Fast Linearized Bregman Iteration for Compressive Sensing and Sparse Denoising

Stanley Osher* Yu Mao† Bin Dong‡ Wotao Yin§

December 11, 2008

We propose and analyze an extremely fast, efficient and simple method for solving the problem:

$$\min\{\|u\|_1 : Au = f, u \in R^n\}.$$

This method was first described in [1], with more details in [2] and rigorous theory given in [3] and [4]. The motivation was compressive sensing, which now has a vast and exciting history, which seems to have started with Candes, et.al. [5] and Donoho, [6]. See [2], [3] and [4] for a large set of references. Our method introduces an improvement called "kicking" of the very efficient method of [1], [2] and also applies it to the problem of denoising of undersampled signals. The use of Bregman iteration for denoising of images began in [7] and led to improved results for total variation based methods. Here we apply it to denoise signals, especially essentially sparse signals, which might even be undersampled.

1 Introduction

Let $A \in R^{m\times n}$, with $n > m$ and $f \in R^m$, be given. The aim of a basis pursuit problem is to find $u \in R^n$ by solving the constrained minimization problem:

$$\min_{u\in R^n}\{J(u)|Au = f\} \tag{1.1}$$

where $J(u)$ is a continuous convex function.

For basis pursuit, we take:

$$J(u) = |u|_1 = \sum_{j=1}^{n} |u_j|. \tag{1.2}$$

We assume that AA^T is invertible. Thus $Au = f$ is underdetermined and has at least one solution, $u = A^T(AA^T)^{-1}f$, which minimizes the ℓ_2 norm. We also assume that $J(u)$ is coercive, i.e., whenever $\|u\| \to \infty$, $J(u) \to \infty$. This implies that the set of all solutions of (1.1) is nonempty and convex. Finally, when $J(u)$ is strictly or strongly convex, the solution of (1.1) is unique.

Basis pursuit arises from many applications. In particular, there has been a recent burst of research in compressive sensing, which involves solving (1.1), (1.2). This was led by Candes et.al. [5], Donoho, [6], and others, see [2], [3] and [4] for extensive references. Compressive sensing guarantees, under appropriate circumstances, that the solution to (1.1), (1.2) gives the sparsest solution satisfying $Au = f$. The problem

*Department of Mathematics, UCLA, Los Angeles, CA 90095 (sjo@math.ucla.edu) This authors research was supported by ONR Grant N000140710810, a grant from the Department of Defense and NIH Grant UH54RR021813

†Department of Mathematics, UCLA, Los Angeles, CA 90095 (ymao29@math.ucla.edu) This authors research was supported by NIH Grant UH54RR021813

‡Department of Mathematics, UCLA, Los Angeles, CA 90095 (bdong@math.ucla.edu) This authors research was supported by NIH Grant UH54RR021813

§Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005 (wotao.yin@rice.edu) This authors research was supported by NSF Grant DMS- and an internal faculty research grant from the Dean of Engineering at Rice University

then becomes one of solving (1.1), (1.2) fast. Conventional linear programming solvers are not tailored for the large scale dense matrices A and the sparse solutions u that arise here. To overcome this, a linearized Bregman iterative procedure was proposed in [1] and analyzed in [2], [3] and [4]. In [2], true, nonlinear Bregman iteration was also used quite successfully for this problem.

Bregman iteration applied to (1.1), (1.2) involves solving the constrained optimization problem through solving a small number of unconstrained optimization problems:

$$\min_u \quad \mu|u|_1 + \frac{1}{2}\|Au-f\|_2^2 \tag{1.3}$$

for $\mu > 0$.

In [2] we used a method called the fast fixed point continuation solver (FPC) [8] which appears to be efficient. Other solvers of (1.3) could be used in this Bregman iterative regularization procedure.

Here we will improve and analyze a linearized Bregman iterative regularization procedure, which, in its original incarnation, [1], [2], involved only a two line code and simple operations and was already extremely fast and accurate.

In addition, we are interested in the denoising properties of Bregman iterative regularization, for signals, not images. The results for images involved the BV norm, which we may discretize for $n\times n$ pixel images as:

$$TV(u) = \sum_{i,j=1}^{n-1} ((u_{i+1,j}-u_{ij})^2+(u_{i,j+1}-u_{ij})^2)^{\frac{1}{2}}. \tag{1.4}$$

We usually regard the success of the ROF TV based model [9]

$$\min_u \quad TV(u)+\frac{\lambda}{2}\|f-u\|^2 \tag{1.5}$$

(we now drop the subscript 2 for the L_2 norm throughout the paper) as due to the fact that images have edges and in fact are almost piecewise constant (with texture added). Therefore, it is not surprising that sparse signals could be denoised using (1.3). The ROF denoising model was greatly improved in [7] and [10] with the help of Bregman iterative regularization. We will do the same thing here using Bregman iteration with (1.3) to denoise sparse signals, with the added touch of undersampling the noisy signals.

The paper is organized as follows: In section 2 we describe Bregman iterative algorithms, as well as the linearized version. We motivate these methods and describe previously obtained theoretical results. In section 3 we introduce an improvement to the linearized version, call "kicking" which greatly speeds up the method, especially for solutions u with a large dynamic range. In section 4 we present numerical results, including sparse recovery for u having large dynamic range, and the recovery of signals in large amounts of noise. In another work in progress [11] we apply these ideas to denoising very blurry and noisy signals remarkably well including sparse recovery for u. By blurry we mean situations where A is perhaps a subsampled discrete convolution matrix whose elements decay to zero with n, e.g. random rows of a discrete Gaussian.

2 Bregman and Linearized Bregman Iterative Algorithms

The Bregman distance [12], based on the convex function J, between points u and v, is defined by

$$D_J^p(u,v) = J(u) - J(v) - \langle p, u-v\rangle \tag{2.6}$$

where $p \in \partial J(v)$ is an element in the subgradient of J at the point v. In general $D_J^p(u,v) \neq D_J^p(v,u)$ and the triangle inequality is not satisfied, so $D_J^p(u,v)$ is not a distance in the usual sense. However it does measure 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. Moreover, if J is convex, $D_J^p(u,v) \geq 0$, if J is strictly convex $D_J^p(u,v) > 0$ for $u \neq v$ and if J is strongly convex, then there exists a constant $a > 0$ such that

$$D_J^p(u,v) \geq a\|u-v\|^2.$$

To solve (1.1) Bregman iteration was proposed in [2] . Given $u^0 = p^0 = 0$, we define:

$$u^{k+1} = \arg\min_{u\in R^n} \quad J(u) - J(u^k) - \langle u - u^k, p^k\rangle + \frac{1}{2}\|Au - f\|^2 \tag{2.7}$$
$$p^{k+1} = p^k - A^T(Au^{k+1} - f).$$

This can be written as

$$u^{k+1} = \arg\min_{u\in R^2} \quad D_J^{p^k}(u,u^k) + \frac{1}{2}\|Au - f\|^2 \quad .$$

It was proven in [2] that if $J(u) \in C^2(\Omega)$ and is strictly convex in Ω, then $\|Au^k - f\|$ decays exponentially whenever $u^k \in \Omega$ for all k. Furthermore, when u^k converges, its limit is a solution of (1.1). It was also proven in [2] that when $J(u) = |u|_1$, i.e. for problem (1.1) and (1.2), or when J is a convex function satisfying some additional conditions, the iteration (2.7) leads to a solution of (1.1) in finitely many steps.

As shown in [2], see also [7], [10], the Bregman iteration (2.7) can be written as:

$$f^{k+1} = f^k + f - Au^k$$
$$u^{k+1} = \arg\min_{u\in R^n} \quad J(u) + \frac{1}{2}\|Au - f^{k+1}\|^2 \tag{2.8}$$

This was referred to as "adding back the residual" in [7] . Here $f^0 = 0, u^0 = 0$. Thus the Bregman iteration uses solutions of the unconstrained problem

$$\min_{u\in R} \; J(u) + \frac{1}{2}\|Au - f\|^2 \tag{2.9}$$

as a solver in which the Bregman iteration applies this process iteratively.

Since there is generally no explicit expression for the solver of (2.7) or (2.8), we turn to iterative methods. The linearized Bregman iteration which we will analyze, improve and use here is generated by

$$u^{k+1} = \arg\min_{u\in R^n} \quad J(u) - J(u^k) - \langle u - u^k, p^k\rangle + \frac{1}{2\delta}\|u - (u^k - \delta A^T(Au^k - f))\|^2$$
$$p^{k+1} = p^k - \frac{1}{\delta}(u^{k+1} - u^k) - A^T(Au^k - f). \tag{2.10}$$

In the special case considered here, where $J(u) = \mu\|u\|_1$, then we have the two line algorithm

$$v^{k+1} = v^k - A^T(Au^k - f) \tag{2.11}$$
$$u^{k+1} = \delta \cdot \text{shrink}(v^{k+1}, \mu) \tag{2.12}$$

where v^k is an auxiliary variable

$$v^k = p^k + \frac{1}{\delta}u^k \tag{2.13}$$

and

$$\text{shrink}(x,\mu) := \begin{cases} x - \mu, & \text{if } x > \mu \\ 0, & \text{if } -\mu \le x \le \mu \\ x + \mu & \text{if } x < -\mu \end{cases}$$

is the soft thresholding algorithm [13] .

This linearized Bregman iterative algorithm was invented in [1] and used and analyzed in [2],[3] and [4]. In fact it comes from the inner-outer iteration for (2.7). In [2] it was shown that the linearized Bregman iteration (2.10) is just one step of the inner iteration for each outer iteration. Here we repeat the arguments also in [2], which begin by summing the second equation in (2.10) arriving at (using the fact that $u^0 = p^0 = 0$):

$$p^k + \frac{1}{\delta}u^k + \sum_{j=0}^{k-1} A^T(Au^j - f) = p^k + \frac{1}{\delta}u^k - v^k = 0, \text{ for } k = 1, 2, \ldots.$$

This gives us (2.12), and allows us to rewrite its first equation as:

$$u^{k+1} = \arg\min_{u\in R^n} \quad J(u) + \frac{1}{2\delta}\|u - \delta v^{k+1}\|^2 \tag{2.14}$$

i.e. we are adding back the "linearized noise", where v^{k+1} is defined in (2.11).

In [2] and [3] some interesting analysis was done for (2.10), (and some for (2.14)). This was done first for $J(u)$ continuously differentiable in (2.10) and the gradient $\partial J(u)$ satisfying

$$\|\partial J(u) - \partial J(v)\|^2 \le \beta\langle \partial J(u) - \partial J(v), u - v\rangle, \tag{2.15}$$

$\forall\, u, v \in R^n$, $\beta > 0$. In [3] it was shown that, if (2.15) is true, then both of the sequences $(u^k)_{k\in N}$ and $(p^k)_{k\in N}$ defined by (2.10) converge for $0 < \delta < \frac{2}{\|AA^T\|}$.

In [4] the authors recently give a theoretical analysis, showing that the iteration in (2.11) and (2.12) converges to the unique solution of

$$\min_{u\in R^n} \quad \mu\|u\|_1 + \frac{1}{2\delta}\|u\|^2 : Au = f \tag{2.16}$$

They also show the interesting result: let S be the set of all solutions of the Basis Pursuit problem (1.1), (1.2) and let

$$u_1 = \operatorname{argmin}_{u\in S}\|u\|^2 \tag{2.17}$$

which is unique. Denote the solution of (2.16) as u^*_μ. Then

$$\lim_{\mu\to\infty}\|u^*_\mu - u_1\| = 0. \tag{2.18}$$

In passing they show that

$$\|u^*_\mu\| \le \|u_1\| \text{ for all } \mu > 0 \tag{2.19}$$

which we will use below.

Another theoretical analysis on Linearized Bregman algorithm is given by Yin in [14], where he shows that Linearized Bregman iteration is equivalent to gradient descent applied to the dual of the problem (2.16) and uses this fact to obtain an elegant convergence proof.

This summarizes the relevant convergence analysis for our Bregman and linearized Bregman models.

Next we recall some results from [7] regarding noise and Bregman iteration.

For any sequence $\{u^k\}, \{p^k\}$ satisfying (2.7) for J continuous and convex, we have, for any $\tilde{u}$

$$D_J^{p^k}(\tilde{u}, u^k) - D_J^{p^{k-1}}(\tilde{u}, u^{k-1}) \le \langle A\tilde{u} - f, Au^{k-1} - f\rangle - \|Au^{k-1} - f\|^2. \tag{2.20}$$

Besides implying that the Bregman distance between u^k and any element $\tilde{u}$ satisfying $A\tilde{u} = f$ is monotonically decreasing, it also implies that, if $\tilde{u}$ is the "noise free" approximation to the solution of (1.1), the Bregman distance between u^k and $\tilde{u}$ diminishes as long as

$$\|Au^k - f\| > \|A\tilde{u} - f\| = \sigma, \text{ where } \sigma \text{ is some measure of the noise} \tag{2.21}$$

i.e., until we get too close to the noisy signal in the sense of (2.21). Note, in [7] we took A to be the identity, but these more general results are also proven there. This gives us a stopping criterion for our denoising algorithm.

In [7] we obtained a result for linearized Bregman iteration, following [15], which states that the Bregman distance between $\tilde{u}$ and u^k diminish as long as

$$\|A\tilde{u} - f\| < (1 - 2\delta\|AA^T\|)\, \|Au^k - f\| \tag{2.22}$$

so we need $0 < 2\delta\|AA^T\| < 1$.

In practice, we will use (2.21) as our stopping criterion.

3 Convergence

We begin with the following simple results for the linearized Bregman iteration or the equivalent algorithm (2.10).

Theorem 3.1. *If $u^k \to u^\infty$, then $Au^\infty = f$.*

Proof. Assume $Au^\infty \neq f$. Then $A^T(Au^\infty - f) \neq 0$ since A^T has full rank. This means that for some i, $(A^T(Au^k - f))_i$ converges to a nonzero value, which means that $v_i^{k+1} - v_i^k$ does as well. On the other hand $\{v^k\} = \{u^k/\delta + p^k\}$ is bounded since $\{u^k\}$ converges and $p^k \in [-\mu, \mu]$. Therefore $\{v_i^k\}$ is bounded, while $v_i^{k+1} - v_i^k$ converges to a nonzero limit, which is impossible. □

Theorem 3.2. *If $u^k \to u^\infty$ and $v^k \to v^\infty$, then u^∞ minimizes $\{J(u) + \frac{1}{2\delta}\|u\|^2 : Au = f\}$.*

Proof. Let $\tilde{J}(u) = J(u) + \frac{1}{2\delta}\|u\|^2$. then

$$\partial \tilde{J}(u) = \partial J(u) + \frac{1}{\delta} u.$$

Since $\partial J(u^k) = p^k = v^k - u^k/\delta$, we have $\partial \tilde{J}(u^k) = v^k$. Using the non-negativity of the Bregman distance we obtain

$$\begin{aligned} \tilde{J}(u^k) &\le \tilde{J}(u_{\rm opt}) - \langle u_{\rm opt} - u^k, \partial \tilde{J}(u^k) \rangle \\ &= \tilde{J}(u_{\rm opt}) - \langle u_{\rm opt} - u^k, v^k \rangle \end{aligned}$$

where $u_{\rm opt}$ minimizes (1.1) with J replaced by $\tilde{J}$, which is strictly convex.

Let $k \to \infty$, we have

$$\tilde{J}(u^\infty) \le \tilde{J}(u_{\rm opt}) - \langle u_{\rm opt} - u^\infty, v^\infty \rangle$$

Since $v^k = A^T \sum_{j=0}^{k-1} A^T(f - Au^j)$, we have $v^\infty \in \mathrm{range}(A^T)$. Since $Au_{\rm opt} = Au^\infty = f$, we have $\langle u_{\rm opt} - u^\infty, v^\infty \rangle = 0$, which implies $\tilde{J}(u^\infty) \le \tilde{J}(u_{\rm opt})$. □

Equation (2.16) (from a result in [3]) implies that u^∞ will approach a solution to (1.1), (1.2), as μ approaches ∞.

The linearized Bregman iteration has the following monotonicity property:

Theorem 3.3. *If $u^{k+1} \neq u^k$ and $0 < \delta < 2/\|AA^T\|$, then*

$$\|Au^{k+1} - f\| < \|Au^k - f\|.$$

Proof. Let

$$u^{k+1} - u^k = \Delta u^k, \quad v^{k+1} - v^k = \Delta v^k.$$

Then the shrinkage operation is such that

$$\Delta u_i^k = \delta q_i^k \Delta v_i^k \tag{3.23}$$

with

$$q_i^k \begin{cases} = 1 & \text{if } u_i^{k+1} u_i^k > 0 \\ = 0 & \text{if } u_i^{k+1} = u_i^k = 0 \\ \in (0,1] & \text{otherwise} \end{cases}$$

Let $Q^k = \mathrm{Diag}\,(q_i^k)$. Then (3.23) can be written as

$$\Delta u^k = \delta Q^k \Delta v^k = \delta Q^k A^T (f - Au^k) \tag{3.24}$$

which implies

$$Au^{k+1} - f = (I - \delta A Q^k A^T)(Au^k - f). \tag{3.25}$$

From (3.23), Q^k is diagonal with $0 \preceq Q^k \preceq I$, so $0 \preceq AQ^kA^T \preceq AA^T$. If we choose $\delta > 0$ such that $\delta AA^T \prec 2I$, then $0 \preceq \delta AQ^kA^T \prec 2I$ or $-I \prec I - \delta AQ^kA^T \preceq I$ which implies that $\|Au^k - f\|$ is not increasing. To get strict decay, we need only show that $AQ^kA^T(Au^k - f) = 0$ is impossible if $u^{k+1} \neq u^k$. Suppose $AQ^kA^T(Au^k - f) = 0$ holds, then from (3.24) we have:

$$\langle \Delta u^k, \Delta v^k \rangle = \delta \langle A^T(f - Au^k), Q^kA^T(f - Au^k) \rangle = 0.$$

By (3.23), this only happens if $u_i^{k+1} = u_i^k$ for all i, which is a contradiction. □

We are still faced with estimating how fast the residual decays. It turns out that if consecutive elements of u do not change sign, then $\|Au - f\|$ decays exponentially. By 'exponential' we mean that the ratio of the residuals of two consecutive iteration converges to a constant, this type of convergence is sometimes called linear convergence. Here we define

$$S_u = \{x \in R^n : \mathrm{sign}(x_i) = \mathrm{sign}(u_i), \forall i\} \tag{3.26}$$

(where $\mathrm{sign}(0) = 0$ and $\mathrm{sign}(a) = a/|a|$ for $a \neq 0$). Then we have the following:

Theorem 3.4. *If $u^k \in S \equiv S_{u_k}$ for $k \in (T_1, T_2)$, then u^k converges to u^*, where $u^* \in \mathrm{argmin}\{\|Au - f\|^2 : u \in S\}$ and $\|Au^k - f\|^2$ decays to $\|Au^* - f\|^2$ exponentially.*

Proof. . Since $u^k \in S$ for $k \in [T_1, T_2]$, we can define $Q \equiv Q^k$ for $T_1 \leq k \leq T_2 - 1$. From (3.23) we see that Q^k is a diagonal matrix consisting of zeros or ones, so $Q = Q^TQ$. Moreover, it is easy to see that $S = \{x | Qx = x\}$.

Following the argument in Theorem 3.3 we have:

$$u^{k+1} - u^k = \Delta u^k = \delta Q \Delta v^k = \delta QA^T(f - Au^k) \tag{3.27}$$

$$Au^{k+1} - f = [I - \delta AQA^T](Au^k - f) \tag{3.28}$$

and

$$-I \prec I - \delta AQA^T \preceq I.$$

Let $R^n = V_0 \oplus V_1$, where V_0 is the null space of AQA^T and V_1 is spanned by the eigenvectors corresponding to the nonzero eigenvalues of AQA^T. Let $Au^k - f = w^{k,0} + w^{k,1}$, where $w^{k,j} \in V_j$ for $j = 0,1$. From (3.28) we have

$$w^{k+1,0} = w^{k,0}$$
$$w^{k+1,1} = [I - \delta AQA^T]w^{k,1}$$

for $T_1 \leq k \leq T_2 - 1$. Since $w^{k,1}$ is not in the null space of AQA^T, then (3.27) and (3.28) imply that $\|w^{k,1}\|$ decays exponentially. Let $w^0 = w^{k,0}$, then $AQA^Tw^0 = 0 \;\; AQQA^Tw^0 \Rightarrow QA^Tw^0 = 0$. Therefore, from (3.27) we have

$$\Delta u^k = \delta Q^TA^T(f - Au^k) = \delta QA^T(w^0 + w^{k,1}) = \delta QA^Tw^{k,1}.$$

Thus $\|\Delta u^k\|$ decays exponentially. This means $\{u^k\}$ forms a Cauchy sequence in S, so it has a limit $u^* \in S$. Moreover

$$Au^* - f = \lim_k (Au^k - f) = \lim_k w^{k,0} + \lim_k w^{k,1} = w^0.$$

Since V_0 and V_1 are orthogonal:

$$\|Au^k - f\|^2 = \|w^{k,0}\|^2 + \|w^{k,1}\|^2 = \|Au^* - f\|^2 + \|w^{k,1}\|^2,$$

so $\|Au^k - f\|^2 - \|Au^* - f\|^2$ decays exponentially. The only thing left to show is that

$$u^* = \mathrm{argmin}(\|Au - f\|^2 : u \in S) = \mathrm{argmin}\{\|Au - f\|^2 : Qu = u\}.$$

This is equivalent to way that $A^T(Au^* - f)$ is orthogonal with the hyperspace $\{u : Qu = u\}$. It's easy to see that since Q is a projection operator, a vector v is orthogonal with $\{u : Qu = u\}$ if and only if $Qv = 0$, thus we need to show $QA^T(Au^* - f) = 0$. This is obvious because we have shown that $Au^* - f = w^0$ and $QA^Tw^0 = 0$. So we find that u^* is the desired minimizer. □

Therefore, instead of decaying exponentially with a global rate, the residual of the linearized Bregman iteration decays in a rather sophisticated manner. From the definition of the shrinkage function we can see that the sign of an element of u will change if and only if the corresponding element of v crosses the boundary of the interval $[-\mu,\mu]$. If μ is relatively large compared with the size of Δv (which is usually the case when applying the algorithm to a compressed sensing problem), then at most iterations the signs of the elements of u will stay unchanged, i.e. u will stay in the subspace S_u defined in (3.26) for a long while. This theorem tells us that under this scenario u will quickly converge to the point u^* that minimizes $\|Au-f\|$ inside S_u, and the difference between $\|Au-f\|$ and $\|Au^*-f\|$ decays exponentially. After u converges to u^*, u will stay there until the sign of some element of u changes. Usually this means that a new nonzero element of u comes up. After that, u will enter a different subspace S and a new converging procedure begins.

The phenomenon described above can be observed clearly in Fig 1. The final solution of u contains five non-zero spikes. Each time a new spike appears, it converges rapidly to the position that minimizes $\|Au-f\|$ in the subspace S_u. After that there is a long stagnation, which means u is just waiting there until the accumulating v brings out a new non-zero element of u. The larger μ is, the longer the stagnation takes. Although the convergence of the residual during each phase is fast, the total speed of the convergence suffers much from the stagnation. The solution of this problem will be described in the next section.

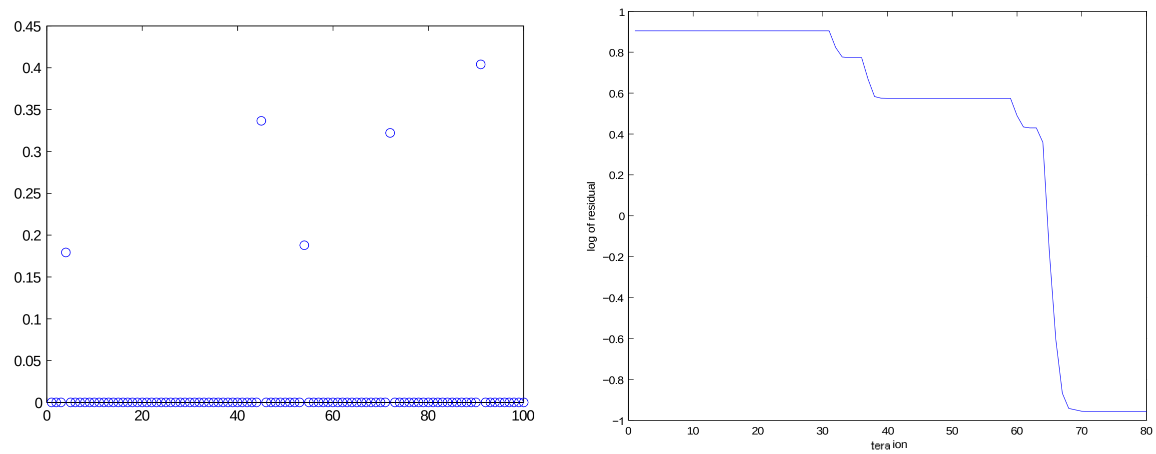

Figure 1: The left figure presents a simple signal with 5 non-zero spikes. The right figure shows how the linearized Bregman iteration converges.

4 Fast Implementation

The iterative formula in Algorithm 1 below gives us the basic linearized Bregman algorithm designed to solve (1.1),(1.2).

Algorithm 1 **Bregman Iterative Regularization**

Initialize: $u=0$, $v=0$.
while "$\|f-Au\|$ not converge" do
  $v^{k+1}=v^k+A^\top(f-Au^k)$
  $u^{k+1}=\delta\cdot\text{shrink}(v^{k+1},\mu)$
end while

This is an extremely concise algorithm, simple to program, involve only matrix multiplication and shrinkage. When A consists of rows of a matrix of a fast transform like FFT which is a common case for compressed sensing, it is even faster because matrix multiplication can be implemented efficiently using the existing fast code of the transform. Also, storage becomes a less serious issue.