

AN $\mathcal{O}(1/K)$ CONVERGENCE RATE FOR THE VARIABLE STEPSIZE BREGMAN OPERATOR SPLITTING ALGORITHM *

WILLIAM W. HAGER[†], MARYAM YASHTINI[‡], AND HONGCHAO ZHANG[§]

Abstract. An earlier paper proved the convergence of a variable stepsize Bregman operator splitting algorithm (BOSVS) for minimizing $\phi(Bu) + H(u)$ where H and ϕ are convex functions, and ϕ is possibly nonsmooth. The algorithm was shown to be relatively efficient when applied to partially parallel magnetic resonance image reconstruction problems. In this paper, the convergence rate of BOSVS is analyzed. When $H(u) = \|Au - f\|^2$ where A is a matrix, it is shown that for an ergodic approximation \mathbf{u}_k obtained by averaging k BOSVS iterates, the error in the objective value $\phi(B\mathbf{u}_k) + H(\mathbf{u}_k)$ is $\mathcal{O}(1/k)$. When the optimization problem has a unique solution u^* , we obtain the estimate $\|\mathbf{u}_k - u^*\| = \mathcal{O}(1/\sqrt{k})$. The theoretical analysis is compared to observed convergence rates for partially parallel magnetic resonance image reconstruction problems where A is a large dense ill conditioned matrix.

Key words. Nonsmooth optimization, convex optimization, BOSVS, ergodic convergence, saddle point problem, variational inequality

1. Introduction. In this paper, we develop a convergence rate analysis for the variable stepsize Bregman operator splitting algorithm (BOSVS) to solve

$$(1.1) \quad \min_{u \in \mathbb{C}^N} \phi(Bu) + H(u),$$

where $\phi(\cdot)$ and $H(\cdot)$ are convex real-valued functions, $\phi : \mathbb{C}^l \rightarrow \mathbb{R}$ is possibly nonsmooth, $H : \mathbb{C}^N \rightarrow \mathbb{R}$ is continuously differentiable, and B is an $l \times N$ matrix. In our target application, magnetic resonance image reconstruction, $u \in \mathbb{C}^N$ denotes an image containing N pixels. Problems with the structure of (1.1) have received much attention during the last decade; references include [4, 8, 11, 13, 14, 18, 20, 21, 26, 33, 39, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52].

Our analysis is based on the following split reformulation of (1.1):

$$(1.2) \quad \min_{u \in \mathbb{C}^N, w \in \mathbb{C}^l} \Phi(u, w) := \phi(w) + H(u) \quad \text{subject to } Bu = w.$$

The paper will focus on the following form for H which arises in image reconstruction problems:

$$(1.3) \quad H(u) = \frac{1}{2} \|Au - f\|^2,$$

where $A \in \mathbb{C}^{M \times N}$ describes the imaging device, $f \in \mathbb{C}^M$ is the measured data, and $\|f\|^2 = \langle f, f \rangle$ where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product. By Lemma 3.1 in [19], if $\phi(w)$ tends to infinity as $\|w\|$ tends to infinity and (1.3) holds, then there exists a

*September 23, 2013. Revised September 30, 2015. This research was partly supported by National Science Foundation grants 1016204, 1115568, 1522629, and 1522654, and by Office of Naval Research grants N00014-11-1-0068 and N00014-15-1-2048.

[†]hager@ufl.edu, <http://people.clas.ufl.edu/hager/>, PO Box 118105, Department of Mathematics, University of Florida, Gainesville, FL 32611-8105. Phone (352) 294-2308. Fax (352) 392-8357.

[‡]myashtini3@math.gatech.edu, <http://people.math.gatech.edu/~myashtini3/>, School of Mathematics, Georgia Institute of Technology, 686 Cherry Street, Atlanta, GA 30332-0160. Phone (404) 894-9469.

[§]hozhang@math.lsu.edu, <https://www.math.lsu.edu/~hozhang/> Department of Mathematics, Louisiana State University, Baton Rouge, LA, 70803-4918, USA. Phone (225) 578-1982.

solution to (1.1). Moreover, if ϕ is strictly convex and the null spaces of A and B only intersect at the origin, then the solution of (1.1) is unique. In partially parallel imaging [9], A is large, dense and ill-conditioned. For efficiency, the evaluation of the product Au must be kept to a minimum when solving (1.1).

The alternating direction method of multipliers (ADMM) [16, 17, 22] is a common approach for split formulations such as (1.2). In this approach, the augmented Lagrangian is successively minimized over one variable, then the other, followed by an update of the multiplier. More precisely, when H has the form (1.3) arising in imaging, the iteration is

$$\begin{aligned} u^{k+1} &= \arg \min \left\{ \frac{1}{2} \|Au - f\|^2 + \frac{\rho}{2} \|w^k - Bu + \rho^{-1}b^k\|^2 : u \in \mathbb{C}^N \right\}, \\ w^{k+1} &= \arg \min \left\{ \phi(w) + \frac{\rho}{2} \|w - Bu^{k+1} + \rho^{-1}b^k\|^2 : w \in \mathbb{C}^l \right\}, \\ b^{k+1} &= b^k - \rho(Bu^{k+1} - w^{k+1}). \end{aligned}$$

The convergence rate of ADMM is $\mathcal{O}(1/k)$ [22, 28]. Notice that in each iteration when u^{k+1} is computed, we need to solve a linear system with matrix $(A^\top A + B^\top B)$ where $^\top$ denotes conjugate transpose. In applications such as partially parallel imaging, this matrix is large and dense, and the algorithm is intractable computationally. In order to deal with the high cost of an ADMM iteration in this context, a linearization technique, the Bregman operator splitting scheme (BOS) was developed [49, 50], where the quadratic term $\|Au\|^2$ was replaced by a linear approximation. BOS can be viewed as a special case of the alternating direction proximal method of multipliers. In this proximal perspective, the BOS update for u^{k+1} is

$$u^{k+1} = \arg \min \left\{ \frac{1}{2} \|Au - f\|^2 + \frac{\rho}{2} \|w^k - Bu + \rho^{-1}b^k\|^2 + \frac{1}{2} \|u - u^k\|_Q^2 : u \in \mathbb{C}^N \right\},$$

where $\|u\|_Q = \langle u, Qu \rangle$ and $Q = \delta I - A^\top A$ with $\delta > 0$ a positive scalar. For this choice of Q , the $A^\top A$ term in $\|Au - f\|^2$ is canceled by the $A^\top A$ term in $\|u - u^k\|_Q^2$, and the intractable term in ADMM is eliminated. The convergence analysis for BOS ultimately requires that δ is greater than the spectral radius of $A^\top A$ (see Theorem 4.2 in [50] or Theorem 5.6 in [36]). When this holds, BOS has the same $\mathcal{O}(1/k)$ convergence rate as ADMM, but for ergodic iterates associated with the original iterates [36]. In the imaging context, BOS is much faster than ADMM, even though the convergence rate is the same, since it is no longer necessary to solve a large dense linear system.

In the variable stepsize Bregman operator splitting algorithm (BOSVS), introduced in [9], the scalar δ in BOS is replaced a parameter δ_k which is evaluated in each iteration through a line search process. Our line search process is loosely related to the SpARSA algorithm [1, 3, 5, 40] where the objective is linearized and a proximal parameter is adjusted in each iteration to satisfy a line search condition. The objective in the BOS update for u^{k+1} is not linear due to the $\|Bu\|^2$ term, but it is partially linearized through the removal of the $\|Au\|^2$ term. Initially, δ_k in BOSVS is given by the Barzilai-Borwein [2] formula

$$\delta_k = \frac{\|A(u^k - u^{k-1})\|^2}{\|u^k - u^{k-1}\|^2},$$

which emerges when $A^\top A$ is approximated by a multiple of the identity matrix. If the line search condition is not satisfied, then δ_k is increased. By carefully controlling

the growth of δ_k , we are able to guarantee convergence while achieving much better performance in partially parallel image reconstruction problems when compared to the performance for a fixed δ greater than the spectral radius of $A^\top A$.

As noted earlier, the convergence analysis of BOS hinged on the positive definiteness of the proximal term, which required that δ is greater than the spectral radius of $A^\top A$. In contrast, in BOSVS δ_k is typically less than the spectral radius of $A^\top A$ and the proximal term is indefinite, which requires a completely new analysis. Our earlier work [9] established the convergence of BOSVS. In this paper, we establish an $\mathcal{O}(1/k)$ convergence rate for the ergodic iterate

$$\mathbf{u}_k = \frac{1}{k} \sum_{j=1}^k u^{j+1},$$

obtained by averaging k BOSVS iterates. If (1.1) has a solution u^* , then we prove that

$$\phi(B\mathbf{u}_k) + H(\mathbf{u}_k) - [\phi(Bu^*) + H(u^*)] = \mathcal{O}(1/k).$$

This is the same convergence rate which holds for the BOS ergodic iterates (see [36, Thm. 6.2]), however, we observe better performance in practice through the use of an indefinite proximal term. If the null spaces of A and B only intersect at the origin and ϕ satisfies a strong convexity condition so that the solution of (1.1) is unique, then we also show that

$$\|\mathbf{u}_k - u^*\| = \mathcal{O}\left(\frac{1}{\sqrt{k}}\right).$$

The numerical experiments reveal that the error in the BOSVS iterates is highly oscillatory, while the error in the ergodic iterates decays nearly monotonically.

The use of an ergodic mean in the analysis of gradient-type and/or proximal point-based methods for convex minimization and monotone variational inequalities dates back to at least the mid-seventies [6]. Applications of ergodic convergence theory include the convergence rate of the Douglas-Rachford alternating direction method [17] established in [22], the complexity result [27] of Monteiro and Svaiter for the hybrid proximal extragradient algorithm, and the $\mathcal{O}(1/k)$ convergence result [30] of Nemirovski for a prox-method. Also, in [24] Lan and Monteiro give an $\mathcal{O}(1/k)$ convergence result for an augmented Lagrangian type algorithm employing Nesterov's optimal method [31]. The analysis assumes that the objective function has a Lipschitz continuous gradient, while we consider a nonsmooth objective in this paper.

Also, we point out the recent work of Chen, Lan, and Ouyang [10] on an accelerated primal-dual method for solving a class of deterministic and stochastic saddle point problems. Their method achieves the same $\mathcal{O}(1/k)$ convergence rate as that of Nesterov's smoothing technique [32]. The error is measured relative to a perturbation-based termination criterion of Monteiro and Svaiter [27]. The algorithm in [10] differs from the BOSVS algorithm in the treatment of the constraint $w = Bu$. BOSVS treats the constraint using an augmented Lagrangian approach (both a penalty and a Lagrange multiplier), while in Algorithm 2 of [10], the constraint is handled through a conjugate function. The algorithm in [10] involves a number of parameters that need to be chosen in a careful, conservative way to obtain the convergence results. Since BOSVS employs a line search, potentially larger and more aggressive steps can be taken, especially in the early iterations. Note that the analogue of the u^k update

in [10] requires the evaluation of ∇H at an auxiliary point, which is not an iterate. Hence, it would seem that if a line search was developed for the u^k update in [10], then the gradient would need to be evaluated at two points in each iteration. The algorithm in [10] seems promising, but the comparison with BOSVS in performance for partially parallel imaging, where ϕ corresponds to total variation regularization, is not clear.

The paper is organized as follows. Section 2 examines some general variational inequality formulations of nonsmooth minimization problems which provide the framework for our analysis, while Section 3 states the BOSVS algorithm and the prior convergence analysis. In Section 4 we derive new properties for BOSVS that enter into the ergodic convergence analysis. The main convergence theory is developed in Section 5, while Section 6 gives a numerical study of the convergence rate of the BOSVS iterates and the ergodic means in partially parallel magnetic resonance image reconstruction.

Notation. A superscript \top denotes conjugate transpose and $\langle \cdot, \cdot \rangle$ is the Euclidean inner product. If x and $y \in \mathbb{C}^N$, then $\langle x, y \rangle = x^\top y$. We let $\|\cdot\|$ denote the Euclidean norm given by $\|x\| = \sqrt{\langle x, x \rangle}$. If $Q \in \mathbb{C}^{N \times N}$ is a Hermitian matrix, then we define

$$\|x\|_Q^2 := \langle x, Qx \rangle.$$

If Q is also positive definite, then $\|\cdot\|_Q$ is a norm. However, in this paper, we also use $\|\cdot\|_Q$ in cases where Q is indefinite. The smallest (most negative) eigenvalue of Q is denoted $\lambda_{\min}(Q)$. $\mathcal{N}(A)$ denotes the null space of the matrix A . Throughout the paper, we let c denote a generic constant which has different values in different equations. This constant is always independent of the iteration number k .

Throughout the paper, we work in \mathbb{C}^N . For each $x \in \mathbb{C}^N$, a real-valued convex function \mathcal{F} has a nonempty subdifferential set $\partial\mathcal{F}(x)$ with the property that for each $s \in \partial\mathcal{F}(x)$ and for each $y \in \mathbb{C}^N$,

$$(1.4) \quad \mathcal{F}(y) \geq \mathcal{F}(x) + \operatorname{Re} \langle s, y - x \rangle.$$

Here Re stands for “real part.” If \mathcal{F} is differentiable, then we also let $\nabla\mathcal{F}(x)$ denote the gradient at x . The real part must be included in (1.4) since the associated support hyperplane inequality should be formulated in \mathbb{R}^{2N} , while we are working in the complex space \mathbb{C}^N .

2. Some variational inequalities. Our analysis centers around some variational inequalities which we now develop. The following result follows from the analysis on page 102 of Nemirovski’s lecture notes [29].

PROPOSITION 2.1. *If $\mathcal{F} : \mathbb{C}^N \rightarrow \mathbb{R}$ and $\mathcal{K} \subset \mathbb{C}^N$ are both convex, then we have*

$$(2.1) \quad \operatorname{Re} \langle y, x - x^* \rangle \geq 0 \text{ for all } x \in \mathcal{K} \text{ and } y \in \partial\mathcal{F}(x)$$

if and only if

$$(2.2) \quad \mathcal{F}(x^*) \leq \mathcal{F}(x) \text{ for all } x \in \mathcal{K}.$$

Let us consider the Lagrangian $\mathcal{L} : \mathbb{C}^N \times \mathbb{C}^l \times \mathbb{C}^l \rightarrow \mathbb{R}$ associated with (1.2):

$$(2.3) \quad \mathcal{L}(u, w, b) = \phi(w) + H(u) - \operatorname{Re} \langle b, Bu - w \rangle,$$

where $b \in \mathbb{C}^l$ is the Lagrange multiplier. Since \mathcal{L} is convex with respect to u and w , it follows that if u^* and w^* are optimal in (1.2), then there exists $b^* \in \mathbb{C}^l$ such that

$$(2.4) \quad \mathcal{L}(u^*, w^*, b^*) = \min\{\mathcal{L}(u, w, b^*) : u \in \mathbb{C}^N, w \in \mathbb{C}^l\}.$$

And conversely, if there exists $u^* \in \mathbb{C}^N$, $w^* \in \mathbb{C}^l$, and $b^* \in \mathbb{C}^l$ such that $w^* = Bu^*$ and (2.4) holds, then u^* and w^* are optimal in (1.2). By Proposition 2.1, (2.4) is equivalent to the inequality

$$(2.5) \quad \operatorname{Re} \langle \nabla H(u) - B^\top b^*, u - u^* \rangle + \operatorname{Re} \langle s + b^*, w - w^* \rangle \geq 0$$

for all $u \in \mathbb{C}^N$, $w \in \mathbb{C}^l$, and $s \in \partial\phi(w)$. Hence, if u^* and w^* are optimal in (1.2), then there exists $b^* \in \mathbb{C}^l$ such that (2.5) holds. And conversely, if there exists $u^* \in \mathbb{C}^N$, $w^* \in \mathbb{C}^l$, and $b^* \in \mathbb{C}^l$ such that $w^* = Bu^*$ and (2.5) holds, then (2.4) holds and u^* and w^* are optimal in (1.2).

Now, let us consider the closely related variational inequality

$$(2.6) \quad \begin{aligned} \operatorname{Re} \langle \nabla H(u) - B^\top b^*, u - u^* \rangle + \operatorname{Re} \langle s + b^*, w - w^* \rangle \\ + \operatorname{Re} \langle Bu^* - w^*, b - b^* \rangle \geq 0 \end{aligned}$$

for all $b \in \mathbb{C}^l$, $u \in \mathbb{C}^N$, $w \in \mathbb{C}^l$, and $s \in \partial\phi(w)$. If (2.6) holds for some $u^* \in \mathbb{C}^N$, $w^* \in \mathbb{C}^l$, and $b^* \in \mathbb{C}^l$, then from the last term in (2.6), we deduce that $Bu^* = w^*$. Hence, the last term in (2.6) vanishes and (2.5) holds. Consequently, u^* and w^* are solutions of (1.2). Conversely, if u^* and w^* are solutions of (1.2), then (2.5) holds, which implies that (2.6) holds since u^* and w^* are feasible in (1.2) and the last term in (2.6) vanishes. Hence, (2.6) is equivalent to the existence of a solution (u^*, w^*) to (1.2).

Finally, let us now consider another related variational inequality: Find $u^* \in \mathbb{C}^N$, $w^* \in \mathbb{C}^l$, and $b^* \in \mathbb{C}^l$ such that

$$(2.7) \quad \begin{aligned} \operatorname{Re} \langle \nabla H(u) - B^\top b, u - u^* \rangle + \operatorname{Re} \langle s + b, w - w^* \rangle \\ + \operatorname{Re} \langle Bu - w, b - b^* \rangle \geq 0 \end{aligned}$$

for all $b \in \mathbb{C}^l$, $u \in \mathbb{C}^N$, $w \in \mathbb{C}^l$, and $s \in \partial\phi(w)$. If (2.7) holds, then we can replace b by $b - b^* + b^*$ in the first two terms and rearrange to obtain (2.6). Conversely, if (2.6) holds, then we can replace b^* by $b^* - b + b$ and rearrange to obtain (2.7). Hence, (2.6) and (2.7) are equivalent. We summarize these observations as follows:

PROPOSITION 2.2. *There exists $u^* \in \mathbb{C}^N$, $w^* \in \mathbb{C}^l$, and $b^* \in \mathbb{C}^l$ satisfying (2.7) if and only if u^* and w^* are optimal in (1.2).*

Note that the variational inequality (2.7) can also be expressed as

$$(2.8) \quad \operatorname{Re} \langle y, x - x^* \rangle \geq 0$$

for all $x = (u, w, b) \in \mathbb{C}^N \times \mathbb{C}^l \times \mathbb{C}^l$ and $y \in F(x)$ where

$$(2.9) \quad F(x) = F(u, w, b) := \begin{bmatrix} \nabla H(u) - B^\top b \\ \partial\phi(w) + b \\ Bu - w \end{bmatrix}.$$

We will show that x^* in (2.8) can be replaced by the ergodic iterate at iteration k if the right side of (2.8) is replaced by $-c/k$. Hence, the ergodic iterates satisfy a

perturbed variational inequality. This perturbed variational formulation becomes the basis for our error analysis.

In contrast to (2.8), He and Yuan in [22] write (2.4) in the form of a variational inequality which in our setting amounts to finding $x^* = (u^*, w^*, b^*) \in \mathbb{C}^N \times \mathbb{C}^l \times \mathbb{C}^l$ such that

$$(2.10) \quad \Phi(u, w) - \Phi(u^*, w^*) + \operatorname{Re} \langle x - x^*, \bar{F}(x^*) \rangle \geq 0$$

for all $x = (u, w, b) \in \mathbb{C}^N \times \mathbb{C}^l \times \mathbb{C}^l$, where

$$\bar{F}(u, w, b) = \begin{bmatrix} -B^\top b \\ b \\ Bu - w \end{bmatrix}.$$

Since $\langle x - x^*, \bar{F}(x) - \bar{F}(x^*) \rangle = 0$, the inequality (2.10) is equivalent to finding $x^* = (u^*, w^*, b^*) \in \mathbb{C}^N \times \mathbb{C}^l \times \mathbb{C}^l$ such that

$$(2.11) \quad \Phi(u, w) - \Phi(u^*, w^*) + \operatorname{Re} \langle x - x^*, \bar{F}(x) \rangle \geq 0$$

for all $x = (u, w, b) \in \mathbb{C}^N \times \mathbb{C}^l \times \mathbb{C}^l$. In [22] that authors show that the ergodic iterates satisfy a perturbed version of the inequality (2.11).

3. BOSVS. Algorithm 3.1 is the BOSVS algorithm of [9]. In BOSVS, the

BREGMAN OPERATOR SPLITTING WITH VARIABLE STEPSIZE (BOSVS)

Parameters: $\tau, \eta > 1$, $\beta, C \geq 0$, $\rho, \delta_{\min} > 0$, $\xi, \sigma \in (0, 1)$, $\delta_0 = 1$.

Starting guess: $w^1 \in \mathbb{C}^l$, $u^1 \in \mathbb{C}^N$, $b^1 \in \mathbb{C}^l$. Set $Q_1 = 0$ and $\Delta_1 = \delta_{\min}$.

For $k = 1, 2, \dots$

Step 1. $\hat{\delta}_k = 1$ if $k = 1$; otherwise, $\hat{\delta}_k$ is given by (3.1).

Step 2. Set $\delta_k = \eta^j \hat{\delta}_k$ where $j \geq 0$ is the smallest integer such that

$Q_{k+1} \geq -\frac{C}{k^2}$ where $Q_{k+1} := \xi_k Q_k + \Omega_k$ with

$\Omega_k := \sigma(\delta_k \|u^{k+1} - u^k\|^2 + \rho \|Bu^{k+1} - w^k\|^2) - \|A(u^{k+1} - u^k)\|^2$, and

$u^{k+1} = \operatorname{argmin}_u \left\{ \delta_k \|u - u^k + \delta_k^{-1} A^\top (Au^k - f)\|^2 + \rho \|w^k - Bu + \rho^{-1} b^k\|^2 \right\}$,

$0 \leq \xi_k \leq \min\{\xi, (1 - k^{-1})^2\}$.

Step 3. If $\delta_k \leq \max\{\delta_{k-1}, \Delta_k\}$, then $\Delta_{k+1} = \Delta_k$; otherwise $\Delta_{k+1} = \tau \Delta_k$.

Step 4. $w^{k+1} = \operatorname{argmin}_w \left\{ \phi(w) + \frac{\rho}{2} \|w - Bu^{k+1} + \rho^{-1} b^k\|^2 + \frac{\beta}{2} \|w - w^k\|^2 \right\}$.

Step 5. $b^{k+1} = b^k - \rho(Bu^{k+1} - w^{k+1})$.

Step 6. If a stopping criterion is satisfied, terminate the algorithm.

End For

variable b^k is the approximation to a multiplier associated with the constraint $Bu = w$ in the split formulation (1.2), the parameter ρ is the penalty for constraint violation in the augmented Lagrangian associated with (1.2), and Step 5 represents a first-order multiplier update. β in Step 4 is a proximal regularization parameter connected with the variable w . Step 3 allows some growth in the lower bounded Δ_k for δ_k when $\delta_k > \delta_{k-1}$. In this paper, $\hat{\delta}_k$ of Step 2 is given by a safe-guarded version of the Barzilai-Borwein [2] formula

$$(3.1) \quad \hat{\delta}_k = \max \left\{ \Delta_k, \frac{\|A(u^k - u^{k-1})\|^2}{\|u^k - u^{k-1}\|^2} \right\}$$

for $k > 1$. For $k = 1$, $\hat{\delta}_k$ can be chosen arbitrarily; we take $\hat{\delta}_1 = 1$. In [9] we proved a uniform bound on δ_k when $\hat{\delta}_k$ is chosen in this way. Typically, we start by taking δ_{\min} small, in which case $\hat{\delta}_k = \|A(u^k - u^{k-1})\|^2 / \|u^k - u^{k-1}\|^2$ for small k , which is less than the largest eigenvalue of $A^\top A$ except in the very special case where $u^k - u^{k-1}$ is aligned with a vector associated with the largest eigenvalue of $A^\top A$. As a result, the associated proximal term in the alternating direction proximal method of multipliers is typically indefinite. Step 2 amounts to a line search which permits potential growth in δ_k . This line search is sufficient to guarantee convergence as shown in [9], but typically δ_k remains smaller than the largest eigenvalue of $A^\top A$. There is a minor change in the initialization of the BOSVS parameter C of Algorithm 3.1 compared to the initializations in [9]. In [9] we indicated $C > 1$ since we thought that larger values of C would cause the line search to terminate more quickly and provide better performance. However, for the convergence theory, we only need $C \geq 0$. For the ergodic convergence results, we need to bound ξ_k away from 1 in Step 2, while this constraint was not needed for the global convergence of BOSVS. The iterates (u^k, w^k, b^k) of the BOSVS algorithm converge to both a solution and a multiplier associated with the constraint in (1.2) as shown in [9].

Results for the BOSVS algorithm established in [9] are summarized in the following theorem. In this section and in the remainder of the paper, we assume that H has the special form (1.3) which arises in imaging.

THEOREM 3.1. *The BOSVS algorithm has the following properties:*

1. *If there exists a solution of (1.2), then the sequence $x^k = (u^k, w^k, b^k)$ generated by BOSVS approaches a limit $x^* = (u^*, w^*, b^*)$ where the first-order optimality conditions for (1.2) are satisfied. That is,*

$$\nabla H(u^*) - B^\top b^* = 0, \quad -b^* \in \partial\phi(w^*), \quad w^* = Bu^*.$$

Moreover, (u^, w^*) is a solution of (1.2) with b^* a corresponding Lagrange multiplier.*

2. *In Step 2 of BOSVS, the integer j is bounded uniformly in k and $\delta_k \leq \eta \|A\|^2 / \sigma$.*
3. *In Step 3 of BOSVS, $\Delta_{k+1} = \Delta_k$ except for at most $\log_\tau(\|A\|^2 / (\sigma\delta_{\min}))$ iterations. Hence, $\delta_k \leq \delta_{k-1}$ except in a finite number of iterations.*
4. *If the sequence $x^k = (u^k, w^k, b^k)$ generated by BOSVS approaches a limit $x^* = (u^*, w^*, b^*)$, then we have*

$$\sum_{k=1}^{\infty} \|u^{k+1} - u^k\|^2 + \|w^{k+1} - w^k\|^2 + \|Bu^{k+1} - w^k\|^2 < \infty.$$

4. Additional properties of BOSVS. We now establish a few additional properties of the BOSVS algorithm:

LEMMA 4.1. *If δ_k is chosen according to Step 2 of the BOSVS algorithm and the matrix G_k is defined by*

$$G_k := \delta_k I - A^\top A,$$

where I is the $N \times N$ identity matrix, then

$$\|u^{k+1} - u^k\|_{G_k}^2 \geq -\tau_k + \theta \|u^{k+1} - u^k\|^2,$$

where $\theta = (1 - \sigma)\delta_{\min}$ and

$$(4.1) \quad \tau_k = \frac{C}{k^2} + \xi_k Q_k + \sigma \rho \|Bu^{k+1} - w^k\|^2.$$

Proof. In Step 2 of BOSVS, δ_k is accepted if

$$(4.2) \quad Q_{k+1} \geq -\frac{C}{k^2},$$

where $Q_{k+1} = \xi_k Q_k + \Omega_k$, $\xi_k \in [0, (1 - k^{-1})^2]$, and Ω_k is defined by

$$\Omega_k = \sigma \left(\delta_k \|u^{k+1} - u^k\|^2 + \rho \|Bu^{k+1} - w^k\|^2 \right) - \|A(u^{k+1} - u^k)\|^2.$$

Hence, the inequality (4.2) can be written as

$$\frac{C}{k^2} + \xi_k Q_k + \sigma \left(\delta_k \|u^{k+1} - u^k\|^2 + \rho \|Bu^{k+1} - w^k\|^2 \right) - \|A(u^{k+1} - u^k)\|^2 \geq 0.$$

We rearrange this to obtain

$$\begin{aligned} & \frac{C}{k^2} + \xi_k Q_k + \left(\delta_k \|u^{k+1} - u^k\|^2 - \|A(u^{k+1} - u^k)\|^2 \right) + \sigma \rho \|Bu^{k+1} - w^k\|^2 \\ & \geq (1 - \sigma)\delta_k \|u^{k+1} - u^k\|^2 \geq (1 - \sigma)\delta_{\min} \|u^{k+1} - u^k\|^2, \end{aligned}$$

since $\delta_k \geq \delta_{\min}$ for all k . In terms of $G_k = \delta_k I - A^\top A$, this inequality can be expressed

$$\frac{C}{k^2} + \xi_k Q_k + \|u^{k+1} - u^k\|_{G_k}^2 + \sigma \rho \|Bu^{k+1} - w^k\|^2 \geq (1 - \sigma)\delta_{\min} \|u^{k+1} - u^k\|^2.$$

Hence, the lemma holds with $\theta = (1 - \sigma)\delta_{\min}$. \square

Numerically, we observe that the error in the iterates generated by BOSVS is highly oscillatory. In the paper, we analyze the convergence rate of an ergodic sequence associated with the BOSVS iterates. Let $x^k = (u^k, w^k, b^k)$ be the BOSVS iterates, and define a modified sequence $\bar{x}^k = (\bar{u}^k, \bar{w}^k, \bar{b}^k)$ where

$$(4.3) \quad \bar{u}^k = u^{k+1}, \quad \bar{w}^k = w^{k+1}, \quad \bar{b}^k = b^k - \rho(Bu^{k+1} - w^k).$$

The associated ergodic mean \mathbf{x}_k is given by

$$(4.4) \quad \mathbf{x}_k := (\mathbf{u}_k, \mathbf{w}_k, \mathbf{b}_k) = \frac{1}{k} \sum_{j=1}^k \bar{x}^j.$$

The next lemma establishes a variational identity for the modified sequence (4.3).

LEMMA 4.2. *The BOSVS sequence $x^k = (u^k, w^k, b^k)$ and the modified sequence \bar{x}^k given in (4.3) satisfy*

$$(4.5) \quad W_k(x^k - \bar{x}^k) = \bar{y}^k$$

for some $\bar{y}^k \in F(\bar{x}^k)$, where

$$W_k := \begin{bmatrix} G_k & 0 & 0 \\ 0 & (\beta + \rho)I & 0 \\ 0 & I & \rho^{-1}I \end{bmatrix}.$$

Here I is the $l \times l$ identity matrix.

Proof. The u -subproblem in Step 2 of BOSVS is equivalent to

$$(4.6) \quad u^{k+1} = \operatorname{argmin}_u \left\{ 2H(u) + \rho \|Bu - w^k - \rho^{-1}b^k\|^2 + \|u - u^k\|_{G_k}^2 \right\},$$

where $H(u) = \frac{1}{2}\|Au - f\|^2$. These problems are equivalent since the objective functions only differ by terms independent of u . Setting the derivative to 0 at u^{k+1} in (4.6), we get

$$\nabla H(u^{k+1}) + \rho B^\top (Bu^{k+1} - w^k - \rho^{-1}b^k) + G_k(u^{k+1} - u^k) = 0.$$

Since $\bar{u}^k = u^{k+1}$ and $\bar{b}^k = b^k - \rho(Bu^{k+1} - w^k)$, this expression can be written as

$$(4.7) \quad \nabla H(\bar{u}^k) - B^\top \bar{b}^k + G_k(\bar{u}^k - u^k) = 0.$$

The first-order optimality condition for Step 4 of BOSVS is

$$s^{k+1} + b^k - \rho(Bu^{k+1} - w^{k+1}) + \beta(w^{k+1} - w^k) = 0$$

for some $s^{k+1} \in \partial\phi(w^{k+1})$. We define $\bar{s}^k := s^{k+1}$ and rewrite this in terms of \bar{b}^k and $\bar{w}^k = w^{k+1}$ to obtain

$$(4.8) \quad \bar{s}^k + \bar{b}^k + (\beta + \rho)(\bar{w}^k - w^k) = 0.$$

The definition of $\bar{b}^k = b^k - \rho(Bu^{k+1} - w^k)$ can be rearranged into the following form:

$$(4.9) \quad (B\bar{u}^k - \bar{w}^k) + (\bar{w}^k - w^k) + \frac{1}{\rho}(\bar{b}^k - b^k) = 0,$$

where again $\bar{w}^k = w^{k+1}$ and $\bar{u}^k = u^{k+1}$. Combine (4.7)–(4.9) to obtain

$$\begin{bmatrix} \nabla H(\bar{u}^k) - B^\top \bar{b}^k - G_k(u^k - \bar{u}^k) \\ \bar{s}^k + \bar{b}^k - (\beta + \rho)(w^k - \bar{w}^k) \\ (B\bar{u}^k - \bar{w}^k) - (w^k - \bar{w}^k) - \rho^{-1}(b^k - \bar{b}^k) \end{bmatrix} = 0,$$

which is the same as (4.5). \square

LEMMA 4.3. *The component $z^k := (w^k, b^k)$ of the BOSVS sequence and the modified sequence $\bar{z}^k := (\bar{w}^k, \bar{b}^k)$ defined in (4.3) satisfy*

$$(4.10) \quad \|z^k - \bar{z}^k\|_D^2 - \|z^{k+1} - \bar{z}^k\|_D^2 = \beta \|w^k - \bar{w}^k\|^2 + \frac{1}{\rho} \|b^k - \bar{b}^k\|^2 \geq 0,$$

where D is a positive definite matrix defined by

$$D := \begin{bmatrix} (\beta + \rho)I & 0 \\ 0 & \rho^{-1}I \end{bmatrix}$$

Proof. After substituting for D , the left side of (4.10) reduces to

$$(4.11) \quad (\beta + \rho)(\|w^k - \bar{w}^k\|^2 - \|w^{k+1} - \bar{w}^k\|^2) + \rho^{-1}(\|b^k - \bar{b}^k\|^2 - \|b^{k+1} - \bar{b}^k\|^2) = (\beta + \rho)\|w^k - w^{k+1}\|^2 + \rho^{-1}(\|b^k - \bar{b}^k\|^2 - \|b^{k+1} - \bar{b}^k\|^2),$$

since $\bar{w}^k = w^{k+1}$. By the definition of \bar{b}^k and by the formula for b^{k+1} in Step 5 of BOSVS, it follows that

$$(4.12) \quad b^{k+1} - \bar{b}^k = \rho(w^{k+1} - w^k).$$

With this substitution in (4.11), we obtain (4.10). \square

LEMMA 4.4. *For all $x = (u, w, b)$, we have*

$$\begin{aligned} & 2\operatorname{Re} \langle x - \bar{x}^k, W_k(x^k - \bar{x}^k) \rangle + \|u - u^k\|_{G_k}^2 + \|z - z^k\|_D^2 + \tau_k \\ & \geq \|u - u^{k+1}\|_{G_k}^2 + \|z - z^{k+1}\|_D^2 + \theta\|u^{k+1} - u^k\|^2, \end{aligned}$$

where τ_k is defined in (4.1).

Proof. By the definition of W_k , we have

$$(4.13) \quad \begin{aligned} & \langle x - \bar{x}^k, W_k(x^k - \bar{x}^k) \rangle = \\ & \langle u - \bar{u}^k, G_k(u^k - \bar{u}^k) \rangle + \langle z - \bar{z}^k, (D + E)(z^k - \bar{z}^k) \rangle \end{aligned}$$

for all $x = (u, z)$ where $z = (w, b)$ and

$$E = \begin{bmatrix} 0 & 0 \\ I & 0 \end{bmatrix}.$$

For any matrix G , we have

$$(4.14) \quad \begin{aligned} & \operatorname{Re} \langle (a - b), G(c - d) \rangle = \\ & \frac{1}{2} \left(\|a - d\|_G^2 - \|a - c\|_G^2 + \|c - b\|_G^2 - \|d - b\|_G^2 \right). \end{aligned}$$

Applying this to the first term on the right side of (4.13) gives

$$\operatorname{Re} \langle u - \bar{u}^k, G_k(u^k - \bar{u}^k) \rangle = \frac{1}{2} \left(\|u - u^{k+1}\|_{G_k}^2 - \|u - u^k\|_{G_k}^2 \right) + \frac{1}{2} \|u^k - u^{k+1}\|_{G_k}^2,$$

since $\bar{u}^k = u^{k+1}$. By Lemma 4.1, it follows that

$$(4.15) \quad \begin{aligned} & \operatorname{Re} \langle u - \bar{u}^k, G_k(u^k - \bar{u}^k) \rangle \geq \frac{1}{2} \left(\|u - u^{k+1}\|_{G_k}^2 - \|u - u^k\|_{G_k}^2 \right) \\ & \quad + \frac{1}{2} \left(\theta\|u^{k+1} - u^k\|^2 - \tau_k \right). \end{aligned}$$

By (4.12), it follows that

$$(D + E)(z^k - \bar{z}^k) = D(z^k - z^{k+1}).$$

With this substitution, the last term in (4.13) becomes $\langle z - \bar{z}^k, D(z^k - z^{k+1}) \rangle$. We apply (4.14) to obtain

$$\begin{aligned} \operatorname{Re} \langle z - \bar{z}^k, D(z^k - z^{k+1}) \rangle &= \frac{1}{2} \left(\|z - z^{k+1}\|_D^2 - \|z - z^k\|_D^2 \right) \\ &\quad + \frac{1}{2} \left(\|z^k - \bar{z}^k\|_D^2 - \|z^{k+1} - \bar{z}^k\|_D^2 \right). \end{aligned}$$

By Lemma 4.3, we have $\|z^k - \bar{z}^k\|_D^2 - \|z^{k+1} - \bar{z}^k\|_D^2 \geq 0$; hence

$$(4.16) \quad \operatorname{Re} \langle z - \bar{z}^k, D(z^k - z^{k+1}) \rangle \geq \frac{1}{2} \left(\|z - z^{k+1}\|_D^2 - \|z - z^k\|_D^2 \right).$$

We combine (4.13), (4.15), and (4.16) to complete the proof. \square

LEMMA 4.5. *In BOSVS, τ_k given in (4.1) satisfies $\tau_k \geq 0$ for each k and*

$$(4.17) \quad \sum_{k=1}^{\infty} \tau_k < \infty.$$

Proof. Since the BOSVS parameter C is nonnegative and $Q_1 = 0$, we have

$$\tau_1 = C + \sigma\rho\|Bu^2 - w^1\|^2 \geq 0.$$

For $k \geq 2$, Step 2 of BOSVS ensures that $Q_k \geq -C/(k-1)^2$. Since $\xi_k \leq (1 - k^{-1})^2$, we have

$$\tau_k \geq \frac{C}{k^2} + \xi_k Q_k \geq \frac{C}{k^2} - \frac{C\xi_k}{(k-1)^2} \geq \frac{C}{k^2} - \frac{C(1 - k^{-1})^2}{(k-1)^2} = 0.$$

Now, consider the series in (4.17):

$$\sum_{k=1}^{\infty} \tau_k = \sum_{k=1}^{\infty} \frac{C}{k^2} + \sum_{k=1}^{\infty} \sigma\rho\|Bu^{k+1} - w^k\|^2 + \sum_{k=1}^{\infty} \xi_k Q_k$$

The first sum on the right is clearly convergent while the second sum is convergent by Theorem 3.1 part 4. To complete the proof, we need to show that the last sum is bound from above. Recall that $Q_{k+1} := \xi_k Q_k + \Omega_k$ and

$$\Omega_k = \sigma(\delta_k\|u^{k+1} - u^k\|^2 + \rho\|Bu^{k+1} - w^k\|^2) - \|A(u^{k+1} - u^k)\|^2.$$

By the uniform bound on δ_k in part 2 of Theorem 3.1, we conclude that for some constant c , we have

$$\Omega_k \leq c(\|u^{k+1} - u^k\|^2 + \|Bu^{k+1} - w^k\|^2) := \bar{\Omega}_k.$$

Working backwards,

$$\begin{aligned}
Q_k &= \xi_{k-1}Q_{k-1} + \Omega_{k-1} \\
&\leq \xi_{k-1}Q_{k-1} + \bar{\Omega}_{k-1} \\
&\leq \xi_{k-1} \left(\xi_{k-2}Q_{k-2} + \bar{\Omega}_{k-2} \right) + \bar{\Omega}_{k-1} \\
&\leq \left(\prod_{j=2}^{k-1} \xi_j \right) \bar{\Omega}_1 + \left(\prod_{j=3}^{k-1} \xi_j \right) \bar{\Omega}_2 + \dots + \bar{\Omega}_{k-1}.
\end{aligned}$$

Since $0 \leq \xi_k \leq \xi$ and $\bar{\Omega}_k \geq 0$, it follows that

$$\xi_k Q_k \leq \xi^{k-1} \bar{\Omega}_1 + \xi^{k-2} \bar{\Omega}_2 + \dots + \xi \bar{\Omega}_{k-1}.$$

Hence,

$$\sum_{k=1}^{\infty} \xi_k Q_k \leq \left(\sum_{k=1}^{\infty} \xi^k \right) \left(\sum_{k=1}^{\infty} \bar{\Omega}_k \right).$$

Since $\xi < 1$, the geometric series in ξ is convergent, and by Theorem 3.1, part 4, the series in $\bar{\Omega}_k$ is convergent. This completes the proof of (4.17). \square

5. Ergodic convergence analysis of BOSVS. Let \mathcal{B}_u and \mathcal{B}_w denote the unit balls centered at u^* and w^* , and let \mathcal{B}_b be the ball with center b^* and radius

$$(5.1) \quad r^* = 1 + \max \{ \|s + b^*\| : s \in \partial\phi(w^*) \}.$$

By [34, Thm. 23.4], $\partial\phi(w^*)$ is a bounded set so r^* is finite. Define $\mathcal{B} = \mathcal{B}_u \times \mathcal{B}_w \times \mathcal{B}_b$. Note that $u^* \in \mathcal{B}_u$, $w^* \in \mathcal{B}_w$, $b^* \in \mathcal{B}_b$, and $x^* \in \mathcal{B}$. We now show that the ergodic means \mathbf{x}_k defined in (4.4) obey a perturbed minimax property and satisfy a perturbed version of the variational inequality (2.8).

THEOREM 5.1. *Suppose BOSVS converges to a solution (u^*, w^*) of (1.2) and an associated multiplier b^* for the constraint. If \mathbf{x}_k is the sequence of ergodic means defined in (4.4), then there exists a constant c independent of k such that*

$$(5.2) \quad \operatorname{Re} \langle y, x - \mathbf{x}_k \rangle \geq -\frac{c}{k} \text{ for all } x \in \mathcal{B} \text{ and } y \in F(x),$$

where F is defined in (2.9). Moreover, we have

$$(5.3) \quad \max\{\mathcal{L}(\mathbf{u}_k, \mathbf{w}_k, b) : b \in \mathcal{B}_b\} - \min\{\mathcal{L}(u, w, \mathbf{b}_k) : u \in \mathcal{B}_u, w \in \mathcal{B}_w\} \leq \frac{c}{k}.$$

Proof. Since BOSVS converges to x^* , it follows that the sequence \bar{x}^k also converges to x^* , which implies that the ergodic sequence \mathbf{x}_k converges to x^* . By Lemmas 4.2 and 4.4, there exists $\bar{y}^j \in F(\bar{x}^j)$ such that

$$\begin{aligned}
2\operatorname{Re} \langle \bar{y}^j, x - \bar{x}^j \rangle &\geq (\|u - u^{j+1}\|_{G_j}^2 - \|u - u^j\|_{G_j}^2) \\
&+ \theta \|u^{j+1} - u^j\|^2 + (\|z - z^{j+1}\|_D^2 - \|z - z^j\|_D^2) - \tau_j
\end{aligned}$$

for all x . Sum this from $j = 1$ to k to obtain

$$\begin{aligned} 2\operatorname{Re} \sum_{j=1}^k \langle \bar{y}^j, x - \bar{x}^j \rangle &\geq \sum_{k=1}^k \left(\|u - u^{j+1}\|_{G_j}^2 - \|u - u^j\|_{G_j}^2 \right) \\ &+ \theta \sum_{k=1}^k \|u^{j+1} - u^j\|^2 + (\|z - z^{k+1}\|_D^2 - \|z - z^1\|_D^2) - \sum_{j=1}^k \tau_j. \end{aligned}$$

We drop the positive terms $\|u^{j+1} - u^j\|^2$ and $\|z - z^{k+1}\|_D^2$ and rearrange to obtain

$$(5.4) \quad \operatorname{Re} \sum_{j=1}^k \langle \bar{y}^j, \bar{x}^j - x \rangle \leq \frac{1}{2} \left[\|z - z^1\|_D^2 + \sum_{j=1}^k \left(\tau_j + \|u - u^j\|_{G_j}^2 - \|u - u^{j+1}\|_{G_j}^2 \right) \right].$$

We now establish a uniform bound for the right side of (5.4); more precisely, it is shown that there exists a constant c , independent of k and $x \in \mathcal{B}$ such that

$$(5.5) \quad \operatorname{Re} \sum_{j=1}^k \langle \bar{y}^j, \bar{x}^j - x \rangle \leq c.$$

For all $x \in \mathcal{B}$, $\|z - z^1\|_D^2$ is bounded uniformly. By Lemma 4.5, the τ_j sum in (5.4) is bounded above, uniformly in k . Now consider the terms in (5.4) subscripted by G_j . Since $G_j = \delta_j I - A^\top A$, it follows that

$$\begin{aligned} \|u - u^j\|_{G_j}^2 - \|u - u^{j+1}\|_{G_j}^2 &= \delta_j (\|u - u^j\|^2 - \|u - u^{j+1}\|^2) \\ &- \|A(u - u^j)\|^2 + \|A(u - u^{j+1})\|^2. \end{aligned}$$

Hence, we have

$$(5.6) \quad \begin{aligned} \sum_{j=1}^k \left(\|u - u^j\|_{G_j}^2 - \|u - u^{j+1}\|_{G_j}^2 \right) &= \delta_1 \|u - u^1\|^2 - \delta_k \|u - u^{k+1}\|^2 \\ &+ \sum_{j=2}^k (\delta_j - \delta_{j-1}) \|u - u^j\|^2 - \|A(u - u^1)\|^2 + \|A(u - u^{k+1})\|^2. \end{aligned}$$

To establish (5.5), we need an upper bound for (5.6). Consequently, the negative terms on the right side can be neglected. Since u^j converges to u^* and $\|u - u^*\| \leq 1$, $\|A(u - u^{k+1})\|^2$ is bounded uniformly in k . Finally, consider the $(\delta_j - \delta_{j-1})\|u - u^j\|^2$ term. By part 2 of Theorem 3.1, $\delta_j \geq 0$ is uniformly bounded, and by part 3 of Theorem 3.1, $\delta_j \leq \delta_{j-1}$ except in a finite number of iterations. The terms with $\delta_j \leq \delta_{j-1}$ are negative and can be neglected. Since u^j converges to u^* and $\|u - u^*\| \leq 1$, the positive terms are bounded uniformly in j , and since there are a finite number of positive terms with $\delta_j > \delta_{j-1}$, (5.6) is bounded uniformly in k , which establishes (5.5).

For any $y \in F(x)$, we have

$$\langle y, \bar{x} - x \rangle = \langle \nabla H(u) - B^\top b, \bar{u} - u \rangle + \langle s + b, \bar{w} - w \rangle + \langle Bu - w, \bar{b} - b \rangle,$$

where s is any element of $\partial\phi(w)$. Since both H and ϕ are convex, they satisfy the monotonicity conditions

$$\operatorname{Re} \langle s, \bar{w} - w \rangle \leq \operatorname{Re} \langle \bar{s}, \bar{w} - w \rangle$$

and

$$\operatorname{Re} \langle \nabla H(u), \bar{u} - u, \rangle \leq \operatorname{Re} \langle \nabla H(\bar{u}), \bar{u} - u \rangle$$

for all $\bar{s} \in \partial\phi(\bar{w})$ and $s \in \partial\phi(w)$. Utilizing these monotonicity conditions, we obtain

$$\begin{aligned} \operatorname{Re} \langle y, \bar{x} - x \rangle &\leq \operatorname{Re} \langle \nabla H(\bar{u}) - B^\top b, \bar{u} - u \rangle \\ &\quad + \operatorname{Re} \langle \bar{s} + b, \bar{w} - w \rangle + \operatorname{Re} \langle Bu - w, \bar{b} - b \rangle. \end{aligned}$$

We replace b by $(b - \bar{b}) + \bar{b}$ in the first two terms on the right side of the inequality and rearrange to deduce that

$$(5.7) \quad \operatorname{Re} \langle y, \bar{x} - x \rangle \leq \operatorname{Re} \langle \bar{y}, \bar{x} - x \rangle$$

for all $y \in F(x)$ and $\bar{y} \in F(\bar{x})$.

We combine the monotonicity relation (5.7) with (5.5) to obtain

$$\operatorname{Re} \sum_{j=1}^k \langle y, \bar{x}^j - x \rangle \leq \operatorname{Re} \sum_{j=1}^k \langle \bar{y}^j, \bar{x}^j - x \rangle \leq c$$

for all $x \in \mathcal{B}$ and $y \in F(x)$. By the definition of the ergodic mean in (4.4), we have

$$\operatorname{Re} \langle y, \mathbf{x}_k - x \rangle = \left(\frac{1}{k} \right) \operatorname{Re} \sum_{j=1}^k \langle y, \bar{x}^j - x \rangle \leq \left(\frac{1}{k} \right) \operatorname{Re} \sum_{j=1}^k \langle \bar{y}^j, \bar{x}^j - x \rangle \leq \frac{c}{k}$$

for all $x \in \mathcal{B}$ and $y \in F(x)$. This yields (5.2).

Next, let us consider the minimax property (5.3). Since $\mathcal{L}(u, w, b)$ is convex in (u, w) and concave (linear) in b , it follows that

$$\begin{aligned} \mathcal{L}(\mathbf{u}_k, \mathbf{w}_k, b) - \mathcal{L}(u, w, \mathbf{b}_k) &= \mathcal{L}\left(\frac{1}{k} \sum_{j=1}^k \bar{u}^j, \frac{1}{k} \sum_{j=1}^k \bar{w}^j, b\right) - \mathcal{L}\left(u, w, \frac{1}{k} \sum_{j=1}^k \bar{b}^j\right) \\ (5.8) \quad &\leq \frac{1}{k} \sum_{j=1}^k \left[\mathcal{L}(\bar{u}^j, \bar{w}^j, b) - \mathcal{L}(u, w, \bar{b}^j) \right]. \end{aligned}$$

By the definition of \mathcal{L} in (2.3), we have

$$(5.9) \quad \mathcal{L}(\bar{u}^j, \bar{w}^j, b) - \mathcal{L}(u, w, \bar{b}^j) = H(\bar{u}^j) - H(u) + \phi(\bar{w}^j) - \phi(w) - \operatorname{Re} \langle b, B\bar{u}^j - \bar{w}^j \rangle + \operatorname{Re} \langle \bar{b}^j, Bu - w \rangle.$$

The convexity of H and ϕ implies that

$$(5.10) \quad H(\bar{u}^j) - H(u) \leq \operatorname{Re} \langle \nabla H(\bar{u}^j), \bar{u}^j - u \rangle,$$

and

$$(5.11) \quad \phi(\bar{w}^j) - \phi(w) \leq \operatorname{Re} \langle \bar{s}^j, \bar{w}^j - w \rangle$$

for every $\bar{s}^j \in \partial\phi(\bar{w}^j)$. In the trailing terms of (5.9), we replace b by $b - \bar{b}^j + \bar{b}^j$ and we rearrange to obtain

$$(5.12) \quad \begin{aligned} & \langle \bar{b}^j, Bu - w \rangle - \langle b, B\bar{u}^j - \bar{w}^j \rangle = \\ & - \langle B^T \bar{b}^j, \bar{u}^j - u \rangle + \langle \bar{b}^j, \bar{w}^j - w \rangle + \langle B\bar{u}^j - \bar{w}^j, \bar{b}^j - b \rangle. \end{aligned}$$

Combine (5.9)–(5.12) to obtain

$$\mathcal{L}(\bar{u}^j, \bar{w}^j, b) - \mathcal{L}(u, w, \bar{b}^j) \leq \operatorname{Re} \langle \bar{y}^j, \bar{x}^j - x \rangle.$$

Hence, (5.5) and (5.8) imply that

$$\mathcal{L}(\mathbf{u}_k, \mathbf{w}_k, b) - \mathcal{L}(u, w, \mathbf{b}_k) \leq \frac{1}{k} \left(\operatorname{Re} \sum_{j=1}^k \langle \bar{y}^j, \bar{x}^j - x \rangle \right) \leq \frac{c}{k},$$

which yields (5.3). \square

Next, we analyze the error in the objective function and the error in the constraint.

THEOREM 5.2. *Suppose BOSVS converges to a solution (u^*, w^*) of (1.2) and an associated multiplier b^* for the constraint. If \mathbf{x}_k is the sequence of ergodic means defined in (4.4), then there exists a constant c independent of k such that*

$$(5.13) \quad \|\mathbf{w}_k - B\mathbf{u}_k\| \leq \frac{c}{\sqrt{k}},$$

$$(5.14) \quad H(\mathbf{u}_k) + \phi(\mathbf{w}_k) - H(u^*) - \phi(w^*) \leq \frac{c}{\sqrt{k}},$$

and

$$(5.15) \quad H(\mathbf{u}_k) + \phi(B\mathbf{u}_k) - H(u^*) - \phi(w^*) \leq \frac{c}{k},$$

Proof. By the Cauchy-Schwarz and triangle inequalities, we have

$$(5.16) \quad \begin{aligned} \|\mathbf{w}_k - B\mathbf{u}_k\|^2 &= \left\| \frac{1}{k} \sum_{j=1}^k (\bar{w}^j - B\bar{u}^j) \right\|^2 = \frac{1}{k^2} \left\| \sum_{j=2}^{k+1} (w^j - Bu^j) \right\|^2 \\ &\leq \frac{1}{k^2} \left(k \sum_{j=2}^{k+1} \|w^j - Bu^j\|^2 \right) \\ &= \frac{1}{k} \sum_{j=2}^{k+1} \|w^j - Bu^j\|^2 \\ &= \frac{1}{k} \left[\sum_{j=2}^{k+1} \|w^{j-1} - Bu^j + (w^j - w^{j-1})\|^2 \right] \\ &\leq \frac{2}{k} \left(\sum_{j=2}^{k+1} \|w^{j-1} - Bu^j\|^2 + \sum_{j=2}^{k+1} \|w^j - w^{j-1}\|^2 \right). \end{aligned}$$

By Theorem 3.1, part 4, the sums in (5.16) are bounded, uniformly in k . This establishes (5.13).

To establish (5.14), we make the choice $u = u^*$, $w = w^*$, and $b = b^*$ in (5.3) to obtain

$$\phi(\mathbf{w}_k) + H(\mathbf{u}_k) - \operatorname{Re} \langle b^*, B\mathbf{u}_k - \mathbf{w}_k \rangle - \phi(w^*) - H(u^*) \leq \frac{c}{k}.$$

Combine this with (5.13) to obtain (5.14).

Finally, let us consider (5.15). Since

$$-\lim_{k \rightarrow \infty} B\mathbf{u}_k = \lim_{k \rightarrow \infty} \mathbf{w}_k = w^*,$$

it follows from [34, Cor. 24.5.1] that for k sufficiently large, there exists $-v_k \in \partial\phi(B\mathbf{u}_k)$ such that the distance from $-v_k$ to $\partial\phi(w^*)$ is less than 1. Choose k sufficiently large and $s_k \in \partial\phi(w^*)$ such that $\|v_k + s_k\| \leq 1$. Consequently, we have

$$\|v_k - b^*\| \leq \|v_k + s_k\| + \|s_k + b^*\| \leq 1 + \|s_k + b^*\| \leq r^*$$

by (5.1). Hence, $v_k \in \mathcal{B}_b$. We apply (5.3) with $b = v_k$, $u = u^*$, and $w = w^* = Bu^*$ to obtain

$$(5.17) \quad \phi(\mathbf{w}_k) + H(\mathbf{u}_k) - \operatorname{Re} \langle v_k, B\mathbf{u}_k - \mathbf{w}_k \rangle - \phi(Bu^*) - H(u^*) \leq \frac{c}{k}.$$

Since $-v_k \in \partial\phi(B\mathbf{u}_k)$ and ϕ is convex, it follows that

$$(5.18) \quad \phi(\mathbf{w}_k) - \operatorname{Re} \langle v_k, B\mathbf{u}_k - \mathbf{w}_k \rangle \geq \phi(B\mathbf{u}_k).$$

Combine (5.17) and (5.18) to obtain (5.15). \square

When a strong convexity assumption holds, there is a convergence result for ergodic means $(\mathbf{u}_k, \mathbf{w}_k)$:

THEOREM 5.3. *Suppose BOSVS converges to a solution (u^*, w^*) of (1.2) and an associated multiplier b^* for the constraint. If $\mathcal{N}(A) \cap \mathcal{N}(B) = \{0\}$, and $\phi(w)$ is strongly convex at w^* in the sense that there exists $\gamma > 0$ such that*

$$(5.19) \quad \phi(w) - \phi(w^*) + \operatorname{Re} \langle b^*, w - w^* \rangle \geq \gamma \|w - w^*\|^2 \text{ for all } w \in \mathcal{B}_w,$$

then (u^, w^*) is the unique solution of (1.2) and there exists a constant c , independent of k , such that the sequence of ergodic means \mathbf{x}_k defined in (4.4) satisfy*

$$(5.20) \quad \|\mathbf{w}_k - w^*\| \leq \frac{c}{\sqrt{k}} \quad \text{and} \quad \|\mathbf{u}_k - u^*\| \leq \frac{c}{\sqrt{k}}.$$

Proof. If (u^*, w^*) is the unique solution of (1.2) in $\mathcal{B}_u \times \mathcal{B}_w$, then due to the convexity of the objective function and constraint, (u^*, w^*) is the unique global minimum. On the other hand, suppose that there exists a second solution (\tilde{u}, \tilde{w}) of (1.2). Since $B\tilde{u} - \tilde{w} = 0$ and $Bu^* = w^*$, the objective function Φ of (1.2) satisfies

$$(5.21) \quad \begin{aligned} & \Phi(\tilde{u}, \tilde{w}) - \Phi(u^*, w^*) = H(\tilde{u}) - H(u^*) \\ & - \operatorname{Re} \langle b^*, B(\tilde{u} - u^*) \rangle + \phi(\tilde{w}) - \phi(w^*) + \operatorname{Re} \langle b^*, \tilde{w} - w^* \rangle, \end{aligned}$$

where H is defined in (1.3). Expanding H in a Taylor series around u^* and taking into account the fact that $B^\top b^* = -\nabla H(u^*)$ by the first-order optimality conditions, we obtain

$$(5.22) \quad H(\tilde{u}) - H(u^*) - \operatorname{Re} \langle b^*, B(\tilde{u} - u^*) \rangle = \frac{1}{2} \|A(\tilde{u} - u^*)\|^2.$$

Combining (5.19), (5.21), and (5.22) yields

$$(5.23) \quad \Phi(\tilde{u}, \tilde{w}) - \Phi(u^*, w^*) \geq \frac{1}{2} \|A(\tilde{u} - u^*)\|^2 + \gamma \|\tilde{w} - w^*\|^2.$$

Since $\Phi(\tilde{u}, \tilde{w}) = \Phi(u^*, w^*)$, we conclude that $\tilde{w} = w^*$ and $\tilde{u} - u^* \in \mathcal{N}(A)$. Since $Bu^* = w^*$, $B\tilde{u} = \tilde{w}$ and $\tilde{w} = w^*$, it follows that $\tilde{u} - u^* \in \mathcal{N}(B)$. Since $\mathcal{N}(A) \cap \mathcal{N}(B) = \{0\}$, we conclude that $\tilde{u} = u^*$. Hence, (u^*, w^*) is the unique solution of (1.2) and u^* is the unique solution of (1.1).

Substituting $x = (u^*, w^*, b^*)$ in (5.3) gives

$$(5.24) \quad \mathcal{L}(\mathbf{u}_k, \mathbf{w}_k, b^*) - \mathcal{L}(u^*, w^*, \mathbf{b}_k) \leq \frac{c}{k}.$$

The left side of (5.24) is

$$(5.25) \quad \phi(\mathbf{w}_k) - \phi(w^*) + H(\mathbf{u}_k) - H(u^*) - \operatorname{Re} \langle b^*, B(\mathbf{u}_k - u^*) - (\mathbf{w}_k - w^*) \rangle$$

since $w^* = Bu^*$. Similar to (5.22), we expand H in a Taylor series around u^* , we drop the linear term by the first-order optimality conditions, and we utilize (5.19) to obtain

$$\mathcal{L}(\mathbf{u}_k, \mathbf{w}_k, b^*) - \mathcal{L}(u^*, w^*, \mathbf{b}_k) \geq \frac{1}{2} \|A(\mathbf{u}_k - u^*)\|^2 + \gamma \|\mathbf{w}_k - w^*\|^2.$$

Therefore, by (5.24), we have

$$(5.26) \quad \frac{1}{2} \|A(\mathbf{u}_k - u^*)\|^2 + \gamma \|\mathbf{w}_k - w^*\|^2 \leq \frac{c}{k}.$$

This implies that

$$(5.27) \quad \|\mathbf{w}_k - w^*\|^2 \leq \frac{c}{\gamma k}.$$

Since $\mathcal{N}(A) \cap \mathcal{N}(B) = \{0\}$, the smallest eigenvalue of $A^\top A + B^\top B$, denoted λ_{\min} , is positive. Hence, we have

$$\begin{aligned} \lambda_{\min} \|\mathbf{u}_k - u^*\|^2 &\leq \|\mathbf{u}_k - u^*\|_{A^\top A + B^\top B}^2 \\ &= \|A(\mathbf{u}_k - u^*)\|^2 + \|B(\mathbf{u}_k - u^*)\|^2 \\ &= \|A(\mathbf{u}_k - u^*)\|^2 + \|B\mathbf{u}_k - \mathbf{w}_k + \mathbf{w}_k - w^*\|^2 \\ &\leq \|A(\mathbf{u}_k - u^*)\|^2 + 2\|B\mathbf{u}_k - \mathbf{w}_k\|^2 + 2\|\mathbf{w}_k - w^*\|^2. \end{aligned}$$

We combine this with (5.13), (5.26), and (5.27) to complete the proof of (5.20). \square

6. Numerical Results. In this section, we numerically explore the convergence speed of the BOSVS algorithm and the sequence of ergodic means using image reconstruction problems that arise in partially parallel imaging (PPI). We compare three algorithms: BOS, BOSVS, and the ergodic version of BOSVS analyzed in this paper. The algorithms were implemented in MATLAB. The PPI reconstruction problems are described in detail in [9]. The regularization term ϕ is total variation regularization (TV), which was first introduced by Rudin, Osher, and Fetami in [35]. Subsequent work on accelerated algorithms for TV regularized problems includes [7, 12, 23, 25, 37, 38]. Total variation regularization is defined by

$$(6.1) \quad \phi(Bu) = \alpha \|u\|_{TV} = \alpha \sum_{i=1}^N \|(\nabla u)_i\|,$$

where $Bu = \nabla u$ and $(\nabla u)_i$ is the vector of finite differences in the image along the coordinate directions at the i -th pixel in the image. The data sets used in our numerical experiments, denoted data 1, data 2, and data 3, correspond to the reconstruction of the PPI images shown in Figure 6.1, panels (a), (b), and (c). The PPI system

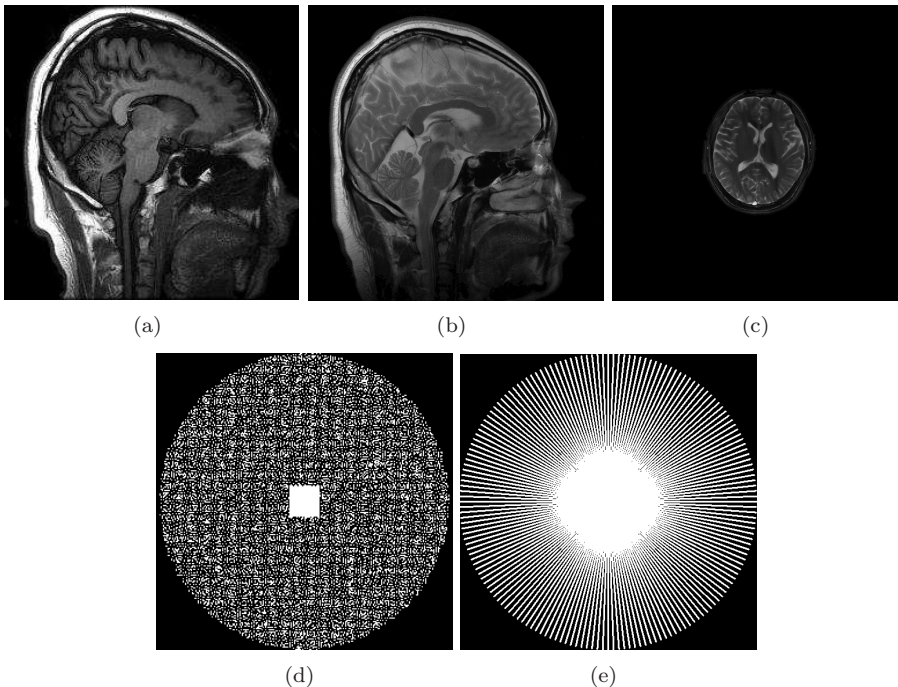


FIG. 6.1. (a) *Reference image for data 1.* (b) *Reference image for data 2.* (c) *Reference image for data 3.* (d) *Poisson random mask used for data 1 and data 2 with 25% undersampling ratio.* (e) *Radial mask used for data 3 with 34% undersampling ratio.*

used in our experiments had 8 coils. For data 1 and data 2, we use a random Poisson mask shown in Figure 6.1 panel (d), while for data 3 we use a radial mask shown in panel (e). In these figures, the illuminated pixels correspond to the components of the Fourier transform that are recorded by the imaging device. The Poisson mask samples 25% of the Fourier coefficients, while the radial mask samples 34% of the

Fourier coefficients. In our experiments, we used the following parameters:

$$\alpha = 0.00001, \quad \rho = 0.0001, \quad \tau = 1.01, \quad \eta = 3, \quad \beta = 0, \quad \delta_{\min} = 0.001, \\ C = 100, \quad \sigma = 0.99, \quad \xi = 0.8.$$

The parameter α , the weight for the total variation regularization term (6.1), is chosen to achieve the best reconstructed image for the solution of (1.1) and the three images in this test set. The penalty parameter ρ is taken relatively small since this value leads to faster convergence of the algorithms. The remaining parameters arise specifically in BOSVS. Their values were based on our personal preferences; for example, σ and ξ should be slightly less than 1 and τ should be slightly greater than 1. The proximal term in w does not seem to provide significant benefit so we set $\beta = 0$.

Our evaluation of performance is based on the following estimates for the optimal objective values obtained by running BOSVS for 100,000 iterations:

$$\begin{aligned} \Phi(u^*, w^*) &= 0.266540 && \text{for data 1,} \\ \Phi(u^*, w^*) &= 1.0525 && \text{for data 2,} \\ \Phi(u^*, w^*) &= 1.047221 && \text{for data 3.} \end{aligned}$$

Figure 6.2 plots the error in the objective function versus the number of multiplications by either A or A^T to achieve that error. Since A is large dense matrix, the number of multiplications by A or by A^T should be proportional to running time of the algorithm if it were coded in a compiled language such as C or Fortran. In computing the ergodic iterates, we skipped the first 10 of the 1000 iterates and the averaging starts at iteration 11.

In Figure 6.2, we see that after about 100 matrix multiplications, the most accurate BOSVS iterates are about an order of magnitude more accurate than the BOS iterates. Thus the indefinite proximal term associated with BOSVS provides a significant improvement in the accuracy of the objective. On the other hand, the BOSVS iterates possess significant oscillation, and the objective error can increase or decrease by an order of magnitude in a few iterations. The ergodic iterates, which possess a guaranteed $\mathcal{O}(1/k)$ convergence rate, eliminate the oscillations. The ergodic iterates are about half and an order of magnitude more accurate than BOS iterates, and significantly less accurate than the most accurate BOSVS iterates.

In Figure 6.3 we plot the logarithm of the error for the ergodic BOSVS iterates as a function of the logarithm of the iteration number. In a log/log plot, a function of the form $y = ck^{-p}$ would appear as a line with slope $-p$. The least square fit to p for the plots shown in Figure 6.3 is 1.6, 1.2, and 0.9 for data 1, data 2, and data 3 respectively. Since our analysis gives an upper bound for the error of ck^{-1} , the observed error roughly follows our established upper bound.

7. Conclusion. The convergence rate of the variable stepsize Bregman operator splitting algorithm (BOSVS) was studied. This algorithm was motivated by the alternating direction method of multipliers [15], the Barzilai-Borwein approximation [2] to a Hessian in the objective function, and a regularization idea loosely related to what is used in the SpaRSA algorithm [1, 3, 5, 40]. The algorithm is equivalent to the alternating direction proximal method of multipliers with an *indefinite* proximal term which is chosen by a line search technique. The analysis utilized the equivalence between the nonsmooth convex optimization problem and a related variational inequality. It was proved that the ergodic mean associated with k iterations of the BOSVS algorithm has an objective error of $\mathcal{O}(1/k)$. Moreover, when the objective

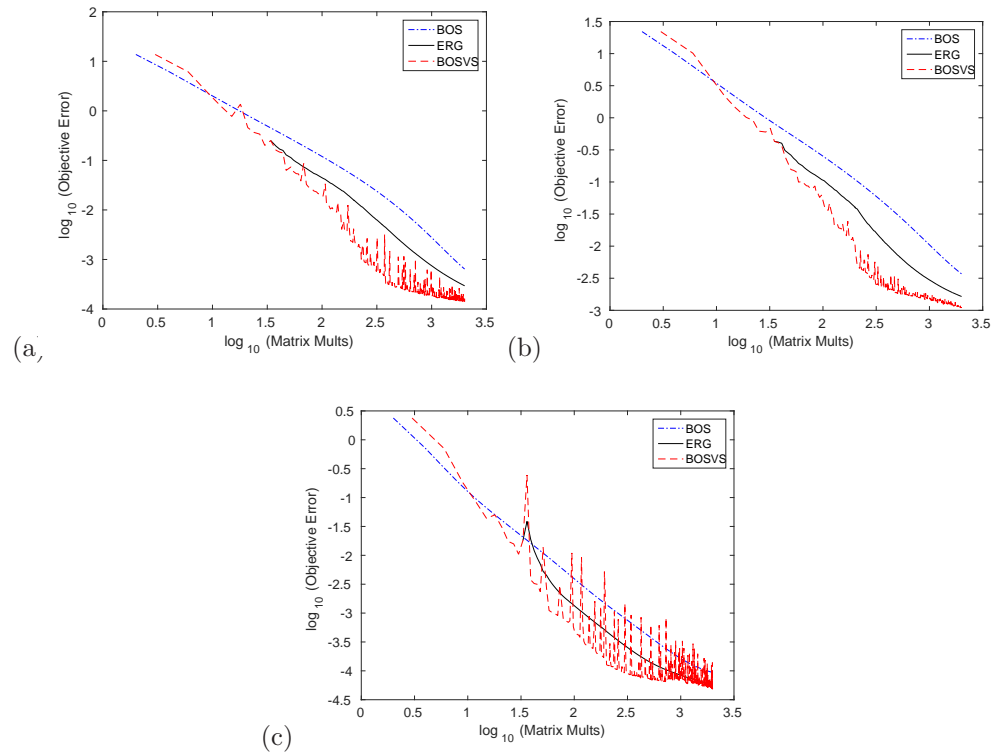


FIG. 6.2. Plots of the objective error as a function of the number of multiplications by A or A^T to achieve the error for data 1, data 2, and data 3 respectively.

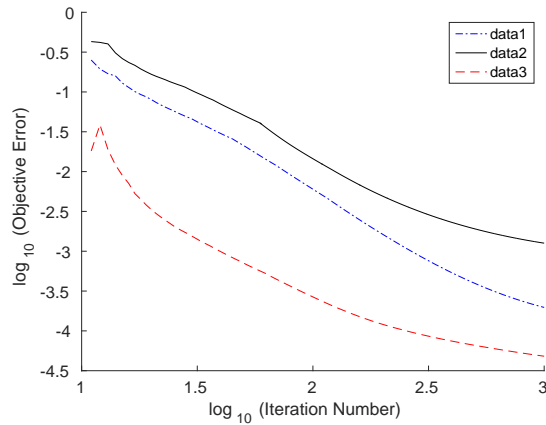


FIG. 6.3. Plots of the objective error for the ergodic BOSVS iterates as a function of the iteration number.

function satisfies a strong convexity condition, then the solution error is $\mathcal{O}(1/\sqrt{k})$. In numerical experiments using image reconstruction problems arising in partially parallel imaging (PPI) with total variation regularization, it was observed that the objective error of the most accurate BOSVS iterates could be an order of magnitude more accurate than the objective error of the BOS iterates where the proximal term is

positive semidefinite. While the error in the BOSVS iterates can oscillate, the error in the ergodic BOSVS iterates decreases nearly monotonically with an guaranteed error bound of $\mathcal{O}(1/k)$.

Acknowledgments. The authors thank Invivo Corporation and Dr. Feng Huang for providing the PPI data used in the paper. The authors thank the reviewers for their constructive comments and suggestions.

REFERENCES

- [1] M. AFONSO, J. BIOCAS-DIAS, AND M. FIGUEIREDO, *Fast image recovery using variable splitting and constrained optimization*, IEEE Trans. Image Process., 19 (2010), pp. 2345–2356.
- [2] J. BARZILAI AND J. M. BORWEIN, *Two point step size gradient methods*, IMA J. Numer. Anal., 8 (1988), pp. 141–148.
- [3] S. BECKER, J. BOBIN, AND E. J. CANDÈS, *NESTA: a fast and accurate first-order method for sparse recovery*, SIAM J. Imaging Sci., 4 (2011), pp. 1–39.
- [4] K. BLOCK, M. UECKER, AND J. FRAHM, *Undersampled radial MRI with multiple coils: Iterative image reconstruction using a total variation constraint*, Magn. Reson. Med., 57 (2007), pp. 1086–1098.
- [5] R. BROUGHTON, I. COOPE, P. RENAUD, AND R. TAPPENDEN, *A box constrained gradient projection algorithm for compressed sensing*, IEEE Trans. Signal Process., 91 (2011), pp. 1985–1992.
- [6] R. E. BRUCK, *On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space*, J. Math. Anal. Appl., 61 (1977), pp. 159–164.
- [7] T. F. CHAN, G. H. GOLUB, AND P. MULET, *A nonlinear primal-dual method for total variation based image restoration*, SIAM J. Optim., 20 (1999), pp. 1964–1977.
- [8] Y. CHEN, W. W. HAGER, F. HUANG, D. T. PHAN, X. YE, AND WOTAO YIN, *Fast algorithms for image reconstruction with application to partially parallel MR imaging*, SIAM J. Imaging Sci., 5 (2012), pp. 90–118.
- [9] Y. CHEN, W. W. HAGER, M. YASHTINI, X. YE, AND H. ZHANG, *Bregman operator splitting with variable stepsize for total variation image reconstruction*, Comput. Optim. Appl., 54 (2013), pp. 317–342.
- [10] Y. CHEN, G. LAN, AND Y. OUYANG, *Optimal primal-dual methods for a class of saddle point problems*, tech. report, 2013, submitted.
- [11] Y. CHEN, M. RAO, Y. TONEGAWA, AND T. WUNDERLI, *Partial regularity for a selective smoothing functional for image restoration*, SIAM J. Math. Anal., 37 (2005), pp. 1098–1116.
- [12] J. DARBON AND M. SIGELLE, *A fast and exact algorithm for total variation minimization*, in Pattern Recognition and Image Analysis, J. S. Marques, N. P. de la Blanca, and P. Pina, eds., vol. 3522 of Lecture Notes in Comput. Sci., Springer-Verlag, Berlin, 2005, pp. 351–359.
- [13] D. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 351–359.
- [14] E. ESSER, X. ZHANG, AND T. CHAN, *A general framework for a class of first order primal-dual algorithms for tv minimization*, SIAM J. Imag. Sci., 3 (2010), pp. 1015–1046.
- [15] D. GABAY, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrange methods: applications to the solution of boundary-valued problems, M. Fortin and R. Glowinski, eds., Amsterdam, 1983, North Holland, pp. 299–331.
- [16] D. GABAY AND B. MERCIER, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Comput. Math. Appl., 2 (1976), pp. 17–40.
- [17] R. GLOWINSKI AND A. MARROCCO, *Sur l’approximation par éléments finis d’ordre un, et la résolution par pénalisation-dualité d’une classe de problèmes de dirichlet nonlinéaires*, RAIRO Analyse Numérique, 9 (1975), pp. 41–76.
- [18] T. GOLDSTEIN AND S. OSHER, *The split Bregman method for L1 regularized problems*, SIAM J. Imag. Sci., 2 (2009), pp. 323–343.
- [19] W. 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*, J. Oper. Res. Soc. China, 3 (2015), pp. 139–162.
- [20] W. W. HAGER, D. T. PHAN, AND H. ZHANG, *Gradient-based methods for sparse recovery*, SIAM J. Imaging Sci., 4 (2011), pp. 146–165.
- [21] E. HALE, W. YIN, AND Y. ZHANG, *A fixed-point continuation method for ℓ_1 -regularized minimization with applications to compressed sensing*, SIAM J. Optim., 19 (2008), pp. 1107–1130.

- [22] B. HE AND X. YUAN, *On the $\mathcal{O}(1/n)$ convergence rate of the Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709.
- [23] Y. HUANG, M. K. NG, AND Y. WEN, *A new total variation method for multiplicative noise removal*, SIAM J. Imag. Sci., 2 (2009), pp. 20–40.
- [24] G. LAN AND R. D. C. MONTEIRO, *Iteration-complexity of first-order augmented Lagrangian methods for convex programming*, tech. report, School of ISyE, Georgia Tech, Atlanta, GA, 30332, USA, 2009.
- [25] Y. LI AND F. SANTOSA, *A computational algorithm for minimizing total variation in image restoration*, IEEE Trans. Image Process., 5 (1996), pp. 987–995.
- [26] MICHAEL LUSTIG, DAVID DONOHO, AND JOHN M. PAULY, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magn. Reson. Med., 58 (2007), pp. 1182–1195.
- [27] R. D. C. MONTEIRO AND B. F. SVAITER, *On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean*, SIAM J. Optim., 20 (2010), pp. 2755–2787.
- [28] ———, *Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers*, SIAM J. Optim., 23 (2013), pp. 475–507.
- [29] A. NEMIROVSKI, *Efficient methods in convex programming*, tech. report, Technion – The Isreal Institute of Technology, Haifa, Israel, 1994.
- [30] A. NEMIROVSKI, *Prox-method with rate of convergence $\mathcal{O}(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems*, SIAM J. Optim., 15 (2005), pp. 229–251.
- [31] YURIH NESTEROV, *Introductory lectures on convex optimization : a basic course*, Applied optimization, Kluwer Academic Publ., Boston, Dordrecht, London, 2004.
- [32] Y. E. NESTEROV, *Smooth minimization of nonsmooth functions*, Math. Program., 103 (2005), pp. 127–152.
- [33] K.P. PRUESSMANN, M. WEIGER, P. BORNERT, AND P. BOESIGER, *Advances in sensitivity encoding with arbitrary k -space trajectories*, Magn. Reson. Med., 46 (2001), pp. 638–651.
- [34] R. T. ROCKAFELLAR, *Convex analysis*, Princeton Univ. Press, 1970.
- [35] L. RUDIN, S. OSHER, AND E. FATEMI, *Non-linear total variation noise removal algorithm*, Physics D., 60 (1992), pp. 259–268.
- [36] RON SHEFI AND MARC TEBoulLE, *Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization*, SIAM J. Optim., 24 (2014), pp. 269–297.
- [37] C. R. VOGEL, *A multigrid method for total variation-based image denoising*, in Computation and Control IV, K. Bowers and J. Lund, eds., vol. 20 of Progress in Systems and Control Theory, Boston, 1995, Birkhauser, pp. 323–331.
- [38] C. R. VOGEL AND M. E. OMAN, *Iterative methods for total variation denoising*, SIAM J. Sci. Comput., 17 (1996), pp. 227–238.
- [39] Y. WANG, J. YANG, W. YIN, AND Y. ZHANG, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM J. Imag. Sci., 1 (2008), pp. 248–272.
- [40] S. J. WRIGHT, R. D. NOWAK, AND M. A. T. FIGUEIREDO, *Sparse reconstruction by separable approximation*, IEEE Trans. Signal Process., 57 (2009), pp. 2479–2493.
- [41] C. WU AND X.-C. TAI, *Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models*, SIAM J. Imaging Sci., 3 (2010), pp. 300–339.
- [42] C. WU, J. ZHANG, AND X.-C. TAI, *Augmented Lagrangian method for total variation restoration with non-quadratic fidelity*, Inverse Problems and Imaging, 5 (2011), pp. 237–261.
- [43] J. YANG, W. YIN, Y. ZHANG, AND Y. WANG, *A fast algorithm for edge-preserving variational multichannel image restoration*, SIAM J. Imag. Sci., 2 (2010), pp. 569–592.
- [44] J. YANG, Y. ZHANG, AND W. YIN, *A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data*, IEEE J. Selected Topics Signal Process., 4 (2010), pp. 288–297.
- [45] M. YASHTINI, W. W. HAGER, Y. CHEN, AND X. YE, *Partially parallel MR image reconstruction using sensitivity encoding*, in 2012 IEEE International Conference on Image Processing, Orlando, 2012, IEEE, pp. 2077–2080.
- [46] X. YE, Y. CHEN, AND F. HUANG, *MR image reconstruction via sparse representation: Modeling and algorithm*, International Conference on Image Processing, Computer Vision, and Pattern Recognition (ICCV), (2009), pp. 10–16.
- [47] ———, *Computational acceleration for MR image reconstruction in partially parallel imaging*, IEEE Trans. Med. Imaging, 30 (2011), pp. 1055–1063.
- [48] L. YING, B. LIU, M.C. STECKNER, G. WU, AND S.-J. LI M. WU, *A statistical approach to SENSE regularization with arbitrary k -space trajectories*, Magn. Reson. Med., 60 (2008), pp. 414–421.

- [49] X. ZHANG, M. BURGER, X. BRESSON, AND S. OSHER, *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, SIAM J. Imaging Sci., 3 (2011), pp. 253–276.
- [50] XIAOQUN ZHANG, MARTIN BURGER, AND STANLEY OSHER, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput., 46 (2011), pp. 20–46.
- [51] M. ZHU AND T.F. CHAN, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, Tech. Report 08-34, CAM UCLA, 2008.
- [52] M. ZHU, S.J. WRIGHT, AND T.F. CHAN, *Duality-based algorithms for total-variation-regularized image restoration*, Comput. Optim. Appl., 47 (2010), pp. 377–400.