Implementation and Parameter Selection for BV-Hilbert Space Regularizations

Jean-François Aujol [*]& Guy Gilboa [†]

UCLA, Departement of Mathematics,
Los Angeles, California 90095
email: {aujol,gilboa}@math.ucla.edu

November 17, 2004

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 ROF and OSV models and opens way for new regularizations methods, that have not been considered yet. We provide a straightforward numerical implementation, following Chambolle's projection algorithm. In order for such schemes to be practical, a systematic method for automatic parameter selection is imperative. The ROF 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 GSZ 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.

[*]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.

[†]Supported by grants from the NSF under contracts ITR ACI-0321917, DMS-0312222, and the NIH under contract P20 MH65166.

A necessary condition for maximal SNR is provided. Lower and upper bounds on the SNR performance of ROF and GSZ type strategies are established, under quite general assumptions.

Key-words: Image restoration, automatic parameter selection, BV, H^{-1}, Hilbert space, SNR, projection, total-variation, denoising.

AMS subject classifications: 46E, 49J, 49M, 65K, 68U.

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 Gaussian noise. We assume that the input image f is composed of the original image s and additive uncorrelated noise n 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

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

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:

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

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. Most of the research in the 90's was focused on the smoothness term. A main contribution made by ROF [30] 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. Similar results were obtained at the time by a closely related PDE method of nonlinear diffusion processes [29]. 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 [22] the attention to the role of the fidelity term has increased. In [22], the author has analyzed the Mathematical properties of the ROF model [30]. 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 [32], followed by [27, 3, 4, 31, 14, 25]. [22] has also raised new theoretical issues [1, 26, 24, 21].

Also, other types for fidelity terms for different noise models were suggested [23, 10].

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 steady-state. A more accurate way to compute the ROF solution is to use dual formulations [11, 7]. Recently, Chambolle has proposed a projection algorithm based on duality to solve the ROF problem [8]. In [4] the authors have proposed a modification of the projection algorithm to solve the OSV problem [27]. In this paper, we will generalize Chambolle's projection algorithm to a large class of functionals.

For a general overview of PDE-based restoration methods, we refer the reader to [2, 6].

We focus our attention on finding the 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 parameter 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.

This aspect of the problem was often disregarded, and the parameter was chosen manually by trial and error. Aiming at achieving automatic denoising algorithms, systematic methods for choosing λ are required. The most famous algorithm for choosing λ is the ROF constrained optimization formulation. We give a generalization to this method for the broader minimization problem:

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

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 estimate the $\mathcal{H}$-norm of white Gaussian noise from its standard deviation σ. Notice that the constrained problem has recently been addressed in [4] in the case of Meyer's G norm. The closely related problem of criteria for the stopping time of nonlinear diffusions was examined by [28] and to some extent by [33]. Physical considerations for solving the stopping time problem for the visco-plastic fluid model were suggested in [17].

One should also mention the approach developed in [13]: in this work, the authors solve the problem

$$\inf_u \|f-u\|_{\mathcal{H}}^2 \text{ subject to } \int_\Omega |Du| = \tau, \tag{6}$$

which is similar to (5). They argue that this formulation leads to less staircase effect. The attention is restricted to $\mathcal{H} = L^2$. Moreover, one needs to estimate τ - the total variation of the noise-free image, which is probably more difficult than estimating the noise variance.

The underlying assumption of ROF [30] is that the denoising process works well, therefore what is filtered is mostly noise: $v \approx n$. A natural condition is then to impose $\mathrm{var}(v) = \mathrm{var}(n) = \sigma^2$. 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 [28] and [33] rely on similar assumptions and "confuse" texture with noise, as shown in [18]. GSZ [18, 20] 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. This condition was then estimated, reaching quite close results to the optimal solution (less than $0.1db$ difference on average, for a collection of natural images). This method can work also on textured images, when the denoising is not very good. 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 [8] 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 will be used in the paper.

2.1 L^2 inner product

2.1.1 Definition

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 will be endowed with the L^2 inner product:

$$\langle u, v \rangle_{L^2} = \sum_{1 \le i,j \le N} u_{i,j} v_{i,j} \tag{7}$$

and the norm

$$\|u\|_{L^2} = \sqrt{\langle u, u\rangle_{L^2}} \tag{8}$$

We will often consider the following subspace of X:

$$X_0 = \{x \in X \ / \ \sum_{i,j} x_{i,j} = 0\} \tag{9}$$

2.1.2 Discrete Fourier transform

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

$$\mathcal{F}(f)(p,q) = F(p,q) = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} f(m,n) e^{-j(2\pi/N)pm} e^{-j(2\pi/N)qn} \tag{10}$$

and the inverse transform is:

$$f(m,n) = \frac{1}{N^2}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1} F(p,q) e^{j(2\pi/N)pm} e^{j(2\pi/N)qn} \tag{11}$$

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:

$$\mathcal{F}(\Delta f)(p,q) = 2\left(\cos\left(\frac{2\pi}{N}p\right) + \cos\left(\frac{2\pi}{N}q\right) - 2\right)\mathcal{F}(f)(p,q) \tag{12}$$

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

$$\mathcal{F}(\Delta^{-1} f)(p,q) = \frac{1}{2\left(\cos\left(\frac{2\pi}{N}p\right) + \cos\left(\frac{2\pi}{N}q\right) - 2\right)}\mathcal{F}(f)(p,q) \tag{13}$$

These basic results will prove useful when we will consider H^{-1} in subsection 2.3.2.

2.1.3 $\mathcal{H}$ Hilbert space

In what follows, we will consider a general family of Hilbert spaces. We consider an operator K such that:

1. K is a linear symmetric operator.
2. $ker(K) \bigcap X_0 = \{0\}$ where we recall that $ker(K) = \{x \in X \ / \ K(x) = 0\}$ and that X_0 is defined by (9)
3. $X_0 \subset dom(K)$ where $dom(K) = \{x \in X \ / \ \|K(x)\|_{L^2} < +\infty\}$.

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

$$\langle f, g \rangle_{\mathcal{H}} = \langle f, Kg \rangle_{L^2} \tag{14}$$

This defines a 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).

2.2 Total variation regularization

Up to now, we have only focused on the data term. Here we get interested in the regularization term.

2.2.1 Definition

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

$$(\nabla u)^1_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < N \\ 0 & \text{if } i = N \end{cases} \tag{15}$$

and

$$(\nabla u)^2_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } j < N \\ 0 & \text{if } j = N \end{cases} \tag{16}$$

The discrete total variation of u is then defined by:

$$J(u) = \sum_{1 \leq i,j \leq N} |(\nabla u)_{i,j}| \tag{17}$$

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

$$\begin{aligned}(\mathrm{div}\,(p))_{i,j} &= \begin{cases} p^1_{i,j} - p^1_{i-1,j} & \text{if } 1 < i < N \\ p^1_{i,j} & \text{if i=1} \\ -p^1_{i-1,j} & \text{if i=N} \end{cases} \\ &+ \begin{cases} p^2_{i,j} - p^2_{i,j-1} & \text{if } 1 < j < N \\ p^2_{i,j} & \text{if j=1} \\ -p^2_{i,j-1} & \text{if j=N} \end{cases}\end{aligned} \tag{18}$$

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

2.2.2 G space:

We are now in position to introduce the discrete version of Meyer's space G [22, 3, 1].

Definition 1

$$G = \{x \in X \ / \ \exists g \in X \times X \ \textit{such that}\ x = \mathrm{div}\,(g)\} \tag{19}$$

and if $x \in G$:

$$\begin{aligned}\|x\|_G &= \inf \{\|g\|_\infty \ / \ x = \mathrm{div}\,(g), \\ &\qquad g = (g^1, g^2) \in X \times X, |g_{i,j}| = \sqrt{(g^1_{i,j})^2 + (g^2_{i,j})^2}\Big\}\end{aligned} \tag{20}$$

where $\|g\|_\infty = \max_{i,j} |g_{i,j}|$.

Moreover, we will denote:

$$G_\mu = \{x \in G \ / \ \|x\|_G \leq \mu\} \tag{21}$$

Remark: The following result is proved in [3]:

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

2.2.3 Convex analysis

We recall that the Legendre-Fenchel transform of J is [15, 8]:

$$J^*(v) = \sup_{u \in X} \left(\langle u, v \rangle_{L^2} - J(u) \right) \tag{22}$$

Since J defined by (17) is homogeneous of degree one (i.e. $J(\lambda u) = \lambda J(u)$ $\forall u$ and $\lambda > 0$), it is then standard (see [15]) that J^* is the indicator function of some closed convex set, which turns out to be the set G_1 defined by (21):

$$J^*(v) = \chi_{G_1}(v) = \begin{cases} 0 & \text{if } v \in G_1 \\ +\infty & \text{otherwise} \end{cases} \tag{23}$$

This result is the key to Chambolle's projection algorithm [8], and it had first been noticed in [9].

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:

$$\|\nabla u\|_{L^2} = \sqrt{\sum_{1 \le i,j \le N} |\nabla u_{i,j}|^2} \tag{24}$$

We can now introduce the H norm:

$$\|u\|_H = \|\nabla u\|_{L^2} \tag{25}$$

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:

$$\langle f, g \rangle_H = \langle f, -\Delta g \rangle_{L^2} \tag{26}$$

2.3.2 H^{-1} space

We consider the polar semi-norm associated with (25):

$$\|v\|_{H^{-1}} = \sup_{\|u\|_H=1} \langle v, u\rangle_{L^2} = \sup_{\|\nabla u\|_{L^2}=1} \langle v, u\rangle_{L^2} \tag{27}$$

This is a discrete version of the H^{-1} norm.

The following result is proved in [4]:

$$\|f\|_{H^{-1}} = \sqrt{\langle -f, \Delta^{-1} f\rangle_{L^2}} \tag{28}$$

Using Parseval identity, one sees that:

$$\begin{aligned} \|f\|^2_{H^{-1}} &= \frac{1}{N^2}\langle -\mathcal{F}(f), \mathcal{F}(\Delta^{-1} f)\rangle_{L^2} \\ &= \frac{1}{N^2} \sum_{(p,q)\neq(0,0)} \frac{1}{2\left(2-\cos\left(\frac{2\pi}{N}p\right)-\cos\left(\frac{2\pi}{N}q\right)\right)} \left(\mathcal{F}(f)(p,q)\right)^2 \end{aligned} \tag{29}$$

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

$$\langle f, g\rangle_{H^{-1}} = (-f, \Delta^{-1} g)_{L^2} \tag{30}$$

Frequency understanding of H^{-1}

1. Using (29), 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 [22] in a more general framework). See also Figure 1 for an intuitive idea of H^{-1} filtering in 1 dimension.

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.

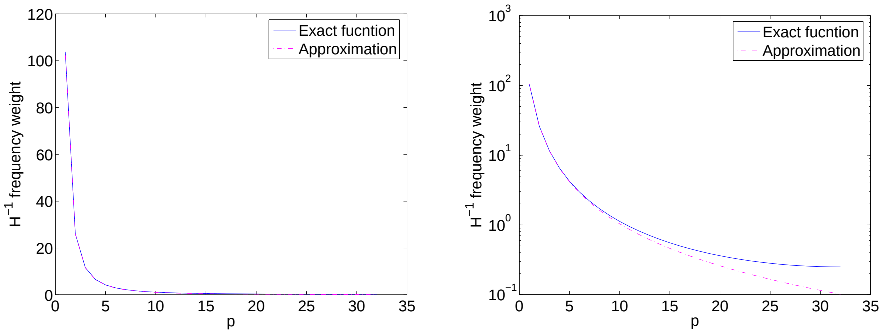

Figure 1: Frequency weight of the H^{-1} norm in one dimension. The exact weight function is $\frac{1}{2(1-\cos(2\pi p/N))}$, which can be approximated, using Taylor expansion, by $\frac{1}{(2\pi p/N)^2}$. On the left both functions are plotted in linear scale. As the difference is quite small, a log 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.

$$\inf_u \left(J(u) + \frac{\lambda}{2}\|f-u\|^2_{\mathcal{H}} \right) \tag{31}$$

All the results of this section have already been proved in the case $\mathcal{H} = L^2$ in [8], and we will draw our inspiration from this paper.

Proposition 2 *Problem (31) admits a unique solution $\hat{u}$.*

Proof. This is a very standard result [9]. The existence comes from the convexity of the functional, and the uniqueness from the fact that $ker(K) \bigcap ker(J) = \{0\}$. $\square$

Proposition 3 *If $\hat{u}$ is the solution of problem (31), then $\hat{v} = f - \hat{u}$ is the solution of the dual problem:*

$$\inf_v \left(\|v-f\|^2_{\mathcal{H}} + \frac{1}{\lambda} J^*(\lambda K v) \right) \tag{32}$$

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 (31), then:

$$0 \in \lambda K\left(\hat{u}-f\right) + \partial J(\hat{u}) \tag{33}$$

i.e.:

$$\lambda K\left(f-\hat{u}\right) \in \partial J(\hat{u}) \tag{34}$$

Hence

$$\hat{u} \in \partial J^*\left(\lambda K\left(f-\hat{u}\right)\right) \tag{35}$$

We then set $\hat{w} = K\left(f-\hat{u}\right)$, and we get:

$$0 \in K^{-1}\hat{w} - f + \partial J^*\left(\lambda \hat{w}\right) \tag{36}$$

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

$$\inf_w \left(\left\|K^{-1}w\right\|_{\mathcal{H}}^2 - 2\left\langle f, w\right\rangle_{L^2} + \frac{1}{\lambda}J^*\left(\lambda w\right) \right) \tag{37}$$

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

$$\left\|K^{-1}w\right\|_{\mathcal{H}}^2 - 2\left\langle f, w\right\rangle_{L^2} = \left\|K^{-1}w - f\right\|_{\mathcal{H}}^2 - \|f\|_{\mathcal{H}}^2 \tag{38}$$

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

$$\inf_w \left(\left\|K^{-1}w - f\right\|_{\mathcal{H}}^2 + \frac{1}{\lambda}J^*\left(\lambda w\right) \right) \tag{39}$$

We now set $\hat{v} = K^{-1}(\hat{w}) = f - \hat{u}$. we therefore get that $\hat{v}$ is a minimizer of (32). □

Since J^* is given by (23), 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 (31) is simply given by:

$$\hat{u} = f - P^{\mathcal{H}}_{K^{-1}G_{1/\lambda}}(f) \tag{40}$$

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

$$\min \left\{ \|\frac{1}{\lambda}K^{-1}\mathrm{div}\left(p\right) - f\|_{\mathcal{H}}^2 : p \;/\; |p_{i,j}|^2 - 1 \leq 0 \;\forall i,j = 1,\dots,N \right\} \tag{41}$$

The Karush-Kuhn-Tucker [12] conditions yield the existence of Lagrange multipliers $\alpha_{i,j} \geq 0$ associated to each constraint in problem (41), such that we have for each i, j:

$$-\left(\nabla\left(\frac{1}{\lambda}K^{-1}\mathrm{div}\,(p)-f\right)\right)_{i,j}+\alpha_{i,j}=0 \tag{42}$$

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

$$\alpha_{i,j}=\left|\left(\nabla\left(\frac{1}{\lambda}K^{-1}\mathrm{div}\,(p)-f\right)\right)_{i,j}\right| \tag{43}$$

We then propose the same kind of semi-implicit gradient descent scheme as in [8]:

$$p^0=0 \tag{44}$$

and

$$p_{i,j}^{n+1}=\frac{p_{i,j}^n+\tau(\nabla(K^{-1}\mathrm{div}\,(p^n)-\lambda f))_{i,j}}{1+\tau|(\nabla(K^{-1}\mathrm{div}\,(p^n)-\lambda f))_{i,j}|} \tag{45}$$

We can now show the following result.

Theorem 1 *If $\tau \leq \frac{1}{8\|K^{-1}\|_{L^2}}$, 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 [8]. The main difference is that here we work with the $\mathcal{H}$ norm instead of the L^2 norm. We denote by $\eta = \frac{p^{n+1}-p^n}{\tau}$. We have:

$$\begin{aligned}
&\left\|K^{-1}p^{n+1}-\frac{f}{\lambda}\right\|_{\mathcal{H}}^2-\left\|K^{-1}p^{n}-\frac{f}{\lambda}\right\|_{\mathcal{H}}^2\\
&=2\tau\left\langle K^{-1}\mathrm{div}\,\eta, K^{-1}\mathrm{div}\,p^n-\frac{f}{\lambda}\right\rangle_{\mathcal{H}}+\tau^2\left\|K^{-1}\mathrm{div}\,\eta\right\|_{\mathcal{H}}^2\\
&=2\tau\left\langle \mathrm{div}\,\eta, K^{-1}\mathrm{div}\,p^n-\frac{f}{\lambda}\right\rangle_{L^2}+\tau^2\left(-K^{-1}\mathrm{div}\,\eta,\mathrm{div}\,\eta\right)_{L^2}\\
&=-2\tau\left\langle \eta, \nabla\left(K^{-1}\mathrm{div}\,p^n-\frac{f}{\lambda}\right)\right\rangle_{L^2}+\tau^2\left\langle -K^{-1}\mathrm{div}\,\eta,\mathrm{div}\,\eta\right\rangle_{L^2}
\end{aligned}$$

We want to show that for τ small enough, the preceding quantity is non positive. Let us denote by $k=\|K^{-1}\|_{L^2}\|\|\mathrm{div}\,\|_{L^2}^2$. We then get:

$$
\begin{aligned}
&\left\|K^{-1}p^{n+1}+\frac{f}{\lambda}\right\|_{\mathcal{H}}^2-\left\|K^{-1}p^{n}+\frac{f}{\lambda}\right\|_{\mathcal{H}}^2\\
&\leq -\tau\left[2\left\langle \eta,-\nabla\left(K^{-1}\mathrm{div}\,p^n+\frac{f}{\lambda}\right)\right\rangle_{L^2}-k\tau\|\eta\|_{L^2}\right]
\end{aligned}
\tag{46}
$$

The rest of the proof is technical : one shows that (46) is non positive if $\tau \leq \frac{1}{k}$ (we refer the reader to [8] for further details).

There remains just to precise $k = \|K^{-1}\|_{L^2}\|\|\mathrm{div}\,\|_{L^2}^2$. It is easy to check that $\|\mathrm{div}\,\|_{L^2}^2 = 8$, and thus $k = 8\|K^{-1}\|_{L^2}$. $\square$

We have therefore shown how to solve problem (31) 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 [30], 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:

$$
\min_u \left\{ J(u) \;/\; \|u-f\|_{L^2}^2 = N^2\sigma^2 \right\} \tag{47}
$$

where N^2 is the size of the image. The equivalent problem we are interested in when restoring an image with our model (31) is then:

$$
\min_u \left\{ J(u) \;/\; \|u-f\|_{\mathcal{H}}^2 = N^2(\rho(\mathcal{H},N,\sigma))^2 \right\} \tag{48}
$$

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. 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:

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

$$(\|f\|_{\mathcal{H}})^* = \|f\|_{\mathcal{H}^*} \tag{49}$$

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 (22)) of $\|f\|_{\mathcal{H}}$.

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

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

□

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 expectancy by $E(Z) = \int z p_Z(z)\, dz$. The following proposition is a straightforward generalization of a result proved in [4]:

Proposition 4

$$E(\|n\|^2_{\mathcal{H}}) = C_{\mathcal{H}} E(\|n\|^2_{L^2}) = C_{\mathcal{H}} N^2 \sigma^2 \tag{50}$$

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

We recall that $Im(K) = \{y \in X \text{ such that } \exists x \in X \text{ such that } y = Kx\}$, and that $P^{L^2}_{Im(K)}(\delta)$ 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 [4] that:

$$C_{H^{-1}} = \frac{1}{N^2} \sum_{(p,q)\neq(0,0)} \frac{1}{2\left(2 - \cos\left(\frac{2\pi}{N}p\right) - \cos\left(\frac{2\pi}{N}q\right)\right)} \tag{51}$$

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

Moreover, it is also shown in [4] that $Var(\|n\|^2_{H^{-1}}) << E(\|n\|^2_{H^{-1}})$. In the rest of the paper, we will therefore make the following approximation:

$$\|n\|^2_{H^{-1}} = C_{H^{-1}} \|n\|^2_{L^2} \tag{52}$$

And more generally, we will assume that:

$$\|n\|^2_{\mathcal{H}} = C_{\mathcal{H}} \|n\|^2_{L^2} \tag{53}$$

Proof of Proposition 4: The proof is similar to the one of Proposition 3.5 in [4], but it is more technical in this general framework. We split the proof into two steps:
Step 1: We begin by computing $E(\|n\|^2_{\mathcal{H}})$. We consider the functional:

$$\inf_{u\in Im(K)} F(u) \tag{54}$$

where $F(u) = \frac{1}{2}\|u\|^2_{\mathcal{H}^*} - \langle n, u\rangle_{L^2}$ (we recall that $\|u\|_{\mathcal{H}^*}$ is defined in Lemma 1). F is convex and lsc (lower semi continuous). Hence (see [16] for instance) there exists u solving problem (54). Moreover, u is characterized by the fact that:

$$\langle K^{-1}u, h\rangle_{L^2} = \langle n, h\rangle_{L^2} \ \forall h \text{ in } Im(K)\ . \tag{55}$$

We denote by δ the image such that $\delta_{0,0} = 1$, and $\delta_{i,j} = 0$ otherwise. We denote by W the solution of the problem: $\inf_{u\in Im(K)} \left(\frac{1}{2}\|u\|^2_{\mathcal{H}^*} - \langle \delta u\rangle_{L^2}\right)$, and thus we have $\langle K^{-1}W, h\rangle_{L^2} = \langle \delta, h\rangle_{L^2}$ for all h in $Im(K)$. We introduce a discrete convolution: $f * g(x,y) = \sum_{i,j} f_{i,j} g_{x-i,y-j}$. It is easy to check that $\delta * f = f * \delta = f$ for all f (in fact, W is the Green function associated to problem (54)).

Let us now consider $u = n * W$; i.e. $u = \sum_{i,j} u_{i,j}$, with $u_{i,j}(x,y) = n_{i,j} W_{x-i,y-j}$. u is the solution of problem(54). If h belongs to $Im(K)$, then (using the fact that the convolution commutes with any linear operator) $\langle K^{-1}u, h\rangle_{L^2} = \langle n * K^{-1}W, \nabla h\rangle_{L^2} = \langle K^{-1}W, \tilde{n} * h\rangle_{L^2} = \langle \delta, \tilde{n} * h\rangle_{L^2} = \langle n * \delta, h\rangle_{L^2} = \langle n, h\rangle_{L^2}$ (we have used the fact that $\langle n * a, b\rangle_{L^2} = \langle a, \tilde{n} * b\rangle_{L^2}$ where $\tilde{n}_{i,j} = -n_{i,j}$).

We notice that $K^{-1}u = n * K^{-1}W$. Hence:
$\|u\|^2_{\mathcal{H}^*} = \langle n * W, n * K^{-1}W\rangle_{L^2} = (n, n * W\rangle_{L^2} = \sum_{k,l} n_{k,l}\ n * W_{k,l}$, i.e. $\|u\|^2_{\mathcal{H}^*} = \sum_{k,l} n_{k,l} \sum_{i,j} n_{i,j} W_{k-i,l-j} = \sum_{i,j,k,l} n_{i,j} n_{k,l} W_{k-i,l-j}$. And thus:

$$
\begin{aligned}
E\left(\|u\|^2_{\mathcal{H}^*}\right) &= E\left(\sum_{i,j,k,l} W_{k-i,l-j} n_{i,j} n_{k,l}\right) \\
&= \sum_{i,j,k,l} W_{k-i,l-j} E\left(n_{i,j} n_{k,l}\right)
\end{aligned}
$$

As the $n_{i,j}$ are independent, we have $E\left(n_{i,j} n_{k,l}\right) = E(n_{i,j})E(n_{k,l})$ whenever $(i,j) \neq (k,l)$. Moreover, $E(n_{i,j}) = 0$, and $E(n^2_{i,j}) = \sigma^2$.
Hence: $E\left(\|u\|^2_{\mathcal{H}^*}\right) = \sigma^2 \sum_{i,j} W_{i-i,j-j}$. We thus get:

$$
E\left(\|u\|^2_{\mathcal{H}^*}\right) = N^2 \sigma^2 W_{0,0} \tag{56}
$$

and from Lemma 1 , we know that $\|n\|_{\mathcal{H}} = \|u\|_{\mathcal{H}^*}$.
Step 2: We now want to compute $E(\|n\|^2_{L^2})$. We have: $\|n\|^2_{L^2} = \sum_{i,j} (n_{i,j})^2$. Hence $E(\|n\|^2_{L^2}) = \sum_{i,j} E\left((n_{i,j})^2\right)$. And we get: $E(\|n\|^2_{L^2}) = N^2\sigma^2$. We then get the result of Proposition 4 with $C_{\mathcal{H}} = W_{0,0} = \|P^{L^2}_{Im(K)}(\delta)\|^2_{\mathcal{H}}$ (thanks to (55) and Lemma 1), where $P^{L^2}_{Im(K)}(\delta)$ is the orthogonal projection of δ over $Im(K)$ with respect to the L^2 inner product. □

4.2 Solving the constrained problem

The problem we are therefore interested in when restoring an image with our model (31) is:

$$
\min_u \left\{ J(u) \ / \ \|u - f\|^2_{\mathcal{H}} = C_{\mathcal{H}} N^2 \sigma^2 \right\} \tag{57}
$$

where $C_{\mathcal{H}}$ is the constant given in Proposition 4. Since σ is less difficult to estimate than λ in (31), it is of practical interest to know how to solve (57) directly. The task is to find $\lambda > 0$ such that $\|P^{\mathcal{H}}_{K^{-1}G_{1/\lambda}}(f)\|^2_{\mathcal{H}} = C_{\mathcal{H}} N^2 \sigma^2$. For $s > 0$, let us set

$$g(s) = \|P^{\mathcal{H}}_{K^{-1}G_{1/s}}(f)\|_{\mathcal{H}} \tag{58}$$

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 [8]. The main difference relies in the use of the $\mathcal{H}$ inner product instead of the L^2 inner product. Let us set $s \le t$. We denote by $v = P^{\mathcal{H}}_{K^{-1}G_{1/s}}(f) = g(s)$ and by $w = P^{\mathcal{H}}_{K^{-1}G_{1/t}}(f) = g(t)$. It is easy to see that $g(s) \ge g(t)$ [8, 3].

Since $K^{-1}G_{1/s}$ is a convex set [5], we have:

$$\langle f - v, x - v \rangle_{\mathcal{H}} \le 0 \tag{59}$$

for every x in $K^{-1}G_{1/s}$. We also have for every y in $K^{-1}G_{1/t}$:

$$\langle f - w, y - w \rangle_{\mathcal{H}} \le 0 \tag{60}$$

We then denote by $\theta = s/t$. We choose $x = w/\theta$ in (59) and $y = \theta v$ in (60). We thus have:

$$\langle f - v, w - \theta v \rangle_{\mathcal{H}} \le 0 \text{ and } \langle f - w, \theta v - w \rangle_{\mathcal{H}} \le 0 \tag{61}$$

Hence

$$\langle \theta v - w, v - w \rangle_{\mathcal{H}} \le 0 \tag{62}$$

i.e.

$$\theta g(s)^2 - (1 + \theta) \langle v, w \rangle_{\mathcal{H}} + g(t)^2 \le 0 \tag{63}$$

Using the Cauchy Schwartz inequality, we get:

$$(g(t) - \theta g(s))(g(t) - g(s)) \le 0 \tag{64}$$

Since $g(s) \ge g(t)$, we deduce that $g(t) \ge \theta g(s)$, i.e. $tg(t) \ge sg(s)$. $\square$

Thanks to Lemma 2 we can propose the following algorithm, in order to solve (57) (similar to the one proposed in [8] to solve (47)). 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

$$v_0 = P^{\mathcal{H}}_{K^{-1}G_{1/\lambda_0}}(f) \tag{65}$$

with the algorithm described in the previous subsection, 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

$$v_{n+1} = P^{\mathcal{H}}_{K^{-1}G_{1/\lambda_{n+1}}}(f) \tag{66}$$

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 4.2 in [8]).

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 (57).*

This closes the generalization of Chambolle's results [8] 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 (31) is the Osher-Sole-Vese model [27].

$$\inf_u \left(J(u) + \frac{\lambda}{2}\|f - u\|^2_{H^{-1}} \right) \tag{67}$$

In [27], the authors write the associated Euler-Lagrange equations, and then compute the solution by solving a fourth order PDE.

Our algorithm (44)-(45) reduces in that case to the one that has been proposed in [4] to solve (67). We can precise Proposition 1:

Proposition 5 *If $\tau \leq \frac{1}{64}$, then $f - \lambda\Delta \operatorname{div} p^n \to \hat{u}$ (solution of (67)) 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 [8]).

Up to now, there had not been proposed any method to solve problem (57) in the case $\mathcal{H} = H^{-1}$, although it has been noticed in [4] that the Osher-Sole-Vese is a very good denoising model. This issue is now addressed by algorithm (65)-(66) (remembering that $C_{\mathcal{H}}$ is given by (51)).

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 [30], 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 [13] 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 $\sigma^2/2$ instead of σ^2.

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

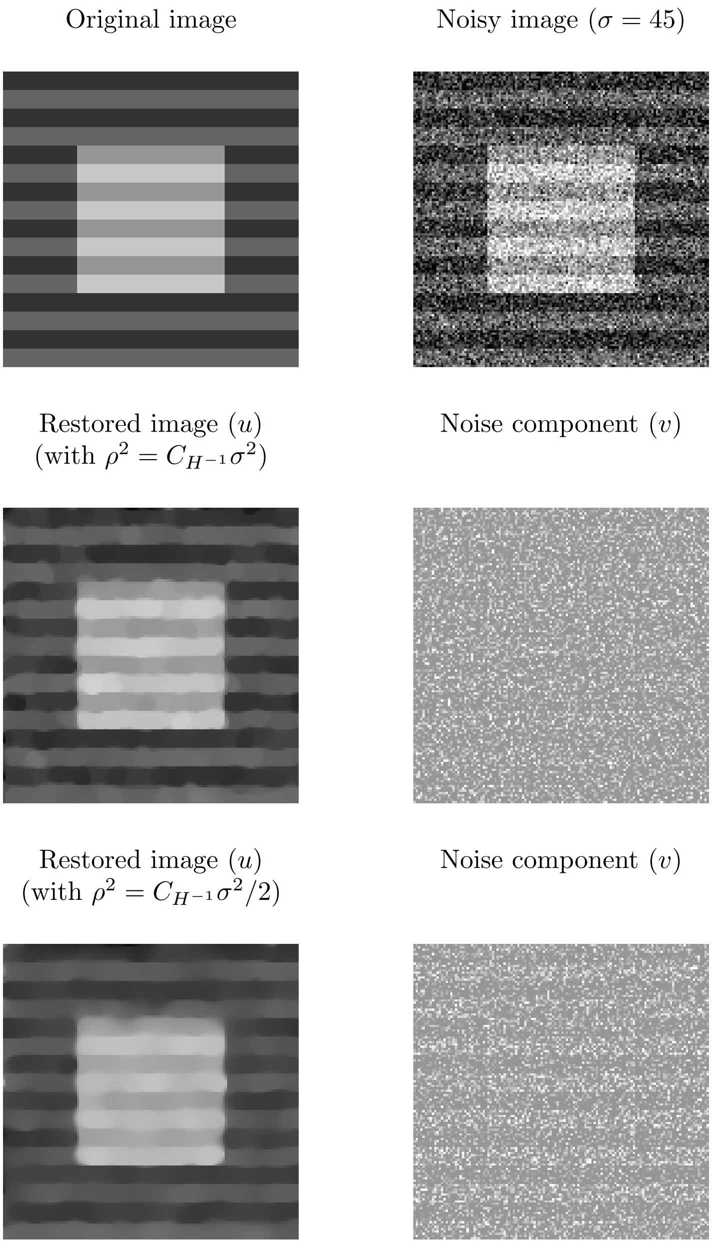

Figure 2: Automatic restoration of a synthetic image

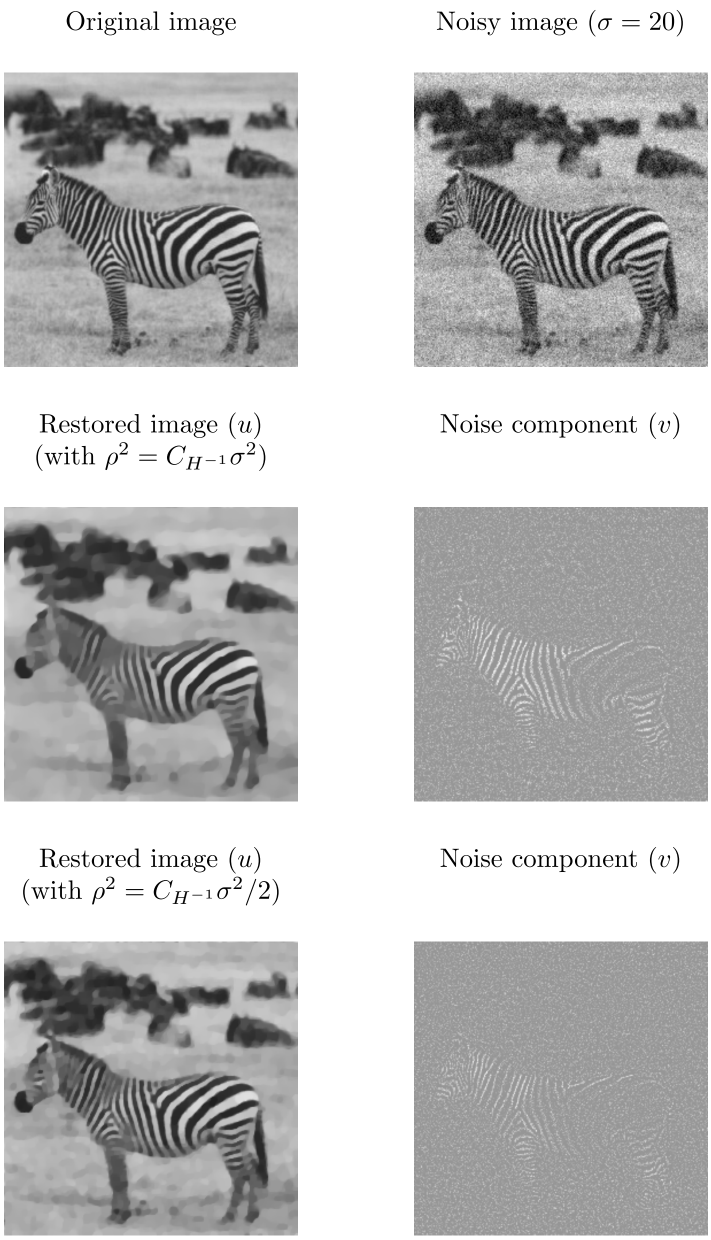

Figure 3: Automatic restoration of a zebra image

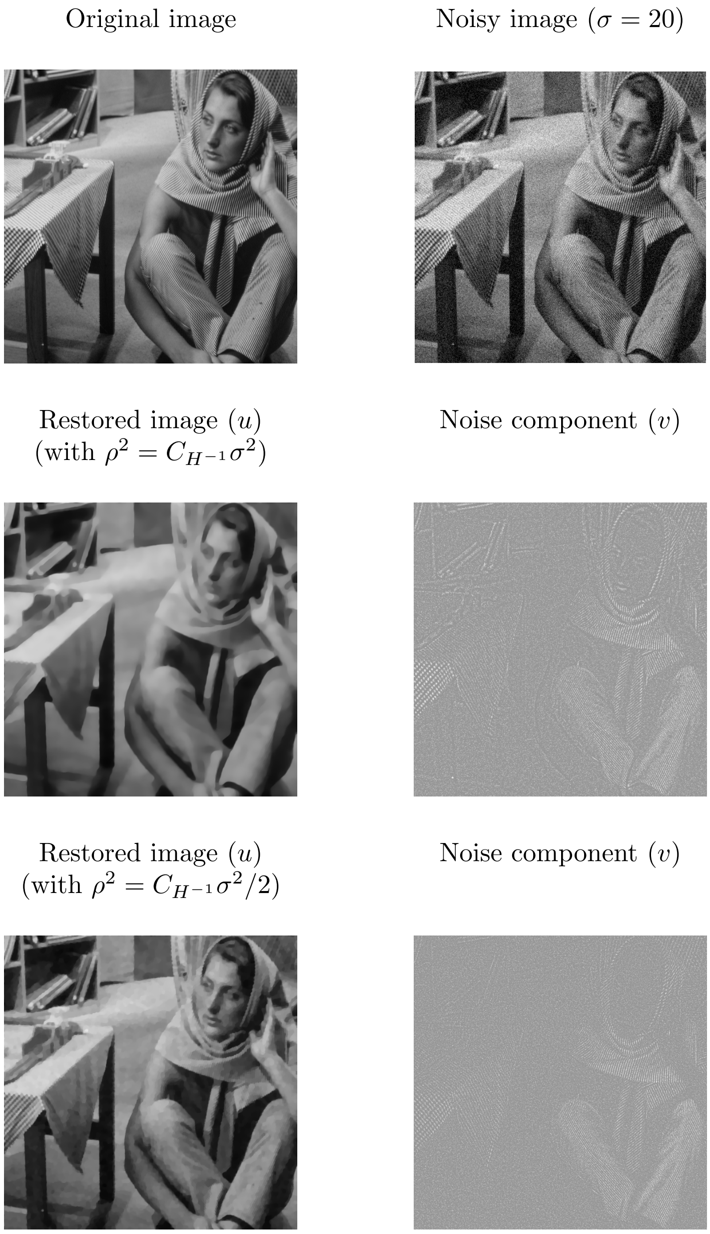

Figure 4: Automatic restoration of Barbara

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*:

$$\mathcal{I}(p,q) \doteq \frac{1}{|\Omega|} < p-\bar{p}, q-\bar{q} >_{\mathcal{H}}, \tag{68}$$

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

$$\mathcal{N}(p) \doteq \mathcal{I}(p,p) = \frac{1}{|\Omega|}\|p-\bar{p}\|^2_{\mathcal{H}}. \tag{69}$$

The above measures become the standard notions of empirical *covariance* and *variance*, respectively, for $\mathcal{H} = L^2$. We will 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

$$\inf_{u,v}\left(J(u) + \frac{\lambda}{2}\mathcal{N}(v)\right), \text{ subject to } f = u + v. \tag{70}$$

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.

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

$$SNR^{\mathcal{H}}(u) \doteq 10\log\frac{\mathcal{N}(s)}{\mathcal{N}(u-s)} = 10\log\frac{\mathcal{N}(s)}{\mathcal{N}(n-v)}, \tag{71}$$

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

$$\mathcal{N}(n) = \rho^2. \tag{72}$$

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

(71), (72) and (1):

$$SNR_0 \doteq SNR(f) = 10 \log \frac{\mathcal{N}(s)}{\mathcal{N}(n)} = 10 \log \frac{\mathcal{N}(s)}{\rho^2}. \tag{73}$$

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:

$$SNR_{opt} \doteq \max_{\lambda} SNR(u_\lambda) \tag{74}$$

where $u = u_\lambda$ attains the minimal energy of (70) 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 will 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 (70). This was proved by Chambolle-Lions [9] in the case $\mathcal{H} = L^2$ and could be generalized to our framework by using Proposition 2 and Lemma 2. We therefore regard 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:

$$\frac{\partial SNR}{\partial \mathcal{N}(v)} = 0. \tag{75}$$

Rewriting $\mathcal{N}(n - v)$ as $\mathcal{N}(n) + \mathcal{N}(v) - 2\mathcal{I}(n, v)$, and using (75) and (71), yields

$$\frac{\partial \mathcal{I}(n, v)}{\partial \mathcal{N}(v)} = \frac{1}{2}. \tag{76}$$

The meaning of this condition may not appear at first glance to be very clear. We therefore resort to our intuition: let us think of an evolutionary process with scale parameter $\mathcal{N}(v)$. We begin with $\mathcal{N}^0(v) = 0$ and increment the norm of v by a small amount $d\mathcal{N}(v)$, so that in the next step $\mathcal{N}^1(v) = d\mathcal{N}(v)$. The residual part of f, v, contains now both part of the noise and part of the signal. As long as in each step the noise is mostly filtered, that is $\frac{\partial \mathcal{I}(n,v)}{\partial \mathcal{N}(v)} > \frac{1}{2}$, then one should keep on with the process and SNR will increase. When we reach the condition of (76), noise and signal are equally filtered and one should therefore stop. If filtering is continued, more signal than noise is filtered (in norm sense) and SNR decreases.

There is also a possibility that the maximum is at the boundaries: If 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 SNR increases monotonically and is maximized when $\mathcal{N}(v) = \mathcal{N}(f)$ (the trivial constant solution $u = \bar{f}$). We will see later (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:

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.

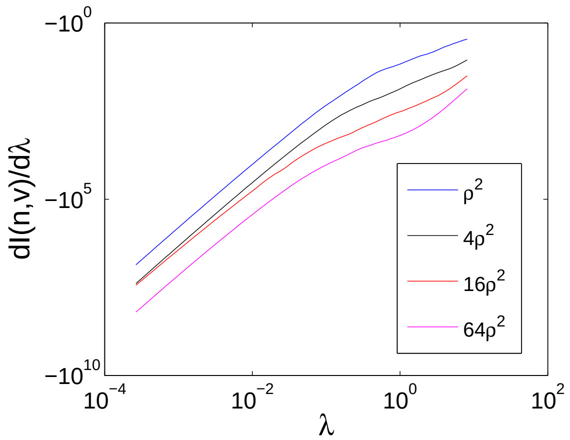

Figure 5: Precomputed term $\partial\mathcal{I}(n,v)/\partial\lambda$ as a function of λ (log scale). Graphs depict plots for different proportion of ρ^2 - 1 : 4 : 16 : 64, from upper curve to lower curve, respectively.

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 are showing 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. It appears that $\mathcal{I}(n,v)$ as a function of λ is almost independent from the underlying image and can be estimated with quite a good accuracy.

First, we need to compute the "statistics" by processing a patch of pure noise and measuring $\mathcal{I}(n,v)$ with respect to λ. This is done a single time for each noise norm ρ^2 and can be regarded as a look-up-table (see Fig. 5). 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:

$$\begin{aligned} \frac{\partial \mathcal{I}(n,v)}{\partial \mathcal{N}(v)} &= \frac{\partial \mathcal{I}(n,v)}{\partial \lambda}\frac{\partial \lambda}{\partial \mathcal{N}(v)} \\ &\approx \frac{\partial \mathcal{I}(n,v)}{\partial \lambda}\Big|_{f=patch}\frac{\partial \lambda}{\partial \mathcal{N}(v)}\Big|_{f=s+n}. \end{aligned} \tag{77}$$

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 was 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 (70) 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:

$$\mathcal{I}(u^s, n) = 0, \quad \mathcal{I}(u^n, s) = 0, \quad \forall \lambda \geq 0. \tag{78}$$

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:

$$\mathcal{I}(u^{s+n}, s) \leq \mathcal{I}(f, s), \quad \mathcal{I}(u^{s+n}, n) \leq \mathcal{I}(f, n), \quad \forall \lambda \geq 0. \tag{79}$$

We are investigating the possibility to characterize the appropriate spaces of s and n such that (78) and (79) are followed. In this paper this question is left open and we resort to the following definition:

Definition 2 ((s, n) pair) *An (s, n) pair consists of two uncorrelated signals s and n which obey conditions (78) and (79).*

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 the appendix. Theorem 3 implies that the denoising process has smoothing properties and consequently, there is no negative correlation between any two elements of U. This basic theorem will be later used to establish several bounds in our performance analysis.

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

$$\mathcal{N}(v) = \rho^2. \tag{80}$$

We define

$$SNR_{\rho^2} \doteq SNR(u)|_{\mathcal{N}(v)=\rho^2}. \tag{81}$$

We denote by (u_{ρ^2}, v_{ρ^2}) the (u, v) pair that obeys (80) and minimizes (70). We will now analyze this method for selecting u in terms of SNR.

Proposition 6 (SNR lower bound) *Imposing (80), for any (s, n) pair SNR_{ρ^2} is bounded from below by*

$$SNR_{\rho^2} \geq SNR_0 - 3dB, \tag{82}$$

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

Proof. From Theorem 3 we have $\mathcal{I}(n, v) \geq 0$, therefore,

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

$\square$

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 (80), 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 (80), $\mathcal{N}(v) = \rho^2$, and from Theorem 3, $\mathcal{I}(u, v) \geq 0$, therefore $\mathcal{N}(u) \leq h\rho^2$. Since $\mathcal{I}(u, s) \geq 0$ (Theorem 3) 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. $\square$

Definition 3 (Regular SNR) *We define the function* $SNR(\mathcal{N}(v))$ *as* regular *if (76) is a sufficient condition for optimality or if the optimum is at the boundaries.*

In Figs. 8, 10 and 12 one can observe that this assumption is valid for both examples of synthetic and natural images (see SNR plot as a function of $\mathcal{N}(v)/\rho^2$).

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 (78). 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 3).

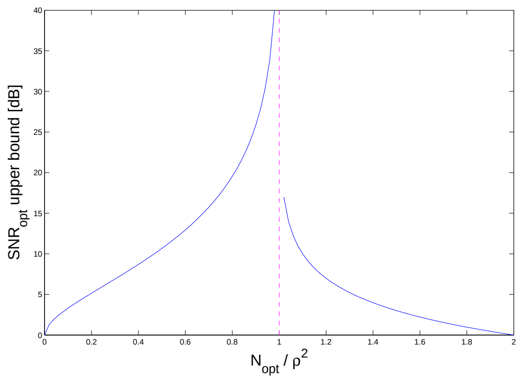

Figure 6: Visualization of Theorem 4: Upper bound of $SNR_{opt} - SNR_0$ as a function of $\mathcal{N}_{opt}/\rho^2$. For $\mathcal{N}_{opt} \to \rho^2$ the bound approaches $+\infty$.

We reach the upper bound by the following inequalities:

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

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 $\frac{\partial \mathcal{I}(n,v)}{\partial \mathcal{N}(v)}|_{\mathcal{N}(v)=0} < \frac{1}{2}$. $\square$

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

$$0 \leq SNR_{opt} - SNR_0 \leq \begin{cases} -10\log(1+\mathcal{N}_{opt}/\rho^2 - 2\sqrt{\mathcal{N}_{opt}/\rho^2}), & 0 \leq \mathcal{N}_{opt} < \rho^2 \\ -10\log(\mathcal{N}_{opt}/\rho^2 - 1), & \rho^2 < \mathcal{N}_{opt} \leq 2\rho^2 \end{cases} \tag{83}$$

Proof. By the SNR definition, (71), and expanding the norm expression, we have

$$SNR_{opt} - SNR_0 = 10\log\left(\frac{\rho^2}{\rho^2 + \mathcal{N}_{opt} - 2\mathcal{I}(n, v_{opt})}\right). \tag{84}$$

For the lower bound we use the relation shown in Proposition 8: $\mathcal{I}(n, v_{opt}) \geq \frac{1}{2}\mathcal{N}_{opt}$. 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. $\square$

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.

Proposition 9 (Energy change as a function of λ) *The energy parts of Eq. (70) vary as a function of λ as follows:*

$$\frac{\partial E_J}{\partial \lambda} > 0, \quad \frac{\partial E_v}{\partial \lambda} \leq 0, \quad \frac{\partial E_u}{\partial \lambda} \geq 0. \tag{85}$$

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

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

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 parameter (Section 5).

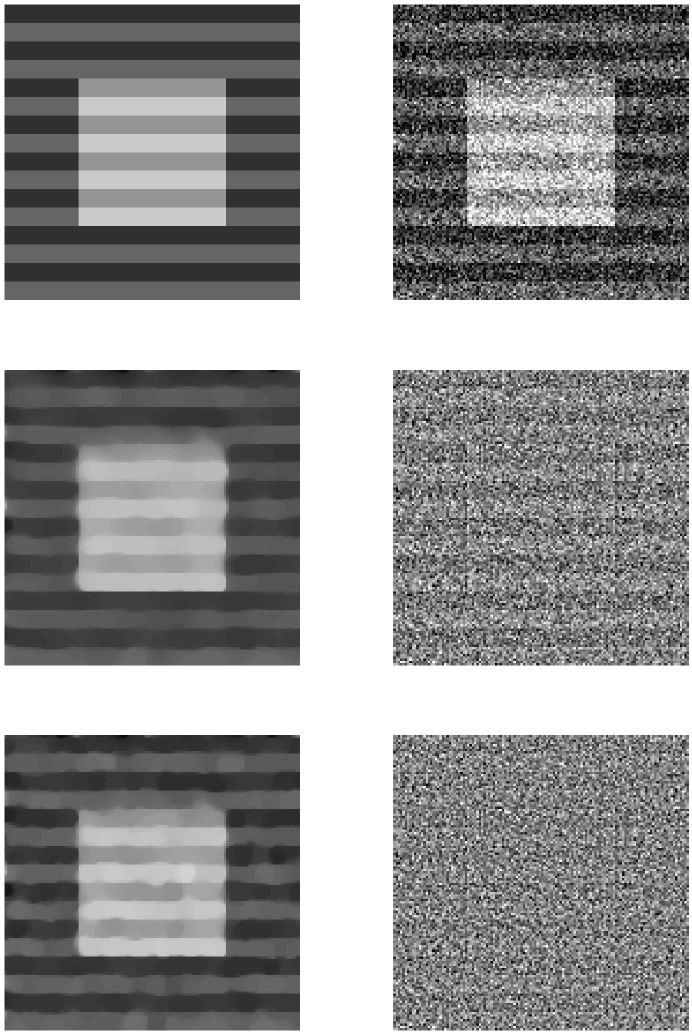

Figure 7: Denoising of a synthetic image. From left to right. Top row: original s, input image f. Second row: u, v of constrained problem $\mathcal{N}(v) = \rho^2$. Bottom row: u, v of SNR based selection. Noise standard deviation is $\sigma = 45$.

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. (76)). This behavior is similar to our experience with $\mathcal{H} = L^2$ (see [18]).

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.

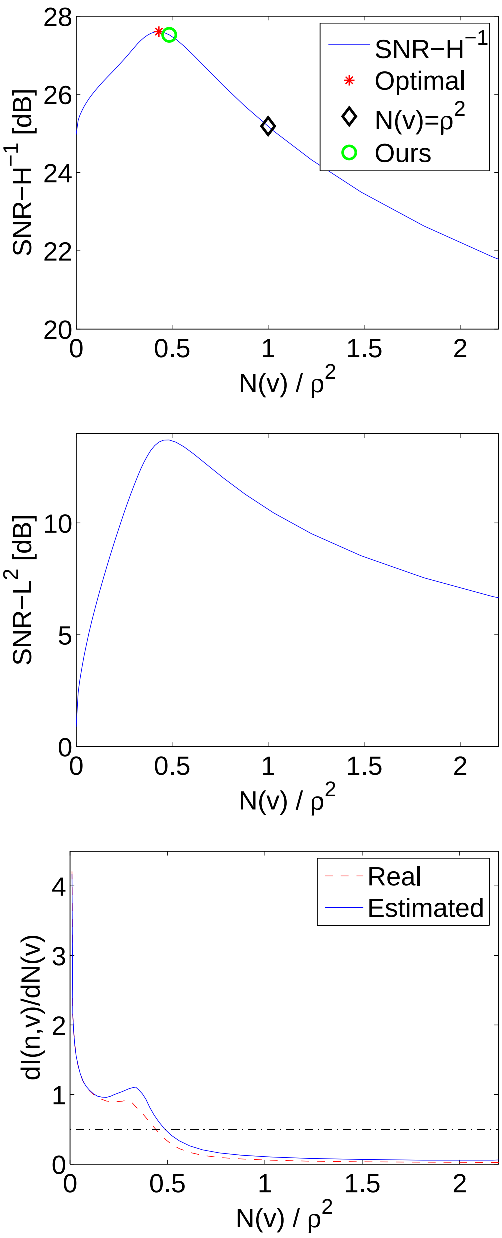

Figure 8: Denoising of synthetic image - SNR and covariance plots. Top: $SNR^{H^{-1}}$ as a function of $\mathcal{N}(v)$ with plots of the optimal, constrained and SNR-based selections ("Ours"). Middle row: a plot of the standard SNR^{L^2}, which behaves quite similarly to $SNR^{H^{-1}}$. Bottom row: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

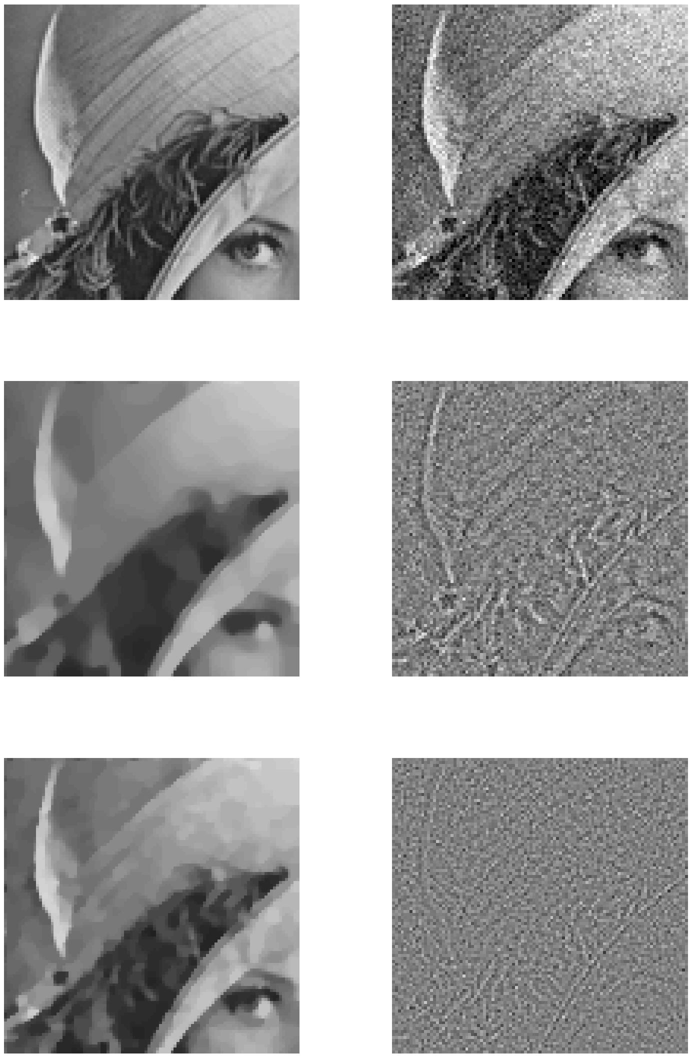

Figure 9: Denoising part of Lena image. From left to right. Top row: original s, input image f. Second row: u, v of constrained problem $\mathcal{N}(v) = \rho^2$. Bottom row: u, v of SNR based selection. Noise standard deviation is $\sigma = 20$.

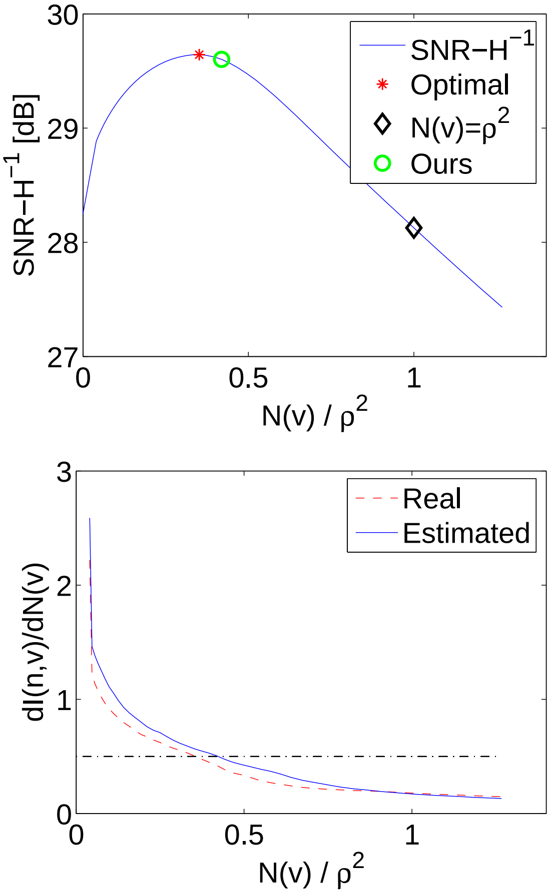

Figure 10: Denoising part of Lena image - SNR and covariance plots. Top: $SNR^{H^{-1}}$ as a function of $\mathcal{N}(v)$ with plots of the optimal, constrained and SNR-based selections ("Ours"). Bottom: estimated $\partial\mathcal{I}(n,v)/\partial\mathcal{N}(v)$ vs. the ground truth.

Figure 11: Denoising Cameraman image. 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. Noise standard deviation is $\sigma = 20$.

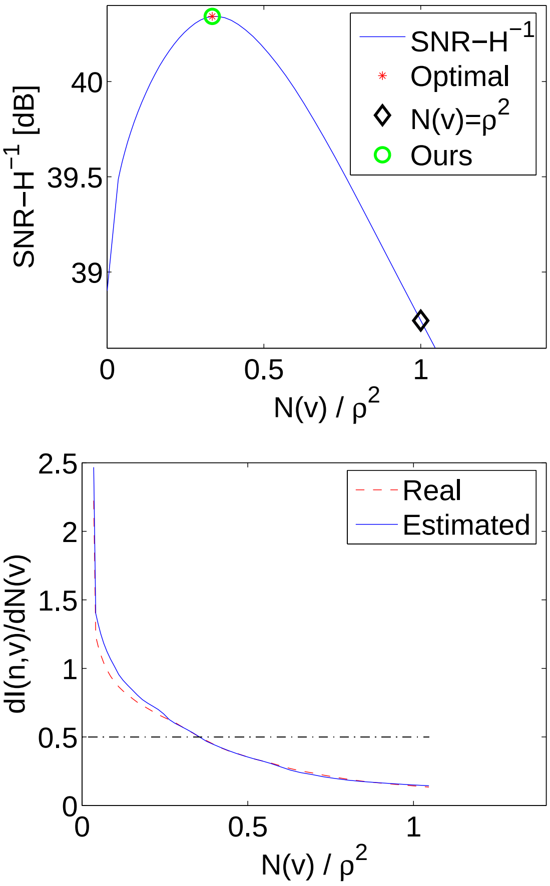

Figure 12: Denoising Cameraman image - SNR and covariance plots. Top: $SNR^{H^{-1}}$ as a function of $\mathcal{N}(v)$ with plots of the optimal, constrained and SNR-based selections ("Ours"). Bottom: estimated $\partial\mathcal{I}(n,v)/\partial\mathcal{N}(v)$ vs. the ground truth.

8 Conclusion

In this paper, we have generalized the ROF [30] restoration model into a BV-Hilbert space restoration model. Based on Chambolle's work [8], and using some results of [4], we have been able to propose projection algorithms to solve our new problem, as well as the constrained problem. We have mathematically proved the convergence of the corresponding algorithms. In particular, this gives a way to automatically denoise an image by using the OSV model [27]. We have also been able to extend the work of [18] to our new framework. This has given us another algorithm to automatically denoise an image, based this time on SNR like optimum criterion.

For the constrained problem, our experiments show that imposing the norm of the residual to be equal to the norm of the noise (as in the ROF model) gives too strong denoising. A better choice, which fits most images, is to choose half of the norm. This gives a fast and efficient automatic denoising. The latter method is more computationally intensive but gives higher quality results, which are very close to the optimal that could be obtained.

In a future work, we intend to propose other $\mathcal{H}$-Hilbert spaces, different than the ones considered in this paper. Choosing a well-suited kernel 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 [19].

Acknowledgements:

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

APPENDIX

A Proof of Theorem 3

Since $\mathcal{I}(q,r) = \mathcal{I}(r,q)$, the matrix is symmetric. The diagonal is the norm of each element, which is non negative. Therefore we have to check the inner product of the 10 elements of the upper right triangle. We consider below all 10 possible signal pairs and show that their inner product is non-negative.

$\mathcal{I}(s,n)$, $\mathcal{I}(f,s)$, $\mathcal{I}(f,n)$

Since s and n are not correlated, we have $\mathcal{I}(s,n)=0$, $\mathcal{I}(f,s)=\mathcal{I}(s+n,s)=\mathcal{N}(s)\geq 0$, $\mathcal{I}(f,n)=\mathcal{I}(s+n,n)=\mathcal{N}(n)\geq 0$.

$\mathcal{I}(u,v)$, $\mathcal{I}(f,u)$, $\mathcal{I}(f,v)$

Once we prove $\mathcal{I}(u,v)\geq 0$, then we readily have $\mathcal{I}(f,u)=\mathcal{I}(u+v,u)=\mathcal{N}(u)+\mathcal{I}(u,v)\geq 0$ and $\mathcal{I}(f,v)=\mathcal{I}(u+v,v)=\mathcal{N}(v)+\mathcal{I}(u,v)\geq 0$.

We follow the proof of Meyer [22]. As the (u,v) decomposition minimizes the energy of Eq. (70), we can write for any function $h\in BV$ and scalar $\epsilon>0$ the following inequality:

$$J(u-\epsilon h)+\lambda\mathcal{N}(v+\epsilon h)\geq J(u)+\lambda\mathcal{N}(v). \tag{86}$$

Replacing $\mathcal{N}(v+\epsilon h)$ by $\mathcal{N}(v)+\epsilon^2\mathcal{N}(h)+2\epsilon\mathcal{I}(v,h)$ we get

$$2\lambda\epsilon\mathcal{I}(v,h)\geq (J(u)-J(u-\epsilon h))-\lambda\epsilon^2\mathcal{N}(h).$$

Replacing h by u and dividing both sides by ϵ we get

$$2\lambda\mathcal{I}(v,u)\geq \frac{1}{\epsilon}\left(J(u)-J(u(1-\epsilon))\right)-\lambda\epsilon\mathcal{N}(u).$$

In the limit as $\epsilon\to 0$, the right term on the right-hand-side vanishes. Since J is a semi-norm and $\epsilon<1$, $\frac{1}{\epsilon}\left(J(u)-J(u(1-\epsilon))\right)=J(u)$ which is non-negative.

$\mathcal{I}(s,u)$, $\mathcal{I}(n,u)$

Let us first examine an equivalent minimization problem to minimizing (70). Since $v=s+n-u$, then u that minimizes E_J is

$$\begin{aligned} u &= \mathrm{argmin}_u\{J(u)+\lambda\mathcal{N}(s+n-u)\}\\ &= \mathrm{argmin}_u\{J(u)+\lambda(\mathcal{N}(s)+\mathcal{N}(n)+\mathcal{N}(u)\\ &\quad +2\mathcal{I}(s,n)-2\mathcal{I}(s,u)-2\mathcal{I}(n,u))\}.\end{aligned}$$

We can disregard expressions that do not involve u and, therefore, the equivalent energy functional to be minimized is:

$$\hat{E}_J(u)=J(u)+\lambda(\mathcal{N}(u)-2\mathcal{I}(s,u)-2\mathcal{I}(n,u)), \tag{87}$$

where $u = \text{argmin}_u\{\hat{E}_J(u)\}$. Since $\mathcal{I}(s,u) + \mathcal{I}(n,u) = \mathcal{I}(f,u) \geq 0$ at least one of the terms $\mathcal{I}(s,u)$ or $\mathcal{I}(n,u)$ must be non-negative. We will now show, by contradiction, that it is not possible that the other term be negative. Let us assume, without loss of generality, that $\mathcal{I}(s,u^{s+n}) \geq 0$ and $\mathcal{I}(n,u^{s+n}) < 0$. We denote the optimal (minimal) energy of (87) with $f = s+n$ as $\hat{E}^*_J|_{f=s+n}$. The energy can be written as

$$\begin{aligned} \hat{E}^*_J|_{f=s+n} &= \hat{E}_J|_{f=s+n}(u^{s+n}) \\ &= J(u^{s+n}) + \lambda(\mathcal{N}(u^{s+n}) - 2\mathcal{I}(s,u^{s+n}) - 2\mathcal{I}(n,u^{s+n})). \end{aligned} \tag{88}$$

On the other hand, according to condition (78), $\mathcal{I}(u^s,n) = 0$ and we have

$$\begin{aligned} &\hat{E}_J|_{f=s+n}(u^s) = J(u^s) + \lambda(\mathcal{N}(u^s) - 2\mathcal{I}(s,u^s)) \\ &= \hat{E}^*_J|_{f=s} \leq \hat{E}_J|_{f=s}(u^{s+n}) = J(u^{s+n}) + \lambda(\mathcal{N}(u^{s+n}) - 2\mathcal{I}(s,u^{s+n})). \end{aligned}$$

In the above final expression, adding the term $-\lambda 2\mathcal{I}(n,u^{s+n})$ we obtain the right hand side of expression (88). Since we assume $\mathcal{I}(n,u^{s+n}) < 0$, we get the following contradiction

$$\hat{E}_J|_{f=s+n}(u^s) < \hat{E}^*_J|_{f=s+n}.$$

Similarly, the opposite case $\mathcal{I}(n,u^{s+n}) \geq 0$ and $\mathcal{I}(s,u^{s+n}) < 0$ is not possible.

$\mathcal{I}(s,v)$, $\mathcal{I}(n,v)$

This follows directly from condition (79) as $\mathcal{I}(f,s) = \mathcal{I}(u,s) + \mathcal{I}(v,s)$ and $\mathcal{I}(f,n) = \mathcal{I}(u,n) + \mathcal{I}(v,n)$.

B Detailed Algorithms

We give below the detailed algorithm that was sketched in Section 5.2 for SNR-based parameter selection. Explanations about parameters and a few remarks appear hereafter.

Main

1. Parameters: $szp, N_p, \lambda^0,\ \lambda_r$.
2. Set $v^0 = 0$, $i = 0$.

3. Read precomputed table $\partial\mathcal{I}(n,v)/\partial\lambda$.

4. Loop

   (a) $i \leftarrow i+1,\ \lambda^i \leftarrow \lambda^{i-1}\lambda_r$.

   (b) Compute u^i by Eq. (70) with λ^i (use p of previous iteration as initial approximation for the projection)

   (c) $v^i \leftarrow f - u^i$.

   (d) $DEI^i \leftarrow DEI^i_{pre} \cdot (\lambda^i - \lambda^{i-1})/(\mathcal{N}(v^i) - \mathcal{N}(v^{i-1}))$. [$DEI$ stands for *Derivative of Estimated Inner product*]

   (e) until ($DEI^i < \frac{1}{2}$ (or ($i = N_p$))

5. Return u^{i-1}

Compute $\frac{\partial\mathcal{I}(n,v)}{\partial\lambda}$

Precomputing a discrete estimation of $\frac{\partial\mathcal{I}(n,v)}{\partial\lambda}$ for a give ρ^2.

1. Parameters: $\rho^2, N_p, szp, \lambda_0, \lambda_r$.

2. $f \leftarrow$ noise patch. $EI^0 \leftarrow 0$. [EI stands for *Estimated Inner product*]

3. Loop ($i \leftarrow 1; i^{++}; i \leq N_p$)

   (a) $\lambda^i \leftarrow \lambda^{i-1}\lambda_r$.

   (b) Compute u^i, v^i as in Main.

   (c) $EI^i \leftarrow < v^i, f >$.

   (d) $DEI^i_{pre} \leftarrow (EI^i - EI^{i-1})/(\lambda^i - \lambda^{i-1})$

4. Return vector DEI_{pre}

Remarks

- Parameters (in brackets are values used for processing natural images):

  1. szp - size of patch (80×80 pixels). It should be noted that in $\mathcal{H}$ space where ρ^2 of noise with variance σ^2 depends on the image size (as in the H^{-1} case), a more accurate precomputation would be on a patch which is the same size of the image. Our experiments

have shown that the difference in the estimation results are very little, and the above fixed size patch is good enough for medium size images.

2. N_p - number of precomputed points, that is different λ values or time-points for indirect method (30). The main loop should do at most N_p iterations.
3. λ^0 - initial λ (1), λ_r - ratio of successive λ (0.9).

It is important to note that this parameters mainly control the step resolution and no tuning is needed for different images. We used the same values, in brackets, for our experiments on natural images.

- In the specific implementation presented here, where the λ values of the Main phase are exactly as in the Precomputing phase, one can actually omit the multiplication and division by $(\lambda^i - \lambda^{i-1})$ in the computation of DEI and DEI_{pre} (we kept it to be consistent with our formulation).

References

[1] G. Aubert and J.F. Aujol. Modeling very oscillating signals. Application to image processing, 2003. INRIA Research Report, to appear in Applied Mathematics and Optimization, http://www.inria.fr/rrrt/rr-4878.html.

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

[3] J.F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *JMIV*, 22(1), January 2005.

[4] J.F. Aujol and A. Chambolle. Dual norms and image decomposition models. *IJCV*, June 2005. in press, http://www.math.ucla.edu/~aujol/publications.html.

[5] H. Brezis. *Analyse fonctionnelle. Théorie et applications.* Mathématiques appliquées pour la maitrise. Masson, 1983.

[6] A. Buades, B. Coll, and J.M. Morel. On image denoising methods, 2004. Preprint CMLA.

[7] J.L. Carter. *Dual methods for total variation-based image restoration*. PhD thesis, UCLA, 2001. Advisor: T.F. Chan.

[8] A. Chambolle. An algorithm for total variation minimization and applications. *JMIV*, 20:89–97, 2004.

[9] A. Chambolle and P.L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(3):167–188, 1997.

[10] T. Chan and S. Esedoglu. Aspects of total variation regularized l^1 function approximation, 2004. CAM report 04-07.

[11] 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*, 20(6):1964–1977, 1999.

[12] P.G. Ciarlet. *Introduction à l'analyse numérique matricelle et à l'optimisation*. Mathématiques appliquées pour la maitrise. Masson, 1982.

[13] P. L. Combettes and J. C. Pesquet. Image restoration subject to a total variation constraint. *IEEE Transactions on Image Processing*, 13(9):1213–1222, 2004.

[14] I. Daubechies and G. Teschke. Variational image restoration by means of wavelets: simultaneous decomposition, deblurring and denoising, 2004. submitted.

[15] I. Ekeland and R. Temam. *Analyse convexe et problmes variationnels*. Etudes Mathématiques. Dunod, 1974.

[16] L.C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, 1991.

[17] 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*, 63(6):1911–1934, 2003.

[18] 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/.

[19] G. Gilboa, N. Sochen, and Y.Y. Zeevi. Texture preserving variational denoising using an adaptive fidelity term. In *Proc. VLSM 2003, Nice, France*, pages 137–144, 2003.

[20] G. Gilboa, N. Sochen, and Y.Y. Zeevi. Estimation of optimal PDE-based denoising in the SNR sense, 2004. submitted.

[21] A. Haddad and Y. Meyer. Variational methods in image processing, 2004. CAM report 04-52.

[22] Yves Meyer. Oscillating patterns in image processing and in some nonlinear evolution equations, March 2001. The Fifteenth Dean Jacquelines B. Lewis Memorial Lectures.

[23] M. Nikolova. A variational approach to remove outliers and impulse noise. *JMIV*, 20(1-2):99–120, 2004.

[24] A. Obereder, S. Osher, and O. Sherzer. On the use of dual norms in bounded variation type regularization, 2004. CAM report 04-35.

[25] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation based image restoration, 2004. CAM report 04-13.

[26] S. Osher and O. Sherzer. G-norm properties of bounded variation regularization, 2004. CAM report 04-23.

[27] 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*, 1(3):349–370, 2003.

[28] M. Navara P. Mrázek. Selection of optimal stopping time for nonlinear diffusion filtering. *IJCV*, 52(2/3):189–203, 2003.

[29] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *PAMI*, 12(7):629–639, 1990.

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

[31] J.L. Starck, M. Elad, and D.L. Donoho. Image decomposition: separation of texture from piecewise smooth content, 2003. To appear in IEEE Transactions on Image Processing.

[32] L.A. Vese and S.J. Osher. Modeling textures with total variation minimization and oscillating patterns in image processing. *Journal of Scientific Computing*, 19:553–572, 2003.

[33] J. Weickert. Coherence-enhancing diffusion of colour images. *IVC*, 17:201–212, 1999.