Noname manuscript No.
(will be inserted by the editor)

Control Subgradient Algorithm for Image ℓ_1 Regularization

Abdelkrim El Mouatasim · Mohammed Wakrim

Received: date / Accepted: date

Abstract We consider the class of subgradient methods for solving minimization of a non-smooth convex function regularized by the discretized ℓ_1 norm models arising in image processing. This class of methods is attractive due to its simplicity, however, they are also known to converge quite slowly.
In this paper, we present a Control Subgradient Algorithm (**CSA**) which preserves the computational simplicity of subgradient methods, but with a convergence rate which is proven to be significantly better, both theoretically and practically. Initial promising numerical results for wavelet-based image deblurring demonstrate the capabilities of CSA.

Keywords Subgradient algorithm · Nesterov algorithm · step-size control · ℓ_1 regularization · image deblurring

1 Introduction

1.1 Problem formulation

The inverse problems model, if supposed linear, is the one that, given an original image $\mathbf{x} \in \mathbb{R}^{\mathbf{P}}$ which we are interested in from observation $\mathbf{y} \in \mathbb{R}^{\mathbf{Q}}$, it evaluates the model response, by applying an operator $\mathbf{A} \in \mathbb{R}^{\mathbf{P}\times\mathbf{Q}}$ to $\mathbf{x}$. When $\mathbf{A}$ is linear, we call these problems linear inverse problems.

The challenge in many linear inverse problems is that they are ill-posed, i.e., either the linear operator does not admit inverse or it is near singular, yielding highly noise sensitive solutions. To cope with the ill-posed nature of these problems, a large number of techniques has been developed, most of them under the ℓ_1 regularization [20] and [22].

Recently, as an approach to linear inverse problems, ℓ_1 based regularization has attracted considerable amount of attention both in the signal and image processing literature [5], including medical imaging, astronomical imaging [27], microscopy [21], remote sensing [23], compressed sensing ([12] and [30]), and, more generally, image restoration [17] and [1].

The standard model in image deblurring assumes that the noisy blurred observed version $\mathbf{y}$, of an original image $\mathbf{x}$, was obtained via

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \beta,$$

where $\mathbf{A}$ is the matrix representation of a convolution and β is the white Gaussian noise with standard deviation σ.
As is common, we adopt the vector notation for images, where the pixels on an $\mathbf{P} \times \mathbf{Q}$ image are stacked into $\mathbf{x}$ an $(\mathbf{PQ})$-vector in, e.g., lexicographic order. In the sequel, we denote by $\mathbf{p} = \mathbf{PQ}$ the number of elements of $\mathbf{x}$, thus $\mathbf{x} \in \mathbb{R}^{\mathbf{p}}$, while $\mathbf{y} \in \mathbb{R}^{\mathbf{q}}$ ($\mathbf{p}$ and $\mathbf{q}$ may be different).
In the particular case of image deconvolution, $\mathbf{A}$ is the matrix representation of a convolution operator. This type of model describes well several physical mechanisms, such as relative motion between the camera

Abdelkrim El Mouatasim
Université Ibn Zohr, Faculté Polydisplinaire, B.P. 284, Ouarzazate 45800, Morocco.
Tel.: +212 671 311 143
Fax: +212 524 885 801
E-mail: a.elmouatasim@uiz.ac.ma

Mohammed Wakrim
Université Ibn Zohr, Facultés des Sciences, Département de Mathématiques, Agadir, Morocco.

and the subject (motion blur), bad focusing (defocusing blur), or a number of other mechanisms [5].
In more general image reconstruction problems, $\mathbf{A}$ represents some linear direct operator, such as tomographic projections (Radon transform), a partially observed (e.g., Fourier) transform, or the loss of part of the image pixels.

1.2 Contributions

The main contribution is a derivation of an accelerated subgradient method that has convergence rate $1/\mathbf{k}^2$. As far as we know, this is a novel kind of result for subgradient methods.

First, we introduce the notation

(i) $\mathbf{x}^t$ the transpose of $\mathbf{x} = (x_1, x_2, ..., x_p) \in \mathbf{E} = \mathbb{R}^p$;
(ii) $||\mathbf{x}||_2 = \sqrt{\mathbf{x}^t\mathbf{x}} = \sqrt{(x_1^2 + ... + x_p^2)}$ the Euclidean norm of $\mathbf{x}$;
(iii) $||.||_e$ stands for the ℓ_e norm (for $e \geq 1$), defined as $||\mathbf{x}||_e = (\sum_i |x_i|^e)^{1/e}$.

Since we are assuming, from the beginning, that images are defined on discrete domains, the definition of total variational TV given by :

$$TV(\mathbf{x}) = \sum_i \sqrt{(\Delta_i^h\mathbf{x})^2 + (\Delta_i^v\mathbf{x})^2},$$

where Δ_i^h and Δ_i^v are linear operators corresponding to horizontal and vertical first order differences, at pixel i, respectively.
At this point, we would like to mention that quite often the ℓ_1 regularizer,

$$\ell_1 - TV(\mathbf{x}) = \sum_i |(\Delta_i^h\mathbf{x})| + |(\Delta_i^v\mathbf{x})|,$$

has been used to approximate $TV(\mathbf{x})$, or even wrongly considered itself as the TV regularizer.
We emphasize here that, our method can by applied to $TV(\mathbf{x})$. However, we will treat only the case,

$$\ell_1(\mathbf{x}) = ||\mathbf{x}||_1,$$

in detail because the treatment for the TV case is not completely analogous [4], [25] and [28].

In image deblurring for example, $\mathbf{A}$ is often chosen as $\mathbf{A} = \mathbf{RW}$ where $\mathbf{R}$ is the blurring matrix and $\mathbf{W}$ contains a wavelet basis. The underlying philosophy here in dealing with the ℓ_1 norm regularization criterion is that most images have a sparse representation in the wavelet domain.

One regularization method that attracted a revived interest and considerable amount of attention in the signal processing literature is ℓ_1 regularization in which one seeks to find the solution of

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{Ax} - \mathbf{y}\|_2^2 + \tau\|\mathbf{x}\|_1 \tag{1}$$

The regularization parameter $\tau > 0$ acts as a trade-off between the number of non zero components of the wavelet representation of the signal and the synthesis error [11].
τ can be tuned by any of the usual rules for parameter estimation, such as cross validation [24], Bayesian method [3] or automatic [10] and L-curve [5,8].

The objective function is convex, although not strictly so. Nevertheless, its minimization represents a significant numerical optimization challenge, owing to the nondifferentiability of $\ell_1(\mathbf{x})$. In the next section, we introduce a new optimization algorithm, fully developed on the discrete domain, which is simple and yet computationally efficient.

The presence of the ℓ_1 term is used to induce sparsity in the optimal solution, of (1), see for instance [9].

Arguably, the most popular methods to solve problem (1) is in the class of iterative shrinkage/thresholding algorithms (ISTA) (for instance [2] as well as [7] and [14]), these methods apply to nonsmooth problems, but they are proximal gradient methods and never explicitly use a subgradient.

1.3 Summary of the paper

In this paper, we propose a new method with two step iterative. It can be used to accelerate subgradient methods for solving an unconstrained convex wavelet based image ℓ_1 regularization problems.

We have implemented the proposed algorithm in MATLAB and compared it with a SpaRSA implementation of the algorithm described in [29], which is the fastest algorithm for solving ℓ_1 regularization problem. Numerical results presented in section 3 show that our algorithm CSA is much faster for image regularization, especially when the blurring kernel is relatively large. Compared with a few other deblurring algorithm which solve different models, including SpaRSA our algorithm consistently generates higher quality images in comparable running times.

2 Control subgradient method

2.1 Subgradient method

We consider the following basic problem of convex non-smooth optimization

$$\min_{\mathbf{x}\in\mathbf{E}} \mathbf{F}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) + \mathbf{h}(\mathbf{x}), \tag{2}$$

where:

$\mathbf{E}$ is a finite-dimensional Euclidean space with the inner product $\mathbf{x}^t\mathbf{y}$ and corresponding norm $\|\mathbf{x}\| = \sqrt{\mathbf{x}^t\mathbf{x}}$,
$\mathbf{f}(\mathbf{x}) = \frac{1}{2}||\mathbf{A}\mathbf{x} - \mathbf{y}||_2^2$,
$\mathbf{h}(\mathbf{x}) = \tau\|\mathbf{x}\|_1$.

It is assumed that solution $\mathbf{x}^*$ of (2) exists. The main intention is to introduce a fast subgradient algorithm, prove it convergence and present some promising results of numerical experiments.

Typically computational algorithms for (2) rely on the use of subgradient oracles which provide at arbitrary point $\mathbf{x}$ the value of objective function $\mathbf{F}(\mathbf{x})$ and some subgradient $\mathbf{g}$ from subdifferental set $\partial\mathbf{F}(\mathbf{x})$.
Recall that a subgradient of $\mathbf{F}$ at $\mathbf{x}$ is any vector $\mathbf{g}$ that satisfies the inequality

$$\mathbf{F}(\mathbf{y}) \geq \mathbf{F}(\mathbf{x}) + \mathbf{g}^t(\mathbf{y} - \mathbf{x}) \forall \mathbf{y}$$

.

The simplest algorithms for (2) are subgradient methods of the kind

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \eta_k\mathbf{g}_k, \quad \mathbf{g}^k \in \partial\mathbf{F}(\mathbf{x}^k), \qquad k = 0, 1, ... \tag{3}$$

which were under intensive study since 1960's. It was shown that (3) converges under very mild conditions for step-size η_k satisfying "divergence series" condition $\sum_k \eta_k = \infty, \quad \eta_k \to +0$.

We define the function **sgn** as follow

$$\mathbf{sgn}(\mathbf{x}) = \begin{cases} 1 \text{ if } & \mathbf{x} > 0 \\ 0 \text{ if } & \mathbf{x} = 0 \\ -1 \text{ if } & \mathbf{x} < 0 \end{cases}$$

Lemma 1 *:* $\mathbf{sgn}(\mathbf{x}) \in \partial\|\mathbf{x}\|_1$.

Proof :

$$\text{Let } \xi_1(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad \mathbf{x} > 0 \\ 1 & \text{if} \quad \mathbf{x} = 0 \\ -1 & \text{if} \quad \mathbf{x} < 0 \end{cases}$$

$$\text{and } \xi_2(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad \mathbf{x} > 0 \\ -1 & \text{if} \quad \mathbf{x} = 0 \\ -1 & \text{if} \quad \mathbf{x} < 0 \end{cases}$$

we have $\xi_1 \in \partial\|\mathbf{x}\|_1$ and $\xi_2 \in \partial\|\mathbf{x}\|_1$ then for all $\mathbf{y}$

$\|\mathbf{y}\|_1 \geq \|\mathbf{x}\|_1 + \xi_1^t(\mathbf{y} - \mathbf{x})$ and $\|\mathbf{y}\|_1 \geq \|\mathbf{x}\|_1 + \xi_2^t(\mathbf{y} - \mathbf{x})$

by adding two inequality in above, we obtain

$$\|\mathbf{y}\|_1 \geq \|\mathbf{x}\|_1 + \frac{(\xi_1 + \xi_2)^T}{2}(\mathbf{y} - \mathbf{x})$$

and $\mathbf{sgn}(\mathbf{x}) = \frac{(\xi_1+\xi_2)}{2}(\mathbf{x})$ then the Lemma is hold.

In this paper we use

$$\mathbf{g}_k = \nabla\mathbf{f}(\mathbf{x}^k) + \tau\mathbf{sgn}(\mathbf{x}^k).$$

However numerical experiments and theoretical analysis demonstrated that this step-size rule results as a rule in slow convergence and further development went along the lines of Quasi-Newton ideas [18], bundle method [6] and variable metric method [13], etc. On the other hand subgradient algorithms still demonstrated quite satisfactory performance under special conditions, for example when $\mathbf{F}^* = \mathbf{F}(\mathbf{x}^*)$ is known or well-estimated so special step-size control rules can be engaged.

2.2 Step size rules

In the subgradient method the step size selection is very different from the standard gradient method. Many different types of step size rules are used. We used three basic step size rules in this paper.

– Square summable but not summable. The step sizes satisfy

$$\eta_k \geq 0, \qquad \sum_{k=1}^{\infty} \eta_k^2 < \infty, \qquad \sum_{k=1}^{\infty} \eta_k = \infty.$$

One typical example is

$$\eta_k = \mathbf{a}/(2 + k), \tag{4}$$

where $\mathbf{a} > 0$.

– Nonsummable diminishing. The step sizes satisfy

$$\eta_k \geq 0, \qquad \lim_{k\to\infty} \eta_k^2 = 0, \qquad \sum_{k=1}^{\infty} \eta_k = \infty.$$

Step sizes that satisfy this condition are called diminishing step size rules. A typical example is

$$\eta_k = \mathbf{b}/\sqrt{2 + k}, \tag{5}$$

where $\mathbf{b} > 0$.

– Step size control: Let us fixed same $\epsilon > 0$

$$\eta_{k+1} = \begin{cases} \frac{1}{2}\eta_k, \text{ if } \mathbf{F}(\mathbf{x}^{k+1}) \leq \mathbf{F}(\mathbf{x}^k) \text{ and } k = 1, \\ \eta_{decr}\eta_k, \text{ if } \mathbf{F}(\mathbf{x}^{k+1}) > \mathbf{F}(\mathbf{x}^k) \text{ and } k > 1 \\ \qquad\qquad \text{and } \epsilon \leq \eta_k, \\ \eta_{incr}\eta_k, \text{ otherwise,} \end{cases} \tag{6}$$

where η_{decr} is a decreasing and η_{incr} is a increasing coefficient of the stepsize η ($0.3 < \eta_{decr} < 1$, $\eta_{incr} > 1$) and $1 < \eta_1$ is initial stepsize with respect to $\mathbf{x}^1$.
We choose $\frac{1}{\eta_{decr}} \leq \eta_{incr}$ then

$$\epsilon < \eta_k \quad \forall k. \tag{7}$$

Also $\sum_{k=1}^{\infty} \eta_k^2 < \infty$ see for instance [13] and [26].

2.3 Control Subgradient Algorithm (CSA)

Subgradient and Nesterov step

1. Select a point $\mathbf{y}_0 \in \mathbf{E}$. Put

$$k = 0, \quad \mathbf{b}_0 = 1, \quad \mathbf{x}^{-1} = \mathbf{y}_0.$$

2. kth iteration.
   a) Compute $\mathbf{g_k} = \nabla \mathbf{f}(\mathbf{y}_k) + \tau \mathbf{sgn}(\mathbf{y}_k)$
   b) Put

$$\begin{aligned} &\mathbf{x}^k = \mathbf{y}_k - \eta_k \mathbf{g}_k, \\ &\mathbf{b}_{k+1} = 0.5(1 + \sqrt{4\mathbf{b}_k^2 + 1}\ ), \\ &\mathbf{y}_{k+1} = \mathbf{x}^k + (\tfrac{\mathbf{b}_k - 1}{\mathbf{b}_{k+1}})(\mathbf{x}^k - \mathbf{x}^{k-1}), \end{aligned} \tag{8}$$

   see for instance [19].

The recalculation of the point $\mathbf{y}_k$ in(8) is done using a "ravine" step, and η_k is the control step (6) .

2.4 Convergence proof

2.4.1 Assumptions

Here we give a proof of some typical convergence results for the control subgradient method.
We assume that there is a minimizer of $\mathbf{F}$, say $\mathbf{x}^*$.

We also make one other assumption on $\mathbf{f}$:
Since $\tau > 0$, we will assume that there exist a real $\mathbf{m} > 0$ and $\mathbf{M} > 0$ such that

$$\mathbf{m} \leq \|\nabla \mathbf{f}(\mathbf{y}_k)\|_2 \leq \mathbf{M} \forall k.$$

Remark 1 One of the stopping criteria of SpaRSA is

$$\|\mathbf{A}^t(\mathbf{A}\mathbf{x}^k - \mathbf{y})\|_2^2 < tolD$$

and $\nabla \mathbf{f}(\mathbf{x}^k) = \mathbf{A}^t(\mathbf{A}\mathbf{x}^k - \mathbf{y})$.

Let $\mathbf{L}$ the be Lipschitz constant of $\nabla \mathbf{f}$, then we have

$$\mathbf{f}(\mathbf{y}_{k+1}) - \mathbf{f}(\mathbf{x}^{k+1}) \geq \frac{1}{2\mathbf{L}} \|\nabla \mathbf{f}(\mathbf{y}_{k+1})\|_2^2 \tag{9}$$

see for instance [19].

Lemma 2 *There exist an integer* $\mathbf{N}$ *such that for* $k \geq \mathbf{N}$ *we have*

$$0.5\eta_k \|\mathbf{g}_k\|_2^2 \leq -\tau\eta_k \|\mathbf{g}_k\|_1 + \frac{1}{2\mathbf{L}} \|\nabla \mathbf{f}(\mathbf{y}_k)\|_2^2 \tag{10}$$

Proof :
There exist two real $\mathbf{m}$ and $\mathbf{M}$ such that

$$\mathbf{m} \leq \|\nabla \mathbf{f}(\mathbf{y}_k)\|_2 \leq \mathbf{M}, \quad \forall k,$$

then there exist also $\mathbf{M}_1$ such that

$$\|\nabla \mathbf{f}(\mathbf{y}_k)\|_1 \leq \mathbf{M}_1, \quad \forall k.$$

We define a function $\mathbf{x} \mapsto \mathbf{ones}(\mathbf{x}) \quad \forall \mathbf{x} \in \mathbb{R}^{\mathbf{n}}$ by

$$\mathbf{ones}(\mathbf{x}) = (1, \cdots, 1),$$

then we have the following inequalities :

$$\begin{aligned} \|\mathbf{g}_k\|_1 &\leq \|\nabla \mathbf{f}(\mathbf{y}_k)\|_1 + \tau \|\mathbf{sgn}(\mathbf{y}_k)\|_1 \\ &\leq \mathbf{M}_1 + \tau \|\mathbf{ones}(\mathbf{y}_k)\|_1 \\ &= \mathbf{M}_1 + \tau \mathbf{n} \end{aligned} \tag{11}$$

Let $\mathbf{G}_1 = \mathbf{M}_1 + \tau \mathbf{n}$, and

$$\begin{aligned} \|\mathbf{g}_k\|_2^2 &\leq (\|\nabla \mathbf{f}(\mathbf{y}_k)\|_2 + \tau \|\mathbf{sgn}(\mathbf{y}_k)\|_2)^2 \\ &\leq (\mathbf{M} + \tau \|\mathbf{ones}(\mathbf{y}_k)\|_2)^2 \\ &= (\mathbf{M} + \tau\sqrt{\mathbf{n}})^2 \end{aligned} \tag{12}$$

Let $\mathbf{G}_2 = (\mathbf{M} + \tau\sqrt{\mathbf{n}})^2$.
Since $\lim_{k \to \infty} \eta_k^2 = 0$ then there exist an integer $\mathbf{N}$ such that for all $k \geq \mathbf{N}$

$$\eta_k^2 \leq [\frac{\mathbf{m}^2}{2\mathbf{L}(0.5\mathbf{G}_2 + \tau\mathbf{G}_1)}]^2$$

$\Rightarrow$

$$\eta_k \leq \frac{\mathbf{m}^2}{2\mathbf{L}(0.5\mathbf{G}_2 + \tau\mathbf{G}_1)}$$

by equations (11) and (12),
$\Rightarrow$

$$\eta_k \leq \frac{\|\nabla \mathbf{f}(\mathbf{y}_k)\|_2^2}{2\mathbf{L}(0.5\|\mathbf{g}_k\|_2^2 + \tau\|\mathbf{g}_k\|_1)}$$

then equation (10) hold.

Theorem 1 *If the sequence* $\{\mathbf{x}_k\}_{\mathbf{k} \geq 0}$ *is constructed by subgradient algorithm, then there exist an integer* $\mathbf{N}$ *and a constants* $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{C}_3$ *such that* $\forall k \geq \mathbf{N}$

a) if we choose η_k *defined in (4), we have*

$$\mathbf{F}(\mathbf{x}_k) - \mathbf{F}(\mathbf{x}^*) \leq \frac{\mathbf{C}_1}{k+3}. \tag{13}$$

b) if we choose η_k defined in (5), we have

$$\mathbf{F}(\mathbf{x}_k) - \mathbf{F}(\mathbf{x}^*) \leq \frac{\mathbf{C}_2}{(k+3)^{\frac{3}{2}}}. \quad (14)$$

c) if we choose η_k defined in (6), we have

$$\mathbf{F}(\mathbf{x}_k) - \mathbf{F}(\mathbf{x}^*) \leq \frac{\mathbf{C}_3}{(k+3)^2}. \quad (15)$$

Proof :

Using inequality (9), we obtain

$$\mathbf{F}(\mathbf{y}_{k+1}) - \mathbf{F}(\mathbf{x}^{k+1}) \geq \tau(\|\mathbf{y}_{k+1}\|_1 - \|\mathbf{x}^{k+1}\|_1) + \tfrac{1}{2\mathbf{L}}\|\nabla\mathbf{f}(\mathbf{y}_{k+1})\|_2^2$$

by the reverse triangle inequality

$$\|\mathbf{y}_{k+1}\|_1 - \|\mathbf{x}^{k+1}\|_1 \geq -\eta_{k+1}\|\mathbf{g}_{k+1}\|_1$$

then

$$\mathbf{F}(\mathbf{y}_{k+1}) - \mathbf{F}(\mathbf{x}^{k+1}) \geq -\tau\eta_{k+1}\|\mathbf{g}_{k+1}\|_1 + \frac{1}{2\mathbf{L}}\|\nabla\mathbf{f}(\mathbf{y}_{k+1})\|_2^2$$

Using inequality (10), we obtain

$$\mathbf{F}(\mathbf{y}_{k+1}) - \mathbf{F}(\mathbf{x}^{k+1}) \geq 0.5\eta_{k+1}\|\mathbf{g}_{k+1}\|_2^2, \qquad \forall k \geq \mathbf{N} \quad (16)$$

Let $\mathbf{p}_k = (\mathbf{b}_k - 1)(\mathbf{x}^{k-1} - \mathbf{x}^k)$, then $\mathbf{y}_{k+1} = \mathbf{x}^k - \mathbf{b}_{k+1}^{-1}\mathbf{p}_k$, and from the definition of subgradient

$$\mathbf{F}(\mathbf{y}_{k+1}) \leq \mathbf{F}(\mathbf{x}^k) - \mathbf{b}_{k+1}^{-1}\mathbf{g}_{k+1}^t\mathbf{p}_k \quad (17)$$

by (16) and (17), we obtain $\forall k \geq \mathbf{N}$

$$0.5\eta_{k+1}\|\mathbf{g}_{k+1}\|_2^2 \leq \mathbf{F}(\mathbf{x}^k) - \mathbf{F}(\mathbf{x}^{k+1}) - \mathbf{b}_{k+1}^{-1}\mathbf{g}_{k+1}^t\mathbf{p}_k. \quad (18)$$

We have also

$$\mathbf{p}_{k+1} - \mathbf{x}^{k+1} = \mathbf{p}_k - \mathbf{x}_k + \mathbf{b}_{k+1}\eta_{k+1}\mathbf{g}_{k+1}.$$

Consequently,

$$\begin{aligned}\|\mathbf{p}_{k+1} - \mathbf{x}^{k+1} - \mathbf{x}^*\|^2 = \|\mathbf{p}_k - \mathbf{x}^k - \mathbf{x}^*\|^2 \\ + 2(\mathbf{b}_{k+1} - 1)\eta_{k+1}\mathbf{g}_{k+1}^t\mathbf{p}_k \\ + 2\mathbf{b}_{k+1}\eta_{k+1}\mathbf{g}_{k+1}^t(\mathbf{x}^* - \mathbf{y}_{k+1}) \\ + \mathbf{b}_{k+1}^2\eta_{k+1}^2\|\mathbf{g}_{k+1}\|^2.\end{aligned} \quad (19)$$

We substitute (16) and (18) into (19) for all $k \geq \mathbf{N}$

$$\begin{aligned}&\|\mathbf{p}_{k+1} - \mathbf{x}^{k+1} - \mathbf{x}^*\|^2 - \|\mathbf{p}_k - \mathbf{x}^k - \mathbf{x}^*\|^2 \\ &\leq 2(\mathbf{b}_{k+1} - 1)\eta_{k+1}\mathbf{g}_{k+1}^t\mathbf{p}_k \\ &- 2\mathbf{b}_{k+1}\eta_{k+1}(\mathbf{F}(\mathbf{x}^{k+1}) - \mathbf{F}(\mathbf{x}^*)) \\ &+ (\mathbf{b}_{k+1}^2 - \mathbf{b}_{k+1})\eta_{k+1}^2\|\mathbf{g}_{k+1}\|^2 \\ &\leq -2\mathbf{b}_{k+1}\eta_{k+1}(\mathbf{F}(\mathbf{x}^{k+1}) - \mathbf{F}(\mathbf{x}^*)) \\ &\quad + 2(\mathbf{b}_{k+1}^2 - \mathbf{b}_{k+1})\eta_{k+1}(\mathbf{F}(\mathbf{x}^k) - \mathbf{F}(\mathbf{x}^{k+1})) \\ &= 2\mathbf{b}_k^2\eta_{k+1}(\mathbf{F}(\mathbf{x}^k) - \mathbf{F}(\mathbf{x}^*)) \\ &- 2\mathbf{b}_{k+1}^2\eta_{k+1}(\mathbf{F}(\mathbf{x}^{k+1}) - \mathbf{F}(\mathbf{x}^*)) \\ &\leq 2\mathbf{b}_k^2\eta_k(\mathbf{F}(\mathbf{x}^k) - \mathbf{F}(\mathbf{x}^*)) \\ &- 2\mathbf{b}_{k+1}^2\eta_{k+1}(\mathbf{F}(\mathbf{x}^{k+1}) - \mathbf{F}(\mathbf{x}^*)).\end{aligned}$$

Thus

$$\begin{aligned}2\mathbf{b}_{k+1}^2\eta_{k+1}(\mathbf{F}(\mathbf{x}^{k+1}) - \mathbf{F}(\mathbf{x}^*)) &\leq 2\mathbf{b}_{k+1}^2\eta_{k+1}(\mathbf{F}(\mathbf{x}^{k+1}) - \mathbf{F}(\mathbf{x}^*)) \\ &\quad + \|\mathbf{p}_{k+1} - \mathbf{x}^{k+1} - \mathbf{x}^*\|^2 \\ &\leq 2\mathbf{b}_k^2\eta_k(\mathbf{F}(\mathbf{x}^k) - \mathbf{F}(\mathbf{x}^*)) \\ &\quad + \|\mathbf{p}_k - \mathbf{x}^k - \mathbf{x}^*\|^2 \\ &\leq 2\mathbf{b}_N^2\eta_{\mathbf{N}}(\mathbf{F}(\mathbf{x}^N) - \mathbf{F}(\mathbf{x}^*)) \\ &\quad + \|\mathbf{p}_N - \mathbf{x}^N - \mathbf{x}^*\|^2.\end{aligned}$$

It remains to observe that $\mathbf{b}_{k+1} \geq 1 + 0.5(k+1)$.

$$\mathbf{C} = 2(2\mathbf{b}_{\mathbf{N}}^2\eta_{\mathbf{N}}(\mathbf{F}(\mathbf{x}^N) - \mathbf{F}(\mathbf{x}^*)) + \|\mathbf{p}_{\mathbf{N}} - \mathbf{x}^{\mathbf{N}} - \mathbf{x}^*\|^2).$$

a) Let

$$\mathbf{C}_1 = \mathbf{C}/\mathbf{a},$$

and $\eta_k = \mathbf{a}/(2+k)$, then (13) hold.

b) Let

$$\mathbf{C}_2 = \mathbf{C}/\mathbf{b},$$

and $\eta_k = \mathbf{b}/\sqrt{k+2}$, then (14) hold.

c) Let

$$\mathbf{C}_3 = \mathbf{C}/\epsilon,$$

and η_k is control step size defined in (6), then (15) hold.

3 Computational experiment

In this section, we report results of experiments aimed which demonstrate the competitive performance of the CSA approach on problems of the form (1), including problems with complex data, and its ability to handle ℓ_1 regularizers and η_k is step size control (6).

Comparing the speed of subgradient method (CSA) with that of the current state of the art method SpaRSA. All the experiments were carried out on a personal computer with an HP Intel(R) Core(TM) i3-3217U CPU processor 1.80GHz, 3.90 Go RAM, x64, using MATLAB(R2008b) for Windows 8.

A less sophisticated criterion makes use of the relative change in objective value at the last step. We terminate at iteration k if

$$\left|\frac{\mathbf{F}(\mathbf{x}^k) - \mathbf{F}(\mathbf{x}^{k-1})}{\mathbf{F}(\mathbf{x}^{k-1})}\right| \leq tol,$$

where $tol \in [1e-4, 1e-6]$.

3.1 Image Deblurring with wavelets

We now present a set of six deblurring experiments illustrating the performance of CSA, the images went through a uniform blur of size 9×9 followed by an additive Gaussian noise with standard deviation σ, the signal-to-noise ratio of the blurred image (BSNR $\equiv var[\mathbf{Ax}]/\sigma^2$) is set to BSNR=40 dB:

1. in the first experiment, the image is the "Cameraman" of size 256×256, corresponding to a noise standard deviation of σ =?,
2. in the second experiment, the image is the "Heart" of size 192×192, corresponding to a noise standard deviation of σ =?,
3. in the third experiment, the image is the "Rice" of size 256×256, corresponding to a noise standard deviation of σ =?,

4. in the fourth experiment, the image is the "Lena" of size 256×256, corresponding to a noise standard deviation of $\sigma =?$,
5. in the fifth experiment, the image is the "Shepp-Logan" phantom of size 256×256, corresponding to a noise standard deviation of $\sigma =?$,
6. in the sixth experiment, the image is the "Kasba" of size 200×200, corresponding to a noise standard deviation of $\sigma =?$.

All pixels of the original images described in the examples were first scaled into the range between 0 and 1.

The original and observed images are given in Figures 1-6.
For these experiments we assume reflexive (Neumann) boundary conditions [16]. We then tested SpaRSA, and CSA for solving problem (1), where $\mathbf{y}$ represents the (vectorized) observed image, and

$$\mathbf{A} = \mathbf{RW},$$

where $\mathbf{R}$ is the matrix representing the blur operator applied via FFT and $\mathbf{W}$ is the inverse of a three stage Haar wavelet transform. The regularization parameter was chosen to be $\tau = 0.0075$, where τ is hand tuned in each case for best improvement in the Signal-to-Noise Ration (SNR) or ISNR, so that the comparison is carried out in the regime that is relevant in practice, and the initial image was the blurred image.

3.2 Computational results

To measure the performance of the resolution enhancement algorithm, a quantitative image quality metric is needed.
ISNR is widely used in image restoration [15] and will be used here. The formula for ISNR is then given as

$$\text{ISNR} \equiv ||\mathbf{y} - \mathbf{x}||^2 / ||\hat{\mathbf{x}} - \mathbf{x}||^2.$$

ISNR is used to evaluate the sharpness of the image. The higher the ISNR value is, the better the quality of the reconstruction image.

It's clear that the performance of such criteria (e.g., in terms of ISNR) does not depend on the optimization algorithms used to implement them, but only on the choice of the type of regularizer.

Table 1 report the number of iterations, MSE, ISNR and CPU times taken in each methods.

Furthermore, some basic experiments suggest the new method CSA is faster than SpaRSA.

Experiments	Methods	CPU time/s	iters	MSE	ISNR/dB
Cameraman	SpaRSA	42.5	398	90.6	7.79
	CSA	19.4	124	89.8	7.83
Heart	SpaRSA	17.2	321	23.2	9.5
	CSA	7.19	91	22.4	9.65
Rice	SpaRSA	38.7	372	58.8	7.87
	CSA	11.4	73	55.3	8.13
Lina	SpaRSA	40	380	70	7.42
	CSA	13.9	90	69.2	7.47
Shepp-Logan phantom	SpaRSA	45	436	75.2	10.7
	CSA	42.4	272	57.7	11.8
Kasba	SpaRSA	22	375	177	7.82
	CSA	9.45	117	171	7.97

Table 1 computational results for deblurring

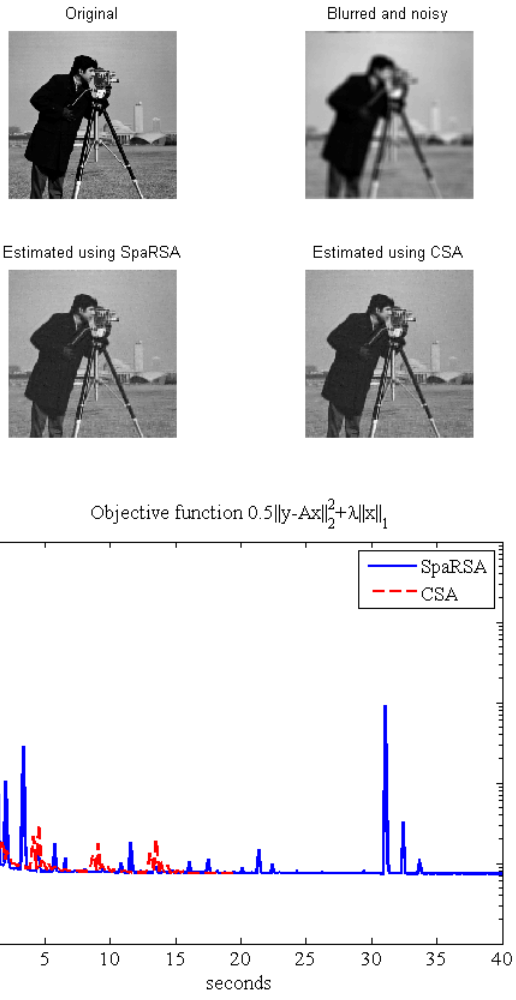

Fig. 1 Cameraman : Origin, blurred & noise, SpaRSA estimated and CSA estimated images

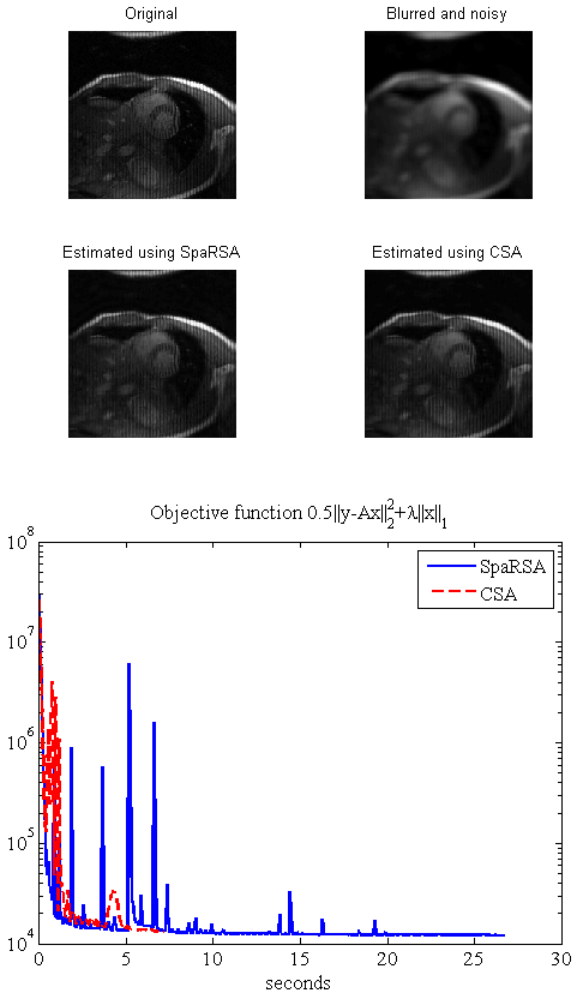

Fig. 2 Heart : Origin, blurred & noise, SpaRSA estimated and CSA estimated images

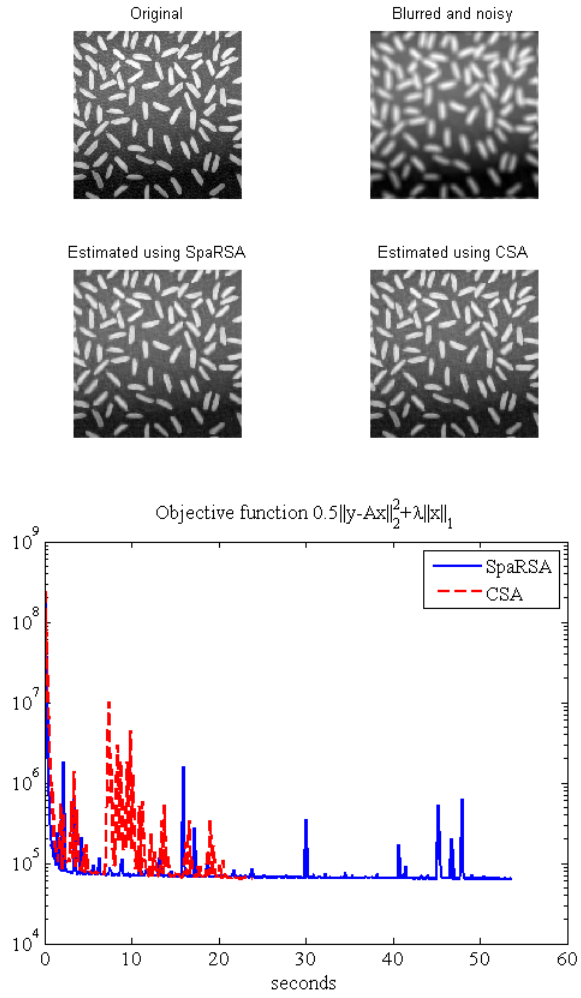

Fig. 3 Rice : Origin, blurred & noise, SpaRSA estimated and CSA estimated images

4 Concluding Remarks

We have introduces a control subgradient algorithm for image ℓ_1 regularization intended to have a low computation complexity and high convergence rate of $1/k^2$.

The theoretical proof of convergence rate and the numerical experiments show that CSA with nonsummable diminishing step size is fast and effective to calculate, and consistently generates higher quality images. This method outperforms SpaRSA in terms of speed and accuracy.

In the future works, we modified CSA for image ℓ_1-TV regularization and perturbation stochastic of CSA image nonconvex regularization with optimal regularization parameter.

Acknowledgements The authors are indebted to the anonymous Reviewers and Editor for many suggestions to improve the original manuscript.

References

1. M. Afonso, J. Bioucas-Dias, and M. Figueiredo, An augmented Lagrangian based method for the constrained formulation of imaging inverse problems, IEEE Transactions on Image Processing, Vol. 20, no. 3, pp 681 - 695, March, 2011.
2. A. Beck and M. Teboulle, A fast iterative shrinkage-thesholding algorithm for linear inverse problems, SIAM J. Imaging Sciences, vol. 2, no. 1, pp. 183-202, 2009.
3. G. Archer and D. M. Titterington, On some Bayesian / regularization methods for image restoration, IEEE Trans. Im. Proc., 4(7): 98995, 1995.
4. S. Becker, J. Bobin and E. J. Candès, NESTA: A Fast and Accurate First-Order Method for Sparse Recovery, SIAM J. Imaging Sciences, vol. 4, no. 1, pp. 1-39, 2011.
5. M. Bertero and P. Boccacci, Introduction to inverse problems in imaging, IOP Publishing, Bristol, UK, 1998.
6. U. Brannlund, K.C. Kiwiel and P.O. Lindberg, A descent proximal level bundle method for convex nondifferentiable optimization, Operations Research Letters, Vol. 17(3), pp. 121-126, 1995.
7. A. Chambolle, R.A. DeVore, N.Y. Lee, and B.J. Lucier, Nonlinear wavelet image processing: variational problems,

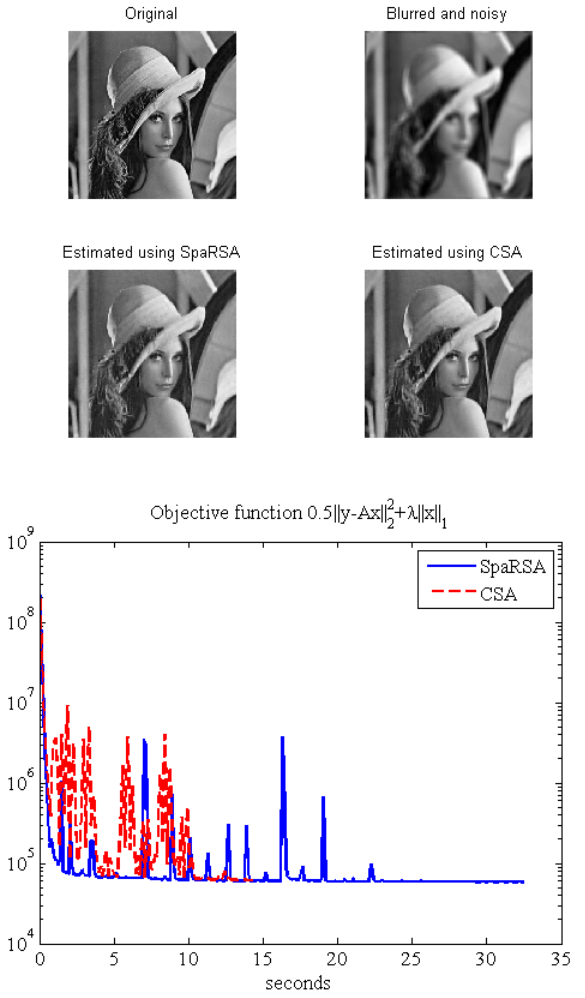

Fig. 4 Lena : Origin, blurred & noise and CSA estimated images

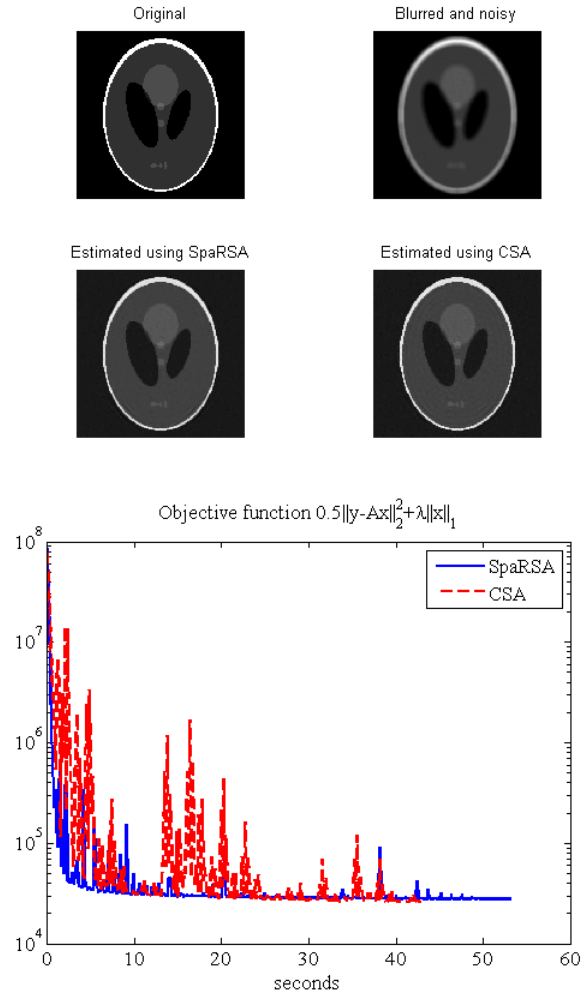

Fig. 5 Shepp-Logan phantom : Origin, blurred & noise and CSA estimated images

compression, and noise removal through wavelet shrinkage, IEEE Trans. Image Processing, vol. 7, pp. 319-335, 1998.

8. T.F. Chan and A.E. Glu, Aspects of total variation regularized L^1 function approximation, SIAM Journal on Applied Mathematics, 65(5):1817-1837, 2005.
9. S.S Chen, D.L. Donoho, and M.A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput., vol. 20, no. 1, pp. 33-61, 1998.
10. C. Clason, B. Jin, and K. Kunisch, A Duality-Based Splitting Method for ℓ^1-TV Image Restoration with Automatic Regularization Parameter Choice, SIAM J. Sci. Comput., 32(3), 14841505, 2010.
11. P. Combettes and V. Wajs, Signal recovery by proximal forwardbackward splitting, SIAM Journal on Multiscale Modeling & Simulation, vol. 4, pp. 11681200, 2005.
12. D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory, vol.52, no. 4, pp. 12891306, Apr. 2006.
13. A. El Mouatasim, Perturbation Stochastic for Global Optimization: Analyses and Application, Scholars Press, ISBN: 978-3-639-66674-8, 2014.
14. M. Figueiredo and R. Nowak, An EM algorithm for wavelet-based image restoration, IEEE Trans. Im. Proc., vol. 12, pp. 906916, 2003.
15. Galbraith, A.E.; Theiler, J.; Thome, K.J.; Ziolkowski, R.W, Resolution enhancement of multilook imagery for the multispectral thermal imager. IEEE Trans. Geosci. Remote Sens. 2005, 43, 19641977.
16. P. C. Hansen, J. G. Nagy, and D. P. OĹeary, Deblurring Images: Matrices, Spectra, and Filtering, Fundam. Algorithms 3, SIAM, Philadelphia, 2006.
17. A. Jain, Fundamentals of digital image processing, Prentice Hall, Englewood Cliffs, 1989.
18. C. Lemaréchal, An extension of Davidon methods to nondifferentiable problems, Mathematical Programming Study, Vol. 3, pp. 95-109, 1975.
19. Y. E. Nesterov, A method for solving the convex programming problem with convergence rate O$(1/k^2)$, Dokl. Akad. Nauk SSSR, 269, pp. 543547, 1983.
20. T. Poggio, V. Torre, and C. Koch, Computational vision and regularization theory, Nature, vol. 317, pp. 314-319, 1985.
21. G. Sluder and D.E. Wolf, Digital microscopy, 3rd ed. New York: Academic, 2007.
22. D. Terzopoulos, Regularization of inverse visual problems involving discontinuities, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, pp. 413-424, 1986.

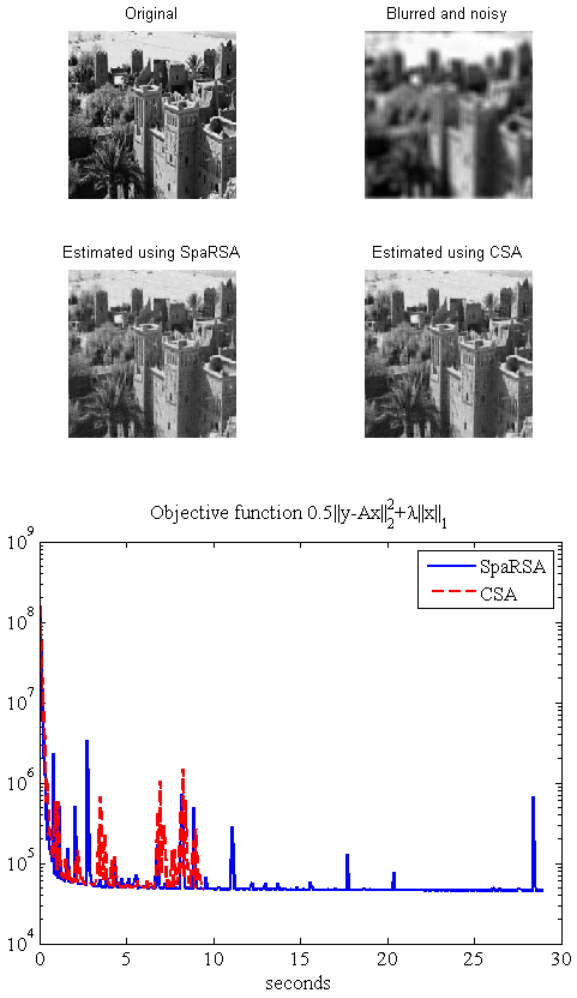

Fig. 6 Kasba : Origin, blurred & noise, SpaRSA estimated and CSA estimated images

23. S. Twomey, Introduction to the mathematics of inversion in remote sensing, New York: Dover, 2002.
24. S. Ramani, Z. Liu, J. Rosen, J. Nielsen, and J.A. Fessler, Regularization Parameter Selection for Nonlinear Iterative Image Restoration and MRI Reconstruction Using GCV and SURE-Based Methods, IEEE Trans Image Process, 21(8): 3659-3672, 2012.
25. L.I. Rudin, S.J. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D, 60:259-268, 1992.
26. S.P. Uryas'ev, New Variable-Metric Algorithms for Non-differentiable Optimization Problems, Journal of Optimization Theory and Applications: vol. 71, no. 2, pp. 359-388, 1991
27. C. Vogel, Computational methods for inverse problems, Philadelphia, PA: SIAM, 2002.
28. Y. Wang, J. Yang, W. Yin and Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, SIAM Journal Imaging Sciences, Vol. 1, No. 3, pp. 248-272, 2008.
29. S. Wright, R. Nowak, M. Figueiredo, Sparse reconstruction by separable approximation, IEEE Transactions on Signal Processing, Vol.57, pp. 2479-2493, 2009.
30. W. Yin, S. Osher, D. Goldfarb and J. Darbon, Bregman iterative algorithms for ℓ_1 minimization with applications to compressed sensing, SIAM Journal Imaging Sciences, Vol. 1, No. 1, pp. 143-168, 2008.