CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.SO-2021-0021

Vol. 3, No. 3, pp. 448-479
September 2022

Convergence on a Symmetric Accelerated Stochastic ADMM with Larger Stepsizes

Jianchao Bai[1], Deren Han[2,*], Hao Sun[3] and Hongchao Zhang[4]

[1] *School of Mathematics and Statistics, the MIIT Key Laboratory of Dynamics and Control of Complex Systems, Northwestern Polytechnical University, Xi'an, 710129, P.R. China.*
[2] *LMIB of the Ministry of Education, School of Mathematical Sciences, Beihang University, Beijing, 100191, P.R. China.*
[3] *School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an, 710129, P.R. China.*
[4] *Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803-4918, USA.*

Received 30 March 2021; Accepted 19 December 2021

Abstract. In this paper, we develop a symmetric accelerated stochastic Alternating Direction Method of Multipliers (SAS-ADMM) for solving separable convex optimization problems with linear constraints. The objective function is the sum of a possibly nonsmooth convex function and an average function of many smooth convex functions. Our proposed algorithm combines both ideas of ADMM and the techniques of accelerated stochastic gradient methods possibly with variance reduction to solve the smooth subproblem. One main feature of SAS-ADMM is that its dual variable is symmetrically updated after each update of the separated primal variable, which would allow a more flexible and larger convergence region of the dual variable compared with that of standard deterministic or stochastic ADMM. This new stochastic optimization algorithm is shown to have ergodic converge in expectation with $\mathcal{O}(1/T)$ convergence rate, where T denotes the number of outer iterations. Our preliminary experiments indicate the proposed algorithm is very effective for solving separable optimization problems from big-data applications. Finally, 3-block extensions of the algorithm and its variant of an accelerated stochastic augmented Lagrangian method are discussed in the appendix.

AMS subject classifications: 65K10, 65Y20, 68W40, 90C25

Key words: Convex optimization, stochastic ADMM, symmetric ADMM, larger stepsize, proximal mapping, complexity.

*Corresponding author. *Email addresses:* `jianchaobai@nwpu.edu.cn` (J. Bai), `handr@buaa.edu.cn` (D. Han), `hsun@nwpu.edu.cn` (H. Sun), `hozhang@math.lsu.edu` (H. Zhang)

1 Introduction

We consider the following structured composite convex optimization problem with linear equality constraints:

$$\min\{f(\boldsymbol{x})+g(\boldsymbol{y})\,|\,\boldsymbol{x}\in\mathcal{X},\ \boldsymbol{y}\in\mathcal{Y},\ A\boldsymbol{x}+B\boldsymbol{y}=\boldsymbol{b}\}, \tag{1.1}$$

where $\mathcal{X}\subset\mathbb{R}^{n_1}$, $\mathcal{Y}\subset\mathbb{R}^{n_2}$ are closed convex subsets, $A\in\mathbb{R}^{n\times n_1}$, $B\in\mathbb{R}^{n\times n_2}$, $\boldsymbol{b}\in\mathbb{R}^n$ are given, $g:\mathcal{Y}\to\mathbb{R}\cup\{+\infty\}$ is a convex but possibly nonsmooth function, and f is an average of N real-valued convex functions:

$$f(\boldsymbol{x})=\frac{1}{N}\sum_{j=1}^{N}f_j(\boldsymbol{x}).$$

We assume that each f_j defined on an open set containing $\mathcal{X}$ is Lipschitz continuously differentiable on $\mathcal{X}$. Problem (1.1) is also referred as the regularized empirical risk minimization in big-data applications [26,35], including classification and regression models in machine learning, where N denotes the sample size and f_j corresponds to the empirical loss. A major difficulty for solving (1.1) is that the sample size N can be very large such that it is often computationally prohibitive to evaluate either the full function value or the gradient of f at each iteration of an algorithm. Hence, it is essential for an effective algorithm, e.g., a stochastic gradient method, to explore the summation structure of f in the objective function.

The augmented Lagrangian function of (1.1) is

$$\mathcal{L}_\beta(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda})=\mathcal{L}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda})+\frac{\beta}{2}\|A\boldsymbol{x}+B\boldsymbol{y}-\boldsymbol{b}\|^2, \tag{1.2}$$

where $\beta>0$ is a penalty parameter, $\boldsymbol{\lambda}$ is the Lagrange multiplier and the Lagrangian of (1.1) is defined as

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda})=f(\boldsymbol{x})+g(\boldsymbol{y})-\boldsymbol{\lambda}^{\mathsf{T}}(A\boldsymbol{x}+B\boldsymbol{y}-\boldsymbol{b}). \tag{1.3}$$

Although the Augmented Lagrangian Method (ALM) can be applied to solve (1.1), it does not take full advantage of the separable structure of (1.1). As a splitting version of ALM, the standard Alternating Direction Method of Multipliers (ADMM, [11,12]) exploits the separable structure of the objective function and performs the following iterations:

$$\begin{cases} \boldsymbol{x}^{k+1}\in\arg\min\limits_{\boldsymbol{x}\in\mathcal{X}}\mathcal{L}_\beta(\boldsymbol{x},\boldsymbol{y}^k,\boldsymbol{\lambda}^k),\\ \boldsymbol{y}^{k+1}\in\arg\min\limits_{\boldsymbol{y}\in\mathcal{Y}}\mathcal{L}_\beta(\boldsymbol{x}^{k+1},\boldsymbol{y},\boldsymbol{\lambda}^k),\\ \boldsymbol{\lambda}^{k+1}=\boldsymbol{\lambda}^k-s\beta\Big(A\boldsymbol{x}^{k+1}+B\boldsymbol{y}^{k+1}-\boldsymbol{b}\Big), \end{cases}$$

where $s\in(0,\frac{1+\sqrt{5}}{2})$ is the stepsize for updating the dual variable $\boldsymbol{\lambda}$.