

一种基于M范数惩罚项的线性化ADMM算法

王博冉

中央民族大学理学院, 北京

收稿日期: 2024年3月25日; 录用日期: 2024年4月22日; 发布日期: 2024年4月29日

摘要

ADMM算法是求解两块可分凸优化问题的经典算法, 主要思想是在增广拉格朗日乘子法的基础上, 利用目标函数关于两块变量的可分性, 降低了求解子问题的计算难度。当增广拉格朗日函数中的惩罚项是M范数时, 求解子问题往往较为困难。因此, 我们在增广拉格朗日函数的基础上, 通过增加一个半正定或正定的临近项, 将M范数的惩罚项变为2范数的惩罚项, 这样, 就可以很快得到子问题的闭形式解。该方法同时具备弱化的惩罚项的条件和半正定临近项的优势, 具有更广的适用性和更高的求解效率。这种改进的新算法可以看成临近点算法, 它的收敛性易于分析, 且无需要较强的假设条件。实验结果表明, 新算法和其他几种主流的高效算法相比, 新算法是可行的。

关键词

变分不等式, 交替方向乘子法, 临近点算法, 全局收敛性

A Linearized ADMM Algorithm Based on M-Norm Penalty Terms

Boran Wang

College of Science, Minzu University of China, Beijing

Received: Mar. 25th, 2024; accepted: Apr. 22nd, 2024; published: Apr. 29th, 2024

Abstract

The ADMM algorithm is a conventional strategy for solving two separable convex optimization problems. Its fundamental idea is to use the objective function on the basis of augmented Lagrangian multiplier method, reducing the computing burden of addressing subproblems. When the penalty term in the augmented Lagrangian function is M -norm, it is generally more difficult to solve subproblems. As a consequence of augmented Lagrange function, we shift the penalty term of the M -norm to the penalty term of the 2-norm by adding a semipositive definite or positive definite proximity

term, allowing us to rapidly find the closed form solution of the subproblem. This approach has the advantages of a weaker penalty term and a semipositive definite proximity term, which allows for a broader range of applications. This improved new algorithm can be viewed as a proximity point algorithm, which is easy to analyze in terms of convergence and does not require strong assumptions. Experimental results show that the new algorithm is feasible compared to several other mainstream efficient algorithms.

Keywords

Variational Inequality, Alternating Direction Method of Multipliers, Proximity Point Algorithm, Global Convergence

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

本文考虑如下具有线性约束的可分凸优化问题:

$$\min \{f(x) + g(y) \mid Ax + By = b, x \in X, y \in Y\} \quad (1)$$

其中 $f: R^{n_1} \rightarrow R$ 和 $g: R^{n_2} \rightarrow R$ 是凸函数(但不一定光滑), $A \in R^{m \times n_1}, B \in R^{m \times n_2}, b \in R^m, X \subseteq R^{n_1}$ 和 $Y \subseteq R^{n_2}$ 是闭凸集。

众所周知, 交替方向乘法是求解(1)的有效方法, 该方法最早由 Gabay 和 Mercier, Glowinski [1] 以及 Marrocco [2] 提出, 该方法是一种求解具有可分离的凸优化问题的重要方法, 由于其收敛速度快、收敛性能好, 在求解可分离凸优化问题上具有简单、灵活、实用性强的特点, 可以将大规模问题拆分成两个甚至多个小规模子问题, 随后交替求解各个小规模子问题, 从而提高了求解的速率, 其优势在于利用对偶上升算法的可分离性, 后来被广泛研究与应用。

为了提高 ADMM 的适用性, 一些学者提出了许多改进的 ADMM 方法。He 等人在文献[3]中提出了不定临近线性化 ADMM。Gao [4]提出了如下一个带不定临近正则项的线性化 ADMM。Fang [5]将增广 Lagrange 函数中的惩罚项采用 M -范数, M 是一个是对称正定矩阵。将 2-范数形式改进为 M -范数形式, 弱化了惩罚项的条件。因为 2-范数形式的惩罚项是矩阵范数 M -范数的一种特殊形式, 所以 M -范数的使用范围更广。

基于上述讨论, 本文结合半正定临近项和基于 M -范数的惩罚项, 提出了一种求解可分凸优化问题(1)的基于 M -范数惩罚项的带半正定临近项的线性化 ADMM 算法。新方法在 y -子问题中引入了一个半正定临近项, 进而将 M -范数惩罚项转化为 2 范数惩罚项。该方法同时具备弱化的惩罚项的条件和半正定临近项的优势, 具有更广的适用性和更高的求解效率。针对所提出的算法, 本文基于变分不等式和最优化解理论给出了严格的收敛性分析以及收敛速率分析, 并通过数值实验验证了算法的有效性。

2. 一种基于 M 范数惩罚项的 ADMM 算法

假设 $f: R^n \rightarrow R \cup +\infty$ 和 $g: R^n \rightarrow R \cup +\infty$ 是合适的、闭的、凸函数。

本文提出的算法如下:

算法 1. 一种基于 M 范数惩罚项的 ADMM 算法

输入: x^0, y^0, λ^0
 计算: $x^{k+1}, y^{k+1}, \lambda^{k+1}$

$$\begin{aligned} x^{k+1} &= \arg \min_x L_\beta(x, y^k, \lambda^k) \\ y^{k+1} &= \arg \min_y L_\beta(x^{k+1}, y, \lambda^k) + 1/2 \|y - y^k\|_D^2 \\ \lambda^{k+1} &= \lambda^k - \beta M(Ax^{k+1} + By^{k+1} - b) \end{aligned}$$

若满足停止准则, 则输出: $x^{k+1}, y^{k+1}, \lambda^{k+1}$, 其中
 $L_\beta(x, y, \lambda) = f(x) + g(y) - \lambda^T(Ax + By - b) + \beta/2 \|Ax + By - b\|_M^2$, M 是一个正定矩阵,
 $D = sI - \beta B^T M B$ ($s \geq \|\beta B^T M B\|$) 是一个半正定矩阵。

3. 收敛性分析

引理 3.1. 设序列 w^k 由算法 1 迭代产生, 令 $u = \begin{pmatrix} x \\ y \end{pmatrix}$, $\theta(u) = f(x) + g(y)$, $w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}$ 。

那么有 $w^{k+1} \in \Omega$,

$$\theta(u) - \theta(u^{k+1}) + (w - w^{k+1}) \left\{ \begin{aligned} &F(w^{k+1}) + \beta \begin{pmatrix} A^T \\ B^T \\ O \end{pmatrix} MB(y^k - y^{k+1}) \\ &+ \begin{pmatrix} \beta B^T M B + D & O \\ O & \frac{1}{\beta} M^{-1} \end{pmatrix} \begin{pmatrix} y^{k+1} - y^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} \end{aligned} \right\} \geq 0 \quad (2)$$

其中 $F(w^{k+1}) = \begin{pmatrix} -A^T \lambda^{k+1} \\ -B^T \lambda^{k+1} \\ Ax^{k+1} + By^{k+1} - b \end{pmatrix}$ 。

引理 3.2. 设序列 w^k 由算法 1 迭代产生, 则

$$w^{k+1} \in \Omega,$$

$$\theta(u) - \theta(u^{k+1}) + (w - w^{k+1}) \left\{ F(w) + \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} MB(y^k - y^{k+1}) \right\} \geq (v - v^{k+1}) G(v - v^{k+1}), \forall w \in \Omega \quad (3)$$

其中 $G = \begin{pmatrix} \beta B^T M B + D & O \\ O & \frac{1}{\beta} M^{-1} \end{pmatrix}$, $v = \begin{pmatrix} y \\ \lambda \end{pmatrix}$ 。

证明: 因为 $(w - w^{k+1})^T (F(w) - F(w^{k+1})) \geq 0, \forall w \in \Omega$, 所以 F 是单调算子。

由(2)可得, (3)成立。

引理 3.3. 设序列 w^k 由算法 1 迭代产生, 其中 $w^* \in \Omega^*$, Ω^* 为变分不等式的解集, 那么有

$$(v^{k+1} - v^*)^T G(v^k - v^{k+1}) \geq (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \quad (4)$$

证明:

$$(v^{k+1} - v^*)^T G(v^k - v^{k+1})$$

$$\begin{aligned}
 &\geq \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) + \beta \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} MB(y^k - y^{k+1}) \\
 &\geq \beta \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix} \begin{pmatrix} A^T \\ B^T \end{pmatrix} MB(y^k - y^{k+1}) \\
 &= \beta \left[(Ax^{k+1} - Ax^*) + (By^{k+1} - By^*) \right]^T B(y^k - y^{k+1}) = (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1})
 \end{aligned}$$

引理 3.4. 设序列 w^k 由算法 1 迭代产生, 那么有:

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq \frac{1}{2} \|y^k - y^{k+1}\|_D^2 - \frac{1}{2} \|y^{k-1} - y^k\|_D^2 \tag{5}$$

证明: 对于算法 1, 由变分不等式可得:

$$\begin{aligned}
 y^{k+1} \in Y, g(y^k) - g(y^{k+1}) + (y^k - y^{k+1})^T - B^T \lambda^{k+1} + D(y^{k+1} - y^k) &\geq 0, \forall y^k \in Y \\
 y^k \in Y, g(y^{k+1}) - g(y^k) + (y^{k+1} - y^k)^T - B^T \lambda^k + D(y^k - y^{k-1}) &\geq 0, \forall y^{k+1} \in Y
 \end{aligned}$$

两式相加可得: $(y^k - y^{k+1})^T B^T (\lambda^k - \lambda^{k+1}) + D[(y^{k+1} - y^k) - (y^k - y^{k-1})] \geq 0$
 所以 $(y^k - y^{k+1})^T B^T (\lambda^k - \lambda^{k+1}) \geq (y^k - y^{k+1})^T D[(y^k - y^{k+1}) - (y^{k-1} - y^k)]$ 。

由柯西不等式可得:

$$(y^k - y^{k+1})^T D[(y^k - y^{k+1}) - (y^{k-1} - y^k)] \geq \frac{1}{2} \|y^k - y^{k+1}\|_D^2 - \frac{1}{2} \|y^{k-1} - y^k\|_D^2$$

故(5)得证。

引理 3.5. 设序列 w^k 由算法 1 迭代产生, 那么有:

$$(v^{k+1} - v^*)^T G(v^k - v^{k+1}) \geq \frac{1}{2} \|y^k - y^{k+1}\|_D^2 - \frac{1}{2} \|y^{k-1} - y^k\|_D^2 \tag{6}$$

证明:

$$\begin{aligned}
 \|v^k - v^*\|_G^2 &= \|(v^{k+1} - v^*) + (v^k - v^{k+1})\|_G^2 \\
 &\geq \|v^{k+1} - v^*\|_G^2 + \|v^k - v^{k+1}\|_G^2 + 2(v^{k+1} - v^*)^T G(v^k - v^{k+1}) \\
 &\geq \|v^{k+1} - v^*\|_G^2 + \|v^k - v^{k+1}\|_G^2 + \|y^k - y^{k+1}\|_D^2 - \|y^{k-1} - y^k\|_D^2
 \end{aligned}$$

显然, (6)式成立。

定理 3.1. 对于任意的 $w^* \in \Omega^*$, 序列 $\|v^{k+1} - v^*\|_G^2 + \|y^k - y^{k+1}\|_D^2$ 是单调递减的。

证明: 由引理 3.5 可得:

$$\|v^k - v^{k+1}\|_G^2 \leq (\|v^k - v^*\|_G^2 + \|y^{k-1} - y^k\|_D^2) - (\|v^{k+1} - v^*\|_G^2 + \|y^k - y^{k+1}\|_D^2)$$

因为 $\|v^k - v^{k+1}\|_G^2 \geq 0$, 所以 $\|v^{k+1} - v^*\|_G^2 + \|y^k - y^{k+1}\|_D^2 \leq \|v^k - v^*\|_G^2 + \|y^{k-1} - y^k\|_D^2$ 。

故序列 $\|v^{k+1} - v^*\|_G^2 + \|y^k - y^{k+1}\|_D^2$ 是单调递减的。

定理 3.2. 设序列 w^k 由算法 1 迭代产生, 那么序列 v^{k_j} 收敛到 v^∞ 。

证明: 因为 $\|v^k - v^{k+1}\|_G^2 \leq (\|v^k - v^*\|_G^2 + \|y^{k-1} - y^k\|_D^2) - (\|v^{k+1} - v^*\|_G^2 + \|y^k - y^{k+1}\|_D^2)$ 。

对上式求和可得: $\sum_{k=1}^\infty \|v^k - v^{k+1}\|_G^2 \leq \|v^1 - v^*\|_G^2 + \|y^0 - y^1\|_D^2$ 。

因为 G 是正定的, 所以 $\sum_{k=1}^{\infty} \|v^k - v^{k+1}\|_G^2$ 是正项级数。

由正项级数收敛的必要条件可知: $\lim_{k \rightarrow \infty} \|v^k - v^{k+1}\|_G^2 = 0$ 。

因为 $0 \leq \|v^k - v^{k+1}\|_G^2 \leq (\|v^k - v^*\|_G^2 + \|y^{k-1} - y^k\|_D^2) - (\|v^{k+1} - v^*\|_G^2 + \|y^k - y^{k+1}\|_D^2)$ 。

所以 $\|v^{k+1} - v^*\|_G^2 \leq \|v^k - v^*\|_G^2 + \|y^{k-1} - y^k\|_D^2 \leq \|v^1 - v^*\|_G^2 + \|y^0 - y^1\|_D^2, \forall k \geq 1$ 。

所以序列 v^k 是有界的并且至少存在一个聚点 v^∞ , 存在一个子列 v^{k_j} 收敛到 v^∞ 。

即 $\lim_{k \rightarrow \infty} v^{k_j} = v^\infty$, 其中 $v^\infty = \begin{pmatrix} y^\infty \\ \lambda^\infty \end{pmatrix} \in V^*$ 。

定理 3.3. w^∞ 是(1)的一个解。

证明: 因为 $\lim_{k \rightarrow \infty} \|v^k - v^{k+1}\|_G^2 = 0$, 所以 $\lim_{k \rightarrow \infty} (y^k - y^{k+1}) = 0$ 和 $\lim_{k \rightarrow \infty} (\lambda^k - \lambda^{k+1}) = 0$ 。

由变分不等式可知:

$$w^{k+1} \in \Omega, \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T \left\{ \begin{array}{l} F(w^{k+1}) + \beta \begin{pmatrix} A^T \\ B^T \\ O \end{pmatrix} MB(y^k - y^{k+1}) \\ + \begin{pmatrix} \beta B^T MB + D & O \\ O & \frac{1}{\beta} M^{-1} \end{pmatrix} \begin{pmatrix} y^{k+1} - y^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} \end{array} \right\} \geq 0, \forall w \in \Omega$$

令上式中 $k \rightarrow \infty$, 可得: $\theta(u) - \theta(u^\infty) + (w - w^\infty)^T F(w^\infty) \geq 0$, 故 w^∞ 是(1)的解。

4. 数值实验

统计中的 LASSO 问题如下:

$$\min_{x, y} \frac{1}{2} \|Ax - b\|_2^2 + \mu \|y\|_1, \text{ s.t. } x = y \tag{7}$$

其中 $A \in R^{m \times n}, b \in R^m, x \in R^n, y \in R^n$ 。

使用算法 1 求解(7)得到 x 子问题: $x^{k+1} = \frac{1}{\delta} [A^T b + \lambda^k + (\delta I - A^T A) y^k]$ 。

并且 y 子问题的闭式解是: $y^{k+1} = \text{shrink} \left(y^k - \frac{1}{\delta} [\lambda^k + M(y^k - x^{k+1})], \frac{\mu}{\beta \delta} \right)$ 。

λ 子问题的解为: $\lambda^{k+1} = \lambda^k - \beta M(x^{k+1} - y^{k+1})$, 其中 $M = \delta I - A^T A, D = A^T A$ 。

本文通过数值实验比较了算法 1 和线性化 DMM 算法的迭代次数。

将每个算法的参数设置如下:

算法 1: $\delta = \|A^T A\| + 0.001, \beta = 1, \mu = 0.1 \|A^T b\|_\infty$ 。

Linearized ADMM: $\delta = \|A^T A\| + 0.001, \beta = 1, \mu = 0.1 \|A^T b\|_\infty, D = \delta I - A^T A$ 。

停止准则是: $\|x^{k+1} - y^{k+1}\| < \varepsilon^{pri}, \|\beta(y^{k+1} - y^k)\| < \varepsilon^{dual}$, 其中 $\varepsilon^{pri} = \sqrt{n} \varepsilon^{abs} + \varepsilon^{rel} \max(\|x^{k+1}\|, \|y^{k+1}\|), \varepsilon^{dual} = \sqrt{n} \varepsilon^{abs} + \varepsilon^{rel} \|y^{k+1}\|$ 。

ε^{pri} 和 ε^{dual} 被设置为 10^{-4} 和 10^{-2} 。

对于给定的维度 $m \times n$, 我们随机生成数据, 如下所示:

$$p = \frac{1}{n}, x^0 = \text{sprandn}(n, 1, p), A = \text{randn}(m, n), b = A * x^0 + \text{sqrt}(0.001) * \text{randn}(m, 1)$$

二种算法的实验结果如表 1 所示:

Table 1. Comparison of the number of iterations between Algorithm 1 and Linearized ADMM

表 1. 算法 1 和 Linearized ADMM 的迭代次数比较

矩阵维数($m \times n$)	算法 1 的迭代次数	Linearized ADMM 的迭代次数
750 × 2500	40	53
900 × 3000	39	52
1050 × 3500	38	51
1200 × 4000	38	51
1500 × 5000	35	48

实验结果表明, 算法 1 的迭代次数少。因此, 算法 1 是有可比较性的。

5. 结论

在本文中, 我们提出了一种基于 M 范数罚项的具有半正定邻近项的广义线性化 ADMM 算法, 对其进行了全局收敛性分析。最后, 通过数值实验验证了算法的有效性。

参考文献

- [1] Gabay, D. and Mercier, B. (1976) A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximation. *Computers & Mathematics with Applications*, **2**, 17-40. [https://doi.org/10.1016/0898-1221\(76\)90003-1](https://doi.org/10.1016/0898-1221(76)90003-1)
- [2] Glowinski, R. and Marrocco, A. (1974) Analyse numérique du champ magnétique d'un alternateur par éléments finis et sur-relaxation ponctuelle non linéaire. *Computer Methods in Applied Mechanics and Engineering*, **3**, 55-85. [https://doi.org/10.1016/0045-7825\(74\)90042-5](https://doi.org/10.1016/0045-7825(74)90042-5)
- [3] He, B., Ma, F. and Yuan, X. (2020) Optimally Linearizing the Alternating Direction Method of Multipliers for Convex Programming. *Computational Optimization and Applications*, **75**, 361-388. <https://doi.org/10.1007/s10589-019-00152-3>
- [4] Gao, B. and Ma, F. (2016) Symmetric ADMM with Positive-Indefinite Proximal Regularization for Linearly Constrained Convex Optimization. <http://www.optimization-online.org>
- [5] Fang, C., Chen, J. and Chen, S. (2022) Image Denoising Algorithm of Compressed Sensing Based on Alternating Direction Method of Multipliers. *International Journal of Modeling, Simulation, and Scientific Computing*, **13**, 2250009. <https://doi.org/10.1142/S179396232250009X>