

正则化Consensus问题的收敛性证明

刘玉洁, 毕文静, 张炎, 俞露, 李伟南

巢湖学院数学与统计学院, 安徽 合肥
Email: 2971303271@qq.com, 1294543746@qq.com

收稿日期: 2021年2月18日; 录用日期: 2021年3月19日; 发布日期: 2021年3月30日

摘要

交替方向乘子法(ADMM算法)是求解可分离凸优化问题的一种有效方法。该算法利用目标函数的可分性, 将原问题拆分成若干个极小化的子问题, 然后交替迭代求解。而一致性(Consensus)问题是求解大数据问题的重要的一种形式, 本文提出了一种正则化的一致性问题的迭代过程, 并在适当的假设下, 证明了其收敛性。

关键词

ADMM, Consensus, 收敛性

Convergence Proof of Regularized Consensus Problems

Yujie Liu, Wenjing Bi, Yan Zhang, Lu Yu, Weinan Li

College of Mathematics and Statistics, Chaohu University, Hefei Anhui
Email: 2971303271@qq.com, 1294543746@qq.com

Received: Feb. 18th, 2021; accepted: Mar. 19th, 2021; published: Mar. 30th, 2021

Abstract

Alternating direction multiplier method (ADMM algorithm) is an effective method to solve separable convex optimization problems. The algorithm USES the separability of the objective function to divide the original problem into several minimization subproblems and then solve them alternately iteratively. Consensus is an important form of solving big data problems. In this paper, a regularized consistency problem is proposed, its iterative process is given, and its convergence is proved under appropriate assumptions.

Keywords

ADMM, Consensus, The Convergence

Copyright © 2021 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

乘子交替方向法(ADMM)是一种求解线性约束凸优化问题的高效算法, 由 Glowinski 等[1]及 Gabay 等[2]分别于 1975 年和 1976 年提出。该算法首先构造原问题的增广拉格朗日函数, 每个迭代步由多个子问题构成, 每个子问题对于增广拉格朗日函数关于 1 块变量进行极小化, 当所有子问题求解结束后更新乘子(对偶变量)。该算法收敛速度较快, 且子问题规模较小、计算量较低, 因而该算法的计算效率较高, 已被广泛应用于求解信息科学、统计学、机器学习等领域的多种优化问题, 如矩阵完整化问题、图像去模糊去噪问题、协方差矩阵校正问题等[3] [4] [5] [6] [7]。斯坦福大学 Boyd 等[6]于 2011 年对 ADMM 进行了综述, 并指出该算法适合求解大规模分布式优化问题。

所谓全局变量一致性优化问题, 即目标函数根据数据分解成 N 子目标函数(子系统), 每个子系统和子数据都可以获得一个参数解, 但是全局解只有一个 z , 于是就可以写成如下优化命题:

$$\begin{aligned} \min \quad & \sum_{i=1}^N f_i(x_i) \\ \text{s.t.} \quad & x_i - z = 0, i = 1, \dots, N \end{aligned}$$

注意, 此时仍是凸函数, 而并不是对参数空间进行划分, 这里是对数据而言, 所以维度一样, 与之前的问题并不太一样。这种问题其实就是所谓的并行化处理, 或分布式处理, 希望从多个分块的数据集中获取相同的全局参数解。

在 ADMM 算法框架下(先返回最初从扩增 lagrangian 导出的 ADMM), 这种问题解法相当明确:

$$\begin{aligned} L_\rho(x_1, \dots, x_N, z, y) &= \sum_{i=1}^N (f_i(x_i) + y_i^T(x_i - z) + (\rho/2)\|x_i - z\|_2^2) \\ \text{s.t.} \quad C &= \{(x_1, \dots, x_N) | x_1 = \dots = x_N\} \end{aligned}$$

从而迭代结果为

$$\begin{aligned} x_i^{k+1} &= \arg \min_{x_i} (f_i(x_i) + y_i^{kT}(x_i - z^k) + (\rho/2)\|x_i - z^k\|_2^2) \\ z^{k+1} &= \frac{1}{N} \sum_{i=1}^N (x_i^{k+1} + (1/\rho)y_i^k) \\ y_i^{k+1} &= y_i^k + \rho(x_i^{k+1} - z^{k+1}) \end{aligned}$$

对 y-update 和 z-update 的和分别求个平均, 易得, 于是可以知道 z-update 步其实可以简化为, 于是上述 ADMM 其实可以进一步化简为如下形式:

$$\begin{aligned} x_i^{k+1} &= \arg \min_{x_i} (f_i(x_i) + y_i^{kT}(x_i - \bar{x}^k) + (\rho/2)\|x_i - \bar{x}^k\|_2^2) \\ y_i^{k+1} &= y_i^k + \rho(x_i^{k+1} - \bar{x}^{k+1}) \end{aligned}$$

这种迭代算法写出来了，并行化那么就是轻而易举了，各个子数据分别并行求最小化，然后将各个子数据的解汇集起来求均值，整体更新对偶变量，然后再继续回带求最小值至收敛。当然也可以分布式部署(hadoop 化)，但是说起来容易，真正工程实施起来又是另外一回事，各个子节点机器间的通信更新是一个需要细细揣摩的问题。

另外，对于全局一致性优化，也需要给出相应的终止迭代准则，与一般的 ADMM 类似，这里的 primal 和 dual 的 residuals 为

$$r^k = (x_1^k - \bar{x}^k, \dots, x_N^k - \bar{x}^k), \quad s^k = -\rho(\bar{x}^k - \bar{x}^{k-1}, \dots, \bar{x}^k - \bar{x}^{k-1})$$

从而 2-norm 为

$$\|r^k\|_2^2 = \sum_{i=1}^N \|x_i - \bar{x}^k\|_2^2, \quad \|s^k\|_2^2 = N\rho^2 \|\bar{x}^k - \bar{x}^{k-1}\|_2^2$$

本文主要考虑正则化的 Consensus 问题，即在目标函数后面加个正则项 $g(z)$ ，并且这个正则项存在偏导数。证明提出的正则化的一致性算法的收敛性，给出了详细证明。

2. 正则化的 Consensus 问题

下面就是要将之前所谈到的经典的机器学习算法并行化起来。想法很简单，就是对全局变量加上正则项即可，

$$\begin{aligned} \min \quad & \sum_{i=1}^N f_i(x_i) + g(z) \\ \text{s.t.} \quad & x_i - z = 0, i = 1, \dots, N \end{aligned}$$

首先构建非增广的 Lagrangian L_0 有

$$L_0(x_1, \dots, x_N, z, y) = \sum_{i=1}^N (f_i(x_i) + y_i^T(x_i - z)) + g(z),$$

从而增广的 Lagrangian L_ρ 有

$$L_\rho(x_1, \dots, x_N, z, y) = \sum_{i=1}^N (f_i(x_i) + y_i^T(x_i - z) + (\rho/2)\|x_i - z\|_2^2) + g(z).$$

因此 ADMM 算法只需要改变下 z-update 步即可

$$\begin{aligned} x_i^{k+1} &= \arg \min_{x_i} (f_i(x_i) + y_i^{kT}(x_i - z^k) + (\rho/2)\|x_i - z^k\|_2^2) \\ z^{k+1} &= \arg \min_z (g(z) + \sum_{i=1}^N -y_i^{kT}z + (\rho/2)\|x_i^{k+1} - z\|_2^2) \\ y_i^{k+1} &= y_i^k + \rho(x_i^{k+1} - z^{k+1}) \end{aligned}$$

同样的，我们仍对 z 做一个平均处理，于是就有

$$z^{k+1} = \arg \min_z (g(z) + (N\rho/2)\|z - \bar{x}^{k+1} - (1/\rho)\bar{y}^k\|_2^2)$$

上述形式都取得是最原始的 ADMM 形式，简化处理，写成 scaled 形式即有

$$\begin{aligned} x_i^{k+1} &= \arg \min_{x_i} (f_i(x_i) + (\rho/2)\|x_i - z^k + u_i^k\|_2^2) \\ z^{k+1} &= \arg \min_z (g(z) + (N\rho/2)\|z - \bar{x}^{k+1} - \bar{u}^k\|_2^2) \\ u_i^{k+1} &= u_i^k + x_i^{k+1} - z^{k+1} \end{aligned}$$

这样对于后续处理问题就清晰明了多了。可以看到如果 $g(z) = \lambda \|z\|_1$ ，即 lasso 问题，那么 z-update 步就用软阈值 operator 即

$$z^{k+1} = S_{\lambda/N\rho}(\bar{x}^{k+1} - (1/\rho)\bar{y}^k)$$

因此，对于大规模数据，要想用 lasso 等算法，只需要对数据做切块(切块也最好切均匀点)，纳入到全局变量一致性的 ADMM 框架中，即可并行化处理。

3. 项收敛性证明

本部分主要说明上述算法的收敛性，在证明之前首先给出一些符号说明及条件假设。

符号说明： 令 (x_1, \dots, x_N, y, z) 为 (x, y, z) ， $f(x) = \sum_{i=1}^n f(x_i)$ 。

假设 1： 函数 f_i, g 是偏导数存在的凸函数。

假设 2： 非增广的 Lagrangian L_0 存在鞍点，即存在 (x^*, y^*, z^*) ，对任意的 (x, y, z) 使得：

$$L_0(x^*, y, z^*) \leq L_0(x^*, y^*, z^*) \leq L_0(x, y^*, z)$$

定理： 在满足上述假设条件 1, 2 和上述算法的迭代过，当 $k \rightarrow \infty$ ，

$$r^k \rightarrow 0, z^{k+1} - z^k \rightarrow 0$$

从而我们有

$$\lim_{k \rightarrow \infty} p^k = p^*$$

证明：

部分 1

因为 (x^*, y^*, z^*) 是函数 L_0 的鞍点，所以有：

$$L_0(x^*, z^*, y^*) \leq L_0(x^{k+1}, z^{k+1}, y^*) \quad (1)$$

再利用 $x^* - z^* = 0$ ，令

$$p^{k+1} = f(x^{k+1}) + g(z^{k+1}), \quad p^* = f(x^*) + g(z^*)$$

从而(1)式有

$$p^* \leq p^{k+1} + y^{*T} r^{k+1} \quad (2)$$

部分 2

通过定义我们知道， x^{k+1} 是 $L_\rho(x, z^k, y^k)$ 的极小值。由于 f 是可偏导的凸函数。所以：

$$0 \in \partial L_\rho(x^{k+1}, z^k, y^k) = \partial f(x^{k+1}) + y^k + \rho(x^{k+1} - z^k)$$

又因为 $y^{k+1} = y^k + \rho r^{k+1}$ ，所以有 $y^k = y^{k+1} - \rho r^{k+1}$ ，从而

$$0 \in \partial L_\rho(x^{k+1}, z^k, y^k) = \partial f(x^{k+1}) + y^{k+1} - \rho(z^{k+1} - z^k)$$

这里表明 x^{k+1} 是下面这个函数的极小值

$$f(x) + y^{k+1} - \rho(z^{k+1} - z^k)x \quad (4)$$

与上面相似的证法可以说明 z^{k+1} 是函数 $g(z) + y^{(k+1)T} z$ 的极小值。结合(4)式有

$$f(x^{k+1}) + y^{k+1} - \rho(z^{k+1} - z^k)x^{k+1} \leq f(x^*) + y^{k+1} - \rho(z^{k+1} - z^k)x^* \quad (5)$$

$$g(z^{k+1}) + y^{(k+1)T} z^{k+1} \leq g(z^*) + y^{(k+1)T} z^* \quad (6)$$

再结合 $x^* - z^* = 0$ ，结合(5)和(6)式有

$$p^{k+1} - p^* \leq -(y^{k+1})^T r^{k+1} - \rho(z^{k+1} - z^k)^T (-r^{k+1} + z^{k+1} - z^*) \quad (7)$$

证明 3

结合(2)和(7)有

$$2(y^{k+1} - y^*)^T r^{k+1} - 2\rho(z^{k+1} - z^k)^T r^{k+1} + 2\rho(z^{k+1} - z^k)^T (z^{k+1} - z^*) \leq 0 \quad (8)$$

对于(8)式中的第一部分利用 $y^{k+1} = y^k + \rho r^{k+1}$ 有

$$2(y^k - y^*)^T r^{k+1} + \rho \|r^{k+1}\|_2^2 + \rho \|r^{k+1}\|_2^2 \quad (9)$$

再利用 $r^{k+1} = (1/\rho)(y^{k+1} - y^k)$ ，因此(9)式中的前两部分有

$$(2/\rho)(y^k - y^*)^T (y^{k+1} - y^k) + (1/\rho) \|y^{k+1} - y^k\|_2^2 + \rho \|r^{k+1}\|_2^2$$

又因为 $y^{k+1} - y^k = (y^{k+1} - y^*) - (y^k - y^*)$ ，从而上式变为

$$(1/\rho) (\|y^{k+1} - y^*\|_2^2 - \|y^k - y^*\|_2^2) + \rho \|r^{k+1}\|_2^2 \quad (10)$$

接下来，整理(8)和(9)式中剩下的部分，即

$$\rho \|r^{k+1}\|_2^2 - 2\rho(z^{k+1} - z^k)^T r^{k+1} + 2\rho(z^{k+1} - z^k)^T (z^k - z^*)$$

再利用 $z^{k+1} - z^* = (z^{k+1} - z^k) - (z^k - z^*)$ ，有

$$\rho \|r^{k+1} - (z^{k+1} - z^k)\|_2^2 + \rho \|z^{k+1} - z^k\|_2^2 + 2\rho(z^{k+1} - z^k)^T (z^k - z^*)$$

再结合 $z^{k+1} - z^k = (z^{k+1} - z^*) - (z^k - z^*)$ ，从而有

$$\rho \|r^{k+1} - (z^{k+1} - z^k)\|_2^2 + \rho (\|z^{k+1} - z^*\|_2^2 - \|z^k - z^*\|_2^2)$$

令

$$U_k = (1/\rho) \|y^k - y^*\|_2^2 + \rho \|z^k - z^*\|_2^2$$

从而(8)可以写成

$$U_k - U_{k+1} \geq \rho \|r^{k+1} - (z^{k+1} - z^k)\|_2^2 \quad (11)$$

由证明 2 中可以看出， z^{k+1} 是函数 $g(z) + y^{(k+1)T} z$ 的极小值， z^k 是函数 $g(z) + y^{(k)T} z$ 的极小值，从而有

$$g(z^{k+1}) + y^{(k+1)T} z^{k+1} \leq g(z^k) + y^{(k+1)T} z^k$$

$$g(z^k) + y^{(k)T} z^k \leq g(z^{k+1}) + y^{(k)T} z^{k+1}$$

结合上面两个不等式，有

$$(y^{k+1} - y^k)^\top (z^{k+1} - z^k) \leq 0$$

又因为 $y^{k+1} - y^k = \rho r^{k+1}$ ，从而得到

$$\rho r^{k+1} (z^{k+1} - z^k) \leq 0 \quad (12)$$

结合(12)式，使得(11)有

$$U_k - U_{k+1} \geq \rho \|r^{k+1}\| + \rho \|z^{k+1} - z^k\|_2^2$$

从而有

$$\sum_{k=0}^{\infty} (\rho \|r^{k+1}\| + \rho \|z^{k+1} - z^k\|_2^2) \leq U_0 \quad (13)$$

从(13)式可以看出 $r^k \rightarrow 0, z^{k+1} - z^k \rightarrow 0, k \rightarrow \infty$ ，从而得到

$$\lim_{k \rightarrow \infty} p^k = p^*$$

4. 总结

通过正则化的一致性问题的研究，我们可以看到并行和分布式部署优化方案的可行性。我们可以切分数据以及相应的目标函数，也可以切分变量到各个子系统上去，分别作优化，甚至我们可以大胆想象对不同数据块用不同的优化算法，结合 Consensus 问题和 ADMM 算法，达到同一个全局变量的优化目的；并且在理论上说明了算法的收敛性。

致 谢

本文的撰写感谢程一元老师的指导。

基金项目

巢湖学院省级大学生创新创业训练计划资助项目(S201910380067)，巢湖学院省级大学生创新创业训练计划资助项目(S201910380065)，巢湖学院校级科研项目(XLY-201903)。

参考文献

- [1] Glowinski, R. and Marroco, A. (1975) Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet nonlinéaires. *Revue Française d'Automatique, Informatique, Recherche Operationnelle. Analyse Numerique*, **9**, 41-46. <https://doi.org/10.1051/m2an/197509R200411>
- [2] Gabay, D. and Mercier, B. (1976) A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximations. *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)
- [3] Lin, F., Jovanovic, M.R. and Georgiou, T.T. (2013) An ADMM Algorithm for Matrix Completion of Partially Known State Covariances. *Decision & Control*, **2013**, 1684-1689.
- [4] Wang, Y.L., Yang, J.F., Yin, W.T., et al. (2008) A New Alternating Minimization Algorithm for Total Variation Image Reconstruction. *SIAM Journal on Imaging Sciences*, **1**, 248-272. <https://doi.org/10.1137/080724265>
- [5] Shen, Y. and Wang, H.Y. (2016) New Augmented Lagrangian-Based Proximal Point Algorithm for Convex Optimization with Equality Constraints. *Journal Optimization Theory and Application*, **171**, 251-261. <https://doi.org/10.1007/s10957-016-0991-1>
- [6] Boyd, S., Parikh, N., Chu, E., et al. (2011) Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found Trends Mach Learning*, **3**, 1-122. <https://doi.org/10.1561/22000000016>
- [7] 申远, 孔玉倩, 纪磊. 一种新参数条件的线性化逐块交替方向乘法[J]. 徐州工程学院学报(自然科学版), 2020, 35(2): 23-32.