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

Control Proximal Gradient Algorithm for Image ℓ_1 Regularization

Abdelkrim El Mouatasim

Received: date / Accepted: date

Abstract We consider Control Proximal Gradient Algorithm (CPGA) for solving minimization of a nonsmooth convex function ℓ_1 regularized least squares by the discretized ℓ_1 norm models arising in image processing. This proximal gradient algorithm with control step size is attractive due to its simplicity, however, the are also known to converge quite slowly.
In this paper, we present a fast control proximal gradient algorithm by adding Nesterov step which preserve the computational simplicity of proximal gradient methods, but with a convergence rate $1/k^2$ which is proven to be significantly better, both theoretically and practically. Initial promising numerical results for images deblurring demonstrate the capabilities of CPGA by comparing with fast iterative shrinkage-thresholding algorithms (FISTA).

Keywords Proximal gradient algorithm · Nesterov algorithm · step-size control · ℓ_1 regularization · image deblurring · FISTA

1 Introduction

Consider a linear inverse model of the form

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \eta \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{\mathbf{P}\times\mathbf{Q}}$ is linear operator, $\mathbf{y} \in \mathbb{R}^{\mathbf{Q}}$ is an observation image, $\mathbf{x} \in \mathbb{R}^{\mathbf{P}}$ is an image of unknowns and $\eta \in \mathbb{R}^{\mathbf{Q}}$ is a noise.

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

The challenge in many linear inverse problems is that they are ill-posed. To cope with the ill-posed nature of these problems, a large number of statistical techniques (least squares regression) has been developed, most of them under the ℓ_1 regularization [11,17, 20], in which one seeks to find the solution of

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

where $\tau > 0$ is regularization parameter acts as a trade-off between the number of non zero components of the wavelet representation of the signal and the synthesis error [8].

It is an unconstrained convex nonsmooth minimization problem due to the presence of the 1-norm. The presence of the ℓ_1 term in the objective function is used to induce sparsity in the optimal solution, of (2) [7,25].

The problem (2) can be in many fields, like statistical and machine learning it is used for feature selection in the lasso [19], astronomical imaging [23], microscopy [18], remote sensing [21], compressed sensing ([9,26]), and, more generally, image restoration [14,1].

Arguably, the most popular methods to solve problem (2) is in the class of iterative shrinkage/thresholding algorithms (ISTA) (for instance [2] as well as [6,12]), these methods apply to nonsmooth problems, but they are proximal gradient methods and never explicitly use a subgradient with control step[11].

In the next section, we introduce a new proximal gradient algorithm, fully developed on the discrete domain, which is simple and yet computationally efficient.

The standard model in image deblurring assumes that the noisy blurred observed version $\mathbf{y}$, of an original image $\mathbf{x}$, was obtained via (1) where $\mathbf{A}$ is the matrix representation of a convolution and η is Gaussian white

noise. 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 and the subject (motion blur), bad focusing (defocusing blur), or a number of other mechanisms [4].

The main contribution is a derivation of an accelerated control proximal gradient method that has convergence rate $1/\mathbf{k}^2$. As far as we know, this is a novel kind of result for proximal gradient methods for solving ℓ_1 regularization image problem.

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) $||.||_1$ stands for the ℓ_1 norm, defined as $||\mathbf{x}||_1 = \sum_i |x_i|$.

We define the function $\mathbf{sgn}$ as follow

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

In this paper, we propose a new method with two step iterative. It can be used to accelerate proximal gradient methods with control step size for solving an unconstrained or bound constraints convex image ℓ_1 regularization problems.

We have implemented the proposed algorithm in MATLAB and compared it with a FISTA implementation of the algorithm described in [2], which is the fastest algorithm for solving ℓ_1 regularization problem. Numerical results presented in section ??? show that our algorithm CPGA is much faster for image ℓ_1 regularization, especially when the blurring kernel is relatively large. Compared with a few other deblurring algorithm which solve different models, including FISTA our algorithm consistently generates higher quality images in comparable running times.

2 Proximal gradient method

2.1 Proximal gradient algorithm

A standard approach for solving the problem (2) is a subgradient method [11]. However, these method have several drawbacks like the are not descent method and they are often slow. In the view of these disadvantages, more efficient method were created for solving the problem (2), it called proximal gradient method.

The proximal gradient method is used for unconstrained optimization problem with an objective function split in the convex smooth part and the convex nonsmooth part

$$\min_{\mathbf{x} \in \mathbf{E}} \mathbf{F}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) + \mathbf{h}(\mathbf{x}), \tag{3}$$

where:

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

It is assumed that solution $\mathbf{x}^*$ of (3) exists.

We define the proximal operator associated with the convex function $\mathbf{h}(\mathbf{x})$ as follows

$$\mathbf{prox_h}(\mathbf{x}) = \arg\min_z (\mathbf{h}(\mathbf{z}) + \frac{1}{2}\|\mathbf{z} - \mathbf{x}\|_2^2) \tag{4}$$

For solving (3), the proximal gradient algorithm has the main iteration with form

$$\mathbf{x^k} = \mathbf{prox}_{\alpha_\mathbf{k},\mathbf{h}}(\mathbf{x^{k-1}} - \alpha_k \nabla f \mathbf{x^{k-1}}), \tag{5}$$

where $\alpha_k > 0$ is step size that can be fixed or determined by line search.

The proximal operator associated to the function $\mathbf{h}$ and α_k, is simply expressed as follows

$$\mathbf{prox}_{\alpha,\mathbf{h}}(\mathbf{x})_i = \begin{cases} x_i - \alpha\tau, \, x_i > \alpha\tau, \\ x_i + \alpha\tau, \, x_i < -\alpha\tau, \\ 0 \qquad\qquad |x_i| \leq \alpha\tau. \end{cases}$$

An equivalent vector formulation of the proximal operator, more suitable for Matlab implementation, is the form

$$\mathbf{prox}_{\alpha,\mathbf{h}}(\mathbf{x}) = \mathbf{sign}(\mathbf{x}) \max[0, |x| - \alpha\tau].$$

In this case ($h(x) = \tau\|x\|_1$), the proximal operator $\mathbf{prox}_{\alpha,\mathbf{h}}$ has a closed form expression called the soft-thresholding or shrinkage function $\mathbf{soft}(\mathbf{x}, \alpha)$.

2.1.1 Fast iterative soft-thresholding algorithm (FISTA)

The iterative soft-thresholding algorithm (ISTA) used a fixed step size at each iteration k

$$\alpha_k = \frac{1}{\mathbf{L}},$$

where $\mathbf{L}$ is the Lipschitz constant of $\nabla\mathbf{f}$.

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

$$\mathbf{x}^k = \mathbf{soft}(\mathbf{y_k} - \frac{\mathbf{1}}{\mathbf{L}}\nabla\mathbf{f}(\mathbf{y_k}), \frac{\tau}{\mathbf{L}}),$$

   b) Put

$$\begin{aligned} \mathbf{b}_{k+1} &= 0.5(1 + \sqrt{4\mathbf{b}_k^2 + 1}\ ), \\ \mathbf{y}_{k+1} &= \mathbf{x}^k + (\tfrac{\mathbf{b}_k - 1}{\mathbf{b}_{k+1}})(\mathbf{x}^k - \mathbf{x}^{k-1}). \end{aligned} \tag{6}$$

2.2 Control proximal gradient algorithm (CPGA)

In this paper we propose the control step size for proximal gradient algorithm defined as follow:

$$\alpha_{k+1} = \begin{cases} \frac{1}{2}\alpha_k, \text{ if } \mathbf{F}(\mathbf{x}^{k+1}) \le \mathbf{F}(\mathbf{x}^k) \text{and } k = 1, \\ \alpha_{decr}\alpha_k, \text{ if } \mathbf{F}(\mathbf{x}^{k+1}) > \mathbf{F}(\mathbf{x}^k) \text{and } k > 1 \\ \qquad\qquad\qquad \& \quad \epsilon \le \alpha_{decr}\alpha_k, \\ \alpha_{incr}\alpha_k, \text{ otherwise,} \end{cases} \tag{7}$$

where α_{decr} is a decreasing and α_{incr} is a increasing coefficient of the step size α, $(0 < \alpha_{decr} < 1 < \alpha_{incr})$ and $1 < \alpha_1$ is initial step size with respect to $\mathbf{x}^1$, and $0 < \epsilon$ is a lower bound of the sequence α_k

$$\epsilon \le \alpha_k \quad \forall k. \tag{8}$$

The main intention is to introduce a accelerate proximal gradient algorithm with control step size, prove it convergence and present some promising results of numerical experiments.
Proximal gradient 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

$$\mathbf{x}^k = \mathbf{soft}(\mathbf{y_k} - \alpha_\mathbf{k}\nabla\mathbf{f}(\mathbf{y_k}), \alpha_\mathbf{k}\tau),$$

   b) Put

$$\begin{aligned} \mathbf{b}_{k+1} &= 0.5(1 + \sqrt{4\mathbf{b}_k^2 + 1}\ ), \\ \mathbf{y}_{k+1} &= \mathbf{x}^k + (\tfrac{\mathbf{b}_k - 1}{\mathbf{b}_{k+1}})(\mathbf{x}^k - \mathbf{x}^{k-1}). \end{aligned} \tag{9}$$

The recalculation of the point $\mathbf{y}_k$ in(9) is done using a "ravine" step, and α_k is the control step (7) .

Lemma 1 *The sequence $(\mathbf{b}_k)_{k\ge1}$ generated by the scheme (9) satisfies the lower bound*

$$\frac{k+1}{2} \le \mathbf{b}_k. \tag{10}$$

Proof : We have

$$0.5(1 + \sqrt{4\mathbf{b}_{k-1}^2 + 1}) \ge 0.5(1 + 2\mathbf{b}_{k-1}^2)$$

then

$$\mathbf{b}_k \ge \frac{1}{2} + \mathbf{b}_{k-1}$$

Therefore, we have $\mathbf{b}_k \ge \frac{k+1}{2}$.

:

3 Convergence rate of CPGA

We assume that there is a minimizer of $\mathbf{F}$, say $\mathbf{x}^*$.

Since the stopping criteria of SpaRSA can be

$$\|\mathbf{f}(\mathbf{x}^k)\|_2^2 < tolD$$

We will assume that there exist a reals $\mathbf{m} > 0$ and $\mathbf{M} > 0$ such that

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

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

$$\mathbf{f}(\mathbf{y}_{k+1}) - \mathbf{f}(\mathbf{x}^{k+1}) \ge \frac{1}{2\mathbf{L}}\|\nabla\mathbf{f}(\mathbf{y}_{k+1})\|_2^2 \tag{11}$$

see for instance [16].

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

$$\begin{aligned} 0.5\|\mathbf{y}_k - \mathbf{soft}(\mathbf{u_k}, \alpha_\mathbf{k}\tau)\|_2^2 \le -\tau\|\mathbf{y}_k - \mathbf{soft}(\mathbf{u_k}, \alpha_\mathbf{k}\tau)\|_1 \\ + \tfrac{1}{2\mathbf{L}}\|\nabla\mathbf{f}(\mathbf{y}_k)\|_2^2 \end{aligned} \tag{12}$$

where $\mathbf{u}_k = \mathbf{y}_k - \alpha_k\nabla f(\mathbf{y}_k)$.

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

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

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

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

We define a function $\mathbf{x} \mapsto \mathbf{ones}(\mathbf{x})$ by

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

and the generalized gradient of F at the point $\mathbf{y}_k$ as

$$\mathbf{g}_k = \frac{\mathbf{y}_k - \mathbf{soft}(\mathbf{y_k} - \alpha_\mathbf{k}\nabla\mathbf{f}(\mathbf{y_k}), \alpha_\mathbf{k}\tau)}{\alpha_k}$$

then we have the following inequalities :

$$\begin{aligned} \|\mathbf{g}_k\|_1 &\le \|\nabla\mathbf{f}(\mathbf{y}_k)\|_1 + \tau\|\mathbf{sgn}(\mathbf{y}_k)\|_1 \\ &\le \mathbf{M}_1 + \tau\|\mathbf{ones}(\mathbf{y}_k)\|_1 \\ &= \mathbf{M}_1 + \tau\mathbf{n} \end{aligned} \tag{13}$$

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

$$\begin{aligned}\|\mathbf{g}_k\|_2^2 &\leq (\|\nabla\mathbf{f}(\mathbf{y}_k)\|_2 + \tau\|\mathbf{sgn}(\mathbf{y}_k)\|_2)^2 \\ &\leq (\mathbf{M} + \tau\|\mathbf{ones}(\mathbf{y}_k)\|_2)^2 \\ &= (\mathbf{M} + \tau\sqrt{\mathbf{n}})^2\end{aligned} \tag{14}$$

Let $\mathbf{G}_2 = (\mathbf{M} + \tau\sqrt{\mathbf{n}})^2$.

We have also $\sum_{k=1}^{\infty} \alpha_k^2 < 0$ [11], then $\lim_{k\to\infty} \alpha_k^2 = 0$. Therefor if we let

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

where $[\mathbf{z}]$ is integer part of $\mathbf{z}$, then there exist an integer $\mathbf{K}$ such that

$$\alpha_k^2 \leq \mathbf{N} \quad \forall k \geq \mathbf{K}$$

So,

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

by equations (13) and (14), we have

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

then equation (12) hold.

Theorem 1 *If the sequence $\{\mathbf{x}_k\}_{\mathbf{k}\geq 0}$ is constructed by accelerated control proximal gradient algorithm, then there exist an integer $\mathbf{K}$ and a constant $\mathbf{C}$ such that $\forall k \geq \mathbf{K}$*

$$\mathbf{F}(\mathbf{x}_k) - \mathbf{F}(\mathbf{x}^*) \leq \mathbf{C}/k^2. \tag{15}$$

Proof :

Using inequality (11), we obtain

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

by the reverse triangle inequality

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

where

$$\mathbf{g}_{k+1} = \frac{\mathbf{y}_{k+1} - \mathbf{soft}(\mathbf{y_{k+1}} - \alpha_{\mathbf{k+1}}\nabla\mathbf{f}(\mathbf{y_{k+1}}), \alpha_{\mathbf{k+1}}\tau)}{\alpha_{k+1}}$$

is the generalized gradient of $\mathbf{F}$ at the point $\mathbf{y}_{k+1}$, then

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

Using inequality (12), we obtain

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

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 generalized gradient

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

by (16) and (17), we obtain

$$0.5\alpha_{k+1}\|\mathbf{g}_{k+1}\|_2^2 \leq \mathbf{F}(\mathbf{x}^k) - \mathbf{F}(\mathbf{x}^{k+1}) - \mathbf{b}_{k+1}^{-1}\mathbf{g}_{k+1}^t\mathbf{p}_k, \ \forall k \geq \mathbf{K}. \tag{18}$$

We have also

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

Consequently,

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

We substitute (16) and (18) into (19) there exist integer $\mathbf{K}$ $\forall k \geq \mathbf{K}$

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

Thus

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

Let

$$\mathbf{C} = \frac{1}{2\epsilon}(2\mathbf{b}_{\mathbf{K}}^2\alpha_{\mathbf{K}}(\mathbf{F}(\mathbf{x}^K) - \mathbf{F}(\mathbf{x}^*)) + \|\mathbf{p}_{\mathbf{K}} - \mathbf{x}^{\mathbf{K}} - \mathbf{x}^*\|^2).$$

Using inequality (10) we obtain $\mathbf{b}_{k+1} \geq 0.5(k+2)$. Therefor,

$$\mathbf{F}(\mathbf{x}^{k+1}) - \mathbf{F}(\mathbf{x}^*) \leq \frac{\mathbf{C}}{(k+2)^2}$$

then (15) hold.

4 Computational experiment

In this section, we report results of experiments aimed which demonstrate the competitive performance of the CPGA on problems of the form (2), including problems with complex data, and its ability to handle ℓ_1 regularizers and α_k is step size control (7).

Comparing CPGA method with that of the current state of the art method FISTA.

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(R2016b) for Windows 8.

We terminate at iteration k if one of this stopping criterion are used $StopCriterion$:

1. A less sophisticated criterion makes use of the relative change in objective value at the last step
$$\left|\frac{\mathbf{F}(\mathbf{x}^k) - \mathbf{F}(\mathbf{x}^{k-1})}{\mathbf{F}(\mathbf{x}^{k-1})}\right| \leq tolerance,$$
where $tolerance = 10^{-5}$.
2. stop when the relative norm of the difference between two consecutive estimates falls below $tolerance = 10^{-5}$
$$\frac{\|\mathbf{x}^k - \mathbf{x}^{k-1}\|}{\|\mathbf{x}^{k-1}\|} \leq tolerance,$$
3. stop when the objective function becomes equal or less than $tolerance$.
$$funval \leq tolerance$$

We first run CPGA by used $StopCriterion = 1$ or $StopCriterion = 2$ then we obtain a last function value $funval_{CPGA}$ that by used for stopping FISTA with $StopCriterion = 3$ and $tolerance = funval_{CPGA}$.

4.1 Image Deblurring

We now present a set of seven images deblurring experiments illustrating the performance of CPGA:

Table 1 Images for deblurring

Experiments	Size	Extension
Man	1024×1024	.tiff
Boat	512×512	.png
Head CT	512×512	.tif
Lena	512×512	.tif
Cameraman	512×512	.tif
Gold hill	512×512	.gif
Patches	256×256	.gif

All pixels of the original images described in th amples were first scaled into the range between (1. The image went through a Gaussian blur of size and standard deviation 4 (applied by the MAT functions imfilter and fspecial) followed by an ad zero-mean white Gaussian noise with standard c tion 10^{-3}. The original and observed images are given in Figures 1-14.
For these experiments we assume reflexive (Neumann) boundary conditions [13]. We then tested FISTA and CPGA for solving problem (2). The regularization parameter was chosen to be $\tau = 0.001$, where τ is hand tuned in each case for best improvement in the Signal-to-Noise Ration (SNR), so that the comparison is carried out in the regime that is relevant in practice, and the initial image was the blurred image.

4.2 Computational results

Table 4.2 reports the number of iterations, ISNR and CPU times taken in each methods, where
$$\mathbf{ISNR} \equiv ||\mathbf{y} - \mathbf{x}||^2 / ||\hat{\mathbf{x}} - \mathbf{x}||^2.$$

Table 2 computational results for deblurring

Experiments	Methods	CPU time/s	iters	MSE	ISNR / dB
Man	FISTA	70.5	116	37.9	6.79
	CPGA	57.9	93	38	6.78
Boat	FISTA	9.06	54	55.5	6.16
	CPGA	6.81	43	55.1	6.2
Head CT	FISTA	34.5	218	38.9	10.4
	CPGA	27.7	175	39.1	10.4
Lena	FISTA	10.4	65	28.3	5.88
	CPGA	8.13	53	28	5.91
Cameraman	FISTA	18.8	116	20.7	9.87
	CPGA	14.8	93	20.9	9.84
Gold hill	FISTA	11.7	69	41.2	5.75
	CPGA	8.73	55	41.2	5.75
Patches	FISTA	4.75	158	138	7.45
	CPGA	3.67	126	138	7.45

Furthermore, all basic experiments suggest the CPGA is faster than FISTA for image deblurring with large scale.

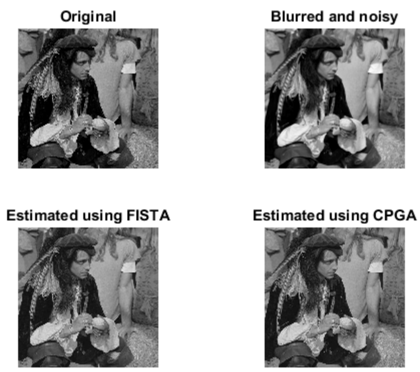

Fig. 1 Man : Origin, blurred & noise and CPGA estimated images

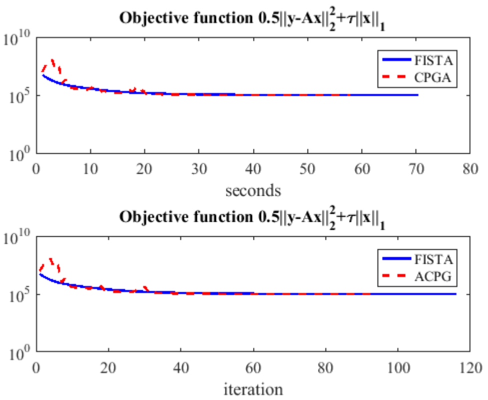

Fig. 2 Man : Evaluation of objective function

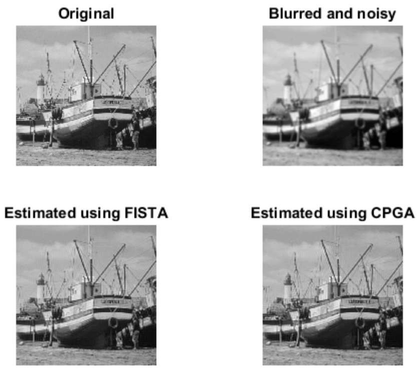

Fig. 3 Boat : Origin, blurred & noise and CPGA esti images

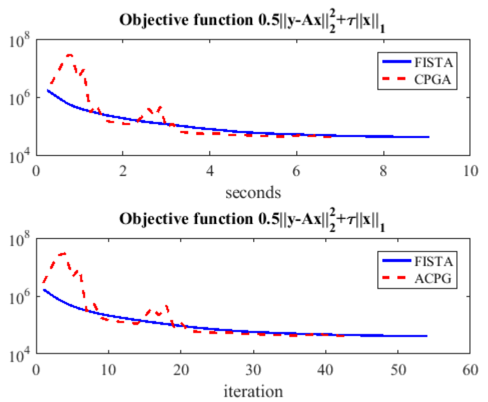

Fig. 4 Boat : Evaluation of objective function

5 Concluding Remarks

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

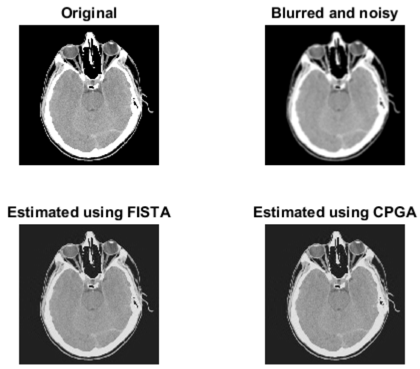

Fig. 5 Head CT : Origin, blurred & noise and CPGA estimated images

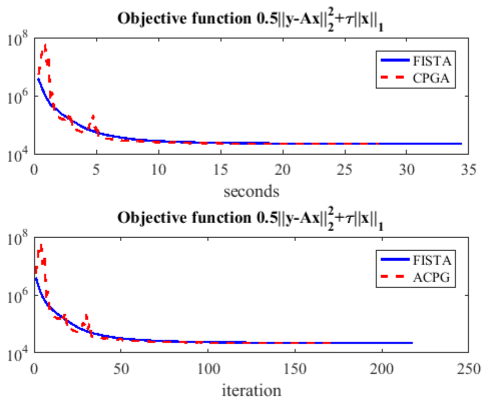

Fig. 6 Head CT : Evaluation of objective function

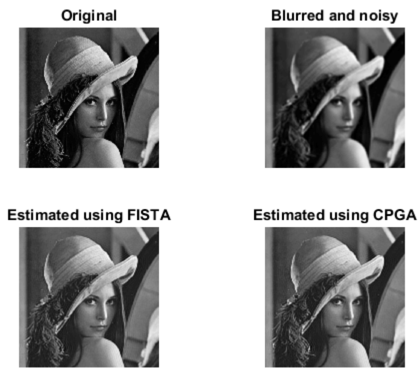

Fig. 7 Lena : Origin, blurred & noise and CPGA estimated images

The theoretical proof of convergence rate and the numerical experiments show that CPGA with control step size is fast and effective to calculate, and consistently generates higher quality images.

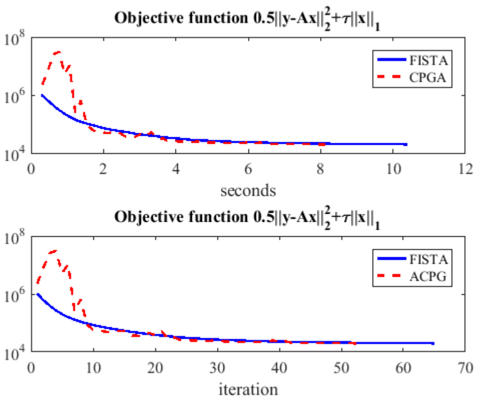

Fig. 8 Lena : Evaluation of objective function

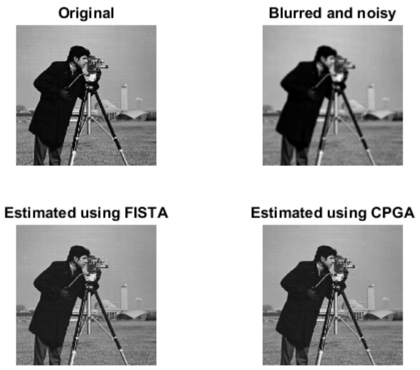

Fig. 9 Cameraman : Origin, blurred & noise and CPGA estimated images

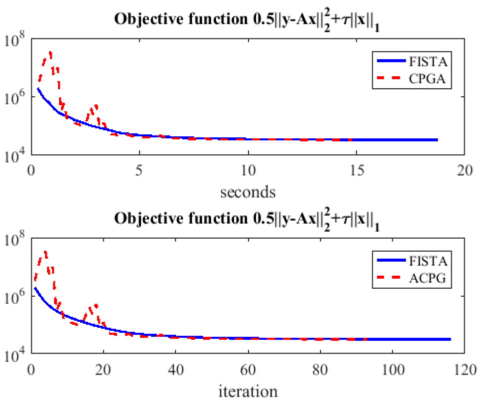

Fig. 10 Cameraman : Evaluation of objective function

In the future work, we add a random perturbation to CPGA for image nonconvex regularization nonsmooth minimization without constraints.

Fig. 11 Gold hill : Origin, blurred & noise and CPGA estimated images

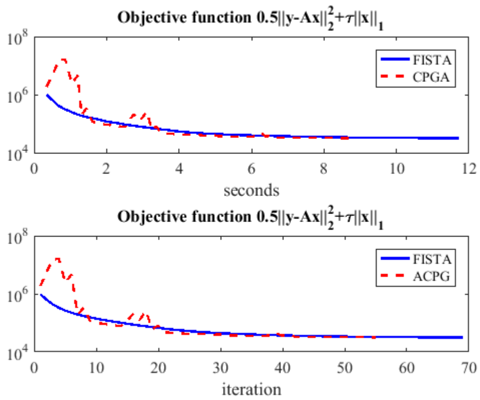

Fig. 12 Gold hill : Evaluation of objective function

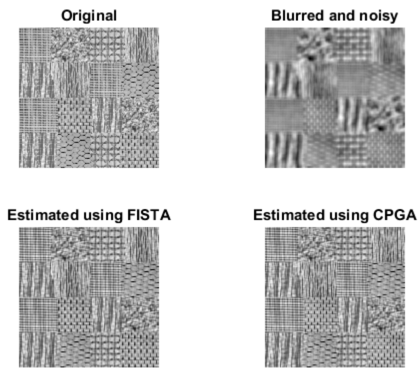

Fig. 13 Patches : Origin, blurred & noise and CPGA estimated images

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-

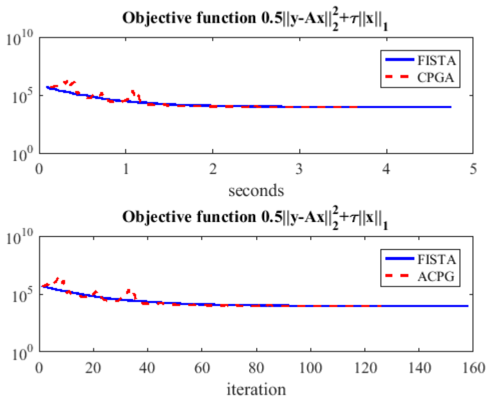

Fig. 14 Patches : Evaluation of objective function

thesholding algorithm for linear inverse problems, SIAM J. Imaging Sciences, vol. 2, no. 1, pp. 183-202, 2009.
3. 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.
4. M. Bertero and P. Boccacci, Introduction to inverse problems in imaging, IOP Publishing, Bristol, UK, 1998.
5. 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.
6. 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 Processing, vol. 7, pp. 319-335, 1998.
7. 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.
8. P. Combettes and V. Wajs, Signal recovery by proximal forwardbackward splitting, SIAM Journal on Multiscale Modeling & Simulation, vol. 4, pp. 1168-1200, 2005.
9. D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory, vol.52, no. 4, pp. 12891306, Apr. 2006.
10. A. El Mouatasim, R. Ellaia and J.E. Souza de Cursi, Random perturbation of variable metric method for unconstraint nonsmooth global optimization, International Journal of Applied Mathematics and Computer Science, vol.16, no.4, pp:463-474, 2006.
11. El Mouatasim, A. and Wakrim, M. (2015) Control Subgradient Algorithm for Image Regularization. Journal of Signal, Image and Video Processing (SViP), 9, 275-283.
12. M. Figueiredo and R. Nowak, An EM algorithm for wavelet-based image restoration, IEEE Trans. Im. Proc., vol. 12, pp. 906-916, 2003.
13. P. C. Hansen, J. G. Nagy, and D. P. OĹeary, Deblurring Images: Matrices, Spectra, and Filtering, Fundam. Algorithms 3, SIAM, Philadelphia, 2006.
14. A. Jain, Fundamentals of digital image processing, Prentice Hall, Englewood Cliffs, 1989.
15. C. Lemaréchal, An extension of Davidon methods to nondifferentiable problems, Mathematical Programming Study, Vol. 3, pp. 95-109, 1975.
16. 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.
17. T. Poggio, V. Torre, and C. Koch, Computational vision and regularization theory, Nature, vol. 317, pp. 314-319, 1985.
18. G. Sluder and D.E. Wolf, Digital microscopy, 3rd ed. New York: Academic, 2007.
19. S. Tao, D. Boley and S. Zhang. Local linear convergence of ISTA and FISTA on the Lasso problem. SIAM Journal of Optimization, 26(1), 313–336, 2016.
20. D. Terzopoulos, Regularization of inverse visual problems involving discontinuities, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, pp. 413-424, 1986.
21. S. Twomey, Introduction to the mathematics of inversion in remote sensing, New York: Dover, 2002.
22. L.I. Rudin, S.J. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D, 60:259-268, 1992.
23. C. Vogel, Computational methods for inverse problems, Philadelphia, PA: SIAM, 2002.
24. 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.
25. S. Wright, R. Nowak, M. Figueiredo, Sparse reconstruction by separable approximation, IEEE Transactions on Signal Processing, Vol.57, pp. 2479-2493, 2009.
26. 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.