Irregular to regular sampling, denoising and deconvolution

Gabriele Facciolo, Andrés Almansa, Jean-François Aujol, Vicent Caselles

To cite this version:

Gabriele Facciolo, Andrés Almansa, Jean-François Aujol, Vicent Caselles. Irregular to regular sampling, denoising and deconvolution. Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal, 2009, 7 (4), pp.1574-1608. ⟨10.1137/080719443⟩. ⟨hal-00988783⟩

HAL Id: hal-00988783
https://hal.science/hal-00988783v1

Submitted on 2 Jul 2018

IRREGULAR TO REGULAR SAMPLING, DENOISING AND DECONVOLUTION

GABRIELE FACCIOLO [*], ANDRÉS ALMANSA [†], JEAN-FRANÇOIS AUJOL [‡], AND VICENT CASELLES [§]

Abstract. We propose a restoration algorithm for band limited images that considers irregular (perturbed) sampling, denoising, and deconvolution. We explore the application of a family of regularizers that allow to control the spectral behavior of the solution combined with the irregular to regular sampling algorithms proposed by H.G. Feichtinger, K. Gröchenig, M. Rauth and T. Strohmer. Moreover, the constraints given by the image acquisition model are incorporated as a set of local constraints. And the analysis of such constraints leads to an early stopping rule meant to improve the speed of the algorithm. Finally we present experiments focused on the restoration of satellite images, where the micro-vibrations are responsible of the type of distortions we are considering here. We will compare results of the proposed method with previous methods and show an extension to zoom.

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 modeled by the following image formation model

$$z(\xi_k) = (h * u)(\xi_k) + n_{\xi_k}, \qquad \xi_k \in \Xi, \tag{1.1}$$

where $\Xi = \{\xi_k\}_{k=1}^{N^2} \subseteq \mathbb{R}^2$ 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 $\hat{h}$ has most of its energy concentrated in the spectral support of u, z is the observed sampled image which is represented as a function $z : \Xi \to \mathbb{R}$, and n_{ξ_k} is, as usual, a white Gaussian noise with zero mean and standard deviation σ.

Reconstructing a signal $u : \mathbb{R}^2 \to \mathbb{R}$ over an infinite support from a finite set of samples $z(\xi_k)$ is not possible without imposing restrictions. As in most works, in order to simplify this problem, we shall assume that the functions h and u are periodic of period N in each direction. Let us denote by Ω_N the interval $[0, N[^2$ and 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 [28] (which may be extended to a continuous periodic function in $\mathbb{R}^2$). Then the samples $(h * u)(\xi_k)$, $\xi_k \in \Xi$, are well defined.

We shall concentrate on the particular case of perturbed sampling and we shall assume that Ξ is a set of N^2 samples, which take the particular form

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

where $\varepsilon : \mathbb{R}^2 \to \mathbb{R}^2$ is a “smooth and small” perturbation function in the sense that $\operatorname{supp} \hat{\varepsilon} \subseteq [-\frac{N}{T_\varepsilon}, \frac{N}{T_\varepsilon}]^2$ for some period $T_\varepsilon > 2$ corresponding to the maximum vibration

---

[*] Dept. Tecnologia, Universitat Pompeu Fabra, Passeig de Circumvalació, 8, 08003 Barcelona, Spain, gabriele.facciolo@upf.edu

[†] TELECOM ParisTech, CNRS UMR 5141 (LTCI), 46 rue Barrault, 75634 Paris cedex 13, France; and ENS Cachan, CNRS UMR 8536 (CMLA), UniverSud, France, Andres.Almansa@TELECOM-ParisTech.fr

[‡] CMLA, ENS Cachan, CNRS, UniverSud, France, Jean-Francois.Aujol@cmla.ens-cachan.fr

[§] Dept. Tecnologia, Universitat Pompeu Fabra, Passeig de Circumvalació, 8, 08003 Barcelona, Spain, vicent.caselles@upf.edu

frequency and the mean amplitude of the perturbation ($[\int_{\Omega_N} |\varepsilon(x)|^2 dx / \int_{\Omega_N} dx]^{\frac{1}{2}}$) 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).

Even knowing the exact sampling geometry Ξ, the blurring kernel h and the statistics of the noise n, the problem of recovering u from z is very ill-posed in a variety of situations that arise in real-world image acquisition systems. The main difficulty comes from the fact that perfect reconstruction from irregular samples can only be ensured either with a sampling rate, which is much larger than in the regular case (relative to the bandwidth of h), or similar to the regular critical sampling rate, but under very constrained sampling geometries that rarely appear in practice. (See [3, Chapter 2] for a discussion on this problem). In any case, even if perfect reconstruction is sometimes possible in theory, with a relatively small and realistic sampling rate, it is still very ill posed, which limits its practical interest whenever noisy measurements are involved. A common and unavoidable strategy to solve this ill-conditioning is regularization. And the typical constrained formulation of a regularization method [47] consists in choosing between all possible solutions of (1.1) the one that minimizes the functional $J(u)$. Modeling the acquisition system (1.1) through constrained minimization eliminates the regularization parameter in classical variational or bayesian formulations, and due to the statistical nature of the noise information, the constraint is expressed as an upper estimate of the noise variance σ^2. The constrained restoration problem becomes

$$\begin{aligned} &\min_u \quad J(u), \\ &\text{subject to} \sum_{\xi_k \in \Xi} |(h * u)(\xi_k) - z(\xi_k)|^2 \leq N^2 \sigma^2, \end{aligned}$$

where the regularizer $J(u)$ embodies our a-priori knowledge of the image, specifying its smoothness properties. The use of the Dirichlet integral $\int_{\Omega_N} |Du|^2\, dx$ is not satisfactory, mainly due to the unability of the previous functional to resolve discontinuities (edges) and oscillatory textured patterns, the information corresponding to high frequencies of z being attenuated by it. Indeed, functions in $W^{1,2}(\Omega)$ (i.e., functions $u \in L^2(\Omega)$ such that $Du \in L^2(\Omega)$) cannot have discontinuities along rectifiable curves. These observations motivated the introduction of Total Variation ($TV(u) = \int_{\Omega_N} |Du|$) in image restoration problems by L. Rudin, S. Osher and E. Fatemi in their work [44]. The a priori hypothesis is that functions of bounded variation (the BV model) ([6, 25]) are a reasonable functional model for many problems in image processing, in particular, for restoration problems ([44]). Typically, functions of bounded variation have discontinuities along rectifiable curves, being continuous in the measure theoretic sense away from discontinuities. The 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] for the definition of functions of bounded variation and their basic properties.

We shall explore in this paper a family of regularizers that takes into account the spectral decay of the Fourier coefficients in the class of images we are looking for. In the case of satellite images, this spectral behavior can be estimated by statistical measures of the decay of Fourier coefficients. The general class of regularizers we consider is

$$J_A(u) = \int_{\Omega_N} |A(D)u| \tag{1.3}$$

where $A(D)u$ is defined by the coefficients of its Fourier series $\mathcal{F}(A(D)u)(\omega) = A(i\omega)\hat{u}(\omega)$, $\omega \in \mathbb{Z}^2$. Note that $J_A(u) < \infty$ imposes a frequency penalization according to the profile $A(i\omega)$. In practice we choose $A(i\omega)$ so that $|A(i\omega)| \sim |\frac{2\pi}{N}\omega|^p$ for large $|\omega|$, $1 \leq p \leq 2$. This is in consonance with the approach of Gröchenig and Strohmer [32] that proposes to incorporate an a-priori decay in the restoration process (see Section 2).

In [4] (see also [5, 13, 43]), the authors proposed a restoration model using a local estimate of the noise variance. The local formulation contributes to reducing the unnaturally-looking aspect of images obtained from global Total Variation based minimization, thus improving the recovery of textures. Following the mentioned proposal we replace the constraint

$$\sum_{\xi_k \in \Xi} |(h * u)(\xi_k) - z(\xi_k)|^2 \leq N^2\sigma^2,$$

by

$$G * |\Delta_\Xi(h * u) - z|^2(\xi_k) \leq \sigma^2, \qquad \forall \xi_k \in \Xi, \tag{1.4}$$

where the sampling operator $\Delta_\Xi : C(\mathbb{R}^2) \to \ell^2(\Xi)$ is given by $\Delta_\Xi(v) = \{v(\xi_k)\}_{k=1}^{N^2}$ and G is a discrete convolution kernel such that $G(\xi) > 0$ for all $\xi \in \Xi$ and $\sum_k G(\xi_k) = 1$.

Combining the two ideas described above, the use of a regularizer that takes into account the spectral decay of images in a certain class (1.3), and the incorporation of the image acquisition model as a set of local constraints (1.4), we propose the following constrained variational model for restoring u

$$\begin{aligned} &\min_u \quad \int_\Omega |A(D)u|, \\ &\text{subject to} \ \left[G * |\Delta_\Xi (h * u) - z|^2\right](\xi_k) \leq \sigma^2 \quad \forall \xi_k \in \Xi. \end{aligned} \tag{1.5}$$

The constrained formulation (1.5) can be solved using the unconstrained formulation

$$\min_u \max_{(\lambda_k) \geq 0} \int_\Omega |A(D)u| + \frac{1}{2} \sum_{\xi_k \in \Xi} \lambda_k \left\{ \left[G * |\Delta_\Xi (h * u) - z|^2\right](\xi_k) - \sigma^2 \right\} \tag{1.6}$$

where $\lambda_k \geq 0$ is a Lagrange multiplier that has to be chosen so that the constraints (1.4) are satisfied. Let us say explicitly that both the blurring kernel h and the sampling grid Ξ (alternatively the grid perturbation function ε) are assumed to be known exactly, and that the only thing known about the noise n_{ξ_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 [35] for a given acquisition device and we shall not address this question here.

The case of recovering an irregularly sampled image on a regular sampling grid was already considered by the second author in [3], but the blurring kernel h 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 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 $\hat{u}$ in the high frequency band $[-1/2, 1/2]^2 \backslash R$ serious theoretical and numerical problems appeared and, actually, restoration errors were most important there. Subsequently, the restoration problem (1.5) was studied in [5] when $J(u)$ is the total variation and the image acquisition model was incorporated as a set of local constraints on a partition of the image obtained as a result of a segmentation. The use of local constraints (1.4) was advocated in [4] and we also adopt this technique here.

Let us finally mention that many numerical algorithms have been proposed to minimize total variation (or similar models) subject to a global constraint as in (1.5) [44, 29, 14, 49, 17, 18, 19, 20, 24, 16, 45]. Imposing local constraints in a partition of the image was proposed in [43] and further developed in [13, 5, 4]. In [38] the authors combined total variation minimization with a set of constraints of type $|\langle h * u - z, \psi \rangle| \leq \tau$ where ψ varies along an orthonormal basis of wavelets (or a family of them) and $\tau > 0$. The aim was also to construct an algorithm which preserves textures and has good denoising properties. As we will do here, these constraints were incorporated using Uzawa's algorithm. In [30], the authors proposed to minimize total variation subject to a family of local constraints which control the local variance of the oscillatory part of the signal. The constraints are introduced via Lagrange multipliers with an approach similar to the one used in [44]. This amounts to adding a spatially varying fidelity term that locally controls the extent of denoising over image regions depending on their content. Besides the fact that we use Uzawa's algorithm and we try to address the problem of deconvolution and denoising of irregularly sampled images, the work [30] is quite similar to our approach. Finally, in [45], the authors proposed a non-convex data attachment term with a larger weight (depending on the inverse of the modulus of the gradient) in flat areas than in textured ones. This goes in the same sense as our model, that is, to be able to denoise flat regions of the image while keeping the oscillations in textured areas.

Let us summarize the contributions of the paper. Our goal is to propose an algorithm for image denoising and deconvolution. The specificity of the considered problem comes from the fact that the images we deal with are irregularly sampled. On the one hand, we carry out a thorough and accurate modelization of the problem based on local constraints (1.4), and propose a family of frequency adaptive regularization functionals (1.3) that incorporates a priori image smoothness model. On the other hand, we prove the convergence of our algorithm. Lastly, the study of the local constraint model combined with standard error propagation allows to derive useful stopping criteria for the algorithm.

Let us finally explain the plan of the paper. In Section 2 we introduce the problem of irregular to regular sampling and we discuss the ACT algorithm of Gröchenig and Strohmer [32]. In Section 3 we introduce our frequency adaptive variational restoration model with local constraints in the continuous setting. The corresponding discrete model is discussed in Section 4 along with the adaptation of a computational improvement introduced by L. Moisan in [39]. In Section 5 we study the existence, uniqueness and numerical approximation to the model introduced in the previous Section. This study is completed in Section 6 where we describe a Quasi-Newton algorithm for the solution of the Euler Lagrange equation corresponding to the energy in (1.6) for fixed values of the Lagrange multipliers (λ_k). In Section 7 we propose a practical stopping condition for the restoration algorithm for the local constraint model. In Section 8 we display experiments concerning restoration and zooming of irregularly sampled images. Section 9 summarizes the main conclusions of this work.

1.1. Preliminaries and notations.

Let us introduce some notations.

For any function $u \in L^2(\Omega_N)$ (assuming periodicity of period N in each direction) we denote its Fourier coefficients as

$$\hat{u}(p,q) = \frac{1}{N^2}\int_{\Omega_N} u(x,y)e^{-2\pi i\frac{(px+qy)}{N}}\,dxdy \qquad \text{for } (p,q) \in \mathbb{Z}^2.$$

As in [40], our plan is to compute a band limited approximation to the solution of the restoration problem. To do so, assume for simplicity that M is an even number and define

$$\mathcal{B}_M := \{u \in L^2(\Omega_N) : \hat{u}\,\text{is supported in } I_M\} \quad \text{where} \quad I_M := \{-\frac{M}{2}+1,\ldots,\frac{M}{2}\}^2.$$

We notice that $\mathcal{B}_M$ is a finite dimensional vector space of dimension M^2 which can be identified with $\mathbb{R}^{M^2}$ by mapping $u \in \mathcal{B}_M$ to the vector $\vec{u} = \left(u(r\frac{N}{M}, l\frac{N}{M})\right)_{r,l=0}^{M-1}$. Moreover, if $u \in \mathcal{B}_M$ we may write

$$u(x,y) = \sum_{-\frac{M}{2}<p,q\leq\frac{M}{2}} \hat{u}(p,q)e^{2\pi i\frac{(px+qy)}{N}}.$$

where

$$\hat{u}(p,q) = \frac{1}{M^2}\sum_{0\leq r,l<M} u\left(r\frac{N}{M}, l\frac{N}{M}\right)e^{-2\pi i\frac{(pr+ql)}{M}}, \qquad -\frac{M}{2} < p,q \leq \frac{M}{2}.$$

Then the values $u\left(r\frac{N}{M}, l\frac{N}{M}\right)$, $0 \leq r,l < M$, can be recovered as the discrete inverse Fourier transform of $\hat{u}(p,q)$. Hence $u \in \mathcal{B}_M$ can also be identified with the vector of Fourier coefficients $\hat{u} \in \mathbb{C}^{M^2}$. Note that we shall mainly study here the critical sampling case $M = N$, however we will keep two different symbols M for the bandwidth and N for the domain size in order to simplify the exposition of certain parts of the algorithm.

We intend to solve the restoration problem in the class of band-limited functions $\mathcal{B}_M$. Later on we will comment on this choice. We will also use the operator notation for the Fourier transform that applied to the function u returns the vector of its Fourier coefficients: $\hat{u} = \mathcal{F}u$, conversely $\mathcal{F}^*\hat{u} = u$ denotes the inverse transform, then $\mathcal{F}^*\mathcal{F} = Id$.

2. Irregular to regular sampling.

Opposed to digital photographs, satellite images are generally not acquired by a squared array of sensors but by a sweeping bar of sensors known as TDI (Time Delay Integrator) [46]. This acquisition geometry called *push-broom* is widely applied in aerospace imaging applications and, nowadays, it provides the highest resolution in earth imaging applications. As a consequence of this progressive acquisition mode, the micro-vibrations of the satellite together with irregularities in sensors position result in perturbed sampling sets. In most cases, the knowledge of certain vibration modes and the analysis of acquired images help to estimate, very accurately, the perturbations in the sampling grid, which can be modeled [3] by

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

for some $q \geq 1$, where $a_k(x)$ are smooth modulation functions and the vibration frequencies ω_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 ω_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 $|\nabla\varepsilon(x)|$ 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 $\omega \in \mathbb{Z}^2$ we define

$$\begin{aligned} \widehat{\varepsilon_i}(\omega) &\sim N(0, \tilde{\sigma}^2) \qquad && \text{if } |\omega| \leq N/T_\epsilon \\ &= 0 && \text{otherwise} \end{aligned} \tag{2.2}$$

where $\tilde{\sigma}$ is chosen in such a way that the standard deviation of $\varepsilon_i(x)$ is A for $i \in \{1, 2\}$. This gives $\tilde{\sigma} = \frac{AT_\epsilon}{2}$ (we have taken the Fourier transform as an isometry). Thus the behavior of the perturbation is characterized by the two parameters "amplitude" A and maximal vibration frequency N/T_ε (or "minimal vibration period" T_ε). The precise values of A and T_ε used in our experiments will be specified in the experiments Section.

There are many works in the literature dealing with the irregular to regular sampling problem. Let us mention that, according to Kadec's theorem [36], we have a perfect recovery of the signal if we consider a perturbed sampling with small perturbations $|\varepsilon(x)| \leq 0.11$. Recall also that Beurling-Landau's theorem [37], ensures perfect reconstruction of a function from its samples for arbitrary *stable sampling sets* [37], but it requires the (lower) sampling density to be larger than 1. These conditions are very restrictive and do not hold true for most of the image restoration applications. For a comparison between several iterative methods we refer the reader to [12, 27, 3].

2.1. The ACT algorithm.

One of the best performing reconstruction methods available for irregular to regular sampling is the ACT algorithm (for Adaptive-weights Conjugate-gradient on Toeplitz-matrix) introduced by Gröchenig *et. al.* in [32]. This method represents a discrete image u as a trigonometric polynomial of order $M/2$ in each variable (for simplicity of notation we shall assume that M is an even number) so that the interpolation at the sampling points $\Xi = \{\xi_k\}_{k=1}^{N^2} \subseteq \mathbb{R}^2$ becomes

$$u(\xi_k) = \sum_{t \in \{-\frac{M}{2}+1, \ldots, \frac{M}{2}\}^2} a_t e^{\frac{2\pi i}{M}\langle t, \xi_k\rangle}, \quad k \in \{1, \ldots, N^2\}. \tag{2.3}$$

Thus, if z represents the irregularly sampled data we may write [32]

$$z = Sa, \quad \text{where} \quad S = ((s_{kt})),\ s_{kt} = e^{\frac{2\pi i}{M}\langle t, \xi_k\rangle}, \tag{2.4}$$

i.e. S is the Vandermonde matrix associated to the trigonometric polynomial in (2.3). Note that S maps $a \in \ell^2(\{-\frac{M}{2}+1, \ldots, \frac{M}{2}\}^2)$ to $z = \{z(\xi_k)\} \in \ell^2(\Xi)$ as given in (2.3). The bandwidth M of the trigonometric polynomial is chosen to be $M \leq N$, so the system (2.4) is expected to be determined or over-determined.

Following [32], the ACT algorithm recovers the coefficients a of the trigonometric polynomial by solving the least squares problem

$$\arg\min_a \|\sqrt{W}(Sa - z)\|^2, \tag{2.5}$$

where the matrix $W = \operatorname{diag}(\{w_k\}_{k=1..N^2})$ assigns weights that are inversely proportional to the sampling density at ξ_k:

$$w_k = \operatorname{area}(V_k) \qquad \text{where} \qquad V_k := (\{x : |x - \xi_k| < |x - \xi_j|, \forall j \neq k\})\,. \tag{2.6}$$

If we interpret the discrete data $z(\xi_k)$ as a piecewise constant function $\sum_k z(\xi_k)\chi_{V_k}$, then the weights w_k guarantee the isometry between the irregular sampling on the grid Ξ and its function representation, thus compensating the local variations in the sampling density. Moreover, by using the weights W, Gröchenig and Strohmer provide an explicit estimate for the rate of convergence of the ACT algorithm [32].

The system of normal equations associated to (2.5) is

$$S^* W S a = S^* W z\,, \tag{2.7}$$

where S^* denotes the adjoint matrix of S (a notation that will be used throughout this paper). Observe that the $M^2 \times M^2$ matrix S^*WS has a Toeplitz structure [32], and thus, S^*WSa is efficiently computed in $\mathcal{O}(M^2 \log_2(M^2))$ time using Fourier methods. Moreover the entries of $T := S^*WS$ and $b := S^*Wz$ can be approximated using the NFFT [41] in $\mathcal{O}(M^2 \log_2(M^2))$ time each [32]. Finally, (2.7) is solved using a *conjugate gradient* (CG) method. The following algorithm summarizes the method.

```
Algorithm I: ACT algorithm for a fixed bandwidth $M$
Requires: $N^2$ irregular samples in vector $z$.
Ensures: $M^2$ regular samples in vector $u$.
  1. Compute $T = S^*WS$ and $b = S^*Wz$ using the NFFT.
  2. Solve $Ta = b$ using conjugate gradients.
  3. Compute the regular samples $u(i\frac{N}{M}, j\frac{N}{M})$ for $(i,j) \in \{0, \ldots, M-1\}^2$ by applying the inverse FFT to $a$.
```

Let us note that in the more realistic cases where T is not invertible or ill-conditioned, the CG solver acts as a regularizer and chooses the minimum norm solution a among those that satisfy (2.7). This is a constrained variational formulation that can be written using a Lagrange multiplier $\lambda > 0$ as the unconstrained minimization problem

$$\min_a \|a\|^2 + \lambda \|Ta - b\|^2.$$

This formulation also applies to the following two variants of the ACT algorithm [32] that incorporate an a-priori spectral decay rate $|\hat{u}(\omega)| \leq L\phi(\omega)$, for some $L > 0$, of the image class to be restored (when available). For satellite images this estimation has been performed by Almansa in [3] and it is given by $\phi(\omega) = (1 + |\frac{2\pi}{N}\omega|)^{-p}$ for some value of p near 1.6.

- The first proposed variant solves $CTa = Cb$ (instead of (2.7)) using the CG algorithm, where $C = diag(\{\phi(\omega)\}_{\omega\in\{-\frac{M}{2}+1,\dots,\frac{M}{2}\}^2})$. Notice that we can write the problem as
$$(\mathrm{ACT}_D) \qquad \min_a \|a\|^2 + \lambda\|C(Ta-b)\|^2, \tag{2.8}$$
for some $\lambda > 0$. Since the weighting matrix C is applied to the residuals $Ta - b$, then it affects the relative cost of the errors by shifting the cost from high frequencies towards the lower ones. This shift is reflected in CG search directions, altering the intermediate solutions so that they will fit the low frequency before the higher ones.
- The second ACT variant solves $Td = b$ where $a = Cd$, using the CG algorithm. Re-writing it as an optimization problem, observe that the weights appear now in the regularity term
$$(\mathrm{ACT}_R) \qquad \min_a \|C^{-1}a\|^2 + \lambda\|Ta-b\|^2. \tag{2.9}$$
The spectral weights C^{-1} are now penalizing the apparition of higher frequencies in the solution a and not in the residual.

In either case, if T is invertible and the CG algorithm converged, then the solution of both variants coincides with the solution of (2.7). But the CG iteration is truncated before its convergence mainly due to the ill-conditioning of the operator T. So, the solutions obtained by the above methods differ because of the different search directions. As it can be observed experimentally, incorporating the spectral decay indeed reduces the restoration errors, specially when applied to the regularity term in (2.9) (see Table 8.1 in Section 8). In that case, it amounts to finding a solution in a class of functions with a particular spectral decay.

REMARK 1. *(Global constraint & stopping condition)* In [32] the authors also proposed to extend the ACT algorithm in order to consider the presence of Gaussian noise n (with standard deviation σ) in the image formation model: $Sa + n = z$. This extension is implemented as a stopping condition for the CG algorithm (Step 2 of Algorithm I)

$$\text{stop CG if} \quad \|Sa - z\|^2 \leq \tau N^2\sigma^2 \quad \text{with} \quad \tau \simeq 1$$

that is designed to avoid the over-fitting of the solution and hence the amplification of the noise. Thus, the ACT with residual-based stop conditions can be seen as a numerical approximation to the minimization of $\|Ca\|^2$ subject to the constraints above.

REMARK 2. *(Fourier vs. spline models)* Notice that the ACT Algorithm is based on the underlying assumption that the data can be represented by a trigonometric polynomial. Other interpolation models like the B-Spline have been used in the literature [8]. In this work we will restrict ourselves to trigonometric polynomials mainly because convolutions are more easily modeled in this setting, but we intend to explore the use of B-splines in the future.

3. A frequency adaptive restoration model with local constraints.

In this Section we sketch main elements of the restoration model adopted in this paper. To simplify the presentation the model is first defined in a continuous setting, and the discretization details are deferred to Section 4. The proposed model is inspired from the ideas presented in the previous Section on the ACT, as well as on a recently proposed ACT+TV extension (next Subsection), and the use of local constraints (Subsection 3.3) as in [4].

3.1. The ACT + TV extension. In [5] the authors proposed to combine the ACT algorithm written as (2.5) with total variation regularization, i.e.,

$$\begin{aligned} &\min_u \int_{\Omega_N} |Du|, \\ &\text{subject to } \|\sqrt{W}(S\hat{u} - z)\|^2 \leq N^2\sigma^2. \end{aligned} \tag{3.1}$$

For convenience, let us refer to this model as ACT+TV. As reported in [5] (see also Table 8.1 in Section 8) we observe an improvement of ACT+TV with respect to the original ACT algorithm in terms of MSE error, this improvement is mainly attributed to the edge preserving ability of the TV regularizer.

3.2. Frequency Adaptive Regularization (FAR). Inspired by (2.9) we propose to integrate the spectral weight priors given by the matrix C into the ACT+TV formulation.

From the ACT+TV formulation we keep: *(i)* the use of the spatial L^1-norm of a (pseudo)differential operator as a regularizer (instead of the L^2-norm $\|C^{-1}a\|$ in (2.9)), because of the better edge-preserving capabilities of the L^1-norm; and *(ii)* the formulation as an optimization problem under constraints, which enables us to choose automatically the regularization parameter λ in (2.9).

From the ACT$_R$ we keep the idea of using a frequency adapted regularizer. For that we define the function $\omega \in \mathbb{Z}^2 \to A(i\omega) \in \mathbb{C} \times \mathbb{C}$. We assume that

$$A(0) = 0, \qquad A(i\omega) \neq 0 \quad \forall \omega \neq 0, \quad \text{and} \quad |A(i\omega)| \leq L\left(1 + \left|\frac{2\pi}{N}\omega\right|^{\kappa}\right), \ \forall \omega \in \mathbb{Z}^2, \tag{3.2}$$

for some $L > 0$, $\kappa \geq 0$.

If $u \in C^\infty(\Omega_N)$ can be extended as a smooth and periodic function to $\mathbb{R}^2$, we define $A(D)u$ by its Fourier coefficients

$$\widehat{A(D)u}(\omega) = A(i\omega)\hat{u}(\omega) \qquad \omega \in \mathbb{Z}^2.$$

We define the regularizer functional

$$J_A(u) = \int_{\Omega_N} |A(D)u| = \int_{\Omega_N} |(A(i\omega)\hat{u}(\omega))^\vee|.$$

Notice that $J_A(u)$ can be defined for any $u \in L^2(\Omega_N)$ such that $A(D)u$ is a Radon measure. In practice this will not be a problem since we only consider band limited functions.

The total variation $J(u) = \int_{\Omega_N} |Du|$ corresponds to the choice $A(i\omega) = \frac{2\pi}{N} i\omega$. Recall that functions with finite total variation are a good model for image restoration since they permit to recover the discontinuities of the image. But, in practice, digital images may exhibit a stronger decay in its Fourier coefficients than $|\frac{2\pi}{N}\omega|^{-1}$ and other functional models can be better suited. The operator $A(D)$ permits us to penalize the frequencies according to the profile $A(i\omega)$. Notice that

$$|A(i\omega)\hat{u}(\omega)| = |\widehat{A(D)u}(\omega)| \leq \int_{\Omega_N} |A(D)u| =: L$$

for all $\omega \in \mathbb{Z}^2$, hence

$$|\hat{u}(\omega)| \leq \frac{L}{|A(i\omega)|}.$$

If $|A(i\omega)| \sim |\frac{2\pi}{N}\omega|^{\kappa}$ for large $|\omega|$, then $|\hat{u}(\omega)|$ decreases as $|\frac{2\pi}{N}\omega|^{-\kappa}$. In this way we can favor a specific decay rate of the Fourier coefficients of u as a prior.

Remark 3. Notice that the frequency adaptive regularizer of ACT_R has the form $C^{-1}\hat{u} = \phi^{-1}(\omega)\hat{u}(\omega)$ where $\phi(\omega) = (1 + |2\pi\omega/N|)^{-k}$, $k \geq 0$. In practice, we consider functions $A(i\omega) = \frac{2\pi i\omega}{N}\psi(\omega)$ where $\psi(\omega) = (1 + |2\pi\omega/N|)^{p}$, $p \geq 0$ ($p = k-1$ gives the same decay). Note that we can write $A(D)u = D\Psi u$, where $\widehat{\Psi u} = \psi\hat{u}$, and $A(D)u$ is the derivative of a filtered version of u. Since $\int |A(D)u|^2 = \sum_{\omega} |A(i\omega)\hat{u}(\omega)|^2$, if we use $\int_{\Omega_N} |A(D)u|^2$ as regularizer, then the phase information disappears. This is not the case when we use the L^1-norm of $A(D)u$. Besides keeping the structure of a regularization operator, when using the L^1-norm, this phase is relevant in practice (we obtain bad results if we dismiss it).

3.3. Local constraints.

Following [4] we will interpret the acquisition model as a set of local constraints (1.4) and we will search for a solution of the restoration problem that minimizes the frequency adaptive regularizer $J_A(u)$ with a data-fitting term derived from the acquisition model and consisting in a set of local constraints (as in (1.5)). For convenience we will describe these constraints directly in the discrete model (see the next Section).

4. A discrete regularization model.

Let us stress here the fact that the regularization functional J_A defined in the previous Section is adapted to the restoration of functions with infinite resolution while its numerical approximation restricts the solution to be in a finite dimensional space. We are going to adopt here the following practical point of view. Since our data consists of a finite set of samples, we are going to reconstruct a sampled version of the image and therefore we work in a finite dimensional space. This reflects the fact that digital images have finite resolution. Usually, the restored image is modeled as a piecewise constant function (the values given on the set of pixels), but we consider here images as bandlimited functions with a finite number frequencies, since this is a reasonable model for restoring digital images. Moreover this model is adapted to compute convolutions and permits to include an a priori decay of the Fourier coefficients.

Based on the above considerations, we propose the following discrete regularization functional, If $u \in \mathcal{B}_M$, then we define

$$J_A^d(u) = \frac{1}{M^2} \sum_{0 \leq r,l < M} \left| A(D)u\left(\frac{rN}{M}, \frac{lN}{M}\right) \right|. \tag{4.1}$$

If $A(i\omega)$ satisfies (3.2), then $J_A(u)$ and J_A^d are seminorms in $\mathcal{B}_M$ and the only function $u \in \mathcal{B}_M$ such that $J_A(u) = 0$ (resp. such that $J_A^d(u) = 0$) is $u =$ constant. Thus $J_A(u)$ and J_A^d are norms in the finite dimensional quotient space $\mathcal{B}_M/\mathbb{R}$, hence they are equivalent. Notice that if $u \in \mathcal{B}_M$ and we define $J_A^{d,k}$ by replacing M by k in (4.1), then $J_A^{d,k}(u) \to J_A(u)$ as $k \to \infty$. Unless we intend to zoom and restore the images, we take $M = N$, where N^2 is the number of data.

Assume that the input data (measurements) consist of N^2 samples $\{z(\xi_k)\}_k$, and let $\bar{\sigma}, \beta > 0$, and $G \in \ell^{\infty}(\mathbb{Z}^2)$ be a discrete, positive, normalized convolution kernel such that $G(r,l) \geq 0$ and $\sum_{(r,l)\in\mathbb{Z}^2} G(r,l) = 1$. Then we propose to minimize the functional

$$\begin{aligned}
&\min_{u\in\mathcal{B}_N} \; J_{\beta,A}(u) := \sum_{0\le r,l<N} \sqrt{\beta^2 + |A(D)u\,(r,l)|^2} \quad, \\
&\text{subject to } \left[G * |\Delta_\Xi\,(h*u) - z|^2\right](\xi_k) \le \bar{\sigma}^2 \quad \forall \xi_k \in \Xi, \qquad (4.2)\\
&\text{and } \sum_{0\le r,l<N} u\,(r,l) = \sum_{\xi_k\in\Xi} w_k z(\xi_k) =: \overline{z}^w.
\end{aligned}$$

To avoid the non differentiability of J_A^d at 0, it is a common approximation to use the regularizer $J_{\beta,A}$ instead.

Therefore we will minimize $J_{\beta,A}(u)$ on $\mathcal{B}_N$ subject to the family of constraints in (4.2). Notice that we have incorporated the image acquisition model (1.1) as a set of local constraints. The convolution of G and $v \in \ell^\infty(\Xi)$ is defined in the usual way by imposing an arbitrary regular grid structure on Ξ, *i.e.* $(G*v)(\xi_k) = \sum_{l\in\mathbb{Z}^2} G(l-k)v(\xi_l)$. In the case of perturbed sampling this regular grid structure may be determined by the original unperturbed grid, otherwise it may be based on a nearest neighbor computation.

Notice that we have used the value $\bar{\sigma} > 0$ as an estimate of the standard deviation of the noise. We will make tests with $\overline{\sigma} = \sigma$ and also with values of $\overline{\sigma}$ different from σ. Also, the effective support of G must permit the statistical estimation of the variance of the noise. In Section 7 we will come back to the noise estimation issue and the choice of $\overline{\sigma}$.

In the rest of the paper, we assume that the blurring kernel h satisfies

$$h \in L^2(\Omega_N), \qquad \operatorname{supp}\hat{h} \subseteq [-\frac{M}{2}, \frac{M}{2}]^2, \qquad \text{and} \qquad \hat{h}(0,0) = 1. \tag{4.3}$$

If $u \in \mathcal{B}_M$, then we can compute $h*u$ using the Fourier transform $\widehat{h*u}\,(p,q) = \hat{h}\,(p,q)\,\hat{u}\,(p,q)$.

The last equality constraint in (4.2) fixes the global mean of u to be the mean of the samples z, weighted by the areas w_k (see (2.6)). This constraint is necessary to assure the uniqueness of the solution, since the data fitting is provided only by inequality constraints, and the solution may be undetermined up to a constant (in the kernel of the regularizer) in areas where the variance of z is smaller than σ^2 [13]. The details of the uniqueness proof are given in Section 5.

Now, our purpose is to prove that the constrained formulation of (4.2) can be solved using Uzawa's method once we guarantee that the assumptions of Uzawa's method [26] hold. But before that, we comment on a improved discretization for approximating $J_A(u)$.

4.1. An improved discretization formula. In this Section we follow the proposal made by Moisan in [39] to improve the discretization of the total variation formula. The basic observation is that, the computation of the Dirichlet integral in $[0,N)^2$ ($\int_{[0,N)^2} |\nabla u|^2$) cannot be done accurately unless we previously zoom the image u by a factor of two. The same argument applies to the case of $\int_{\Omega_N} |A(D)u|^2$, but not for $\int_{\Omega_N} \sqrt{\beta^2 + |A(D)u|^2}$ with $\beta \ge 0$. Indeed, in the last case, an exact computation would involve an infinite number of samples. However, as Moisan has shown for the TV case [39], doubling the number of variables leads to a good approximation of the above integral, being a good compromise between precision and algorithmic efficiency.

Let us analyze the implications of Moisan's discretization for our restoration model. To do so, we need to introduce some notation. For each $M \in \mathbb{N}$, we denote by X_M the Euclidean space $\mathbb{R}^{M\times M}$. The Euclidean scalar product and the norm in X_M will be denoted by $\langle\cdot,\cdot\rangle_{X_M}$ and $\|\cdot\|_{X_M}$, respectively, but in absence of ambiguities we will omit the subindex. X_M represents the space of images $\mathcal{B}_M$ sampled in the regular grid $\{0,\ldots,M-1\}\times\{0,\ldots,M-1\}$ (or given by its Fourier coefficients $\hat{u}\in\ell^2(I_M)$. Let us introduce the operator

$$P : X_N \to X_{2N}\,, \quad P\left\{u(k,l)\right\} = \left\{\bar{u}\left(\tfrac{r}{2},\tfrac{l}{2}\right)\right\}_{r,l\in\{0,\ldots,2N-1\}}\,,$$

where $\{u(r,l)\}_{r,l\in\{0,\ldots,N-1\}}\in X_N$ and $\bar{u}$ is the function of $\mathcal{B}_N$ defined by the samples $\{u(r,l)\}_{r,l\in\{0,\ldots,N-1\}}$. Observe that $\widehat{Pu}(p,q) = \mathbf{1}_{I_N}(p,q)\hat{u}(p,q)$. So we may consider the operator $A(D)$ as acting on X_{2N} or as acting on X_N. Notice that if $u\in X_N$ we may write $A(D)Pu = PA(D)u$. From now on we will avoid (except in ambiguous cases) the use of subindexes to specify the function spaces of norms and scalar products, and the function space should be clear from the context.

Thus, our final restoration model is

$$\begin{aligned}&\min_{u\in X_N}\; K_{\beta,A}(u) := \sum_{0\le r,l<2N}\sqrt{\beta^2 + \left|A(D)Pu\,(r,l)\right|^2}\quad,\\ &\text{subject to }\; \left[G * \left|\Delta_\Xi\,(h*u) - z\right|^2\right](\xi_k) \le \bar{\sigma}^2 \quad \forall \xi_k\in\Xi\,,\\ &\text{and }\; \sum_{0\le r,l<N} u\,(r,l) = \sum_{\xi_k\in\Xi} w_k z(\xi_k) =: \overline{z}^w.\end{aligned} \tag{4.4}$$

5. The well-posedness of the model and its numerical solution.

Proposition 5.1. *Assume that* (4.3) *holds. Then there exists a unique minimum* $u\in X_N$ *of (4.4).*

Proof. Let u_m be a minimizing sequence of (4.4). Since $A(D)Pu_m$ is bounded in X_{2N}, and $\omega = 0$ is the only vanishing frequency for $A(i\omega)$ we deduce that $v_m := u_m - \widehat{u_m}(0,0)$ is bounded in X_N. Now, since $\widehat{u_m}(0,0)$ is constrained to be $\overline{z}^\omega$, we have that u_m is bounded in X_N. By extracting a subsequence, if necessary, we may assume that $u_n \to u$. It is immediate to see that u satisfies the constraints. Since $K_{\beta,A}$ is lower semicontinuous, we have that u is a minimum of (4.4).

Now, let u_1, u_2 be two minima of (4.4). If $A(D)Pu_1 \ne A(D)Pu_2$, letting $\overline{u} = \frac{u_1+u_2}{2}$, then the strict convexity of $K_{\beta,A}$ proves that $K_{\beta,A}(\overline{u}) < \inf_{u\in X_N} K_{\beta,A}$, a contradiction. Thus $A(D)Pu_1 = A(D)Pu_2$ and we have uniqueness modulo constants, i.e., $u_1 - u_2 = c$ for some $c\in\mathbb{R}$. Since $\hat{u_1}(0,0) = \hat{u_2}(0,0)$ we deduce that $c = 0$, and therefore, $u_1 = u_2$. ☐

Remark 4. Proposition 5.1 is also true if instead of assuming the average constraint in (4.4) we assume that $\inf_{c\in\mathbb{R}} G*(z-c)^2 > \bar{\sigma}^2$. This can be proved as in [13, 17]. But in that case, we should also use a different (gradient descent based) algorithm to minimize (4.4) as described in [13].

From now on, we assume that the constraints are qualified, that is there is $u\in X_N$ such that

$$\hat{u}(0,0) = \overline{z}^w \qquad \text{and} \qquad \left[G * \left|\Delta_\Xi(h*u) - z\right|^2\right](\xi_k) < \bar{\sigma}^2, \qquad \forall \xi_k\in\Xi, \tag{5.1}$$

which implies that the set of functions satisfying the constraints is non-empty.

We prove that the solution of (4.4) can be computed by adapting Uzawa's algorithm. Let $\mu > 0$ and $\lambda = (\lambda_k)_{k=1}^{N^2} \geq 0$. Define the Lagrangian function

$$\mathcal{L}^\mu(u, \{\lambda\}) = K_{\beta,A}(u) + \mu\,(\hat{u}(0,0) - \overline{z}^w)^2 + \sum_{\xi_k \in \Xi} \frac{\lambda_k}{2}([G * |\Delta_\Xi(h * u) - z|^2](\xi_k) - \bar{\sigma}^2).$$

To adapt Uzawa's algorithm we need the following result which can be proved as in the proof of Proposition 5.1.

Theorem 5.2. *For each* $\lambda = (\lambda_k)_{k=1}^{N^2} \geq 0$*, there is a unique solution* u *of*

$$\min_{u \in X_N} \mathcal{L}^\mu(u, \{\lambda\}).$$

Proof. Since $\mathcal{L}^\mu(u, \{\lambda\})$ is lower semicontinuous in u, it suffices to prove that any minimizing sequence u_n is bounded. Since $\mathcal{L}^\mu(u_m, \{\lambda\})$ is bounded, we know that $A(D)(u_m)$ and $(\widehat{u_m}(0,0) - \overline{z}^w)^2$ are bounded. The boundedness of $A(D)(u_m)$ implies that $u_m - \widehat{u_m}(0,0)$ is bounded. Since $(\widehat{u_m}(0,0) - \overline{z}^w)^2$ is bounded, then $\widehat{u_m}(0,0)$ is also bounded. □

We solve (4.4) with Uzawa's algorithm.

Algorithm II: Restoration with local constraints

1. Choose any set of values $\lambda_k^0 \geq 0$, $k = 1, \ldots, N^2$, and $\mu^0 > 0$.
   Iterate from $p = 0$ until convergence of λ^p the following steps:
2. With the values of λ^p, μ^p solve:

$$u_p = \arg\min_u \mathcal{L}^{\mu^p}(u, \{\lambda^p\}). \tag{5.2}$$

3. Update μ and λ in the following way:

$$\mu^{p+1} = \mu^p + 1,$$

$$\lambda_k^{p+1} = \max(\lambda_k^p + \rho_p([G * |\Delta_\Xi(h * u_p) - z|^2](\xi_k) - \bar{\sigma}^2), 0) \quad \forall \xi_k, \tag{5.3}$$

   where $0 < \rho_* \leq \rho_p \leq \rho^*$.

Proposition 5.3. *Assume that there exists* $u \in X_N$ *such that* $\hat{u}(0,0) = \overline{z}^w$ *and* $z(\xi_k) = h*u(\xi_k)$ $\forall \xi_k \in \Xi$*. Then Uzawa's algorithm converges to the solution of (4.4).*

To prove Proposition 5.3 we need to reformulate problem (4.4) as

$$\min_{u \in X_N} \max_{\lambda \geq 0, \alpha_+, \alpha_- \geq 0} \mathcal{L}(u, \{\lambda\}, \alpha_+, \alpha_-), \tag{5.4}$$

where $\lambda = (\lambda_k)_{k=1}^{N^2}$, $\alpha_+, \alpha_- \geq 0$,

$$\mathcal{L}(u, \{\lambda\}, \alpha_+, \alpha_-) = K_{\beta,A}(u) + \textstyle\sum_{\xi_k \in \Xi} \lambda_k([G * |\Delta_\Xi(h * u) - z|^2](\xi_k) - \bar{\sigma}^2)$$
$$+ \alpha_+ \varphi_+(u) + \alpha_- \varphi_-(u),$$

and

$$\varphi_+(u) := \hat{u}(0,0) - \overline{z}^w \qquad \text{and} \qquad \varphi_-(u) = -\hat{u}(0,0) + \overline{z}^w,$$

so that the equality constraint is written as the two inequalities $\varphi_+(u) \leq 0$, $\varphi_-(u) \leq 0$.

Since, by Proposition 5.1, problem (4.4) has a solution, the classical existence result of saddle points (see [26]) proves the existence of a solution of (5.4). Indeed the following result is classical and can be found, for instance, in [26] (Theorems 4 and 6, pp. 59-61) or [21] (Theorem 9.3.2).

THEOREM 5.4. *Assume that* (5.1) *holds. Let* u *be the solution of (4.4). Then there are* $(\{\lambda\}, \alpha_+, \alpha_-) \geq 0$ *such that* $(u, \{\lambda\}, \alpha_+, \alpha_-)$ *is a solution of (5.4), i. e., a saddle point of* $\mathcal{L}(\cdot,\cdot,\cdot,\cdot)$. *If* $(u, \{\lambda\}, \alpha_+, \alpha_-)$ *is a solution of (5.4), then* u *is a solution of (4.4).*

Since we will need it below, let us compute the gradient of $K_{\beta,A}(u)$. For any $v \in X_N$ we have

$$\begin{aligned}\langle \nabla K_{\beta,A}(u), v\rangle_{X_N} &= \left\langle \frac{A(D)Pu}{\sqrt{\beta^2 + |A(D)Pu|^2}}, A(D)Pv \right\rangle_{X_{2N}} \\ &= \left\langle P^* A(D)^* \left(\frac{A(D)Pu}{\sqrt{\beta^2 + |A(D)Pu|^2}} \right), v \right\rangle_{X_N}\end{aligned}$$

for each $v \in X_N$ vanishing on the boundary of $\{0, \ldots, N-1\}^2$. Thus, we may write

$$\nabla K_{\beta,A}(u) = A(D)^* P^* \left(\frac{A(D)Pu}{\sqrt{\beta^2 + |A(D)Pu|^2}} \right) \in X_N.$$

Now, we notice that $P^* f$ is just the restriction operator (subsampling operator) that considers only the samples of $f \in X_{2N}$ in the grid $\{0, \ldots, N-1\}^2$.

Finally, using this and the last two formulas, we deduce that

$$\frac{A(D)Pu}{\sqrt{\beta^2 + |A(D)Pu|^2}} \cdot \nu^{\{0,\ldots,N-1\}^2} = 0$$

where $\nu^{\{0,\ldots,N-1\}^2}$ is the discrete normal.

As usual, we denote by $\|v\|_q = \left(\sum_{i,j=1}^{N^2} |v(i,j)|^q \right)^{1/q}$ for any $v \in X_N$, $1 \leq q < \infty$. We denote $\|v\|_\infty = \max_{(i,j)\in\{1,\ldots,N\}} |v(i,j)|$. And for simplicity in the cases where there is no ambiguity, we shall omit the subindexes for the L_2-norm, then $\|u\|_2 = \|u\|$.

Proof of Proposition 5.3. Let us write $Q(u) = (\widehat{u}(0,0) - \overline{z}^w)^2$, $R(u) = G * |\Delta_\Xi(h * u) - z|^2$. To adapt the convergence proof of Uzawa's method to our case, we need to prove that

(a) If U is a bounded subset of X_N then there is a constant $\alpha > 0$ such that

$$\langle \nabla K_{\beta,A}(u) - \nabla K_{\beta,A}(v), u - v\rangle + \mu \langle \nabla Q(u) - \nabla Q(v), u - v\rangle \geq \alpha \|u - v\|^2$$

for all $u, v \in U$.

(b) $R(u)$ is Lipschitz on bounded sets of X_N and

(c) the sequence u_p constructed in Step 2 of the above algorithm is bounded in X_N.

To prove (a) we use the inequality [48]

$$\left\langle \frac{\xi}{\sqrt{\beta^2 + |\xi|^2}} - \frac{\xi'}{\sqrt{\beta^2 + |\xi'|^2}}, \xi - \xi' \right\rangle \geq \beta^2 \frac{|\xi - \xi'|^2}{(\beta^2 + |\xi|^2 + |\xi'|^2)^{3/2}} \qquad \forall \xi, \xi' \in \mathbb{R}^k$$

with $k = 2$ and we compute

$$\begin{aligned}\langle \nabla K_{\beta,A}(u) - \nabla K_{\beta,A}(v), u - v\rangle &\geq \sum_{(r,l)} \frac{|A(D)Pu(r,l)) - A(D)Pv(r,l)|^2}{(\beta^2 + |A(D)Pu(r,l)|^2 + |A(D)Pv(r,l)|^2)^{3/2}} \\ &\geq \alpha \sum_{(r,l)} |A(D)Pu(r,l) - A(D)Pv(r,l)|^2\end{aligned}$$

where $(r,l) \in \{0, 1, \ldots, 2N-1\}$, the constant $\alpha > 0$ depends on the bound for U, and

$$\langle \nabla Q(u) - \nabla Q(v), u - v\rangle = 2(\hat{u}(0,0) - \hat{v}(0,0))^2.$$

Then (a) follows as a consequence of the two previous inequalities.
(b) Assume that $U \subseteq X_N$ is a bounded set. Let $u, \overline{u} \in U$. Since $\|G\|_1 \leq 1$, we have

$$\begin{aligned}\|R(u) - R(v)\| &\leq \|G\|_1 \|(h*u - z)^2 - (h*v - z)^2\| \\ &\leq 2\|z\|_\infty \|h*(u-v)\| + \|h*(u+v)\|_\infty \|h*(u-v)\| \\ &\leq C\|u - v\|\end{aligned}$$

where C is a constant depending on the norms of h and z and on the bound for U.
(c) To prove that $\{u_p\}_p$ is bounded we observe that

$$\mathcal{L}^\mu(u_p, \{\lambda^p\}) \leq \mathcal{L}^\mu(u, \{\lambda^p\}), \qquad \forall u \in X_N, \tag{5.5}$$

for all p. Choosing $u \in X_N$ such that $\hat{u}(0,0) = \overline{z}^w$ and $z = \Delta_\Xi(h*u)$, we obtain that

$$K_{\beta,A}(u_p) + \mu^p Q(u_p) \leq K_{\beta,A}(u),$$

hence $\{u_p\}_p$ is bounded in X_N.

Now, we can adapt the proof of Uzawa's method to our case (see Theorem 5 in [26], Sect. 3.1). Since u_p satisfies (5.5) we have

$$\langle \nabla K_{\beta,A}(u_p), u - u_p\rangle + \mu^p \langle \nabla Q(u_p), u - u_p\rangle + \langle \lambda^p, R(u) - R(u_p)\rangle \geq 0 \qquad \forall u \in X_N. \tag{5.6}$$

Let u^* be the solution of problem (4.4). Since, by Theorem 5.4, we have

$$\mathcal{L}(u^*, \{\lambda\}, \alpha_+, \alpha_-) \leq \mathcal{L}(u, \{\lambda\}, \alpha_+, \alpha_-) \qquad \forall u \in X_N,$$

we also have

$$\begin{aligned}&\langle \nabla K_{\beta,A}(u^*), u - u^*\rangle + \langle \lambda, R(u) - R(u^*)\rangle \\ &+\alpha_+(\varphi_+(u) - \varphi_+(u^*)) + \alpha_-(\varphi_-(u) - \varphi_-(u^*)) \geq 0\end{aligned} \tag{5.7}$$

for all $u \in X_N$. Since u^* is a solution of (4.4), we have that

$$\nabla Q(u^*) = 0,$$

and we can add $\mu^p \langle \nabla Q(u^*), u - u^*\rangle$ to the inequality (5.7). Taking $u = u^p$ in this form of the second inequalities, and $u = u^*$ in (5.6) and adding both of them we obtain

$$\begin{aligned}&\langle \nabla K_{\beta,A}(u_p) - \nabla K_{\beta,A}(u^*), u_p - u^*\rangle + \mu^p \langle \nabla Q(u_p) - Q(u^*), u_p - u^*\rangle \\ &-\alpha_+(\varphi_+(u_p) - \varphi_+(u^*)) - \alpha_-(\varphi_-(u_p) - \varphi_-(u^*)) + \langle \lambda^p - \lambda, R(u_p) - R(u^*)\rangle \leq 0\end{aligned}$$

Since

$$\langle \nabla K_{\beta,A}(u_p) - \nabla K_{\beta,A}(u^*), u_p - u^* \rangle \geq \alpha \|A(D)Pu_p - A(D)Pu^*\|^2$$

and

$$\begin{aligned} &\mu^p \langle \nabla Q(u_p) - Q(u^*), u_p - u^* \rangle - \alpha_+(\varphi_+(u_p) - \varphi_+(u^*)) - \alpha_-(\varphi_-(u_p) - \varphi_-(u^*)) \\ &= 2\mu_p(\widehat{u_p}(0,0) - \widehat{u^*}(0,0))^2 - (\alpha_+ - \alpha_-)(\widehat{u_p}(0,0) - \widehat{u^*}(0,0)) \\ &\geq \mu^p(\widehat{u_p}(0,0) - \widehat{u^*}(0,0))^2 \end{aligned}$$

for p large enough, we have

$$\begin{aligned} \langle \lambda^p - \lambda, R(u_p) - R(u^*) \rangle &\leq -\alpha \|A(D)Pu_p - A(D)Pu^*\|^2 - \mu^p(\widehat{u_p}(0,0) - \widehat{u^*}(0,0))^2 \\ &\leq -\alpha_0 \|u_p - u^*\|. \end{aligned} \tag{5.8}$$

Now, the proof follows in a standard way. Let us give the details for the sake of completeness. Using (5.3), we have

$$\|\lambda^{p+1} - \lambda\|_{\leq} \|\lambda^p - \lambda + \rho_p(R(u_p) - R(u^*))\|.$$

Taking squares, we have

$$\|\lambda^{p+1} - \lambda\|^2 \leq \|\lambda^p - \lambda\|^2 + 2\rho_p \langle \lambda^p - \lambda, R(u_p) - R(u^*) \rangle + \rho_p^2 \|R(u_p) - R(u)\|^2.$$

Using (5.8) and (b), we have

$$\|\lambda^{p+1} - \lambda\|^2 \leq \|\lambda^p - \lambda\|^2 - 2\alpha_0 \rho_p \|u_p - u^*\|^2 + \rho_p^2 L^2 \|u_p - u^*\|^2.$$

for some $L > 0$. If we choose ρ_p such that

$$2\alpha_0 \rho_p - L^2 \rho_p^2 \geq \gamma > 0,$$

that is, $0 < \rho_* \leq \rho_p \leq \rho^*$, we have

$$\|\lambda^{p+1} - \lambda\|^2 \leq \|\lambda^p - \lambda\|^2 - \gamma \|u_p - u^*\|^2.$$

Then we deduce that $\|\lambda^p - \lambda\|$ is decreasing and, thus, has a limit $\ell \geq 0$. Then letting $p \to \infty$ we have that $\|u_p - u^*\| \to 0$. ☐

6. A Quasi-Newton algorithm for the solution of (5.2)**.** The purpose of this Section is to explain the algorithm used to solve problem (5.2) in Algorithm II. For convenience, let us denote the convolution and irregular sampling operators, as $\mathcal{S}u = \Delta_\Xi(h * u)$ for any $u \in X_N$.

Observe that the Euler-Lagrange equation corresponding to (5.2) is

$$A(D)^* \left(P^* \frac{A(D)Pu}{\sqrt{\beta^2 + |A(D)Pu|^2}} \right) + 2\mu(\hat{u}(0,0) - \overline{z}^w) + \mathcal{S}^*(G * \lambda)(\mathcal{S}u - z) = 0. \tag{6.1}$$

To shorten our expressions, let us define the following operators:

$$\mathcal{M}u = (\sqrt{G * \lambda})\mathcal{S}u, \qquad \text{so that} \qquad \mathcal{M}^*u = \mathcal{S}^*(\sqrt{G * \lambda u}),$$

$$\mathcal{N}u = \hat{u}(0,0), \qquad b = \mathcal{S}^*((G * \lambda)z) + 2\mu\overline{z}^w,$$

$$\mathcal{A}[v](u) = A(D)^*\, P^* \left(\frac{A(D)Pu}{\sqrt{\beta^2 + |A(D)Pv|^2}} \right)$$

and

$$\mathcal{T}[v](u) = \mathcal{A}[v](u) + 2\mu\mathcal{N}u + \mathcal{M}^*\mathcal{M}u.$$

where $u, v \in X_N$.

We want to solve (6.1) with a fixed point iteration:

$$\mathcal{A}[u^t](u^{t+1}) + 2\mu\mathcal{N}u^{t+1} + \mathcal{M}^*\mathcal{M}u^{t+1} = b. \tag{6.2}$$

The rest of this Section is devoted to show that such a fixed point algorithm converges to the minimizer of (5.2). Notice that a proof based on convex analysis (the half-quadratic regularization approach) can be found in [17] in the continuous case, or in [20, 9] for the discrete case. Further analysis can be found in [1, 23, 49]. Here, we will extend the proof proposed by Chan and Mulet in [19]. The advantage of such an approach is that only basic algebra is needed. Moreover, the linear convergence rate of this algorithm can be shown explicitly. The difference with the approach in [19] relies on the fact that the operator $\mathcal{M}$ is not assumed to be invertible. In our case the presence of the mean constraint $\mu\mathcal{N}$ permits to prove the same convergence result as in [19], without an invertibility hypothesis on $\mathcal{M}$.

Remark 5. Note that computing $\mathcal{M}^*\mathcal{M}$ in (6.2) entails the computation of an operator with a Toeplitz structure. This is efficiently computed in $\mathcal{O}(N^2 \log_2 N^2)$ steps, as in the ACT algorithm (Section 2.1).

Remark 6. As mentioned in Remark 3, we may write $A(D)u = D\Psi u$. This permits to use the change of variables $v = \Psi u$ and write the constrained restoration problem (4.4) in terms of v. Then the regularizer coincides with the total variation applied to v. The solution of (4.4) can be recovered as $u = \Psi^{-1}v$. This represents no change in the developments of this paper and we keep the notation $A(D)u$.

6.1. Existence of u^t and its boundedness.

The sequence u^t will be defined iteratively using (6.2).

Proposition 6.1. *The equation (6.2) has a unique solution $u^{t+1} \in X_N$ which is the minimizer of*

$$\mathcal{E}(u) = \left\| \frac{A(D)Pu}{(\beta^2 + |A(D)Pu^t|^2)^{1/4}} \right\|^2 + \frac{1}{2}\|\mathcal{M}u - z'\|^2 + \mu\|\mathcal{N}u - \overline{z}^w\|^2 \tag{6.3}$$

where $\left\| \frac{A(D)Pu}{(\beta^2+|A(D)Pu^t|^2)^{1/4}} \right\|^2 = \sum_{(r,l)} \left| \frac{A(D)Pu(r,l)}{(\beta^2+|A(D)Pu^t(r,l)|^2)^{1/4}} \right|$ *and* $z' = \sqrt{G * \lambda}z$.

Proof. It is standard that (6.3) admits a unique solution $u^{t+1} \in X_N$. Moreover, (6.2) is the Euler-Lagrange equation associated to (6.3) and solutions of (6.2) are minimizers of (6.3). □

Proposition 6.2. (*i*) *There exists $K_0 > 0$ such that*

$$\|A(D)Pu^t\| \le K_0.$$

(ii) $\mathcal{T}[u^t]$ *is a bounded coercive operator. Indeed we have*

$$\langle \mathcal{T}[u^t]u, u\rangle \geq \alpha \|u\|^2 \tag{6.4}$$

for some $\alpha > 0$ *independent of* t.

(iii) The sequence u^t *is uniformly bounded.*

Proof. Since u^{t+1} is a minimizer of $\mathcal{E}(u)$, we have $\mathcal{E}(u^{t+1}) \leq \mathcal{E}(0) = \frac{1}{2}\|z'\|^2 + \mu\|\overline{z}^w\|^2$, and thus:

$$\left\| \frac{1}{(\beta^2 + |A(D)Pu^t|^2)^{1/4}} A(D)Pu^{t+1} \right\|^2 \leq \mathcal{E}(0) = \frac{1}{2}\|z'\|^2 + \mu\|\overline{z}^w\|^2.$$

We have

$$\left| \frac{A(D)Pu^{t+1}}{(\beta^2 + |A(D)Pu^t|^2)^{1/4}} \right|^2 \geq \frac{\left|A(D)Pu^{t+1}\right|^2}{\|\sqrt{\beta^2 + |A(D)Pu^t|^2}\|_\infty}$$

Thus $\left\|A(D)Pu^{t+1}\right\|^2 \leq \mathcal{E}(0)\left\|\sqrt{\beta^2 + |A(D)Pu^t|^2}\right\|_\infty \leq \mathcal{E}(0)\sqrt{\beta^2 + \|A(D)Pu^t\|_\infty^2}$. But since we deal with finite dimensional spaces, there exists $L > 0$ which does not depend on u^t such that $\|A(D)Pu^t\|_\infty \leq L\|A(D)Pu^t\|$. Hence we deduce that

$$\left\|A(D)Pu^{t+1}\right\|^2 \leq \mathcal{E}(0)\sqrt{\beta^2 + L^2\|A(D)Pu^t\|^2}$$

Assume that $\|A(D)Pu^t\| \leq K$. Using (6.1), to get that $\left\|A(D)Pu^{t+1}\right\| \leq K$, it is sufficient to choose $K > 0$ large enough so that

$$\mathcal{E}(0)\sqrt{\beta^2 + L^2K^2} \leq K^2.$$

(ii) The boundedness of $\mathcal{T}[u^t]$ is immediate and we omit its proof. Let us prove that $\mathcal{T}_t$ is a coercive operator. Using the bounds in Step *(i)*, we have

$$\left\langle \frac{1}{\sqrt{\beta^2 + |A(D)Pu^t|^2}} A(D)Pu, A(D)Pu \right\rangle \geq \frac{1}{\sqrt{\beta^2 + L^2K_0^2}}\|A(D)Pu\|^2.$$

Hence

$$\langle \mathcal{A}[u^t]u, u\rangle \geq \frac{1}{\sqrt{\beta^2 + L^2K_0^2}}\|A(D)Pu\|^2 \geq \alpha_0\|u_0\|^2,$$

where we wrote $u = u_0 + c$, with $c = \mathcal{N}u$ and $\mathcal{N}u_0 = 0$, and we used the fact that $A(D)P$ is a linear operator whose kernel are the constants. Using $\mathcal{A}[u^t]c = 0$ and $\langle u_0, c\rangle = 0$, we get

$$\begin{aligned}\langle \mathcal{T}[u^t](u_0 + c), u_0 + c\rangle &\geq \langle \mathcal{A}[u^t](u_0 + c), u_0 + c\rangle + 2\mu\langle \mathcal{N}(u_0 + c), u_0 + c\rangle \\ &\geq \langle \mathcal{A}[u^t]u_0, u_0\rangle + 2\mu c^2 \\ &\geq \alpha_0\|u_0\|^2 + 2\mu c^2.\end{aligned}$$

Thus, we deduce (6.4).

(iii) From (6.4), we know that $\langle \mathcal{T}[u^t]u^{t+1}, u^{t+1}\rangle \geq \alpha\|u^{t+1}\|^2$. But from (6.1), we know that $\langle \mathcal{T}[u^t]u^{t+1}, u^{t+1}\rangle = \langle b, u^{t+1}\rangle \leq \|b\|\|u^{t+1}\|$. We deduce that $\|u^{t+1}\| \leq \frac{\|b\|}{\alpha}$. $\square$

6.2. Convergence of the fixed point algorithm. For simplicity, given $\lambda = (\lambda_k)_{k=1}^{N^2}$, we write $\mathcal{L}(u) = \mathcal{L}^{\mu}(u, \{\lambda\})$. Recall that

$$\nabla_u \mathcal{L}(u) = \mathcal{T}[u](u) - 2\mu \overline{z}^w - \mathcal{M}^* z'.$$

Let us finally define

$$\mathcal{G}(v,u) = \mathcal{L}(u) + \langle v-u, \nabla_u \mathcal{L}(u)\rangle + \frac{1}{2}\langle v-u, \mathcal{T}[u](v-u)\rangle$$

Proposition 6.3. *The following inequality holds for any* $u, v \in X_N$*:*

$$\mathcal{L}(v) \leq \mathcal{G}(v,u) \tag{6.5}$$

Proof. We follow the proof in [19]. Since

$$\mathcal{G}(v,u) - \mathcal{L}(v) = \mathcal{L}(u) - \mathcal{L}(v) + \langle v-u, \nabla_u \mathcal{L}(u)\rangle + \frac{1}{2}\langle v-u, \mathcal{T}[u](v-u)\rangle,$$

standard computations lead to

$$\mathcal{G}(v,u) - \mathcal{L}(v) = \sum_{(r,l)} \left(a - \overline{a} + \frac{1}{2a}(\overline{a}^2 - a^2) \right)$$

with $a = \sqrt{\beta^2 + |A(D)Pu(r,l)|^2}$ and $\overline{a} = \sqrt{\beta^2 + |A(D)Pv(r,l)|^2}$ where $(r,l) \in \{0, 1, \ldots, 2N-1\}$. Since $a, \overline{a} > 0$, and $a - \overline{a} + \frac{1}{2a}(\overline{a}^2 - a^2) = \frac{(a-\overline{a})^2}{2a} \geq 0$, we have that $\mathcal{G}(v,u) - \mathcal{L}(v) \geq 0$. ◻

Proposition 6.4. (i) *The function* u^{t+1} *defined by (6.2) is such that:*

$$u^{t+1} = \text{argmin}_v \mathcal{G}(v, u^t) \tag{6.6}$$

i.e.: $0 = \nabla_u \mathcal{L}(u^t) + \mathcal{T}[u^t](u^{t+1} - u^t)$.

(ii) *We have* $\lim_{t \to +\infty} \|u^{t+1} - u^t\| = 0$.

Proof. (i) Let us denote by $\bar{u} = \text{argmin}_v \mathcal{G}(v, u^t)$. We thus have $0 = \nabla_u \mathcal{L}(u^t) + \mathcal{T}[u^t](\bar{u} - u^t)$. And this last equation is precisely equation (6.2), which implies that $\bar{u} = u^{t+1}$.

(ii) From (6.5) and (6.6), we have $\mathcal{L}(u^{t+1}) \leq \mathcal{G}(u^{t+1}, u^t) \leq \mathcal{G}(u^t, u^t) \leq \mathcal{L}(u^t)$, i.e. $(\mathcal{L}(u^t))$ is decreasing. Now, from (6.5) and (6.6), we have:

$$\begin{aligned}
\mathcal{L}(u^{t+1}) &\leq \mathcal{G}(u^{t+1}, u^t) \\
&= \mathcal{L}(u^t) + \langle u^{t+1} - u^t, \nabla_u \mathcal{L}(u^t)\rangle + \frac{1}{2}\langle u^{t+1} - u^t, \mathcal{T}[u^t](u^{t+1} - u^t)\rangle \\
&= \mathcal{L}(u^t) - \frac{1}{2}\langle u^{t+1} - u^t, \mathcal{T}[u^t](u^{t+1} - u^t)\rangle
\end{aligned}$$

Using (6.4), we deduce: $\frac{1}{2}\alpha \|u^{t+1} - u^t\|^2 \leq \frac{1}{2}\langle u^{t+1} - u^t, \mathcal{T}[u^t](u^{t+1} - u^t)\rangle \leq \mathcal{L}(u^t) - \mathcal{L}(u^{t+1})$ and (ii) follows. ◻

We are now in position to state a convergence result.

Theorem 6.5. *The sequence* u^t *defined by (6.2) converges to the solution of (5.2).*

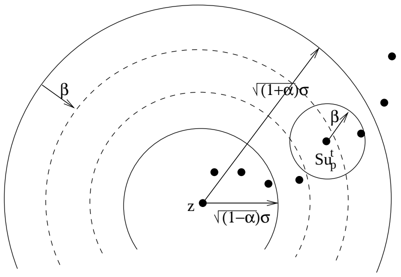


FIG. 7.1. *The relaxed optimization problem allows solutions within a band. From the depicted sequence of solutions* u_p^t *(black dots) the ones that are truly satisfying the constraints are those inside the dotted ring, since it accounts for the approximation error* η. $\mathcal{S}$ *stands for the irregular sampling and convolution operators,* Su_p^t *is also a sequence defined over the irregular samples and not over the solution space.*

Proof. From Proposition 6.2.(*iii*), we know that u^t is uniformly bounded. There exists u such that we can extract a convergent subsequence, which we still denote by u^t, with $u^t \to u$ as $t \to +\infty$. From Proposition 6.4.(*ii*) , we know that u^{t+1} is also convergent and $u^{t+1} \to u$ as $t \to +\infty$. Letting $t \to +\infty$ in equation (6.2), we deduce that u is the solution of (6.1) (and thus of (5.2)). By uniqueness of the solution of (5.2), we conclude that the whole sequence u^t converges to u, solution of (5.2). □

We end this Section by stating a result about the convergence rate. We denote by $\tilde{u}$ the solution of Problem (5.2). We use the following notations:

$$\gamma_t := \frac{\mathcal{G}(\tilde{u}, u^t) - \mathcal{L}(\tilde{u})}{\frac{1}{2}\langle \tilde{u} - u^t, \mathcal{T}[u^t](\tilde{u} - u^t)\rangle}$$

and

$$\eta := 1 - \lambda_{\min}(\mathcal{T}[\tilde{u}]^{-1}\nabla_u^2 \mathcal{L}(\tilde{u}))$$

where $\lambda_{\min}(M)$ denotes the smallest eigenvalue of the matrix M; in particular if M is positive definite then $\lambda_{\min}(M) > 0$

PROPOSITION 6.6.

1. $\mathcal{L}(u^{t+1}) - \mathcal{L}(\tilde{u}) \leq \gamma_t(\mathcal{L}(u^t) - \mathcal{L}(\tilde{u}))$.
2. $\eta < 1$ *and* $0 \leq \gamma_t \leq \eta$, *for* t *sufficiently large. In particular,* $\mathcal{L}(u^t)$ *has a linear convergence rate of at most* η.
3. u^t *is r-linearly convergent with a convergence rate of at most* $\sqrt{\eta}$.

Proof. We refer the interested reader to the proof of Theorem 6.1 in [19] which can easily be extended to our case. □

7. Band constraints and stopping conditions.

Coming back to the optimization problem (4.4), since both the functional $K_{\beta,A}(u)$ and the constraints are convex, and the constraint's feasible set V does not contain the absolute minimun of $K_{\beta,A}(u)$, then the solution lies in the boundary of V. In practice, computing a solution in ∂V is not only computationally too expensive (due to the size of the problems), but also unnecessary because of the noise. Since we rely on noise estimates that have a certain accuracy, exceeding this accuracy in the data fitting is useless (as we will see later in this Section). Moreover, as it has been observed in all numerical experiments [14, 16, 17, 18, 19, 20, 24, 29, 10, 31, 44, 49], using total variation as regularizer in denoising or restoration generally carries some loss of texture and it is not desirable to compute the solution that (absolutely) minimizes the TV but to keep a solution with a slightly higher TV value in order to avoid the loss of textures.

As a consequence, to avoid this degradation, the rule of thumb has been ever since to remove less noise than noise is actually present in the image. Gilboa studied this in [10] and concluded that in terms of SNR the optimal selection of $\bar{\sigma}$ is between 0.7 and 0.8 times the value of σ. In what follows we will modify the constraints to account for this change

$$\left[G * |\Delta_{\Xi}(h * u) - z|^2\right](\xi_k) \leq \bar{\sigma}^2 \quad \forall \xi_k \in \Xi\,,$$

where $\bar{\sigma} < \sigma$. That is, in order to keep more texture in u we do not remove all noise and we could write $z = \mathcal{S}(u_0 + \mathcal{S}^{-1}(n - \bar{n})) + \bar{n}$ where $\bar{n}$ is a noise with variance $\bar{\sigma}^2$ and we identify $u_0 + \mathcal{S}^{-1}(n - \bar{n})$ with u.

Motivated by this observation and by the fact that, due to noise, there is always some uncertainty in the surroundings of ∂V, we will avoid the computational overhead of getting exactly to ∂V by stopping the algorithm as soon as the solution is close to it. This is the notion behind the *band constraint*

$$(1 - \alpha)\bar{\sigma}^2 \leq \left[G * |\Delta_{\Xi}(h * u) - z|^2\right](\xi_k) \leq (1 + \alpha)\bar{\sigma}^2 \quad \forall \xi_k \in \Xi, \tag{7.1}$$

with $\alpha > 0$. The constraint described by equation (7.1) is clearly non-convex, and therefore it cannot be integrated in the method presented here. But since Uzawa's algorithm always pushes the solution towards the boundary of the feasible set, then (at least in practice) it can be used to stop Uzawa's loop by testing if (7.1) is fulfilled.

In what follows we will see that even considering relaxed constraints like (7.1), imposing all of the local constraints simultaneously is not statistically correct, since each constraint's estimator behaves as a random variable. Then we will see how this relaxation of constraints is used to early stop Uzawa's iterations and how this helps to improve the efficiency of our implementation.

In our experiments, we have chosen $\bar{\sigma} = 0.8\sigma$ and α such that $0.8(1 + \alpha) < 1$.

7.1. Expected number of satisfied *band* constraints. Let us summarize the arguments of [4] and adapt them to the case of band constraints. Each local constraint relies on a local estimate of the residual variance of the form

$$S_G(\xi_k) = \left[G * |\Delta_{\Xi}(h * u) - z|^2\right](\xi_k) = \left[G * |\bar{n}|^2\right](\xi_k)\,, \tag{7.2}$$

where G is a Gaussian or uniform window centered at the interest point and n_k denotes a zero mean Gaussian noise and variance $\bar{\sigma}^2$ (recall that we are going to remove only a noise of variance $\bar{\sigma}^2 < \sigma^2$). Since S_G is a random variable itself, the number of satisfied constraints is estimated by the probability $P[S_G \leq (1 + \alpha)\bar{\sigma}^2]$ in the case of the ball shaped constraint, or $P[(1 - \alpha)\bar{\sigma}^2 \leq S_G \leq (1 + \alpha)\bar{\sigma}^2]$ for the band constraint. Observe that a constraint of the type $S_G = \bar{\sigma}^2$ (that is in practice imposed when solving exactly (4.4)) was already doomed to failure since it has zero probability to occur $P[S_G = \bar{\sigma}^2] = 0$.

Using the Central Limit Theorem gives only a loose estimate of the probability distribution of S_G. To improve the estimation of the expected number of satisfied constraints let us simplify S_G. By approximating the discrete convolution with G (of standard deviation $\tilde{r}$) by a mean over a disk I of radius $r = 2\tilde{r}$ we can define a simpler estimator $S_I = \frac{1}{|I|}\sum_{k \in I} \bar{n}_k^2$.

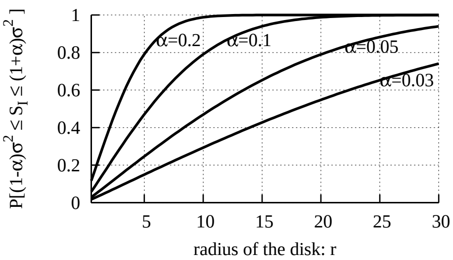


FIG. 7.2. *Relation between the expected number of satisfied constraints and the radius r of the disk used for the local noise estimation. The graph shows the curves computed for different widths of the band α.*

Then, for the case of the ball constraint, the expected number of satisfied constraints is the number of pixels times the following probability

$$P[S_I \leq (1+\alpha)\bar{\sigma}^2] = P\left[\frac{1}{|I|}\sum_{k\in I}\bar{n}_k^2 \leq (1+\alpha)\bar{\sigma}^2\right] = P\left[\sum_{k\in I}\left(\frac{\bar{n}_k}{\bar{\sigma}}\right)^2 \leq (1+\alpha)|I|\right].$$

Notice that in the rightmost equation $\sum_{k\in I}\left(\frac{\bar{n}_k}{\bar{\sigma}}\right)^2$ is a sum of $|I|$ squared normalized Gaussian random variables, so it follows a chi-square distribution with $|I|$ degrees of freedom ($\chi^2(|I|)$). And the probability can be computed using the incomplete gamma function $\Gamma(a,x) = \int_x^\infty t^{a-1}e^{-t}dt$

$$P[S_I \leq (1+\alpha)\bar{\sigma}^2] = P\left[\chi^2(|I|) \leq (1+\alpha)|I|\right] = \Gamma\left(\frac{(1+\alpha)|I|}{2}, \frac{|I|}{2}\right).$$

In the case of the band constraint the expected number of satisfied constraints $N(\alpha,r)$ is the number of pixels times the following probability

$$P\left[(1-\alpha)\bar{\sigma}^2 \leq S_I \leq (1+\alpha)\bar{\sigma}^2\right] = \Gamma\left(\frac{(1+\alpha)|I|}{2}, \frac{|I|}{2}\right) - \Gamma\left(\frac{(1-\alpha)|I|}{2}, \frac{|I|}{2}\right). \quad (7.3)$$

Equation (7.3) expresses the expected proportion of satisfied constraints as a function of the radius of the disk r ($|I| = \pi r^2$) and the width of the band α. We plot in Figure 7.2 this function, for different values of α to give an intuition of its behavior. Notice that the expected number of satisfied constraints decreases as we reduce the band width α or the radius r. This permits to determine one parameter as a function of the other two, i.e. using a disk of radius $r = 13$ (or a Gaussian with standard deviation 7.0) and defining a band of width $0.2\bar{\sigma}^2$ ($\alpha = 0.1$) gives 89% of satisfied constraints. In practice, either we specify α, r and then the expected number of satisfied constraints is $N(\alpha,r)$, or we give α and N_α and we compute r so that $N(\alpha,r) = N_\alpha$. We have taken the second option in the experiments displayed in Section 8.

REMARK 7. Observe that in (7.2) the estimation of the noise variance corresponds to the case when the mean of the random variable is zero. Indeed, the global mean is enforced to be zero in (4.4). We should also impose that the local means are zero with a new set of constraints, otherwise S_G will be an overestimation of the noise variance. Adding the local mean constraint $\sum_{\xi_k\in\Xi}|G*(\Delta_\Xi(h*u)-z)|^2(\xi_k) = 0$ in (4.4) adapts to the formalism developed in this paper. But to avoid the computational overhead of its implementation, and since the overestimation plays in favor of the relaxation arguments presented earlier in this Section, we will not include it in the present formulation.

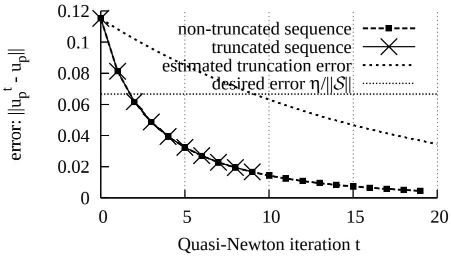


Fig. 7.3. *Truncation of the Quasi-Newton method. The first sequence shows the error evolution along the Quasi-Newton iterations, without applying the stopping condition for the CG algorithm. The second sequence was obtained stopping the CG with the empirical bound, observe that the two sequences are indistinguishable. The third sequence depicts the estimated error used to effectively stop the Quasi-Newton iterations, as soon as the desired error is achieved.*

Table 7.1

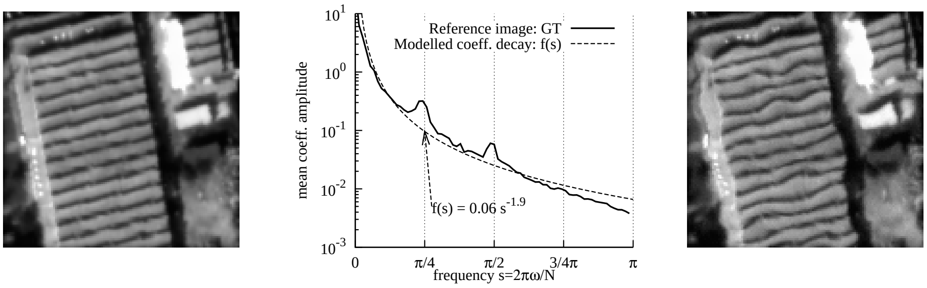
Relation between the width of the constraint band and the restoration time. Reducing the width of the band increases the computational cost of the restoration algorithm. In the table the value of η *was selected according to the rule* $\eta = \frac{1}{3}(\sqrt{1+\alpha\bar{\sigma}} - \sqrt{1-\alpha\bar{\sigma}})$*, and* $\bar{\sigma} = 1$*. All the reported times correspond to experiments ran on a* 1.6*Ghz CPU restoring a* 256×256 *pixels image.*

Band Width	**Effective Band**		**Total**
parameter: α	$\sqrt{(1-\alpha)\bar{\sigma}} + \eta$	$\sqrt{(1+\alpha)\bar{\sigma}} - \eta$	**running time**
0.60	0.84	1.05	3 min 3 sec
0.34	0.93	1.04	3 min 37 sec
0.22	0.96	1.03	4 min 58 sec
0.10	0.98	1.02	18 min

7.2. Practical stopping conditions for an efficient implementation. Using (7.3) we may derive a practical rule to stop Uzawa's loop. Indeed, the user specifies α and the proportion of constraints N_α/N that must lie within the band of width α, and the algorithm deduces the radius r of the kernel G such that $N(\alpha, r) = N_\alpha$. Then we iterate the Uzawa's loop in Algorithm II until the number of pixels that satisfy the constraint (7.3) is at least N_α.

The truncation error of the Quasi-Newton has three sources: *(i)* truncation of the Quasi-Newton iterations themselves, *(ii)* truncation of the nested CG loop, and *(iii)* propagation of the CG error along QN iterations. Here we summarize how to estimate and control the combination of the three errors for a given (global) target error bound on the QN result $\|u_p - u_p^t\|^2 \leq \eta/\|\mathcal{S}\|$. Using standard error propagation analysis [22] and the knowledge that QN is at least linearly convergent we can estimate the global error determining *(ii)* and bound the inverses of the operators $\mathcal{T}[u^t]$ and their dependence on u^t. In our case these bounds are estimated empirically, and the CG error is approximated by its residual. Figure 7.3 shows that this procedure is quite effective in practice. First our CG stopping condition makes the truncated QN sequence indistinguishable from the non-truncated QN sequence (i.e. the one with CG iterated until exact convergence is reached, and thus not affected by CG truncation errors *(ii)* and their propagation *(iii)*). This is because the actual QN truncation error (with either CG truncated or not) is considerably over-estimated by our error propagation analysis as shown in Figure 7.3. This also means that the desired error is achieved much faster than predicted by the our error bounds.

Let us now take into account the truncation error of the Quasi-Newton method in the determination of the band constraints. Assume that we are computing Quasi-



(a) Reference Image.

(b) Mean Fourier coefficient amplitude for the reference image.

(c) Noisy and irregularly sampled image.

FIG. 8.1. *Image used in the denoising experiment. At left: the reference image is non-aliased, and it has 149 × 149 pixels. It was multiplied by a smooth window vanishing on the borders in order to avoid periodization artifacts (not shown). In this image the mean Fourier coefficient amplitudes decay like* $(2\pi\omega/N)^{-1.9}$ *(center). At right, is shown the perturbed sampled image. The perturbations have an amplitude of* $A = 0.88$ *pixels (standard deviation of* $\varepsilon(x)$*), and were simulated according to* (2.2) *as colored noise with the spectral content inside* $[-0.5/T_\varepsilon, 0.5/T_\varepsilon]^2$ *for* $T_\varepsilon = 10$.

Newton solution u_p and we have controlled the errors $\|\mathcal{S}u_p - \mathcal{S}u_p^t\| \le \|\mathcal{S}\|\|u_p - u_p^t\| \le \eta$, where u_p^t denotes the solution obtained at the t-th iteration of the method. Then we may erode the band by η as seen in Figure 7.1. In this way we ensure that if we stop the Quasi-Newton solution with the criterion $\|\mathcal{S}\|\|u_p - u_p^t\| \le \eta$, then the truncated solution u_p satisfies (7.3). Clearly η must satisfy the inequality $\eta < \frac{1}{2}(\sqrt{1+\alpha\sigma} - \sqrt{1-\alpha\sigma})$ and we have taken $\eta := \frac{1}{3}(\sqrt{1+\alpha\sigma} - \sqrt{1-\alpha\sigma})$, which equalizes the widths of the three bands in Figure 7.1. Figure 7.1 illustrates the band and its reduced version for a single constraint.

Finally, we notice that the computational complexity of the algorithm increases as we reduce the width of the band. Indeed, taking $\alpha \to 0$ makes it harder to satisfy the constraints. And to illustrate this we display in Table 7.1 the computation times corresponding to different values of α.

8. Experiments. We will test the proposed algorithm in three contexts, first in the irregular to regular sampling and denoising task, then in Subsection 8.2 we add the deconvolution, and finally in Subsection 8.3 we consider the full restoration problem, with deconvolution, denoising, and zooming.

We compare the results obtained by the ACT Algorithm [32] (when applicable), the algorithm described in [5] (ACT+TV), and the proposed algorithm discussing different choices of $A(D)$ in each context. In addition we will also consider a reformulation of our local constraint algorithm based on the minimization of the L^2-norm of the pseudo-differential operator $A(D)$. The L^2-norm formulation is faster and leads to a linear system (which is solved directly with Conjugate Gradient, avoiding the need of a fixed point loop), and it is derived by replacing the regularizer $K_{A,\beta}$ in (4.4) by

$$\mathrm{L}_A(u) := \sum_{0 \le r,l \le 2N} |A(D)Pu(r,l)|^2. \tag{8.1}$$

All the experiments were performed with simulated images. The perturbations $\varepsilon(x)$ were generated according to the model (2.2), with an amplitude $A = 0.88$ (stan-

Table 8.1

Comparison of the algorithms in the irregular to regular sampling and denosing task. These results correspond to the restoration of the image shown in Figure 8.1, corrupted by a white Gaussian noise with standard deviation $\sigma = 1$*. The error column is obtained by comparing the restored image* u *with the ground truth* u_0*, where* $RMSE(u, u_0) = (\frac{1}{N}\|u - u_0\|^2)^{1/2}$*. Its values evidence that all the algorithms achieve errors similar to the noise variance. In all the experiments the power of the removed noise was set to be* $\frac{1}{N}\|\mathcal{S}u - z\|^2 \sim 0.908$*, a little below the noise level* $\sigma^2 = 1$*, but this allows to improve both the result's RMSE and the visual quality of the restored images.*

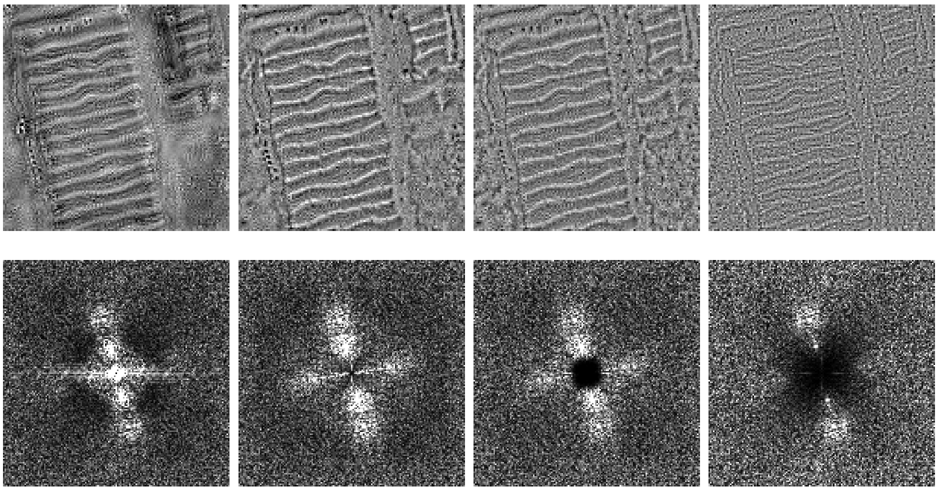
Algorithm	**Regularizer**	**RMSE**
ACT [32] or (2.5)		1.354
ACT$_D$ (2.8) residual preconditioning		1.121
ACT$_R$ (2.9) regularity preconditioning		1.049
ACT+TV [5] or (3.1)		0.961
QN+TV $\lvert A(i\omega)\rvert = \lvert\frac{2\pi}{N}\omega\rvert$	$K_{\beta,A}(u)$ eq. (4.4)	0.874
QN+FAR $\lvert A(i\omega)\rvert = \lvert\frac{2\pi}{N}\omega\rvert^{1.6}$		0.776
QN+FAR $\lvert A(i\omega)\rvert = \lvert\frac{2\pi}{N}\omega\rvert^{1.9}$		0.757
L^2-norm of the gradient	$\mathrm{L}_A(u)$ eq. (8.1)	0.880
L^2-norm of $A(D)u$ with $\lvert A(i\omega)\rvert = \lvert\frac{2\pi}{N}\omega\rvert^{1.9}$		0.773

dard deviations of $\varepsilon(x)$), and where supp $\hat{\varepsilon} \subseteq [-\frac{0.5}{T_\varepsilon}, \frac{0.5}{T_\varepsilon}]^2$ for $T_\varepsilon = 10$. The perturbed samples z were computed very accurately (usually 10^{-8}) by approximating the irregular sampling formula (2.3) with the transposed NFFT [41]. Finally, we added a white noise of standard deviation σ to the irregular samples. In the experiment displayed here, we have taken $\sigma = 1$ gray levels (i.e.the noise power is 890 times smaller than the image power, $SNR = 29.5dB$). The perturbed image shown in Figure 8.1, which corresponds to the denoising experiment, was simulated according to this procedure.

To quantify the errors we adopt the classical root mean squared error measure $\mathrm{RMSE}(u, u_0) = (\frac{1}{N}\|u - u_0\|^2)^{1/2}$ against the ground truth image (denoted as u_0), and the *method noise* for a qualitative analysis. The method noise was originally aimed at comparing denoising algorithms. It consists in subtracting the restored image to the noisy one, and studying the remaining noise. In our context assuming an image formation model like (1.1) and denoting u the image obtained by a restoration algorithm, the method noise becomes $(z - \mathcal{S}u)$, where z are the noisy samples and $\mathcal{S}$ stands for the irregular sampling and convolution operators. Since the restoration is expected to recover the original image $u \simeq u_0$, the method noise should be as similar to a white noise as possible. In addition, since we would like the original image u_0 not to be altered by the restoration method, the method noise should not exceed the actual noise variance, which justifies our local constraints approach.

8.1. Denoising. Observe in Table 8.1 that the proposed algorithm outperforms (in terms of RMSE) the ACT and ACT+TV [4] algorithms, in the denoising experiment. Also notice in Figure 8.2 how the method noise of the proposed algorithm retains less structure, meaning that the method removes just the noise with less alteration of the texture.

The variants of the algorithm based on the minimization of an L^2-norm (8.1), give results that are comparable with the ones obtained with the FAR regularizer, and in both reported cases the L^2 results are indistinguishable from the ones obtained by FAR. This observation advocates for the local formulation of the constraints and it can be justified by the density of the samples in this denosing experiment, and by the lack of both deconvolution and zoom. We can conclude that, for the denoising (without either deconvolution nor zoom) the advantage of a nonlinear regularizer over the L^2 one is marginal, mostly because no spectral extrapolation is needed. However



(a) ACT$_R$
eq. (2.9) regularizer precondition
RMSE= 1.049

(b) ACT+TV
eq. (3.1)
RMSE= 0.961

(c) QN+TV
eq. (4.4) with
$|A(i\omega)| \sim |\frac{2\pi\omega}{N}|^1$
RMSE= 0.864

(d) QN+FAR
eq. (4.4) with
$|A(i\omega)| \sim |\frac{2\pi\omega}{N}|^{1.9}$
RMSE= 0.757

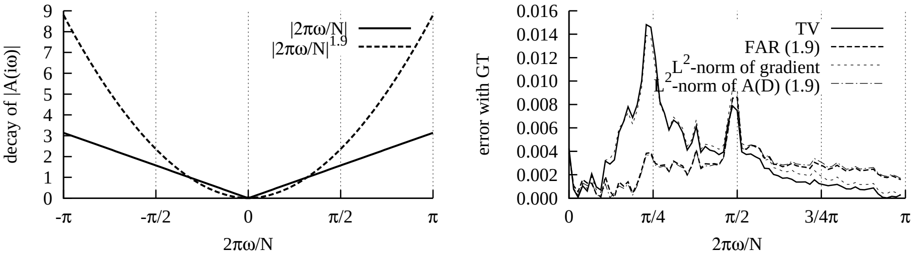
FIG. 8.2. *Method noise of the different algorithms in the experiment of irregular to regular sampling plus denoising. The images in the first row display the method noise* ($z - Su$) *for different methods, less visible structure indicates a better reconstruction. The images display the grayscale range* $[-3, 3]$ *scaled to* $[0, 255]$ *(the full grayscale range of the image* z *is* $[0, 155]$*). In the second row we show the corresponding Fourier transforms, the spectrum highlights the structures that are barely visible in the spatial domain.*

minimizing the L^2-norm is not expected to perform well in tasks that entail a spectral extrapolation like deconvolving or zooming.

The imposition of the spectral profile produces a consistent improvement in all the cases ($|\frac{2\pi}{N}\omega|$ vs. $|\frac{2\pi}{N}\omega|^{1.6}$ vs. $|\frac{2\pi}{N}\omega|^{1.9}$), and imposing the profile corresponding to the coefficient decay of the reference image (Figure 8.1) produces the best results. We can study the performance of the spectral profiles by analyzing the frequency distribution of the errors of the restored images with respect to the reference image (GT), shown in Figure 8.3. There we can see that using the linear profile (total variation), the low frequencies are heavily penalized and most of the errors come from them, but imposing the profile corresponding to this image ($|\frac{2\pi}{N}\omega|^{1.9}$) we reduce the errors in the low frequency range.

Finally, let us spend a word to note the impact of the practical stopping conditions proposed in Section 7, since they allow a significant speed up of the algorithm reducing the execution time from 200 sec (stopping after 50 Uzawa's iterations) to 30 sec for images of size 149 × 149 pixels.

8.2. Denoising and deconvolution. We consider in this Section the denoising and deconvolution of irregularly sampled images. For that we include in our image formation model the MTF corresponding to SPOT 5 HRG (High Resolution Geometric) satellite with *Hipermode* sampling [33]. Shortly, Hipermode is a push-broom acquisition mode that uses two shifted bars of sensors to sample on a double-density



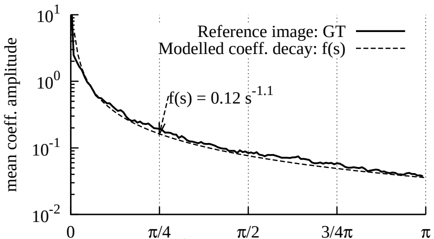
(a) Spectral profiles.

(b) Errors in the restored images.

FIG. 8.3. *Denoising experiments with different selections of spectral profiles $A(i\omega)$. Using the spectral profile that fits the model for the ground truth image, reduces the error, especially in the mid-low frequencies.*

(a) Reference Image.

(b) Noisy, blurred and irregularly sampled image.



(c) Fourier coefficient decay of the reference image.

FIG. 8.4. *Image used in the deconvolution experiment. The Fourier coefficients' amplitudes in the reference image (*257×257 *pixels) decay as* $(2\pi\omega/N)^{-1.1}$.

Table 8.2
Comparison of the algorithms in the irregular to regular sampling and deconvolution task.

Algorithm	**Regularizer**	**RMSE**
ACT+TV [5]		9.035
QN+FAR $\lvert A(i\omega)\rvert = \lvert\frac{2\pi}{N}\omega\rvert^{1.6}$	$K_{\beta,A}(u)$ eq. (4.4)	8.751
QN+TV $\lvert A(i\omega)\rvert = \lvert\frac{2\pi}{N}\omega\rvert^{1}$		8.542
QN+FAR $\lvert A(i\omega)\rvert = \lvert\frac{2\pi}{N}\omega\rvert^{0.4}$		8.779
QN+FAR $\lvert A(i\omega)\rvert \sim \lvert\frac{2\pi}{N}\omega\rvert^{0.4}$ for $\lvert\omega\rvert \geq \frac{N}{4}$ and $\lvert A(i\omega)\rvert = \lvert\frac{2\pi}{N}\omega\rvert^{1}$ for $\lvert\omega\rvert <= \frac{N}{4}$		8.337
regular QN+FAR $\lvert A(i\omega)\rvert = \lvert\frac{2\pi}{N}\omega\rvert^{1.1}$	$K_{\beta,A}(u)$ eq. (4.4)	8.393
L^2-norm of the gradient	$\mathrm{L}_A(u)$ eq. (8.1)	9.112
L^2-norm of $A(D)u$ with $\lvert A(i\omega)\rvert \sim \lvert\frac{2\pi}{N}\omega\rvert^{0.4}$ for $\lvert\omega\rvert \geq \frac{N}{4}$ and $\lvert A(i\omega)\rvert = \lvert\frac{2\pi}{N}\omega\rvert^{1}$ for $\lvert\omega\rvert <= \frac{N}{4}$		9.044

grid. The MTF associated to this system is modeled by

$$\hat{h}(p,q) = \underbrace{\mathrm{sinc}_\pi\left(\frac{2p}{N}\right)\mathrm{sinc}_\pi\left(\frac{2q}{N}\right)e^{-\beta_1\left|\frac{p}{N}\right|}}_{\text{sensor integration blur}}\ \underbrace{e^{-\frac{\alpha_1}{N}\sqrt{p^2+q^2}}}_{\substack{\text{optics \& aper-}\\\text{ture diffraction}}}\ \underbrace{\mathrm{sinc}_\pi\left(\frac{p}{N}\right)}_{\text{motion blur}},\ -\frac{N}{2} < p,q \leq \frac{N}{2},$$

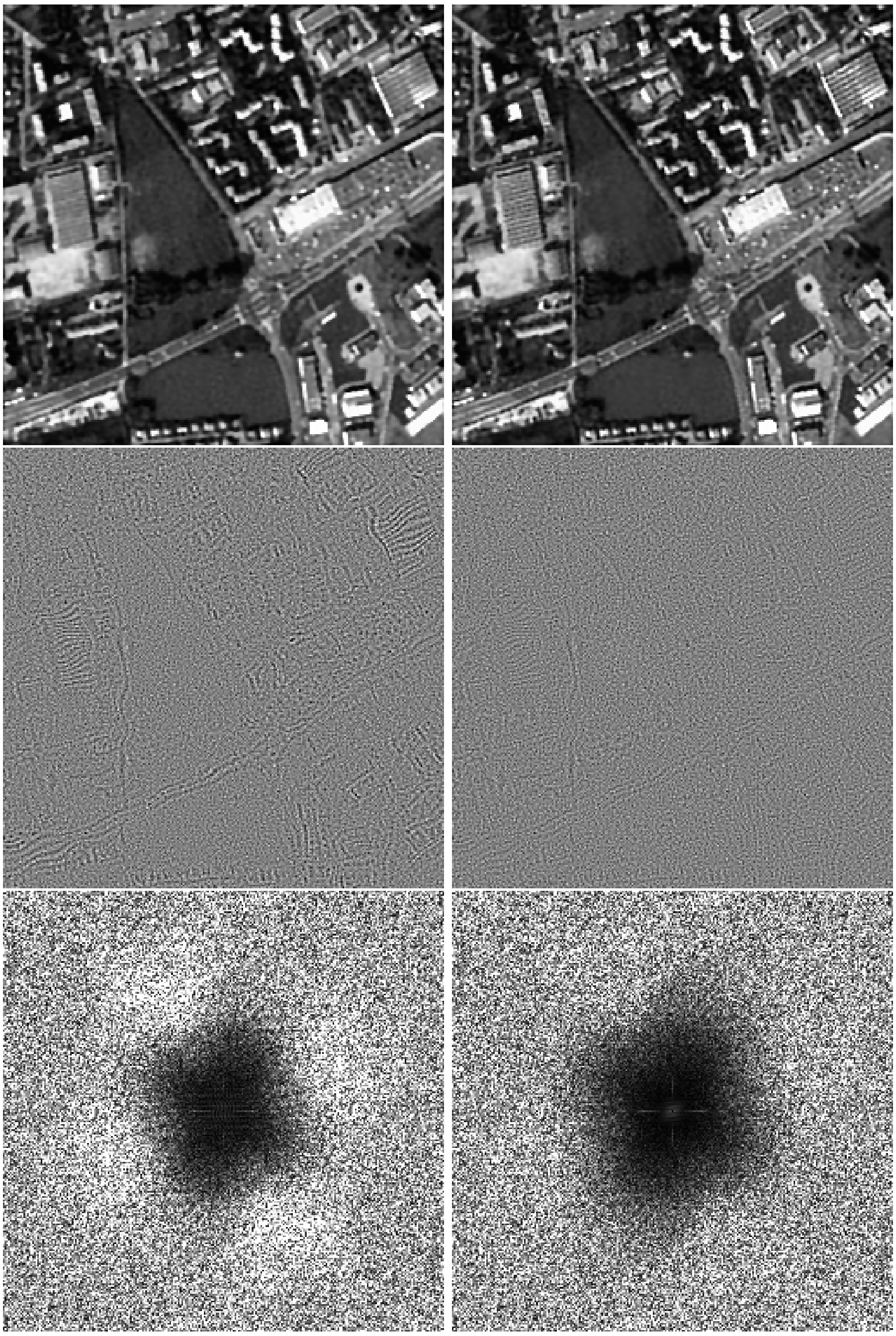
where $\alpha_1 = 3.73$, $\beta_1 = 1.75$, and where $\mathrm{sinc}_\pi(x) = \frac{\sin(\pi x)}{\pi x}$ if $x \neq 0$ and $\mathrm{sinc}_\pi(x) = 1$ if $x = 0$. This function has its first zero crossing at frequency $N/2 = 1/2Hz$, while at frequency $1/4Hz$ the power of the MTF is only 1%, meaning that outside the spectral support $[-N/2, N/2]^2$ (or $[-1/2, 1/2]^2$ Hz) of the filter there is almost no information. To simulate a blurred, noisy and irregularly sampled image with $N \times N$ samples we start by applying the MTF $\hat{h}$ to the image $u_0 \in X_N$. The samples z (Figure 8.4) are then obtained by sampling $h * u_0$ on the irregular grid and adding the noise.

In Figure 8.5 and Table 8.2 we report the results of the restoration experiments corresponding to deconvolution with the Hipermode MTF. The small variability in the results is an indicator that the MTF $\hat{h}$ makes the restoration harder per-se, this is also confirmed by restoring a regularly sampled image and observing that in that case the error is similar to the irregular one (see Table 8.2). Anyway, the proposed algorithm outperforms ACT+TV [5] mainly thanks to the local formulation of the constraints (see Figure 8.6).

We observed that manipulating the spectral profile $A(D)$ does not produce improvements consistent with the denoising case (see Figure 8.7). This is due, on one hand, to the fact that the reference image has a spectral decay different from the target image class ($|\frac{2\pi}{N}\omega|^{1.6}$). Indeed the Fourier coefficients of the reference image decay as ($|\frac{2\pi}{N}\omega|^{1.1}$), which explains the good performance of the total variation (see Table 8.2). On the other hand, in Figure 8.7(b) we see that the total variation controls the error in the low frequency range, but in the high frequency range it is too conservative and does not promote the spectral extrapolation.

The previous observation motivates the following experiment, building a profile that combines, the decay of the total variation in the low frequency range, with a decay similar to $|\frac{2\pi}{N}\omega|^{0.4}$ in the high frequency range. The result of this experiment is shown in Table 8.2 and its profile is depicted in Figure 8.7(b), there we can confirm the desired effect.

Remark 8. The results exhibited in Tables 8.2 and 8.3 confirm that the regularizers based on the minimization of L^2-norm, are not adequate to restore images with



Global Constraint (RMSE=9.035) Local Constraints (RMSE=8.542)

Fig. 8.5. *Restoration with deconvolution computed with ACT+TV [5] (left), and with the proposed algorithm (right) that solves* (4.4) . *In the first row are shown the restored images. In the second row, the method errors, that are re-scaled from [-5,5] (the range of the image is [0,255]). The third row is shown the Fourier transform of the method noise (it should resemble the white noise).*

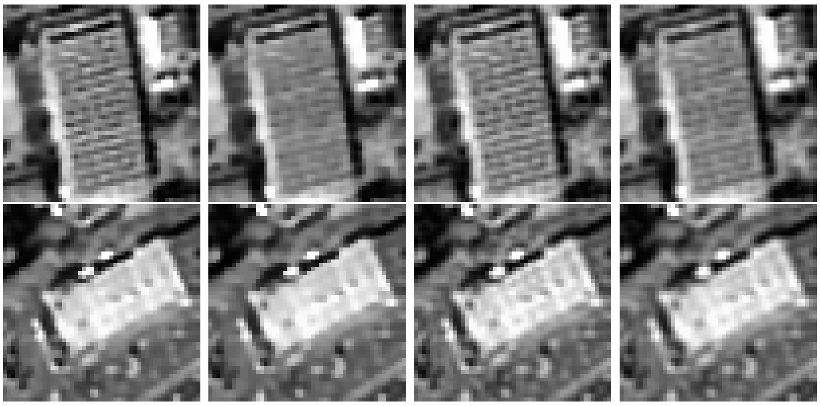
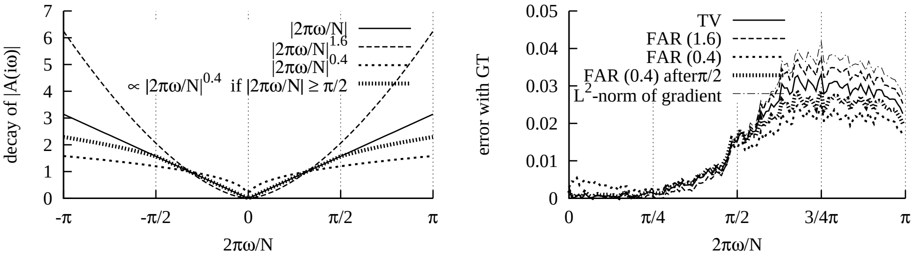


FIG. 8.6. *Two details of images deconvolved using total variation and L^2-norm of the gradient as regularizers, both with local and global constraints. From left to right is shown the result of the proposed algorithm* (4.4)*, ACT+TV [5] (global constraint), the L^2-norm with local constraints and with a global constraint. Note that the last two results are more blurred that the total variation ones, and as seen in Table 8.2 have a higher RMSE.*



(a) Spectral profiles.

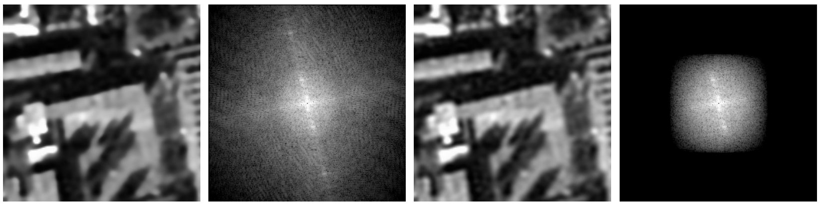
(b) Errors in the restored images.

FIG. 8.7. *Deconvolution experiments with different selections of spectral profiles $A(i\omega)$. Using the spectral profile that fits the model for the ground truth image, reduces the error.*

deconvolution and/or zoom, due to the inability of this type of regularizers to extrapolate the spectrum. This difference is also illustrated in Figure 8.6 and in Figure 8.8 for the case of deconvolution and zoom.

8.3. Extension to zooming. Zooming requires to interpolate and restore the image while preserving and enhancing the shapes, this can be seen as a spectrum extrapolation problem. The basic idea is to fit in as much as this is possible, the low frequency components of the restored and zoomed image to the original data, and to extrapolate the spectrum to the rest of the frequency domain by means of the regularization functional. The regularization allows to recover some high frequencies, which is indeed much more convenient than just filling them with zeros, a technique which is known to produce ringing.

Since the FAR regularizer allows to control the spectral behavior of the solution, in particular the extrapolated part, it will allow to improve the zoom results. Let us first extend the formulation (4.4) to consider the restoration of irregularly sampled



TV result (RMSE=6.079) L^2-norm result (RMSE=6.270)

FIG. 8.8. *Deconvolution and zoom using TV and the L^2-norm of the gradient as regularizers. The images show a detail of the zoomed images and its spectrum. Note that the L^2 regularizer recovers the low frequencies but is unable to extrapolate to higher ones.*

images with a zoom of factor n

$$\begin{aligned} &\min_{u \in X_{nN}} \; K^n_{\beta,A}(u) := \sum_{0 \le r,l < 2nN} \sqrt{\beta^2 + |A(D)Pu\,(r,l)|^2} \quad , \\ &\text{subject to } \left[G * |\Delta_\Xi \left(h * p * u \right) - z|^2 \right] (\xi_k) \le \sigma^2 \quad \forall \xi_k \in \Xi \qquad (8.2) \\ &\text{and} \sum_{0 \le r,l < nN} u\,(r,l) = \sum_{\xi_k \in \Xi} w_k z(\xi_k). \end{aligned}$$

The zoomed and restored image u is a vector of size $nN \times nN$ (we recall that the size of z is $N \times N$), and p is a spectral projector (e.g. $\hat{p} = \chi_R$ or a prolate function) on a low-band region R which depends both on the MTF and the sampling set. In the context we are considering here, $\hat{p}$ will be different from zero in the frequency band corresponding to the resolution of the data $[-1/2, 1/2]^2$ and the constraint is saying that the data is explained by the lower frequency part $h * p * u$ of $h * u$. The regularization functional $K^n_{\beta,A}(u)$ penalizes the oscillations that may appear when we extrapolate the high frequencies in the spectral region $[-n/2, n/2]^2 \setminus [-1/2, 1/2]^2$. Let us mention that, as discussed in [3] 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, if we want to restore and zoom the image u by a factor n, $\hat{p}$ will be different from zero on the region $R = [-1/2, 1/2]^2$ and zero on $[-n/2, n/2]^2 \backslash R$. This is a way to impose that the restored image fits the data z at low frequencies and the high ones are extrapolated via the minimization of $K^n_{\beta,A}(u)$. This minimization problem (8.2) with $\hat{p}(\omega) = \mathbf{1}_{I_N}(\omega)$ is a direct extension of the oversampling and denoising method introduced by Malgouyres and Guichard [40] to the more general case of irregular to regular sampling, deconvolution, denoising and oversampling.

The experiments shown here correspond to a 2X zoom, the images were simulated using the procedure described for the deconvolution case, with a filter $\hat{h}$ extended with zeros up to a double size, and where the irregular sampling is performed at a double spacing. The resulting image is shown in Figure 8.9. Since the restored image and the reference image u_0 have the same size, they can be directly compared. In Figure 8.9 are shown the distorted and the reference image as well as two restorations.

Let us first comment on the stair-casing effect that is noticeable in the bottom left image of Figure 8.9. It is a common observation that the total variation *introduces* a stair-casing effect in the restored images, but let us point out that in our case where

Table 8.3

Deconvolution denoising and 2X zoom experiments.

Algorithm	**Regularizer**	**RMSE**
ACT+TV [5]		6.236
QN+FAR $\|A(i\omega)\| \sim \|\frac{2\pi}{N}\omega\|^{1.6}$ for $\|\omega\| > \frac{N}{4}$	$K_{\beta,A}(u)$ eq (4.4)	5.957
QN+TV $\|A(i\omega)\| = \|\frac{2\pi}{N}\omega\|^{1}$		5.965
QN+FAR $\|A(i\omega)\| \sim \|\frac{2\pi}{N}\omega\|^{0}$ for $\|\omega\| > \frac{N}{4}$ and $\|A(i\omega)\| = \|\frac{2\pi}{N}\omega\|^{1}$ for $\|\omega\| <= \frac{N}{4}$		5.996
L^2-norm of the gradient	$\mathrm{L}_A(u)$ eq. (8.1)	6.270
L^2-norm of $A(D)u$ with $\|A(i\omega)\| \sim \|\frac{2\pi}{N}\omega\|^{1.6}$ for $\|\omega\| \geq \frac{N}{4}$ and $\|A(i\omega)\| = \|\frac{2\pi}{N}\omega\|^{1}$ for $\|\omega\| <= \frac{N}{4}$		6.270

the derivatives are computed analytically this effect may not appear. It appears in the bottom left image of Figure 8.9 due to the poor discretization of the total variation. Notice that the same image processed with a finer approximation as proposed in Section 4.1 does not exhibit this artifact, see bottom right image in Figure 8.9. As we mentioned above, this effect is product of the coarse discretization of the total variation and was negligible in the cases of restoration without zoom. But when zooming it is important to avoid this effect since it produces un-natural looking images in spite of the fact that the RMSE errors of both images are similar.

Since the spectrum is extrapolated the quality of the zoomed image is not affected by the penalty in the frequencies imposed by the regularization term. In Table 8.3 we display the results obtained with different penalization profiles. Notice that the results are better in the case of L^1-norm than L^2-norm.

In contrast with the previous applications, in the present case, not removing all the noise leads to some artifacts. The noise is defined over the original grid (be either regular or irregular), but any residual of the original noise becomes a low frequency colored noise in the zoomed image that is visible as artifacts.

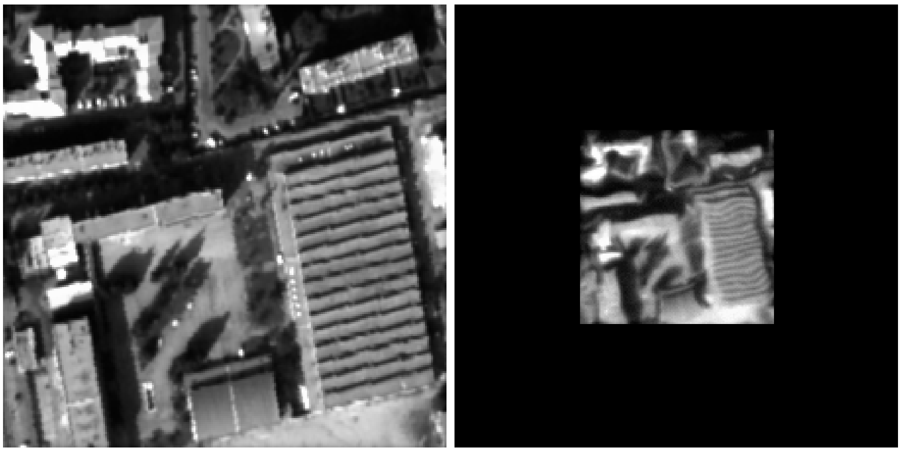
9. Conclusions.

We have proposed a model for the restoration of band limited images that considers irregular (perturbed) sampling, denoising, and deconvolution. In addition a theoretically sound discretization of this model, and an algorithm that solves the resulting discrete optimization problem efficiently using automatically optimized thresholds and stopping conditions for a prescribed output precision level, are presented.

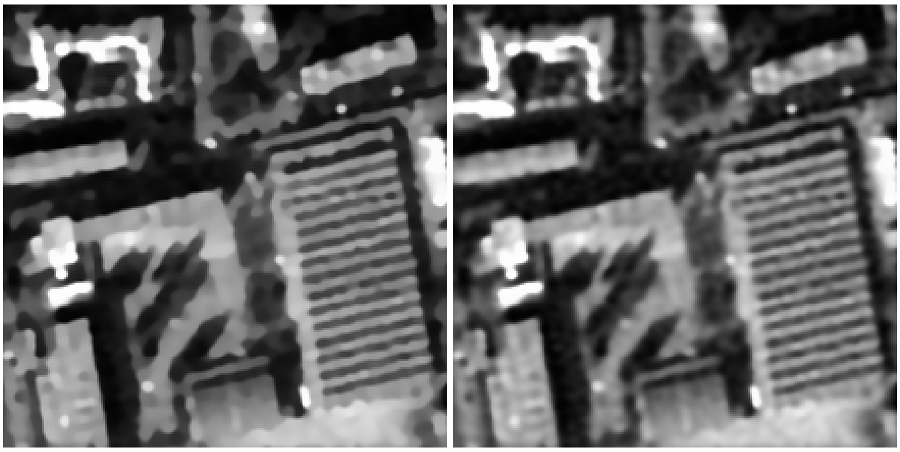
An experimental evaluation (which actually requires all three modeling, discretization, and algorithmical elements) shows that our restoration model, and its accurate discrete approximation, actually improves the performance of previous methods in terms of both signal to noise ratio (Tables 8.1, 8.2 and 8.3) and adequacy of the statistical properties of the computed solution with respect to the assumed image acquisition model. As we can see in Figures 8.2 and 8.5 our approach permits to recover part of the structure lost with a global constraint leading to a method noise that resembles more the postulated gaussian white noise. As observed by Nikolova in [34], most denoising methods tend to produce a method noise with a distribution which largely differs from the postulated gaussian white noise hypothesis.

The key ingredients that make our restoration model (Section 3) produce good results are the combination of

1. the modeling of the image acquisition system through local constraints as in [4, 5],
2. the use of a special kind of *frequency adapted regularization* in the spirit of



Reference image Irregular samples



Zoomed with QN+FAR using (4.2) Zoomed with QN+FAR using (4.4)

FIG. 8.9. *Deconvolution denoising and zoom. We display the QN+FAR results using a spectral decay like:* $(\frac{2\pi\omega}{N})^{1.6}$*, which coincides with the decay of the reference image. The stair-casing effect in* (c) *(RMSE: 6.003) is due to the poor discretization of the regularizer in* (4.2)*. In* (d) *(RMSE: 5.957) we observe the result obtained with a finer discretization of the regularizer* (4.4)*.*

the ACT algorithm proposed by H.G. Feichtinger, K. Gröchenig, M. Rauth and T. Strohmer [32], and

3. the minimization of an L^1 (instead of L^2-norm) of this frequency-adapted pseudo-differential operator. However the L^2-norm produces competitive results in the case of denoising without deconvolution nor zoom (see Table 8.1).

All three aspects together collaborate to provide a better model of both, image acquisition and a priori image regularity, that leads to better accordance of the statistical properties of our solution with respect to the postulated models and the reference image. An accurate discrete approximation to this model (Section 4), both allows the use of convenient and efficient numerical algorithms (Section 6), and leads to well-behaved solutions (Section 5). Finally the study of the local constraints (Section 7) leads to a trade-off between the precision of the noise estimation and the number of constraints that can be ensured, that is used to derive the stopping criterion for the different stages of the algorithm.

Experiments (Section 8) highlighting the importance of the modeling and discretization aspects described earlier are presented, namely: the quality boost provided by the local formulation of the constraints (Figure 8.2), the importance of the discretization of the regularizer (Figure 8.9) and the usefulness of the frequency adapted regularizer in favoring certain spectral behaviors of the solutions (Figure 8.2 and 8.5).

Our prospects for future work include adapting our algorithm to use splines as the underlying interpolation model, and the introduction of anti-aliasing filters to handle the case of aliased data. The exploitation of self-similarity proporties of most natural images would also be useful to increase resolution while reducing noise and aliasing effects.

Acknowledgement. Vicent Caselles acknowledges partial support by the PNPGC project, reference MTM2006-14836. Gabriele Facciolo and Andrés Almansa acknowledge support by the ALFA project, No. AML//19.0902/97/0666/II-0366-FA. Andrés Almansa acknowledges support from CNES contracts R-S06/ OT-001-040 and R-S06/ OT-004-024 and by ECOS-Sud project U06E01. Andrés Almansa and Jean-François Aujol acknowledge support by the French “Agence Nationale de la Recherche” (ANR), under grant FREEDOM (ANR07-JCJC-0048-01), “Films, REstauration Et DOnnées Manquantes”.

REFERENCES

[1] R. Acar and C. Vogel, *Analysis of total variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229.

[2] A. Aldroubi and K. Gröchenig, *Non-uniform sampling and reconstruction in shift-invariantspaces*, SIAM Rev., 43 (2001), pp. 585–620.

[3] A. Almansa, *Echantillonnage, Interpolation et Détection. Applications en Imagerie Satellitaire*, Ph.D. thesis, Ecole Normale Supérieure de Cachan, France, 2002.

[4] A. Almansa, C. Ballester, V. Caselles, and G. Haro, *A TV based restoration model with local constraints*, J. Sci. Comput., 34 (2008), pp. 209–236.

[5] A. Almansa, V. Caselles, G. Haro, and B. Rougé, *Restoration and zoom of irregularly sampled, blurred and noisy images by accurate total variation minimization with local constraints*, Multiscale Model. Simul., 5 (2006), pp. 235–272.

[6] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Math. Monogr., Oxford Science Publications, Oxford, 2000.

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

[8] M. Arigovindan, M. Suhling, P. Hunziker, and M. Unser, *Variational image reconstruction from arbitrarily spaced samples: a fast multiresolution spline solution*, IEEE Trans. Image Process., 14 (2005), pp 450–460.

[9] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing*, Applied Mathematical Sciences, Vol 147, Springer-Verlag, 2002.

[10] J. F. Aujol and G. Gilboa, *Constrained and SNR-based Solutions for TV-Hilbert Space Image Denoising*, J. Math. Imaging Vision, 26 (2006), pp. 217–237.

[11] J.F. Aujol, G. Gilboa, T. Chan and S. Osher, *Structure-Texture Image Decomposition – Modeling, Algorithms, and Parameter Selection*, Int. J. Comput. Vision, 67 (2006), pp. 111–136.

[12] P. Azmi and F. Marvasti, *Comparison between several iterative methods of recovering signals from nonuniformly spaced samples*, Sampl. Theory Signal Image Process., 1 (2002), pp. 207–224.

[13] M. Bertalmío, V. Caselles, B. Rougé, and A. Solé, *TV based image restoration with local constraints*, J. Sci. Comput., 19 (2003), pp. 95–122.

[14] L. Blanc-Féraud, P. Charbonnier, G. Aubert and M. Barlaud, *Nonlinear image processing: modeling and fast algorithm for regularization with edge detection*, in Proceedings of the IEEE International Conference on Image Processing, (1995), pp. 474–477.

[15] H. Brezis, *Operateurs Maximaux Monotones*, North Holland, Amsterdam, 1973.

[16] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.

[17] A. Chambolle and P.L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.

[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 and P. Mulet, *On the convergence of the Lagged Diffusivity Fixed Point Method in total variation image restoration*, SIAM J. Numer. Anal., 36 (1999), pp. 354–367.

[20] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, *Deterministic edge-preserving regularization in computed imaging*, IEEE Trans. Image Process., 6 (1997), pp. 298–311.

[21] P.G. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation*, Dunod, Paris, 1998.

[22] G. Dahlquist and A. Bjorck, *Numerical Methods*, Dover Publications, Incorporated, 2003.

[23] D. Dobson and C. Vogel, *Convergence of an iterative method for total variation denoising*, SIAM J. Numer. Anal., 34 (1997), pp. 1779–1791.

[24] S. Durand, F. Malgouyres and B. Rougé, *Image Deblurring, Spectrum Interpolation and Application to Satellite Imaging*, ESAIM Control Optim. Calc. Var., 5 (2000), pp. 445–475.

[25] L.C. Evans and R.F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, Boca Raton, FL, 2000.

[26] P. Faurre, *Analyse numérique. Notes d'optimisation*, École Polytechnique. Ed. Ellipses, 1988.

[27] H.G. Feichtinger, C. Cenker, and M. Herrmann, *Iterative algorithms in irregular sampling: A first comparison of methods*, in Proceedings of the IEEE International Conference on Communication, Computer and Power, (1991), pp. 483–489.

[28] C. Gasquet and P. Witomski, *Analyse de Fourier et Applications*, Ed. Masson, 1990.

[29] D. Geman and G. Reynolds, *Constrained Image Restoration and Recovery of Discontinuities*, IEEE Trans. Pattern Anal. Machine Intell., 14 (1992), pp. 367–383.

[30] G. Gilboa, N. Sochen, and Y.Y. Zeevi, *Texture Preserving Variational Denoising Using an Adaptive Fidelity Term*, 2nd IEEE Workshop on Variational, Geometric, and Level Set Methods in Computer Vision, (2003), pp. 137–144.

[31] G. Gilboa, N. Sochen, and Y.Y. Zeevi, *Variational denoising of partly-textured images by spatially varying constraints*, IEEE Trans. Image Process., 15 (2006), pp. 2281–2289.

[32] K. Gröchenig and T. Strohmer, *Numerical and theoretical aspects of non-uniform sampling of band-limited images*, in Theory and Practice of Nonuniform Sampling, F. Marvasti, editor, Kluwer/Plenum, 2000.

[33] O. Hagolle, J.-M. Martinuzzi and C. Latry, *How to double the spatial resolution of a push-broom instrument*, International Geoscience and Remote Sensing Symposium. IGARSS '94. Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation, 3 (1994), pp. 1553–1555.

[34] M. Nikolova, *Model distortions in bayesian map reconstruction*, Inverse Probl. Imaging, 1 (2007), pp. 399–422.

[35] A. Jalobeanu, L. Blanc-Feraud, and J. Zerubia, *Estimation of blur and noise parameters in remote sensing*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, (2002), pp. 249–256.

[36] M.I. Kadec, *The exact value of the Paley-Wiener constant*, Sov. Math. Dokl., 5 (1964), pp. 559–561.

[37] H.J. Landau, *Necessary density conditions for sampling and interpolation of certain entire fuctions*, Acta Math., 117 (1967), pp. 37–52.

[38] S. Lintner and F. Malgouyres, *Solving a variational image restoration model which involves L^∞ constraints*, Inverse Problems, 20 (2004), pp. 815–831.

[39] L. Moisan, *How to discretize the Total Variation of an image?*, in Proceedings of the International Congress on Industrial and Applied Mathematics, (2007).

[40] F. Malgouyres and F. Guichard, *Edge direction preserving image zooming: a mathematical and numerical analysis*, SIAM J. Numer. Anal., 39 (2001), pp. 1–37.

[41] D. Potts, G. Steidl, and M. Tasche, *Fast Fourier transforms for nonequispaced data: A tutorial*, in Modern Sampling Theory: Mathematics and Applications, J.J. Benedetto and P. Ferreira (Eds.),Chapter 12, pp. 249–274, 2001.

[42] M. Rauth, *Gridding geophysical potential fields from noisy scattered data*, Ph.D. thesis, University of Vienna, 1998.

[43] L. Rudin and S. Osher, *Total Variation based Image Restoration with Free Local Constraints*, in Proc. of the IEEE International Conference on Image Processing, (1994), pp. 31–35.

[44] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*,

Physica D, 60 (1992), pp. 259–268.

[45] O. Scherzer and M. Grasmair and H. Grossauer and M. Haltmeier and F. Lenzen, *in Variational Methods in Imaging (Applied Mathematical Sciences)*, Springer, Chapter 5, 2008.

[46] J.R. Schott, *Remote Sensing: The Image Chain Approach*, Oxford University Press, 1997.

[47] A.N. Tikhonov and V.Y. Arsenin, *Solutions of Ill-Posed Problems*, V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York, 1977.

[48] N. Trudinger, *On the analyticity of generalized minimal surfaces*, Bull. Austral. Math. Soc., 5 (1971), pp. 315–320.

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