RESTORATION AND ZOOM OF IRREGULARLY SAMPLED, BLURRED AND NOISY IMAGES BY ACCURATE TOTAL VARIATION MINIMIZATION WITH LOCAL CONSTRAINTS

By

Andrés Almansa

Vicent Caselles

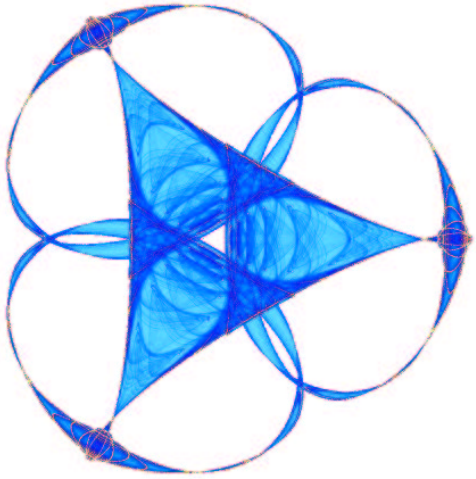
Gloria Haro

and

Bernard Rougé

IMA Preprint Series # 2088

(January 2006)



INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

UNIVERSITY OF MINNESOTA

400 Lind Hall

207 Church Street S.E.

Minneapolis, Minnesota 55455–0436

Phone: 612/624-6066 Fax: 612/626-7370

URL: http://www.ima.umn.edu

RESTORATION AND ZOOM OF IRREGULARLY SAMPLED, BLURRED AND NOISY IMAGES BY ACCURATE TOTAL VARIATION MINIMIZATION WITH LOCAL CONSTRAINTS.

ANDRÉS ALMANSA *, VICENT CASELLES †, GLORIA HARO ‡, AND BERNARD ROUGÉ §

Abstract. We propose an algorithm to solve a problem in image restoration which considers several different aspects of it, namely: irregular sampling, denoising, deconvolution, and zooming. Our algorithm is based on an extension of a previous image denoising algorithm proposed by A. Chambolle using total variation, combined with irregular to regular sampling algorithms proposed by H.G. Feichtinger, K. Gröchenig, M. Rauth and T. Strohmer. Finally we present some experimental results and we compare them with those obtained with the algorithm proposed by K. Gröchenig et al.

Key words. Image restoration, Total Variation, variational methods, satellite images.

AMS subject classifications. 68U10, 65K10, 65J20, 94A08

1. Introduction. A general image acquisition system may be modelled by the following image formation model

$$g(\lambda_k) = (h * u)(\lambda_k) + n_{\lambda_k}, \qquad \lambda_k \in \Lambda, \tag{1.1}$$

where $\Lambda = \{\lambda_k\}_{k=1}^{N^2} \subseteq \mathbb{R}^2$ is a finite set of regular or irregular samples, $u : \mathbb{R}^2 \to \mathbb{R}$ is the ideal undistorted image, $h : \mathbb{R}^2 \to \mathbb{R}$ is a blurring kernel whose Fourier spectrum (MTF) has most of its energy concentrated in $[-1/2, 1/2]^2$, g is the observed sampled image which is represented as a function $g : \Lambda \to \mathbb{R}$, and n_{λ_k} is, as usual, a white Gaussian noise with zero mean and standard deviation σ.

Let us denote by Ω_N the interval $[0, N[^2$. We shall concentrate in the particular case of perturbed sampling and we shall assume that Λ is a set of N^2 samples which take the particular form

$$\Lambda = \mathbb{Z}^2 \cap \Omega_N + \varepsilon(\mathbb{Z}^2 \cap \Omega_N) \tag{1.2}$$

where $\varepsilon : \mathbb{R}^2 \to \mathbb{R}^2$ is a "smooth and small" perturbation function in the sense that $\operatorname{supp} \hat{\varepsilon} \subseteq [-\frac{1}{T_\varepsilon}, \frac{1}{T_\varepsilon}]^2$ for some period $T_\varepsilon > 2$ corresponding to the maximum vibration frequency and the amplitude A, which can be measured as the maximum of the standard deviation of each component of the vector $\varepsilon(x)$, is small with respect to 1 pixel (we refer to Section 2 for a model (2.1) of this perturbation and also for a general overview of irregular sampling aspects).

As in most of works, in order to simplify this problem, we shall assume that the functions h and u are periodic of period N in each direction. That amounts to neglecting some boundary effects. Therefore, we shall assume that h, u are functions defined in Ω_N. To fix ideas, we assume that $h, u \in L^2(\Omega_N)$, so that $h*u$ is a continuous function in Ω_N [29] (which may be extended to a continuous periodic function in $\mathbb{R}^2$) and the samples $h * u(\lambda_k)$, $\lambda_k \in \Lambda$, have sense.

*InCo, Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay

†Departament de Tecnologia, Universitat Pompeu Fabra, Barcelona, Spain

‡Departament de Tecnologia, Universitat Pompeu Fabra, Barcelona, Spain, now at Institute for Mathematics and its Applications, University of Minnesota, USA.

§Centre National d'Etudes Spatiales (CNES), Toulouse, France

In a general setting where sampling is not necessarily perturbed, the well-posedness of the sampling set is measured by its density. Let us recall that a sampling set $\Lambda \subseteq \Omega$ is said to be γ-dense if

$$\bigcup_{\lambda_k \in \Lambda} D_\gamma(\lambda_k) = \Omega \tag{1.3}$$

i.e. if the image domain Ω can be covered by disks $D_\gamma(\lambda_k)$ of radius γ that are centered on the sampling points λ_k. The value $\upsilon := 2\gamma$ will be called the *"maximal gap"* of the sampling set Λ whenever γ is the minimal value such that Λ is γ-dense.

Our problem is to recover as much as possible of u, from our knowledge of the sampling geometry Λ, the blurring kernel h, the statistics of the noise n, and the irregular samples g.

The case of recovering an irregularly sampled image on a regular sampling grid was already considered by the first author in [2], but the MTF was assumed to be an ideal window (with Nyquist frequency cutoff), i.e., $\hat{h} = \chi_{[-1/2,1/2]^2}$. Different numerical algorithms were tested in the case where the sampling set is perturbed according to (1.2) and the period of the maximum vibration frequency satisfies $T_\varepsilon > 2$, and they worked relatively well only within a low-frequency spectral region $R \subseteq [-\alpha, \alpha]^2$, where $\alpha \approx \frac{1}{2} - 1/T_\varepsilon$. When attempting to recover $\hat{u}$ on the high frequency band $[-1/2, 1/2]^2 \backslash R$ serious theoretical and numerical problems appeared and, actually, restoration errors were most important in that band.

A second problem considered in [2] was the study of how to circumvent the aliasing artifacts when the sampling is assumed to be regular, but the MTF $\hat{h}$ is not restricted to $[-1/2, 1/2]^2$. As the author concluded (see also [3]), an optimal spectral region R similar to $[-1/2, 1/2]^2$ could be found (depending on h, and a statistical model of the image u and the noise n) where sampled coefficients contain a minimal amount of noise and aliasing. Then, following [38], Almansa proposed to extrapolate the restored information on R to a region R' containing $[-1/2, 1/2]^2$ and R by minimizing the total variation.

We consider in this paper the following variational model for restoring u, based on the full image formation model given by equation (1.1):

$$\begin{aligned} &\min_u \quad J(u), \\ &\text{subject to } \| w \cdot (\Delta_\Lambda \cdot (p * h * u) - g) \|^2_{\ell^2(\Lambda)} \leq N^2 \sigma^2, \end{aligned} \tag{1.4}$$

or its unconstrained formulation

$$\min_u \frac{1}{2\lambda} \| w \cdot (\Delta_\Lambda \cdot (p * h * u) - g) \|^2_{\ell^2(\Lambda)} + J(u), \tag{1.5}$$

where $\Delta_\Lambda : C(\mathbb{R}^2) \to \ell^2(\Lambda)$ is given by $\Delta_\Lambda(v) = \{v(\lambda_k)\}_{k=1}^{N^2}$, $J(u) = \int_{\Omega_N} |\nabla u|$ is the total variation of u, $w : \Lambda \to \mathbb{R}$ is a weighting function acting as a preconditionner (typically is the Voronoï area of the sampling set Λ, see Section 2.1), and p is a spectral projector (e.g. $\hat{p} = \chi_R$ or a prolate function, see Section 6) on a low-band region R which depends both on the MTF and the irregularity of the sampling set. Notice that the image model (1.1) has been introduced as a constraint in (1.4) expressing that the σ^2 is an upper estimate of the variance of the noise. The constraint has been incorporated in (1.4) via a Lagrange multiplier λ^{-1} (with $\lambda > 0$), although

one can also see λ as a penalization parameter. The total variation penalizes the oscillations that may appear when we extrapolate the high frequencies in the spectral region containing $[-1/2, 1/2]^2$. Let us mention that, as discussed in [2] in the context of regular sampling, the right choice of the spectral region R permits to reduce the aliasing effects, but we shall not consider this problem here. For us $\hat{p}$, the Fourier transform of p, will be different from zero on the region R and zero on $[-1/2, 1/2]^2 \backslash R$. This is a way to impose that the restored image fits the data g at low frequencies and the high ones are extrapolated via the total variation minimization.

The use of total variation for image restoration problems was originally proposed by Rudin, Osher and Fatemi in [36]. Functions with finite total variation, usually called bounded variation functions [4, 25, 46], are a reasonable functional model for many problems in image processing, in particular, for image restoration problems [36]. Typically, functions of bounded variation in the plane have discontinuities along rectifiable curves, being continuous in some sense (in the measure theoretic sense) away from discontinuities. These discontinuities could be identified with edges. The ability of total variation regularization to recover edges is one of the main features which advocates for the use of this model (its ability to describe textures is less clear, some textures can be recovered, but up to a certain scale of oscillation). We refer to [25, 46] for the definition of functions of bounded variation and its basic properties.

Many numerical algorithms have been proposed to minimize total variation subject to constraints (we refer to [6] for a survey). The particular case of total variation image denoising, for which the image formation model (1.1) simplifies to $g = u + n$, has been widely studied numerically [18, 19, 20, 21, 44]. Still, in a recent work [16], A. Chambolle proposed a very interesting algorithm to solve it for which the convergence could be proved (see [7] for a related work) . In this work, we shall extend Chambolle's result to the case of total variation image restoration under the constraint given by the general image formation model (1.1). Let us say explicitly that both the PSF h (or the MTF $\hat{h}$) and the sampling grid Λ (alternatively the grid perturbation function ε) are assumed to be known exactly, and that the only thing known about the noise n_{λ_k} is that it is a white Gaussian noise with zero mean and known variance σ^2. Several methods exist to estimate all these parameters for a given acquisition device and we shall not address this question here.

As we have just said, the main focus of the present work will be to tackle the full image restoration model (1.1). For that, we study the case of irregular to regular bandlimited sampling, and we observe that the ideas of the ACT algorithm proposed by Feichtinger, Gröchenig, Strohmer and Rauth [27, 32, 41] fit perfectly within our framework. Indeed, they can be easily combined with our extension of Chambolle's ideas to write a TV-regularized irregular to regular sampling algorithm. The proposed variational model (1.5) includes not only irregular sampling but also deconvolution, denoising, antialiasing and zoom. All these features, except the antialiasing, will be considered in this work. We shall prove the feasibility of a numerical algorithm, which combines Gröchenig and Chambolle's techniques, for solving (1.5). Let us mention that, as far as we know, in most of the irregular sampling literature, the only regularizers considered are linear, akin to Tikhonov regularization.

Let us finally explain the plan of the paper. In Section 2 we introduce the problem of irregular to regular sampling and we review the ACT algorithm of Gröchenig and Strohmer [32]. Section 3 explains our extension of Chambolle's algorithm [16] for total variation based image denoising to the more general case of image restoration which permits to include irregular to regular sampling, deconvolution, denoising, and

zoom. The variational approach to the general image restoration problem formulated in (1.5) is studied in Section 4. In Section 5 we discuss a method to incorporate the constraint of the image acquisition model (1.1) in a local way. The inclusion of zoom in the general algorithm is studied in more detail in Section 6. Finally, in Section 7 we display some experimental results. Section 8 summarizes the main conclusions of this work.

2. Irregular to regular sampling. Micro-vibrations of satellites together with irregularities in sensors position result in irregular sampling sets in satellite imaging. In most cases, the knowledge of certain vibration modes and the analysis of acquired images help to estimate with high accuracy the perturbations in the sampling grid, which can be modeled [2] by

$$\varepsilon(x) = \sum_{k=1}^{q} a_k(x) cos(2\pi\langle \xi_k, x\rangle + \phi_k), \qquad x \in \mathbb{R}^2, \tag{2.1}$$

for some $q \geq 1$, where $a_k(x)$ are smooth modulation functions and the vibration frequencies ξ_k are an order of magnitude (or even more) below the Nyquist frequency of the sampling rate. The bound on the modulation functions is inversely proportional to ξ_k and the number of vibration modes is small. This results in smooth and small perturbations with $|\varepsilon(x)|$ no larger than a few pixels, and perturbation slope $|\varepsilon'(x)|$ no larger than about one tenth of a pixel per pixel. As a consequence these perturbations are hardly noticeable and we should talk of *perturbed sampling* rather than irregular sampling in those cases. Even if the image distortion is not evident from a geometrical point of view it is very important to correct the perturbations in image registration applications where a sub-pixel accuracy is necessary.

In order to be less dependent on a particular physical instrument, in our experiments we used a simplified version of this model which still captures its main characteristics, namely the perturbation function $\varepsilon = (\varepsilon_1, \varepsilon_2)$ is simulated as a discrete colored noise, i.e. for $\xi \in (\mathbb{Z}/N)^2$ we define

$$\begin{array}{ll} \widehat{\varepsilon_i}(\xi) \sim N(0, \bar{\sigma}^2) & \text{if } |\xi| \leq 1/T_\epsilon \\ = 0 & \text{otherwise} \end{array} \tag{2.2}$$

where $\bar{\sigma}$ is chosen in such a way that the standard deviation of $\varepsilon_i(x)$ is A for $i \in \{1, 2\}$. This gives $\bar{\sigma} = \frac{AT_\epsilon}{2}$ (we have taken the Fourier transform as an isometry). Thus the behavior of the perturbation is governed by the two parameters "amplitude" A and maximal vibration frequency $1/T_\varepsilon$ (or "minimal vibration period" T_ε), even though these two parameters obviously do not completely determine ε. The precise values of A and T_ε will be specified in Section 7.

There exist many works on the literature dealing with the irregular sampling problem. However, many of them are iterative algorithms which are adapted to well-conditioned problems and small data sets but are inadequate to more realistic problems with large sampling sets and their generalization to the two dimensional case is not always evident. A comparison between several iterative methods can be found in [9, 28]. In [2] the author makes a survey of other techniques [27, 32, 41, 1] which are well-suited for the two dimensional case and also proposes a pseudo-inverse algorithm.

2.1. ACT algorithm. One of the best performing reconstruction methods available for irregular to regular sampling is the ACT algorithm, initially developed by

Feichtinger et al. [27], and further analyzed, refined and generalized by Gröchenig and Strohmer [32], and Rauth [41]. The method intelligently combines an accelerated version of the *frame iteration* derived from the proof of Kadec's theorem [45, 34], with *adaptive weights* in order to improve the condition number of the problem, a *conjugate gradient* iteration which accelerates convergence, and the formulation of the problem as a *Toeplitz system* in order to gain structure and, thus, numerical efficiency. Furthermore, the preparation steps before the conjugate gradient iteration can start can benefit from the USFFT (for unequally spaced fast Fourier transform) algorithm developed by Beylkin [13, 14].

More precisely, the algorithm is based on a representation of a discrete image $\{u(i,j)\}_{i,j=0}^{N-1}$ as a discretization of a trigonometric polynomial of order $N/2$ (for simplicity of notation we shall assume that N is an even number) in each variable ([27, 32])

$$u(x) = \sum_{n\in\{-\frac{N}{2},\dots,\frac{N}{2}-1\}^2} a_n e^{\frac{2\pi i}{N}\langle n,x\rangle}, \tag{2.3}$$

so that the interpolation conditions at the irregular sample points $\Lambda = \{\lambda_k\}_{k=1}^{N^2} \subseteq \mathbb{R}^2$ become

$$g_k = u(\lambda_k) = \sum_{n\in\{-\frac{N}{2},\dots,\frac{N}{2}-1\}^2} a_n e^{\frac{2\pi i}{N}\langle n,\lambda_k\rangle}, \quad k \in \{1,\dots,N^2\}, \tag{2.4}$$

or, equivalently, in matrix form

$$g = Sa, \quad \text{where} \quad S = ((s_{kn})),\ s_{kn} = e^{\frac{2\pi i}{N}\langle n,\lambda_k\rangle}, \tag{2.5}$$

i.e. S is the Vandermonde matrix associated to the trigonometric polynomial in (2.4). Note that S maps $a \in \ell^2(\{-\frac{N}{2},\dots,\frac{N}{2}-1\}^2)$ to $\{u(\lambda_k)\} \in \ell^2(\Lambda)$ defined in (2.4). Let us write ℓ^2_N instead of $\ell^2(\{-\frac{N}{2},\dots,\frac{N}{2}-1\}^2)$. Thus, the problem is reduced to solve the linear system of equations (2.5). But, if Λ contains some regions with extremely dense sampling or large gaps, then the system will not be well balanced. In order to improve the condition number, the k-th equation is multiplied by a weight

$$w_k = \text{area}\left(\{x : |x-\lambda_k| < |x-\lambda_j|,\, \forall j \neq k\}\right), \tag{2.6}$$

which is inversely proportional to the sampling density at λ_k. Thus, the use of weights w_k compensates the local variations in the sampling density. Moreover, the adaptive weights method provides explicit estimates for the rate of convergence and, therefore, gives useful stopping criteria. In addition, instead of solving the linear system (2.5) directly, it will be more convenient to solve the system of normal equations

$$S^*WSa = S^*Wg \tag{2.7}$$

because the $N^2 \times N^2$ matrix $T := S^*WS$ (where $W = \text{diag}(\{w_k\}_{k=1..N^2})$) is always a square matrix which has a Toeplitz structure [27, 32], so that the multiplication Ta can be efficiently computed in $N^2\log_2(N^2)$ time using Fourier methods. Defining $T = S^*WS = (t_{n-n'})_{nn'}$ and $b = S^*Wg$ in equation (2.7), the non-harmonic series

$$t_n = \sum_{k=1}^{N^2} e^{-\frac{2\pi i}{N}\langle n,\lambda_k\rangle} w_k \tag{2.8}$$

$$b_n = \sum_{k=1}^{N^2} e^{-\frac{2\pi i}{N}\langle n,\lambda_k\rangle} w_k g_k, \tag{2.9}$$

$n \in \{-\frac{N}{2}, \dots, \frac{N}{2} - 1\}^2$, can be approximated using the USFFT [13] in $CN^2 \log_2(N^2)$ time each, where C is a constant, which is inversely proportional to the required precision. The theory developed in [27, 32] also permits to consider the case of more irregular samples than frequencies but we shall not need this here.

Let us observe that, if T is not invertible, the solution of (2.7) is replaced by the minimum norm solution of the problem

$$\min_a \|Ta - b\|^2_{\ell^2_N}, \tag{2.10}$$

and we shall understand that this choice is always taken when T is not invertible.

The overall procedure can be summarized in the following algorithm.

Algorithm 1: ACT algorithm

REQUIRES: N^2 irregular samples in vector g.

ENSURES: N^2 regular samples in vector u.

1. Compute $T = S^*WS$ and $b = S^*Wg$ using the USFFT.
2. Solve $Ta = b$ using conjugate gradients.
3. Compute the regular samples $u(i, j)$ for $(i, j) \in \{0, \dots, N - 1\}^2$ by applying the inverse FFT to a.

The convergence rate of the CG algorithm is determined by the condition number $\kappa = \text{cond}(T)$, or the ratio of the largest to the smallest eigenvalue of T. More precisely at each iteration the approximation error is decreased by a factor $\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$ [30]. Gröchenig [31] and Gröchenig and Strohmer [32] provided a useful characterization of the condition number of T in the 2-dimensional case,

PROPOSITION 2.1 (ACT convergence rate). *If the sampling set is γ-dense with*

$$\gamma < \frac{\log 2}{2\pi},$$

then the matrix T is invertible, its condition number is

$$\kappa \leq \frac{4}{(2 - e^{2\pi\gamma})^2}, \tag{2.11}$$

and Algorithm 1 converges to the exact solution.

Remarks:

1. The invertibility of T does not depend on the choice of the weights w_k. Indeed, S^*WS and S^*S have the same kernel and the matrix T will be invertible if and only if its kernel is $\{0\}$. Proposition 2.1 gives us a sufficient condition for the invertibility of T which depends only on the maximal gap $\upsilon = 2\gamma$.
2. The estimate (2.11) on κ is only valid if we are using adaptive weights.
3. As a consequence of using this kind of weights, the rate of convergence is independent of possible clustering shapes in the set of samples and depends only on the maximal gap between samples.
4. Increasing the number of samples (oversampling) improves the condition number (2.11).
5. The sampling density of a regular grid $\Gamma := \{n_1 e_1 + n_2 e_2 : n_1, n_2 \in \mathbb{Z}\}$ (generated by a basis e_1, e_2 of $\mathbb{R}^2$) is the number of samples per unit area and is given by $\rho(\Gamma) := \frac{1}{\det(\Gamma)}$. One would like to satisfy Shannon's sampling theorem with as least samples as possible, i.e., with the lowest possible value of $\rho(\Gamma)$, a number which is called the critical sampling rate (see [3] for a

more precise definition). Now, concerning Proposition 2.1, we note that the sampling has to be much more dense than the critical sampling rate (which is one in our case), for the algorithm to ensure convergence to the exact result, a condition which does not hold in the case of satellite imaging.[1] Nevertheless, even when T is not invertible (and has an infinite condition number), the CG iteration chooses among the minimizers of $\|Ta - b\|_{\ell^2_N}$ the one of minimal norm, i.e. $a = T^+ b$, where T^+ is the pseudo-inverse of T.

6. In our particular case, we are working with perturbed sampling sets. The perturbation in satellite imaging is of the form (2.1), producing smooth and small perturbations over the uniform grid. Thus, the distribution of sampling points is uniform and the use of weights w_k is not relevant in that case (in practice we fix $w_k = 1$ for all k).

On the other hand, if we know the a-priori spectral decay rate of the image

$$|\hat{u}(\omega)| \leq C\phi(\omega),$$

and the known transfer function, we can regularize the solution by imposing this decay rate (typically $\phi(\omega) = (1 + |\omega|)^{-r}$ for $r \in [1, 2]$, or a combination of these natural decay rates). As proposed in [32], if T is not invertible or is ill-conditioned, the a priori information about the frequency decay can be incorporated by using one of the following two algorithms:
(i) we compute the minimum norm least squares solution of the modified system $DTa = Db$, i.e., the minimum norm solution of the problem

$$\min_a \|DTa - Db\|_{\ell^2_N}, \tag{2.12}$$

where $D = \operatorname{diag}\left(\{\phi(\frac{2\pi}{N}n)\}_{n\in\{-\frac{N}{2},\dots,\frac{N}{2}-1\}^2}\right)$ is a diagonal matrix containing the corresponding values of $\phi(\omega)$. In this way we shall obtain the solution $a = (DT)^+ Db$, where $(DT)^+$ is the pseudo-inverse of DT.
(ii) we compute the minimum norm solution of the problem

$$\min_c \|TDc - b\|_{\ell^2_N}, \tag{2.13}$$

where $a = Dc$; in this case the solution is $a = D(TD)^+ b$.

In either case the solution coincides with $T^{-1}b$, if T is invertible. But in more realistic cases where T is not invertible or ill-conditioned, the minimum norm solution of (2.13) is the one which best follows the prescribed spectral decay rate, i.e. $D^{-1}a = c$ is of minimal norm, thus penalizing large Fourier coefficients in high frequencies. Similarly, but more subtly, the minimization of residuals corresponding to high frequencies in (2.12) (related to high frequency coefficients of a) will be neglected –after multiplication by D which makes them smaller– in favor of the minimization of $\|a\|_{\ell^2_N}$ leading also to solutions with smaller high frequency coefficients. This is a heuristic approach in consonance with the experiments [32] (see also the experiments in Section 7).

Note (still following [32]) that Algorithm 1 can be adapted to solve (2.12) and (2.13), although we have only implemented the first approach: it suffices to replace

[1] The bound on the density γ can be significantly relaxed in the one-dimensional case and in the case of a "separable" two-dimensional perturbation [32], which is a quite good approximation to perturbed sampling by push-broom acquisition devices such as a vibrating satellite.

the solution of $Ta = b$ by conjugate gradients in step 2 by the solution of $DTa = Db$ by preconditioned conjugate gradients. In contrast to the original purpose of the preconditioned conjugate gradients, here the matrix D is not acting as a preconditioner (actually it is possible that we decrease the rate of convergence using D) but to obtain an improved and regularized solution via incorporating a priori information about the frequency decay [32].

This discussion explains why the ACT algorithm of Gröchenig and Strohmer [32] provides good approximations to the exact solution, even when the convergence conditions in Proposition 2.1 are not satisfied (as in the case of satellite imaging), while at the same time the convergence rate is not that good.

3. Generalization of Chambolle's algorithm. We shall denote by Ω the image domain, which is assumed to be a rectangle in $\mathbb{R}^2$. For simplicity, we shall write

$$J(u) := \begin{cases} \displaystyle\int_\Omega |Du| & \text{if} \quad u \in BV(\Omega) \\ \\ +\infty & \text{if} \quad u \in L^2(\Omega) \setminus BV(\Omega), \end{cases} \tag{3.1}$$

where $BV(\Omega)$ denotes the space of functions of bounded variation in Ω (see [25, 46]).

In [16], Chambolle proposed a numerical algorithm to solve the total variation approach to image denoising, which can be written as a constrained minimization problem

$$\begin{aligned} &\min \quad J(u) \\ &\text{with} \quad \int_\Omega |u-g|^2\,dx \le \sigma^2|\Omega|. \end{aligned} \tag{3.2}$$

The constraint corresponds to the assumption that σ is an upper bound of the standard deviation of the noise. The constraint is a way to incorporate the simplified image acquisition model given by $g = u + n$, reflecting the fact that the ideal image u has been distorted by a Gaussian white noise n. In practice, problem (3.2) is solved via the following unconstrained minimization problem

$$\min_u \quad J(u) + \frac{1}{2\beta}\int_\Omega (u-g)^2\,dx, \tag{3.3}$$

where $\beta^{-1} > 0$ is some Lagrange multiplier. As it has been observed in many places (see [17, 43, 6]), problem (3.3) has a unique solution in $BV(\Omega)$. Moreover, for any $\beta > 0$, the solution u of (3.3) satisfies the Euler-Lagrange equation

$$u + \beta\partial J(u) \ni g, \tag{3.4}$$

where $\partial J(u)$ denotes the subdifferential of J at u, i.e.,

$$\partial J(u) = \{w \in L^2(\Omega) : J(v) - J(u) \ge \langle w, v-u\rangle, \forall v \in L^2(\Omega)\}.$$

We recall that $u \to \partial J(u)$ is a multivoque operator, this explains the $\ni$ sign in (3.4) instead of the more classical equality sign.

In [16], the author proposed an algorithm to solve the discrete version of (3.4) and he proved its convergence. This algorithm was also proposed in [11], though in

a different context. Our purpose is to extend this algorithm to be able to apply it in the more general case where the image formation model is given by (1.1) (see Section 4).

Even if we shall need them only in the discrete case, let us first develop the basic arguments in a continuous framework. For that, let Q be a bounded linear self-adjoint operator in $L^2(\Omega)$, and assume that Q is positive definite, hence invertible. We want to solve the equation

$$Q(u) + \beta \partial J(u) \ni b, \tag{3.5}$$

where $b \in L^2(\Omega)$, and $\beta > 0$, which is nothing else than the Euler-Lagrange equation for the minimization problem

$$\min_u \quad J(u) + \frac{1}{2\beta} \int_\Omega (Q^{1/2}(u-g))^2 \, dx, \tag{3.6}$$

for $g = Q^{-1}b$. We observe that, under our assumption of invertibility of Q, existence and uniqueness of solutions of (3.6) has been proved in several papers, let us mention [17, 38, 43, 6] to quote a few of them. Equation (3.5) could be written in a more classical way as

$$\begin{aligned} &Q(u) - \beta \mathrm{div}\Big(\frac{Du}{|Du|}\Big) = b, \\ &[\xi, \nu^\Omega] = 0, \end{aligned} \tag{3.7}$$

where ξ represents the vector field $\frac{Du}{|Du|}$, since $\partial J(u)$ is nothing else than the operator $-\mathrm{div}\Big(\frac{Du}{|Du|}\Big)$ coupled with Neumann boundary conditions, and $\nu^\Omega(x)$ denotes the outer unit normal at the point $x \in \partial\Omega$ (see, for instance, [23], vol. 5). For a precise characterization of $\partial J(u)$ we refer to [5, 43] (see also [6]) where it was introduced to give a more classical PDE sense to $\partial J(u)$.

Our next proposition extends the result in [16] to the case of a general operator Q as above.

PROPOSITION 3.1. *Let Q be a bounded linear self-adjoint operator in $L^2(\Omega)$. Assume Q is positive definite, hence invertible. We define the sets*

$$\mathcal{V} := \{\xi \in L^\infty(\Omega, \mathbb{R}^2) : \|\xi\|_\infty \le 1, \, [\xi \cdot \nu^\Omega]|_{\partial\Omega} = 0\},$$

and

$$K := \{\mathrm{div}\, \xi : \xi \in L^\infty(\Omega, \mathbb{R}^2), \|\xi\|_\infty \le 1, \, [\xi \cdot \nu^\Omega]|_{\partial\Omega} = 0\}.$$

Let u be the solution of (3.5). Then $u = Q^{-1}(b + \beta \mathrm{div}\, \xi)$, where ξ is the solution of the variational problem

$$\min_{\{\xi \in \mathcal{V}\}} \|\mathrm{div}\, \xi + \beta^{-1} b\|_Q \tag{3.8}$$

and $\|\cdot\|_Q$ is the norm in $L^2(\Omega)$ defined as, $\|\overline{w}\|_Q = \|Q^{-1/2}\overline{w}\|$. In other words, $\mathrm{div}\, \xi$ is the projection of $-\beta^{-1}b$ onto the convex set K in the Hilbert space norm $\|\cdot\|_Q$.

Proof. Following the arguments in [16], let us write (3.5) as

$$\frac{1}{\beta}(b - Q(u)) \in \partial J(u). \tag{3.9}$$

Now, this implies that

$$u \in \partial J^*(\frac{1}{\beta}(b - Q(u))), \tag{3.10}$$

where J^* denotes the Legendre-Fenchel transform of J given by

$$J^*(v) = \sup_u \{\langle v, u\rangle - J(u)\},$$

since the relation $v \in \partial J(w)$ is equivalent to $w \in \partial J^*(v)$. Let

$$\tilde{b} = Q^{-1}b, \qquad w = \tilde{b} - u,$$

so that

$$Q(u) + Q(w) = b. \tag{3.11}$$

Then we may write (3.10) as

$$u \in \partial J^*(\frac{1}{\beta}Q(w)), \tag{3.12}$$

which, after multiplication by $\beta^{-1}Q$ becomes

$$\beta^{-1}Q(u) \in \beta^{-1}Q \circ \partial J^*(\frac{1}{\beta}Q(w)). \tag{3.13}$$

Since Q is a bounded self-adjoint operator in $L^2(\Omega)$, we have $\partial(J^* \circ (\beta^{-1}Q)) = (\beta^{-1}Q) \circ \partial J^* \circ (\beta^{-1}Q)$ [24], hence we may write (3.13) as

$$\beta^{-1}b - \beta^{-1}Qw = \beta^{-1}Q(u) \in \partial(J^* \circ (\beta^{-1}Q)))(w). \tag{3.14}$$

Now, we observe that (3.14) corresponds to the Euler-Lagrange equations of the variational problem

$$\min_{\overline{w}} \quad \frac{1}{2}\langle Q\overline{w}, \overline{w}\rangle - \langle b, \overline{w}\rangle + \beta(J^* \circ (\beta^{-1}Q))(\overline{w}). \tag{3.15}$$

Notice that, since J is homogeneous of degree one (i.e. $J(\lambda u) = \lambda J(u)$ for every u and any $\lambda > 0$), then J^* is the indicator function of a convex set K ([24]):

$$J^*(v) = \delta_K(v) =:= \begin{cases} 0 & \text{if} \quad v \in K \\ \\ +\infty & \text{if} \quad v \notin K. \end{cases} \tag{3.16}$$

Observe that $K := \{\operatorname{div}\xi : \xi \in \mathcal{V}\}$. Then we may write (3.15) as

$$\min_{\{\overline{w}:\beta^{-1}Q(\overline{w})\in K\}} \quad \frac{1}{2}\langle Q\overline{w}, \overline{w}\rangle - \langle b, \overline{w}\rangle \tag{3.17}$$

Writing $\beta^{-1}Q(\overline{w}) \in K$ as $\beta^{-1}Q(\overline{w}) = -\operatorname{div}\xi$ for some $\xi \in \mathcal{V}$, we may write (3.17) as

$$\min_{\{\xi\in\mathcal{V}\}} \quad \frac{\beta}{2}\langle Q^{-1}\operatorname{div}\xi, \operatorname{div}\xi\rangle + \langle \operatorname{div}\xi, Q^{-1}b\rangle, \tag{3.18}$$

after dividing by β the energy functional. Now, observing that

$$\frac{\beta}{2}\|Q^{-1/2}\mathrm{div}\,\xi+\beta^{-1}Q^{-1/2}b\|^2=\frac{\beta}{2}\langle Q^{-1}\mathrm{div}\,\xi,\mathrm{div}\,\xi\rangle+\langle\mathrm{div}\,\xi,Q^{-1}b\rangle+\frac{\beta}{2}\|\beta^{-1}Q^{-1/2}b\|^2$$

we may write the minimization problem (3.18) as

$$\min_{\{\xi\in\mathcal{V}\}}\|\mathrm{div}\,\xi+\beta^{-1}b\|_Q. \tag{3.19}$$

□

3.1. The discrete case formulation. We denote by X the Euclidean space $\mathbb{R}^{M\times M}$. The Euclidean scalar product and the norm in X will be denoted by $\langle\cdot,\cdot\rangle_X$ and $\|\cdot\|_X$, respectively. Then the image $u\in X$ is the vector $u=(u(i,j))_{i,j=0}^{M-1}$, and the vector field ξ is the map $\xi:\{0,\dots,M-1\}\times\{0,\dots,M-1\}\to\mathbb{R}^2$. If $u\in X$, the discrete gradient is a vector in $Y=X\times X$ given by

$$\nabla^{+,+}u:=(\nabla_x^+u,\nabla_y^+u),$$

where

$$\nabla_x^+u(i,j)=\begin{cases}u(i+1,j)-u(i,j) & \text{if} \quad i<M-1\\ 0 & \text{if} \quad i=M-1,\end{cases} \tag{3.20}$$

$$\nabla_y^+u(i,j)=\begin{cases}u(i,j+1)-u(i,j) & \text{if} \quad j<M-1\\ 0 & \text{if} \quad j=M-1\end{cases} \tag{3.21}$$

for $i,j=0,\dots,M-1$. Other choices of the gradient are possible, this one will be convenient for the developments below.

The Euclidean scalar product in Y is defined in the standard way by

$$\langle p,q\rangle_Y=\sum_{0\le i,j\le M-1}(p_{i,j}^1q_{i,j}^1+p_{i,j}^2q_{i,j}^2)$$

for every $p=(p^1,p^2)$, $q=(q^1,q^2)\in Y$. The norm of $p=(p^1,p^2)\in Y$ is, as usual, $\|p\|_Y=\langle p,p\rangle_Y^{1/2}$. We denote the euclidean norm of a vector $v\in\mathbb{R}^2$ by $|v|$. Then the discrete total variation is

$$J_d(u)=\|\nabla^{+,+}u\|_Y=\sum_{0\le i,j\le M-1}|\nabla^{+,+}u(i,j)|. \tag{3.22}$$

We have

$$J_d(u)=\sup_{p\in Y,\,|p_{i,j}|\le1\,\forall(i,j)}\langle p,\nabla^{+,+}u\rangle_Y. \tag{3.23}$$

When necessary we shall write X_M,Y_M, J_d^M instead of X,Y, J_d to stress the dependence on M.

By analogy with the continuous setting, we introduce a discrete divergence $\mathrm{div}^{-,-}$ as the dual operator of $\nabla^{+,+}$, i.e., for every $p\in Y$ and $u\in X$ we have

$$\langle-\mathrm{div}^{-,-}\,p,u\rangle_X=\langle p,\nabla^{+,+}u\rangle_Y.$$

One can easily check that $\mathrm{div}^{-,-}$ is given by

$$\begin{aligned}\mathrm{div}^{-,-}p(i,j) &= \begin{cases} p^1(i,j)-p^1(i-1,j) & \text{if} \quad 0<i<M-1 \\ p^1(i,j) & \text{if} \quad i=0 \\ -p^1(i-1,j) & \text{if} \quad i=M-1 \end{cases} \\ &+ \begin{cases} p^2(i,j)-p^2(i,j-1) & \text{if} \quad 0<j<M-1 \\ p^2(i,j) & \text{if} \quad j=0 \\ -p^2(i,j-1) & \text{if} \quad j=M-1 \end{cases}\end{aligned} \tag{3.24}$$

for every $p=(p^1,p^2)\in Y$. In this setting, we have

$$J_d^*(v)=\delta_{K_d}(v) =:= \begin{cases} 0 & \text{if} \quad v\in K_d \\ \\ +\infty & \text{if} \quad v\notin K_d, \end{cases} \tag{3.25}$$

where

$$K_d := \{\mathrm{div}^{-,-}\, p : |p_{i,j}|\leq 1,\ \forall\, i,j\in\{0,\dots,M-1\}\}.$$

Let us also denote by Q a self-adjoint positive definite matrix acting on X, and let $b\in X$. Our purpose is to solve the equation

$$Q(u)+\beta\partial J_d(u)\ni b \tag{3.26}$$

for some $u\in X$ which corresponds to the Euler-Lagrange equation for the minimization problem in X:

$$\min_{u\in X}\quad J_d(u)+\frac{1}{2\beta}\|Q^{1/2}(u-g)\|_X^2 \tag{3.27}$$

for $g=Q^{-1}b$. Again, under our assumptions on Q, there exists a unique solution of (3.27) ([17, 38, 43, 6]). Proceeding as in the continuous framework we have to solve the problem

$$\min_{\{\xi\in\mathcal{V}_d\}}\quad \frac{\beta}{2}\langle Q^{-1}\mathrm{div}^{-,-}\,\xi,\mathrm{div}^{-,-}\,\xi\rangle_X+\langle\mathrm{div}^{-,-}\,\xi,Q^{-1}b\rangle_X, \tag{3.28}$$

which may be written as

$$\min_{\{\xi\in\mathcal{V}_d\}}\|\mathrm{div}^{-,-}\,\xi+\beta^{-1}b\|_Q, \tag{3.29}$$

where

$$\mathcal{V}_d=\{\xi\in Y : |\xi(i,j)|^2-1\leq 0,\ \forall\, i,j\in\{0,\dots,M-1\}\}$$

and

$$\|\overline{w}\|_Q=\|Q^{-1/2}\overline{w}\|_X,\quad \overline{w}\in X.$$

The solution of (3.29) is the unique projection of $-\beta^{-1}b$ onto the convex set $K_d\subseteq X$, when X is endowed with the norm $\|\cdot\|_Q$. We shall denote this projection by $\Pi_{K_d}^Q(-\beta^{-1}b)$.

As in [16], the Karush-Kuhn-Tucker Theorem ([22], Theorem 9.2-4) yields the existence of Lagrange multipliers $\alpha^*(i,j) \geq 0$ for the constraints $\xi \in \mathcal{V}_d$, such that we have for each $(i,j) \in \{0,\dots,M-1\}^2$

$$\nabla^{+,+}[Q^{-1}\mathrm{div}^{-,-}\,\xi + Q^{-1}(\beta^{-1}b)](i,j) - \alpha^*(i,j)\xi(i,j) = 0, \tag{3.30}$$

with either $\alpha^*(i,j) > 0$ and $|\xi(i,j)| = 1$, or $\alpha^*(i,j) = 0$ and $|\xi(i,j)| \leq 1$. In the later case, we have $\nabla^{+,+}[Q^{-1}\mathrm{div}^{-,-}\,\xi + Q^{-1}(\beta^{-1}b)](i,j) = 0$. In any case, we have

$$\alpha^*(i,j) = |\nabla^{+,+}[Q^{-1}\mathrm{div}^{-,-}\,\xi + Q^{-1}(\beta^{-1}b)](i,j)|. \tag{3.31}$$

Let $\nu > 0$, $\xi^0 = 0$, $p \geq 0$. We solve (3.30) using the following gradient descent (or fixed point) algorithm

$$\begin{aligned}\xi^{p+1}(i,j) = \xi^p(i,j) + \nu\nabla^{+,+}[Q^{-1}\mathrm{div}^{-,-}\,\xi^p + Q^{-1}(\beta^{-1}b)](i,j)\\ -\nu|\nabla^{+,+}[Q^{-1}\mathrm{div}^{-,-}\,\xi^p + Q^{-1}(\beta^{-1}b)](i,j)|\xi^{p+1}(i,j),\end{aligned} \tag{3.32}$$

hence

$$\xi^{p+1}(i,j) = \frac{\xi^p(i,j) + \nu\nabla^{+,+}[Q^{-1}\mathrm{div}^{-,-}\,\xi^p + Q^{-1}(\beta^{-1}b)](i,j)}{1 + \nu|\nabla^{+,+}[Q^{-1}\mathrm{div}^{-,-}\,\xi^p + Q^{-1}(\beta^{-1}b)](i,j)|}. \tag{3.33}$$

Observe that $|\xi^p(i,j)| \leq 1$ for all $i,j \in \{0,\dots,M-1\}$ and every $p \geq 0$.

THEOREM 3.2. *In the discrete framework, assuming that $\nu < \frac{1}{8\|Q^{-1/2}\|^2}$, ξ^p has a subsequence which converges to a solution ξ of (3.30) where α^* is given by (3.31). The function $\xi \in Y$ is a solution of (3.29). Moreover $div^{-,-}\,\xi^p$ converges to the unique projection $\Pi^Q_{K_d}(-\beta^{-1}b)$.*

In particular, we deduce that $w^p = -\beta Q^{-1}(\mathrm{div}^{-,-}\,\xi^p)$ converges to a solution $w = -\beta Q^{-1}(\mathrm{div}^{-,-}\,\xi)$ of the discrete analogous of (3.17).

Finally, the solution u of (3.26) may be recovered from (3.11), i.e.,

$$u = Q^{-1}(b + \beta\mathrm{div}^{-,-}\,\xi).$$

Proof. For simplicity, let us denote ∇ and div instead of $\nabla^{+,+}$ and $\mathrm{div}^{-,-}$. Let us fix $p \geq 0$, and let

$$\eta = \frac{\xi^{p+1} - \xi^p}{\nu}, \quad \rho \in Y,$$

where $\rho(i,j) = |\nabla Q^{-1}(\mathrm{div}\,\xi^p + (\beta^{-1}b))(i,j)|\xi^{p+1}(i,j)$, so that

$$\eta = \nabla Q^{-1}(\mathrm{div}\,\xi^p + (\beta^{-1}b)) - \rho.$$

Then

$$\begin{aligned}&\|Q^{-1/2}(\mathrm{div}\,\xi^{p+1} + \beta^{-1}b)\|_X^2 = \|Q^{-1/2}(\mathrm{div}\,\xi^p + \nu\mathrm{div}\,\eta + \beta^{-1}b)\|_X^2\\ &= \|Q^{-1/2}(\mathrm{div}\,\xi^p + \beta^{-1}b)\|_X^2 + \nu^2\|Q^{-1/2}\mathrm{div}\,\eta\|_X^2 + 2\nu\langle Q^{-1/2}(\mathrm{div}\,\xi^p + \beta^{-1}b), Q^{-1/2}\mathrm{div}\,\eta\rangle_X\\ &\leq \|Q^{-1/2}(\mathrm{div}\,\xi^p + \beta^{-1}b)\|_X^2 + \kappa^2\nu^2\|\eta\|_Y^2 - 2\nu\langle\nabla Q^{-1}(\mathrm{div}\,\xi^p + \beta^{-1}b), \eta\rangle_Y,\end{aligned}$$

where κ is the norm of the operator $Q^{-1/2}\mathrm{div} : Y \to X$, that will be estimated later on. Since

$$\begin{aligned}
&2\langle \nabla Q^{-1}(\mathrm{div}\,\xi^p + \beta^{-1}b), \eta\rangle_Y \\
&= \langle \nabla Q^{-1}(\mathrm{div}\,\xi^p + \beta^{-1}b), \eta\rangle_Y + \langle \nabla Q^{-1}(\mathrm{div}\,\xi^p + \beta^{-1}b), \eta\rangle_Y \\
&= \langle \nabla Q^{-1}(\mathrm{div}\,\xi^p + \beta^{-1}b), \nabla Q^{-1}(\mathrm{div}\,\xi^p + \beta^{-1}b) - \rho\rangle_Y + \langle \eta + \rho, \eta\rangle_Y \\
&= \|\nabla Q^{-1}(\mathrm{div}\,\xi^p + \beta^{-1}b)\|_Y^2 + \|\eta\|_Y^2 - \langle \nabla Q^{-1}(\mathrm{div}\,\xi^p + \beta^{-1}b) - \eta, \rho\rangle_Y \\
&= \|\nabla Q^{-1}(\mathrm{div}\,\xi^p + \beta^{-1}b)\|_Y^2 + \|\eta\|_Y^2 - \|\rho\|_Y^2,
\end{aligned}$$

combining the last two computations, we have

$$\begin{aligned}
\|Q^{-1/2}(\mathrm{div}\,\xi^{p+1} + \beta^{-1}b)\|_X^2 \le \|Q^{-1/2}(\mathrm{div}\,\xi^p + \beta^{-1}b)\|_X^2 + (\kappa^2\nu - 1)\nu\|\eta\|_Y^2 \\
+\nu(\|\rho\|_Y^2 - \|\nabla Q^{-1}(\mathrm{div}\,\xi^p + \beta^{-1}b)\|_Y^2).
\end{aligned} \tag{3.34}$$

Observe that

$$\|\rho\|_Y \le \|\nabla Q^{-1}(\mathrm{div}\,\xi^p + \beta^{-1}b)\|_Y \|\xi^{p+1}\|_Y \le \|\nabla Q^{-1}(\mathrm{div}\,\xi^p + \beta^{-1}b)\|_Y.$$

By choosing $\nu < \frac{1}{\kappa^2}$ we obtain

$$\|Q^{-1/2}(\mathrm{div}\,\xi^{p+1} + \beta^{-1}b)\|_X^2 \le \|Q^{-1/2}(\mathrm{div}\,\xi^p + \beta^{-1}b)\|_X^2,$$

i.e., the sequence $\|Q^{-1/2}(\mathrm{div}\,\xi^p + \beta^{-1}b)\|_X^2$ is decreasing. Let

$$m := \lim_p \|Q^{-1/2}(\mathrm{div}\,\xi^p + \beta^{-1}b)\|_X^2,$$

and let $\overline{\xi}$ be the limit of a convergent subsequence $\{\xi^{p_j}\}$ of $\{\xi^p\}$. Modulo a extraction of a subsequence, we may assume that $\{\xi^{p_j+1}\}$ converges to a vector $\overline{\xi}' \in Y$. Observe that

$$\|Q^{-1/2}(\mathrm{div}\,\overline{\xi}' + \beta^{-1}b)\|_X^2 = \|Q^{-1/2}(\mathrm{div}\,\overline{\xi} + \beta^{-1}b)\|_X^2 = m.$$

Letting $p_j \to \infty$ in (3.33) we obtain

$$\overline{\xi}'(i,j) = \frac{\overline{\xi}(i,j) + \nu\nabla[Q^{-1}\mathrm{div}\,\overline{\xi} + Q^{-1}(\beta^{-1}b)](i,j)}{1 + \nu|\nabla[Q^{-1}\mathrm{div}\,\overline{\xi} + Q^{-1}(\beta^{-1}b)](i,j)|}. \tag{3.35}$$

If we repeat here the above computations or, simply, we let $p_j \to \infty$ in (3.34) we obtain

$$\begin{aligned}
\|Q^{-1/2}(\mathrm{div}\,\overline{\xi}' + \beta^{-1}b)\|_X^2 \le \|Q^{-1/2}(\mathrm{div}\,\overline{\xi} + \beta^{-1}b)\|_X^2 + (\kappa^2\nu - 1)\nu\|\overline{\eta}\|_Y^2 \\
+ \nu(\|\overline{\rho}\|_Y^2 - \|\nabla Q^{-1}(\mathrm{div}\,\overline{\xi} + \beta^{-1}b)\|_Y^2),
\end{aligned}$$

where

$$\overline{\eta} := \frac{\overline{\xi}' - \overline{\xi}}{\nu},$$

and

$$\overline{\rho} := |\nabla Q^{-1}(\mathrm{div}\,\overline{\xi} + \beta^{-1}b)|\overline{\xi}'.$$

Hence

$$(\kappa^2\nu - 1)\|\overline{\eta}\|_Y^2 + \nu(\|\overline{\rho}\|_Y^2 - \|\nabla Q^{-1}(\operatorname{div}\overline{\xi} + \beta^{-1}b)\|_Y^2) = 0$$

If $\nu < \frac{1}{\kappa^2}$, then we deduce that $\overline{\eta} = 0$, that is, $\overline{\xi}' = \overline{\xi}$. Hence

$$\nabla Q^{-1}(\operatorname{div}\overline{\xi} + \beta^{-1}b) - |\nabla Q^{-1}(\operatorname{div}\overline{\xi} + \beta^{-1}b)|\overline{\xi} = 0,$$

which is nothing else than the Euler-Lagrange equation (3.30). Let $\overline{\alpha}^*(i,j) = |\nabla Q^{-1}(\operatorname{div}\overline{\xi} + \beta^{-1}b)(i,j)|$. Observe that $\langle\overline{\alpha}^*\overline{\xi}, \overline{\xi}\rangle \geq \langle\overline{\alpha}^*\overline{\xi}, \xi'\rangle$ for any $\xi' \in \mathcal{V}_d$. This implies that $t \in [0,1] \to \|\operatorname{div}(\overline{\xi} + t(\xi' - \overline{\xi})) + \beta^{-1}b\|_Q^2$ has a minimum at $t = 0$. That is, $\overline{\xi}$ is a minimum of (3.29). Since the projection of $-\beta^{-1}b$ on K_d (when X is endowed with the norm $\|\cdot\|_Q$) is unique, we deduce that $\operatorname{div}\overline{\xi}$ coincides with $\Pi_{K_d}^Q(-\beta^{-1}b)$ and also that all the sequence $\operatorname{div}\xi^p$ converges to $\Pi_{K_d}^Q(-\beta^{-1}b)$.

Finally, since

$$\|\operatorname{div}\xi\|_X^2 \leq 8\|\xi\|_Y^2$$

for every $\xi \in Y$, we have

$$\kappa^2 \leq 8\|Q^{-1/2}\|^2,$$

where $\|Q^{-1/2}\|$ is the norm of $Q^{-1/2}$ on X. Thus, if $\nu < \frac{1}{8\|Q^{-1/2}\|^2}$, we also have that $\nu < \frac{1}{\kappa^2}$. □

A similar result has been proved in [8] (see also [33]) in the context of regularization methods used for $u + v$ decompositions. Minimization of (3.6), or its discrete version (3.27), can be also accomplished with gradient descent methods on regularized functionals F_ϵ which Γ-converge to (3.6) [43]. Other numerical methods can be found in [17, 18, 19, 20, 21, 38, 44].

Algorithm 2: Computing the solution of (3.26)

Assume that the datum b and the matrix Q are given, $\beta > 0$.

1. Solve $w = Q^{-1}b$ by CG up to precision θ_g.
2. Choose an initialization ξ^0, take by default $\xi^0 = 0$.
   Set $p = 0$.
3. While $|\operatorname{div}\xi^{p+1} - \operatorname{div}\xi^p| > \theta_z$ iterate in p:
   (a) Solve $v_p = Q^{-1}(\operatorname{div}\xi^p)$ by CG up to precision θ_z', using as initial condition v_{p-1} for $p > 0$, or 0 if $p = 0$.
   (b) Compute ξ^{p+1} from equation (3.33), i.e.

$$\begin{aligned} G_p &:= \nabla[v_p + \beta^{-1}w] \\ \alpha_p^* &:= |G_p| \\ \xi^{p+1} &:= [\xi^p + \nu G_p]/(1 + \nu\alpha_p^*). \end{aligned}$$

   Let $\tilde{\xi} = \xi^{p^*}$, where p^* is the last iteration where convergence is ensured.
4. Solve $u = Q^{-1}(b + \beta\operatorname{div}\tilde{\xi})$ by CG up to precision θ_u, with initial condition $w + \beta v_{p^*-1}$.

Since Algorithm 2 will be used in combination with Algorithm 3, we shall comment on the parameters used after introducing Algorithm 3 in Section 4.

4. TV-regularized irregular to regular sampling, deconvolution, denoising and zoom. Let $\alpha \geq 1$ be an integer representing the oversampling factor. Assume that $g \in \ell^2(\Lambda)$ is a vector representing the N^2 irregular samples. Our plan is to compute a band limited approximation to the solution of the restoration problem (1.5). For that we define

$$\mathcal{B}_\alpha := \{u \in L^2(\Omega_N) : \hat{u} \,\text{is supported in}\, \{-\frac{\alpha N}{2}, \dots, \frac{\alpha N}{2} - 1\}^2\}$$

If $u \in \mathcal{B}_\alpha$, then u is determined by a vector of $\alpha N \times \alpha N$ regular samples that, with a slight abuse of notation, we shall also call u. Let $F = FFT/(\alpha N)^2$, where $FFT = (e^{-\frac{2\pi i}{\alpha N}\langle j,n\rangle})_{nj}$, so that $a = Fu$ is defined by

$$a_n = \frac{1}{(\alpha N)^2} \sum_{j \in \{0,\dots,\alpha N-1\}^2} u(j/\alpha) e^{-\frac{2\pi i}{\alpha N}\langle j,n\rangle},$$

where

$$n \in I_\alpha := \{-\frac{\alpha N}{2}, \dots, \frac{\alpha N}{2} - 1\}^2.$$

Then the inverse operator F^{-1} is

$$u(j/\alpha) = \sum_{n \in I_\alpha} a_n e^{\frac{2\pi i}{\alpha N}\langle j,n\rangle}, \quad j \in \{0, \dots, \alpha N - 1\}^2$$

and we denote it as $F^{-1} = IFFT$. However the adjoint operator is $F^* = IFFT/(\alpha N)^2$, and therefore $F^*F = I/(\alpha N)^2$.

Similarly, let us call S_Λ, or simply S, the irregular sampling operator mapping $a \in \ell^2(\{-\frac{N}{2}, \dots, \frac{N}{2} - 1\}^2)$ to $\{v(\lambda_k)\}_{k=1}^{N^2} \in \ell^2(\Lambda)$ where v is the trigonometric polynomial whose Fourier coefficients are a, i.e., $a = Fv$, and

$$v(\lambda_k) = (Sa)_k = \sum_{n \in \{-\frac{N}{2}, \dots, \frac{N}{2}-1\}^2} a_n e^{\frac{2\pi i}{N}\langle \lambda_k, n\rangle},$$

$k \in \{1, \dots, N^2\}$. Let $P = \text{diag}(\{\hat{p}(\frac{2\pi n}{\alpha N})\}_{n \in I_\alpha})$, $H = \text{diag}(\{\hat{h}(\frac{2\pi n}{\alpha N})\}_{n \in I_\alpha})$ be the diagonal operators containing along their diagonals the Fourier coefficients of the spectral projector p and the blurring kernel h, respectively. We recall that the spectral projector vanishes for frequencies in $I_\alpha \setminus \{-\frac{N}{2}, \dots, \frac{N}{2} - 1\}^2$. Thus, the notation SPb, $b = (b_n)_{n \in I_\alpha}$, is understood as $S(Pb|_{\{-\frac{N}{2},\dots,\frac{N}{2}-1\}^2})$. Finally, $W = \text{diag}(\{w_k\}_{k \in \{1,\dots,N^2\}})$ is a diagonal matrix containing the weights corresponding to the sampling points $\lambda_k \in \Lambda$ (see (2.6)).

We notice that $\mathcal{B}_\alpha$ is a finite dimensional vector space of dimension $(\alpha N)^2$ which can be identified with $X_{\alpha N}$. Both $J(u) = \int_{\Omega_N} |\nabla u|$ and $J_d^{\alpha N}(u)$ are norms on the quotient space $\mathcal{B}_\alpha/\mathbb{R}$, hence they are equivalent. As above, for simplicity, we shall write J_d instead of $J_d^{\alpha N}$. We shall write $X = X_{\alpha N}$ and with a slight abuse of notation we shall indistinctly write $u \in \mathcal{B}_\alpha$ or $u \in X$.

Now, based on the continuous model (1.5), in the discrete case we shall consider the functional

$$\min_{u \in \mathcal{B}_\alpha} \frac{1}{2} \|W(SHPFu - g)\|^2_{\ell^2(\Lambda)} + \lambda J_d(u), \tag{4.1}$$

for which existence is easily proved ([17, 38, 43, 6]). Uniqueness is only guaranteed when the operator $u \in X \to WSHPFu \in \ell^2(\Lambda)$ is injective ([17, 38, 43, 6]), and this is not the case if we include zooming, i.e., when $\alpha > 1$. When $\alpha = 1$ we take $P = I$ and uniqueness is guaranteed if $H(n) \neq 0$ for $n \in \{-\frac{N}{2}, \dots, \frac{N}{2} - 1\}^2$, S is injective ([27, 32]) and the weights $w_k \neq 0$ for all $k \in \{1, \dots, N^2\}$. Unfortunately we do not know of any condition ensuring the injectivity of S in all cases of interest in satellite imaging. Kadec's $\frac{1}{4}$ theorem and its extensions to the two-dimensional case cannot ensure the stable invertibility of S if the perturbation $|\varepsilon(x)|$ is larger than 0.11 (or 0.25 in the separable case). This fails short to be the case in satellite imaging where perturbations can be of this order or slightly larger. For larger perturbations, or more irregular (non-perturbed) sampling geometries we could use the results of Gröchenig and Strohmer [31, 32] (such as the one reproduced in proposition 2.1). However, these results only ensure the stable invertibility of $S|_B$ where B is an even smaller frequency band than I_1, *i.e.* only when the number of available irregular samples λ_k is significantly higher than the number of Fourier coefficients a_n. Similarly, Beurling-Landau type results only provide a sufficient condition for invertibility of S under a similar constraint. For a thorough discussion on conditions under which S is known to be invertible and well conditioned see [2]. But all these results search for a stronger property (stability) than injectivity, or are stated in the infinite domain case. If we only search for injectivity when we have a finite number of samples in an interval more useful conditions have to be found.

If $\alpha > 1$ and u_1, u_2 are two solutions of (4.1), then also $\frac{u_1+u_2}{2}$ is also a solution and we have

$$J_d(u_1 + u_2) = J_d(u_1) + J_d(u_2),$$

and, since $|\nabla^{+,+}(u_1 + u_2)(i,j)| \leq |\nabla^{+,+}u_1(i,j)| + |\nabla^{+,+}u_2(i,j)|$, we deduce that

$$|\nabla^{+,+}(u_1 + u_2)(i,j)| = |\nabla^{+,+}u_1(i,j)| + |\nabla^{+,+}u_2(i,j)|,$$

for all $(i,j) \in \{0, \dots, N-1\}^2$. This implies that for any $(i,j) \in \{0, \dots, N-1\}^2$ such that $\nabla^{+,+}u_1(i,j) \neq 0$, there is a value of $\eta_{ij} \geq 0$ such that $\nabla^{+,+}u_2(i,j) = \eta_{ij}\nabla^{+,+}u_1(i,j)$, and similarly with u_1 and u_2 interchanged. This type of argument is taken from [38]. A different argument but leading to the same conclusion was considered in [10]. Observe that this is a weak uniqueness result in the sense that two different solutions share a weak form of the topographic map, i.e., the normal to the level lines of one of them is also normal to the level lines of the other (see [10, 38]).

Finally, the Euler-Lagrange equation corresponding to (4.1) is readily written as

$$0 \in F^*P^*H^*S^*W^2SHPFu - F^*P^*H^*S^*W^2g + \lambda\partial J_d(u), \tag{4.2}$$

where $\partial J_d(u)$ is the subdifferential of $J_d(u)$. Now define

$$\begin{aligned} T &:= S^*W^2S/(\alpha N)^2, \quad T' := P^*H^*THP, \\ \hat{r} &:= S^*W^2g/(\alpha N)^2, \quad \hat{r}' := P^*H^*\hat{r}, \end{aligned} \tag{4.3}$$

where the $(\alpha N)^2$ normalization factor is included to absorb the corresponding factor from $F^{-1} = F^*/(\alpha N)^2$. Recall that in this equation T and $\hat{r}$ can be computed from g and Λ in $O(N^2 \log N^2)$ time by means of the NFFT [40], where N^2 is the number of samples in Λ. In addition T has Toeplitz structure and the products of T by a vector can be computed in $O(N^2 \log N^2)$ time using the FFT. Therefore products of T' by a vector also take $O(N^2 \log N^2)$ time, since P and H are diagonal.

Summarizing, the Euler-Lagrange equation (4.2) can be written as

$$0 \in F^{-1}T'Fu - F^{-1}\hat{r}' + \lambda \partial J_d(u), \tag{4.4}$$

and can be solved by looking for a steady state solution of the evolution problem

$$\frac{\partial u}{\partial t} \in -[F^{-1}T'Fu - F^{-1}\hat{r}' + \lambda \partial J_d(u)]. \tag{4.5}$$

For that, we use Euler's implicit scheme, also known as Crandall-Liggett's iteration scheme,

$$u_{m+1} - u_m \in -\tau[F^{-1}T'Fu_{m+1} - F^{-1}\hat{r} + \lambda \partial J_d(u_{m+1})], \tag{4.6}$$

which can be re-written as

$$Qu_{m+1} + \tau\lambda \partial J_d(u_{m+1}) \ni b_m, \tag{4.7}$$

where

$$\begin{aligned} Q &= F^{-1}UF \\ U &= I + \tau T' \\ b_m &= u_m + \tau r' \\ r' &= F^{-1}\hat{r}'. \end{aligned} \tag{4.8}$$

Now for any positive τ, Q is hermitian and positive definite (hence invertible) because T is hermitian positive semidefinite. Thus, we can solve equation (4.7) by applying the extension of Chambolle's technique presented in Section 3, i.e, we solve

$$Qu_{m+1} - \tau\lambda \operatorname{div} \xi = b_m, \tag{4.9}$$

where ξ is the fixed point of

$$G(\xi) := \nabla[\tau\lambda Q^{-1}(\operatorname{div} \xi) + Q^{-1}b_m] = \alpha^* \xi, \tag{4.10}$$

with the Lagrange multipliers α^* chosen in such a way that the constraint $|\xi| \leq 1$ holds. This fixed point is obtained using the iterative scheme

$$\begin{aligned} \alpha_p^* &= |G(\xi^p)| \\ \xi^{p+1} &= [\xi^p + \nu G(\xi^p)]/(1 + \nu\alpha_p^*), \end{aligned} \tag{4.11}$$

with ν satisfying the requirements of Theorem 3.2. In this case $\beta = \tau\lambda$, and $\|Q^{-1/2}\|^2 = \frac{1}{1+\tau r(T')}$, where $r(T')$ denotes the spectral radius of T'. The condition $\nu < \frac{1}{8\|Q^{-1/2}\|^2}$ can be written as

$$\nu < 0.125(1 + \tau r(T'))$$

and it is satisfied in particular if $\nu < 0.125$.

The two inversions of Q that appear in equation (4.10) can be efficiently solved by CG as in the ACT algorithm, since each product of Q by a vector only involves multiplication by Toeplitz or diagonal matrices. In addition, for low values of τ, the

matrix Q is better conditioned than T, which means that CG should converge faster. The condition number of Q can be bounded by

$$\mathrm{cond}(Q) = \frac{\lambda_{\max}(Q)}{\lambda_{\min}(Q)} = \frac{\lambda_{\max}(T')\tau + 1}{\lambda_{\min}(T')\tau + 1} \leq \tau\lambda_{\max}(T') + 1 \tag{4.12}$$

In the particular case $P = H = I$ (no deconvolution and no zoom), $T = T'$, and according to [27], $\lambda_{\max}(T)$ can be bounded, at least in the separable case, by $(1+\upsilon)^4$, if υ is the "maximal gap" of the sampling set.

Putting all these pieces together, we have the following algorithm:

Algorithm 3: TV based restoration by solving (4.4)

1. Compute $\hat{r}$ and (the first row and column in) T by means of the NFFT
2. Choose an initial condition $u_0 = 0$, or a better guess based on the irregular samples g.
   Set $m = 0$.
3. While $|u_{m+1} - u_m| > \theta_u\ \tau$ iterate in m:
   (a) Update $b_m = u_m + \tau r'$.
   (b) Solve $Qu + \lambda\tau\partial J_d(u) \ni b_m$ using Algorithm 2. Let u_{m+1} be the solution.

When using Algorithm 2 in Step 3.(b), better initializations in some of the Steps (of Algorithm 2) can be used. For instance, in Step 1 of Algorithm 2 we solve the system $Q^{-1}b_m$ which we may initialize with the $Q^{-1}b_{m-1}$ if $m > 0$, or with 0 if $m = 0$.

In order to reduce the number of parameters so that the user only needs to specify the TV regularization parameter λ and the precision θ_u of the final solution u, we used the following criteria: The convergence of Chambolle's fixed point iteration in step 3.(b) of Algorithm 2 (used in Step 3 of Algorithm 3) imposes $\nu = 0.125$ and the maximal value of $\tau = \lambda^{-1}$. We set the tolerance $\theta_z = \theta_u/2$ since Q^{-1} is contractive (i.e. this operation does not increase the error already present in $b_m + \lambda\tau\mathrm{div}\,\tilde{\xi}^{m+1}$) and we want an error of order θ_u in u_m. As we want a precision θ_z in $\mathrm{div}\,\xi^{p+1}$ (see step 3.(b) in Algorithm 2), we set $\theta_g = \theta_u/2$ and $\theta'_z = \theta_z/10$. The reason for the tolerance $\theta'_z = \theta_z/10$ in 3.(a) of Algorithm 2 is that, from numerical experimentation, we know that the number of iterations in p done in step 3.(b) of Algorithm 2 does not exceed 10.

For a full justification of Algorithm 3, we prove the convergence of the iterations (4.6) as $m \to \infty$ to a solution of (4.1). This type of result was proved in the continuous case by L. Vese in [43], Theorem 5.4. Our proof will be based on the techniques in [15].

PROPOSITION 4.1. *Let $X = X_{\alpha N}$. For any $u_0 \in X$, the sequence u_m converges to a solution u of (4.1) as $m \to \infty$.*

Proof. For convenience of notation, let us write

$$G(u) := \frac{1}{2}\|W(SHPFu - g)\|^2_{\ell^2(\Lambda)}.$$

Then for each $m \geq 0$ u_{m+1} is a solution of

$$\frac{u_{m+1} - u_m}{\tau} + \lambda\partial J_d(u_{m+1}) + \partial G(u_{m+1}) \ni 0. \tag{4.13}$$

Let $\overline{u} \in X$ be a solution of the variational problem (4.1). Then we have

$$\lambda\partial J_d(\overline{u}) + \partial G(\overline{u}) \ni 0,$$

and, therefore,

$$u_{m+1} - \overline{u} + \tau\lambda(\partial J_d(u_{m+1}) - \partial J_d(\overline{u})) + \tau(\partial G(u_{m+1}) - \partial G(\overline{u})) \ni u_m - \overline{u}. \quad (4.14)$$

Multiplying (4.14) by $u_{m+1}-\overline{u}$ and using that both ∂J_d and ∂G are maximal monotone operators, we have

$$|u_{m+1} - \overline{u}|_X^2 \leq \langle u_m - \overline{u}, u_{m+1} - \overline{u}\rangle_X \leq |u_m - \overline{u}|_X |u_{m+1} - \overline{u}|_X,$$

hence

$$|u_{m+1} - \overline{u}|_X \leq |u_m - \overline{u}|_X, \quad (4.15)$$

and, thus, u_m is bounded in X. Hence, we may extract a subsequence u_{m_j} such that u_{m_j} converges to some function u^* in X.

Now, multiplying (4.13) by $\frac{u_{m+1}-u_m}{\tau}$, taking scalar products, and adding the identities obtained we have

$$\begin{aligned}\sum_{m=0}^{p} \left\| \frac{u_{m+1} - u_m}{\tau} \right\|_X^2 &= \frac{1}{\tau} \sum_{m=0}^{p} \langle \partial(\lambda J_d + G)(u_{m+1}), u_m - u_{m+1}\rangle_X \\ &\leq \frac{1}{\tau} \sum_{m=0}^{p} ((\lambda J_d + G)(u_m) - (\lambda J_d + G)(u_{m+1})).\end{aligned}$$

Hence,

$$\tau \sum_{m=0}^{p} \left\| \frac{u_{m+1} - u_m}{\tau} \right\|_X^2 + (\lambda J_d + G)(u_{p+1}) \leq (\lambda J_d + G)(u_0). \quad (4.16)$$

This estimate implies that $u_{m+1} - u_m \to 0$ as $m \to \infty$.

Letting $m \to \infty$ in (4.13) along the sequence m_i we have that u^* is a solution of

$$\lambda \partial J_d(u) + \partial G(u) \ni 0.$$

In other words, u^* is a solution of (4.1). Replacing $\overline{u}$ by u^* in (4.15), we have

$$|u_{m+1} - u^*|_X \leq |u_m - u^*|_X, \quad (4.17)$$

for all m. Together with $u_{m_i} \to u^*$, this implies that $u_m \to u^*$ as $m \to \infty$. □

This result proves that, in spite of not having a strong form of uniqueness for (4.1), the solutions of the discrete in time gradient descent converge to a solution of it.

5. Constrained image restoration: working with local constraints. As noticed in [12], the use of a global constraint in image restoration problems based on total variation regularization may be inadequate in images which contain very different regions. The usual formulation of the constrained total variation approach to image restoration is based on the minimization problem

$$\begin{aligned}&\min_{u \in X} J_d(u) \\ &\sum_{\lambda_k \in \Lambda} |SHPFu(\lambda_k) - g(\lambda_k)|^2 \leq \sigma^2 N^2,\end{aligned} \quad (5.1)$$

where the constraint incorporates our knowledge on the image acquisition model (1.1) in an integral way. The unconstrained formulation of the problem is:

$$\min_{u\in X}\max_{\gamma\geq 0}\ \mathcal{L}_g(u,\gamma) \tag{5.2}$$

where

$$\mathcal{L}_g(u,\gamma) := J_d(u) + \frac{\gamma}{2}\left[\frac{1}{N^2}\sum_{\lambda_k\in\Lambda}|SHPFu(\lambda_k) - g(\lambda_k)|^2 - \sigma^2\right]$$

and $\gamma \geq 0$ is a Lagrange multiplier. The appropriate value of γ can be computed using Uzawa's algorithm [26] so that the constraint in (5.1) is satisfied. The parameter γ is such that γ^{-1} can be identified with the penalization parameter λ in (4.1). Recall that in this case λ controls the importance of the regularization term, and if we set this parameter to be large, then homogeneous zones are well denoised while highly textured regions will loose a great part of its structure. On the contrary, if λ is set to be small, texture will be kept but noise will remain in homogeneous regions. On the other hand, as the authors of [12] observed, if we use the constrained formulation (5.1), or equivalently (5.2), then the Lagrange multiplier does not produce satisfactory results since we do not keep textures and denoise flat regions simultaneously, and they proposed to incorporate the image acquisition model as a set of local constraints, adapted to the different regions of the image.

Following [12], in order to adapt the restoration of irregularly sampled images to the use of local constraints on the data fidelity term, we assume that $\{O_1, ..., O_r\}$ is a partition of Λ into regions. We define

$$I_{O_l}(u) := \frac{1}{|O_l|}\sum_{\lambda_k\in O_l}|SHPFu(\lambda_k) - g(\lambda_k)|^2 \tag{5.3}$$

where $|O_l|$ denotes the number of pixels in O_l. The problem (5.1) is replaced by

$$\begin{aligned}&\min_{u\in X} J_d(u)\\ &\text{subject to } I_{O_l}(u) \leq \sigma^2, \qquad \forall\, l = 1, ..., r.\end{aligned} \tag{5.4}$$

The unconstrained formulation of this problem is

$$\min_{u\in X}\max_{\gamma_l\geq 0}\ \mathcal{L}(u,\{\gamma_l\}) \tag{5.5}$$

where

$$\mathcal{L}(u,\{\gamma_l\}) := J_d(u) + \frac{1}{2}\sum_{l=1}^{r}\gamma_l(I_{O_l}(u) - \sigma^2),$$

where γ_l, $l = 1, \ldots, r$ is a set Lagrange multipliers associated to the constraints in (5.4). Let us recall that the partition $\{O_l\}_{l=1}^r$ can be obtained with the simplified version [35] of the Mumford-Shah segmentation algorithm [39] applied to the image g. The segmentation algorithm is stopped to ensure that regions O_l have an area of at least 200 pixels to ensure that the sum in (5.3) gives a good estimate of σ^2 ([12]).

The strategy would be to use Uzawa's method (see [26]) to solve problem (5.5). To guarantee that the assumptions of Uzawa's method hold we shall use a gradient

descent strategy. For that, let $\tau > 0$, $v \in X$. At each step we have to solve a problem like

$$\begin{aligned} &\min_{u\in X} \frac{1}{2\tau}\|u-v\|_X^2 + J_d(u) \\ &\text{subject to } I_{O_l}(u) \le \sigma^2, \qquad \forall\, l = 1, ..., r, \end{aligned} \tag{5.6}$$

whose unconstrained formulation is

$$\min_{u\in X} \max_{\gamma_l \ge 0} \mathcal{L}^\tau(u, \{\gamma_l\}; v) \tag{5.7}$$

where

$$\mathcal{L}^\tau(u, \{\gamma_l\}; v) := \frac{1}{2\tau}\|u-v\|_X^2 + J_d(u) + \frac{1}{2}\sum_{l=1}^{r} \gamma_l (I_{O_l}(u) - \sigma^2).$$

Note that for fixed values of $\{\gamma_l\}$, the minimization with respect to u in (5.7) amounts to solving a problem of type (4.7) if we recall how (4.7) was derived from (4.1), and we observe that we may write

$$\sum_{l=1}^{r} \gamma_l I_{O_l}(u) = \sum_{\lambda_k \in \Lambda} \omega(\lambda_k) |SHPFu(\lambda_k) - g(\lambda_k)|^2 = \|W(SHPFu - g)\|_{\ell^2(\Lambda)}$$

where $W = \mathrm{diag}(\{\sqrt{\omega(\lambda_k)}\}_{k=1,\dots,N^2})$, where $\omega(\lambda_k) = \sum_{l=1}^{r} \frac{\gamma_l}{|O_l|}\chi_{O_l}(\lambda_k)$, and $\lambda = 1$ in (4.1). So, the structure of the algorithm adapted to the case of local constraints is basically the same, except for the actualization of W, T' and $\hat{r}'$ each time that the values of γ_l change. Recall that in the previous case $W = I$, and T', $\hat{r}'$ defined in (4.3) were computed only at the beginning.

Thus, as regards the numerical implementation, the algorithm is:

Algorithm 4: TV based restoration algorithm with local constraints

1. Set $u_0 = 0$ or a better guess based on the irregular samples g.
   Set $m = 0$.
2. Use Uzawa's algorithm to solve the problem

$$\min_{u\in X} \max_{\gamma_l \ge 0} \mathcal{L}^\tau(u, \{\gamma_l\}; u_m), \tag{5.8}$$

   that is:
   (a) Choose any set of values $\gamma_l^0 \ge 0$, $l = 1, \dots, r$.
   Iterate from $p = 0$ until convergence of γ_l^p the following steps:
   (b) With the values of γ_l^p solve

$$\min_u \mathcal{L}^\tau(u, \{\gamma_l^p\}; u_m)$$

   using Algorithm 2 with $\lambda = 1$, $W = \mathrm{diag}(\{\sqrt{\omega(\lambda_k)}\}_{k=1,\dots,N^2})$, where $\omega(\lambda_k) = \sum_{l=1}^{r} \frac{\gamma_l^p}{|O_l|}\chi_{O_l}(\lambda_k)$, starting with the initial condition u^p (if $p = 0$, we initialize it with u_m). Let u^{p+1} be the solution obtained.
   (c) Update γ_l, $\forall\, l = 1, ..., r$, in the following way:

$$\gamma_l^{p+1} = \max\left(\gamma_l^p + \rho(I_{O_l}(u^{p+1}) - \sigma^2), 0\right).$$

Let u_{m+1} be the solution of (5.8). Stop when $\|u_{m+1} - u_m\|_X \leq \theta_u \tau$.

In practice, to reduce computations, step 2.(b) is performed by solving 10 iterations of Step 3 in Algorithm 2, without observing great differences in the global convergence.

Let us say some final words about the convergence of Algorithm 4.

PROPOSITION 5.1. *(i) Assume that $g \in SHPF(X)$. Then Uzawa's algorithm in Step 2 of Algorithm 4 converges.*
(ii) Assume that u_0 satisfies the constraints. Then u_m tends to a solution u of (5.4) as $m \to \infty$.

Proof. (i) Since $\tau > 0$, the assumptions of Uzawa's algorithm [26] are satisfied once we prove that $I_{O_l}(u)$ is Lipschitz on bounded sets of X and we observe that the sequence u^p constructed in Step 2.(b) is bounded in X. Assume that $U \subseteq X$ is a bounded set. Let $u, \overline{u} \in U$, and denote $\overline{H} = SHPF$. Then

$$\begin{aligned} |O_l||I_{O_l}(u) - I_{O_l}(\overline{u})| &\leq 2\|g\|_{\ell^2(\Lambda)}\|\overline{H}u - \overline{H}\overline{u}\|_{\ell^2(\Lambda)} + |\langle \overline{H}(u-\overline{u}), \overline{H}(u+\overline{u})\rangle| \\ &\leq C\|u - \overline{u}\|_X \end{aligned}$$

where C is a constant depending on the norm of $\overline{H}$ and on the bound for U. Now, to prove that $\{u^p\}_p$ is bounded we observe that

$$\mathcal{L}^\tau(u^p, \{\gamma_l^p\}; u_m) \leq \mathcal{L}^\tau(u, \{\gamma_l^p\}; u_m), \qquad \forall u \in X,$$

for all p. Choosing $u \in X$ such that $g = \overline{H}u$, we obtain that

$$\frac{1}{2\tau}\|u^p - v\|_X^2 + J_d(u^p) \leq \frac{1}{2\tau}\|u - v\|_X^2 + J_d(u),$$

hence $\{u^p\}_p$ is bounded in X. Now, (i) follows from Theorem 5 in [26], Sect. 3.1.
(ii) Let C be the convex set of functions $u \in X$ such that $I_{O_l}(u) \leq \sigma^2$. Let $\delta_C(u) = 0$ if $u \in C$ and $= +\infty$, otherwise. Now, we observe that, by (i), u_{m+1} are solutions of (5.6) with $v = u_m$ and, since the domain of J_d is all X, we have $\partial(J_d + \delta_C) = \partial J_d + \partial\delta_C$ ([15], Corollary 2.11), and we may write

$$\frac{u_{m+1} - u_m}{\tau} + \partial J_d(u_{m+1}) + \partial\delta_C(u_{m+1}) \ni 0. \tag{5.9}$$

Now, the proof follows by proceeding as in the proof of Proposition 4.1. □

The assumption $g \in SHPF(X)$ in (i) holds if $H(n) \neq 0$ for $n \in \{-\frac{N}{2}, \ldots, \frac{N}{2} - 1\}^2$ and S is bijective. Concerning (ii), in practice we take u_0 as the solution of the ACT algorithm with g as set of irregular samples.

6. Zoom. In this section, we analyze in more detail the method introduced in section 4 to increase the resolution of the image g.

If we denote by α the zoom factor, then the zoomed and restored image u is a vector of size $\alpha N \times \alpha N$ (we recall that the size of g is $N \times N$). The minimization problem (4.1) with $\hat{p}_d = \chi_{\{-\frac{N}{2}, \ldots, \frac{N}{2}-1\}^2}$ (we denote $\hat{p}_d$ the discrete version of $\hat{p}$) is a direct extension of the oversampling and denoising method introduced by Malgouyres and Guichard [37, 38] to the more general case of regular sampling, deconvolution, denoising and oversampling.

The basic idea is to fit, inasmuch as this is possible, the low frequency components of the restored and zoomed image to the original data, and to extrapolate the

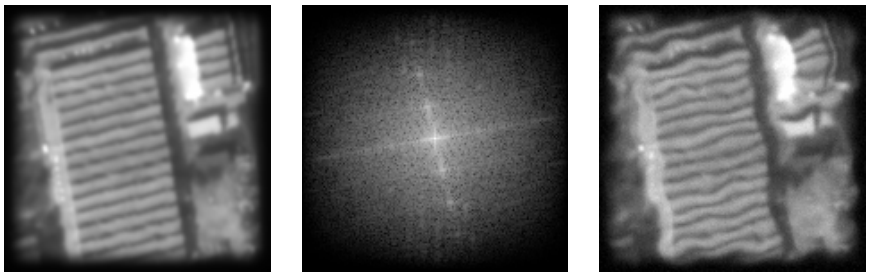


FIG. 7.1. Reference image (left), its Fourier spectrum (middle) and an example of perturbed and noised image (right). *The reference image has 149 × 149 pixels and was multiplied by a smooth window on the borders in order to avoid periodization artifacts. The perturbation is a colored noise with standard deviation of $A = 1$ pixels, and spectral contents inside $[-\frac{1}{T_\varepsilon}, \frac{1}{T_\varepsilon}]^2$ for $T_\varepsilon = 20$.*

spectrum to the rest of the frequency domain by means of the total variation. This regularization allows to reconstruct some high frequencies, which is indeed much more convenient than just filling them with zeros, a technique which is known to produce ringing. For that we require that $\hat{p}_d \neq 0$ in the spectral region $\{-\frac{N}{2}, \dots, \frac{N}{2} - 1\}^2$ and $\hat{p}_d = 0$ in $I_\alpha \backslash \{-\frac{N}{2}, \dots, \frac{N}{2} - 1\}^2$ so that we do not impose any restriction at high frequencies and the extrapolation there is done by means of the total variation, which limits the oscillations. The characteristic function $\hat{p}_d = \chi_{\{-\frac{N}{2}, \dots, \frac{N}{2}-1\}^2}$ fulfils both conditions. We also used a prolate function but we have not noticed great changes in the experiments.

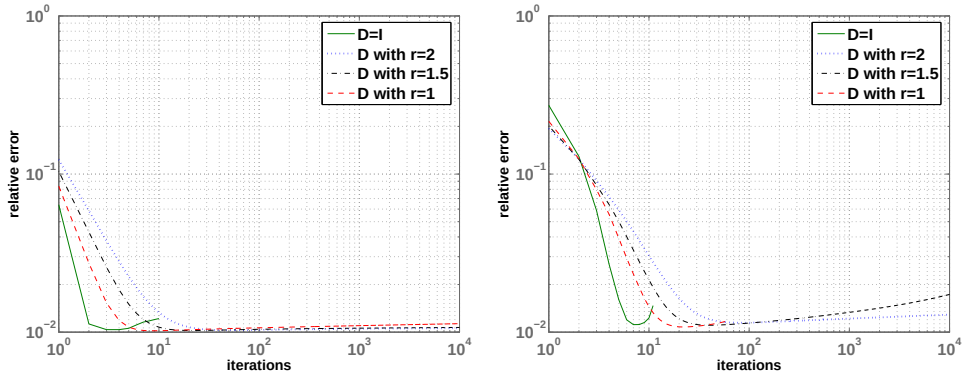
As it is proved in [38], the pure zoom problem (without denoising) admits a solution but uniqueness is not guaranteed. We only have the weak uniqueness result mentioned in Section 4. The case of zooming with denoising has been considered in [37].

7. Experiments.

7.1. Irregular to regular sampling and denoising. In order to test and compare the performance of the different algorithms, we worked with the reference image displayed in Figure 7.1 which is of size 149 × 149. Then, using (2.2), we simulated different perturbations $\varepsilon(x)$ such that $\operatorname{supp} \hat{\varepsilon} \subseteq [-\frac{1}{T_\varepsilon}, \frac{1}{T_\varepsilon}]^2$ for $T_\varepsilon = 20$, for different standard deviations A, typically, 0.25 and 1. Given the perturbation, we simulated the perturbed images g with a high precision (usually 10^{-8}) using the transposed NFFT [40]. Finally, we added some white noise to the irregular samples with standard deviation σ, for different values such as $\sigma = 0.64$ (i.e. 10^{-2} times smaller than the standard deviation of the image, $SNR = 40dB$), and $\sigma = 2$ gray levels. In this way, we obtained perturbed images such as the one displayed in Figure 7.1 with $A = 1$ and $\sigma = 2$.

The aim of this section is to compare the results of our algorithm with those obtained with the original ACT algorithm, i.e., without the regularization produced by the total variation. In both cases we have used different types of weights $D = \operatorname{diag}\left(\{\phi(\frac{2\pi}{N} n)\}_{n \in \{-\frac{N}{2}, \dots, \frac{N}{2}-1\}^2}\right)$ in order to regularize the solution (recall that $\phi(\omega) = (1 + |\omega|)^{-r}$ for $r \in [1, 2]$).

Figure 7.2 displays the evolution of the relative error (in L^2 norm) along the iterations of the CG in the ACT algorithm (Algorithm 1). Observing the graphics in Figure 7.2, we can say that the ACT algorithm attains a level of relative error which



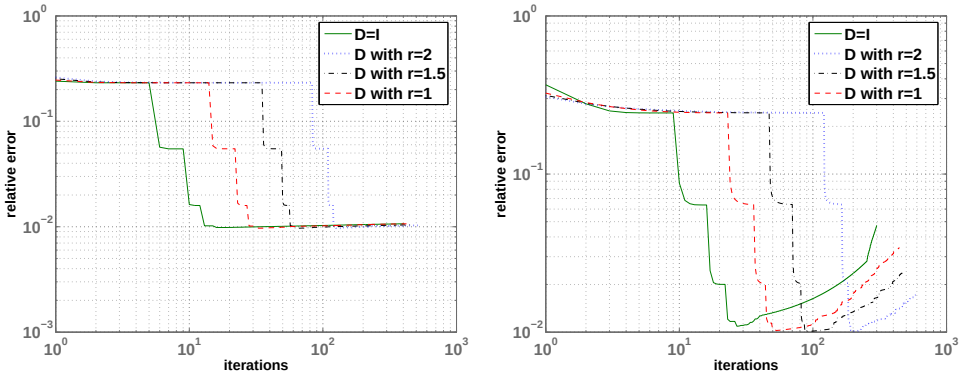
(a) $A = 0.25$, $\sigma = 0.64$ (b) $A = 1$, $\sigma = 0.64$

FIG. 7.2. Convergence graphics of the ACT algorithm. *We display the relative error (in L^2 norm) with respect to the reference image in Figure 7.1 along iterations of the ACT algorithm (with different type of weights) for different values of perturbation amplitude A (0.25 and 1) and noise level $\sigma = 0.64$. The weights impose a spectral decay $(1 + |\omega|)^{-r}$, we test values $r = 0$ (no weights $D = I$) and $r = 1, 1.5$, and 2.*

agrees with the SNR. The presence of the weights that regularize the solution and impose a specific spectral decay make the algorithm more stable. We can also observe that, if we do not use weights, the ACT algorithm begins to diverge after reaching the minimum error and the rate of divergence grows as the standard deviation A of the perturbation becomes larger. In addition, using weights, the relative error is slightly smaller, and the errors in the high frequency band are significantly decreased. However, the presence of weights makes the algorithm to have a poorer convergence rate. This decrease in the convergence rate when using weights is also observed in the case of the proposed algorithm, which combines the ACT algorithm with total variation regularization (see Figure 7.3). Notice that the horizontal axis in Figure 7.3 corresponds to iterations of the CG in step 4 of Algorithm 2 which is called in step 3 of Algorithm 3 (and thus for each time iteration m). We also notice that, in Figure 7.3, the error is strongly reduced at some specific moments, they correspond to a change from time step m to $m+1$. This algorithm seems to be more stable than the ACT algorithm for small perturbation amplitudes.

Figure 7.4 displays some examples of restoration with both algorithms, the ACT and the TV based algorithm 3, for the perturbed image in Figure 7.1. We also display the root mean square error (RMSE) in space, and the error images in space and Fourier domains, with respect to the reference image (the restored images always correspond to a final iteration before the algorithm begins to diverge).

Finally, let us say some words on the execution times. Our experiments were made with a Pentium M, with a 2GHz processor and 2GB of RAM. The experiments displayed in Figure 7.2 on the convergence of the ACT algorithm took between 20 seconds when $A = 0.25$, $\sigma = 0.64$ to 1 minute when $A = 1$, $\sigma = 0.64$. The experiments displayed in Figure 7.3 on the convergence of the combined ACT algorithm with total variation regularization took between 10 seconds in the case of no weights to 12 seconds in the case of weights for both cases $A = 0.25$, $\sigma = 0.64$ and $A = 1$, $\sigma = 0.64$.



(a) $A = 0.25$, $\sigma = 0.64$ (b) $A = 1$, $\sigma = 0.64$

FIG. 7.3. Convergence graphics of the TV based algorithm 3. *We display the relative error (in L^2 norm) with respect to the reference image in Figure 7.1 along the iterations of the proposed TV based algorithm 3 (with different type of weights) for different values of perturbation amplitude A (0.25 and 1) and noise level $\sigma = 0.64$. The weights impose a spectral decay of type $(1 + |\omega|)^{-r}$, and we tested the values $r = 0$ (no weights $D = I$) and $r = 1, 1.5,$ and 2.*

The experiments displayed in Figure 7.6 on the convergence of the ACT algorithm including deconvolution were quite slow, they took around 30 minutes in the case of using weights (when we use no weights the algorithm started to diverge after some iterations). The experiments displayed in Figure 7.7 on the convergence of the TV based algorithm 3 for deconvolution took around 50 seconds when there are no weights and 1 minute 20 seconds when we use weights (the number of time steps is 25). The typical execution times per iteration are 1 second in the case of the ACT algorithm and 2 seconds in the case of the TV based algorithm 3.

7.2. Adding deconvolution.

7.2.1. Test of convergence. After extending the ACT algorithm to include deconvolution, we proceed to numerical tests of convergence, both for the ACT and the TV based deconvolution algorithms.

Let us first explain how to extend the ACT algorithm to the deconvolution case. In that case, the system to solve is

$$g = SHa,$$

where $a = Fu$ are the Fourier coefficients of the restored image u. Then, as in the original ACT algorithm, it is more convenient to solve the system of normal equations

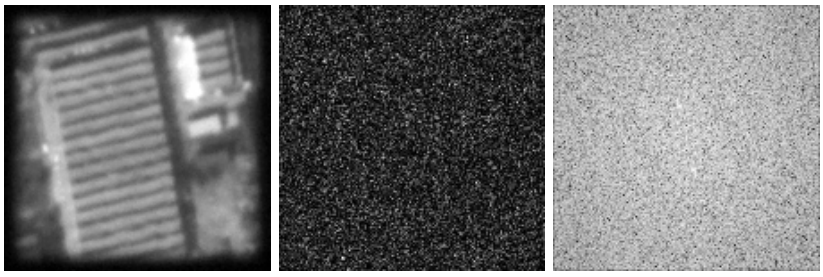
$$H^*S^*WSHa = H^*S^*Wg, \tag{7.1}$$

where W is as in Section 2.1 a matrix of weights (or to compute the minimum norm solution of the optimization problem $\min_a \|\overline{T}a - \overline{b}\|^2_{\ell^2_N}$, where $\overline{T} = H^*TH$ is still a Toeplitz matrix, and $\overline{b} = H^*b$). Notice that the ACT algorithm does not need structural changes in order to include deconvolution.

To simulate our data we use the modulation transfer function corresponding to SPOT 5 HRG satellite with Hipermode sampling (see Appendix 9 and [2] for more

(a) Results with ACT algorithm (space RMSE = 2.35)

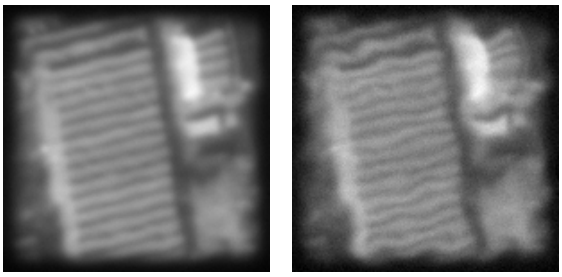
(b) Results with ACT algorithm using weights D with $r = 2$ (space RMSE = 2.07)



(c) Results with TV minimization (space RMSE = 1.53)

(d) Results with TV minimization using weights D with $r = 2$ (space RMSE = 1.53)

FIG. 7.4. Irregular to regular sampling with the ACT and the TV based algorithms. *The perturbation amplitude is $A = 1$, and the noise level is $\sigma = 2$. The left column shows the restored image; the middle column the error in the spatial domain; the right column the error in the Fourier domain.*



(a) $A = 0.25$, $\sigma = 0.64$ (b) $A = 1$, $\sigma = 2$

FIG. 7.5. Filtered, perturbed and noised images. *To construct these images we have used the reference image displayed in Figure 7.1. We filtered this image with filter (7.2). The perturbation is a colored noise with a standard deviation of A pixels, and spectral contents inside* $[-\frac{1}{T_\varepsilon}, \frac{1}{T_\varepsilon}]^2$ *for* $T_\varepsilon = 20$. *Finally, we added a white noise of standard deviation* σ.

details):

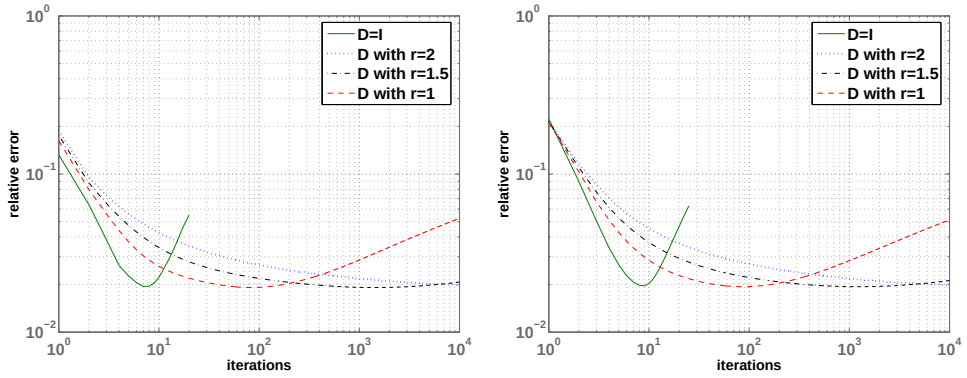
$$\hat{h}(\xi,\eta) = e^{-4\pi\beta_1|\xi|}e^{-4\pi\alpha\sqrt{\xi^2+\eta^2}}\,\mathrm{sinc}(2\xi)\,\mathrm{sinc}(2\eta)\,\mathrm{sinc}(\xi), \qquad \xi,\eta \in [-1/2, 1/2], \tag{7.2}$$

where $\mathrm{sinc}(\xi) = \sin(\pi\xi)/(\pi\xi)$, $\alpha = 0.58$, and $\beta_1 = 0.14$. Then we filter the reference image given in Figure 7.1 with the filter (7.2) and use the same perturbation in sampling as in the previous section. Finally, we add some white noise of standard deviation σ to obtain the images 7.5(a), where $A = 0.25$, $\sigma = 0.64$, and 7.5(b), where $A = 1$, $\sigma = 2$.

The relative errors (in L^2 norm) along iterations for both algorithms can be seen in Figures 7.6 and 7.7. From these experiments, we conclude that the TV based algorithm 3 is always, for small and large values of A, stable in the deconvolution case, unlike the ACT algorithm. We also note that the use of weights does not improve the solution, still, it reduces the rate of convergence. Some restored images, obtained with both methods, and their corresponding errors in space and frequency are displayed in Figure 7.8.

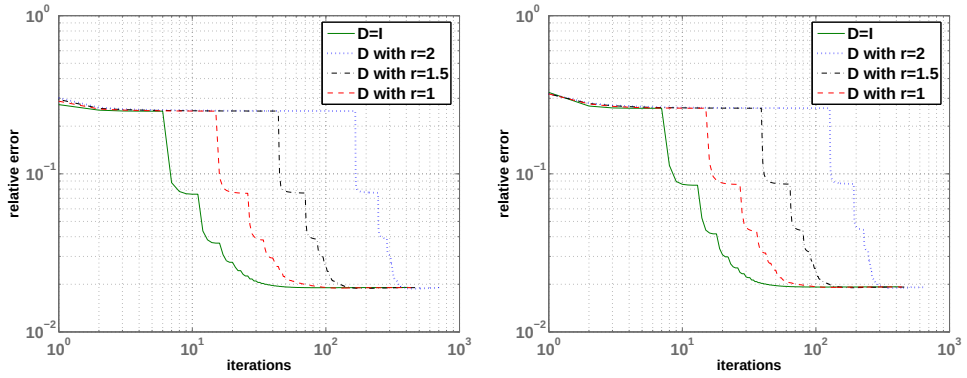
7.2.2. Experiments using a global regularization parameter. In this section we display some examples of the TV based deconvolution algorithm when we use a single regularization parameter, i.e., when we use Algorithm 3. For that, we use the reference image displayed in Figure 7.9(a) and we degrade it with the modulation transfer function (7.2). Then we apply a perturbation in sampling which has the same characteristics as the one used in previous examples ($T_\varepsilon = 20$), and we add white noise of standard deviation $\sigma = 1$. Figure 7.9(b) is generated with a large perturbation ($A = 1$).

Some results of the TV based deconvolution algorithm for different values of the regularization parameter λ are displayed in Figure 7.10. These examples show the drawbacks of using a global regularization parameter. Figure 7.10(a) corresponds to the restoration with the optimal λ such that the constraint $\sum_{k=1}^{N^2} |SHPFu^p(\lambda_k) - g(\lambda_k)|^2 = N^2\sigma^2$ is verified but the result is not satisfactory because it is difficult to get denoised smooth regions and keep the textures at the same time (see the detail in Figure 7.11(a)). On the other hand, notice that most textures are lost when using



(a) $A = 0.25$, $\sigma = 0.64$ (b) $A = 1$, $\sigma = 0.64$

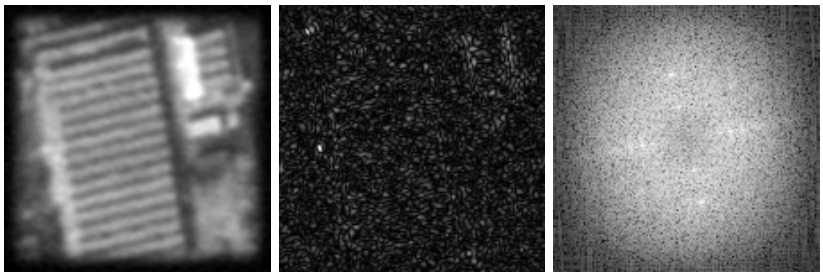
FIG. 7.6. Convergence graphics of the ACT algorithm extended to include deconvolution. *We display the relative error (in L^2 norm) with respect to the reference image in Figure 7.1 along iterations of the ACT algorithm with deconvolution (with different type of weights) for different values of the perturbation amplitude A (0.25 and 1), and noise level $\sigma = 0.64$. We used the weights to impose a spectral decay of $(1 + |\omega|)^{-r}$, and we tested the values $r = 0$ (no weights $D = I$), $r = 1, 1.5$, and 2.*



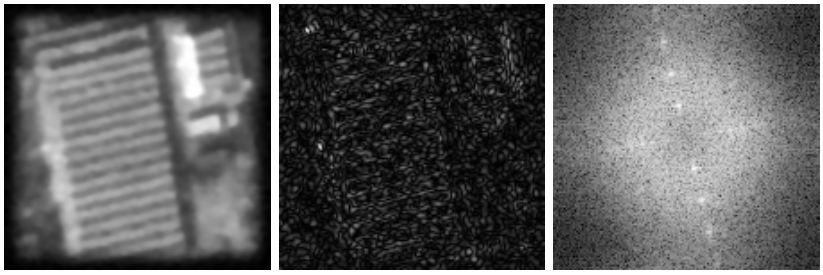
(a) $A = 0.25$, $\sigma = 0.64$ (b) $A = 1$, $\sigma = 0.64$

FIG. 7.7. Convergence graphics of the TV based algorithm 3 (deconvolution case). *We display the relative error (in L^2 norm) with respect to the reference image in Figure 7.1 along iterations of the proposed TV based algorithm (with different type of weights) for different values of the perturbation amplitude A (0.25 and 1), and noise level $\sigma = 0.64$. We use the weights to impose a spectral decay of $(1 + |\omega|)^{-r}$, and we tested the values $r = 0$ (no weights $D = I$), $r = 1, 1.5$, and 2.*

a large value of λ (see Figure 7.10(b) and its detail in 7.11(b)) and how much noise is present if we use a smaller value of λ (see Figure 7.10(c) and its detail in 7.11(c)). In next subsection we will see how these problems can be solved by incorporating the constraints in a local way. We also display in Figure 7.12(a) the result of the ACT algorithm extended to the deconvolution case and using regularization weights



(a) ACT algorithm (using weights D with $r = 2$), $A = 1$, $\sigma = 2$ (RMSE = 2.02)



(b) TV minimization (no weights, $D = I$), $A = 1$, $\sigma = 2$ (RMSE = 1.89)

FIG. 7.8. Irregular to regular sampling, deconvolution and denoising with the ACT algorithm and the TV based algorithm 3. *The perturbation amplitude is $A = 1$, and the noise level is $\sigma = 2$. The left column displays the restored image; the middle column the error in the spatial domain; the right column the error in the Fourier domain.*

which impose a spectral quadratic decay. We notice that the root mean square error (RMSE, measured in space) of the result is larger than in the TV based algorithm. Moreover, visually, the restore image is not as good as the ones displayed in Figure 7.10 (obtained with the TV based algorithm).

7.2.3. Experiments using local parameters. Figure 7.12(b) displays the restored image corresponding to the perturbed image in Figures 7.9(b). This result has been obtained using Algorithm 4, i.e., we have used local data fidelity parameters. Notice that the textures in buildings are recovered and noise is not so evident as in the case of TV restoration with a small global parameter (see Figure 7.13). We also note that the RMS error is reduced with respect to the global parameter case (see Figure 7.10) and, specially, with respect to the ACT algorithm (see Figure 7.12(a)).

In order to test the performance of the Lagrange multipliers to impose the local constraints we have measured in three regions O_l the evolution of the integrals

$$I_{O_l}(u^p) = \frac{1}{|O_l|} \sum_{\lambda_k \in O_l} |SHPFu^p(\lambda_k) - g(\lambda_k)|^2$$

along the iterations p and m of Algorithm 4. The two regions O_l, $l = 1, 2$, are displayed in Figure 7.14(b). We have chosen a textured region (region 1) and a more regular one (region 2). The segmentation of image 7.9(b) can be seen in Figure 7.14(a). Figure

(a) reference image (b) filtered, perturbed and noised image

FIG. 7.9. Reference image and a filtered, perturbed and noised image. *The reference image has 237 × 237 pixels. The perturbation in sampling is a colored noise with standard deviation of* $A = 1$ *pixels, and spectral contents inside* $[-\frac{1}{T_\varepsilon}, \frac{1}{T_\varepsilon}]^2$ *for* $T_\varepsilon = 20$. *The filtered, perturbed and noised image with white noise of standard deviation* $\sigma = 1$ *is also displayed.*

7.15 displays the two graphics corresponding to the evolution of I_{O_l} for each each one of the two regions marked in Figure 7.14(b). For each of these regions, the local constraint $I_{O_l} \leq \sigma^2$ is satisfied and the convergence is quite fast. We observe here that I_{O_l} is computed on the pixels which are interior to the region O_l (we dismissed the pixels on the boundary of O_l). We recall here that for each region O_l we have $I_{O_l}(u) \leq \sigma^2$ and the Lagrange multiplier γ_l satisfies $\gamma_l(I_{O_l}(u) - \sigma^2) = 0$, that is, $I_{O_l}(u) = \sigma^2$ if $\gamma_l > 0$. While the evolution of I_{O_l} really approaches 1 in the case of a more regular area, it is slightly below 1 (0.85) in the case of the textured region.

7.3. Adding zoom. Finally we display an example with the full restoration model: irregular to regular sampling, deconvolution, denoising and zoom. Zoom is added by using a projector P as described in section 6. We also tested a prolate function as a projector instead of the ideal window without noticeable changes in the experiments. Figure 7.16 displays the result corresponding to the restoration and zooming by a factor 2 of image 7.9(b). The result has been obtained with Algorithm 4 with a Lagrange multiplier for each region of the segmentation displayed in Figure 7.14(a).

8. Conclusions. We have proposed an algorithm to solve a general formulation of the image restoration problem which considers several different aspects of it, namely: irregular sampling, denoising, deconvolution, and zooming. Our algorithm is based on an extension of a previous algorithm proposed by A. Chambolle [16] for Total Variation based image denoising, combined with irregular to regular sampling algorithms proposed by Gröchenig and his coauthors [27, 32].

Regarding the irregular to regular sampling and denoising problems, we have compared our approach with the results obtained with the ACT algorithm of Gröchenig

(a) restored with $\lambda = 0.1$ (RMSE = 8.05)

(b) restored with $\lambda = 0.5$ (RMSE = 9.86)

(c) restored with $\lambda = 0.01$ (RMSE = 7.79)

FIG. 7.10. Restored images with a global TV regularization parameter. *Some examples of restoration of the image displayed in Figure 7.9(b) for three different values of λ. Figure 7.10(a) corresponds to the restoration with the optimal $\lambda(= 0.1)$ such that the constraint $\sum_{k=1}^{N^2} |SHPFu^p(\lambda_k) - g(\lambda_k)|^2 = N^2\sigma^2$ is verified. (b) We notice that noise is reduced and textures (high frequencies) are lost if we use a large value of λ. (c) If a small value of λ is used, noise is still present in non-textured regions.*

et al. [27, 32]. The ACT algorithm makes use of Fourier weights in order to regularize and stabilize the solution. In our approach, the regularization is done by Total Variation and the inclusion of Fourier weights does not really benefit the algorithm. Whereas both approaches manage to control the artifacts on high frequencies, TV minimization was found to be a better denoiser, and better reduce errors for fair

(a) restored with $\lambda = 0.1$

(b) restored with $\lambda = 0.5$

(c) restored with $\lambda = 0.01$

FIG. 7.11. A detail of the restored images of Figure 7.10. *(a), (b) and (c) represent a detail of Figures 7.10.(a), (b), (c), respectively.*

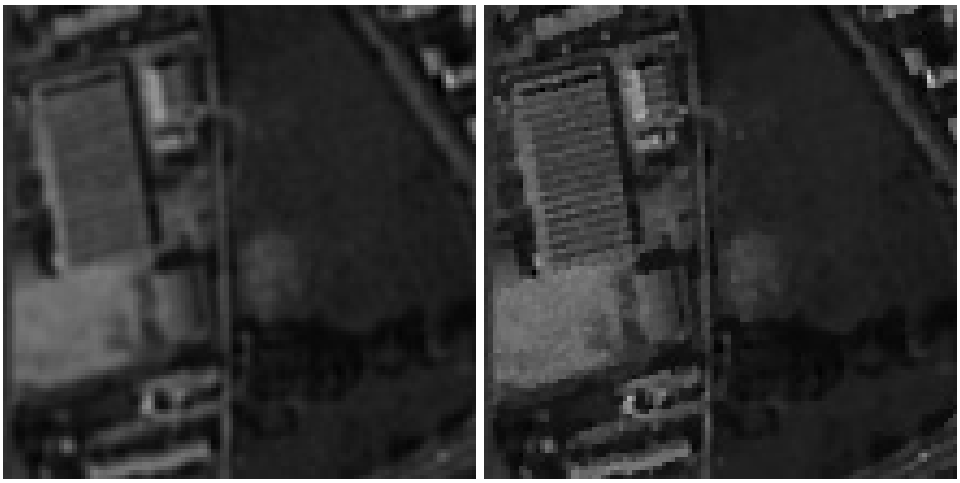
noise levels and perturbation amplitudes. In particular, the error measures behave better. However, the RMS errors of both algorithms are comparable in the case of small perturbations and small noise levels, if we use the regularized version of the ACT algorithm.

To test the proposed algorithm in the case of deconvolution (with irregular to regular sampling and denoising) we have compared it with a direct extension of the ACT algorithm which includes deconvolution. We conclude that the TV based algorithm is more stable in this case. On the other hand, the ACT algorithm starts to diverge after

(a) ACT algorithm extended to the deconvolution case (RMSE = 10.16)

(b) TV based algorithm with local data fidelity parameters (RMSE = 7.36)

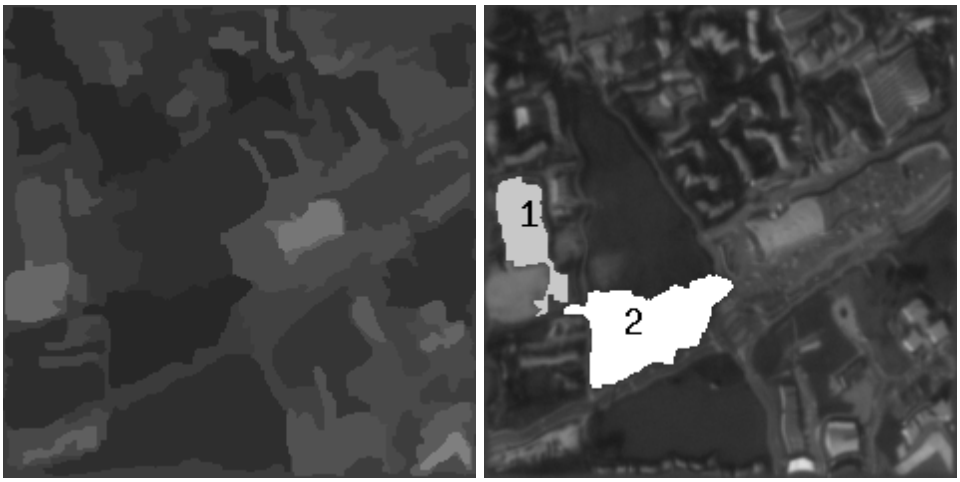
FIG. 7.12. Restored images with the ACT algorithm and TV based algorithm 4. *Both images are restorations of image in Figure 7.9(b). Left: result of the ACT algorithm extended to the deconvolution case (uses regularization weights D with $r = 2$). Right: TV based algorithm with local data fidelity parameters (Algorithm 4).*



(a) Restored with the ACT algorithm extended to the deconvolution case

(b) Restored with the TV based algorithm with local data fidelity parameters

FIG. 7.13. A detail of the restored images of Figure 7.12. *(a), (b) represent a detail of Figures 7.12.(a), (b), respectively.*



(a) segmentation of image 7.9(b) (b) three segmented regions to study

FIG. 7.14. Segmented image and the three regions under study. *The left image displays the segmentation of image 7.9(b) obtained using the simplified Mumford-Shah functional. To ensure that the segmented regions have an area greater than 200 pixels, the image 7.9(b) was filtered with a grain filter with area threshold of 200 pixels. The right image displays three regions in image 7.9(b) where we shall follow the evolution of the constraint.*

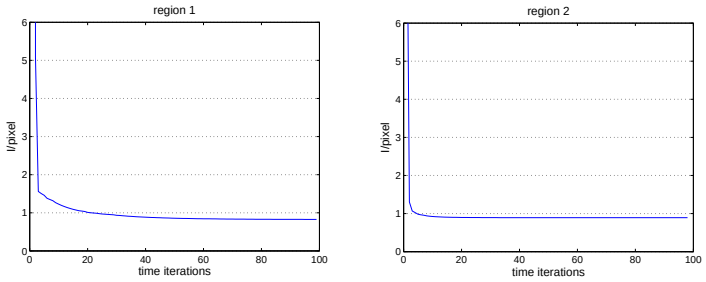


FIG. 7.15. Convergence along iterations of I_{O_l} for two different regions. *Recall that the restored image is displayed Figure 7.9(b) and* $\sigma^2 = 1$. *The regions under study are displayed in Figure 7.14(b). The evolution of* I_{O_l} *corresponds to the restored image 7.12(b).*

reaching the minimum, thus, it becomes crucial to choose a good stopping criterion. In addition, the root mean square errors are smaller with the TV based algorithm, specially if we use local constraints.

In our first experiments with the proposed TV based algorithm we use a global TV regularization parameter λ. We have seen that for large values of λ the image is well denoised, but high frequencies, or textures, are lost. In contrast, for smaller values of λ textures are recovered, but the noise is becomes visible. To overcome this difficulty, following [12], the data fidelity term is incorporated as a set of local constraints.

Finally, we have displayed some results with the full restoration model including zoom. To conclude, we can say that, in general, TV based restoration is more ro-

FIG. 7.16. Restored and zoomed image obtained using local data fidelity parameters. *The image is a zoomed restoration of image in Figure 7.9(b).*

bust and more versatile since many different aspects of the problem can be included. Moreover, the more stable, regularized, version of the ACT algorithm needs the a priori knowledge of the spectral decay of the image.

As a future work, a better error estimate is needed to stop the CG iterations at the right point when inverting Q. Currently, we perform many more iterations than needed, because our error estimate is not optimal. On the other hand, in our experiments we have always worked with a MTF whose support is contained in $[-1/2, 1/2]^2$. When this is not true, antialiasing has to be included into the model and our algorithm can be extended to incorporate it. For that, it suffices to adapt the projector to an appropriate region where sampled coefficients contain a minimal amount of aliasing, as proposed in [2, 3].

Acknowledgements.

The authors would like to thank François Alter for his programming help in the initial stages of this project. We would like to thank also Coloma Ballester and Laura Igual for their help. The first author was partially supported by Programa de Desarrollo Tecnológico in Uruguay (PDT S/C/OP/17/01). The second and third authors acknowledge partial support by the Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya and by PNPGC project, reference BFM2003-02125. The third author was supported by a grant from the Spanish Ministerio de Ciencia y Tecnología, reference FP2000-5801 and by the Institute for Mathematics and its Applications (University of Minnesota). A grant by IBRO (Inter-

national Brain Research Organisation) and the organization of the 2004 Symposium "Representation of Reality by Brain and Machines" allowed a closer collaboration of the first and third authors.

9. Modulation Transfer Function. We describe here a simple model for the Modulation Transfer Function of a general satellite. More details can be found in [42] and [2] where specific examples of MTF for different acquisition systems are shown.

Recall that the MTF, that we denote by $\hat{h}$, is the Fourier transform of the impulse response of the system. Let $(\xi, \eta) \in [-1/2, 1/2]$ denote the coordinates in the frequency domain. There are different parts in the acquisition system that contribute to the global transfer function:

- **Sensors:** every sensor has a sensitive region where all the photons that arrive are integrated. This region can be approximated by a unit square $[-c/2, c/2]^2$ where c is the distance between consecutive sensors. Its impulse response is then the convolution of two pulses, one in each spatial direction. The corresponding transfer function also includes the effect of the conductivity (diffusion of information) between neighbouring sensors, which is modeled by an exponential decay factor, thus:

$$\hat{h}_S(\xi, \eta) = \operatorname{sinc}(\xi c) \operatorname{sinc}(\eta c) e^{-2\pi\beta_1 c|\xi|} e^{-2\pi\beta_2 c|\eta|},$$

  where $\operatorname{sinc}(\xi) = \sin(\pi\xi)/(\pi\xi)$.
- **Optical system:** it is considered as an isotropic low-pass filter

$$\hat{h}_O(\xi, \eta) = e^{-2\pi\alpha c\sqrt{\xi^2+\eta^2}}.$$

- **Motion:** each sensor counts the number of photons that arrive to its sensitive region during a certain time of acquisition. During the sampling time the system moves a distance τ and so does the sensor; this produces a motion blur effect in the motion direction (d_1, d_2):

$$\hat{h}_M(\xi, \eta) = \operatorname{sinc}(\langle(\xi, \eta), (d_1, d_2)\rangle \tau).$$

Finally, the global MTF is the product of each of these intermediate transfer functions modeling the different aspects of the satellite:

$$\hat{h}(\xi, \eta) = \hat{h}_S \hat{h}_O \hat{h}_M.$$

REFERENCES

[1] A. Aldroubi and K. Gröchenig. Non-uniform sampling and reconstructionin shift-invariant spaces. *SIAM Rev.*, 2001.

[2] A. Almansa. *Echantillonnage, Interpolation et Détéction. Applications en Imagerie Satellitaire.* PhD thesis, Ecole Normale Supérieure de Cachan, 94235 Cachan cedex, France, December 2002.

[3] A. Almansa, S. Durand, and B. Rouge. Measuring and improving image resolution by adaptation of the reciprocal cell. *Journal of Mathematical Imaging and Vision*, 21(3):235–279, November 2004.

[4] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems.* Oxford Mathematical Monographs, 2000.

[5] F. Andreu, C. Ballester, Caselles V., and J.M. Mazón. Minimizing total variation flow. *Diff. and Int. Equations*, 14:321–360, 2001.

[6] F. Andreu-Vaillo, V. Caselles, and J.M. Mazón. *Parabolic Quasilinear Equations Minimizing Linear Growth Functionals.* Birkhauser Verlag, 2004.

[7] J. F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *Journal of Mathematical Imaging and Vision*, 22(1), 2005.
[8] J.F. Aujol and G. Gilboa. Implementation and parameter selection for bv-hilbert space regularizations. *UCLA CAM-Reports 04-66*, 2004.
[9] P. Azmi and F. Marvasti. Comparison between several iterative methods of recovering signals from nonuniformly spaced samples. *International Journal of Sampling Theory in Signal and Image Processing*, 1(3):207–224, 2002.
[10] G. Bellettini, V. Caselles, and M. Novaga. The total variation flow in r^n. *J. Diff. Equations*, 184:475–525, 2002.
[11] A. Bermúdez and C. Moreno. Duality methods for solving variational inequalities. *Comput. Math. Appl.*, 7(1):43–58, 1981.
[12] M. Bertalmío, V. Caselles, B. Rougé, and A. Solé. TV based image restoration with local constraints. *Journal of Scientific Computing*, 19(1-3):95–122, 2003.
[13] G. Beylkin. On the fast Fourier transform of functions with singularities. *ACHA*, 2:363–381, 1995.
[14] G. Beylkin. On applications of unequally spaced fast Fourier transform. In *Mathematical Geophysics Summer School*, Stanford, August 1998.
[15] H. Brezis. *Operateurs Maximaux Monotones*. North Holland, Amsterdam, 1973.
[16] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004.
[17] A. Chambolle and P.L. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76:167–188, 1997.
[18] R.H. Chan, T.F. Chan, and H.M. Zhou. Continuation method for total variation denoising problems. Technical Report 95-18, University of California, Los Angeles, 1995.
[19] T.F. Chan, G.H. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM Journal on Scientific Computing*, 20(6):1964–1977, 1999.
[20] T.F. Chan and P. Mulet. On the convergence of the Lagged Diffusivity Fixed Point Method in total variation image restoration. *SIAM Journal on Numerical Analysis*, 36(2):354–367, 1999.
[21] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311, 1997.
[22] P.G. Ciarlet. *Introduction à l'analyse numérique matricielle et à l'optimisation*. Dunod, Paris, 1998.
[23] R. Dautray and J.L. Lions. *Analyse Mathématique et Calcul Numérique pour les sciences et les techniques, vol. 5*. Masson, 1988.
[24] I. Ekeland and R. Temam. *Convex Analysis and variational Problems*. Studies in Advanced Math., CRC Press, 1992.
[25] L.C. Evans and R.F. Gariepy. *Measure Theory and Fine Properties of Functions*. 2000.
[26] P. Faurre. *Analyse numérique. Notes d'optimisation*. École Polytechnique. Ed. Ellipses, 1988.
[27] H. G. Feichtinger, K. Gröchenig, and T. Strohmer. Efficient numerical methods in non-uniform sampling theory. *Numerische Mathematik*, 69:423–440, 1995.
[28] H.G. Feichtinger, C. Cenker, and M. Herrmann. Iterative algorithms in irregular sampling: A first comparison of methods. In *Conf. ICCCP'91*, pages 483–489, March 1991.
[29] C. Gasquet and P. Witomski. *Analyse de Fourier et Applications*. Ed. Masson, 1990.
[30] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, third edition edition, 1996. isbn 0801854148.
[31] K. Gröchenig. Reconstruction algorithms in irregular sampling. *Math. Comp.*, 59(181–1924), 1992.
[32] K. Gröchenig and T. Strohmer. Numerical and theoretical aspects of non-uniform sampling of band-limited images. In F. Marvasti, editor, *Theory and Practice of Nonuniform Sampling*. Kluwer/Plenum, 2000.
[33] T. Chan J.F. Aujol, G. Gilboa and S. Osher. Structure-texture decomposition by a tv-gabor model. *To appear in IJCV*.
[34] M. I. Kadec. The exact value of the Payley-Wiener constant. *Soviet Math. Doklady*, 5(559–561), 1964.
[35] G. Koepfler, C. Lopez, and J. M. Morel. A multiscale algorithm for image segmentation by variational method. *SIAM Journal on Numerical Analysis*, 31(1):282–299, 1994.
[36] S. Osher L. Rudin and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
[37] F. Malgouyres. Total Variation based oversampling of noisy images. In *Proc. on Scale-Space*

2001, Lecture Notes in Computer Science 2106, pages 111–122, 2001.

[38] F. Malgouyres and F. Guichard. Edge direction preserving image zooming: a mathematical and numerical analysis. *Journal on Numerical Analysis*, 39(1):1–37, 2001.

[39] D. Mumford and J. J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–684, 1989.

[40] D. Potts. NFFT. http://www.math.mu-luebeck.de/potts/nfft/, 2001.

[41] M. Rauth. *Gridding geophysical potential fields from noisy scattered data*. PhD thesis, University of Vienna, May 1998.

[42] B. Rougé. Théorie de l'échantillonage et sattelites d'observation de la terre. Analyse de Fourier et traitement d'images. Journées X-UPS 1998. Centre de Mathématiques – Ecole Polytechnique. France., 1998.

[43] L. Vese. A study in the bv space of a denoising-deblurring variational problem. *Applied Mathematics and Optimization*, 44:131–161, 2001.

[44] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. *SIAM Journal on Scientific Computing*, 17(1):227–238, 1996.

[45] R. M. Young. *Introduction to Nonharmonic Fourier Series*. Academic Press, 1980.

[46] W. P. Ziemer. *Weakly Differentiable Functions*. GTM 120, Springer Verlag, 1989.