Institute for Computational Mathematics
Hong Kong Baptist University

ICM Research Report
09-16

Solving Constrained Total-Variation Image Restoration and Reconstruction Problems via Alternating Direction Methods

Michael Ng[1], Pierre Weiss[2] and Xiao-Ming Yuan[3]

Abstract In this paper, we study alternating direction methods for solving constrained total-variation image restoration and reconstruction problems. Alternating direction methods can be implementable variants of the classical augmented Lagrangian method for optimization problems with separable structures and linear constraints. The proposed framework allows us to solve problems of image restoration, impulse noise removal, inpainting and image cartoon+texture decomposition. As the constrained model is employed, we only need to input the noise level and the estimation of the regularization parameter is not required in these imaging problems. Experimental results for such imaging problems are presented to illustrate the effectiveness of the proposed method. We show that the alternating direction method is very efficient for solving image restoration and reconstruction problems.

1 Introduction

Digital image restoration and reconstruction play an important role in various areas of applied sciences such as medical and astronomical imaging, film restoration, image and video coding and many others [47, 9, 58, 6, 43]. Most imaging systems capture an image x and return a degraded data x^0. A common model of the degradation process is the following :

$$x^0 = Ax + b \tag{1}$$

where b is an additive noise and $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transform (e.g. a convolution by a blurring kernel followed by a sub-sampling). Recovering x from x^0 is usually an ill-posed inverse problem and it should be regularized. Since the work of Rudin, Osher and Fatemi [57], the regularization methods based on total variation (TV) have known an important success, mostly due to their ability to preserve edges in the image.

Most of these techniques share the same formulation. They can be written as :

$$\begin{array}{rl} \min & |||\nabla x|||_1 + \tau \left|\left|Ax - x^0\right|\right|_N \\ \text{subject to} & x \in \mathbb{R}^n \end{array} \tag{2}$$

or the equivalent constrained form :

$$\begin{array}{rl} \min & |||\nabla x|||_1 \\ \text{subject to} & x \in \mathbb{R}^n,\ \left|\left|Ax - x^0\right|\right|_N \le \alpha \end{array} \tag{3}$$

[1]Centre for Mathematical Imaging and Vision and Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong. Research supported in part by HKRGC grants and HKBU FRGs.

[2]Centre for Mathematical Imaging and Vision, Hong Kong Baptist University, Kowloon Tong, Hong Kong.

[3]Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong. Research supported in part by an HKRGC grant and NSFC grant 10701055.

In these equations n is the pixels number. $\nabla : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ is a discrete version of the gradient. $|||\cdot|||_1$ represents a norm on $\mathbb{R}^n \times \mathbb{R}^n$. $||\cdot||_N$ is a norm or a semi-norm on $\mathbb{R}^m$. $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transform. α and τ are positive real numbers which measure the trade-off between the fit to x^0 and the amount of regularization.

Content of the paper A number of numerical methods have been proposed for solving instances of problem (2), see e.g. [57, 41, 54, 62, 10, 36, 37, 35, 42, 22, 39, 63, 29, 23]. To our knowledge, very few works tackle the constrained problems (3). This is surprising since choosing a reasonable value of α is usually much easier than finding a suitable value of τ. A natural choice proposed in the early papers [57] consists in setting $\alpha = ||b||_N$ or $\alpha = (1+\epsilon)\,||b||_N$ where ϵ is a small value.

Our aim in this work is to provide fast schemes for solving the constrained problems (3). We restrict our attention to a special case of (3) : we assume that $A = SH$ where H is a circulant matrix (e.g. a convolution by a blurring kernel) and S has a very simple structure described in Section 4. This structure is however sufficiently general to cover the cases of sub-sampling (super-resolution problems) and loss of partial information (inpainting problems). The proposed algorithms allow to treat the case $||\cdot||_N = ||\cdot||_p$, the classical l^p-norms, but also the norms associated to the Sobolev spaces of negative index $W^{-1,p}$. They thus allow to solve problems of image cartoon+texture decomposition, but also reconstruction with negative norms. Such problems are very demanding and not tackled well yet in the literature. The computational times are competitive with some recently proposed approaches.

By reformulating the problem (3) into some optimization problems with favorably separable structures, we notice that the problem (3) can be solved efficiently via the well-developed Alternating Direction Methods (ADMs). In short, ADMs are implementable variants of the classical Augmented Lagrangian method for optimization problems with separable structures and linear constraints, and they have been intensively studied in the community of optimization. We refer the reader to Section 3 for the detailed delineation of ADM. Overall, the aim of this paper is to develop ADM-based numerical algorithms for solving the constrained problem (3).

Related work Though this work was begun in early 2009 independently, the ADM method was introduced earlier in image processing in [18], where the author provides some equivalence with the so-called split-Bregman algorithm [29]. However, in this paper the authors only consider the case where the projection onto the constraint can be computed explicitly. In this paper, we will consider a more general constraint. Recently, Wu et al. [66] extended the augmented Lagrangian method to total variation restoration models with non-quadratic fidelities. They applied their method to TV restoration with l^1 and Kullback-Leibler (KL) fidelities. They showed the algorithm can have closed form solution and can also be efficiently solved via FFT implementation. The main differences between our paper and the Wu paper are that (i) the constrained model is studied and (ii) inpainting and image cartoon+texture decomposition are also considered and tested.

To our knowledge very few works tackle the case of constrained problems when the projection onto the set of constraints cannot be computed exactly. A common method consists in solving a sequence of unconstrained problems of type (2) in order to satisfy the constraint asymptotically. However, using this technique, a lot of time is lost by solving many instances of problem (3). Another technique used in [11] consists in using subgradient projections. Once again, this technique is known to have a bad asymptotic behavior. Moreover, it is not straightforward to compute the subgradient of the negative norms used in image cartoon+texture decomposition. Recently, the authors of [50] proposed a dual first order method. However, this technique can be used only when the objective function is strongly convex.

Outline of the paper The outline of this paper is as follows. In Section 2, we introduce the notations used in this paper. In Section 3, the Alternating Directions Method (ADM) based on the augmented Lagrangian is presented. Some theoretical aspects of this algorithm are presented. In Section 4, we present how to apply the ADM for solving problem (3) using different norms. The main difficulty consists in introducing correct equality constraints. In Section 5, numerical examples are given to demonstrate the effectiveness of the proposed method. Conclusions are made in Section 6.

2 Notations

In this work, we consider optimization problems in the finite dimensional vector space $\mathbb{R}^n$ ($n \in \mathbb{N}$). $\langle \cdot, \cdot \rangle$ denotes the canonical inner product and $||\cdot||$ denotes the associated norm : $||x|| = \sqrt{\langle x, x\rangle}$.

Now let $x \in \mathbb{R}^n$ be a 2D image. Then $n = n_1 \cdot n_2$ is the total number of pixels and n_1 and n_2 denote the number of pixels in the horizontal and vertical directions respectively. Let $i \in \{1, 2, ..., n_1\}$ and $j \in \{1, 2, ..., n_2\}$. $x(i, j)$ denotes the intensity value of pixel (i, j). The theory developed in this paper extends in a trivial manner to color images. We omit this case for the clarity of exposition. $\{e_{i,j}\}_{i,j}$ is the canonical basis of $\mathbb{R}^n$.

The notation $||\cdot||_p$ refers to the standard definition of an l^p-norm. In particular $||\cdot||_2 = ||\cdot||$. We also define another p-norm $|||\cdot|||_p$ on $\mathbb{R}^n \times \mathbb{R}^n$. Let $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^n$. We define $|y| = \sqrt{y_1^2 + y_2^2} \in \mathbb{R}^n$ and :

$$|||y|||_p = ||(|y|)||_p \, .$$

Note that $|||y|||_2 = ||y||$.

Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. A^* denotes its conjugate transpose. A^T denotes its transpose. We define $||A||$ as :

$$||A|| = \sup_{x \neq 0} \left(\frac{||Ax||}{||x||} \right).$$

This corresponds to the definition of the largest singular value of A. I denotes the identity matrix. $\mathcal{F}$ denotes the Discrete Fourier Transform (DFT) and $\mathcal{F}^{-1}$ denotes its inverse. Sometimes, we will also write $\hat{x} = \mathcal{F}x$.

Let us define the discrete gradient operator. $\partial_1 : \mathbb{R}^n \to \mathbb{R}^n$ and $\partial_2 : \mathbb{R}^n \to \mathbb{R}^n$ are linear applications. They design the discretized derivatives in the horizontal and vertical directions respectively. The gradient operator is denoted

$\nabla := \begin{bmatrix} \partial_1 \\ \partial_2 \end{bmatrix}$. All the experiments carried out in this paper are done using circular boundary conditions :

$$(\partial_1 x)(i,j) = \begin{cases} x(i+1,j) - x(i,j) & \text{if } i \leq n_1 \\ x(n_1,j) - x(1,j) & \text{if } i = n_1 \end{cases}$$

and

$$(\partial_2 x)(i,j) = \begin{cases} x(i,j+1) - x(i,j) & \text{if } j \leq n_2 \\ x(i,n_2) - x(1,j) & \text{if } j = n_2. \end{cases}$$

The interest of using circular boundary conditions is that the differential operators are diagonalized by the DFT. We have :

$$\partial_1 = \mathcal{F}^{-1} D_1 \mathcal{F} \text{ and } \partial_2 = \mathcal{F}^{-1} D_2 \mathcal{F}$$

where $D_1 \in \mathbb{C}^{n \times n}$ and $D_2 \in \mathbb{C}^{n \times n}$ are diagonal matrices. The counterpart of periodic boundary conditions is that it can lead to some artifacts on the image borders. Let us point out that L. Moisan proposed an arguably better discretization of the gradient, also based on a convolution in [46]. In this work we make use of the described operators for simplicity.

The operator div : $\mathbb{R}^n \to \mathbb{R}^n$ is a matrix which to an image x associates a discretization of its divergence. It defined by div $= -\nabla^T = -\left[\partial_1^T, \partial_2^T\right]$. The Laplacian of $x \in \mathbb{R}^n$ is defined by $\Delta x = \mathrm{div}(\nabla x)$

In all the paper $H : \mathbb{R}^n \to \mathbb{R}^n$ will design a circulant matrix (e.g. the blurring operator) which can be written as :

$$H = \mathcal{F}^{-1} D_H \mathcal{F}$$

where $D_H \in \mathbb{C}^{n \times n}$ is a diagonal matrix.

All the functions considered in this text are extended-real valued functions $f : \mathbb{R}^n \to [-\infty, +\infty]$ [4]. The domain of f is defined by $\mathrm{dom}(f) = \{x \in E, f(x) < +\infty\}$. Let $K \subseteq \mathbb{R}^n$. χ_K denotes the indicator function of K :

$$\chi_K(x) = \begin{cases} 0 & \text{if } x \in K \\ \infty & \text{otherwise} \end{cases}$$

Let $X \subseteq \mathbb{R}^n$ be a convex closed set. The projection on X is defined for $x^0 \in \mathbb{R}^n$ by :

$$\Pi_X(x^0) = \begin{array}{cc} \arg\min & \frac{1}{2}\left|\left|x - x^0\right|\right|^2 \\ \text{subject to} & x \in X \end{array}$$

We will also make use of the proximal operator made popular in signal processing in papers like [12]. Let f be a convex closed function. The proximal operator of f is defined for $x^0 \in \mathbb{R}^n$ by :

$$\mathrm{prox}_f\left(\ x^0\ \right) = \begin{array}{cc} \arg\min & f(x) + \frac{1}{2}\left|\left|x - x^0\right|\right| \\ \text{subject to} & x \in \mathbb{R}^n \end{array}$$

The proximal operator of $\beta\,|||\cdot|||_1$ will be denoted $\mathrm{shrink}_\beta(\cdot)$. For $y^0 = (y_1^0, y_2^0) \in \mathbb{R}^n \times \mathbb{R}^n$ it is defined by :

$$\mathrm{shrink}_\beta\left(y^0\right) = \mathrm{prox}_{\beta\,|||\cdot|||_1}\left(\ y^0\ \right) = \begin{array}{cc} \arg\min & \beta\,|||y|||_1 + \frac{1}{2}\left|\left|y - y^0\right|\right|^2 \\ \text{subject to} & y \in \mathbb{R}^n \times \mathbb{R}^n \end{array}$$

[4]this avoids having to consider constrained optimization problems

This operator has a closed form expression. Let $|y^0| \in \mathbb{R}^n$ be defined by $|y^0| = \sqrt{(y_1^0)^2 + (y_2^0)^2}$. Then

$$(\text{shrink}_\beta\left(y^0\right))(i,j) = y^0 - \min(\beta, |y^0|) \cdot \frac{y^0}{|y^0|}.$$

where $\frac{y^0}{|y^0|}$ should be taken as 0 on places where $|y^0| = 0$.

3 Review of ADM methods

In this section, we briefly review the well-known Alternating Direction Method (ADM) which has been well studied in the areas of convex programming and variational inequalities, see e.g. [5, 16, 17, 21, 24, 25, 26, 31, 33, 40, 60, 67].

We consider the following well-structured optimization problem that motivated the original presence of ADM in [25, 26, 31].

$$\begin{array}{ll} \min & f_1(x) + f_2(y) \\ s.t. & Bx + Cy = b, \\ & x \in X, y \in Y, \end{array} \tag{4}$$

where $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ are given convex sets, $f_1 : X \to \mathbb{R}$ and $f_2 : Y \to \mathbb{R}$ are closed proper convex functions ; $B \in \mathbb{R}^{l\times n}$ and $C \in \mathbb{R}^{l\times m}$ are given matrices ; and $b \in \mathbb{R}^l$ is a given vector. Obviously, the particular problem (3) falls into a special case of (4).

3.1 Algorithmic framework of ADM

By attaching the Lagrangian multiplier $\lambda \in \mathbb{R}^l$ to the linear constraint, the Augmented Lagrangian (AL) function of (4) is

$$\mathcal{L}(x, y, \lambda) := f_1(x) + f_2(y) + \langle \lambda, Bx + Cy - b\rangle + \frac{\beta}{2} \left|\left|Bx + Cy - b\right|\right|^2, \tag{5}$$

where $\beta > 0$ is the penalty parameter for the violation of the linear constraints, see e.g. [51]. Hence, with a given initial λ^0, the AL method approaches the solutions of the original problem (4), denoted by (x^*, y^*), by solving iteratively the following auxiliary problem :

$$\left\{ \begin{array}{lccl} (x^{k+1}, y^{k+1}) & \in & \arg\min & \mathcal{L}(x, y, \lambda^k) \\ & & \text{subject to} & x \in X \\ & & & y \in Y \\ \lambda^{k+1} & = & \lambda^k + \beta(Bx^{k+1} + Cy^{k+1} - b) & \end{array} \right. \tag{6}$$

Despite its popularity for solving generic optimization problem with equality constraints, it is not beneficial to implement directly the original AL method (6) to solve the particular problem (4). In fact, the nice separable structure emerging in both the objective function and the constraint of (4) is completely ignored by doing so, and thus x^{k+1} and y^{k+1} are required to be minimized simultaneously in (6). This pitfall, however, can be favorably overcame by the ADM. More specifically, to approach a solution of (4), the ADM solves the following sub problem at each iteration :

$$\left\{\begin{array}{lcl} x^{k+1} & \in & \begin{array}{cc} \arg\min & \mathcal{L}(x, y^k, \lambda^k) \\ \text{subject to} & x \in X \end{array} \\ y^{k+1} & \in & \begin{array}{cc} \arg\min & \mathcal{L}(x^{k+1}, y, \lambda^k) \\ \text{subject to} & y \in Y \end{array} \\ \lambda^{k+1} & = & \lambda^k + \beta(Bx^{k+1} + Cy^{k+1} - b) \end{array}\right. . \tag{7}$$

Note that ADM inherits the algorithmic framework of the AL method, but with the improvement of minimizing x^{k+1} and y^{k+1} serially via solving two lower-dimensional sub problems. In this sense, ADM is a variant of AL with practical features for solving structured problems like (4).

Hence, the ADM algorithmic framework for solving (4) as follows :

Algorithm 1: Alternating Direction Method for solving (4)

Input: The maximal number of iterations N ; The starting point $x^0 \in \text{dom}(f_1)$, $y^0 \in \text{dom}(f_2)$ and $\lambda^0 \in \mathbb{R}^l$; The initial value of β.

Output: x^N, an estimate of an element of x^* satisfying some stopping criterion.

begin

  for $k = 0$ *to* $N-1$ **do**

    Step 1. $x^{k+1} = \begin{array}{cc} \arg\min & \mathcal{L}(x, y^k, \lambda^k) \\ \text{subject to} & x \in X \end{array}$.

    Step 2. $y^{k+1} = \begin{array}{cc} \arg\min & \mathcal{L}(x^{k+1}, y, \lambda^k) \\ \text{subject to} & y \in Y \end{array}$.

    Step 3. $\lambda^{k+1} = \lambda^k + \beta(Bx^{k+1} + Cy^{k+1} - b)$.

  end

end

Remark 1. For a clear exposition of the main idea, we only present the algorithmic framework of the basic ADM. Some more advanced ADM type methods are also applicable for solving (4). For example, the general ADM in [31] modifies Step 3 of the basic ADM with a relaxation parameter in the interval $(0, \frac{\sqrt{5}+1}{2})$; and the method in [67] develops descent directions based on the iterates generated by the basic ADM.

3.2 Variational inequality reformulation

Let ∂f_1 and ∂f_2 denote the respective subdifferentials of f_1 and f_2 ; let $\Omega := X \times Y \times \mathbb{R}^l$. Then, it is easy to verify that solving (4) is equivalent to finding $(x^*, y^*, \lambda^*) \in \Omega$ such that

$$\left\{\begin{array}{l} (x - x^*)^T(\partial f_1(x^*) + B^T\lambda^*) \geq 0, \\ (y - y^*)^T(\partial f_2(y^*) + C^T\lambda^*) \geq 0, \quad \forall\, (x, y, \lambda) \in \Omega. \\ (\lambda - \lambda^*)^T(Bx^* + Cy^* - b) = 0, \end{array}\right. \tag{8}$$

where ∂f_i is understood as any element of the corresponding subdifferential without ambiguity. Hence, we have the following variational inequality (VI) reformulation of (4) : Find $u^* \in \Omega$ such that

$$(u - u^*)^T F(u) \geq 0, \qquad \forall\, u' \in \Omega,$$

where

$$u = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix} \qquad \text{and} \qquad F(u) = \begin{pmatrix} \partial f_1(x) + B^T\lambda \\ \partial f_2(x) + C^T\lambda \\ Bx + Cy - b \end{pmatrix}.$$

3.3 Stopping criterion

In this subsection, we specify an implementable stopping criterion for applying the ADM.

Note that the iterate $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ generated by ADM satisfies

$$\left\{ \begin{array}{l} (x - x^{k+1})^T \left\{\partial f_1(x^{k+1}) + B^T\lambda^{k+1} + \beta B^T C(y^k - y^{k+1})\right\} \geq 0,\ \forall x \in X, \\ (y - y^{k+1})^T \left\{\partial f_2(y^{k+1}) + C^T\lambda^{k+1}\right\} \geq 0,\ \forall y \in Y, \\ (Bx^{k+1} + Cy^{k+1} - b) = \frac{1}{\beta}(\lambda^{k+1} - \lambda^k). \end{array} \right. \tag{9}$$

Therefore, (9) enable us to design an easy-for-check stopping criterion for implementing ADM, see also [68]. More specifically, based on (8) and (9), it is clear that $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is a solution of (8) if and only if $y^k = y^{k+1}$ and $\lambda^k = \lambda^{k+1}$. This observation motivates us to develop a stopping criterion for implementing ADM in the following manner :

$$\max\{ey^k, ez^k\} \leq \epsilon, \tag{10}$$

where $\epsilon > 0$ and

$$ey^k := \beta\|B^T C(y^k - y^{k+1})\| \qquad \text{and} \qquad ez^k := \frac{1}{\beta}\|(\lambda^k - \lambda^{k+1})\|. \tag{11}$$

3.4 A strategy of adjusting penalty parameter

Theoretically, ADM type methods are convergent for any constant $\beta > 0$, see e.g. [?, 25, 26, 31]. In practice, however, it is necessary to adjust the values of β dynamically in order to achieve better numerical performance. We refer to, e.g. [33, 34, 40], for some existing strategies of adjusting the variable sequence of penalty parameter $\{\beta_k\}$ self-adaptively. Convergence of ADM with variable penalty parameters can also been found there.

The property implied by (9) not only motivates us to present the stopping criterion (10)-(11) for implementing ADM, but also induces us to present the following strategy of adjusting β with the purpose of balancing ey^k and ez^k iteratively :

$$\beta_{k+1} = \left\{ \begin{array}{ll} \frac{2}{3}\beta_k & \text{if } ez^k \leq 0.9\epsilon, \\ \beta_k & \text{otherwise.} \end{array} \right.$$

In Section 5, we will use this strategy to improve the convergence of the ADM method.

4 Application of the ADM for constrained TV problems

In this section, we focus on applying the ADM to solve some interesting scenarios of the model (3) in image restoration and reconstruction.

4.1 $TV - l^2$ deconvolution problems

In this paragraph, we show how to apply the ADM method for solving the constrained $TV - l^2$ model. This model is adapted to the restoration of a blurry image with additive Gaussian noise. The problem under consideration is the following :

$$\begin{array}{rl} \min & |||\nabla x|||_1 \\ \text{subject to} & x \in \mathbb{R}^n, \left|\left|Hx - x^0\right|\right|^2 \leq \alpha \end{array} \tag{12}$$

where $H : \mathbb{R}^n \to \mathbb{R}^n$ is a spatially invariant blur. We also assume that the set $K = \{x \in \mathbb{R}^n, \left|\left|Hx - x^0\right|\right|^2 \leq \alpha\}$ is non-empty, which ensures that there exists a convex set of minimizers.

To apply the preceding theory, the first step consists in splitting the minimization of $|||\nabla x|||_1$ and the projection onto K. In this paragraph we present an efficient solution for the $TV - l^2$ problem. An alternative will be presented in the next section.

Problem (12) can be rewritten as :

$$\begin{array}{rl} \min & |||y|||_1 \\ \text{subject to} & x \in K \\ & y \in \mathbb{R}^n \times \mathbb{R}^n, y = \nabla x \end{array} \tag{13}$$

This problem fits the framework (4) by choosing : $f_1(x) = \chi_K(x)$, $f_2(y) = |||y|||_1$, $B = \nabla$, $C = -I$ and $b = 0$.

The augmented Lagrangian functional associated to this problem is defined by :

$$\mathcal{L}(x, y, \lambda) = |||y|||_1 + \langle \lambda, \nabla x - y \rangle + \frac{\beta}{2} ||\nabla x - y||^2 + \chi_K(x)$$

The main difficulty in order to apply the ADM method consists in solving steps 1 and 2 in Algorithm 1. Let us show that it can be achieved efficiently.

Step 1 amounts to solving :

$$\text{Find } x^{k+1} \in \quad \begin{array}{rl} \arg\min & \langle \lambda^k, \nabla x - y^k \rangle + \frac{\beta}{2} \left|\left|\nabla x - y^k\right|\right|^2 \\ \text{subject to} & x \in \mathbb{R}^n, \left|\left|Hx - x^0\right|\right|^2 \leq \alpha \end{array}$$

This is a constrained least square problem. Its solution can be computed efficiently if the singular value decompositions of the matrices H and ∇ are given explicitly [30, p.580]. This is the case for our problem. We describe such an algorithm in the appendix. In large dimensions, the overall cost of this algorithm is dominated by two DFTs.

Step 2 amounts to solving

$$\begin{array}{rrl} y^{k+1} \in & \arg\min & \||y\||_1 - \langle \lambda^k, y\rangle + \frac{\beta}{2} \left|\left| y - \nabla x^{k+1} \right|\right|^2 \\ & \text{subject to} & y \in \mathbb{R}^n \times \mathbb{R}^n \\ \Leftrightarrow \quad y^{k+1} = & \arg\min & \||y\||_1 + \frac{\beta}{2} \left|\left| y - \left(\nabla x^{k+1} + \frac{\lambda^k}{\beta}^2 \right) \right|\right|^2 \\ & \text{subject to} & y \in \mathbb{R}^n \times \mathbb{R}^n \end{array}$$

$$\Leftrightarrow \quad y^{k+1} = \mathrm{shrink}_{\frac{1}{\beta}} \left(\nabla x^{k+1} + \frac{\lambda^k}{\beta} \right).$$

We see that both Steps 1 and 2 can be computed explicitly : the ADM is thus implementable.

4.2 $TV - l^p$ super-resolution, inpainting and deconvolution problems

Image deconvolution, inpainting and super-resolution are prototypical image reconstruction problems. In this paper we focus on the case where a single image has been convoluted and suffered some loss of information. This image formation model can be written as :

$$x^0 = SHx + b \tag{14}$$

where

- H is a spatially invariant convolution (or more generally a circulant matrix).
- b is an additive white noise. We will focus only on the cases of Laplacian, Gaussian and uniform noise in this paper. We assume that the noise distribution is the same for every pixel of the image domain. This hypothesis is not absolutely necessary but simplifies the notations.
- $S : \mathbb{R}^n \to \mathbb{R}^m$ is a rectangular matrix with a special structure. We assume that $S = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_m \end{bmatrix}$ and that the lines S_k of S satisfy $S_k = s_k e_{i(k),j(k)}$ with $s_k \in \mathbb{R}$. This structure, despite its simplicity, covers important applications :
  - **Vignetting and inpainting.** If S is square and diagonal, it can represent a variable lighting of the scene : the intensity of every pixel k is multiplied by a factor $s_k \in \mathbb{R}$. This effect often occurs both in traditional and digital photography and is called vignetting. If we set $s_k = 0$ for $k \in K$ and $s_k = 1$ on the complementary of K then the degradation (14) represents a loss of information on K. This equation can be used to model the degradation occurring in inpainting problems.
  - **Sub-sampling.** Let $x \in \mathbb{R}^n$, suppose that n_1 and n_2 are multiple of $s \in \mathbb{N}$. Let $i \in \{1, 2, ..., \frac{n_1}{s}\}$ and $j \in \{1, 2, ..., \frac{n_2}{s}\}$. Now let us define S as :

    $$(Sx)(i, j) = x(i \cdot s, j \cdot s).$$

    This mapping can be represented as a matrix S with the structure described earlier. It is a sub-sampling by a factor s in the vertical and horizontal directions.

– Clearly, we can also consider cases where vignetting, sub-sampling and loss of information occur all together.

The problem of image reconstruction we tackle in this paper consists in recovering x from x^0 knowing, S, H and the statistical properties of the noise. The variational formulation of this problem is the following :

$$\text{Find } x^* \in \begin{array}{ll} \arg\min & |||\nabla x|||_1 \\ \text{subject to} & x \in \mathbb{R}^n, \left|\left|SHx - x^0\right|\right|_p \leq \alpha \end{array} \tag{15}$$

In the case of impulsive or Laplacian noise p should be equal to one [49], while it should be equal to 2 for Gaussian noise and to ∞ for uniform noise. We assume that the set $\{x \in \mathbb{R}^n, \left|\left|SHx - x^0\right|\right|_p \leq \alpha\}$ is non-empty. This is a sufficient condition to ensure existence of a solution.

Now, let us show how the ADM method can be applied to solve problem (15). For this problem, solving Step 1 of Algorithm 1 is too complicated if we use the same strategy as in (13). We use two equality constraints instead of using just one. Problem (15) can be rewritten as :

$$\text{Find } x^* \in \begin{array}{ll} \arg\min & |||w|||_1 \\ \text{subject to} & x \in \mathbb{R}^n \\ & w \in \mathbb{R}^n \times \mathbb{R}^n, w = \nabla x \\ & z \in \mathbb{R}^n, z = Hx, \left|\left|Sz - x^0\right|\right|_p \leq \alpha \end{array} \tag{16}$$

Let $Z = \{z \in \mathbb{R}^n, \left|\left|Sz - x^0\right|\right|_p \leq \alpha\}$. This problem can be casted in the framework (4) using the following notations : $y = \begin{pmatrix} w \\ z \end{pmatrix}$, $f_1(x) = 0$, $f_2(y) = |||w|||_1 + \chi_Z(z)$, $B = \begin{bmatrix} \nabla \\ H \end{bmatrix}$, $C = \begin{bmatrix} -I \\ -I \end{bmatrix}$ and $b = 0$.

The augmented Lagrangian of this problem writes :

$$\mathcal{L}(x, y, \lambda) = f_2(y) + \langle \lambda, Bx - y \rangle + \frac{\beta}{2} ||Bx - y||^2 .$$

Step 1 amounts to solving :

$$\text{Find } x^{k+1} \in \begin{array}{ll} \arg\min & \langle \lambda^k, Bx - y^k \rangle + \frac{\beta}{2} \left|\left|Bx - y^k\right|\right|^2 \\ \text{subject to} & x \in \mathbb{R}^n \end{array}$$

x^k is thus the solution of a linear system which can be diagionalized by the DFT. This makes no difficulty and allows to avoid the technicalities of the previous example due to the constrained least square problem.

Step 2 amounts to solving :

$$\text{Find } y^{k+1} \in \begin{array}{ll} \arg\min & f_2(y) + \langle \lambda^k, Bx^{k+1} - y \rangle + \frac{\beta}{2} \left|\left|y - Bx^{k+1}\right|\right|^2 \\ \text{subject to} & y \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \end{array}$$

The solution of this problem is unique due to strong convexity of $||\cdot||^2$ and we can continue as follows :

$$\begin{array}{rcll} y^{k+1} & = & \arg\min & f_2(y) + \langle \lambda^k, Bx^{k+1} - y\rangle + \frac{\beta}{2}\left|\left|y - Bx^{k+1}\right|\right|^2 \\ & & \text{subject to} & y \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \\ & = & \arg\min & f_2(y) + \frac{\beta}{2}\left|\left|y - \left(Bx^{k+1} + \frac{\lambda^k}{\beta}\right)\right|\right|^2 \\ & & \text{subject to} & y \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \\ & = & \multicolumn{2}{l}{\mathrm{prox}_{f_2/\beta}\left(\; Bx^{k+1} + \frac{\lambda^k}{\beta} \; \right).} \end{array}$$

Then we can write $\begin{pmatrix} w^k \\ z^k \end{pmatrix} = Bx^{k+1} + \frac{\lambda^k}{\beta}$ and decompose $\mathrm{prox}_{f_2/\beta}\left(\; \cdot \; \right)$ as :

$$\mathrm{prox}_{f_2/\beta}\begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} \mathrm{prox}_{|||\cdot|||_1/\beta}\left(\; w \; \right) \\ \mathrm{prox}_{\chi_K/\beta}\left(\; z \; \right) \end{pmatrix}$$

We already proved that $\mathrm{prox}_{|||\cdot|||_1/\beta}\left(\; w \; \right) = \mathrm{shrink}_{1/\beta}(w)$. Now let us focus on

$$\begin{array}{rcll} \mathrm{prox}_{\chi_K/\beta}\left(\; z \; \right) & = & \multicolumn{2}{l}{\Pi_K(z)} \\ & = & \arg\min & \frac{1}{2}\left|\left|z' - z\right|\right|^2 \\ & & \text{subject to} & z' \in \mathbb{R}^n \\ & & & \left|\left|Sz' - x^0\right|\right|_p \leq \alpha \end{array}$$

Due to the structure of S, this operation is a projection onto a weighted l^p-ball. This can be computed explicitly or very accurately. We refer the reader to the appendix of [64] for a detailed implementation.

4.3 Image cartoon+texture decomposition problems

Despite a wide amount of publications related to total variation reconstruction problems, the practical results obtained using this prior on natural images are usually not as good as those obtained using harmonic analysis transforms for instance. One possible reason is that natural images do not seem to be of bounded total variation [32] as they contain details at very small scales.

To our belief, one of the main interests of total variation is that it allows to simplify the image content by removing erratic information. This idea was formalized by Y. Meyer who proposed new total variation based models in his book [45]. There he describes variational methods which aim is to decompose an image x^0 into the sum of two components : a "cartoon" u and a "texture" v. This kind of decomposition has a large number of potential applications : image segmentation, texture analysis, separate compression or inpainting of the cartoon and the textured parts... The basic idea of Y. Meyer is to define two functionals f_1 and f_2 and to look for the solution of :

$$\begin{array}{rll} u^* = & \arg\min & f_1(u) + f_2(v) \\ & \text{subject to} & u \in \mathbb{R}^n, v \in \mathbb{R}^n \\ & & x^0 = u + v \end{array}$$

This model should lead to a good decomposition if the two following conditions are fulfilled simultaneously :

– $f_1(v) \gg f_2(v)$ if v is an "oscillating" component.
– $f_1(u) \ll f_2(u)$ if u is a "cartoon" component.

In his work [45], Y. Meyer proposes to choose $f_1(u) = TV(u)$ and $f_2(v) = ||v||_G$ where $||\cdot||_G$ is a norm on the negative Sobolev space $W_0^{-1,\infty}$. In the discrete setting, this norm is simply the dual of the total variation. One of the reasons for the choice of this norm is a nice result of functional analysis (see [45, 1] for more precise results and definitions) :

$$\text{If} \quad v_n \rightharpoonup 0 \text{ then } \quad ||v_n||_G \to 0,$$

where $\rightharpoonup$ denotes the weak convergence. This indicates that an oscillating function has a low G-norm.

From a numerical point of view, this model was investigated in [55, 2, 28, 64]. However, we believe that the numerical schemes used in order to solve decomposition problems still lack precision or efficiency (see e.g. [65]).

After Meyer's model was proposed, a number of variants and new decomposition models were proposed (see e.g. [56, 59, 4]). Recently the dual norms were used not only for image decomposition but also for image reconstruction [14, 53, 52]. These kinds of techniques can be of interest when the noise distribution is not white or when only the "cartoon" part of the image should be retrieved. The numerical experiments led in [52] also show that they allow to capture white noise better than the l^2-norm for geometric images.

Our aim in this paragraph is to show that the ADM method allows to solve these kinds of problems efficiently. We investigate only the total variation based models. The ADM can clearly be successful for other related problems and we hope the given numerical results will encourage people using this technique.

4.3.1 Dual norm and its computation

Let us describe the problem under consideration. We refer the reader to [3] for more detailed comments. Let

$$X = \{x \in \mathbb{R}^n, \sum_{i,j} x(i,j) = 0\}.$$

We define the following norm on X :

$$||x||_{-1,p} = \begin{array}{cc} \min & |||g|||_p \\ \text{subject to} & g = (g_1, g_2) \in \mathbb{R}^n \times \mathbb{R}^n, \text{div} g = x \end{array}$$

This is possible since div is surjective from $\mathbb{R}^n \times \mathbb{R}^n$ to X. In the case $p = \infty$, this norm corresponds to the discrete version of Meyer's G-norm [2]. In the case $p = 2$, it corresponds to the discrete H^{-1}-norm proposed in [56]. The case $p = 1$ was already investigated numerically in [64] and led to good practical decomposition results. In the continuous setting, this norm would be adapted to the Sobolev space of negative index $W_0^{-1,1}$.

The computation of such dual norms is not straightforward. However, it is important to have an estimate in order to assess - e.g. - the empirical convergence rates of iterative algorithms. A dichotomic solution is proposed in [3], but the authors report that the computation is not accurate. Computation of such dual norms can be achieved efficiently using the ADM. It suffices to remark that :

$$\begin{array}{rrl} ||x||_{-1,p} = & \min & |||g|||_p \\ & \text{subject to} & g \in \mathbb{R}^n \times \mathbb{R}^n, g = y \\ & & y \in \mathbb{R}^n \times \mathbb{R}^n, \mathrm{div} y = x \end{array}$$

This problem is a simple instance of (4). At each iteration, the proximal operator of l^p-norms must be computed, which can be done explicitly for $p \in \{1, 2, \infty\}$. We do not detail this approach further and turn to the more demanding reconstruction problem.

4.3.2 Reconstruction problem

The class of problems we consider in this paper write :

$$\begin{array}{rrl} \text{Find } x^* \in & \arg\min & |||\nabla x|||_1 \\ & \text{subject to} & \left|\left|SHx - x^0\right|\right|_{-1,p} \leq \alpha \end{array} \tag{17}$$

This problem is clearly more complicated than the previous ones and applying the ADM method is tricky. The first remark in order to apply it consists in rewriting (17) as :

$$\begin{array}{rl} \arg\min & |||\nabla x|||_1 \\ \text{subject to} & x \in \mathbb{R}^n \\ & g \in G, \mathrm{div} g = SHx - x^0, \end{array} \tag{18}$$

where $G = \{g \in \mathbb{R}^m \times \mathbb{R}^m, |||g|||_p \leq \alpha\}$. This is possible since all terms in the problem are convex. Problem (18) can be tackled directly by the ADM method. However the subproblems 1 and 2 in this case are still difficult convex problems and can only be solved by iterative methods. In order to make these subproblems easier, we need to introduce other linear equality constraints. Problem (18) is also equivalent to :

$$\begin{array}{rl} \arg\min & |||w|||_1 \\ \text{subject to} & x \in \mathbb{R}^n \\ & y \in \mathbb{R}^m \times \mathbb{R}^m, \mathrm{div} y = Sz - x^0 \\ & g \in G, g = y \\ & w \in \mathbb{R}^n \times \mathbb{R}^n, w = \nabla x \\ & z \in \mathbb{R}^n, z = Hx \end{array} \tag{19}$$

By setting : $f_1(x, y) = 0$, $f_2(g, w, z) = |||w|||_1 + \chi_G(g)$ and

$$A = \begin{bmatrix} \nabla & 0 \\ H & 0 \\ 0 & \mathrm{div} \\ 0 & I \end{bmatrix} \quad B = \begin{bmatrix} 0 & -I & 0 \\ 0 & 0 & -I \\ 0 & 0 & -S \\ -I & 0 & 0 \end{bmatrix} \quad \text{and} \quad b = \begin{pmatrix} 0 \\ 0 \\ -x^0 \\ 0 \end{pmatrix},$$

problem (19) rewrites :

$$\begin{array}{rl} \arg\min & f_1(x, y) + f_2(g, w, z) \\ \text{subject to} & A(x, y) + B(g, w, z) = b \end{array}$$

This is clearly an instance of the class of problems (4). The introduction of 4 equality constraints is the most straightforward way to ensure that both subproblems 1 and 2 can be solved exactly. The augmented Lagrangian associated to problem (19) writes :

$$\begin{aligned}\mathcal{L}((x,y),(g,w,z),\lambda) &= f_2(g,w,z) + \langle \lambda, A(x,y) + B(g,w,z) - b\rangle \\ &\quad + \frac{\beta}{2}\left|\left|A(x,y) + B(g,w,z) - b\right|\right|^2,\end{aligned}$$

where $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. Now let us show that the ADM method can be used efficiently for solving (19).

Step 1 amounts to minimizing $\mathcal{L}((x,y),(g^k,w^k,z^k),\lambda^k)$ w.r.t (x,y). The solution (x^{k+1},y^{k+1}) satisfies :

$$\beta A^T A(x^{k+1},y^{k+1}) = -A^T\left(\lambda + \beta(B(g^k,w^k,z^k) - b)\right), \tag{20}$$

where

$$A^T A = \begin{bmatrix} \nabla^T\nabla + H^T H & 0 \\ 0 & \nabla\nabla^T + I \end{bmatrix}.$$

The resolution of (20) can be done in $O(n\log(n))$ operations using the DFT.

Step2 amounts to minimizing $\mathcal{L}((x^{k+1},y^{k+1}),(g,w,z),\lambda^k)$ w.r.t (g,w,z). This writes :

$$\begin{aligned}(g^{k+1},w^{k+1},z^{k+1}) \in \quad & \arg\min \quad |||w|||_1 + \langle \lambda_1^k, \nabla x^{k+1} - w\rangle + \tfrac{\beta}{2}\left|\left|\nabla x^{k+1} - w\right|\right|^2 \\ & \qquad + \chi_G(g) + \langle \lambda_4^k, y^{k+1} - g\rangle + \tfrac{\beta}{2}\left|\left|y^{k+1} - g\right|\right|^2 \\ & \qquad + \langle \lambda_2^k, Hx^{k+1} - z\rangle + \tfrac{\beta}{2}\left|\left|Hx^{k+1} - z\right|\right|^2 \\ & \qquad + \langle \lambda_3^k, \mathrm{div} y^{k+1} - Sz + x^0\rangle + \tfrac{\beta}{2}\left|\left|\mathrm{div} y^{k+1} - Sz + x^0\right|\right|^2 \\ & \text{subject to} \quad w \in \mathbb{R}^n \times \mathbb{R}^n \\ & \qquad g \in \mathbb{R}^m \times \mathbb{R}^m \\ & \qquad z \in \mathbb{R}^n\end{aligned}$$

To solve this problem we first remark that the variables w, g and z are not coupled. Using the same strategy as in the previous examples it is easy to get :

$$w^{k+1} = \mathrm{shrink}_{1/\beta}\left(\nabla x^{k+1} + \frac{\lambda_1^k}{\beta}\right)$$

and

$$g^{k+1} = \Pi_G\left(y^{k+1} + \frac{\lambda_4^k}{\beta}\right).$$

Finally z^{k+1} is the solution of the following diagonal linear system :

$$\beta(I + S^T S)z^{k+1} = \lambda_2^k + \beta H x^{k+1} + S^T\left(\lambda_3^k + \beta\mathrm{div} y^{k+1} + \beta x^0\right).$$

We see that every step of the ADM has an explicit solution.

4.4 Metric change

We showed that every step of the ADM method could be performed exactly for various imaging problems of interest. In practice, we observed that it is important to refine the ADM scheme in order to get better performances. There are two natural ways of improving the performances of the scheme : changing the metric and using an adaptive "step-size" $\{\beta^k\}$ instead of a constant β. We already described the idea of adjusting β dynamically in Section 3.4.

We only describe the idea of using a different metric for the TV-l^p problem considered in Section 4.2. The other examples can be generalized easily. The TV-l^p problem is formulated as in (16). By choosing two symmetric definite positive matrices $W_\nabla : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ and $W_H : \mathbb{R}^n \to \mathbb{R}^n$, problem (16) can be reformulated as :

$$\begin{array}{lll} \text{Find } \ x^* \in & \arg\min & |||w|||_1 \\ & \text{subject to} & x \in \mathbb{R}^n \\ & & w \in \mathbb{R}^n \times \mathbb{R}^n, W_\nabla(w - \nabla x) = 0 \\ & & z \in \mathbb{R}^n, W_H(z - Hx) = 0, \left|\left|Sz - x^0\right|\right|_p \leq \alpha \end{array} \tag{21}$$

The augmented Lagrangian associated to this problem is defined by :

$$\begin{aligned} \mathcal{L}(x, w, z) = |||w|||_1 &+ \langle \lambda_1, W_\nabla(w - \nabla x)\rangle + \frac{\beta}{2} \left|\left|W_\nabla(w - \nabla x)\right|\right|_2^2 \\ &+ \langle \lambda_2, W_H(z - Hx)\rangle + \frac{\beta}{2} \left|\left|W_H(z - Hx)\right|\right|_2^2 . \end{aligned} \tag{22}$$

We can define a new scalar product $\langle \cdot, \cdot \rangle_W$ on $\mathbb{R}^{2n} \times \mathbb{R}^{2n}$:

$$\langle (w_1, z_1), (w_2, z_2)\rangle_W := \langle (w_1, z_1), (W_\nabla w_2, W_H z_2)\rangle_W$$

and a norm on $\mathbb{R}^{2n} \times \mathbb{R}^{2n}$: $||x||_W := \sqrt{\langle x, x\rangle_W}$. Doing so and using the same notations as in Section 4.2, the augmented Lagrangian (22) rewrites :

$$\mathcal{L}(x, y) = |||w|||_1 + \langle \lambda, Bx + Cy - b\rangle_W + \frac{\beta}{2} \left|\left|Bx + Cy - b\right|\right|_W^2 . \tag{23}$$

The ADM method can be shown to converge in any metric and the possibility of modifying W_H and W_∇ can thus lead to better numerical performances [31]. In this work, we simply choose :

$$W_H = \delta I \quad \text{and} \quad W_\nabla = I \tag{24}$$

where $\delta > 0$. A good value of δ can be found by the trial and error approach, by comparing the convergence behaviors. We found out that the “optimal” value of δ is independent of the images, but depends on the image restoration and reconstruction problem.

5 Numerical experiments

We divide the numerical section into two parts. In the first one, we study the effectiveness of the algorithms. In the second one, we show some numerical results of the image restoration and reconstruction models discussed in Section 4.

5.1 Convergence behavior

5.1.1 Comparisons on TV-l^2

The first experiment consists in comparing three approaches on the TV-l^2 deconvolution problem :

- The ADM approach based on the secular equation in Section 4.1.
- The ADM approach using the splitting of the gradient and the convolution operator in Section 4.2.
- The Nesterov approach described in [64], which can be regarded as an improvement of the gradient descent method. This technique can be used only when the projection onto the constraints set can be done explicitly.

The experimental protocol is as follows. We convolve the 512×512 Barbara image with a uniform kernel of size 5×5. We add Gaussian noise to the image. We run each method and compute the different characteristics of the methods.

According to Figure 1, we see that the ADM approach using the secular equation is the most efficient. We use the SNR as a comparison criterion as one of the interests of this method is that it leads to visually satisfactory solutions very fast. In this example, the ADM method based on the secular equation gives stable solutions in around 1 second on a laptop for a 512×512 image.

5.1.2 Comparisons on TV-l^1

In the following tests, we use the TV-l^1 deconvolution problem (deconvolution with impulse noise).

Dynamic β We first intend to show that the variable adjustment of β proposed in Section 3.4 allows to pick values β^k which are sub-optimal, but reasonable and *automatic*. In figure (5.1.2), we see that a near optimal, fixed value of β would be 1000 as with this value, the algorithm converges very quickly to a high SNR. In this example, the variable strategy described in (3.4) first leads to a value of β^k close to 100 and increases regularly up to 400. In practice, it would thus be better to replace these values of β^k with a fixed $\beta = 1000$, but using the described strategy, the tuning is done automatically and leads to satisfactory results in all cases we tested. Therefore, this technique should be adopted as it has a very low cost at each iteration.

Metric change In this experiment, we show that changing the metric, as described in paragraph Section 4.4 is *crucial* in order to get a good convergence. We show it through the TV-l^1 deconvolution problem with impulse noise. We use the variable strategy described in Section 3.4 for choosing β. Then we test different values of δ in the convergence speed. Figure (5.1.2) shows the convergence behavior of the algorithm for various choices of δ. For instance, if the user chooses $\delta = 1$, the convergence is very slow (at least 10^4 iterations are required to get a visually stable solution). For $\delta = 100$ (which is the experimental optimal value of δ), a "good" solution is obtained in no more than 100 iterations. This is comparable to state-of-the-art results for the unconstrained version of the TV-l^1 problem.

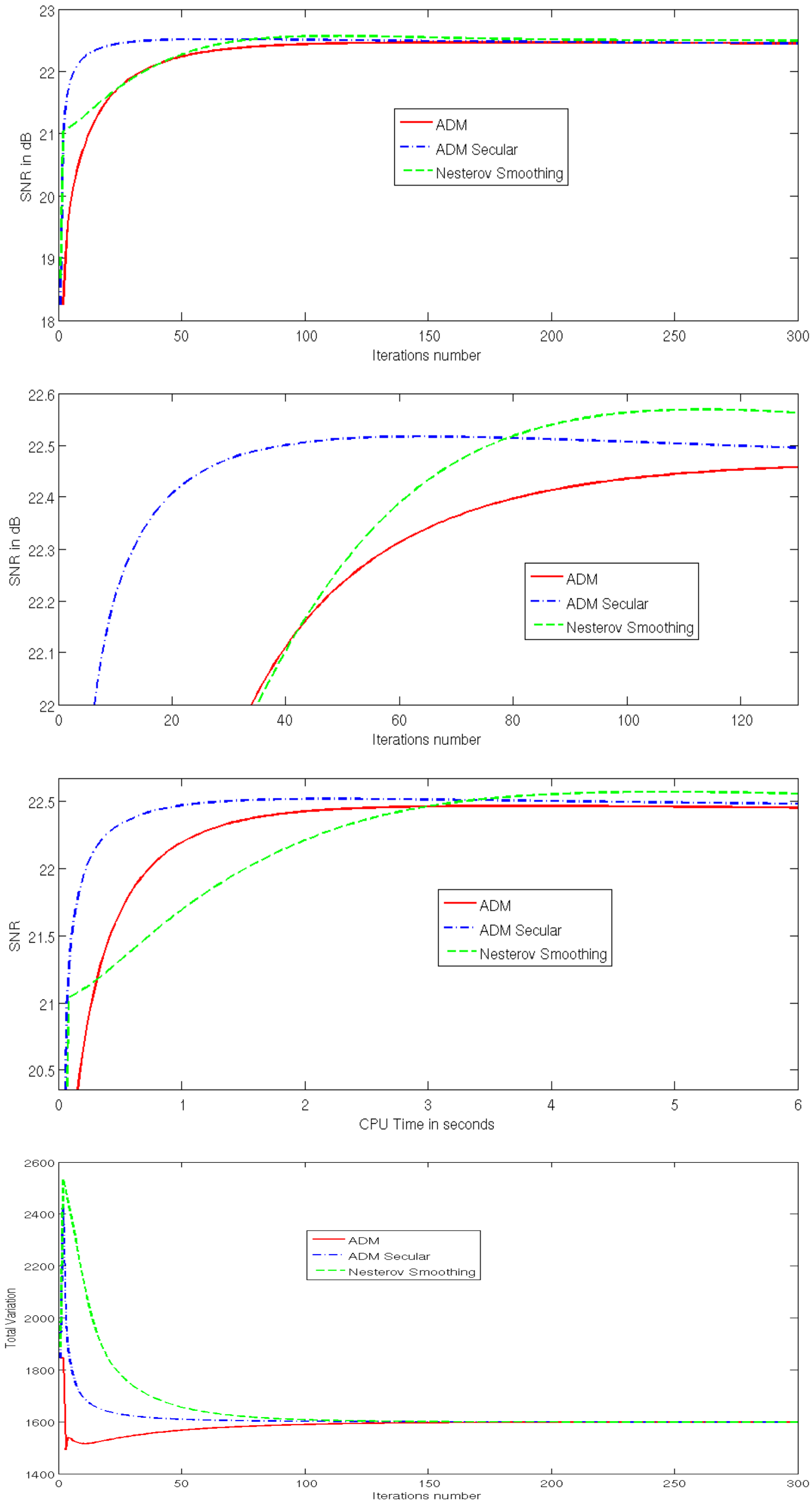

FIG. 1 – From top to bottom : SNR w.r.t. iterations ; a zoom on SNR w.r.t. iterations ; SNR w.r.t. computing times ; and total variation w.r.t. iterations.

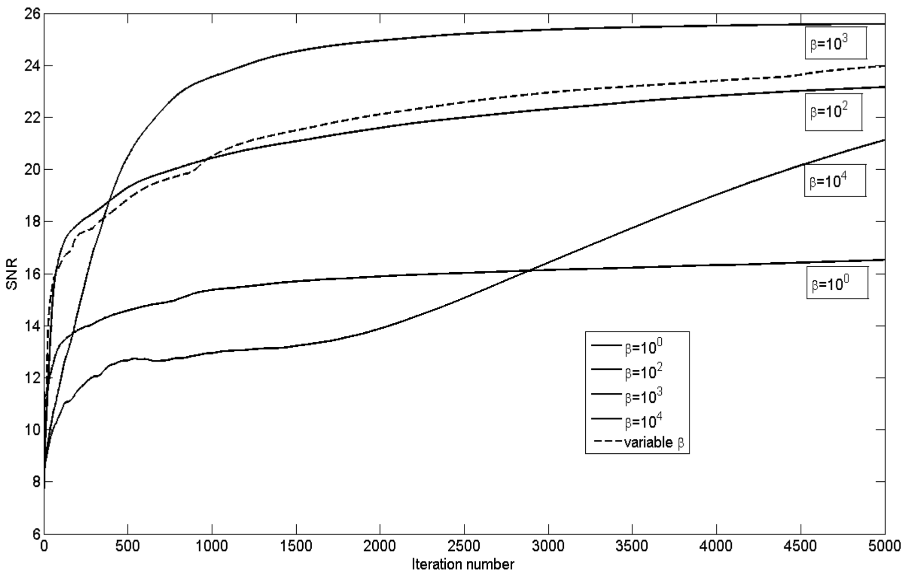

FIG. 2 – SNR w.r.t. iterations for different fixed values of β and a dynamic choice of β as described in Section 3.4.

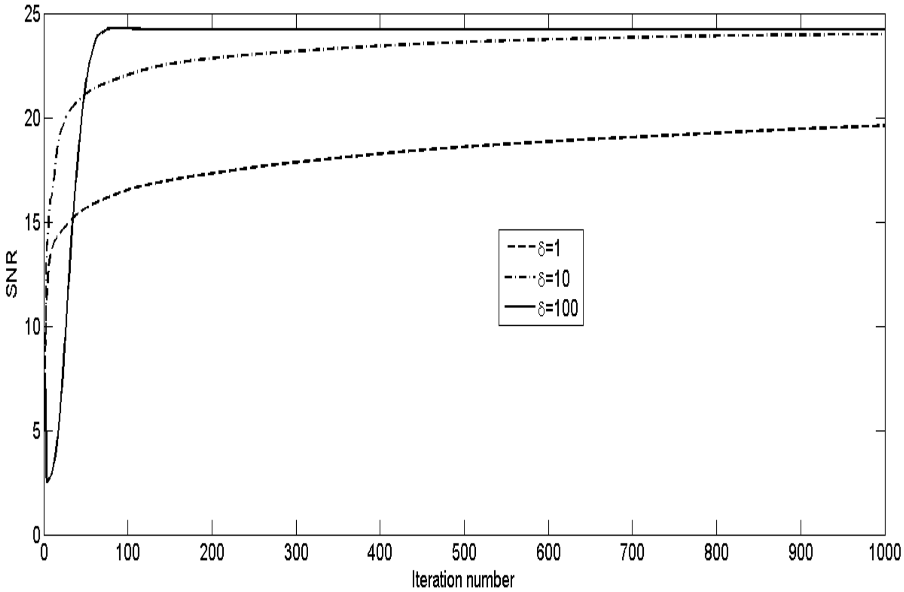

FIG. 3 – SNR w.r.t. iterations for different fixed values of δ.

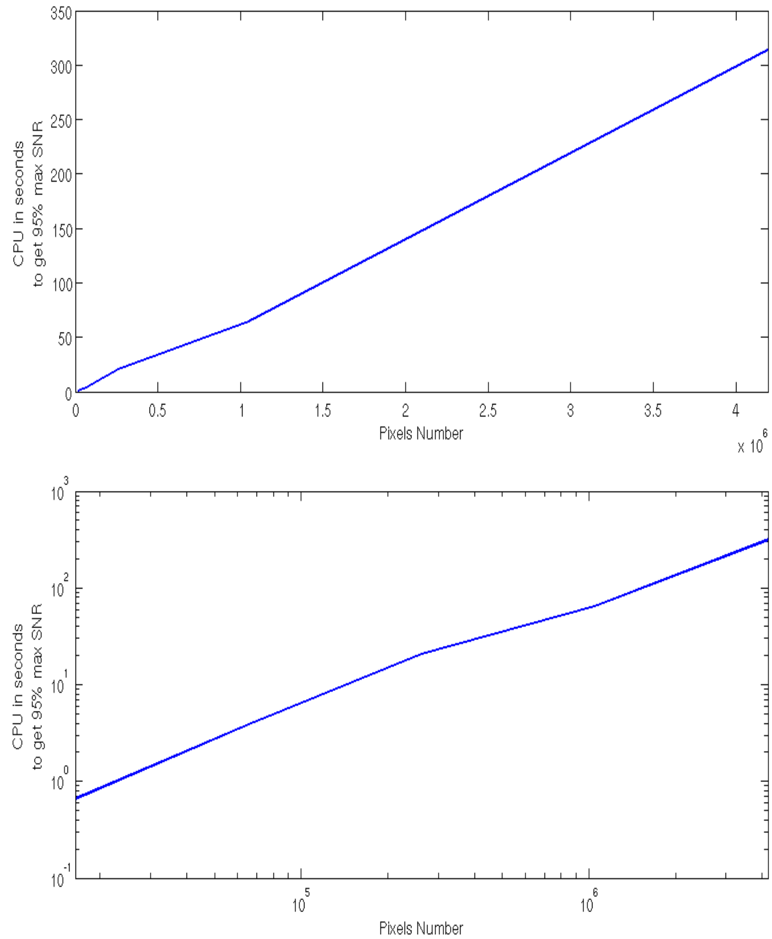

FIG. 4 – CPU time in seconds in order to get 95% of the maximum SNR w.r.t. number of pixels.

Computing times with respect to problem sizes This experiment is designed in order to show experimentally that the ADM method scales well with the problem dimensions. In the test, we proceed as follows. We convolve images of different sizes with a uniform kernel of size 5×5. We use the Matlab function *imnoise* with 30% of the pixels corrupted (i.e., we set randomly 30% of the pixels to be 0 or 255). We run each method and compute the time necessary to get a solution which has a SNR higher than 95% of the highest SNR obtainable with the method. According to these curves, the computing times seem to increase almost linearly with the problem size. Using the theoretical iteration complexity, we could expect an increase as $O(n \log n)$.

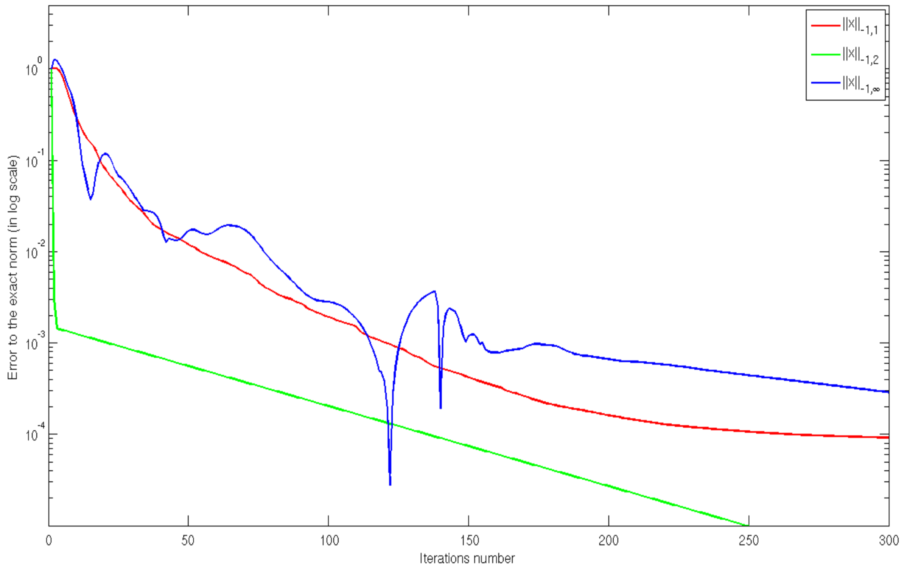

FIG. 5 – Computing the dual norm w.r.t. number of iterations.

5.1.3 Dual norm computation

In this experiment, we show in Figure 5 how fast the ADM method converges in order to compute the dual norms introduced in Section 4.3.1. Here we use the Barbara image and compute its dual norm. No more than 100 iterations are necessary in order to obtain a 10^{-2} precision (around 5 seconds on a laptop). The behavior for the $W^{-1,\infty}$-norm is not as satisfactory in the first iterations as for the other norms. However, the computational results are still acceptable by using the ADM method.

5.2 Image Restoration and Reconstruction Results

In this subsection, we employ the ADM method to illustrate the quality of restored and reconstructed images in the imaging problems stated in Section 4. Figure 6 shows an example of image upsampling using the total variation. Figure 7 shows an example of deconvolution with impulse noise. Figure 8 shows some examples of image texture+cartoon decomposition. The fingerprint example shows that the method permits to remove the epidermal ridges independently of the light intensity. These three examples demonstrate the ADM method can restore, reconstruct and decompose images quite effectively and efficiently.

6 Conclusion

In this paper, we showed that the ADM method based on an augmented Lagrangian formulation is an efficient tool for solving many *constrained* TV-

Fig. 6 – Image upsampling experiment. From left to right and top to bottom : original image (1024x1024), image convoluted, downsampled by 4 in each directions with an additive Gaussian noise, nearest neighbor interpolation (14.5 dB), bilinear interpolation (SNR 15.4dB), sinus cardinal interpolation (19 dB), TV-l^2 (SNR 19.7dB, 20 seconds).

FIG. 7 – Image deconvolution with 50% impulse noise. (top-left) : original image (460x360), (top-right) : image convoluted by a 10x10 uniform kernel, (bottom-left) : blurred + noisy image, (bottom-right) : TV-l^1 (SNR=19.5dB, 15 seconds).

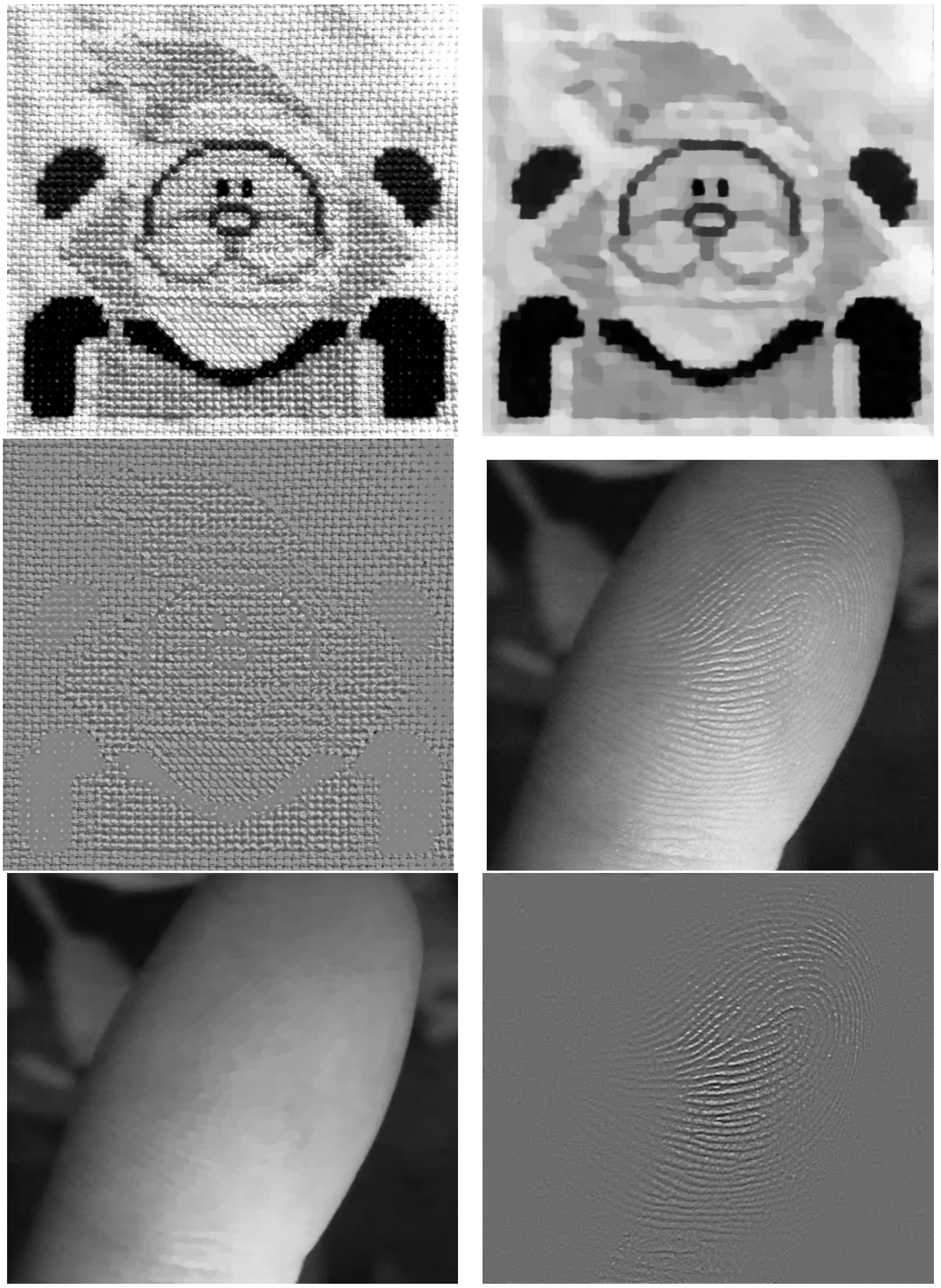

FIG. 8 – Examples of image decomposition with the $TV - W^{-1,1}$ model.

based image restoration and reconstruction problems (denoising, deconvolution, super-resolution, inpainting, texture+cartoon decompositions). The method scales well with the problems sizes and allows to get satisfactory solutions in times comparable to 1 second for 512×512 pixels images.

From a theoretical point of view, the method is known to be convergent. However, the convergence rate can be arbitrarily low. The behaviour in the first iterations still seems to lack theoretical foundations and some future work will consist in trying to get a better understanding of this feature. It also seems important to find an automatic way to choose the metric since this modification can affect the convergence behavior a lot.

7 Appendix

7.1 Solving the constrained least squares problem

We describe how to solve the following constrained least square problem :

$$\text{Find } \bar{x} \in \begin{array}{ll} \arg\min & \langle \lambda, \nabla x - y \rangle + \frac{\beta}{2} \|\nabla x - y\|^2 \\ \text{subject to} & x \in \mathbb{R}^n \\ & \|Hx - x^0\|^2 \leq \alpha \end{array} \tag{25}$$

Similar ideas are proposed in [30, p.580].

First, note that there exists a Lagrange multiplier $\bar{\delta} \in [0, +\infty[$ such that the solution of (25) satisfies :

$$\bar{x} \in \begin{array}{ll} \arg\min & \langle \nabla^T \lambda, x \rangle + \frac{\beta}{2} \|\nabla x - y\|^2 + \bar{\delta} \|Hx - x^0\|^2 \\ \text{subject to} & x \in \mathbb{R}^n \end{array}$$

The solution of this problem is characterized by the linear system :

$$(\beta \nabla^T \nabla + \bar{\delta} H^T H) \bar{x} = -\nabla^T \lambda + \beta \nabla^T y + \bar{\delta} H^T x^0. \tag{26}$$

We assume that the blur H is spatially invariant and periodic boundary conditions are used for the discrete differential operator. Thus the above matrices are diagonalized by the Fourier transform : $H = \mathcal{F}^{-1} D_H \mathcal{F}$, $\nabla^T \nabla = \mathcal{F}^{-1} D_\Delta \mathcal{F}$ and $H^T H = \mathcal{F}^{-1} D_{HH} \mathcal{F}$ where D_H, D_Δ and D_{HH} are diagonal matrices. Moreover, the diagonal elements of D_{HH} and D_Δ are positive. Therefore, we obtain

$$(\beta \nabla^T \nabla + \bar{\delta} H^T H) = \mathcal{F}^{-1} (\beta D_\Delta + \bar{\delta} D_{HH}) \mathcal{F}.$$

The main issue is to find the Lagrange parameter $\bar{\delta}$. We denote $b = -\frac{1}{\beta} \nabla^T \lambda + \nabla^T y$ and $c = H^T x^0$. The solution of (26) is characterized by :

$$(\beta D_\Delta + \bar{\delta} D_{HH}) \hat{\bar{x}} = \beta \hat{b} + \bar{\delta} \hat{c}.$$

where $\hat{x} = \mathcal{F} x$.

Let $I_\Delta = \{i, D_\Delta(i,i) = 0\}$, $I_H = \{i, D_{HH}(i,i) = 0\}$ and $I_{\Delta,H} = I_\Delta \Cap I_H$. In order to find $\bar{\delta}$, we need to distinguish two cases. Either the constraint in equation (25) is inactive and $\bar{\delta} = 0$, either it is active and $\bar{\delta} > 0$. Let us detail the two cases :

Case $\bar{\delta} = 0$ In this case, the problem simplifies to finding :

$$\begin{array}{rl} \arg\min & \left|\left|Hx - x^0\right|\right|^2 \\ \text{subject to} & x \in \mathbb{R}^n \\ & \mathcal{F}^{-1} D_\Delta \mathcal{F} x = b \end{array}$$

Denoting $\hat{x} = \mathcal{F}x$ and as the Fourier transform is an isometry, this can be rewritten :

$$\begin{array}{rl} \arg\min & ||D_H \hat{x} - \hat{x}^0||^2 \\ \text{subject to} & \hat{x} \in \mathbb{C}^n \\ & D_\Delta \hat{x} = \hat{b} \end{array}$$

One solution of this problem is given by :

$$\hat{x}(i) = \begin{cases} 0 & \text{if } \ i \in I_{H,\Delta} \\ \frac{\hat{x}^0}{D_H(i,i)} & \text{if } \ i \in I_\Delta \setminus I_H \\ \frac{\hat{b}}{D_\Delta(i,i)} & \text{if } \ i \in I_\Delta^c \end{cases} \tag{27}$$

In the case $\bar{\delta} = 0$, as $||Hx - x^0||^2 = ||D_H \hat{x} - \hat{x}^0||^2$, we get that :

$$||H\bar{x} - x^0||^2 = \sum_{i \in I_{H,\Delta}} |\hat{x}^0(i)|^2 + \sum_{i \in I_\Delta^c} \left| \frac{D_H(i,i)\hat{b}}{D_\Delta(i,i)} - \hat{x}^0 \right|^2.$$

Case $\bar{\delta} > 0$ In that case, the problem consists in finding $\bar{\delta}$. Let x_δ be any element satisfying :

$$(\beta D_\Delta + \delta D_{HH})\hat{x}_\delta = \beta\hat{b} + \delta\hat{c}.$$

We are looking for $\bar{\delta}$ such that :

$$\Psi(\bar{\delta}) = ||D_H \hat{x}_{\bar{\delta}} - \hat{x}^0||^2 = \alpha. \tag{28}$$

Ψ can be rewritten :

$$\Psi(\delta) = \sum_{i \in I_0} \left(|\hat{x}^0(i)|^2\right) + \sum_{i \in I_0^c} \left(\left| \frac{D_H(i,i)(\beta\hat{b}(i) + \delta\hat{c}(i))}{\beta D_\Delta(i,i) + \delta D_{HH}(i,i)} - \hat{x}^0(i) \right|^2 \right).$$

where $I_0 = \{i, D_\Delta(i,i) = 0 \text{ and } D_{HH}(i,i) = 0\}$. Ψ can be shown to be twice differentiable, convex and decreasing. Equation (28) can thus be solved using any root finding method. We propose a Newton technique as follows :

We settled everything to design an algorithm in order to solve Step 1 in Section 4.1. It writes as follows :

Algorithm 2: Solving the constrained least squares problem (4.1)

Input: $\lambda \in R^n$, $y \in \mathbb{R}^n \times \mathbb{R}^n$;
$\beta \in \mathbb{R}_+$, $\alpha \in \mathbb{R}_+^*$;
$D_H \in \mathbb{R}^n$ and $D_\Delta \in \mathbb{R}^n$ (the diagonal elements) ;
A stopping criterion ϵ (10^{-15} in our experiments).

Output: $x \in \mathbb{R}^n$ a solution of (25)

begin
- Set $b = -\frac{1}{\beta}\nabla^T \lambda - \nabla^T y$ and $c = H^T x^0$;
- Compute $\hat{b}$ and $\hat{c}$;
- Compute $\alpha_0 = \sum_{i \in I_{H,\Delta}} |\hat{x}^0(i)|^2 + \sum_{i \in I_\Delta^c} \left| \frac{D_H(i,i)\hat{b}}{D_\Delta(i,i)} - \hat{x}^0 \right|^2$;
- **if** $\alpha_0 \leq \alpha$ **then**
  - Set $\hat{x}$ according to equation (27) ;
  - Compute $x = \mathcal{F}^{-1}\hat{x}$.
- **else**
  - Use Newton method in order to compute $\bar{\delta}$. ;
  - Set $\delta \leq \bar{\delta}$ (e.g. $\delta = 0$ or an approximation) ;
  - **while** $|\Psi(\delta) - \alpha| \geq \epsilon$ **do**
    - $\delta \leftarrow \delta + \frac{\alpha - \Psi(\delta)}{\Psi'(\alpha)}$;
  - **end**
  - Set $\hat{x}(I_0^c) = \frac{(\beta\hat{b}(I_0^c) + \delta\hat{c}(I_0^c))}{(\beta D_\Delta + \delta D_{HH})(I_0^c)}$;
  - Set $\hat{x}(I_0) = 0$;
  - Set $x = \mathcal{F}^{-1}\hat{x}$;
- **end**

end

In practice, if the initial value of δ is taken as the value at the previous iteration, no more than 4 Newton iterations are necessary to get a very high precision.

Références

[1] G. Aubert and J.-F. Aujol, Modeling very oscillating signals. Application to image processing, *Applied Mathematics and Optimization*, volume 51, number 2, pages 163-182, March/April 2005.

[2] J.F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, Image decomposition into a bounded variation component and an oscillating component, *J. Math. Imaging Vision*, 22 (2005), pp.71-88.

[3] J.F. Aujol and A. Chambolle, Dual norms and image decomposition models, *International Journal of Computer Vision*, 63 (1) (2005), pp. 85-104.

[4] J.-F. Aujol, G. Gilboa, T. Chan and S. Osher, Structure-Texture Image Decomposition - Modeling, Algorithms, and Parameter Selection, *International Journal of Computer Vision*, 67 (1) (2006), pp. 111-136.

[5] , A. Auslender and M. Teboulle, Entropic proximal decomposition methods for convex programs and variational inequalities, Mathematical Programming, Ser. A, 91 (2001), pp. 33-47.

[6] M. R. Banham and A. K. Katsaggelos, Digital image restoration, *IEEE Signal Processing Magazine*, vol. 14, pp. 24-41, 1997.

[7] T. Chan, G. Golub and P. Mulet, A nonlinear primal-dual method for total variation-based image restoration, *SIAM J. Sci. Comput.*, 20 (1999), pp. 1964–1977.

[8] A. Chambolle and P.-L. Lions, Image recovery via total variation minimization and related problems *Numerische Mathematik*, 76 (1997), pp. 167–188.

[9] C. L. Chan, A. K. Katsaggelos, and A. V. Sahakian, Image sequence filtering in quantum-limited noise with applications to low-dose fluoroscopy, *IEEE Trans. Medical Imaging*, 12 (1993), pp. 610-621.

[10] T. Chan and P. Mulet, On the convergence of the lagged diffusivity fixed point method in total variation image restoration. *SIAM J. Numer. Anal.*, 36 (1999), pp. 354–367.

[11] P. L. Combettes and J. C. Pesquet, Image restoration subject to a total variation constraint, *IEEE Trans. Image Process.*, 13 (2004), pp. 1295-1304.

[12] P. L. Combettes and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Modeling and Simulation*, 4 (4) (2005), pp. 1168-1200.

[13] P. L. Combettes and J.-C. Pesquet, A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery, *IEEE Journal of Selected Topics in Signal Processing*, 1 (4) (2007), pp. 564-574.

[14] I. Daubechies and G. Teschke Variational image restoration by means of wavelets : simultaneous decomposition, deblurring and denoising. *Appl. Comput. Harmon. Anal.*, 19(1)(2005), pp. 1-16.

[15] J. Douglas and H. H. Rachford, On the numerical solution of heat conduction problems in two or three space variables, *Trans. Amer. Math. Soc.*, 82 (1956), pp. 421-439.

[16] J. Eckstein and D. P. Bertsekas, On the Douglas-Rachford splitting method and the proximal points algorithm for maximal monotone operators, *Math. Program.*, 55 (1992), pp. 293–318.

[17] J. Eckstein and M. Fukushima, Some reformulation and applications of the alternating direction method of multipliers, Large Scale Optimization : State of the Art, W.W.Hager et al. eds., Kluwer Academic Publishers, (1994), pp.115–134.

[18] E. Esser, Applications of Lagrangian-Based Alternating Direction Methods and Connections to Split Bregman, *UCLA Cam report*, April 2009

[19] E. Esser, X. Zhang and T. Chan, A General Framework for a Class of First Order Primal-Dual Algorithms for TV Minimization, *UCLA CAM Report*, August 2009.

[20] G. Facciolo, A. Almansa, J.-F. Aujol and V. Caselles, Irregular to regular sampling, denoising and deconvolution *SIAM Journal on Multiscale Modeling and Simulation*, 7 (4) (2009), pp. 1574-1608.

[21] M. Fortin and R. Glowinski, Eds., Augmented Lagrangian Methods : Applications to the solution of Boundary-Valued Problems, North-Holland, Amsterdam, 1983.

[22] H. Fu, M. Ng, M. Nikolova and J. Barlow, Efficient minimization methods of mixed $l_1 - l_1$ and $l_2 - l_1$ norms for image restoration, *SIAM J. Sci. Comput.*, 27 (2006), pp. 1881–1902.

[23] M. Fornasier and C. Schoenlieb, Subspace correction methods for total variation and l1-minimization *to appear in SIAM J. Numer. Anal.*, 2009.

[24] M. Fukushima, Application of the alternating direction method of multipliers to separable convex programming problems, *Comput. Optim. Appl.*, 2 (1992), pp. 93–111.

[25] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite-element approximations, *Comput. Optim. Appl.*, 2 (1976), pp.17–40.

[26] D. Gabay, Applications of the method of multipliers to variational inequalities, Augmented Lagrange Methods : Applications to the Solution of Boundary-valued Problems, M. Fortin and R. Glowinski, eds., North Holland, Amsterdam, The Netherlands, (1983), pp. 299–331.

[27] D. Mumford and B. Gidas Stochastic Models for Generic Images, *Quarterly of Applied Mathematics*, LIV(1) (2001), pp. 85 - 111.

[28] D. Goldfarb and W. Yin, Second-order cone programming methods for total variation based image restoration, *SIAM J. Sci. Comput.*, 27 (2005), pp. 622-645.

[29] T. Goldstein and S. Osher The Split Bregman Method for L1-Regularized Problems, *SIAM J. Imaging Sci.*, (2009)

[30] G.H. Golub and C.F. Van Loan, Matrix computations *The John hopkins university press*, third edition 1996.

[31] R. Glowinski, Numerical Methods for Nonlinear Variational Problems, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[32] Y. Gousseau and J.-M. Morel, Are natural images of bounded variation? *SIAM Journal of Mathematical Analysis*, 33(3) (2001), pp. 634-648.

[33] B. S. He, L-Z Liao, D.R. Han and H. Yang, A new inexact alternating directions method for monotone variational inequalities, *Math. Program.* Ser. A, 92 (2002), pp.103–118.

[34] B.S. He and H. Yang, Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities, *Oper. Res. Letters*, 23 (1998), pp. 151-161.

[35] M. Hintermüller, K. Ito and K. Kunisch, The primal-dual active set strategy as a semismooth Newton's method, *SIAM J. Optim.*, 13 (2003), pp. 865–888.

[36] M. Hintermüller and K. Kunisch, Total bounded variation regularization as a bilaterally constrained optimization problem, *SIAM J. Appl. Math.*, 64 (2004), pp. 1311–1333.

[37] M. Hintermüller and G. Stadler, An infesible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration, *SIAM J. Sci. Comput.*, 28 (2006), pp. 1–23.

[38] J. Yang, Alternating direction method applied to TV image reconstruction from partial convolution data, *working paper*, 2009.

[39] Y. Huang, M. Ng, Y. Wen, A fast total variation minimization method for image restoration, *SIAM Multiscale Modeling and Simulation*, 7 (2008), pp. 774–795.

[40] S. Kontogiorgis and R. R. Meyer, A variable-penalty alternating directions method for convex optimization, *Math. Program.*, 83 (1998), pp. 29–53.

[41] D. Krishnan, P. Lin and X. Tai, An efficient operator splitting method for noise removal in images, *Commun. Comput. Phys.*, 1 (2006), pp. 847–858.

[42] Y. Li and F. Santosa, A computational algorithm for minimizing total variation in image reconstruction, *IEEE Trans. Image Process.*, 5 (1996), pp. 987–995.

[43] Z. K. Liu and J. Y. Xiao, Restoration of blurred TV pictures caused by uniform linear motion, *Comput. Vision, Graphics, Image Proc.,* 44 (1988), pp. 30–34.

[44] F. Malgouyres and F. Guichard, Edge direction preserving image zooming : a mathematical and numerical analysis, *SIAM J. on Numerical Analysis*, 39 (1) (2001), pp. 1-37.

[45] Y. Meyer, Oscillating patterns in image processing and in some nonlinear evolution equations, *The Fifteenth Dean Jaqueline B. Lewis Memorial Lectures*, 2001.

[46] L. Moisan, How to discretize the Total Variation of an image ?, *Proceedings of ICIAM'07, Proc. Appl. Math. Mech.*, 7 (1) (2007), pp. 1041907-1041908.

[47] M. R. Nagel, Introduction to evaluation of motion-degraded images, *Proc. of NASA Electronics Research Center Seminar,* Cambridge, MA, Dec. 3-5, 1968.

[48] Y. Nesterov, Gradient methods for minimizing composite objective function, *CORE discussion paper*, 2007.

[49] M. Nikolova, A variational approach to remove outliers and impulse noise, *Journal of Mathematical Imaging and Vision*, 20 (1-2) (2004), pp. 99-120.

[50] J. Fadili and G. Peyré$\frac{1}{2}$, Total Variation Projection with First Order Schemes, *HAL Research Report*, April 2009.

[51] J. Nocedal and S. J. Wright, Numerical Optmization, Springer-Verlag, New York, Berlin, Heidelberg, 1999.

[52] Y. Kim and L.A. Vese, Image recovery using functions of bounded variation and Sobolev spaces of negative differentiability, *Inverse Problems and Imaging*, 3(2009), pp. 43–68

[53] L. H. Lieu and Luminita A. Vese, Image Restoration and Decomposition via Bounded Total Variation and Negative Hilbert-Sobolev Spaces, *Applied Math. Optimi.*, 58 (2008), pp. 167-193.

[54] M. Lysaker and X. Tai, Noise removal using smoothed normals and surface fitting, *IEEE Trans. Image Process.*, 13 (2004), pp. 1345–1357.

[55] S.J. Osher, A. Sole and L. A. Vese, Image decomposition and restoration using total variation minimization and the H-1 norm, *Multiscale Model. Simul.*, 1(3) (2003), pp. 349-370.

[56] S.J. Osher and L. A. Vese, Modeling textures with total variation minimization and oscillating patterns, *J. Sci. Comput.*, (2003), pp. 553-572.

[57] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D*, 60 (1992), pp. 259–268.

[58] C. H. Slump, Real-time image restoration in diagnostic X-ray imaging, the effects on quantum noise, *Proc, 11th IAPR Int. Conf. on Pattern Recognition,* vol. II, Conference B : Pattern Recognition Methodology and Systems, pp. 693-696, 1992.

[59] J.-L. Starck, M. Elad and D. L. Donoho, Image decomposition via the combination of sparse representation and a variational approach, *IEEE Trans. Image Process.*, 14 (2005), pp. 1570-1582.

[60] P. Tseng, Applications of a splitting algorithm to decomposition in convex programming and variational inequalities, *SIAM J. Con. Optim.*, 29 (1991), pp.119–138.

[61] E. Van Den Berg, M. Schmidt, M.P. Friedlander and K. Murphy, Group sparsity via linear-time projection, *Technical report, University of British Columbia*, June 2008.

[62] C. Vogel and M. Oman, Iterative method for total variation denoising, *SIAM J. Sci. Comput.*, 17 (1996), pp. 227–238.

[63] Y. Wang, J. Yang, W. Yin and Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, *SIAM J. Imaging Sciences*, 1 (2008), pp. 248-272.

[64] P. Weiss, L. Blanc-Féraud and G. Aubert, Efficient schemes for total variation minimization under constraints in image processing, *SIAM J. Scientific Computing*, 31 (2009), pp. 2047-2080.

[65] P. Weiss, L. Blanc-Féraud, A proximal method for inverse problems in image processing, *Proceedings of EUSIPCO*, Glasgow, 2009.

[66] C. Wu, J. Zhang and X. Tai, Augmented Largrangian method for total variation restoration with non-quadratic fidelity. *CAM-09-82*, UCLA CAM Report, 2009.

[67] C. H. Ye and X. M. Yuan, A descent method for structured monotone variational inequalities, *Optim. Methods Soft.*, 22(2) (2007) pp. 329-338.

[68] X. M. Yuan, Alternating direction methods for sparse covariance selection, manuscript, 2009.