]. UCLA CAM Report 15-62, 2015.
[7] LI G Y, PONG T K. Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems[J]. Mathematical Programming, 2016, 159(1/2): 374-401.
[8] WANG F H, XU Z B, XU H K. Convergence of multiblock Bregman ADMM for nonconvex composite problems[J]. Mathematics, 2015, arXiv: 1505.03063vl: 1-25.
[9] GUO K, HAND R, WUT T. Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints[J]. International Journal of Computer Mathematics, 2016: 1-17.
[10] BOLTE J, DANIILIDIS A, LEY O, et al. Characterizations of Lojasiewicz inequalities and applications[J]. arXiv: 0802.0826vl, 2008: 1-48.
[11] ATTOUCH H, BOLTE J, REDONT P, et al. A Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality[J]. Mathematics of Operations Research, 2010, 35(2): 438-457.
[12] BOLTE J, SABACH S, TEBOULLE M. Proximal alternating linearized minimization for nonconvex and nonsmooth problem[J]. Mathematical Programming, 2013, 146(1/2): 459-494.

(责任编辑:米慧芝)

(上接第 421 页 Continue from page 421)

[14] JIANG X Z, JIAN J B. A sufficient descent Dai-Yuan type nonlinear conjugate gradient method for unconstrained optimization problems[J]. Nonlinear Dynamics, 2013, 72(1/2): 101-112.
[15] 江羡珍,简金宝,马国栋.具有充分下降性的两个共轭梯度法[J].数学学报,2014,57(2):365-372.
JIANG X Z, JIAN J B, MA G D. Two conjugate gradient methods with sufficient descent property[J]. Acta Mathematica Sinica, 2014, 57(2): 365-372.
[16] 马国栋,简金宝,江羡珍.一个具有下降性的改进 Fletcher-Reeves 共轭梯度法[J].应用数学学报,2015,38(1):89-97.
MA G D, JIAN J B, JIANG X Z. An improved Fletcher-Reeves conjugate gradient method with descent property[J]. Acta Mathematicae Applicatae Sinica, 2015, 38 (1): 89-97.
[17] ZHANG B X, ZHU Z B, LI S A. A modified spectral conjugate gradient projection algorithm for total variation image restoration[J]. Applied Mathematics Letters, 2014, 27: 26-35.
[18] ZHANG B X, ZHU Z B, WANG S. A simple primal-dual method for total variation image restoration[J]. Journal of Visual Communication and Image Representation, 2016, 38: 814-823.
[19] ZHANG B X, ZHU Z B. A modified quasi-Newton diagonal update algorithm for total variation denoising problems and nonlinear monotone equations with applications in compressive sensing[J]. Numerical Linear Algebra with Applications, 2015, 22(3): 500-522.
[20] 曹莉,王伟.求解非线性不适定问题的连续 Landweber 型正则化方法[J].黑龙江大学自然科学学报,2014,31 (2):165-170.
CAO L, WANG W. Continuous Landweber type regularization method for solving nonlinear ill-posed problem[J]. Journal of Natural Science of Heilongjiang University, 2014, 31(2): 165-170.
[21] VOGEL C R. Computational Methods for Inverse Problems[M]. Philadelphia: Society for Industrial and Applied Mathematics, 2011: 1-11.

(责任编辑:尹 闯)