INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 20, Number 4, Pages 518–537

doi: 10.4208/ijnam2023-1022

CONVERGENCE RATE ANALYSIS OF ACCELERATED FORWARD-BACKWARD ALGORITHM WITH GENERALIZED NESTEROV MOMENTUM SCHEME

YIZUN LIN[1], SI LI[2,*], AND YUNZHONG ZHANG[3]

Abstract. Nesterov's accelerated forward-backward algorithm (AFBA) is an efficient algorithm for solving a class of two-term convex optimization models consisting of a differentiable function with a Lipschitz continuous gradient plus a nondifferentiable function with a closed form of its proximity operator. It has been shown that the iterative sequence generated by AFBA with a modified Nesterov's momentum scheme converges to a minimizer of the objective function with an $o\left(\frac{1}{k^2}\right)$ convergence rate in terms of the function value (FV-convergence rate) and an $o\left(\frac{1}{k}\right)$ convergence rate in terms of the distance between consecutive iterates (DCI-convergence rate). In this paper, we propose a more general momentum scheme with an introduced power parameter $\omega \in (0,1]$ and show that AFBA with the proposed momentum scheme converges to a minimizer of the objective function with an $o\left(\frac{1}{k^{2\omega}}\right)$ FV-convergence rate and an $o\left(\frac{1}{k^{\omega}}\right)$ DCI-convergence rate. The generality of the proposed momentum scheme provides us a variety of parameter selections for different scenarios, which makes the resulting algorithm more flexible to achieve better performance. We then employ AFBA with the proposed momentum scheme to solve the smoothed hinge loss ℓ_1-support vector machine model. Numerical results demonstrate that the proposed generalized momentum scheme outperforms two existing momentum schemes.

Key words. Nesterov's momentum, forward-backward algorithm, convergence rate, support vector machine.

1. Introduction

In this paper, we consider fast algorithm with a generalized Nesterov momentum scheme for solving a class of two-term optimization problems of the form

$$\min_{x\in\mathbb{R}^n}\{f(x)+g(x)\}, \tag{1}$$

where $f:\mathbb{R}^n\to\mathbb{R}$ is a convex and differentiable function with a Lipschitz continuous gradient, $g:\mathbb{R}^n\to\mathbb{R}$ is a proper lower-semicontinuous convex function which may not be differentiable. This two-term optimization model has important applications in machine learning (e.g. LASSO regression, support vector machine) [23, 25, 26], image processing (e.g. image denoising, image restoration) [10, 11, 12], compressed sensing [13, 24] and so on.

The possible nondifferentiability of g in model (1) precludes the use of classical gradient type algorithms. Under these circumstances, the Forward-Backward Algorithm (FBA) [16, 20] was developed to solve the model when the proximity operator of g has a closed-form. The FBA is easily-implemented and robust. However, for large scale ill-conditioned problems, it has been shown to be too slow no matter in practice or in the sense of asymptotic rate of convergence [3, 5]. To address this issue, various modifications of FBA have been developed [3, 4, 9]. One of the most popular strategies is the utilization of momentum technique, such as Nesterov's

Received by the editors on November 25, 2022 and, accepted on April 13, 2023.
2000 *Mathematics Subject Classification.* 49M37, 65K05, 90C25.
*Corresponding author: Si Li; E-mail: sili@gdut.edu.cn.

momentum scheme [19]. Beck and Teboulle showed that FBA has an $O\left(\frac{1}{k}\right)$ convergence rate in terms of the function value (FV-convergence rate), and FBA with Nesterov's momentum (Fast Iterative Shrinkage-Thresholding Algorithm, FISTA) can improve the FV-convergence rate to $O(\frac{1}{k^2})$. However, the convergence of the iterative sequence generated by FISTA is unclear in their work [3]. Chambolle and Dossal proved in [6] not only the $O(\frac{1}{k^2})$ FV-convergence rate but also the convergence of the iterative sequence for the momentum accelerated forward-backward algorithm with a new setting of momentum parameters (AFBA-CD). Later, Attouch and Peypouquet showed that AFBA-CD can actually achieve an $o\left(\frac{1}{k^2}\right)$ FV-convergence rate and an $o\left(\frac{1}{k}\right)$ convergence rate in terms of the distance between consecutive iterates (DCI-convergence rate) [1]. Although AFBA-CD is theoretically guaranteed to be faster than FISTA, it does not always give a distinguishingly improved performance on practical applications.

In this work, we propose a more general setting of momentum parameters in the Accelerated Forward-Backward Algorithm (AFBA). A power parameter $\omega \in (0, 1]$ is introduced in our momentum scheme. We shall show that the setting of momentum parameters in [6] is a special case of the proposed generalized scheme with $\omega = 1$. More importantly, the iterative sequence generated by AFBA with the generalized momentum scheme converges to a minimizer of the objective function with an $o\left(\frac{1}{k^{2\omega}}\right)$ FV-convergence rate and an $o\left(\frac{1}{k^{\omega}}\right)$ DCI-convergence rate. This result provides a wider class of momentum algorithms with various convergence rates. Numerical results demonstrate that the proposed momentum scheme outperforms the existing momentum schemes used in [3] and [6] for classification problems using Support Vector Machine (SVM).

We organize this paper in six sections. In section 2, we describe the accelerated forward backward algorithm and three types of momentum schemes, including two existing schemes and the proposed generalized scheme. We analyze in section 3 the convergence of the iterative sequence and both the FV-convergence rate and the DCI-convergence rate for AFBA with the proposed momentum scheme. In section 4, we formulate the smoothed hinge loss ℓ_1-SVM model as the two-term optimization model (1), and then employ AFBA to solve this model. Section 5 presents the numerical results for comparison of the proposed momentum scheme with the other two schemes mentioned in section 2. Section 6 offers a conclusion.

2. Accelerated forward-backward algorithm

In this section, we first review the Accelerated Forward-Backward Algorithm (AFBA) for solving model (1) and two existing momentum schemes. Inspired by these two schemes, we then propose a more general setting of momentum parameters. To better describe the iteration scheme of AFBA, we recall the definition of proximity operator of a convex function [18]. For $x, y \in \mathbb{R}^n$, the inner product is defined by $\langle x, y\rangle := \sum_{i=1}^n x_i y_i$, and the corresponding ℓ_2 norm is given by $\|x\| := \langle x, x\rangle^{\frac{1}{2}}$.

Definition 2.1. *Let $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper convex function. The proximity operator of ψ at $x \in \mathbb{R}^n$ is defined by*

$$\operatorname{prox}_{\psi}(x) := \underset{u\in\mathbb{R}^n}{\operatorname{argmin}} \left\{\frac{1}{2}\|u - x\|^2 + \psi(u)\right\}. \tag{2}$$

Throughout this paper, we define $F := f + g$ and

$$T := \operatorname{prox}_{\beta g} \circ (\mathcal{I} - \beta\nabla f), \tag{3}$$