Estimation of the Optimal Variational Parameter via SNR Analysis

Guy Gilboa[1] *, Nir A. Sochen[2] **, and Yehoshua Y. Zeevi[3] * * *

[1] Department of Mathematics, UCLA, Los Angeles, CA 90095, USA.
[2] Department of Applied of Mathematics, Tel-Aviv Univ., Tel-Aviv 69978, Israel.
[3] Department of Electrical Engineering, Technion, Haifa 32000, Israel.

Abstract. We examine the problem of finding the optimal weight of the fidelity term in variational denoising. Our aim is to maximize the signal to noise ratio (SNR) of the restored image. A theoretical analysis is carried out and several bounds are established on the performance of the optimal strategy and a widely used method, wherein the variance of the residual part equals the variance of the noise. A necessary condition is set to achieve maximal SNR. We provide a practical method for estimating this condition and show that the results are sufficiently accurate for a large class of images, including piecewise smooth and textured images.

1 Introduction

Variational methods have been increasingly applied for purposes of image denoising and restoration (for some examples see [3, 6, 8, 11, 12]). The basic concept is to view the restoration process as a task of energy minimization. Classically, the restored image is a minimization of a weighted sum of two fundamental energy terms:

$$E(u) = E_{smooth}(u) + \lambda E_{fidelity}(u, f), \tag{1}$$

where u is the restored image, and f is the input (noisy) image. 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. The underlying assumption is that the original clean image is smoother than the noisy image. By minimizing both terms we seek a compromise between a smooth solution (often in the TV sense, so edges are preserved) 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

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

** Supported by MUSCLE, a European Network of Excellence funded by the EC 6th Framework IST Programme, the Israeli Ministry of Science, the Israel Science Foundation, the Tel-Aviv University fund and the Adams Center.

* * * Supported by the Ollendorf Minerva Center, the Fund for the Promotion of Research at the Technion and the Israel Academy of Science.

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 for the problem at hand is therefore imperative. A similar problem has been investigated in regularization theory, in the context of operator inversion by Tikhonov-type methods (e.g. [4, 9]). As we are concerned with denoising of images (therefore our operator is the identity and the regularization preserves edges), different approaches should be used. In our field of PDE-base image processing, the problem was seriously addressed by only a few researchers: by [11] for total-variation denoising and by [7] and [13] for a closely related problem of finding the right stopping time in nonlinear scale-space. We refer in this paper only to the variational setting, but our method has shown to be very effective also for selecting the proper stopping time [5].

An analysis of the optimal parameter choice from SNR perspective is presented. We examine the widely used denoising strategy of [11] where the weight of the fidelity term is set such that the variance of the residual part equals that of the noise. Lower bound on the SNR performance of this strategy is established as well as a proof of non existence of an upper bound. Examples which illustrate worst- and best-case scenarios are presented and discussed.

Next, we derive a necessary condition for optimality in the SNR sense. From a theoretical viewpoint, this facilitates the computation of upper and lower bounds of the optimal strategy. From a practical viewpoint, the condition suggests the numerical method that should be followed for the purpose of maximizing the SNR of the filtered image. An algorithm for parameter calculation is suggested based on the above condition, resulting in fairly accurate estimates.

2 SNR Bounds for the Scalar Φ Process

2.1 Denoising Model, Definitions and Assumptions

We assume that the input signal f is composed of the original signal s and additive uncorrelated noise n of variance σ^2. Our aim is to find a decomposition u, v such that u approximates the original signal s and v is the residual part of f:

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

We accomplish that by finding the minimum to the following energy

$$\tilde{E}_\Phi(u) = \int_\Omega \left(\Phi(|\nabla u|) + \tilde{\lambda}(f-u)^2 \right) d\Omega. \tag{3}$$

Φ is assumed to be convex in this paper. Some of the following results, though, can also apply to the more general case of monotonically increasing Φ. The standard condition $\int_\Omega f d\Omega = \int_\Omega u d\Omega$ is set, (corresponding to the Neumann boundary condition of the evolutionary equations). Then $\int_\Omega v dx dy = 0$, rescaling $\tilde{\lambda}$ by the area of the domain $|\Omega|$: $\lambda = \tilde{\lambda}|\Omega|$, we get

$$E_\Phi(u, v) = \int_\Omega \Phi(|\nabla u|) d\Omega + \lambda V(v), \quad f = u + v. \tag{4}$$

where $V(q)$ is the variance of a signal q: $V(q) \doteq \frac{1}{|\Omega|}\int_\Omega (q-\bar{q})^2 d\Omega$, and $\bar{q}$ is the mean value: $\bar{q} \doteq \frac{1}{|\Omega|}\int_\Omega q d\Omega$. The covariance of two signals is defined as: $\text{cov}(q,r) \doteq \frac{1}{|\Omega|}\int_\Omega (q-\bar{q})(r-\bar{r})d\Omega$. We remind the identity $V(q+r) = V(q) + V(r) + 2\text{cov}(q,r)$.

Let us denote u^z as the solution of (4) for $f = z$. For example, u^s is the solution where $f = s$. The decorrelation assumption is taken also between s and n with respect to the Φ process:

$$\text{cov}(u^s, n) = 0, \quad \text{cov}(u^n, s) = 0, \quad \forall \lambda \geq 0. \tag{5}$$

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

$$\text{cov}(u^{s+n}, s) \leq \text{cov}(f, s), \quad \text{cov}(u^{s+n}, n) \leq \text{cov}(f, n), \quad \forall \lambda \geq 0. \tag{6}$$

Both of the above assumptions were verified numerically on a collection of natural images. We are investigating the possibility to characterize in an analytical manner the appropriate spaces of s and n such that (5) and (6) are followed. In this paper this question is left open and we resort to the following definition:

Definition 1 **((s,n) pair).** *An (s,n) pair consists of two uncorrelated signals s and n which obey conditions (5) and (6).*

Theorem 1. *For any (s,n) pair and an increasing Φ ($\Phi'(q) > 0, \forall q \geq 0$) the covariance matrix of $U = (f, s, n, u, v)^T$ has only non-negative elements.*

For proof see the appendix. Theorem 1 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.

We define the Signal-to-Noise Ratio (SNR) of the recovered signal u as

$$\text{SNR}(u) \doteq 10 \log \frac{V(s)}{V(u-s)} = 10 \log \frac{V(s)}{V(n-v)}, \tag{7}$$

where $\log \doteq \log_{10}$. The initial SNR of the input signal, denoted by SNR_0, where no processing is carried out ($u = f$, $v = 0$), is according to (7) and (2):

$$SNR_0 \doteq \text{SNR}(f) = 10 \log \frac{V(s)}{V(n)} = 10 \log \frac{V(s)}{\sigma^2}. \tag{8}$$

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

$$SNR_{opt} \doteq \max_\lambda \text{SNR}(u_\lambda) \tag{9}$$

where $u = u_\lambda$ attains the minimal energy of (4) with weight parameter λ (for a given f, v is implied). We denote by (u_{opt}, v_{opt}) the decomposition pair (u, v) that reaches SNR_{opt}, and define $V_{opt} \doteq V(v_{opt})$.

Equivalently, the desired variance could be set as $V(v) = P$, where P is some constant, and then (4) is reformulated to a constrained convex optimization problem

$$\min_u \int_\Omega \Phi(|\nabla u|) d\Omega \text{ subject to } V(v) = P. \tag{10}$$

In this formulation λ is viewed as a Lagrange multiplier. The value λ can be computed using the Euler-Lagrange equations and the pair (u, v):

$$\lambda = \frac{1}{P} \int_\Omega \operatorname{div} \left(\Phi' \frac{\nabla u}{|\nabla u|} \right) v d\Omega. \tag{11}$$

The problem then transforms to which value P should be imposed.

The strategy of [11] is to assume $v \approx n$ and therefore impose

$$V(v) = \sigma^2. \tag{12}$$

We define

$$SNR_{\sigma^2} \doteq \mathrm{SNR}(u)|_{V(v)=\sigma^2}. \tag{13}$$

We denote by $(u_{\sigma^2}, v_{\sigma^2})$ the (u, v) pair that obeys (12) and minimizes (4). We will now analyze this method for selecting u in terms of SNR.

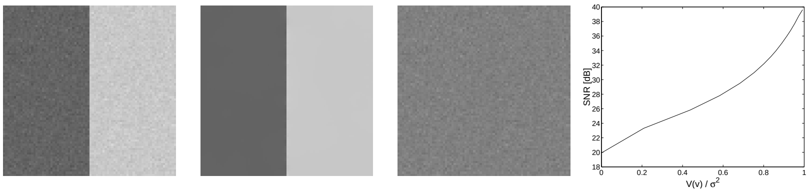

Fig. 1. Approaching best-case scenario in piece-wise constant images. In this example SNR increases by almost $20dB$ from 19.9dB to 39.6dB (variance of noise is $\approx \frac{1}{100}$ of the input noise). From left: f, u, v, SNR as a function of $V(v)/\sigma^2$.

Proposition 1 (SNR lower bound) *Imposing (12), for any (s, n) pair SNR_{σ^2} is bounded from below by*

$$SNR_{\sigma^2} \geq SNR_0 - 3dB, \tag{14}$$

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

Proof. From Theorem 1 we have $\mathrm{cov}(n, v) \geq 0$, therefore,

$$\begin{aligned} SNR_{\sigma^2} &= 10 \log \tfrac{V(s)}{V(n-v)} \\ &\geq 10 \log \tfrac{V(s)}{V(n)+V(v)} \\ &= 10 \log \tfrac{V(s)}{2\sigma^2} \\ &= SNR_0 - 3dB. \quad \square \end{aligned}$$

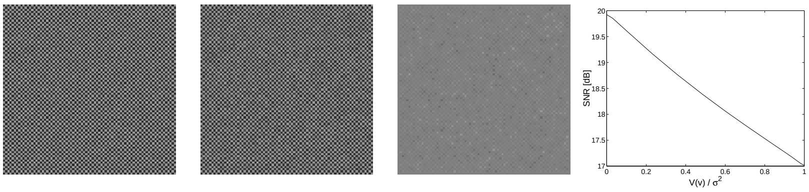

Fig. 2. Approaching worst-case scenario in a checkered-board image. SNR decreases by almost $3dB$ from 19.9dB to 17.0dB. From left: f, u, v, SNR as a function of $V(v)/\sigma^2$.

The lower bound of proposition 1 is reached only in the very rare and extreme case where $\mathrm{cov}(n, v) = 0$. This implies that only parts of the signal were filtered out and no denoising was performed.

Proposition 2 (SNR upper bound) *Imposing (12), then there does not exist an upper bound* $0 < M < \infty$*, where* $SNR_{\sigma^2} \leq SNR_0 + M$*, 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 $V(s) = h\sigma^2$, $0 < h < 1$. Then $SNR_0 = 10 \log h$. As signal and noise are not correlated we have $V(f) = V(s)+V(n) = (1+h)\sigma^2$. We can write $V(f)$ also as $V(u+v) = V(u)+V(v)+2\mathrm{cov}(u, v)$. From (12), $V(v) = \sigma^2$, and from Theorem 1, $\mathrm{cov}(u, v) \geq 0$, therefore $V(u) \leq h\sigma^2$. Since $\mathrm{cov}(u, s) \geq 0$ (Theorem 1) we get $V(u - s) \leq 2h\sigma^2$. This yields $SNR_{\sigma^2} \geq 10 \log \frac{1}{2}$ and

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

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

Simulations that illustrate worst- and best-case scenarios are presented in Figs. 1 and 2. A signal that consists of a single very contrasted step function is shown in Fig. 1. This example illustrates a best-case scenario for an edge preserving Φ. SNR resulting from the PDE-based denoising is greatly increased (by $\sim 20dB$). Note that this case approximates an ideal decomposition $u \approx s$, $v \approx n$ which differs from the simple case used in the proof of Proposition 2. A worst-case scenario is illustrated in Fig. 2 by means of the Checkered-board example. A very oscillatory signal s is being denoised and, in the process, is heavily degraded. The reduction in SNR, compared to SNR_0, is $\sim 2.9dB$, close to the theoretical $3dB$ bound.

2.2 Condition for optimal SNR

We will now develop a necessary condition for the optimal SNR. As discussed, we have a single degree of freedom of choosing $V(v)$. We therefore regard SNR as a function $\mathrm{SNR}(V(v))$ and assume that it is smooth. A necessary condition for the maximum in the range $V(v) \in (0, V(f))$ is:

$$\frac{\partial\,\mathrm{SNR}}{\partial V(v)} = 0. \tag{15}$$

Rewriting $V(n-v)$ as $V(n) + V(v) - 2\mathrm{cov}(n,v)$, and using (15) and (7), yields

$$\frac{\partial \mathrm{cov}(n,v)}{\partial V(v)} = \frac{1}{2}. \tag{16}$$

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 $V(v)$. We begin with $V^0(v) = 0$ and increment the variance of v by a small amount $dV(v)$, so that in the next step $V^1(v) = dV(v)$. 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 $\frac{\partial \mathrm{cov}(n,v)}{\partial V(v)} > \frac{1}{2}$, then one should keep on with the process and SNR will increase. When we reach the condition of (16), noise and signal are equally filtered and one should therefore stop. If filtering is continued, more signal than noise is filtered (in terms of variance) 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 $\frac{\partial \mathrm{cov}(n,v)}{\partial V(v)}|_{V(v)=0} < \frac{1}{2}$ and $SNR_{opt} = SNR_0$. The other extreme case is when SNR increases monotonically and is maximized when $V(v) = V(f)$ (the trivial constant solution $u = \bar{f}$). We will see later (Proposition 3) that this can only happen when SNR_0 is negative or, equivalently, when $V(s) < \sigma^2$.

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

1. Set $\mathrm{cov}^0(n,v) = 0$, $V^0(v) = 0$, $i = 1$.
2. $V^i(v) \leftarrow V^{i-1}(v) + dV(v)$. Compute $\mathrm{cov}^i(n,v)$.
3. If $\frac{\mathrm{cov}^i(n,v) - \mathrm{cov}^{i-1}(n,v)}{dV(v)} \le \frac{1}{2}$ then stop.
4. $i \leftarrow i + 1$. Goto step 2.

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

Definition 2 (Regular SNR). *We define the function* $\mathrm{SNR}(V(v))$ *as* regular *if (16) is a sufficient condition for optimality or if the optimum is at the boundaries.*

Proposition 3 (Range of optimal SNR) *If SNR is regular, then for any* (s,n) *pair* $0 \le V_{opt} \le 2\sigma^2$.

Proof. Let us first show the relation $\mathrm{cov}(n,v) \le \sigma^2$: $\mathrm{cov}(n,f) = \mathrm{cov}(n, n+s) = V(n) + \mathrm{cov}(n,s) = \sigma^2$. On the other hand $\mathrm{cov}(n,f) = \mathrm{cov}(n, u+v) = \mathrm{cov}(n,u) + \mathrm{cov}(n,v)$. The relation is validated by using $\mathrm{cov}(n,u) \ge 0$ (Theorem 1).

We reach the upper bound by the following inequalities:

$$\sigma^2 \geq \mathrm{cov}(n,v)|_{V_{opt}} = \int_0^{V_{opt}} \frac{\partial \mathrm{cov}(n,v)}{\partial V(v)} dV(v) \geq \int_0^{V_{opt}} \frac{1}{2} dV(v) = \frac{1}{2} V_{opt}.$$

The inequality on the right is based on that $\frac{\partial \mathrm{cov}(n,v)}{\partial V(v)} \geq \frac{1}{2}$ for $V(v) \in (0, V_{opt})$.

The lower bound $V_{opt} = 0$ is reached whenever $\frac{\partial \mathrm{cov}(n,v)}{\partial V(v)}|_{V(v)=0} < \frac{1}{2}$. □

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

$$0 \leq SNR_{opt} - SNR_0 \leq \begin{cases} -10\log(1 + V_{opt}/\sigma^2 - 2\sqrt{V_{opt}/\sigma^2}), & 0 \leq V_{opt} < \sigma^2 \\ -10\log(V_{opt}/\sigma^2 - 1), & \sigma^2 < V_{opt} \leq 2\sigma^2 \end{cases} \quad (17)$$

Proof. By the SNR definition, (7), and expanding the variance expression, we have

$$SNR_{opt} - SNR_0 = 10\log(\frac{\sigma^2}{\sigma^2 + V_{opt} - 2\mathrm{cov}(n, v_{opt})}). \quad (18)$$

For the lower bound we use the relation shown in Proposition 3: $\mathrm{cov}(n, v_{opt}) \geq \frac{1}{2} V_{opt}$. For the upper bound we use two upper bounds on $\mathrm{cov}(n, v_{opt})$ and take their minimum. The first one, $\mathrm{cov}(n, v_{opt}) \leq \sigma\sqrt{V_{opt}}$, is a general upper bound on covariance. The second relation, $\mathrm{cov}(n, v_{opt}) \leq \sigma^2$, is outlined in Proposition 3. □

A plot of the upper bound of the optimal SNR with respect to V_{opt}/σ^2 is depicted in Fig. 3, left.

In practice, the flow is not performed by directly increasing $V(v)$, but by decreasing the value of λ. Therefore, it is instructive to check how $V(v)$ varies, as well as the other energies, as λ varies. In the next proposition we show that as λ decreases the total energy strictly decreases, $E_v(v) \doteq V(v)$ increases and $E_u(u) \doteq \int_\Omega \Phi(|\nabla u|) d\Omega$ decreases.

Proposition 4 (Energy change as a function of λ) *The energy parts of Eq. (4) vary as a function of* λ *as follows:*

$$\frac{\partial E_\Phi}{\partial \lambda} > 0, \quad \frac{\partial E_v}{\partial \lambda} \leq 0, \quad \frac{\partial E_u}{\partial \lambda} \geq 0. \quad (19)$$

For proof see [5].

3 Estimating $\mathrm{cov}(n, v)$

The term $\mathrm{cov}(n, v)$ is unknown, as we do not know the noise, and therefore should be estimated. We are showing here for the first time a representation of denoising by a family of curves which connects the variance of the noise, λ and $\mathrm{cov}(n, v)$

of pure noise. This can be regarded as some sort of nonlinear statistics of noise with respect to a specific Φ process. It appears that $\mathrm{cov}(n, v)$ 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 $\mathrm{cov}(n, v)$ with respect to λ. This is done a single time for each noise variance and can be regarded as a look-up-table (see Fig. 3, right). For each processed image the behavior of λ with respect to $V(v)$ is measured. Combining the information, it is possible to approximate how $\mathrm{cov}(n, v)$ behaves with respect to $V(v)$. In other words, this is simply the chain-rule for differentiation:

$$\begin{aligned} \frac{\partial \mathrm{cov}(n,v)}{\partial V(v)} &= \frac{\partial \mathrm{cov}(n,v)}{\partial \lambda} \frac{\partial \lambda}{\partial V(v)} \\ &\approx \frac{\partial \mathrm{cov}(n,v)}{\partial \lambda}\Big|_{f=patch} \frac{\partial \lambda}{\partial V(v)}\Big|_{f=s+n}. \end{aligned} \tag{20}$$

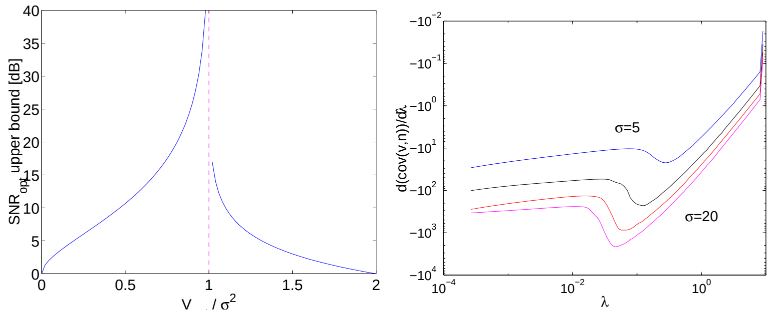

Fig. 3. Left: Visualization of Theorem 2: Upper bound of $SNR_{opt} - SNR_0$ as a function of V_{opt}/σ^2. For $V_{opt} \to \sigma^2$ the bound approaches ∞. Right: Precomputed function $\partial \mathrm{cov}(n, v)/\partial\lambda$ plotted as a function of λ (log scale). Graphs depict plots for values of σ: $5, 10, 15, 20$, from upper curve to lower curve, respectively.

3.1 Experimental results

We compare our method for finding λ with the standard method of imposing (12) and with the optimal λ, which maximizes the SNR. Six classical benchmark images are processed: Cameraman, Lena, Boats, Barbara, Toys and Sailboat. The summary of the results is shown in Table 1. Our method is quite close to the optimal denoising (less than $0.1dB$ difference on average) and performs better than the method of [11].

We used $\Phi(s) = \sqrt{1 + s^2}$, which can be viewed as the Vogel-Oman [12] regularization of TV [11] with $\epsilon = 1$ or the Charbonnier [2] process. The image grey-level range is $1 : 256$ so edges are well preserved. Other details about this experiment can be found in [5].

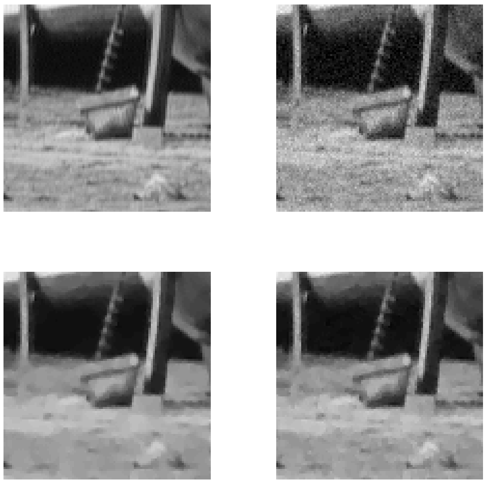

Fig. 4. Part of Boats image. Top (left to right): s, f. Bottom (left to right): u by standard method ($V(v) = \sigma^2$), u by our estimation method. More textural information is preserved by our method.

In Fig. 4 we show example results of processing the Boats image. The main visual difference from the standard method is that textural information is better preserved, as we approach the optimal λ. In Fig. 5 the terms SNR(u) and $\frac{\partial \mathrm{cov}(n,v)}{\partial \lambda}$ are plotted as functions of the normalized variance $V(v)/\sigma^2$. It is apparent that the SNR is smooth and behaves regularly, in accordance with our assumptions. An interesting phenomenon is that the covariance derivative estimation tends to be more accurate near the critical value of $\frac{1}{2}$. Naturally, this is advantageous to our algorithm. We currently have no explanation for this behavior.

4 Conclusion

Most image denoising processes are quite sensitive to the choice and fine tuning of various parameters. This is a major obstacle for fully automatic algorithms. This problem motivated us to develop a criterion for the optimal choice of the fidelity weight parameter in variational denoising. Our criterion is to maximize the SNR of the resultant image. Bounds on the SNR as well as on the optimal variance are obtained. We demonstrate our method on a series of benchmark images and show that the performance is only slightly worse than optimal (less than 0.1dB difference).

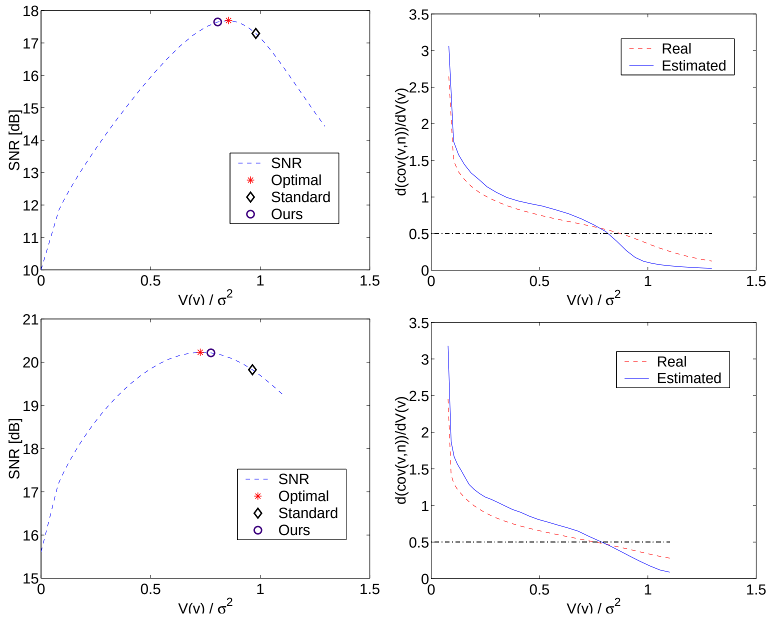

Fig. 5. SNR as a function of $V(v)/\sigma^2$ (left). $d\text{cov}(n,v)/dV(v)$ as a function of $V(v)/\sigma^2$ (right), as computed by our estimation method (solid) and the ground truth (dashed). Graphs depict processing of Toys (top) and Boats (bottom).

We should comment that the SNR criterion is not always in accordance with human-based quality evaluations. Other, more sophisticated criteria, may also be applied for parameter selection using the spirit of the method presented here.

The basic ingredients of the proposed method, namely the covariance condition (16) and its estimation (20), are quite universal and do not depend on the specific denoising algorithm. The method was generalized for selecting the stopping time in nonlinear diffusion [5] and for regularizations based on BV and Hilbert-space norms [1].

A Proof of Theorem 1

We present the main steps of the proof. A full version is given in [5]. Since $\text{cov}(q,r) = \text{cov}(r,q)$, the matrix is symmetric. The diagonal is the variance of each element, which is non negative. Therefore we have to consider all 10 possible signal pairs and show that their covariance is non-negative.

$\mathbf{cov}(\boldsymbol{s},\boldsymbol{n})$, $\mathbf{cov}(\boldsymbol{f},\boldsymbol{s})$, $\mathbf{cov}(\boldsymbol{f},\boldsymbol{n})$. Since s and n are not correlated, we have $\text{cov}(s,n) = 0$, $\text{cov}(f,s) = \text{cov}(s+n,s) = V(s) \geq 0$, $\text{cov}(f,n) = \text{cov}(s+n,n) = V(n) \geq 0$.

Image	SNR_0	SNR_{opt}	SNR_{σ^2}	SNR_{ours}
Cameraman	15.86	19.56	19.32	19.50
Lena	13.47	18.19	17.65	18.18
Boats	15.61	20.23	19.83	20.22
Barbara	14.73	16.86	16.21	16.64
Toys	10.00	17.69	17.29	17.65
Sailboat	10.36	15.51	15.16	15.48
Average difference from SNR_{opt}	4.67	0.00	0.43	0.06

Table 1. Denoising results of several images widely used in image processing. The original images were degraded by additive white Gaussian noise ($\sigma = 10$) prior to their processing.

$\mathbf{cov}(\boldsymbol{u}, \boldsymbol{v})$, $\mathbf{cov}(\boldsymbol{f}, \boldsymbol{u})$, $\mathbf{cov}(\boldsymbol{f}, \boldsymbol{v})$. Once we prove $\text{cov}(u, v) \geq 0$, then we readily have $\text{cov}(f, u) = \text{cov}(u + v, u) = V(u) + \text{cov}(u, v) \geq 0$ and $\text{cov}(f, v) = \text{cov}(u + v, v) = V(v) + \text{cov}(u, v) \geq 0$.

We follow the spirit of the proof of Meyer [8]. As the (u, v) decomposition minimizes the energy of Eq. (4), we can write for any function $h \in BV$ and scalar $\epsilon > 0$ the following inequality:

$$\int_\Omega \Phi(|\nabla(u - \epsilon h)|)d\Omega + \lambda V(v + \epsilon h) \geq \int_\Omega \Phi(|\nabla u|)d\Omega + \lambda V(v). \qquad (21)$$

Replacing $V(v + \epsilon h)$ by $V(v) + \epsilon^2 V(h) + 2\epsilon \text{cov}(v, h)$ and then changing h to u and dividing both sides by ϵ we get

$$2\lambda \text{cov}(v, u) \geq \frac{1}{\epsilon} \int_\Omega \left(\Phi(|\nabla u|) - \Phi(|\nabla(u - \epsilon u)|)\right) d\Omega - \lambda \epsilon V(u).$$

In the limit as $\epsilon \to 0$, the right term on the right-hand-side vanishes. Since Φ is increasing, the term in the integral is non-negative.

$\mathbf{cov}(\boldsymbol{s}, \boldsymbol{u})$, $\mathbf{cov}(\boldsymbol{n}, \boldsymbol{u})$. By writing $V(v)$ as $V(s+n-u)$, expanding the variance expression and omitting expressions that do not involve u, we can reach the following minimization problem equivalent to minimizing (4): $u = \text{argmin}_u\{\hat{E}_\Phi(u)\}$ where

$$\hat{E}_\Phi(u) = \int_\Omega \Phi(|\nabla u|)d\Omega + \lambda(V(u) - 2\text{cov}(s, u) - 2\text{cov}(n, u)). \qquad (22)$$

Since $\text{cov}(s, u) + \text{cov}(n, u) = \text{cov}(f, u) \geq 0$ at least one of the terms $\text{cov}(s, u)$ or $\text{cov}(n, u)$ 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 $\text{cov}(s, u^{s+n}) \geq 0$ and $\text{cov}(n, u^{s+n}) < 0$. We denote the optimal (minimal) energy of (22) with $f = s+n$ as $\hat{E}^*_\Phi|_{f=s+n}$. The energy can be written as

$$\begin{aligned}\hat{E}^*_\Phi|_{f=s+n} &= \hat{E}_\Phi|_{f=s+n}(u^{s+n}) \\ &= \int_\Omega \Phi(|\nabla u^{s+n}|)d\Omega + \lambda(V(u^{s+n}) - 2\mathrm{cov}(s,u^{s+n}) - 2\mathrm{cov}(n,u^{s+n})).\end{aligned} \tag{23}$$

On the other hand, according to condition (5), $\mathrm{cov}(u^s,n)=0$ and we have

$$\begin{aligned}&\hat{E}_\Phi|_{f=s+n}(u^s) = \int_\Omega \Phi(|\nabla u^s|)d\Omega + \lambda(V(u^s) - 2\mathrm{cov}(s,u^s)) \\ &= \hat{E}^*_\Phi|_{f=s} \le \hat{E}_\Phi|_{f=s}(u^{s+n}) = \int_\Omega \Phi(|\nabla u^{s+n}|)d\Omega + \lambda(V(u^{s+n}) - 2\mathrm{cov}(s,u^{s+n})).\end{aligned}$$

In the above final expression, adding the term $-\lambda 2\mathrm{cov}(n,u^{s+n})$ we obtain the right hand side of expression (23). Since we assume $\mathrm{cov}(n,u^{s+n})<0$, we get the following contradiction: $\hat{E}_\Phi|_{f=s+n}(u^s) < \hat{E}^*_\Phi|_{f=s+n}$. Similarly, the opposite case $\mathrm{cov}(n,u^{s+n}) \ge 0$ and $\mathrm{cov}(s,u^{s+n})<0$ is not possible.

$\mathbf{cov}(s,v)$, $\mathbf{cov}(n,v)$. This follows directly from condition (6) as $\mathrm{cov}(f,s) = \mathrm{cov}(u,s)+\mathrm{cov}(v,s)$ and $\mathrm{cov}(f,n)=\mathrm{cov}(u,n)+\mathrm{cov}(v,n)$. $\square$

References

1. J.F. Aujol, G. Gilboa, "Implementation and parameter selection for BV-Hilbert space regularizations", UCLA CAM Report 04-66, November 2004.
2. P. Charbonnier, L. Blanc-Feraud, G. Aubert, M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging", Proc. IEEE ICIP '94, Vol.2, 168-172, 1994.
3. R. Deriche, O. Faugeras, "Les EDP en traitement des images et vision par ordinateur", Traitement du Signal, 13(6), 1996.
4. H.W. Engl, H. Gfrerer, "A posteriori parameter choice for general regularization methods for solving linear ill-posed problems", Appl. Numer. Math, 4(5), 395 - 417, 1988.
5. G. Gilboa, N. Sochen, Y.Y. Zeevi, "Estimation of optimal PDE-based denoising in the SNR sense", CCIT report No. 499, Technion, August 2004.
6. G. Gilboa, N. Sochen, Y.Y. Zeevi, "Texture preserving variational denoising using an adaptive fidelity term", Proc. VLSM 2003, Nice, France, pp. 137-144, 2003.
7. P. Mrázek, M. Navara, "Selection of optimal stopping time for nonlinear diffusion filtering", IJCV, v. 52, no. 2/3, pp. 189-203, 2003.
8. Y. Meyer, *Oscillatory Patterns in Image Processing and Nonlinear Evolution Equations*, Vol. 22 University Lecture Series, AMS, 2001.
9. S. Pereverzev, E. Schock, "Morozov's discrepancy principle for Tikhonov regularization of severely ill-posed problems in finite-dimensional subspaces", Numer. Funct. Anal. Optim, 21, pp. 901-916, 2000.
10. P. Perona, J. Malik, "Scale-space and edge detection using anisotropic diffusion", PAMI 12(7), pp. 629-639, 1990.
11. L. Rudin, S. Osher, E. Fatemi, "Nonlinear Total Variation based noise removal algorithms", Physica D 60 259-268,'92.
12. R.V. Vogel and M.E. Oman, "Iterative methods for total variation denoising", SIAM J. Scientific Computing, 17(1):227–238, 1996.
13. J. Weickert, "Coherence-enhancing diffusion of colour images", IVC, Vol. 17, 201-212, 1999.