J Math Imaging Vis 26: 217–237, 2006

DOI: 10.1007/s10851-006-7801-6

Constrained and SNR-Based Solutions for TV-Hilbert Space Image Denoising

JEAN-FRANÇOIS AUJOL
Department of Mathematics, UCLA, Los Angeles, California 90095, USA
CMLA, UMR CNRS 8536, ENS Cachan, France

GUY GILBOA
Department of Mathematics, UCLA, Los Angeles, California 90095, USA

Published online: 14 August 2006

Abstract. We examine the general regularization model which is based on total-variation for the structural part and a Hilbert-space norm for the oscillatory part. This framework generalizes the Rudin-Osher-Fatemi and the Osher-Sole-Vese models and opens way for new denoising or decomposition methods with tunable norms, which are adapted to the nature of the noise or textures of the image. We give sufficient conditions and prove the convergence of an iterative numerical implementation, following Chambolle's projection algorithm.

In this paper we focus on the denoising problem. In order to provide an automatic solution, a systematic method for choosing the weight between the energies is imperative. The classical method for selecting the weight parameter according to the noise variance is reformulated in a Hilbert space sense. Moreover, we generalize a recent study of Gilboa-Sochen-Zeevi where the weight parameter is selected such that the denoised result is close to optimal, in the SNR sense. A broader definition of SNR, which is frequency weighted, is formulated in the context of inner products. A necessary condition for maximal SNR is provided. Lower and upper bounds on the SNR performance of the classical and optimal strategies are established, under quite general assumptions.

Keywords: image restoration, BV, H^{-1}, hilbert space, SNR, projection, total-variation, denoising

1. Introduction

Regularization of images by variational methods has shown to achieve impressive results and is today an increasing field of interest in image processing. In this paper we are concerned with the classical denoising problem of image degraded by additive white noise. We assume that the input image f is composed of the original image s and additive uncorrelated noise n with zero mean and of variance σ^2:

$$f = s + n. \tag{1}$$

The aim is to find a decomposition u, v such that u approximates the original signal s and v is the residual part of f:

$$f = u + v. \tag{2}$$

Our regularization is based on finding u that minimizes the following energy

$$E(u, f) = \int_\Omega |Du| + \frac{\lambda}{2}\|f - u\|_{\mathcal{H}}^2, \tag{3}$$

where the left term on the right-hand-side is the total-variation energy and the right term is the square norm

of a Hilbert space $\mathcal{H}$. The specific definition of the spaces appears in Section 2.

In the following we explain the basic concept of variational denoising and review the recent main contributions. Classically, there are two basic measures, often referred to as energy terms, that are to be jointly minimized:

$$E(u, f) = E_{smooth}(u) + \lambda E_{fidelity}(u, f). \qquad (4)$$

E_{smooth} is a smoothing term which rewards smooth signals and penalizes oscillatory ones. $E_{fidelity}$ accounts for fidelity, or closeness, to the input image f. This type of formula dates back to Tikhonov regularization [45] (with an identity blur operator). In the image-processing field, however, the more successful approaches are based on nonlinear methods. A main contribution made by Rudin, Osher and Fatemi (ROF) [42] was to consider the total-variation energy, which does not penalize sharp edges over any other monotone signals, thus allowing piecewise smooth solutions, which considerably reduces image blurring in the denoising process. Most of the research in the 90's was focused on the smoothness term. One should note that similar results were obtained at the time by Perona and Malik using a closely related PDE method of nonlinear diffusion processes [40]. However, the method was not related to a norm, and in the original paper was not convex. Many variations of such smoothness terms followed. The fidelity term, though, was mostly based on the L^2 norm.

Following Meyer's work [32] the attention to the role of the fidelity term has increased. In [32], the author has analyzed the mathematical properties of the ROF model [42]. He has suggested the use of other functional spaces which would suit more the oscillating patterns of an image (and which would thus capture the noise more efficiently). This has led to new image decomposition and denoising algorithms. The first work in this direction was [46], followed by [38, 4, 5, 9, 44, 21, 36]. [32] has also raised new theoretical issues [2, 37, 35, 30, 20].

Also, other types of fidelity terms for impulsive noise models were suggested [34, 14].

One significant contribution of this paper is to suggest a large family of fidelity terms, which can be designed for various purposes and noise models by selecting a proper linear operator. Once it is chosen, this framework guarantees a unique solution, offers the numerical way to implement the regularization, and suggests how to select the weight parameter between the energies.

Another important matter is the way the solution of (3) is computed. In the standard method, one derives the associated Euler-Lagrange equations, embeds them into a dynamical scheme which is iterated to a steady-state. A more accurate way to compute the ROF solution is to use dual formulations [15, 11] or projections onto convex sets [18, 17]. Recently, Chambolle has proposed a projection algorithm based on duality to solve the ROF problem [12]. In [5] the authors have proposed a modification of the projection algorithm to solve the Osher-Sole-Vese (OSV) problem [38]. In this paper, we generalize Chambolle's projection algorithm to a large class of functionals. For a new computational approach based on second-order cone programming see [49].

For a detailed overview of PDE-based restoration methods see [3]. For a recent review on denoising methods we refer the reader to [10].

We focus our attention on finding the weight parameter λ, an important component of the basic regularization equation (4). By minimizing both terms of (4) we seek a compromise between a non-oscillatory solution and one which is "close enough" to the original image. Any minimization of one of the terms by itself leads to degenerate solutions which are not interesting (a constant or the input noisy image). The appropriate compromise then highly depends on λ, the weight between these two energies. When it is too low, the restored image is over-smoothed. When it is too high, u still contains too much noise. Finding the right value of λ is therefore an important part of solving the denoising problem.

Less attention was given to this aspect of the problem in variational image processing, and many times the parameter was chosen manually by trial and error. Aiming at achieving automatic denoising algorithms, systematic methods for choosing λ are required. A similar problem has been investigated in regularization theory, in the context of operator inversion by Tikhonov-type methods. As we are concerned with denoising of images (therefore our operator is the identity and the regularization preserves edges) many of the techniques do not apply for our purpose. In this paper we generalize for our framework two methods, that fit image denoising. The first one is very classical, came from regularization theory and was used in [42]. The second one is very recent, less simple but more accurate, and was proposed for an L^2 fidelity in [28, 29].

Many statistical approaches have been proposed (see [22, 3] and references herein), but in this paper we focus on deterministic approaches. The classical algorithm for choosing λ dates back to Morozov

discrepancy principle [33] and for the $TV - L^2$ minimization is often referred to as the ROF constrained formulation [42]. See [19] for a general approach on the use of noise properties in set theoretic estimation. For approaches such as generalized cross validation and the L-curve, we refer to Vogel's survey of regularization parameter selection methods in [47]. We give a generalization to the ROF constrained formulation for the broader minimization problem:

$$\inf_u \int_\Omega |Du| \text{ subject to } \|f - u\|_{\mathcal{H}}^2 = |\Omega|\rho_{\mathcal{H}}^2, \qquad (5)$$

where $\rho_{\mathcal{H}}^2$ is the normalized square $\mathcal{H}$-norm ("$\mathcal{H}$-variance") of the noise. A generalization of Chambolle's projection algorithm is used to compute the solution. λ is being iteratively updated such that the $\mathcal{H}$-norm of $v = f - u$ equals that of the noise. To this end, we need to compute the $\mathcal{H}$-norm of white Gaussian noise from its standard deviation σ. Notice that the constrained problem has recently been addressed in [5] in the case of Meyer's G norm. The closely related problem of criteria for the stopping time of nonlinear diffusions was examined by [39] and to some extent by [48]. Physical considerations for solving the stopping time problem for the visco-plastic fluid model were suggested in [24].

The underlying assumption of the classical algorithm is that the denoising process works well, therefore what is filtered is mostly noise: $v \approx n$. A natural condition is then to impose $\text{var}(v) = \text{var}(n) = \sigma^2$. When the image is partly textured, though, parts of the textures are also filtered out and v contains both noise and texture. Imposing the above condition in these cases often causes oversmoothing of textures. The criteria of [39] and [48] rely on similar assumptions and "confuse" texture with noise, as shown in [28]. Gilboa-Sochen-Zeevi (GSZ) [28, 25] addressed this problem recently. Their method was to base the selection of λ on the Signal to Noise Ratio (SNR) criterion. The optimal solution was defined as the one that maximizes the SNR. A necessary condition for optimal SNR was formulated. Note that this condition is equivalent to minimizing the L^2 error. This condition was then estimated, reaching close results to the optimal solution. This method can work also on textured images, when the denoising is not very good. In our examples we apply the algorithms to the OSV model, where currently the authors are not aware of any suggested mechanism for selecting λ.

Under some general assumptions related to the denoising process and the non-correlation of signal and noise, GSZ provide bounds for the $TV - L^2$ model. We generalize these results, both with respect to the maximum SNR estimations and with respect to the SNR performance bounds.

The plan of the paper is as follows: We first introduce notations in Section 2. We propose a generalization of Chambolle's projection algorithm to solve (3) in Section 3. We can then generalize Chambolle's approach [12] for solving the constrained problem (5) in Section 4. This provides us with a new automatic restoration algorithm based on the variance of the noise. We propose another automatic restoration algorithm based on SNR like optimum condition in Section 5. This improves the numerical results of Section 4. In Section 6 we provide theoretical estimates on the SNR performance of the methods. Experimental results comparing the two selection criteria for λ are presented in Section 7. We conclude the paper with some final remarks in Section 8.

2. Notations

In this section we introduce the main definitions and mathematical spaces that are used in the paper.

2.1. L^2 Inner Product

Definition 1. In this paper, we consider only the discrete case (for the sake of clarity). The image is a two dimension vector of size $N \times N$. We denote by X the Euclidean space $\mathbb{R}^{N\times N}$. The space X is endowed with the L^2 inner product: $\langle u, v\rangle_{L^2} = \sum_{1\le i,j\le N} u_{i,j}v_{i,j}$, and the norm $\|u\|_{L^2} = \sqrt{\langle u, u\rangle_{L^2}}$. We shall often consider the following subspace of X:

$$X_0 = \{x \in X \,/\, \sum_{i,j} x_{i,j} = 0\} \qquad (6)$$

2.1.2. Discrete Fourier Transform We recall that the DFT of a given discrete image $(f(m,n))$ $(0 \le m \le N-1$ and $0 \le n \le N-1)$ is given by $(0 \le p \le N-1$ and $0 \le q \le N-1)$:

$$\begin{aligned}&\mathcal{F}(f)(p,q) = F(p,q)\\ &= \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} f(m,n)e^{-\jmath(2\pi/N)pm}e^{-\jmath(2\pi/N)qn}\end{aligned} \qquad (7)$$

and the inverse transform is:

$$f(m,n) = \frac{1}{N^2}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1} F(p,q)e^{J(2\pi/N)pm}e^{J(2\pi/N)qn} \tag{8}$$

Moreover, we also have $\|\mathcal{F}(f)\|_{L^2}^2 = N^2\|f\|_{L^2}^2$ and $\langle \mathcal{F}(f), \mathcal{F}(g)\rangle_{L^2} = N^2\langle f, g\rangle_{L^2}$.

We have $\Delta f(m,n) = f(m+1,n) + f(m-1,n) + f(m,n+1) + f(m,n-1) - 4f(m,n)$. Standard computations lead to:

$$\begin{aligned}&\mathcal{F}(\Delta f)(p,q)\\ &= 2\left(\cos\left(\frac{2\pi}{N}p\right) + \cos\left(\frac{2\pi}{N}q\right) - 2\right)\mathcal{F}(f)(p,q)\end{aligned} \tag{9}$$

We deduce that, if f has zero mean, then for $(p,q) \neq (0,0)$

$$\begin{aligned}&\mathcal{F}(\Delta^{-1} f)(p,q)\\ &= \frac{1}{2\left(\cos\left(\frac{2\pi}{N}p\right) + \cos\left(\frac{2\pi}{N}q\right) - 2\right)}\mathcal{F}(f)(p,q)\end{aligned} \tag{10}$$

These basic results are proved to be useful when applied to the H^{-1} case of section 2.3.2.

2.1.3. $\mathcal{H}$ Hilbert Space In what follows, we will consider a general family of Hilbert spaces. We consider K a linear symmetric positive operator from A to L^2, where A is either X_0 or L^2 (we recall that X_0 is defined by (6)). In the case when $A = X_0$, then we extend K to the whole L^2 by setting $K(x) = +\infty$ if $x \in L^2 \backslash X_0$. Notice that under these assumptions we can define K^{-1} on $ImK = \{z \in L^2$ such that $\exists x \in A$ with $z = K(x)\}$.

If f and g are in X_0, then let us define:

$$\langle f, g\rangle_{\mathcal{H}} = \langle f, Kg\rangle_{L^2} \tag{11}$$

This defines an inner product on $X_0 = \{x \in X \,/\, \sum_{i,j} x_{i,j} = 0\}$.

Examples

1. When $K = Id$, then $\mathcal{H} = L^2$.
2. When $K = -\Delta$, then $\mathcal{H} = H$ (see subsection 2.3.1).
3. When $K = -\Delta^{-1}$, then $\mathcal{H} = H^{-1}$ (see subsection 2.3.2).

For other choices of K (based on Gabor functions), and an application to image decomposition, we refer the reader to [8, 7].

2.2. Total Variation Regularization

Up to now, we have focused only on the fidelity term. Here we are interested in the smoothing term.

Definition 1. To define a discrete total variation, we introduce a discrete version of the gradient operator. ∇u is given by: $(\nabla u)_{i,j} = ((\nabla u)^1_{i,j}, (\nabla u)^2_{i,j})$ with

$$(\nabla u)^1_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < N \\ 0 & \text{if } i = N \end{cases} \tag{12}$$

and

$$(\nabla u)^2_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } j < N \\ 0 & \text{if } j = N \end{cases} \tag{13}$$

The discrete total variation of u is then defined by:

$$J(u) = \sum_{1\le i,j\le N} |(\nabla u)_{i,j}| \tag{14}$$

We also introduce a discrete version of the divergence operator. We define it by analogy with the continuous setting by $\text{div} = -\nabla^*$ where ∇^* is the adjoint of ∇: that is, for every $p \in X \times X$ and $u \in X$, $\langle -\text{div}\, p, u\rangle_{L^2} = \langle p, \nabla u\rangle_{L^2}$. It is easy to check that:

$$\begin{aligned}(\text{div}\,(p))_{i,j} = &\begin{cases} p^1_{i,j} - p^1_{i-1,j} & \text{if } 1 < i < N \\ p^1_{i,j} & \text{if } i = 1 \\ -p^1_{i-1,j} & \text{if } i = N \end{cases}\\ + &\begin{cases} p^2_{i,j} - p^2_{i,j-1} & \text{if } 1 < j < N \\ p^2_{i,j} & \text{if } j = 1 \\ -p^2_{i,j-1} & \text{if } j = N \end{cases}\end{aligned} \tag{15}$$

By analogy with the continuous setting, we define a discrete Laplacian operator by setting $\Delta u = \text{div}\,(\nabla u)$. From now on, we will use these discrete operators.

2.2.1. G Space We are now in position to introduce the discrete version of Meyer's space G [32, 4, 2].

Definition 1.

$$G = \{x \in X \,/\, \exists g \in X \times X \text{ such that } x = \operatorname{div}(g)\} \tag{16}$$

and if $x \in G$:

$$\begin{aligned}\|x\|_G = \inf\{\|g\|_\infty \,/\, & x = \operatorname{div}(g),\\ & g = (g^1, g^2) \in X \times X, |g_{i,j}|\\ & = \sqrt{(g^1_{i,j})^2 + (g^2_{i,j})^2}\Big\}\end{aligned} \tag{17}$$

where $\|g\|_\infty = \max_{i,j} |g_{i,j}|$. Moreover, we will denote:

$$G_\mu = \{x \in G \,/\, \|x\|_G \le \mu\} \tag{18}$$

Remark The following result is proved in [4]:

Proposition 1. *The space G identifies with the following subspace:* $X_0 = \{v \in X \,/\, \sum_{i,j} v_{i,j} = 0\}$.

2.2.2. Convex Analysis We recall that the Legendre-Fenchel transform of J is [23, 12]:

$$J^*(v) = \sup_{u \in X} (\langle u, v\rangle_{L^2} - J(u)) \tag{19}$$

Since J defined by (14) is homogeneous of degree one (i.e. $J(\lambda u) = \lambda J(u)\ \forall u$ and $\lambda > 0$), it is then standard (see [23]) that J^* is the indicator function of some closed convex set, which turns out to be the set G_1 defined by (18):

$$J^*(v) = \chi_{G_1}(v) = \begin{cases} 0 & \text{if } v \in G_1 \\ +\infty & \text{otherwise} \end{cases} \tag{20}$$

This result is the key to Chambolle's projection algorithm [12], and it had first been noticed in [13].

We close this section by giving examples of classical Hilbert spaces which are in the class of $\mathcal{H}$.

2.3. H and H^{-1} Inner Product

2.3.1. H Space We use the following norm: $\|\nabla u\|_{L^2} = \sqrt{\sum_{1 \le i,j \le N} |\nabla u_{i,j}|^2}$. We can now introduce the H norm:

$$\|u\|_H = \|\nabla u\|_{L^2} \tag{21}$$

This is a norm on the space $X_0 = \left\{u \in X, \sum_{i,j} u_{i,j} = 0\right\}$. It is associated with the inner product: $\langle f, g\rangle_H = \langle f, -\Delta g\rangle_{L^2}$.

2.3.2. H^{-1} Space We consider the polar semi-norm associated with (21):

$$\|v\|_{H^{-1}} = \sup_{\|u\|_H = 1} \langle v, u\rangle_{L^2} = \sup_{\|\nabla u\|_{L^2} = 1} \langle v, u\rangle_{L^2} \tag{22}$$

This is a discrete version of the H^{-1} norm. The following result is proved in [5]:

$$\|f\|_{H^{-1}} = \sqrt{\langle -f, \Delta^{-1} f\rangle_{L^2}} \tag{23}$$

Using Parseval identity, one sees that:

$$\begin{aligned}\|f\|^2_{H^{-1}} &= \tfrac{1}{N^2} \langle -\mathcal{F}(f), \mathcal{F}(\Delta^{-1} f)\rangle_{L^2}\\ &= \tfrac{1}{N^2} \sum_{(p,q) \ne (0,0)} \frac{1}{2\left(2 - \cos\left(\frac{2\pi}{N} p\right) - \cos\left(\frac{2\pi}{N} q\right)\right)} (\mathcal{F}(f)(p,q))^2\end{aligned} \tag{24}$$

We can thus define a inner product on H^{-1} by setting:

$$\langle f, g\rangle_{H^{-1}} = (-f, \Delta^{-1} g)_{L^2} \tag{25}$$

Frequency Understanding of H^{-1}

1. Using (24), one sees that the H^{-1} norm differs from the L^2 norm by the fact that the frequencies are weighted: therefore, H^{-1} owes much more importance to the low frequencies. This is the reason why an oscillating pattern has a small H^{-1} norm (as shown in [32] in a more general framework). See also Figure 1 for an intuitive idea of H^{-1} filtering in 1 dimension. Although naive, this explains why the OSV model [38] is very good at denoising piecewise constant images. This is also the reason why the OSV model cannot split an image into its geometric component and its texture component when the texture contains some low frequencies. These two properties have been checked numerically in [5]. Fig. 2
2. The H^{-1} inner product is easily computed thanks to the discrete Fourier transform.

Now that we have presented the different notations, we are in position to introduce our model and a method to solve it.

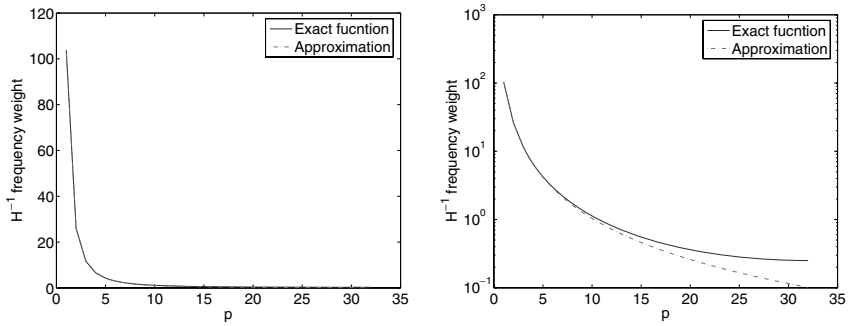

Figure 1. Frequency weight of the H^{-1} norm in one dimension. The exact weight function is $\frac{1}{2(1-\cos(2\pi p/N))}$, which can be approximated, using Taylor expansion, by $\frac{1}{(2\pi p/N)^2}$. On the left both functions are plotted in linear scale. As the difference is quite small, a log scale plot is shown on the right. It is apparent that the approximation is quite accurate for the low frequency range. For this graph we used $N = 64$.

3. A Projection Algorithm

In this section, we are interested in solving the following problem.

$$\inf_u \left(J(u) + \frac{\lambda}{2}\|f-u\|_{\mathcal{H}}^2 \right) \tag{26}$$

All the results of this section have already been proved in the case $\mathcal{H} = L^2$ in [12], and we will draw our inspiration from this paper.

Proposition 2. *Problem (26) admits a unique solution $\hat{u}$.*

Proof: This is a very standard result [13]. The existence comes from the convexity of the functional, and the uniqueness from the fact that K is positive on X_0. □

Although the next result can be seen as an application of the Fenchel-Rockafellar duality formula (see [41]), we give below an elementary proof.

Proposition 3. *If $\hat{u}$ is the solution of problem (26), then $\hat{v} = f - \hat{u}$ is the solution of the dual problem:*

$$\inf_v \left(\|v-f\|_{\mathcal{H}}^2 + \frac{1}{\lambda} J^*(\lambda K v) \right) \tag{27}$$

Proof: We first recall that $\|f-u\|_{\mathcal{H}}^2 = \langle f-u, K(f-u)\rangle_{L^2}$. If $\hat{u}$ is a minimizer of (26), then: $0 \in \lambda K(\hat{u}-f) + \partial J(\hat{u})$, i.e.: $\lambda K(f-\hat{u}) \in \partial J(\hat{u})$. Hence $\hat{u} \in \partial J^*(\lambda K(f-\hat{u}))$. We then set $\hat{w} = K(f-\hat{u})$, and we get:

$$0 \in K^{-1}\hat{w} - f + \partial J^*(\lambda \hat{w}) \tag{28}$$

We then deduce that $\hat{w}$ is the minimizer of:

$$\inf_w \left(\left\|K^{-1}w\right\|_{\mathcal{H}}^2 - 2\langle f, w\rangle_{L^2} + \frac{1}{\lambda} J^*(\lambda w) \right) \tag{29}$$

Since $\langle f, w\rangle_{L^2} = \left\langle f, K^{-1}w\right\rangle_{\mathcal{H}}$, we have:

$$\left\|K^{-1}w\right\|_{\mathcal{H}}^2 - 2\langle f, w\rangle_{L^2} = \left\|K^{-1}w - f\right\|_{\mathcal{H}}^2 - \|f\|_{\mathcal{H}}^2 \tag{30}$$

Thus $\hat{w}$ is the minimizer of:

$$\inf_w \left(\left\|K^{-1}w - f\right\|_{\mathcal{H}}^2 + \frac{1}{\lambda} J^*(\lambda w) \right) \tag{31}$$

We now set $\hat{v} = K^{-1}(\hat{w}) = f - \hat{u}$. we therefore get that $\hat{v}$ is a minimizer of (27). □

Since J^* is given by (20), we deduce that $\hat{v} = P^{\mathcal{H}}_{K^{-1}G_{1/\lambda}}(f)$, where $P^{\mathcal{H}}_{K^{-1}G_{1/\lambda}}(f)$ is the orthogonal projection of f over $K^{-1}G_{1/\lambda}$ with respect to the $\mathcal{H}$ inner product. Hence, the solution $\hat{u}$ of problem (26) is simply given by:

$$\hat{u} = f - P^{\mathcal{H}}_{K^{-1}G_{1/\lambda}}(f) \tag{32}$$

A possible algorithm to compute $\hat{u}$ is therefore to compute $P^{\mathcal{H}}_{K^{-1}G_{1/\lambda}}(f)$.

We now describe our method to compute this projection (this is just an adaptation of Chambolle's method

[12]). Computing $P^{\mathcal{H}}_{K^{-1}G_{1/\lambda}}(f)$ amounts to finding:

$$\min\left\{\|\frac{1}{\lambda}K^{-1}\mathrm{div}\,(p)-f\|^2_{\mathcal{H}} : p \,/\, |p_{i,j}|^2 - 1 \le 0\ \forall i,j=1,\ldots,N\right\} \quad (33)$$

The Karush-Kuhn-Tucker [16, 31] conditions yield the existence of Lagrange multipliers $\beta_{i,j} \ge 0$ associated to each constraint in problem (33), such that we have for each i, j:

$$-\left(\nabla\left(\frac{1}{\lambda}K^{-1}\mathrm{div}\,(p)-f\right)\right)_{i,j}+\beta_{i,j}|p_{i,j}| = 0 \quad (34)$$

with either $\beta_{i,j} > 0$ and $|p_{i,j}| = 1$, or $\beta_{i,j} = 0$ and $|p_{i,j}| < 1$. In any case, we get:

$$\beta_{i,j} = \left|\left(\nabla\left(\frac{1}{\lambda}K^{-1}\mathrm{div}\,(p)-f\right)\right)_{i,j}\right| \quad (35)$$

We then propose the same kind of semi-implicit gradient descent scheme as in [12]:

$$p^0 = 0 \quad (36)$$

and

$$p^{n+1}_{i,j} = \frac{p^n_{i,j}+\tau(\nabla(K^{-1}\mathrm{div}\,(p^n)-\lambda f))_{i,j}}{1+\tau|(\nabla(K^{-1}\mathrm{div}\,(p^n)-\lambda f))_{i,j}|} \quad (37)$$

We can now state the following result.

Theorem 1. *If $\tau \le \frac{1}{8\|K^{-1}\|_{L^2}}$, then $\frac{1}{\lambda}K^{-1}\mathrm{div}\,p^n \to \hat{v}$ as $n \to \infty$, and $f - \frac{1}{\lambda}K^{-1}\mathrm{div}\,p^n \to \hat{u}$ as $n \to \infty$.*

Proof: It is very similar to the proof of Theorem 3.1 in [12]. The main difference is that here we work with the $\mathcal{H}$ norm instead of the L^2 norm. For the detailed proof, we refer the interested reader to [6]. Notice that a similar result has recently been obtained independently in [1]. □

We have therefore shown how to solve problem (26) when we know the correct value of the Lagrange multiplier λ. We now focus on how to automatically tune λ in the case of image denoising.

4. The Constrained Problem

The idea of minimizing the total variation for image denoising, suggested in [42], assumes that the observed image f is the addition of an image with little oscillations (typically piecewise smooth) s and a random Gaussian noise n, of estimated variance σ^2. It is then suggested to recover the original image by trying to solve the problem:

$$\min_u\left\{J(u) \,/\, \|u-f\|^2_{L^2} = N^2\sigma^2\right\} \quad (38)$$

where N^2 is the size of the image. The equivalent problem we are interested in when restoring an image with our model (26) is then:

$$\min_u\left\{J(u) \,/\, \|u-f\|^2_{\mathcal{H}} = \alpha N^2(\rho(\mathcal{H},N,\sigma))^2\right\} \quad (39)$$

where $N\rho(\mathcal{H},N,\sigma)$ is the $\mathcal{H}$ norm of an image (of size N^2) of a white Gaussian noise with standard deviation σ. The classical (ROF) case is with $\alpha = 1$. We give an estimation of $N\rho(\mathcal{H},N,\sigma)$ in the following subsection.

4.1. $\mathcal{H}$ Norm of a White Gaussian Noise

We will need the following lemma (which is a standard result in convex analysis [41]):

Lemma 1. *If $f \in \mathcal{H}\bigcap\mathcal{H}^*$, then*

$$(\|f\|_{\mathcal{H}})^* = \|f\|_{\mathcal{H}^*} \quad (40)$$

where $\mathcal{H}^*$ is the Hilbert space whose inner product is defined by $\langle u,v\rangle_{\mathcal{H}^*} = \langle u, K^{-1}v\rangle$ and $(\|f\|_{\mathcal{H}})^*$ is the Legendre-Fenchel transform (see (19)) of $\|f\|_{\mathcal{H}}$.

Proof: Let us denote by $L(u) = \frac{1}{2}\|u\|^2_{\mathcal{H}}$. Then we have:

$$\begin{aligned} L^*(v) &= \sup_{u\in X}\left(\langle u,v\rangle_{L^2} - L(u)\right) \\ &= \sup_{u\in X}\left(-\frac{1}{2}\|Ku-v\|^2_{\mathcal{H}^*}+\frac{1}{2}\|v\|^2_{\mathcal{H}^*}\right) \\ &= \frac{1}{2}\|v\|^2_{\mathcal{H}^*} \end{aligned}$$

□

For the sake of clarity, we will assume periodic boundary conditions in the rest of this subsection. We assume that n is an image of white Gaussian noise; i.e.

for all (i, j), $n_{i,j}$ follows a Gaussian probability density function: $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$ where σ^2 is the variance of the noise. We recall that if Z has a probability density function p_Z, than we denote its expectation by $E(Z) = \int z p_Z(z)\, dz$. The following proposition is a straightforward generalization of a result proved in [5]:

Proposition 4.

$$E(\|n\|_{\mathcal{H}}^2) = C_{\mathcal{H}} E(\|n\|_{L^2}^2) = C_{\mathcal{H}} N^2 \sigma^2 \tag{41}$$

with $C_{\mathcal{H}} = \|P_{Im(K)}^{L^2}(\delta)\|_{\mathcal{H}}^2$, *where* δ *is defined by* $\delta_{0,0} = 1$ *and* $\delta_{i,j} = 0$ *otherwise.*

We recall that $Im(K) = \{y \in X$ such that $\exists x \in X$ such that $y = Kx\}$, and that $P_{Im(K)}^{L^2}(\delta)$ is the orthogonal projection of δ over $Im(K)$ with respect to the L^2 inner product.

Remarks

1. When $\mathcal{H} = L^2$, then $C_{L^2} = 1$.
2. When $\mathcal{H} = H^{-1}$, then it is shown in [5] that:

$$C_{H^{-1}} = \frac{1}{N^2} \sum_{(p,q)\neq(0,0)} \frac{1}{2\left(2 - \cos\left(\frac{2\pi}{N}p\right) - \cos\left(\frac{2\pi}{N}q\right)\right)} \tag{42}$$

3. When $\mathcal{H} = H$, then $C_H = 2$.

Moreover, it is also shown in [5] that $Var(\|n\|_{H^{-1}}^2) << E(\|n\|_{H^{-1}}^2)$. In the rest of the paper, we will therefore make the following approximation:

$$\|n\|_{H^{-1}}^2 = C_{H^{-1}} \|n\|_{L^2}^2 \tag{43}$$

And more generally, we will assume that:

$$\|n\|_{\mathcal{H}}^2 = C_{\mathcal{H}} \|n\|_{L^2}^2 \tag{44}$$

Equations (43) and (44) are consequences of the statistical law of large numbers and hold asymptotically, when $N \to +\infty$.

Proof of Proposition 4. The proof is similar to the one of Proposition 3.5 in [5], but it is more technical in this general framework. We refer the interested reader to [6] for a detailed proof. □

4.2. Solving the Constrained Problem

The problem we are therefore interested in when restoring an image with our model (26) is:

$$\min_u \left\{ J(u) \;/\; \|u - f\|_{\mathcal{H}}^2 = C_{\mathcal{H}} N^2 \sigma^2 \right\} \tag{45}$$

where $C_{\mathcal{H}}$ is the constant given in Proposition 4. Since σ is less difficult to estimate than λ in (26), it is of practical interest to know how to solve (45) directly. The task is to find $\lambda > 0$ such that $\|P_{K^{-1}G_{1/\lambda}}^{\mathcal{H}}(f)\|_{\mathcal{H}}^2 = C_{\mathcal{H}} N^2 \sigma^2$. For $s > 0$, let us set

$$g(s) = \|P_{K^{-1}G_{1/s}}^{\mathcal{H}}(f)\|_{\mathcal{H}} \tag{46}$$

The following lemma states the main properties of g (we denote by $\bar{f}$ the mean of f).

Lemma 2. *The function* $g(s)$ *maps* $[0, +\infty)$ *onto* $[0, \|f - \bar{f}\|_{\mathcal{H}}]$. *It is non-increasing, while the function* $s \mapsto sg(s)$ *is non-decreasing.*

Proof: The proof is very close to the one of Lemma 4.1 in [12]. The main difference relies in the use of the $\mathcal{H}$ inner product instead of the L^2 inner product. For a detailed proof, we refer the reader to [6]. □

Thanks to Lemma 2 we can propose the following algorithm, in order to solve (45) (similar to the one proposed in [12] to solve (38)). We assume $\sqrt{C_{\mathcal{H}}} N\sigma$ is between 0 and $\|f - \bar{f}\|_{\mathcal{H}}$. We need to find a value $\tilde{\lambda}$ for which $g(\tilde{\lambda}) = \sqrt{C_{\mathcal{H}}} N\sigma$.

Algorithm

1. *Initialization:* Choose any arbitrary $\lambda_0 > 0$, and compute

$$v_0 = P_{K^{-1}G_{1/\lambda_0}}^{\mathcal{H}}(f) \tag{47}$$

   with the algorithm described in the previous section, as well as $g_0 = g(\lambda_0) = \|v_0\|_{\mathcal{H}}$.
2. *Iterations:* Given λ_n and g_n, then let $\lambda_{n+1} = \frac{g_n}{\sqrt{C_{\mathcal{H}}} N\sigma} \lambda_n$ and compute

$$v_{n+1} = P_{K^{-1}G_{1/\lambda_{n+1}}}^{\mathcal{H}}(f) \tag{48}$$

   as well as $g_{n+1} = g(\lambda_{n+1}) = \|v_{n+1}\|_{\mathcal{H}}$.

From Lemma 2, it is easy to deduce the following result (the proof is exactly the same as the one of Theorem 2 in [12]).

Theorem 2. *As $n \to +\infty$, we have $g_n \to \sqrt{C_{\mathcal{H}}} N\sigma$ while $u_n = f - v_n$ converges to the unique solution of (45).*

This closes the generalization of Chambolle's results [12] to our new model. We now turn our attention to a particular case when these new general results lead to a new automatic denoising algorithm.

4.3. Application to the Osher-Sole-Vese Algorithm

In the specific case when $K = -\Delta^{-1}$, i.e. when $\mathcal{H} = H^{-1}$, then our model (26) is the Osher-Sole-Vese model [38].

$$\inf_u \left(J(u) + \frac{\lambda}{2} \| f - u \|^2_{H^{-1}} \right) \tag{49}$$

In [38], the authors write the associated Euler-Lagrange equations, and then compute the solution by solving a fourth order PDE.

Our algorithm (36)–(37) reduces in that case to the one that has been proposed in [5] to solve (49). We therefore reformulate Theorem 1 in the following proposition:

Proposition 5. *If $\tau \leq \frac{1}{64}$, then $f - \lambda \Delta \operatorname{div} p^n \to \hat{u}$ (solution of (49)) as $n \to \infty$.*

In fact, we have checked numerically that the algorithm converges as long as $\tau < \frac{1}{32}$ (which is twice the theoretical bound, and which has already been noticed in the case $\mathcal{H} = L^2$ in [12]).

Up to now, there have not been proposed any method to solve problem (45) in the case $\mathcal{H} = H^{-1}$, although it has been noticed in [5] that the Osher-Sole-Vese is a very good denoising model. This issue is now addressed by algorithm (47)–(48) (remembering that $C_{\mathcal{H}}$ is given by (42)).

In practice, we have checked numerically that we can considerably increase the convergence speed of the algorithm by choosing $\lambda_0 = 1$ and by updating λ each 20 iterations.

Numerical examples. We give some numerical examples on Figures 2 to 4.

In practice, we have checked that using the right σ leads to a too strong denoising: the denoised image is then oversmoothed. In fact, as it is also the case with the ROF model [42], the value of λ computed from σ leads to a residual which has the same $\mathcal{H}$ norm as the original noise (see for instance [17] where the authors address this problem by imposing the value of the total variation of the restored image instead of the norm of the noise). Unfortunately, as always, the denoising model is not perfect: therefore, before getting rid of all the noise, our algorithm also removes some of the textures and edges. Visually, we prefer a less denoised image with more details. This question will be addressed in the following section. Anticipating on these results, we will see that a good numerical choice is to use $\alpha = 1/2$ instead of $\alpha = 1$ in (39).

In the next section, we derive a different algorithm for automatic denoising.

5. SNR-Based Parameter Selection

5.1. Definitions

In this section we use a slightly different definition of the inner-product and norm. We also treat these quantities as continuous functions with respect to the parameter λ. Therefore, different notations are used. We omit the dependency on λ for brevity. We define $\mathcal{I}(\cdot, \cdot)$ to be the normalized, zero-mean *inner-product* (we recall that $\bar{p}$ stands for the mean of p):

$$\mathcal{I}(p, q) \doteq \frac{1}{|\Omega|} < p - \bar{p}, q - \bar{q} >_{\mathcal{H}}, \tag{50}$$

and consequently we define $\mathcal{N}(\cdot)$ to be the normalized, zero-mean square of a *norm*:

$$\mathcal{N}(p) \doteq \mathcal{I}(p, p) = \frac{1}{|\Omega|} \| p - \bar{p} \|^2_{\mathcal{H}}. \tag{51}$$

The above measures become the standard notions of empirical *covariance* and *variance*, respectively, for $\mathcal{H} = L^2$. In the following we refer to $\mathcal{N}$ for short as "norm" and not "the square of the normalized norm". Note that in the discrete setting of this paper $|\Omega| = N^2$.

Our problem can be written as

$$\inf_{u,v} \left(J(u) + \frac{\lambda}{2} \mathcal{N}(v) \right), \text{ subject to } f = u + v. \tag{52}$$

We prefer to specifically write v in the minimization problem, though it is implied by u and f, as it turns out to have a significant part in our analysis below.

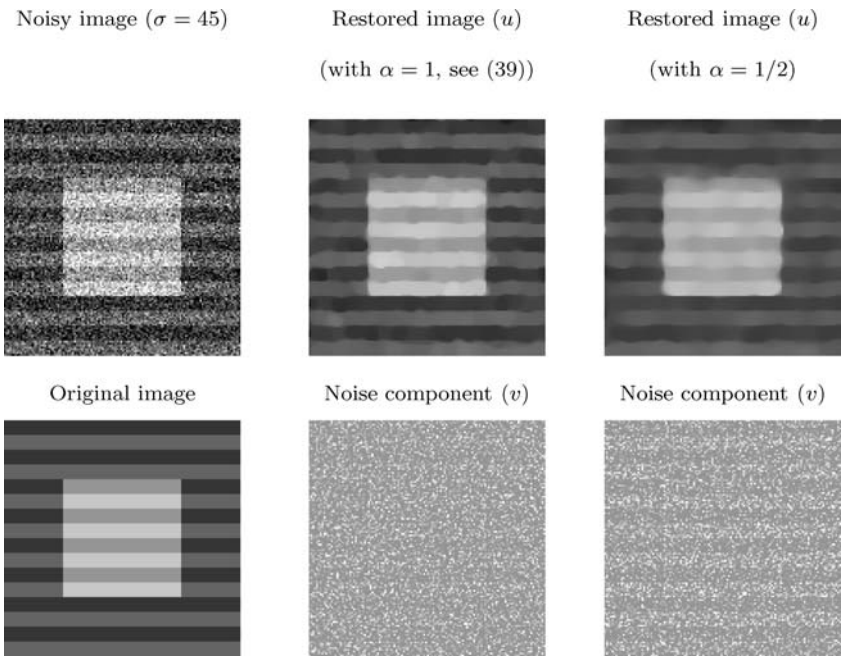

Figure 2. Automatic restoration of a synthetic image. First column: noisy image (corrupted by white gaussian noise with standard deviation $\sigma = 45$) and original image. Second column: automatic restoration with $\alpha = 1$ (one sees that the noise component contains more information than just the noise). Third column: : automatic restoration with $\alpha = 1/2$ (one sees that such a value is large enough to remove the noise).

The $\mathcal{H}$ Signal-to-Noise Ratio ($SNR^{\mathcal{H}}$) of the recovered signal u is defined as

$$SNR^{\mathcal{H}}(u) \doteq 10 \log \frac{\mathcal{N}(s)}{\mathcal{N}(u-s)} = 10 \log \frac{\mathcal{N}(s)}{\mathcal{N}(n-v)}, \tag{53}$$

where $\log \doteq \log_{10}$. We usually omit the $\mathcal{H}$ superscript. The (square) norm of the noise is

$$\mathcal{N}(n) = \rho^2. \tag{54}$$

For $\mathcal{H} = L^2$ we have $\rho^2 = \sigma^2$. The initial SNR of the input signal, denoted by SNR_0, where no processing is carried out ($u = f$, $v = 0$), is according to (53), (54) and (1):

$$SNR_0 \doteq SNR(f) = 10 \log \frac{\mathcal{N}(s)}{\mathcal{N}(n)} = 10 \log \frac{\mathcal{N}(s)}{\rho^2}. \tag{55}$$

For $\mathcal{H} = L^2$ we reach the standard SNR definition: $SNR_0 = 10 \log(\text{var}(s)/\sigma^2)$, where var denotes the variance.

Let us define the optimal SNR of a certain process applied to an input image f as:

$$SNR_{opt} \doteq \max_{\lambda} SNR(u_\lambda) \tag{56}$$

where $u = u_\lambda$ attains the minimal energy of (52) with weight parameter λ for a given f. We denote by (u_{opt}, v_{opt}) the decomposition pair (u, v) that reaches SNR_{opt}, and define $\mathcal{N}_{opt} \doteq \mathcal{N}(v_{opt})$.

5.2. *Condition for Optimal SNR*

We now develop a necessary condition for the optimal SNR. Imposing a specific value for the norm of v, $\mathcal{N}(v) = P$, in the constrained problem amounts to choosing λ in (52). This was proved by Chambolle-Lions [13] in the case $\mathcal{H} = L^2$ and could be generalized to our framework by using Proposition 2 and Lemma 2. We therefore regard the SNR as a function $SNR(\mathcal{N}(v))$ and assume that it is smooth. A necessary condition for the maximum in the range $\mathcal{N}(v) \in (0, \mathcal{N}(f))$ is:

$$\frac{\partial SNR}{\partial \mathcal{N}(v)} = 0. \tag{57}$$

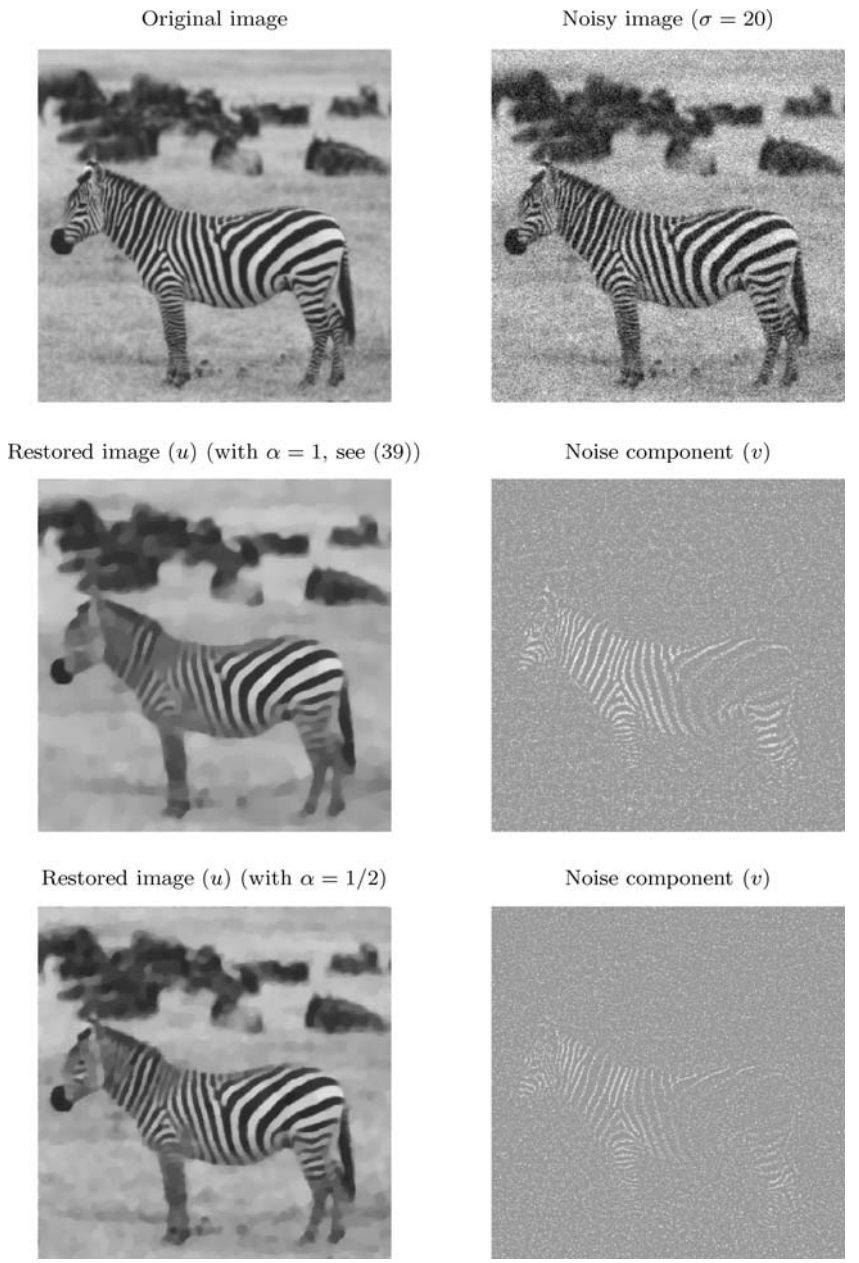

Figure 3. Automatic restoration of a zebra image. First line: original image and noisy image (corrupted by white gaussian noise with standard deviation $\sigma = 20$). Second line: automatic restoration with $\alpha = 1$ (one sees that the noise component contains more information than just the noise). Third line: : automatic restoration with $\alpha = 1/2$ (one sees that such a value is large enough to remove the noise).

Rewriting $\mathcal{N}(n - v)$ as $\mathcal{N}(n) + \mathcal{N}(v) - 2\mathcal{I}(n, v)$, and using (57) and (53), yields

$$\frac{\partial \mathcal{I}(n, v)}{\partial \mathcal{N}(v)} = \frac{1}{2}. \tag{58}$$

More specifically, as long as $\frac{\partial \mathcal{I}(n,v)}{\partial \mathcal{N}(v)} > \frac{1}{2}$, the SNR increases. When the condition of (58) is reached, noise and signal are equally filtered and the SNR is at a local maximum. If filtering is continued, more signal than noise is filtered (in the $\mathcal{H}$-norm sense) and the SNR decreases.

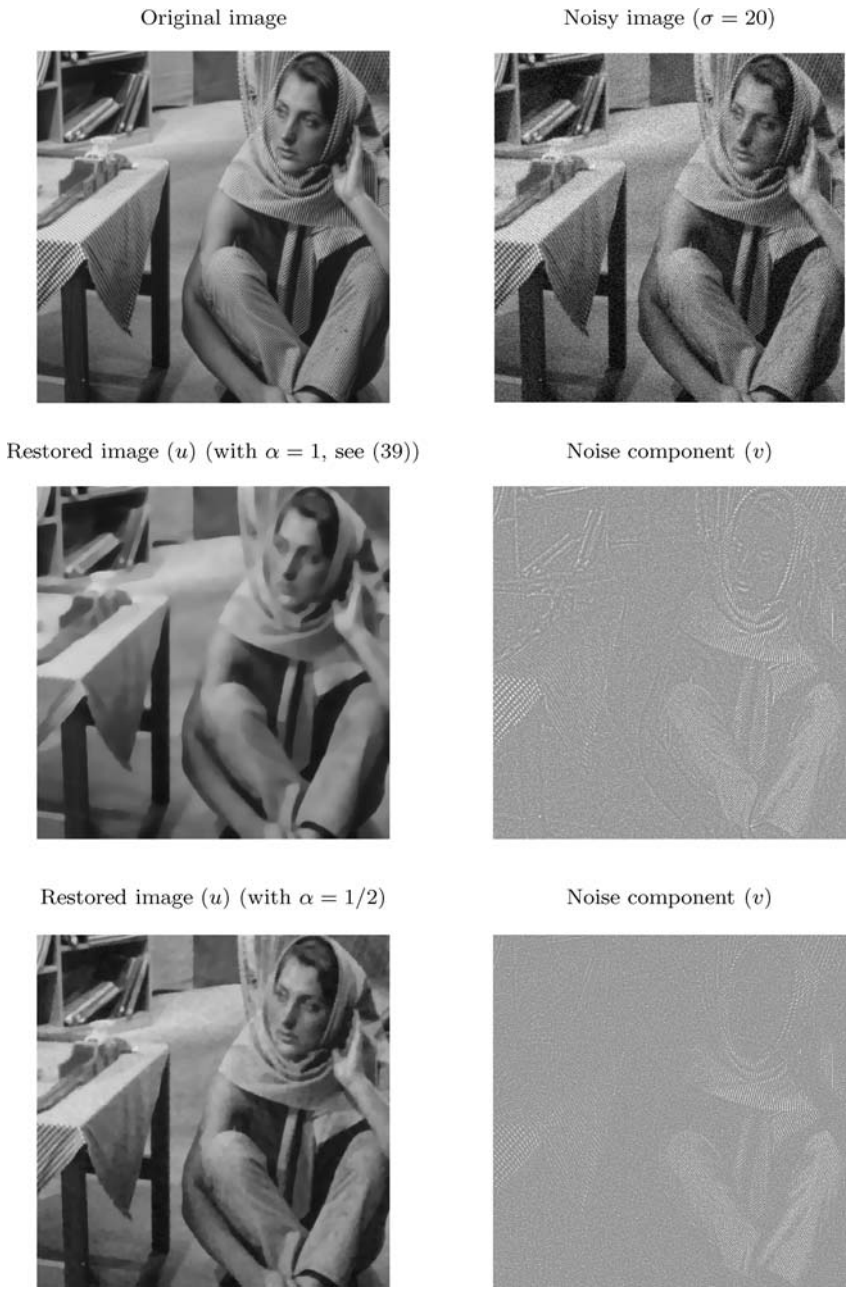

Figure 4. Automatic restoration of Barbara First line: original image and noisy image (corrupted by white gaussian noise with standard deviation $\sigma = 20$). Second line: automatic restoration with $\alpha = 1$ (one sees that the noise component contains more information than just the noise). Third line: : automatic restoration with $\alpha = 1/2$ (one sees that such a value is large enough to remove the noise).

There is also a possibility that the maximum is at the boundaries: If the SNR is dropping from the beginning of the process we have $\frac{\partial \mathcal{I}(n,v)}{\partial \mathcal{N}(v)}|_{\mathcal{N}(v)=0} < \frac{1}{2}$ and $SNR_{opt} = SNR_0$. The other extreme case is when the SNR increases monotonically and is maximized when $\mathcal{N}(v) = \mathcal{N}(f)$ (the trivial constant solution $u = \bar{f}$). We see in the following (Proposition 8) that this can only happen when SNR_0 is negative or, equivalently, when $\mathcal{N}(s) < \rho^2$.

In light of these considerations, provided that one can estimate $\mathcal{I}(n, v)$, our basic numerical algorithm should be as follows:

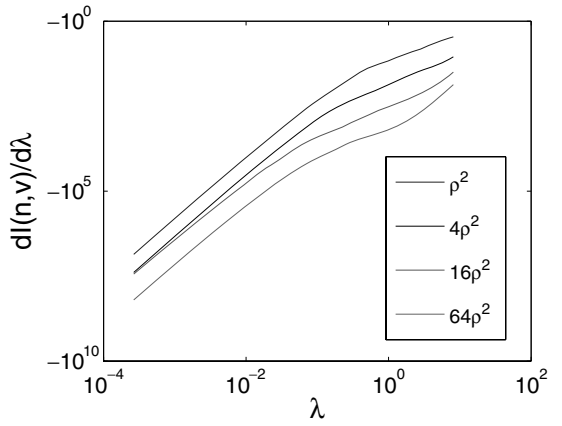

Figure 5. Precomputed term $\partial\mathcal{I}(\tilde{n}, v)/\partial\lambda$ as a function of λ (log scale), $\mathcal{H} = H^{-1}$. Graphs depict plots for different proportion of $\rho^2 - 1 : 4 : 16 : 64$, from upper curve to lower curve, respectively.

1. Set $\mathcal{I}^0(n, v) = 0, \mathcal{N}^0(v) = 0, i = 1$.
2. $\mathcal{N}^i(v) \leftarrow \mathcal{N}^{i-1}(v) + d\mathcal{N}(v)$. Compute $\mathcal{I}^i(n, v)$.
3. If $\frac{\mathcal{I}^i(n,v) - \mathcal{I}^{i-1}(n,v)}{d\mathcal{N}(v)} \leq \frac{1}{2}$ then stop.
4. $i \leftarrow i + 1$. Goto step 2.

In the next section we suggest a method to approximate the inner product term.

5.3. *Estimating* $\mathcal{I}(n, v)$

The term $\mathcal{I}(n, v)$ is unknown, as we do not know the noise, and therefore should be estimated. We show below a representation of denoising by a family of curves which connects the norm of the noise, λ and $\mathcal{I}(n, v)$ of pure noise. This can be regarded as some sort of nonlinear statistics of noise with respect to a specific energy functional.

Our observation is that the extent of filtering of additive noise, with respect to λ, is not affected much by the underlying image s. What mainly affects the denoising performance is the extent of filtering of s. In the linear case, the decoupling of the filtering of s and n, in the case of additive noise, is very clear: $h * f = h * (s + n) = h * s + h * n$, where h is the filtering kernel, and $*$ denotes convolution. The filtering of n depends solely on the filtering kernel h and not on s. We follow a significant observation made in [25] for the L^2 fidelity term in which this decoupling property can well approximate ROF-type regularizations. Here this concept is extended to the general case of Hilbert space fidelity terms. Note that one may view a wide range of nonlinear scale-spaces as a concatenation of short time filtering kernels (with some small error) [43]. Here a relatively simple decoupling mechanism is applied successfully also to our nonlinear framework. Currently, we investigate the possibility to obtain an analytic expression for the approximation error.

In the proposed decoupling technique, we measure in advance the extent of filtering of noise with the same variance. As all our quantities (norm, inner-product, SNR) are computed by averaging over many pixels ("measurements"), one is not restricted to the original noise, and can alternatively use a generated noise with similar statistics, which is denoted in the following as $\tilde{n}$.

First, we need to compute the "statistics" by processing a patch of pure noise $\tilde{n}$ while measuring $\mathcal{I}(\tilde{n}, v)$ with respect to λ. As the instance of $\tilde{n}$ is generated by us it is completely known. This is done a single time for each noise norm ρ^2 and can be regarded as a look-up-table (see Fig. 5). In general one may repeat this procedure several times for each ρ^2 and average the results to increase accuracy (reduce the variance of the error), however we observed that for a patch of a few thousands pixels a single iteration can suffice. For each processed image the behavior of λ with respect to $\mathcal{N}(v)$ is measured. Combining the information, it is possible to approximate how $\mathcal{I}(n, v)$ behaves with respect to $\mathcal{N}(v)$. The connection is done through the chain-rule for differentiation:

$$\begin{aligned}\frac{\partial\mathcal{I}(n, v)}{\partial\mathcal{N}(v)} &= \frac{\partial\mathcal{I}(n, v)}{\partial\lambda}\frac{\partial\lambda}{\partial\mathcal{N}(v)} \\ &\approx \frac{\partial\mathcal{I}(\tilde{n}, v)}{\partial\lambda}\Big|_{f=\tilde{n}}\frac{\partial\lambda}{\partial\mathcal{N}(v)}\Big|_{f=s+n}.\end{aligned} \tag{59}$$

In the next section we provide performance bounds, based on SNR analysis, for the constrained problem presented in Section 4 and for the optimal parameter, which is estimated in Sections 5. Experimental results comparing the suggested methods are shown in Section 7.

6. SNR Performance Bounds

Let us denote u^z as the solution of (52) for $f = z$. For example, u^s is the solution where $f = s$.

For the purpose of this analysis, two assumptions are made with respect to s, n and the regularization process. They were tested numerically for the cases $\mathcal{H} = L^2$ and $\mathcal{H} = H^{-1}$ for different signals s and white Gaussian noise n.

First we have an orthogonality assumption of s and n which is taken with respect to the regularization:

$$\mathcal{I}(u^s, n) = 0, \quad \mathcal{I}(u^n, s) = 0, \quad \forall \lambda \geq 0. \tag{60}$$

We further assume the process applied to $f = s + n$ does not amplify or sharpen either s or n. This can be formulated in terms of inner product as follows:

$$\begin{aligned} &\mathcal{I}(u^{s+n}, s) \leq \mathcal{I}(f, s), \\ &\mathcal{I}(u^{s+n}, n) \leq \mathcal{I}(f, n), \\ &\forall \lambda \geq 0. \end{aligned} \tag{61}$$

The above assumptions could be understood as some extension of the linear case. In fact, for linear filtering Assumption (60) and the fact that the regularizing filter is not amplifying the signal (in the $\mathcal{H}$ sense) is sufficient to *prove* Assumption (61): Let us assume that our regularizer is a linear filter h such that: $u = f * h = T_h f$, where $*$ denotes convolution and T_h is the associated linear operator. Commonly this would be some low-pass filter which attenuates high frequencies. The requirement for not amplifying the signal in the $\mathcal{H}$ sense can be compactly stated by a unit bound on the norm of the linear operator: $\|T_h\|_{\mathcal{H}} \leq 1$. In the linear case Assumption (60) reads: $\mathcal{I}(h * s, n) = 0$, $\mathcal{I}(h * n, s) = 0$. Statistical independence between s and n (which is a standard assumption regarding additive noise) is a sufficient condition for these relations to hold. As in our more general nonlinear case h is controlled by a parameter (such as λ) which defines the extent of filtering and can degenerate to the identity filter ($\lambda \to \infty$) and therefore this assumption implies also $\mathcal{I}(s, n) = 0$. The relations stated in Assumption (61) can be shown by:

$$\begin{aligned} \mathcal{I}(u^{s+n}, s) &= \mathcal{I}(h * (s + n), s) = \mathcal{I}(h * s, s) + \mathcal{I}(h * n, s) \\ &= \mathcal{I}(h * s, s) \qquad\quad \leq \|T_h\|_{\mathcal{H}} \mathcal{N}(s) \\ &\leq N(s) \qquad\qquad\quad = \mathcal{I}(s + n, s) \\ &= \mathcal{I}(f, s). \end{aligned}$$

Similarly the relation $\mathcal{I}(u^{s+n}, n) \leq \mathcal{I}(f, n)$ can be shown. This analysis holds, for instance, for the corresponding linear case of our framework, in which the functional to be minimized is $\tilde{J}(u) + \frac{\lambda}{2}\mathcal{N}(v)$, where $\tilde{J}(u) = \sum_{1 \leq i,j \leq N} |(\nabla u)_{i,j}|^2$.

We have strong indications that (60) and (61) are general properties which may hold in the nonlinear case for any $\mathcal{H}$ within our framework whenever n is an independent noise (not necessarily Gaussian). Yet, this is only a conjecture at this stage. In this paper the question is left open and we resort to the following definition:

Definition 2 ((s, n) pair). An (s, n) pair consists of two signals s and n which obey assumptions (60) and (61).

Theorem 3. *For any (s, n) pair the inner product matrix of $U = (f, s, n, u, v)^T$ has only non-negative elements.*

For proof see [6]. Theorem 6 implies that the denoising process has smoothing properties and consequently, there is no negative correlation between any two elements of U. This basic theorem is used to establish several bounds in our performance analysis.

The constrained problem of Section 4 can be formulated in our context as imposing

$$\mathcal{N}(v) = \rho^2. \tag{62}$$

We define

$$SNR_{\rho^2} \doteq SNR(u)|_{\mathcal{N}(v) = \rho^2}. \tag{63}$$

We denote by (u_{ρ^2}, v_{ρ^2}) the (u, v) pair that obeys (62) and minimizes (52). We now analyze this method for selecting u in terms of SNR.

Proposition 6 (SNR lower bound). *Imposing (62), for any (s, n) pair SNR_{ρ^2} is bounded from below by*

$$SNR_{\rho^2} \geq SNR_0 - 3dB, \tag{64}$$

where we use the customary notation $3dB$ for $10\log_{10}(2)$.

Proof: From Theorem 6 we have $\mathcal{I}(n, v) \geq 0$, therefore,

$$\begin{aligned} SNR_{\rho^2} &= 10\log\tfrac{\mathcal{N}(s)}{\mathcal{N}(n-v)} \\ &\geq 10\log\frac{\mathcal{N}(s)}{\mathcal{N}(n) + \mathcal{N}(v))} \\ &= 10\log\frac{\mathcal{N}(s)}{2\rho^2} \\ &= SNR_0 - 3dB. \end{aligned}$$

□

The lower bound of proposition 6 is reached only in the very rare and extreme case where $\mathcal{I}(n, v) = 0$. This implies that only parts of the signal were filtered out and no denoising was performed.

Proposition 7 (SNR upper bound). *Imposing (62), then there does not exist an upper bound* $0 < M < \infty$, *where* $SNR_{\rho^2} \leq SNR_0 + M$, *that is valid for any given* (s, n) *pair.*

Proof: To prove this we need to show only a single case where the SNR cannot be bounded. Let us assume $\mathcal{N}(s) = h\rho^2$, $0 < h < 1$. Then $SNR_0 = 10 \log h$. As signal and noise are not correlated we have $\mathcal{N}(f) = \mathcal{N}(s) + \mathcal{N}(n) = (1 + h)\rho^2$. We can write $\mathcal{N}(f)$ also as $\mathcal{N}(u + v) = \mathcal{N}(u) + \mathcal{N}(v) + 2\mathcal{I}(u, v)$. From (62), $\mathcal{N}(v) = \rho^2$, and from Theorem 6, $\mathcal{I}(u, v) \geq 0$, therefore $\mathcal{N}(u) \leq h\rho^2$. Since $\mathcal{I}(u, s) \geq 0$ (Theorem 6) we get $\mathcal{N}(u - s) \leq 2h\rho^2$. This yields $SNR_{\rho^2} \geq 10 \log \frac{1}{2}$ and

$$SNR_{\rho^2} - SNR_0 \geq 10 \log \frac{1}{2h}.$$

Thus, for any M we can choose a sufficiently small h where the bound does not hold. □

Definition 3 (Regular SNR). We define the function $SNR(\mathcal{N}(v))$ as *regular* if (58) is a sufficient condition for optimality or if the optimum is at the boundaries.

Regular *SNR* basically means that the denoising process has two phases (where the evolution is with respect to $\mathcal{N}(v)$). First the regularization improves the quality of the image as mostly noise is being filtered (*SNR* increases). At some point any additional regularization causes degradation of the result as more signal than noise is being filtered. This definition permits either of these phases to be of "zero duration" but not to have multiple switches between phases (meaning the SNR function would have more than a single local maximum). In Figs. 8, 10 and 12 one can observe that this regularity is valid for the synthetic and natural image examples (see *SNR* plot as a function of $\mathcal{N}(v)/\rho^2$). Our next proposition and theorem on the range and bound of *SNR*, analyze the case of a regular *SNR*.

Proposition 8 (Range of optimal *SNR*). *If SNR is regular, then for any* (s, n) *pair* $0 \leq \mathcal{N}_{opt} \leq 2\rho^2$.

Proof: Let us first show the relation $\mathcal{I}(n, v) \leq \rho^2$: $\mathcal{I}(n, f) = \mathcal{I}(n, n + s) = \mathcal{N}(n) + \mathcal{I}(n, s) = \rho^2$, using (60). On the other hand $\mathcal{I}(n, f) = \mathcal{I}(n, u + v) = \mathcal{I}(n, u) + \mathcal{I}(n, v)$. The relation is validated by using $\mathcal{I}(n, u) \geq 0$ (Theorem 6).

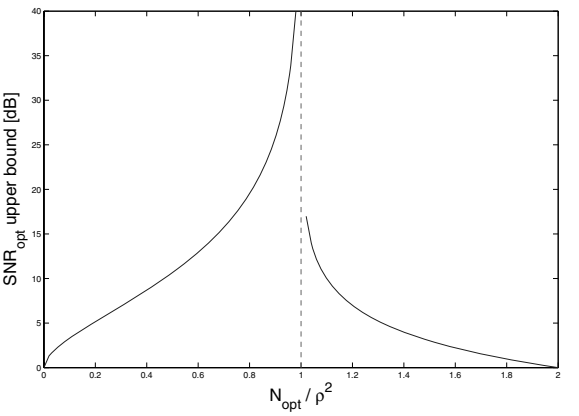

Figure 6. Visualization of Theorem 6: Upper bound of $SNR_{opt} - SNR_0$ as a function of $\mathcal{N}_{opt}/\rho^2$. For $\mathcal{N}_{opt} \to \rho^2$ the bound approaches $+\infty$.

Table 1. Denoising results in terms of $SNR^{H^{-1}}$ of the examples presented in Figures 7, 9 and 11. SNR_{ρ^2} is the result of imposing $\mathcal{N}(v) = \rho^2$ (Section 4). SNR_{est} is the result of our estimation of the optimal result (Section 5).

Image	SNR_0	SNR_{opt}	SNR_{ρ^2}	SNR_{est}
Synthetic	24.97	27.60	25.19	27.52
Lena	28.25	29.64	28.13	29.60
Cameraman	38.90	40.34	38.75	40.34
Average difference from SNR_{opt}	1.82	0.00	1.84	0.04

We reach the upper bound by the following inequalities:

$$\begin{aligned}\rho^2 &\geq \mathcal{I}(n, v)|_{v=v_{opt}} \\ &= \int_0^{\mathcal{N}_{opt}} \frac{\partial \mathcal{I}(n, v)}{\partial \mathcal{N}(v)} d\mathcal{N}(v) \geq \int_0^{\mathcal{N}_{opt}} \frac{1}{2} d\mathcal{N}(v) \\ &= \frac{1}{2}\mathcal{N}_{opt}.\end{aligned}$$

The inequality on the right is based on that $\frac{\partial \mathcal{I}(n,v)}{\partial \mathcal{N}(v)} \geq \frac{1}{2}$ for $\mathcal{N}(v) \in (0, \mathcal{N}_{opt})$.

The lower bound $\mathcal{N}_{opt} = 0$ is reached whenever $\frac{\partial \mathcal{I}(n,v)}{\partial \mathcal{N}(v)}|_{\mathcal{N}(v)=0} < \frac{1}{2}$. □

Theorem 4 (Bound on optimal SNR). *If SNR is regular, then for any* (s, n) *pair and* $\mathcal{N}_{opt} \in \{[0, \rho^2), (\rho^2, 2\rho^2]\}$,

$$\begin{aligned}&0 \leq SNR_{opt} - SNR_0 \\ &\leq \begin{cases} -10 \log(1 + \mathcal{N}_{opt}/\rho^2 - 2\sqrt{\mathcal{N}_{opt}/\rho^2}), & 0 \leq \mathcal{N}_{opt} < \rho^2 \\ -10 \log(\mathcal{N}_{opt}/\rho^2 - 1), & \rho^2 < \mathcal{N}_{opt} \leq 2\rho^2 \end{cases} \quad (65)\end{aligned}$$

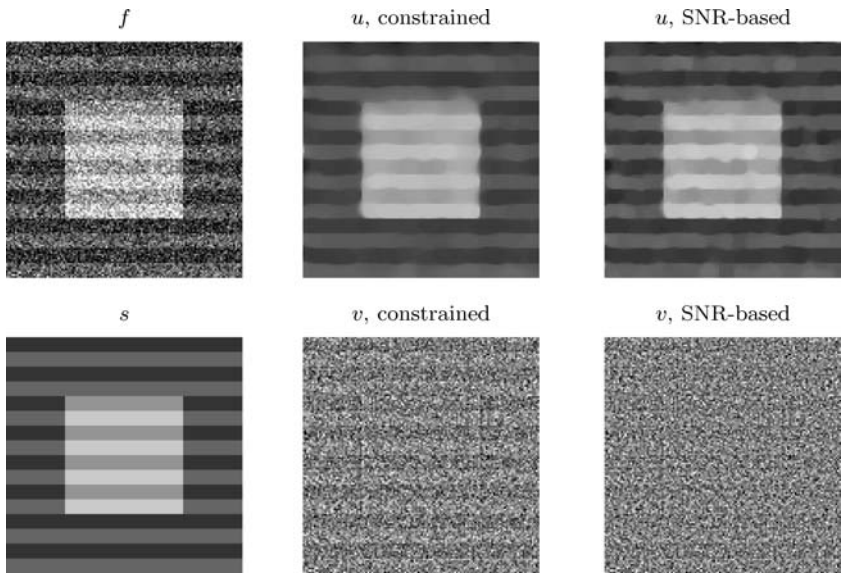

Figure 7. Denoising of a synthetic image ($\sigma = 45$). Middle column: solution of the constrained problem $\mathcal{N}(v) = \rho^2$. Right column: solution of SNR based result.

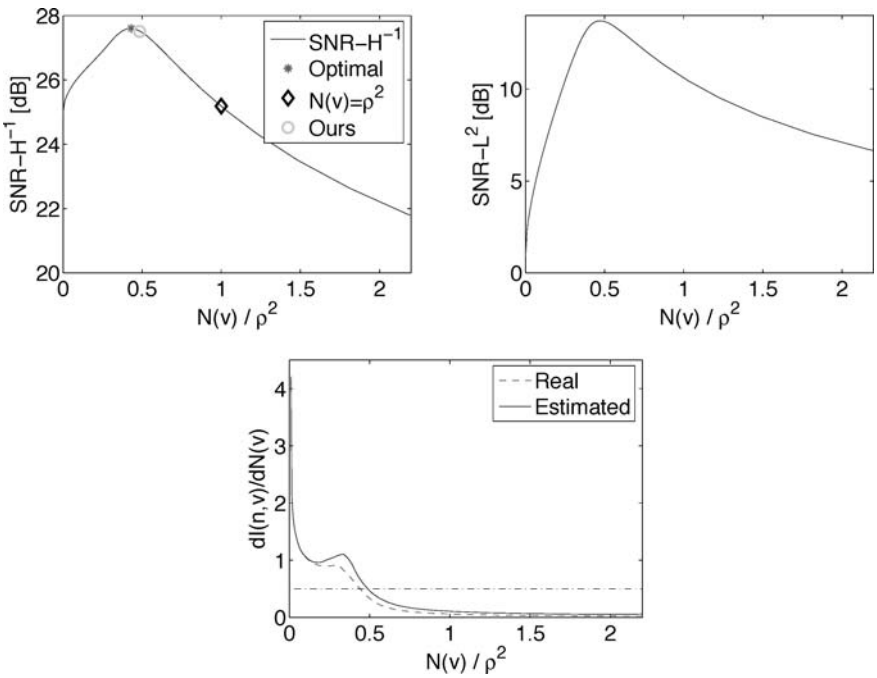

Figure 8. Denoising of a synthetic image (Fig. 7) - *SNR* and inner-product plots. Top (left): $SNR^{H^{-1}}$ as a function of $\mathcal{N}(v)$ with plots of the optimal, constrained and *SNR*-based selections ("Ours"). Top (right): a plot of the standard SNR^{L^2}, which behaves quite similarly to $SNR^{H^{-1}}$. Bottom row: estimated $\partial\mathcal{I}(n, v)/\partial\mathcal{N}(v)$ vs. the ground truth.

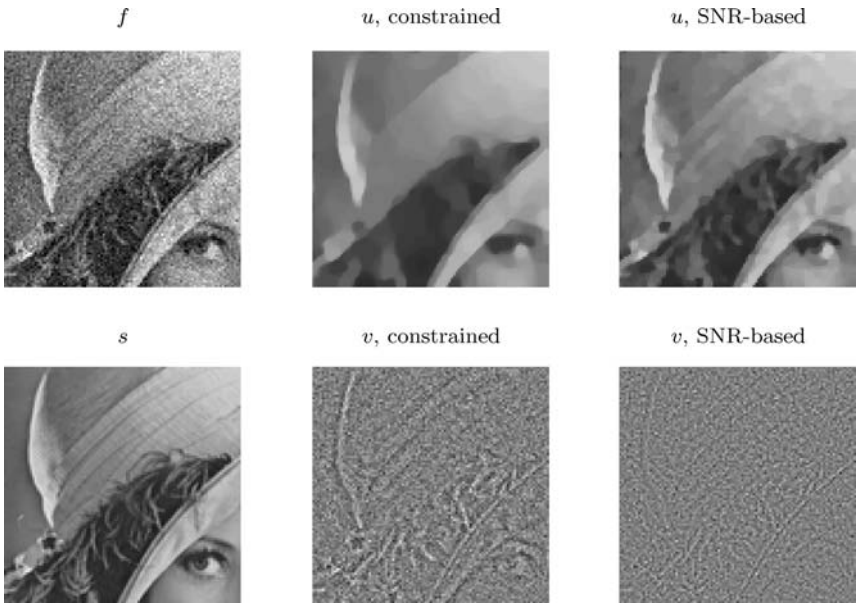

Figure 9. Denoising part of Lena image ($\sigma = 20$). Middle column: solution of the constrained problem $\mathcal{N}(v) = \rho^2$. Right column: solution of SNR based result.

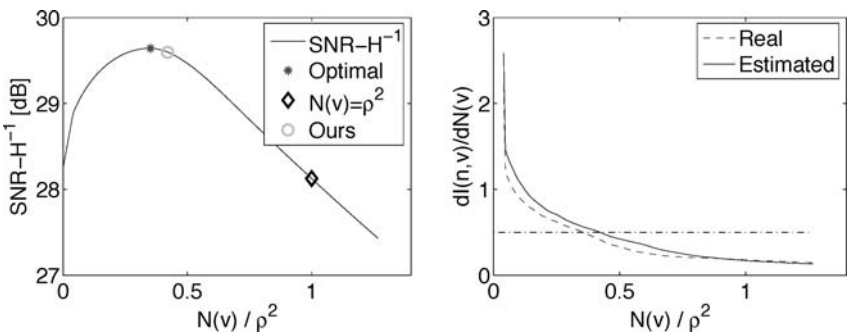

Figure 10. Denoising part of Lena image - SNR and covariance plots. Left: $SNR^{H^{-1}}$ as a function of $\mathcal{N}(v)$ with plots of the optimal, constrained and SNR-based selections ("Ours"). Right: estimated $\partial\mathcal{I}(n, v)/\partial\mathcal{N}(v)$ vs. the ground truth.

Proof: By the SNR definition, (53), and expanding the norm expression, we have

$$SNR_{opt} - SNR_0 = 10 \log\left(\frac{\rho^2}{\rho^2 + \mathcal{N}_{opt} - 2\mathcal{I}(n, v_{opt})}\right). \quad (66)$$

For the lower bound we use the relation shown in Proposition 6: $\mathcal{I}(n, v_{opt}) \geq \frac{1}{2}\mathcal{N}_{opt}$. For the upper bound we use two upper bounds on $\mathcal{I}(n, v_{opt})$ and take their minimum. The first one, $\mathcal{I}(n, v_{opt}) \leq \rho\sqrt{\mathcal{N}_{opt}}$, is the Cauchy-Schwartz inequality. The second relation, $\mathcal{I}(n, v_{opt}) \leq \rho^2$, is outlined in Proposition 8. □

A plot of the upper bound of the optimal *SNR* with respect to $\mathcal{N}_{opt}/\rho^2$ is depicted in Fig. 6.

In practice, the flow is not performed by directly increasing $\mathcal{N}(v)$, but by decreasing the value of λ. Therefore, it is instructive to check the vary of $\mathcal{N}(v)$, as well as the other energies, with respect to a vary in λ. In the next proposition we show that as λ decreases the total energy $E_J \doteq E_u + \frac{\lambda}{2}E_v$ strictly decreases, $E_u(u) \doteq J(u)$ decreases and $E_v(v) \doteq \mathcal{N}(v)$ increases.

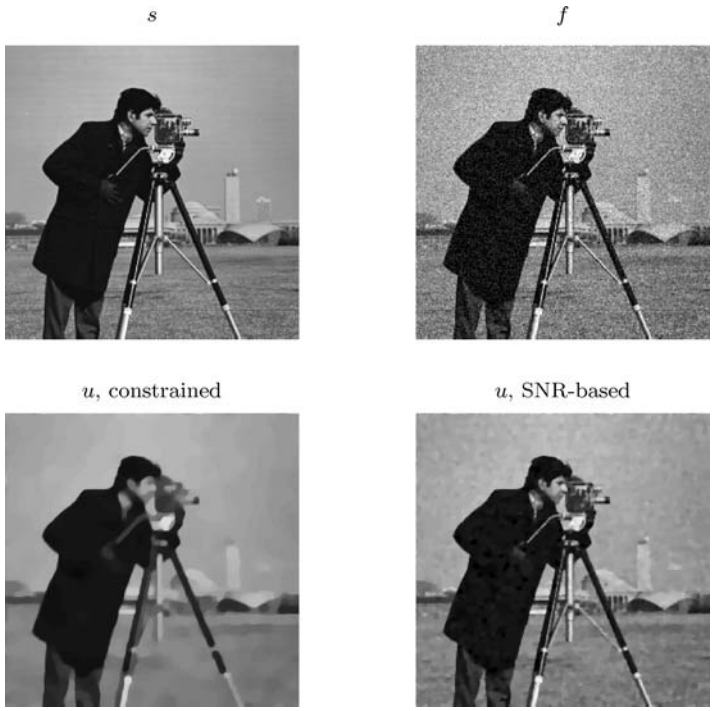

Figure 11. Denoising Cameraman image ($\sigma = 20$). From left to right. Top: original s, input image f. Bottom: u of constrained problem $\mathcal{N}(v) = \rho^2$, u of our SNR based selection.

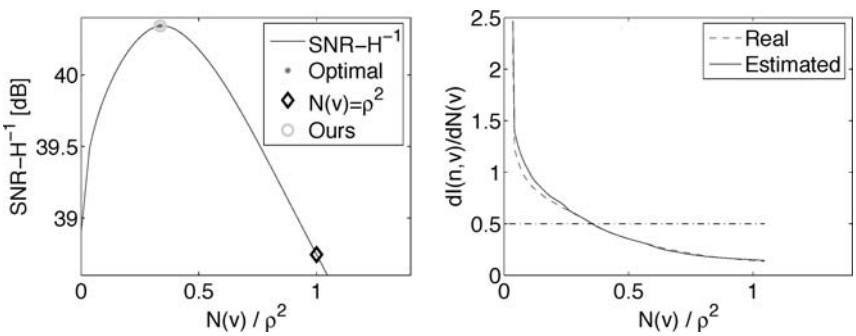

Figure 12. Denoising Cameraman image - SNR and covariance plots. Left: $SNR^{H^{-1}}$ as a function of $\mathcal{N}(v)$ with plots of the optimal, constrained and SNR-based selections ("Ours"). Right: estimated $\partial\mathcal{I}(n, v)/\partial\mathcal{N}(v)$ vs. the ground truth.

Proposition 9 (Energy change as a function of λ). *The energy parts of Eq. (52) vary as a function of λ as follows:*

$$\frac{\partial E_J}{\partial \lambda} > 0, \quad \frac{\partial E_v}{\partial \lambda} \leq 0, \quad \frac{\partial E_u}{\partial \lambda} \geq 0. \tag{67}$$

The proof is a consequence of Lemma 2.

We have given a mathematical analysis of our approach and shown performance bounds with respect to the $\mathcal{H}$-*SNR* criterion. In the next section we illustrate the proposed methods with numerical examples.

7. Experimental Comparison of the Methods

We have tested our algorithms for automatic parameter selection on both synthetic and natural images. To each original image white Gaussian noise of standard deviation σ was added ($\sigma = 45$, 20, 20 for Figs. 7, 9, 11, respectively). We display the result of imposing

$\mathcal{N}(v) = \rho^2$ and our estimated optimal denoising result (in the $\mathcal{H}$-SNR sense).

In Fig. 7, a synthetic image with a large square and stripes is processed. The stripes are better preserved with the estimated optimal approach. In Fig. 8 plots of $SNR^{H^{-1}}$, SNR^{L^2} and estimated and real $\partial\mathcal{I}(n, v)/\partial\mathcal{N}(v)$ are shown as a function of $\mathcal{N}(v)/\rho^2$. As seen also visually, the result of imposing $\mathcal{N}(v) = \rho^2$ is not very close to the optimal parameter choice. This phenomenon is observed also in the case of natural images (Figs. 10 and 12). A better choice of imposing a specific value for $\mathcal{N}(v)$ is about $\frac{1}{2}\rho^2$. See the examples in Section 4. In Fig. 8 one can observe that the behavior of $SNR^{H^{-1}}$ is similar to that of the classical SNR^{L^2}. Specifically, the maximum is obtained in similar values of $\mathcal{N}(v)$. At the bottom of Fig. 8 it is shown that the estimated value of $\partial\mathcal{I}(n, v)/\partial\mathcal{N}(v)$ is quite similar to the real value. We plot the $\frac{1}{2}$ mark (dash-dot line), that indicates optimal SNR (see Eq. (58)). This behavior is similar to our experience with $\mathcal{H} = L^2$ (see [28]).

Similar results are obtained with a part of Lena image, Figs. 9, 10, and with the Cameraman image, Figs. 11, 12.

Table 1 summarizes the performance results, in terms of $SNR^{H^{-1}}$, of the processed images.

8. Discussion and Conclusion

In this paper the ROF [42] model is generalized into a BV-Hilbert space model, where the total-variation semi-norm is used for the structural part and a tunable Hilbert-space norm, controlled by a liner operator K, is used for the oscillatory part. Provided K is symmetric and positive, the model attains a unique solution. We proposed numerical implementations to solve the model, as well as the constrained problem, based on the projection algorithm of [12] and using some results of [5]. We have proved convergence of the corresponding algorithms.

This type of regularization method can be used for different purposes. In modern variational image-processing it is widely accepted that the total-variation energy is highly adequate to capture the structural part of an image. However, for the complementary energy of the oscillatory part there are many models which depend on the class of the image, type of the noise and/or texture, and the purpose of the regularization. Our general model allows a mathematically well-founded way to design and implement new types of such energies, by using different K operators. In this paper we based our numerical examples on the Osher-Sole-Vese model [38], where the H^{-1} norm is used ($K = -\Delta^{-1}$). As part of a larger study of the decomposition problem [8], the authors, together with Chan and Osher, proposed a TV-Gabor model for structure-texture decomposition, based on the ideas and theory presented here.

In this paper, however, our focus was given to the denoising problem, and more specifically, how to choose automatically λ, the weight parameter between the TV and Hilbert-space energies. We have suggested two basic methods for this. The first one relies on the constrained problem, where one imposes a norm constraint on the oscillatory part. For the OSV model, our experiments show that imposing the square norm of the residual to be equal to the square norm of the noise (as in the ROF model) gives too strong denoising. This is due to the high penalty of the H^{-1} norm for the very low frequency components of the noise (see Fig. 1). A better choice, which fits most images, is to choose half of the square norm. This gives a fast and automatic denoising with quite good results in most cases.

We have also been able to extend the work of [28] to our new framework. This has given us a more accurate algorithm to automatically denoise an image, based this time on SNR like optimum criterion. The basic idea is to aim at finding λ which maximizes the SNR of the denoised image. This is achieved by using an optimality condition which is based on the $\mathcal{H}$ inner product of the oscillatory part and an estimate of the noise. This method is more computationally intensive but gives higher quality results, which are very close to the optimal that could be obtained. Some SNR bounds and performance analysis, are given with respect to the classical constrained problem and the optimal strategy. A more detailed version of this paper can be found in [6].

In a future work, other $\mathcal{H}$-Hilbert spaces may be proposed, different than the ones considered in this paper for denoising purposes. Choosing a well-suited kernel for colored noise, for instance, could lead us to new and better adaptive frequency denoising algorithms. Another direction we want to explore is spatial adaptivity as it was recently done in [27, 26].

Acknowledgments

We would like to thank Prof. Stanley Osher and Prof. Tony Chan for their encouragement and support.

J.-F. Aujol was supported by Grants from the NSF under contracts DMS-9973341, ACI-0072112,

INT-0072863, the ONR under contract N00014-03-1-0888, the NIH under contract P20 MH65166, and the NIH Roadmap Initiative for Bioinformatics and Computational Biology U54 RR021813 funded by the NCRR, NCBC, and NIGMS. G. Gilboa was supported by Grants from the NSF under contracts ITR ACI-0321917, DMS-0312222, and the NIH under contract P20 MH65166.

References

1. A. Almansa, V. Caselles, G. Haro, and B. Rouge, "Restoration and zoom of irregularly sampled, blurred and noisy images by accurate total variation minimization with local constraints," accepted in *SIAM Journal on Multiscale Modelling and Simulation*.
2. G. Aubert, and J.F. Aujol, "Modeling very oscillating signals. Application to image processing," *Applied Mathematics and Optimization*, Vol. 51, No. 2, pp. 163–182, March/April 2005.
3. G. Aubert and P. Kornprobst, *"Mathematical Problems in Image Processing*, Vol. 47, of *Applied Mathematical Sciences,"* Springer-Verlag, 2002.
4. J.F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, "Image decomposition into a bounded variation component and an oscillating component," *JMIV*, Vol. 22, No. 1, pp. 71–88, January 2005.
5. J.F. Aujol and A. Chambolle, "Dual norms and image decomposition models," *International Journal on Computer Vision*, Vol. 63, No. 1, pp. 85–104, June 2005.
6. J.F. Aujol and G. Gilboa, "Implementation and parameter selection for BV-Hilbert space regularizations," UCLA CAM Report 04-66, ftp://ftp.math.ucla.edu/pub/camreport/cam04-66.pdf, 2004.
7. J.F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture decomposition by a TV-Gabor model," In *VLSM 05*, 2005.
8. J.F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition — modeling, algorithms, and parameter selection," UCLA CAM Report 05-10, ftp://ftp.math.ucla.edu/pub/camreport/cam05-10.pdf, International Journal of Computer Vision, Vol. 67, No. 1, pp. 111–136, April 2006.
9. J. Bect, G. Aubert, L. Blanc-Féraud, and A. Chambolle, "A l^1 unified variational framwork for image restoration," In *ECCV 2004*, May 2004.
10. A. Buades, B. Coll, and J.M. Morel, "A review of image denoising algorithms, with a new one," *SIAM Journal on Multiscale Modelling and Simulation*, Vol. 4, No. 2, pp. 490–530, 2005.
11. J.L. Carter, *"Dual methods for total variation-based image restoration,"* PhD thesis, UCLA, Advisor: T.F. Chan, 2001.
12. A. Chambolle, "An algorithm for total variation minimization and applications," *JMIV*, Vol. 20, pp. 89–97, 2004.
13. A. Chambolle and P.L. Lions, "Image recovery via total variation minimization and related problems," *Numerische Mathematik*, Vol. 76, No. 3, pp. 167–188, 1997.
14. T. Chan and S. Esedoglu, "Aspects of total variation regularized l^1 function approximation," CAM report 04-07, to appear in SIAM Journal on Applied Mathematics, 2004.
15. T.F. Chan, G.H. Mulet, and P. Mulet, "A nonlinear primal-dual method for total variation minimization and related problems," *SIAM Journal of Scientific Computing*, Vol. 20, No. 6, pp. 1964–1977, 1999.
16. P.G. Ciarlet, *"Introduction à l'analyse numérique matricelle et à l'optimisation*," Mathématiques appliquées pour la maitrise. Masson, 1982.
17. P. L. Combettes and J. C. Pesquet, "Image restoration subject to a total variation constraint," *IEEE Transactions on Image Processing*, Vol. 13, No. 9, pp. 1213–1222, 2004.
18. P.L. Combettes and J. Luo, "An adaptative level set method for nondifferentiable constrained image recovery," *IEEE Transactions on Image Processing*, Vol. 11, No. 11, pp. 1295–1304, 2002.
19. P.L. Combettes and H.J. Trussel, "The use of noise properties in set theoretic estimation," *IEEE Transactions on Signal Processing*, Vol. 39, No. 7, pp. 1630–1641, 1991.
20. P.L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Journal on Multiscale Modeling and Simulation*, Vol. 4, 2005.
21. I. Daubechies and G. Teschke, "Variational image restoration by means of wavelets: simultaneous decomposition, deblurring and denoising," *Applied and Computational Harmonic Analysis*, Vol. 19, No. 1, pp. 1–16, 2005.
22. Y. Dodge, *"Statistical Data Analysis based on L1-norm and Related Methods*," North-Holland, Amsterdam, 1987.
23. I. Ekeland and R. Temam, *"Convex Analysis and Variational Problems*," Amsterdam: North Holland, 1976.
24. I.A. Frigaard, G. Ngwa, and O. Scherzer, "On effective stopping time selection for visco-plastic nonlinear diffusion filters used in image denoising," *SIAM Journal on Applied Mathematics*, Vol. 63, No. 6, pp. 1911–1934, 2003.
25. G. Gilboa, N. Sochen, and Y.Y. Zeevi, "Estimation of optimal PDE-based denoising in the SNR sense," *To appear in IEEE Transactions on Image Processing*, 2006.
26. G. Gilboa, N. Sochen, and Y.Y. Zeevi, "Variational denoising of partly-textured images by spatially varying constraints," *To appear in IEEE Transactions on Image Processing*, 2006.
27. G. Gilboa, N. Sochen, and Y.Y. Zeevi, "Texture preserving variational denoising using an adaptive fidelity term," In *Proc. VLSM 2003, Nice, France*, pp. 137–144, 2003.
28. G. Gilboa, N. Sochen, and Y.Y. Zeevi, "Estimation of optimal PDE-based denoising in the SNR sense," CCIT report No. 499, Technion, August, see http://www.math.ucla.edu/∼gilboa/. 2004.
29. G. Gilboa, N. Sochen, and Y.Y. Zeevi, "Estimation of the optimal variational parameter via SNR analysis," In *Scale-Space*, of *Lecture Notes in Computer Science*, Vol. 3459, pp. 230–241, April 2005.
30. A. Haddad and Y. Meyer, "Variational methods in image processing," UCLA CAM report, 04-52, 2004.
31. J.B. Hiriart-Urruty and C. Lemarechal, *"Convex Analysis ans Minimisation Algorithms I*, of *Grundlehren der mathematischen Wissenschaften*," Springer-Verlag, Vol. 305, 1993.
32. Yves Meyer, "Oscillating patterns in image processing and in some nonlinear evolution equations," The Fifteenth Dean Jacquelines B. Lewis Memorial Lectures, March 2001.
33. V. A. Morosov, "On the solution of functional equations by the method of regularization," *Soviet Math. Dokl.*, Vol. 7, pp. 414–417, 1966.

34. M. Nikolova, "A variational approach to remove outliers and impulse noise," *JMIV*, Vol. 20, No. 1–2, pp. 99–120, 2004.
35. A. Obereder, S. Osher, and O. Sherzer, "On the use of dual norms in bounded variation type regularization," UCLA CAM report 04-35, 2004.
36. S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation based image restoration," *Multiscale Modeling and Simulations*, Vol. 4, No. 2, pp. 460–489, 2005.
37. S. Osher and O. Sherzer, "G-norm properties of bounded variation regularization," UCLA CAM report 04-23, 2004.
38. S.J. Osher, A. Sole, and L.A. Vese, "Image decomposition and restoration using total variation minimization and the H^{-1} norm," *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, Vol. 1, No. 3, pp. 349–370, 2003.
39. M. Navara P. Mrázek, "Selection of optimal stopping time for nonlinear diffusion filtering," *IJCV*, Vol. 52, No. 2/3, pp. 189–203, 2003.
40. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *PAMI*, Vol. 12, No. 7, pp. 629–639, 1990.
41. T. Rockafellar, *"Convex Analysis*, of *Grundlehren der mathematischen Wissenschaften*," Princeton University Press, second edition, Vol. 224, 1983.
42. L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, Vol 60, pp. 259–268, 1992.
43. Alon Spira, Ron Kimmel, and Nir Sochen, "Efficient Beltrami flow using a short time kernel," In *Scale-Space*, of *Lecture Notes in Computer Science*, Vol. 2695, pp. 511–522, January 2003.
44. J.L. Starck, M. Elad, and D.L. Donoho. "Image decomposition: separation of texture from piecewise smooth content," *IEEE Transactions on Image Processing*, Vol. 14, No. 10, pp. 1570–1582, 2005.
45. A.N. Tikhnonv and V.Y. Arsenin, *"Solutions of Ill-posed problems*, W.H. Winston," Washington D.C, 1977.
46. L.A. Vese and S.J. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *Journal of Scientific Computing*, Vol. 19, pp. 553–572, 2003.
47. C.R. Vogel, *"Computational methods for inverse problems*," Philadelphia, PA. SIAM, 2002.
48. J. Weickert, "Coherence-enhancing diffusion of colour images," *IVC*, Vol. 17, pp. 201–212, 1999.
49. Wotao Yin, Donald Goldfarb, and Stanley Osher, "Total variation based image cartoon-texture decomposition," UCLA CAM Report 05-27, 2005.

Jean-François Aujol graduated from "1' Ecole Normale Supérieure de Cachan" in 2001. He was a PH'D student in Mathematics at the University of Nice-Sophia-Antipolis (France). He was a member of the J.A. Dieudonné Laboratory at Nice, and also a member of the Ariana research group (CNRS/INRIA/UNSA) at Sophia-Antipolis (France). His research interests are calculus of variations, nonlinear partial differential equations, numerical analysis and mathematical image processing (and in particular classification, texture, decomposition model, restoration). He is Assistant Researcher at UCLA (Math Department).

Guy Gilboa is currently an Assistant Researcher (postdoctoral position) at the Department of Mathematics, UCLA, hosted by Prof. Stanley Osher. He received his Ph.D. in Electrical Engineering from the Technion—Israel Institute of Technology, in 2004. He previously worked for three years at Intel Development Center, Haifa, Israel, in the design of processors.

His main research interests are related to variational and PDE-based processes applied to image enhancement, denoising and decomposition.