

ACCELERATED BREGMAN OPERATOR SPLITTING WITH BACKTRACKING

YUNMEI CHEN AND XIANQI LI*

Department of Mathematics
University of Florida
Gainesville, FL 32611, USA

YUYUAN OUYANG

Department of Mathematical Sciences
Clemson University
Clemson, SC 29634, USA

EDUARDO PASILIAO

Munitions Directorate
Air Force Research Laboratory
Eglin AFB, FL 32542-9998, USA

(Communicated by Hao-Min Zhou)

ABSTRACT. This paper develops two accelerated Bregman Operator Splitting (BOS) algorithms with backtracking for solving regularized large-scale linear inverse problems, where the regularization term may not be smooth. The first algorithm improves the rate of convergence for BOSVS [5] in terms of the smooth component in the objective function by incorporating Nesterov's multi-step acceleration scheme under the assumption that the feasible set is bounded. The second algorithm is capable of dealing with the case where the feasible set is unbounded. Moreover, it allows more aggressive stepsize than that in the first scheme by properly selecting the penalty parameter and jointly updating the acceleration parameter and stepsize. Both algorithms exhibit better practical performance than BOSVS and AADMM [21], while preserve the same accelerated rate of convergence as that for AADMM. The numerical results on total-variation based image reconstruction problems indicate the effectiveness of the proposed algorithms.

1. Introduction. The main purpose of this paper is to develop accelerated Bregman Operator Splitting algorithms with backtracking for solving the following convex composite optimization problem

$$(1.1) \quad \min_{u \in U} \left\{ \frac{1}{2} \|Au - f\|^2 + \psi(Bu) \right\},$$

2010 *Mathematics Subject Classification.* Primary: 92C55; Secondary: 90C25.

Key words and phrases. Bregman operator splitting, accelerated ADMM, convex optimization, Barzilai-Borwein stepsize, backtracking, total variation image reconstruction.

The first author is supported by NSF grant DMS-1319050 and AFOSR/Eglin (FA8651-08-D-0108). The second author was partially supported by UFII fellowship and AFRL Mathematical Modeling and Optimization Institute.

* Corresponding author: Xianqi Li.

where $U \in \mathbb{C}^n$ is a closed convex set, $A \in \mathbb{C}^{m \times n}$ is a matrix, $f \in \mathbb{C}^m$ is a vector, $\psi : \mathbb{C}^{d \times n} \rightarrow \mathbb{R}$ is a proper convex and Lipschitz continuous function, $B : \mathbb{C}^{d \times n \times n} \rightarrow \mathbb{C}^{d \times n}$ is a linear operator, and $\|\cdot\|$ is the 2-norm in Euclidean space.

The problem (1.1) has attracted much attention in recent years because of its applications in various disciplines, such as signal and image processing [1, 2, 4, 22, 27, 28], compressive sensing [7, 12, 18], and machine learning [26, 17, 3]. In those fields, the first term in (1.1) represents the data fidelity and the second term is a regularization term. For instance, the total variation based image reconstruction is often modeled by (1.1), where $u = (u_1, \dots, u_n)^T$ is the n -vector form of the image u , the matrix A depends on the physics of the data acquisition, f is the measurement, and

$$(1.2) \quad \psi(Bu) = \alpha \|u\|_{TV} = \alpha \sum_{i=1}^n \|(\nabla u)_i\|$$

is the discrete form of the total variation of u . In (1.2), $(\nabla u)_i$ is a discrete gradient (finite differences along the coordinate directions) of u at the i -th voxel, and α is a parameter, which balances the fidelity term and the regularization term.

There have been a large number of algorithms developed for solving problem (1.1) (see, e.g. [5, 13, 12, 16, 23, 25, 27, 29, 30, 31, 32]). Among them, the alternating direction method of multiplier (ADMM) [8, 9] is one of the most extensively used techniques. By introducing an auxiliary variable $w \in \mathbb{C}^{d \times n}$, the problem (1.1) is equivalent to the following constrained optimization problem:

$$(1.3) \quad \min_{u \in U, w \in \mathbb{C}^{d \times n}} \frac{1}{2} \|Au - f\|^2 + \psi(w) \quad \text{subject to} \quad w = Bu.$$

The augmented Lagrangian function associated with (1.3) is

$$(1.4) \quad \mathcal{L}(u, w, \lambda) = \frac{1}{2} \|Au - f\|^2 + \psi(w) + \langle \lambda, Bu - w \rangle + \frac{\rho}{2} \|Bu - w\|^2.$$

The ADMM minimizes (1.4) by iteratively updating (u, w, λ) as follows:

$$(1.5) \quad \begin{cases} u_{k+1} = \arg \min_{u \in U} \left\{ \frac{1}{2} \|Au - f\|^2 + \frac{\rho}{2} \left\| Bu - w_k + \frac{\lambda_k}{\rho} \right\|^2 \right\} \\ w_{k+1} = \arg \min_{w \in \mathbb{C}^{d \times n}} \left\{ \psi(w) + \frac{\rho}{2} \left\| Bu_{k+1} - w + \frac{\lambda_k}{\rho} \right\|^2 \right\} \\ \lambda_{k+1} = \lambda_k - \rho(w_{k+1} - Bu_{k+1}). \end{cases}$$

In many applications, the dimension of the matrix A could be very large and A might be also dense and ill-conditioned. In those cases, the high computational cost for solving the u -subproblem hindered the applicability of ADMM. Recently, there have been active researches on improving the theoretical and practical performance of ADMM. One type of efforts has been focused on developing backtracking strategies to search for larger stepsize with convergence guaranteed [4, 5, 13]. The other effort is on improving the iteration complexity of the ADMM algorithm by incorporating Nesterov's multi-step acceleration schemes [15, 6, 10, 19, 11, 20, 21]. To have more insight from those two types of efforts, below we will discuss a few closely related works.

In [29, 30], the u -subproblem of (1.5) is solved by Bregman operator splitting (BOS) technique, i.e. linearized ADMM. It updates u_{k+1} of (1.5) by linearizing

$\frac{1}{2}\|Au - f\|^2$ at u_k and adding a proximal term, i.e.

$$(1.6) \quad u_{k+1} = \arg \min_{u \in U} \left\{ \langle A^T(Au_k - f), u \rangle + \frac{\eta}{2}\|u - u_k\|^2 + \frac{\rho}{2} \left\| Bu - w_k + \frac{\lambda_k}{\rho} \right\|^2 \right\}.$$

It has been shown in [30] that the BOS algorithm converges to an optimal solution of (1.1) with the constant stepsize $1/\eta \leq 1/\|A^T A\|$. In [4], this stepsize is replaced by the Barzilai-Borwein (BB) stepsize, i.e. $\eta = \eta_k^{BB}$, where

$$(1.7) \quad \eta_k^{BB} = \frac{\|A(u_k - u_{k-1})\|^2}{\|u_k - u_{k-1}\|^2}.$$

Their experimental results showed that the objective function value decreased much faster by taking $\eta = \eta_k^{BB}$ than $\eta = \|A^T A\|_2$. However, the BB stepsize $1/\eta_k^{BB} \geq 1/\|A^T A\|_2$ violates the convergence condition of the BOS algorithm. Thus, the convergence of the algorithm is not guaranteed. The improvement was made in [5], where a scheme of BOS with variable stepsize (BOSVS) was developed to improve the performance of ADMM for solving (1.1) with guaranteed convergence. In this work a line search strategy was presented for searching for a better stepsize. The stepsize rule starts with a safeguarded BB stepsize and gradually increases the nominal stepsize η_k until the termination condition is satisfied. With a good choice of parameters in the line search conditions, more aggressive stepsize is allowed, especially at the early iterates. The global convergence of the iterates of the BOSVS was established in [5]. Later on, it was shown in [14] that the objective function at the average of the BOSVS iterates converge to an optimal value with the rate of $O(1/k)$, where k is the number of iterations.

To improve the iteration complexity of ADMM, in [21] an accelerated ADMM (AADMM) was developed by incorporating Nesterov's fast gradient scheme [20]. The AADMM solves a class of convex composite optimization problems with linear equality constraints, which includes the problem (1.3) as a special case. AADMM improves the rate of convergence of ADMM (or BOS) in terms of the smooth component in the objective function from $O(1/k)$ to $O(1/k^2)$. The accelerated rate of convergence is achieved by the aggregated iterates rather than the average of the iterates in [14]. To improve the practical performance of AADMM, a simple backtracking strategy is incorporated in the algorithm. The idea of the backtracking technique is to search for an underestimated Lipschitz constants L_k at the iteration k in consideration of the dependence of the stepsize on L_k . The backtracking procedure in [21] starts with a relatively smaller L_k , and properly selects involved parameters to solve u_{k+1} . If an employed line search condition is violated, then L_k is doubled and used as the new L_k to solve u_{k+1} . This procedure is repeated until the utilized line search condition holds.

Motivated by the aforementioned work, we propose two accelerated Bregman Operator Splitting (BOS) schemes with backtracking for solving the problem (1.1). The first proposed algorithm improves the convergence rate of BOSVS in terms of the smooth component in the objective function by incorporating Nesterov's multi-step acceleration scheme [20] under the assumption that the feasible set is bounded. The second one can deal with the situation when the feasible set is unbounded. By jointly computing the acceleration parameter and stepsize, the monotonicity condition on the nominal stepsize required in BOSVS [5] and the first proposed algorithm can be removed in the second algorithm. Combining with the good choices of the penalty parameters ρ for updating u , w and λ in (1.5) (

possibly different ρ 's in u, w, λ subproblems), more aggressive stepsize is allowed in the second proposed algorithm. Instead of searching the local Lipschitz constant to satisfy the conservative line search condition in AADMM [21], the proposed algorithms utilize the product of the acceleration parameter and a safeguarded Barzilai-Borwein (BB) choice as the initial stepsize, then gradually increase it until a more relaxed line search conditions than that in [21] is satisfied. The proposed algorithms are capable of hunting for more aggressive stepsize via conducting fewer number of line searches. Meanwhile, the proposed algorithms preserve the same accelerated rate of convergence as that for AADMM.

In [21], a center piece of the theoretical analysis is for the convergence rate of AADMM in terms of the dependence on the Lipschitz constant of the smooth objective. A simple backtracking scheme is proposed in [21] to demonstrate that prior knowledge of the Lipschitz constant is not a definitive requirement. In this work most of the convergence analyses is devoted to maintaining the accelerated convergence results under aggressive stepsize strategy, which was not touched in either [5] or [21]. The problem (1.1) is a special case of the problem of interest in [21] which is called the unconstrained composite optimization (UCO) problem there. The backtracking technique in [21] is designed to solve UCO with bounded primal and dual feasible sets. In this manuscript, the proposed ABOSVS II algorithm is able to solve (1.1) with an unbounded primal feasible set without any feasibility residue, which was not discussed in [21]. It seems that the technical details for studying (1.1) with unbounded primal feasible set is nontrivial comparing with the analysis in [21]. Moreover, in this work we use a different termination criterion from the one for AADMM in [21] to obtain the same accelerated rate of convergence. In [21], the convergence analysis is based on the estimation of the duality gap function for its corresponding saddle point problem. In this work, the termination criterion is based on the error between the objective function values at the aggregated iterates and the optimal solution. Consequently, we are able to conduct a different proof of the accelerated rate of convergence when the feasible set is unbounded by observing the relationship between the Lipschitz continuity of the function ψ in (1.1) and the boundedness of its subgradients. We believe that the strategy in the proof of the case with unbounded feasible sets provides a relatively simpler alternative of the proof of accelerated convergence results in [21]. Our experimental results show that the proposed algorithms outperform several state-of-the-art algorithms on total variation based image reconstruction problems.

1.1 Outline of the paper

Our paper is organized as follows. Section 2 presents the proposed algorithms, namely Accelerated BOSVS-I (ABOSVS-I) and Accelerated BOSVS-II (ABOSVS-II), for solving the type of problems (1.1). Section 3 studies the convergence analysis for the proposed algorithms. Section 4 is devoted to numerical experiments and comparisons with state-of-the-art algorithms on total-variation based image reconstruction problems. The last section draws the conclusion for this paper.

1.2 Notation and terminologies

The Euclidean inner product of two column vectors $x, y \in \mathbb{C}^n$ is denoted by $\langle x, y \rangle = x^T y$, where the superscript T denotes the conjugate transpose. Assume u^* is an optimal solution of (1.1). Define $D_{u^*} := \|u_1 - u^*\|$, $D_{u^*, B} := \|B(u_1 - u^*)\|$,

$D_{\lambda^*} := \|\lambda_1 - \lambda^*\|$, $D_{U,B} := \sup_{u_1, u_2 \in U} \|Bu_1 - Bu_2\|$, $D_V := \sup_{v_1, v_2 \in V} \|v_1 - v_2\|$ for any compact sets U and V . For a convex function $F : \mathbb{C}^n \rightarrow \mathbb{R}$, ∂F represents the subdifferential of F .

2. Proposed Algorithms. In this section, we present the frameworks of the ABOSVS-I and ABOSVS-II. The first algorithm incorporates Nesterov's multi-step acceleration scheme to improve the rate of convergence of BOSVS under the assumption that the feasible set is bounded. While the second one is capable of dealing with the case where the feasible set is unbounded. In the second algorithm the acceleration parameter and the stepsize are updated jointly and the penalty parameters ρ in (1.5) are chosen differently for updating u -, w - and λ . For convenience, We use ρ_i^u, ρ_i^w , and ρ_i^λ to replace ρ in those three subproblems, respectively. For ABOSVS-I, with assumption of bounded feasible set, ρ can be chosen to be the same constant in (1.5) without affecting the convergence rate, i.e.

$$(2.1) \quad \rho_i^u = \rho_i^w = \rho_i^\lambda = \rho > 0.$$

In the proposed algorithms, the initial choice of the nominal step η_i is a safeguarded BB choice:

$$(2.2) \quad \eta_{0,1} \geq \eta_{\min} \quad \text{and} \quad \eta_{0,i} = \max \{ \eta_{\min}, \eta_i^{BB} \} \quad \text{for } i > 1.$$

where η_i^{BB} is defined in (1.7). Now we present the scheme of ABOSVS-I in Algorithm 1 under the assumption that the feasible set U is bounded.

From ABOSVS-I, we can see that if $\alpha_i \equiv 1$, then $u_i^{md} = u_i$ and the aggregate points $u_{i+1}^{ag}, w_{i+1}^{ag}, \lambda_{i+1}^{ag}$ are exactly the iterates u_{i+1}, w_{i+1} , and λ_{i+1} , respectively. In this case, ABOSVS-I becomes BOSVS with a minor modification. For ABSOVS-I, the derivation of the accelerated convergence rate relies on the asymptotic monotonicity of η_i/α_i for $i = 1, 2, \dots$. Whenever η_i/α_i is not monotone decreasing, $\eta_{\min}$ is increased by a factor $\tau > 1$ in step 3 for $i = 1, 2, \dots$. Hence, if the monotonicity of η_i/α_i violates continuously, then η_i/α_i will approach a constant, which is usually smaller than $\|A^T A\|_2$, after a finite number of iterations. Moreover, it should be note that from $Q_{i+1} = \beta_i Q_i + \Gamma_i$, we have

$$(2.3) \quad \sum_{i=2}^k (\beta_i - 1) Q_i - Q_{k+1} = - \sum_{i=1}^k \Gamma_i.$$

where β_i and the condition on Q_i can be chosen jointly as long as that $-\sum_{i=1}^k \Gamma_i$ is a finite positive number.

ABOSVS-II can deal with the cases where the feasible set is either bounded or unbounded. When the feasible set is unbounded, the penalty parameters ρ_i^u, ρ_i^w , and ρ_i^λ have to be chosen differently from those for bounded feasible set. More precisely, if the feasible set U is unbounded, we choose

$$(2.4) \quad \rho_i^u = \rho_i^w = \frac{K\eta_i}{\eta_{\min}} \quad \text{and} \quad \rho_i^\lambda = \frac{\eta_{\min}}{K\eta_i},$$

where K is the total number of iterations. If the feasible set U is bounded, we choose

$$(2.5) \quad \rho_i^u = \rho_i^w = \frac{i\eta_i}{\eta_{\min}} \quad \text{and} \quad \rho_i^\lambda = \frac{\eta_{\min}}{i\eta_i},$$

Algorithm 1 Accelerated BOSVS I (ABOSVS-I)

Choose $C \geq 0, \sigma > 1, \rho > 0, \tau > 1, \eta_{0,1} \geq \eta_{\min} > 0$, and $u_1 \in U$. Set $\alpha_1 = 1$, $u_1^{ag} = u_1$, $w_1^{ag} = w_1 = Bu_1$, $\lambda_1^{ag} = \lambda_1 = 0$, and $Q_1 = 0$.

For $i = 1, 2, \dots, k$,

// Backtracking

1. Set $\eta_i = \alpha_i \eta_{0,i}$, where $\eta_{0,i}$ is from (2.2). Compute

$$u_i^{md} = (1 - \alpha_i)u_i^{ag} + \alpha_i u_i,$$

$$u_{i+1} = \arg \min_{u \in U} \left\{ \langle A^T(Au_i^{md} - f), u \rangle + \frac{\eta_i}{2} \|u - u_i\|^2 + \frac{\rho_i^u}{2} \left\| Bu - w_i + \frac{\lambda_i}{\rho_i^u} \right\|^2 \right\},$$

$$\Gamma_i = \frac{\eta_i}{\alpha_i} \|u_{i+1} - u_i\|^2 + \frac{\rho_i^u}{\alpha_i} \|Bu_{i+1} - w_i\|^2 - \|A(u_{i+1} - u_i)\|^2,$$

$$Q_{i+1} = \beta_i Q_i + \Gamma_i, \quad \text{where } 0 \leq \beta_i \leq (1 - \frac{1}{i})^2.$$

2. If $Q_{i+1} < -C/i^2$, then replace $\eta_{0,i}$ by $\sigma \eta_{0,i}$ and return to step 1.
3. Update the safeguard threshold $\eta_{\min}$ by $\tau \eta_{\min}$ if $\eta_i/\alpha_i > \eta_{i-1}/\alpha_{i-1}$, for $i > 1$.

// Updating iterates

4. Compute

$$u_{i+1}^{ag} = (1 - \alpha_i)u_i^{ag} + \alpha_i u_{i+1},$$

$$w_{i+1} = \arg \min_{w \in \mathbb{C}^{d \times n}} \left\{ \psi(w) + \frac{\rho_i^w}{2} \left\| w - Bu_{i+1} - \frac{\lambda_i}{\rho_i^w} \right\|^2 \right\},$$

$$w_{i+1}^{ag} = (1 - \alpha_i)w_i^{ag} + \alpha_i w_{i+1},$$

$$\lambda_{i+1} = \lambda_i - \rho_i^\lambda (w_{i+1} - Bu_{i+1}),$$

$$\lambda_{i+1}^{ag} = (1 - \alpha_i)\lambda_i^{ag} + \alpha_i \lambda_{i+1},$$

$$\alpha_{i+1} = \frac{\alpha_i \left(-\alpha_i + \sqrt{\alpha_i^2 + 4} \right)}{2}.$$

EndFor

Output $(u_{k+1}^{ag}, w_{k+1}^{ag})$.

in ABOSVS-II by jointly updating the acceleration parameter and the stepsize. The scheme of ABOSVS-II is presented in Algorithm 2.

Two remarks are in place. First, step 3 from Algorithm 1 is not required anymore in Algorithm 2. Second, in Algorithm 2, step 2 actually can be written as

$$(2.6) \quad \text{If } Q_{i+1} < -C, \text{ then } \eta_{0,i} := \sigma \eta_{0,i}, \text{ return to step 1}$$

by setting $\beta_i = 1$ for $i > 1$. Obviously, it also satisfies the condition we impose on $-\sum_{i=1}^k \Gamma_i$ from (2.3). Additionally, in (2.6), we can also set β_i such that $(1 - 1/i^2) \leq \beta_i \leq 1$.

Algorithm 2 Accelerated BOSVS II (ABOSVS-II)

Choose $C \geq 0, \sigma > 1, \eta_{0,1} \geq \eta_{\min}$, u_1 and w_1 such that $w_1 = Bu_1$. Set $\alpha_1 = 1$, $u_1^{ag} = u_1$, $w_1^{ag} = w_1$, $\lambda_1^{ag} = \lambda_1 = 0$, and $i = 1$.

For $i = 1, 2, \dots, k$,

// *Backtracking*

1. Set $\eta_i = \alpha_i \eta_{0,i}$, where $\eta_{0,i}$ is from (2.2). Solve α_i from $\frac{1}{\alpha_{i-1} \eta_{i-1}} = \frac{1-\alpha_i}{\alpha_i \eta_i}$ for $i > 1$. Update ρ_i^u, ρ_i^w and ρ_i^λ by (2.4) or (2.5). Compute $u_i^{md} = (1 - \alpha_i)u_i^{ag} + \alpha_i u_i$.
 $u_{i+1} = \arg \min_{u \in U} \left\{ \langle A^T(Au_i^{md} - f), u \rangle + \frac{\eta_i}{2} \|u - u_i\|_2^2 + \frac{\rho_i^u}{2} \|Bu - w_i + \frac{\lambda_i}{\rho_i^u}\|^2 \right\}.$
2. If $\sum_{j=1}^i \Gamma_j < -C$,
 where $\Gamma_j := \|u_{j+1} - u_j\|^2 + \frac{\rho_j^u}{\eta_j} \|Bu_{j+1} - w_j\|^2 - \frac{\alpha_j}{\eta_j} \|A(u_{j+1} - u_j)\|^2$,
 then $\eta_{0,i} := \sigma \eta_{0,i}$, return to step 1.

// *Updating iterates*

3. Compute the iterates $u_{i+1}^{ag}, w_{i+1}, w_{i+1}^{ag}, \lambda_{i+1}, \lambda_{i+1}^{ag}$ from step 4 in Algorithm 1.

EndFor

Output $(u_{k+1}^{ag}, w_{k+1}^{ag})$.

3. Convergence analysis. In this section, we focus on proving the convergence properties of the proposed algorithms. We start with two lemmas describing properties regarding the line search scheme in Algorithms 1 and 2. Lemma 1 below shows that the safeguard stepsize threshold $\eta_{\min}$ in step 3 of ABOSVS-I will stop increasing after a finite number of iteration.

Lemma 3.1. *The replacement of $\eta_{\min}$ by $\tau \eta_{\min}$ in step 3 of ABOSVS-I can occur in at most a finite number of iterations, denoted by N_0 .*

Proof. See Lemma 3.2 (II) of [5]. $\square$

Lemma 3.2. *In steps 2 and 3 of Algorithm 1 and step 2 of Algorithm 2, the number of line search, denoted by l , is less than or equal to $\lceil \log_\sigma(\frac{\|A^T A\|}{\eta_{\min}}) \rceil$, where $\lceil x \rceil$ is the smallest integer greater than or equal to x for any $x \in \mathbb{R}$.*

Proof. The proof for Algorithm 1 can be obtained by a similar argument, we only give the proof for Algorithm 2. By step 2 of Algorithm 2 (see also the remark on (2.6)), the line search stops when $Q_{i+1} \geq -C$. Noting (2.3) (with $\beta_i = 1$ for $i > 1$) and the definition of Γ_j in Algorithm 2, the condition $Q_{i+1} \geq -C$ is equivalent to

$$Q_i + \|u_{i+1} - u_i\|^2 + \frac{\rho_i^u}{\eta_i} \|Bu_{i+1} - w_{i+1}\|^2 - \frac{\alpha_i}{\eta_i} \|A(u_{i+1} - u_i)\|^2 \geq -C.$$

After rearranging terms, the above relation can be reformulated to

$$\eta_i(C + Q_i) + \eta_i \|u_{i+1} - u_i\|^2 + \rho_i^u \|Bu_{i+1} - w_{i+1}\|^2 - \alpha_i \|A(u_{i+1} - u_i)\|^2 \geq 0,$$

or the equivalent form

$$\eta_i \geq \frac{\alpha_i \|A(u_{i+1} - u_i)\|^2}{\|u_{i+1} - u_i\|^2} - \frac{\rho_i^u \|Bu_{i+1} - w_{i+1}\|^2}{\|u_{i+1} - u_i\|^2} - \frac{\eta_i(C + Q_i)}{\|u_{i+1} - u_i\|^2}.$$

Noting the two negative terms at the right hand side of the above line search stopping criterion, we observe that it is satisfied when $\eta_i \geq \frac{\alpha_i \|A(u_{i+1} - u_i)\|^2}{\|u_{i+1} - u_i\|^2}$. By the definition of η_i , such condition is satisfied after l rounds of line search, as long as

$$\sigma^l \leq \frac{\|A(u_{i+1} - u_i)\|^2}{\eta_{0,i} \|u_{i+1} - u_i\|^2}.$$

Noting that $\|A(u_{i+1} - u_i)\|^2 \leq \|A^T A\| \|u_{i+1} - u_i\|^2$ and that $\eta_{0,i} \geq \eta_{\min}$, we have $l \leq \lceil \log_{\sigma}(\frac{\|A^T A\|}{\eta_{\min}}) \rceil$. □

The following lemma plays a primary role for the convergence analysis of the proposed Algorithms. Since we have $\rho_i^w = \rho_i^u$ for both proposed algorithms, in the following proof ρ_i^w is replaced by ρ_i^u for the purpose of a unified analysis. Throughout this section, we use notations $u_i^e = u_i - u$ and $w_i^e = w_i - w$ for $i \geq 1$. For convenience, we denote $\frac{1}{2} \|Au - f\|^2$ by $H(u)$.

Lemma 3.3. *For all $u \in U$ and all $w \in \mathbb{C}^{d \times n}$, the iterates $\{(u_i^{ag}, w_i^{ag})\}_{i \geq 1}$ generated by Algorithms 1 and 2 satisfy*

$$\begin{aligned}
 & [H(u_{i+1}^{ag}) + \psi(w_{i+1}^{ag}) - H(u) - \psi(w)] \\
 & - (1 - \alpha_i)[(H(u_i^{ag}) + \psi(w_i^{ag}) - H(u) - \psi(w))] \\
 \leq & \alpha_i \left\{ \frac{\eta_i}{2} (\|u_i^e\|^2 - \|u_{i+1}^e\|^2) - \frac{1}{2} (\rho_i^u \|w_i - Bu_{i+1}\|^2 \right. \\
 & + \eta_i \|u_{i+1} - u_i\|^2 - \alpha_i \|A(u_{i+1} - u_i)\|^2) \\
 & + \frac{\rho_i^u}{2} (\|w_i^e\|^2 - \|w_{i+1}^e\|^2) + \frac{1}{2\rho_i^\lambda} (\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2) \\
 & - \frac{\rho_i^u - \rho_i^\lambda}{2} \|w_{i+1} - Bu_{i+1}\|^2 + \rho_i^u \langle w_i - Bu_{i+1}, w - Bu \rangle \\
 & \left. - \langle \lambda_i, w - Bu \rangle \right\}.
 \end{aligned} \tag{3.1}$$

Proof. Since H is differentiable, we have

$$(3.2) \quad H(u_{i+1}^{ag}) = H(u_i^{md}) + \int_0^1 \langle \nabla H(u_i^{md} + \tau(u_{i+1}^{ag} - u_i^{md})), u_{i+1}^{ag} - u_i^{md} \rangle d\tau.$$

Applying the definition of $H(u)$ and u_{i+1}^{ag} to the above equation, and observing the relationship $u_{i+1}^{ag} - u_i^{md} = \alpha_i(u_{i+1} - u_i)$, we have

(3.3)

$$\begin{aligned}
H(u_{i+1}^{ag}) &= H(u_i^{md}) + \int_0^1 \langle A^T A(u_i^{md} + \tau(u_{i+1}^{ag} - u_i^{md})) - A^T f, u_{i+1}^{ag} - u_i^{md} \rangle d\tau \\
&= H(u_i^{md}) + \int_0^1 \langle A^T (Au_i^{md} - f), u_{i+1}^{ag} - u_i^{md} \rangle d\tau \\
&\quad + \int_0^1 \tau \|A(u_{i+1}^{ag} - u_i^{md})\|^2 d\tau \\
&= H(u_i^{md}) + \langle \nabla H(u_i^{md}), u_{i+1}^{ag} - u_i^{md} \rangle + \frac{1}{2} \|A(u_{i+1}^{ag} - u_i^{md})\|^2 \\
&= H(u_i^{md}) + (1 - \alpha_i) \langle \nabla H(u_i^{md}), u_i^{ag} - u_i^{md} \rangle \\
&\quad + \alpha_i \langle \nabla H(u_i^{md}), u_{i+1} - u_i^{md} \rangle + \frac{\alpha_i^2}{2} \|A(u_{i+1} - u_i)\|^2 \\
&= (1 - \alpha_i) (H(u_i^{md}) + \langle \nabla H(u_i^{md}), u_i^{ag} - u_i^{md} \rangle) + \alpha_i (H(u_i^{md}) \\
&\quad + \langle \nabla H(u_i^{md}), u_{i+1} - u_i^{md} \rangle) + \frac{\alpha_i^2}{2} \|A(u_{i+1} - u_i)\|^2.
\end{aligned}$$

Here by the convexity of $H(u)$ and (3.3), we have

$$\begin{aligned}
H(u_{i+1}^{ag}) &= (1 - \alpha_i) (H(u_i^{md}) + \langle \nabla H(u_i^{md}), u_i^{ag} - u_i^{md} \rangle) \\
&\quad + \alpha_i (H(u_i^{md}) + \langle \nabla H(u_i^{md}), u - u_i^{md} \rangle) \\
&\quad + \alpha_i \langle \nabla H(u_i^{md}), u_{i+1} - u \rangle + \frac{\alpha_i^2}{2} \|A(u_{i+1} - u_i)\|^2 \\
(3.4) \quad &\leq (1 - \alpha_i) H(u_i^{ag}) + \alpha_i H(u) \\
&\quad + \alpha_i \langle \nabla H(u_i^{md}), u_{i+1} - u \rangle + \frac{\alpha_i^2}{2} \|A(u_{i+1} - u_i)\|^2, \forall u \in U.
\end{aligned}$$

By (3.4) and the convexity of ψ , we can calculate the following difference.

$$\begin{aligned}
&[H(u_{i+1}^{ag}) + \psi(w_{i+1}^{ag}) - H(u) - \psi(w)] \\
&\quad - (1 - \alpha_i) [(H(u_i^{ag}) + \psi(w_i^{ag}) - H(u) - \psi(w))] \\
&= (H(u_{i+1}^{ag}) + \psi(w_{i+1}^{ag}) - H(u) - \psi(w)) \\
&\quad - (1 - \alpha_i) (H(u_i^{ag}) + \psi(w_i^{ag}) - H(u) - \psi(w)) \\
&= (H(u_{i+1}^{ag}) - (1 - \alpha_i) H(u_i^{ag}) - \alpha_i H(u)) \\
(3.5) \quad &\quad + (\psi(w_{i+1}^{ag}) - (1 - \alpha_i) \psi(w_i^{ag}) - \alpha_i \psi(w)) \\
&\leq \alpha_i \left\{ \langle \nabla H(u_i^{md}), u_{i+1}^e \rangle + \frac{\alpha_i}{2} \|A(u_{i+1} - u_i)\|^2 \right. \\
&\quad \left. + (\psi(w_{i+1}) - \psi(w)) \right\} \\
&\leq \alpha_i \left\{ \langle \nabla H(u_i^{md}), u_{i+1}^e \rangle + \frac{\alpha_i}{2} \|A(u_{i+1} - u_i)\|^2 + \langle s_{i+1}, w_{i+1}^e \rangle \right\}.
\end{aligned}$$

where $s_{i+1} \in \partial\psi(w_{i+1})$. On the other hand, by the first-order optimality conditions for the sequence $(u_{i+1}, w_{i+1}, \lambda_{i+1})$ generated by Algorithm 1 and 2, for all $u \in U$

and $w \in \mathbb{C}^{d \times n}$, we have

$$(3.6) \quad \begin{cases} \langle \nabla H(u_i^{md}), u_{i+1}^e \rangle + \eta_i \langle u_{i+1} - u_i, u_{i+1}^e \rangle \\ + \rho_i^u \langle Bu_{i+1} - w_i + (\rho_i^u)^{-1} \lambda_i, Bu_{i+1}^e \rangle \leq 0 \\ \langle s_{i+1}, w_{i+1}^e \rangle - \langle \lambda_i, w_{i+1}^e \rangle + \rho_i^u \langle w_{i+1} - Bu_{i+1}, w_{i+1}^e \rangle \leq 0 \\ \lambda_{i+1} = \lambda_i - \rho_i^\lambda (w_{i+1} - Bu_{i+1}), \end{cases}$$

Using the relationship $2\langle a - b, a - c \rangle = -\|b - c\|^2 + \|a - c\|^2 + \|a - b\|^2$ and the definition of u_{i+1}^e, w_{i+1}^e , the above equation can be rewritten as

$$(3.7) \quad \begin{cases} \frac{\eta_i}{2} (\|u_{i+1} - u_i\|^2 + \|u_{i+1}^e\|^2 - \|u_i^e\|^2) + \langle \nabla H(u_i^{md}), u_{i+1}^e \rangle \\ + \rho_i^u \langle Bu_{i+1}^e - w_i^e, Bu_{i+1}^e \rangle - \rho_i^u \langle w - Bu, Bu_{i+1}^e \rangle + \langle \lambda_i, Bu_{i+1}^e \rangle \leq 0 \\ \langle s_{i+1}, w_{i+1}^e \rangle + \rho_i^u \langle w_{i+1}^e - Bu_{i+1}^e, w_{i+1}^e \rangle - \langle \lambda_i, w_{i+1}^e \rangle \\ + \rho_i^u \langle w - Bu, w_{i+1}^e \rangle \leq 0 \\ \lambda_{i+1} = \lambda_i - \rho_i^\lambda (w_{i+1} - Bu_{i+1}). \end{cases}$$

Substituting (3.7) to (3.5), we have

$$(3.8) \quad \begin{aligned} & [H(u_{i+1}^{ag}) + \psi(w_{i+1}^{ag}) - H(u) - \psi(w)] \\ & - (1 - \alpha_i)[(H(u_i^{ag}) + \psi(w_i^{ag}) - H(u) - \psi(w))] \\ & \leq \alpha_i \left\{ \frac{\eta_i}{2} (\|u_i^e\|^2 - \|u_{i+1}^e\|^2) - \frac{1}{2} (\eta_i \|u_{i+1} - u_i\|^2 \right. \\ & \quad - \alpha_i \|A(u_{i+1} - u_i)\|^2) + \underbrace{\rho_i^u \langle w_i^e - Bu_{i+1}^e, Bu_{i+1}^e \rangle}_{(I)} \\ & \quad + \underbrace{\langle w_{i+1}^e - Bu_{i+1}^e, \lambda_i \rangle}_{(II)} - \underbrace{\rho_i^u \langle w_{i+1}^e - Bu_{i+1}^e, w_{i+1}^e \rangle}_{(III)} \\ & \quad \left. + \underbrace{\rho_i^u \langle w - Bu, Bu_{i+1}^e - w_{i+1}^e \rangle}_{(IV)} \right\}. \end{aligned}$$

To give a further estimation of (3.8), next we focus on estimating terms (I)-(IV).

$$\begin{aligned} (I) &= \rho_i^u \langle Bu_{i+1}^e, w_i^e \rangle - \rho_i^u \|Bu_{i+1}^e\|^2 \\ &= -\frac{\rho_i^u}{2} \|w_i^e - Bu_{i+1}^e\|^2 + \frac{\rho_i^u}{2} \|w_i^e\|^2 - \frac{\rho_i^u}{2} \|Bu_{i+1}^e\|^2 \\ &= -\frac{\rho_i^u}{2} \|w_i - Bu_{i+1}\|^2 + \frac{\rho_i^u}{2} \|w_i^e\|^2 - \frac{\rho_i^u}{2} \|Bu_{i+1}^e\|^2 \\ &\quad + \rho_i^u \langle w_i - Bu_{i+1}, w - Bu \rangle - \frac{\rho_i^u}{2} \|w - Bu\|^2, \\ (II) &= \langle \lambda_i, w_{i+1} - Bu_{i+1} \rangle - \langle \lambda_i, w - Bu \rangle \\ &= \frac{1}{\rho_i^\lambda} \langle \lambda_i, \lambda_i - \lambda_{i+1} \rangle - \langle \lambda_i, w - Bu \rangle \\ &= \frac{1}{2\rho_i^\lambda} (\|\lambda_i - \lambda_{i+1}\|^2 - \|\lambda_{i+1}\|^2 + \|\lambda_i\|^2) - \langle \lambda_i, w - Bu \rangle \\ &= \frac{\rho_i^\lambda}{2} \|w_{i+1} - Bu_{i+1}\|^2 + \frac{1}{2\rho_i^\lambda} (\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2) - \langle \lambda_i, w - Bu \rangle, \end{aligned}$$

$$\begin{aligned}
(\text{III}) &= -\rho_i^u \|w_{i+1}^e\|^2 + \rho_i^u \langle Bu_{i+1}^e, w_{i+1}^e \rangle \\
&= -\frac{\rho_i^u}{2} \|w_{i+1}^e - Bu_{i+1}^e\|^2 - \frac{\rho_i^u}{2} \|w_{i+1}^e\|^2 + \frac{\rho_i^u}{2} \|Bu_{i+1}^e\|^2 \\
&= -\frac{\rho_i^u}{2} \|w_{i+1} - Bu_{i+1}\|^2 - \frac{\rho_i^u}{2} \|w_{i+1}^e\|^2 + \frac{\rho_i^u}{2} \|Bu_{i+1}^e\|^2 \\
&\quad + \rho_i^u \langle w_{i+1} - Bu_{i+1}, w - Bu \rangle - \frac{\rho_i^u}{2} \|w - Bu\|^2,
\end{aligned}$$

$$\text{and } (\text{IV}) = \rho_i^u \langle Bu_{i+1} - w_{i+1}, w - Bu \rangle + \rho_i^u \|w - Bu\|^2.$$

Combining those above terms, we have

$$\begin{aligned}
(\text{I}) + (\text{II}) + (\text{III}) + (\text{IV}) &= -\frac{\rho_i^u}{2} \|w_i - Bu_{i+1}\|^2 + \frac{\rho_i^u}{2} (\|w_i^e\|^2 - \|w_{i+1}^e\|^2) \\
&\quad + \frac{1}{2\rho_i^\lambda} (\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2) - \frac{\rho_i^u - \rho_i^\lambda}{2} \|w_{i+1} - Bu_{i+1}\|^2 \\
&\quad + \rho_i^u \langle w_i - Bu_{i+1}, w - Bu \rangle - \langle \lambda_i, w - Bu \rangle.
\end{aligned}$$

Applying the above summation to (3.8), we obtain (3.1). $\square$

The following lemma presents an important property of the out $(u_{k+1}^{ag}, w_{k+1}^{ag})$ generated by Algorithm 1.

Lemma 3.4. *Suppose that the parameters ρ_i^u, ρ_i^w , and ρ_i^λ in Algorithm 1 satisfy (2.1), the output $(u_{k+1}^{ag}, w_{k+1}^{ag})$ generated by Algorithm 1 satisfies*

$$\begin{aligned}
(3.9) \quad & \frac{1}{\alpha_k^2} [H(u_{k+1}^{ag}) + \psi(w_{k+1}^{ag}) - H(u^*) - \psi(w^*)] \\
& \leq \sum_{i=1}^k \frac{\eta_i}{2\alpha_i} (\|u_i - u^*\|^2 - \|u_{i+1} - u^*\|^2) + \sum_{i=1}^k \frac{\rho}{2\alpha_i} (\|w_i - w^*\|^2 \\
& \quad - \|w_{i+1} - w^*\|^2) + \sum_{i=1}^k \frac{1}{2\rho\alpha_i} (\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2) + \sum_{i=2}^k \frac{(1-\beta_i)C}{2(i-1)^2} + \frac{C}{2k^2}.
\end{aligned}$$

Proof. Dividing both sides of (3.1) by α_i^2 , using (2.1) and step 1 in Algorithm 1, we have

$$\begin{aligned}
(3.10) \quad & \frac{1}{\alpha_i^2} [H(u_{i+1}^{ag}) + \psi(w_{i+1}^{ag}) - H(u) - \psi(w)] \\
& - \frac{(1-\alpha_i)}{\alpha_i^2} [(H(u_i^{ag}) + \psi(w_i^{ag}) - H(u) - \psi(w))] + \frac{(\beta_i Q_i + \Gamma_i)}{2} \\
& \leq \frac{\eta_i}{2\alpha_i} (\|u_i^e\|^2 - \|u_{i+1}^e\|^2) + \frac{\rho}{2\alpha_i} (\|w_i^e\|^2 - \|w_{i+1}^e\|^2) + \frac{1}{2\rho_i^\lambda \alpha_i} (\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2) \\
& \quad + \frac{\rho}{\alpha_i} \langle w_i - Bu_{i+1}, w - Bu \rangle - \frac{1}{\alpha_i} \langle \lambda_i, w - Bu \rangle + \frac{(\beta_k - 1)Q_k}{2} + \frac{Q_k}{2}.
\end{aligned}$$

Since $Q_{i+1} = \beta_i Q_i + \Gamma_i$, $0 \leq \beta_i \leq (1 - \frac{1}{i})^2$, and $Q_i \geq -\frac{C}{(i-1)^2}$, we obtain

$$\begin{aligned}
 (3.11) \quad & \frac{1}{\alpha_i^2} [H(u_{i+1}^{ag}) + \psi(w_{i+1}^{ag}) - H(u) - \psi(w)] - \\
 & - \frac{(1 - \alpha_i)}{\alpha_i^2} [(H(u_i^{ag}) + \psi(w_i^{ag}) - H(u) - \psi(w))] + \frac{Q_{i+1}}{2} \\
 & \leq \frac{\eta_i}{2\alpha_i} (\|u_i^e\|^2 - \|u_{i+1}^e\|^2) + \frac{\rho}{2\alpha_i} (\|w_i^e\|^2 - \|w_{i+1}^e\|^2) + \frac{1}{2\rho_i^\lambda \alpha_i} (\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2) \\
 & + \frac{\rho}{\alpha_i} \langle w_i - Bu_{i+1}, w - Bu \rangle - \frac{1}{\alpha_i} \langle \lambda_i, w - Bu \rangle + \frac{(1 - \beta_i)C}{2(i-1)^2} + \frac{Q_i}{2}.
 \end{aligned}$$

Setting $u = u^*$ and $w = w^*$, by the relationship $\frac{1}{\alpha_i^2} = \frac{1 - \alpha_{i+1}}{\alpha_{i+1}^2}$, $Q_{i+1} \geq -\frac{C}{i^2}$, and $\alpha_1 = 1$, we have (3.9) after summing (3.11) from $i = 1$ to k . $\square$

Now we are ready to prove the accelerated convergence rate of the ABOSVS-I algorithm when the feasible sets U and $V := \text{dom } \psi^*$ are compact, where ψ^* is the convex conjugate of $\psi(\cdot)$.

Theorem 3.5. *The output $(u_{k+1}^{ag}, w_{k+1}^{ag})$ generated by Algorithm 1 satisfies*

$$\begin{aligned}
 (3.12) \quad & H(u_{k+1}^{ag}) + \psi(Bu_{k+1}^{ag}) - H(u^*) - \psi(Bu^*) \\
 & \leq \frac{2\sigma \|A^T A\| D_U^2 + 2C_\kappa + 2C_1}{(k+1)^2} + \frac{8D_V^2}{\rho(k+1)} + \frac{4\rho D_{U,B}^2}{k+1},
 \end{aligned}$$

where C_κ and C_1 , which are independent of k , are finite nonnegative numbers.

Proof. If $\eta_{\min} \geq \|A^T A\|$, by the definition of (2.2), $\eta_{0,i} \equiv \eta_{\min}$, which yields $\Gamma_i > 0$. Thus, there is no line search needed based on the backtracking strategy in steps 1 and 2. We have to exclude this case by setting $\eta_{0,1} < \|A^T A\|$, which can be easily satisfied.

Next we show several important properties of the sequence $\{\alpha_i\}$ for $i \geq 1$. The update of α_{i+1} in step 4 of Algorithm 1 is actually equivalent to solving α_{i+1} from $\frac{1}{\alpha_i^2} = \frac{1 - \alpha_{i+1}}{\alpha_{i+1}^2}$, based on which, we have

$$(3.13) \quad \frac{1}{\alpha_i} = \frac{\sqrt{1 - \alpha_{i+1}}}{\alpha_{i+1}} \leq \frac{1 - \frac{\alpha_{i+1}}{2}}{\alpha_{i+1}} = \frac{1}{\alpha_{i+1}} - \frac{1}{2},$$

$$(3.14) \quad \text{and} \quad \frac{1}{\alpha_i} = \frac{\sqrt{1 - \alpha_{i+1}}}{\alpha_{i+1}} \geq \frac{1 - \alpha_{i+1}}{\alpha_{i+1}} \geq \frac{1}{\alpha_{i+1}} - 1 \quad \text{for } i \geq 1.$$

Summing (3.13) and (3.14) from $i = 1$ to k and by $\alpha_1 = 1$, easily we can derive

$$(3.15) \quad \alpha_i \leq \frac{2}{i+1} \quad \text{and} \quad 0 < \frac{1}{\alpha_{i+1}} - \frac{1}{\alpha_i} \leq 1 \quad \text{for } i \geq 1.$$

In the following proof, we estimate the terms on the right hand side of (3.9). In view of lemma 3.1, there is N_0 such that $\frac{\eta_{i+1}}{\alpha_{i+1}} \leq \frac{\eta_i}{\alpha_i}$ for any $i \geq N_0 + 1$. Then the

first term of the right hand side of (3.9) can be estimated by

$$\begin{aligned}
 & \sum_{i=1}^k \frac{\eta_i}{2\alpha_i} (\|u_i - u^*\|^2 - \|u_{i+1} - u^*\|^2) \\
 &= \frac{\eta_1}{2\alpha_1} \|u_1^e\|^2 - \sum_{i=1}^{k-1} \left(\frac{\eta_i}{2\alpha_i} - \frac{\eta_{i+1}}{2\alpha_{i+1}} \right) \|u_{i+1} - u^*\|^2 - \frac{\eta_k}{2\alpha_k} \|u_{k+1} - u^*\|^2 \\
 (3.16) \quad &= \frac{\eta_1}{2\alpha_1} \|u_1 - u^*\|^2 - \left(\sum_{i=1}^{N_0} + \sum_{i=N_0+1}^{k-1} \right) \left(\frac{\eta_i}{2\alpha_i} - \frac{\eta_{i+1}}{2\alpha_{i+1}} \right) \|u_{i+1} - u^*\|^2 \\
 &\quad - \frac{\eta_k}{2\alpha_k} \|u_{k+1} - u^*\|^2 \\
 &\leq \frac{\eta_1}{2} \|u_1 - u^*\|^2 + \frac{C_\kappa}{2} \leq \frac{\sigma \|A^T A\|}{2} D_U^2 + \frac{C_\kappa}{2},
 \end{aligned}$$

where

$$C_\kappa = \begin{cases} 0 & : \text{if } \sum_{i=1}^{N_0} \left(\frac{\eta_i}{\alpha_i} - \frac{\eta_{i+1}}{\alpha_{i+1}} \right) \|u_{i+1} - u^*\|^2 \geq 0 \\ C_0 & : \text{if } \sum_{i=1}^{N_0} \left(\frac{\eta_i}{\alpha_i} - \frac{\eta_{i+1}}{\alpha_{i+1}} \right) \|u_{i+1} - u^*\|^2 < 0, \end{cases}$$

since U is a compact set, clearly C_κ is a finite nonnegative number.

By the optimality condition of u -subproblem in step 4, we have $\lambda_i - \rho_i^u(w_{i+1} - Bu_{i+1}) \in \partial\psi(w_{i+1}) \subseteq \text{dom}\psi^* = V$, which indicates that

$$\begin{aligned}
 & \|w_{i+1} - w^*\|^2 \leq 2\|w_{i+1} - Bu_{i+1}\|^2 + 2\|Bu_{i+1} - w^*\|^2 \\
 (3.17) \quad & \leq 2\frac{D_V^2}{(\rho_i^u)^2} + 2\|Bu_{i+1} - Bu^*\|^2 \leq 2\frac{D_V^2}{\rho^2} + 2D_{U,B}^2.
 \end{aligned}$$

By the second inequality of (3.15), $\alpha_1 = 1$, and (3.17), we have

$$\begin{aligned}
 & \sum_{i=1}^k \frac{\rho}{2\alpha_i} (\|w_i - w^*\|^2 - \|w_{i+1} - w^*\|^2) \\
 (3.18) \quad &= \frac{\rho}{2\alpha_1} \|w_1 - w^*\|^2 + \frac{\rho}{2} \sum_{i=1}^{k-1} \left(\frac{1}{\alpha_{i+1}} - \frac{1}{\alpha_i} \right) \|w_{i+1} - w^*\|^2 \\
 &\quad - \frac{\rho}{2\alpha_k} \|w_{k+1} - w^*\|^2 \\
 &\leq \frac{\rho}{2} \sum_{i=1}^k \|w_i - w^*\|^2 \leq \frac{\rho k}{2} \left(\frac{2D_V^2}{\rho^2} + 2D_{U,B}^2 \right).
 \end{aligned}$$

Also,

$$\begin{aligned}
 & \sum_{i=1}^k \frac{1}{2\rho\alpha_i} (\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2) \\
 (3.19) \quad &= \frac{1}{2\rho\alpha_1} \|\lambda_1\|^2 + \frac{1}{2\rho} \sum_{i=1}^{k-1} \left(\frac{1}{\alpha_{i+1}} - \frac{1}{\alpha_i} \right) \|\lambda_{i+1}\|^2 - \frac{1}{2\rho\alpha_k} \|\lambda_{k+1}\|^2 \\
 &\leq \frac{1}{2\rho} \sum_{i=1}^k \|\lambda_i\|^2 \leq \frac{k}{2\rho} D_V^2.
 \end{aligned}$$

By (3.16), (3.18), and (3.19), we have

$$(3.20) \quad \begin{aligned} & H(u_{k+1}^{ag}) + \psi(w_{k+1}^{ag}) - H(u^*) - \psi(w^*) \\ & \leq \alpha_k^2 \left(\frac{\sigma \|A^T A\|}{2} D_U^2 + \frac{C_\kappa}{2} + \frac{\rho k}{2} \left(\frac{2D_V^2}{\rho^2} + 2D_{U,B}^2 \right) + \frac{k}{2\rho} D_V^2 + \frac{C_1}{2} \right). \end{aligned}$$

where $C_1 = \sum_{i=2}^k \frac{(1-\beta_i)C}{(i-1)^2} + \frac{C}{k^2}$, which is a finite nonnegative number.

To obtain (3.12), we also need the following estimation:

$$(3.21) \quad \begin{aligned} & -\frac{1}{\alpha_k^2} \langle \lambda, w_{k+1}^{ag} - Bu_{k+1}^{ag} \rangle = -\langle \lambda, \sum_{i=1}^k \frac{1}{\alpha_i} (w_{i+1} - Bu_{i+1}) \rangle = \\ & -\sum_{i=1}^k \frac{1}{\alpha_i} \langle \lambda, w_{i+1} - Bu_{i+1} \rangle = -\sum_{i=1}^k \frac{1}{\rho \alpha_i} \langle \lambda, \lambda_i - \lambda_{i+1} \rangle \leq \\ & \sum_{i=1}^k \frac{1}{2\rho \alpha_i} (\|\lambda_i - \lambda\|^2 - \|\lambda_{i+1} - \lambda\|^2) \leq \frac{k}{2\rho} D_V^2, \forall \lambda \in V, \end{aligned}$$

where the first equality was obtained by summing the sequences of $\{u_{i+1}^{ag} - Bu_{i+1}^{ag}\}$ for $i \geq 1$ and using the same technique in Lemma 3.4 and the last inequality was derived by using the same process as that in (3.19).

By the convexity of $\psi(\cdot)$, for $\forall \lambda \in V$, we have

$$(3.22) \quad \begin{aligned} & H(u_{k+1}^{ag}) + \psi(Bu_{k+1}^{ag}) - H(u^*) - \psi(Bu^*) \\ & \leq H(u_{k+1}^{ag}) + \psi(Bu_{k+1}^{ag}) - H(u^*) - \psi(Bu^*) \\ & \quad + \sup_{\lambda \in V} [\psi(w_{k+1}^{ag}) - \psi(Bu_{k+1}^{ag}) - \langle \lambda, w_{k+1}^{ag} - Bu_{k+1}^{ag} \rangle] \\ & = H(u_{k+1}^{ag}) + \psi(w_{k+1}^{ag}) - H(u^*) - \psi(Bu^*) + \sup_{\lambda \in V} -\langle \lambda, w_{k+1}^{ag} - Bu_{k+1}^{ag} \rangle \end{aligned}$$

By the first inequality of (3.15), combining (3.20), (3.21) and (3.22), we obtain (3.12). $\square$

Next we focus on analyzing the accelerated convergence rate of ABOSVS-II algorithm. First we need to establish an important lemma similarly as lemma 3.4 before giving the accelerated convergence rate of ABOSVS-II.

Lemma 3.6. *Suppose that the parameters ρ_i^u, ρ_i^w , and ρ_i^λ in Algorithm 2 satisfy (2.4) or (2.5), the output $(u_{k+1}^{ag}, w_{k+1}^{ag})$ generated by Algorithm 2 satisfy*

$$(3.23) \quad \begin{aligned} & \frac{1}{\alpha_k \eta_k} [H(u_{k+1}^{ag}) + \psi(w_{k+1}^{ag}) - H(u^*) - \psi(w^*)] \\ & \leq \sum_{i=1}^k \frac{1}{2} (\|u_i - u^*\|^2 - \|u_{i+1} - u^*\|^2) + \sum_{i=1}^k \frac{\rho_i^u}{2\eta_i} (\|w_i - w^*\|^2 \\ & \quad - \|w_{i+1} - w^*\|^2) + \sum_{i=1}^k \frac{1}{2\rho_i^\lambda \eta_i} (\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2) \\ & \quad - \sum_{i=1}^k \frac{(\rho_i^u - \rho_i^\lambda)}{2(\rho_i^\lambda)^2 \eta_i} \|\lambda_{i+1} - \lambda_i\|^2 - \sum_{i=1}^k \frac{\Gamma_i}{2}. \end{aligned}$$

Proof. Dividing by $\alpha_k \eta_k$ from both sides of (3.1), we have

$$\begin{aligned}
 & \frac{1}{\alpha_i \eta_i} [H(u_{i+1}^{ag}) + \psi(w_{i+1}^{ag}) - H(u) - \psi(w)] \\
 & - \frac{(1 - \alpha_i)}{\alpha_i \eta_i} [(H(u_i^{ag}) + \psi(w_i^{ag}) - H(u) - \psi(w))] \\
 (3.24) \quad & \leq \frac{1}{2} (\|u_i^e\|^2 - \|u_{i+1}^e\|^2) + \frac{1}{2\rho_i^\lambda \eta_i} (\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2) \\
 & + \frac{\rho_i^u}{2\eta_i} (\|w_i^e\|^2 - \|w_{i+1}^e\|^2) - \frac{\rho_i^u - \rho_i^\lambda}{2(\rho_i^\lambda)^2 \eta_i} \|\lambda_i - \lambda_{i+1}\|^2 \\
 & + \frac{\rho}{\eta_i} \langle w_i - Bu_{i+1}, w - Bu \rangle - \frac{1}{\eta_i} \langle \lambda_i, w - Bu \rangle - \frac{\Gamma_i}{2}.
 \end{aligned}$$

Setting $u = u^*$ and $w = w^*$, by the relationship $\frac{1}{\alpha_i \eta_i} = \frac{1 - \alpha_{i+1}}{\alpha_{i+1} \eta_{i+1}}$, and $\alpha_1 = 1$, we get (3.23) after summing (3.24) from $i = 1$ to k . $\square$

Now we analyze the accelerated convergence rate of the ABSOVS-II algorithm when the feasible sets U is unbounded. In this case, the total number of iterations has to be fixed in advance and chosen based on a worst-case complexity analysis. Additionally, since ψ is Lipschitz continuous, for all $w \in \mathbb{C}^{d \times n}$, we have $\|\xi\| \leq \Delta$, for $\forall \xi \in \partial\psi(w)$.

Theorem 3.7. Suppose that the parameters ρ_i^u, ρ_i^w , and ρ_i^λ in Algorithm 2 satisfy (2.4), then the output $(u_{k+1}^{ag}, w_{k+1}^{ag})$ generated by Algorithm 2 satisfies

$$(3.25) \quad \|w_K^{ag} - Bu_K^{ag}\| \leq \frac{4\sigma \|A^T A\|}{\eta_{\min} K} \Delta,$$

$$\begin{aligned}
 & H(u_K^{ag}) + \psi(Bu_K^{ag}) - H(u^*) - \psi(Bu^*) \\
 (3.26) \quad & \leq \frac{2\sigma \|A^T A\| (D_{u^*}^2 + C)}{K^2} + \frac{2\sigma \|A^T A\| D_{u^*, B}^2}{\eta_{\min} K} + \frac{4\sigma \|A^T A\|}{\eta_{\min} K} \Delta^2.
 \end{aligned}$$

Proof. Since $\eta_{\min} \leq \eta_{0,1}$, we have $\eta_{\min} \leq \|A^T A\|$. Then, by definition of $\eta_{0,i}$, we get

$$(3.27) \quad \eta_{\min} \leq \eta_{0,i} \leq \|A^T A\|.$$

Denote $\sigma^l \eta_{0,i}$ by η_i' , where l is the number of line searches in step 3 of ABOSVS-II. By $\frac{1}{\alpha_i \eta_i} = \frac{1 - \alpha_{i+1}}{\alpha_{i+1} \eta_{i+1}}$ and the definition of η_i , we have

$$(3.28) \quad \frac{1}{\alpha_i \sqrt{\eta_i'}} = \frac{\sqrt{1 - \alpha_{i+1}}}{\alpha_{i+1} \sqrt{\eta_{i+1}'}} \leq \frac{1 - \frac{1}{2}\alpha_{i+1}}{\alpha_{i+1} \sqrt{\eta_{i+1}'}} \leq \frac{1}{\alpha_{i+1} \sqrt{\eta_{i+1}'}} - \frac{1}{2\eta_{i+1}'}$$
 for $i \geq 1$.

Then, by induction we can get, with $\alpha_1 = 1$,

$$\left(\frac{1}{\sqrt{\eta_1'}} + \frac{1}{2} \sum_{i=2}^k \frac{1}{\sqrt{\eta_i'}} \right)^2 \leq \frac{1}{\alpha_k^2 \eta_k'},$$

which implies

$$(3.29) \quad \alpha_k \eta_k \leq \frac{1}{\left(\frac{1}{\sqrt{\eta_1'}} + \frac{1}{2} \sum_{i=2}^k \frac{1}{\sqrt{\eta_i'}} \right)^2} \leq \frac{4\sigma \|A^T A\|}{(k+1)^2} \text{ for } k \geq 1,$$

where we used (3.27) and the definition of η_i' . This estimate is crucial for obtaining the accelerated rate of convergence. Also, since

$$(3.30) \quad \frac{1}{\alpha_i \sqrt{\eta_i'}} = \frac{\sqrt{1 - \alpha_{i+1}}}{\alpha_{i+1} \sqrt{\eta_{i+1}'}} \geq \frac{1 - \alpha_{i+1}}{\alpha_{i+1} \sqrt{\eta_{i+1}'}} = \frac{1}{\alpha_{i+1} \sqrt{\eta_{i+1}'}} - \frac{1}{\eta_{i+1}'},$$

by induction again, we obtain $\frac{1}{\alpha_k \sqrt{\eta_k'}} \leq \frac{k}{\sqrt{\eta_{\min}}}$, i.e. $\alpha_k \eta_k \geq \frac{\eta_{\min}}{k^2}$, which implies

$$(3.31) \quad \frac{1}{\eta_k} \leq \frac{k}{\eta_{\min}} \quad \text{for } k \geq 1.$$

Next we estimate the terms on the right hand side of (3.23). By the definition of ρ_i^u and ρ_i^λ , and (3.31), it is clear that $\rho_i^u \geq \rho_i^\lambda$ for $i \geq 1$. Thus,

$$(3.32) \quad \sum_{i=1}^k \frac{(\rho_i^u - \rho_i^\lambda)}{2(\rho_i^\lambda)^2 \eta_i} \|\lambda_{i+1} - \lambda_i\|^2 \geq 0.$$

Also, using the definition of ρ_i^u and ρ_i^λ and the non-increasing property of the sequences $\{\frac{\rho_i^u}{2\eta_i}\}$ and $\{\frac{1}{2\rho_i^\lambda \eta_i}\}$ for $i = 1, 2, \dots, k$, we have

$$(3.33) \quad \begin{aligned} & \sum_{i=1}^k \frac{1}{2} (\|u_i - u^*\|^2 - \|u_{i+1} - u^*\|^2) \\ &= \frac{1}{2} \|u_1 - u^*\|^2 - \frac{1}{2} \|u_{k+1} - u^*\|^2 \leq \frac{1}{2} D_{u^*}^2, \end{aligned}$$

$$(3.34) \quad \begin{aligned} & \sum_{i=1}^k \frac{\rho_i^u}{2\eta_i} (\|w_i - w^*\|^2 - \|w_{i+1} - w^*\|^2) \\ &= \frac{\rho_1^u}{2\eta_1} \|w_1 - w^*\|^2 - \sum_{i=1}^{k-1} \left(\frac{\rho_i^u}{2\eta_i} - \frac{\rho_{i+1}^u}{2\eta_{i+1}} \right) \|w_{i+1} - w^*\|^2 - \frac{\rho_k^u}{2\eta_k} \|w_{k+1} - w^*\|^2 \\ &\leq \frac{K}{2\eta_{\min}} \|w_1 - w^*\|^2 = \frac{K}{2\eta_{\min}} \|Bu_1 - Bu^*\|^2 \leq \frac{K}{2\eta_{\min}} D_{u^*, B}^2, \end{aligned}$$

$$(3.35) \quad \begin{aligned} & \text{and } \sum_{i=1}^k \frac{1}{2\rho_i^\lambda \eta_i} (\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2) \\ &= \frac{1}{2\rho_1^\lambda \eta_1} \|\lambda_1\|^2 - \sum_{i=1}^{k-1} \left(\frac{1}{2\rho_i^\lambda \eta_i} - \frac{1}{2\rho_{i+1}^\lambda \eta_{i+1}} \right) \|\lambda_{i+1}\|^2 - \frac{1}{2\rho_k^\lambda \eta_k} \|\lambda_{k+1}\|^2 \\ &= \frac{K}{2\eta_{\min}} (\|\lambda_1\|^2 - \|\lambda_{k+1}\|^2) \leq 0, \end{aligned}$$

By (3.32), (3.33), (3.34), (3.35), and the fact that $-\sum_{i=1}^k \Gamma_i \leq C$, we have

$$(3.36) \quad \begin{aligned} & \frac{1}{\alpha_k \eta_k} [H(u_{k+1}^{ag}) + \psi(w_{k+1}^{ag}) - H(u^*) - \psi(w^*)] \\ &\leq \frac{D_{u^*}^2 + C}{2} + \frac{K D_{u^*, B}^2}{2\eta_{\min}}. \end{aligned}$$

Next we focus on estimating $\|w_{k+1}^{ag} - Bu_{k+1}^{ag}\|$ in order to obtain (3.26). Similarly as (3.21), for $\forall \lambda \in \mathbb{C}^{d \times n}$, we can obtain

$$\begin{aligned} \frac{1}{\alpha_k \eta_k} (w_{k+1}^{ag} - Bu_{k+1}^{ag}) &= \sum_{i=1}^k \frac{1}{\eta_i} (w_{i+1} - Bu_{i+1}) \\ (3.37) \quad &= \sum_{i=1}^k \frac{1}{\eta_i} (w_{i+1} - Bu_{i+1}) = \sum_{i=1}^k \frac{1}{\rho_i^\lambda \eta_i} (\lambda_i - \lambda_{i+1}). \end{aligned}$$

Thus, by $\frac{1}{\alpha_k \eta_k} (w_{k+1}^{ag} - Bu_{k+1}^{ag}) = \sum_{i=1}^k \frac{1}{\rho_i^\lambda \eta_i} (\lambda_i - \lambda_{i+1})$, we have

$$(3.38) \quad w_{k+1}^{ag} - Bu_{k+1}^{ag} = \frac{K \alpha_k \eta_k}{\eta_{\min}} (\lambda_1 - \lambda_{k+1}).$$

By setting $k = K - 1$, we have

$$(3.39) \quad \|w_{k+1}^{ag} - Bu_{k+1}^{ag}\| = \frac{4\sigma \|A^T A\|}{\eta_{\min} K} \|\lambda_{k+1}\|.$$

The only left work is to bound the term $\|\lambda_{k+1}\|$. By the optimality condition of w -subproblem in step 3 of Algorithm 2, we have

$$0 \in \partial\psi(w_{i+1}) - \lambda_i + \rho_i^w (w_{i+1} - Bu_{i+1}) \quad \forall i \geq 1,$$

or

$$\lambda_{i+1} = \lambda_i - \rho_i^w (w_{i+1} - Bu_{i+1}) \in \partial\psi(w_{i+1}) \quad \forall i \geq 1.$$

Therefore we have $\|\lambda_{i+1}\| \leq \Delta$, which leads to

$$(3.40) \quad \|w_{k+1}^{ag} - Bu_{k+1}^{ag}\| = \frac{4\sigma \|A^T A\|}{\eta_{\min} K} \Delta.$$

By the convexity of $\psi(\cdot)$, we have

$$\begin{aligned} (3.41) \quad &H(u_K^{ag}) + \psi(Bu_K^{ag}) - H(u^*) - \psi(w^*) \leq H(u_K^{ag}) + \psi(w_K^{ag}) \\ &+ \langle \xi, Bu_K^{ag} - w_K^{ag} \rangle - H(u^*) - \psi(w^*) \leq H(u_K^{ag}) + \psi(w_K^{ag}) \\ &\|\xi\| \|Bu_K^{ag} - w_K^{ag}\| - H(u^*) - \psi(w^*), \end{aligned}$$

where $\xi \in \partial\psi(Bu_K^{ag})$. Then, by (3.36), (3.40), (3.41), we have (3.26). $\square$

The following theorem gives the accelerated convergence rate of the ABOSVS-II algorithm when the feasible sets U is compact. Since its proof is similar with that for Theorem 3.7, we just present the convergence rate result without proof.

Theorem 3.8. *Suppose that the parameters ρ_i^u, ρ_i^w , and ρ_i^λ in Algorithm 2 satisfy (2.5), then the output $(u_{k+1}^{ag}, w_{k+1}^{ag})$ generated by Algorithm 2 satisfies*

$$\begin{aligned} (3.42) \quad &H(u_{k+1}^{ag}) + \psi(Bu_{k+1}^{ag}) - H(u^*) - \psi(Bu^*) \\ &\leq \frac{2\sigma \|A^T A\| (D_U^2 + C)}{(k+1)^2} + \frac{10\sigma \|A^T A\| D_V^2}{\eta_{\min} (k+1)} + \frac{4\sigma \|A^T A\| D_{U,B}^2}{\eta_{\min} (k+1)}. \end{aligned}$$

4. Numerical Results. In this section, we conduct several experiments on synthetic data and the data from partially parallel imaging (PPI) to examine the performance of the proposed algorithms. We also compare them with several state-of-the-art algorithms. All the algorithms are implemented in MATLAB, R2015a on a computer with a 2.6GHz Intel i5 processor.

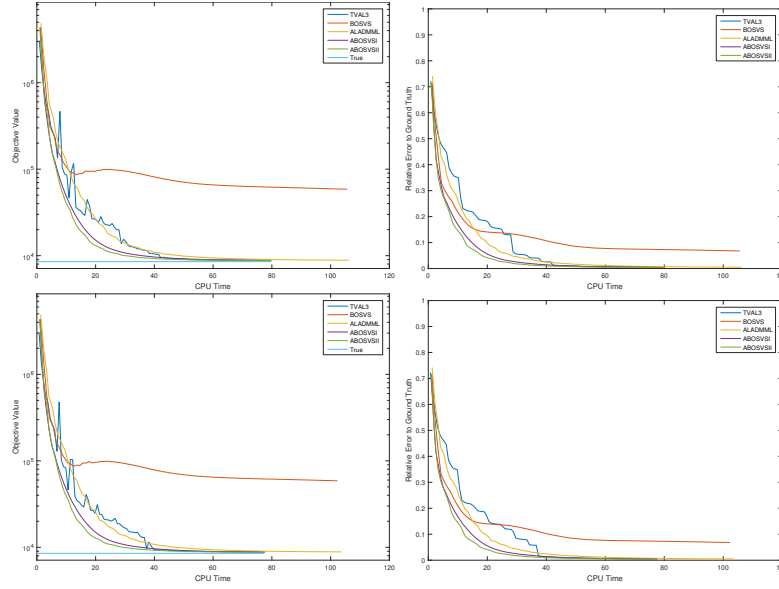


FIGURE 1. The objective function values and relative error vs. CPU time for first instance

4.1. Total-variation based image reconstruction. In this subsection, we present the numerical results on solving the following TV based image reconstruction problem:

$$(4.1) \quad \min_{u \in U} \frac{1}{2} \|Au - f\|_2^2 + \alpha \|u\|_{TV}$$

where $U := \{u \in \mathbb{R}^n : l_* \leq u^{(i)} \leq u_* \text{ for all } i = 1, \dots, n\}$ or $\mathbb{R}^n$, the entries of $A \in \mathbb{R}^{m \times n}$ are randomly generated from a normal distribution $\mathcal{N}(0, 1)$ or a Bernoulli distribution that takes equal probability for the value 1 and -1 , respectively. The measurements f are generated by $f = Au_{true} + \varepsilon$, where u_{true} is the true image, ε is the Gaussian noise with distribution $\mathcal{N}(0, \sigma)$. We apply ABSOVS-I, ABSOVS-II, BOSVS¹, TVAL3² and ALADMML³ to solve (4.1).

We consider two instances of this problem. In the first instance, we set $U := \{u \in \mathbb{R}^n : 0 \leq u^{(i)} \leq 1 \text{ for all } i = 1, \dots, n\}$. The dimension of A is 22500×22500 . The true image u_{true} is a 150×150 Shepp-Logan phantom [24] generated by MATLAB with intensities in $[0, 1]$. Moreover, we set the standard derivation $\sigma = 10^{-2}$. We run 100 iterations to compare the performance of ABSOVS-I, ABSOVS-II, BOSVS, TVAL3 and ALADMML. In the second instance, we have $U := \mathbb{R}^n$. The dimension of A is 11250×22500 . u_{true} is a brain image [4] of dimension 150×150 . The σ is the same as that in first instance and 200 iterations are executed in this case.

The parameters setting for both instances is provided in Table 1. The decreasing of the objective function and the relative error of the reconstruction $\tilde{u}$, defined by $\frac{\|\tilde{u} - u_{true}\|}{\|u_{true}\|}$, are presented in Figure 1 and 2, respectively. In the first instance, we

¹Code is available at <http://people.math.gatech.edu/xye33/software/BOSVS.zip>

²Code is available at <http://www.caam.rice.edu/~optimization/L1/TVAL3>

³AADMML [21] with backtracking techniques. Code is available at <https://github.com/o3y/m.aadmm.2014>

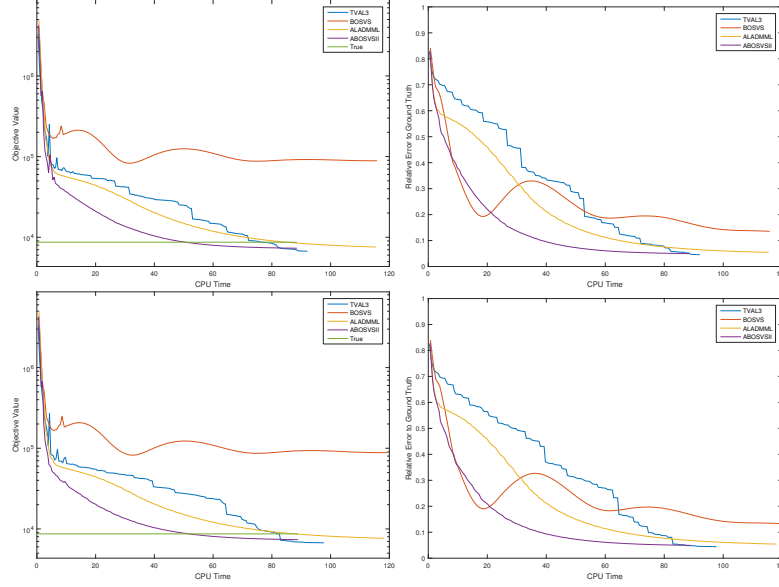


FIGURE 2. The objective function values and relative error vs. CPU time for second instance

TABLE 1. Parameter settings for ABOSVS-I and ABOSVS-II in 4.1. Note that ρ and τ are only used in the first instance

Parameters	α	σ	$\eta_{\min}$	C	β_k	ρ	τ
Values	10	2	$\ A\ ^2/10$	100	$1/k$	1	2

can see that the proposed algorithms perform much better than BOSVS, especially ABOSVS-II. Also, it is evident that ABOSVS-I and ABOSVS-II outperform TVAL3 and ALADMML in this case. In the second instance, ABOSVS-II and ALADMML achieve better performance than BOSVS, and it is also evident that ABOSVS-II outperforms TVAL3 and ALADMML.

4.2. Partially parallel imaging. In this section, we apply the proposed algorithm to two PPI datasets denoted data1 and data2 and compare the performance of ABOSVS-II with ALADMML and TVAL3, and BOSVS. Details of PPI reconstruction problems can be found in [4]. The underlying image can be reconstructed by solving the following optimization problem:

$$(4.2) \quad \min_{u \in \mathbb{C}^n} \left\{ \frac{1}{2} \sum_{l=1}^L \|\mathcal{F}_p(s_l \odot u) - f_l\|_2^2 + \alpha \|u\|_{TV} \right\}$$

where $\mathcal{F}_p$ is the undersampled Fourier transform defined by $\mathcal{F}_p := P\mathcal{F}$, and $\mathcal{F}$ is the Fourier transform, P is a binary matrix representing the undersampling pattern, s_l is the sensitivity map for the l -th channel, and f_l is the measurement. The symbol $\odot$ is the Hadamard product between two vectors. For notation simplicity, let

$$A = [\mathcal{F}_p S_1; \mathcal{F}_p S_2; \dots; \mathcal{F}_p S_L], \quad f = [f_1; f_2; \dots; f_L],$$

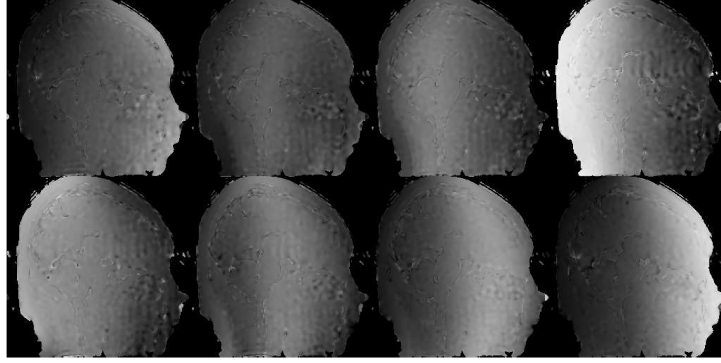


FIGURE 3. Sensitivity maps of data1

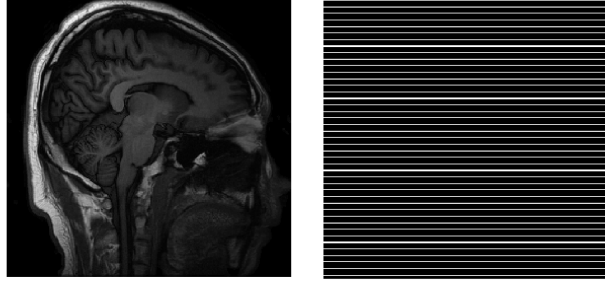


FIGURE 4. Left: True image; Right: Cartesian mask for data 1

where $S_l := \text{diag}(s_l) \in \mathbb{C}^{n \times n}$ is the diagonal matrix with $s_l \in \mathbb{C}^n$ on the diagonal, $l = 1, 2, \dots, L$. Then the above optimization problem can be rewritten as

$$(4.3) \quad \min_{u \in \mathbb{C}^n} \left\{ \frac{1}{2} \|Au - f\|_2^2 + \alpha \|u\|_{TV} \right\}$$

which clearly can be solved by the proposed algorithm. Figure 3 and 6 show the sensitivity maps of data1 and data2, respectively. Figure 4 and 7 show the true images and sampling pattern corresponding to data1 and data2, respectively. The Cartesian mask only samples 18% of the Fourier coefficients. In this experiment, the measurements $\{f_l\}$ are generated by

$$f_l = P(FS_l u_{true} + \varepsilon_l^{re}/\sqrt{2} + \varepsilon_l^{im}/\sqrt{-2}), \quad \text{for } l = 1, \dots, L,$$

where ε_l^{re} and ε_l^{im} are the noise with entries independently generated from distribution $N(0, 10^{-4}\sqrt{n}I_n)$. The size of data1 is $n = 256 \times 256$ and the size of data2 is $n = 512 \times 512$. The parameter settings for ABOSVS-II are given in Table 2.

For data1 and data2, the performance of BOSVS, TVA13, ALADMML, and ABOSVS-II is shown in Table 3 and 4 in terms of the objective function value, relative error, and CPU time in seconds. From Table 3 and 4, we can see that ABOSVS-II, ALADMML converge much faster than BOSVS, and the performance

TABLE 2. Parameter settings for ABOSVS-II in 4.2

Parameters	α	σ	$\eta_{\min}$	C
Values	$10^{-10}n$	2	$\ A\ ^2/5$	100

TABLE 3. Comparison of objective function value, relative error, and CPU time in seconds using data1

Algorithms	Objective value	Relative error	CPU
BOSVS	15.46841	0.0596	2528.0
TVAL3	56.7263	0.050	228.6
ALADMML	15.46810	0.0499	66.5
ABOSVS-II	15.44535	0.0479	49.7

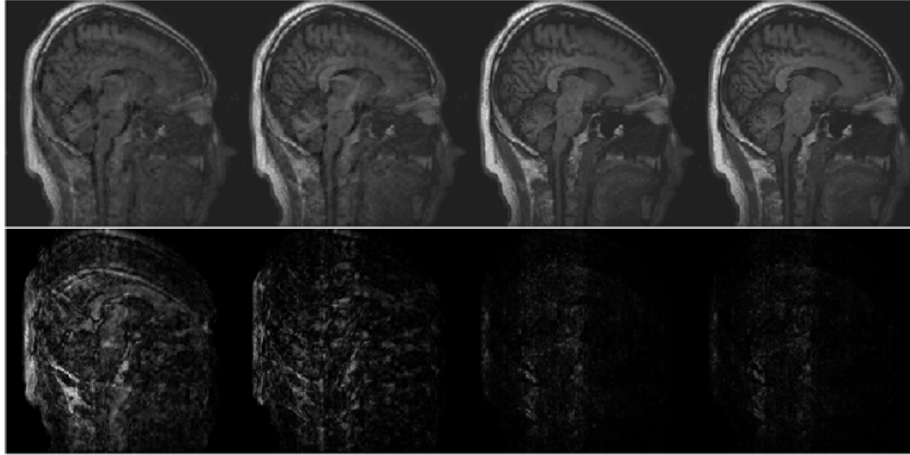


FIGURE 5. Comparison of 400 iterations of BOSVS, TVAL3, ALADMML, and ABOSVS-II in partially parallel image reconstruction (Top) and their differences with true image (Bottom) using data1. From left to right: BOSVS, TVAL3, ALADMML, and ABOSVS-II

of TVAL3 is between ALADMML and BOSVS. ABOSVS-II has the best performance among them.

The reconstructed images by those three algorithms and the differences between the reconstructed images and the true images are shown in Figure 5 and 8 for comparison. Clearly, the images recovered by ABOSVS-II, ALADMML, and TVAL3 have much sharper resolution than that by BOSVS, and the image recovered by ABOSVS-II is slightly better than that by ALADMML and better than that by TVAL3.

5. Conclusions. In this paper, we propose two accelerated Bregman Operator Splitting schemes with backtracking for solving regularized large-scale linear inverse problems. The proposed schemes, named ABOSVS-I and ABOSVS-II, are well

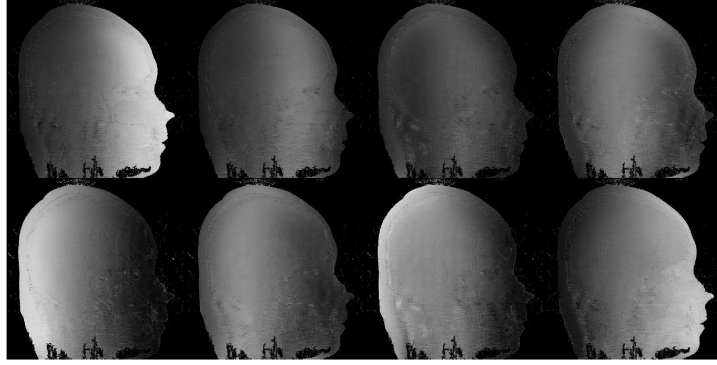


FIGURE 6. Sensitivity maps of data2

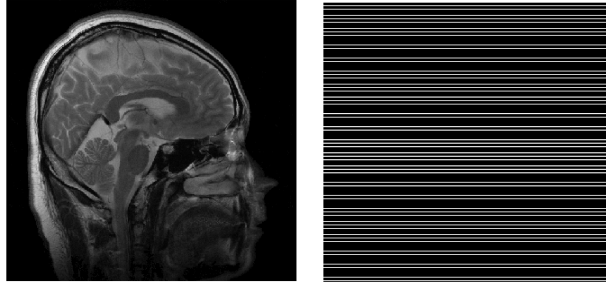


FIGURE 7. Left: True image; Right: Cartesian mask for data 2

TABLE 4. Comparison of objective function value, relative error, and CPU time in seconds using data2

Algorithms	Objective value	Relative error	CPU
BOSVS	5.007e+2	0.0646	7448.4
TVAL3	1.704e+3	0.050	681.2
ALADMML	5.006e+2	0.0499	180.0
ABOSVS-II	4.970e+2	0.0497	144.7

suited for solving total-variation based large-scale linear inverse problems, especially when the matrix in the fidelity term is large, dense and ill-conditioned. ABOSVS-I, which improves the convergence rate of BOSVS in terms of the smooth component in the objective function, can handle the case where the feasible set is bounded. ABOSVS-II can deal with the case where the feasible set is unbounded. For both cases of bounded and unbounded feasible sets, we employ an error between the objective function values at the aggregated iterates and at a solution of problem (1.3) to estimate the rate of convergence instead of using the duality gap technique in AADMM. Both algorithms maintain the same accelerated rate of convergence

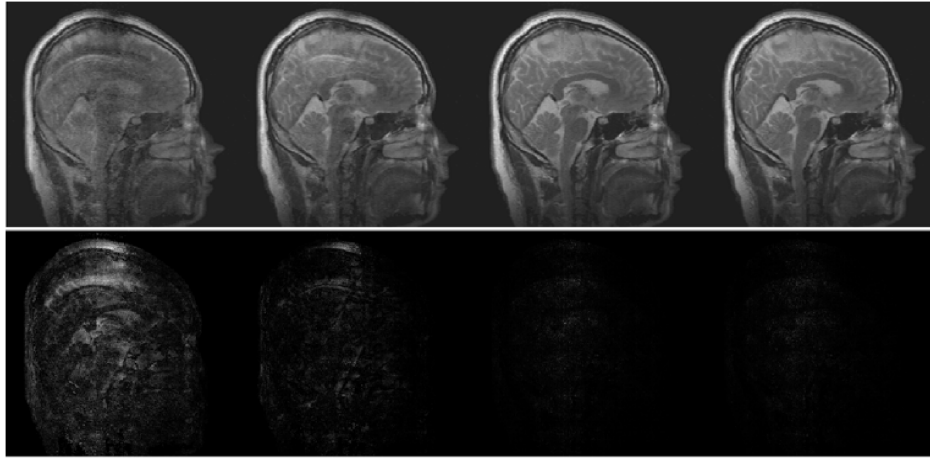


FIGURE 8. Comparison of 400 iterations of BOSVS, TVAL3, AL-ADMML, and ABOSVS-II in partially parallel image reconstruction (Top) and their differences with true image (Bottom) using data2. From left to right: BOSVS, TVAL3, ALADMML, and ABOSVS-II.

as that for AADMM. Experimental results show that the proposed algorithms are promising for large-scale image reconstruction problems.

Acknowledgments. The authors would like to thank Invivo Philips of Gainesville, FL for providing the PPI brain scan datasets.

REFERENCES

- [1] A. Beck and M. Teboulle, Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems, *IEEE Transactions on Image Processing* **18.11** (2009): 2419-2434.
- [2] K. Block, M. Uecker and J. Frahm, Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint, *Magnetic resonance in medicine*, **57** (2007), 1086-1098.
- [3] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends® in Machine Learning*, **3** (2011), 1-122.
- [4] Y. Chen, W. Hager, F. Huang, D. Phan, X. Ye and W. Yin, Fast algorithms for image reconstruction with application to partially parallel MR imaging, *SIAM Journal on Imaging Sciences*, **5** (2012), 90-118.
- [5] Y. Chen, W. Hager, M. Yashtini and X. Ye, Bregman operator splitting with variable stepsize for total variation image reconstruction, *Computational Optimization and Applications*, **54** (2013), 317-342.
- [6] W. Deng and W. Yin, On the global and linear convergence of the generalized alternating direction method of multipliers, *Journal of Scientific Computing*, **66** (2012), 1-28.
- [7] D. Donoho, Compressed sensing, *IEEE Transactions on Information Theory*, **52** (2006), 1289-1306.
- [8] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximation, *Computers & Mathematics with Applications*, **2** (1976), 17-40.
- [9] R. Glowinski and A. Marroco, Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires, *Revue*

- française d'automatique, informatique, recherche opérationnelle. *Analyse numérique*, **9** (1975), 41–76.
- [10] D. Goldfarb, S. Ma and K. Scheinberg, Fast multiple-splitting algorithms for convex optimization, *SIAM Journal on Optimization*, **22** (2012), 533–556.
 - [11] T. Goldstein, B. O'Donoghue, S. Setzer and R. Baraniuk, Fast alternating direction optimization methods, *SIAM Journal on Imaging Sciences*, **7** (2014), 1588–1623.
 - [12] T. Goldstein and S. Osher, The Split Bregman Method for L1-Regularized Problems, *SIAM Journal on Imaging Sciences*, **2** (2009), 323–343.
 - [13] W. Hager, C. Ngo, M. Yashtini and H. Zhang, An alternating direction approximate Newton algorithm for ill-conditioned inverse problems with application to parallel MRI, *Name of the Journal*, **3** (2015), 139–162.
 - [14] W. Hager, M. Yashtini and H. Zhang, An $\mathcal{O}(1/k)$ Convergence Rate for the Variable Stepsize Bregman Operator Splitting Algorithm, *SIAM Journal on Numerical Analysis*, **54** (2016), 1535–1556.
 - [15] M. Hong and Z. Luo, On the linear convergence of the alternating direction method of multipliers, [arXiv:1208.3922](https://arxiv.org/abs/1208.3922).
 - [16] C. Li, W. Yin, H. Jiang and Y. Zhang, An efficient augmented Lagrangian method with applications to total variation minimization, *Computational Optimization and Applications*, **56.3** (2013), 507–530.
 - [17] Q. Liu, J. Luo, S. Wang M. Xiao and M. Ye, An augmented Lagrangian multi-scale dictionary learning algorithm, *EURASIP Journal on Advances in Signal Processing*, **2011** (2011), 1–16.
 - [18] M. Lustig, D. Donoho and J. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging, *Magnetic resonance in medicine*, **58** (2007), 1182–1195.
 - [19] R. Monteiro and B. Svaiter, Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers, *SIAM Journal on Optimization*, **23** (2013), 475–507.
 - [20] Y. Nesterov, Smooth minimization of non-smooth functions, *Mathematical programming*, **103** (2005), 127–152.
 - [21] Y. Ouyang, Y. Chen, G. Lan and E. Pasiliao, An accelerated linearized alternating direction method of multipliers, *SIAM Journal on Imaging Sciences*, **8** (2015), 644–681.
 - [22] K. Pruessmann, M. Weiger, P. Börnert and P. Boesiger, Advances in sensitivity encoding with arbitrary k-space trajectories, *Magnetic Resonance in Medicine*, **46** (2001), 638–651.
 - [23] K. Scheinberg, D. Goldfarb and X. Bai, Fast first-order methods for composite convex optimization with backtracking, *Foundations of Computational Mathematics*, **14** (2014), 389–417.
 - [24] L. Shepp and B. Logan, The Fourier reconstruction of a head section, *IEEE Transactions on Nuclear Science*, **21** (1974), 21–43.
 - [25] Y. Wang, J. Yang, W. Yin and Y. Zhang A new alternating minimization algorithm for total variation image reconstruction, *SIAM Journal on Imaging Sciences*, **1** (2008), 248–272.
 - [26] B. Wohlberg, Efficient Algorithms for Convolutional Sparse Representations, *IEEE Transactions on Image Processing*, **25** (2016), 301–315.
 - [27] J. Yang, Y. Zhang and W. Yin, A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data, *IEEE Journal of Selected Topics in Signal Processing*, **4** (2010), 288–297.
 - [28] X. Ye, Y. Chen and F. Huang, Computational acceleration for MR image reconstruction in partially parallel imaging, *IEEE Transactions on Medical Imaging*, **30** (2011), 1055–1063.
 - [29] X. Zhang, M. Burger, X. Bresson and S. Osher, Bregmanized nonlocal regularization for deconvolution and sparse reconstruction, *SIAM Journal on Imaging Sciences*, **3** (2010), 253–276.
 - [30] X. Zhang, M. Burger and S. Osher, A unified primal-dual algorithm framework based on Bregman iteration, *Journal of Scientific Computing*, **46** (2011), 20–46.
 - [31] M. Zhu and T. Chan, An efficient primal-dual hybrid gradient algorithm for total variation image restoration, *UCLA CAM Report*, (2008), 08–34.
 - [32] M. Zhu, S. Wright and T. Chan, Duality-based algorithms for total-variation-regularized image restoration, *Computational Optimization and Applications*, **47** (2010), 377–400.

Received 1111 2016; revised 0424 2017.