TOWARDS A UNIFIED VIEW OF ESTIMATION: VARIATIONAL VS. STATISTICAL

Hamid Krim *A. Ben Hamza*

Department of Electrical and Computer Engineering
North Carolina State University, Raleigh, NC 27695-7914, USA
E-mails: {ahk, abhamza}@eos.ncsu.edu

ABSTRACT

A connection between the maximum *a posteriori* (MAP) estimation and the variational formulation based on the minimization of a given variational integral subject to some noise constraints is established in this paper. A MAP estimator which uses a Markov or a maximum entropy random field model for the prior distribution can be viewed as a minimizer of a variational problem. Inspired by the maximum entropy principle, a nonlinear variational filter called *improved entropic gradient descent flow* is proposed. It minimizes a hybrid functional between the neg-entropy variational integral and the total variation subject to some noise constraints. Simulation results showing a much improved performance of the proposed filter in the presence of Gaussian and Laplacian noise are analyzed and illustrated.

1. INTRODUCTION

Linear filtering techniques have been used in many image processing applications and their popularity mainly stems from their mathematical tractability and their efficiency in the presence of additive Gaussian noise. 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. Among the class of Bayesian image estimation methods for example, the MAP estimator using Markov or maximum entropy random field priors [1, 2] has proven to be a powerful approach to image restoration. However, a major limitation in the use of MAP estimation is the lack of practical and robust method for choosing the prior distribution and its corresponding energy function.

In recent years, variational methods and partial differential equations (PDE) based methods [3, 4, 5] have been introduced for a variety of purposes inluding image segmentation, mathematical morphology and image denoising. This last topic will be the focus of the present paper. The problem of denoising has been addressed using a number of different techniques including wavelets [6] and nonlinear median based filters [7].

In this paper, we present a variational approach to MAP estimation. The key idea behind this approach is to avoid assumptions about the prior distribution in MAP estimation. Insipred by the maximum entropy principle, we propose a nonlinear PDE based filter called *improved entropic gradient descent flow*.

In the next section we outline the MAP estimation. In Section 3, we formulate a variational approach to MAP estimation. In section 4, an entropic variational approach to MAP estimation is given, and an improved entropic gradient descent flow is proposed. Finally, in section 5, we provide experimental results to show a much improved performance of the proposed gradient descent flows in image denoising.

This work was supported by an AFOSR grant F49620-98-1-0190 and by ONR-MURI grant JHU-72798-S2 and by NCSU School of Engineering

2. PROBLEM FORMULATION

Consider the additive noise model

$$u_0 = u + \eta, \tag{1}$$

where u is the original image $u : \Omega \to \mathbb{R}$, and Ω is a nonempty, bounded, open set in $\mathbb{R}^2$ (usually Ω is a rectangle in $\mathbb{R}^2$). The noise process η is i.i.d., and u_0 is the observed image. The objective is to recover u, knowing u_0 and also some statistics of η. Throughout, $\boldsymbol{x} = (x_1, x_2)$ denotes a pixel location in Ω, $|\cdot|$ denotes the Euclidean norm and $||\cdot||$ denotes the L^2-norm.

One commonly used Bayesian approach to estimate the image u is the maximum a posteriori (MAP) estimation method which incorporates prior information. Denote 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 a Markov random field which is characterized by its Gibbs distribution

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

where Z in the normalizing term called em partition function, λ is a constant known as the temperature in physical systems terminology. For large λ, the prior probability becomes flat, and for small λ, the prior probability has sharp modes. $\mathcal{F}$ is called the *energy function* and has the form $\mathcal{F}(u) = \sum_{c\in\mathcal{C}} V_c(u)$, where $\mathcal{C}$ denotes the set of cliques for the MRF, and V_c is a potential function defined on a clique.

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 positive constant and σ^2 is the noise variance. 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. However, a major problem limiting its utility is the lack of practical and robust method for selecting the 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, reducing to the maximum likelihood solution as λ goes to infinity. Another difficulty using the MAP estimator is the lack of stability and uniqueness of the solution when the energy function $\mathcal{F}$ is not convex.

3. A VARIATIONAL APPROACH TO MAP ESTIMATION

According to the noise model (1), the main goal of the image denoising problem is to estimate the original image u based on the observed image u_0 and any knowledge of the noise statistics. This leads to solve the following noise-constrained optimization problem

$$\begin{array}{ll} \min\limits_{u} & \mathcal{F}(u) \\ \text{s.t.} & \|u-u_0\|^2 = \sigma^2 \end{array} \tag{4}$$

where $\mathcal{F}$ is a given functional which is often a criterion of smoothness of the reconstructed image.

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 Lagrange multiplier λ is often estimated or chosen *a prioiri*.

Equations (3) and (5) show a close connection between image recovery via MAP estimation and image recovery via optimized variational integrals. Indeed, Eq. (3) can be written in integral form as Eq. (5).

Much like the choice of the prior distribution in MAP estimation, a critical issue in variational problems is the choice of the functional $\mathcal{F}$. In the latter case, much insight based on desired physical average behavior has been developed. The classical functionals (also called *variational integrals*) used in image denoising are the Dirichlet and the total variation integrals defined respectively as follows

$$\mathcal{D}(u) = \frac{1}{2}\int_{\Omega} |\nabla u|^2 d\boldsymbol{x}, \quad \text{and} \quad TV(u) = \int_{\Omega} |\nabla u| d\boldsymbol{x}, \tag{6}$$

where ∇u stands for the gradient of the image u.

A generalization of these functionals is the variational integral defined as

$$\mathcal{F}(u) = \int_{\Omega} F(|\nabla u|) d\boldsymbol{x}, \tag{7}$$

where $F : \mathbb{R}^+ \to \mathbb{R}$ is a given smooth function called *variational integrand* or *Lagrangian*

Using Eq. (7), we define the following functional

$$\mathcal{L}(u) = \int_{\Omega} \left(F(|\nabla u|) + \frac{\lambda}{2}|u-u_0|^2 \right) d\boldsymbol{x}, \tag{8}$$

thus, the optimization problem (5) becomes

$$\hat{u} = \arg\min_{u \in X} \mathcal{L}(u), \tag{9}$$

where X is an appropriate image space of smooth functions like $C^1(\overline{\Omega})$, or the space $BV(\Omega)$ of image functions with bounded variation, or the Sobolev space $H^1(\Omega) = W^{1,2}(\Omega)$.

The most important first-order necessary condition to be satisfied by any minimizer of the variational integral $\mathcal{L}$ is the vanishing of its first variation $\delta\mathcal{L}(u; v)$ at u in direction of v, that is

$$\delta\mathcal{L}(u; v) = \frac{d}{d\epsilon}\mathcal{L}(u + \epsilon v)\Big|_{\epsilon=0} = 0, \tag{10}$$

and a solution u of (10) is called a *weak extremal* of $\mathcal{L}$.

By means of the fundamental lemma of the calculus of variations, relation (10) yields the Euler-Lagrange equation as a necessary condition to be satisfied by any minimizer of $\mathcal{L}$. This Euler-Lagrange equation is given by

$$-\nabla \cdot \left(\frac{F'(|\nabla u|)}{|\nabla u|} \nabla u \right) + \lambda(u - u_0) = 0, \quad \text{in } \Omega \tag{11}$$

and an image u satisfying (11) (or equivalently $\nabla\mathcal{L}(u) = 0$) is called an *extremal* of $\mathcal{L}$.

Proposition 1 *Let $\lambda = 0$, and S be a convex set of image space X. If the Lagrangian F is nonnegative convex and of class C^1, then every weak extremal of $\mathcal{L}$ is a minimizer of $\mathcal{L}$ on S.*

Proof. The convexity of F yields

$$F(y) \geq F(x) + F'(x)(y - x), \quad \forall x, y \in \mathbb{R}_+. \tag{12}$$

By assumption u is a weak extremal of $\mathcal{L}$, ie. $\delta\mathcal{L}(u; v) = 0$ for all $v \in S$. This implies that $F'(|\nabla u|) = 0$. Therefore, using (12) we obtain

$$\int_{\Omega} F(|\nabla v|) d\boldsymbol{x} \geq \int_{\Omega} F(|\nabla u|) d\boldsymbol{x}.$$

This concludes the proof. ■

Proposition 2 *Let $\lambda = 0$, and S be a convex set of image space X. If the Lagrangian F is nonegative convex and of class C^1 such that $F'(0) \geq 0$, then the global minimizer of $\mathcal{L}$ is a constant image.*

Proof. Using (12), it follows that $F(|\nabla u|) \geq F(0)$. Thus the constant image is a minimizer of $\mathcal{L}$. Since S is convex, it follows that this minimizer is global. ■

Proposition 3 *Let $\lambda > 0$, and S be a convex set of image space X. If the Lagrangian F is nonnegative strictly convex and of class C^1, then an extremal u of $\mathcal{L}$ is the unique minimizer of $\mathcal{L}$ on S.*

Proof. Since $u \longmapsto \frac{\lambda}{2}|u - u_0|^2$ is strictly convex when $\lambda > 0$, then the functional $\mathcal{L}(u)$ is strictly convex on S, that is

$$\mathcal{L}(v) > \mathcal{L}(u) + \nabla\mathcal{L}(u) \cdot (v - u).$$

By assumption u is an extremal of $\mathcal{L}$, thus $\mathcal{L}(v) > \mathcal{L}(u)$, for all $v \neq u$. ■

Using the Euler-Lagrange variational principle, the minimizer of (9) can be interpreted as the steady state solution to the following PDE called *gradient descent flow*

$$u_t = \nabla \cdot (g(|\nabla u|)\nabla u) - \lambda(u - u_0), \quad \text{in } \Omega \times \mathbb{R}_+ \tag{13}$$

where g is the diffusion function given by $g(z) = F'(z)/z$, with $z > 0$, and assuming homogeneous Neumann boundary conditions.

4. ENTROPIC VARIATIONAL APPROACH

The maximum entropy criterion is an important principle in statistics for modeling the prior probability $p(u)$, and it has been used with success in many applications of image processing [2]. Suppose the available information by way of moments of some known functions $m_r(u)$, where $r \in \{1, \ldots, s\}$. The maximum entropy principle suggests that a good choice of the prior probability is the one that has the maximum entropy or equivalently has the minimum neg-entropy

$$\begin{array}{ll} \min\limits_{u} & \int p(u) \log p(u) du \\ \text{s.t.} & \int p(u) du = 1 \\ & \int m_r(u) p(u) du = \mu_r, \quad r = 1, \ldots, s. \end{array} \tag{14}$$

Using Lagrange's theorem, the solution of (14) is given by

$$p(u) = \frac{1}{Z} \exp\left\{ -\sum_{r=1}^{s} \lambda_r m_r(u) \right\}, \tag{15}$$

where λ_r's are the Lagrange multipliers, and Z is the partition function. Thus, the maximum entropy distribution $p(u)$ given by Eq. (15) can be used as a model for the prior distribution in MAP estimation.

4.1. Entropic gradient descent flow

Motivated by the good performance of the maximum entropy method in the probabilistic approach to image denoising, we define the neg-entropy variational integral as

$$\mathcal{H}(u) = \int_\Omega H(|\nabla u|) d\boldsymbol{x} = \int_\Omega |\nabla u| \log |\nabla u| d\boldsymbol{x}, \tag{16}$$

where $H(z) = z \log(z),\ z \geq 0$. Note that $-H(z) \to 0$ as $z \to 0$.

It follows from the inequality $z \log(z) \leq z^2$ that

$$|\mathcal{H}(u)| \leq \int_\Omega |\nabla u|^2 d\boldsymbol{x} \leq \|u\|^2_{H^1(\Omega)} < \infty, \quad \forall u \in H^1(\Omega),$$

where $\|\cdot\|_{H^1(\Omega)}$ denotes the H^1-norm. Thus the neg-entropy variational integral $\mathcal{H} : H^1(\Omega) \to \mathbb{R}$ is well defined. Clearly, the Lagrangian H is strictly convex, and coercive, i.e. $H(z) \to +\infty$ as $|z| \to +\infty$. It follows from Proposition 3, that for $\lambda > 0$, the minimization problem

$$\hat{u} = \arg \min_{u \in H^1(\Omega)} \int_\Omega \left(|\nabla u| \log |\nabla u| + \frac{\lambda}{2} |u - u_0|^2 \right) d\boldsymbol{x}$$

has a unique solution provided that $|\nabla u| \geq 1$.

Using the Euler-Lagrange variational principle, it follows that the entropic gradient descent flow is given by

$$u_t = \nabla \cdot \left(\frac{1 + \log |\nabla u|}{|\nabla u|} \nabla u \right) - \lambda (u - u_0), \quad \text{in } \Omega \times \mathbb{R}_+, \tag{17}$$

assuming homogeneous Neumann boundary conditions.

Proposition 4 *Let u be an image. The neg-entropy variational integral and the total variation satisfy the following inequality*

$$\mathcal{H}(u) \geq TV(u) - 1.$$

Proof. Since the neg-entropy H is a convex function, the Jensen inequality yields

$$\begin{aligned} \int_\Omega H(|\nabla u|) d\boldsymbol{x} &\geq H\left(\int_\Omega |\nabla u| d\boldsymbol{x} \right) \\ &= H\Big(TV(u) \Big) \\ &= TV(u) \log TV(u), \end{aligned}$$

and using the inequality $z \log(z) \geq z - 1$ for $z \geq 0$, we conclude the proof. ■

Fig. 1 illustrates a visual comparison between some of the variational integrands discussed in this paper.

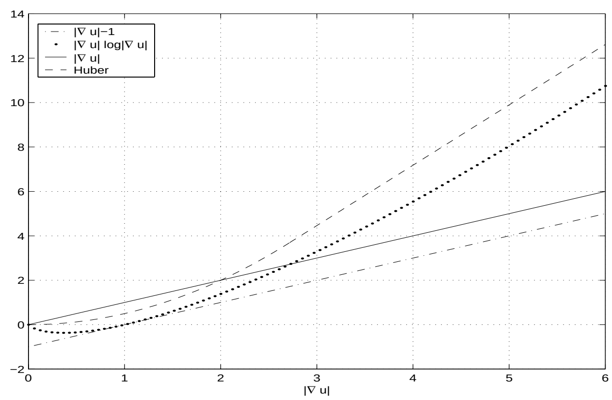

Fig. 1. Visual comparison of some variational integrands

4.2. Improved entropic gradient descent flow

From Fig. 1, one may define a hybrid functional between the neg-entropy variational integral and the total variation as follows

$$\widetilde{\mathcal{H}}(u) = \begin{cases} \mathcal{H}(u) & \text{if } |\nabla u| \leq e \\ TV(u) & \text{otherwise.} \end{cases}$$

Note that the functional $\widetilde{\mathcal{H}}$ is not differentiable when the Euclidean norm of ∇u is equal to e (i.e. Euler number: $e = \lim_{n \to \infty} (1 + 1/n)^n \approx 2.71$). This difficulty is overcome if we replace $\widetilde{\mathcal{H}}$ with the following functional $\mathcal{H}_{TV}$ defined as

$$\mathcal{H}_{TV}(u) = \begin{cases} \mathcal{H}(u) & \text{if } |\nabla u| \leq e \\ 2\ TV(u) - |\Omega| e & \text{otherwise,} \end{cases} \tag{18}$$

where $|\Omega|$ denotes the Lebesque measure of the image domain Ω. In the numerical implementation of our algorithms, we may assume without loss of generality that $\Omega = (0, 1) \times (0, 1)$, so that $|\Omega| = 1$. Note that $\mathcal{H}_{TV} : H^1(\Omega) \to \mathbb{R}$ is well defined, differentiable, weakly lower semicontinuous, and coercive. Using the Euler-Lagrange variational principle, it follows that the improved entropic gradient descent flow is given by

$$u_t = \nabla \cdot \left(\frac{H'_{TV}(|\nabla u|)}{|\nabla u|} \nabla u \right) - \lambda (u - u_0), \quad \text{in } \Omega \times \mathbb{R}_+, \tag{19}$$

assuming homogeneous Neumann boundary conditions. H'_{TV} is the derivative of the function $H_{TV} : \mathbb{R}^+ \to \mathbb{R}$ defined as

$$H_{TV}(z) = \begin{cases} z \log(z) & \text{if } z \leq e \\ 2z - e & \text{otherwise.} \end{cases}$$

5. SIMULATION 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 [8]. The regularization parameter (or Lagrange multiplier) λ for the proposed gradient descent flows is chosen to be proportional to signal-to-noise ratio (SNR) in all the experiments.

In order to evaluate the performance of the proposed gradient descent flows in the presence of Gaussian noise, the image shown in Fig. 1(a) has been corrupted by Gaussian white noise with $\mathrm{SNR} = 4.79$ db. Fig. 2 displays the results of filtering the noisy image shown in Fig. 2(b) by Huber, 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 Laplacian noise is somewhat heavier than the Gaussian noise. Moreover, the Laplace distribution is similar to Huber's least favorable distribution [6] (for the no process noise case), at least in the tails. To demonstrate the application of the proposed gradient descent flows to image denoising, qualititive comparisons are performed to show a much improved performance of these techniques. Fig. 3(b) shows a noisy image contaminated by Laplacian white noise with $\mathrm{SNR} = 3.91$ db. The filtered images are shown in Fig. 3. Note 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.

Fig. 2. Filtering results for Gaussian noise: (a) Original image, (b) Noisy image, (c) Huber flow, (d) Entropic flow, (e) Total Variation flow, and (f) Improved Entropic flow.

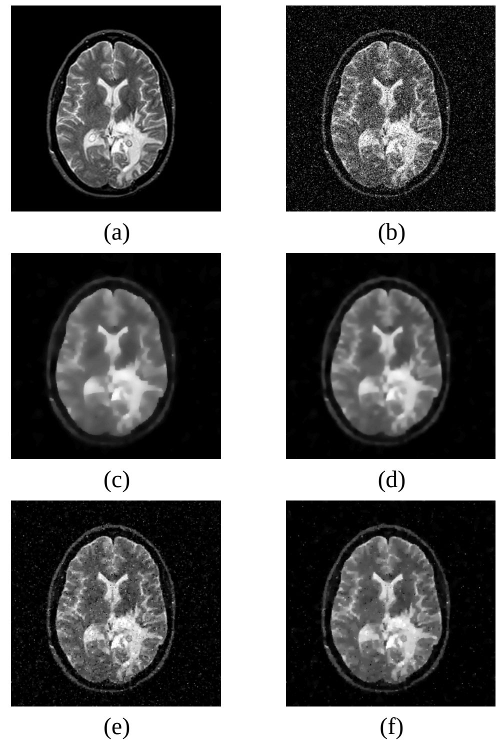

Fig. 3. Filtering results for Laplacian noise: (a) Original image, (b) Noisy image, (c) Huber flow, (d) Entropic flow, (e) Total Variation flow, and (f) Improved Entropic flow.

6. REFERENCES

[1] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 6, pp. 721-741, 1984.

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

[3] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 7, pp. 629-639, 1990.

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

[5] J Morel and S. Solemini, *Variational Methods in Image Segmentation*, Birkhauser, Boston, 1995.

[6] H. Krim and I.C. Schick, "Minimax description length for signal denoising and optimized representation," *IEEE Trans. Information Theory*, vol 45, no. 3, pp. 898-908, 1999.

[7] A. Ben Hamza, P. Luque, J. Martinez, and R. Roman "Removing noise and preserving details with relaxed median filters," *J. Math. Imaging and Vision*, vol. 11, no. 2, pp. 161-177, 1999.

[8] A. Ben Hamza and H. Krim, "A variational approach to MAP Estimation for image denoising," EMMCVPR, September 2001.