Estimation of Optimal PDE-based Denoising in the SNR Sense

Guy Gilboa, Nir Sochen and Yehoshua Y. Zeevi

***Abstract*— This paper is concerned with finding the best PDE-based denoising process, out of a set of possible ones. We focus either on finding the proper weight of the fidelity term in the energy minimization formulation, or on determining the optimal stopping time of a nonlinear diffusion process.**

A necessary condition for achieving maximal SNR is stated, based on the covariance of the noise and the residual part. We provide two practical alternatives for estimating this condition, by observing that the filtering of the image and the noise can be approximated by a decoupling technique, with respect to the weight or time parameters. Our automatic algorithm obtains quite accurate results on a variety of synthetic and natural images, including piecewise smooth and textured ones. We assume that the statistics of the noise were previously estimated. No *a-priori* knowledge regarding the characteristics of the clean image is required .

A theoretical analysis is carried out, where several SNR performance bounds are established for the optimal strategy and for a widely used method, wherein the variance of the residual part equals the variance of the noise.

I. Introduction

The use of Partial Differential Equations (PDE's) 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. For some comprehensive studies and background on the subject see [1], [23], [4], [17], [22] and the references therein. Invariably, these methods require the determination of a significant 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. The fact that the signal is affected is clear, since an image without noise is also altered in the process. The PDE's are constructed to reduce noise level at a faster rate than the piecewise smooth image parts are affected. Yet, the process must be stopped before the structure of the image has been modified too much, for example textured segments have become smooth.

G. Gilboa is with the Department of Mathematics, UCLA, 520 Portola Plaza, Los Angeles, CA 90095, USA, e-mail: gilboa@math.ucla.edu

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

Y.Y. Zeevi is with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Technion city, Haifa 32000, Israel. e-mail: zeevi@ee.technion.ac.il

It is thus important to determine what is the optimal point of stopping the process. This question is imperative, but, surprisingly, was seriously addressed in the context of PDE-based image processing only by a few studies [11], [18], [24].

We derive a necessary condition for optimality in the Signal-to-Noise Ratio (SNR) sense. From a practical viewpoint, the condition suggests a numerical method that should be followed for the purpose of maximizing the SNR of the filtered image. Two algorithms for the parameter calculation are proposed, based on the above condition, yielding fairly accurate estimates. From a theoretical viewpoint, this facilitates the computation of upper and lower bounds of the optimal strategy.

Next, we present an analysis of the optimal parameter from a SNR viewpoint. We also examine the popular denoising strategy, based on Morozov's discrepancy principle [13], used in the field of regularization theory. This method was most notably used in variational image processing in the seminal Rudin-Osher-Fatemi paper [18]. The selection of the weight of the fidelity term is such that the variance of the residual part equals that of the noise. A 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.

We demonstrate our method and show its advantages with respect to the methods of [18], [11] and [24].

Our main focus in this paper is on variational denoising (Sections II-V). In Section II we present the variational denoising model and derive the optimality condition. Two practical methods are provided for the approximation of this condition in Section III. In Section IV an analysis of the SNR performance is carried out, where lower and upper bounds are established. In Section V we present numerical results on a set of benchmark images. Similar methods are applied to diffusion-like processes in Section VI. A detailed comparison to other stopping criteria is presented. The comparison is carried out from both theoretical and empirical viewpoints. Conclusions and future directions are discussed in Section VII. Proofs and details of the algorithms are provided in the appendix. A short version of the ideas presented here can be found in [10].

II. Denoising Model and Optimality Condition

We try to solve the additive noise model, where the input signal f is composed of the original signal s and additive uncorrelated noise n of variance σ^2:

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

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 this decomposition by minimizing the following energy functional:

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

Φ is assumed to be convex. For a convex Φ the solution (u, v) exists and is unique [1]. More explanations and examples regarding this type of regularization can be seen e.g. in [18], [6], [1], [21], [22], [8], [16]. Some of the following results are also applicable to the more general case of monotonically increasing Φ. This holds as long as Φ is regularized so that a minimizer exists (such as in the discrete case or by convolving the gradient) and is unique. For the sake of simplicity we remain in the convex framework.

The condition $\int_\Omega f d\Omega = \int_\Omega u d\Omega$ is set, (corresponding to the Neumann boundary condition of the evolutionary equations). This yields $\int_\Omega v d\Omega = 0$. 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

$$\text{cov}(q, r) \doteq \frac{1}{|\Omega|} \int_\Omega (q - \bar{q})(r - \bar{r}) d\Omega.$$

Note that these quantities are based on the empirical definitions, and therefore could be measured for a given image. We recall the identity

$$V(q + r) = V(q) + V(r) + 2\text{cov}(q, r).$$

The SNR of the recovered signal u is defined as

$$SNR(u) \doteq 10 \log \frac{V(s)}{V(u - s)} = 10 \log \frac{V(s)}{V(n - v)}, \tag{5}$$

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 (5) and (1):

$$SNR_0 \doteq SNR(f) = 10 \log \frac{V(s)}{V(n)} = 10 \log \frac{V(s)}{\sigma^2}. \tag{6}$$

A. Condition for Optimal SNR

We proceed by developing a necessary condition for the optimal SNR. In this convex problem we have a single degree of freedom of choosing $V(v)$ [1], [3]. We therefore can regard the SNR as a function, $SNR(V(v))$, and assume that it is smooth (see examples of SNR functions of different images in Fig. 8). A necessary condition for the maximum in the range $V(v) \in (0, V(f))$ is:

$$\frac{\partial SNR}{\partial V(v)} = 0. \tag{7}$$

Rewriting $V(n - v)$ as $V(n) + V(v) - 2\text{cov}(n, v)$, and using (7) and (5), yields

$$\frac{\partial \text{cov}(n, v)}{\partial V(v)} = \frac{1}{2}. \tag{8}$$

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 the SNR will increase. When we reach the condition of (8), the noise and the signal are equally filtered and one should therefore stop. If filtering is continued, more signal than noise is filtered (in terms of variance) and the SNR decreases.

There is also a possibility to have the maximum at the boundaries: If the 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 the optimal SNR is SNR_0. The other extreme case is when the SNR increases monotonically and is maximized for $V(v) = V(f)$ (the trivial constant solution $u = \bar{f}$). We shall 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.

We will now suggest two ways to approximate the covariance term.

III. Estimating The Optimal Solution

In order to approximate $\text{cov}(n, v)$, 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 normally we do not know in advance which regions of the image are smooth and which are not.

Our observation is that the extent of filtering of additive noise, with respect to λ, is not affected much by the underlying

image s. What mainly affects the denoising performance is the extent of filtering of s. This property is very natural in the linear case: $h * f$=$h * (s + n)$=$h * s + h * n$, where h is the filtering kernel, and $*$ denotes convolution. We show that in some sense a similar decoupling can be applied to the nonlinear case. Currently, we investigate the possibility to obtain an analytic expression for the approximation error.

Fig. 1. Illustration of the direct (patch) method. Left: input image f. Right: a patch of pure noise with statistics similar to n is attached to the right side of f.

A. Direct Estimation

We assume that we have access to a source of a synthetic noise generator. 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 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-priori* or could be well estimated beforehand). [See Fig. 1.] Knowing, for this patch, both v and n, we can compute their covariance. Note that although $cov(n, v)$ is estimated based on the patch, $V(v)$ is measured in the usual way based on the original image domain.

B. Indirect Estimation

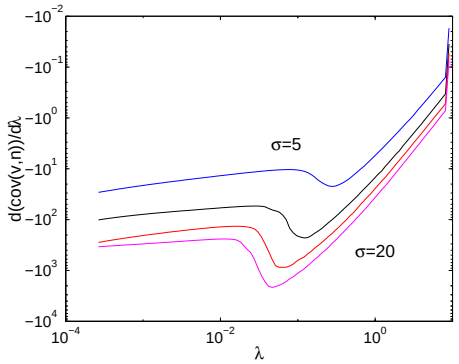

Fig. 2. Precomputed function for indirect estimation. $\partial \text{cov}(\tilde{n}, v)/\partial\lambda$ is plotted as a function of λ (log scale). Graphs depict plots for values of σ: $5, 10, 15, 20$, from upper curve to lower curve, respectively.

Another way of estimating $\text{cov}(n, v)$ 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, and which cause some side affects on the processed image near the patch and which affect the computations carried within the patch, are avoided.

The idea is to separate the computation into two phases. A patch of noise $\tilde{n}$ with similar statistics to n is processed and $\text{cov}(\tilde{n}, v)$ is measured with respect to λ. For the case of white Gaussian noise, only σ^2 should be estimated in order to generate $\tilde{n}$. Then the input image is processed and the behavior of λ with respect to $V(v)$ is measured. Combining the information, it is possible to approximate how $\text{cov}(n, v)$ behaves with respect to $V(v)$. In other words, we use the chain-rule for differentiation:

$$\begin{aligned} \frac{\partial \text{cov}(n,v)}{\partial V(v)} &= \frac{\partial \text{cov}(n,v)}{\partial \lambda}\frac{\partial \lambda}{\partial V(v)} \\ &\approx \frac{\partial \text{cov}(\tilde{n},v)}{\partial \lambda}|_{f=\tilde{n}}\frac{\partial \lambda}{\partial V(v)}|_{f=s+n}. \end{aligned} \tag{9}$$

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. 2). The second term is computed while the image is being processed.

In this scheme we rely on a very simplistic assumption that we can estimate the behavior of $\text{cov}(n, v)$ of any image based on the very degenerate case where the image is simply pure noise. Quite extraordinarily, our numerical experiments show that the estimations are not so far from the ground truth (see Fig. 8, right side). A more comprehensive approach may accommodate the computation of the function $\frac{\partial \text{cov}(n,v)}{\partial \lambda}$ based on a representative collection of natural images.

Numerical examples of both estimation methods are shown in Section V.

IV. SNR Bounds for the Scalar Φ Process

In this section we address a few theoretical issues and provide some bounds on the standard and optimal methods.

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:

$$\text{cov}(u^s, n) = 0, \quad \text{cov}(u^n, s) = 0, \quad \forall\lambda \geq 0. \tag{10}$$

We further assume that 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:

$$\begin{aligned} &\text{cov}(u^{s+n}, s) \leq \text{cov}(f, s), \quad \text{cov}(u^{s+n}, n) \leq \text{cov}(f, n), \\ &\forall\lambda \geq 0. \end{aligned} \tag{11}$$

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 (10) and (11) 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 (10) and (11).

Theorem 1: For any (s,n) pair and a convex increasing Φ ($\Phi'(q) > 0, \forall q > 0$) 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 that, 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.

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

$$SNR_{opt} \doteq \max_{\lambda} SNR(u_\lambda), \tag{12}$$

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

$$\min_u \int_\Omega \Phi(|\nabla u|) d\Omega \quad \text{subject to } V(v) = P. \tag{13}$$

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):

$$\lambda = \frac{1}{P} \int_\Omega \text{div}\left(\Phi'(|\nabla u|)\frac{\nabla u}{|\nabla u|}\right) v d\Omega. \tag{14}$$

The problem then transforms to which value P should be imposed (see [3], [1] for details).

A popular denoising strategy ([13], [18]) is to assume $v \approx n$ and therefore impose

$$V(v) = \sigma^2. \tag{15}$$

We define

$$SNR_{\sigma^2} \doteq SNR(u)|_{V(v)=\sigma^2}, \tag{16}$$

and denote by $(u_{\sigma^2}, v_{\sigma^2})$ the (u,v) pair that obeys (15) and minimizes (4). We shall now analyze this method for selecting u in terms of SNR.

Proposition 1 (SNR lower bound): Imposing (15), for any (s,n)-pair, SNR_{σ^2} is bounded from below by

$$SNR_{\sigma^2} \geq SNR_0 - 3dB, \tag{17}$$

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

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

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

■

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

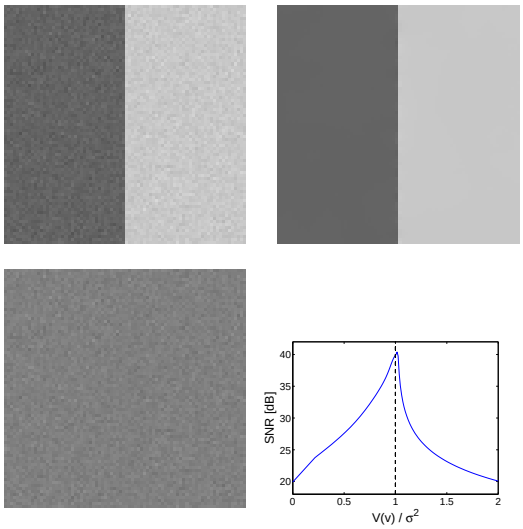

Fig. 3. Approaching best-case scenario in piece-wise constant images. In this example, for $V(v) = \sigma^2$, the SNR increases by almost 20dB from 19.9dB to 39.6dB (the variance of the noise is $\approx \frac{1}{100}$ of the input noise). Top: f (left), u (right). Bottom: v (left), SNR as a function of $V(v)/\sigma^2$ (right). In this case $V_{opt} = 1.02\sigma^2$, $SNR_{opt} = 40.2$dB.

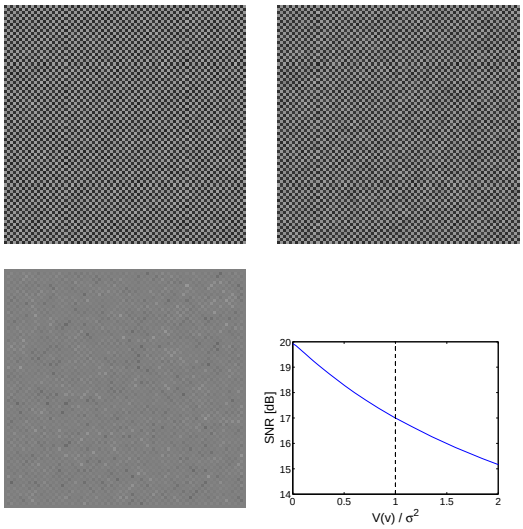

Fig. 4. Approaching worst-case scenario in a checkered-board image. For $V(v) = \sigma^2$, the SNR decreases by almost $3dB$ from 19.9dB to 17.0dB. Top: f (left), u (right). Bottom: v (left), SNR as a function of $V(v)/\sigma^2$ (right).

Proposition 2 (SNR upper bound): Imposing (15), 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 $SNR_0 = 10\log h$. Since the 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 (15), $V(v) = \sigma^2$, 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. 3 and 4. A signal that consists of a single very contrasted step function is shown in Fig. 3. 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. 4 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.

A. Regular SNR

Our experience shows that in these well-behaved denoising processes the SNR does not oscillate and has a single maximum. We use this significant observation for our estimation procedures and would like to assume this property also for the theoretical analysis. Let us define first the SNR regularity:

Definition 2 (Regular SNR): We define the function $SNR(V(v))$ as *regular* if (8) is a sufficient condition for optimality or if the optimum is at the boundaries.

Proposition 3 (Range of optimal SNR): If the 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:

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

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

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

$$0 \leq SNR_{opt} - SNR_0 \leq \begin{cases} -10\log(1 + V_{opt}/\sigma^2 - 2\sqrt{V_{opt}/\sigma^2}), & 0 \leq V_{opt} < \sigma^2 \\ -10\log(V_{opt}/\sigma^2 - 1), & \sigma^2 < V_{opt} \leq 2\sigma^2 \end{cases} \quad (18)$$

Proof: By the SNR definition, (5), and expanding the variance expression, we have

$$SNR_{opt} - SNR_0 = 10\log\left(\frac{\sigma^2}{\sigma^2 + V_{opt} - 2\text{cov}(n, v_{opt})}\right). \quad (19)$$

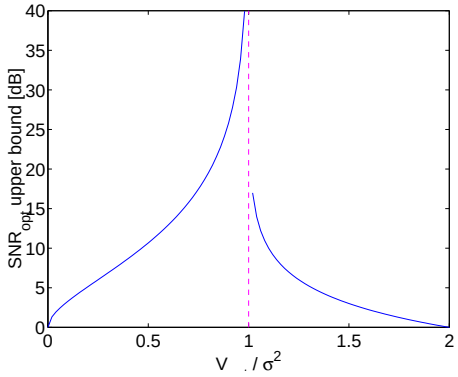

Fig. 5. Visualization of Theorem 2: Upper bound of $SNR_{opt} - SNR_0$ as a function of V_{opt}/σ^2. For $V_{opt} \to \sigma^2$ the bound approaches ∞.

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. 5.

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 change of $V(v)$, as well as the other energies,with respect to a change in λ. In the next proposition we show that as λ decreases the total energy E_Φ strictly decreases, the energy term $E_v(v) \doteq V(v)$ increases whereas the energy term $E_u(u) \doteq \int_\Omega \Phi(|\nabla u|) d\Omega$ decreases.

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

$$\frac{\partial E_\Phi}{\partial \lambda} > 0, \quad \frac{\partial E_v}{\partial \lambda} \leq 0, \quad \frac{\partial E_u}{\partial \lambda} \geq 0. \quad (20)$$

The proof is in the appendix.

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	N/A	0.43	0.06	0.06

TABLE I

COMPARISON OF METHODS PRESENTED IN SECTION III: 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.

V. VARIATIONAL DENOISING EXPERIMENTS

We compare our two methods for finding λ with the standard method of imposing (15) and with the optimal λ,

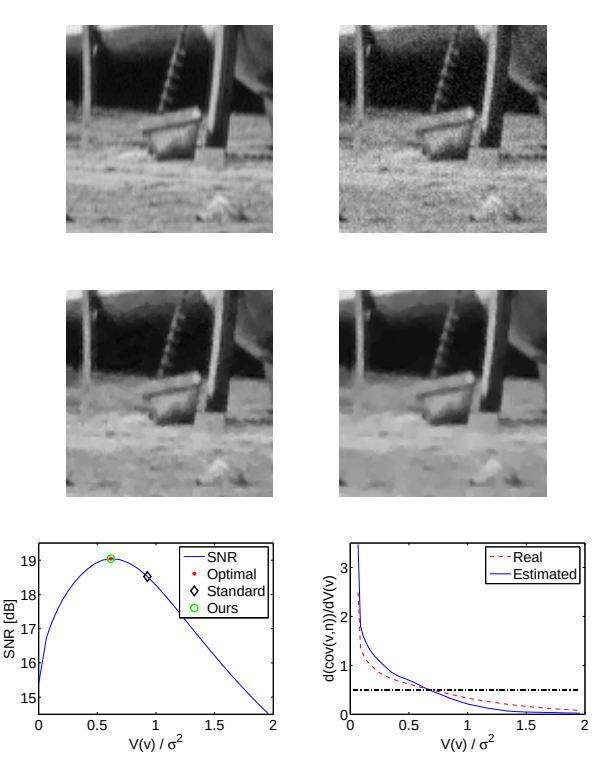

Fig. 6. Denoising part of the Boats image. Top row: s (left), f (right). Middle row: u by our direct estimation (left), u by standard method ($V(v) = \sigma^2$, right). Bottom row: SNR (left), and $\partial \text{cov}(n, v)/\partial V(v)$ (right) as a function of $V(v)/\sigma^2$.

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 I. Both of our methods are quite close to the optimal denoising (less than $0.1dB$ difference on average) and perform better than the standard scheme.

In Figures 6 and 7 results are shown for the direct and indirect estimations, respectively. Qualitatively, the proposed method (with both estimation techniques) tends to better preserve the textural information than the standard method.

We used $\Phi(s) = \sqrt{1+s^2}$, which can be viewed as the Vogel-Oman [21] regularization of TV [18] with $\epsilon = 1$ or the Charbonnier [5] process. The image grey-level range is $1:256$ so edges are well preserved. Other details about this experiment can be found in the appendix.

In Fig. 8 the terms $SNR(u)$ and $\frac{\partial \text{cov}(n,v)}{\partial \lambda}$ 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 point where $\frac{\partial \text{cov}(n,v)}{\partial V(v)} = \frac{1}{2}$. Naturally, this is advantageous to our algorithm. We currently have no explanation for this behavior.

VI. Evolutionary Flows

The process of the estimation of the optimal solution can be similarly formulated in evolutionary flows that do not have a fidelity term, e.g. [15], [23], [24], [11], [9]. We refer the reader

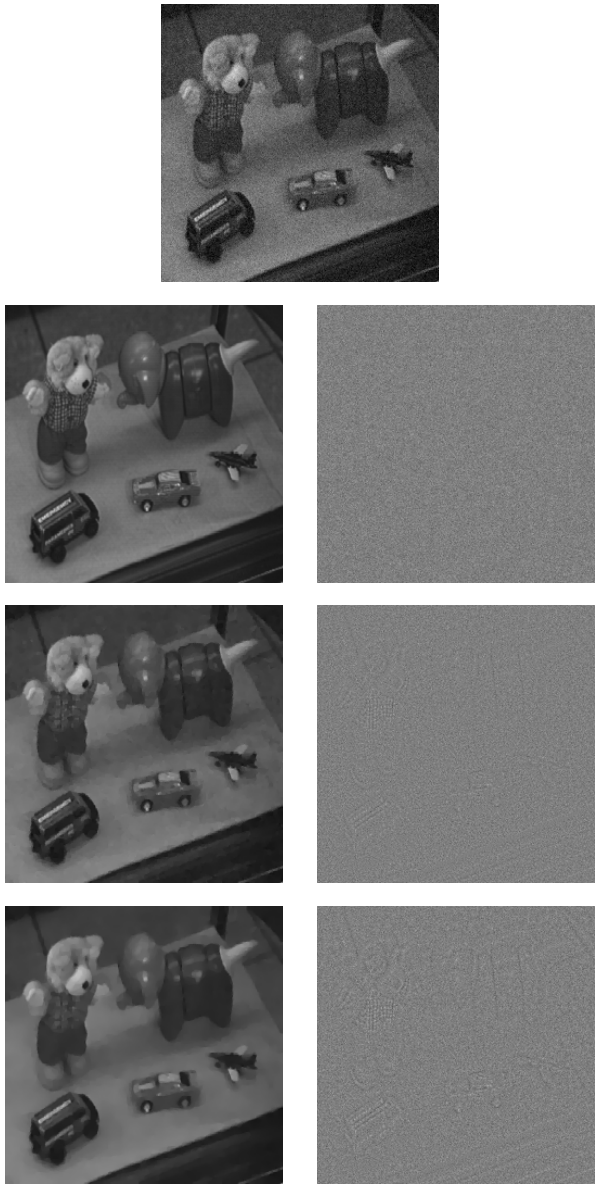

Fig. 7. Top row: f. Second row: s (left), n. Third row: u (left), v by our indirect estimation. Bottom row: u (left), v by standard method ($V(v) = \sigma^2$).

to [16], [19] to learn more on the close connections between variational denoising and nonlinear diffusion methods, and the similar role of the weight and time parameters. In the evolutionary case one has to select the best stopping time T. Our definitions are changed somewhat, but essentially have the same sense. The process is

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

where $c(|\nabla u|)$ is the adaptive diffusion coefficient. For convex processes one has to validate that $\int_0^s c(q)qdq$ is convex [1]. 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. For example, the

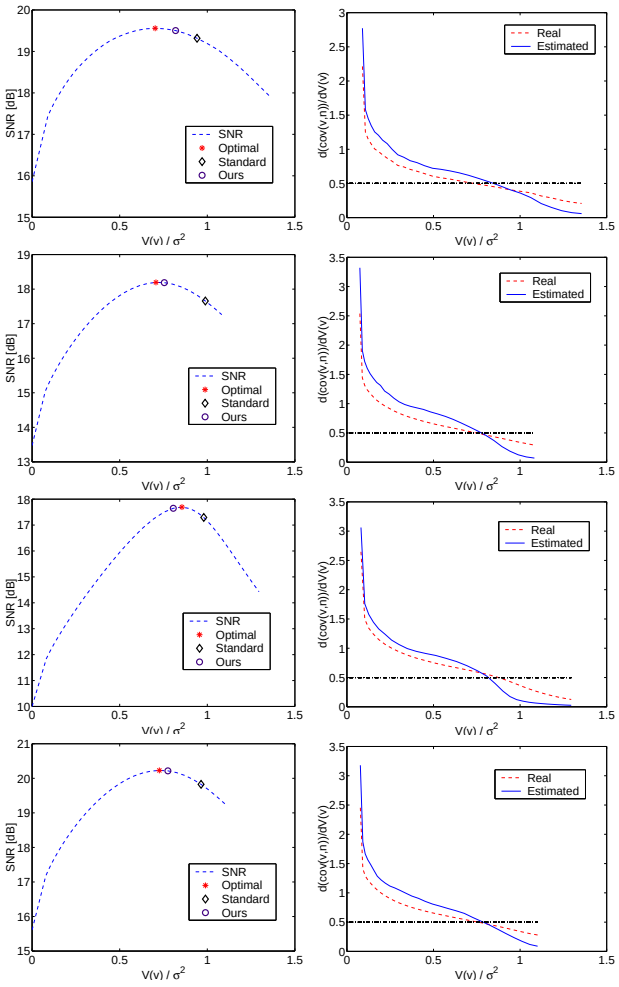

Fig. 8. SNR as a function of $V(v)/\sigma^2$ (left). $d\text{cov}(n,v)/dV(v)$ as a function of $V(v)/\sigma^2$ (right), as computed by indirect estimation (solid) and the ground truth (dashed). Graphs depict processing of the following natural images (from top): Cameraman, Lena, Toys, Boats.

indirect estimation of Eq. (9), for evolutionary flows is

$$\frac{\partial \text{cov}(n,v)}{\partial V(v)} \approx \frac{\partial \text{cov}(\tilde{n},v)}{\partial t}|_{f=\tilde{n}} \frac{\partial t}{\partial V(v)}|_{f=s+n}. \quad (22)$$

The detailed algorithm for implementing this method is in the appendix. Note that we use t (time) as the scale parameter (and not $V(v)$) as, though $V(v(t))$ almost always increases with time, we cannot guarantee it.

A. Comparison to Previous Stopping Mechanisms

A comprehensive study of the stopping time problem is discussed in [11]. Here we relate to the most recent method proposed by Mrazek and Navara [11] and the more classical one suggested by Weickert in [24].

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

$$T = \text{argmin}_t \{\text{corr}(u(t), v(t))\}, \quad (23)$$

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

$$\frac{V(u(T))}{V(f)} = \frac{1}{1 + V(n)/V(s)} \quad (24)$$

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

$$V(v) = V(n) - 2\text{cov}(u,v). \quad (25)$$

All three methods of imposing (15), [11] and [24], 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 [24] the process is often stopped considerably before the optimal time.

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 and fine details.

The main advantage of the method proposed in [11] is that no knowledge of the noise variance is required. It is also easy to compute, without any need for estimations. It is, however, 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 (23) may produce undesirable results. Actually, its performance in terms of SNR, cannot be bound from below such as is determined by Proposition 1. One can construct examples where the stopping time should be near $t = 0$, whereas $\text{corr}(u,v)$ decreases for a very long duration. This can be illustrated, for example, by the checkered-board image. The curves of the SNR function and the correlation are depicted in Fig. 10. In a more realistic example of processing the Barbara image (Fig. 11), the results are not as extreme, but image is considerably over-smoothed.

The method of [24] is similar in its spirit to imposing (15). Here, though, the term $2\text{cov}(u,v)$ is being deducted, resulting in an early stopping of the process (especially when u and v are highly correlated as in the case of textured images). In any case, the stopping time 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 [24] and [11] are illustrated in In Figs. 11, 12 and 13. 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 implicit assumption of both [24] and [11], which regards v as mostly containing noise. In partly textured images, v contains both noise and texture. In the case of [24], the term $\text{cov}(u,v)$ is large, and the process stops too early. In the case of [11], the consequences are more severe and $\text{corr}(u,v)$ is minimal only when the texture is smoothed out (see Fig. 12 for a plot of the correlation function). In terms of SNR, applying the method of [11] to this image results in a drop of more than $3dB$ below SNR_0. The SNR results are: $SNR_0 = 14.73$,

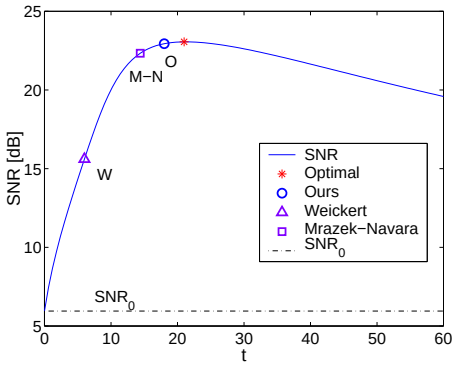

Fig. 9. Processing a step image (as in Fig. 3). SNR plot as a function of t. Stopping time is sufficiently close to the optimal selection by both methods of Mrazek-Navara and ours.

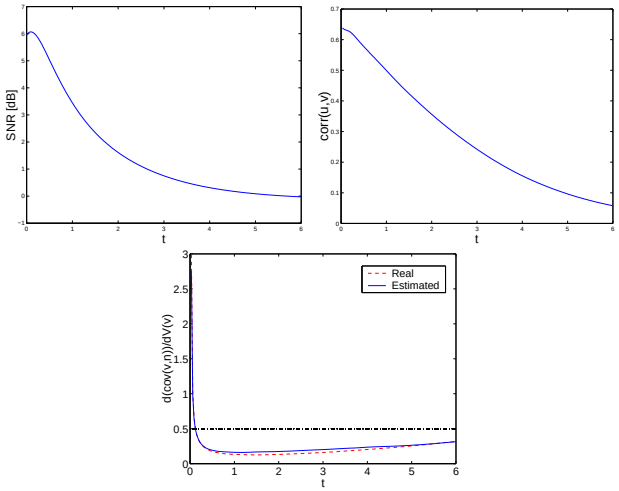

Fig. 10. A checkered-board image (medium contrast) with noise: Top: Left - SNR as a function of t, right - $\mathrm{corr}(u, v)$ as a function of t, bottom - $d\mathrm{cov}(n, v)/dV(v)$ as a function of t. Whereas the criterion of Eq. (23) cannot be used in this example (no local minimum near 0), our estimation of the general criterion stated in Eq. (8) works well also on highly textured signals (stopping time is $T = 0.12$ versus the optimal $T_{opt} = 0.09$).

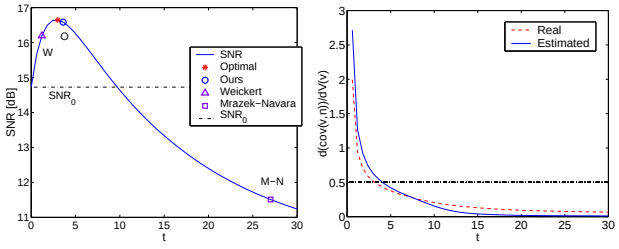

Fig. 11. Processing Barbara image. Left: SNR plot as a function of t. Right: $d\mathrm{cov}(n, v)/dV(v)$ as a function of t.

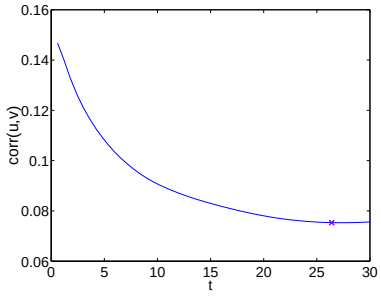

Fig. 12. Processing Barbara image. $\mathrm{corr}(u, v)$ as a function of t. The minimum is marked with 'X'. As seen in the SNR plot, the minimum correlation is not attained near the time with the largest SNR.

Fig. 13. 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 (24). Bottom left: Mrazek-Navara (23), right - our method of direct estimation.

$SNR_{opt} = 16.65$, $SNR_W = 16.19$, $SNR_{MN} = 11.51$, $SNR_{GSZ} = 16.59$, which stand for the SNR of the input image, the optimal denoising, the method of [24] , the method of [11] and our direct estimation method (attaching a "noise patch"), respectively. For the image results see Fig. 13.

VII. Conclusion

Most image denoising processes are quite sensitive to the choice and fine tuning of various parameters. In order to reach fully automatic denoising procedures, systematic methodologies for determining the appropriate parameters of a given image are a prerequisite. This problem motivated us to develop a new method 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 evolutionary processes.

Our criterion is to maximize the SNR, resulting from the application of a PDE-based denoising process. We provide two practical alternatives for estimating this condition, by observing that the filtering of the noise with respect to the weight or the time parameters is in some sense decoupled of the filtering of the clean image. Thus, we can study the behavior of a noise with similar statistics with respect to the nonlinear filtering process and utilize it for the approximation. This is done without assuming any knowledge of the clean image. Our method yields with sufficient accuracy the first local maximum of the SNR with respect to the variance of the residual part v (or time in nonlinear diffusion). In principle, there can be additional local maxima at larger scales with higher values of SNR. In practice, however, we have not encountered a natural image nor managed to generate a synthetic one, wherein the SNR depicts more than a single maximum. Our experience leads us to the empirical conclusion that such cases with peculiar SNR are quite rare in convex PDE-based processing.

We compare the performance of our algorithm with the performance of those obtained by means of previously proposed algorithms [18], [11], [24] and demonstrate that our method achieves better results on a series of benchmark images.

Bounds on the SNR of the optimal strategy (which we estimate) and the one used by ROF [18] are presented. These are proved for all signal and noise pairs which obey a strong decorrelation property (10) and a non-enhancement property (11), with respect to the process. Further studies may extend this framework by finding new bounds and relations or, perhaps, by using more relaxed assumptions. At this stage, we have not found a numerical example where these assumptions are violated.

We should also comment that the SNR criterion is not always in accord with human-based subjective criteria of quality evaluations. For the purpose of achieving this, other, more sophisticated criteria, may also be applied for parameter selection using the spirit of the methods presented here. For a recent report applying this method to a generalized Hilbert-space SNR we refer the reader to [2]. Thus, whereas the criterion developed and applied in this study yields sufficiently promising results, it may be further elaborated and, perhaps, combined with additional criteria under the variational framework.

We have restricted the analysis, for practical reasons, to examination of the widely-studied classical case of additive white Gaussian noise. Filtering other types of additive and uncorrelated noise may be analyzed in a similar manner. Generalizations to other regularization processes, and to non-stationary spatially varying parameters [8], are under current investigation.

Acknowledgements

Guy Gilboa acknowledges support by grants from the NSF under contracts ITR ACI-0321917, DMS-0312222, and the NIH under contract P20 MH65166. Nir Sochen acknowledges support by MUSCLE, Multimedia Understanding through Semantics, Computation and Learning, 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. Yehoshua Zeevi acknowledges support by the Ollendorf Minerva Center, the Fund for the Promotion of Research at the Technion and the HASSIP Research Network program HPRN-CT-2002-00285, funded by the European Commission.

APPENDIX

PROOF OF THEOREM 1

The covariance matrix of $U = (f, s, n, u, v)^T$ has 25 elements. 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 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.

$\mathbf{cov}(s, n)$, $\mathbf{cov}(f, s)$, $\mathbf{cov}(f, n)$

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

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

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 [12]. 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:

$$\int_\Omega \Phi(|\nabla(u-\epsilon h)|)d\Omega+\lambda V(v+\epsilon h) \geq \int_\Omega \Phi(|\nabla u|)d\Omega+\lambda V(v). \tag{26}$$

Replacing $V(v + \epsilon h)$ by $V(v) + \epsilon^2 V(h) + 2\epsilon\text{cov}(v, h)$ 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.

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

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, therefore, the equivalent energy functional to be minimized is:

$$\hat{E}_\Phi(u) = \int_\Omega \Phi(|\nabla u|)d\Omega + \lambda(V(u) - 2\text{cov}(s,u) - 2\text{cov}(n,u)), \tag{27}$$

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 (27) with $f = s+n$ as $\hat{E}^*_\Phi|_{f=s+n}$. The energy can be written as

$$\begin{aligned} \hat{E}^*_\Phi|_{f=s+n} &= \hat{E}_\Phi|_{f=s+n}(u^{s+n}) \\ &= \textstyle\int_\Omega \Phi(|\nabla u^{s+n}|)d\Omega + \lambda(V(u^{s+n}) \\ &\quad -2\text{cov}(s,u^{s+n}) - 2\text{cov}(n,u^{s+n})). \end{aligned} \tag{28}$$

On the other hand, according to condition (10), $\text{cov}(u^s, n) = 0$ 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 (28). 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 opposite case $\text{cov}(n,u^{s+n}) \geq 0$ and $\text{cov}(s,u^{s+n}) < 0$ is not possible.

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

This follows directly from condition (11) 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 $0 < \epsilon < \lambda_0$, we have

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

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

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

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 (29)).
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 (29)).
   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 - $10\times$(image length) pixels, indirect 80×80 pixels).
  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).
  4) DT - time between consecutive timepoints. (We used $DT = 0.6$, 3 iterations of $dt = 0.2$ (where $dt < CFL$)).

  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.
- Discrete covariance:

$$< q, r > \equiv \frac{1}{N} \sum_{k,l} (q(k,l) - \bar{q})(r(k,l) - \bar{r}), \qquad (29)$$

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

[2] J.F. Aujol, G. Gilboa, "Implementation and parameter selection for BV-Hilbert space regularizations", UCLA CAM Report 04-66, November 2004.

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

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] G. Gilboa, N. Sochen, Y.Y. Zeevi, "Image enhancement and denoising by complex diffusion processes", PAMI 26(8), pp. 1020-1036, 2004.

[10] G. Gilboa, N. Sochen, Y.Y. Zeevi, "Estimation of the optimal variational parameter via SNR analysis", Scale-Space 2005, LNCS 3459, R. Kimmel, N. Sochen, J. Weickert (Eds.), pp. 230–241, 2005.

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

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

[13] V. A. Morozov, "On the Solution of Functional Equations by the Method of Regularization", Soviet Math. Dokl., Vol. 7, pp. 414–417, 1966.

[14] 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.

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

[16] E. Radmoser, O. Scherzer and J. Weickert, "Scale-space properties of nonstationary iterative regularization methods", Journal of Visual Communication and Image Representation, Vol. 8, pp. 96-114, 2000.

[17] B M ter Haar Romeny Ed., *Geometry Driven Diffusion in Computer Vision*, Kluwer Academic Publishers, 1994.

[18] L. Rudin, S. Osher, E. Fatemi, "Nonlinear Total Variation based noise removal algorithms", Physica D 60 259-268, 1992.

[19] G. Steidl, J. Weickert, T. Brox, P. Mrazek, M. Welk, "On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs", SIAM J. Numer. Anal. 42(2), pp. 686-658, 2004.

[20] L.A. Vese, S.J. Osher, "Modeling textures with total variation minimization and oscillatory patterns in image processing", J. Scientific Computing, 19 , pp. 553-572, 2003.

[21] R.V. Vogel and M.E. Oman, "Iterative methods for total variation denoising", SIAM J. Scientific Computing, 17(1):227–238, 1996.

[22] 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.

[23] J. Weickert, *Anisotropic Diffusion in Image Processing*, ECMI Series, Teubner, Stuttgart, 1998.

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