

A Fast Symmetric Alternating Direction Method of Multipliers

Gang Luo^{1,*} and Qingzhi Yang^{2,1}

¹ School of Mathematical Sciences and LPMC, Nankai University, Tianjin 300071, China

² School of Mathematics and Statistics, Kashi University, Kashi 844006, China

Received 17 October 2018; Accepted (in revised version) 20 May 2019

Abstract. In recent years, alternating direction method of multipliers (ADMM) and its variants are popular for the extensive use in image processing and statistical learning. A variant of ADMM: symmetric ADMM, which updates the Lagrange multiplier twice in one iteration, is always faster whenever it converges. In this paper, combined with Nesterov's accelerating strategy, an accelerated symmetric ADMM is proposed. We prove its $\mathcal{O}(\frac{1}{k^2})$ convergence rate under strongly convex condition. For the general situation, an accelerated method with a restart rule is proposed. Some preliminary numerical experiments show the efficiency of our algorithms.

AMS subject classifications: 90C25, 90C30, 49M29, 65B99

Key words: Nesterov's accelerating strategy, alternating direction method of multipliers, symmetric ADMM, separable linear constrained optimization.

1. Introduction

In this paper, we consider the following convex minimization problem with a linear constrain and a separable objective function:

$$\begin{cases} \min_{x,y} f(x) + g(y) \\ \text{s.t. } Ax + By = c, \end{cases} \quad (1.1)$$

where A, B are linear maps and f, g are convex functions. Problem (1.1) has found numerous applications in statistic and image processing. The augment Lagrangian formulation of (1.1) is

$$\max_{\lambda} \min_{x,y} f(x) + g(y) - \langle \lambda, Ax + By - c \rangle + \frac{\rho}{2} \|Ax + By - c\|^2, \quad (1.2)$$

where λ is the dual variable and ρ is a penalty parameter. Solving (1.1) via (1.2) is exactly the augmented Lagrangian method by Hestenes [16] and Powell [22]. For

*Corresponding author. Email address: luogangnk@gmail.com (G. Luo)

the separable structure in object function in above class problem, alternating direction method of the multipliers (ADMM) algorithm [9], one variant of ALM, is preferred. It minimizes (1.2) on x, y alternatively, then updates the dual variable λ . Readers can refer to the review paper [2] for some applications of ADMM in statistics and machine learning fields.

The iterative scheme of ADMM on (1.1) is

$$\begin{cases} x^{k+1} = \operatorname{argmin}_x f(x) - \langle \lambda^k, Ax \rangle + \frac{\rho}{2} \|Ax + By^k - c\|^2, \\ y^{k+1} = \operatorname{argmin}_y g(y) - \langle \lambda^k, By \rangle + \frac{\rho}{2} \|Ax^{k+1} + By - c\|^2, \\ \lambda^{k+1} = \lambda^k - \rho(Ax^{k+1} + By^{k+1} - c), \end{cases} \quad (1.3)$$

where x, y and Lagrange multiplier λ are updated in each iteration. ADMM is shown to be equivalent to the Douglas-Rachford splitting method (DRSM) [5] on the dual problem of (1.1). The convergence of ADMM under general condition is guaranteed for two block situation, and a proof can be found in [2]. The above algorithm can be easily extended to solve linear constrained minimization with three or more separated block objective function, while in these cases, its convergence is no longer guaranteed under general conditions [4].

Apply another famous splitting method: Peaceman-Rachford splitting method (PRSM) [21] on the dual problem of (1.1), we get a variation of ADMM and its iterative scheme is

$$\begin{cases} x^{k+1} = \operatorname{argmin}_x f(x) - \langle \lambda^k, Ax \rangle + \frac{\rho}{2} \|Ax + By^k - c\|^2, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \rho(Ax^{k+1} + By^k - c), \\ y^{k+1} = \operatorname{argmin}_y g(y) - \langle \lambda^{k+\frac{1}{2}}, By \rangle + \frac{\rho}{2} \|Ax^{k+1} + By - c\|^2, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \rho(Ax^{k+1} + By^{k+1} - c). \end{cases} \quad (1.4)$$

This algorithm is called symmetric ADMM (sADMM) for it updates λ twice in one iteration.

Different from ADMM, symmetric ADMM requires more to ensure its convergence [12], but it shows a faster convergence than ADMM in numerical computing. In [12], a contractive step size $a \in (0, 1)$ was introduced to the dual variable updating step to ensure the convergence of the algorithm

$$\begin{cases} x^{k+1} = \operatorname{argmin}_x f(x) - \langle \lambda^k, Ax \rangle + \frac{\rho}{2} \|Ax + By^k - c\|^2, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - a\rho(Ax^{k+1} + By^k - c), \\ y^{k+1} = \operatorname{argmin}_y g(y) - \langle \lambda^{k+\frac{1}{2}}, By \rangle + \frac{\rho}{2} \|Ax^{k+1} + By - c\|^2, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - a\rho(Ax^{k+1} + By^{k+1} - c). \end{cases} \quad (1.5)$$

ADMM and sADMM are the first-order algorithms. In [12, 14, 15], their $\mathcal{O}(\frac{1}{k})$ convergences were established. In practice, they can converge slowly to reach a high accuracy,