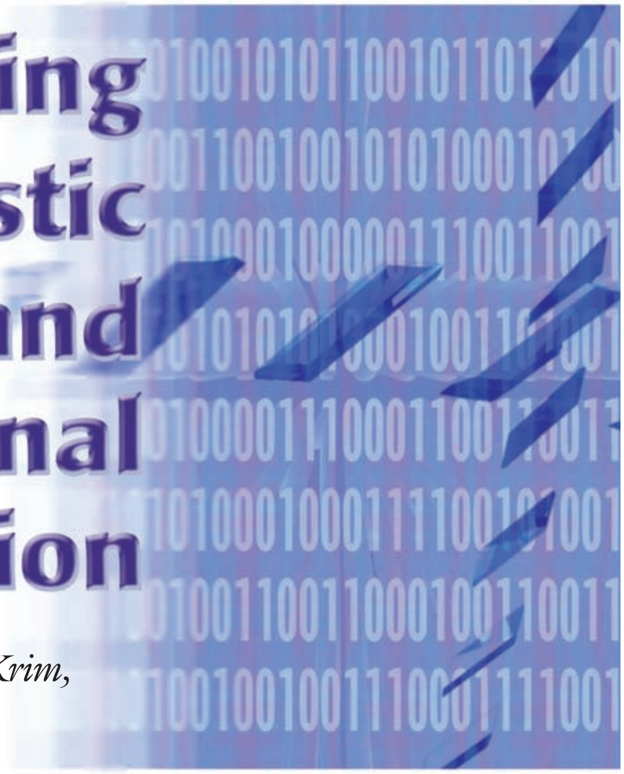
Unifying Probabilistic and Variational Estimation



A. Ben Hamza, Hamid Krim, and Gozde B. Unal

A maximum a posteriori (MAP) estimator using a Markov or a maximum entropy random field model for a prior distribution may be viewed as a minimizer of a variational problem. Using notions from robust statistics, a variational filter referred to as a Huber gradient descent flow is proposed. It is a result of optimizing a Huber functional subject to some noise constraints and takes a hybrid form of a total variation diffusion for large gradient magnitudes and of a linear diffusion for small gradient magnitudes. Using the gained insight, and as a further extension, we propose an information-theoretic gradient descent flow which is a result of minimizing a functional that is a hybrid between a negentropy variational integral and a total variation. Illustrating examples demonstrate a much improved performance of the approach in the presence of Gaussian and heavy tailed noise.

In this article, we present a variational approach to MAP estimation with a more qualitative and tutorial emphasis. The key idea behind this approach is to use geometric insight in helping construct regularizing functionals and avoiding a subjective choice of a prior in MAP estimation. Using tools from robust statistics and information theory, we show that we can extend this strategy and develop two gradient descent flows for image denoising with a demonstrated performance.

Introduction

Linear filtering techniques abound in many image processing applications and their popularity mainly stems from their mathematical simplicity and their efficiency in the presence of additive Gaussian noise. A mean filter, for example, is the optimal filter for Gaussian noise in the sense of minimum mean square error. Linear filters, however, tend to blur sharp edges, destroy lines and other fine image details, fail to effectively remove heavy tailed noise, and perform poorly in the presence of signal-dependent noise. This led to a search for nonlinear filtering alternatives. The research effort on nonlinear median-based filtering, for example, has resulted in remarkable results and has highlighted some new promising research avenues [1]. On account of its simplicity, its edge preservation property and its robustness to impulsive noise, the standard median filter remains among the favorites for image processing applications [1]. The median filter, however, often tends to remove fine details in the image, such as thin lines and corners [1]. In recent years, a variety of median-type filters such as stack filters and weighted median filters [1] have been developed to overcome this drawback. In spite of an improved performance, the solutions would clearly benefit from the regularizing power of a prior on the underlying information of interest.

1053-5888/02/$17.00©2002IEEE

Linear filtering techniques abound in many image processing applications and their popularity mainly stems from their mathematical simplicity and their efficiency in the presence of additive Gaussian noise.

Among Bayesian image estimation methods which enjoy such regularizations, the MAP estimator with a Markov or a maximum entropy random field prior [2]-[4] has proven to be a powerful approach to image restoration. The limitation in using MAP estimation is the difficulty of systematically and reliably choosing a prior distribution and its corresponding optimizing energy function and in some cases of the resulting computational complexity.

In recent years, variational methods and partial differential equation (PDE) based methods [5], [6] have been introduced to explicitly account for intrinsic geometry to address a variety of problems including image segmentation, mathematical morphology, and image denoising [7], [8]. The latter will be the focus of the present article. The problem of denoising has been addressed using a number of different techniques including wavelets [9], order-statistics-based filters [1], PDE-based algorithms [7], [8], and variational approaches [10]-[12]. In particular, a large number of PDE-based methods have particularly been proposed to tackle the problem of image denoising [7] with a good preservation of edges. Much of the appeal of PDE-based methods lies in the availability of a vast arsenal of mathematical tools which at the very least act as a key guide in achieving numerical accuracy as well as stability. PDEs or gradient descent flows are generally a result of variational problems using the Euler-Lagrange principle [13]. One popular variational technique used in image denoising is the total variation based approach. It was developed in [6] to overcome the basic limitations of all smooth regularization algorithms, and a variety of numerical methods have also recently been developed for solving total variation minimization problems [6], [14].

Image Analysis: Two Perspectives

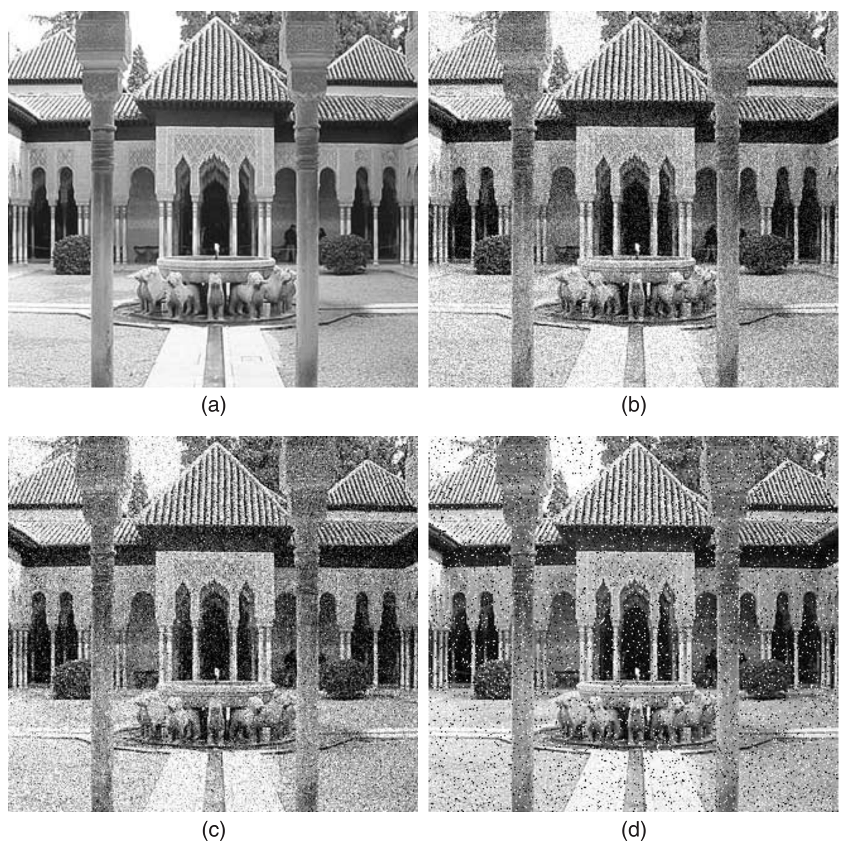
Problem Statement

In all real applications, measurements are perturbed by noise. In the course of acquiring, transmitting, or processing a digital image, for example, the noise-induced degradation may be dependent or independent of data. The noise is usually described by its probabilistic model, e.g., Gaussian noise is characterized by two moments. Application-dependent, a degradation often yields a resulting signal/image observation model, and the most commonly used is the additive one

$$u_0 = u + \eta, \qquad (1)$$

where the observed image u_0 includes the original signal u and the independent and identically distributed (i.i.d) noise process η. Fig. 1 depicts an image contaminated by three types of noise: Gaussian, Laplacian, and impulsive.

Image denoising refers to the process of recovering an image contaminated by noise (see Fig. 2). The challenge of the problem of interest lies in faithfully recovering the



▲ *1. Image denoising problem: (a) original image, (b) Gaussian noise, (c) Laplacian noise, and (d) impulsive noise.*

underlying signal/image u from u_0, and furthering the estimation by making use of any prior knowledge/assumptions about the noise process η. This goal is graphically and succinctly described in Fig. 2.

Model-Based Approach

In a probabilistic setting, the image denoising problem is usually solved in a discrete domain, and in this case an image is expressed by a random matrix $u = (u_{ij})$ of gray levels. To account for prior probabilistic information we may have for u, a technique of choice is that of a maximum a posteriori estimation. Denoting by $p(u)$ the prior distribution for the unknown image u, the MAP estimator is given by

$$\hat{u} = \arg\max_{u} \{\log p(u_0|u) + \log p(u)\}, \tag{2}$$

where $p(u_0|u)$ denotes the conditional probability of u_0 given u.

A general model for the prior distribution $p(u)$ is that of a Markov random field (MRF) which is characterized by its Gibbs distribution given by [2]

$$p(u) = \frac{1}{Z}\exp\left\{-\frac{\mathcal{F}(u)}{\lambda}\right\},$$

where Z is a partition function and λ is a constant known as the temperature in the terminology of physical systems. $\mathcal{F}$ is called the energy function and has the form $\mathcal{F}(u) = \sum_{c \in \mathcal{C}} V_c(u)$, where $\mathcal{C}$ denotes a set of cliques (i.e., set of connected pixels) for the MRF, and V_c is a potential function defined on a clique. We may define the cliques to be adjacent pairs of horizontal and vertical pixels. Note that for large λ, the prior probability becomes flat, and for small λ, the prior probability exhibits sharp modes.

MRFs have been extensively used in computer vision particularly for image restoration, and it has been established that Gibbs distributions and MRFs are equivalent (e.g., see [2]). In other words, if a problem is defined in terms of local potentials then there is a simple way of formulating the problem in terms of MRFs. If the noise process η is i.i.d. Gaussian, then we have

$$p(u_0|u) = K\exp\left(-\frac{|u-u_0|^2}{2\sigma^2}\right),$$

where K is a normalizing positive constant, σ^2 is the noise variance, and $|\cdot|$ stands for the Euclidean norm or for the absolute value in the case of a scalar. Thus, the MAP estimator in (2) yields

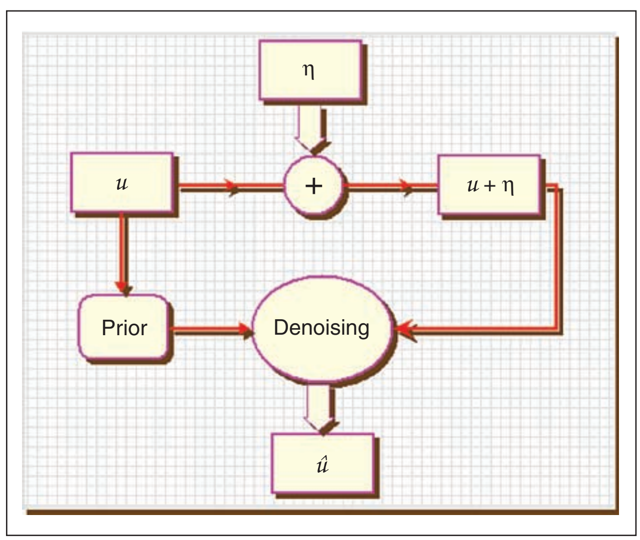
$$\hat{u} = \arg\min_{u}\left\{\mathcal{F}(u) + \frac{\lambda}{2}|u-u_0|^2\right\}. \tag{3}$$

Image estimation using MRF priors has proven to be a powerful approach to restoration and reconstruction of high-quality images. Its major drawback, besides its computational load, is the difficulty in systematically selecting a practical and reliable prior distribution. The Gibbs prior parameter λ is also of particular importance since it controls the balance of influence of the Gibbs prior and that of the likelihood. If λ is too small, the prior will tend to have an over-smoothing effect on the solution. Conversely, if it is too large, the MAP estimator may be unstable and it reduces to the maximum likelihood solution as λ goes to infinity. Another difficulty in using a MAP estimator is the nonuniqueness of the solution when the energy function $\mathcal{F}$ is not convex.

Image estimation using MRF priors has proven to be a powerful approach to restoration and reconstruction of high-quality images.

A Variational/Nonparametric Approach to MAP Estimation

Unknown prevailing statistics or underlying signal/image/noise models often make a "target" desired performance quantitatively less well defined. Specifically, it may be qualitative in nature (e.g., preserve high gradients in a geometric setting or determine a worst case noise distribution in a statistical estimation setting with a number of interpretations) and may not necessarily be tractably assessed by an objective and optimal performance measure. The formulation of such qualitative goals is typically carried out by way of adapted functionals which upon being optimized, achieve the stated goal, e.g., a monotonically decreasing functional of gradient modifying a diffusion [5]. This approach is the so-called variational approach. It is commonly formulated in a continuous domain which enjoys a large



▲ 2. Block diagram of image denoising process.

Image filtering refers to the process of recovering an image contaminated by noise.

arsenal of analytical tools, and hence offers a greater flexibility. An image is defined as a real-valued function $u{:}\Omega \rightarrow \mathbb{R}$, and Ω is a nonempty, bounded, open set in $\mathbb{R}^2$ (usually Ω is a rectangle in $\mathbb{R}$). Throughout, $\mathrm{x} = (x_1, x_2)$ denotes a pixel location in Ω, and $\|\cdot\|$ denotes the L^2-norm. While the ultimate overall objective in the aforementioned formulation may coincide with that of a probabilistic formulation, namely the recovery of an underlying desired signal u, it is herein often implicit and embedded in an energy functional to be optimized. Generally, the construction of an energy functional is based on some characteristic quantity specified by the task at hand (gradient for segmentation, Laplacian for smoothing, etc.). This energy functional is oftentimes coupled to a regularizing force/energy to rule out a great number of solutions and to also avoid any degenerate solution.

When considering the signal model (1), our goal may be succinctly stated as one of estimating the underlying image u based on an observation u_0 and/or any potential knowledge of the noise statistics to further regularize the solution. This yields the following fidelity-constrained optimization problem:

$$\begin{aligned} &\min_u \quad \mathcal{F}(u) \\ &\text{s.t.} \quad \|u - u_0\|^2 = \sigma^2 \end{aligned} \tag{4}$$

where $\mathcal{F}$ is a given functional which often defines, as noted above, the particular emphasis on the features of the achievable solution. In other words, we want to find an optimal solution that yields the smallest value of the objective functional among all solutions that satisfy the constraints. Using Lagrange's theorem, the minimizer of (4) is given by

$$\hat{u} = \arg\min_u \left\{ \mathcal{F}(u) + \frac{\lambda}{2}\|u - u_0\|^2 \right\}, \tag{5}$$

where λ is a nonnegative parameter chosen so that the constraint $\|u_0 - u\|^2 = \sigma^2$ is satisfied. In practice, the parameter λ is often estimated or chosen a priori.

Equations (3) and (5) show a close connection between image recovery via MAP estimation and image recovery via optimization of variational integrals. One may in fact reexpress (3) in an integral form similar to that of (5).

A critical issue, however, is the choice of the variational integral $\mathcal{F}$, which, as discussed later, is often driven by geometric arguments. Among the better known functionals (also called *variational integrals*) in image denoising are the Dirichlet and the total variation integrals defined, respectively, as

$$\mathcal{D}(u) = \frac{1}{2}\int_\Omega |\nabla u|^2\, dx \text{ and } TV(u) = \int_\Omega |\nabla u| dx,$$

where ∇u denotes the gradient of the image u.

A generalization of these functionals is the variational integral given by

$$\mathcal{F}(u) = \int_\Omega F(|\nabla u|)dx, \tag{6}$$

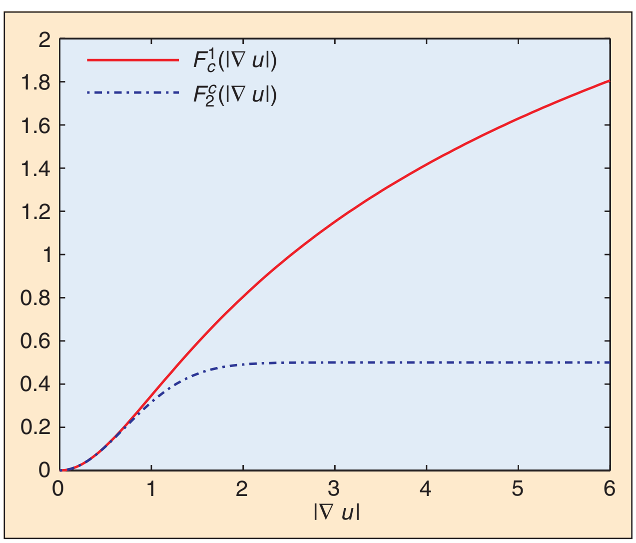
where $F{:}\mathbb{R}^+ \rightarrow \mathbb{R}$ is a given smooth function called a variational integrand or Lagrangian [13]. Using (6), we hence define a functional

$$\begin{aligned} \mathcal{L}(u) &= \mathcal{F}(u) + \frac{\lambda}{2}\|u - u_0\|^2 \\ &= \int_\Omega \left(F(|\nabla u|) + \frac{\lambda}{2}|u - u_0|^2 \right) dx, \end{aligned} \tag{7}$$

which by the formulation in (5) becomes

$$\hat{u} = \arg\min_{u \in X} \mathcal{L}(u), \tag{8}$$

where X is an appropriate image space of smooth functions.



▲ *3. Anisotropic Lagrangians.*

Numerical Solution: Gradient Descent Flows

To solve the optimization problem (8), a variety of iterative methods such as gradient descent [6], or fixed point method [15], may be applied.

The first-order necessary condition to be satisfied by any minimizer of the functional $\mathcal{L}$ given by (7) is its vanishing first variation (or vanishing gradient). Using the fundamental lemma of the calculus of variations, this vanishing first variation yields an Euler-Lagrange equation as a necessary condition to be satisfied by minimizers of $\mathcal{L}$. In mathematical terms, the Euler-Lagrange equation is given by

▲ 4. Filtering results: (a) original image, (b) heat flow, (c) Perona-Mailk flow, and (d) TV flow.

$$-\text{div}\left(\frac{F'(|\nabla u|)}{|\nabla u|}\nabla u\right)+\lambda(u-u_0)=0, \text{ in } \Omega, \tag{9}$$

where "div" stands for the divergence operator. An image u satisfying (9) is called an extremal of $\mathcal{L}$.

Using the Euler-Lagrange variational principle, the minimizer of (8) may be interpreted as the steady state solution to the following nonlinear elliptic PDE called gradient descent flow

$$u_t=\text{div}(g(|\nabla u|)\nabla u)-\lambda(u-u_0), \text{ in } \Omega\times\mathbb{R}_+,$$

where $g(z)=F'(z)/z$, with $z>0$, and assumed homogeneous Neumann boundary conditions.

Illustrative Cases

The following examples illustrate the close connection between optimization problems of variational integrals and boundary value problems for partial differential equations in a no noise constraint case (i.e., setting $\lambda=0$).

Heat Equation

$u_t=\Delta u$ is the gradient descent flow for the Dirichlet variational integral $\mathrm{D}(u)$.

It is important to point out that the Dirichlet functional tends to smooth out sharp jumps because it controls the second derivative of image intensity i.e., its "spatial acceleration," and it diffuses the intensity values isotropically. Fig. 4(b) shows this blurring effect on a clean image depicted in Fig. 4(a).

Perona-Malik Equation

It has been shown in [16] that the Perona-Malik (PM) diffusion $u_t=\text{div}(g(|\nabla u|)\nabla u)$ is the gradient descent flow for the variational integral

$$\mathcal{F}_c(u)=\int_\Omega F_c(|\nabla u|)dx,$$

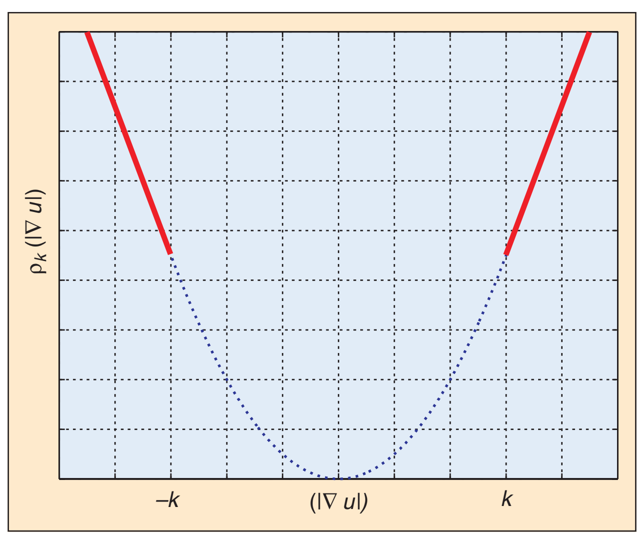
with sample Lagrangians $F_c^1(z)=c^2\log(1+z^2/c^2)$ or $F_c^2(z)=c^2(1-\exp(-z^2/c^2))$, see Fig. 3, where $z\in\mathbb{R}^+$ and c is a tuning positive constant.

A minimization of such functionals encourages the smoothing of homogenuous/small gradient regions and the preservation of edges/high gradient regions. Note that ill-posedness of this formulation was addressed in a number of papers (e.g., see [16]). A result of applying the PM flow with F_c^1 to the original image in Fig. 4(a) is illustrated in Fig. 4(c). It is worth noting how the diffusion takes place throughout the homogeneous regions and not across the edges.

Curvature Flow

$u_t=\text{div}(\nabla u/|\nabla u|)$ corresponds to the total variation integral.

While limiting spurious oscillations, TV optimization preserves sharp jumps as is often encountered in "blocky" signals/images. Fig. 4(d) illustrates the output of the TV flow.



▲ 5. Huber function.

The performance of a filter clearly depends on the filter type, the properties of signals/images, and the characteristics of the noise.

Robust Variational Approach

Robustness for Unknown Statistics

In robust estimation, for example, a case where even the noise statistics are *not* precisely known [17], [9] arises. In this case, a reasonable strategy would be to assume that the noise is a member of some set, or of some class of parametric families, and to pick the worst case density (least favorable, in some sense) member of that set, and obtain the best signal reconstruction for it. Huber's ε-contaminated normal set $\mathcal{P}_\varepsilon$ is defined as [17]

$$\mathcal{P}_\varepsilon = \{(1-\varepsilon)\Phi + \varepsilon H : H \in \mathcal{S}\},$$

where Φ is the standard normal distribution, $\mathcal{S}$ is the set of all probability distributions symmetric with respect to the origin and $\varepsilon \in [0,1]$ is the known fraction of "contamination." Huber found that the least favorable distribution in $\mathcal{P}_\varepsilon$ which maximizes the asymptotic variance (or, equivalently, minimizes the Fisher information) is Gaussian in the center and Laplacian in the tails. The transition between the two depends on the fraction of contamination ε, i.e., larger fractions correspond to smaller switching points and vice versa.

For the set $\mathcal{P}_\varepsilon$ of ε-contaminated normal distributions, the least favorable distribution has a density function $f_H(z) = ((1-\varepsilon)/\sqrt{2\pi}\exp(-\rho_k(z))$ (e.g., see [17]), where ρ_k is the Huber M-estimator cost function (see Fig. 5) given by

$$\rho_k(z) = \begin{cases} \dfrac{z^2}{2} & \text{if } |z| \le k \\ k|z| - \dfrac{k^2}{2} & \text{otherwise.} \end{cases}$$

Here k is a positive constant determined by the fraction of contamination ε [17].

Motivated by the robustness of the Huber M-filter in a probabilistic setting [1] and its resilience to impulsive noise, we propose a variational filter which, when accounting for these properties, leads to the following energy functional:

$$\mathcal{R}_k(u) = \int_\Omega \rho_k(|\nabla u|)dx.$$

Note that the Huber variational integral is a hybrid of the Dirichlet variational integral ($\rho_k(|\nabla u|) \propto |\nabla u|^2/2$ as $k \to \infty$) and of the total variation integral ($\rho_k(|\nabla u|) \propto |\nabla u|$ as $k \to 0$).

Using the Euler-Lagrange variational principle, a Huber gradient descent flow is obtained as

$$u_t = \operatorname{div}(g_k(|\nabla u|)\nabla u) - \lambda(u - u_0), \text{ in } \Omega \times \mathbb{R}_+,$$

where g_k is the Huber M-estimator weight function

$$g_k(z) = \frac{\rho_k'(z)}{z} = \begin{cases} 1 & \text{if } |z| \le k \\ \dfrac{k}{|z|} & \text{otherwise.} \end{cases}$$

For large k, this flow yields an isotropic diffusion (heat equation when $\lambda = 0$), and for small k, it corresponds to total variation gradient descent flow (curvature flow when $\lambda = 0$).

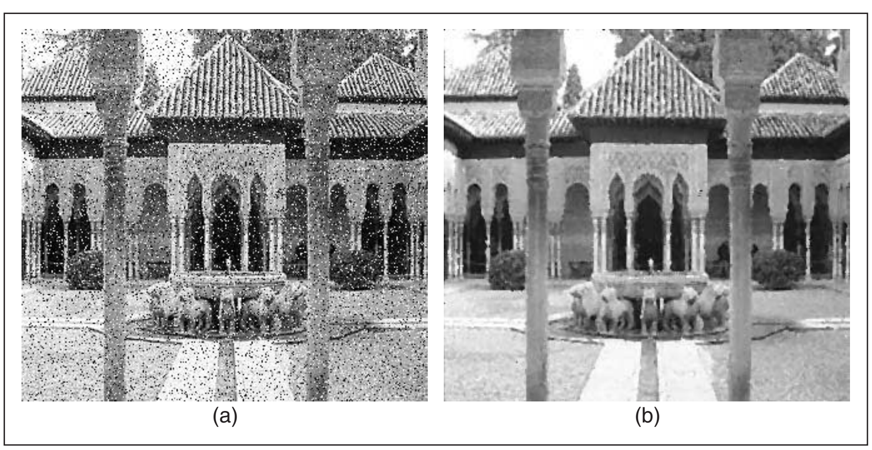
It is worth pointing out that in the case of no noise constraint (i.e., setting $\lambda = 0$), the Huber gradient descent flow yields a robust anisotropic diffusion [18] obtained by replacing the diffusion functions proposed in [5] with robust M-estimator weight functions [17], [1].

PM Equation: An Estimation-Theoretic Perspective

In a similar spirit as above, one may proceed to justify the PM equation from a specific statistical model. Assuming an image $u = (u_{ij})$ as a random matrix with i.i.d. elements, the output of the Log-Cauchy filter [21] is defined as a solution to the maximum log-likelihood estimation problem for a Cauchy distribution with dispersion c and estimation parameter θ. In other words, the output of a Log-Cauchy filter is the solution to the following robust estimation problem [21]:

$$\min_\theta \sum_{i,j} \log\left(c^2 + (u_{ij} - \theta)^2\right) = \min_\theta \sum_{i,j} F_c(u_{ij} - \theta),$$

where the cost function F_c coincides with the Lagrangian function which yields the PM equation. Hence, in the



6. *Log-Cauchy filtering: (a) impulsive noise and (b) filtered image.*

probabilistic setting the PM flow corresponds to the Log-Cauchy filter. Fig. 6 illustrates the performance of the Log-Cauchy filter in removing impulsive noise.

Information-Based Functionals

Information Theoretic Approach

In the previous section, we proposed a least favorable distribution as a result of exercising our ignorance in describing that of an image gradient within a domain. Another effective way is to adopt a criterion which bounds such a case, namely that of entropy. The maximum entropy criterion is indeed an important principle in statistics for modeling the prior probability $p(u)$ of a process u and has been used with success in numerous image processing applications [3]. The term is often associated with qualifying the selection of a distribution subject to some moments constraints (e.g., mean, variance, etc.), that is, the available information is described by way of moments of some known functions $m_r(u)$ with $r = 1, \ldots, s$. Coupling the finiteness of $m_r(u)$, for example, with the maximum entropy condition of the data suggests a most random model $p(u)$ with the corresponding moments constraints as a most adapted model (equivalently minimizing negentropy see [19]).

$$\begin{aligned}\min_u \quad & \int p(u)\log p(u)du \\ \text{s.t.} \quad & \int p(u)du = 1 \\ & \int m_r(u)p(u)du = \mu_r, \; r = 1, \ldots, s.\end{aligned} \tag{10}$$

Using Lagrange's theorem, the solution of (10) is given by

$$p(u) = \frac{1}{Z}\exp\left\{-\sum_{r=1}^{s}\lambda_r m_r(u)\right\}, \tag{11}$$

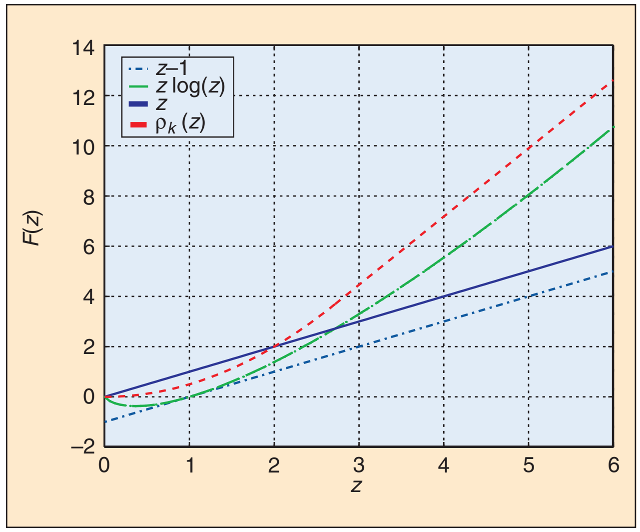
where λ_r s are the Lagrange multipliers, and Z is a partition function. The resulting model $p(u)$ given by (11) may hence be used as a prior in a MAP estimation formulation.

Entropic Gradient Descent Flow

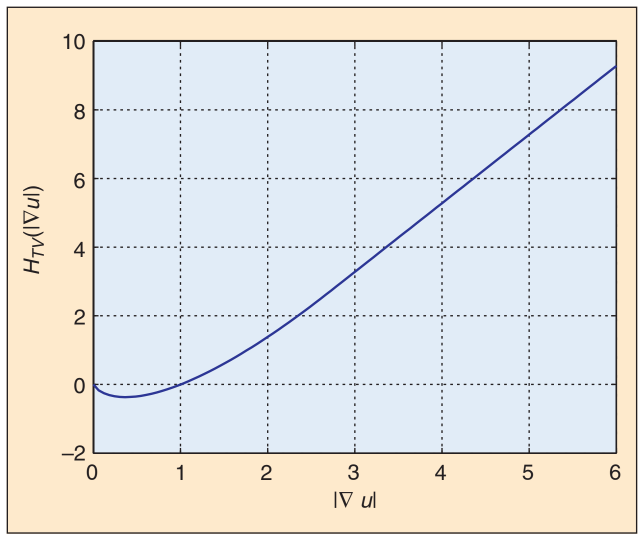
Motivated by the good performance of the maximum entropy principle in image/signal analysis applications and inspired by its rationale, we may naturally adapt it to describe the distribution of a gradient throughout an image. Specifically, the large gradients should coincide with tail events of this distribution, while the small and medium ones representing the smooth regions, form the mass of the distribution. Towards that end, we write

$$\mathcal{H}(u) = \int_\Omega H(|\nabla u|)dx = \int_\Omega |\nabla u| \log |\nabla u| dx.$$

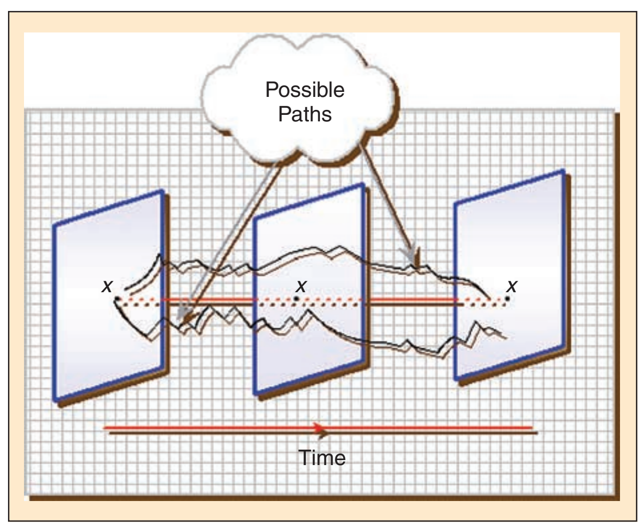
Calling upon the Euler-Lagrange variational principle again, the following entropic gradient descent flow results:



▲ 7. Visual comparison of some variational integrands.



▲ 8. Improved entropic Lagrangian.



▲ 9. A particle (pixel) may diffuse over many possible paths, and an average is usually computed.

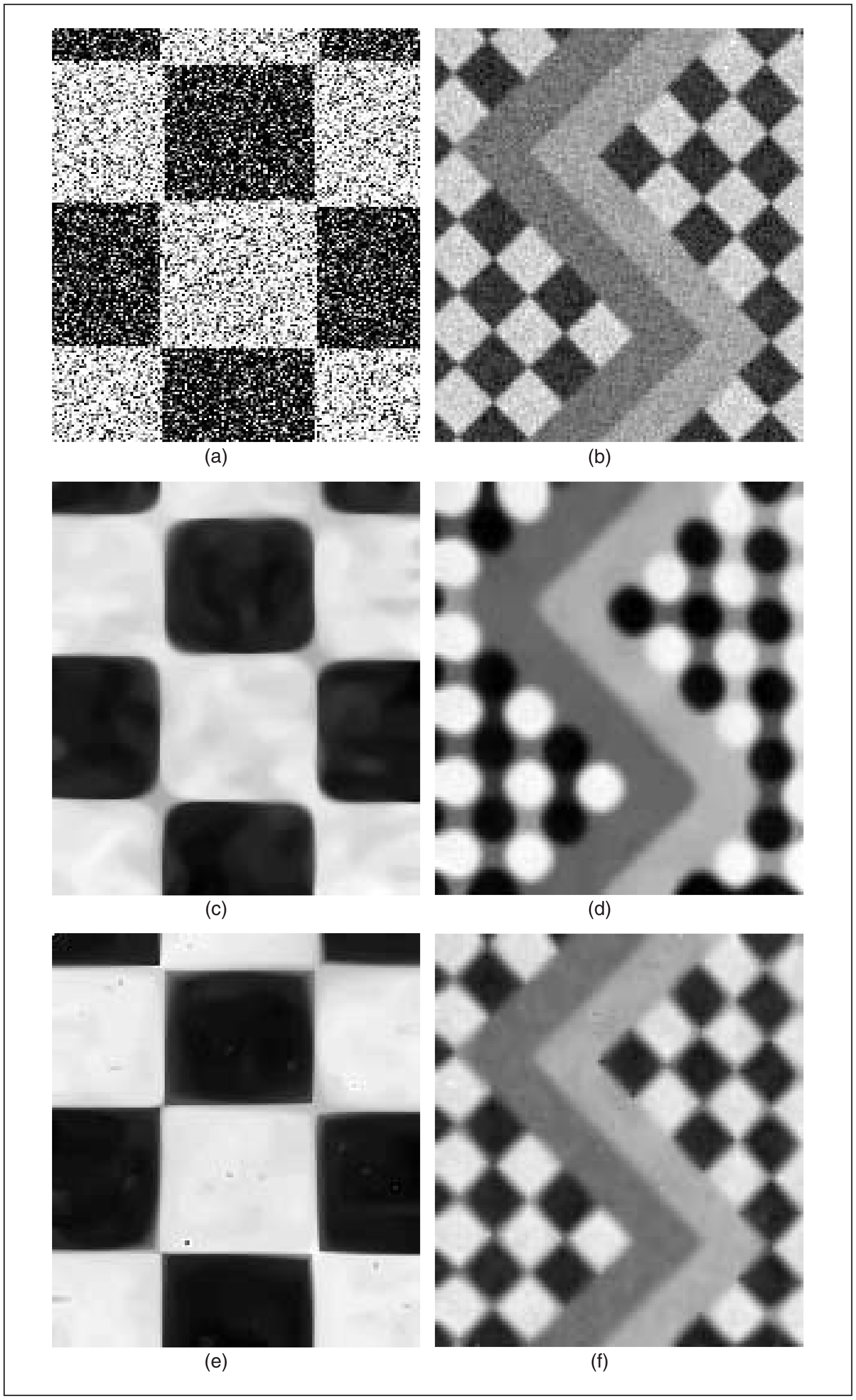
$$u_t = \mathrm{div}\left(\frac{1+\log|\nabla u|}{|\nabla u|}\nabla u\right) - \lambda(u-u_0), \text{ in } \Omega\times\mathbb{R}_+,$$

with homogeneous Neumann boundary conditions. In addition, this energy spread of the gradient energy may be related to that sought by the total variation method, which in contrast allows for additional higher gradients.

Improved Entropic Gradient Descent Flow

To summarize and for a comparison, we show in Fig. 7 the behavior of the variational integrands we have discussed in this article. It can be readily shown [20] that a differentiable hybrid functional between the negentropy variational integral and the total variation may be defined as

$$\widetilde{\mathcal{H}}(u) = \begin{cases} \mathcal{H}(u) & \text{if} |\nabla u| \le e \\ 2\ TV(u) - \mathrm{meas}(\Omega)e & \text{otherwise,} \end{cases}$$



▲ 10. Outputs of polygonal flows.

yielding an improved gradient descent flow. The quantity meas(Ω) denotes the Lebesgue measure of the image domain Ω, and $e = \exp(1)$. Fig. 8 depicts the Lagrangian corresponding to the improved entropic flow.

Physical Basis of Diffusion

In contrast to the macroscopic scale which reflects the large number of particles process typically modeled by a PDE and where large scale regularization is ill-posed, a microscopic scale approach may also be adopted.

The physical notion of diffusion pertains to the net transport of particles across a unit surface being proportional to the gradient of the material density normal to the unit area. A similar interpretation may be given to a diffusion of an image by modeling the motion of pixels along a Brownian (or random walk on a discrete lattice) trajectory (see Fig. 9). Invoking the microscopic scale of diffusion by way of modeling the particles trajectories helps clarify the dynamics which are often important to propose a creative solution.

A probabilistic view of the above evolution equations may hence be achieved with a careful interpretation of an image (a two-dimensional function) as a density of particles (image pixels). Specifically, a diffusion applied to an image is tantamount to evolving a probability density function of a process for which a particle trajectory (i.e., microscopic scale) is captured by a stochastic differential equation (SDE) [22]. The corresponding macroscopic process is modeled by a PDE and which for the linear case is the heat equation. This interpretation is not only intuitive, but also provides a powerful framework for possibly novel diffusions with unique properties. In [22], a nonlinear diffusion was developed for investigating the PM equation in this light, and thereby resolving a long standing problem of unknown stopping time for unconstrained nonlinear diffusion equations. A similar approach was used in developing polygonal flows important in preserving man-made shapes in images [24].

In Fig. 10, the top row shows two simple images with polygonal structures, namely rectangles and diamonds, corrupted by additive Gaussian noise. The middle row shows the results of applying the geometric heat flow (also known as curve shortening flow) [23], which acts on the image level curves, to the noisy images. The geometric blurring (i.e., the rounding effects) caused by the geometric heat flow can be overcome by using information on the orientation of the salient image lines. It follows that modified geometric heat flows can be designed for specific structures, and the corresponding results are illustrated in the bottom row of Fig. 10. It is important to note that the geometric diffusion is slowed down along important structures (i.e the rectangles and diamonds shapes) [24].

▲ *11. Filtering results for Gaussian noise: (a) original image, (b) noisy image, (c) Huber, (d) entropic, (e) total variation, and (f) improved entropic.*

Experimental Results

This section presents simulation results where Huber, entropic, total variation, and improved entropic gradient descent flows are applied to enhance images corrupted by Gaussian and Laplacian noise.

The performance of a filter clearly depends on the filter type, the properties of signals/images, and the characteristics of the noise. The choice of criteria by which to measure the performance of a filter presents certain difficulties and only gives a partial picture of reality. To assess the performance of the proposed denoising methods, a mean square error (MSE) between the filtered and the original image is evaluated and used as a quantitative measure of performance of the proposed techniques. The regularization parameter (or Lagrange multiplier) λ for the pro-

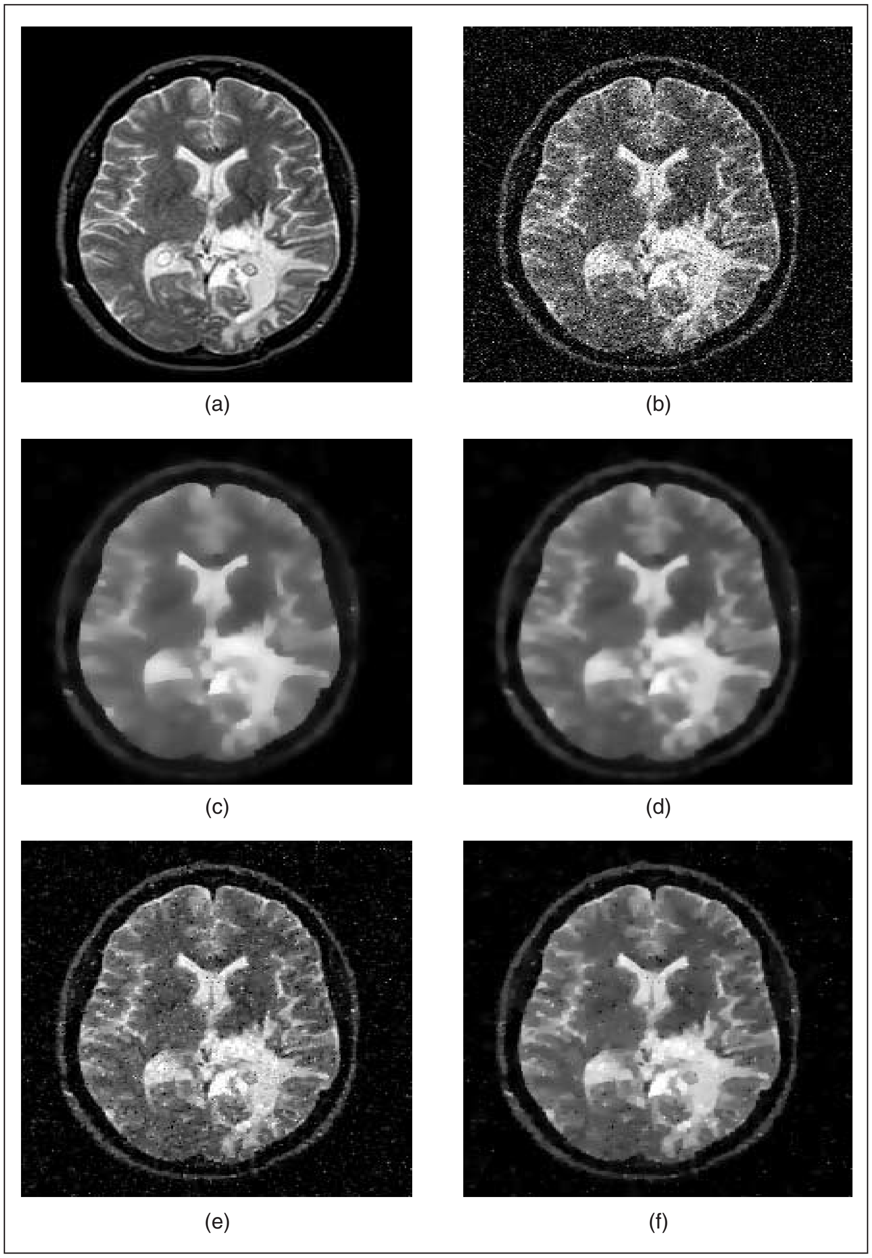
Table 1. MSE Computations for Gaussian noise.

PDE	MSE		
	SNR = 4.79	SNR = 3.52	SNR = 2.34
Huber	234.1499	233.7337	230.0263
Entropic	205.0146	207.1040	205.3454
TV	247.4875	263.0437	402.0660
Improved Entropic	121.2550	137.9356	166.4490

posed gradient descent flows is chosen to be proportional to signal-to-noise ratio (SNR) in all the experiments.

To evaluate the performance of the proposed gradient descent flows in the presence of Gaussian noise, the image shown in Fig. 11(a) has been corrupted by Gaussian white noise with SNR = 4.79 db. Fig. 11 displays the results of filtering the noisy image shown in Fig. 11(b) by Huber with optimal $k = 1.345$, entropic, total variation and improved entropic gradient descent flows. Qualitatively, we observe that the proposed techniques are able to suppress Gaussian noise while preserving important features in the image. The resulting MSE computations are depicted in Table 1.

The Laplacian noise is somewhat heavier than the Gaussian noise. Moreover, the Laplace distribution is similar to Huber's least favorable distribution [17] at least in the tails. To demonstrate the application of the proposed gradient descent flows to image denoising, qualitative and quantitative comparisons are performed to show a much improved performance of these techniques. Fig. 12(b) shows a noisy image contaminated by Laplacian white noise with SNR = 3.91 db. The MSE's results obtained by applying the proposed techniques to the noisy image are shown in Table 2. Note that from Fig. 12 it is clear that the improved entropic gradient descent flow outperforms the other flows in removing Laplacian noise. Comparison of these images clearly indicates that the improved entropic gradient descent flow preserves well the image structures while removing heavy tailed noise.



▲ *12. Filtering results for Laplacian noise: (a) original image, (b) noisy image, (c) Huber, (d) entropic, (e) total variation, and (f) improved entropic.*

Table 2. MSE Computations for Laplacian Noise.

PDE	MSE		
	SNR = 6.33	SNR = 3.91	SNR = 3.05
Huber	237.7012	244.4348	248.4833
Entropic	200.5266	211.4027	217.3592
TV	138.4717	176.1719	213.1221
Improved Entropic	104.4591	170.2140	208.8639

Acknowledgment

This work was supported in part by an AFOSR grant F49620-98- 1-0190 and NSF.

A. Ben Hamza received his degrees in applied mathematics. From March 2000 to February 2001, he was a research associate with the Electrical and Computer Engineering Department at North Carolina State University, Raleigh, where he is currently pursuing the Ph.D. degree. His research interests include nonlinear probabilistic and variational filtering, information-theoretic measures, and computer vision.

Hamid Krim received his degrees in electrical engineering. As a Member of Technical Staff at AT&T Bell Labs, he has worked in the areas of telephony and digital communication systems/subsystems. Following an NSF postdoctoral fellowship at Foreign Centers of Excellence, LSS/University of Orsay, Paris, France, he became a Research Scientist at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge. He is presently on the faculty in the Electrical and Computer Engineering Department, North Carolina State University, Raleigh. His research interests are in statistical signal and image analysis and mathematical modeling with a keen emphasis on applied problems. He is a Senior Member of the IEEE and an Associate Editor for the *IEEE Transactions on Signal Processing*.

Gozde B. Unal received her B.S. degree in electrical engineering from Middle East Technical University, Ankara, in 1996 and her M.S. degree in electrical engineering from the Bilkent University, Ankara, Turkey, in 1998. She is pursuing her Ph.D. in the Electrical and Computer Engineering Department at North Carolina State University, Raleigh, and is a research assistant with the Vision, Information and Statistical Signal Theories and Applications group. Her research interests include curve evolution theory with connections to information theory and probability theory, application of curve evolution techniques to various image and video processing problems such as image smoothing, image and texture segmentation, object tracking, and computer vision problems such as object recognition.

References

[1] J. Astola and P. Kuosmanen, *Fundamentals of Nonlinear Digital Filtering*. Boca Raton, FL: CRC, 1997.

[2] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 721-741, July 1984.

[3] H. Stark, Ed., *Image Recovery: Theory and Application*. New York: Academic, 1987.

[4] S.C. Zhu and D. Mumford, "Prior learning and Gibbs reaction-diffusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 1236-1244, Nov. 1997.

[5] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 629-639, July 1990.

[6] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259-268, 1992.

[7] D. Tschumperle and R. Deriche, "Diffusion PDEs on vector-valued images," *IEEE Signal Processing Mag.,* vol. 19, pp. 16-25, Sept. 2002.

[8] I. Pollak, "Segmentation and restoration via nonlinear multiscale filtering," *IEEE Signal Processing Mag.,* vol. 19, pp. 26-36, Sept. 2002.

[9] H. Krim and I.C. Schick, "Minimax description length for signal denoising and optimized representation," *IEEE Trans. Inform. Theory*, vol. 45, pp. 898-908, Apr. 1999.

[10] I. Pollak, A.S. Willsky, and H. Krim, "Image segmentation and edge enhancement with stabilized inverse diffusion equations," *IEEE Trans. Image Processing*, vol. 9, pp. 256-266, Feb. 2000.

[11] J. Morel and S. Solemini, *Variational Methods in Image Segmentation*. Boston, MA: Birkhaüser, 1995.

[12] C. Samson, L. Blanc-Feraud, G. Aubert, and J. Zerubia, "A variational model for image classification and restoration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 460-472, May 2000.

[13] M. Giaquinta and S. Hildebrandt, *Calculus of Variations I: The Lagrangian Formalism*. New York: Springer-Verlag, 1996.

[14] A. Chambolle and P.L. Lions, "Image recovery via total variation minimization and related problems," *Num. Math.*, vol. 76, pp. 167-188, 1997.

[15] K. Ito and K. Kunisch, "Restoration of edge-flat-grey scale images," *Inverse Problems*, vol. 16, no. 4, pp. 909-928, Aug. 2000.

[16] Y.L. You, W. Xu, A. Tannenbaum, and M. Kaveh, "Behavioral analysis of anisotropic diffusion in image processing," *IEEE Trans. Image Processing*, vol. 5, pp. 1539-1553, Nov. 1996.

[17] P. Huber, *Robust Statistics*. New York: Wiley, 1981.

[18] M.J. Black, G. Sapiro, D.H. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. Image Processing*, vol. 7, pp. 421-432, Mar. 1998.

[19] H. Krim, "On the distributions of optimized multiscale representations," *ICASSP-97*, vol. 5, pp. 3673 -3676, 1997.

[20] A. Ben Hamza and H. Krim, "A variational approach to maximum a posteriori estimation For Image denoising," Lecture Notes in Comput. Sci., vol. 2134, pp. 19-34, Sept. 2001.

[21] A. Ben Hamza and H. Krim, "Image denoising: A nonlinear robust statistical approach," *IEEE Trans. Signal Processing*, vol. 49, pp. 3045-3054, Dec. 2001.

[22] H. Krim and Y. Bao, "Smart nonlinear diffusion: A probabilistic approach," submitted for publication.

[23] M.A. Grayson, "The heat equation shrinks embedded plane curves to round points," *J. Differ. Geometry*, vol. 26, pp. 285-314, 1987.

[24] G.B. Unal, H. Krim, A. Yezzi, "Stochastic differential equations and geometric flows," *IEEE Trans. Image Processing*, to be published.