

# THE SPLIT BREGMAN METHOD FOR L1 REGULARIZED PROBLEMS

TOM GOLDSTEIN, STANLEY OSHER

**Abstract.** The class of l1-regularized optimization problems has received much attention recently because of the introduction of “compressed sensing,” which allows images and signals to be reconstructed from small amounts of data. Despite this recent attention, many l1-regularized problems still remain difficult to solve, or require techniques that are very problem-specific. In this paper, we show that Bregman iteration can be used to solve a wide variety of constrained optimization problems. Using this technique, we propose a “Split Bregman” method, which can solve a very broad class of l1-regularized problems. We apply this technique to the ROF functional for image denoising, and to a compressed sensing problem that arises in Magnetic Resonance Imaging.

**Key words.** Constrained Optimization, l1 regularization, compressed sensing, total variation denoising.

**1. Introduction.** The category of l1-regularized problems includes many important problems in engineering, computer, and imaging science. The general form for such problems is

$$(1.1) \quad \min_u |\Phi(u)| + H(u)$$

where  $|\cdot|$  denotes the l1-norm, and both  $|\Phi(u)|$  and  $H(u)$  are convex functions. Many important problems in imaging science (and other computational areas) can be posed as l1-regularized optimization problems. Some common examples of this include the following:

$$(1.2) \quad \text{“TV/ROF Denoising: ” } \min_u \|u\|_{BV} + \frac{\mu}{2} \|u - f\|_2^2$$

$$(1.3) \quad \text{“Basis Pursuit/Compressed Sensing: ” } \min_u J(u) + \frac{\mu}{2} \|Au - f\|_2^2$$

where  $J(u)$  is some regularizing functional, usually in the form of a BV or Besov norm.

The Rudin-Osher-Fatemi (ROF) functional (1.2), despite its simple form, has proved to be very difficult to minimize by conventional methods. Total variation based image restoration was first introduced in [22]. In that paper, the authors propose to minimize this energy using a gradient projection method. While this approach is simple, the non-linearity and poor conditioning of the problem make this approach very slow. Several authors have proposed improved time-stepping schemes that result in better performance, such as those presented in [26, 15]. A more efficient class of solvers are those based on Newton’s method. One such algorithm was presented in [9], in which the preconditioned conjugate gradient method is used to invert the Hessian at each step. A somewhat more efficient implementation of a second-order method was proposed by Vogel et. al. in [25], where an algebraic multigrid preconditioner is used to accelerate the method.

Still, the most efficient solver for the ROF problem was proposed by Darbon and Sigelle in [10]. In this work, the ROF functional was approximated using an anisotropic BV norm. It was shown that, using this formulation, the resulting problem could be solved very quickly using graph cuts [3].

Problems of the form (1.3) have received a lot of attention recently because of the introduction of compressed sensing techniques, which allow high resolution images and signals to be reconstructed from a small amount of data [7, 6, 11]. This formulation has been shown to be useful in many areas, including medical imaging [16, 17], radar [1], and other signal processing applications. Compressed sensing is based on the idea that a signal can be reconstructed from a very small number of measurements, provided that these measurements are obtained in the correct basis. The particular application of compressed sensing which we will focus on is MR image reconstruction, or “Sparse MRI” [16, 14]. The goal of sparse MRI is to solve

$$\min_u J(u) \text{ such that } \|\mathcal{F}u - f\|_2 = 0.$$

Here,  $f$  represents the so-called “compressed sensing data,” which consists of samples of the Fourier transform of the unknown image.  $\mathcal{F}$  comprises a subset of the rows of a Fourier matrix,  $u$  represents the unknown image that we wish to reconstruct, and  $J(\cdot)$  is a properly chosen l1-regularization term. The unconstrained formulation for this problem was introduced in [14], where a Bregman iterative approach [4] was used to obtain solutions to “denoising” problems of the form

$$(1.4) \quad \min_u J(u) \text{ such that } \|Fu - f\|_2 < \sigma.$$

Because of the presence of an l1-regularization term, optimization problems of the form (1.4) are still very difficult to solve. Several authors have applied classical optimization schemes, such as interior point methods, to problems of these forms. In [23], CS problems were reformulated as quadratic programming problems which were then solved using the code “l1\_ls,” which is claimed to be one of the most efficient solvers for general compressed sensing problems. Another notable interior-point approach is the code “l1-magic,” which formulates a CS problem as a second order cone program, and enforces inequality constraints using a logarithmic barrier potential [7].

In the special case where the CS problem can be written in the form

$$(1.5) \quad \min_u |u| \text{ such that } \|Au - f\|_2 < \sigma$$

a relatively new class of methods can be used that reduce the CS problem to set of simpler problems using linearization. The first of these methods is the “Fixed Point Continuation” method (FPC) introduced in [13] by Hale, Yin, and Zhang. This method solves the unconstrained problem

$$(1.6) \quad \min_u |u| + \frac{\mu}{2} \|Au - f\|_2^2$$

by iteratively performing gradient descent steps.

By applying the Bregman iteration scheme in [21, 14], it is also possible to solve the constrained problem (1.5) using FPC/Bregman. This is done by iteratively solving the unconstrained problem (1.6), and then modifying the value of  $f$  used in the next iteration.

Rather than solving the unconstrained problem and then performing a Bregman update separately, these two steps were elegantly combined in the Linearized Bregman Algorithm [30]. Linearized Bregman solves the problem (1.5) by iteratively solving

$$\begin{aligned} v^{k+1} &= v^k + A^T(f - Au^k) \\ u^{k+1} &= \delta * \mathit{shrink}(v^{k+1}, 1/\mu) \end{aligned}$$

for  $k = 1, 2, \dots, n$ . The algorithm is terminated when the denoising constraint in (1.5) is met. It was shown in [5, 20] that  $u^n$  will be a suitable approximation to the solution of (1.5) provided that appropriate values are chosen for the parameters  $\mu$  and  $\delta$ .

While the FPC and Linearized Bregman algorithms are extremely efficient, they can only solve problems which can be put into the form (1.6). Because the gradient operator is not invertible, these algorithms cannot be used to solve problems involving the BV norm. Also, these schemes cannot solve optimization problems involving multiple l1-regularization terms. For these reasons, it is difficult to apply the FPC and linearized Bregman methods to image processing problems. For example, it has been noted by many authors [16, 24] that optimal MRI reconstruction is obtained using a combination of both the BV norm and Besov norm,  $B_{1,1}$ . In this case, this problem can be written

$$\min_u \|u\|_{BV} + \|u\|_H + \frac{\mu}{2} \|\mathcal{F}u - f\|_2^2$$

where  $\|\cdot\|_{BV} = \|\nabla u\|_1$  is the ‘‘bounded variation’’ norm, and  $\|u\|_H = \|Hu\|_1$  denotes the ‘‘Besov’’ norm with respect to the Haar wavelet transform. We will later address this particular form of the compressed sensing problem.

In this paper, we will present a general technique that can be used to solve most common l1-regularized problems efficiently. Furthermore, the optimization scheme we employ can be generalized to solve a very broad range of equality-constrained optimization problems, some of which may be difficult to solve by existing techniques.

The contents of this paper are organized as follows: We begin with a brief discussion of constrained optimization problems and classical penalty function methods. We then review the concept of Bregman iteration, and use it to derive a general principle for solving constrained optimization problems. We will then define the ‘‘Split Bregman’’ method, and show how it can be used to solve the general l1-regularized optimization problem (1.1). Finally, we will apply the Split Bregman technique to TV denoising and compressed sensing problems to demonstrate its efficiency. The latter may involve BV, Besov, or multiple regularizers.

**1.1. Constrained Optimization Problems.** Consider a convex energy functional,  $E$ , and some linear function,  $A : R^n \rightarrow R^m$ . We wish to solve the generalized constrained optimization problem

$$(1.7) \quad \min_u E(u) \text{ s.t. } Au = b.$$

This problem can be very difficult to solve directly if  $E$  is non-differentiable.

In order to make (1.7) simpler to solve, we wish to convert it into an unconstrained optimization problem. One common method for doing this is to use a penalty function/continuation method, which approximates (1.7) by a problem of the form

$$(1.8) \quad \min_u E(u) + \frac{\lambda_k}{2} \|Au - b\|_2^2.$$

where  $\lambda_1 < \lambda_2 < \dots < \lambda_N$  is an increasing sequence of penalty function weights [2, 19]. In order to enforce that  $H(u) \approx 0$ , we must choose  $\lambda_N$  to be extremely large.

Unfortunately, for many problems, choosing a large value for  $\lambda$  makes (1.8) extremely difficult to solve numerically. We often wish to solve (1.8) by a Newton-type method, which requires us to invert the Hessian of the objective function. However, as  $\lambda_k \rightarrow \infty$ , the condition number of the Hessian approaches infinity, making it impractical to use fast iterative methods (such as Conjugate Gradient or Gauss-Seidel

methods)[2, 19]. Also, for many applications,  $\lambda_k$  must be increased in very small steps, making the method less efficient.

In the next section, we will show that Bregman iteration can be used to reduce (1.7) to a short sequence of unconstrained problems. In this sense, Bregman iteration is an alternative to conventional penalty function methods.

**2. Bregman Iteration.** Bregman iteration is a concept that originated in functional analysis for finding extrema of convex functionals [4]. Bregman iteration was first used in image processing by Osher et al. in [21], where it was applied to the ROF model for TV denoising. Bregman iteration has also been applied to solve the basis pursuit problem in [30, 5, 20], and was subsequently applied to medical imaging problems in [14]. Rather than focus on specific applications, we will present here a general formulation of this technique.

We begin with the concept of “Bregman Distance.” The Bregman Distance associated with a convex function  $E$  at the point  $v$  is

$$D_E^p(u, v) = E(u) - E(v) - \langle p, u - v \rangle$$

where  $p$  is in the subgradient of  $E$  at  $v$ . Clearly, this is not a distance in the usual sense because it is not in general symmetric. However, it does measure closeness in the sense that  $D_E^p(u, v) \geq 0$ , and  $D_E^p(u, v) \geq D_E^p(w, v)$  for  $w$  on the line segment between  $u$  and  $v$ .

Again, consider two convex energy functionals,  $E$  and  $H$ , defined over  $R^n$  with  $\min_{u \in R^n} H(u) = 0$ . The associated unconstrained minimization problem is

$$(2.1) \quad \min_u E(u) + \lambda H(u).$$

We can modify this problem by iteratively solving

$$(2.2) \quad u^{k+1} = \min_u D_E^p(u, u^k) + \lambda H(u)$$

$$(2.3) \quad = \min_u E(u) - \langle p^k, u - u^k \rangle + \lambda H(u)$$

as was suggested by Bregman in [4].

For simplicity, we will assume that  $H$  is differentiable. In this case, we have that  $0 \in \partial(D_E^p(u, u^k) + \lambda H(u))$  where this sub-differential is evaluated at  $u^{k+1}$ . Since  $p^{k+1} \in \partial E(u^{k+1})$  at this location, we have that

$$p^{k+1} = p^k - \nabla H(u^{k+1}).$$

In [21], the authors analyze the convergence of Bregman iterative schemes. In particular, it is shown that, under fairly weak assumptions on  $E$  and  $H$ , that  $H(u^k) \rightarrow 0$  as  $k \rightarrow \infty$ .

Two particular convergence results from [21] are especially relevant here, and so we restate them.

**THEOREM 2.1.** *Assume that  $E$  and  $H$  are convex functionals, and that  $H$  is differentiable. We also assume that solutions to the sub-problems in (2.2) exist. We then have*

- 1) *Monotonic decrease in  $H$ :  $H(u^{k+1}) \leq H(u^k)$*
- 2) *Convergence to a minimizer of  $H$ :  $H(u^k) \leq H(u^*) + J(u^*)/k$*

In addition to these convergence results, Bregman iteration has several nice denoising properties which are discussed and proved in [21] and [14].

**2.1. Constrained Optimization via Bregman Iteration.** In this section, we present a method for solving a wide variety of constrained optimization problems without using continuation. We first show that Bregman iteration can be used to solve the constrained problem (1.7). We will then discuss a simplified form of Bregman iteration, which is equivalent to “adding the noise back” as is done with ROF denoising [21]. Finally, we will discuss the convergence properties of this method.

We wish to solve

$$(2.4) \quad \min_u E(u) \text{ such that } Au = b$$

for some linear operator  $A$  and vector  $b$ . To apply formula (2.1), we make this into an unconstrained problem using a quadratic penalty function:

$$(2.5) \quad \min_u E(u) + \frac{\lambda}{2} \|Au - b\|_2^2$$

For small  $\lambda$ , the penalty function does not accurately enforce the constraint. The conventional solution to this problem is to let  $\lambda \rightarrow \infty$ . Rather, we apply the Bregman iteration (2.2), and iteratively minimize:

$$(2.6) \quad u^{k+1} = \min_u D_E^p(u, u^k) + \frac{\lambda}{2} \|Au - b\|_2^2$$

$$(2.7) \quad = \min_u E(u) - \langle p^k, u - u^k \rangle + \frac{\lambda}{2} \|Au - b\|_2^2$$

$$(2.8) \quad p^{k+1} = p^k - \lambda A^T (Au^{k+1} - b)$$

Bregman iterations of this form were considered in [30] and [21]. Here, it is shown that, when  $A$  is linear, this seemingly complicated iteration is equivalent to the simplified method:

$$(2.9) \quad u^{k+1} = \min_u E(u) + \frac{\lambda}{2} \|Au - b^k\|_2^2$$

$$(2.10) \quad b^{k+1} = b^k + b - Au^k$$

In other words, we simply add the error in the constraint back to the right hand side. This is the analog of “adding back the noise” in the ROF model for TV denoising [21].

Because of the equivalence of (2.6-2.8) and (2.9-2.10), and the convergence results of theorem 2.1, we have that

$$(2.11) \quad \lim_{k \rightarrow \infty} Au^k = b$$

where convergence is in the 2-norm sense. In other words, for large  $k$ , the iterates  $u^k$  satisfy the constraint condition to an arbitrarily high degree of accuracy.

We now need to show that a solution,  $u^*$ , of  $Au = b$  obtained through (2.9-2.10) is indeed a solution to the original constrained problem (2.4). Note that Bregman iteration was used to solve a constrained optimization problem in [30]. In this paper, the authors assume a specific form for the operator  $A$ . Here, we broaden this result and present a very simple proof.

**THEOREM 2.2.** *Let  $H : R^n \rightarrow R$  be convex. Let  $A : R^n \rightarrow R^m$  be linear. Consider the algorithm (2.9-2.10). Suppose that some iterate,  $u^*$ , satisfies  $Au^* = b$ . Then  $u^*$  is a solution to the original constrained problem (2.4).*

Proof: Let  $u^*$  and  $b^*$  be such that  $Au^* = b$ , and

$$(2.12) \quad u^* = \min_u E(u) + \frac{\lambda}{2} \|Au - b^*\|_2^2$$

Let  $\hat{u}$  be a true solution to (2.4). Then  $Au^* = b = A\hat{u}$ , which implies that

$$(2.13) \quad \|Au^* - b^*\|_2^2 = \|A\hat{u} - b^*\|_2^2$$

Because  $u^*$  satisfies (2.12), we have

$$(2.14) \quad E(u^*) + \frac{\lambda}{2} \|Au^* - b^*\|_2^2 \leq E(\hat{u}) + \frac{\lambda}{2} \|A\hat{u} - b^*\|_2^2.$$

Finally, note that (2.13) and (2.14) together imply

$$E(u^*) \leq E(\hat{u}).$$

Because  $\hat{u}$  satisfies the original optimization problem, this inequality can be sharpened to an equality, showing that  $u^*$  solves (2.4).

□

This shows that, provided (2.9-2.10) converges in the sense of (2.11), the iterates  $u^k$  will get arbitrarily close to a solution of the original constrained problem (1.7). Note the generality of the above theorem. The proof does not explicitly use the linearity of  $A$ , and in fact this condition is not required for the theorem to hold. The application of this method to problems in which  $A$  is not linear will be a subject of further research.

**2.2. Advantages of Bregman Iteration.** This Bregman iteration technique has several advantages over tradition penalty function/continuation methods. First, Bregman iteration converges very quickly when applied to certain types of objective functions, especially for problems where  $E$  contains an l1-regularization term. For an explanation of why this is true, see the attached appendix. When Bregman iteration converges quickly, we only need to solve a small number of unconstrained problems.

The second (and perhaps most significant) advantage of Bregman iteration over continuation methods is that the value of  $\lambda$  in (2.1) remains constant. We can therefore choose a value for  $\lambda$  that minimizes the condition number of the sub-problems, resulting in fast convergence for iterative optimization methods, such as Newton or Gauss-Seidel.

Bregman iteration also avoids the problem of numerical instabilities that occur as  $\lambda \rightarrow \infty$  that arise when using continuation methods.

### 3. Split Bregman - A Better formulation for l1 Regularized Problems.

We will now apply the Bregman framework to solve the general l1-regularized optimization problem (1.1). In the discussion that follows, we shall assume  $H(\cdot)$  and  $|\Phi(\cdot)|$  to be convex functionals. We shall also assume  $\Phi(\cdot)$  to be differentiable.

The key to our method is that we will “de-couple” the l1 and l2 portions of the energy in (1.1). This split formulation follows that proposed in [27], where a similar technique is applied to l1-regularized deconvolutions. Rather than considering (1.1), we will consider the problem

$$(3.1) \quad \min_{u,d} |d| + H(u) \text{ such that } d = \Phi(u)$$

This problem is clearly equivalent to (1.1). To solve this problem, first convert it into an unconstrained problem:

$$(3.2) \quad \min_{u,d} |d| + H(u) + \frac{\lambda}{2} \|d - \Phi(u)\|_2^2$$

This is where our method departs from [27]. If we let  $E(u, d) = |d| + H(u)$ , and define  $A(u, d) = d - \Phi(u)$ , then we can see that (3.2) is simply an application of (2.5). To enforce the constraint condition we now plug this problem into the above Bregman formulation (2.6-2.8).

$$(3.3) \quad (u^{k+1}, d^{k+1}) = \min_{u,d} D_E^p(u, u^k, d, d^k) + \frac{\lambda}{2} \|d - \Phi(u)\|_2^2$$

$$(3.4) \quad = \min_{u,d} E(u, d) - \langle p_u^k, u - u^k \rangle - \langle p_d^k, d - d^k \rangle + \frac{\lambda}{2} \|d - \Phi(u)\|_2^2$$

$$(3.5) \quad p_u^{k+1} = p_u^k - \lambda(\nabla\Phi)^T(\Phi u^{k+1} - d^{k+1})$$

$$(3.6) \quad p_d^{k+1} = p_d^k - \lambda(d^{k+1} - \Phi u^{k+1})$$

When we apply the simplification presented in (2.9-2.10), we get the elegant two-phase algorithm

*The Split Bregman Iteration*

$$(3.7) \quad (u^{k+1}, d^{k+1}) = \min_{u,d} |d| + H(u) + \frac{\lambda}{2} \|d - \Phi(u) - b^k\|_2^2$$

$$(3.8) \quad b^{k+1} = b^k + (\Phi(u^{k+1}) - d^{k+1})$$

We have reduced the l1-regularized problem (1.1) to a sequence of unconstrained optimization problems and Bregman updates. It may not be immediately clear why this algorithm is so effective. We will see in the next section that this formulation of the problem is much easier to compute than the conventional formulation for an l1 regularized problem.

**3.1. Iterative Minimization.** In order to implement the algorithm (3.7), we must be able to solve the problem

$$(3.9) \quad (u^{k+1}, d^{k+1}) = \min_{u,d} |d| + H(u) + \frac{\lambda}{2} \|d - \Phi(u) - b^k\|_2^2$$

Because of the way that we have “split” the l1 and l2 components of this functional, we can perform this minimization efficiently by iteratively minimizing with respect to  $u$  and  $d$  separately. The two steps we must perform are

$$\text{Step 1 : } u^{k+1} = \min_u H(u) + \frac{\lambda}{2} \|d^k - \Phi(u) - b^k\|_2^2$$

$$\text{Step 2 : } d^{k+1} = \min_d |d| + \frac{\lambda}{2} \|d - \Phi(u^{k+1}) - b^k\|_2^2$$

The speed of the Bregman splitting method is largely dependent on how fast we can solve each of these two subproblems.

To solve Step 1, note that because we have “de-coupled”  $u$  from the l1 portion of the problem, the optimization problem that we must solve for  $u^k$  is now differentiable. We can thus use a wide variety of optimization techniques to solve this problem. The particular method used to solve this optimization problem depends on the exact nature of  $H$ , but for many common problems either Gauss-Seidel or Fourier transform methods can be used. For rare problems in which  $\Phi$  has little structure, a few steps of a conjugate gradient method can be used to approximately solve this problem.

In Step 2 of the above algorithm, there is no coupling between elements of  $d$ . We can explicitly compute the optimal value of  $d$  using shrinkage operators. We simply compute

$$d_j^{k+1} = \mathit{shrink}(\Phi(u)_j + b_j^k, 1/\lambda)$$

where

$$\mathit{shrink}(x, \gamma) = \frac{x}{|x|} * \max(|x| - \gamma, 0).$$

this shrinkage is extremely fast, and requires only a few operations per element of  $d^{k+1}$ .

**3.2. Implementation of the Proposed Algorithm.** When we place the iterative minimization scheme into the process described in (3.7), we get the following:

**Generalized Split Bregman Algorithm**

```

While  $\|u^k - u^{k-1}\|_2 > \mathit{tol}$ 
  for n = 1 to N
     $u^{k+1} = \min_u H(u) + \frac{\lambda}{2} \|d^k - \Phi(u) - b^k\|_2^2$ 
     $d^{k+1} = \min_d |d| + \frac{\lambda}{2} \|d - \Phi(u^{k+1}) - b^k\|_2^2$ 
  end
   $b^{k+1} = b^k + (\Phi(u^{k+1}) - d^{k+1})$ 
end

```

We have found that it is not desirable to solve the first subproblem in (3.7) to full convergence. Intuitively, the reason for this is that if the error in our solution for this subproblem is small compared to  $\|b^k - b^*\|_2$ , then this extra precision will be “wasted” when the Bregman parameter is updated. In fact, we have found empirically that for many applications optimal efficiency is obtained when only one iteration of the inner loop is performed (i.e. N=1 in the above algorithm). Even when we only solve for  $u^{k+1}$  approximately (e.g. by using a few steps of an iterative method), the above algorithm still converges.

To understand why the Split Bregman algorithm is so robust to numerical imprecision, we must examine the results of theorem 2.2. Using this theorem, it is easy to show that any fixed point of the split Bregman algorithm is indeed a minimizer of the original constrained problem (3.1), even if we use inexact iterative methods for each subproblem. Let  $(u^*, b^*)$  be a fixed point of (3.7-3.8) that also satisfies (3.7). The fixed point satisfies  $b^* = b^* + \Phi u^* - d^*$ , which implies that  $d^* = \Phi u^*$ . This results, together with (3.7), satisfies the conditions of theorem 2.2, which shows that  $(u^*, b^*)$  is a solution of the constrained problem (3.1).

**4. Applications.** We will now illustrate how to use the Split Bregman framework by discussing several applications.

**4.1. TV Denoising.** TV denoising is considered to be one of the best denoising models, but also one of the hardest to compute. In this section, we will show how the Split Bregman technique can be used to solve this problem in a way that is not only simple, but also extremely efficient. Furthermore, this model can solve the isotropic TV minimization problem (a superior TV model which cannot be solved using the popular graph cuts method). We will treat this as a 2 dimensional problem to demonstrate that the Split Bregman method applies to problems with more than one l1 regularization term.

We begin by addressing the anisotropic problem

$$(4.1) \quad \min_u |\nabla_x u| + |\nabla_y u| + \frac{\mu}{2} \|u - f\|_2^2$$

To apply Bregman splitting, we first replace  $\nabla_x u$  by  $d_x$  and  $\nabla_y u$  by  $d_y$ . This yields the constrained problem

$$\min_u |d_x| + |d_y| + \frac{\mu}{2} \|u - f\|_2^2, \text{ such that } d_x = \nabla_x u \text{ and } d_y = \nabla_y u$$

To weakly enforce the constraints in this formulation, we add penalty function terms as was done in (3.2). This yields

$$\min_{d_x, d_y, u} |d_x| + |d_y| + \frac{\mu}{2} \|u - f\|_2^2 + \frac{\lambda}{2} \|d_x - \nabla_x u\|_2^2 + \frac{\lambda}{2} \|d_y - \nabla_y u\|_2^2.$$

Finally, we strictly enforce the constraints by applying the Bregman iteration (3.7) to get

$$\min_{d_x, d_y, u} |d_x| + |d_y| + \frac{\mu}{2} \|u - f\|_2^2 + \frac{\lambda}{2} \|d_x - \nabla_x u - b_x^k\|_2^2 + \frac{\lambda}{2} \|d_y - \nabla_y u - b_y^k\|_2^2.$$

where the proper values of  $b_x^k$  and  $b_y^k$  are chosen through Bregman iteration.

To solve this minimization problem, we will apply the iterative minimization approach (3.10), which requires us to solve the subproblem

$$u^{k+1} = \min_u \frac{\mu}{2} \|u - f\|_2^2 + \frac{\lambda}{2} \|d_x^k - \nabla_x u - b_x^k\|_2^2 + \frac{\lambda}{2} \|d_y^k - \nabla_y u - b_y^k\|_2^2$$

which has the optimality condition.

$$(4.2) \quad (\mu I - \lambda \Delta) u^{k+1} = \mu f + \lambda \nabla_x^T (d_x^k - b_x^k) + \lambda \nabla_y^T (d_y^k - b_y^k)$$

In order to achieve optimal efficiency, we wish to use a fast iterative algorithm to get approximate solutions to this system. Because the system is strictly diagonally dominant, the most natural choice is the Gauss-Seidel method. The Gauss-Seidel solution to this problem can be written component-wise as  $u_{i,j}^{k+1} = G_{i,j}^k$  where

$$G_{i,j}^k = \frac{\lambda}{\mu + 4\lambda} (u_{i+1,j}^k + u_{i-1,j}^k + u_{i,j+1}^k + u_{i,j-1}^k + d_{x,i-1,j}^k - d_{x,i,j}^k + d_{y,i,j-1}^k - d_{y,i,j}^k - b_{x,i-1,j}^k + b_{x,i,j}^k - b_{y,i,j-1}^k + b_{y,i,j}^k) + \frac{\mu}{\mu + 4\lambda} f_{i,j}.$$

Using this solver, the Split Bregman algorithm is written

**Split Bregman Anisotropic TV Denoising**

Initialize:  $u^0 = f$ , and  $d_x^0 = d_y^0 = b_x^0 = b_y^0 = 0$

While  $\|u^k - u^{k-1}\|_2 > tol$

$$u^{k+1} = G^k$$

$$d_x^{k+1} = shrink(\nabla_x u^{k+1} + b_x^k, 1/\lambda)$$

$$d_y^{k+1} = shrink(\nabla_y u^{k+1} + b_y^k, 1/\lambda)$$

$$b_x^{k+1} = b_x^k + (\nabla_x u^{k+1} - d_x^{k+1})$$

$$b_y^{k+1} = b_y^k + (\nabla_y u^{k+1} - d_y^{k+1})$$

end

note that the “for” loop in the Generalized Split Bregman algorithm is absent here. We have found that this algorithm attains optimal efficiency when this loop is executed only once per iteration, and have therefore removed the loop for clarity. It may be necessary to include this loop in applications where high precision results are needed.

The isotropic TV model can also be minimized using the split Bregman technique. In this case, we wish to solve

$$\min_u \sum_i \sqrt{(\nabla_x u)_i^2 + (\nabla_y u)_i^2} + \frac{\mu}{2} \|u - f\|_2^2$$

Just as we did for the anisotropic problem, we will split the l1 and l2 components of the problem by setting  $d_x \approx \nabla_x u$  and  $d_y \approx \nabla_y u$ . The split Bregman formulation of the problem then becomes:

$$\min_{u, d_x, d_y} \|(d_x, d_y)\|_2 + \frac{\mu}{2} \|u - f\|_2^2 + \frac{\lambda}{2} \|d_x - \nabla_x u - b_x\|_2^2 + \frac{\lambda}{2} \|d_y - \nabla_y u - b_y\|_2^2$$

where

$$(4.3) \quad \|(d_x, d_y)\|_2 = \sum_{i,j} \sqrt{d_{x,i,j}^2 + d_{y,i,j}^2}$$

Note that the  $d_x$  and  $d_y$  variables do not decouple as they did in the anisotropic case. This changes the way in which we must treat these variables. In order to apply the iterative minimization procedure to this problem, we must solve the subproblem

$$(d_x^{k+1}, d_y^{k+1}) = \min_{d_x, d_y} \|(d_x, d_y)\|_2 + \frac{\lambda}{2} \|d_x - \nabla_x u - b_x\|_2^2 + \frac{\lambda}{2} \|d_y - \nabla_y u - b_y\|_2^2$$

Despite the fact that the variables  $d_x$  and  $d_y$  do not decouple as they did in the anisotropic case, we can still explicitly solve the minimization problem for  $(d_x^{k+1}, d_y^{k+1})$  using a generalized shrinkage formula [27]:

$$d_x^{k+1} = \max(s^k - 1/\lambda, 0) \frac{\nabla_x u^k + b_x^k}{s^k}$$

$$d_y^{k+1} = \max(s^k - 1/\lambda, 0) \frac{\nabla_y u^k + b_y^k}{s^k}$$

where

$$(4.4) \quad s^k = \sqrt{|\nabla_x u^k + b_x^k|^2 + |\nabla_y u^k + b_y^k|^2}$$

If we apply Bregman iteration to this problem, we get the minimization algorithm for the isotropic TV functional:

**Split Bregman Isotropic TV Denoising**

Initialize:  $u^0 = f$ , and  $d_x^0 = d_y^0 = b_x^0 = b_y^0 = 0$

While  $\|u^k - u^{k-1}\|_2 > tol$

$$u^{k+1} = G_{i,j}^k$$

$$d_x^{k+1} = \max(s^k - 1/\lambda, 0) \frac{\nabla_x u^k + b_x^k}{s^k}$$

$$d_y^{k+1} = \max(s^k - 1/\lambda, 0) \frac{\nabla_y u^k + b_y^k}{s^k}$$

$$b_x^{k+1} = b_x^k + (\nabla_x u^{k+1} - d_x^{k+1})$$

$$b_y^{k+1} = b_y^k + (\nabla_y u^{k+1} - d_y^{k+1})$$

end

where  $s^k$  is defined above by (4.4).

**4.2. Fast Compressed Sensing for Image Reconstruction.** Compressed sensing (CS) is an emerging area of Medical Imaging, and many people predict that it will some day be a commonplace tool for radiologists. Consequently, fast algorithms for this problem are extremely desirable.

The exact formulation of the CS optimization problem depends somewhat on the application being considered. For demonstration purposes, we shall focus on the application of CS for Sparse Magnetic Resonance Imaging (MRI). We choose this application not only because of the great success of CS in this field, but also because the difficulties of this problem allow us to demonstrate the versatility of the Split Bregman method.

The general form for the Sparse MRI reconstruction problem is presented and discussed in [16], [14] and [23]. We choose to write this problem in the form

$$(4.5) \quad \min_u J(u) \text{ such that } \|R\mathcal{F}u - f\|_2^2 < \sigma^2$$

where  $\mathcal{F}$  represents the Fourier transform matrix,  $f$  represents the observed “ $k$ -space” data, and  $\sigma$  represents the variance of the signal noise. The matrix  $R$  represents a “row selector” matrix, which comprises a subset of the rows of an identity matrix. Also,  $J(u)$  represents some l1 regularization term.

In [30] it was shown that, using a Bregman iteration technique, the problem (4.5) could be reduced to a sequence of unconstrained problems of the form

$$(4.6) \quad u^{k+1} = \min_u J(u) + \frac{\mu}{2} \|R\mathcal{F}u - f^k\|_2^2$$

$$(4.7) \quad f^{k+1} = f^k + f - R\mathcal{F}u^{k+1}$$

It is this unconstrained problem that we wish to solve using the split Bregman technique.

To be concrete about the regularization term that we are using, we will now choose a specific form for  $J(\cdot)$ . Several authors have observed that superior image reconstructions occur when a hybrid of Total-Variation and Besov regularizers are used. Following these authors, we choose  $J(u) = \|u\|_{BV} + \|u\|_{B_{1,1}} = |\nabla u| + |Wu|$ , where  $W$  represents the discrete Haar orthogonal wavelet transform. While it may seem that the BV and Haar regularizers are very similar, we have found that the inclusion of multiple regularizers helps to ensure accurate reconstruction of smooth

images. Since the inclusion of such “overcomplete” transforms has little effect on the convergence rate of the Split Bregman algorithm, and both transforms can be evaluated quickly, there is little additional cost to including such regularizers.

Note that  $u$  can now take on complex values, and so we must be precise about our notation. In the following, we have  $|v| = \|v\|_1 = \sum_i \sqrt{v_i^H v_i}$  where  $v_i^H$  denotes the Hermitian transpose of the vector  $v_i$ . Using this definition of the l1 norm, we may define  $\|v\|_{BV} = |\nabla v| = \sum_i \sqrt{|\nabla_x v|^2 + |\nabla_y v|^2}$ .

To apply the Split Bregman method to this problem, we first make the replacements  $w \leftarrow Wu$ ,  $d_x \leftarrow \nabla_x u$ , and  $d_y \leftarrow \nabla_y u$ . The split formulation of the problem then becomes

$$\begin{aligned} \min_{u, d_x, d_y, w} \quad & \|(d_x, d_y)\|_2 + |w| + \frac{\mu}{2} \|R\mathcal{F}u - f\|_2^2 + \frac{\lambda}{2} \|d_x - \nabla_x u - b_x\|_2^2 \\ & + \frac{\lambda}{2} \|d_y - \nabla_y u - b_y\|_2^2 + \frac{\gamma}{2} \|w - Wu - b_w\|_2^2 \end{aligned}$$

Where we have used the short hand notation

$$\|(d_x, d_y)\|_2 = \sum_{i,j} \sqrt{|d_{x,i,j}|^2 + |d_{y,i,j}|^2}$$

which is the complex analog of (4.3). Note that  $d_x$ ,  $d_y$ ,  $b_x$ , and  $b_y$  are now complex valued, and are obtained by applying the difference operator to the real and complex parts of  $u$  separately.

We then decompose this minimization into subproblems using the iterative minimization procedure (3.10). We may use the generalized shrinkage formula (4.4) to solve for the optimal values of  $d_x$  and  $d_y$ , and the standard shrinkage formula to solve for the optimal value of  $w$ . To find the optimal value of  $u$ , we must solve the optimization sub-problem

$$u^{k+1} = \min_u \frac{\mu}{2} \|\mathcal{F}u - f\|_2^2 + \frac{\lambda}{2} \|d_x^k - \nabla_x u - b_x^k\|_2^2 + \frac{\lambda}{2} \|d_y^k - \nabla_y u - b_y^k\|_2^2 + \frac{\gamma}{2} \|w^k - Wu - b_w^k\|_2^2$$

Because this subproblem is differentiable, optimality conditions for  $u^{k+1}$  are easily derived. By differentiating with respect to  $u$  and setting the result equal to zero, we get the update rule

$$(\mu\mathcal{F}^T R^T R\mathcal{F} + \lambda\nabla_x^T \nabla_x + \lambda\nabla_y^T \nabla_y + \gamma W^T W)u^{k+1} = rhs^k$$

where

$$rhs^k = \mu\mathcal{F}^T Rf + \lambda\nabla_x^T (d_x^k - b_x) + \lambda\nabla_y^T (d_y^k - b_y) + \gamma W^T (w - b_w)$$

represents the right hand side in the above equation.

We now take advantage of the identities  $\nabla^T \nabla = -\Delta$ ,  $W^T W = I$ , and  $\mathcal{F}^T = \mathcal{F}^{-1}$  to get

$$(\mu\mathcal{F}^T R^T R\mathcal{F} - \lambda\Delta + \gamma I)u^{k+1} = rhs^k$$

Therefore, the system that must be inverted to solve for  $u^{k+1}$  is circulant. We can thus write the system as  $\mathcal{F}^{-1} K \mathcal{F}$ , where  $K$  is the diagonal operator

$$K = (\mu R^T R - \lambda \mathcal{F} \Delta \mathcal{F}^{-1} + \gamma I)$$

Because of the circulant structure of this system, we can solve for the optimal value of  $u^{k+1}$  using only two Fourier transforms.

When we put all of these elements together, we get the following algorithm:

#### Unconstrained CS Optimization Algorithm

Initialize:  $u^0 = \mathcal{F}^{-1}f$ , and  $d_x^0 = d_y^0 = w^0 = b_x^0 = b_y^0 = b_w^0 = 0$

While  $\|u^k - u^{k-1}\|_2 > tol$

$$u^{k+1} = \mathcal{F}^{-1}K^{-1}\mathcal{F}rh_s^k$$

$$d_x^{k+1} = \max(s^k - 1/\lambda, 0) \frac{\nabla_x u^k + b_x^k}{s^k}$$

$$d_y^{k+1} = \max(s^k - 1/\lambda, 0) \frac{\nabla_y u^k + b_y^k}{s^k}$$

$$w^{k+1} = shrink(Wu^{k+1} + b_w^k, 1/\gamma)$$

$$b_x^{k+1} = b_x^k + (\nabla_x u^{k+1} - d_x^{k+1})$$

$$b_y^{k+1} = b_y^k + (\nabla_y u^{k+1} - d_y^{k+1})$$

$$b_w^{k+1} = b_w^k + (Wu^{k+1} - w^{k+1})$$

end

where  $s^k$  is defined in (4.4).

Note that this algorithm only solves the unconstrained CS problem (4.6). To solve the constrained problem (4.5), we must replace  $f$  by  $f^k$  in the above algorithm. After approximately solving each unconstrained problem, we must apply the Bregman update rule  $f^{k+1} = f^k + f - R\mathcal{F}u^{k+1}$ . When we embed the unconstrained algorithm inside of this outer Bregman update, we get

#### Constrained CS Optimization Algorithm

Initialize:  $u^0 = \mathcal{F}^{-1}f$ , and  $d_x^0 = d_y^0 = w^0 = b_x^0 = b_y^0 = b_w^0 = 0$

While  $\|R\mathcal{F}u^k - f\|_2^2 > \sigma^2$

For i= 1 to N

$$u^{k+1} = \mathcal{F}^{-1}K^{-1}\mathcal{F}rh_s^k$$

$$d_x^{k+1} = \max(s^k - 1/\lambda, 0) \frac{\nabla_x u^k + b_x^k}{s^k}$$

$$d_y^{k+1} = \max(s^k - 1/\lambda, 0) \frac{\nabla_y u^k + b_y^k}{s^k}$$

$$w^{k+1} = shrink(Wu^{k+1} + b_w^k, 1/\gamma)$$

$$b_x^{k+1} = b_x^k + (\nabla_x u^{k+1} - d_x^{k+1})$$

$$b_y^{k+1} = b_y^k + (\nabla_y u^{k+1} - d_y^{k+1})$$

$$b_w^{k+1} = b_w^k + (Wu^{k+1} - w^{k+1})$$

end

$$f^{k+1} = f^k + f - R\mathcal{F}u^{k+1}$$

end

The speed of this algorithm will largely depend on how many times the outer “while” loop needs to be executed. When parameter values are properly chosen, it has been found that the outer loop of this algorithm only needs to be executed a small number of times. Also, for imaging applications it is not necessary to solve each unconstrained sub-problem entirely to numerical precision. As a result, this algorithm is very fast when properly chosen parameter values are used.

## 5. Numerical Results.

**5.1. TV Denoising Results.** We will now examine the efficiency of the split Bregman approach using time trials. The split Bregman algorithm was implemented in C++, and compiled on a UNIX platform using the g++ compiler. Time trials were generated on an Intel Core 2 Duo desktop PC (E6850, 3.00 GHz).

We tested our method on two images: The first was a  $256 \times 256$  synthetic image of two overlapping squares. The second was a  $512 \times 512$  representation of the famous test image “Lena.” Both images were contaminated with noise ( $\sigma = 15$ ). Denoising parameters were  $\mu = .05$ ,  $\lambda = 0.1$  for both images. In general, we have found that choosing  $\lambda = 2\mu$  usually results in good convergence. Iterations were terminated when the condition  $\|u^k - u^{k-1}\|/\|u^k\| < 5 * 10^{-3}$  was met. Results of time trials for the isotropic ROF algorithm are shown in table 5.1. For comparison, we also report the computation time of a graph-cuts based solver [29, 12], which minimizes an anisotropic TV functional using either a 4-neighbor stencil, or a less anisotropic 16-point stencil. Note that in addition to outperforming the graph cuts algorithm in terms of speed, the Split Bregman method can minimize the isotropic functional, and is thus less likely to introduce artifacts into the image.

TABLE 5.1  
ROF Computation Times (sec)

Image	Split Bregman	Graph Cuts(4 point)	Graph Cuts(16 point)
$256 \times 256$ Blocks	0.0732	0.214	0.468
$512 \times 512$ Lena	0.2412	0.709	1.51

The convergence speed of the Split Bregman method can also be observed in figure 5.1, where we have plotted the error vs. iteration number for the test image “Lena” as described above. To compare our method to another unregularized iterative method, we have also plotted the convergence curve for the gradient descent algorithm [8], which is based on the dual formulation of the ROF functional. We have found that both methods converge quickly for the first several iterations, but then slow down as the exact solution is reached. This stalling effect, however, is considerably less severe for the Split Bregman method.

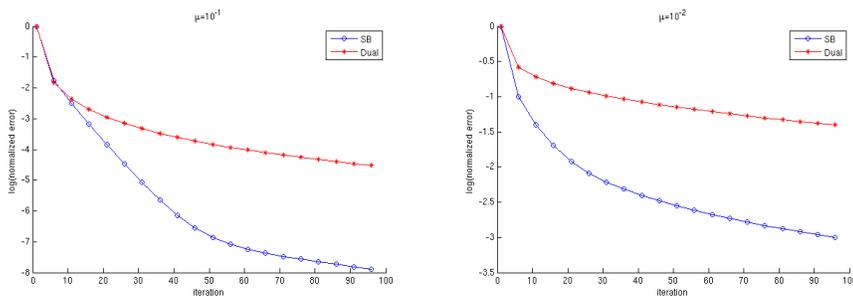


FIG. 5.1. Error vs. iteration number for the Split Bregman ROF minimization algorithm, and for an iterative method based on the dual problem [8]. The error at iteration  $k$  is defined as  $\|u^k - u^*\|$  where  $u^k$  is the approximation at iteration  $k$ , and  $u^*$  is the exact solution. Convergence results are for the test image “Lena” with Gaussian noise ( $\sigma = 15$ ). (top left) Results for  $\mu = 0.1$ . (top right) Results for  $\mu = 0.01$ .

In addition to being fast, the split Bregman method has the advantage that, before

convergence is reached, the intermediate images are smooth. Most of the image noise is eliminated during the first 10 iterations of the algorithm, and subsequent iterations serve to flatten “staircases.” For both the geometric image and “Lena,” almost 50 iterations of the Split Bregman algorithm were needed to reach convergence within 0.5 intensity units per pixel. However, because intermediate results of the algorithm are smooth, it is extremely difficult to discern between the results after 10 iterations, and the results after full convergence. This is illustrated using the image of Lena in Figure 5.2. This property is also illustrated in Figure 5.3, where we show the test image, as well as its cross sections, at various level of convergence.

Note that the speed of this algorithm depends strongly on the fact that we have used Bregman iteration, rather than continuation, to enforce a constraint. Had we followed a more conventional path, and let  $\lambda \rightarrow \infty$ , the problem (4.2) would have become ill-conditioned, and the Gauss-Seidel method would have stalled.



FIG. 5.2. *Split Bregman ROF denoising of Lena. (top left) Original image. (top right) Noise contaminated with  $\sigma = 25$ . (bottom left) Denoised with 10 iterations of Split Bregman. (bottom right) Denoised with 50 iterations of Split Bregman.*

**5.2. Compressed Sensing Results.** We now discuss the efficiency of the Split Bregman compressed sensing algorithm using BV and/or  $B_{1,1}$  as regularizers. We tested the Split Bregman algorithm on two sample images, one synthetic and one real. Two different formulations of the CS problem were tested. First, the compressed sensing reconstruction was performed using only a Besov regularizer. Next, we reconstructed images using the hybrid Besov+BV algorithm discussed above. In the case where only the Besov regularizer is used, the optimization problem can be put into a particularly simple form, which allows it to be solvable by the Linearized Bregman method. In this case, we compare the Split Bregman algorithm to the Linearized Bregman algorithm with “kicking” [20] for efficiency.

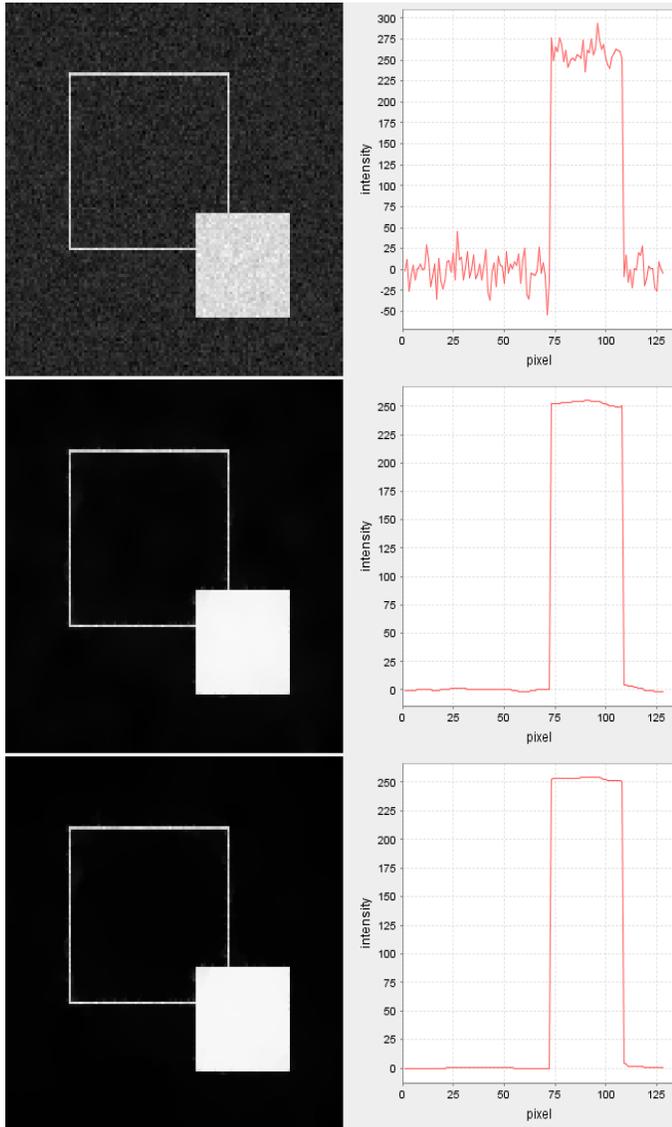


FIG. 5.3. *Split Bregman ROF denoising of the geometric test image. A cross section of each image is displayed on the right. (top) Noisy image with  $\sigma = 15$ . (middle) Denoised with 10 iterations of Split Bregman. (bottom) Denoised with 50 iterations of Split Bregman.*

In table 5.2, we present the number of iterations of each algorithm required to reach the convergence criteria for each task described below. Note that with both the Split Bregman and Linearized Bregman algorithm, two Fourier transforms are required per iteration. Because the Fourier transform is by far the most time consuming step in each algorithm, we found that both algorithms require a similar amount of time to execute one iteration.

*Example 1:* We generated a sample synthetic image with dimensions  $128 \times 128$ , and pixel intensities ranging from 0 to 255. The image consists of two overlapping geometric shapes, and has an extremely sparse gradient and Haar wavelet transform.

The synthetic image has signal in both its real and imaginary components, however for display purposes we show the absolute value of the image. For our first test, no noise was added to the CS data. The objective of this test was to recover the exact image using only 50% of the  $k$ -space data (sampled at random). The algorithm was run until  $\|\mathcal{F}u - f^k\|_2/128^2 < 10^{-3}$ . For the second test, the CS data was contaminated with noise ( $\sigma = 25$ ), and each algorithm was run until the stopping criteria  $\|\mathcal{F}u - f^k\|_2/128^2 < \sigma$  was met.

*Example 2:* To demonstrate the effectiveness of the split Bregman method in medical imaging, we tested the algorithm on real MRI data. For this purpose, we used a cross-sectional image of a saline phantom. In the original image acquisition, the entire  $k$ -space was sampled. To generate the CS data, we randomly and uniformly selected 30% of the  $k$ -space samples. Note that, while ideal for compressed sensing, this type of sampling is not practical for most MRI applications. This is because most pulse sequences acquire  $k$ -space data in some sort of geometric pattern, such as a spiral [18, 24, 16]. Because the focus of this paper is on numerics, and not on the details of image acquisition, we choose uniform random sampling for simplicity.

TABLE 5.2  
*Iteration Counts for CS Algorithms*

Algorithm	Geometric	Geometric, $\sigma = 25$	MRI Phantom
Linearized, Besov	367	35	51
Split, Besov	10	5	40
Split, Besov+BV	17	10	35

Several results from table 5.2 are particularly noteworthy. First, we found that the linearized Bregman algorithm was fairly efficient for large  $\sigma$ , but converged very slowly for  $\sigma = 10^{-3}$ . The reason for this is that the linearized Bregman algorithm slowed down considerably as the method approached convergence. Also, to attain this high level of accuracy, fairly extreme values for the method parameters had to be chosen, which resulted in slow performance. For all tasks involving noisy data, the split Bregman method was comparable to, or even faster than, linearized Bregman. One reason for this is that the Split Bregman method is very tolerant of large values of  $\mu$  in equation (4.6). Because we were able to choose large values for this parameter, the number of outer “while” loops of the Split Bregman CS algorithm was very small (8 or less for problems involving the phantom test image). Also, the inner “for” loop had to be executed only 5 times between each outer loop iteration. These convergence results are demonstrated using the MR phantom in Figure 5.4.

Note that the most significant advantage of the split Bregman method over other compressed sensing methods is its versatility. The split Bregman method can minimize energies involving the BV norm, which most CS algorithms (such as linearized Bregman and FPC) cannot. Also, the split Bregman method can solve optimization problems involving multiple regularization terms.

**6. Conclusion.** In this paper, we introduce the “Split Bregman” framework for solving  $l_1$ -regularized optimization problems. By applying the method to image denoising and compressed sensing problems, we showed that this method is a very efficient solver for many problems that are difficult to solve by other means. Besides its speed, our algorithm has several advantages: Because the Split Bregman algorithm

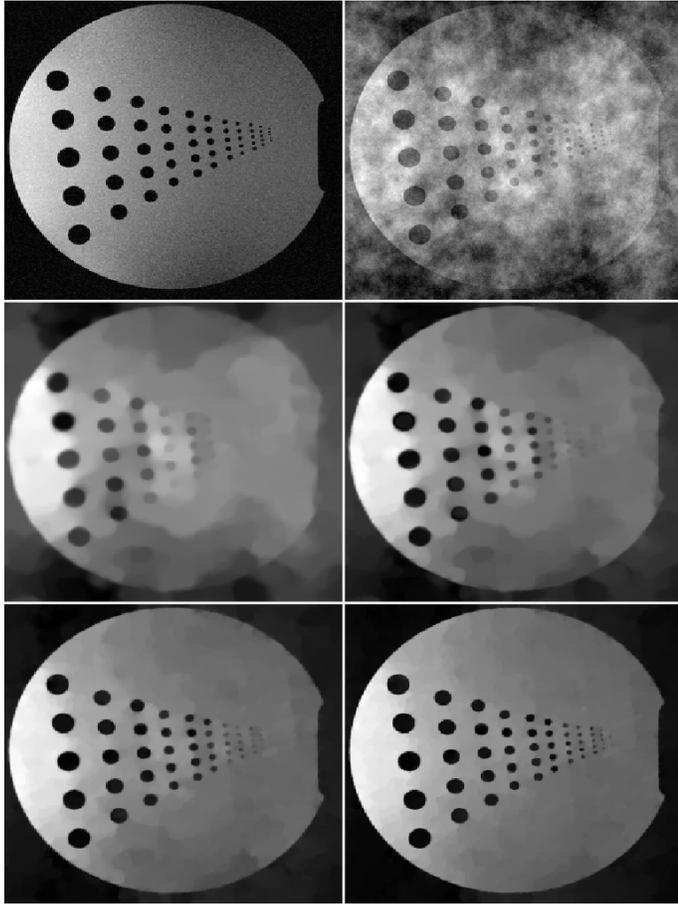


FIG. 5.4. *Split Bregman compressed sensing reconstruction of an MR image using 30% of the  $k$ -space data. (top left) Original image reconstructed using the full  $k$ -space. (top right) Image reconstructed using 30% of  $k$ -space data. Reconstruction was done using the conventional method, which fills in missing samples with zeros. (middle left) Results of Split Bregman algorithm after 10 inner iterations (2 outer iterations). (middle right) Results of Split Bregman algorithm after 20 inner iterations (4 outer iterations). (bottom left) Results after 30 inner iterations. (bottom right) Optimal results obtained after 40 inner iterations.*

makes extensive use of Gauss-Seidel and Fourier transform methods, it is easily parallelizable. Also, it has a relatively small memory footprint compared to second order methods that require explicit representations of the Hessian matrix. Both of these characteristics make Split Bregman a practical algorithm for large scale problems. Finally, the method is easy to code.

**Appendix: Fast Convergence of Bregman Iteration.** Here we give an informal, but intuitive, explanation of why Bregman iteration converges so quickly for  $l_1$  based problems. This is a generalization of what was done in [28], where Bregman iteration was applied to soft thresholding for Wavelet denoising.

Consider the optimization problem

$$\min_u |u|_\epsilon \text{ such that } Au = f$$

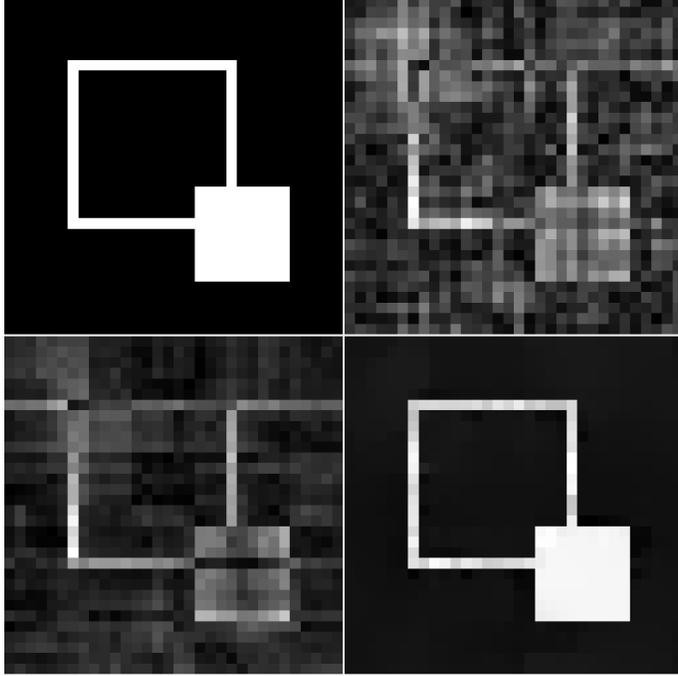


FIG. 5.5. Synthetic image recovered from 50% of  $k$ -space data. The  $k$ -space data was contaminated with noise ( $\sigma = 15$ ) prior to recovery. (top left) Original image reconstructed using the full  $k$ -space. (top right) Image reconstructed using 50% of  $k$ -space data. Reconstruction was done using the conventional method, which fills in missing samples with zeros. (bottom left) CS reconstruction using only the Besov regularizer. (bottom right) CS results using the hybrid TV+Besov regularizer. Note the superiority of the results when the hybrid regularizer is used.

where

$$|u|_\epsilon = \sum_i \sqrt{u_j^2 + \epsilon}$$

is a “smoothed out” variant of the l1 norm. We will need this smoothness to perform the analysis below. The Bregman update rule for this problem is

$$(6.1) \quad p^{k+1} - p^k + A^T(Au^{k+1} - f) = 0$$

Now, because  $J(\cdot) = |\cdot|_\epsilon$  is convex, and  $p^k$  and  $p^{k+1}$  are gradients of this functional at  $u^k$  and  $u^{k+1}$ , the mean value theorem tells us that

$$p^{k+1} - p^k = D^{k+\frac{1}{2}}(u^{k+1} - u^k)$$

where  $D^{k+\frac{1}{2}}$  is a diagonal matrix such that

$$D_{i,i}^{k+\frac{1}{2}} = \epsilon \left( (u_i^{k+\frac{1}{2}})^2 + \epsilon \right)^{-3/2}$$

for some  $u^{k+\frac{1}{2}}$  between  $u^k$  and  $u^{k+1}$ .

Applying this formula to (6.1) yields

$$D^{k+\frac{1}{2}}(u^{k+1} - u^k) + A^T(Au^{k+1} - f) = 0$$

We now let  $Q^{k+\frac{1}{2}} = (D^{k+\frac{1}{2}})^{-1}$ , multiply by  $AQ^{k+\frac{1}{2}}$ , and rearrange to get

$$(6.2) \quad Au^{k+1} - f = \left( I + AQ^{k+\frac{1}{2}}A^T \right)^{-1} (Au^k - f)$$

This equation gives us some insight into why Bregman iteration behaves as it does. When  $u_i^{k+\frac{1}{2}}$  is large compared to  $\epsilon$ , which occurs at “spikes” (or edges in the BV case),  $Q_{i,i}^{k+\frac{1}{2}}$  is large, and  $u_i^k$  converges rapidly. Small values of  $u^k$  take more iterations to settle down. An important result of this is that for problems with shocks or edges, Bregman iteration puts these features in the right place almost immediately, unlike the continuation based method [27]. Resolving discontinuities is usually the most difficult part of any imaging task, and Bregman iteration is extremely well suited for this task.

**Acknowledgment.** We thank Jie Zheng for his helpful discussions regarding MR image processing. This publication was made possible by the support of the National Science Foundation’s GRFP program, as well as ONR grant N000140710810 and the Department of Defense.

#### REFERENCES

- [1] R. Baraniuk and P. Steeghs. Compressive radar imaging. *Radar Conference, 2007 IEEE*, pages 128–133, 2007.
- [2] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence*, 23:1222 – 1239, 2005.
- [4] L Bregman. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex optimization. *USSR Computational Mathematics and Mathematical Physics*, 7:200–217, 1967.
- [5] JF Cai, S. Osher, and Z Shen. Linearized Bregman iterations for compressed sensing. *UCLA CAM Report*, 08-06.
- [6] E. J. Candes and J. Romberg. Signal recovery from random projections. *Proceedings of SPIE Computational Imaging III*, 5674:76 – 86, 2005.
- [7] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52:489 – 509, 2006.
- [8] A. Chambolle. Total variation minimization and a class of binary mrf models. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 136 – 152. Springer, 2005.
- [9] Tony F. Chan, Gene H. Golub, and Pep Mulet. A nonlinear primal-dual method for total variation-based image restoration. In *ICAOS '96 (Paris, 1996)*, volume 219, pages 241–252, Berlin, Germany / Heidelberg, Germany / London, UK / etc., 1996. Springer-Verlag.
- [10] J Darbon and M Sigelle. A fast and exact algorithm for total variation minimization. *IbPRIA*, 3522:351–359, 2005.
- [11] D Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52:1289–1306, 2006.
- [12] D. Goldfarb and W. Yin. Parametric maximum flow algorithms for fast total variation minimization. *CAAM technical report, TR07-09*, 2008.
- [13] Elaine Hale, Wotao Yin, and Yin Zhang. A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing. *CAAM Technical Report, TR07*, 2007.
- [14] Lin He, Ti-Chiun Chang, and Stanley Osher. Mr image reconstruction from sparse radial samples by using iterative refinement procedures. *Proceedings of the 13th annual meeting of ISMRM*, page 696, 2006.
- [15] Y Li and F Santosa. An affine scaling algorithm for minimizing total variation in image enhancement. *Tech Repost 12/94, Center for theory and simulation in science and engineering, Cornell University*, (TR94-1470):24, 1994.

- [16] M Lustig, D Donoho, and J Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*, 58:1182–1195, 2007.
- [17] M Lustig, J.H. Lee, D.L. Donoho, and J.M. Pauly. Faster imaging with randomly perturbed undersampled spirals and l1 reconstruction. *Proc. Of the ISMRM*, 2005.
- [18] GL Marseille, R de Beer, M Fuderer, AF Mehlkopf, and D van Ormondt. Nonuniform phase-encode distributions for MRI scan-time reduction. *J Magn Reson*, 111:70–75, 1996.
- [19] J Nocedal and S. Wright. *Numerical Optimization*. Springer Verlag, 2006.
- [20] A Osher, Y Mao, B Dong, and W Yin. Fast linearized Bregman iterations for compressed sensing and sparse denoising. *UCLA CAM Report*, 08-37.
- [21] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. An iterative regularization method for total variation-based image restoration. *MMS*, 4:460–489, 2005.
- [22] L Rudin, S Osher, and E Fatemi. Nonlinear total variation based noise removal algorithms. *Physica. D.*, 60:259–268, 1992.
- [23] Kim S, K Koh, M Lustig, S Boyd, and D Gerinvesky. A method for large-scale l1-regularized least squares problems with applications in signal processing and statistics. *Tech. Report, Dept. of Electrical Engineering, Stanford University*, 2007.
- [24] J Trzasko, A Manduca, and E Borisch. Sparse mri reconstruction via multiscale l0-continuation. *Statistical Signal Processing*, 26–29:176 – 180, 2007.
- [25] C. Vogel. A multigrid method for total variation-based image denoising. *Computation and Control IV, Progress in Systems and Control Theory*, 20, Birkhauser, 1995.
- [26] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. *SIAM Journal on Scientific Computing*, 17(1):227–238, 1996.
- [27] Y Wang, W Yin, and Y Zhang. A fast algorithm for image deblurring with total variation regularization. *CAAM Technical Reports*, 2007.
- [28] J Xu and S Osher. Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising. *IEEE Transactions on Image Processing*, 16:534–544, February 2007.
- [29] W. Yin. Pgc: A preflow-push based graph-cut solver. version 2.32.
- [30] W Yin, S Osher, D Goldfarb, and J Darbon. Bregman iterative algorithms for l1-minimization with applications to compressed sensing. *Siam J. Imaging Science*, 1:142–168, 2008.