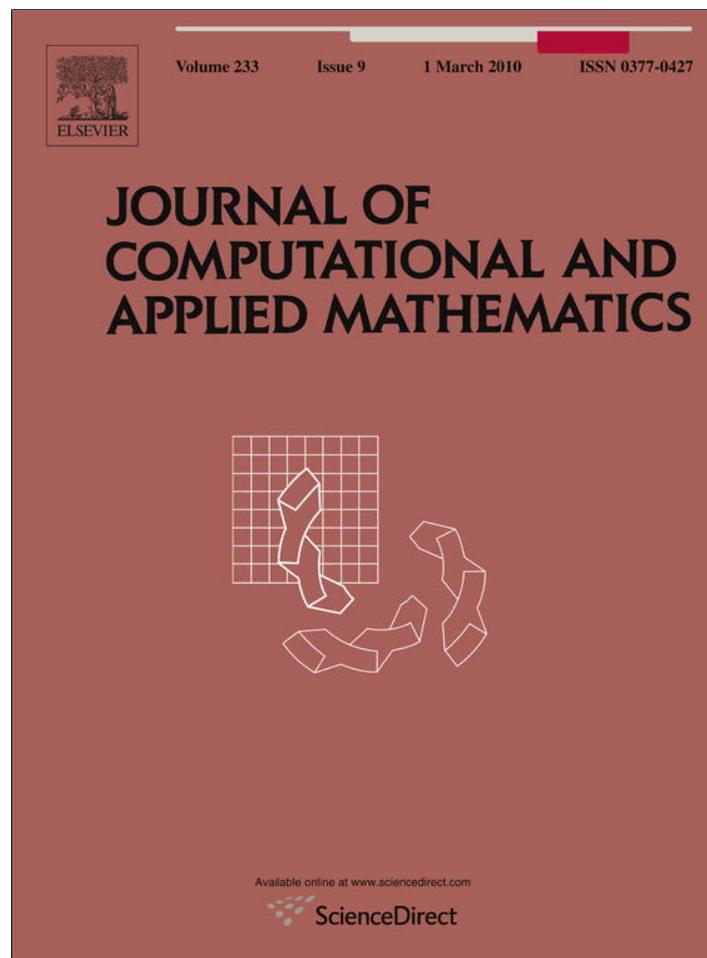


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

A reduced Newton method for constrained linear least-squares problems

Benedetta Morini^{a,*}, Margherita Porcelli^b, Raymond H. Chan^c^a *Dipartimento di Energetica "S. Stecco", Università di Firenze, via C. Lombroso 6/17, 50134 Firenze, Italie*^b *Dipartimento di Matematica "Ulisse Dini", Università di Firenze, Viale Morgagni 67/a, 50134 Firenze, Italie*^c *Department of Mathematics, The Chinese University of Hong Kong, Shatin, NT, Hong Kong*

ARTICLE INFO

Article history:

Received 11 November 2008

Received in revised form 5 October 2009

Keywords:

Bound-constrained linear least-squares problems

Image processing

Newton method

Active set strategy

ABSTRACT

We propose an iterative method that solves constrained linear least-squares problems by formulating them as nonlinear systems of equations and applying the Newton scheme. The method reduces the size of the linear system to be solved at each iteration by considering only a subset of the unknown variables. Hence the linear system can be solved more efficiently. We prove that the method is locally quadratic convergent. Applications to image deblurring problems show that our method gives better restored images than those obtained by projecting or scaling the solution into the dynamic range.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Consider solving the constrained least-squares problem:

$$\min_{l \leq x \leq u} q(x) = \min_{l \leq x \leq u} \left\{ \frac{1}{2} \|Ax - b\|^2 + \frac{1}{2} \lambda^2 \|Bx\|^2 \right\}, \quad (1)$$

where $\|\cdot\|$ is the 2-norm, $A, B \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\lambda \in \mathbb{R}$, $l \in (\mathbb{R} \cup \{-\infty\})^n$, $u \in (\mathbb{R} \cup \{+\infty\})^n$ are given, and $m \geq n$. The constraint $l \leq x \leq u$ is to be interpreted entry-wise, and we will denote the i th entry of any vector v by v_i . We assume that $A^T A + \lambda^2 B^T B$ is positive definite. Thus, problem (1) is a strictly convex problem and has a unique solution x^* for any vector b , [1, p. 195].

Problem (1) arises in many practical applications such as contact problems, control problems, and intensity modulated radiotherapy problems; see for example [1,2]. One major application is in image deblurring using Tikhonov approach; see [3, p.358] and [4, p.255]. In this case, A will be the blurring operator, b the observed image, B the regularization operator, λ^2 the regularization parameter, and x the restored image to be sought. The constraints represent the dynamic range of the image. For 8-bit gray-scale images, we have $l_i = 0$ and $u_i = 255$ for all $1 \leq i \leq n$. In practice, one may just solve the unconstrained problem and project or scale the solution into the dynamic range. In principle, this should give an inferior solution as the solution will not be a minimizer of the constrained problem. In fact, we will demonstrate in our numerical examples that the constrained problem (1) indeed gives substantially better restored images.

The Tikhonov approach, though a preferred method by engineers because of its speed, can oversmooth images, especially when B is a high order differential operator. An alternative is to use the total-variation (TV) regularization proposed in [5]. However, deblurring images under TV regularization can be slow. Recently, in [6], the authors proposed to separate the deblurring problem with TV regularization into two parts: (i) a TV denoising part and (ii) a Tikhonov deblurring part where our method can be applied.

* Corresponding author. Tel.: +39 0554224137.

E-mail addresses: benedetta.morini@unifi.it (B. Morini), margherita.porcelli@math.unifi.it (M. Porcelli), rchan@math.cuhk.edu.hk (R.H. Chan).

There are many methods for finding the constrained minimizer of (1); see for examples [7–14]. In this paper, we propose a Reduced Newton Method for solving (1). Our idea comes from a recently proposed Interior Newton-like Method [9] which is based on formulating the Karush–Kuhn–Tucker conditions for (1) as a system of nonlinear equations. The nonlinear system is then solved by a Newton-like method. The inner linear system in each Newton iteration may be ill-conditioned when the iterate approaches a solution on the boundary of the feasible set. Thus in [9], a scaling matrix was introduced to precondition the system.

Here instead of preconditioning the inner linear system, we consider its subsystem which corresponds only to components of the iterate that are not close to the boundary. The advantage is that the subsystem is smaller and less ill-conditioned. Moreover, in some applications, such as deblurring astronomical images, where most part of the image are black, the subsystem is small and much easier to solve than the original system. We are able to prove the local quadratic convergence of the algorithm, and the numerical examples in image deblurring show that it speeds up the solution of these problems as it requires less number of Newton's steps and inner linear iterations than the interior Newton-like method in [9].

The outline of the paper is as follows. In Section 2, we recall the method in [9]. Our algorithm is introduced in Section 3 and its proof of local convergence is given in Section 4. In Section 5, we discuss the globalization strategies. Then in Section 6, we illustrate the efficiency of our algorithm for image deblurring problems. Conclusions are given in Section 7.

2. Interior Newton-like method

In this section, we briefly recall the method in [9]. To simplify the notations and without loss of generality, we assume that for all i , $-\infty < l_i$ and $u_i < \infty$, which is indeed the case in image deblurring. Let

$$g(x) := \nabla q(x) = A^T(Ax - b) + \lambda^2 B^T Bx.$$

If x^* solves problem (1), then it satisfies the Karush–Kuhn–Tucker conditions:

$$g_i(x^*) \begin{cases} = 0 & \text{if } l_i < x_i^* < u_i, \\ \leq 0 & \text{if } x_i^* = u_i, \\ \geq 0 & \text{if } x_i^* = l_i, \end{cases} \quad i = 1, \dots, n.$$

Equivalently x^* solves a system of nonlinear equations [15]:

$$D(x)g(x) = 0, \tag{2}$$

where $D(x) = \text{diag}(d_1(x), \dots, d_n(x))$ has entries

$$d_i(x) = \begin{cases} u_i - x_i & \text{if } g_i(x) < 0, \\ x_i - l_i & \text{if } g_i(x) > 0, \\ \min\{x_i - l_i, u_i - x_i\} & \text{if } g_i(x) = 0. \end{cases} \tag{3}$$

Let x^k be a strictly feasible vector, i.e. $l < x^k < u$. If the Newton method is used to solve (2), then the Newton equation takes the form

$$[D(x^k)(A^T A + \lambda^2 B^T B) + E(x^k)]p^k = -D(x^k)g(x^k), \tag{4}$$

where the coefficient matrix is obtained by formal application of the product rule. In particular, $E(x) = \text{diag}(e_1(x), \dots, e_n(x))$ where $e_i(x) = g_i(x) \frac{\partial}{\partial x_i} d_i(x)$, $i = 1, \dots, n$. We note that the partial derivative of $d_i(x)$ may not exist everywhere. However, one can still develop quadratic convergent methods to the solution x^* by following [16] and letting

$$e_i(x) = \begin{cases} |g_i(x)| & \text{if } |g_i(x)| < \min\{x_i - l_i, u_i - x_i\}^2 \\ & \text{or } \min\{x_i - l_i, u_i - x_i\} < |g_i(x)|^2, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Let us partition the index set $\{1, \dots, n\}$ into the sets

$$\mathcal{C}_I = \{i \in \{1, \dots, n\} : l_i < x_i^* < u_i\}, \tag{6}$$

$$\mathcal{C}_A = \{i \in \{1, \dots, n\} : x_i^* = l_i \text{ and } g_i^* > 0, \text{ or } x_i^* = u_i \text{ and } g_i^* < 0\}, \tag{7}$$

$$\mathcal{C}_D = \{i \in \{1, \dots, n\} : x_i^* = l_i \text{ or } x_i^* = u_i \text{ and } g_i^* = 0\}. \tag{8}$$

The set \mathcal{C}_I represents the indices of the inactive components of x^* , while the sets \mathcal{C}_A and \mathcal{C}_D contain the indices of the active components with and without strict complementarity, respectively.

The Newton equation (4) may be difficult to solve when x^k approaches a solution x^* on the boundary. In fact, if $i \in \mathcal{C}_D$ and $\{x^k\}$ is such that $\lim_{k \rightarrow \infty} x^k = x^*$, then we may have $\lim_{k \rightarrow \infty} d_i(x^k) = 0$. Since $\lim_{k \rightarrow \infty} e_i(x^k) = 0$, the i th row of $(D(x^k)(A^T A + \lambda^2 B^T B) + E(x^k))$ may tend to zero, i.e.

$$\lim_{k \rightarrow \infty} \|(D(x^k)(A^T A + \lambda^2 B^T B) + E(x^k))^{-1}\| = \infty.$$

Since $D(x)$ is invertible in the interior of the box $[l, u]$, one may think that one could, instead of solving (4), solve $M(x^k)p^k = -g(x^k)$ where

$$M(x) = A^T A + \lambda^2 B^T B + D(x)^{-1} E(x). \tag{9}$$

However, for $i \in \mathcal{C}_A$, we have $\lim_{k \rightarrow \infty} d_i(x^k) = 0$ while $\lim_{k \rightarrow \infty} e_i(x^k) \neq 0$, i.e. $\lim_{k \rightarrow \infty} \|M(x^k)\| = \infty$.

To overcome these pitfalls, the Newton equation (4) is restated in [9] as

$$[S(x^k)M(x^k)S(x^k)]\tilde{p}^k = -S(x^k)g(x^k), \tag{10}$$

where $\tilde{p}^k = S(x^k)^{-1}p^k$, $S(x) = W(x)^{\frac{1}{2}}D(x)^{\frac{1}{2}}$, and

$$W(x) = \text{diag} \left(\frac{1}{d_1(x) + e_1(x)}, \dots, \frac{1}{d_n(x) + e_n(x)} \right).$$

It is proved in [9] that the scaling matrix $S(x)$ is invertible for any $l < x < u$ and $\|S(x)\| \leq 1$. Moreover, $S(x)M(x)S(x)$ is symmetric positive definite for $l < x < u$, and its inverse is uniformly bounded.

Strict feasibility of the iterates is guaranteed by setting

$$x_{k+1} = x_k + \max\{\sigma, 1 - \|P(x^k + p^k) - x^k\|\} (P(x^k + p^k) - x^k), \tag{11}$$

where $\sigma \in (0, 1)$ and $P(v) = \max\{l, \min\{v, u\}\}$ is the projection map onto the box $[l, u]$.

3. A reduced Newton method

In this section we present a new iterative procedure for the solution of problem (1). We still apply the Newton method to the nonlinear system (2) but we use a different way to form the Newton equation. We saw in the last section that the drawback of solving $M(x^k)p^k = -g(x^k)$ directly for p^k is that, if $\lim_{k \rightarrow \infty} x^k = x^*$, then $\lim_{k \rightarrow \infty} e_i(x^k)/d_i(x^k) = \infty$ for $i \in \mathcal{C}_A$, and hence $\lim_{k \rightarrow \infty} \|M(x^k)\| = \infty$. Here we present one way to overcome the drawback. It is an alternative to (10) and can reduce the dimension of the system. Our aim is to exclude the components p_i^k corresponding to the active constraints at x^* in solving the system. This leads to the idea of combining the solution of the system $M(x^k)p^k = -g(x^k)$ with a strategy for identifying the active components of the current iterate. We describe our algorithm below.

We will use the following notations. For any function f , the notation f^k is used to denote $f(x^k)$. For any $v \in \mathbb{R}^n$ and $\mathcal{K} \subset \{1, \dots, n\}$, we write either $v_{\mathcal{K}}$ or $(v)_{\mathcal{K}}$ for the subvector of v having components $v_i, i \in \mathcal{K}$. Further, if $V = (v_{ij}) \in \mathbb{R}^{n \times n}$, we denote either by $V_{\mathcal{K}, \mathcal{L}}$ or $(V)_{\mathcal{K}, \mathcal{L}}$ the submatrix of V with elements $v_{ij}, i \in \mathcal{K}, j \in \mathcal{L}$.

Algorithm 3.1 (Reduced Newton Method).

1. *Initialization:*
Choose $l < x^0 < u, \delta > 0, \sigma \in (0, 1)$. Set $k = 0$.
2. *Termination criteria:*
If x^k is a stationary point of (1) i.e. it solves (2): stop.
3. *Identification of the active set:*
Set $g^k = \nabla q(x^k) = A^T(Ax^k - b) + \lambda^2 B^T Bx^k$,
 $\delta_k = \min\{\delta, \sqrt{\|P(x^k - g^k) - x^k\|}\}$,
 $\mathcal{A}_k = \{i \in \{1, \dots, n\} : x_i^k - l_i \leq \delta_k \text{ or } u_i - x_i^k \leq \delta_k\}, \mathcal{J}_k = \{1, \dots, n\} \setminus \mathcal{A}_k$.
4. *Search direction calculation:*
Compute a search direction $p^k \in \mathbb{R}^n$ such that

$$p_i^k = \begin{cases} u_i - x_i^k & \text{if } u_i - x_i^k \leq \delta_k, \\ l_i - x_i^k & \text{if } x_i^k - l_i \leq \delta_k, \end{cases} \quad \text{for } i \in \mathcal{A}_k,$$

and $(p^k)_{\mathcal{J}_k}$ solves

$$(M^k)_{\mathcal{J}_k \mathcal{J}_k} (p^k)_{\mathcal{J}_k} = -(g^k)_{\mathcal{J}_k} - (M^k)_{\mathcal{J}_k \mathcal{A}_k} (p^k)_{\mathcal{A}_k} \tag{12}$$

where the matrix $M^k = M(x^k)$ is given in (9).

5. *Enforce strict feasibility:*

Set

$$\hat{p}^k = \max\{\sigma, 1 - \|P(x^k + p^k) - x^k\|\} (P(x^k + p^k) - x^k). \tag{13}$$

6. *Form the new iterate:*

Set $x^{k+1} = x^k + \hat{p}^k$.

In each iteration of the Reduced Newton method the identification of the active set is performed by checking the closeness of x^k to the boundary. The scalar δ is assumed to be sufficiently small such that $\delta < \min_i |u_i - l_i|/2$. This is to avoid ambiguity in the definition of \mathcal{A}_k ; see [17]. After the identification of the active set, the linear system (12) restricted to the inactive components of x^k is solved. Clearly, if $\mathcal{A}_k \neq \emptyset$ then the size of (12) is smaller than that of (10) and by Cauchy interlace theorem [18, p. 396] the coefficient matrix $(M^k)_{\mathcal{I}_k \mathcal{I}_k}$ is better conditioned than M^k since the former is a submatrix of the latter. Then, in Steps 5–6 we force x^{k+1} to be in the interior of the box $[l, u]$. This way the inverse of D is computable at the new iterate and the matrix M can be formed at x^{k+1} .

We will prove the following three theoretical results about the Reduced Newton method in the next section. First, in a neighborhood of the solution x^* , the set \mathcal{A}_k coincides with the set of active constraints at x^* . Second, the matrix $(M^k)_{\mathcal{I}_k \mathcal{I}_k}$ and its inverse are bounded above in a neighborhood of the solution x^* of (1). Third, starting from an initial guess sufficiently near to x^* , $\{x_k\}$ converges to x^* quadratically. The local convergence behavior of the method is identical to that shown in [9]. However our method of proof is different from [9] as here the analysis on $(M^k)_{\mathcal{I}_k \mathcal{I}_k}$ and its inverse and the provided reduction in the distance to x^* is based on the information given by the active set identification strategy.

Finally, it is worth noting that in applications such as astronomical imaging, comparatively fewer pixels are nonzero. Thus, eventually the system (12) is small and requires fewer iterations than if we are solving the full system (10).

4. Local convergence analysis

In this section, we prove the local quadratic convergence of our Reduced Newton Method. For any $\rho > 0$, we let $\mathbf{B}_\rho(x^*)$ be the ball centered at x^* with radius ρ .

Lemma 4.1. *Let x^* be the solution of (1). Then there exist positive constants ρ_1 and α such that for all $x^k \in \mathbf{B}_{\rho_1}(x^*)$ we have*

$$\|x^k - x^*\| \leq \alpha \|P(x^k - g^k) - x^k\|. \tag{14}$$

Proof. Since the Hessian matrix $\nabla^2 q(x^*) = A^T A + \lambda^2 B^T B$ is positive definite, x^* satisfies the strong second order sufficiency condition for (1). Then [19, Theorem 3.7] shows that there exists a constant $\alpha > 0$ such that (14) holds for all x in a sufficiently small neighborhood of x^* . \square

Let us define the active constraints set at x^* as

$$\mathcal{A}_* = \{i \in \{1, \dots, n\} : x_i^* \in [l_i, u_i]\},$$

and denote by \mathcal{I}_* the complement of \mathcal{A}_* in $\{1, \dots, n\}$, i.e. $\mathcal{I}_* = \{1, \dots, n\} \setminus \mathcal{A}_*$. The following lemma shows that if x^k is sufficiently close to x^* then the set \mathcal{A}_k is the same as \mathcal{A}_* .

Lemma 4.2. *Let x^* be the solution of (1). Then there exists a positive constant ρ_2 such that if $x^k \in \mathbf{B}_{\rho_2}(x^*)$ then $\mathcal{A}_k = \mathcal{A}_*$.*

Proof. The proof follows [17, Lemma 9.4, Lemma 9.7] closely. First we show that $\mathcal{A}_k \subseteq \mathcal{A}_*$. Define

$$\nu = \min\{\min\{x_i^* - l_i, u_i - x_i^*\} \mid i \in \mathcal{I}_*\} > 0, \tag{15}$$

i.e., ν is the smallest distance of the inactive components x_i^* to the boundary of the feasible set $[l, u]$. Since $\|P(x^k - g^k) - x^k\|$ is continuous (see [20, p. 450]), there exists $\rho_2 \leq \nu/4$ sufficiently small so that if $x^k \in \mathbf{B}_{\rho_2}(x^*)$ then $\sqrt{\|P(x^k - g^k) - x^k\|} \leq \nu/4$.

Let $x^k \in \mathbf{B}_{\rho_2}(x^*)$ and $i \in \mathcal{A}_k$. Then $x_i^k - l_i \leq \delta_k$ or $u_i - x_i^k \leq \delta_k$. If $x_i^k - l_i \leq \delta_k$, then the strict feasibility of x^k yields

$$|x_i^k - l_i| = x_i^k - l_i \leq \delta_k \leq \sqrt{\|P(x^k - g^k) - x^k\|} \leq \frac{\nu}{4}.$$

Then by noting $|x_i^k - x_i^*| \leq \|x^k - x^*\| \leq \nu/4$, we obtain

$$|x_i^* - l_i| \leq |x_i^* - x_i^k| + |x_i^k - l_i| \leq \frac{\nu}{2},$$

i.e., we have $i \in \mathcal{A}_*$ in view of (15). The same result is obtained if $u_i - x_i^k \leq \delta_k$.

Now we show that $\mathcal{A}_* \subseteq \mathcal{A}_k$. Reduce ρ_2 if needed so that $\rho_2 < \rho_1$, where ρ_1 is defined in Lemma 4.1. Then if $x^k \in \mathbf{B}_{\rho_2}(x^*)$ we get

$$|x_i^k - x_i^*| \leq \alpha \|P(x^k - g^k) - x^k\|, \tag{16}$$

for all $i \in \{1, \dots, n\}$. Let $i \in \mathcal{A}_*$ be any fixed index. Then $x_i^* = l_i$ or $x_i^* = u_i$. If $x_i^* = l_i$, then by (16) we have

$$x_i^k - l_i = |x_i^k - x_i^*| \leq \alpha \|P(x^k - g^k) - x^k\|.$$

Since $\|P(x^k - g^k) - x^k\| \rightarrow 0$ for $x^k \rightarrow x^*$ and $\delta_k = O(\sqrt{\|P(x^k - g^k) - x^k\|})$ in view of the definition of δ_k , we have $\alpha \|P(x^k - g^k) - x^k\| \leq \delta_k$ for all x^k sufficiently close to x^* . Therefore we obtain $x_i^k - l_i \leq \delta_k$, i.e. $i \in \mathcal{A}_k$. The case where $x_i^* = u_i$ can be studied in an analogous way and therefore we can conclude that $\mathcal{A}_* = \mathcal{A}_k$ for all $x^k \in \mathbf{B}_{\rho_2}(x^*)$. \square

In the next two lemmas we explore the properties of the matrices $D(x)M(x)$ and $M(x)$ both at and near x^* when restricted onto the inactive set.

Lemma 4.3. *Let x^* be the solution of (1). The matrices $(D(x^*)M(x^*))_{\mathcal{I}_* \mathcal{I}_*}$ and $M(x^*)_{\mathcal{I}_* \mathcal{I}_*}$ are nonsingular.*

Proof. It can be seen easily that

$$\begin{aligned} (D(x^*)M(x^*))_{\mathcal{I}_* \mathcal{I}_*} &= (D(x^*)(A^T A + \lambda^2 B^T B) + E(x^*))_{\mathcal{I}_* \mathcal{I}_*} \\ &= D(x^*)_{\mathcal{I}_* \mathcal{I}_*} (A^T A + \lambda^2 B^T B)_{\mathcal{I}_* \mathcal{I}_*} + E(x^*)_{\mathcal{I}_* \mathcal{I}_*}. \end{aligned}$$

If $i \in \mathcal{I}_*$, then $d_i(x^*) \neq 0$ and $e_i(x^*) = 0$. Hence,

$$(D(x^*)M(x^*))_{\mathcal{I}_* \mathcal{I}_*} = D(x^*)_{\mathcal{I}_* \mathcal{I}_*} (A^T A + \lambda^2 B^T B)_{\mathcal{I}_* \mathcal{I}_*},$$

and $(D(x^*)M(x^*))_{\mathcal{I}_* \mathcal{I}_*}$ is nonsingular because $D(x^*)_{\mathcal{I}_* \mathcal{I}_*}$ is nonsingular and $(A^T A + \lambda^2 B^T B)_{\mathcal{I}_* \mathcal{I}_*}$ is positive definite.

Similarly we get that $M(x^*)_{\mathcal{I}_* \mathcal{I}_*} = (A^T A + \lambda^2 B^T B)_{\mathcal{I}_* \mathcal{I}_*}$, i.e. $M(x^*)_{\mathcal{I}_* \mathcal{I}_*}$ is positive definite. \square

Lemma 4.4. *Let x^* be the solution of (1). Then there exist positive constants ϵ , κ_1 , κ_2 and κ_3 such that if $x^k \in \mathbf{B}_\epsilon(x^*)$ then the matrices $(D^k M^k)_{\mathcal{I}_k \mathcal{I}_k}$ and $(M^k)_{\mathcal{I}_k \mathcal{I}_k}$ are nonsingular, and*

$$\|(M^k)_{\mathcal{I}_k \mathcal{I}_k}^{-1}\| \leq \kappa_1, \tag{17}$$

$$\|(D^k)_{\mathcal{I}_k \mathcal{I}_k}^{-1}\| \leq \kappa_2, \tag{18}$$

$$\|(D^k M^k)_{\mathcal{I}_k \mathcal{I}_k}^{-1}\| \leq \kappa_3. \tag{19}$$

Proof. Let $x^k \in \mathbf{B}_{\rho_2}(x^*)$ where ρ_2 is defined in Lemma 4.2. Then $\mathcal{A}_k = \mathcal{A}_*$ and $\mathcal{I}_k = \mathcal{I}_*$. Hence

$$(D^k M^k)_{\mathcal{I}_k \mathcal{I}_k} = (D^k M^k)_{\mathcal{I}_* \mathcal{I}_*} = (D^k)_{\mathcal{I}_* \mathcal{I}_*} (A^T A + \lambda^2 B^T B)_{\mathcal{I}_* \mathcal{I}_*} + (E^k)_{\mathcal{I}_* \mathcal{I}_*}. \tag{20}$$

Note that the matrix $(E^k)_{\mathcal{I}_* \mathcal{I}_*}$ is positive semidefinite and $(D^k)_{\mathcal{I}_* \mathcal{I}_*}$ is positive definite; thus $(D^k M^k)_{\mathcal{I}_k \mathcal{I}_k}$ is nonsingular. Concerning $(M^k)_{\mathcal{I}_k \mathcal{I}_k}$, it is positive definite since x^k is strictly feasible.

Now we prove (17). Let v be an arbitrary vector in \mathbb{R}^n . Then we have

$$\begin{aligned} \|v\| \|(M^k)_{\mathcal{I}_* \mathcal{I}_*} v\| &\geq \|v^T (M^k)_{\mathcal{I}_* \mathcal{I}_*} v\| \\ &= v^T (A^T A + \lambda^2 B^T B)_{\mathcal{I}_* \mathcal{I}_*} v + v^T ((D^k)^{-1} E^k)_{\mathcal{I}_* \mathcal{I}_*} v \\ &\geq v^T (A^T A + \lambda^2 B^T B)_{\mathcal{I}_* \mathcal{I}_*} v \geq \frac{\|v\|^2}{\|(A^T A + \lambda^2 B^T B)_{\mathcal{I}_* \mathcal{I}_*}^{-1}\|}. \end{aligned}$$

Thus $\|(M^k)_{\mathcal{I}_* \mathcal{I}_*} v\| \geq \|v\| / \|(A^T A + \lambda^2 B^T B)_{\mathcal{I}_* \mathcal{I}_*}^{-1}\|$, and the inequality (17) holds with $\kappa_1 = \|(A^T A + \lambda^2 B^T B)_{\mathcal{I}_* \mathcal{I}_*}^{-1}\|$.

To prove (18), let $\nu > 0$ be the scalar defined in (15) and let $\epsilon < \rho_2$. If $x^k \in \mathbf{B}_\epsilon(x^*)$ then by Lemma 4.2 we have that $\|x^k - x^*\| \leq \nu/4$. Take $i \in \mathcal{I}_*$. Note that $d_i(x^k) = x_i^k - l_i$ or $d_i(x^k) = u_i - x_i^k$. In the first case we have

$$|d_i(x^k)| = |x_i^* - l_i| - |x_i^k - x_i^*| > \nu - \frac{\nu}{4} = \frac{3\nu}{4}.$$

Proceeding analogously in the other case we have that $|d_i(x^k)|$ is bounded below for all $i \in \mathcal{I}_*$. Then there exists a constant κ_2 such that $\|(D^k)_{\mathcal{I}_k \mathcal{I}_k}^{-1}\| = \|(D^k)_{\mathcal{I}_* \mathcal{I}_*}^{-1}\| \leq \kappa_2$.

Finally, note that (20) implies

$$\|(D^k M^k)_{\mathcal{I}_k \mathcal{I}_k}^{-1}\| \leq \|(D^k)_{\mathcal{I}_k \mathcal{I}_k}^{-1}\| \|(M^k)_{\mathcal{I}_k \mathcal{I}_k}^{-1}\|.$$

Hence, using (17) and (18) we get (19). \square

The above lemma implies the boundedness of $\|(M^k)_{\mathcal{I}_k \mathcal{I}_k}\|$ when $x^k \in \mathbf{B}_\epsilon(x^*)$. In particular, by (18) and (20) it follows that

$$\|(M^k)_{\mathcal{I}_k \mathcal{I}_k}\| \leq k_4$$

for some scalar k_4 .

We next prove that the directional vector p^k in Step 4 of Algorithm 3.1 provides a quadratic reduction in the distance to x^* . We note however that the next iterate so formed may not satisfy the feasibility constraints.

Lemma 4.5. *Let x^* be the solution of (1). Then there exist positive constants ϵ and γ such that if $x^k \in \mathbf{B}_\epsilon(x^*)$, then*

$$\|x^k + p^k - x^*\| \leq \gamma \|x^k - x^*\|^2, \tag{21}$$

where the vector p^k is formed in Step 4 of Algorithm 3.1.

Proof. Let ϵ be the scalar defined in Lemma 4.4 and $x^k \in \mathbf{B}_\epsilon(x^*)$; thus $\mathcal{A}_k = \mathcal{A}_*$. Letting $\bar{x}^{k+1} = x^k + p^k$, we have

$$(D^k M^k)_{\mathcal{I}_k \mathcal{N}} (\bar{x}^{k+1} - x^k) = -(D^k g^k)_{\mathcal{I}_k}, \tag{22}$$

where $\mathcal{N} = \{1, \dots, n\}$. Subtracting the equality

$$(D^k M^k)_{\mathcal{I}_k \mathcal{N}} (x^* - x^k) = -(D(x^*)g(x^*))_{\mathcal{I}_k},$$

from (22), we obtain

$$(D^k M^k)_{\mathcal{I}_k \mathcal{N}} (\bar{x}^{k+1} - x^*) = r_{\mathcal{I}_k}^k, \tag{23}$$

with $r_{\mathcal{I}_k}^k$ defined as

$$r_{\mathcal{I}_k}^k = -(D^k M^k)_{\mathcal{I}_k \mathcal{N}} (x^* - x^k) - (D^k g^k)_{\mathcal{I}_k} + (D(x^*)g(x^*))_{\mathcal{I}_k}.$$

Since $\mathcal{A}_k = \mathcal{A}_*$, we have that $\bar{x}_i^{k+1} = x_i^k + p_i^k = x_i^*$ for all $i \in \mathcal{A}_k$ and $\|\bar{x}^{k+1} - x^*\| = \|(\bar{x}^{k+1} - x^*)_{\mathcal{I}_k}\|$. Also, by $\mathcal{N} = \mathcal{I}_k \cup \mathcal{A}_k$ we get

$$\begin{aligned} (D^k M^k)_{\mathcal{I}_k \mathcal{N}} (\bar{x}^{k+1} - x^*) &= (D^k M^k)_{\mathcal{I}_k \mathcal{I}_k} (\bar{x}^{k+1} - x^*)_{\mathcal{I}_k} + (D^k M^k)_{\mathcal{I}_k \mathcal{A}_k} (\bar{x}^{k+1} - x^*)_{\mathcal{A}_k} \\ &= (D^k M^k)_{\mathcal{I}_k \mathcal{I}_k} (\bar{x}^{k+1} - x^*)_{\mathcal{I}_k}, \end{aligned}$$

and (23) takes the form

$$(D^k M^k)_{\mathcal{I}_k \mathcal{I}_k} (\bar{x}^{k+1} - x^*)_{\mathcal{I}_k} = r_{\mathcal{I}_k}^k.$$

By Lemma 4.4, the matrix $(D^k M^k)_{\mathcal{I}_k \mathcal{I}_k}$ is nonsingular and hence (19) yields

$$\|\bar{x}^{k+1} - x^*\| \leq \kappa_3 \|r_{\mathcal{I}_k}^k\|. \tag{24}$$

Now we find an upper bound for $\|r_{\mathcal{I}_k}^k\|$. Let i be an index in \mathcal{I}_k and note that

$$\begin{aligned} r_i^k &= -d_i(x^k)(A^T A + \lambda^2 B^T B)_{i \mathcal{N}} (x^* - x^k) - e_i(x^k)(x_i^* - x_i^k) - d_i(x^k)g_i^k + d_i(x^*)g_i(x^*) + d_i(x^k)g_i(x^*) - d_i(x^k)g_i(x^*) \\ &= d_i(x^k)(g_i(x^*) - g_i(x^k) - (A^T A + \lambda^2 B^T B)_{i \mathcal{N}} (x^* - x^k)) + g_i(x^*)(d_i(x^*) - d_i(x^k)) - e_i(x^k)(x_i^* - x_i^k). \end{aligned}$$

Since

$$g_i(x^*) - g_i^k = (A^T A + \lambda^2 B^T B)_{i \mathcal{N}} (x^* - x^k), \tag{25}$$

and $g_i(x^*) = 0$, it follows $|r_i^k| = |e_i(x^k)(x_i^* - x_i^k)|$. Further, by (5), we get

$$\begin{aligned} |r_i^k| &\leq |e_i(x^k)| |(x^* - x^k)_i| \leq |g_i^k| |(x^* - x^k)_i| = |g_i(x^*) - g_i^k| |(x^* - x^k)_i| \\ &= |((A^T A + \lambda^2 B^T B)(x^k - x^*))_i| |(x^* - x^k)_i|. \end{aligned}$$

So, denoting the cardinality of the set \mathcal{I}_k by $\omega_{\mathcal{I}_k}$, we obtain

$$\|r_{\mathcal{I}_k}^k\| \leq \sqrt{\omega_{\mathcal{I}_k}} \|A^T A + \lambda^2 B^T B\| \|x^* - x^k\|^2. \tag{26}$$

Thus, by (24) and (26), the Lemma is proved by setting $\gamma = \sqrt{\omega_{\mathcal{I}_k}} \kappa_3 \|A^T A + \lambda^2 B^T B\|$. \square

Finally, we show that by enforcing the feasibility constraints using condition (13), the sequence $\{x^k\}$ still has quadratic convergence.

Theorem 4.1. *Let x^* be the solution of (1). If the initial guess x^0 is sufficiently close to x^* , then the sequence $\{x^k\}$ generated by Algorithm 3.1 converges quadratically to x^* .*

Proof. If x^0 is sufficiently close to x^* then the scalar $\max\{\sigma, 1 - \|P(x^0 + p^0) - x^0\|\}$ used in (13) is close to one. This fact is crucial for fast convergence.

Let x^0 be sufficiently close to x^* so that

$$\max\{\sigma, 1 - \|P(x^0 + p^0) - x^0\|\} = 1 - \|P(x^0 + p^0) - x^0\|, \tag{27}$$

and (21) is valid, i.e.

$$\|x^0 + p^0 - x^*\| \leq \gamma \|x^0 - x^*\|^2. \tag{28}$$

Then x^1 has the form

$$\begin{aligned} x^1 - x^* &= x^0 + \hat{p}^0 - x^* = x^0 + (1 - \|P(x^0 + p^0) - x^0\|)(P(x^0 + p^0) - x^0) - x^* \\ &= P(x^0 + p^0) - \|P(x^0 + p^0) - x^0\|(P(x^0 + p^0) - x^0) - x^*. \end{aligned}$$

Using the non-expansion property of the projection map and (28) we get

$$\begin{aligned} \|x^1 - x^*\| &\leq \|P(x^0 + p^0) - x^*\| + \|P(x^0 + p^0) - x^0\|^2 \\ &\leq \|x^0 + p^0 - x^*\| + \|x^0 + p^0 - x^0\|^2 \\ &\leq \gamma \|x^0 - x^*\|^2 + 2\|x^0 + p^0 - x^*\|^2 + 2\|x^0 - x^*\|^2 \\ &\leq \gamma \|x^0 - x^*\|^2 + 2\gamma \|x^0 - x^*\|^4 + 2\|x^0 - x^*\|^2. \end{aligned}$$

Then we have that $\|x^1 - x^*\| = O(\|x^0 - x^*\|^2)$. The proof can be completed using standard induction arguments. \square

5. Globalization strategies

In Algorithm 3.1, the components \hat{p}_i^k with $i \in \mathcal{A}_k$ force the corresponding components of x^k to get closer to the boundary. If the resulting step does not produce a reduction in the objective function q , it is necessary to change \mathcal{A}_k . A rapid change of such set can be obtained as follows. Let $p^{k,C}$ be the generalized Cauchy step; see equations (29) and (31) in [9]. We accept the step \hat{p}^k if it satisfies the condition

$$\frac{1}{2}(\hat{p}^k)^T M^k \hat{p}^k + (\hat{p}^k)^T g^k \geq \beta \left\{ \frac{1}{2}(\hat{p}^{k,C})^T M^k \hat{p}^{k,C} + (\hat{p}^{k,C})^T g^k \right\}, \tag{29}$$

for a fixed scalar $\beta \in (0, 1)$. Otherwise, we take the step $p^{k,C}$. The analysis conducted in [9] shows that each limit point of the sequence $\{x^k\}$ generated is a stationary point of (1). Since the problem is strictly convex, each limit point of the sequence $\{x^k\}$ is a global minimum of (1). Uniqueness of the solution to problem (1) implies that $\lim_{k \rightarrow \infty} x^k = x^*$.

6. Numerical experiments

In this section, we show the efficiency of our method by applying it to deblurring problems. In our tests, we use four 256-by-256 gray images shown in Fig. 1. The Satellite image is from the US Air Force Phillips Laboratory, the Church, Eagle and Bridge images are taken from [21]. The dimensions of the least-squares problem (1) are $m = n = 65\,536$. We choose these images because they have different numbers of pixels with values either close or equal to 0 or 255. Specifically, there are 89.81%, 22.54%, 12.71% and 11.75% active pixels in the true Satellite, Church, Eagle and Bridge images respectively. Due to these features, it is easier to illustrate the difference in the quality of the restored images before and after we enforce the constraints. We will also see that many constraints are active during the iterations of the Newton Reduced method, and hence the size of the system to be solved is considerably reduced.

In (1), we choose the blurring matrix $A \in \mathbb{R}^{n \times n}$ to be the out-of-focus blur with radius 3 and the regularization matrix $B \in \mathbb{R}^{n \times n}$ to be the gradient matrix. Hence $B^T B$ is the two-dimensional discrete Laplacian matrix. For both matrices, we employ the Neumann boundary conditions [22], which usually gives less artifacts at the boundary. The use of such boundary conditions means that $A^T A + \lambda^2 B^T B$ is a block-Toeplitz-plus-Hankel matrix with Toeplitz-plus-Hankel blocks. The observed image b is such that $b = Ax_{true} + \eta r$ where x_{true} is the true image, r is a random vector with entries distributed as standard normal, η is the level of noise. The constraints are such that $l_i = 0$ and $u_i = 255$ for $i = 1, \dots, n$.

The restored images were obtained in double precision using MATLAB 7.0 on an Intel Xeon (TM) 3.4 GHz, 1 GB RAM. Three levels of noise, $\eta = 1, 2, 3$, were tested. The procedures we compare numerically are the following:

1. The projection (P) method:
 - Solve the unconstrained problem

$$\min q(x) = \frac{1}{2} \|Ax - b\|_2^2 + \frac{1}{2} \lambda^2 \|Bx\|_2^2,$$
 i.e. solve $(A^T A + \lambda^2 B^T B)x = A^T b$.
 - Project the solution onto the box $[l, u]$.
 - Round the pixel values of the projected solution to integers.
2. The Interior Newton-like (IN) algorithm in [9] (see Section 2):
 - Apply the P method and let x_p be the computed solution.
 - Perturb x_p to form a strictly feasible vector x_p^{sf} .
 - Apply the Interior Newton-like method starting from $x_0 = x_p^{sf}$.
 - Round the pixel values of the solution to integers.
3. The Reduced Newton (RN) algorithm given by Algorithm 3.1:
 - Apply the P method and let x_p be the computed solution.
 - Perturb x_p to form a strictly feasible vector x_p^{sf} .
 - Apply Algorithm 3.1 starting from $x_0 = x_p^{sf}$.
 - Round the pixel values of the solution to integers.



Fig. 1. The true images.

The starting point for the IN and the RN methods needs to be strictly feasible. Therefore, the vector x_p^{sf} is formed by projecting the solution x_p of the P method onto the box $[w, 254w]$, with $w = (1, \dots, 1)^T \in \mathbb{R}^n$.

The linear systems for all three procedures are solved by the conjugate gradient (CG) method. Regarding the Reduced Newton method, the matrix–vector products required to solve (12) can be computed by exploiting the Toeplitz-like structure of M^k as follows. Letting $I \in \mathbb{R}^{n \times n}$ be the identity matrix, $\mathcal{N} = \{1, \dots, n\}$, and using the notation of Section 3, it is easy to see that

$$(M^k)_{J_k J_k} = (I)_{J_k \mathcal{N}} M^k (I)_{\mathcal{N} J_k}, \quad (M^k)_{J_k \mathcal{A}_k} = (I)_{J_k \mathcal{N}} M^k (I)_{\mathcal{N} \mathcal{A}_k}.$$

Since M^k is a Toeplitz-plus-Hankel matrix with Toeplitz-plus-Hankel block, the multiplication can be done via fast cosine transform in $O(n \log n)$ operations; see [23,22].

In the tests, the parameter $\sigma = 0.9995$ is used in (11) and (13) while the parameter $\beta = 0.3$ is employed in (29). The parameter λ is chosen by trial and error such that it maximizes the Peak Signal to Noise Ratio (PSNR) [24] value of the reconstructed image. In the RN method, we set $\delta = 1$ in the active set strategy. For both the IN and RN methods, a successful termination at the point x^k is declared if any one of the following conditions is satisfied:

- (i) $q^{k-1} - q^k < \tau(1 + q^{k-1})$,
- (ii) $q^{k-1} - q^k < \sqrt{\tau}(1 + q^{k-1})$ and $\|x_{\text{int}}^k - x_{\text{int}}^{k-1}\|_\infty = 0$,
- (iii) $\|P(x^k + g^k) - x^k\| \leq n\sqrt{\tau}$.

Here x_{int} is obtained from x after rounding its entries to integers and $\tau = 10^{-8}$.

Table 1 displays the results when we have performed the tests three times for each η and computed the average numbers of nonlinear Newton iterations (N) and CG iterations (L) performed and the PSNR value of the recovered images. We also give the average percentage of pixels that are active in the images restored. We see that the percentages in the restored images by the P method are far away from those of the true Satellite and Church images and that the IN and RN methods overestimate the number of active pixels for the last two images. The table shows also that the PSNR values attained by the IN and RN methods are close. These values are 0.9–4.5 dB higher than those attained by the P algorithm for the Satellite, Church and Eagle images. For the Bridge image, the gain in the value of PSNR obtained by IN and RN methods over the P method is between 0.5–0.8 dB. Since an increase of 1 dB translates roughly to 10% reduction in the relative error, we see that

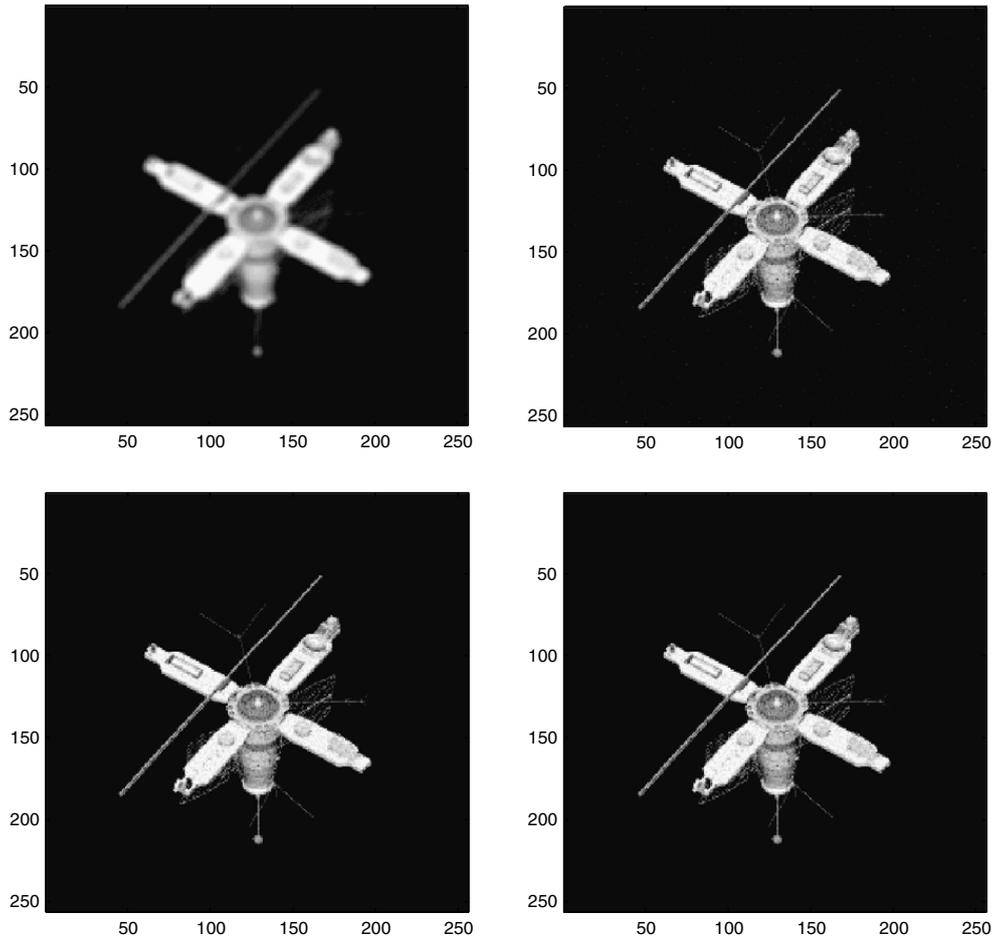


Fig. 2. Restoration results for $\eta = 1$: (top-left) Blurred image PSNR = 25.26 dB, (top-right) P algorithm PSNR = 29.64 dB, (bottom-left) IN algorithm PSNR = 34.50 dB, (bottom-right) RN algorithm PSNR = 34.47 dB.

Table 1

Number of linear and nonlinear iterations, PSNR value and active pixel percentage of the restored images.

Image	η	P method			IN method				RN method			
		L	PSNR	Active pixel (%)	N	L	PSNR	Active pixel (%)	N	L	PSNR	Active pixel (%)
Satellite	1	197.0	29.63	46.34	11.0	90.7	34.40	82.14	6.0	45.3	34.37	83.27
	2	145.0	26.81	45.93	11.0	87.0	31.45	81.67	6.0	45.3	31.43	82.05
	3	94.0	26.55	45.81	10.0	68.3	30.00	80.07	6.0	42.0	29.98	80.78
Church	1	102.0	29.69	15.41	20.3	220.7	31.14	21.44	5.0	45.3	31.13	24.54
	2	64.0	27.81	15.55	8.0	51.3	29.01	24.16	5.0	42.6	28.99	24.41
	3	48.0	26.95	15.57	6.7	35.0	28.03	22.85	5.0	38.3	28.01	23.87
Eagle	1	72.0	32.18	12.75	7.0	42.0	33.36	18.25	5.0	37.7	33.34	20.02
	2	47.0	30.63	12.86	6.0	29.7	31.70	18.48	5.0	31.0	31.68	20.02
	3	47.0	29.50	12.97	6.0	29.3	30.44	22.87	5.0	33.7	30.44	20.20
Bridge	1	184.0	26.69	14.78	48.0	872.7	27.51	18.21	5.0	51.0	27.52	22.23
	2	105.0	24.00	14.79	17.7	188.3	25.15	20.40	5.0	45.0	25.10	22.89
	3	75.0	23.43	15.07	8.3	58.0	23.95	22.60	5.0	42.3	23.95	23.09

projecting the image onto the dynamic range is not a good method and satisfying the constraints is necessary in getting a substantially better image.

From Table 1, we also see that the solution to the projection method x_p^{sf} is a good initial guess for both IN and RN methods, so both methods converge reasonably fast. However, the RN method converges faster than the IN method and the average numbers of linear and nonlinear iterations required by the RN method are actually fairly constant between different runs. Concerning the Church and Bridge images, the IN method converges very slowly for some values of η . This is because the Cauchy step is taken at most iterations in the IN method (see Section 5). On the contrary, in all our experiments with the RN method we observed that the active set settles down in the first few iterations and the steps generated provide a rapid decrease in the value of the objective function so that Cauchy steps are rarely needed. We give the restored Satellite and Church images in Figs. 2–5.

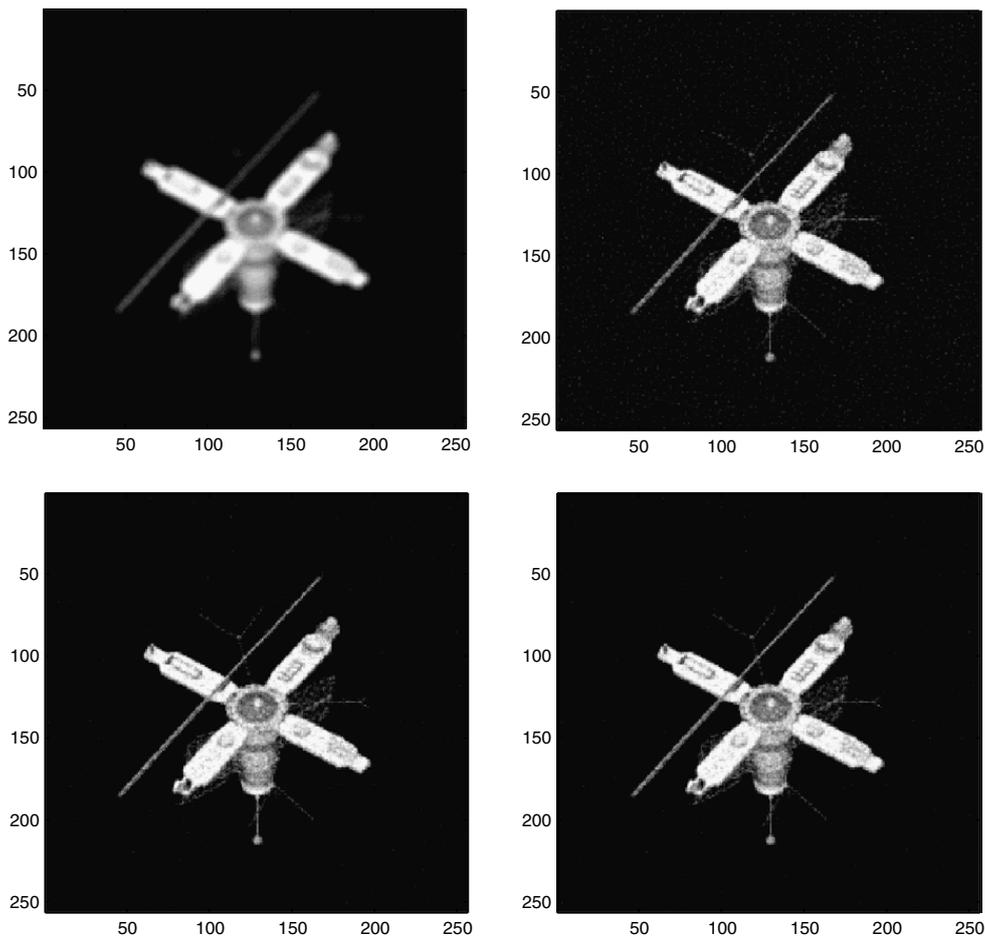


Fig. 3. Restoration results for $\eta = 3$: (top-left) Blurred image PSNR = 25.26 dB, (top-right) P algorithm PSNR = 26.53 dB, (bottom-left) IN algorithm PSNR = 29.97 dB, (bottom-right) RN algorithm PSNR = 29.94 dB.

Table 2
Elapsed time in seconds.

Image	η	Elapsed time		
		P method	IN method	RN method
Satellite	1	85.72	166.51	130.90
	2	61.34	130.02	102.84
	3	54.36	112.89	95.56
Church	1	48.98	184.16	81.24
	2	45.28	84.92	77.12
	3	43.65	71.48	73.54
Eagle	1	27.58	70.38	70.08
	2	42.23	68.69	68.20
	3	40.49	63.93	66.70
Bridge	1	73.81	551.14	102.68
	2	57.36	191.58	98.74
	3	47.58	85.21	79.08

In Table 2, we show the average time in seconds required by the three algorithms. Remarkably, the computational overhead of the IN and RN algorithms is low and our RN method is typically more efficient than the IN method. In this regard, note that the size of the linear systems arising in the RN method can be reduced significantly, e.g. in the Satellite image the size can be reduced by about 80%.

An alternative to the Tikhonov approach in (1) is to use the total-variation (TV) regularization proposed in [5]. Recently, in [6] the deblurring problem with TV regularization was separated into two parts: (i) a TV denoising part which can be solved by many methods, and (ii) a Tikhonov-like deblurring part that minimizes: $\frac{1}{2} \|Ax - b\|^2 + \lambda^2 \|x - b\|^2$. We note that by setting $\bar{x} = x - b$, the second part takes the form (1)

$$\min_{l-b \leq \bar{x} \leq u-b} \frac{1}{2} \|A\bar{x} - (b - Ab)\|^2 + \lambda^2 \|\bar{x}\|^2 \tag{30}$$

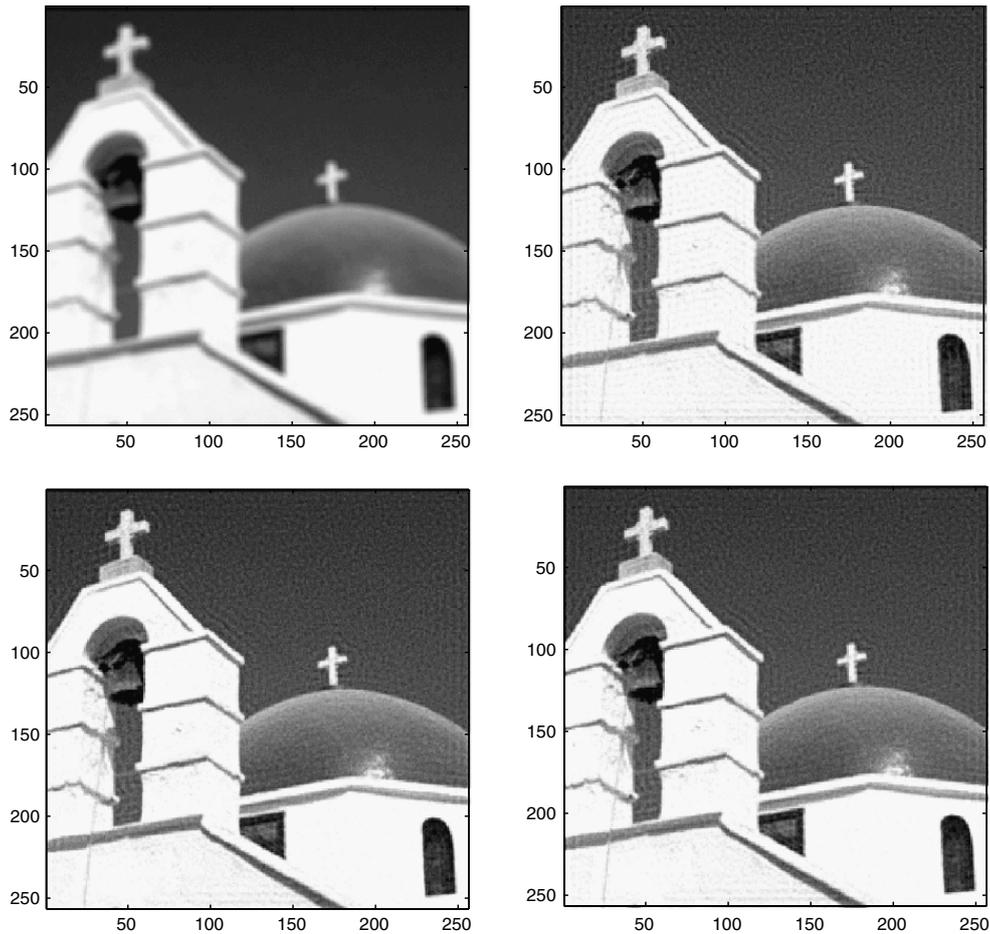


Fig. 4. Restoration results for $\eta = 1$: (top-left) Blurred image PSNR = 24.29 dB, (top-right) P algorithm PSNR = 29.70 dB, (bottom-left) IN algorithm PSNR = 31.17 dB, (bottom-right) RN algorithm PSNR = 31.12 dB.

Table 3
Results for problem (30).

Image	η	P method		RN method			
		L	PSNR	N	L	PSNR	Pixel (%)
Satellite	1	222.0	28.58	6.0	47.3	34.34	84.70
	2	121.0	26.86	6.0	43.7	31.22	83.24
	3	84.0	26.27	6.0	39.7	29.65	81.86
Church	1	101.0	29.41	5.0	41.3	30.51	25.07
	2	74.0	27.41	5.0	36.0	28.27	25.15
	3	39.0	26.67	5.0	27.0	27.17	23.31
Eagle	1	72.0	32.68	4.0	24.0	33.32	19.36
	2	38.0	31.25	4.0	18.0	31.63	19.31
	3	39.0	30.26	4.0	18.0	30.67	19.22
Bridge	1	170.0	26.28	5.0	47.7	27.05	21.54
	2	82.0	24.00	5.0	38.0	24.44	21.53
	3	76.0	22.99	5.0	40.0	23.43	23.16

which can readily be solved by our RN method. Table 3 shows the results obtained by our RN method by conducting the experiments in a way analogous to what we have done above. We see again from the table that our method gives better images than projection method, especially when the number of active pixels is high.

7. Conclusion

In this paper, we propose a Reduced Newton method that solves constrained linear least-squares problems and we prove its local quadratic convergence. We have applied it to deblurring problems and found that it gives better restored images than those obtained by projecting the images onto the constraint set. Moreover, for images with many black or white pixels, such as astronomical images, the inner linear iteration matrices from our method will be small and hence easier to solve.

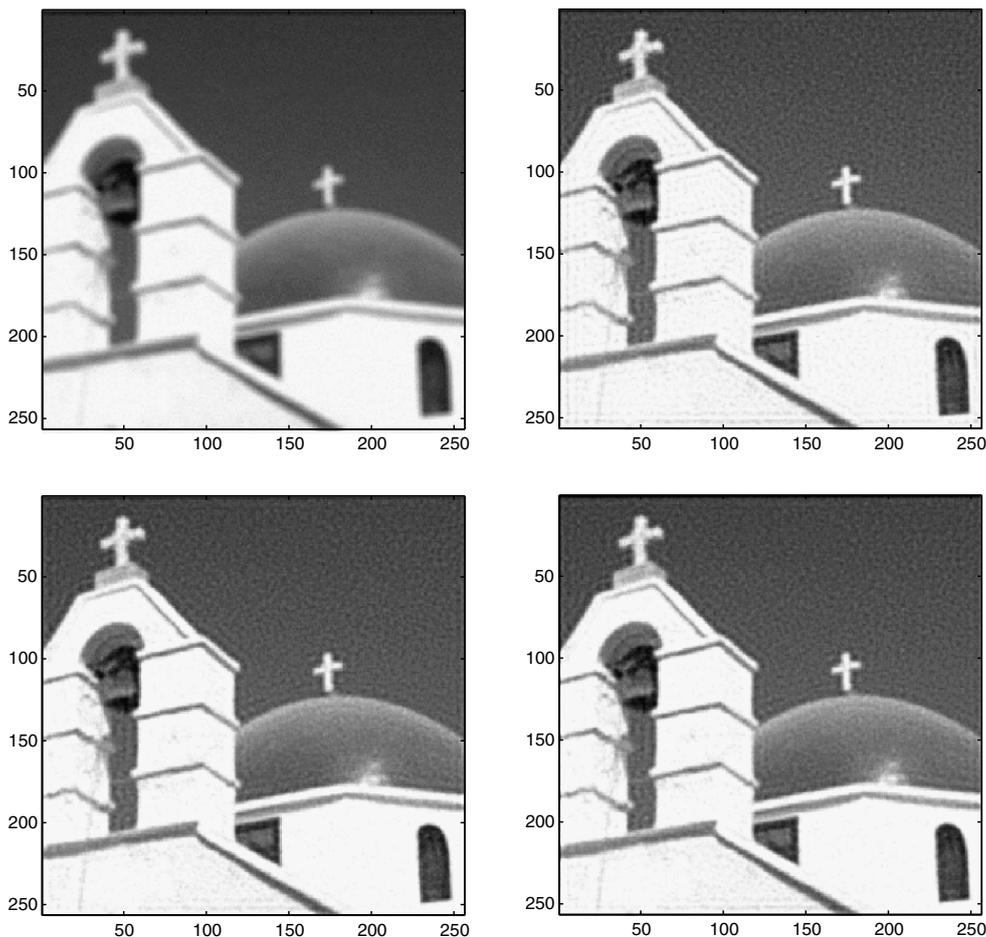


Fig. 5. Restoration results for $\eta = 3$: (top-left) Blurred image PSNR = 24.29 dB, (top-right) P algorithm PSNR = 26.92 dB, (bottom-left) IN algorithm PSNR = 28.01 dB, (bottom-right) RN algorithm PSNR = 28.00 dB.

Acknowledgements

The first and second authors' work were supported by "Progetti di Ricerca di Interesse Nazionale" PRIN 2007, MIUR, and GNCS/INDAM.

The third author's work was supported by HKRGC Grant CUHK 400508 and CUHK DAG 2060257.

References

- [1] A. Björck, Numerical Methods for Least Squares Problems, SIAM, 1996.
- [2] M. Ehrgott, I. Winz, Interactive decision support in radiation therapy treatment planning, *OR Spectrum* 30 (2008) 311–329.
- [3] R. Gonzalez, R. Woods, Digital Image Processing, 3rd ed., Prentice Hall, NJ, 2008.
- [4] M. Ng, Iterative Methods for Toeplitz Systems, Oxford University Press, Oxford, 2004.
- [5] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D* 60 (1992) 259–268.
- [6] Y. Huang, M. Ng, Y. Wen, A fast total variation minimization method for image restoration, *SIAM Journal on Multiscale Modeling and Simulation* 7 (2008) 774–795.
- [7] S. Bellavia, J. Gondzio, B. Morini, Regularization and preconditioning of KKT systems arising in nonnegative least-squares problems, *Numerical Linear Algebra with Applications* 16 (2009) 39–61.
- [8] S. Bellavia, J. Gondzio, B. Morini, Computational experience with numerical methods for nonnegative least-squares problems, Report 09/04, Dipartimento di Energetica S. Stecco, Università di Firenze, (submitted for publication).
- [9] S. Bellavia, M. Macconi, B. Morini, An interior Newton-like method for nonnegative least-squares problems with degenerate solution, *Numerical Linear Algebra with Applications* 13 (2006) 825–846.
- [10] D. Chen, R. Plemmons, Nonnegativity constraints in numerical analysis, in: A. Bultheel, R. Cools (Eds.) Conference Proceedings of the Symposium on the Birth of Numerical Analysis, Leuven Belgium, October 2007, World Scientific Press, 2009, (in press).
- [11] M. Friedlander, K. Hatz, Computing nonnegative tensor factorizations, *Optimization Methods and Software* 23 (2008) 631–647.
- [12] W. Hager, B. Mair, H. Zhang, An Affine-scaling interior-point CBB method for box-constrained optimization, *Mathematical Programming* 119 (2009) 1–32.
- [13] B. Kim, Numerical optimization methods for image restoration. Ph.D. Thesis, Stanford University, 2002.
- [14] S. Morigi, L. Reichel, F. Sgallari, F. Zama, An iterative method for linear discrete ill-posed problems with box constraints, *Journal of Computational and Applied Mathematics* 198 (2007) 505–520.
- [15] T. Coleman, Y. Li, An interior trust-region approach for nonlinear minimization subject to bounds, *SIAM Journal on Optimization* 6 (1996) 418–445.
- [16] M. Heinkenschloss, M. Ulbrich, S. Ulbrich, Superlinear and quadratic convergence of affine-scaling interior-point Newton methods for problems with simple bounds without strict complementarity assumptions, *Mathematical Programming* (1999).

- [17] C. Kanzow, An active set-type Newton method for constrained nonlinear systems, in: M.C. Ferris, O.L. Mangasarian, J.S. Pang (Eds.), *Complementarity: Applications, Algorithms and Extensions*, Kluwer Academic Publishers, 2001, pp. 179–200.
- [18] G. Golub, C. Van Loan, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, London, 1996.
- [19] C. Kanzow, A. Klug, On affine-scaling interior-point Newton methods for nonlinear minimization with bound constraints, *Computational Optimization and Applications* 35 (2006) 177–197.
- [20] A. Conn, N. Gould, P. Toint, *Trust-region methods*, in: *SMPS/SIAM Series on Optimization*, 2000.
- [21] Berkeley Segmentation Dataset: Images, <http://www.eecs.berkeley.edu/Research/Projects/CSvision/grouping/segbench/BSDS300/html/dataset/images.html>.
- [22] M. Ng, R. Chan, W. Tang, A fast algorithm for deblurring models with Neumann boundary conditions, *SIAM Journal on Scientific Computing* 21 (2000) 851–866.
- [23] R.H. Chan, X.Q. Jin, An introduction to iterative Toeplitz solvers, in: *Fundamentals of Algorithms*, SIAM, 2007.
- [24] A. Bovik, *Handbook of Image and Video Processing*, Academic Press, 2000.