CCIT Report #499
August 2004

Estimation of Optimal PDE-based Denoising in the SNR Sense

Guy Gilboa, Nir Sochen and Yehoshua Y. Zeevi

G. Gilboa and Y.Y. Zeevi are with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Technion city, Haifa 32000, Israel. e-mail: gilboa@tx.technion.ac.il, zeevi@ee.technion.ac.il

N. Sochen is with the Department of Applied Mathematics, University of Tel-Aviv, Tel-Aviv 69978, Israel e-mail: sochen@math.tau.ac.il

Abstract

This paper is concerned with the problem of finding the best of a set of possible solutions of a PDE-based denoising process. We focus on either finding the proper weight of the fidelity term in the energy minimization formulation or the optimal stopping time of a nonlinear diffusion process. 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. An optimality condition is set to achieve maximal SNR, under quite general assumptions. We provide two practical alternatives of estimating this condition and show that the results are sufficiently accurate for a large class of images, including piecewise smooth and textured images.

I. INTRODUCTION

The use of Partial Differential Equations (PDEs) to regularize images is becoming a very active field of research. The elegance of the formulation, frequently via the calculus of variations, and the good results, attract researchers and users alike. Invariably, these methods require the determination of a parameter in the process. This parameter is the time, or number of iterations, in diffusion like processes, or the weight of the fidelity term of the energy functional, in the calculus of variations approach. In both cases, a simplification of the image is achieved via a parameter dependent PDE. It is desirable that the "true" signal will not be degraded in the process of this simplification while noise is removed. In fact, both noise AND signal are being altered in the process. That the signal is changed is clear since an image without noise is also altered in the process. The PDEs are constructed to reduce noise faster than the alteration of piecewise smooth images. Yet, the process must be stopped when too much of the signal is altered, either because there is very few noise left, or because the image contains texture.

It is thus important to decide on the optimal point to stop the process. This question is pertinent in image processing but to our surprise was addressed by only few researchers in the nonlinear diffusion context [5], [10], [3].

We present in this paper an analysis of the optimal parameter choice from a Signal to Noise (SNR) perspective. We examine the very popular denoising strategy (suggested in [8]) 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. Two algorithms for "on the fly" parameter calculation are suggested based on the above condition, which give fairly accurate estimates.

We demonstrate our method and show its superiority with respect to the methods of [8], [5] and [10].

The paper is organized as follows: In Section II we present the model and discuss the lower and upper SNR bounds that can be obtained in a Φ- process regularization [2]. An optimality condition is derived in terms of the variances and covariances of the signal, noise and their estimations. The covariance of the noise and the residual part is a key ingredient and is the focus of Section III. Two practical methods are provided for its approximation. In Section IV we explain that similar arguments, used for the Φ-process regularization, can be applied to diffusion-like processes. We define the concepts and quantities that link between the two types of regularization techniques and which are relevant to our analysis. A detailed comparison to other stopping criteria is presented. The comparison is carried out from a theoretical and an empirical points of view. A table that compares the results on benchmark images is presented. We present our conclusions in Section V.

II. SNR Bounds for the Scalar Φ Process

A. Denoising Model, Definitions and Assumptions

We assume that the input signal f is composed of the original signal s and additive white Gaussian noise n of variance σ^2:

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

It is assumed that s and n are uncorrelated. 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:

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

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:

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

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 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

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

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 (1):

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

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

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

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):

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

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

A popular denoising strategy ([8]) is to assume $v \approx n$ and therefore impose

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

We define

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

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.

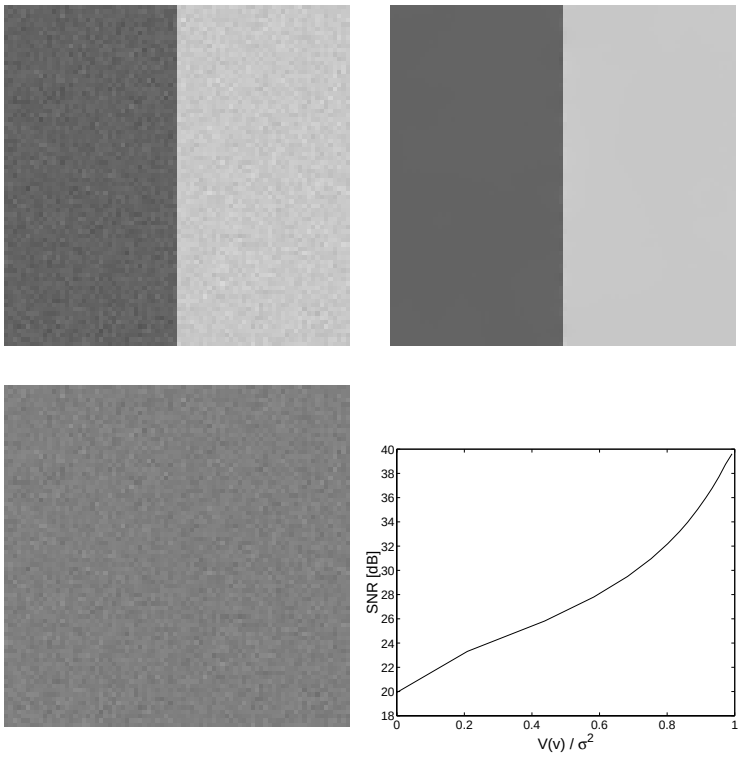

Fig. 1. Approaching best-case scenario in piece-wise constant images. In this example SNR increases by almost $20dB$ from 19.9dB to 39.6dB (variance of noise is $\approx \frac{1}{100}$ of the input noise). Top: f (left), u (right). Bottom: v (left), SNR as a function of $V(v)/\sigma^2$ (right).

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 $10\log_{10}(2)$.

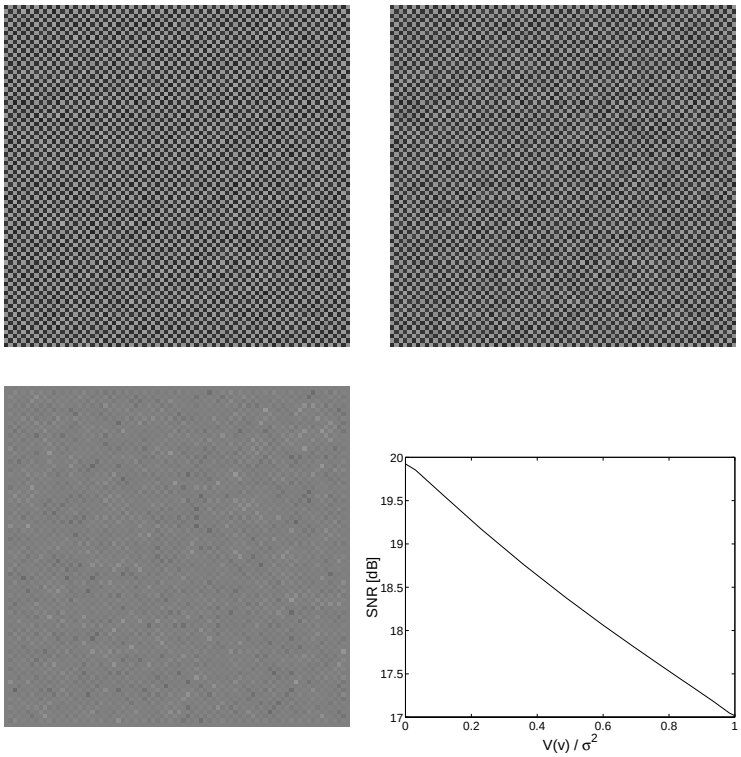

Fig. 2. Approaching worst-case scenario in a checkered-board image. SNR decreases by almost $3dB$ from 19.9dB to 17.0dB. Top: f (left), u (right). Bottom: v (left), SNR as a function of $V(v)/\sigma^2$ (right).

Proof: From Theorem 1 we have cov$(n, v) \geq 0$, therefore,

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

■

The lower bound of proposition 1 is reached only in the very rare and extreme case where 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 gain 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 $V(u+v) = V(u)+V(v)+2\text{cov}(u, v)$. From (12), $V(v) = \sigma^2$, and from Theorem 1, $\text{cov}(u, v) \geq 0$, therefore $V(u) \leq h\sigma^2$. Since $\text{cov}(u, s) \geq 0$ (Theorem 1) we get $V(u - s) \leq 2h\sigma^2$. This yields $SNR_{\sigma^2} \geq 10 \log \frac{1}{2}$ and

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

Thus, 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.

B. 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 $SNR(V(v))$ and assume that it is smooth. A necessary condition for the maximum in the range $V(v) \in (0, V(f))$ is:

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

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

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

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 $\frac{\partial \text{cov}(n,v)}{\partial V(v)} > \frac{1}{2}$, 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 $\frac{\partial \text{cov}(n,v)}{\partial V(v)}|_{V(v)=0} < \frac{1}{2}$ 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 $\text{cov}(n, v)$, our basic numerical algorithm should be as follows:

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

In the next section we suggest two ways to approximate the covariance term.

Definition 2—Regular SNR: We define the function $SNR(V(v))$ 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 $0 \leq V_{opt} \leq 2\sigma^2$.

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

We reach the upper bound by the following inequalities:

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

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

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

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

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

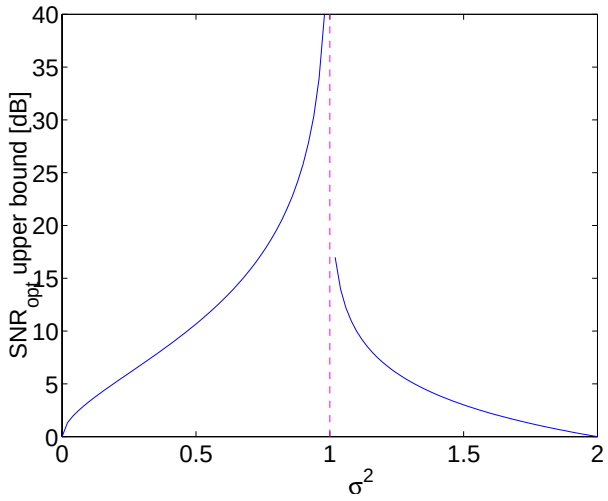

Fig. 3. Visualization of Theorem 2: Upper bound of $SNR_{opt} - SNR_0$ as a function of V_{opt}/σ^2. For $V_{opt} \to \sigma^2$ the bound approaches ∞.

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

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

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

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 the vary of $V(v)$, 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 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:

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

The proof is in the appendix.

III. ESTIMATING COV(n, v)

In order to estimate cov(n, v), we need an estimate of the noise. We may try to use only segments of the image where we have high confidence that we are able to distinguish between the noise and the image. These are typically the smooth regions. The problem is that generally we do not know in advance which regions of the image are smooth and which are not. There may be ways to overcome this problem by preprocessing the image (see [4] for an approach suitable for distinguishing between textured and smooth regions by a Φ process).

A. Direct Estimation

Here we adopt a different method, assuming that we have access to a source of a synthetic white Gaussian noise. Instead of finding regions in the image where we can estimate the noise, we simply extend the image with a "noise patch". This patch is simply an extension of the image in one direction, by a constant function with additive noise of variance σ^2 (as previously mentioned, we assume the noise variance is known *a-priory* or could be well estimated beforehand). [See Fig. 5.] Knowing, for this patch, both v and n, we can compute their covariance.

B. Indirect Estimation

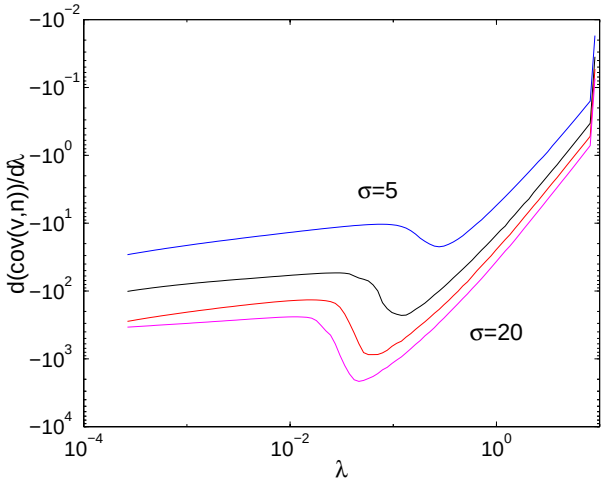

Fig. 4. Precomputed function for indirect estimation. ∂cov$(n, v)/\partial\lambda$ is plotted as a function of λ (log scale). Graphs depict plots for values of σ: $5, 10, 15, 20$, from upper curve to lower curve, respectively.

Another way of estimating cov(n, v) is by an indirect manner, which does not rely on physically attaching a synthetic patch to the image. Consequently, some minor inferences, which may

occur on the image-patch boundary, causing some side affects on the processed image near the patch and affecting the computations carried within the patch, are avoided.

The idea is to separate the computation to two phases. A patch of noise is processed and cov(n, v) is measured with respect to λ. Then the input image is processed and the behovior of λ with respect to $V(v)$ is measured. Combining the information, it is possible to approximate how cov(n, v) behaves with respect to $V(v)$. In other words, this is simply the chain-rule for differentiation:

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

The first term on the right-hand-side is a precomputed function, or in the discrete case of λ can be regarded as a look-up table (see Fig. 4). The second term is computed on the fly as the image is being processed.

In this scheme we rely on a very simplistic assumption that we can estimate the behavior of cov(n, v) of any image based on the very degenerate case where the image is simply pure noise. Quite extraordinarily, our numerical experiments show that our estimations are not so far from the ground truth (see Fig. 7, right side). A more comprehensive approach could accommodate the computation of the function $\frac{\partial \text{cov}(n,v)}{\partial \lambda}$ based on a representative collection of natural images.

IV. Evolutionary Flows

This idea can be similarly implemented in evolutionary flows that do not have a fidelity term. In this case one has to select the best stopping time T. Our definitions have to be changed somewhat, but essentially have the same meaning. The process is

$$u_t = \text{div}(c(|\nabla u|)\nabla u, \quad u|_{t=0} = f. \tag{21}$$

We define $v(x;t) = f(x) - u(x;t)$. In this formulation $dV(v)$ is defined as $dV(v(t)) = V(v(t)) - V(v(t-dt))$. Other similar changes in notations are straightforward. The detailed algorithm for

A. Comparison to Previous Stopping Mechanisms

A comprehensive study of the stopping time problem is reported in [5]. Here we relate to the most recent method proposed by Mrazek and Navara [5] and the more classical one discussed by Weickert in [10].

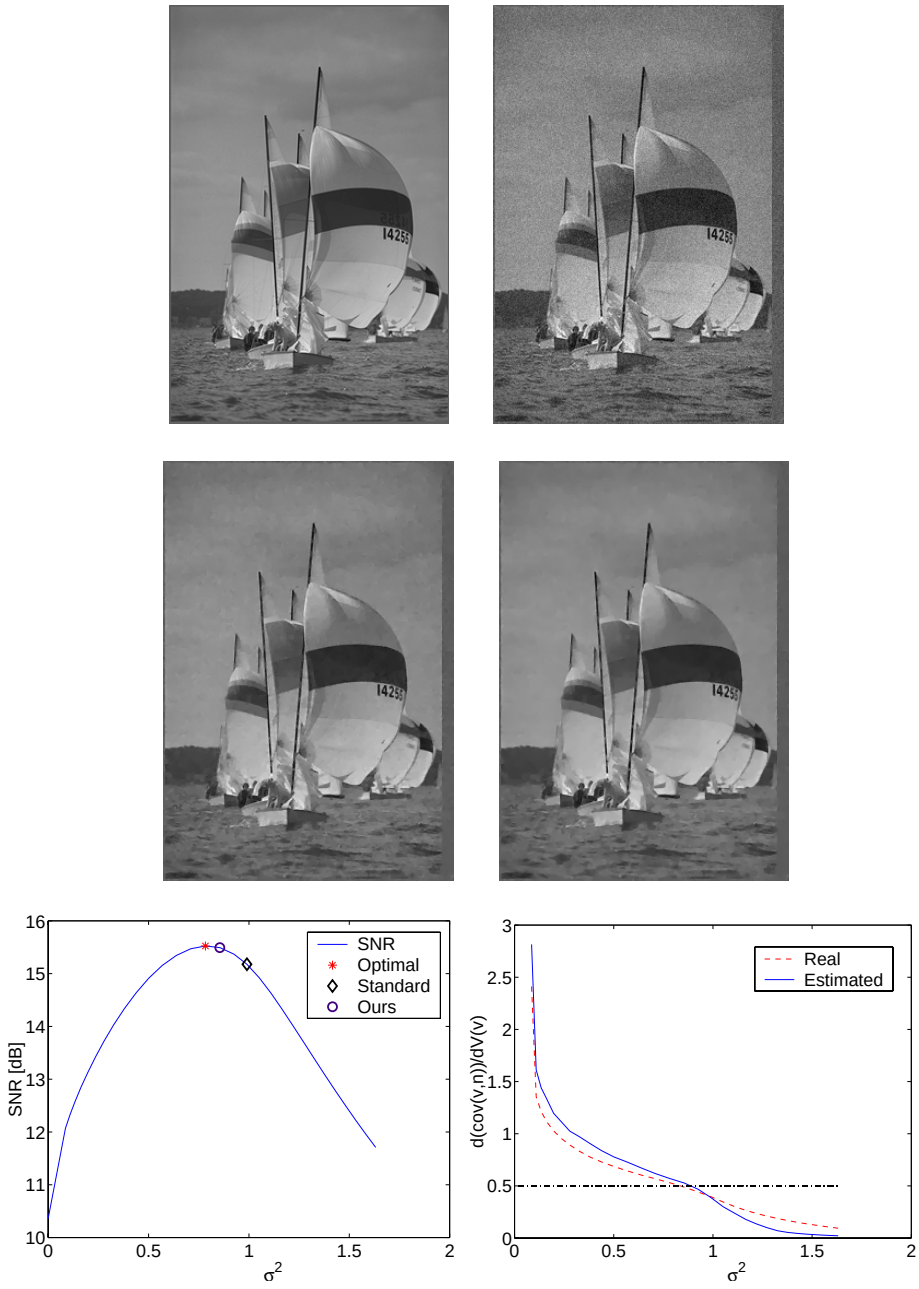

Fig. 5. Top row: s (left), f (right). Second row: u by our direct estimation (left), u by standard method ($V(v) = \sigma^2$, right). On the right of f and u the noise patch can be observed. Bottom row: SNR (left), and $\partial \text{cov}(n, v)/\partial V(v)$ (right) as a function of $V(v)/\sigma^2$.

Image	SNR_0	SNR_{opt}	SNR_{σ^2}	Ours SNR_{dir}	Ours SNR_{ind}
Cameraman	15.86	19.56	19.32	19.50	19.50
Lena	13.47	18.19	17.65	18.13	18.18
Boats	15.61	20.23	19.83	20.16	20.22
Barbara	14.73	16.86	16.21	16.73	16.64
Toys	10.00	17.69	17.29	17.66	17.65
Sailboat	10.36	15.51	15.16	15.48	15.48
Average difference from SNR_{opt}	4.67	0.00	0.43	0.06	0.06

TABLE I

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.

The former aims at finding the point in time of minimal correlation between u and v:

$$T = \text{argmin}_t \left\{\text{corr}(u(t), v(t))\right\}, \tag{22}$$

where

$$\text{corr}(u, v) \doteq \frac{\text{cov}(u, v)}{\sqrt{V(v)V(u)}}.$$

The underlying assumption of the method is that v carries most of the noise at the beginning of the denoising process. As $\text{corr}(s, n) = 0$ it is argued that a reasonable decomposition would be at a time where the correlation between u and v is minimal (in practice, the first local minimum is sought).

Weickert's method requires that

$$\frac{V(u(T))}{V(f)} = \frac{1}{1 + V(n)/V(s)} \tag{23}$$

or equivalently $V(u) = V(s)$, which can also be written as

$$V(v) = V(n) - 2\text{cov}(u, v). \tag{24}$$

All three methods of imposing (12), [5] and [10], work well on piecewise smooth images (without fine-scale features). In all three methods the decomposition is near $V(v) = V(n)$, which approaches the optimal decomposition in these cases. Using the method of [10] the process is often stopped too early.

The other approaches differ from each other and from our proposed method in the non-ideal cases of most natural images, where images contain textured regions.

The main advantage of the method proposed in [5] is that no knowledge of the noise variance is needed. It is also easy to compute, without any need for estimations. However it is not always practical to use this method for all classes of images. If the denoising process smoothes also some significant components of the signal such, that we cannot assume $v \approx n$, the stopping criterion of (22) may produce undesirable results. Actually, its performance in terms of SNR, cannot be bounded from below such as is determined by Proposition 1. One can construct examples where the stopping time should be near $t = 0$, whereas corr(u, v) is decreasing for a very long duration. This can be illustrated, for example, by the checkered-board image. The graphs of the SNR and correlation are depicted in Fig. 9. In a more realistic example of processing the Barbara image (Fig. 10), the results are not as extreme, but image is considerably over-smoothed.

The method of [10] is similar in its spirit to imposing (12). Here, though, the term 2cov(u, v) is being deducted, resulting in an early stopping of the process (esp. when u and v are highly correlated as in the case of textured images). In any case, the stopping is in the 'safe' regime $V(v) \leq \sigma^2$ (and thus its performance has a lower bound).

The differences between our method and those of [10] and [5] are illustrated in In Figs. 10, 11 and 12. The Barbara image, contaminated by additive white Gaussian noise ($\sigma = 10$) is processed by the nonlinear diffusion equation (21), with $c(s) = 1/\sqrt{1 + s^2}$. The image contains smooth regions and highly textured ones. This breaks the underlying assumption, used by both [10] and [5], which regards v as mostly containing noise. In partly textured images, v contains both noise and texture. In the case of [10], the term cov(u, v) is large, and the process stops too early. In the case of [5], the consequences are more severe and corr(u, v) is minimal only when the texture is smoothed out (see Fig. 11 for a plot of the correlation function). In terms of SNR, applying the method of [5] to this image results in a drop of more than $3dB$ below SNR_0. The SNR results are: $SNR_0 = 14.73$, $SNR_{opt} = 16.65$, method of [10] $SNR_W = 16.19$, method

of [5] $SNR_{MN} = 11.51$. In the case of our direct estimation method (using a 'noise patch') $SNR_{GSZ} = 16.59$.

V. 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 scale of interest, a significant parameter in PDE-based denoising, represented by the weight of the fidelity term λ in the variational formulation, or by the stopping time T in evlutionary processes. Our criterion is to maximize the SNR obtained as a result of the application of PDE-based denoising. Bounds on the SNR as well as on the optimal variance are obtained. We compare the performance of our algorithm with those obtained by means of previously proposed algorithms and demonstrate that ours achieves better results on a series of benchmark images.

Our suggested method finds with sufficient accuracy the first local maximum of the SNR with respect to the variance of the residual part v. In principle, there can be further local maxima at larger scales with greater SNR. In practice, however, we have not managed to produce an example, of either a synthetic or a natural image, where SNR has more than a single maximum. Current experience leads us to believe that these cases are quite rare in convex PDE-based processing.

We should also 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 methods presented here.

We examine the classical case of additive white Gaussian noise. Filtering other types of noise which are additive and uncorrelated with the signal, could be performed in a similar manner. Generalizations to other regularization processes, and to spatially variant parameters [4], are under current investigation.

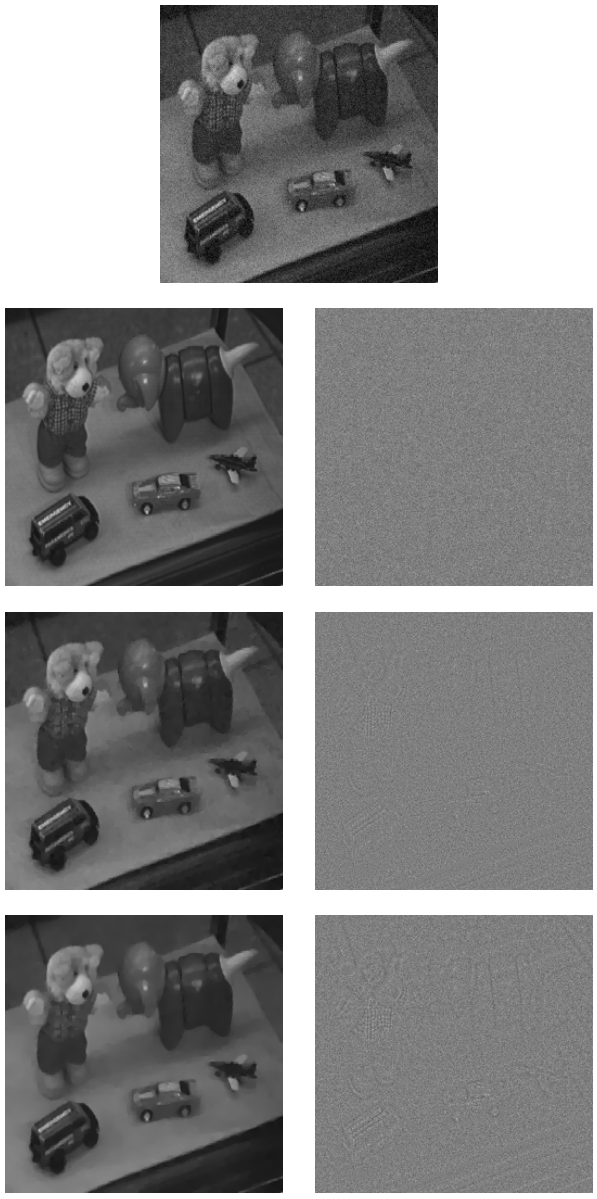

Fig. 6. Top row: f. Second row: s (left), n. Third row: u (left), v by our indirect estimation. Bottom row: u (left), v by standard method ($V(v) = \sigma^2$).

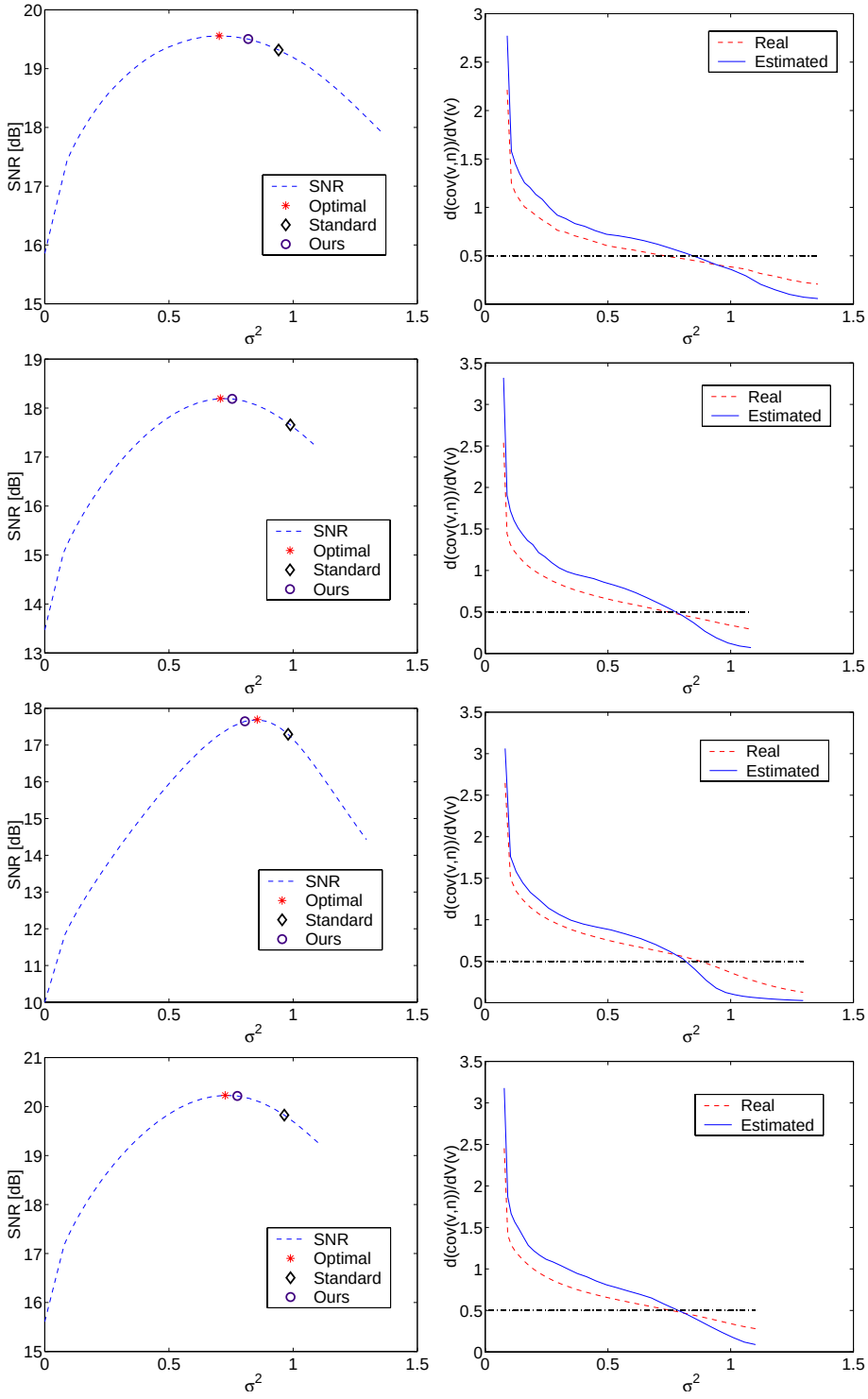

Fig. 7. SNR as a function of $V(v)/\sigma^2$ (left). $d\mathrm{cov}(n, v)/dV(v)$ as a function of $V(v)/\sigma^2$ (right), as computed by indirect estimation (solid) and the ground truth (dashed). Graphs depict processing of the following natural images (from top): Cameraman, Lena, Toys, Boats.

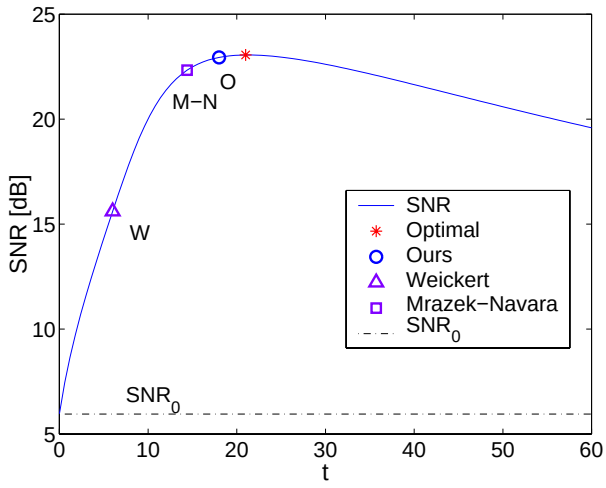

Fig. 8. Processing a step image (as in Fig. 1). Left: SNR plot as a function of t. Right: $d\text{cov}(n,v)/dV(v)$ as a function of t. Stopping time is sufficiently close to the optimal selection by both methods of Mrazek-Navara and ours.

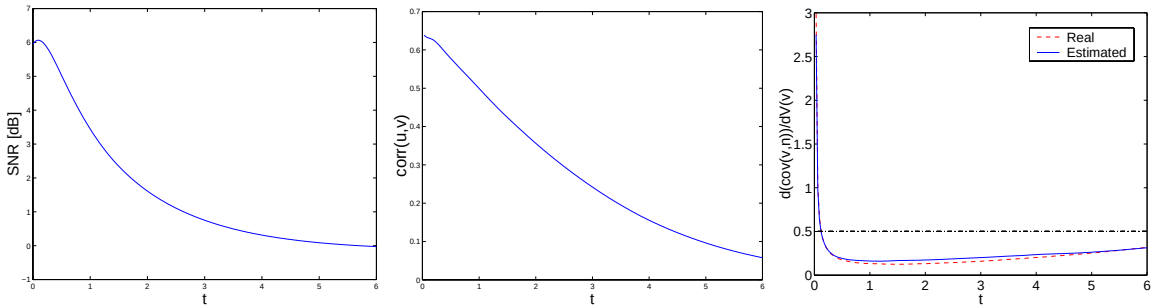

Fig. 9. A checkered-board image (medium contrast) with noise: Left - SNR as a function of t, middle - corr(u,v) as a function of t, right - $d\text{cov}(n,v)/dV(v)$ as a function of t. Whereas the criterion of Eq. (22) cannot be used in this example (no local minimum near 0), our estimation of the general criterion stated in Eq. (16) works well also on highly textured signals (stopping time is $T = 0.12$ versus the optimal $T_{opt} = 0.09$).

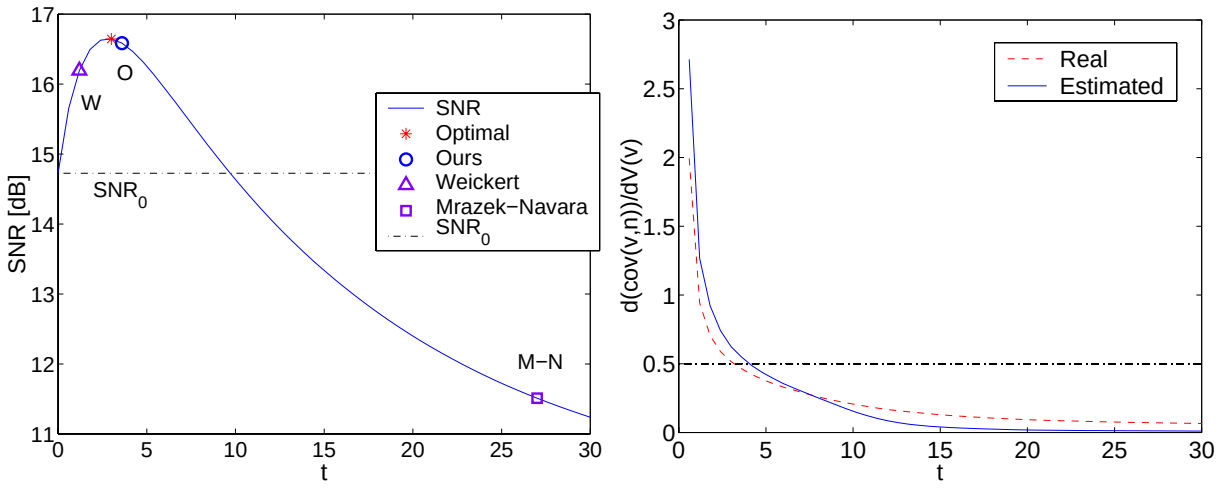

Fig. 10. Processing Barbara image. Left: SNR plot as a function of t. Right: $d\text{cov}(n, v)/dV(v)$ as a function of t.

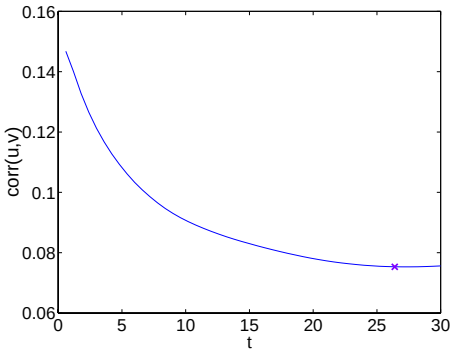

Fig. 11. Processing Barbara image. $\text{corr}(u, v)$ as a function of t. The minimum is marked with 'X'. As seen in the SNR plot, the minimum correlation is not attained near the time with largest SNR.

Fig. 12. Effects of stopping criterion on processing results of different stopping times, processing Barbara image (head part is shown). Top left: noisy image f; Right - Weickert's method (23). Bottom left: Mrazek-Navara (22), right - our method of direct estimation.

APPENDIX

PROOF OF THEOREM 1

Since cov(q,r) = cov(r,q), the matrix is symmetric. The diagonal is the variance of each element, which is non negative. Therefore we have to check the covariance of the 10 elements of the upper right triangle.

We recall the identity

$$\text{cov}(q+r,s+t) = \text{cov}(qs) + \text{cov}(qt) + \text{cov}(rs) + \text{cov}(rt).$$

In the sequel we 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 cov$(s,n) = 0$, cov(f,s) = cov$(s+n,s) = V(s) \geq 0$, cov(f,n) = cov$(s+n,n) = V(n) \geq 0$.

$\mathbf{cov}(\boldsymbol{u},\boldsymbol{v})$, $\mathbf{cov}(\boldsymbol{f},\boldsymbol{u})$, $\mathbf{cov}(\boldsymbol{f},\boldsymbol{v})$

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

We follow the spirit of the proof of Meyer [6]. 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:

$$\int_\Omega \Phi(|\nabla(u-\epsilon h)|)d\Omega + \lambda V(v+\epsilon h) \geq \int_\Omega \Phi(|\nabla u|)d\Omega + \lambda V(v). \tag{25}$$

Replacing $V(v+\epsilon h)$ by $V(v) + \epsilon^2 V(h) + 2\epsilon$cov$(v,h)$ we get

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

Replacing h by 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.

$\mathbf{cov}(\boldsymbol{s},\boldsymbol{u})$, $\mathbf{cov}(\boldsymbol{n},\boldsymbol{u})$

Let us first examine an equivalent minimization problem to minimizing (4). Since $v = s + n - u$, then u that minimizes E_Φ is

$$\begin{aligned} u &= \text{argmin}_u\{\textstyle\int_\Omega \Phi(|\nabla u|)d\Omega + \lambda V(s+n-u)\} \\ &= \text{argmin}_u\{\textstyle\int_\Omega \Phi(|\nabla u|)d\Omega + \lambda(V(s)+V(n)+V(u) \\ &\quad +2\text{cov}(s,n) - 2\text{cov}(s,u) - 2\text{cov}(n,u))\}. \end{aligned}$$

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

$$\hat{E}_\Phi(u) = \int_\Omega \Phi(|\nabla u|)d\Omega + \lambda(V(u) - 2\text{cov}(s,u) - 2\text{cov}(n,u)), \tag{26}$$

where $u = \text{argmin}_u\{\hat{E}_\Phi(u)\}$. 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 (26) with $f = s+n$ as $\hat{E}^*_\Phi|_{f=s+n}$. The energy can be written as

$$\begin{aligned} \hat{E}^*_\Phi|_{f=s+n} &= \hat{E}_\Phi|_{f=s+n}(u^{s+n}) \\ &= \textstyle\int_\Omega \Phi(|\nabla u^{s+n}|)d\Omega + \lambda(V(u^{s+n}) - 2\text{cov}(s,u^{s+n}) - 2\text{cov}(n,u^{s+n})). \end{aligned} \tag{27}$$

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

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

In the above final expression, adding the term $-\lambda 2\text{cov}(n,u^{s+n})$ we obtain the right hand side of expression (27). Since we assume $\text{cov}(n,u^{s+n}) < 0$, we get the following contradiction

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

Similarly, the opposit case $\text{cov}(n,u^{s+n}) \geq 0$ and $\text{cov}(s,u^{s+n}) < 0$ is not possible.

$\mathbf{cov}(\boldsymbol{s},\boldsymbol{v})$, $\mathbf{cov}(\boldsymbol{n},\boldsymbol{v})$

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

PROOF OF PROPOSITION 4

Proof:

Part I: E_Φ

Let us define $(u_{\lambda_0}, v_{\lambda_0})$ as the solution for E_Φ with $\lambda = \lambda_0$. Then for any $\lambda = \lambda_0 - \epsilon$, where $0 < \epsilon < \lambda$, we have

$$\begin{aligned} E_\Phi|_{\lambda_0} &= \int_\Omega \Phi(|\nabla u_{\lambda_0}|)d\Omega + \lambda_0 V(v_{\lambda_0}) \\ &> \int_\Omega \Phi(|\nabla u_{\lambda_0}|)d\Omega + (\lambda_0 - \epsilon)V(v_{\lambda_0}) \\ &\geq \min_{(u,v)} \int_\Omega \Phi(|\nabla u|)d\Omega + (\lambda_0 - \epsilon)V(v) \\ &= E_\Phi|_{\lambda_0 - \epsilon}. \end{aligned}$$

Part II: $E_u,\ E_v$

We examine both energies together and show that the only possible option is that E_u decreases and E_v increases as λ decreases. Let us state the four possible options as λ decreases:

(a) E_u is increasing and E_v is nondecreasing.

(b) E_u is nonincreasing and E_v is decreasing.

(c) E_u is increasing and E_v is decreasing.

(d) E_u is nonincreasing and E_v is nondecreasing.

Option (a) is contradicted by setting the pair $(u_{\lambda_0}, v_{\lambda_0})$ in the energy with $\lambda = \lambda_0 - \epsilon$, reaching the contradiction $E_u(u_{\lambda_0}) + (\lambda_0 - \epsilon)E_v(v_{\lambda_0}) < E_\Phi|_{\lambda_0 - \epsilon}$. Option (b) is contradicted by setting the pair $(u_{\lambda_0 - \epsilon}, v_{\lambda_0 - \epsilon})$ in the energy with $\lambda = \lambda_0$, reaching the contradiction $E_u(u_{\lambda_0 - \epsilon}) + \lambda_0 E_v(v_{\lambda_0 - \epsilon}) < E_\Phi|_{\lambda_0}$.

Option (c) is somewhat more subtle. We assume that $E_v(v_{\lambda_0 - \epsilon})$ decreases by some measure $K_\epsilon > 0$. Then E_u must be bounded by $E_u(v_{\lambda_0 - \epsilon}) < E_u(v_{\lambda_0}) + \epsilon K_\epsilon$ (else we reach an immediate contradiction similar to option (a)). In this case we get the following inequalities

$$\begin{aligned} &E_u(u_{\lambda_0 - \epsilon}) + \lambda_0 E_v(v_{\lambda_0 - \epsilon}) \\ &< E_u(u_{\lambda_0}) + \epsilon K_\epsilon + \lambda_0 E_v(v_{\lambda_0 - \epsilon}) \\ &= E_u(u_{\lambda_0}) + \epsilon K_\epsilon + (\lambda_0 - \epsilon)(E_v(v_{\lambda_0}) - K_\epsilon) + \epsilon(E_v(v_{\lambda_0}) - K_\epsilon) \\ &= E_u(u_{\lambda_0}) + \lambda_0 E_v(v_{\lambda_0}) - (\lambda_0 - \epsilon)K_\epsilon. \end{aligned}$$

Since the term $(\lambda_0 - \epsilon)K_\epsilon$ is positive we reach the contradiction $E_u(u_{\lambda_0-\epsilon}) + \lambda_0 E_v(v_{\lambda_0-\epsilon}) < E_\Phi|_{\lambda_0}$.

Option (d) is, therefore, the only valid one. ■

DETAILED ALGORITHMS

We give below the general algorithm that covers both denoising methods (energy-based / time-flow) and both estimations (direct / indirect). When there is a difference in the algorithm we write the energy-based first and the time-flow second in curly brackets: {Energy}{Flow}. Explanations about parameters and a few remarks appear hereafter.

Main

1) Parameters: $szp, N_p, \{\lambda^0,\ \lambda_r\}\{DT\}$.
2) Set $Ecov^0 = 0$, $v^0 = 0$, $i = 0$.
3) Initialize according to method.
4) Loop
   a) $i \leftarrow i + 1,\ \{\lambda^i \leftarrow \lambda^{i-1}\lambda_r\}\{\}$.
   b) Compute u^i by {Eq. (4) with λ^i (use u^{i-1} as initial approximation)}{Eq. (21), evolving u^{i-1} by DT}.
   c) $v^i \leftarrow f - u^i$.
   d) $DEcov^i \leftarrow$ Estimated covariance derivative according to method.
   e) until ($DEcov^i < \frac{1}{2}$ (or ($i = N_p$))
5) (If direct method, remove patch from u)
6) Return u^{i-1}

Direct method

Initialization: adding a patch to the right of the image.

1) $mc \leftarrow$ mean value of right column of image.
2) $n_p(k,l) \leftarrow$ patch of random noise with variance σ^2.
3) $f_p(k,l) \leftarrow mc + n_p(k,l)$
4) $f \leftarrow [f\, f_p]$ (concatenate patch to right of image). We define $\Omega = \Omega_0 \cup \Omega_p$, where Ω_0 contains the input image and Ω_p contains the patch.

Estimation of covariance:

1) $v_p^i|_{\Omega_p} \leftarrow f|_{\Omega_p} - u^i|_{\Omega_p}$.
2) $Ecov^i \leftarrow < v_p^i, n_p >$ (discrete covariance, see (28)).
3) $DEcov^i \leftarrow (Ecov^i - Ecov^{i-1})/(V(v^i) - V(v^{i-1}))$.

Indirect method

Precomputing a discrete estimation of $\frac{\partial \text{cov}(n,v)}{\partial\{\lambda\}\{t\}}$.

1) Parameters: $\sigma^2, N_p, szp, \{\lambda_0, \lambda_r\}\{DT\}$.
2) $f \leftarrow$ noise patch.
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) $Ecov^i \leftarrow < v^i, f >$ (see (28)).
   d) $DEcov^i_{pre} \leftarrow (Ecov^i - Ecov^{i-1})/\{(\lambda^i - \lambda^{i-1})\}\{DT\}$
4) Return vector $DEcov_{pre}$

Estimation of covariance:

1) $DEcov^i \leftarrow DEcov^i_{pre} \cdot \{(\lambda^i - \lambda^{i-1})\}\{DT\}/(V(v^i) - V(v^{i-1}))$.

Remarks

- Parameters (in brackets are values used for processing natural images):
  1) szp - size of patch (direct - $10\cdot$(image length) pixels, indirect 80×80 pixels).
  2) N_p - number of precomputed points (different λ values or time-points) for indirect method. The main loop should do at most N_p iterations.
  3) λ^0 - initial λ (1), λ_r - ratio of successive λ (0.9).
  4) DT - time between consecutive timepoints. (We used $DT = 0.6$, 3 iterations of $dt = 0.2$ (where $dt < CFL$)).
- Discrete covariance:

$$< q, r > \equiv \frac{1}{N} \sum_{k,l} (q(k,l) - \bar{q})(r(k,l) - \bar{r}), \tag{28}$$

where N is the number of pixels in q (or r).

- With regard to the indirect method, in the specific implementation presented here, where the λ values / time points of the Main phase are exactly as in the Precomputing phase, one can actually omit the multiplication and division by $\{(\lambda^i - \lambda^{i-1})\}\{DT\}$ in the computation of $DEcov$ and $DEcov_{pre}$ (we kept it to be consistent with our formulation).

REFERENCES

[1] T.F. Chan, J. Shen, "A good image model eases restoration - on the contribution of Rudin-Osher-Fatemi's BV image model", IMA preprints 1829, Feb. 2002.

[2] R. Deriche, O. Faugeras, "Les EDP en traitement des images et vision par ordinateur", Traitement du Signal, 13(6), 1996.

[3] I.C. Dolcetta, R. Ferretti, "Optimal stopping time formulation of adaptive image filtering", Appl. Math. Opt. 43 (2001), 245-258.

[4] 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.

[5] P. Mrázek, M. Navara, "Selection of optimal stopping time for nonlinear diffusion filtering", IJCV, v. 52, no. 2/3, pp. 189-203, 2003.

[6] Y. Meyer, *Oscillatory Patterns in Image Processing and Nonlinear Evolution Equations*, Vol. 22 University Lecture Series, AMS, 2001.

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

[8] L. Rudin, S. Osher, E. Fatemi, "Nonlinear Total Variation based noise removal algorithms", Physica D 60 259-268,'92.

[9] L.A. Vese, S.J. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing", UCLA CAM Report 02-19, May 2002.

[10] J. Weickert, "Coherence-enhancing diffusion of colour images", IVC, Vol. 17, 201-212, 1999.

[11] J. Weickert, "A review of nonlinear diffusion filtering", B. ter Haar Romeny, L. Florack, J. Koenderink, M. Viergever (Eds.), Scale-Space Theory in Computer Vision, LNCS 1252, Springer, Berlin, pp. 3-28, 1997.