Journal of Computational and Applied Mathematics 309 (2017) 342–358

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

An algorithm twisted from generalized ADMM for multi-block separable convex minimization models

Jin Jiang Wang [a,b], Wen Song [a,c,*]

[a] Department of Mathematics, Harbin Institute of Technology, Harbin 150001, China

[b] College of Computer Science and Information Engineering, Harbin Normal University, Harbin 150025, China

[c] School of Mathematical Sciences, Harbin Normal University, Harbin 150025, China

ARTICLE INFO

Article history:
Received 28 October 2015
Received in revised form 22 January 2016

Keywords:
Separable convex optimization
ADMM
Proximal ADMM
Augmented Lagrangian method

ABSTRACT

The alternating direction method with multipliers (ADMM) has been one of most powerful and successful methods for solving a two-block linearly constrained convex minimization model whose objective function is the sum of two functions without coupled variables. It is known that the numerical efficiency is inherited for a large number of applications, but the convergence is not guaranteed if the ADMM is directly extended to a multiple-block convex minimization model whose objective function has more than two functions. This viewpoint was in fact the motivation for developing efficient algorithms that cannot only preserve the numerical advantages of the direct extension of ADMM but also guarantee convergence. One way is to correct the output of the direct extension of ADMM slightly via a simple correction step, and the other is to employ a simple proximal to solve inexactly each subproblem in the direct extension of ADMM. In this paper, in order to solve the multi-block separable convex minimization model efficiently, we present a method which is a combination of the above two ways, that is, we first solve each subproblem with a simple proximal, then we correct the output via a simple correction step. Theoretically, we derive global convergence results for this method and establish a worst-case $O(1/k)$ iteration complexity. Numerically, the efficiency of this method can be showed by testing the problem of recovering low-rank and sparse components of matrices from incomplete and noisy observation.

1. Introduction

In this paper, we consider the following convex optimization problem with m block variables and the objective being the sum of m $(m \geq 2)$ separable convex functions:

$$\min\left\{\sum_{i=1}^{m}\theta_i(x_i)\;\middle|\;\sum_{i=1}^{m}A_ix_i=b,\ x_i\in\mathcal{X}_i,\ i=1,2,\ldots,m\right\},\tag{1}$$

where $\mathcal{X}_i \subset \mathcal{R}^{n_i}$ $(i=1,2,\ldots,m)$ are convex sets; $A_i \in \mathcal{R}^{l\times n_i}$, $b\in\mathcal{R}^l$ and $\theta_i : \mathcal{R}^{n_i} \to (-\infty,+\infty]$ $(i=1,2,\ldots,m)$ are lower semicontinuous proper convex (not necessarily smooth) functions. Throughout, we assume that the solution set of (1)

* Corresponding author at: Department of Mathematics, Harbin Institute of Technology, Harbin 150001, China.
E-mail addresses: wxuew_jj@163.com (J.J. Wang), wsong@hrbnu.edu.cn (W. Song).

http://dx.doi.org/10.1016/j.cam.2016.02.001

is nonempty. This model has numerous applications in many fields, such as the latent variable Gaussian graphical model selection in [1], the quadratic discriminant analysis model in [2] and the robust principal component analysis model with noisy and incomplete data in [3,4], and so on.

The augmented Lagrangian function for problem (1) is defined as

$$\mathcal{L}_\beta(x_1, x_2, \dots, x_m, \lambda) = \sum_{i=1}^{m} \theta_i(x_i) - \lambda^T\Big(\sum_{i=1}^{m} A_i x_i - b\Big) + \frac{\beta}{2}\Big\|\sum_{i=1}^{m} A_i x_i - b\Big\|^2, \tag{2}$$

where $\lambda \in \mathcal{R}^l$ is the Lagrangian multiplier for the equality constraint and $\beta > 0$ is a penalty parameter. By setting

$$\begin{cases} u = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}, \qquad \theta(u) = \theta_1(x_1) + \cdots + \theta_m(x_m), & \text{(a)} \\ \mathcal{A} = (A_1, A_2, \dots, A_m), \qquad \mathcal{U} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_m, & \text{(b)} \end{cases} \tag{3}$$

the problem (1) can be rewritten as

$$\min\Big\{\theta(u) \;\Big|\; \mathcal{A}u = b,\ u \in \mathcal{U}\Big\}, \tag{4}$$

and thus the augmented Lagrangian function (2) can be rewritten as

$$\mathcal{L}_\beta(u, \lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b) + \frac{\beta}{2}\|\mathcal{A}u - b\|^2. \tag{5}$$

Sometimes, problems similar to (4) are addressed by means of generalized inverses, which also have other interesting applications (see [5–11]).

The augmented Lagrangian method (ALM) proposed in [12,13] is a classical method for solving (4). Its iterative scheme is described as

$$\begin{cases} u^{k+1} = \arg\min_{u \in \mathcal{U}} \mathcal{L}_\beta(u, \lambda^k), & \text{(a)} \\ \lambda^{k+1} = \lambda^k - \beta(\mathcal{A}u^{k+1} - b). & \text{(b)} \end{cases} \tag{6}$$

Although the u-subproblem in (6) provides an ideal input for updating the variable λ, its solvability critically depends on the properties of θ_i $(i = 1, 2, \dots, m)$. In this setting, the standard augmented Lagrangian algorithm (6) is not very attractive because the minimizations of θ_i in the subproblem (6)(a) are strongly coupled through the term $\frac{\beta}{2}\|\mathcal{A}u - b\|^2$ and hence the subproblems are not likely to be easier to solve than the original problem (1). The conventional alternating direction method of multiplier (ADMM) for solving (1) in the case when $m = 2$ was proposed in [14] (see also [15–20]) and its iterative scheme can be described as below:

$$\begin{cases} x_1^{k+1} = \arg\min_{x_1 \in \mathcal{X}_1} \mathcal{L}_\beta(x_1, x_2^k, \lambda^k), & \text{(a)} \\ x_2^{k+1} = \arg\min_{x_2 \in \mathcal{X}_2} \mathcal{L}_\beta(x_1^{k+1}, x_2, \lambda^k), & \text{(b)} \\ \lambda^{k+1} = \lambda^k - \alpha_0\beta(A_1x_1^{k+1} + A_2x_2^{k+1} - b), & \text{(c)} \end{cases} \tag{7}$$

where α_0 is called step length. Unlike the classical augmented Lagrangian method, the ADMM essentially decouples the functions θ_1 and θ_2. In many situations, this decoupling makes it possible to exploit the individual structure of the θ_1 and θ_2 so that each of (7)(a) and (7)(b) may be computed in an efficient and perhaps highly parallel manner. Fortin and Glowinski [21] and Glowinski [22] have developed an outstanding global convergence analysis for the conventional ADMM (7) with any $\alpha_0 \in (0, (1 + \sqrt{5})/2)$, noticeably in particular $\alpha_0 = 1.618$.

Due to the extreme simplicity and efficiency of the ADMM, it is natural to extend the ADMM (7) directly to the problem (1) with $m \geq 3$ as the following form:

$$\begin{cases} x_1^{k+1} = \arg\min_{x_1 \in \mathcal{X}_1} \mathcal{L}_\beta(x_1, x_2^k, \dots, x_m^k, \lambda^k), & \text{(a)} \\ \quad\vdots \qquad \vdots & \\ x_i^{k+1} = \arg\min_{x_i \in \mathcal{X}_i} \mathcal{L}_\beta(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i, x_{i+1}^k, \dots, x_m^k, \lambda^k), & \text{(b)} \\ \quad\vdots \qquad \vdots & \\ x_m^{k+1} = \arg\min_{x_m \in \mathcal{X}_m} \mathcal{L}_\beta(x_1^{k+1}, \dots, x_{m-1}^{k+1}, x_m, \lambda^k), & \text{(c)} \\ \lambda^{k+1} = \lambda^k - \alpha_0\beta(A_1x_1^{k+1} + A_2x_2^{k+1} + \cdots + A_mx_m^{k+1} - b). & \text{(d)} \end{cases} \tag{8}$$