Proceedings of the IASTED International Conference
Visualization, Imaging and Image Processing (VIIP 2012)
July 3 - 5, 2012 Banff, Canada

FAST DUAL-BASED LINEARIZED BREGMAN ALGORITHM FOR COMPRESSIVE SENSING OF DIGITAL IMAGES

Jie Yan
Department or Electrical and Computer Engineering
University of Victoria
Victoria, BC, Canada V8W 3P6
jyan@ece.uvic.ca

Wu-Sheng Lu
Department of Electrical and Computer Engineering
University of Victoria
Victoria, BC, Canada V8W 3P6
wslu@ece.uvic.ca

ABSTRACT
A central problem in compressive sensing is the recovery of a sparse signal using a relatively small number of linear measurements. The basis pursuit (BP) has been a successful formulation for this signal reconstruction problem. Among other things, linearized Bregman (LB) methods proposed recently are found effective to solve BP. In this paper, we present a fast linearized Bregman algorithm applied to a dual formulation that accelerates the conventional LB iterations considerably. Performance of the proposed algorithm is evaluated and compared with the conventional LB algorithm in compressive sampling of 1-D sparse signals and digital images.

KEY WORDS
Linearized Bregman, FISTA, compressive sensing

1 Introduction

A central problem in compressive sensing [1–3] (CS) is the recovery of a sparse signal from a relatively small number of linear measurements. A successful approach in the current CS theory deals with this signal reconstruction problem by means of nonsmooth convex programming (NCP). A representative formulation in the NCP setting examines the equality constrained problem

$$\min_{\mathbf{x}} \ J(\mathbf{x}) \ \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b} \tag{1}$$

where $J(\mathbf{x})$ is a continuous (but non-differentiable) objective function. In particular when $J(\mathbf{x}) = \|\mathbf{x}\|_1$, (1) becomes the well known basis pursuit problem [4]. Another representative NCP formulation is associated with the unconstrained ℓ_1-ℓ_2 problem

$$\min_{\mathbf{x}} \ \lambda\|\mathbf{x}\|_1 + \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 \tag{2}$$

where $\|\cdot\|$ denotes the ℓ_2 norm and parameter λ regularizes signal sparsity while taking signal fidelity into account.

Concerning the computational aspects of the problem, a rich variety of algorithms is now available. With $J(\mathbf{x}) = \|\mathbf{x}\|_1$, (1) can be solved by linear programming (LP) for real-valued data or by second-order cone programming (SOCP) for complex-valued data. Reliable LP and SOCP solvers are easy to find, but they are not tailored for CS problems involving large-scale data such as digital images. Gradient-based algorithms for problem (2) that are especially suited for large-scale CS problems have been developed [5]. Of particular interest are those based on proximal-point functions in conjunction with iterative shrinkage techniques. These include the fast iterative shrinkage-thresholding algorithm (FISTA) and monotone FISTA (MFISTA) [6]. For a solution of (2) to be a good approximate solution of (1), parameter λ in (2) must be sufficiently small that inevitably slows down the FISTA as a large number of iterations are required for the algorithm to converge. In [7–9], solution methods for problem (1) based on Bregman distance [10] are proposed. These methods are known as linearized Bregman (LB) algorithms that are suited for large-scale problems and shown to be able to identify global minimizer of (1) efficiently. In addition, the LB algorithm is shown to be equivalent to a gradient descent algorithm applied to a dual formulation [9].

In this paper, we propose a fast dual-based LB algorithm for problem (1) with $J(\mathbf{x}) = \|\mathbf{x}\|_1$. The algorithm's acceleration is made possible by enhancing each gradient descent iteration in a way similar to that employed in FISTA. Unlike FISTA, however, the algorithm is carried out for a dual problem, making the selection and adjustment of the regularization parameter rather straightforward. Performance and complexity of the proposed algorithm are evaluated and compared with the conventional LB algorithm [7] by applying them to CS reconstruction of 1-D sparse signals. In addition, performance of the proposed algorithm in dealing with large-scale data is demonstrated by accurately reconstructing several test images.

2 Linearized Bregman Methods

The Bregman distance [10] with respect to a convex function $J(\cdot)$ between points $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$D_J^{\mathbf{p}}(\mathbf{u}, \mathbf{v}) = J(\mathbf{u}) - J(\mathbf{v}) - \langle \mathbf{p}, \mathbf{u} - \mathbf{v} \rangle \tag{3}$$

where $\mathbf{p} \in \partial J(\mathbf{v})$, the subdifferential of J at $\mathbf{v}$. The linearized Bregman (LB) method was proposed in [7], and its convergence and optimality properties were investigated in [8] and [11]. An LB algorithm for problem (1) as

DOI: 10.2316/P.2012.782-022

presented in [7] is sketched below in Algorithm 1, where we have adopted the notation of [12] for presentation consistency.

Several important results of the LB method, which are most relevant to the new development described in this paper, are summarized as follows.

Proposition 1 ([8]). *Suppose $J(\cdot)$ is convex and continuously differentiable, and its gradient satisfies*

$$\|\nabla J(\mathbf{u}) - \nabla J(\mathbf{v})\|^2 \leq \beta\langle \nabla J(\mathbf{u}) - \nabla J(\mathbf{v}), \mathbf{u}-\mathbf{v}\rangle \quad (4)$$

for $\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^N$. Then the sequence $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ generated by Algorithm 1 with $0 < \tau < \frac{2}{\mu\|\mathbf{A}\mathbf{A}^T\|}$ converges. The limit of $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ is the unique solution of

$$\min_{\mathbf{x}} \; J(\mathbf{x}) + \frac{1}{2\mu}\|\mathbf{x}\|^2 \;\text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}. \quad (5)$$

However, note that Proposition 1 is not applicable when $J(\cdot) = \|\cdot\|_1$ because the ℓ_1-norm is not differentiable. For the ℓ_1 norm case, we have the following propositions.

Proposition 2 ([11]). *Let $J(\cdot) = \|\cdot\|_1$. Then the sequence $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ generated by Algorithm 1 with $0 < \tau < \frac{1}{\mu\|\mathbf{A}\mathbf{A}^T\|}$ converges to the unique solution of problem (5). Let $\mathcal{S}$ be the set of all solutions of problem (1) when $J(\mathbf{x}) = \|\mathbf{x}\|_1$ and define $\mathbf{x}_1$ as the unique minimum ℓ_2-norm solution among all the solutions in $\mathcal{S}$, i.e., $\mathbf{x}_1 = \text{argmin}_{\mathbf{x}\in\mathcal{S}}\|\mathbf{x}\|^2$. Denote the solution of (5) to be $\mathbf{x}_\mu$. Then $\|\mathbf{x}_\mu\| \leq \|\mathbf{x}_1\|$ for all $\mu > 0$ and $\lim_{\mu\to\infty}\|\mathbf{x}_\mu - \mathbf{x}_1\| = 0$.*

Algorithm 1 LB ([7])

1: Input: $\mathbf{x}^0 = \mathbf{p}^0 = \mathbf{0}$, $\mu > 0$ and $\tau > 0$.
2: **for** $k = 0, 1, ...K$ **do**
3: $\quad \mathbf{x}^{k+1} = \text{argmin}_{\mathbf{x}}\{D_J^{\mathbf{p}^k}(\mathbf{x}, \mathbf{x}^k) + \tau\langle \mathbf{A}^T(\mathbf{A}\mathbf{x}^k - \mathbf{b}), \mathbf{x}\rangle + \frac{1}{2\mu}\|\mathbf{x}-\mathbf{x}^k\|^2\}$;
4: $\quad \mathbf{p}^{k+1} = \mathbf{p}^k - \tau\mathbf{A}^T(\mathbf{A}\mathbf{x}^k - \mathbf{b}) - \frac{1}{\mu}(\mathbf{x}^{k+1} - \mathbf{x}^k)$;
5: **end for**

3 A Fast Dual-Based Linearized Bregman Algorithm

3.1 Lagrangian Dual of Problem (5)

Recently, the LB method is shown to be equivalent to a gradient descent algorithm applied to the Lagrangian dual of (5) [9], which assumes the form

$$\max_{\mathbf{y}}\min_{\mathbf{x}} \; J(\mathbf{x}) + \frac{1}{2\mu}\|\mathbf{x}\|^2 - \langle \mathbf{y}, \mathbf{A}\mathbf{x} - \mathbf{b}\rangle. \quad (6)$$

By defining

$$\tilde{\mathbf{x}} = \underset{\mathbf{x}}{\text{argmin}}\{J(\mathbf{x}) + \frac{1}{2\mu}\|\mathbf{x}\|^2 - \langle \mathbf{y}, \mathbf{A}\mathbf{x} - \mathbf{b}\rangle\}, \quad (7)$$

problem (6) can be expressed as

$$\min_{\mathbf{y}} \; E(\mathbf{y}) = -\{J(\tilde{\mathbf{x}}) + \frac{1}{2\mu}\|\tilde{\mathbf{x}}\|^2 - \langle \mathbf{y}, \mathbf{A}\tilde{\mathbf{x}} - \mathbf{b}\rangle\}. \quad (8)$$

It is known that $E(\mathbf{y})$ is continuously differentiable with its gradient $\nabla E(\mathbf{y}) = \mathbf{A}\tilde{\mathbf{x}} - \mathbf{b}$. If $J(\cdot) = \|\cdot\|_1$, then ∇E is Lipschitz continuous with the smallest Lipschitz constant $L = \mu\|\mathbf{A}\mathbf{A}^T\|$. Consequently, the dual problem can be solved by means of gradient-based techniques such as limited-memory BFGS, conjugate gradient, and Nesterov's methods, possibly in conjunction with efficient line search (e.g., Barzilai-Borwein) techniques.

3.2 Fast Algorithm Optimizing the Lagrangian Dual

In the rest of the paper, we focus on the ℓ_1 case, i.e., $J(\cdot)$ is assumed to be $\|\cdot\|_1$. Because $E(\mathbf{y})$ is convex with Lipschitz continuous $\nabla E(\mathbf{y})$, it follows that

$$E(\mathbf{y}) \leq E(\mathbf{y}^k) + \langle \mathbf{y} - \mathbf{y}^k, \nabla E(\mathbf{y}^k)\rangle + \frac{L}{2}\|\mathbf{y} - \mathbf{y}^k\|^2 \quad (9)$$

for any $\mathbf{y}$ and $\mathbf{y}^k$. In a steepest descent method [13], iterate $\mathbf{y}^k$ is updated to $\mathbf{y}^{k+1}$ with

$$\mathbf{y}^{k+1} = \mathbf{y}^k - \alpha_k\nabla E(\mathbf{y}^k) \quad (10)$$

where $\alpha_k > 0$ is a scalar step size. Note that iterate $\mathbf{y}^{k+1}$ may be interpreted as the solution to a quadratic problem

$$\mathbf{y}^{k+1} = \underset{\mathbf{y}}{\text{argmin}}\, H(\mathbf{y}, \mathbf{y}^k)$$

where

$$H(\mathbf{y}, \mathbf{y}^k) = E(\mathbf{y}^k) + \langle \mathbf{y} - \mathbf{y}^k, \nabla E(\mathbf{y}^k)\rangle + \frac{1}{2\alpha_k}\|\mathbf{y} - \mathbf{y}^k\|^2.$$

By comparing the equation above with (9), we see that the quadratic function $H(\mathbf{y}, \mathbf{y}^k)$ serves as a reasonable approximation of $E(\mathbf{y})$ at $\mathbf{y} = \mathbf{y}^k$ if α_k is set to $1/L$ where $L = \mu\|\mathbf{A}\mathbf{A}^T\|$. Thus, at the $(k+1)$th iteration, we compute

$$\mathbf{y}^{k+1} = \mathbf{y}^k - \frac{1}{L}\nabla E(\mathbf{y}^k) = \mathbf{y}^k - \frac{1}{L}(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}) \quad (11)$$

where $\mathbf{x}^{k+1}$ is computed by

$$\begin{aligned}\mathbf{x}^{k+1} &= \underset{\mathbf{x}}{\text{argmin}}\{\|\mathbf{x}\|_1 + \frac{1}{2\mu}\|\mathbf{x}\|^2 - \langle \mathbf{y}^k, \mathbf{A}\mathbf{x} - \mathbf{b}\rangle\} \\ &= \underset{\mathbf{x}}{\text{argmin}}\{\|\mathbf{x}\|_1 + \frac{1}{2\mu}\|\mathbf{x} - \mu\mathbf{A}^T\mathbf{y}^k\|^2\}.\end{aligned}$$

By defining $\mathcal{T}_\alpha : \mathbb{R}^N \to \mathbb{R}^N$ as the soft-shrinkage operator, i.e., $\mathcal{T}_\alpha(\mathbf{z}) = \text{sgn}(\mathbf{z}) \circ \max\{|\mathbf{z}| - \alpha, 0\}$, we have

$$\mathbf{x}^{k+1} = \mathcal{T}_\mu(\mu\mathbf{A}^T\mathbf{y}^k) = \mu\mathcal{T}_1(\mathbf{A}^T\mathbf{y}^k). \quad (12)$$

If we choose the initial iterate $\mathbf{y}^0 = \frac{1}{L}\mathbf{b}$, the implementation is described in Algorithm 2. Such iteration corresponds to the conventional gradient-descent method, and is known to possess a worst-case convergence rate of $O(1/k)$ where k refers to the number of iterations. Convergence proof of Algorithm 2 is provided in the Appendix. In addition, the equivalent between the dual-based LB method and the conventional LB in Algorithm 1 has been established in [9, 12].

In [9], Yin considered several techniques such as line search, Barzilai-Borwein step and limited memory BFGS (L-BFGS) to accelerate the classical gradient descent method. When working on this project, we have become aware of a very recent manuscript [12] that deals with the CS problem in the dual space by utilizing the acceleration technique proposed by Nesterov [14]. On the other hand, Beck and Teboulle devise a faster method called FISTA [6] where at each iteration a smartly chosen point is introduced. While both FISTA and Nesterov's method are proven to converge with the same rate, the two schemes are remarkably different both conceptually and computationally [6]. Since FISTA is a proximal subgradient algorithm, it is also simpler than Nesterov's method.

Algorithm 2 Dual-Based LB

Input: $\mu > 0$, $L = \mu\|\mathbf{A}\mathbf{A}^T\|$ and $\mathbf{y}^0 = \frac{1}{L}\mathbf{b}$.
for $k = 0, 1, ...$ **do**
  $\mathbf{x}^{k+1} = \mu\mathcal{T}_1(\mathbf{A}^T\mathbf{y}^k)$;
  $\mathbf{y}^{k+1} = \mathbf{y}^k - \frac{1}{L}(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b})$;
end for

Inspired by Beck and Teboulle [6], we propose a fast iteration scheme by carrying out FISTA type of iterations in the dual space. Specifically, we perform gradient projection with a new iterate $\mathbf{z}^{k+1}$ as

$$\mathbf{y}^{k+1} = \mathbf{z}^{k+1} - \frac{1}{L}\nabla E(\mathbf{z}^{k+1}) \tag{13}$$

where

$$\mathbf{z}^{k+1} = \mathbf{y}^k + \frac{t_k - 1}{t_{k+1}}(\mathbf{y}^k - \mathbf{y}^{k-1}). \tag{14}$$

In particular, the balancing parameter t_k has an iterative formula

$$t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2 \tag{15}$$

starting from the initial $t_0 = 0$. The main difference between (13) and (10) is that the current iteration is not employed on the previous point $\mathbf{y}^k$, but rather at the point $\mathbf{z}^{k+1}$ which uses a very specific linear combination of the preceding two points $\{\mathbf{y}^k, \mathbf{y}^{k-1}\}$. Obviously the requested additional computation for the fast algorithm is clearly marginal. However, the new iteration possesses a faster convergence speed of $O(1/k^2)$ as opposed to the conventional $O(1/k)$. The specific formula of the linear combination (14) and the computation of parameter t_k in (15) emerge from the recursive relation that has been established in [6] in the framework of fast iterative shrinkage algorithm (FISTA).

Naturally it follows from $\nabla E(\mathbf{y}) = \mathbf{A}\tilde{\mathbf{x}} - \mathbf{b}$ that we have

$$\nabla E(\mathbf{z}^{k+1}) = \mathbf{A}\mathbf{x}^{k+1} - \mathbf{b} \tag{16}$$

where, as suggested by (7), $\mathbf{x}^{k+1}$ is obtained by

$$\begin{aligned}\mathbf{x}^{k+1} &= \operatorname*{argmin}_{\mathbf{x}}\{\|\mathbf{x}\|_1 + \frac{1}{2\mu}\|\mathbf{x}\|^2 - \langle \mathbf{z}^{k+1}, \mathbf{A}\mathbf{x} - \mathbf{b}\rangle\} \\ &= \mathcal{T}_\mu(\mu\mathbf{A}^T\mathbf{z}^{k+1}) = \mu\mathcal{T}_1(\mathbf{A}^T\mathbf{z}^{k+1})\end{aligned}$$

In this way, we are ready to summarize the iteration procedures described above in Algorithm 3 as the fast dual-based linearized Bregman algorithm. The convergence complexity of our proposed algorithm is further analyzed in the next section.

Algorithm 3 Fast Dual-Based LB

Input: $\mu > 0$, $L = \mu\|\mathbf{A}\mathbf{A}^T\|$, $\mathbf{y}^{-1} = \mathbf{y}^0 = \frac{1}{L}\mathbf{b}$ and $t_0 = 0$,
for $k = 0, 1, ..., K$ **do**
  $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$;
  $\mathbf{z}^{k+1} = \mathbf{y}^k + \frac{t_k - 1}{t_{k+1}}(\mathbf{y}^k - \mathbf{y}^{k-1})$;
  $\mathbf{x}^{k+1} = \mu\mathcal{T}_1(\mathbf{A}^T\mathbf{z}^{k+1})$;
  $\mathbf{y}^{k+1} = \mathbf{z}^{k+1} - \frac{1}{L}(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b})$;
end for

3.3 Convergence of the Fast Dual-Based Linearized Bregman Algorithm

In this section, we sketch a proof to show that Algorithm 3 converges at a rate of $O(1/k^2)$. The proof is based on the fact [6] that if $\{a_k, b_k\}$ are positive sequences of reals satisfying

$$a_k - a_{k+1} \geq b_{k+1} - b_k \text{ and } a_1 + b_1 \leq c \tag{17}$$

for some $c > 0$ and $k \geq 1$, then $a_k < c$.

Let $\{\mathbf{y}^k\}$ be a sequence generated by Algorithm 3, $a_k = 2t_k^2 v_k/L$, $b_k = \|\mathbf{u}_k\|^2$, $c = \|\mathbf{y}^0 - \mathbf{y}^*\|^2$ with $v_k = E(\mathbf{y}^k) - E(\mathbf{y}^*)$ and $\mathbf{u}_k = t_k\mathbf{y}^k - (t_k - 1)\mathbf{y}^{k-1} - \mathbf{y}^*$. It can be proved that (17) is satified (readers are referred to [6] for more details). Hence $a_k < c$ which implies that

$$E(\mathbf{y}^k) - E(\mathbf{y}^*) < \frac{L(\|\mathbf{y}^0 - \mathbf{y}^*\|^2)}{2t_k^2}. \tag{18}$$

It can also be verified that the sequence t_k produced by Algorithm 3 satisfies $t_k \geq (k+1)/2$, which in conjunction with (18) shows that for $k \geq 1$

$$E(\mathbf{y}^k) - E(\mathbf{y}^*) < \frac{2L\|\mathbf{y}^0 - \mathbf{y}^*\|^2}{(k+1)^2}. \tag{19}$$

Hence $\mathbf{y}^k$ is an ϵ-optimal solution with respect to the dual function $E(\mathbf{y})$ if $k > \lceil C/\sqrt{\epsilon} - 1\rceil$ where $C = \sqrt{2L}\|\mathbf{y}^0 - \mathbf{y}^*\|^2$. As $\{\mathbf{y}^k\}$ converges to $\mathbf{y}^*$, the sequence $\{\mathbf{x}^k\}$ converges to $\mathbf{x}_\mu$, the unique minimizer of (5) with the same rate of $O(1/k^2)$.

We remark that our algorithm has the advantage over FISTA in the sense that FISTA is only limited to minimize the unconstrained ℓ_1-ℓ_2 problem [15]. That said, it takes a large number of iterations for FISTA to converge to a solution that satisfies equality constraints $\mathbf{Ax} = \mathbf{b}$. However unlike FISTA, the proposed Algorithm 3 is associated with a dual problem of (5). As a result, the method is able to efficiently deal with equality constrained CS problem with fast converging speed.

4 Performance Evaluation

4.1 Compressive Sensing of 1-D Signals

In the first set of examples, a partial DCT matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$ was used as the measurement matrix whose M rows were chosen randomly from an $N \times N$ DCT matrix with $N = 4 \times 10^3$, 2×10^4 and 5×10^4 respectively, and $M = 0.5N$. In each case, a K-sparse test signal $\mathbf{x}^* \in \mathbb{R}^N$, with $K = 0.05N$ and $0.02N$ respectively, was constructed by assigning K values that are randomly drawn from $\mathcal{U}(-1, 1)$ (i.e., `2*rand(K,1)-1`) to K randomly selected locations in an otherwise zero vector of length N. We remark that partial DCT matrix is known to be efficient for compressive sensing, and both $\mathbf{Ax}$ and $\mathbf{A}^T\mathbf{u}$ can be carried out efficiently by fast DCT or the inverse DCT. The observed data $\mathbf{b}$ was set to $\mathbf{b} = \mathbf{Ax}^*$.

Algorithm 3 was implemented and compared with the conventional LB method [7]. The measurement matrix constructed above implies that $L = \mu$ where μ was set to 10 in the simulation. The algorithms were terminated when $\|\mathbf{Ax}^k - \mathbf{b}\|/\|\mathbf{b}\| < 10^{-5}$ or the number of iterations exceeds 10^4. The performance of the algorithms was measured in terms of number of iterations (NoI) and CPU time using a PC laptop with a 2.67GHz Intel quad-core processor. The results are summarized in Tables 1 and 2, where the reconstructed signal is denoted as $\mathbf{x}_p$, which clearly indicate improved performance offered by Algorithm 3 relative to the conventional LB method.

N	M	$\|\mathbf{x}^*\|_0$	NoI	$\frac{\|\mathbf{x}_p-\mathbf{x}^*\|}{\|\mathbf{x}^*\|}$	time (s)
4000	2000		4011	1.0355e-5	10.7
20000	10000	$0.05N$	10000+	N/A	90.2+
50000	25000		10000+	N/A	238.4+
4000	1000		7096	1.1380e-5	17.3
20000	5000	$0.02N$	10000+	N/A	84.6+
50000	12500		10000+	N/A	223.1+

Table 1: Conventional LB [7]

In addition, Fig. 1 illustrates the number of iterations for Algorithm 3 to achieve a precision of $\|\mathbf{Ax}^k - \mathbf{b}\|/\|\mathbf{b}\| < 10^{-5}$ versus μ from 1 to 100 where the parameters were set to $N = 5 \times 10^4$, $M = 0.5N$, $K = 0.02N$. It is observed that the number of iterations increases approximately linearly with respect to μ. Unlike parameter λ involved in the ℓ_1-ℓ_2 unconstrained problem (2) that needs to be tuned diligently, Fig. 1 indicates that the number of iterations w.r.t. μ for a given solution accuracy is rather

N	M	$\|\mathbf{x}^*\|_0$	NoI	$\frac{\|\mathbf{x}_p-\mathbf{x}^*\|}{\|\mathbf{x}^*\|}$	time (s)
4000	2000		220	1.0247e-5	0.6
20000	10000	$0.05N$	1004	7.2738e-6	9.1
50000	25000		759	9.7291e-6	19.0
4000	1000		346	1.0635e-5	0.8
20000	5000	$0.02N$	1680	6.2055e-6	14.7
50000	12500		1202	1.0571e-5	27.4

Table 2: Fast Dual-Based LB (Algorithm 3)

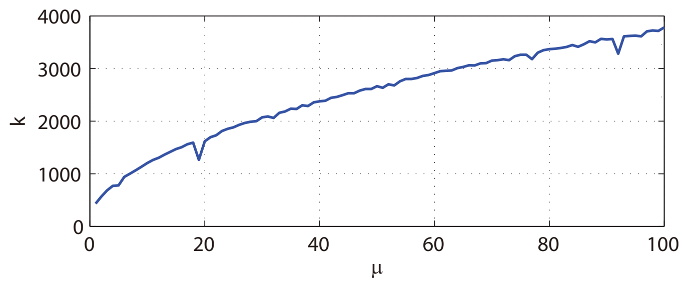

Figure 1: Number of iterations required by Algorithm 3 (with $N = 5 \times 10^4$, $M = 0.5N$, $K = 0.02N$) versus parameter μ.

predicatable. In effect, the iteration number required by Algorithm 3 for a highly accurate solution remains fairly small relative to that required by FISTA.

4.2 Compressive Sensing of a Synthetic Image

To evaluate the proposed Algorithm 3 for large-scale data, we applied it to a test image $\mathbf{X}^*$ of size 512×512 (see Fig. 2(a)) which was produced by retaining its $K = 7 \times 10^3$ largest (9-level 2-D Haar) wavelet coefficients of an original image known as "man". Thus $\mathbf{X}^*$ is sparse in the wavelet domain as 97.33% of its wavelet coefficients are zero. Image $\mathbf{X}^*$ was then normalized so that its components are in between 0 and 1.

To apply Algorithm 3, we adopted a sampling matrix to measure the wavelet coefficients of the image. The measurement matrix $\mathbf{A}$ was a partial 2-D DCT matrix of size $M \times N$ with $M = \lceil 0.2N \rceil$ and $N = 512^2$. The M rows were chosen randomly from an $N \times N$ 2-D DCT matrix. We remark that $\mathbf{A}$ needs not to be explicitly produced or stored as any matrix-vector product involving $\mathbf{A}$ can be carried out by fast 2-D DCT. Parameter μ was set to 100. Note that $\|\mathbf{AA}^T\| = 1$, hence $L = \mu$. The algorithm was terminated as soon as the relative constraint error $\|\mathbf{Ax}^k - \mathbf{b}\|/\|\mathbf{b}\|$ falls below 10^{-2}. It took the proposed fast algorithm 217 iterations (41.1 seconds) to converge. The relative reconstruction error as measured by $(\|\mathbf{X}_p - \mathbf{X}^*\|_2)/\|\mathbf{X}^*\|_2$ was found to be 0.0116, where $\mathbf{X}_p$ represents the reconstructed image and $\|\cdot\|_2$ denotes the matrix Frobenius norm. By comparison, a total of 3168 iterations (601.6 seconds) were needed for the conventional LB algorithm to produce a reconstructed image of a relative reconstruction error 0.0118. The original and the reconstructed images are illustrated in Fig. 2. The visual difference between the two is hardly noticeable.

4.3 Compressive Sensing of Natural Images

Reconstructions of several 256×256 natural images were also carried out to demonstrate efficiency of the proposed algorithm. Since natural images are sparse in the wavelet domain, in the simulations, observed data $\mathbf{b}$ were obtained by sampling (9-level 2-D Haar) wavelet coefficients of the image under a partial 2-D DCT matrix of size $M \times N$ with $N = 65536$. The number of measurement M was specified as 20000, and parameter μ was set to 100. The algorithm was terminated when $\|\mathbf{A}\mathbf{x}^k - \mathbf{b}\|/\|\mathbf{b}\| < 10^{-2}$.

The number of iterations (NoI), the relative reconstruction error, and the CPU time required for reconstruction of a number of digital images are listed in Tables 3 and 4 for the conventional LB method and the proposed fast algorithm, respectively. It can be seen that the proposed algorithm converges with number of iterations significantly less than those obtained from the conventional algorithm. As a matter of fact, it takes less than 10% of the time for the fast dual-based LB algorithm to achieve similar reconstruction performance compared with the conventional LB method.

Images	**NoI**	$\frac{\|\mathbf{X}_p - \mathbf{X}^*\|_2}{\|\mathbf{X}^*\|_2}$	**time (s)**
camera	6951	0.0988	167.4
lena	7547	0.1311	185.1
barbara	6302	0.1721	147.6
fruits	6928	0.0716	163.9
boat	6641	0.1013	152.6
circles	3016	0.0112	69.1
building	5269	0.0695	117.0
crosses	7562	0.0146	173.8
bird	8654	0.0428	225.1

Table 3: Image Reconstruction by conventional LB

5 Conclusion

A fast dual-based linearized Bregman algorithm has been proposed. Our analysis is focused on the Lagrangian dual function for which a fast iterative scheme is developed in identifying its global minimizer. This method accelerates the linearized Bregman method and shares a convergence rate of $O(1/k^2)$. Experimental results are presented to demonstrate the superiority of the proposed algorithm compared with the conventional LB method for CS recovery of images.

(a)

(b)

Figure 2: (a) Synthesized image “man” with 97.33% zero wavelet coefficients; (b) Reconstructed image “man” with 20% of DCT sampled coefficients by fast dual-based LB algorithm with 217 iterations.

Images	**NoI**	$\frac{\|\mathbf{X}_p - \mathbf{X}^*\|_2}{\|\mathbf{X}^*\|_2}$	**time (s)**
camera	515	0.1014	14.7
lena	526	0.1335	12.7
barbara	478	0.1723	11.5
fruits	510	0.0751	12.1
boat	500	0.1039	12.2
circles	249	0.0097	6.0
building	438	0.0728	10.7
crosses	401	0.0126	9.8
bird	597	0.0475	14.7

Table 4: Image Reconstruction by Fast Dual-Based LB

Appendix: Convergence Proof of Dual-Based LB Algorithm

We prove in the following that the global convergence rate of Algorithm 2 is $O(1/k)$.

Since $E(\mathbf{y})$ is convex, it follows that

$$E(\mathbf{y}) \geq E(\mathbf{y}^k) + \langle \mathbf{y} - \mathbf{y}^k, \nabla E(\mathbf{y}^k) \rangle. \tag{20}$$

The above inequality together with (9) when $\mathbf{y} = \mathbf{y}^{k+1}$ produces

$$\begin{aligned} & E(\mathbf{y}) - E(\mathbf{y}^{k+1}) \\ & \geq \langle \mathbf{y} - \mathbf{y}^{k+1}, \nabla E(\mathbf{y}^k) \rangle - \frac{L}{2}\|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 \\ & = \frac{L}{2}\|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 + L\langle \mathbf{y} - \mathbf{y}^k, \mathbf{y}^k - \mathbf{y}^{k+1} \rangle. \end{aligned} \tag{21}$$

In particular, by substituting $\mathbf{y} = \mathbf{y}^*$ (the global minimizer of $E(\mathbf{y})$) and $\mathbf{y} = \mathbf{y}^k$ in (21) respectively, the following two inequalities hold,

$$\begin{aligned} & E(\mathbf{y}^*) - E(\mathbf{y}^{k+1}) \\ & \geq \frac{L}{2}(\|\mathbf{y}^* - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y}^* - \mathbf{y}^k\|^2), \end{aligned} \tag{22}$$

and

$$E(\mathbf{y}^k) - E(\mathbf{y}^{k+1}) \geq \frac{L}{2}\|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2. \tag{23}$$

Clearly, summing inequality (22) over $k = 0, \cdots, K-1$ produces

$$KE(\mathbf{y}^*) - \sum_{k=0}^{K-1} E(\mathbf{y}^{k+1}) \geq \frac{L}{2}(\|\mathbf{y}^* - \mathbf{y}^K\|^2 - \|\mathbf{y}^* - \mathbf{y}^0\|^2). \tag{24}$$

In a similar way, we multiply (23) by k and sum over $k = 0, \cdots, K-1$, then

$$\sum_{k=0}^{K-1} k(E(\mathbf{y}^k) - E(\mathbf{y}^{k+1})) \geq \frac{L}{2}\sum_{k=0}^{K-1} k\|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2. \tag{25}$$

Since it can be shown that

$$\begin{aligned} & \sum_{k=0}^{K-1} k(E(\mathbf{y}^k) - E(\mathbf{y}^{k+1})) \\ & \quad = \sum_{k=0}^{K-1} \left(kE(\mathbf{y}^k) - (k+1)E(\mathbf{y}^{k+1}) + E(\mathbf{y}^{k+1})\right) \\ & \quad = -KE(\mathbf{y}^K) + \sum_{k=0}^{K-1} E(\mathbf{y}^{k+1}), \end{aligned}$$

then we have

$$-KE(\mathbf{y}^K) + \sum_{k=0}^{K-1} E(\mathbf{y}^{k+1}) \geq \frac{L}{2}\sum_{k=0}^{K-1} k\|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2. \tag{26}$$

The sum of (24) and (26) produces

$$KE(\mathbf{y}^*) - KE(\mathbf{y}^K) \geq -\frac{L}{2}\|\mathbf{y}^* - \mathbf{y}^0\|^2 + c$$

where $c \geq 0$. Therefore,

$$E(\mathbf{y}^k) - E(\mathbf{y}^*) \leq \frac{L\|\mathbf{y}^0 - \mathbf{y}^*\|^2}{2k} \tag{27}$$

At this point, we have shown that Algorithm 2 shares a convergence rate of $O(1/k)$. That is, $\mathbf{y}^k$ is an ϵ-optimal solution if $k \geq \lceil C/\epsilon \rceil$ with $C = L\|\mathbf{y}^0 - \mathbf{y}^*\|^2/2$. When $\{\mathbf{y}^k\}$ converges to $\mathbf{y}^*$, with the same rate the sequence $\{\mathbf{x}^k\}$ converges to $\mathbf{x}_\mu$, the unique minimizer of (5). In addition, we observe that the sequence of function values $\{E(\mathbf{y}^k)\}$ produced by Algorithm 2 is non-increasing, as shown by (23). Further, if we define the Lagrangian function for (5) as

$$\mathcal{L}_\mu(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}\|_1 + \frac{1}{2\mu}\|\mathbf{x}\|^2 - \langle \mathbf{y}, \mathbf{A}\mathbf{x} - \mathbf{b} \rangle, \tag{28}$$

then based on Algorithm 2, we have

$$E(\mathbf{y}^k) = -\mathcal{L}_\mu(\mathbf{x}^{k+1}, \mathbf{y}^k), \tag{29a}$$

$$E(\mathbf{y}^*) = -\mathcal{L}_\mu(\mathbf{x}_\mu, \mathbf{y}^*). \tag{29b}$$

Hence,

$$\mathcal{L}_\mu(\mathbf{x}_\mu, \mathbf{y}^*) - \mathcal{L}_\mu(\mathbf{x}^{k+1}, \mathbf{y}^k) \leq \frac{L\|\mathbf{y}^0 - \mathbf{y}^*\|^2}{2k}. \tag{30}$$

Thus, $(\mathbf{x}^{k+1}, \mathbf{y}^k)$ is an ϵ-optimal solution to problem (5) with respect to the Lagrangian function if $k \geq \lceil C/\epsilon \rceil$ with $C = L\|\mathbf{y}^0 - \mathbf{y}^*\|^2/2$.

Acknowledgement

The authors are grateful to the Natural Sciences and Engineering Research Council of Canada (NSERC) for supporting this work.

References

[1] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Info. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[2] D. Donoho, “Compressed sensing,” *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] E. Candès and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?” *IEEE Trans. Info. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[4] S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[5] M. Zibulevsky and M. Elad, “L1-L2 optimization in signal and image processing,” *Signal Processing Magazine, IEEE*, vol. 27, no. 3, pp. 76–88, 2010.

[6] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[7] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, “Bregman iterative algorithms for l1-minimization with applications to compressed sensing,” *SIAM J. Imaging Sci*, vol. 1, no. 1, pp. 143–168, 2008.

[8] J. Cai, S. Osher, and Z. Shen, “Linearized Bregman iterations for compressed sensing,” *Math. Comp*, vol. 78, no. 267, pp. 1515–1536, 2009.

[9] W. Yin, “Analysis and generalizations of the linearized Bregman method,” *SIAM Journal on Imaging Sciences*, vol. 3, no. 4, pp. 856–877, 2010.

[10] L. Bregman, “The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming,” *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 3, pp. 200–217, 1967.

[11] J. Cai, S. Osher, and Z. Shen, “Convergence of the linearized Bregman iteration for l1-norm minimization,” *Mathematics of Computation*, vol. 78, no. 268, pp. 2127–2136, 2009.

[12] B. Huang, S. Ma, and D. Goldfarb, “Accelerated linearized Bregman method,” *Arxiv preprint arXiv:1106.5413*, 2011.

[13] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. Springer-Verlag New York Inc, 2007.

[14] Y. Nesterov, “A method for solving the convex programming problem with convergence rate $O(1/k^2)$,” in *Soviet Mathematics Doklady*, vol. 27, no. 2, 1983, pp. 372–376.

[15] J. Yan and W.-S. Lu, “Power-iterative strategy for ℓ_p-ℓ_2 optimization for compressive sensing: towards global solution,” in *Asilomar 2011*, Pacific Grove, CA, Nov. 2011.