Comput Optim Appl (2013) 54:317–342
DOI 10.1007/s10589-012-9519-2

Bregman operator splitting with variable stepsize for total variation image reconstruction

Yunmei Chen · William W. Hager · Maryam Yashtini · Xiaojing Ye · Hongchao Zhang

Received: 25 January 2012 / Published online: 4 December 2012

Abstract This paper develops a Bregman operator splitting algorithm with variable stepsize (BOSVS) for solving problems of the form $\min\{\phi(Bu) + 1/2\|Au - f\|_2^2\}$, where ϕ may be nonsmooth. The original Bregman Operator Splitting (BOS) algorithm employed a fixed stepsize, while BOSVS uses a line search to achieve better efficiency. These schemes are applicable to total variation (TV)-based image reconstruction. The stepsize rule starts with a Barzilai-Borwein (BB) step, and increases the nominal step until a termination condition is satisfied. The stepsize rule is related

This research was partly supported by National Science Foundation grants 1115568 and 1016204, and by Office of Naval Research grant N00014-11-1-0068.

Y. Chen · W.W. Hager · M. Yashtini (✉)
Department of Mathematics, University of Florida, P.O. Box 118105, Gainesville, FL 32611-8105, USA
e-mail: myashtini@ufl.edu
url: http://www.math.ufl.edu/~myashtini

Y. Chen
e-mail: yun@math.ufl.edu
url: http://www.math.ufl.edu/~yun

W.W. Hager
e-mail: hager@ufl.edu
url: http://www.math.ufl.edu/~hager

X. Ye
School of Mathematics, Georgia Institute of Technology, 686 Cherry Street, Atlanta, GA 30332-0160, USA
e-mail: xye33@math.gatech.edu
url: http://people.math.gatech.edu/~xye33

H. Zhang
Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803-4918, USA
e-mail: hozhang@math.lsu.edu
url: https://www.math.lsu.edu/~hozhang

to the scheme used in SpaRSA (Sparse Reconstruction by Separable Approximation). Global convergence of the proposed BOSVS algorithm to a solution of the optimization problem is established. BOSVS is compared with other operator splitting schemes using partially parallel magnetic resonance image reconstruction problems. The experimental results indicate that the proposed BOSVS algorithm is more efficient than the BOS algorithm and another split Bregman Barzilai-Borwein algorithm known as SBB.

Keywords Total variation image reconstruction · Bregman operator splitting · Barzilai-Borwein stepsize · SpaRSA · Convergence analysis · Magnetic resonance imaging

1 Introduction

Image reconstruction has been widely applied to many applications in computer science and engineering. The general form for this problem minimizes a functional involving two terms, a fidelity term H related to the data and a regularization term ϕ:

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

where $u = (u_1, \ldots, u_N)^{\mathsf{T}} \in \mathbb{C}^N$, u_i is the intensity of the i-th pixel in the image, N is the number of pixels in the image, $\phi(\cdot)$ and $H(\cdot)$ are convex, real-valued functions, $\phi : \mathbb{C}^{d\times N} \to \mathbb{R}$ is possibly nondifferentiable, $H : \mathbb{C}^N \to \mathbb{R}$ is continuously differentiable, $B \in \mathbb{C}^{d\times N\times N}$, and Bu is a $d \times N$ matrix:

$$(Bu)_{ij} = \sum_{k=1}^{N} b_{ijk}u_k$$

Since ϕ and H are finite-valued functions, it follows from their convexity that they are continuous [16, Corollary 10.1.1].

In this paper, we focus a specific version of this problem, which arises in imaging and signal processing, where H is quadratic:

$$\min_{u\in\mathbb{C}^N} \phi(Bu) + \frac{1}{2}\|Au - f\|^2. \tag{1.1}$$

Here $A \in \mathbb{C}^{M\times N}$ is a possibly large and ill-conditioned matrix, $f \in \mathbb{C}^M$, and $\|\cdot\|$ is the 2-norm (Euclidean norm). In imaging applications, the matrix A describes the imaging device, f is the measured data, and d is the dimension of the image. For the examples in this paper, $d = 2$. For discrete total variation regularization, $Bu = \nabla u$ where $(\nabla u)_i$ is a discrete gradient (finite differences along the coordinate directions) of $u \in \mathbb{C}^N$ at the i-th pixel in the image, and

$$\phi(Bu) = \alpha\|u\|_{TV} = \alpha\sum_{i=1}^{N} \big\|(\nabla u)_i\big\|. \tag{1.2}$$

Here $\alpha > 0$ is a parameter corresponding to the weight of the regularization term $\|u\|_{TV}$ relative to the fidelity term $\frac{1}{2}\|Au - f\|^2$. Hence, discrete total variation regularization is based on the sum of the Euclidean norms of the discrete gradients at each pixel.

The problem (1.1) has received considerable attention [9, 21, 23, 25–28] due to its application in signal and image processing including partially parallel magnetic resonance (MR) imaging (PPI) [3, 5, 15, 23, 24], compressed sensing [7, 9, 14], and radar [1]. If A is the identity matrix or a blurring matrix, then solving (1.1) yields a restored clean image u from an observed noisy or blurry image f. Image denoising by total variation regularization was first introduced by Rudin, Osher, and Fetami in [17]. One of the significant advantages of TV regularization is that it allows discontinuities in the restored image; hence, it is able to preserve edges while suppressing noise [18]. Despite the simplicity of the objective function in (1.1), the lack of smoothness makes the TV-regularized problem difficult. Rudin et al. in their pioneering work [17] exploited the gradient projection method to minimize the energy. This method was observed to converge slowly due to the nonlinearity and poor conditioning of the problem. Subsequent papers [4, 6, 12, 13, 19, 20] presented algorithms with accelerated convergence.

The algorithm to be analyzed in this paper is based on ideas given in [24] where a variable splitting and the alternating direction method of multipliers are combined with a Barzilai/Borwein (BB) approximation to the Hessian. Due to differences between the algorithm in [24] and the algorithm that we analyze, including differences in the formula for the BB parameter and in the treatment of proximal terms, we provide a short review of the ideas. First, the objective function in (1.1) is split by introducing an auxiliary variable w and adding a constraint:

$$\min \phi(w) + \frac{1}{2}\|Au - f\|^2 \quad \text{subject to} \quad w = Bu, \quad u \in \mathbb{C}^N, \quad w \in \mathbb{C}^{d\times N}. \tag{1.3}$$

Note that splitting techniques can be found in [9, 21, 23]. The augmented Lagrangian associated with (1.3) is

$$\mathcal{L}^\rho(u, w, b) = \phi(w) + \frac{1}{2}\|Au - f\|^2 + \langle b, Bu - w\rangle + \frac{\rho}{2}\|Bu - w\|^2, \tag{1.4}$$

where $b \in \mathbb{C}^{d\times N}$ is a Lagrange multiplier, and $\langle\cdot,\cdot\rangle$ is the Euclidean inner product. If b^k is the current approximation to the multiplier, then the gradient version of the method of multipliers [11] minimizes the Lagrangian $\mathcal{L}^\rho(u, w, b^k)$ in (1.4) with respect to (u, w) and then updates the multiplier:

$$\begin{cases} \left(u^{k+1}, w^{k+1}\right) = \arg\min\limits_{u,w} \mathcal{L}^\rho\left(u, w, b^k\right) \\ b^{k+1} = b^k + \rho\left(Bu^{k+1} - w^{k+1}\right). \end{cases}$$

We complete the square in $\mathcal{L}^\rho(u, w, b^k)$ to write this as

$$\begin{cases} \left(u^{k+1}, w^{k+1}\right) = \arg\min\limits_{u,w} \left\{\phi(w) + \frac{1}{2}\|Au - f\|^2 + \frac{\rho}{2}\left\|Bu - w + \rho^{-1}b^k\right\|^2\right\} \\ b^{k+1} = b^k + \rho\left(Bu^{k+1} - w^{k+1}\right). \end{cases}$$

We now modify the (u, w) minimization to ensure uniqueness as well as a fast iteration. The first modification is to insert a proximal term in w:

$$\begin{cases} \left(u^{k+1}, w^{k+1}\right) = \arg\min\limits_{u,w} \left\{ \phi(w) + \dfrac{1}{2}\|Au - f\|^2 + \dfrac{\rho}{2}\left\|Bu - w + \rho^{-1}b^k\right\|^2 \right. \\ \qquad\qquad \left. + \dfrac{\beta}{2}\left\|w - w^k\right\|^2 \right\} \\ b^{k+1} = b^k + \rho\left(Bu^{k+1} - w^{k+1}\right), \end{cases} \tag{1.5}$$

where $\beta > 0$. The second modification involves the term:

$$H(u) = \frac{1}{2}\|Au - f\|^2. \tag{1.6}$$

The expansion of H in a Taylor series around u^k can be expressed

$$H(u) = H\left(u^k\right) + \left\langle \nabla H\left(u^k\right), u - u^k \right\rangle + \frac{1}{2}\left\langle u - u^k, \nabla^2 H\left(u - u^k\right) \right\rangle, \tag{1.7}$$

where the Hessian $\nabla^2 H = A^\mathsf{T} A$. In magnetic resonance imaging and other applications, $A^\mathsf{T} A$ can be large and dense. We make the approximation $\nabla^2 H \approx \delta_k I$. With this substitution, (1.7) gives

$$H(u) \approx \frac{1}{2}\left\|Au^k - f\right\|^2 + \left\langle A^\mathsf{T}\left(Au^k - f\right), u - u^k \right\rangle + \frac{\delta_k}{2}\left\|u - u^k\right\|^2. \tag{1.8}$$

In this paper, we exploit the Barzilai-Borwein (BB) choice for δ_k given by [2]

$$\begin{aligned} \delta_k &= \arg\min_{\delta} \left\|\nabla H\left(u^k\right) - \nabla H\left(u^{k-1}\right) - \delta\left(u^k - u^{k-1}\right)\right\|^2 \\ &= \frac{\|A(u^k - u^{k-1})\|^2}{\|u^k - u^{k-1}\|^2}. \end{aligned} \tag{1.9}$$

By completing square, (1.8) can be written

$$H(u) \approx \frac{1}{2}\left\|Au^k - f\right\|^2 - \frac{1}{2\delta_k}\left\|A^\mathsf{T}\left(Au^k - f\right)\right\|^2 + \frac{\delta_k}{2}\left\|u - u^k + {\delta_k}^{-1}A^\mathsf{T}\left(Au^k - f\right)\right\|^2.$$

With this replacement in (1.5), we obtain

$$\begin{cases} \left(u^{k+1}, w^{k+1}\right) = \arg\min\limits_{u,w} \left\{ \phi(w) + \dfrac{\delta_k}{2}\left\|u - u^k + {\delta_k}^{-1}A^\mathsf{T}\left(Au^k - f\right)\right\|^2 \right. \\ \qquad\qquad \left. + \dfrac{\rho}{2}\left\|Bu - w + \rho^{-1}b^k\right\|^2 + \dfrac{\beta}{2}\left\|w - w^k\right\|^2 \right\} \\ b^{k+1} = b^k + \rho\left(Bu^{k+1} - w^{k+1}\right). \end{cases} \tag{1.10}$$

To further simplify the minimization, we utilize the alternating direction method of multipliers (ADMM) introduced by Gabay in [8]. This approximates the (u, w)-minimization by the alternating minimization over u with w fixed and over w with u fixed. That is

$$\begin{cases} u^{k+1} = \arg\min_u \left\{ \delta_k \left\| u - u^k + \delta_k^{-1} A^{\mathsf{T}} \left(A u^k - f \right) \right\|^2 + \rho \left\| B u - w^k + \rho^{-1} b^k \right\|_2^2 \right\} \\ w^{k+1} = \arg\min_w \left\{ \phi(w) + \dfrac{\rho}{2} \left\| w - B u^{k+1} - \rho^{-1} b^k \right\|^2 + \dfrac{\beta}{2} \left\| w - w^k \right\|^2 \right\} \\ b^{k+1} = b^k + \rho \left(B u^{k+1} - w^{k+1} \right). \end{cases} \tag{1.11}$$

This algorithm with $\delta_k = \delta$, independent of k, is a special case of the general algorithm Algorithm A1 in [26] corresponding to the following parameter choices: $Q_1 = \delta I - A^{\mathsf{T}} A$, $Q_2 = \beta I$, and $C = 1/\rho$. Convergence follows from Theorem 4.2 of [26] for $\delta > \|A^{\mathsf{T}} A\|$. The idea of changing δ in each iteration using the BB choice (1.9) appears in [24]. In [5] this split Bregman BB scheme is referred to as SBB. For the BB choice of δ_k, we have $\delta_k \leq \|A^{\mathsf{T}} A\|$ while the convergence theory in [26] requires $\delta > \|A^{\mathsf{T}} A\|$. Nonetheless, the numerical performance for SBB was better than that of other algorithms based on splittings.

In this paper we prove theoretically global convergence when δ_k is chosen in accordance with a line search criterion. Numerically, we obtain significantly better performance compared with the fixed stepsize and BB stepsize schemes. We let BOSVS denote this Bregman operator splitting algorithm with our variable stepsize.

1.1 Outline of the paper

This paper is organized as follows. In Sect. 2, we present BOSVS and discuss its implementation in the context of total variation regularization. Section 3 gives a convergence analysis for BOSVS. The relation between the line search in BOSVS and that of SpaRSA is explained in Sect. 4. Section 5 studies the performance of BOSVS relative to that of either SBB or the fixed step size $\delta_k = \|A^{\mathsf{T}} A\|$ for Algorithm A1 of [26] using test problem that arise in partially parallel MR imaging (PPI). Finally, some concluding remarks are given in Sect. 6.

1.2 Notation

For any matrix M, $\mathcal{N}(M)$ is the null space of M. The superscript $^{\mathsf{T}}$ denotes the conjugate transpose. For any vectors x and $y \in \mathbb{C}^N$, $\langle x, y \rangle$ is the Euclidean inner product. When x and y are column vectors, $\langle x, y \rangle = x^{\mathsf{T}} y$. The norm $\|\cdot\|$ is the 2-norm (Euclidean norm) given by $\|x\| = \sqrt{\langle x, x \rangle}$. For matrices X and $Y \in \mathbb{C}^{m \times n}$, their inner product is

$$\langle X, Y \rangle = \sum_{i=1}^{n} \langle X_i, Y_i \rangle$$

where X_i and Y_i are the i-th columns of X and Y respectively. The adjoint B^* of B is given by

$$\langle X, Bu\rangle = \left\langle B^*X, u\right\rangle$$

for all $X \in \mathbb{C}^{d\times N}$ and $u \in \mathbb{C}^N$. For a differentiable function $f : \mathbb{C}^N \to \mathbb{R}$, ∇f is the gradient of f, a row vector. More generally, $\partial f(x)$ denotes the subdifferential set at x.

2 Proposed algorithm

The algorithm that we analyze in this paper appears in Fig. 1. Note that the updates of (1.11) appear in Steps 2, 4, and 5 of BOSVS. The new feature in BOSVS is the line search strategy in Step 2. δ_k increases until the condition in Step 2 is satisfied. For the numerical experiments, the initial choice of δ_k is a safeguarded BB choice

$$\delta_{0,k} = \max\left\{\delta_{\min}, \frac{\|A(u^k - u^{k-1})\|^2}{\|u^k - u^{k-1}\|^2}\right\}. \tag{2.1}$$

BREGMAN OPERATOR SPLITTING WITH VARIABLE STEPSIZE (BOSVS)

Given $\tau, C, \eta > 1$, $\beta, \rho, \delta_{\min} > 0$, $\delta_0 = 1$, $\sigma \in (0,1)$, and starting guess u^1, w^1, and b^1. Set $k = 1$ and $Q_1 = 0$.

Step 1. Choose $\delta_{0,k} \geq \delta_{\min}$.

Step 2. Set $\delta_k = \eta^j \delta_{0,k}$ where $j \geq 0$ is the smallest integer such that $Q_{k+1} \geq -\frac{C}{k^2}$ where
$Q_{k+1} := \bar{\eta}_k Q_k + \Delta_k$, $0 \leq \bar{\eta}_k \leq (1-k^{-1})^2$ for $k > 1$,
$\Delta_k := \sigma(\delta_k \|u^{k+1} - u^k\|^2 + \rho\|Bu^{k+1} - w^k\|^2) - \|A(u^{k+1} - u^k)\|^2$,
and $u^{k+1} = \arg\min_u \{\delta_k \|u - u^k + \delta_k^{-1} A^{\mathsf{T}}(Au^k - f)\|^2 + \rho\|Bu - w^k + \rho^{-1} b^k\|^2\}$.

Step 3. If $\delta_k > \delta_{k-1}$, then $\delta_{\min}$ is replaced by $\tau\delta_{\min}$.

Step 4. $w^{k+1} = \arg\min_w \{\phi(w) + \frac{\rho}{2}\|w - Bu^{k+1} - \rho^{-1} b^k\|^2 + \frac{\beta}{2}\|w - w^k\|^2\}$.

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.

Step 7. Set $k = k+1$ and go to step 1.

Fig. 1 BOSVS

The stepsize rule for δ_k in Step 2 of BOSVS is related but different from the stepsize rule introduced in the SpaRSA algorithm [10, 22] (Sparse Reconstruction by Separable Approximation). More precisely, in Sect. 3 we show that a monotone SpaRSA step applied to the u-minimization of (1.11) is acceptable when

$$\left\| A\left(u^{k+1}-u^k\right)\right\|^2 \le (2-\sigma)\delta_k \left\|u^{k+1}-u^k\right\|^2 + \rho \left\| B\left(u^{k+1}-u^k\right)\right\|^2, \tag{2.2}$$

while Step 2 amounts to requiring that

$$\left\| A\left(u^{k+1}-u^k\right)\right\|^2 \le \sigma\delta_k \left\|u^{k+1}-u^k\right\|^2 + \sigma\rho \left\| Bu^{k+1}-w^k\right\|^2 + \bar{\eta}_k Q_k + \frac{C}{k^2}.$$

Observe that $\Delta_k \ge 0$ if

$$\delta_k \ge \frac{\|A(u^{k+1}-u^k)\|^2}{\sigma \|u^{k+1}-u^k\|^2}, \tag{2.3}$$

and in this case, we have

$$Q_{k+1} = \bar{\eta}_k Q_k + \Delta_k \ge \bar{\eta}_k Q_k \ge -\bar{\eta}_k C/(k-1)^2 \ge -C/k^2$$

since $\bar{\eta}_k \le (k-1)^2/k^2$. Hence, when

$$\delta_k \ge \|A\|^2/\sigma \ge \frac{\|A(u^{k+1}-u^k)\|^2}{\sigma \|u^{k+1}-u^k\|^2},$$

Step 2 terminates with $j = 0$.

The convergence analysis relies on the asymptotic monotonicity of δ_k. Whenever δ_k is not monotone decreasing, $\delta_{\min}$ is increased by a factor $\tau > 1$ in Step 3. Hence, if the monotonicity of δ_k continues to be violated, then we eventually have

$$\delta_{\min} > \frac{\|A\|^2}{\sigma} \ge \frac{\|A(u^k-u^{k-1})\|^2}{\sigma \|u^k-u^{k-1}\|^2}. \tag{2.4}$$

If $\delta_{0,k}$ is given by (2.1), then $\delta_{0,k} = \delta_{\min}$ whenever $\delta_{\min} > \|A\|^2$. Moreover, if (2.4) holds, then (2.3) holds, and by our previous observation, Step 2 stops with $j = 0$. Hence, within a finite number of iterations, either δ_k is monotone decreasing or δ_k becomes constant and equal to $\delta_{\min}$. If $\delta_{\min}$ is large enough, then the convergence analysis in [25] could be applied. In practice, we find that the algorithm reaches a regime where $\delta_{k+1} \le \delta_k$ for each k sufficiently large; as a result, δ_k approaches a constant which is usually smaller than $\|A\|^2$, and which leads to faster convergence than is possible with a constant value of δ_k that is greater than $\|A\|^2$. On the other hand, numerical experiments in Sect. 5 indicate that if δ_k is simply set to the BB parameter in each iteration, then the iterates may not converge to a solution of (1.3).

For total variation regularization, the minimization in Steps 2 and 4 of BOSVS can be implemented using Fourier transforms and soft shrinkage as pointed out in [24]. For completeness, we summarize these implementation details. By the first-order optimality conditions associated with the optimization problem for u^{k+1}, we have

$$\big(\rho B^*B + \delta_k I\big)u^{k+1} = \delta_k u^k - A^{\mathsf{T}}\big(Au^k - f\big) + \rho B^*\big(w^k - \rho^{-1}b^k\big). \tag{2.5}$$

If the image satisfies periodic boundary conditions and if we use total variation regularization, then the matrix B^*B is block circulant; hence, it can be diagonalized by the Fourier transform matrix $\mathcal{F}$ as noted in [21, p. 252]. If $\Delta = \mathcal{F}B^*B\mathcal{F}^{\mathsf{T}}$ is the diagonalization, then

$$u^{k+1} = \mathcal{F}^{\mathsf{T}}(\rho\Delta + \delta_k I)^{-1}\mathcal{F}\big[\delta_k u^k - A^{\mathsf{T}}\big(Au^k - f\big) + \rho B^*\big(w^k - \rho^{-1}b^k\big)\big].$$

For total variation regularization, the solution of

$$w^{k+1} = \arg\min_w \left\{\phi(w) + \frac{\rho}{2}\big\|w - Bu^{k+1} - \rho^{-1}b^k\big\|^2 + \frac{\beta}{2}\big\|w - w^k\big\|^2\right\}$$

is obtained by soft shrinkage [21]:

$$w^{k+1} = \mathsf{shrink}\left\{\frac{\rho(Bu^{k+1} + \rho^{-1}b^k) + \beta w^k}{\rho + \beta}, \frac{\rho + \beta}{\alpha}\right\},$$

where $\mathsf{shrink}(t, \mu) = \frac{t}{\|t\|}\max\{\|t\| - \frac{1}{\mu}, 0\}$ with the convention $(\frac{0}{\|0\|} = 0)$.

3 Convergence analysis

In this section we prove that the sequence (u^k, w^k, b^k) generated by BOSVS converges to a solution of (1.3). We begin with an existence result.

Lemma 3.1 *If $\mathcal{N}(A) \cap \mathcal{N}(B) = \{0\}$ and $\phi(w)$ tends to infinity as $\|w\|$ tends to infinity, then there exists a solution to* (1.1)*, or equivalently to* (1.3)*.*

Proof Suppose that u^k is a minimizing sequence for (1.1). Hence there exists a positive κ such that

$$\phi\big(Bu^k\big) \le \phi\big(Bu^k\big) + \frac{1}{2}\big\|Au^k - f\big\|^2 \le \kappa, \tag{3.1}$$

for all k. Since $\phi(Bu^k)$ is uniformly bounded in k, it follows that $\|Bu^k\|$ is uniformly bounded in k. Write $u^k = b^k + q^k$ where $b^k \in \mathcal{N}(B)$ and q^k is orthogonal to $\mathcal{N}(B)$. Since $\|Bu^k\| = \|Bq^k\|$ and q^k is orthogonal to $\mathcal{N}(B)$, $\|q^k\|$ is bounded uniformly in k.

Now write $u^k = a^k + p^k$ where $a^k \in \mathcal{N}(A)$ and p^k is orthogonal to $\mathcal{N}(A)$; likewise write $f = f_\parallel + f_\perp$ where $f_\parallel$ lies in the range of A and $f_\perp$ is orthogonal to the range of A. We have

$$\left\|Au^k - f\right\|^2 = \left\|Ap^k - f_\parallel\right\|^2 + \|f_\perp\|^2 = \left\|A\left(p^k - z\right)\right\|^2 + \|f_\perp\|^2$$

where $Az = f_\parallel$ and z is orthogonal to $\mathcal{N}(A)$. Combining this with (3.1) gives

$$\frac{1}{2}\left\|A\left(p^k - z\right)\right\|^2 \le \kappa - \frac{1}{2}\|f_\perp\|^2 - \phi\left(Bq^k\right).$$

Since q^k is bounded uniformly in k, it follows that $\|A(p^k - z)\|$ is bounded uniformly in k, and since $p^k - z$ is orthogonal to $\mathcal{N}(A)$, we conclude that $\|p^k - z\|$ is bounded uniformly in k; hence, $\|p^k\|$ is bounded uniformly in k.

Recall that $u^k = a^k + p^k$ where $a^k \in \mathcal{N}(A)$ and p^k is orthogonal to $\mathcal{N}(A)$. Let us further write $a^k = c^k + r^k$ where $c^k \in \mathcal{N}(B)$ and r^k is orthogonal to $\mathcal{N}(B)$. By (3.1),

$$\kappa \ge \phi\left(Bu^k\right) = \phi\left(B\left(c^k + r^k + p^k\right)\right) = \phi\left(B\left(r^k + p^k\right)\right).$$

Hence, $\|B(p^k + r^k)\|$ is bounded uniformly in k. Since p^k is bounded and r^k is orthogonal to $\mathcal{N}(B)$, we conclude that $\|r^k\|$ is bounded uniformly in k.

Consider the following minimization problem:

$$\gamma = \min_{\substack{a \in \mathcal{N}(A) \\ \|a\|=1}} \min_{b \in \mathcal{N}(B)} \|a - b\|.$$

Since the minimization is over a compact set and $\mathcal{N}(A) \cap \mathcal{N}(B) = \{0\}$, it follows that $\gamma > 0$.

Since $r^k = a^k - c^k$ where $a^k \in \mathcal{N}(A)$ and $c^k \in \mathcal{N}(B)$, we have

$$\left\|r^k\right\| = \left\|a^k - c^k\right\| = \left\|a^k\right\| \left\| \frac{a^k}{\|a^k\|} - \frac{c^k}{\|a^k\|} \right\| \ge \left\|a^k\right\| \gamma.$$

Since $\|r^k\|$ is bounded uniformly in k and $\gamma > 0$, we conclude that $\|a^k\|$ is bounded uniformly in k. Consequently, $\|u^k\| = \|a^k + p^k\|$ is bounded uniformly in k. By continuity of the objective function, a minimizer exists. □

Lemma 3.2

(I) *In Step* 2 *of the* BOSVS *algorithm*, $j \le \lceil \log_\eta(\frac{\|A^\mathsf{T}A\|}{\sigma\delta_{\min}}) \rceil$, *where* $\lceil x \rceil$ *is the smallest integer greater than or equal to* x *for any* $x \in \mathbb{R}$.

(II) *Moreover, the replacement of* $\delta_{\min}$ *by* $\tau\delta_{\min}$ *in Step* 3 *of* BOSVS *can occur in at most a finite number of iterations*.

Proof The line search condition in Step 2 of BOSVS is satisfied by the smallest j for which

$$\frac{C}{k^2} + \bar{\eta}_k Q_k + \sigma\left(\eta^j \delta_{0,k}\left\|u^{k+1} - u^k\right\|^2 + \rho\left\|Bu^{k+1} - u^k\right\|^2\right) \ge \left\|A\left(u^{k+1} - u^k\right)\right\|^2.$$

We rearrange this to obtain

$$\eta^j \geq \frac{\sigma^{-1}}{\delta_{0,k}} \frac{\|A(u^{k+1}-u^k)\|^2}{\|u^{k+1}-u^k\|^2} - \frac{\frac{C}{k^2}+\bar{\eta}_k Q_k + \sigma\rho\|Bu^{k+1}-w^k\|^2}{\sigma\delta_{0,k}\|u^{k+1}-u^k\|^2}. \tag{3.2}$$

Hence, the smallest j in Step 2 will be the first nonnegative integer where (3.2) holds. Neglecting the negative term and taking logarithm to base η, we have

$$j \leq \left\lceil \log_\eta \left(\frac{\sigma^{-1}\|A^\mathsf{T}A\|}{\delta_{\min}} \right) \right\rceil,$$

since $\delta_{0,k} \geq \delta_{\min}$ and $\|A^\mathsf{T}A\| \geq \|A(u^{k+1}-u^k)\|^2/\|u^{k+1}-u^k\|^2$.

Note that (II) holds since Step 2 terminates at $j=0$ and $\delta_k = \delta_{\min}$ when $\delta_{\min}$ is larger than $\frac{\|A^\mathsf{T}A\|}{\sigma}$ as we showed earlier. □

Lemma 3.3 *If $s^k \in \partial\phi(w^k)$ and $\overline{s} \in \partial\phi(\overline{w})$, then*

$$\langle s^k - \overline{s}, w^k - \overline{w}\rangle \geq 0. \tag{3.3}$$

Moreover, if $\lim_{k\to\infty}\langle s^k-\overline{s}, w^k-\overline{w}\rangle = 0$, then

(a) $\lim_{k\to\infty}\phi(w^k) - \phi(\overline{w}) - \langle \overline{s}, w^k-\overline{w}\rangle = 0$, *and*
(b) $\lim_{k\to\infty}\phi(\overline{w}) - \phi(w^k) - \langle s^k, \overline{w}-w^k\rangle = 0$.

Proof For a convex function ϕ, we have

$$\phi(w^k) - \phi(\overline{w}) - \langle \overline{s}, w^k-\overline{w}\rangle \geq 0, \tag{3.4}$$

and

$$\phi(\overline{w}) - \phi(w^k) - \langle s^k, \overline{w}-w^k\rangle \geq 0. \tag{3.5}$$

Add (3.4) and (3.5) to obtain (3.3). If (3.3) approaches to zero as k goes to infinity, then both (3.4) and (3.5) approach zero. □

Theorem 3.4 *If there exists a solution of* (1.1)*, then the sequence (u^k, w^k, b^k) generated by* BOSVS *approaches a point (u^*, w^*, b^*) where the first-order optimality conditions for* (1.3) *are satisfied. Moreover, (u^*, w^*) is a solution of* (1.3) *and u^* is a solution of* (1.1).

Proof The first-order optimality conditions for the sequence $(u^{k+1}, w^{k+1}, b^{k+1})$ generated by BOSVS are

$$\begin{cases} 0 = \delta_k\left(u^{k+1}-u^k+\delta_k{}^{-1}A^\mathsf{T}(Au^k-f)\right) + \rho B^*\left(Bu^{k+1}-w^k+\rho^{-1}b^k\right) \\ 0 = s^{k+1} + \rho\left(w^{k+1}-Bu^{k+1}-\rho^{-1}b^k\right) + \beta\left(w^{k+1}-w^k\right) \\ b^{k+1} = b^k + \rho\left(Bu^{k+1}-w^{k+1}\right), \end{cases} \tag{3.6}$$

where $s^{k+1} \in \partial\phi(w^{k+1})^{\mathsf{T}}$. Let $\overline{u}$ denote a solution of (1.1), and define $\overline{w} = B\overline{u}$. The first-order optimality condition for (1.3) can be written

$$0 = \overline{s} - \overline{b}, \quad 0 = A^{\mathsf{T}}(A\overline{u} - f) + B^*\overline{b}, \quad 0 = B\overline{u} - \overline{w} \tag{3.7}$$

for some $\overline{s} \in \partial\phi(\overline{w})^{\mathsf{T}}$. Rearrange (3.7) to obtain

$$\begin{cases} 0 = \delta_k\big(\overline{u} - \overline{u} + \delta_k^{-1}A^{\mathsf{T}}(A\overline{u} - f)\big) + \rho B^*\big(B\overline{u} - \overline{w} + \rho^{-1}\overline{b}\big) \\ 0 = \overline{s} + \rho\big(\overline{w} - B\overline{u} - \rho^{-1}\overline{b}\big) + \beta(\overline{w} - \overline{w}) \\ \overline{b} = \overline{b} + \rho(B\overline{u} - \overline{w}), \end{cases} \tag{3.8}$$

which shows that $(\overline{u}, \overline{w}, \overline{b})$ is a fixed point of the BOSVS algorithm.

Denote the errors by $u_e^k = u^k - \overline{u}$, $w_e^k = w^k - \overline{w}$, $b_e^k = b^k - \overline{b}$, and $s_e^k = s^k - \overline{s}$. Subtract (3.8) from (3.6) to obtain

$$\begin{cases} 0 = \delta_k\big(u_e^{k+1} - u_e^k + \delta_k{}^{-1}A^{\mathsf{T}}Au_e^k\big) + \rho B^*\big(Bu_e^{k+1} - w_e^k + \rho^{-1}b_e^k\big) \\ 0 = s_e^{k+1} + \rho\big(w_e^{k+1} - Bu_e^{k+1} - \rho^{-1}b_e^k\big) + \beta\big(w_e^{k+1} - w_e^k\big) \\ b_e^{k+1} = b_e^k + \rho\big(Bu_e^{k+1} - w_e^{k+1}\big). \end{cases} \tag{3.9}$$

Form the inner product between the three equations of (3.9) and u_e^{k+1}, w_e^{k+1}, and b_e^k respectively to obtain

$$\begin{aligned} 0 &= \delta_k\big\langle u_e^{k+1}, u_e^{k+1} - u_e^k\big\rangle + \big\langle u_e^{k+1}, A^{\mathsf{T}}Au_e^k\big\rangle + \rho\big\langle Bu_e^{k+1}, Bu_e^{k+1} - w_e^k + \rho^{-1}b_e^k\big\rangle \\ 0 &= \big\langle s_e^{k+1}, w_e^{k+1}\big\rangle + \rho\big\langle w_e^{k+1} - Bu_e^{k+1} - \rho^{-1}b_e^k, w_e^{k+1}\big\rangle + \beta\big\langle w_e^{k+1}, w_e^{k+1} - w_e^k\big\rangle \\ \big\langle b_e^k, b_e^{k+1}\big\rangle &= \big\langle b_e^k, b_e^k\big\rangle + \rho\big\langle b_e^k, Bu_e^{k+1} - w_e^{k+1}\big\rangle. \end{aligned}$$

We exploit the equality $\langle a, a-b\rangle = 1/2(\langle a, a\rangle + \langle a-b, a-b\rangle - \langle b, b\rangle)$ to obtain

$$\begin{cases} \dfrac{\delta_k}{2}\big(\|u_e^{k+1}\|^2 + \|u^{k+1} - u^k\|^2 - \|u_e^k\|^2\big) + \big\langle Au_e^{k+1}, Au_e^k\big\rangle \\ \quad = \rho\big\langle Bu_e^{k+1}, w_e^k - \rho^{-1}b_e^k - Bu_e^{k+1}\big\rangle \\ \dfrac{\beta}{2}\big(\|w_e^{k+1}\|^2 + \|w^{k+1} - w^k\|^2 - \|w_e^k\|^2\big) + \big\langle s_e^{k+1}, w_e^{k+1}\big\rangle \\ \quad = \big\langle b_e^k, w_e^{k+1}\big\rangle + \rho\big\langle Bu_e^{k+1} - w_e^{k+1}, w_e^{k+1}\big\rangle \\ \dfrac{1}{2\rho}\big(\|b_e^{k+1}\|^2 - \|b_e^k\|^2\big) = \dfrac{1}{2\rho}\|b_e^{k+1} - b_e^k\|^2 + \big\langle b_e^k, Bu_e^{k+1} - w_e^{k+1}\big\rangle. \end{cases} \tag{3.10}$$

Insert $b_e^{k+1} - b_e^k = \rho(Bu_e^{k+1} - w_e^{k+1})$ in the last equality to get

$$\frac{1}{2\rho}\big(\|b_e^{k+1}\|^2 - \|b_e^k\|^2\big) = \frac{\rho}{2}\|Bu_e^{k+1} - w_e^{k+1}\|^2 + \big\langle b_e^k, Bu_e^{k+1} - w_e^{k+1}\big\rangle. \tag{3.11}$$

Add the equations in (3.10), while taking into account (3.11), to obtain

$$\begin{aligned}&\frac{\delta_k}{2}\left(\left\|u_e^{k+1}\right\|^2+\left\|u^{k+1}-u^k\right\|^2-\left\|u_e^k\right\|^2\right)+\left\langle Au_e^{k+1},Au_e^k\right\rangle+\left\langle s_e^{k+1},w_e^{k+1}\right\rangle\\&\quad+\frac{\beta}{2}\left(\left\|w_e^{k+1}\right\|^2+\left\|w^{k+1}-w^k\right\|^2-\left\|w_e^k\right\|^2\right)+\frac{\rho^{-1}}{2}\left(\left\|b_e^{k+1}\right\|^2-\left\|b_e^k\right\|^2\right)\\&=-\frac{\rho}{2}\left\|Bu_e^{k+1}-w_e^k\right\|^2-\frac{\rho}{2}\left(\left\|w_e^{k+1}\right\|^2-\left\|w_e^k\right\|^2\right).\end{aligned}$$

Rearrange this equation to obtain

$$\begin{cases}\frac{\delta_k}{2}\left\|u_e^{k+1}\right\|^2+\left(\frac{\beta+\rho}{2}\right)\left\|w_e^{k+1}\right\|^2+\frac{\rho^{-1}}{2}\left\|b_e^{k+1}\right\|^2+\frac{\delta_k}{2}\left\|u^{k+1}-u^k\right\|^2\\\quad+\left\langle Au_e^{k+1},Au_e^k\right\rangle+\frac{\beta}{2}\left\|w^{k+1}-w^k\right\|^2+\left\langle s_e^{k+1},w_e^{k+1}\right\rangle\\\quad+\frac{\rho}{2}\left\|Bu_e^{k+1}-w_e^k\right\|^2\\=\frac{\delta_k}{2}\left\|u_e^k\right\|^2+\left(\frac{\beta+\rho}{2}\right)\left\|w_e^k\right\|^2+\frac{\rho^{-1}}{2}\left\|b_e^k\right\|^2.\end{cases}\tag{3.12}$$

Substituting

$$\left\langle Au_e^{k+1},Au_e^k\right\rangle=\frac{1}{2}\left\|Au_e^{k+1}\right\|^2+\frac{1}{2}\left\|Au_e^k\right\|^2-\frac{1}{2}\left\|A\left(u^{k+1}-u^k\right)\right\|^2,$$

in (3.12) gives

$$\begin{cases}\delta_k\left\|u_e^{k+1}\right\|^2+(\beta+\rho)\left\|w_e^{k+1}\right\|^2+\rho^{-1}\left\|b_e^{k+1}\right\|^2\\\quad+\delta_k\left\|u^{k+1}-u^k\right\|^2-\left\|A\left(u^{k+1}-u^k\right)\right\|^2+\left\|Au_e^{k+1}\right\|^2+\left\|Au_e^k\right\|^2\\\quad+\beta\left\|w^{k+1}-w^k\right\|^2+2\left\langle s_e^{k+1},w_e^{k+1}\right\rangle+\rho\left\|Bu_e^{k+1}-w_e^k\right\|^2\\=\delta_k\left\|u_e^k\right\|^2+(\beta+\rho)\left\|w_e^k\right\|^2+\rho^{-1}\left\|b_e^k\right\|^2.\end{cases}\tag{3.13}$$

which by definition of Δ_k, can be equivalently written as

$$\begin{cases}\delta_k\left\|u_e^{k+1}\right\|^2+(\beta+\rho)\left\|w_e^{k+1}\right\|^2+\rho^{-1}\left\|b_e^{k+1}\right\|^2+\Delta_k+\bar{\eta}_kQ_k\\\quad+(1-\sigma)\delta_k\left\|u^{k+1}-u^k\right\|^2+\epsilon_\rho\left\|Bu_e^{k+1}-w_e^k\right\|^2\\\quad+\left\|Au_e^{k+1}\right\|^2+\left\|Au_e^k\right\|^2+\beta\left\|w^{k+1}-w^k\right\|^2+2\left\langle s_e^{k+1},w_e^{k+1}\right\rangle\\=\delta_k\left\|u_e^k\right\|^2+(\beta+\rho)\left\|w_e^k\right\|^2+\rho^{-1}\left\|b_e^k\right\|^2+(\bar{\eta}_k-1)Q_k+Q_k.\end{cases}$$

where $\epsilon_\rho = (1-\sigma)\rho$. Since $Q_{k+1} = \Delta_k + \bar{\eta}_k Q_k$, $\bar{\eta}_k \le 1$ and $-\frac{C}{(k-1)^2} \le Q_k$, the equality above yields the following relation:

$$\begin{cases} \delta_k \left\| u_e^{k+1} \right\|^2 + (\beta+\rho) \left\| w_e^{k+1} \right\|^2 + \rho^{-1} \left\| b_e^{k+1} \right\|^2 + Q_{k+1} \\ \quad + (1-\sigma)\delta_k \left\| u^{k+1} - u^k \right\|^2 + \epsilon_\rho \left\| B u_e^{k+1} - w_e^k \right\|^2 \\ \quad + \left\| A u_e^{k+1} \right\|^2 + \left\| A u_e^k \right\|^2 + \beta \left\| w^{k+1} - w^k \right\|^2 + 2\langle s_e^{k+1}, w_e^{k+1} \rangle \\ \quad \le \delta_k \left\| u_e^k \right\|^2 + (\beta+\rho) \left\| w_e^k \right\|^2 + \rho^{-1} \left\| b_e^k \right\|^2 + Q_k + \dfrac{(1-\bar{\eta}_k)C}{(k-1)^2}. \end{cases} \tag{3.14}$$

By Lemma 3.3, $\langle s^{k+1} - \overline{s}, w^{k+1} - \overline{w} \rangle = \langle s_e^{k+1}, w_e^{k+1} \rangle \ge 0$. We drop from the left side the following nonnegative terms

$$\left\| u^{k+1} - u^k \right\|^2, \quad \left\| A u_e^{k+1} \right\|^2, \quad \left\| A u_e^k \right\|^2, \quad \left\| w^{k+1} - w^k \right\|^2,$$
$$\langle s_e^{k+1}, w_e^{k+1} \rangle, \quad \left\| B u_e^{k+1} - w_e^k \right\|^2.$$

After dropping these terms, we obtain

$$\begin{aligned} &\delta_k \left\| u_e^{k+1} \right\|^2 + (\beta+\rho) \left\| w_e^{k+1} \right\|^2 + \rho^{-1} \left\| b_e^{k+1} \right\|^2 + Q_{k+1} \\ &\quad \le \delta_k \left\| u_e^k \right\|^2 + (\beta+\rho) \left\| w_e^k \right\|^2 + \rho^{-1} \left\| b_e^k \right\|^2 + Q_k + \frac{(1-\bar{\eta}_k)C}{(k-1)^2}. \end{aligned}$$

By Lemma 3.2, $\delta_k \ge \delta_{k+1}$ for all k sufficiently large. It follows that

$$\begin{aligned} &\delta_{k+1} \left\| u_e^{k+1} \right\|^2 + (\beta+\rho) \left\| w_e^{k+1} \right\|^2 + \rho^{-1} \left\| b_e^{k+1} \right\|^2 + Q_{k+1} \\ &\quad \le \delta_k \left\| u_e^k \right\|^2 + (\beta+\rho) \left\| w_e^k \right\|^2 + \rho^{-1} \left\| b_e^k \right\|^2 + Q_k + \frac{(1-\bar{\eta}_k)C}{(k-1)^2} \end{aligned} \tag{3.15}$$

for all k sufficiently large. Hence, for any j sufficiently large and $k \ge j$ we conclude that

$$\begin{aligned} &\delta_{\min} \left\| u_e^k \right\|^2 + (\beta+\rho) \left\| w_e^k \right\|^2 + \rho^{-1} \left\| b_e^k \right\|^2 + Q_k \\ &\quad \le \delta_k \left\| u_e^k \right\|^2 + (\beta+\rho) \left\| w_e^k \right\|^2 + \rho^{-1} \left\| b_e^k \right\|^2 + Q_k \\ &\quad \le \delta_j \left\| u_e^j \right\|^2 + (\beta+\rho) \left\| w_e^j \right\|^2 + \rho^{-1} \left\| b_e^j \right\|^2 + Q_j + \sum_{i=j}^{k-1} \frac{(1-\bar{\eta}_i)C}{(i-1)^2}. \end{aligned} \tag{3.16}$$

This shows that the sequence (u^k, w^k, b^k) is bounded since $\bar{\eta}_i \in (0,1)$, $\sum_{i=j}^{k-1} \frac{(1-\bar{\eta}_i)C}{(i-1)^2}$ is finite, and $Q_k \ge -C/(k-1)^2$.

Let $\epsilon = (1-\sigma)\delta_{\min}$, we sum the inequality (3.14) from $k = j$ to infinity to obtain

$$\begin{cases} \displaystyle\sum_{k=j}^{\infty} \epsilon \left\| u^{k+1} - u^k \right\|^2 + \sum_{k=j}^{\infty} \epsilon_\rho \left\| B u_e^{k+1} - w_e^k \right\|^2 \\ \quad + \displaystyle\sum_{k=j}^{\infty} \beta \left\| w^{k+1} - w^k \right\|^2 + \sum_{k=j}^{\infty} (\delta_k - \delta_{k+1}) \left\| u_e^{k+1} \right\|^2 \\ \quad + \displaystyle\sum_{k=j}^{\infty} \left\| A u_e^{k+1} \right\|^2 + \sum_{k=j}^{\infty} \left\| A u_e^k \right\|^2 + 2 \sum_{k=j}^{\infty} \left\langle s_e^{k+1}, w_e^{k+1} \right\rangle \\ \quad \le \delta_j \left\| u_e^j \right\|^2 + (\beta + \rho) \left\| w_e^j \right\|^2 + \rho^{-1} \left\| b_e^j \right\|^2 + Q_j + \displaystyle\sum_{k=j}^{\infty} \frac{(1-\bar{\eta}_k) C}{(k-1)^2} < \infty. \end{cases} \tag{3.17}$$

Suppose that j is large enough that $\delta_k \ge \delta_{k+1}$ for all $k \ge j$. Since

$$\sum_{k=j}^{\infty} (\delta_k - \delta_{k+1}) \left\| u_e^{k+1} \right\|^2 \ge 0,$$

we drop this term and additional nonnegative terms on the left side of (3.17) to obtain the relation

$$\sum_{k=j}^{\infty} \left(\epsilon \left\| u^{k+1} - u^k \right\|^2 + \epsilon_\rho \left\| B u_e^{k+1} - w_e^k \right\|^2 + \beta \left\| w^{k+1} - w^k \right\|^2 \right) < \infty.$$

This implies that

$$\lim_{k \to \infty} \left(\epsilon \left\| u^{k+1} - u^k \right\| + \epsilon_\rho \left\| B u_e^{k+1} - w_e^k \right\| + \beta \left\| w^{k+1} - w^k \right\| \right) = 0. \tag{3.18}$$

Moreover, by the last equality in (3.9) and the triangle inequality, we obtain

$$\begin{aligned} \left\| b^{k+1} - b^k \right\| = \left\| b_e^{k+1} - b_e^k \right\| &= \rho \left\| B u_e^{k+1} - w_e^{k+1} \right\| \\ &\le \rho \left\| B u_e^{k+1} - w_e^k \right\| + \rho \left\| w_e^{k+1} - w_e^k \right\|. \end{aligned}$$

Combining this with (3.18) gives

$$\lim_{k \to \infty} \left\| b^{k+1} - b^k \right\| = 0. \tag{3.19}$$

We just showed in (3.16) that the sequence (u^k, w^k, b^k) is bounded. Consider any subsequence, also denoted (u^k, w^k, b^k) for convenience, converging to a limit $(u^\infty, w^\infty, b^\infty)$. Choose any $s^k \in \partial\phi(w^k)^\mathsf{T}$. By Theorem 23.4 of [16], $\partial\phi(w^\infty)$ is bounded, and by Corollary 24.5.1 of [16], for any $\lambda > 0$, there exists $\mu > 0$ with the following property: If $\|w - w^\infty\| \le \mu$, then for each $s \in \partial\phi(w)$, the distance from s to $\partial\phi(w^\infty)$ is at most λ. Choose L large enough that $\|w^k - w^\infty\| \le \mu$ for all $k \ge L$.

It follows that s^k is bounded uniformly. Hence, there exists a subsequence converging to a limit s^∞. Again, for convenience, let (u^k, w^k, b^k, s^k) denote a subsequence converging to $(u^\infty, w^\infty, b^\infty, s^\infty)$. According to Theorem 24.4 in [16], $s^\infty \in \partial\phi(w^\infty)^\mathsf{T}$.

The first-order optimality conditions (3.6) for the sequence (u^k, w^k, b^k) is

$$\begin{cases} 0 = \delta_k\big(u^{k+1} - u^k\big) + A^\mathsf{T}\big(Au^k - f\big) + \rho B^*\big(Bu^{k+1} - w^k + \rho^{-1}b^k\big) \\ 0 = s^{k+1} + \rho\big(w^{k+1} - Bu^{k+1} - \rho^{-1}b^k\big) + \beta\big(w^{k+1} - w^k\big) \\ 0 = -b^{k+1} + b^k + \rho\big(Bu^{k+1} - w^{k+1}\big). \end{cases} \tag{3.20}$$

By (3.18) and (3.19), we have

$$\lim_{k\to\infty} \big\|u^{k+1} - u^k\big\| = 0, \quad \lim_{k\to\infty} \big\|w^{k+1} - w^k\big\| = 0, \quad \text{and} \quad \lim_{k\to\infty} \big\|b^{k+1} - b^k\big\| = 0.$$

Let k tend to infinity in (3.20) while taking values associated with the convergent subsequence to obtain

$$\begin{cases} 0 = A^\mathsf{T}\big(Au^\infty - f\big) + \rho B^*\big(Bu^\infty - w^\infty + \rho^{-1}b^\infty\big) \\ 0 = s^\infty + \rho\big(w^\infty - Bu^\infty - \rho^{-1}b^\infty\big) \\ 0 = Bu^\infty - w^\infty, \end{cases}$$

which simplifies to

$$0 = A^\mathsf{T}\big(Au^\infty - f\big) + B^*b^\infty, \quad 0 = s^\infty - b^\infty, \quad 0 = Bu^\infty - w^\infty,$$

where $s^\infty \in \partial\phi(w^\infty)^\mathsf{T}$. Hence, the limit point $(u^\infty, w^\infty, b^\infty)$ satisfies the first-order optimality conditions for (1.3).

The proof of this theorem started with an arbitrary extreme point $(\overline{u}, \overline{w}, \overline{b})$. Let us now consider the specific extreme point $\overline{u} = u^\infty$, $\overline{w} = w^\infty$, and $\overline{b} = b^\infty$ that is the limit of a convergent subsequence $(u^{k_l}, w^{k_l}, b^{k_l})$, $l \geq 1$. We substitute $j = k_l$ in (3.16) and let l tend to infinity to deduce that

$$0 = \lim_{k\to\infty} u_e^k = \lim_{k\to\infty} w_e^k = \lim_{k\to\infty} b_e^k.$$

This shows that the entire sequence (u^k, w^k, b^k) converges to $(u^\infty, w^\infty, b^\infty)$. Since ϕ is convex, (u^∞, w^∞) is a solution of (1.3) and u^∞ is a solution of (1.1). □

4 Connection with SpaRSA

We now study the relation between the stepsize condition in Step 2 of BOSVS and a monotone SpaRSA step [10, 22]. SpaRSA is designed to minimize a function of the form $\Phi(u) = F(u) + \psi(u)$. In each iteration, F is linearized while ψ is treated exactly. In a monotone SpaRSA step,

$$u^{k+1} = \arg\min\big\{\nabla F\big(u^k\big)v + \psi(v) + \delta_k\big\|v - u^k\big\|^2 : v \in \mathbb{C}^N\big\},$$

where $\delta_k = \eta^j \delta_0$ with $j \geq 0$ the smallest integer such that

$$\Phi\left(u^{k+1}\right) \leq \Phi\left(u^k\right) - \sigma\delta_k \left\|u^{k+1} - u^k\right\|^2. \tag{4.1}$$

Let us take $F(u) = \|Au - f\|^2$ and $\psi(u) = \rho\|Bu - w^k + \rho^{-1}b^k\|^2$. In this case, u^{k+1} in SpaRSA is given by

$$\begin{aligned} u^{k+1} &= \arg\min\left\{2(Au)^{\mathsf{T}}\left(Au^k - f\right) + \rho\left\|Bu - w^k + \rho^{-1}b^k\right\|^2 + \delta_k\left\|u - u^k\right\|^2 : u \in \mathbb{C}^N\right\} \\ &= \arg\min\left\{\delta_k\left\|u - u^k + \delta_k^{-1}A^{\mathsf{T}}\left(Au^k - f\right)\right\|^2 + \rho\left\|Bu - w^k + \rho^{-1}b^k\right\|^2 : u \in \mathbb{C}^N\right\} \end{aligned}$$

Hence, the formula for u^{k+1} in SpaRSA is exactly the same as the formula for u^{k+1} in Step 2 of BOSVS.

Now let us compare the SpaRSA criterion for the choice of δ_k to the criterion in Step 2 of BOSVS. The SpaRSA criterion (4.1) reduces to

$$\begin{aligned} &\left\|Au^{k+1} - f\right\|^2 + \rho\left\|Bu^{k+1} - w^k + \rho^{-1}b^k\right\|^2 \\ &\quad \leq \left\|Au^k - f\right\|^2 + \rho\left\|Bu^k - w^k + \rho^{-1}b^k\right\|^2 - \sigma\delta_k\left\|u^{k+1} - u^k\right\|^2. \end{aligned}$$

We simplify this inequality to obtain

$$\begin{cases} \left\|Au^{k+1}\right\|^2 + \rho\left\|Bu^{k+1}\right\|^2 + \sigma\delta_k\left\|u^{k+1} - u^k\right\|^2 \leq \left\|Au^k\right\|^2 + \rho\left\|Bu^k\right\|^2 \\ -2\rho\left\langle B\left(u^{k+1} - u^k\right), -w^k + \rho^{-1}b^k\right\rangle + 2\left\langle A\left(u^{k+1} - u^k\right), f\right\rangle. \end{cases} \tag{4.2}$$

The first-order optimality conditions for u^{k+1} are

$$\delta_k\left(u^{k+1} - u^k + \delta_k{}^{-1}A^{\mathsf{T}}\left(Au^k - f\right)\right) + \rho B^*\left(Bu^{k+1} - w^k + \rho^{-1}b^k\right) = 0.$$

We compute the inner product with $u^{k+1} - u^k$ to obtain

$$\begin{aligned} &\delta_k\left\|u^{k+1} - u^k\right\|^2 + \left\langle A\left(u^{k+1} - u^k\right), Au^k - f\right\rangle \\ &\quad + \rho\left\langle B\left(u^{k+1} - u^k\right), Bu^{k+1} - w^k + \rho^{-1}b^k\right\rangle = 0. \end{aligned}$$

Rearrange this equality to obtain

$$\begin{cases} -\rho\left\langle B\left(u^{k+1} - u^k\right), -w^k + \rho^{-1}b^k\right\rangle + \left\langle A\left(u^{k+1} - u^k\right), f\right\rangle \\ \quad = \delta_k\left\|u^{k+1} - u^k\right\|^2 + \left\langle Au^{k+1}, Au^k\right\rangle + \rho\left\|Bu^{k+1}\right\|^2 - \left\|Au^k\right\|^2 \\ \qquad - \rho\left\langle Bu^{k+1}, Bu^k\right\rangle. \end{cases} \tag{4.3}$$

Finally, incorporate (4.3) in (4.2) to obtain

$$\left\|A\left(u^{k+1} - u^k\right)\right\|^2 \leq (2 - \sigma)\delta_k\left\|u^{k+1} - u^k\right\|^2 + \rho\left\|B\left(u^{k+1} - u^k\right)\right\|^2. \tag{4.4}$$

Thus for a monotone SpaRSA step, δ_k is acceptable when (4.4) holds, while for the BOSVS algorithm, δ_k is acceptable when

$$\left\|A\left(u^{k+1}-u^k\right)\right\|^2 \leq \sigma\left(\eta^j \delta_{0,k}\left\|u^{k+1}-u^k\right\|^2+\rho\left\|Bu^{k+1}-u^k\right\|^2\right)+\bar{\eta}_k Q_k+\frac{C}{k^2}.$$

5 Numerical experiments

We test the proposed algorithm using data from an emerging magnetic resonance (MR) imaging technology known as partially parallel imaging, or PPI. This is an MR imaging technique that uses multiple radio frequency (RF) coil arrays with a separate receiver channel for each RF coil. A set of multi-channel Fourier (k-space) data from each RF coil array is acquired simultaneously. The imaging process is accelerated by only acquiring part of the k-space data. Partial data acquisition increases the spacing between regular subsequent read-out lines, thereby reducing scan time. However, this reduction in the number of recorded Fourier components leads to aliasing artifacts in images.

Sensitivity encoding (SENSE) is one of the most commonly used PPI approaches to remove the aliasing artifacts and reconstruct high quality images. The fundamental equations of SENSE are as follows: In a PPI system consisting of L coil arrays, the undersampled data f_l from the l-th channel is related to the underlying image u^* by

$$P\mathcal{F}\left(s_l \odot u^*\right)=f_l, \quad l=1,\ldots,L, \tag{5.1}$$

where $\mathcal{F}$ is the Fourier transform, P is a binary matrix representing the undersampling pattern (mask) (e.g. Fig. 2), and $s_l \in \mathbb{C}^N$ is the sensitivity map for the l-th channel. The symbol $\odot$ is the Hadamard (or componentwise) product between two vectors. The underlying image can be recovered by solving an optimization problem of the form:

$$\min_{u\in\mathbb{C}^N}\left\{\Psi(u):=\alpha\|u\|_{TV}+\frac{1}{2}\sum_{l=1}^{L}\left\|\mathcal{F}_p(s_l\odot u)-f_l\right\|^2\right\}, \tag{5.2}$$

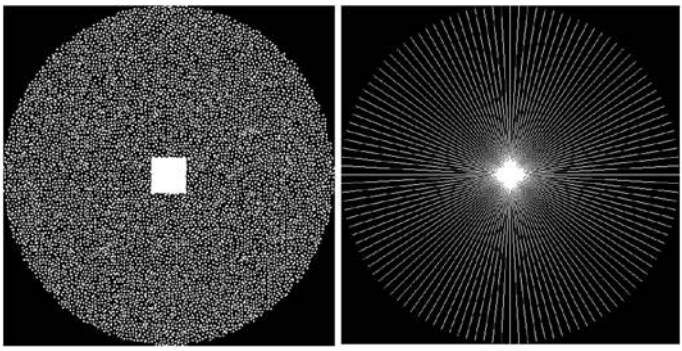


Fig. 2 *Left*: Poisson random mask used for data1 and data2, with 25 % undersampling ratio. *Right*: radial mask used for data3, with 34 % undersampling ratio

Table 1 The data acquisition parameters for three datasets

Data acquisition parameters	data1	data2	data3
Relaxation Time (TR)	3060 ms	3000 ms	53.5 ms
Echo Time (TE)	126 ms	85 ms	34 ms
Field of View (FOV)	220 mm^2	205 mm^2	220 mm^2
Size	$512 \times 512 \times 8$	$500 \times 512 \times 8$	$256 \times 512 \times 8$
Slice Thickness (ST)	5 mm	5 mm	5 mm
Flip Angle (FA)	90°	90°	75°

where $\mathcal{F}_p$ is the undersampled Fourier transform defined by $\mathcal{F}_p := P\mathcal{F}$ and α is a given parameter which weighs the importance of the regularization term $\|.\|_{TV}$ to the fidelity term. Let

$$A = [\mathcal{F}_p S_1; \mathcal{F}_p S_2; \cdots; \mathcal{F}_p S_L], \qquad f = [f_1; f_2; \ldots; f_L], \tag{5.3}$$

where $S_l := \mathrm{diag}(s_l) \in \mathbb{C}^{N\times N}$ is the diagonal matrix with $s_l \in \mathbb{C}^N$ on the diagonal, $l = 1, 2, \ldots, L$, and $[X; Y]$ denotes the matrix obtained by stacking X above Y. Problem (5.2) has the form of (1.1) which can be solved using the BOSVS algorithm. For more details concerning the data structure and model formulation, we refer readers to [24] and references therein.

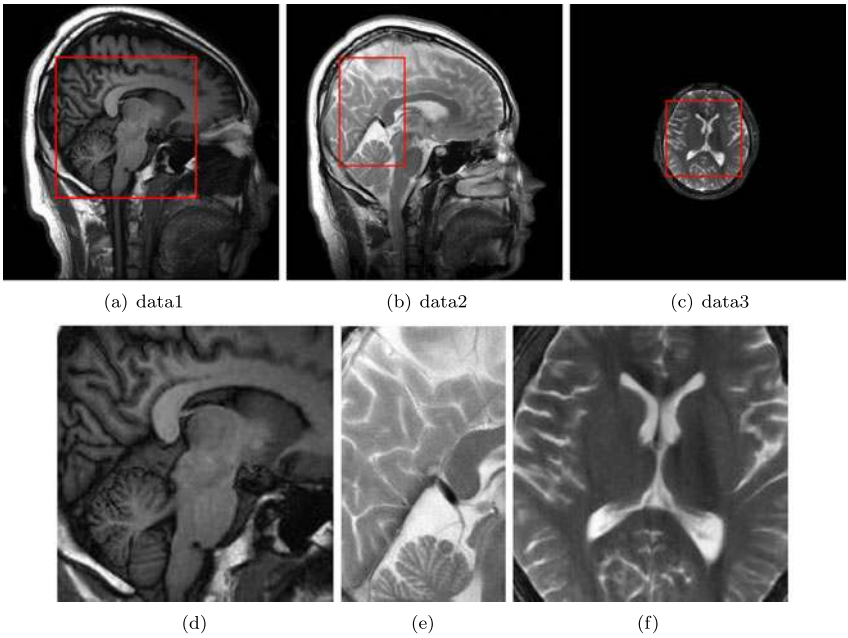
5.1 Test problems

In this section, we apply the proposed algorithm to three PPI datasets denoted data1, data2 and data3. These were obtained from commercially available eight-channel ($L = 8$) PPI machines.

Data1 was acquired from a 3T GE system (GE Healthcare, Waukesha, Wisconsin). Data2 was acquired from a 3T Phillips scanner (Phillips, Best, Netherlands) using T2-weighted turbo spin echo (T2 TSE) sequence. Data3 was acquired from a 1.5T Siemens Symphony system (Siemens Medical Solutions, Erlangen, Germany). The data acquisition parameters for these three datasets are given in Table 1. These three datasets were fully acquired; that is, all the Fourier components were acquired to obtain an image u with intensities scaled to the interval [0, 1] (see Fig. 3) along with estimated sensitivity maps $\{s_l\}_{l=1}^L$. We use the masks P shown in Fig. 2 with 25 % undersampling for data1 and data2 and 34 % undersampling for data3. The artificially undersampled data f_l was given by $P\mathcal{F}(s_l \odot u) + n_l$, where n_l is complex valued white Gaussian noise with standard deviation $\bar{\sigma}$ for both of the real and imagery parts (we set $\bar{\sigma} = 0.7 \times 10^{-3}$ for all the experiments). These artificially undersampled data $\{f_l\}_{l=1}^L$ were used to test the performance of three different line search schemes:

(a) The fixed stepsize $\delta_k = \|A^{\mathsf{T}}A\|$ corresponding to the BOS type Algorithm A1 in [26] with $Q_1 = \delta I - A^{\mathsf{T}}A$, $Q_2 = \beta I$, and $C = 1/\rho$.
(b) The BB stepsize $\delta_k = \|A(u^k - u^{k-1})\|^2 / \|u^k - u^{k-1}\|^2$ given in [5] for the SBB algorithm.
(c) The variable stepsize employed in the BOSVS algorithm.

We used TV-regularization to recover the images.



(a) data1 (b) data2 (c) data3

(d) (e) (f)

Fig. 3 *Top row*: reference images corresponding to data1, data2, and data3, respectively. *Bottom row*: (**d**)–(**f**) zoomed into the box in (**a**)–(**c**), respectively

Table 2 Parameter values for three algorithms: BOS, SBB and BOSVS. α and ρ are used in all of the algorithms where τ, η, $\delta_{\min}$, σ, C, and $\bar{\eta}_k$ are specific to BOSVS

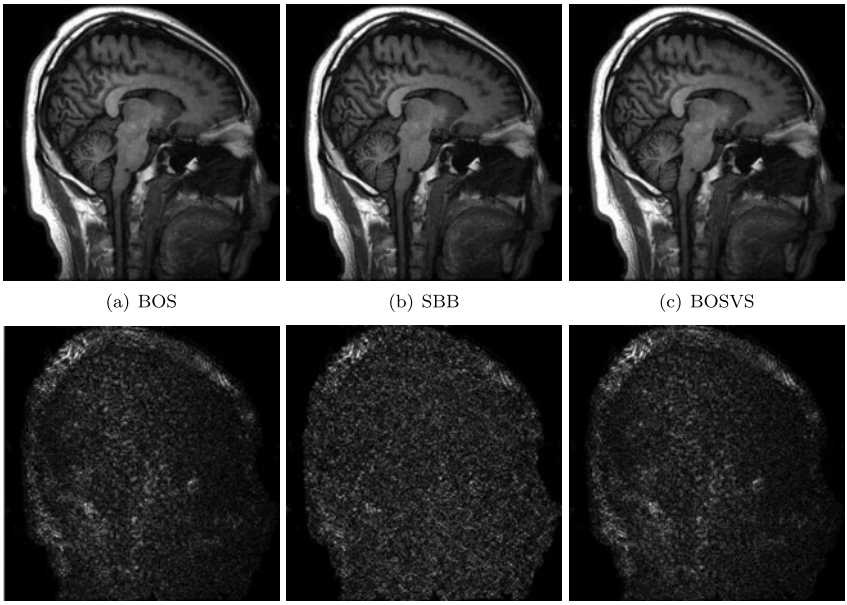
Algorithms parameters	α	ρ	β	τ	η	$\delta_{\min}$	σ	C	$\bar{\eta}_k$
Considered values	10^{-4}	10^{-2}	1	2	3	0.001	0.99999	100	$1/k$

5.2 Parameter setting

In all of our experiments, we consider the parameter values given in Table 2. Theoretically the choice of ρ does not effect the convergence of the tested algorithms. This is also demonstrated by our experiments since the results are not sensitive to ρ for a large range. Therefore in all experiments we set ρ to a moderate value 10^{-2}. It is worth noting that the image reconstructions depend on the choice of α in (1.2). The initial guess u^1, b^1 and w^1 were set to 0 for all algorithms. Our algorithm comparisons are based on the value of objective function, the CPU time in seconds, the number of matrix-vector multiplications, and the relative error in the image defined by $\|u^k - u^*\|/\|u^*\|$ where u^* is the underlying image.

We ran the BOS algorithm for 1000 iterations to obtain the following high accuracy estimates for the optimal objective function value Ψ in (5.2):

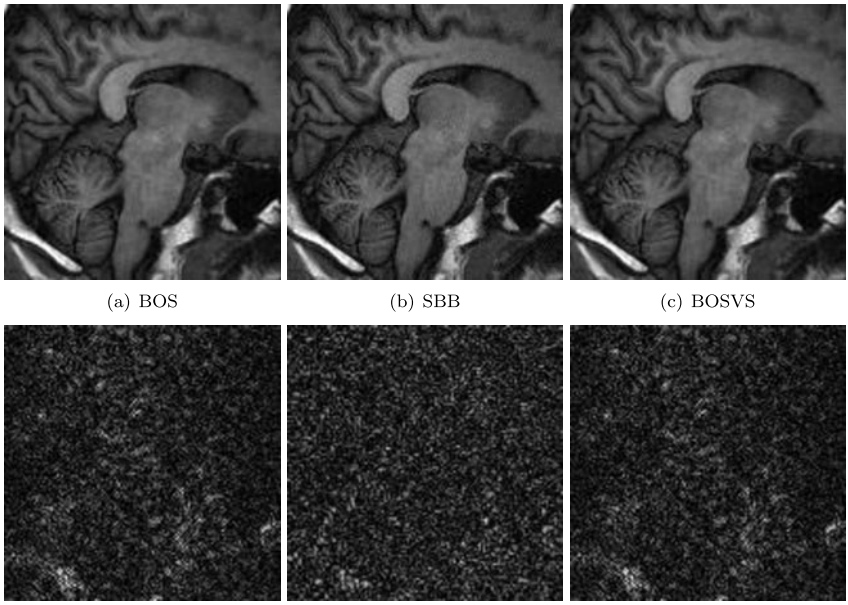
1. For data1: $\Psi^* = 0.50400636019746$.
2. For data2: $\Psi^* = 1.814917047692733$.
3. For data3: $\Psi^* = 1.189609461117846$.



(a) BOS (b) SBB (c) BOSVS

(d) BOS (e) SBB (f) BOSVS

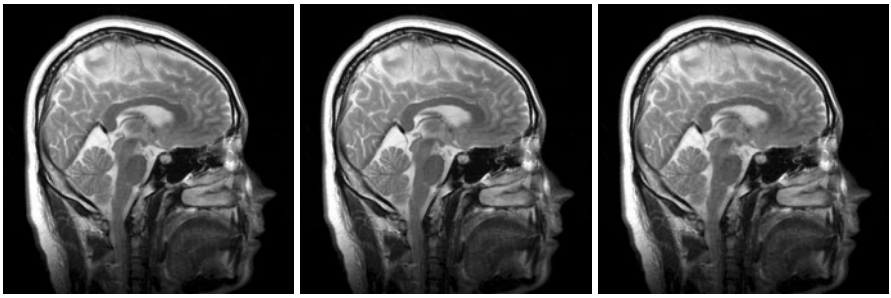
Fig. 4 *Top row*: (**a**)–(**c**) reconstructed image of data1 by BOS, SBB and BOSVS, respectively. *Bottom row*: (**d**)–(**f**) absolute errors of reconstructions u (shown in *top row*) to the reference image (shown in Fig. 3(a)), i.e., $|u - u^*|$, respectively. All images are shown with the same brightening scale



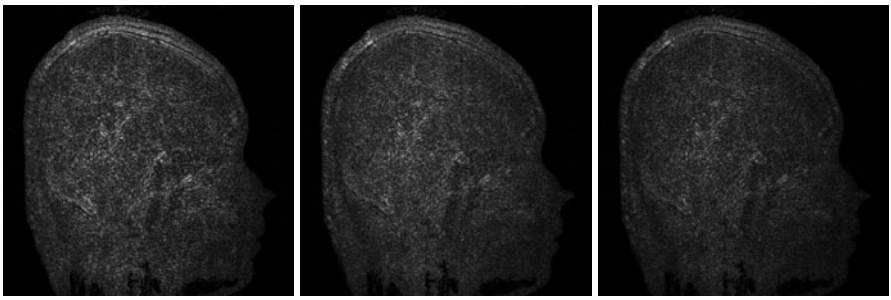
(a) BOS (b) SBB (c) BOSVS

(d) BOS (e) SBB (f) BOSVS

Fig. 5 *Top row*: (**a**)–(**c**) zoomed in to the box of reconstructed images of data1 (shown in Fig. 4(a)–(c)) by BOS, SBB and BOSVS, respectively. *Bottom row*: (**d**)–(**f**) absolute errors of reconstructed images in the top row to the image (shown in Fig. 3(d)). All images are shown with the same brightening scale



(a) BOS (b) SBB (c) BOSVS



(d) BOS (e) SBB (f) BOSVS

Fig. 6 *Top row*: (**a**)–(**c**) reconstructed image of data2 by BOS, SBB and BOSVS, respectively. *Bottom row*: (**d**)–(**f**) absolute errors of reconstructions u (shown in *top row*) to the reference image (shown in Fig. 3(b)), i.e., $|u - u^*|$, respectively. All images are shown with the same brightening scale

For all algorithms tested in our experiments, we terminate the computation when the absolute error of the objective function value satisfies the stopping criterion:

$$\left|\Psi\left(u^k\right) - \Psi^*\right| < 10^{-5}.$$

Note that, all data sets are normalized such that the intensities of reference images have range [0, 1]. All algorithms tested in this section were implemented in the MATLAB programming environment (Version R2010a). The experiments were performed on a HP Pavilion laptop, Model G71-340US, with Intel Dual Core, and Windows 7 operating system.

5.3 Numerical results

The reconstructed images from the three PPI datasets are shown in Figs. 4, 6, and 8 respectively. For a better comparisons, Figs. 5, 7, and 9 show a portion of each image in higher resolution. The figures correspond to the reconstructed images that satisfy the stopping criteria of Sect. 5.2 or the image at iteration 500 if the stopping criteria were not satisfied. The absolute error of the reconstructed image appear in the second row or each figure. Since all images are normalized in [0 1], the pixels with the zero absolute error are black, and the pixels with the highest error are white. All

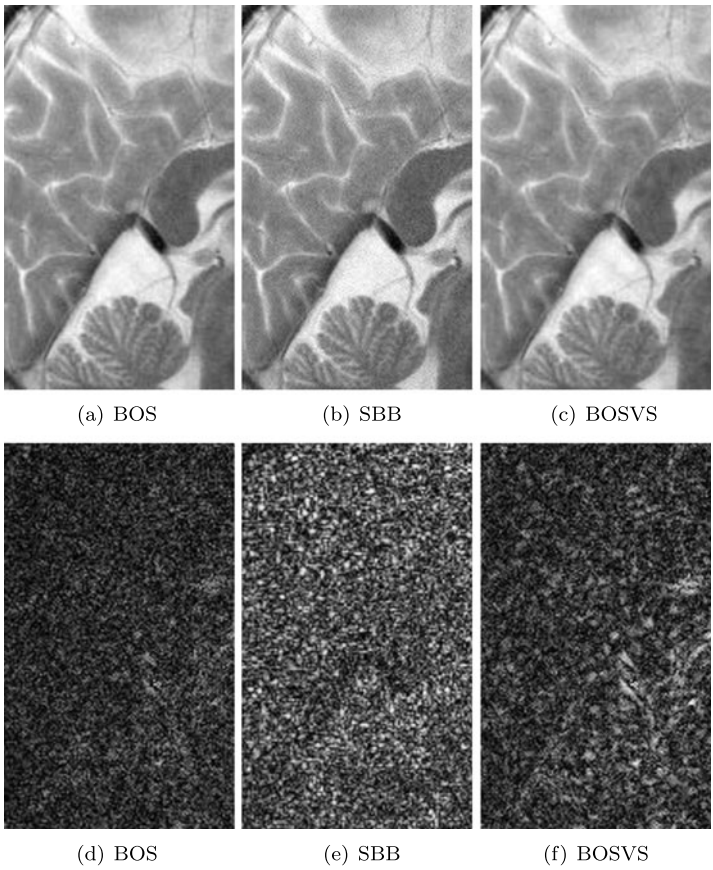
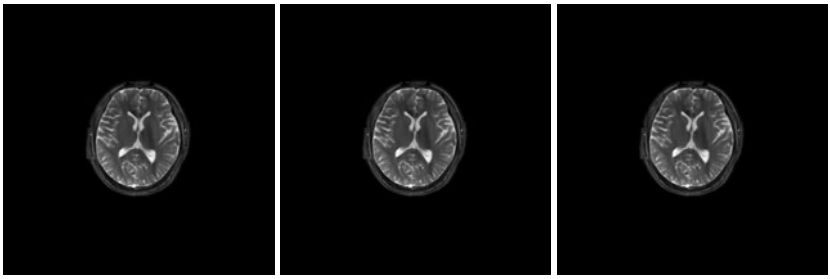


Fig. 7 *Top row*: (**a**)–(**c**) zoomed in to the box of reconstructed images of data2 (shown in Fig. 6(a)–(c)) by BOS, SBB and BOSVS, respectively. *Bottom row*: (**d**)–(**f**) absolute errors of reconstructed images in the *top row* to the image (shown in Fig. 3(e)). All images are shown with the same brightening scale

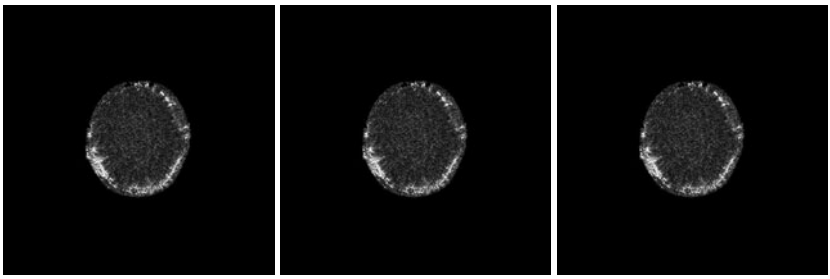
three methods adequately recovered the image in the sense that most details and fine structures were accurately recovered.

Table 3 shows the comparisons between BOS, SBB, and BOSVS. Both BOS and BOSVS eventually satisfy the stopping criteria, however, BOS is much slower than BOSVS. The SBB algorithm did not satisfy the stopping criteria for data1 and data2, and it did not converge. Hence, for SBB we report in Table 3 the minimum objective function value and corresponding error in data1 and data2.

The plots of objective function value versus iteration number are given in Figs. 10, 11, and 12. In all three cases, the objective function values for BOS decay monotonically, however, the convergence speed is slower than that of BOSVS and SBB. SBB at first decreases the objective value quickly, but asymptotically, the algorithm may not converge. Initially, the BOSVS and SBB iterates are identical; asymptotically, BOSVS converges monotonically. The Figures also plot the value of j in Step 2 of BOSVS as a function of iteration number. The initial values of j are 0,

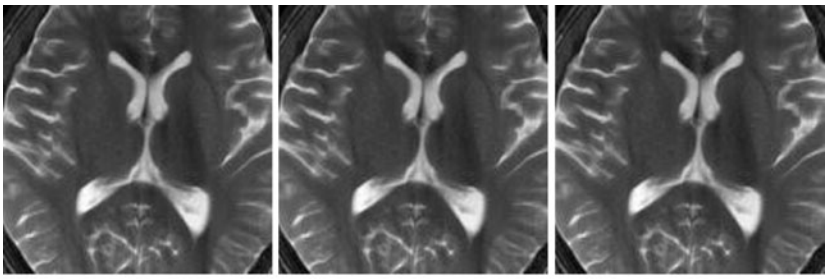


(a) BOS (b) SBB (c) BOSVS

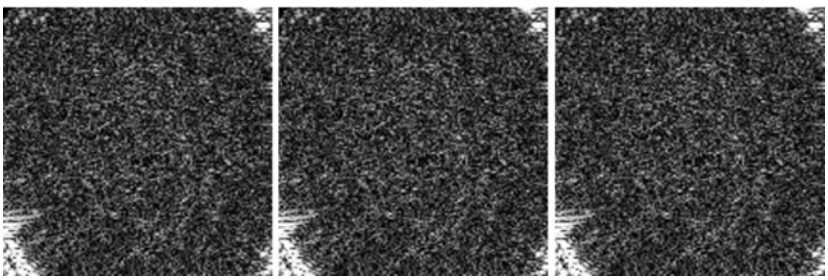


(d) BOS (e) SBB (f) BOSVS

Fig. 8 *Top row*: (**a**)–(**c**) reconstructed image of data3 by BOS, SBB and BOSVS, respectively. *Bottom row*: (**d**)–(**f**) absolute errors of reconstructions u (shown in *top row*) to the reference image (shown in Fig. 3(c)), i.e., $|u - u^*|$, respectively. All images are shown with the same brightening scale



(a) BOS (b) SBB (c) BOSVS

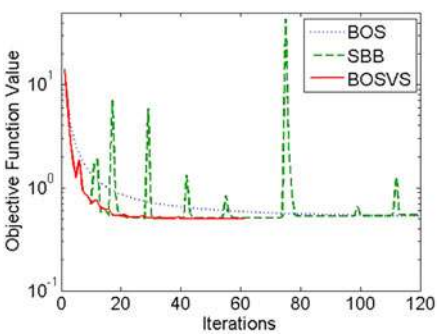


(d) BOS (e) SBB (f) BOSVS

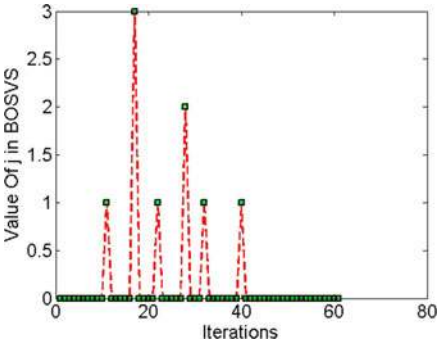
Fig. 9 *Top row*: (**a**)–(**c**) zoomed in to the box of reconstructed images of data3 (shown in Fig. 8(a)–(c)) by BOS, SBB and BOSVS, respectively. *Bottom row*: (**d**)–(**f**) absolute errors of reconstructed images in the *top row* to the image (shown in Fig. 3(f)). All images are shown with the same brightening scale

Table 3 Comparison of objective function value, relative error $\|u^k - u^*\|/\|u^*\|$, CPU time in seconds, and the number of matrix-vector products (Ax) for BOS, SBB, and BOSVS. SBB is unable to satisfy the stopping criterion for data1 and data2; hence, for these two images, we give the minimum objective value that SBB achieves along with the corresponding relative error and the number of matrix-vector products

Algorithms	Objective value	Relative error	CPU	Ax
	data1			
BOS	0.50401527	0.06432442	93.50	712
SBB	0.50407064	0.06567046	∞	123
BOSVS	0.50399763	0.06044609	23.54	192
	data2			
BOS	1.81491239	0.05268322	348	494
SBB	1.81751525	0.04972965	∞	105
BOSVS	1.81492665	0.05103128	117	168
	data3			
BOS	1.18961102	0.02122704	87	136
SBB	1.18960946	0.02100478	32	51
BOSVS	1.18960946	0.02100478	32	51

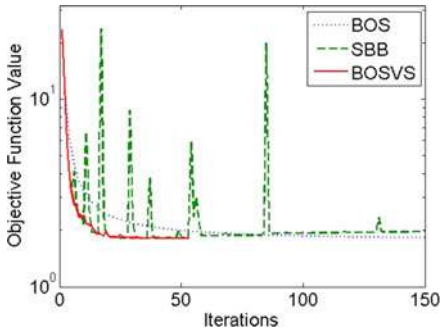


(a) Objective function value versus iteration for data1.

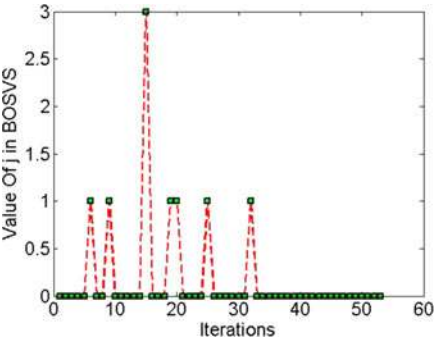


(b) Value of j in Step 2 of BOSVS versus iteration number.

Fig. 10 Performance results for data1

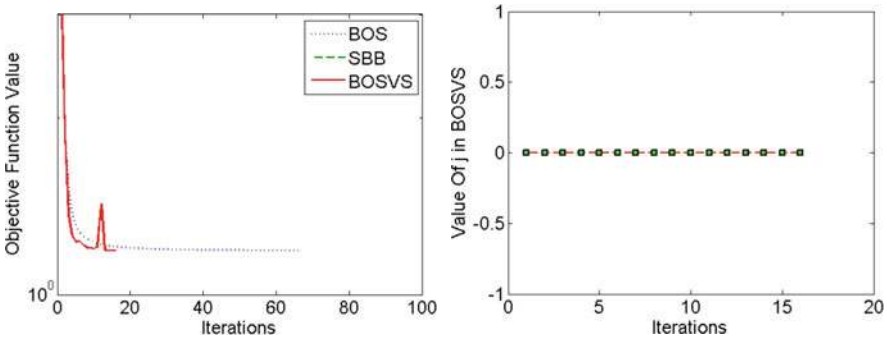


(a) Objective function value versus iteration for data2



(b) Value of j in Step 2 of BOSVS versus iteration number using data2.

Fig. 11 Performance results for data2



(a) Objective function value versus iteration for data3.

(b) Value of j in Step 2 of BOSVS versus iteration number using data3.

Fig. 12 Performance results for data3

and asymptotically, j becomes 0. In between, there can be a few iterations where j is nonzero. Since we took $\bar{\eta}_k = 1/k$ (see Table 2), it follows that both $\bar{\eta}_k$ and $-\frac{C}{k^2}$ approach 0 as k increases. Referring to Step 2, that can cause j to increase. In our numerical results, we noticed that the value of C has an important role in performance. If C was 20, the objective function had jumps like the SBB algorithm. The best values for C in our experiments were between 80 and 120.

6 Conclusion

A variable stepsize Bregman operator splitting algorithm (BOSVS) was introduced and analyzed. Our implementation utilizes a BB iteration in the updates of the image. Global convergence is established. BOSVS is well suited for total variation-based image reconstruction, and problems where the matrix in the fidelity term is large and complex. Experimental results are given for partially parallel magnetic resonance imaging. Comparisons are given with a BOS type algorithm in [26] based on a fixed step size and an algorithm based on a BB step as given in [5]. BOSVS was found to be asymptotically efficient and robust for large-scale image reconstruction.

Acknowledgements The authors thank Invivo Corporation and Dr. Feng Huang for providing the PPI data used in the paper.

References

1. Baraniuk, R., Steeghs, P.: Compressive radar imaging. In: 2007 IEEE Radar Conference, pp. 128–133 (2007)
2. Barzilai, J., Borwein, J.M.: Two point step size gradient methods. IMA J. Numer. Anal. **8**, 141–148 (1988)
3. Block, K., Uecker, M., Frahm, J.: Undersampled radial MRI with multiple coils: iterative image reconstruction using a total variation constraint. Magn. Reson. Med. **57**, 1086–1098 (2007)
4. Chan, T.F., Golub, G.H., Mulet, P.: A nonlinear primal-dual method for total variationbased image restoration. SIAM J. Optim. **20**, 1964–1977 (1999)

5. Chen, Y., Hager, W.W., Huang, F., Phan, D.T., Ye, X., Yin, W.: Fast algorithms for image reconstruction with application to partially parallel MR imaging. SIAM J. Imaging Sci. **5**, 90–118 (2012)
6. Darbon, J., Sigelle, M.: A fast and exact algorithm for total variation minimization. In: Lecture Notes in Comput. Sci. IbPRIA, vol. 3522, pp. 351–359. Springer, Berlin (2005)
7. Donoho, D.: Compressed sensing. IEEE Trans. Inf. Theory **52**, 351–359 (2006)
8. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrange Methods: Applications to the Solution of Boundary-Valued Problems, pp. 299–331. North Holland, Amsterdam (1983)
9. Goldstein, T., Osher, S.: The split Bregman method for L1 regularized problems. SIAM J. Imaging Sci. **2**, 323–343 (2009)
10. Hager, W.W., Phan, D.T., Zhang, H.: Gradient-based methods for sparse recovery. SIAM J. Imaging Sci. **4**, 146–165 (2011)
11. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303–320 (1969)
12. Huang, Y., Ng, M.K., Wen, Y.: A new total variation method for multiplicative noise removal. SIAM J. Imaging Sci. **2**, 20–40 (2009)
13. Li, Y., Santosa, F.: An affine scalling algorithm for minimizing total variation in image enhancement. Tech. Rep. TR94-1470, Cornell Theory Center, Cornell University, Ithaca, NY (1994)
14. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn. Reson. Med. **58**, 1182–1195 (2007)
15. Pruessmann, K., Weiger, M., Bornert, P., Boesiger, P.: Advances in sensitivity encoding with arbitrary k-space trajectories. Magn. Reson. Med. **46**, 638–651 (2001)
16. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
17. Rudin, L., Osher, S., Fatemi, E.: Non-linear total variation noise removal algorithm. Physica D **60**, 259–268 (1992)
18. Strong, D., Chan, T.: Edge preserving and scale-dependent properties of total variation regularization. Inverse Probl. **19**, 165–187 (2003)
19. Vogel, C.R.: A multigrid method for total variation-based image denoising. In: Bowers, K., Lund, J. (eds.) Computation and Control IV. Progess in Systems and Control Theory, vol. 20, pp. 323–331. Birkhauser, Boston (1995)
20. Vogel, C.R., Oman, M.E.: Iterative methods for total variation denoising. SIAM J. Sci. Comput. **17**, 227–238 (1996)
21. Wang, Y., Yang, J., Yin, W., Zhang, Y.: A new alternating minimization algorithm for total variation image reconstruction. SIAM J. Imaging Sci. **1**, 248–272 (2008)
22. Wright, S.J., Nowak, R.D., Figueiredo, M.A.T.: Sparse reconstruction by separable approximation. IEEE Trans. Signal Process. **57**, 2479–2493 (2009)
23. Yang, J., Zhang, Y., Yin, W.: A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data. IEEE J. Sel. Top. Signal Process. **4**, 288–297 (2010)
24. Ye, X., Chen, Y., Huang, F.: Computational acceleration for MR image reconstruction in partially parallel imaging. IEEE Trans. Med. Imaging **30**, 1055–1063 (2011)
25. Zhang, X., Burger, M., Bresson, X., Osher, S.: Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. SIAM J. Imaging Sci. **3**, 253–276 (2011)
26. Zhang, X., Burger, M., Osher, S.: A unified primal-dual algorithm framework based on Bregman iteration. J. Sci. Comput. **46**, 20–46 (2011)
27. Zhu, M., Chan, T.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. Tech. Rep. 08-34, CAM UCLA (2008)
28. Zhu, M., Wright, S., Chan, T.: Duality-based algorithms for total-variation-regularized image restoration. Comput. Optim. Appl. **47**, 377–400 (2010)