SIAM J. IMAGING SCIENCES
Vol. 1, No. 1, pp. 143–168

Bregman Iterative Algorithms for ℓ_1-Minimization with Applications to Compressed Sensing*

Wotao Yin[†], Stanley Osher[‡], Donald Goldfarb[§], and Jerome Darbon[‡]

Abstract. We propose simple and extremely efficient methods for solving the basis pursuit problem $\min\{\|u\|_1 : Au = f, u \in \mathbb{R}^n\}$, which is used in compressed sensing. Our methods are based on Bregman iterative regularization, and they give a very accurate solution after solving only a very small number of instances of the unconstrained problem $\min_{u\in\mathbb{R}^n} \mu\|u\|_1 + \frac{1}{2}\|Au - f^k\|_2^2$ for given matrix A and vector f^k. We show analytically that this iterative approach yields exact solutions in a finite number of steps and present numerical results that demonstrate that as few as two to six iterations are sufficient in most cases. Our approach is especially useful for many compressed sensing applications where matrix-vector operations involving A and $A^\top$ can be computed by fast transforms. Utilizing a fast fixed-point continuation solver that is based solely on such operations for solving the above unconstrained subproblem, we were able to quickly solve huge instances of compressed sensing problems on a standard PC.

Key words. ℓ_1-minimization, basis pursuit, compressed sensing, iterative regularization, Bregman distances

AMS subject classifications. 49, 90, 65

DOI. 10.1137/070703983

1. Introduction.

Let $A \in \mathbb{R}^{m\times n}$, $f \in \mathbb{R}^m$, and $u \in \mathbb{R}^n$. The basis pursuit problem [23] solves the constrained minimization problem

$$\text{(Basis Pursuit)} \quad \min_u\{\|u\|_1 : Au = f\} \tag{1.1}$$

to determine an ℓ_1-minimal solution u_{opt} of the linear system $Au = f$, typically underdetermined; i.e., $m < n$ (in many cases, $m \ll n$), and $Au = f$ has more than one solution.

The basis pursuit problem (1.1) arises in the applications of compressed sensing (CS). A recent burst of research in CS was led by Candès and Romberg [12], Candès, Romberg, and Tao [14], Candès and Tao [16], Donoho [35], Donoho and Tanner [36], Tsaig and Donoho [88], and others [80, 86]. The fundamental principle of CS is that, through optimization, the sparsity of a signal can be exploited for recovering that signal from incomplete measurements of it. Let the vector $\bar{u} \in \mathbb{R}^n$ denote a highly sparse signal (i.e., $k = \|\bar{u}\|_0 := |\{i : \bar{u}_i \neq 0\}| \ll n$).

*Received by the editors September 28, 2007; accepted for publication (in revised form) January 29, 2008; published electronically March 20, 2008.
http://www.siam.org/journals/siims/1-1/70398.html

[†]Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005 (wotao.yin@rice.edu). This author's research was supported by an internal faculty research grant from the Dean of Engineering at Rice University.

[‡]Department of Mathematics, UCLA, Los Angeles, CA 90095 (sjo@math.ucla.edu, jerome@math.ucla.edu). The research of these authors was supported by ONR grant N000140710810.

[§]Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027 (goldfarb@columbia.edu). This author's research was supported in part by NSF grant DMS 06-06712, ONR grant N00014-03-0514, and DOE grant GE-FG01-92ER-25126.

This principle states that one can encode $\bar{u}$ by a linear transform $f = A\bar{u} \in \mathbb{R}^m$ for some m greater than k but much smaller than n, and then recover $\bar{u}$ from f by solving (1.1). It is proved that the recovery is perfect; i.e., the solution $u_{\text{opt}} = \bar{u}$ for any $\bar{u}$ whenever k, m, n, and A satisfy certain conditions (e.g., see [13, 32, 39, 44, 80, 100, 101]). While these conditions are computationally intractable to check, it was found in [15, 16] and other work that the types of matrices A allowing a high compression ratio (i.e., $m \ll n$) include random matrices with independent and identically distributed (i.i.d.) entries and random ensembles of orthonormal transforms (e.g., matrices formed from random sets of rows of the matrices corresponding to Fourier and cosine transforms).

Recent applications of ℓ_1-minimization can be found in [51, 84, 91, 92] for compressive imaging, [61, 68, 70, 69, 96] for MRI and CT, [3, 4, 50, 54, 76, 93] for multisensor networks and distributive sensing, [63, 65, 66, 77, 87] for analog-to-information conversion, [83] for biosensing, [97] for microarray processing, and [24, 25, 98, 99] for image decomposition and computer vision tasks. ℓ_1-minimization also has applications in image inpainting and missing data recovery; see [42, 82, 101], for example. Also nonconvex quasi–ℓ_p-norm approaches for $0 \leq p < 1$ have been proposed by Chartrand [20, 21] and Chartrand and Yin [22].

Problem (1.1) can be transformed into a linear program and then solved by conventional linear programming solvers. However, such solvers are not tailored for the matrices A that are large-scale and completely dense, or are formed by rows taken from orthonormal matrices corresponding to fast transforms so that Ax and $A^\top x$ can be computed by fast transforms. This, together with the fact that f may contain noise in certain applications, makes solving the unconstrained problem

$$\min_u \mu\|u\|_1 + \frac{1}{2}\|Au - f\|_2^2 \tag{1.2}$$

more preferable than solving the constrained problem (1.1) (e.g., see [26, 27, 29, 37, 41, 49, 58, 62, 89]). Hereafter, we use $\|\cdot\| \equiv \|\cdot\|_2$ to denote the 2-norm. In section 2.1, we give a review of recent numerical methods for solving (1.2). Because (1.2) also allows the constraint $Au = f$ to be relaxed, it is used when the measurement f is contaminated by encoding errors such as noise. However, when there is no encoding error, one must assign a tiny value to μ to heavily weigh the fidelity term $\|Au - f\|^2$ in order for $Au = f$ to be nearly satisfied. Furthermore, one can show that the solution of (1.2) never equals that of (1.1) unless they both have the trivial solution $\mathbf{0}$. In this paper, we introduce a simple method based on Bregman iterative regularization [73], which we review in section 2.2, for finding a solution of problem (1.1) by solving only a small number of instances of the unconstrained problem (1.2). Our numerical algorithm, based on this iterative method, calls the fast fixed-point continuation solver FPC [55, 56] of (1.2), which involves only matrix-vector multiplications (or fast linear transforms) and componentwise shrinkages (defined in (2.4)). Using a moderate value for the penalty parameter μ, we were able to obtain a very accurate solution to the original basis pursuit problem (1.1) for a very small multiple of the cost of solving a single instance of (1.2).

Our results can also be generalized to the constrained problem

$$\min_u \{J(u) : Au = f\} \tag{1.3}$$

for other types of convex functions J (refer to section 5). Specifically, a solution of (1.3) can be obtained through a finite number of the Bregman iterations of

$$\min_u \mu J(u) + \frac{1}{2}\|Au - f\|^2. \tag{1.4}$$

In addition, in section 5.3, we also introduce a two-line algorithm (given in (5.19) and (5.20)) also involving only matrix-vector multiplication and shrinkage operators that generates a sequence $\{u^k\}$ that converges rapidly to an approximate solution of the basis pursuit problem (1.1). In fact, the numerical experiments in [34] indicate that this algorithm converges to a true solution if the parameter μ is large enough. Finally, preliminary experiments indicate that our algorithms are robust with respect to a certain amount of noise. This is also implied by our theoretical results stated in Theorems 2.1 and 5.5.

The rest of the paper is organized as follows. In section 2, we summarize the existing methods for solving the unconstrained problem (1.2) and provide some background on our Bregman iterative regularization scheme. The main Bregman iterative algorithm is described in section 3.1; its relationship to some previous work [95] is presented in section 3.2; and its convergence is analyzed in section 3.3. Numerical results are presented in section 4. Finally, we extend our results to more general classes of problems in section 5, including a description and analysis of our linearized Bregman iterative scheme, and conclude the paper in section 6.

2. Background.

2.1. Solving the unconstrained problem (1.2).

Several recent algorithms can efficiently solve (1.2) with large-scale data. The authors of GPSR [48], Figueiredo, Nowak, and Wright [49], reformulate (1.2) as a box-constrained quadratic program, to which they apply the gradient projection method with Barzilai–Borwein steps. The algorithm ℓ_1_ℓ_s [64] by Kim et al. [62] was developed for an ℓ_1-regularization problem equivalent to (1.2). The authors apply an interior-point method to a log-barrier formulation of (1.2). The main step in each interior-point iteration, which involves solving a system of linear equations, is accelerated by using a preconditioned conjugate gradient method, for which the authors developed an efficient preconditioner. In the code SPGL1 [90], Van den Berg and Friedlander apply an iterative method for solving the LASSO problem [85], which minimizes $\|Au - f\|$ subject to $\|u\|_1 \leq \sigma$, by using an increasing sequence of σ-values in their algorithm to accelerate the computation. In [71], Nesterov proposes an accelerated multistep gradient method with an error convergence rate $O(1/k^2)$. Under some conditions, the greedy approach StOMP [37] by Donoho, Tsaig, Drori, and Starck can also quickly solve (1.2).

A method widely used by many researchers to solve (1.2) or general ℓ_1-minimization problems of the form

$$\min_u \mu\|u\|_1 + H(u) \tag{2.1}$$

for convex and differentiable functions $H(\cdot)$ is an iterative procedure based on shrinkage (also called soft thresholding; see (2.4) below). This type of method was independently proposed and analyzed by Figueiredo and Nowak in [46, 72] under the expectation-minimization framework for wavelet-based deconvolution, De Mol and Defrise [31] for wavelet inversion, Bect et al.

in [5] using an auxiliary variable and the idea from Chambolle's projection method [17], Elad in [38] and Elad et al. [40] for sparse representation and other related problems, Daubechies, Defrise, and De Mol in [29] through an optimization transfer technique, Combettes and Pesquet [26] using operator-splitting, Hale, Yin, and Zhang [55] also using operator-splitting combined with a continuation technique in their code FPC [56], Darbon and Osher [27] through an implicit PDE approach, and others. In addition, related applications and algorithms can be found in Adeyemi and Davies [1] for image sparse representation, Bioucas-Dias [6] for wavelet-based image deconvolution using a Gaussian scale mixture model, Bioucas-Dias and Figueiredo for a recent "two-step" shrinkage-based algorithm [7], Blumensath and Davies [8] for solving a cardinality constrained least-squares problem, Chambolle et al. [19] for image denoising, Daubechies, Fornasier, and Loris [30] for a direct and accelerated projected gradient method, Elad, Matalon, and Zibulevsky in [41] for image denoising, Fadili and Starck [43] for sparse representation-based image deconvolution, Figueiredo and Nowak [47] for image deconvolution, Figueiredo, Bioucas-Dias, and Nowak [45] for wavelet-based image denoising using majorization-minimization algorithms, and Reeves and Kingsbury [78] for image coding.

While all of these authors used different approaches, they all developed or used algorithms based on the iterative scheme

$$u^{k+1} \leftarrow \arg\min_u \mu\|u\|_1 + \frac{1}{2\delta^k}\left\|u - (u^k - \delta^k \nabla H(u^k))\right\|^2 \tag{2.2}$$

for $k = 0, 1, \ldots,$ starting from a point u^0. The parameter δ^k is positive and serves as the step size at iteration k. Since the unknown variable u is componentwise separable in problem (2.2), each of its components u_i can be independently obtained by the shrinkage operation, which is also referred to as soft thresholding:

$$u_i^{k+1} = \text{shrink}((u^k - \delta^k \nabla H(u^k))_i, \mu\delta^k), \ i = 1, \ldots, n, \tag{2.3}$$

where for $y, \alpha \in \mathbb{R}$, we define

$$\text{shrink}(y, \alpha) := \text{sgn}(y)\max\{|y| - \alpha, 0\} = \begin{cases} y - \alpha, & y \in (\alpha, \infty), \\ 0, & y \in [-\alpha, \alpha], \\ y + \alpha, & y \in (-\infty, -\alpha). \end{cases} \tag{2.4}$$

Among the several approaches that can be used to derive (2.2), one of the simplest is the following: first, $H(u)$ is approximated by its first-order Taylor expansion at u^k, $H(u^k) + \langle \nabla H(u^k), u - u^k\rangle$. Then, since this approximation is only accurate for u close to u^k, an ℓ_2-penalty term $\|u - u^k\|^2/(2\delta^k)$ is added to the objective; the resulting step is

$$u^{k+1} \leftarrow \arg\min_u \mu\|u\|_1 + H(u^k) + \langle \nabla H(u^k), u - u^k\rangle + \frac{1}{2\delta^k}\|u - u^k\|^2, \tag{2.5}$$

which is equivalent to (2.2) because their objectives differ by only a constant. It is easy to see that the larger the δ^k, the larger the allowable distance between u^{k+1} and u^k. It was proved in [55] that $\{u^k\}$ given by (2.2) converges to an optimum of (1.4) at a q-linear[1] rate

[1] q stands for "quotient"; $\{x^k\}$ converges to x^* q-linearly if $\lim_k \|x^{k+1} - x^*\|/\|x^k - x^*\|$ exists and is less than 1.

under certain conditions on H and δ^k. Under weaker conditions, they also established r-linear convergence of $\{u^k\}$ based on previous work by Pang [74] and Luo and Tseng [67] on gradient projection methods. Furthermore, it was also proved in [55] that under mild conditions, the support and signs of u^k converge finitely; that is, there exists a finite number K such that $\{i : u^k \neq 0\} = \{i : u_{\text{opt}} \neq 0\}$ and $\text{sgn}(u^k) = \text{sgn}(u_{\text{opt}})$ for all $k > K$, where u_{opt} denotes the solution of (1.4). However, an estimate or bound for K is not known.

To improve the efficiency of the iterations (2.2), various techniques have been applied to (2.2), including generalizing (2.3) by using more parameters [41], performing line searches [49], and using a decreasing sequence of μ-values [55]. The last technique is called path following or continuation. While our algorithm does not depend on using a specific code, we chose to use FPC [56], one of the fastest codes, to solve each subproblem in (2.2).

In [27], [94], and other work, the iterative procedure (2.2) is adapted for solving the total variation regularization problem

$$\min_u \mu TV(u) + H(u), \tag{2.6}$$

where $TV(u)$ denotes the total variation of u (see [102] for a definition of $TV(u)$ and its properties). Specifically, the regularization term $\mu\|u\|_1$ in (2.2) is replaced by $\mu TV(u)$, yielding

$$u^{k+1} \leftarrow \arg\min_u \mu TV(u) + \frac{1}{2\delta^k}\left\|u - (u^k - \delta^k \nabla H(u^k))\right\|^2. \tag{2.7}$$

Each subproblem (2.7) can be efficiently solved, for example, by one of the recent graph/network-based algorithms [18, 28, 53]. In [27] Darbon and Osher also studied an algorithm obtained by replacing $\mu TV(u)$ in (2.7) by its Bregman distance (see section 2.2) and proved that if $H(u) = 0.5\|Au - f\|^2$, then $\{u^k\}$ converges to the solution of $\min_u\{TV(u) : Au = f\}$. Their algorithm and results are described in section 5.3. In the next subsection, we give an introduction to Bregman iterative regularization.

2.2. Bregman iterative regularization. Bregman iterative regularization was introduced by Osher et al. [73] in the context of image processing; it was then extended to wavelet-based denoising [95], nonlinear inverse scale space in [10, 11], and compressed sensing in MR imaging [59]. The authors of [73] extend the Rudin–Osher–Fatemi [81] model

$$\min_u \mu \int |\nabla u| + \frac{1}{2}\|u - b\|^2, \tag{2.8}$$

where u is an unknown image, b is typically an input noisy measurement of a clean image $\bar{u}$, and μ is a tuning parameter, into an iterative regularization model by using the Bregman distance (2.10) below based on the total variation functional:

$$J(u) = \mu\, TV(u) = \mu \int |\nabla u|. \tag{2.9}$$

Specifically, the Bregman distance [9] based on a convex functional $J(\cdot)$ between points u and v is defined as

$$D_J^p(u, v) = J(u) - J(v) - \langle p, u - v\rangle, \tag{2.10}$$

where $p \in \partial J(v)$ is some subgradient in the subdifferential of J at the point v. Because $D_J^p(u,v) \neq D_J^p(v,u)$ in general, $D_J^p(u,v)$ is not a distance in the usual sense. However, it measures the closeness between u and v in the sense that $D_J^p(u,v) \geq 0$ and $D_J^p(u,v) \geq D_J^p(w,v)$ for all points w on the line segment connecting u and v.

Instead of solving (2.8) once, the Bregman iterative regularization procedure of Osher et. al. [73] solves a sequence of convex problems

$$u^{k+1} \leftarrow \min_u D_J^{p^k}(u,u^k) + \frac{1}{2}\|u-b\|^2 \tag{2.11}$$

for $k = 0, 1, \ldots$, starting with $u^0 = \mathbf{0}$ and $p^0 = \mathbf{0}$ (hence, for $k = 0$, one solves the original problem (2.8)). Since $\mu TV(u)$ is not differentiable everywhere, the subdifferential of $\mu TV(u)$ may contain more than one element. However, from the optimality of u^{k+1} in (2.11), it follows that $\mathbf{0} \in \partial J(u^{k+1}) - p^k + u^{k+1} - b$; hence, they set

$$p^{k+1} := p^k + b - u^{k+1}.$$

The difference between (2.8) and (2.11) is in the use of regularization. While (2.8) regularizes u by directly minimizing its total variation, (2.11) regularizes u by minimizing the total variation-based Bregman distance of u to a previous solution u^k.

In [73] two key results for the sequence $\{u^k\}$ generated by (2.11) were proved. First, $\|u^k - b\|$ converges to 0 monotonically; second, u^k also gets closer to $\bar{u}$, the *unknown* noiseless image, monotonically in terms of the Bregman distance $D_{TV}^{p^k}(\bar{u}, u^k)$, at least while $\|u^k - b\| \geq \|\bar{u} - b\|$. Numerical results in [73] demonstrate that for μ sufficiently large, this simple iterative procedure remarkably improves denoising quality over the original model (2.8).

Interestingly, not only for the first iteration $k = 0$, but for all k, the new problem (2.11) can be reduced to the original problem (2.8) with the input $b^{k+1} := b + (b^k - u^k)$ starting with $b^0 = u^0 = \mathbf{0}$; i.e., the iterations (2.11) are equivalent to

$$u^{k+1} \leftarrow \min_u J(u) + \frac{1}{2}\|u - b^{k+1}\|^2, \quad \text{where } b^{k+1} = b + (b^k - u^k), \tag{2.12}$$

and can be carried out using any existing algorithms for (2.8).

The iterative procedure (2.12) has an intriguing interpretation: Let ω represent the noise in b, i.e., $b = \bar{u} + \omega$, and let μ be large. At $k = 0$, $b^k - u^k = \mathbf{0}$, so (2.12) decomposes the input noisy image b into $u^1 + v^1$. Since μ is large, the resulting image u^1 is oversmoothed (by total variation minimization) so it does not contain any noise. Consequently, u^1 can be considered to be a portion of the original clean image $\bar{u}$. The residual $v^1 = b - u^1 = (\bar{u} - u^1) + \omega$, hence, is the sum of the unrecovered "good" signal $(\bar{u} - u^1)$ and the "bad" noise ω. We wish to recover $(\bar{u} - u^1)$ from v^1. Intuitively, one would next consider letting v^1 be the new input for (2.8) and solving (2.8). However, Bregman iterative regularization turns out to be both better and "nonintuitive": it adds v^1 back to the original input b. The new input of (2.12) in the second iteration is

$$b + v^1 = (u^1 + v^1) + v^1 = u^1 + 2(\bar{u} - u^1) + 2\omega,$$

which, compared to the original input $b = u^1 + (\bar{u} - u^1) + \omega$, contains twice as much of both the unrecovered "good" signal $\bar{u} - u^1$ and the "bad" noise ω. What is remarkable is that the new decomposition u^2 is a better approximation of $\bar{u}$ than u^1 (for μ large enough); one explanation is that u^2 not only inherits u^1 but also captures a part of $(\bar{u} - u^1)$, the previously uncaptured "good" signal. Of course, as the convergence results indicate, u^k will eventually pick up the noise ω since $\{u^k\}$ converges to $b = \bar{u} + \omega$. However, a high quality image can be found among the sequence $\{u^k\}$: the image u^k that has $\|u^k - b\|$ closest to $\|\bar{u} - b\|$ is cleaner and has a higher contrast than the best image that could be obtained by solving (2.8) one single time, with the best μ.

Formally Bregman iterative regularization applied to the problem

$$\min_u J(u) + H(u) \tag{2.13}$$

is given as Algorithm 1 in which the Bregman distance $D_J^{p^k}(\cdot, \cdot)$ is defined by (2.10).

Algorithm 1 (Bregman iterative regularization).

Require: $J(\cdot)$, $H(\cdot)$
1: Initialize: $k = 0$, $u^0 = \mathbf{0}$, $p^0 = \mathbf{0}$.
2: **while** "not converge" **do**
3: $\quad u^{k+1} \leftarrow \arg\min_u D_J^{p^k}(u, u^k) + H(u)$
4: $\quad p^{k+1} \leftarrow p^k - \nabla H(u^{k+1}) \in \partial J(u^{k+1})$
5: $\quad k \leftarrow k + 1$
6: **end while**

We conclude this section by citing some useful convergence results from [73] that are used in section 3.3.

Assumption 1. $J(\cdot)$ is convex, $H(\cdot)$ is convex and differentiable, and the solutions u^{k+1} in step 3 of Algorithm 1 exist.

Theorem 2.1. *Under Assumption 1, the iterate sequence $\{u^k\}$ satisfies the following:*

1. *Monotonic decrease in H: $H(u^{k+1}) \leq H(u^{k+1}) + D_J^{p^k}(u^{k+1}, u^k) \leq H(u^k)$.*
2. *Convergence to the original in H with exact data: If $\tilde{u}$ minimizes $H(\cdot)$ and $J(\tilde{u}) < \infty$, then $H(u^k) \leq H(\tilde{u}) + J(\tilde{u})/k$.*
3. *Convergence to the original in D with noisy data: Let $H(\cdot) = H(\cdot; f)$ and suppose $H(\tilde{u}; f) \leq \delta^2$ and $H(\tilde{u}; g) = 0$ (f, g, $\tilde{u}$, and δ represent noisy data, noiseless data, perfect recovery, and noise level, respectively). Then $D_J^{p^{k+1}}(\tilde{u}, u^{k+1}) < D_J^{p^k}(\tilde{u}, u^k)$ as long as $H(u^{k+1}; f) > \delta^2$.*

3. Bregman iterations for basis pursuit.

3.1. Formulations.

The main purpose of this paper is to show that the Bregman iterative procedure is a simple but very efficient method for solving the basis pursuit problem (1.1), as well as a broader class of problems of the form (1.3), in both theory and practice. Below we first give the details of the algorithm, describe our motivation, and then prove that in a finite number of iterations, u^k becomes a minimizer of $\|u\|_1$ among $\{u : Au = f\}$.

We solve the constrained problem (1.1) by applying Algorithm 1 to (1.2) for $J(u) = \mu\|u\|_1$

and $H(u) = \frac{1}{2}\|Au - f\|^2$:

Version 1:

$$u^0 \leftarrow \mathbf{0},\ p^0 \leftarrow \mathbf{0}, \tag{3.1}$$

For $k = 0, 1, \dots$ do

$$u^{k+1} \leftarrow \arg\min_u D_J^{p^k}(u, u^k) + \frac{1}{2}\|Au - f\|^2, \tag{3.2}$$

$$p^{k+1} \leftarrow p^k - A^\top(Au^{k+1} - f); \tag{3.3}$$

Version 2:

$$f^0 \leftarrow \mathbf{0},\ u^0 \leftarrow \mathbf{0}, \tag{3.4}$$

For $k = 0, 1, \dots$ do

$$f^{k+1} \leftarrow f + (f^k - Au^k), \tag{3.5}$$

$$u^{k+1} \leftarrow \arg\min_u J(u) + \frac{1}{2}\left\|Au - f^{k+1}\right\|^2. \tag{3.6}$$

Given u^k and p^k in Version 1, u^{k+1} satisfies the first-order optimality condition:

$$\mathbf{0} \in \partial J(u^{k+1}) - p^k + \nabla H(u^{k+1}) = \partial J(u^{k+1}) - p^k + A^\top(Au^{k+1} - f).$$

Therefore,

$$p^{k+1} = p^k - A^\top(Au^{k+1} - f) \in \partial J(u^{k+1}); \tag{3.7}$$

hence, $D_J^{p^{k+1}}(u, u^{k+1})$ is well defined. Clearly, if $u_i^{k+1} = 0$, then one can pick any $p_i^{k+1} \in [-1, 1]$ and still have a well defined $D_J^{p^{k+1}}(u, u^{k+1})$. However, the choice of p^{k+1} in (3.7) is not only simple but also crucial for the sequence $\{u^k\}$ to converge to the minimizer u_{opt} of the constrained problem (1.1).

Theorem 3.1. *The Bregman iterative procedure Version* 1 (3.1)–(3.3) *and Version* 2 (3.4)–(3.6) *are equivalent in the sense that* (3.2) *and* (3.6) *have the same objective functions (up to a constant) for all k.*

Proof. Let u^k and $\bar{u}^k$ denote the solutions to Versions 1 and 2, respectively. The initialization (3.1) gives $D_J^{p^0}(u, u^0) = J(u)$, while (3.4) gives $f^1 = f$. Therefore, at iteration $k = 0$, (3.2) and (3.6) solve the same optimization problem,

$$\min_u J(u) + \frac{1}{2}\|Au - f\|^2.$$

We note that this problem, as well as those for all other iterations k, may have more than one solution. We do *not* assume that in this case, u^1 (Version 1) is equal to $\bar{u}^1$ (Version 2). Instead, we use the fact from [55] that $A^\top(f - Au)$ is constant for all optimal solutions u; i.e., $A^\top(f - Au^1) = A^\top(f - A\bar{u}^1)$. According to (3.3), $p^0 = \mathbf{0}$, and $f = f^1$, we have

$$p^1 = p^0 - A^\top(Au^1 - f) = A^\top(f - Au^1) = A^\top(f - A\bar{u}^1) = A^\top(f^1 - A\bar{u}^1).$$

Next, we use induction on $p^k = A^\top(f^k - A\bar{u}^k)$. Given $p^k = A^\top(f^k - A\bar{u}^k)$, we will show the following: (i) the optimization problems in (3.2) and (3.6) at iteration k are equivalent,

(ii) $A^\top(Au^{k+1}-f) = A^\top(A\bar{u}^{k+1}-f)$, and (iii) $p^{k+1} = A^\top(f^{k+1}-A\bar{u}^{k+1})$. Clearly, part (i) proves the theorem.

Part (i): From the induction assumption it follows that

$$\begin{aligned} D_J^{p^k}(u,u^k) + \frac{1}{2}\|Au-f\|^2 &= J(u) - \langle p^k, u\rangle + \frac{1}{2}\|Au-f\|^2 + C_1 \\ &= J(u) - \langle f^k - A\bar{u}^k, Au\rangle + \frac{1}{2}\|Au-f\|^2 + C_2 \\ &= J(u) + \frac{1}{2}\|Au - (f + (f^k - A\bar{u}^k))\|^2 + C_3 \\ &= J(u) + \frac{1}{2}\|Au - f^{k+1}\|^2 + C_3, \end{aligned}$$

where C_1, C_2, and C_3 stand for terms constant in u; hence, (3.2) and (3.6) have the same objective function (up to a constant).

Part (ii): $A^\top(Au^{k+1}-f) = A^\top(A\bar{u}^{k+1}-f)$ follows from part (i) and the result in [55].

Part (iii): It follows from the induction assumption, as well as (3.3), (3.5), and part (ii), that

$$\begin{aligned} p^{k+1} = p^k - A^\top(Au^{k+1}-f) &= p^k - A^\top(A\bar{u}^{k+1}-f) \\ &= A^\top(f^k - A\bar{u}^k) - A^\top(A\bar{u}^{k+1}-f) \\ &= A^\top\left(f + (f^k - A\bar{u}^k) - A\bar{u}^{k+1}\right) \\ &= A^\top(f^{k+1} - A\bar{u}^{k+1}). \quad ■ \end{aligned} \tag{3.8, 3.9}$$

Remark. When J is not strictly convex, the subproblems in Versions 1 and 2 may both have more than one solution. The above proof shows, however, that even if Versions 1 and 2 generate different intermediate solutions at a certain iteration, they remain equivalent thereafter.

Each iteration of (3.6) is an instance of (1.2), which can be solved by the code FPC [56]. Although our convergence result below holds for any strictly positive μ, we choose μ so that (1.2) is solved efficiently by FPC and the total time of the Bregman iterations is nearly optimal.

3.2. Motivation.

In [95], Xu and Osher applied Bregman iterative regularization to wavelet-based denoising. Briefly, they considered

$$\min_u \mu\|u\|_{1,1} + \frac{1}{2}\|f-u\|_{L^2}^2, \tag{3.10}$$

where $\|u\|_{1,1}$ is the Besov norm defined in [33]; if $u = \sum_j \tilde{u}_j\psi_j$ and $f = \sum_j \tilde{f}_j\psi_j$, for a wavelet basis $\{\psi_j\}$, they solved

$$\min_{\{\tilde{u}_j\}} \mu\sum_j |\tilde{u}_j| + \frac{1}{2}\sum_j |\tilde{f}_j - \tilde{u}_j|^2.$$

It was observed in [95] and elsewhere that this minimization procedure is equivalent to shrinkage; i.e., $\tilde{u}_j = \text{shrink}(\tilde{f}_j, \mu)$, for all j, where $\text{shrink}(\cdot,\cdot)$ is defined in (2.4).

What is interesting is that Bregman iterations gives

$$\tilde{u}_j^k = \begin{cases} \tilde{f}_j, & |\tilde{f}_j| > \frac{\mu}{k-1}, \\ k\tilde{f}_j - \mu \,\mathrm{sign}(\tilde{f}_j), & \frac{\mu}{k} \leq |\tilde{f}_j| \leq \frac{\mu}{k-1}, \\ 0, & |\tilde{f}_j| \leq \frac{\mu}{k}. \end{cases} \tag{3.11}$$

So soft shrinkage becomes firm shrinkage [52] with thresholds $\tau^{(k)} = \frac{\mu}{k}$ and $\tau^{(k-1)} = \frac{\mu}{(k-1)}$.

In [10, 11] the concept of nonlinear inverse scale space was introduced and analyzed, which is basically the limit of Bregman iteration as k and μ increase with $\frac{k}{\mu} \to t$. This iterative Bregman procedure then approaches hard thresholding:

$$\tilde{u}_j(t) = \begin{cases} \tilde{f}_j, & |\tilde{f}_j| > \frac{1}{t}, \\ 0, & |\tilde{f}_j| \leq \frac{1}{t}. \end{cases} \tag{3.12}$$

For Bregman iterations it takes

$$k_j = \text{ smallest integer } \geq \frac{\mu}{|\tilde{f}_j|} \tag{3.13}$$

iterations to recover $\tilde{u}_j(k) = \tilde{f}_j$ for all $k \geq k_j$. This means that spikes return in decreasing order of magnitude and sparse data comes back very quickly.

Next, we consider the trivial example of minimizing $\|u\|_1$ subject to $a^\top u = f$, where $\mathbf{0} \neq a \in \mathbb{R}^n$ and $f \in \mathbb{R}$. Obviously, the solution is $u_{\text{opt}} = (f/a_j)\mathbf{e}_j$, where $\mathbf{e}_j$ is the jth unit vector and a_j is the component of a with the largest magnitude. Without loss of generality, we suppose $a \geq \mathbf{0}$, $f > 0$, and the largest component of a is $a_1 > 0$, which is strictly larger than the rest (to avoid solution nonuniqueness); hence, $u_{\text{opt}} = (f/a_1)\mathbf{e}_1$. Let $f^k > 0$; then the solution of the Bregman iterative subproblem

$$\min_u \mu\|u\|_1 + \frac{1}{2}(a^\top u - f^k)^2$$

is given by

$$u^k = \begin{cases} \mathbf{0}, & \mu \geq f^k a_1, \\ \frac{f^k a_1 - \mu}{a_1^2}\mathbf{e}_1, & 0 < \mu < f^k a_1. \end{cases} \tag{3.14}$$

The Bregman iterations (3.6) start with $f^1 = f$. If $\mu \geq f^1 a_1$, then $u^1 = \mathbf{0}$ so $f^2 = f + (f^1 - a^\top \mathbf{0}) = 2f$; hence, as long as u^i remains $\mathbf{0}$, $f^{i+1} = (i+1)f$. Therefore, we have $u^j = \mathbf{0}$ and $f^{j+1} = (j+1)f$ for $j = 1, \ldots, J$, for

$$J = \max\{k : \mu \geq f^k a_1\} = \left\lfloor \frac{\mu}{f a_1} \right\rfloor.$$

If $\mu < f^1 a_1$, $J = 0$. In both cases, $u_1^{J+1} = ((J+1)f a_1 - \mu)/a_1^2$ so

$$f^{J+2} = f + (f^{J+1} - a^\top u^{J+1}) = (J+2)f - a_1 \frac{(J+1)f a_1 - \mu}{a_1^2} = f - \frac{\mu}{a_1}$$

and

$$u^{J+2} = \frac{f^{J+2}a_1 - \mu}{a_1^2}\mathbf{e}_1 = \frac{f}{a_1}\mathbf{e}_1;$$

i.e., $u^{J+2} = u_{\text{opt}}$. Therefore, the Bregman iterations give an exact solution in

$$\left\lfloor \frac{\mu}{f \max_i\{|a_i|\}} \right\rfloor + 2$$

steps for any problem with a one-dimensional signal f.

We believe that these simple examples help explain why our procedure works so well in compressed sensing applications.

3.3. Convergence results. In this section, we show that the Bregman iterative regularization (3.1)–(3.3) (or, equivalently, (3.4)–(3.6)) described in section 3.1 generates a sequence of solutions $\{u^k\}$ that converges to an optimum u_{opt} of the basis pursuit problem (1.1) in a finite number of steps; that is, there exists a K such that every u^k for $k > K$ is a solution of (1.1). The analytical results of this section are generalized to many other types of ℓ_1 and related minimization problems in section 5.

We divide our analysis into two theorems. The first theorem shows that if u^k satisfies the linear constraints $Au^k = f$, then it minimizes $J(\cdot) = \mu\|\cdot\|_1$; the second theorem proves that such a u^k is obtained for a finite k.

Theorem 3.2. *Suppose an iterate u^k from (3.2) satisfies $Au^k = f$; then u^k is a solution of the basis pursuit problem (1.1).*

Proof. For any u, by the nonnegativity of the Bregman distance, we have

$$J(u^k) \leq J(u) - \langle u - u^k, p^k\rangle \tag{3.15}$$
$$= J(u) - \langle u - u^k, A^\top(f^k - Au^k)\rangle \tag{3.16}$$
$$= J(u) - \langle Au - Au^k, f^k - Au^k\rangle \tag{3.17}$$
$$= J(u) - \langle Au - f, f^k - f\rangle, \tag{3.18}$$

where the first equality follows from (3.9).

Therefore, u^k satisfies $J(u^k) \leq J(u)$ for any u satisfying $Au = f$; hence, u^k is an optimal solution of the basis pursuit problem (1.1). ■

Theorem 3.3. *There exists a number $K < \infty$ such that any u^k, $k \geq K$, is a solution of the basis pursuit problem (1.1).*

Proof. Let (I_+^j, I_-^j, E^j) be a partition of the index set $\{1, 2, \dots, n\}$, and define

$$U^j := U(I_+^j, I_-^j, E^j) = \{u : u_i \geq 0,\ i \in I_+^j;\ u_i \leq 0,\ i \in I_-^j;\ u_i = 0,\ i \in E^j\}, \tag{3.19}$$

$$H^j := \min_u \left\{\frac{1}{2}\|Au - f\|^2 : u \in U^j\right\}. \tag{3.20}$$

There are a finite number of distinct partitions (I_+^j, I_-^j, E^j), and the union of all possible U^j's is $\mathcal{H}$, the entire space of u.

At iteration k, let (I_+^k, I_-^k, E^k) be defined in terms of p^k as follows:

$$I_+^k = \{i : p^k = \mu\},\ I_-^k = \{i : p^k = -\mu\},\ E^k = \{i : p^k \in (-\mu, \mu)\}. \tag{3.21}$$

In light of the definition (3.19) and the fact that $p^k \in \partial J(u^k) = \partial(\mu\|u^k\|_1)$, we have $u^k \in U^k$. To apply Theorem 2.1, we let $\tilde{u}$ satisfy $H(\tilde{u}) = \frac{1}{2}\|A\tilde{u} - f\|^2 = 0$. Using this $\tilde{u}$ in statement 2 of Theorem 2.1, we see that for each j with $H^j > 0$ there is a sufficiently large K_j such that u^k is not in U^j for $k \geq K_j$. Therefore, letting $K := \max_j\{K_j : H^j > 0\}$, we have $H(u^k) = 0$ for $k \geq K$. That is, $Au^k = f$ for $k \geq K$.

Therefore, it follows from (3.3) that $p^K = p^{K+1} = \cdots$, and then from (3.5) that $f^{K+1} = f^{K+2} = \cdots$. Because the minimizers of Bregman iterations (3.2) and (3.6) are not necessarily unique, the u^k for $k > K$ are not necessarily the same. Nevertheless, it follows from Theorem 3.2 that all u^k for $k > K$ are optimal solutions of the basis pursuit problem (1.1). ■

Both Theorems 3.2 and 3.3 can be extended to a Bregman iterative scheme in which μ takes on varying values $\{\mu^k\}$ as long as this sequence is bounded above. Suppose $J(u) = J^k(u) = \mu^k\|u\|_1$ and $p^k \in \partial J^k(u)$ at the kth Bregman iteration and $J(u) = J^{k+1}(u) = \mu^{k+1}\|u\|_1$ at iteration $k+1$; then, the subproblem (3.2) becomes

$$\text{Version 1:} \quad u^{k+1} \leftarrow \min_u J^{k+1}(u) - \frac{\mu^{k+1}}{\mu^k}\langle p^k, u\rangle + \frac{1}{2}\|Au - f\|^2,$$

where we replace the Bregman distance of J in (3.2) by that of J^{k+1} between u and u^k, so

$$p^{k+1} = \frac{\mu^{k+1}}{\mu^k}p^k - A^\top(Au^{k+1} - f) = -\mu^{k+1}\sum_{j=1}^{k+1}\frac{A^\top(Au^j - f)}{\mu^j}.$$

Using the above identity, the subproblem (3.6), equivalent to (3.2), becomes

$$\text{Version 2:} \quad u^{k+1} \leftarrow \min_u J^{k+1}(u) + \frac{1}{2}\left\|Au - f^{k+1}\right\|^2$$

for

$$f^{k+1} = f + \mu^{k+1}\sum_{j=1}^{k}\frac{f - Au^j}{\mu^j}$$

or

$$f^{k+1} = f + \frac{\mu^{k+1}}{\mu^k}(f^k - Au^k), \; u^0 = f^0 = \mathbf{0}.$$

We plan to explore varying μ to improve the efficiency of our code.

3.4. Equivalence to the augmented Lagrangian method. After we initially submitted this paper, we found that the Bregman iterative method Algorithm 1 is equivalent to the well-known augmented Lagrangian method (also known as the method of multipliers), which was introduced by Hestenes [60] and Powell [75] and was later generalized by Rockafellar [79].

To solve the constrained optimization problem

$$\min_u s(u), \text{ subject to } c_i(u) = 0, \; i = 1, \dots, m, \tag{3.22}$$

the augmented Lagrangian method minimizes the augmented Lagrangian function

$$L(u; \lambda^k, \nu) := s(u) + \sum_{i=1}^m \lambda_i^k c_i(u) + \frac{1}{2}\sum_{i=1}^m \nu_i c_i^2(u) \tag{3.23}$$

with respect to u at each iteration k, and uses the minimizer u^{k+1} to update the multipliers

$$\lambda_i^{k+1} \leftarrow \lambda_i^k + \nu_i c_i(u^{k+1}). \tag{3.24}$$

The equivalence between this method and Version 1, (3.1)–(3.3), can be seen by letting

$$\begin{aligned} s(u) &= J(u), \\ c = \begin{bmatrix} c_1 \\ \cdots \\ c_m \end{bmatrix} &= Au - f, \\ p^k &= -A^\top \lambda^k, \\ \nu_i &\equiv 1 \ \forall i. \end{aligned}$$

Then, we have

$$\begin{aligned} L(u; \lambda^k, \nu) &= J(u) + \langle \lambda^k, Au \rangle + \frac{1}{2}\|Au - f\|^2 + C_1 \\ &= J(u) - \langle p^k, u \rangle + \frac{1}{2}\|Au - f\|^2 + C_1 \\ &= \text{the objective function in (3.2)} + C_2, \end{aligned}$$

where C_1 and C_2 are constant in u, and also (3.24) yields (3.3). Therefore, whenever $u^{\mathbf{0}} = 0$ and $\lambda^0 = \mathbf{0}$, the augmented Lagrangian method is equivalent to Version 1 (3.1)–(3.3). This inspires us to study and apply techniques and results of the augmented Lagrangian method to our problem in the future. Finally, we note that Bregman iterative regularization is generally not equivalent to the augmented Lagrangian method when the constraints are not linear.

4. Numerical results.

In this section, we demonstrate the effectiveness of Algorithm 1 for solving the basis pursuit problem (1.1), where the constraints $Ax = f$ are underdetermined linear equations and f is generated from a sparse signal $\bar{u}$ that has $\|\bar{u}\|_0 \ll n$, where $\|\bar{u}\|_0$ is defined as the number of nonzeros in u.

Our numerical experiments used two types of A matrices: orthogonalized Gaussian matrices whose elements were generated from i.i.d. normal distributions $\mathcal{N}(0, 1)$ (**randn(m,n)** in MATLAB) and whose rows were orthogonalized by QR decompositions, and partial discrete cosine transform (DCT) matrices whose m rows were chosen randomly from the $n \times n$ DCT matrix. These matrices are known to be efficient for compressed sensing in the sense of allowing a good compression ratio m/n with a high probability and have been widely used by researchers in their numerical experiments. We orthogonalized the rows of A because the subproblem solver FPC tends to be more numerically stable with such A's.

In two sets of experiments, we set the number of nonzeros in each of the *original sparse signals* $\bar{u}$ equal to $0.1m$ and $0.2m$ using **round(0.1*m)** and **round(0.2*m)**, respectively, in MATLAB. Given the number of nonzeros $\|\bar{u}\|_0$, an *original sparse signal* $\bar{u} \in \mathbb{R}^n$ was generated by randomly selecting the locations of these nonzeros, and sampling each of these nonzero elements from $\mathcal{N}(0, 4)$ (**2*randn** in MATLAB). Then, f was computed as $A\bar{u}$. When $\|\bar{u}\|_0$ is small enough, we expect the basis pursuit problem (1.1), which we solved using Algorithm 1, to yield a solution $u_{\text{opt}} = \bar{u}$ from the inputs A and f.

Table 1

Experiment results using 20 *random instances for each configuration of* $(n, m, \|\bar{u}\|_0)$.

Results of Algorithm 1	
Bregman stopping tol.	$\|Au^k - f\|/\|f\| < 10^{-5}$
Subproblem solver	FPC-basic, ver 1. Stopping tol: **xtol**$= 10^{-4}$, **gtol**$= 10^{-3}$

		Orthogonalized Gaussian matrices								
		stopping itr. k			relative error $\|u^k - \bar{u}\|/\|\bar{u}\|$			time (sec.)		
		mean	std.	max	mean	std.	max	mean	std.	max
m	n	$\|\bar{u}\|_0/m = 0.1$								
256	512	2.0	0.0	2	2.16e-08	1.37e-08	4.90e-08	0.1	0.0	0.3
512	1024	2.0	0.2	3	2.42e-08	4.53e-08	1.56e-07	0.3	0.0	0.4
1024	2048	2.1	0.4	4	2.74e-07	1.21e-06	5.42e-06	1.1	0.2	2.1
2048	4096	2.2	0.9	6	3.45e-07	1.17e-06	5.11e-06	5.0	1.8	12.4
m	n	$\|\bar{u}\|_0/m = 0.2$								
256	512	2.6	2.5	13	6.11e-07	2.48e-06	1.11e-05	0.2	0.2	0.8
512	1024	2.0	0.2	3	7.48e-08	1.08e-07	4.19e-07	0.5	0.1	0.7
1024	2048	2.5	1.8	10	7.51e-07	2.26e-06	8.93e-06	2.7	1.9	10.4
2048	4096	2.2	0.4	3	7.85e-08	3.26e-07	1.46e-06	10.4	2.1	14.4
		Partial DCT matrices								
m	n	$\|\bar{u}\|_0/m = 0.1$								
512	1024	2.3	1.0	6	9.80e-07	2.99e-06	1.09e-05	0.1	0.0	0.2
2048	4096	2.4	0.7	4	3.57e-08	1.40e-07	6.30e-07	0.4	0.1	0.6
16384	32768	2.0	0.0	2	2.06e-06	2.83e-06	9.19e-06	4.7	0.4	6.3
524288	1048576	2.0	0.0	2	2.33e-07	1.05e-07	3.69e-07	205.4	1.5	208.6
m	n	$\|\bar{u}\|_0/m = 0.2$								
512	1024	2.4	1.1	7	6.16e-08	9.19e-08	4.00e-07	0.2	0.1	0.4
2048	4096	2.5	0.6	4	1.06e-06	3.16e-06	1.04e-05	0.8	0.2	1.2
16384	32768	2.0	0.0	2	1.14e-06	2.10e-06	7.98e-06	8.5	0.2	8.9
524288	1048576	2.0	0.0	2	1.91e-07	7.00e-08	2.97e-07	423.6	2.0	428.3

We used the fast MATLAB code FPC, basic version 1.0, to solve the unconstrained subproblem (1.2) at each Bregman iteration. This basic version does *not* use any line search techniques to speed up convergence. The reader may use more recent versions of FPC or other solvers of (1.2) such as GPSR [48], ℓ_1_ℓ_s [64], and SPGL1 [90] to repeat the experiments.

While the full Gaussian matrices were explicitly stored in memory, the partial DCT matrices were implicitly stored as fast transforms for which matrix-vector multiplications of the form Ax and $A^\top x$ were computed by the MATLAB commands **dct(x)** and **idct(x)**, respectively. Therefore, we were able to test partial DCT matrices of much larger sizes than Gaussian matrices. The dimensions m-by-n of these matrices are given in the first two columns of Table 1.

Our code was written in MATLAB and was run on a Linux (version 2.6.9) workstation with a 1.8GHz AMD Opteron CPU and 3GB memory. The MATLAB version was 7.1.

The computational results given in Table 1 were obtained using the stopping tolerance

$$\frac{\|Au^k - f\|}{\|f\|} < 10^{-5}, \tag{4.1}$$

which was sufficient to give a small error $\|u^k - \bar{u}\|/\|\bar{u}\| < O(10^{-6})$. The total number of Bregman iterations and the running time of Algorithm 1 heavily depend on μ. Throughout our experiments, we used

$$\mu = \frac{0.02}{\sqrt{\|\bar{u}\|_0}}, \tag{4.2}$$

because μ bounds the maximum residual $\|Au - f\|$ up to a constant factor according to the following: for any minimizer $u_{\text{opt}} \approx \bar{u}$ of the unconstrained subproblem (1.2),

$$\|Au_{\text{opt}} - f\|^2 = \|A^\top(Au_{\text{opt}} - f)\|^2 \lessapprox O(\|\bar{u}\|_0)\,\|A^\top(Au_{\text{opt}} - f)\|_\infty^2 \le O(\|\bar{u}\|_0)\mu^2. \tag{4.3}$$

The first equality in (4.3) follows from $AA^\top = I$ for all orthogonalized Gaussian and partial DCT matrices, the second approximate inequality "$\lessapprox$" is an improved estimate over the inequality $\|\cdot\| \le \sqrt{n}\|\cdot\|_\infty$ in $\mathbb{R}^n$, and the last inequality follows from the optimality of u to (1.2).

For dense Gaussian matrices A, our code was able to solve large-scale problem instances with more than 8 million nonzeros in A (e.g., $n \times m = 4096 \times 2048 = 2^{23} > 8 \times 10^6$) in 11 seconds on average over 20 random instances. For implicit partial DCT matrices A, our code was able to handle problems with matrices of dimension $2^{19} \times 2^{20}$ in less than eight minutes.

It is easy to see that the solver FPC was itself very efficient at solving the subproblem (1.2) for the assigned values of μ in (4.2). However, to yield solutions by a single call to FPC with errors as small as those produced by the Bregman iterations, one needs to use a much smaller value of μ. We tried straight FPC on the same set of test problems using values of μ that were 100 times smaller than those used in the Bregman procedure. This produced solutions with relative errors that were more than 10 times larger, while requiring longer running times. However, we *cannot* conclude that the Bregman iterative procedure accelerates FPC, since the best set of parameters for FPC to run with a tiny μ-value can be very different from those for a normal μ-value, but they are not known to us.

It is interesting that Bregman iterations yield very accurate solutions even if the subproblems are *not* solved as accurately. In other words, our approach can tolerate errors in p^k and u^k to a certain extent. To see this, notice that the stopping tolerances **xtol** (relative error between two subsequence inner iterates) and **gtol** (violation of optimality conditions) for the subproblem solver FPC are much larger (see Table 1 for their values) than the relative errors of the final Bregman solutions. The reason for this remains a subject of further study.

Finally, to compare the Bregman iterative procedure based on the solver FPC with other recent ℓ_1 algorithms such as StOMP [37], one can refer to the CPU times of FPC in the comparative study [57] and multiply these times by the average numbers of Bregman iterations.

5. Extensions.

In this section we present extensions of our results in section 3 to more general convex functionals $J(\cdot)$ and $H(\cdot)$ and describe a linearized Bregman iterative regularization scheme.

5.1. Finite convergence.

Let $J(\cdot)$ and $H(\cdot)$ denote two convex functionals defined on a Hilbert space $\mathcal{H}$. Moreover, we assume that there exists a $\tilde{u} \in \mathcal{H}$ that minimizes $H(\cdot)$ such that $J(\tilde{u}) < \infty$. Consider the minimization problem

$$\min_{u \in \mathcal{H}} J(u) + H(u). \tag{5.1}$$

Below we study the iterates $\{u^k\}$ of the Bregman iterative procedure Algorithm 1 applied to (5.1) assuming that a solution always exists in step 3.

Theorem 5.1. *Let $J(u)$ be convex and $H(u) = h(Au - f)$ for some nonnegative convex and differentiable function $h(\cdot)$ that vanishes only at $\mathbf{0}$, and assume that $J(\cdot)$ and $H(\cdot)$ satisfy the above assumptions. In* a finite number of iterations, *Algorithm* 1 *returns a solution of*

$$\min_u \{J(u) : Au = f\} \tag{5.2}$$

under the following conditions. There exists a collection $\mathcal{U} = \{U^j\}$ satisfying

1. *$\mathcal{H} = \cup_{U \in \mathcal{U}} U$;*
   define $H^j := \min_u\{H(u) : u \in U^j\}$;
2. *if $H^j = 0$, then a minimum of $\{H(u) : u \in U^j\}$ can be attained;*
3. *$\{U^j \in \mathcal{U} : H^j > 0\}$ is a finite subcollection;*
4. *there exists a rule to associate each u^k with a $U^{j_k} \ni u^k$ from $\mathcal{U}$ so that if $H^{j_k} = 0$, then $D_J^{p^k}(u, u^k) = 0 \Leftrightarrow u \in U^{j_k}$.*

Theorem 5.1 generalizes Theorems 3.2 and 3.3 in section 3; therefore, its proof is similar to those of Theorems 3.2 and 3.3, and the only differences are the following:

1. In the generalized case, $\nabla H(u^k) = A^* \nabla h(Au^k - f)$ and $p^k = A^* \sum_{i=1}^k \nabla h(Au^i - f)$.
2. Equations (3.15)–(3.18) need to be replaced by

$$\begin{aligned} J(u^k) &\le J(u) - \langle u - u^k, p^k \rangle \\ &= J(u) - \left\langle u - u^k, A^* \sum_{i=1}^k \nabla h(Au^i - f) \right\rangle \\ &= J(u) - \left\langle Au - Au^k, \sum_{i=1}^k \nabla h(Au^i - f) \right\rangle \\ &= J(u) - \left\langle Au - f, \sum_{i=1}^k \nabla h(Au^i - f) \right\rangle. \end{aligned}$$

3. The last condition in Theorem 5.1 generalizes (3.21) and (3.19).

5.2. Bregman iterations for strictly convex functions. We now consider a strictly convex $J(u) \in C^2(\Omega)$, for a compact set $\Omega \subset \mathbb{R}^n$, without the homogeneity assumption (see (5.4) below) that we previously used; we also assume that the sequence $\{u^k\}$ lies in Ω. We have a unique element

$$p(u) \in \partial J(u) \tag{5.3}$$

for all u, and, in general,

$$J(u) \not\equiv \langle u, p(u) \rangle. \tag{5.4}$$

Again we wish to solve the constrained minimization problem

$$\min_u \{J(u) : Au = f\}. \tag{5.5}$$

Our procedure will be, as before, the Bregman iterations (3.2) or (3.6).

Theorem 5.2. *If* $\|A^\top u\| \geq \delta \|u\|$, $\delta > 0$, *then* $\|Au^k - f\|$ *decays exponentially to zero with* k, *and* $w = \lim_{k\to\infty} u^k$ *solves* (5.5).

Proof. Following (3.7), we have

$$p^{k+1} - p^k + A^\top(Au^{k+1} - f) = 0. \tag{5.6}$$

By the strict convexity of $J \in C^2(\Omega)$ and the compactness of Ω, there exist $\epsilon > 0$, independent of k, and a positive definite matrix $Q_{k+\frac{1}{2}}$ with $\varepsilon I \prec Q_{k+\frac{1}{2}} \prec \frac{1}{\varepsilon} I$; i.e., both $Q_{k+\frac{1}{2}} - \varepsilon I$ and $\frac{1}{\varepsilon} I - Q_{k+\frac{1}{2}}$ are strictly positive definite, for some $\varepsilon > 0$ with

$$p^{k+1} - p^k = Q_{k+\frac{1}{2}}(u^{k+1} - u^k). \tag{5.7}$$

This leads us to

$$u^{k+1} - u^k + Q^{-1}_{k+\frac{1}{2}} A^\top (Au^{k+1} - f) = 0, \tag{5.8}$$

$$(I + AQ^{-1}_{k+\frac{1}{2}} A^\top)(Au^{k+1} - f) = (Au^k - f), \tag{5.9}$$

or

$$\begin{aligned} (Au^{k+1} - f) &= (I + AQ^{-1}_{k+\frac{1}{2}} A^\top)^{-1}(Au^k - f) \\ &= -\prod_{j=0}^{k} (I + AQ^{-1}_{j+\frac{1}{2}} A^\top)^{-1} f, \end{aligned} \tag{5.10}$$

and hence

$$\|Au^k - f\| \leq \left(\frac{1}{1 + \epsilon\delta^2}\right)^k \|f\|. \tag{5.11}$$

By the nonnegativity of the Bregman distance, we have, letting $\tilde{u}$ satisfy $A\tilde{u} = f$,

$$\begin{aligned} J(u^k) &\leq J(\tilde{u}) - \langle \tilde{u} - u^k, p^k \rangle \\ &= J(\tilde{u}) + \left\langle \tilde{u} - u^k, \sum_{j=0}^{k} A^\top(Au^j - f) \right\rangle \\ &= J(\tilde{u}) - \left\langle Au^k - f, \sum_{j=0}^{k} Au^j - f \right\rangle; \end{aligned} \tag{5.12}$$

thus, by (5.11), taking the limit as $k \to \infty$, we obtain

$$J(w) \leq J(\tilde{u}) \text{ with } Aw = f. \quad \blacksquare \tag{5.13}$$

5.3. Linearized Bregman iterations. In [27], Darbon and Osher combined the fixed-point iterations (2.2) with Bregman iterations for solving the image deblurring/deconvolution problem

$$\min_u \{TV(u) : Au = f\}. \tag{5.14}$$

Let $J(u) = \mu TV(u)$. Their Bregman iterations are

$$u^{k+1} \leftarrow D_J^{p^k}(u, u^k) + \frac{1}{2\delta}\left\| u - \left(u^k - \delta A^\top (Au^k - f)\right)\right\|^2, \quad k = 1, 2, \ldots, \tag{5.15}$$

which are different from the fixed-point iterations (2.2) by the use of regularization. While (2.2) minimizes J, (5.15) minimizes the Bregman distance $D_J^{p^k}$ based on J. On the other hand, (5.15) differs from (3.2) by replacing the fidelity term $\|Au - f\|^2/2$ by the sum of its first-order approximation at u^k and an ℓ_2-proximity term at u^k, which are the last three terms in (2.5). This sum is identical to a constant plus the last term in (5.15).

The sequence $\{p^k\}$ in (5.15) is chosen iteratively according the optimality conditions for (5.15):

$$\mathbf{0} = p^{k+1} - p^k + \frac{1}{\delta}\left(u^{k+1} - \left(u^k - \delta A^\top (Au^k - f)\right)\right), \tag{5.16}$$

so each p^{k+1} is uniquely determined from p^k, u^k, and u^{k+1} at the end of iteration k. By noticing that $p^0 = \mathbf{0}$ and $u^0 = \mathbf{0}$, we obtain from (5.16) that

$$p^{k+1} = p^k - A^\top (Au^k - f) - \frac{(u^{k+1} - u^k)}{\delta} = \cdots = \sum_{j=0}^{k} A^\top (f - Au^j) - \frac{u^{k+1}}{\delta}. \tag{5.17}$$

Therefore, $\{p^k\}$ can be computed on the fly. In addition, iterating (5.15) is very simple because it is a componentwise separable problem.

Motivated by basis pursuit, we consider the case for which $J(u) = \mu\|u\|_1$. Then, letting

$$v^k = \sum_{j=0}^{k} A^\top (f - Au^j), \tag{5.18}$$

each linearized Bregman iteration (5.15) after rearrangement yields

$$u_i^{k+1} \leftarrow \delta \,\mathrm{shrink}(v_i^k, \mu), \quad i = 1, \ldots, n, \tag{5.19}$$

$$v^{k+1} \leftarrow v^k + A^\top (f - Au^{k+1}). \tag{5.20}$$

This is an extremely fast algorithm, very simple to program, involving only matrix multiplication and scalar shrinkage.

In [34] we will discuss this method in detail, both in terms of its convergence and speed of execution. We have the following key results under the assumption that $J \in C^2$ is strictly convex over a compact set $\Omega \supset \{u^k\}$ (although $\mu TV(\cdot)$ is not strictly convex, it can be approximated by the strictly convex perturbed functional $\mu \int \sqrt{|\nabla u|^2 + \varepsilon}$ for $\varepsilon > 0$).

Theorem 5.3. *Let J be strictly convex and u_{opt} be an optimal solution of* $\min\{J(u) : Au = f\}$. *Then if $u^k \to w$, we have*

$$J(w) \leq J(u_{\text{opt}}) + \frac{1}{\delta}\langle w, u_{\text{opt}} - w\rangle, \tag{5.21}$$

and $\|Au^k - f\|$ decays exponentially in k if

$$I - \frac{\delta}{2}AA^\top \tag{5.22}$$

is strictly positive definite.

Proof. We have

$$u^{k+1} = \arg\min J(u) - J(u^k) - \langle u - u^k, p^k\rangle + \frac{1}{2\delta}\left\|u - u^k + \delta A^\top(Au^k - f)\right\|^2 \tag{5.23}$$

which, by (5.17) and (5.18), becomes

$$\begin{aligned} u^{k+1} = \arg\min\ & J(u) - J(u^k) - \langle u - u^k, v^{k-1}\rangle \\ & + \frac{1}{\delta}\langle u - u^k, u^k\rangle + \frac{1}{2\delta}\left\|u - u^k + \delta A^\top(Au^k - f)\right\|^2. \end{aligned}$$

By nonnegativity of the Bregman distance, we have

$$\begin{aligned} J(u^k) &\leq J(u_{\text{opt}}) - \langle u_{\text{opt}} - u^k, p^k\rangle \\ &= J(u_{\text{opt}}) - \langle u_{\text{opt}} - u^k, v^{k-1}\rangle + \frac{1}{\delta}\langle u_{\text{opt}} - u^k, u^k\rangle \\ &= J(u_{\text{opt}}) - \left\langle f - Au^k, \sum_{j=0}^{k-1}(f - Au^j)\right\rangle + \frac{1}{\delta}\langle u_{\text{opt}} - u^k, u^k\rangle. \end{aligned}$$

We will show that $\|f - Au^j\|$ decays exponentially with j; then the middle term in the last right-hand side above will vanish and the results follow.

We have

$$p^{k+1} - p^k + \frac{1}{\delta}(u^{k+1} - u^k) = -A^\top(Au^k - f). \tag{5.24}$$

By the strict convexity of J, there exists a symmetric positive definite operator $Q_{k+\frac{1}{2}}$ such that

$$\left(Q_{k+\frac{1}{2}} + \frac{1}{\delta}I\right)(u^{k+1} - u^k) = -A^\top(Au^k - f). \tag{5.25}$$

Then

$$u^{k+1} - u^k = -\left(Q_{k+\frac{1}{2}} + \frac{1}{\delta}I\right)^{-1} A^\top(Au^k - f) \tag{5.26}$$

and

$$Au^{k+1} - f = \left(I - A\left(Q_{k+\frac{1}{2}} + \frac{1}{\delta}I\right)^{-1} A^\top\right)(Au^k - f). \tag{5.27}$$

To have exponential decay in k of $\|Au^k - f\|$, we need the maximum eigenvalue of $(I - A(Q_{k+\frac{1}{2}} + \frac{1}{\delta}I)^{-1}A^\top)$ to be strictly less than 1 or, equivalently, the minimum and maximum eigenvalues of $A(Q_{k+\frac{1}{2}} + \frac{1}{\delta}I)^{-1}A^\top$ to be strictly positive and less than 2, respectively. The former requirement follows from the strict positive definiteness of $A(Q_{k+\frac{1}{2}} + \frac{1}{\delta}I)^{-1}A^\top$. To have the latter, we note that $A(Q_{k+\frac{1}{2}} + \frac{1}{\delta}I)^{-1}A^\top \prec \delta AA^\top$, which follows from the Sherman–Morrison–Woodbury formula; hence, it suffices to have $\delta AA^\top \prec 2I$, i.e., (5.22). ■

We comment that the rate of exponential decay in k of $\|Au^k - f\|$ depends on the value of μ even when δ satisfies the condition $I - \frac{\delta}{2}AA^\top \succ 0$. Since $Q_{k+\frac{1}{2}}$ is linear in μ (assuming u^k and u^{k+1} are fixed), we let $Q_{k+\frac{1}{2}} = \mu\overline{Q}_{k+\frac{1}{2}}$. When $\mu \gg \frac{1}{\delta}$, $(Q_{k+\frac{1}{2}} + \frac{1}{\delta}I)^{-1}$ is dominated by $\frac{1}{\mu}\overline{Q}_{k+\frac{1}{2}}^{-1}$, so the minimum eigenvalue of $A(Q_{k+\frac{1}{2}} + \frac{1}{\delta}I)^{-1}A^\top$ diminishes linearly to 0 in μ.

If we let μ be very large, then w approaches a minimizer of $\|u\|_1$ subject to $Au = f$. We also have a simple estimate from [27].

Theorem 5.4. *If $\delta A^\top A < I$, then*

$$\|Au^{k+1} - f\|^2 + \left(\frac{1}{\delta} - \|A^\top A\|\right)\|u^{k+1} - u^k\|^2 \le \|Au^k - f\|^2.$$

Proof. Since the Bregman distance used in (5.23) is nonnegative, we have

$$\|u^{k+1} - u^k + \delta A^\top(Au^k - f)\|^2 \le \|\delta A^\top(Au^k - f)\|^2, \tag{5.28}$$

$$\|u^{k+1} - u^k\|^2 + 2\delta\langle u^{k+1} - u^k, A^\top(Au^k - f)\rangle \le 0, \tag{5.29}$$

$$\|u^{k+1} - u^k\|^2 + \delta\|Au^{k+1} - f\|^2 - \delta\langle u^{k+1} - u^k, A^\top A(u^{k+1} - u^k)\rangle \le \delta\|Au^k - f\|^2, \tag{5.30}$$

or

$$\|Au^{k+1} - f\|^2 + \left(\frac{1}{\delta}I - \|A^\top A\|\right)\|u^{k+1} - u^k\|^2 \le \|Au^k - f\|^2. \tag{5.31}$$ ■

Finally we have a result which typifies the effectiveness of our linearized Bregman iteration in the presence of noisy data. Our argument below follows that of [2].

Theorem 5.5. *Let $J(\tilde{u})$ and $\|\tilde{u}\|$ be finite and $I - 2\delta AA^\top$ be strictly positive definite. Then the generalized Bregman distance*

$$\tilde{D}_J^{p^k}(\tilde{u}, u^k) = J(\tilde{u}) - J(u^k) - \langle \tilde{u} - u^k, p^k\rangle + \frac{1}{2\delta}\|\tilde{u} - u^k\|^2$$

diminishes with increasing k as long as $\|A\tilde{u} - f\| < (1 - 2\delta\|AA^\top\|)\|Au^k - f\|$.

Proof. Using (5.24) and the fact that $\langle u^{k+1} - u^k, p^{k+1} - p^k\rangle \geq 0$, we have

$$\|p^{k+1} - p^k\|^2 \leq -\langle p^{k+1} - p^k, A^\top(Au^k - f)\rangle$$

so

$$\|p^{k+1} - p^k\| \leq \|A^\top(A^k u - f)\|. \tag{5.32}$$

Also, following [27], we have for any $\tilde{u}$ for which $J(\tilde{u})$, $\|\tilde{u}\|$ are both finite

$$\begin{aligned} D_J^{p^{k+1}}(\tilde{u}, u^{k+1}) - D_J^{p^k}(\tilde{u}, u^k) + \frac{1}{2\delta}\|\tilde{u} - u^{k+1}\|^2 - \frac{1}{2\delta}\|\tilde{u} - u^k\|^2 \\ \leq \left\langle p^{k+1} - p^k + \frac{1}{\delta}(u^{k+1} - u^k),\ u^{k+1} - u^k \right\rangle \\ + \left\langle p^{k+1} - p^k + \frac{1}{\delta}(u^{k+1} - u^k), u^k - \tilde{u} \right\rangle. \end{aligned} \tag{5.33}$$

Using (5.24), the first term on the right-hand side of (5.33) equals

$$\delta\|A^\top(Au^k - f)\|^2 - \delta\langle A^\top(Au^k - f), p^{k+1} - p^k\rangle \leq 2\delta\|A^\top(Au^k - f)\|^2. \tag{5.34}$$

The second term on the right-hand side of (5.33) equals

$$\langle -A^\top(Au^k - f), u^k - \tilde{u}\rangle = -\|Au^k - f\|^2 + \langle Au^k - f, A\tilde{u} - f\rangle. \tag{5.35}$$

Adding (5.34) and (5.35) gives us

$$\begin{aligned} D_J^{p^{k+1}}(\tilde{u}, u^{k+1}) - D_J^{p^k}(\tilde{u}, u^k) + \frac{1}{2\delta}\|\tilde{u} - u^{k+1}\|^2 - \frac{1}{\delta}\|\tilde{u} - u^k\|^2 \\ \leq 2\delta\|A^\top(Au^k - f)\|^2 - \|Au^k - f\|^2 + \langle Au^k - f, A\tilde{u} - f\rangle. \end{aligned} \tag{5.36}$$

This means that this generalized Bregman distance $\tilde{D}_J^{p^k}(u, u^k)$ between $\tilde{u}$ and u^k diminishes in k as long as

$$\|A\tilde{u} - f\| < (1 - 2\delta\|AA^\top\|)\|Au^k - f\|, \tag{5.37}$$

i.e., as long as $\|Au^k - f\|$ is not too small compared with $\|A\tilde{u} - f\|$ for $\tilde{u}$, the "denoised" solution. Of course if $\tilde{u}$ is a solution of the basis pursuit problem, then this generalized Bregman distance monotonically decreases in k. ■

6. Conclusion.

For solving the basis pursuit problem (1.1), which is traditionally formulated and solved as a constrained linear program, we show that a simple Bregman iterative scheme applied to its unconstrained Lagrangian relaxation (1.2) yields an exact solution in a finite number of iterations. Using a moderate value of the relaxation penalty parameter μ, very few iterations are required for most problem instances.

Solving (1.1) via (1.2) enables one to use recently developed fast codes designed to solve (1.2) that require only matrix-vector products and thus take advantage of available fast transforms. As a result, we are able to solve to high accuracies huge compressed sensing problems on a standard PC.

Our discovery that certain types of constrained problems can be exactly solved by iteratively solving a sequence of unconstrained subproblems generated by a Bregman iterative regularization scheme is new. We extend this result in several ways. One yields even simpler iterations (5.19) and (5.20). We hope that our discovery and its extensions will lead to efficient algorithms for even broader classes of problems.

Acknowledgments. We want to thank Bin Dong (UCLA), Elaine Hale (Rice), Yu Mao (UCLA), and Yin Zhang (Rice) for their helpful discussions and comments. We would like to acknowledge the anonymous referees for making helpful suggestions.

REFERENCES

[1] T. Adeyemi and M. Davies, *Sparse representations of images using overcomplete complex wavelets*, in 2005 IEEE/SP 13th Workshop on Statistical Signal Processing, 2005, pp. 865–870.

[2] M. Bachmayr, *Iterative Total Variation Methods for Nonlinear Inverse Problems*, Master's thesis, Johannes Kepler Universität, Linz, Austria, 2007.

[3] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, *Compressive wireless sensing*, in Proceedings of the Fifth International Conference on Information Processing in Sensor Networks (IPSN), Nashville, TN, ACM, New York, 2006, pp. 134–142.

[4] D. Baron, M. Wakin, M. Duarte, S. Sarvotham, and R. Baraniuk, *Distributed Compressed Sensing*, preprint, 2005.

[5] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, *A ℓ_1-unified variational framework for image restoration*, in Proceedings of the 8th European Conference on Computer Vision, Prague, Lecture Notes in Comput. Sci. 3024, Springer-Verlag, Berlin, 2004, pp. 1–13.

[6] J. Bioucas-Dias, *Bayesian wavelet-based image deconvolution: A GEM algorithm exploiting a class of heavy-tailed priors*, IEEE Trans. Image Process., 15 (2006), pp. 937–951.

[7] J. Bioucas-Dias and M. Figueiredo, *Two-step algorithms for linear inverse problems with non-quadratic regularization*, in Proceedings of the IEEE International Conference on Image Processing (ICIP 2007), San Antonio, TX, 2007.

[8] T. Blumensath and M. Davies, *Iterative thresholding for sparse approximations*, J. Fourier Anal. Appl., to appear.

[9] L. Bregman, *The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming*, Comput. Math. Math. Phys., 7 (1967), pp. 200–217.

[10] M. Burger, G. Gilboa, S. Osher, and J. Xu, *Nonlinear inverse scale space methods*, Commun. Math. Sci., 4 (2006), pp. 179–212.

[11] M. Burger, S. Osher, J. Xu, and G. Gilboa, *Nonlinear inverse scale space methods for image restoration*, in Variational Geometric and Level Set Methods in Computer Vision, Lecture Notes in Comput. Sci. 3752, Springer-Verlag, Berlin, 2005, pp. 25–36.

[12] E. Candès and J. Romberg, *Quantitative robust uncertainty principles and optimally sparse decompositions*, Found. Comput. Math., 6 (2006), pp. 227–254.

[13] E. Candès and J. Romberg, *Sparsity and incoherence in compressive sampling*, Inverse Problems, 23 (2007), pp. 969–985.

[14] E. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

[15] E. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4203–4215.

[16] E. Candès and T. Tao, *Near optimal signal recovery from random projections: Universal encoding strategies*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5425.

[17] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.

[18] A. Chambolle, *Total Variation Minimization and a Class of Binary MRF models*, Technical report UMR CNRS 7641, Ecole Polytechnique, Palaiseau, France, 2005.

[19] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier, *Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Process., 7 (1998), pp. 319–335.

[20] R. Chartrand, *Exact reconstructions of sparse signals via nonconvex minimization*, IEEE Signal Process. Lett., 14 (2007), pp. 707–710.

[21] R. Chartrand, *Nonconvex compressed sensing and error correction*, in Proceedings of the 32nd International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2007, pp. 889–892.

[22] R. Chartrand and W. Yin, *Iteratively reweighted algorithms for compressive sensing*, Proceedings of ICASSP'08, submitted.

[23] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput., 20 (1998), pp. 33–61.

[24] T. Chen, T. Huang, W. Yin, and X. S. Zhou, *A new coarse-to-fine framework for* 3*D brain MR image registration*, in Computer Vision for Biomedical Image Applications, Lecture Notes in Comput. Sci. 3765, Springer-Verlag, Berlin, 2005, pp. 114–124.

[25] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. Huang, *Total variation models for variable lighting face recognition*, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI), 28 (2006), pp. 1519–1524.

[26] P. L. Combettes and J.-C. Pesquet, *Proximal thresholding algorithm for minimization over orthonormal bases*, SIAM J. Optim., 18 (2007), pp. 1351–1376.

[27] J. Darbon and S. Osher, *Fast Discrete Optimization for Sparse Approximations and Deconvolutions*, preprint, 2007.

[28] J. Darbon and M. Sigelle, *Image restoration with discrete constrained total variation, Part* I*: Fast and exact optimization*, J. Math. Imaging Vision, 26 (2006), pp. 261–276.

[29] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457.

[30] I. Daubechies, M. Fornasier, and I. Loris, *Accelerated Projected Gradient Method for Linear Inverse Problems with Sparsity Constraints*, http://arxiv.org/abs/0706.4297 (2007).

[31] C. De Mol and M. Defrise, *A note on wavelet-based inversion algorithms*, Contemp. Math., 313 (2002), pp. 85–96.

[32] R. A. DeVore, *Deterministic constructions of compressed sensing matrices*, J. Complexity, 23 (2007), pp. 918–925.

[33] R. A. DeVore and B. J. Lucier, *Wavelets*, in Acta Numerica 1992, Acta Numer., Cambridge University Press, Cambridge, UK, 1992, pp. 54–81.

[34] B. Dong, Y. Mao, S. Osher, and W. Yin, *Fast Linearized Bregman Iteration for Compressed Sensing and Related Problems*, working paper, 2007.

[35] D. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[36] D. Donoho and J. Tanner, *Neighborliness of randomly projected simplices in high dimensions*, Proc. Natl. Acad. Sci. USA, 102 (2005), pp. 9452–9457.

[37] D. Donoho, Y. Tsaig, I. Drori, and J.-C. Starck, *Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit*, IEEE Trans. Inform. Theory, submitted.

[38] M. Elad, *Why simple shrinkage is still relevant for redundant representations?*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5559–5569.

[39] M. Elad and A. Bruckstein, *A generalized uncertainty principle and sparse representations in pairs of bases*, IEEE Trans. Inform. Theory, 48 (2002), pp. 2558–2567.

[40] M. Elad, B. Matalon, J. Shtok, and M. Zibulevsky, *A wide-angle view at iterated shrinkage algorithms*, in Proceedings of the SPIE (Wavelet XII), San Diego, CA, 2007.

[41] M. Elad, B. Matalon, and M. Zibulevsky, *Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization*, Appl. Comput. Harmon. Anal., 23 (2007), pp. 346–367.

[42] M. Elad, J.-C. Starck, P. Querre, and D. Donoho, *Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)*, Appl. Comput. Harmon. Anal., 19 (2005), pp. 340–358.

[43] M. J. Fadili and J. L. Starck, *Sparse representation-based image deconvolution by iterative thresholding*, in Proceedings of Astronomical Data Analysis (ADA'06), Marseille, France, 2006.

[44] A. Feuer and A. Nemirovski, *On sparse representation in pairs of bases*, IEEE Trans. Inform. Theory, 49 (2003), pp. 1579–1581.

[45] M. Figueiredo, J. Bioucas-Dias, and R. Nowak, *Majorization-minimization algorithms for wavelet-based image restoration*, IEEE Trans. Image Process., 16 (2007), pp. 2980–2991.
[46] M. Figueiredo and R. Nowak, *An EM algorithm for wavelet-based image restoration*, IEEE Trans. Image Process., 12 (2003), pp. 906–916.
[47] M. Figueiredo and R. Nowak, *A bound optimization approach to wavelet-based image deconvolution*, in Proceedings of the IEEE International Conference on Image Processing (ICIP), 2005.
[48] M. Figueiredo, R. Nowak, and S. J. Wright, *GPSR: Gradient projection for sparse reconstruction*, IEEE J. Selected Topics Image Process., to appear.
[49] M. Figueiredo, R. Nowak, and S. J. Wright, *Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems*, IEEE J. Selected Topics in Signal Processing, to appear.
[50] M. Fornasier and H. Rauhut, *Recovery Algorithms for Vector Valued Data with Joint Sparsity Constraints*, preprint, 2006; available online at http://arxiv.org/abs/math/0608124.
[51] L. Gan, *Block compressed sensing of natural images*, in Proceedings of the International Conference on Digital Signal Processing (DSP), Cardiff, UK, 2007.
[52] H.-Y. Gao and A. G. Bruce, *WaveShrink with firm shrinkage*, Statist. Sinica, 7 (1997), pp. 855–874.
[53] D. Goldfarb and W. Yin, *Parametric Maximum Flow Algorithms for Fast Total Variation Minimization*, CAAM Technical report TR07-09, Rice University, Houston, TX, 2007.
[54] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, *Atoms of All Channels, Unite! Average Case Analysis of Multi-channel Sparse Recovery Using Greedy Algorithms*, http://hal.inria.fr/inria-00146660 (2007).
[55] E. Hale, W. Yin, and Y. Zhang, *A Fixed-Point Continuation Method for ℓ_1-Regularization with Application to Compressed Sensing*, CAAM Technical report TR07-07, Rice University, Houston, TX, 2007.
[56] E. Hale, W. Yin, and Y. Zhang, *FPC: A Fixed-Point Continuation Method for ℓ_1-Regularization*, http://www.caam.rice.edu/∼optimization/L1 (2007).
[57] E. Hale, W. Yin, and Y. Zhang, *A numerical study of fixed-point continuation applied to compressed sensing*, IEEE Trans. Inform. Theory, submitted.
[58] E. T. Hale, W. Yin, and Y. Zhang, *Fixed-point continuation for ℓ_1-minimization: Methodology and convergence*, SIAM J. Optim., submitted.
[59] L. He, T.-C. Chang, S. Osher, T. Fang, and P. Speier, *MR Image Reconstruction by Using the Iterative Refinement Method and Nonlinear Inverse Scale Space Methods*, UCLA CAM Report 06-35, University of California, Los Angeles, 2006.
[60] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl., 4 (1969), pp. 303–320.
[61] H. Jung, J. Ye, and E. Kim, *Improved k-t BLAST and k-t SENSE using FOCUSS*, Phys. Med. Biol., 52 (2007), pp. 3201–3226.
[62] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, *A Method for Large-Scale ℓ_1-Regularized Least Squares Problems with Applications in Signal Processing and Statistics*, http://www.stanford.edu/∼boyd/l1_ls.html (2007).
[63] S. Kirolos, J. Laska, M. Wakin, M. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R. Baraniuk, *Analog-to-information conversion via random demodulation*, in Proceedings of the IEEE Dallas Circuits and Systems Workshop (DCAS), Dallas, TX, 2006.
[64] K. Koh, S.-J. Kim, and S. Boyd, *ℓ_1_ℓ_s: A Simple MATLAB Solver for ℓ_1-Regularized Least Squares Problems*, http://www.stanford.edu/∼boyd/l1_ls/ (2007).
[65] J. Laska, S. Kirolos, M. Duarte, T. Ragheb, R. Baraniuk, and Y. Massoud, *Theory and implementation of an analog-to-information converter using random demodulation*, in Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS), New Orleans, LA, 2007.
[66] J. Laska, S. Kirolos, Y. Massoud, R. Baraniuk, A. Gilbert, M. Iwen, and M. Strauss, *Random sampling for analog-to-information conversion of wideband signals*, in Proceedings of the IEEE Dallas Circuits and Systems Workshop, Dallas, TX, 2006.
[67] Z.-Q. Luo and P. Tseng, *On the linear convergence of descent methods for convex essentially smooth minimization*, SIAM J. Control Optim., 30 (1992), pp. 408–425.
[68] M. Lustig, D. Donoho, and J. M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magnetic Resonance in Medicine, 58 (2007), pp. 1182–1195.

[69] M. Lustig, J. Santos, D. Donoho, and J. Pauly, *k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity*, in Proceedings of the 14th Annual Meeting of the International Society for Magnetic Resonance in Medicine, Seattle, WA, 2006.

[70] M. Lustig, J. Santos, J.-H. Lee, D. Donoho, and J. Pauly, *Application of Compressed Sensing for Rapid MR Imaging*, preprint, Stanford University, Stanford, CA, 2007.

[71] Y. Nesterov, *Gradient Methods for Minimizing Composite Objective Function*, CORE Discussion Paper 2007/76, http://www.optimization-online.org (2007).

[72] R. Nowak and M. Figueiredo, *Fast wavelet-based image deconvolution using the EM algorithm*, in Proceedings of the 35th Asilomar Conference on Signals, Systems, and Computers, Monterey, CA, 2001.

[73] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterated regularization method for total variation-based image restoration*, Multiscale Model. Simul., 4 (2005), pp. 460–489.

[74] J.-S. Pang, *A posteriori error bounds for the linearly-constrained variational inequality problem*, Math. Oper. Res., 12 (1987), pp. 474–484.

[75] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization, R. Fletcher, ed., Academic Press, New York, 1972, pp. 283–298.

[76] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, *Decentralized compression and predistribution via randomized gossiping*, in Proceedings of the International Conference on Information Processing in Sensor Networks (IPSN), Nashville, TN, 2006.

[77] T. Ragheb, S. Kirolos, J. Laska, A. Gilbert, M. Strauss, R. Baraniuk, and Y. Massoud, *Implementation models for analog-to-information conversion via random sampling*, in Proceedings of the Midwest Symposium on Circuits and Systems (MWSCAS), 2007.

[78] T. H. Reeves and N. G. Kingsbury, *Overcomplete image coding using iterative projection-based noise shaping*, in Proceedings of the 2002 International Conference on Image Processing, Vol. 3, 2002, pp. 24–28.

[79] R. T. Rockafellar, *A dual approach to solving nonlinear programming problems by unconstrained optimization*, Math. Programming, 5 (1973), pp. 354–373.

[80] M. Rudelson and R. Vershynin, *Geometric approach to error correcting codes and reconstruction of signals*, Int. Math. Res. Not., 64 (2005), pp. 4019–4041.

[81] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[82] I. Selesnick, R. Van Slyke, and O. Guleryuz, *Pixel recovery via ℓ_1 minimization in the wavelet domain*, in Proceedings of the 2004 International Conference on Image Processing, Vol. 3, 2004, pp. 1819–1822.

[83] M. Sheikh, O. Milenkovic, and R. Baraniuk, *Compressed Sensing DNA Microarrays*, Rice ECE Department Technical report TREE 0706, Rice University, Houston, TX, 2007.

[84] D. Takhar, J. Laska, M. Wakin, M. Duarte, D. Baron, S. Sarvotham, K. Kelly, and R. Baraniuk, *A new compressive imaging camera architecture using optical-domain compression*, in Proceedings of Computational Imaging IV at the SPIE Symposium on Electronic Imaging, San Jose, CA, 2006.

[85] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B, 58 (1996), pp. 267–288.

[86] J. Tropp, *Just relax: Convex programming methods for identifying sparse signals*, IEEE Trans. Inform. Theory, 51 (2006), pp. 1030–1051.

[87] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, *Random filters for compressive sampling and reconstruction*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Toulouse, France, 2006.

[88] Y. Tsaig and D. L. Donoho, *Extensions of compressed sensing*, Signal Processing, 86 (2006), pp. 549–571.

[89] E. Van den Berg and M. P. Friedlander, *In Pursuit of a Root*, UBC Computer Science Technical report TR-2007-16, University of British Columbia, Vancouver, 2007.

[90] E. Van den Berg and M. P. Friedlander, *SPGL1: A solver for sparse reconstruction*, http://www.cs.ubc.ca/labs/scl/spgl1 (2007).

[91] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, *An architecture for compressive imaging*, in Proceedings of the IEEE International Conference on Image Processing (ICIP), Atlanta, GA, 2006.

[92] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, *Compressive imaging for video representation and coding*, in Proceedings of the Picture Coding Symposium (PCS), Beijing, China, 2006.

[93] W. Wang, M. Garofalakis, and K. Ramchandran, *Distributed sparse random projections for refinable approximation*, in Proceedings of the International Conference on Information Processing in Sensor Networks (IPSN), Nashville, TN, 2007.

[94] Y. Wang, W. Yin, and Y. Zhang, *A Fast Fixed-Point Algorithm for Convex Total Variation Regularization*, working paper, Rice University, Houston, TX, 2007.

[95] J. Xu and S. Osher, *Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising*, IEEE Trans. Image Process., 16 (2006), pp. 534–544.

[96] J. Ye, *Compressed sensing shape estimation of star-shaped objects in Fourier imaging*, IEEE Signal Process. Lett., 14 (2007), pp. 750–753.

[97] W. Yin, T. Chen, X. S. Zhou, and A. Chakraborty, *Background correction for cDNA microarray image using the* $TV + L^1$ *model*, Bioinformatics, 21 (2005), pp. 2410–2416.

[98] W. Yin, D. Goldfarb, and S. Osher, *A comparison of three total variation-based texture extraction models*, J. Visual Commun. Image Representation, 18 (2007), pp. 240–252.

[99] W. Yin, D. Goldfarb, and S. Osher, *The total variation regularized* L^1 *model for multiscale decomposition*, Multiscale Model. Simul., 6 (2007), pp. 190–211.

[100] Y. Zhang, *A Simple Proof for Recoverability of* ℓ_1*-Minimization: Go Over or Under?*, CAAM Technical report TR05-09, Rice University, Houston, TX, 2005.

[101] Y. Zhang, *When Is Missing Data Recoverable?*, CAAM Technical report TR06-15, Rice University, Houston, TX, 2006.

[102] W. P. Ziemer, *Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation*, Grad. Texts in Math., Springer-Verlag, New York, 1989.