Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

A Modified Newton Projection Method for $\ell 1$ -Regularized Least Squares Image Deblurring

(Article begins on next page)

28 April 2024

A modified Newton Projection method for ℓ_1-regularized Least Squares Image Deblurring

G. Landi*

May 19, 2014

Abstract

In recent years, ℓ_1-regularized least squares have become a popular approach to image deblurring due to the edge-preserving property of the ℓ_1-norm. In this paper, we consider the nonnegatively constrained quadratic program reformulation of the ℓ_1-regularized least squares problem and we propose to solve it by an efficient modified Newton projection method only requiring matrix-vector operations. This approach favors nonnegative solutions without explicitly imposing any constraints in the ℓ_1-regularized least squares problem. Experimental results on image deblurring test problems indicate that the developed approach performs well in comparison with state-of-the-art methods.

1 Introduction

Image deblurring is an important inverse problem arising in many image processing applications in which an unknown image $\mathbf{x} \in \mathbb{R}^n$ has to be estimated from noisy observations $\mathbf{b} \in \mathbb{R}^m$ defined by

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

where $\boldsymbol{\eta}$ is the unknown white Gaussian noise vector and $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the discretized linear blur operator. As usual, images are assumed to be represented as vectors, by storing the pixel values in some (e.g. lexicographical) order.

Least squares optimization with ℓ_1-regularization is now the state-of-the-art approach to image restoration. In fact, in imaging problems, the presence of edges often causes the prior distribution of the unknown image $\mathbf{x}$ to be not Gaussian and leads to outliers in the regularization term. In this context, the ℓ_1-norm, which is less sensitive to outliers, is advantageous compared to the ℓ_2-norm employed in Tikhonov-like regularization, and can be effectively used to promote sparse solutions [50, 48, 42, 11]. For these reason, the ℓ_1-regularized least squares problem

$$\min_{\mathbf{x}} \phi(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1, \quad \lambda > 0 \tag{2}$$

*Department of Mathematics, University of Bologna, e-mail: germana.landi@unibo.it

has recently attracted considerable attention both in the context of imaging inverse problems [26, 58, 41, 1] and of compressed sensing [14, 17, 21, 52, 15, 18, 19, 16]. (For simplicity of notation, hereafter $\| \cdot \|$ will denote the ℓ_2-norm $\| \cdot \|_2$.)

In this work, we focus on ℓ_1-regularized least squares problems (2) arising in image deblurring applications where the observation operator $\mathbf{A}$ describes spatially invariant blur [6]. Under periodic boundary conditions, $\mathbf{A}$ is a block circulant with circulant blocks (BCCB) matrix and matrix-vector products can be efficiently performed via the FFT [32]. We suppose that $m \geq n$ and that $\mathbf{A}^H \mathbf{A}$ has full rank. This assumption is satisfied in a variety of practical applications and is frequently used in image deblurring and in the literature of imaging inverse problems (see, for example, [32, 27, 33, 4, 58, 28, 23, 29]).

A variety of algorithms have been proposed in the literature for the solution of (2), especially in compressed sensing. The state-of-the-art methods for (2) are probably gradient descent-type methods since their computational cost is mainly due to matrix-vector products with $\mathbf{A}$ and $\mathbf{A}^H$. This class of methods includes the following popular methods: IST [20], TwIST [10], SparSA [56], FISTA [4], NESTA [5] and AdaptiveSPARSA [43]. SPGL1 [53] and GPSS [41] are gradient projection-type methods for the equivalent Lasso formulation of (2).

Fixed-point continuation methods [30, 31], as well as methods based on Bregman iterations [25, 57] and variable splitting, as SALSA [1] and C-SALSA [2], have also been recently proposed.

The dual formulation of (2) has been considered in the literature; in [27] and [33], a Newton projection method and an active-set-type method are respectively proposed to solve the dual problem.

A different approach to handle the nondifferentiability of the objective function of (2) consists in reformulating the nondifferentiable unconstrained problem (2) as a constrained differentiable one. The disadvantage of this strategy is that, usually, the number of variable is doubled. The interior point method of [34] solves a quadratic programming formulations of (2) obtained by introducing an auxiliary variable; it has been shown to be efficient for large scale problems since a preconditioned gradient method is used to compute the search direction.

An equivalent formulation of (2) as a quadratic programming problem with nonnegativity constraints can be done by splitting $\mathbf{x}$ into two nonnegative variables representing its positive and negative parts. Several methods have been proposed for the solution of the resulting bound constrained quadratic program as gradient projection [22], interior point [23] and Newton projection [45, 47, 46] methods.

The Newton projection methods (also referred to as Two-Metric projection methods) can be interpreted as scaled gradient projection methods where the scaling matrix is the inverse of a *modified* Hessian [8, 24, 9]. In particular, only the variables in the working set are scaled by the *inverse of the corresponding submatrix of the Hessian*, where the working set is defined as the complementary of the set of the null variables with positive gradient. Under proper hypothesis on the scaling matrix, the global convergence and local superlinear convergence

of Newton projection methods con be proved [8, 24].

In the literature, Newton-like projection methods have been recently developed which are tailor-made for image restoration problems formulated as ℓ_2-regularized optimization problems under nonnegativity constraints, both in the case of Gaussian and Poisson noise [39, 37, 38]. In these methods, the Hessian matrix is suitably approximated so that its inversion can be performed in the Fourier space at a very low computational cost. Then, the variables in the working set are scaled by the *corresponding submatrix of the inverse approximated Hessian*. It is worth mentioning that, using this scaling, the superlinear convergence rate of Newton-based methods is lost. However, this Newton-like projection methods have been shown to perform well compared to accelerated gradient projection algorithms for ℓ_2-regularized least squares problems [12, 36].

This work aims at developing a new efficient and effective ℓ_1-based method for the restoration of images corrupted by blur and Gaussian noise. In particular, we consider the analysis formulation of the ℓ_1-regularized least squares problem for the restoration of images which are quite sparse in the pixel representation as, for example, medical images and astronomical images. This approach has been introduced in [35, 23] and recently in [58]. Observe that in astronomical imaging, the ℓ_1-norm penalty is used as a flux conservation constraint. Therefore, the contribution of this paper is twofold. First, we show, by numerical evidence, that the ℓ_1-regularized least squares problem can be effectively used as a mathematical model of the problem of restoring images degraded by blur and Gaussian noise. Second, we develop an efficient Newton-like projection method for its solution. In the proposed approach, problem (2) is firstly formulated as a nonnegatively constrained quadratic program by splitting the variable $\mathbf{x}$ into the positive and negative parts. Then, the quadratic program is solved by a special purpose Newton-like projection method where a fair regularized approximation to the Hessian matrix is proposed so that products of its inverse and vectors can be computed at low computational cost. As a result, the only operations required for the search direction computation are matrix-vector products involving $\mathbf{A}$ and $\mathbf{A}^H$. Since the developed method uses a fair modification to the Hessian matrix, in the sequel, it will be referred to as Modified Newton Projection (MNP) method. Moreover, the regularization strategy, used to improve the conditioning of the Hessian matrix, penalizes negative entries in $\mathbf{x}$ so that nonnegative solutions are favored. The convergence of the MNP method is proved even if MNP slows the convergence rate of Newton-type methods. Even if the size of the problem is doubled, the low computational cost per iteration and less iterative steps make MNP quite efficient. It is worth mentioning that a similar approach is described in [46]; however, the projectionL1 method of [46] uses, as the classical Two-Metric projection methods, a suitable sub-matrix of the Hessian for scaling the gradient.

The performance of MNP is evaluated on some image restoration problems and is compared with that of the state-of-the-art methods. The results of the comparative study show that MNP is competitive and in some cases is also able to outperform the state-of-the-art methods in terms of computational complexity and achieved accuracy.

3

The paper is organized as follows. In section 2, the quadratic program formulation of (2) is derived. In section 3, the MNP method is presented and its convergence is analyzed. In this section, the computation of the search direction is also discussed. In section 4, the numerical results are presented. Conclusions are given in section 5.

2 The Modified Newton projection method

Before introducing our MNP method, we need to reformulate the ℓ_1-regularized least squares problem (2) as a nonnegatively constrained quadratic program (NCQP). This strategy for dealing with the nondifferentiability of the ℓ_1-norm is quite classic in the literature and has been previously adopted by several authors [13, 23, 22, 45, 46].

2.1 Nonnegatively constrained quadratic program formulation

If we split $\mathbf{x}$ into its positive and negative parts, that is

$$\mathbf{x} = \mathbf{u} - \mathbf{v}$$

where

$$\mathbf{u} = \max(\mathbf{x}, 0), \quad \mathbf{v} = \max(-\mathbf{x}, 0),$$

then, we obtain the following nonnegatively constrained quadratic program formulation of the original ℓ_1-regularized least squares problem (2):

$$
\begin{aligned}
\min_{(\mathbf{u}, \mathbf{v})} \ & \mathcal{F}(\mathbf{u}, \mathbf{v}) = \frac{1}{2} \|\mathbf{A}(\mathbf{u} - \mathbf{v}) - \mathbf{b}\|^2 + \lambda \mathbf{1}^H \mathbf{u} + \lambda \mathbf{1}^H \mathbf{v} \\
\text{s.t.} \ & \mathbf{u} \geq 0 \\
& \mathbf{v} \geq 0
\end{aligned}
\tag{3}
$$

where $\mathbf{1}$ denotes the n-dimensional column vector of ones. In [13, 22, 44] it is shown that problems (2) and (3) share the same solutions and that at the solution of (3) either u_i or v_i or both are equal to zero.

The gradient of $\mathcal{F}(\mathbf{u}, \mathbf{v})$ is defined by

$$
\nabla_{(\mathbf{u}, \mathbf{v})} \mathcal{F}(\mathbf{u}, \mathbf{v}) = \begin{bmatrix} \mathbf{A}^H \mathbf{A}(\mathbf{u} - \mathbf{v}) - \mathbf{A}^H \mathbf{b} + \lambda \mathbf{1} \\ -\mathbf{A}^H \mathbf{A}(\mathbf{u} - \mathbf{v}) + \mathbf{A}^H \mathbf{b} + \lambda \mathbf{1} \end{bmatrix}.
\tag{4}
$$

Henceforth, we will denote by $\mathbf{y}$ and $\mathbf{g}$ the $2n$-dimensional vectors

$$
\mathbf{y} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} \mathbf{g_u} \\ \mathbf{g_v} \end{bmatrix}
$$

where $\mathbf{g_u}$ and $\mathbf{g_v}$ are respectively the partial derivatives of $\mathcal{F}$ with respect to $\mathbf{u}$ and $\mathbf{v}$.

Observe that, even if by reformulating (2) as a NCQP we have doubled the problem size, the computation of the objective function and its gradient values indeed requires only one multiplication by $\mathbf{A}$ and one by $\mathbf{A}^H$.

2.2 Hessian approximation

The MNP method is basically a Newton-based method where the search direction computation requires the inversion of the Hessian matrix. Unfortunately, the Hessian $\mathbf{H}$ of $\mathcal{F}(\mathbf{u}, \mathbf{v})$

$$\mathbf{H} = \begin{bmatrix} \mathbf{A}^H\mathbf{A} & -\mathbf{A}^H\mathbf{A} \\ -\mathbf{A}^H\mathbf{A} & \mathbf{A}^H\mathbf{A} \end{bmatrix} \tag{5}$$

is a positive semidefinite matrix. The idea underlying the proposed approach is to substitute the Hessian with a nonsingular approximation. In MNP, we modify $\mathbf{H}$ by adding a small perturbation to its negative part. More precisely, we use the following Hessian approximation:

$$\mathbf{H}_\tau = \mathbf{H} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tau\mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^H\mathbf{A} & -\mathbf{A}^H\mathbf{A} \\ -\mathbf{A}^H\mathbf{A} & \mathbf{A}^H\mathbf{A} + \tau\mathbf{I} \end{bmatrix} \tag{6}$$

where τ is a positive parameter and $\mathbf{I}$ and $\mathbf{0}$ are respectively the identity and zero matrix of size n. There are several reasons for this choice of the Hessian modification. Firstly, an explicit formula for the inverse of $\mathbf{H}_\tau$ can be derived, secondly the search direction is computable at moderate cost and finally, negative part of images are penalized. Naturally, other nonsingular Hessian modifications are possible. Adding the positive constant τ to all elements of the diagonal $\mathbf{H}$ results in an Hessian modification such that the multiplication of its inverse by a vector requires more computational cost. Moreover, our numerical experiments indicate that this Hessian approximation tends to be more sensitive to the choice of the parameter τ. On the other hand, adding a small perturbation to the positive part of $\mathbf{H}$ would result in a penalization of the positive part of images and in less accurate reconstructions.

Proposition 2.1. *Assume that $\mathbf{A}^H\mathbf{A}$ is nonsingular. Then, $\mathbf{H}_\tau$ is nonsingular and its inverse is*

$$\mathbf{M}_\tau = \frac{1}{\tau} \begin{bmatrix} \tau(\mathbf{A}^H\mathbf{A})^{-1} + \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} \end{bmatrix}. \tag{7}$$

Proof. We prove that $\mathbf{H}_\tau\mathbf{M}_\tau = \mathbf{M}_\tau\mathbf{H}_\tau = \mathbf{I}_{2n}$ where $\mathbf{I}_{2n}$ is the identity matrix of size $2n$.
We have

$$\begin{aligned}
\mathbf{H}_\tau\mathbf{M}_\tau &= \frac{1}{\tau} \begin{bmatrix} \mathbf{A}^H\mathbf{A} & -\mathbf{A}^H\mathbf{A} \\ -\mathbf{A}^H\mathbf{A} & \mathbf{A}^H\mathbf{A} + \tau\mathbf{I} \end{bmatrix} \begin{bmatrix} \tau(\mathbf{A}^H\mathbf{A})^{-1} + \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} \end{bmatrix} \\
&= \frac{1}{\tau} \begin{bmatrix} \tau\mathbf{I} + \mathbf{A}^H\mathbf{A} - \mathbf{A}^H\mathbf{A} & \mathbf{A}^H\mathbf{A} - \mathbf{A}^H\mathbf{A} \\ -\tau\mathbf{I} - \mathbf{A}^H\mathbf{A} + \mathbf{A}^H\mathbf{A} + \tau\mathbf{I} & -\mathbf{A}^H\mathbf{A} + \mathbf{A}^H\mathbf{A} + \tau\mathbf{I} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}.
\end{aligned}$$

Similarly, we have $\mathbf{M}_\tau \mathbf{H}_\tau = \mathbf{I}_{2n}$ and this concludes the proof. $\qquad\square$

Proposition 2.2. *The inverse Hessian approximation $\mathbf{M}_\tau$ is a symmetric positive definite matrix.*

Proof. Let $\mathbf{z}, \mathbf{w} \in \mathbb{R}^n$. We have

$$\begin{bmatrix} \mathbf{z}^H & \mathbf{w}^H \end{bmatrix} \frac{1}{\tau} \begin{bmatrix} \tau(\mathbf{A}^H\mathbf{A})^{-1} + \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} = \mathbf{z}^H(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{z} + \frac{1}{\tau}\|\mathbf{z} + \mathbf{w}\|^2 \geq 0. \quad (8)$$

Equality holds in (8) if and only if $\mathbf{z} = 0$ and $\mathbf{z} + \mathbf{w} = 0$, i.e. if and only if $\mathbf{z} = 0$ and $\mathbf{w} = 0$. This concludes the proof. $\qquad\square$

Remark 2.1. The developed strategy for approximating the Hessian matrix $\mathbf{H}$ is closely related to the method of Lavrentiev regularization [40, 49, 51, 6]. In fact, in Lavretiev regularization, a regularized solution of a linear ill-posed problem with positive semidefinite linear operator $\mathbf{K}$ is obtained as the solution of a slightly modified equation whose linear operator is $\mathbf{K} + \tau\mathbf{I}$. Therefore, the Hessian approximation $\mathbf{H}_\tau$ is indeed a Lavretiev-type regularized approximation to $\mathbf{H}$ where only the last n variables are penalized. In this way, large values of the last n components of the search direction are penalized.

2.3 Algorithm

Firstly, let us introduce the basic notation that enables us to formalize the description of the MNP method. In the sequel, $\mathcal{A}(\mathbf{y})$ will indicate the set of indices [8, 9, 54]:

$$\mathcal{A}(\mathbf{y}) = \left\{ i \mid 0 \leq y_i \leq \varepsilon \text{ and } g_i > 0 \right\}, \quad \varepsilon = \min\{\bar{\varepsilon}, \|\mathbf{y} - [\mathbf{y} - \mathbf{g}]^+\|\}$$

where $\bar{\varepsilon}$ is a small positive parameter and $[\cdot]^+$ denotes the projection on the positive orthant. Finally, let $\mathbf{E}$ and $\mathbf{F}$ denote the diagonal matrices [54] such that

$$\{\mathbf{E}(\mathbf{y})\}_{ii} = \left\{ \begin{array}{ll} 1, & i \notin \mathcal{A}(\mathbf{y}); \\ 0, & i \in \mathcal{A}(\mathbf{y}); \end{array} \right.$$
$$\mathbf{F}(\mathbf{y}) = \mathbf{I}_{2n} - \mathbf{E}(\mathbf{y}).$$

Given a feasible initial iterate $\mathbf{y}^{(0)} \in \mathbb{R}^{2n}$, the MNP method is defined by the general iteration

$$\mathbf{y}^{(k+1)} = [\mathbf{y}^{(k)} - \alpha^{(k)}\mathbf{p}^{(k)}]^+, \quad \mathbf{p}^{(k)} = \mathbf{S}^{(k)}\mathbf{g}^{(k)} \qquad (9)$$

where

$$\mathbf{S}^{(k)} = \mathbf{E}(\mathbf{y}^{(k)})\mathbf{M}_\tau\mathbf{E}(\mathbf{y}^{(k)}) + \mathbf{F}(\mathbf{y}^{(k)}) \qquad (10)$$

and the step-length $\alpha^{(k)}$ is determined by the Armijo rule along the projection arc [8, 9]. That is, $\alpha^{(k)}$ is the first number of the sequence $\{2^{-m}\}_{m \in \mathbb{N}}$ such that

$$\mathcal{F}\big(\mathbf{y}^{(k)}\big) - \mathcal{F}\big(\mathbf{y}^{(k)}(2^{-m})\big) \geq$$
$$\beta \left(2^{-m} \sum_{i \notin \mathcal{A}^{(k)}} g_i^{(k)} p_i^{(k)} + \sum_{i \in \mathcal{A}^{(k)}} g_i^{(k)} \Big(y_i^{(k)} - y_i^k(2^{-m}) \Big) \right) \quad (11)$$

where $\mathbf{y}^{(k)}(2^{-m}) = [\mathbf{y}^{(k)} - 2^{-m}\mathbf{p}^{(k)}]^+$, $\beta \in (0, \frac{1}{2})$. For easier notation, in the following, $\mathbf{E}^{(k)}$, $\mathbf{F}^{(k)}$ and $\mathcal{A}^{(k)}$ will denote respectively the diagonal matrices $\mathbf{E}(\mathbf{y}^{(k)})$ and $\mathbf{F}(\mathbf{y}^{(k)})$ and the index set $\mathcal{A}(\mathbf{y}^{(k)})$.

Remark 2.2. In the Two-Metric projection method originally proposed by Bertsekas [8, 9, 24], the scaling matrix is

$$\mathbf{S}^{(k)} = \left(\mathbf{E}^{(k)} \mathbf{H} \mathbf{E}^{(k)} + \mathbf{F}^{(k)} \right)^{-1}. \quad (12)$$

Using this scaling, the standard Newton-like projection methods attain the superlinear convergence rate of Newton-like methods. However, the inversion of a submatrix of $\mathbf{H}$, required in (12), is often impracticable in image deblurring applications. Therefor, in practice, a conjugate gradient version of the Newton-like projection method has to be used where the CG iterations are terminated when the relative residual becomes smaller than a given tolerance. The resulting approximate Newton-CG projection method may slow the convergence rate.

On the other hand, the scaling matrix (10) involves the inverse of the whole approximated Hessian matrix $\mathbf{H}_\tau$ for which an explicit formula is given in proposition 2.1. As a result, the search direction of MNP can be computed very quickly. However, the scaling (10) may significantly slow the convergence rate and MNP may have worse convergence properties.

2.4 Convergence analysis

As proved in [8, 24], the convergence of Newton-like projection methods only requires the scaling matrices $\mathbf{S}^{(k)}$ to be positive definite matrices with uniformly bounded eigenvalues. In particular, the global convergence property of these methods can be proved under the general following assumptions [8].

A1 The gradient $\nabla_{(\mathbf{u},\mathbf{v})}\mathcal{F}$ is Lipschitz continuous on each bounded set of $\mathbb{R}^{2n}$.

A2 There exist positive scalars c_1 and c_2 such that

$$c_1 \|\mathbf{y}\|^2 \leq \mathbf{y}^H \mathbf{S}^{(k)} \mathbf{y} \leq c_2 \|\mathbf{y}\|^2, \quad \forall \mathbf{y} \in \mathbb{R}^{2n}, \quad k = 0, 1, \dots$

The key convergence result is provided in Proposition 2 of [7] which is restated here for the shake of completeness.

Proposition 2.3. *[7, Proposition 2] Let $\{\mathbf{y}^{(k)}\}$ be a sequence generated by iteration (9) where $\mathbf{S}^{(k)}$ is a positive definite symmetric matrix which is diagonal with respect to $\mathcal{A}^{(k)}$ and α^k is computed by the Armijo rule along the projection arc. Under assumptions A1 and A2 above, every limit point of a sequence $\{\mathbf{y}^{(k)}\}$ is a critical point with respect to problem (3).*

Since the objective $\mathcal{F}$ of (3) is twice continuously differentiable, it satisfies assumption A1.

The inverse Hessian approximation $\mathbf{M}_\tau$ is a symmetric positive definite matrix (proposition 2.2) and hence, the scaling matrix $\mathbf{S}^{(k)}$ defined by (10) is a positive definite symmetric matrix which is diagonal with respect to $\mathcal{A}^{(k)}$. Therefore, the global convergence of the MNP method is guaranteed if $\mathbf{S}^{(k)}$ verifies assumption A2.

Proposition 2.4. *Given a positive parameter $\tau > 0$, there exist two positive scalars c_1^τ and c_2^τ such that*

$$c_1^\tau \|\mathbf{y}\|^2 \le \mathbf{y}^H \mathbf{S}^{(k)} \mathbf{y} \le c_2^\tau \|\mathbf{y}\|^2, \quad \forall \mathbf{y} \in \mathbb{R}^{2n}, \quad k = 0, 1, \dots$$

Proof. Since $\mathbf{M}_\tau$ is positive definite, then

$$\sigma_1^\tau \|\mathbf{y}\|^2 \le \mathbf{y}^H \mathbf{M}_\tau \mathbf{y} \le \sigma_{2n}^\tau \|\mathbf{y}\|^2, \quad \forall \mathbf{y} \in \mathbb{R}^{2n}. \tag{13}$$

where σ_1^τ and σ_{2n}^τ are the largest and smallest eigenvalue of $\mathbf{M}_\tau$, respectively. We have

$$\mathbf{y}^H \mathbf{S}^{(k)} \mathbf{y} = \mathbf{y}^H \big(\mathbf{E}^{(k)} \mathbf{M}_\tau \mathbf{E}^{(k)} + \mathbf{F}^{(k)}\big)\mathbf{y} = \big(\mathbf{E}^{(k)} \mathbf{y}\big)^H \mathbf{M}_\tau \big(\mathbf{E}^{(k)} \mathbf{y}\big) + \mathbf{y}^H \mathbf{F}^{(k)} \mathbf{y}.$$

From (13) it follows that

$$\sigma_1^\tau \|\mathbf{E}^{(k)} \mathbf{y}\|^2 + \mathbf{y}^H \mathbf{F}^{(k)} \mathbf{y} \le \big(\mathbf{E}^{(k)} \mathbf{y}\big)^H \mathbf{M}\big(\mathbf{E}^{(k)} \mathbf{y}\big) + \mathbf{y}^H \mathbf{F}^{(k)} \mathbf{y}$$
$$\le \sigma_{2n}^\tau \|\mathbf{E}^{(k)} \mathbf{y}\|^2 + \mathbf{y}^H \mathbf{F}^{(k)} \mathbf{y}.$$

Moreover we have

$$\mathbf{y}^H \mathbf{F}^{(k)} \mathbf{y} = \sum_{i \in \mathcal{A}^{(k)}} y_i^2, \quad \|\mathbf{E}^{(k)} \mathbf{y}\|^2 = \sum_{i \notin \mathcal{A}^{(k)}} y_i^2.$$

Hence, we obtain

$$\sigma_1^\tau \sum_{i \notin \mathcal{A}^{(k)}} y_i^2 + \sum_{i \in \mathcal{A}^{(k)}} y_i^2 \le \mathbf{y}^H \mathbf{S}^{(k)} \mathbf{y} \le \sigma_{2n}^\tau \sum_{i \notin \mathcal{A}^{(k)}} y_i^2 + \sum_{i \in \mathcal{A}^{(k)}} y_i^2$$

and

$$\min\{\sigma_1^\tau, 1\} \|\mathbf{y}\|^2 \le \mathbf{y}^H \mathbf{S}^{(k)} \mathbf{y} \le \max\{\sigma_{2n}^\tau, 1\} \|\mathbf{y}\|^2.$$

The thesis immediately follows by setting

$$c_1^\tau = \min\{\sigma_1^\tau, 1\} \quad c_2^\tau = \max\{\sigma_{2n}^\tau, 1\}.$$

$\square$

Propositions 2.3 and 2.4 ensure the global convergence of the MNP method.

2.5 Implementation of the search direction computation

The computation of the search direction $\mathbf{p}^{(k)}$ requires the multiplication of a vector by $\mathbf{M}_\tau$. Let $\begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} \in \mathbb{R}^{2n}$ be a given vector, then it immediately follows that

$$\mathbf{M}_\tau \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} = \frac{1}{\tau} \begin{bmatrix} \tau(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{z} + \mathbf{z} + \mathbf{w} \\ \mathbf{z} + \mathbf{w} \end{bmatrix} = \begin{bmatrix} (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{z} + (\mathbf{z} + \mathbf{w})/\tau \\ (\mathbf{z} + \mathbf{w})/\tau \end{bmatrix}. \quad (14)$$

In many image restoration applications, the blurring matrix $\mathbf{A}$ is severely ill-conditioned and computing $(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{z}$ requires a regularization strategy. Therefore, in our implementation, a Tikhonov-like technique is employed in the search direction computation by approximating the matrix-vector product (14) with

$$\mathbf{M}_\tau \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} \approx \begin{bmatrix} (\mathbf{A}^H\mathbf{A} + \gamma\mathbf{I})^{-1}\mathbf{z} + (\mathbf{z} + \mathbf{w})/\tau \\ (\mathbf{z} + \mathbf{w})/\tau \end{bmatrix} \quad (15)$$

where γ is a positive parameter. The inversion of $\mathbf{A}^H\mathbf{A} + \gamma\mathbf{I}$ can be efficiently performed in the Fourier space with computational complexity of two Fast Fourier Transforms as follows. In fact, assuming periodic boundary conditions, $\mathbf{A}$ can be factorized as

$$\mathbf{A} = \mathbf{U}^*\mathbf{D}\mathbf{U}$$

where $\mathbf{U}$ is the two dimensional unitary Discrete Fourier Transform (DFT) matrix and $\mathbf{D}$ is the diagonal matrix containing the eigenvalues of $\mathbf{A}$. Thus

$$(\mathbf{A}^H\mathbf{A} + \gamma\mathbf{I})^{-1} = \mathbf{U}^*(|\mathbf{D}|^2 + \gamma\mathbf{I})^{-1}\mathbf{U}.$$

The products involving $\mathbf{U}$ and $\mathbf{U}^*$ can be performed by using the FFT algorithm at the cost of $O(n\log_2 n)$ operations while the inversion of the diagonal matrix $(|\mathbf{D}|^2 + \gamma\mathbf{I})$ has the cost $O(n)$.

3 Numerical results

In this section, we present the numerical results of several image restoration test problems. The numerical experiments aim at illustrating the performance of MNP compared with several state-of-the-art methods as SALSA [1], SPARSA [56], FISTA [4], NESTA [5], CGIST [26] and the Split Bregman method [25, 57]. In our comparative study, we also consider first and second-order methods solving, as MNP, the quadratic program (3) such as the nonmonotonic version of GPSR using the BarzilaiBorwein technique for the step-length selection [22], the original Two-Metric Projection (TMP) method of Gafni and Bertsekas [24, 8, 45] employing the scaling (12) and a modified Newton projection method obtained by regularizing the full diagonal of the Hessian matrix (5). This last method, which will be indicated as MNP2, requires the inversion of the Hessian approximation:

$$\mathbf{H}_\tau = \begin{bmatrix} \mathbf{A}^H\mathbf{A} + \tau\mathbf{I} & -\mathbf{A}^H\mathbf{A} \\ -\mathbf{A}^H\mathbf{A} & \mathbf{A}^H\mathbf{A} + \tau\mathbf{I} \end{bmatrix}.$$

It can be prove (the proof is similar to that of proposition 2.1) that the inverse of $\mathbf{H}_\tau$ is defined as

$$
\mathbf{M}_\tau = \frac{1}{\tau} \begin{bmatrix} \mathbf{A}^H\mathbf{A} + \tau\mathbf{I} & \mathbf{A}^H\mathbf{A} \\ \mathbf{A}^H\mathbf{A} & \mathbf{A}^H\mathbf{A} + \tau\mathbf{I} \end{bmatrix} \begin{bmatrix} (2\mathbf{A}^H\mathbf{A} + \tau\mathbf{I})^{-1} & \mathbf{0} \\ \mathbf{0} & (2\mathbf{A}^H\mathbf{A} + \tau\mathbf{I})^{-1} \end{bmatrix}.
\tag{16}
$$

Formula (16) indicates that, in MNP2, the computation of the search direction requires more computational effort than in MNP.

Finally, MNP has also been compared with the OWLQN method of Andrew and Gao [3] which is probably one of the most effective method proposed for general large-scale ℓ_1-regularized learning.

3.1   Overall assessment of the considered methods

The Matlab source code of the considered methods, when made publicly available by the authors, has been used in the numerical experiments. The OWLQN method is coded, by the authors, in C; a Matlab version, following the instruction provided in [3], has been implemented.

In TMP, the linear system for the search direction computation has been solved by the Conjugate Gradient method. The relative residual tolerance has been fixed at 0.01 and a maximum number of 100 iterations has been allowed. We remark that lower values for the CG relative tolerance produce restorations of worse quality since the Hessian matrix is ill-conditioned.

The numerical experiments have been executed on a Sun Fire V40z server consisting of four 2.4GHz AMD Opteron 850 processors with 16GB RAM using Matlab 7.10 (Release R2010a).

For all the considered methods, the initial iterate $\mathbf{x}^{(0)}$ has been chosen to be the zero image whose pixel values are all equal to zero. The regularization parameter λ of (2) has been heuristically chosen.

MNP depends on the τ and γ parameters. After a wide experimentation, we heuristically found that good values for these parameters are

$$
\tau = 50\lambda, \quad \gamma = \lambda.
\tag{17}
$$

These values have been used in all the presented numerical experiments. It is worth remarking that the values (17) for τ and γ are not optimal for each experiments and that the numerical results could be further improved. However, with this parameter setting, the MNP algorithm only depends on the regularization parameter λ. We also observe that the SALSA method depends on a penalty parameter μ [1] and that the Split Bregman method depends on a "splitting" regularization coefficient ν [25, 57]. In all the presented experiments, the values of these parameters have been handtuned for the best mean squared error reduction. The default parameters of the other methods, suggested by the authors have been chosen.

The termination criterion of all the considered methods is based on the relative change in the objective function at the last step. In particular, the

10

Figure 1: Experiment 1: deblurring of the *satellite* image. Left: original image; right: observed image (NL $= 5 \cdot 10^{-3}$).

methods iteration is terminated when

$$|\phi(\mathbf{x}^{(k+1)}) - \phi(\mathbf{x}^{(k)})| \leq tol_\phi \ \phi(\mathbf{x}^{(k+1)}) \tag{18}$$

where tol_ϕ is a small positive parameter. The selection of a fair stopping criterium in image deblurring application is a critical issue. In our numerical experiments, this criterium, also used in [56, 5] appears to provide sufficiently accurate solutions while avoiding excessive computational cost. Moreover, the use of a stopping criterion based on the norm of the projected gradient or the duality gap can be problematic since it can confuse lack of progress for optimality.

A maximum number of 500 iterations has been allowed for each method.

The quality of the restorations provided by the compared methods has been measured by using the Peak Signal-to-Noise Ratio (PSNR) values.

All pixels of the original images described in the following experiments have been first scaled into the range between 0 and 1.

3.2 Experiment 1: the satellite image

In the first experiment, the famous 256×256 *satellite* image has been considered. The satellite image is a good test image for ℓ_1-based image restoration because it has many pixels with value equal to zero. The observed image of figure 1 has been generated by convolving the original image, also shown in figure 1, with a Gaussian PSF with variance equal to 2, obtained with the code `psfGauss` from [32], and then by adding Gaussian noise with noise level equal to $5 \cdot 10^{-3}$. (The noise level NL is defined as NL $:= \|\boldsymbol{\eta}\|/\|\mathbf{A}\mathbf{x}_{\text{original}}\|$ where $\mathbf{x}_{\text{original}}$ is the original image.) In figure 1 and following, the satellite image intensities are displayed in negative gray-scale. For this test problem, the μ and ν parameters of the SALSA and Split Bregman methods have been set to $\mu = 5\lambda$ and $\nu = 0.025\lambda$, respectively.

Table 1 reports the PSNR values, the objective values, the CPU times in seconds, the number of performed iterations and the percentage of negative pixel

Method	tol_ϕ	PSNR	Obj	Time	It	Neg %
MNP	10^{-3}	2.92e+01	3.54e+01	0.42	10	0.16
	10^{-5}	2.94e+01	3.53e+01	0.80	17	0.00
SALSA	10^{-3}	2.83e+01	3.51e+01	0.09	6	42.03
	10^{-5}	2.92e+01	3.52e+01	0.63	35	39.19
GPSR	10^{-3}	2.80e+01	3.54e+01	0.34	16	0.00
	10^{-5}	2.89e+01	3.53e+01	0.78	37	0.00
Bregman	10^{-3}	2.85e+01	4.09e+01	0.87	48	44.60
	10^{-5}	2.83e+01	3.90e+01	3.46	200	44.72
CGIST	10^{-3}	2.72e+01	3.56e+01	0.43	15	0.00
	10^{-5}	2.91e+01	3.53e+01	3.44	114	0.00
SPARSA	10^{-3}	2.70e+01	3.57e+01	0.24	10	0.00
	10^{-5}	2.93e+01	3.53e+01	0.85	47	0.00
FISTA	10^{-3}	2.77e+01	3.54e+01	0.37	16	0.00
	10^{-5}	2.92e+01	3.53e+01	1.67	75	0.00
NESTA	10^{-3}	2.70e+01	3.70e+01	4.21	70	44.74
	10^{-5}	2.79e+01	3.77e+01	38.94	933	44.67
MNP2	10^{-3}	2.69e+01	3.57e+01	0.80	20	0.02
	10^{-5}	2.88e+01	3.53e+01	8.41	150	0.00
TMP	10^{-3}	2.83e+01	3.54e+01	2.61	7	0.00
	10^{-5}	2.88e+01	3.53e+01	17.83	75	0.00
OWLQN	10^{-3}	2.78e+01	3.54e+01	0.39	10	0.00
	10^{-5}	2.85e+01	3.53e+01	3.60	65	0.00

Table 1: Experiment 1: numerical results.

values in the restored images. All the results are averaged over ten runs of each method. The numerical results given in table 1 have been obtained by using the stopping tolerance values $tol_\phi = 10^{-3}$ and $tol_\phi = 10^{-5}$. Smaller values of tol_ϕ do not improve the visual quality of the restored images even if they produce more accurate solutions to the optimization problem (2). The information in table 1 is summarized in figure 2 where the PSNR values versus time are plotted for all the considered methods. The results in table 1 and figure 2 show that MNP is able to provide good quality restorations at low computational effort. They also indicate that the inversion of the modified Hessian adopted in MNP2 requires more computational effort.

In figure 3, the performance of MNP is compared to the performance of the first-order and second order methods (NESTA has been omitted from these comparisons because of its high computational time, see table 1). Since PSNR is inversely proportionally to the Mean Squared Error (MSE), figure 3 illus-
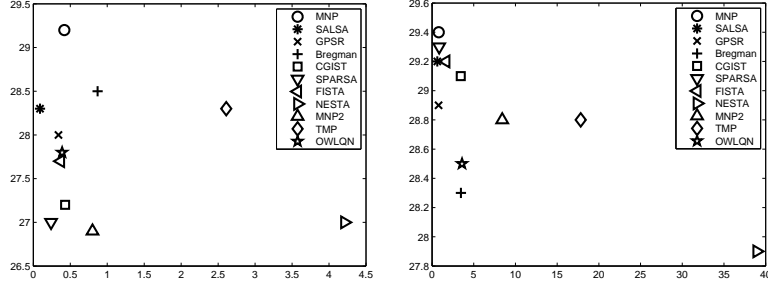
Figure 2: Experiment 1: PSNR values versus time (in seconds) obtained at $tol_\phi = 10^{-3}$ (left) and $tol_\phi = 10^{-5}$ (right).

trates, in a semi-logarithmic scale, the MSE behavior and the decreasing of the objective function versus time. For the sake of readability, all the methods are not represented in the same picture. The convergence rates of the considered second-order methods (MNP, MNP2, TMP and OWLQN) have also been compared. For these methods, figure 4 shows, in a semi-logarithmic scale, the behavior of the gradient norm values versus time; from the graphs it is evident that MNP, MNP2 and TMP decrease the gradient norm slower than OWLQN. As expected, the scaling of the gradient direction employed by MNP and MNP2 reduces the local convergence rate; moreover, in TMP, the use of an iterative solver for the search direction computation leads to slower convergence rate.
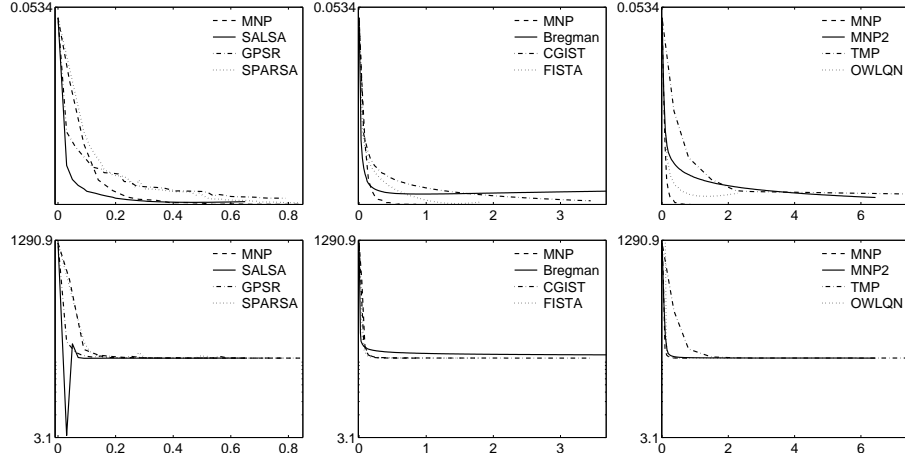

Figure 3: Experiment 1: comparison between the considered methods. Top line: relative error histories versus time; bottom line: objective function decrease versus time (in seconds).
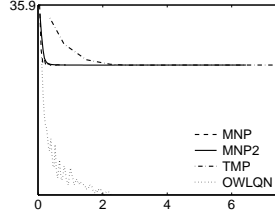
13

Figure 4: Experiment 1: comparison between MNP and the second-order methods (MNP2, TMP, OWLQN): gradient norm reduction versus time (in seconds).

MNP $(tol = 10^{-3})$ SALSA   Bregman
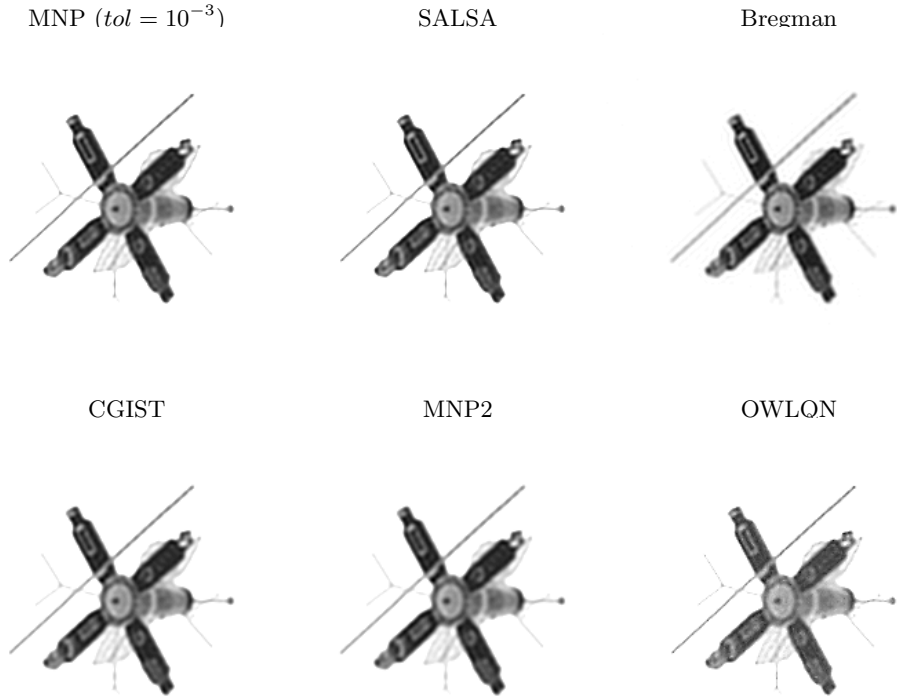
CGIST   MNP2   OWLQN

Figure 5: Experiment 1: restorations provided by some of the considered methods with $tol_\phi = 10^{-5}$. For MNP, the restoration at $tol_\phi = 10^{-3}$ is depicted.

Figure 5 compares the image obtained by MNP at $tol_\phi = 10^{-3}$ with those provided by SALSA, Split Bregman, CGIST, MNP2 and OWLQN at $tol_\phi = 10^{-5}$. (For these methods, the images obtained at $tol_\phi = 10^{-3}$ are not shown since their visual quality is clearly poor.) In the Bregman image, for a better

14

visualization, the pixels with negative values image has been set to zero. The images restored by SPARSA and FISTA are not shown because they are practically indistinguishable from those obtained by SALSA and MNP. The visual quality of the GPSR, NESTA and TMP images is inferior and therefore these images are not displayed (see also their PSNR values in table 1). Observe that the MNP image is visually comparable to the SALSA image while the CGIST image has an inferior visual quality even if its PSNR value is comparable to that of SALSA. Moreover, the MNP2 and OWLQN images are not comparable, in terms of visual quality, to the MNP image; in particular, the MNP2 image seems to be too smooth.

The restored images obtained by SALSA, NESTA and the Split Bregman methods have pixels with negative entries. They are displayed in figure 6 where the black pixels correspond to nonnegative pixels values of the reconstructions.
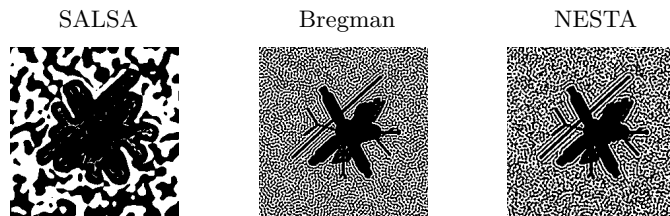

Figure 6: Experiment 1: pixels with negative entries. The black pixels correspond to pixels with nonnegative values. The white pixels correspond to pixels with negative values.

Regularization	PSNR	neg %
ℓ_1	2.94e+01	0
ℓ_2	2.75e+01	44
TV	2.96e+01	45

Table 2: Experiment 1: comparison between ℓ_1-norm, ℓ_2-norm and TV regularizers.

Finally, in order to assess the performance of the image restoration criterium (2), we have compared the ℓ_1-norm regularizer with the ℓ_2-norm and the Total Variation (TV) regularizers, which are widely-used in image restoration. The Lagged Diffusivity Fixed Point method [55] has been considered for the solution of the TV-regularized least squares problem. The ℓ_2-regularized least squares problem has been solved by inverting, in the Fourier space, the linear system of its first order conditions. The MNP method is used for the solution of (2). Table 2 reports the PSNR values and the percentage of negative pixels; figure 7
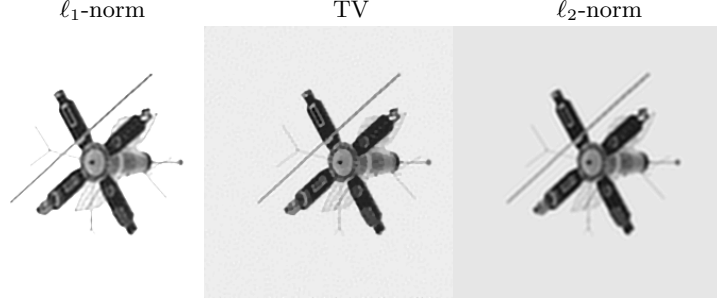
Figure 7: Experiment 1: comparison between ℓ_1-norm, ℓ_2-norm and TV regularizers.

displays the restored images. The superiority of the ℓ_1-norm regularized image is evident both in terms of PSNR values and visual quality improvement. In fact, edges are preserved by the restoration model (2) without any staircasing or blurring effects which are respectively evident in the TV-regularized and ℓ_2-norm regularized images.

3.3   Experiment 2: the Flintstones image

In the second experiment, the 512×512 *Flintstones* image (figure 8) has been used. This image is a good test image since has a nice mixture of detail and flat regions. The original Flintstones image has been degraded by two blurring operators, shown in figure 8, and by adding varying amounts of Gaussian noise. In this experiment, MNP has been compared to SALSA, Split Breg-
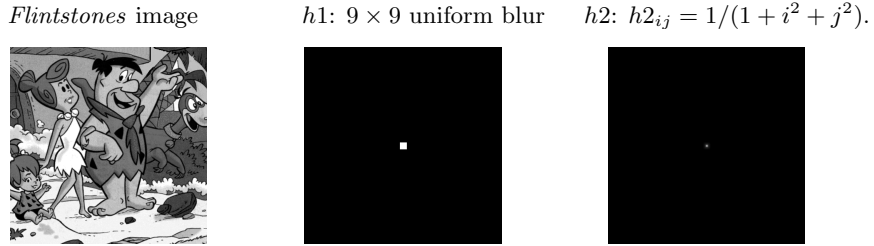

Figure 8: Experiment 2: original *Flintstones* image and blur operators.

man, FISTA, MNP2 and OWLQN. The reasons why we limit our comparative study to these methods are the following. From experiment 1, it is evident that SALSA and FISTA are able to achieve good quality results in little time. Moreover, Split Bregman and OWLQN have been considered because they are

well-known methods for general ℓ_1 minimization not only proposed for image restoration applications. Finally, MNP2 has been considered because closely-related to MNP.

We have fixed the values $\mu = 2.5\lambda$ and $\nu = \lambda$ for the SALSA and Split Bregman parameter. The numerical results are summarized in table 3 where the level of the added noise (first column), the PSNR values, the objective values and the CPU times in seconds are reported for each blurring operator. The reported numerical values are the average over 10 run of each experiment and have been obtained using the tolerance values $tol_\phi = 10^{-3}$ since smaller values of tol_ϕ do not improve the visual quality of the restored images. Figure 9 depicts the MNP restorations from the degraded images with highest noise level (NL=0.01). They asses the ability of the image restoration model (2) to remove blur and noise from the degraded images.

Figure 9: Experiment 2: Top line: degraded image (NL=0.01). Bottom line: MNP restorations. Left column: blur $h1$; right column: blur $h2$.

3.4   Experiment 3

This last experiment is based on four famous test images with different features: the 256×256 *Text* image, the 358×358 *Moon* image, the 364×364 *Concord* image and the 490×490 *Spine* image (see figure 10). These images are example images from the Matlab distribution. They have been corrupted by the blurring operators depicted in figure 8 and then Gaussian noise has been added with the same noise level values used in experiment 2. For each noise level, 10 Gaussian

NL	Method	Blur $h1$			Blur $h2$		
		PSNR	Obj	Time	PSNR	Obj	Time
0.001	MNP	3.40e+01	1.39e+03	1.67	2.49e+01	4.21e+02	1.08
	SALSA	3.25e+01	1.39e+03	0.76	2.49e+01	4.19e+02	0.51
	BREGMAN	3.72e+01	1.40e+03	0.20	2.73e+01	4.21e+02	0.24
	FISTA	2.65e+01	1.39e+03	1.47	2.14e+01	4.22e+02	1.78
	MNP2	2.45e+01	1.40e+03	3.42	2.25e+01	4.21e+02	7.41
	OWLQN	2.83e+01	1.39e+03	1.42	2.18e+01	4.22e+02	1.61
0.0025	MNP	3.16e+01	2.08e+03	0.96	2.42e+01	6.99e+02	1.36
	SALSA	2.91e+01	2.06e+03	0.39	2.41e+01	6.98e+02	0.51
	BREGMAN	3.16e+01	2.10e+03	0.18	2.45e+01	7.03e+02	0.24
	FISTA	2.57e+01	2.08e+03	1.22	2.10e+01	7.02e+02	1.54
	MNP2	2.34e+01	2.09e+03	3.31	2.06e+01	7.05e+02	3.04
	OWLQN	2.56e+01	2.08e+03	0.98	2.11e+01	7.03e+02	1.09
0.005	MNP	2.94e+01	2.42e+03	0.77	2.38e+01	9.78e+02	0.92
	SALSA	2.83e+01	2.40e+03	0.40	2.35e+01	9.76e+02	0.51
	BREGMAN	2.87e+01	2.46e+03	0.27	2.31e+01	9.85e+02	0.32
	FISTA	2.56e+01	2.42e+03	1.32	2.07e+01	9.82e+02	1.34
	MNP2	2.31e+01	2.44e+03	3.67	2.00e+01	9.86e+02	2.75
	OWLQN	2.55e+01	2.42e+03	0.97	2.11e+01	9.82e+02	1.11
0.0075	MNP	2.84e+01	2.48e+03	0.65	2.33e+01	1.12e+03	1.18
	SALSA	2.84e+01	2.48e+03	0.59	2.31e+01	1.12e+03	0.51
	BREGMAN	2.72e+01	2.53e+03	0.32	2.23e+01	1.13e+03	0.40
	FISTA	2.50e+01	2.49e+03	0.96	2.07e+01	1.12e+03	1.34
	MNP2	2.30e+01	2.51e+03	3.02	1.99e+01	1.13e+03	2.73
	OWLQN	2.54e+01	2.49e+03	0.84	2.10e+01	1.12e+03	1.10
0.01	MNP	2.83e+01	2.60e+03	0.47	2.28e+01	1.26e+03	1.36
	SALSA	2.73e+01	2.55e+03	0.60	2.27e+01	1.25e+03	0.52
	BREGMAN	2.64e+01	2.60e+03	0.40	2.15e+01	1.27e+03	0.41
	FISTA	2.49e+01	2.56e+03	0.96	2.05e+01	1.26e+03	1.26
	MNP2	2.29e+01	2.58e+03	3.07	1.97e+01	1.27e+03	2.77
	OWLQN	2.54e+01	2.56e+03	0.84	2.10e+01	1.26e+03	1.11

Table 3: Experiment 2: numerical results.

noise vectors have been obtained with different noise realizations. This amounts to 100 experiments for each image and to 400 experiments in total.

Also in this experiment, MNP has been compared to SALSA, Split Bregman, FISTA, MNP2 and OWLQN. The numerical results have been obtained by using the tolerance value $tol_\phi = 10^{-3}$. The μ and ν parameters of the SALSA and Split Bregman methods have been set to the same values of the previous experiment.

Let PSNR$_{\text{METHOD}}$ denote the PSNR value given by a method among those

considered (MNP, SALSA, Split Bregman, FISTA, MNP2 and OWLQN) and let $\text{TIME}_{\text{METHOD}}$ denote the corresponding CPU time, in seconds. The performance of the methods has been compared in terms of percentage of experiments producing a reconstructed image such that

$$\text{PSNR}_{\text{METHOD}} \geq C_1 \cdot \text{PSNR}_{\text{MAX}}, \quad C_1 = 1, 0.9, 0.8$$

and

$$\text{TIME}_{\text{METHOD}} \leq C_2 \cdot \text{TIME}_{\text{MIN}}, \quad C_2 = 1, 3, 5$$

where, for each experiment, PSNR_{MAX} is the maximum PSNR value and TIME_{MIN} is the minimum time. For each image and for the total 400 experiments, the corresponding percentage values are reported in tables 4 and 5. Figure 10 depicts the MNP restorations from the images degraded by the $h1$ blurring operator and Gaussian noise with NL=0.005.

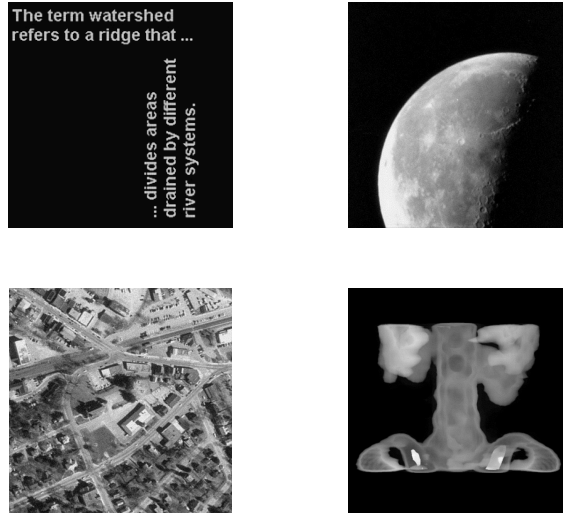Figure 10: Experiment 3: MNP restorations for $h1$ blurring operator and NL=0.005. From left to right and from top to bottom: *Text, Moon, Concord* and *Spine* images.

4   Conclusions

This paper describes a new approach to the solution of ℓ_1-regularized least squares problems whose matrix is supposed to be overdetermined and full-rank. This approach solves the nonnegatively constrained quadratic programming reformulation of the original least squares problem by a modified Newton projection method where the Hessian matrix is approximated so that it can be

Image	C_1	MNP	SALSA	BREG.	FISTA	MNP2	OWQLN
	1	70	10	20	0	0	0
Text	0.9	82	89	35	43	0	6
	0.8	99	100	75	90	31	83
	1	50	0	10	0	20	20
Moon	0.9	80	70	20	70	80	80
	0.8	99	90	50	100	100	100
	1	20	60	10	0	0	10
Concord	0.9	100	100	20	50	48	60
	0.8	100	100	60	90	90	90
	1	0	21	20	0	29	30
Spine	0.9	0	60	51	90	89	100
	0.8	53	80	60	100	100	100
	1	35.00	22.75	15.00	0.00	12.25	15.00
Total	0.9	65.50	79.75	31.50	63.25	54.25	61.50
	0.8	87.75	92.50	61.25	95.00	80.25	93.25

Table 4: Experiment 3: percentage of experiments such that $\mathrm{PSNR_{METHOD}} \geq C_1 \cdot \mathrm{PSNR_{MAX}}$.

Image	C_2	MNP	SALSA	BREG.	FISTA	MNP2	OWQLN
	1	3	80	20	0	0	0
Text	0.9	34	96	48	1	0	10
	5	47	100	100	54	0	55
	1	0	82	10	2	0	9
Moon	3	45	90	59	48	40	90
	5	86	100	60	90	40	90
	1	0	15	86	0	0	0
Concord	3	26	100	100	0	0	19
	5	38	100	100	30	6	32
	1	0	9	32	10	0	52
Spine	0.9	30	70	80	70	70	70
	5	60	70	97	70	70	70
	1	0.75	46.50	37.50	2.50	0.00	15.75
Total	0.9	34.75	89.00	70.25	28.75	27.50	48.25
	5	58.00	92.50	89.25	61.00	31.00	63.50

Table 5: Experiment 3: percentage of experiments such that $\mathrm{Time_{METHOD}} \leq C_2 \cdot \mathrm{Time_{MIN}}$.

efficiently inverted in the Fourier space. The developed MNP method favors nonnegative solutions of the ℓ_1-regularized least squares problem without explicitly imposing any constraints in the optimization problem. Thus, MNP can also be applied to image restoration problems whose solution may have negative components as in compressed sensing when $\mathbf{x}$ represents the coefficient vector of the image under some basis. A comparative study with some state-of-the-art methods has been performed in order to evaluate the potential of the described approach. The numerical results show that MNP is competitive with the considered methods in terms of PSNR values while SALSA is often the fastest method.

5   Acknowledgments

References

[1] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. Image Process.*, 19:2345–2356, 2010.

[2] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Process.*, 20(3), 2011.

[3] G. Andrew and J. F. Gao. Scalable training of l1-regularized log-linear models. In *Proc. 24th Intl. Conf. Mach. Learn.*, 2007.

[4] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

[5] S. Becker, J. Bobin, and E. J. Candès. NESTA: A fast and accurate first-order method for sparse recovery. *SIAM J. Imaging Sciences*, pages 1–39, 2011.

[6] M. Bertero and P. Boccacci. *Introduction to Inverse Problems in Imaging.* IOP Publishing, Bristol, 1998.

[7] D. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, 1982.

[8] D. Bertsekas. Projected Newton methods for optimization problem with simple constraints. *SIAM J. Control Optim.*, 20(2):221–245, 1982.

[9] D. Bertsekas. *Nonlinear Programming.* Athena Scientific, (2nd Edition), 1999.

[10] J. Bioucas-Dias and M. Figueiredo. A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Process.*, 16(12):29923004, 2007.

[11] P. Bloomfield and W. Steiger. *Least Absolute Deviations: Theory, Applications, and Algorithms.* Progress in probability and statistics. Birkhäuser, 1983.

[12] S. Bonettini, G. Landi, E. Loli Piccolomini, and L. Zanni. Scaling techniques for gradient projection-type methods in astronomical image deblurring. *Int. J. Comput. Math.*, 90(1):9–29, 2013.

[13] Y. Cai, Y. Sun, Y. Cheng, J. Li, and S. Goodison. Fast implementation of l1 regularized learning algorithms using gradient descent methods. In *Proc 10th SIAM International Conference on Data Mining (SDM)*, pages 862–871, 2010.

[14] E. Candès and J. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Found. Comput. Math.*, 6:227–254, 2006.

[15] E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inv. Prob.*, 23:969–985, 2007.

[16] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.*, 59(8):12071223, 2005.

[17] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489509, 2006.

[18] E. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory*, 52(12):5406 – 5425, 2006.

[19] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20:33–61, 1998.

[20] I. Daubechies, M. De Friese, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.*, 57:14131457, 2004.

[21] D. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):12891306, 2006.

[22] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Top. Signal Process.*, 1:586–597, 2007.

[23] H. Fu, M. K. Ng, M. Nikolova, and J. L. Barlow. Efficient minimization methods of mixed l2-l1 and l1-l1 norms for image restoration. *SIAM J. Sci. Comput.*, 27(6):1881–1902, 2006.

[24] E.M. Gafni and D.P. Bertsekas. Two-metric projection methods for constrained optimization. *SIAM J. Control. Optim.*, 22:936–964, 1984.

[25] T. Goldstein and S. Osher. The split Bregman method for l1 regularized problems. *SIAM J. Imag. Sci.*, 2(2):323–343, 2009.

[26] T. Goldstein and S. Setzer. High-order methods for basis pursuit. Technical Report CAM report 10-41, UCLA, 2010.

[27] P. Gong and C. Zhang. A fast dual projected Newton method for l1-regularized least squares. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume Two*, IJCAI'11, pages 1275–1280, 2011.

[28] M. Guerquin-Kern, J.-C. Baritaux, and M. Unser. Efficient image reconstruction under sparsity constraints with application to MRI and bioluminescence tomography. In *ICASSP*, pages 5760–5763. IEEE, 2011.

[29] X. Guo, F. Li, and M. K. Ng. A fast l1-TV algorithm for image restoration. *SIAM J. Sci. Comput.*, 31(3):2322–2341, 2009.

[30] E. T. Hale, W. Yin, and Y. Zhang. A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing. Technical Report CAAM TR07-07, Rice University, Houston, TX, 2007.

[31] E. T. Hale, W. Yin, and Y. Zhang. Fixed-point continuation for l1-minimization: Methodology and convergence. *SIAM J. Optim.*, 19:1107–1130, 2008.

[32] P. C. Hansen, J. Nagy, and D. P. O'Leary. *Deblurring images. Matrices, Spectra and Filtering*. SIAM, 2006.

[33] J. Kim and H. Park. Fast active-set-type algorithms for l1-regularized linear regression. In *In Proc. AISTAT*, pages 397–404, 2010.

[34] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An interior-point method for large-scale l1-regularized least squares. *IEEE J. Sel. Top. Signal Process.*, 1:606–617, 2007.

[35] S.-S. Kuo and R.J. Mammone. Image restoration by convex projections using adaptive constraints and the l1 norm. *Trans. Sig. Proc.*, 40(1), 1992.

[36] G. Landi and E. Loli Piccolomini. Quasi-Newton projection methods and the discrepancy principle in image restoration. *Appl. Math. Comput.*, 218(5):2091–2107, 2011.

[37] G. Landi and E. Loli Piccolomini. An efficient method for nonnegatively constrained Total Variation-based denoising of medical images corrupted by poisson noise. *Comp. Med. Imag. and Graph.*, 36(1):38–46, 2012.

[38] G. Landi and E. Loli Piccolomini. An improved Newton projection method for nonnegative deblurring of Poisson-corrupted images with Tikhonov regularization. *Numerical Algorithms*, 60(1):169–188, 2012.

[39] G. Landi and E. Loli Piccolomini. NPTool: a Matlab software for nonnegative image restoration with Newton projection methods. *Num. Alg.*, 62(3):487–504, 2013.

[40] M. M. Lavrentiev. *Some Improperly Posed Problems of Mathematical Physics.* New York:Springer-Verlag, 1967.

[41] I. Loris, M. Bertero, C. De Mol, R. Zanella, and L. Zanni. Accelerating gradient projection methods for l1-constrained signal recovery by steplength selection rules. *Appl. Comput. Harmon. Anal.*, 27(2):247–254, 2009.

[42] M. Nikolova. Minimizers of cost-functions involving nonsmooth data-fidelity terms. application to the processing of outliers. *SIAM J. Numer. Anal.*, 40:965994, 2002.

[43] W. W. Hager D. T. Phan and H. Zhang. Gradient-based methods for sparse recovery. *SIAM J. Imaging Sci.*, 4(1):146–165, 2011.

[44] M. Schmidt. *Graphical Model Structure Learning with L1-Regularization.* PhD thesis, University of British Columbia, 2010.

[45] M. Schmidt, G. Fung, and R. Rosales. Fast optimization methods for l1 regularization: A comparative study and two new approaches. In *Proceedings of the 18th European conference on Machine Learning*, ECML '07, pages 286–297. Springer-Verlag, 2007.

[46] M. Schmidt, G. Fung, and R. Rosales. Optimization methods for l1-regularization. Technical Report TR-2009-19, UBC, 2009.

[47] M. Schmidt, D. Kim, and S. Sra. *Optimization for Machine Learning*, chapter Projected Newton-type Methods in Machine Learning. MIT Press, 2011.

[48] R. Tibshirani T. Hastie and J. Friedman. *The Elements of Statistical Learning.* Springer Series in Statistics. New York: Springer-Verlag, 2001.

[49] U. Tautenhahn. On the method of Lavrentiev regularization for nonlinear illposed problems. *Inv. Prob.*, 18:191–207, 2002.

[50] R. Tibshirani. Regression shrinkage and selection via the Lasso. *J. Roy. Statist. Soc., ser. B*, 58(1):267288, 1996.

[51] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems*. John Wiley, New York, 1977.

[52] J. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inf. Theory*, 52(3):10301051, 2006.

[53] E. van den Berg and M. P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.*, 31:890–912, 2008.

[54] C. R. Vogel. *Computational Methods for Inverse Problems*. SIAM, Philadelphia, PA, USA, 2002.

[55] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. *SIAM J. Sci. Comput.*, 17:227–238, 1996.

[56] S. Wright, R. Nowak, and M. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.*, 57(7):24792493, 2009.

[57] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for l1-minimization with applications to compressed sensing. *SIAM J. Imaging Sci.*, 1:143–168, 2008.

[58] J. J. Zhang and B. Morini. Solving regularized linear least-squares problems by alternating direction methods with applications to image restoration. *ETNA*, 40:356–372, 2013.