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 Weihong Guo)

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

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

2010 Mathematics Subject Classification. Primary: 90C25, 90C52; Secondary: 65Y20, 68W40.

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, NSF/DMS 1719932 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.

©2017 American Institute of Mathematical Sciences
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, $\psi : \mathbb{C}^{d \times n} \to \mathbb{R}$ is a proper convex and Lipschitz continuous function, $B : \mathbb{C}^{d \times n \times n} \to \mathbb{C}^{d \times n}$ 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 $u = (u_1, ..., u_n)^T$ 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

\[
\psi(Bu) = \alpha \|u\|_{TV} = \alpha \sum_{i=1}^n \|\nabla u_i\|
\]

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 $\alpha$ 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:

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

The augmented Lagrangian function associated with (1.3) is

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

The ADMM minimizes (1.4) by iteratively updating $(u, w, \lambda)$ as follows:

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

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.

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

It has been shown in [30] that the BOS algorithm converges to an optimal solution of (1.1) with the constant stepsize \( \frac{1}{\eta} \leq \frac{1}{\| A^T A \|} \). 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 \( \frac{1}{\| A^T A \|_2} \) 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 \( \eta_k \) 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 \( \rho \) for updating \( u, w \) and \( \lambda \) in (1.5)
(possibly different \( \rho \)'s in \( u, w, \lambda \) 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-Borwen (BB) choice as the initial stepsize, then gradually increase it until a more relaxed line search condition 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 \( \psi \) 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 $F : \mathbb{C}^n \to \mathbb{R}$, $\partial F$ 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 $\rho$ in (1.5) are chosen differently for updating $u$, $w$ and $\lambda$. For convenience, We use $\rho_u^i$, $\rho_w^i$, and $\rho_\lambda^i$ to replace $\rho$ in those three subproblems, respectively. For ABOSVS-I, with assumption of bounded feasible set, $\rho$ can be chosen to be the same constant in (1.5) without affecting the convergence rate, i.e.

$$\rho_u^i = \rho_w^i = \rho_\lambda^i = \rho > 0.$$  

In the proposed algorithms, the initial choice of the nominal step $\eta_i$ is a safeguarded BB choice:

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

where $\eta_i^{BB}$ 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 $u_{i+1}^{ag} = u_i$ and the aggregate points $v_{i+1}^{ag}$, $w_{i+1}^{ag}$, $\lambda_{i+1}^{ag}$ are exactly the iterates $u_{i+1}$, $w_{i+1}$, and $\lambda_{i+1}$, respectively. In this case, ABOSVS-I becomes BOSVS with a minor modification. For ABOSVS-I, the derivation of the accelerated convergence rate relies on the asymptotic monotonicity of $\eta_i/\alpha_i$ for $i = 1, 2, \ldots, \infty$. Whenever $\eta_i/\alpha_i$ is not monotone decreasing, $\eta_{\text{min}}$ is increased by a factor $\tau > 1$ in step 3 for $i = 1, 2, \ldots$. Hence, if the monotonicity of $\eta_i/\alpha_i$ violates continuously, then $\eta_i/\alpha_i$ 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

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

where $\beta_i$ 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 $\rho_u^i$, $\rho_w^i$, and $\rho_\lambda^i$ have to be chosen differently from those for bounded feasible set. More precisely, if the feasible set $U$ is unbounded, we choose

$$\rho_u^i = \rho_w^i = \frac{K\eta_i}{\eta_{\text{min}}} \text{ and } \rho_\lambda^i = \frac{\eta_{\text{min}}}{K\eta_i},$$

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

$$\rho_u^i = \rho_w^i = \frac{\eta_i}{\eta_{\text{min}}} \text{ and } \rho_\lambda^i = \frac{\eta_{\text{min}}}{\eta_i},$$

in ABOSVS-II by jointly updating the acceleration parameter and the stepsize. The scheme of ABOSVS-II is presented in Algorithm 2.
Algorithm 1 Accelerated BOSVS I (ABOSVS-I)

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

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

// Backtracking
1. Set $\eta_i = \alpha_i \eta_{0,i}$, where $\eta_{0,i}$ is from (2.2). Compute
   
   $u_{md}^i = (1 - \alpha_i) u_{ag}^i + \alpha_i u_i$,
   
   $u_{i+1} = \arg\min_{u \in U} \left\{ \langle A^T (A u_{md}^i - f), u \rangle + \frac{\eta_i}{2} \| u - u_i \|^2 + \frac{\rho_i^u}{2} \left\| Bu - w_i + \frac{\lambda_i}{\rho_i^u} \right\|^2 \right\}$,

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

   $Q_{i+1} = \beta_i Q_i + \Gamma_i$, 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 $\eta_{\text{min}}$ by $\tau \eta_{\text{min}}$ if $\eta_i / \alpha_i > \eta_{i-1} / \alpha_{i-1}$, for $i > 1$.

// Updating iterates
4. Compute

   $u_{ag}^{i+1} = (1 - \alpha_i) u_{ag}^i + \alpha_i u_i$,
   
   $w_{i+1} = \arg\min_{w \in C^d \times n} \left\{ \psi(w) + \frac{\rho_w^u}{2} \left\| w - Bu_{i+1} - \frac{\lambda_i}{\rho_w^u} \right\|^2 \right\}$,

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

   $\lambda_{i+1} = \lambda_i - \rho_w^u \left( w_{i+1} - Bu_{i+1} \right)$,

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

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

EndFor

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

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

$$Q_{i+1} < -C,$$ then $\eta_{0,i} := \sigma \eta_{0,i}$, 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 $\beta_i$ such that $(1 - 1/i^2) \leq \beta_i \leq 1$.
Algorithm 2 Accelerated BOSVS II (ABOSVS-II)

Choose $C \geq 0$, $\sigma > 1$, $\eta_{0,1} \geq \eta_{\text{min}}$, $u_{1}$ 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 $i = 1, 2, \ldots, k$,

// Backtracking
1. Set $\eta_{i} = \alpha_{i}\eta_{0,i}$, where $\eta_{0,i}$ is from (2.2). Solve $\alpha_{i}$ from $\frac{1}{\alpha_{i-1} \eta_{i-1}} = \frac{1-\alpha_{i}}{\alpha_{i} \eta_{i}}$ for $i > 1$. Update $\rho_{i}^{u}, \rho_{i}^{w}$ and $\rho_{i}^{\lambda}$ by (2.4) or (2.5). Compute $u_{i}^{md} = (1-\alpha_{i})u_{i}^{ag} + \alpha_{i}u_{i}$.

$u_{i+1} = \arg \min_{u \in C} \left\{ (A^{T}(Au_{i}^{md} - f), u) + \frac{\rho_{i}^{u}}{2} \|u - u_{i}\|^{2} + \frac{\rho_{i}^{w}}{2} \|Bu - w_{i} - \lambda_{i} w_{i}^{ag}\|^{2} \right\}$.

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 $\eta_{\text{min}}$ in step 3 of ABOSVS-I will stop increasing after a finite number of iteration.

Lemma 3.1. The replacement of $\eta_{\text{min}}$ by $\tau \eta_{\text{min}}$ 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].

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 $\lceil \log_{\sigma}(2 \|A^{T}A\|) \rceil$, 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 $\Gamma_{j}$ 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 \( \eta_i \), 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 \( \| A(u_{i+1} - u_i) \|^2 \leq \| AT \| \| u_{i+1} - u_i \|^2 \) and that \( \eta_{0,i} \geq \eta_{\text{min}} \), we have
\[ l \leq \left\lfloor \log_\sigma (\frac{\| AT \|}{\eta_{\text{min}}}) \right\rfloor. \]

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 \( \rho_i^w \) is replaced by \( \rho_i^u \) for the purpose of a unified analysis. Throughout this section, we use notations \( u_i^c = u_i - u \) and \( w_i^c = w_i - w \) 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 \( \{ (u_i^{ag}, u_i^{ag}) \}_{i \geq 1} \) generated by Algorithms 1 and 2 satisfy

\[
[H(u_i^{ag}) + \psi(w_i^{ag}) - H(u) - \psi(w)]
- (1 - \alpha_i)(H(u_i^{ag}) + \psi(u_i^{ag}) - H(u) - \psi(u)) \leq \alpha_i \left\{ \frac{\eta_i}{2}(\| u_i^c \|^2 - \| u_i^{ag} \|^2) - \frac{1}{2}\rho_i^u\| w_i - Bu_{i+1} \|^2 
+ \eta_i\| u_{i+1} - u_i \|^2 - \alpha_i\| A(u_{i+1} - u_i) \|^2 
+ \frac{\rho_i^u}{2}(\| w_i^c \|^2 - \| u_i^{ag} \|^2) + \frac{1}{2\rho_i^u}(\| \lambda_i \|^2 - \| \lambda_{i+1} \|^2) 
- \rho_i^u\| w_{i+1} - Bu_{i+1} \|^2 + \rho_i^u\| w_i - Bu_{i+1}, w - Bu \n- \langle \lambda_i, w - Bu \rangle \right\}
\]

**Proof.** Since \( H \) is differentiable, we have
\[ H(u_{i+1}^{ag}) = H(u_i^{ag}) + \int_0^1 \langle \nabla H(u_i^{ag} + \tau(u_{i+1}^{ag} - u_i^{ag})), u_{i+1}^{ag} - u_i^{ag} \rangle \, d\tau. \]
Applying the definition of \( H(u) \) and \( u_{i+1}^{ag} \) to the above equation, and observing the relationship \( u_{i+1}^{ag} - u_i^{ag} = \alpha_i(u_{i+1} - u_i) \), we have
\[ H(u_{i+1}^{ag}) = H(u_i^{ag}) + \int_0^1 \langle AT(Au_i^{ag} + \tau(u_{i+1}^{ag} - u_i^{ag})), u_{i+1}^{ag} - u_i^{ag} \rangle \, d\tau = H(u_i^{md}) + \int_0^1 \langle A^T(Au_i^{md} - f), u_{i+1}^{ag} - u_i^{ag} \rangle \, d\tau + \int_0^1 \tau\| A(u_{i+1}^{ag} - u_i^{ag}) \|^2 \, d\tau = H(u_i^{md}) + \langle \nabla H(u_i^{md}), u_{i+1}^{ag} - u_i^{ag} \rangle + \frac{1}{2}\| A(u_{i+1}^{ag} - u_i^{ag}) \|^2 \]
\[ H(u_i^{md}) + (1 - \alpha_i)\langle \nabla H(u_i^{md}), u_i^{ag} - u_i^{md} \rangle \\
+ \alpha_i \langle \nabla H(u_i^{md}), u_{i+1} - u_i^{md} \rangle + \frac{\alpha_i^2}{2} \| A(u_{i+1} - u_i) \|^2 \\
\leq (1 - \alpha_i)H(u_i^{md}) + \alpha_i H(u) \\
+ \alpha_i \langle \nabla H(u_i^{md}), u_{i+1} - u_i \rangle + \frac{\alpha_i^2}{2} \| A(u_{i+1} - u_i) \|^2, \forall u \in U. \]

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

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

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

\[ \begin{align*}
\langle \nabla H(u_i^{md}), u_i^{ag} \rangle + \eta \langle u_{i+1} - u_i, u_{i+1} \rangle \\
+ \rho \langle B(u_{i+1} - u_i), \rho_i^{(2)}(u_{i+1} - u_i) \rangle \leq 0 \\
\langle s_{i+1}, w_i^{ag} \rangle + \langle \lambda_i, w_i^{ag} \rangle + \rho_i^{(2)}(w_i^{ag} - Bu_{i+1}, w_i^{ag}) \leq 0 \\
\lambda_{i+1} = \lambda_i - \rho_i^{(2)}(w_i^{ag} - Bu_{i+1}), \\
\end{align*} \]

Using the relationship \( 2(a - b, a - c) = -\|b - c\|^2 + \|a - c\|^2 + \|a - b\|^2 \) and the definition of \( u_i^{ag}, w_i^{ag} \), the above equation can be rewritten as
To give a further estimation of (3.8), next we focus on estimating terms (I)-(IV).

\[
\begin{aligned}
\{ & \frac{\eta_i}{2} (\|u_{i+1} - u_i\|^2 + \|u_{i+1}^c\|^2 - \|u_i^c\|^2) + \langle \nabla H(u_i^{md}, u_{i+1}^c) \\
& + \rho_i^u (Bu_{i+1}^c - w_{i+1}^c, Bu_{i+1}^c) - \rho_i^v (w - Bu, Bu_{i+1}^c) + \langle \lambda_i, Bu_{i+1}^c \rangle \leq 0 \\
\} \\
\{ & (s_{i+1}, w_{i+1}^c) + \rho_i^u \langle w_{i+1}^c - Bu_{i+1}^c, w_{i+1}^c \rangle - \langle \lambda_i, w_{i+1}^c \rangle \\
& + \rho_i^v (w - Bu, w_{i+1}^c) \leq 0 \\
\}
\]  

(3.7)

Substituting (3.7) to (3.5), we have

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

(3.8)

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

(I) = \rho_i^u \langle Bu_{i+1}^c, w_{i+1}^c \rangle - \rho_i^u \|Bu_{i+1}^c\|^2

\[
\begin{aligned}
& = -\frac{\rho_i^u}{2} \|w_{i+1}^c - Bu_{i+1}^c\|^2 + \frac{\rho_i^u}{2} \|w_{i+1}^c\|^2 - \frac{\rho_i^u}{2} \|Bu_{i+1}^c\|^2 \\
& = -\frac{\rho_i^u}{2} \|w_i - Bu_{i+1}\|^2 + \frac{\rho_i^u}{2} \|w_{i+1}^c\|^2 - \frac{\rho_i^u}{2} \|Bu_{i+1}^c\|^2 \\
& + \rho_i^v \langle w_i - Bu_{i+1}, w - Bu \rangle - \frac{\rho_i^v}{2} \|w - Bu\|^2.
\end{aligned}
\]

(II) = \langle \lambda_i, w_{i+1} - Bu_{i+1} \rangle - \langle \lambda_i, w - Bu \rangle

\[
\begin{aligned}
& = \frac{1}{\rho_i^u} \langle \lambda_i, \lambda_i - \lambda_{i+1} \rangle - \langle \lambda_i, w - Bu \rangle \\
& = \frac{1}{2\rho_i^u} (\|\lambda_i - \lambda_{i+1}\|^2 - \|\lambda_{i+1}\|^2 - \|\lambda_{i}\|^2) - \langle \lambda_i, w - Bu \rangle \\
& = \frac{\rho_i^\lambda}{2} \|w_{i+1} - Bu_{i+1}\|^2 + \frac{\rho_i^\lambda}{2} (\|\lambda_{i}\|^2 - \|\lambda_{i+1}\|^2) - \langle \lambda_i, w - Bu \rangle \\
\end{aligned}
\]

(III) = \rho_i^u \|w_{i+1}^c\|^2 + \rho_i^v \langle Bu_{i+1}^c, w_{i+1}^c \rangle

\[
\begin{aligned}
& = \frac{\rho_i^u}{2} \|w_{i+1}^c - Bu_{i+1}^c\|^2 - \frac{\rho_i^u}{2} \|w_{i+1}^c\|^2 + \frac{\rho_i^u}{2} \|Bu_{i+1}^c\|^2 \\
& = \frac{\rho_i^u}{2} \|w_{i+1} - Bu_{i+1}\|^2 - \frac{\rho_i^u}{2} \|w_{i+1}^c\|^2 + \frac{\rho_i^u}{2} \|Bu_{i+1}^c\|^2 \\
& + \rho_i^v \langle w_{i+1} - Bu_{i+1}, w - Bu \rangle - \frac{\rho_i^v}{2} \|w - Bu\|^2, \\
\end{aligned}
\]

and (IV) = \rho_i^u \langle Bu_{i+1} - w_{i+1}, w - Bu \rangle + \rho_i^v \|w - Bu\|^2.
Combining those above terms, we have

\[
(I) + (II) + (III) + (IV) = -\frac{\rho_i^2}{2}\|w_i - Bu_{i+1}\|^2 + \frac{\rho_i^2}{2}(\|w_i^\alpha\|^2 - \|w_{i+1}^\alpha\|^2)
\]

\[+ \frac{1}{2\rho_i^2}(\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2) - \frac{\rho_i^2 - \rho_i^\lambda}{2}\|w_{i+1} - Bu_{i+1}\|^2
\]

\[+ \rho_i^\lambda \langle w_i - Bu_{i+1}, w - Bu \rangle - \langle \lambda_i, w - Bu \rangle.
\]

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

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

**Lemma 3.4.** Suppose that the parameters \(\rho_i^e, \rho_i^w, \) and \(\rho_i^\lambda\) in Algorithm 1 satisfy (2.1), the output \((u_{k+1}^{ag}, w_{k+1}^{ag})\) generated by Algorithm 1 satisfies

\[
\frac{1}{\alpha_i^2}[H(u_{k+1}^{ag}) + \psi(w_{k+1}^{ag}) - H(u) - \psi(w)]
\]

\[
\leq \sum_{i=1}^{k} \frac{\eta_i}{2\alpha_i}(\|u_i - u^*\|^2 - \|u_{i+1} - u^*\|^2) + \sum_{i=1}^{k} \rho_i(\|w_i - w^*\|^2
\]

\[+ \|u_{i+1} - u^*\|^2) + \sum_{i=1}^{k} \frac{1}{2\rho_i^2}(\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2) + \sum_{i=2}^{k} (1 - \beta_i)C_a + \frac{C}{2k^2}.
\]

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

\[
\frac{1}{\alpha_i^2}[H(u_{k+1}^{ag}) + \psi(w_{k+1}^{ag}) - H(u) - \psi(w)]
\]

\[= \frac{(1 - \alpha_i)}{\alpha_i^2}[(H(u_{k+1}^{ag}) + \psi(w_{k+1}^{ag}) - H(u) - \psi(w))] + \frac{(\beta_i Q_i + \Gamma_i)}{2}
\]

\[\leq \frac{\eta_i}{2\alpha_i}(\|u_i^\alpha\|^2 - \|u_{i+1}^\alpha\|^2) + \frac{\rho_i}{2\alpha_i}(\|w_i^\alpha\|^2 - \|w_{i+1}^\alpha\|^2)
\]

\[+ \frac{1}{2\rho_i^2}(\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2)
\]

\[+ \frac{\rho_i}{\alpha_i} \langle w_i - Bu_{i+1}, w - Bu \rangle - \frac{1}{\alpha_i} \langle \lambda_i, w - Bu \rangle + \frac{(\beta_k - 1)C_k}{2} + \frac{Q_k}{2}.
\]

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

\[
\frac{1}{\alpha_i^2}[H(u_{k+1}^{ag}) + \psi(w_{k+1}^{ag}) - H(u) - \psi(w)] -
\]

\[= \frac{(1 - \alpha_i)}{\alpha_i^2}[(H(u_{k+1}^{ag}) + \psi(w_{k+1}^{ag}) - H(u) - \psi(w))] + \frac{Q_{i+1}}{2}
\]

\[\leq \frac{\eta_i}{2\alpha_i}(\|u_i^\alpha\|^2 - \|u_{i+1}^\alpha\|^2) + \frac{\rho_i}{2\alpha_i}(\|w_i^\alpha\|^2 - \|w_{i+1}^\alpha\|^2)
\]

\[+ \frac{1}{2\rho_i^2}(\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2)
\]

\[+ \frac{\rho_i}{\alpha_i} \langle w_i - Bu_{i+1}, w - Bu \rangle - \frac{1}{\alpha_i} \langle \lambda_i, w - Bu \rangle + \frac{(1 - \beta_i)C}{2(i-1)^2} + \frac{Q_i}{2}.
\]

Setting \(u = u^*\) and \(w = w^*\), by the relationship \(\frac{1}{\alpha_i} = \frac{1 - \alpha_i - 1}{\alpha_i^{ag}}\), \(Q_{i+1} \geq -\frac{C}{i^2}\), and \(\alpha_1 = 1\), we have (3.9) after summing (3.11) from \(i = 1\) to \(k\). \(\blacksquare\)
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 $\psi^*$ is the convex conjugate of $\psi(\cdot)$.

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

$$
H(u_{k+1}^a) + \psi(Bu_{k+1}^a) - H(u^*) - \psi(Bu^*)
\leq 2\sigma A^T A D_u^2 + 2C_\kappa + 2C_1 + \frac{8D_u^2}{(k+1)^2} + \frac{4\rho D_{U,B}^2}{k+1},
$$

where $C_\kappa$ and $C_1$, which are independent of $k$, are finite nonnegative numbers.

**Proof.** If $\eta_{\text{min}} \geq \|A^T A\|$, by the definition of (2.2), $\eta_{0,i} \equiv \eta_{\text{min}}$, 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 $\eta_{0,i} < \|A^T A\|$, which can be easily satisfied.

Next we show several important properties of the sequence $\{\alpha_i\}$ for $i \geq 1$. The update of $\alpha_{i+1}$ in step 4 of Algorithm 1 is actually equivalent to solving $\alpha_{i+1}$ from

$$
\frac{1}{\alpha_i} = 1 - \frac{\alpha_{i+1}}{\alpha_{i+1}},
$$

based on which, we have

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

and

$$
\frac{1}{\alpha_i} = \frac{\sqrt{1 - \alpha_{i+1}}}{\alpha_{i+1}} \geq 1 - \frac{\alpha_{i+1}}{\alpha_{i+1}} \geq \frac{1}{\alpha_{i+1}} - 1 \text{ for } i \geq 1.
$$

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

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

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

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

where

$$
C_\kappa = \begin{cases} 0 & \text{if } \sum_{i=1}^{N_0} \left( \frac{\eta_i}{\alpha_i} - \frac{\eta_{i+1}}{\alpha_{i+1}} \right) \|u_{i+1} - u^*\|^2 \geq 0, \\ C_0 & \text{if } \sum_{i=1}^{N_0} \left( \frac{\eta_i}{\alpha_i} - \frac{\eta_{i+1}}{\alpha_{i+1}} \right) \|u_{i+1} - u^*\|^2 < 0, \end{cases}
$$
By (3.16), (3.18), and (3.19), we have

By the second inequality of (3.15), $u^*$ is a compact set, clearly for $i \in \{1, 2\}$ 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).


\begin{align*}
\|w_{i+1} - w^*\|^2 &\leq 2\|w_{i+1} - Bu_{i+1}\|^2 + 2\|Bu_{i+1} - w^*\|^2 \\
&\leq 2\frac{D^2_U}{\rho^2} + 2\|Bu_{i+1} - Bu^*\|^2 \leq \frac{2D^2_U}{\rho^2} + 2D^2_{U,B}.
\end{align*}

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

\begin{align*}
\sum_{i=1}^{k} \frac{\rho}{2\alpha_i} (\|w_i - w^*\|^2 - \|w_{i+1} - w^*\|^2) \\
= \frac{\rho}{2\alpha_1} \|w_1 - w^*\|^2 + \frac{\rho}{2} \sum_{i=1}^{k-1} \left( \frac{1}{\alpha_i} - \frac{1}{\alpha_{i+1}} \right) \|w_{i+1} - w^*\|^2 \\
- \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^2_U}{\rho^2} + 2D^2_{U,B} \right).
\end{align*}

Also,

\begin{align*}
\sum_{i=1}^{k} \frac{1}{2\rho\alpha_i} (\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2) \\
= \frac{1}{2\rho\alpha_1} \|\lambda_1\|^2 + \frac{1}{2\rho} \sum_{i=1}^{k-1} \left( \frac{1}{\alpha_i} - \frac{1}{\alpha_{i+1}} \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^2_V.
\end{align*}

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

\begin{align*}
H(u^{ag}_{k+1}) + \psi(u^{ag}_{k+1}) - Bu^* - \psi(w^*) \\
\leq \alpha_k^2 \frac{\sigma^2}{2} \left( \frac{1}{\alpha_k} \left| \frac{A^T A}{\alpha_k} \right| + \frac{C_1}{2} \right) + \frac{\rho k}{2} \left( \frac{2D^2_U}{\rho^2} + 2D^2_{U,B} \right) + \frac{k}{2\rho} D^2_V + C_1.
\end{align*}

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

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

\begin{align*}
- \frac{1}{\alpha_k} \langle \lambda, Bu^{ag}_{k+1} - Bu^* \rangle &\leq - \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 &\leq - \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^2_V, \forall \lambda \in V,
\end{align*}

where the first equality was obtained by summing the sequences of $\{u^{ag}_{i+1} - Bu^*\}$ 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
\[
H(u_{k+1}^a) + \psi(Bu_{k+1}^a) - H(u^*) - \psi(Bu^*)
\leq H(u_{k+1}) + \psi(Bu_{k+1}) - H(u^*) - \psi(Bu^*)
+ \sup_{\lambda \in V} [\psi(w_{k+1}^a) - \psi(Bu_{k+1}^a) - (\lambda, w_{k+1}^a - Bu_{k+1}^a)]
\]
\[
= H(u_{k+1}) + \psi(u_{k+1}^a) - H(u^*) - \psi(Bu^*) + \sup_{\lambda \in V} (\lambda, w_{k+1}^a - Bu_{k+1}^a)
\]

(3.22)

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

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 $\rho^u_\lambda$, $\rho^\psi_\lambda$, and $\rho^\psi_\lambda$ in Algorithm 2 satisfy (2.4) or (2.5), the output $(u_{k+1}^a, w_{k+1}^a)$ generated by Algorithm 2 satisfy

\[
\frac{1}{\alpha_k \eta_k}[H(u_{k+1}^a) + \psi(w_{k+1}^a) - H(u^*) - \psi(w^*)]
\leq \sum_{i=1}^k \frac{1}{2} \|u_i - u^*\|^2 - \|u_{i+1} - u^*\|^2 + \sum_{i=1}^k \frac{\rho^u_{\lambda_i}}{2\eta_i} (\|w_i - w^*\|^2)
\leq \|w_{i+1} - w^*\|^2 + \sum_{i=1}^k \frac{1}{2\rho^u_{\lambda_i}} (\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2)
\]

(3.23)

\[
- \sum_{i=1}^k \frac{(\rho^u_{\lambda_i} - \rho^\psi_{\lambda_i})}{2(\rho^\psi_{\lambda_i})^2\eta_i} \|\lambda_i - \lambda_i\|^2 - \sum_{i=1}^k \frac{\Gamma_i}{2}.
\]

**Proof.** Dividing by $\alpha_k \eta_k$ from both sides of (3.1), we have
\[
\frac{1}{\alpha_k \eta_i}[H(u_{i+1}^a) + \psi(w_{i+1}^a) - H(u) - \psi(w)]
- \frac{1}{\alpha_i \eta_i} [(H(u_i^a) + \psi(w_i^a) - H(u) + \psi(w))]
\leq \frac{1}{2} (\|u_i^a\|^2 - \|u_{i+1}^c\|^2 + \frac{1}{2\rho^u_{\lambda_i}} (\|\lambda_i\|^2 - \|\lambda_{i+1}\|^2)
+ \frac{\rho^u_{\lambda_i}}{2\eta_i} (\|\lambda_i\|^2 - \|\lambda_i\|^2)
+ \frac{\rho^u_{\lambda_i}}{2(\rho^\psi_{\lambda_i})^2\eta_i} \|\lambda_i - \lambda_i\|^2
+ \frac{\rho^\psi_{\lambda_i}}{\eta_i} (w_i - Bu_{i+1}, w - Bu) - \frac{1}{\eta_i} (\lambda_i, w - Bu) - \frac{\Gamma_i}{2}.
\]

(3.24)

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

Now we analyze the accelerated convergence rate of the ABOSVS-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 $\psi$ is Lipschitz continuous, for all $w \in C^{d \times n}$, we have $\|\xi\| \leq \Delta$, for $\forall \xi \in \partial \psi(w)$. 

---

**Inverse Problems and Imaging** Volume 11, No. 6 (2017), 1047–1070
Theorem 3.7. Suppose that the parameters $\rho_i^u, \rho_i^w$, and $\rho_i^\lambda$ in Algorithm 2 satisfy (2.4), then the output $(u_{k+1}^ag, w_{k+1}^ag)$ generated by Algorithm 2 satisfies

$$
\|u_{k+1}^ag - Bu_{k+1}^ag\| \leq \frac{4\sigma \|AT\|}{\eta_{\min K}} \Delta ,
$$

where

$$
H(u_{k+1}^ag) + \psi(Bu_{k+1}^ag) - H(u^*) - \psi(Bu^*)
$$

$$
\leq 2\sigma \|AT\|(D_{u^*}^2 + C) + \frac{2\sigma \|AT\|D_{w^*}^2}{\eta_{\min K}} + \frac{4\sigma \|AT\|}{\eta_{\min K}} \Delta^2.
$$

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

$$
\eta_{\min} \leq \eta_{0,i} \leq \|AT\|.
$$

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

$$
\frac{1}{\alpha_i \sqrt{\eta_i}} = \frac{\sqrt{1-\alpha_i+1}}{\alpha_i+1} \leq \frac{1 - \frac{1}{2}\alpha_i+1}{\alpha_i+1 \sqrt{\eta_i+1}} \leq \frac{1}{\alpha_i+1 \sqrt{\eta_i+1}} - \frac{1}{2\eta_{i+1}} \text{ for } i \geq 1.
$$

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

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

which implies

$$
\alpha_i \eta_k \leq \frac{1}{\sqrt{\eta_i}} + \frac{1}{2} \sum_{i=2}^{k} \frac{1}{\sqrt{\eta_k}} \leq \frac{4\sigma \|AT\|}{(k+1)^2} \text{ for } k \geq 1,
$$

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

$$
\frac{1}{\alpha_i \sqrt{\eta_i}} = \frac{\sqrt{1-\alpha_i+1}}{\alpha_i+1 \sqrt{\eta_i+1}} \geq \frac{1 - \alpha_i+1}{\alpha_i+1 \sqrt{\eta_i+1}} = \frac{1}{\alpha_i+1 \sqrt{\eta_i+1}} - \frac{1}{\alpha_i+1 \sqrt{\eta_i+1}}
$$

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

$$
\frac{1}{\eta_k} \leq \frac{k}{\eta_{\min}} \text{ for } k \geq 1.
$$

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

$$
\sum_{i=1}^{k} \left(\frac{\rho_i^u - \rho_i^\lambda}{2(\rho_i^\lambda)^{\sigma_i}}\right) \|\lambda_{i+1} - \lambda_i\|^2 \geq 0.
$$
Also, using the definition of \( \rho^u_i \) and \( \rho^k \), and the non-increasing property of the sequences \( \{ \frac{\varphi^u_{k_i}}{2\eta_i} \} \) for \( i = 1, 2, \ldots, k \), we have

\[
\sum_{i=1}^{k} \frac{1}{2} \left( \| u_i - u^* \|^2 - \| u_{i+1} - u^* \|^2 \right) = \frac{1}{2} \| u_1 - u^* \|^2 - \frac{1}{2} \| u_k - u^* \|^2 \leq \frac{1}{2} D^2_u,
\]

(3.33)

\[
\sum_{i=1}^{k} \frac{\rho^u_i}{2\eta_i} (\| w_i - u^* \|^2 - \| w_{i+1} - u^* \|^2)
\]

(3.34)

\[
= \frac{\rho^u_i}{2\eta_i} \| w_1 - u^* \|^2 - \sum_{i=1}^{k-1} \left( \frac{\rho^u_i}{2\eta_i} - \frac{\rho^u_{i+1}}{2\eta_{i+1}} \right) \| w_{i+1} - u^* \|^2 - \frac{\rho^u_k}{2\eta_k} \| w_{k+1} - u^* \|^2 \leq \frac{K}{2\eta_{\min}} \| w_1 - u^* \|^2 \leq \frac{K}{2\eta_{\min}} D^2_{u^*},
\]

\[
\sum_{i=1}^{k} \frac{1}{2\rho^i_k \eta_i} (\| \lambda_i \|^2 - \| \lambda_{i+1} \|^2)
\]

(3.35)

\[
= \frac{1}{2\rho^i_k \eta_i} \| \lambda_1 \|^2 - \sum_{i=1}^{k-1} \left( \frac{1}{2\rho^i_k \eta_i} - \frac{1}{2\rho^i_{k+1} \eta_{k+1}} \right) \| \lambda_{i+1} \|^2 - \frac{1}{2\rho^k_k \eta_k} \| \lambda_{k+1} \|^2 \leq \frac{K}{2\eta_{\min}} (\| \lambda_1 \|^2 - \| \lambda_{k+1} \|^2) \leq 0,
\]

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

\[
\frac{1}{\alpha_k \eta_k} \left[ H(u^g_{k+1}) + \psi(w^g_{k+1}) - H(u^*) - \psi(u^*) \right] \leq \frac{D^2_u}{2} + \frac{KD^2_{u^*}}{2\eta_{\min}}.
\]

(3.36)

Next we focus on estimating \( \| u^g_{k+1} - Bu^g_{k+1} \| \) in order to obtain (3.36). Similarly as (3.31), for \( \forall \lambda \in \mathbb{C}^{d \times n} \), we can obtain

\[
\frac{1}{\alpha_k \eta_k} (u^g_{k+1} - Bu^g_{k+1}) = \sum_{i=1}^{k} \frac{1}{\eta_i} (u_{i+1} - Bu_{i+1})
\]

(3.37)

\[
= \sum_{i=1}^{k} \frac{1}{\eta_i} (u_{i+1} - Bu_{i+1}) = \sum_{i=1}^{k} \frac{1}{\rho^i \eta_i} (\lambda_i - \lambda_{i+1}).
\]

Thus, by \( \frac{1}{\alpha_k \eta_k} (u^g_{k+1} - Bu^g_{k+1}) = \sum_{i=1}^{k} \frac{1}{\rho^i \eta_i} (\lambda_i - \lambda_{i+1}) \), we have

\[
\frac{k \alpha_k \eta_k}{\eta_{\min}} (\lambda_1 - \lambda_{k+1}).
\]

(3.38)

By setting \( k = K - 1 \), we have

\[
\| u^g_{k+1} - Bu^g_{k+1} \| = \frac{4\sigma \| A^T A \|}{\eta_{\min} K} \| \lambda_{k+1} \|.
\]

(3.39)
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
\[ 0 \in \partial \psi(w_{i+1}) - \lambda_i + \rho_i^w(w_{i+1} - Bu_{i+1}) \forall i \geq 1, \]
or
\[ \lambda_{i+1} = \lambda_i - \rho_i^w(w_{i+1} - Bu_{i+1}) \in \partial \psi(w_{i+1}) \forall i \geq 1. \]
Therefore we have $\|\lambda_{i+1}\| \leq \Delta$, which leads to
\[ (3.40) \quad \|w_{k+1}^{ag} - Bu_{k+1}^{ag}\| = \frac{4\sigma\|A^TA\|}{\eta_{\min}K} \Delta. \]
By the convexity of $\psi(\cdot)$, we have
\[ H(u_{k+1}^{ag}) + \psi(Bu_{k+1}^{ag}) - H(u^*) - \psi(u^*) \leq H(u_{k}^{ag}) + \psi(u_{k}^{ag}) \]
\[ + \left< \xi, Bu_{k}^{ag} - Bu_{k+1}^{ag} \right> - H(u^*) - \psi(u^*), \]
where $\xi \in \partial \psi(Bu_{k+1}^{ag})$. Then, by (3.36), (3.40), (3.41), we have (3.26).

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 $\rho_i^u, \rho_i^w$, and $\rho_i^\lambda$ in Algorithm 2 satisfy (2.5), then the output $(u_{k+1}^{ag}, w_{k+1}^{ag})$ generated by Algorithm 2 satisfies
\[ (3.42) \quad \|u_{k+1}^{ag} - Bu_{k+1}^{ag}\| \leq \frac{2\sigma\|A^TA\|(D^2_v + C)}{(k+1)^2} + \frac{10\sigma\|A^TA\|D^2_v}{\eta_{\min}(k+1)} + \frac{4\sigma\|A^TA\|D^2_{U,B}}{\eta_{\min}(k+1)}. \]

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.

4.1. **Total-variation based image reconstruction.** In this subsection, we present the numerical results on solving the following TV based image reconstruction problem:
\[ (4.1) \quad \min_{u \in U} \frac{1}{2}\|Au - f\|^2 + \alpha\|u\|_{TV} \]
where $U := \{u \in \mathbb{R}^n : l_s \leq u^{(i)} \leq u_s \text{ for all } i = 1, \ldots, n\}$ 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 $f = Au_{\text{true}} + \varepsilon$, where $u_{\text{true}}$ is the true image, $\varepsilon$ is the Gaussian noise with distribution $\mathcal{N}(0, \sigma)$. We apply ABOVS-I, ABOVS-II, BOSVS \(^1\), TVAL3 \(^2\) and ALADMM \(^3\) to solve (4.1).

We consider two instances of this problem. In the first instance, we set $U := \{u \in \mathbb{R}^n : 0 \leq u^{(i)} \leq 1 \text{ for all } i = 1, \ldots, n\}$. The dimension of $A$ is $22500 \times 22500$. The

---

\(^1\)Code is available at [https://math.gsu.edu/xye/software/MPSEG_pub.zip](https://math.gsu.edu/xye/software/MPSEG_pub.zip)

\(^2\)Code is available at [http://www.caam.rice.edu/~optimization/L1/TVAL3](http://www.caam.rice.edu/~optimization/L1/TVAL3)

true image $u_{true}$ is a 150 $\times$ 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 ABOSVS-I, ABOSVS-II, BOSVS, TVAL3 and ALADMML. In the second instance, we have $U := \mathbb{R}^n$. The dimension of $A$ is 11250 $\times$ 22500. $u_{true}$ is a brain image [4] of dimension 150 $\times$ 150. The $\sigma$ 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 $\hat{u}$, defined by $\frac{\|\hat{u} - u_{true}\|}{\|u_{true}\|}$, are presented in Figure 1 and 2, respectively. In the first instance, we 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.

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

**Table 1.** Parameter settings for ABOSVS-I and ABOSVS-II in 4.1. Note that $\rho$ and $\tau$ are only used in the first instance

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\alpha$</th>
<th>$\sigma$</th>
<th>$\eta_{min}$</th>
<th>$C$</th>
<th>$\beta_k$</th>
<th>$\rho$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>10</td>
<td>2</td>
<td>$|A|^2/10$</td>
<td>100</td>
<td>1/k</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
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 ALADMM and TVAL3, and BOSVS. Details of PPI reconstruction problems can be found in [4]. The underlying image can be reconstructed

---

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

**Figure 3.** Sensitivity maps of data1
Figure 4. Left: True image; Right: Cartesian mask for data 1

Table 2. Parameter settings for ABOSVS-II in 4.2

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\alpha$</th>
<th>$\sigma$</th>
<th>$\eta_{\min}$</th>
<th>$C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>$10^{-10}$</td>
<td>2</td>
<td>$|A|^2/5$</td>
<td>100</td>
</tr>
</tbody>
</table>

by solving the following optimization problem:

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

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

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

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, $l = 1, 2, \ldots, L$. Then the above optimization problem can be rewritten as

\[
\min_{u \in \mathbb{C}^n} \left\{ \frac{1}{2} \| Au - f \|^2_2 + \alpha \|u\|_{TV} \right\}
\]

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

\[
f_l = P(FS_l u_{\text{true}} + \varepsilon_{l}^{re}/\sqrt{2} + \varepsilon_{l}^{im}/\sqrt{-2}), \quad \text{for} \quad l = 1, \ldots, L,
\]

where $\varepsilon_{l}^{re}$ and $\varepsilon_{l}^{im}$ are the noise with entries independently generated from distribution $\mathcal{N}(0, 10^{-4}\sqrt{n}I_n)$. 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, ALADMMML, 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, ALADMMML converge much faster than BOSVS, and the performance of TVA13 is between ALADMMML and BOSVS. ABOSVS-II has the best performance among them.
Table 3. Comparison of objective function value, relative error, and CPU time in seconds using data1

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Objective value</th>
<th>Relative error</th>
<th>CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOSVS</td>
<td>15.46841</td>
<td>0.0596</td>
<td>2528.0</td>
</tr>
<tr>
<td>TVAL3</td>
<td>56.7263</td>
<td>0.050</td>
<td>228.6</td>
</tr>
<tr>
<td>ALADMML</td>
<td>15.46810</td>
<td>0.0499</td>
<td>66.5</td>
</tr>
<tr>
<td>ABOSVS-II</td>
<td>15.44535</td>
<td>0.0479</td>
<td>49.7</td>
</tr>
</tbody>
</table>

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

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

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Objective value</th>
<th>Relative error</th>
<th>CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOSVS</td>
<td>5.007e+2</td>
<td>0.0646</td>
<td>7448.4</td>
</tr>
<tr>
<td>TVAL3</td>
<td>1.704e+3</td>
<td>0.050</td>
<td>681.2</td>
</tr>
<tr>
<td>ALADMML</td>
<td>5.006e+2</td>
<td>0.0499</td>
<td>180.0</td>
</tr>
<tr>
<td>ABOSVS-II</td>
<td>4.970e+2</td>
<td>0.0497</td>
<td>144.7</td>
</tr>
</tbody>
</table>

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 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 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.
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.

REFERENCES


Received November 2016; revised June 2017.

E-mail address: yun@math.ufl.edu
E-mail address: xianqili@ufl.edu
E-mail address: yuyuano@clemson.edu
E-mail address: eduardo.pasiliao@us.af.mil