Exact Distribution of Edge-Preserving MAP Estimators for Linear Signal Models with Gaussian Measurement Noise

Jeffrey A. Fessler, *Member, IEEE,* Hakan Erdoğan, *Member, IEEE,* and Wei Biao Wu

***Abstract*—We derive the exact statistical distribution of maximum *a posteriori* (MAP) estimators having edge-preserving non-Gaussian priors. Such estimators have been widely advocated for image restoration and reconstruction problems. Previous investigations of these image recovery methods have been primarily empirical; the distribution we derive enables theoretical analysis. The signal model is linear with Gaussian measurement noise. We assume that the energy function of the prior distribution is chosen to ensure a unimodal posterior distribution (for which convexity of the energy function is sufficient), and that the energy function satisfies a uniform Lipschitz regularity condition. The regularity conditions are sufficiently general to encompass popular priors such as the generalized Gaussian Markov random field prior and the Huber prior, even though those priors are not everywhere twice continuously differentiable.**

***Index Terms*—Bayesian methods, image reconstruction, image restoration.**

I. Introduction

MANY papers have described edge-preserving methods for image reconstruction and image restoration based on the Bayesian formalism with non-Gaussian priors. Recent examples in this journal include [1]–[4]. Maximum *a posteriori* (MAP) estimators for non-Gaussian priors are nonlinear and defined implicitly (lacking an explicit expression). Therefore, in virtually all such papers, the evaluation of the *performance* of such methods has been investigated only empirically. This paper contributes a step toward an *analytical* understanding of edge-preserving MAP estimators by deriving their probability distribution functions (pdf's).

We attempted to analyze the mean and covariance of such implicitly defined estimators in [5] using linear approximations to the gradient of the objective function. However, non-Gaussian priors have nonquadratic energy functions that induce significant nonlinearities, rendering inaccurate the approximations in [5]. The mean and covariance analysis in [5] accommodates general measurement models. In this paper, we focus on linear Gaussian measurements. Remarkably, this simple restriction allows us to derive the *exact* (conditional) pdf for MAP estimators having a broad class of non-Gaussian priors. This explicit form for the pdf may be useful in analyzing the statistical properties of MAP estimation methods. The pdf may also be useful for Bayesian techniques such as Gibbs sampling and Markov chain Monte Carlo, e.g., [6] and [7].

This paper complements the work of Abbey *et al.* [8], [9], who developed an *approximate* pdf for maximum likelihood and MAP estimates under a general *nonlinear* signal model. (See also [10].) By considering *linear* signal models, we obtain the *exact* pdf, even for non-Gaussian priors.

Manuscript received October 28, 1998; revised October 24, 1999. This work was supported in part by the NIH under Grants CA-60711 and CA-54362 and by the Whitaker Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Patrick L. Combettes.

J. A. Fessler is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122 USA (e-mail: fessler@umich.edu).

H. Erdoğan is with IBM T.J. Watson Research Labs, Yorktown Heights, NY 10598 USA.

W. B. Wu is with the Department of Statistics, University of Michigan, Ann Arbor, MI 48109-1027 USA.

Publisher Item Identifier S 1057-7149(00)04699-6.

II. Problem

Let $X \in \mathbb{R}^n$ denote the unknown image vector, and let $Y \in \mathbb{R}^m$ denote the observed measurement vector (e.g. a noisy blurry image in image restoration, or a sinogram in image reconstruction). We assume a linear Gaussian measurement model, i.e., the conditional distribution of Y given $X = x$ is normal

$$f(y|x) = \frac{1}{\sqrt{2\pi|\boldsymbol{\Pi}|}} \exp\left(-\frac{1}{2}(y - \boldsymbol{A}x)'\boldsymbol{\Pi}^{-1}(y - \boldsymbol{A}x)\right) \tag{1}$$

where $\boldsymbol{A}$ is a known $m \times n$ system matrix and $|\boldsymbol{\Pi}|$ denotes the determinant of a known measurement noise covariance matrix $\boldsymbol{\Pi}$, assumed to be symmetric positive definite. Assume that the prior distribution for X has the usual Gibbs form

$$f(x) = ce^{-R(x)} \tag{2}$$

where c is a constant independent of x and $R(x)$ is an energy function that discourages image roughness. For edge-preserving image recovery, typically $R(x)$ is a nonquadratic function [1]–[4], usually composed of functions that increase less rapidly than quadratic functions [see (23)]. If $R(x)$ were quadratic, then both the prior (2) for X and the measurement model (1) would be Gaussian, so the posterior distribution $f(x|y)$ would also be Gaussian. We focus on the non-Gaussian case here, for which no explicit form for $f(x|y)$ has been previously found to our knowledge.

Suppose we observe $Y = y$**.** Under the above assumptions, the MAP estimator $\hat{X}$ for X, i.e., the maximizer of the posterior distribution $f(x|y)$, is equivalent, by Bayes rule, to the maximizer of $f(y|x)f(x)$, or equivalently to the minimizer of $\Phi(x) \triangleq -\log f(y|x) - \log f(x)$. Thus, the MAP estimator

for (1), (2) minimizes the following regularized weighted least squares objective function:

$$\hat{X} = \arg\min_x \Phi(x),$$
$$\Phi(x) = \tfrac{1}{2}(y - \boldsymbol{A}x)'\boldsymbol{\Pi}^{-1}(y - \boldsymbol{A}x) + R(x) \qquad (3)$$

ignoring constants independent of x, where "$'$" denotes matrix or vector transpose.

Our goal is to characterize statistically the implicitly defined estimator $\hat{X} = \hat{X}(Y)$ defined in (3), even though there is no explicit expression for $\hat{X}(\cdot)$. There are two reasonable choices at this point. We could try to find the unconditional distribution $f(\hat{x})$ or find the conditional distribution $f(\hat{x}|\tilde{x})$ for some true image $X = \tilde{x}$ of interest. One can find the former (in principle) from the latter by applying total probability

$$f(\hat{x}) = \int f(\tilde{x}|\tilde{x}) f(\tilde{x})\, d\tilde{x}.$$

A devout Bayesian might focus on the unconditional distribution $f(\hat{x})$ but such a Bayesian would need to have faith that the prior distribution $f(x)$ properly reflects the *global* prior characteristics of the unknown image. The fairly simple energy functions $R(x)$ that are typically used in practice generally only capture *local* properties of natural images.[1] Such priors are useful for MAP estimation, but may be ill-suited for global ensemble statistics.[2] Thus, we focus on the conditional pdf $f(\hat{x}|\tilde{x})$. By studying this distribution for various true images $X = \tilde{x}$ of interest, one could examine analytically how MAP estimates vary relative to $\tilde{x}$ as a function of hyperparameters, system models, noise levels, etc.

III. Solution

Our main result is the expression for the pdf $f(\hat{x}|\tilde{x})$ given in (12) below. Our derivation is complicated by our goal of imposing minimal restrictions on the nature of the energy function $R(x)$, so that the result is as widely applicable as possible to the cornucopia of priors that have been proposed. If we were to assume that $R(x)$ is strictly convex and twice continuously differentiable, then the proof of the main result (12) would be fairly straightforward. However, there are many energy functions of interest for edge-preserving image recovery that do not satisfy those regularity conditions [1]–[4], so such a simpler proof would be of less interest. For example, one popular prior uses an energy function formed from Huber *potential functions* (see, e.g., [1] and (23)) defined by

$$\psi_{\mathrm{Huber}}(t) \triangleq \begin{cases} t^2/2, & |t| \le \delta \\ \delta|t| - \delta^2/2, & |t| > \delta \end{cases}$$

for some user-selected parameter δ that controls the degree of edge preservation. This function is convex but not strictly convex (see Fig. 1). The derivative $\dot{\psi}(t)$ of the potential function is called the *influence function*, and plays a key role in its edge-preserving properties and in our pdf formula. As illustrated in Fig. 1, $\dot{\psi}_{\mathrm{Huber}}(t)$ is not differentiable at the two points $t = \pm\delta$, i.e., $\psi_{\mathrm{Huber}}(t)$ is not globally twice differentiable. Similarly, the generalized Gaussian prior [1] has a potential function defined by $\psi(t) = |t|^p/2$ for $p \in (1, 2]$. As illustrated in Fig. 1, this function is not twice differentiable at $t = 0$ for $p < 2$.

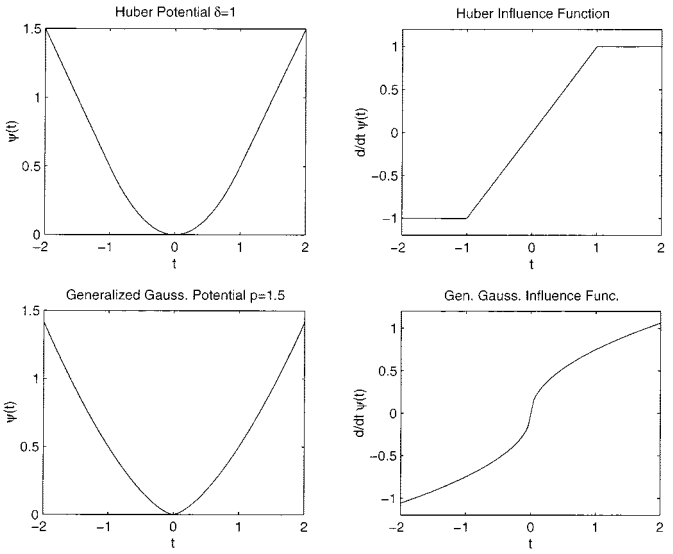

Fig. 1. Huber and Generalized Gaussian potential functions $\psi(t)$, and derivatives $\dot{\psi}(t)$.

Remarkably, despite these "irregularities" in the Huber potential function and generalized Gaussian potential function, the pdf result (12) is indeed applicable to MAP estimators having priors based on these types of potential functions. However, rigorously proving that generality requires a more technical treatment than would be needed if only globally twice continuously differentiable energy functions were of interest.

A. The Basic Idea

Before delving into technicalities, we first present the general idea behind the derivation. Under regularity assumptions A2 and A3 below, the (unconstrained) MAP estimate is a stationary point of the gradient of the objective function Φ, where the column gradient of Φ is given by

$$\nabla'\Phi(x) = -\boldsymbol{A}'\boldsymbol{\Pi}^{-1}(y - \boldsymbol{A}x) + \nabla' R(x).$$

Thus, $\hat{X}$ is related implicitly to the measurement Y via

$$0 = -\boldsymbol{A}'\boldsymbol{\Pi}^{-1}(Y - \boldsymbol{A}\hat{X}) + \nabla' R(\hat{X}). \qquad (4)$$

Rearranging (4) leads to the following *transformation*:

$$Z = h(\hat{X}) \qquad (5)$$

where $Z \in \mathbb{R}^n$, h: $\mathbb{R}^n \to \mathbb{R}^n$ and

$$Z \triangleq \boldsymbol{A}'\boldsymbol{\Pi}^{-1}Y \qquad (6)$$
$$h(x) \triangleq \boldsymbol{A}'\boldsymbol{\Pi}^{-1}\boldsymbol{A}x + \nabla' R(x). \qquad (7)$$

[1] See [11] for an interesting exception.

[2] If the priors used in imaging were truly global priors, then conditional mean estimation should be more appropriate than MAP estimation under a squared error loss function.

It follows from (6) and (1) that, conditioned on $X = \hat{x}$, the random vector Z also has a Gaussian distribution, with the following mean and covariance:

$$\boldsymbol{\mu}_{z|\tilde{x}} \triangleq \boldsymbol{A}'\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{y|\tilde{x}} \tag{8}$$

$$\boldsymbol{\Pi}_{z|\tilde{x}} \triangleq \boldsymbol{A}'\boldsymbol{\Pi}^{-1}\boldsymbol{A} \tag{9}$$

where $\boldsymbol{\mu}_{y|\tilde{x}} \triangleq \boldsymbol{A}\tilde{x}$ is the conditional mean of Y given $X = \tilde{x}$.

Equation (5) describes a functional relationship between the MAP estimate $\hat{X}$ and the random vector Z having a known Gaussian pdf $f(z|\tilde{x})$. The problem of finding $f(\hat{x}|\tilde{x})$ thus becomes a "transformation of random variables" problem. The remainder of this section deals primarily with the technical aspects of showing that the transformation (5) leads to the pdf $f(\hat{x}|\tilde{x})$ in (12).

B. General Transformations

To prove (12) under the general conditions of interest, we need the following theorem, which generalizes the usual such formulas found in engineering probability texts.

Theorem 1: (See [12] and [13] for proofs.)

Let g: $\mathbb{R}^n \rightarrow \mathbb{R}^n$ be one-to-one and assume that $h = g^{-1}$ is continuous. Assume that, on an open set $\mathcal{V} \subseteq \mathbb{R}^n$, h is continuously differentiable with Jacobian[3] $|\nabla h(x)| \triangleq |\det\{(\partial/\partial x_j)h_i(x)\}|$.

Suppose random vector Z has pdf $f(z)$ and

$$P[Z \in h(\mathcal{V}^c)] = \int_{\mathcal{V}^c} f(z)\,dz = 0 \tag{10}$$

where $\mathcal{V}^c$ denotes the set complement (in $\mathbb{R}^n$) of $\mathcal{V}$, and $h(\mathcal{A}) = \{y \in \mathbb{R}^n\colon y = h(x), x \in \mathcal{A}\}$. Then, the pdf of $X = g(Z)$ is given by

$$|\nabla h(x)| f(h(x)), \qquad x \in \mathcal{V} \tag{11}$$

and is zero otherwise.

C. General Case for MAP pdf

Now we apply Theorem 1 to find $f(\hat{x}|\tilde{x})$. So that the problem is well defined and the analysis is tractable, we make the following assumptions.

- A1) $\boldsymbol{A}$ has full column rank.
- A2) The energy function $R(x)$ is chosen such that, for any $y \in \mathbb{R}^m$, the negative log posterior Φ has a unique stationary point that globally minimizes Φ.
- A3) $R(x)$ is continuously differentiable[4] on $\mathbb{R}^n$.
- A4) $R(x)$ is twice continuously differentiable on an open set $\mathcal{V} \subseteq \mathbb{R}^n$.
- A5) The Lebesgue measure of $h(\mathcal{V}^c)$ is zero, where h is defined in (7).

Assumptions A2)–A5) are trivially satisfied by all globally twice continuously differentiable convex energy functions $R(x)$, such as the large family described in [14]. The additional generality afforded by these assumptions will be used in the corollaries following the next theorem to address energy functions that are not globally twice continuously differentiable, such as those illustrated in Fig. 1.

Theorem 2: Under assumptions A1–A5 above, the conditional pdf of the (unconstrained) MAP estimator defined by (3) is

$$\boxed{f(\hat{x}|\tilde{x}) = \frac{|\nabla^2\Phi(\hat{x})|}{\sqrt{2\pi|\boldsymbol{F}|}}\ \exp\left(\frac{-q(\hat{x};\tilde{x})'\boldsymbol{F}^{-1}q(\hat{x};\tilde{x})}{2}\right)} \tag{12}$$

for $\hat{x} \in \mathcal{V}$ and is 0 elsewhere, where

$$\nabla^2\Phi(x) = \boldsymbol{A}'\boldsymbol{\Pi}^{-1}\boldsymbol{A} + \nabla^2 R(x) \tag{13}$$

is the Hessian[5] of the objective function (defined on $\mathcal{V}$),

$$\boldsymbol{F} \triangleq \boldsymbol{A}'\boldsymbol{\Pi}^{-1}\boldsymbol{A} \tag{14}$$

is the Fisher information matrix for estimating x under the model (1), and

$$q(\hat{x};\tilde{x}) \triangleq \boldsymbol{A}'\boldsymbol{\Pi}^{-1}\boldsymbol{A}(\hat{x} - \tilde{x}) + \nabla' R(\hat{x}) \tag{15}$$

where $\nabla' R(x)$ is the column gradient of $R(x)$.

Proof: By A2 and A3, the MAP estimate (3) satisfies the transformation relationship given in (5). To apply (11) to (5) we must verify the conditions of Theorem 1.

Ignoring constants, we can write the objective function (3) in terms of the random vector Z defined in (6)

$$\Phi(x) \equiv \tfrac{1}{2}\, x'\boldsymbol{A}'\boldsymbol{\Pi}^{-1}\boldsymbol{A}x - x'Z + R(x). \tag{16}$$

By A2, for each Z there is a single $\hat{X}$ that minimizes Φ. Thus there is an (implicit) function g: $\mathbb{R}^n \rightarrow \mathbb{R}^n$ for which

$$\hat{X} = g(Z) = g(\boldsymbol{A}'\boldsymbol{\Pi}^{-1}Y). \tag{17}$$

We show that g is one-to-one by contradiction. Suppose there exists $Z_1 \neq Z_2$ such that $g(Z_1) = g(Z_2) = \hat{X}$. Then, by (5) $Z_1 = h(\hat{X}) = Z_2$, contradicting $Z_1 \neq Z_2$. Furthermore, since $g(h(\hat{X})) = \hat{X}$ and $h(g(Z)) = Z$, h is one-to-one with inverse $g = h^{-1}$. Although there is no explicit expression for $g = h^{-1}$ in general, we can nevertheless find $f(\hat{x}|\tilde{x})$ using Theorem 1.

By A3, h is continuous over $\mathbb{R}^n$. By A4, h is continuously differentiable over the open set $\mathcal{V}$. Since the Lebesgue measure of $h(\mathcal{V}^c)$ is zero by A5, and since Z has a Gaussian distribution, $P[Z \in h(\mathcal{V}^c)]$ is zero. Thus, all the conditions of Theorem 1 are satisfied, and we can apply (11) to (5) to conclude

$$f(\hat{x}|\tilde{x}) = |\nabla h(\hat{x})|\ \ f(z|\tilde{x})\Big|_{z=h(\hat{x})} \tag{18}$$

for $\hat{x} \in \mathcal{V}$ and is zero elsewhere. Everything on the right hand side of (18) has an explicit expression, i.e., the dependence of (17) on $g = h^{-1}$ has disappeared in (18). In particular, since from (7)

$$\nabla h(x) = \boldsymbol{A}'\boldsymbol{\Pi}^{-1}\boldsymbol{A} + \nabla^2 R(x) = \nabla^2\Phi(x) \tag{19}$$

[3]We use $|\boldsymbol{F}|$ to denote the absolute value of the determinant of a matrix $\boldsymbol{F}$.

[4]This condition precludes the absolute value potential function (Laplacian prior).

[5]$[\nabla^2\Phi(x)]_{ij} \triangleq (\partial^2/\partial x_i \partial x_j)\Phi(x)$ for $x \in \mathcal{V}$.

for $x \in \mathcal{V}$, we arrive at the pdf expression

$$f(\hat{x}|\tilde{x}) = |\nabla^2\Phi(\hat{x})| \; f(z|\tilde{x})\Big|_{z=h(\hat{x})} = \frac{|\nabla^2\Phi(\hat{x})|}{\sqrt{2\pi|\boldsymbol{\Pi}_{z|\tilde{x}}|}} \exp\left(\frac{-q(\hat{x};\tilde{x}'\boldsymbol{\Pi}_{z|\hat{x}}^{-1}\, q(\hat{x};\tilde{x})}{2}\right) \quad (20)$$

for $\hat{x} \in \mathcal{V}$, where $\boldsymbol{\Pi}_{z|\hat{x}}$ is defined in (9) [cf. (14))], $\nabla^2\Phi(\hat{x})$ is defined in (13), and

$$q(\hat{x};\tilde{x}) \triangleq h(\hat{x}) - \boldsymbol{\mu}_{z|\tilde{x}}.$$

Thus, (12) is the conditional pdf of the MAP estimator. □

As a sanity check, one can consider the case of a Gaussian prior, for which the energy function is quadratic: $R(x) = \frac{1}{2}\, x'\boldsymbol{R}x$. In this case, $\hat{X}$ is linear in Y, so it has a Gaussian distribution with easily computed moments. Substituting into (12) yields the expected Gaussian distribution for $\hat{X}$.

D. Practical Special Cases

In principle, we could leave it to the reader to verify conditions A2)–A5) for his or her own favorite energy functions. However, establishing condition A5 is nontrivial in general for energy functions of interest such as those based on the Huber prior or Generalized Gaussian prior. In this section, we provide corollaries that show that the conditions of Theorem 2 hold under most cases of interest.

We first lay some groundwork that helps cover the case of influence functions such as that of the Huber potential function shown in Fig. 1. The Huber influence function is continuous, but not differentiable. However, $\dot{\psi}_{\text{Huber}}$ does satisfy a *Lipschitz condition*, a property that is stronger than continuity, but weaker than differentiability. Fortunately, this weaker condition is adequate to establish conditions A4)–A5).

Definition 1: A function $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is said to satisfy a uniform Lipschitz condition (of order one) on a compact set $\mathcal{C} \subseteq \mathbb{R}^n$ if there exists a real number $M_{\mathcal{C}} < \infty$ such that

$$\|f(u) - f(v)\| \le M_{\mathcal{C}}\, \|u - v\|, \quad \forall\, u, v \in \mathcal{C}. \quad (21)$$

Define $B(x,r)$ to be the open ball in $\mathbb{R}^n$ of radius r centered at $x \in \mathbb{R}^n$. It follows from the Lipschitz condition (21) on f that

$$f(B(x,r) \cap \mathcal{C}) \subseteq B(f(x), M_{\mathcal{C}}\, r), \quad \forall x \in \mathcal{C}. \quad (22)$$

The following Lemma will help establish A4. (See [15, Prop. 2.2] for a closely related argument for Hausdorff measures.)

Lemma 1: Let $\mathcal{A} \subset \mathcal{C} \subset \mathbb{R}^n$ satisfy $|\mathcal{A}| = 0$, where $|\cdot|$ denotes Lebesgue measure, and $\mathcal{C}$ is a compact set. If $f\colon \mathbb{R}^n \to \mathbb{R}^n$ satisfies a uniform Lipschitz condition on $\mathcal{C}$, then $|f(\mathcal{A})| = 0$. Thus, functions satisfying uniform Lipschitz conditions map zero measure sets into zero measure sets.

Proof: Since $\mathcal{A}$ has measure zero by assumption, for any $\varepsilon > 0$ there exists a countable covering of open balls with centers $\{x_i\}$ and radii $\{r_i\}$ such that $\mathcal{A} \subseteq \cup_i B(x_i, r_i)$ and $\sum_i |B(x_i, r_i)| < \varepsilon / M_{\mathcal{C}}^n$. Thus by (22) and by the subadditivity of Lebesgue measure

$$\begin{aligned} |f(\mathcal{A})| &= |f(\mathcal{A} \cap \mathcal{C})| \le \sum_i |f(B(x_i, r_i) \cap \mathcal{C})| \\ &\le \sum_i |B(f(x_i), M_{\mathcal{C}} r_i)| \\ &= M_{\mathcal{C}}^n \sum_i |B(f(x_i), r_i)| < \varepsilon. \end{aligned}$$

Since ε was arbitrary, this shows that $|f(\mathcal{A})| = 0$.

The Lipschitz condition (21) applies to many energy functions of interest, i.e., $\nabla' R(x)$ is often uniform Lipschitz over $\mathbb{R}^n$. The following Corollary, proven in Appendix A, specializes Theorem 2 to a fairly general form for the energy function that covers many cases of interest.

Corollary 1: Suppose A1) holds and that the energy function has the following form:

$$R(x) = \sum_{k=1}^{K} \psi_k([\boldsymbol{C}x]_k) \quad (23)$$

where

$$[\boldsymbol{C}x]_k \triangleq \sum_{j=1}^{n} c_{kj} x_j$$

and where the potential functions $\{\psi_k\}$ satisfy the following conditions.

- Each ψ_k is convex and differentiable on $\mathbb{R}$ (and hence continuously differentiable).
- Each ψ_k is twice continuously differentiable everywhere on $\mathbb{R}$ *except* possibly for a finite number n_k of points, say $t_{k,1}, \cdots, t_{k,n_k}$, where the second derivative of $\psi_k(t)$ is undefined.
- Each influence function $\dot{\psi}_k$ satisfies a uniform Lipschitz condition on any bounded interval $[-c, c]$ for any $c > 0$.

Then all the conditions of Theorem 2 hold, and (12) is the conditional distribution of the MAP estimator.

The preceding Corollary covers cases such as the Huber potential function, since the Huber influence function is differentiable everywhere except at $t = \pm\delta$, and satisfies a uniform Lipschitz condition over $\mathbb{R}$.

However, the proof of Corollary 1 is inapplicable to the Generalized Gaussian prior. As illustrated in Fig. 1, the Generalized Gaussian influence function is not uniform Lipshitz over any interval of the form $[-c, c]$ since its derivative is unbounded near $t = 0$ for $p < 2$. However, the Generalized Gaussian influence function *is* uniform Lipshitz over intervals of the form $[-c, -\varepsilon]$ and $[\varepsilon, c]$ for $c > \varepsilon > 0$. This turns out to be adequate regularity to establish the conditions of Theorem 2, as shown by the following Corollary, proven in Appendix B.

Corollary 2: Suppose A1 holds and that the energy function has the form (23), where the potential functions $\{\psi_k\}$ satisfy the following conditions.

- Each ψ_k is convex and continuously differentiable on $\mathbb{R}$.
- Each $\psi_k(t)$ is twice continuously differentiable everywhere on $\mathbb{R}$ *except* possibly at $t = 0$.

- Each influence function $\dot{\psi}_k$ satisfies a uniform Lipschitz condition on any bounded interval of the form $[\varepsilon, c]$ and $[-c, -\varepsilon]$ for any $c > \varepsilon > 0$. for any $\varepsilon > 0$.

Then all the conditions of Theorem 2 hold, and (12) is the conditional distribution of the MAP estimator.

We note without proof that one can combine the general approach to the Proofs of Corollary 1 and 2 to formulate sufficient conditions for the ψ_k's that encompass an even wider class of priors. For an energy function that has the form (23), it is sufficient that the potential functions ψ_k be convex and continuously differentiable over $\mathbb{R}$, twice continuously differentiable everywhere except at a finite number of "problem" points, and that the influence functions $\dot{\psi}_k$ be Lipschitz on closed intervals that do not contain those points. We are unaware of any differentiable convex potential functions that fail to satisfy these conditions.

The Proofs of Corollary 1 and 2 do not use the fact that h is one-to-one, and only weakly use the fact that $\mathcal{V}^c$ is a collection of hyperplanes. We conjecture that if these properties were used fully, then one could eliminate the Lipschitz conditions on the ψ_k's and simply assume that each ψ_k is convex and continuously differentiable and twice differentiable almost everywhere.

IV. Non-Gaussian Measurements

The above development relies fairly heavily on the assumption that Y has a Gaussian distribution, since in the Gaussian case we can easily find the pdf of $Z = \boldsymbol{A}'\boldsymbol{\Pi}^{-1}Y$ for use in (18).

When Y is non-Gaussian, (3) defines a penalized weighted least-squares (PWLS) estimator, rather than a MAP estimator. We can determine the pdf $f(\hat{x}|\tilde{x})$ of this PWLS estimator even for non-Gaussian measurement noise if $\boldsymbol{A}$ is invertible (and hence square). When $\boldsymbol{A}$ is invertible, we can rewrite (4) as

$$Y = h_2(\hat{X})$$

where

$$h_2(x) \triangleq \boldsymbol{A}x + \boldsymbol{\Pi}\boldsymbol{A}^{-T}\nabla' R(x)$$

so similar arguments as above lead to

$$f(\hat{x}|\tilde{x}) = |\nabla h_2(\hat{x})| \; f(y|\tilde{x})\Big|_{y=h_2(\hat{x})}. \tag{24}$$

Again all terms on the right hand side have explicit expressions.

V. Simple 2-D Example

The simplest nontrivial example of (12) is when $\boldsymbol{A} = \boldsymbol{I}_2$, $\boldsymbol{\Pi} = \sigma^2\boldsymbol{I}_2$ where $\boldsymbol{I}_2$ is the 2×2 identity matrix, $\tilde{x} = \boldsymbol{\mu}_{y|\tilde{x}} = [0 \; \mu]'$, and where

$$R(x) = \beta\psi(x_1 - x_2)$$

for some symmetric, convex potential function ψ. One can consider this case as representing an image consisting of two neighboring pixels. If $\beta = 0$, then the pixels are estimated independently by $\hat{x}_j = y_j, j = 1, 2$. Including a prior with energy function $R(x)$ yields MAP estimates for the two pixel values that are encouraged to be similar. By choosing $\mu \neq 0$ we can study the case where the two pixels straddle an "edge," since their mean values differ, thereby investigating the edge-preserving properties of the prior.

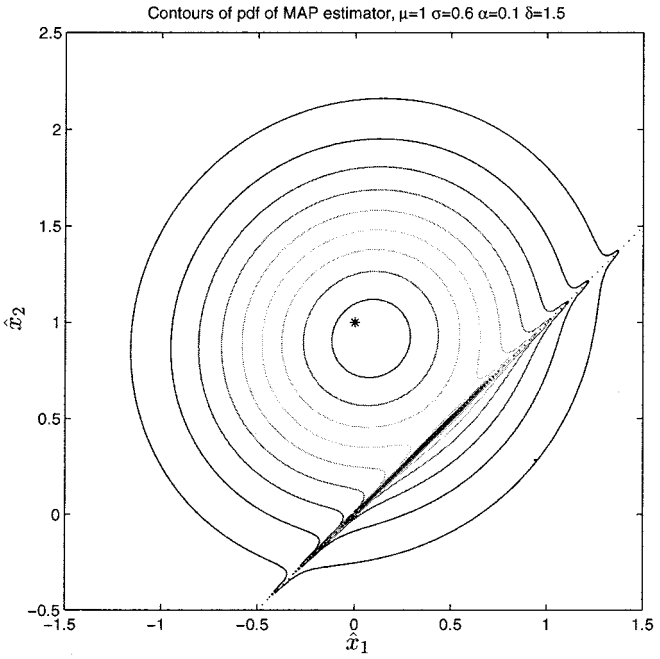

Fig. 2. Contours of conditional pdf (12) of MAP estimator with generalized Gaussian prior. There is one mode near the true parameter $\tilde{x} = [0 \; 1]'$, and a second mode near the average parameter $\hat{x} = [0.5 \; 0.5]'$.

To compute the pdf of the MAP estimator, note that

$$\nabla' R(x) = \beta \begin{bmatrix} 1 \\ -1 \end{bmatrix} \dot{\psi}(x_1 - x_2)$$

$$\nabla^2 R(x) = \beta \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \ddot{\psi}(x_1 - x_2).$$

Thus, from (13), the Hessian of Φ is

$$\nabla^2\Phi(x) = \sigma^{-2}\boldsymbol{I}_2 + \beta \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \ddot{\psi}(x_1 - x_2)$$

$$= \sigma^{-2}\left(\boldsymbol{I}_2 + \alpha \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \ddot{\psi}(x_1 - x_2)\right)$$

where $\alpha \triangleq \beta\sigma^2$, and the determinant of that Hessian is

$$|\nabla^2\Phi(x)| = \sigma^{-4}[1 + 2\alpha\ddot{\psi}(x_1 - x_2)].$$

Also, $\boldsymbol{\mu}_{z|\tilde{x}} = [0 \; \mu/\sigma^2]'$ and $\boldsymbol{\Pi}_{z|\tilde{x}} = \sigma^{-2}\boldsymbol{I}_2$. Substituting into (12) and simplifying

$$f(\hat{x}|\tilde{x}) = \frac{1 + 2\alpha\ddot{\psi}(\hat{x}_1 - \hat{x}_2)}{2\pi\sigma^2} \cdot \exp\left(-\frac{1}{2\sigma^2}\left\|x + \alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix} \ddot{\psi}(\hat{x}_1 - \hat{x}_2) - \begin{bmatrix} 0 \\ \mu \end{bmatrix}\right\|^2\right).$$

As a numerical example, we computed $f(\hat{x}|\tilde{x})$ for the case $\sigma = 0.6$, $\mu = 1$, $\alpha = 0.1$. We used a generalized Gaussian energy function [1]: $\psi(t) = \frac{1}{2}|t|^{1.5}$, so the above pdf is zero on $\mathcal{V}^c = \{\hat{x}\colon \hat{x}_1 = \hat{x}_2\}$. Fig. 2 shows contours of the pdf $f(\hat{x}|\tilde{x})$. There is one mode of the pdf near the true value $\tilde{x} = [0 \; 1]'$, indicated by the asterisk. Estimates that lie near this mode correspond to "preserved edges," since the pixel estimates are closer to the truth than to each other. However, there is another mode near $\hat{x} = [0.5 \; 0.5]'$, because for measurement realizations for which $y_1 \approx y_2$, the energy function of the prior encourages

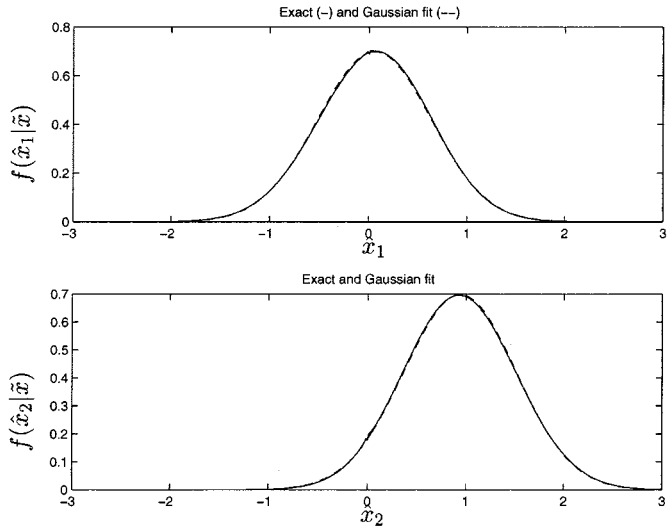

Fig. 3. Marginal conditional pdfs of the components of the MAP estimator, computed numerically from the joint pdf shown in Fig. 2. These marginal pdfs are surprisingly Gaussian, despite the highly non-Gaussian joint pdf.

$\hat{x}_1$ and $\hat{x}_2$ to be approximately equal. In fact, there is a ridge of higher relative probability near the line $\hat{x}_1 = \hat{x}_2$ indicated by the dotted line in Fig. 2. This ridge is induced by the non-Gaussian prior; for a Gaussian prior the contours would be elliptical and centered at the conditional mean.

From the contours of the exact pdf shown in Fig. 2, it would appear that simple characterizations (i.e., Gaussian approximations) of the joint pdf are nontrivial. Whether the covariance matrix of $\hat{X}$ can be approximated analytically via extensions of [5] remains an open question. Fig. 3 shows the marginal distributions of $f(\hat{x}_j|\tilde{x})$ computed by numerical integration of the joint pdf. Despite the complicated structure of the joint pdf, the marginal pdfs are remarkably similar to Gaussian pdfs. Thus it may be possible to find simple approximations for the marginal means and variances.

VI. Conclusion

Our main result is (12), an explicit exact expression for the conditional pdf of MAP estimators for edge-preserving priors. The expression is surprisingly simple given that edge-preserving MAP estimators are defined implicitly and can be highly nonlinear. Limitations of the result include the restriction to Gaussian measurements, and the regularity assumptions for the energy function. Nevertheless, the explicit expression may prove useful in better understanding the properties of edge-preserving image recovery methods.

Appendix A
Proof of Corollary 1

A2) holds by the assumption that the potential functions are convex, so Φ is strictly convex. A3 holds by the assumption that the potential functions are continuously differentiable. For A4), consider $\mathcal{V} \subseteq \mathbb{R}^n$ defined by

$$\mathcal{V}^c \triangleq \bigcup_{k=1}^{K} \bigcup_{i=1}^{n_k} \{x \in \mathbb{R}^n : [\boldsymbol{C}x]_k = t_{k,i}\}.$$

The set $\mathcal{V}^c$ is a finite union of hyperplanes and thus is a closed set. Therefore $\mathcal{V}$ is an open set on which $R(x)$ is twice continuously differentiable, so A4) holds.

To complete the proof, we must establish A5) by showing that $|h(\mathcal{V}^c)| = 0$, where h was defined in (7). Because $\mathcal{V}^c$ is a finite union of hyperplanes, $|\mathcal{V}^c| = 0$, so we can apply Lemma 1. Define $\mathcal{A}_l = \mathcal{V}^c \cap [-l, l]^n$ for $l \in \mathbb{N}$. Since $h(\mathcal{V}^c) = \cup_{l=1}^{\infty} h(\mathcal{A}_l)$, by subadditivity of Lebesgue measure $|h(\mathcal{V}^c)| \leq \sum_{l=1}^{\infty} |h(\mathcal{A}_l)|$, so it suffices to show that $|h(\mathcal{A}_l)| = 0$.

For the energy function $R(x)$ given in (23) above, the function h has the form

$$h(x) = \boldsymbol{F}x + \nabla' R(x) = \boldsymbol{F}x + \sum_{k=1}^{K} c_k \dot{\psi}([\boldsymbol{C}x]_k)$$

where c_k denotes the transpose of the kth row of $\boldsymbol{C}$, i.e. $[\boldsymbol{C}x]_k = c_k' x$. Since each $\dot{\psi}_k$ is uniform Lipschitz over bounded intervals, the function h is uniform Lipschitz over $[-l, l]^n$, since for $x \in [-l, l]^n$, $|[\boldsymbol{C}x]_k| \leq c \triangleq \max_k \sum_j |c_{kj}| l$. Thus by Lemma 1, $|h(\mathcal{A}_l)| = 0$. □

Appendix B
Proof of Corollary 2

Conditions A2)–A4) hold as shown in the Proof of Corollary 1 with the set $\mathcal{V}^c$ in this case defined by

$$\mathcal{V}^c \triangleq \bigcup_{k=1}^{K} \mathcal{H}_k, \quad \text{where } \mathcal{H}_k \triangleq \{x \in \mathbb{R}^n : [\boldsymbol{C}x]_k = 0\}.$$

To show A5, again define $\mathcal{A}_l \triangleq \mathcal{V}^c \cap [-l, l]^n$, so that again it suffices to show that $|h(\mathcal{A}_l)| = 0$, noting that $|\mathcal{A}_l| = 0$.

For any $\boldsymbol{I} \subseteq \{1, \cdots, K\}$, denote $\mathcal{I}^c = \{1, \cdots, K\} - \mathcal{I}$ and define

$$G_{\mathcal{I}}(x) \triangleq \boldsymbol{F}x + \sum_{k \in \mathcal{I}^c} c_k \dot{\psi}_k([\boldsymbol{C}x]_k) + \sum_{k \in \mathcal{I}} c_k \dot{\psi}_k(0)$$

and[6]

$$\mathcal{Z}_{\mathcal{I}} = \bigcap_{k \in \mathcal{I}} \mathcal{H}_k$$

Then for $x \in \mathcal{Z}_{\mathcal{I}}$, clearly $h(x) = G_{\mathcal{I}}(x)$.

The function $G_{\mathcal{I}}$ is uniform Lipschitz on the compact set

$$\mathcal{C}_l(\mathcal{I}, \varepsilon) \triangleq \bigcap_{k \in \mathcal{I}^c} \{x \in \mathbb{R} : |[\boldsymbol{C}x]_k| \geq \varepsilon\} \cap [-l, l]^n.$$

Thus, by Lemma 1, $|G_{\mathcal{I}}(\mathcal{A}_l \cap \mathcal{Z}_{\mathcal{I}} \cap \mathcal{C}_l(\mathcal{I}, \varepsilon))| = 0$ for all $\varepsilon > 0$, since $|\mathcal{A}_l| = 0$. But $\mathcal{A}_l \cap \mathcal{Z}_{\mathcal{I}} \cap \mathcal{C}_l(\mathcal{I}, \varepsilon) \subset \mathcal{Z}_{\mathcal{I}}$, and $h = G_{\mathcal{I}}$ on $\mathcal{Z}_{\mathcal{I}}$, so $|h(\mathcal{A}_l \cap \mathcal{Z}_{\mathcal{I}} \cap \mathcal{C}_l(\mathcal{I}, \varepsilon))| = 0$. Defining

$$\mathcal{C}_l(\mathcal{I}) \triangleq \bigcup_{m=1}^{\infty} \mathcal{C}_l(\mathcal{I}, 1/m)$$

then by subadditivity of Lebesgue measure

$$|h(\mathcal{A}_l \cap \mathcal{Z}_{\mathcal{I}} \cap \mathcal{C}_l(\mathcal{I}))| \leq \sum_{m=1}^{\infty} |h(\mathcal{A}_l \cap \mathcal{Z}_{\mathcal{I}} \cap \mathcal{C}_l(\mathcal{I}, 1/m))|$$

[6]By convention, if $\mathcal{I} = \phi$, then $\bigcap_{k \in \mathcal{I}} \mathcal{H}_k = \mathbb{R}^n$.

which is zero. Finally, since

$$C_l(\mathcal{I}) = \bigcap_{k \in \mathcal{I}^c} \{x \in \mathbb{R}^n : |[\boldsymbol{C}x]_k| > 0\} \cap [-l, l]^n = \bigcap_{k \in \mathcal{I}^c} \mathcal{H}_k^c \cap [-l, l]^n$$

one can easily show that

$$\bigcup_{\mathcal{I} \subseteq \{1,2,\cdots,K\}} \mathcal{Z}_{\mathcal{I}} \cap \mathcal{C}_l(\mathcal{I}) = [-l, l]^n.$$

Thus, again by subadditivity,

$$|h(\mathcal{A}_l)| \leq \sum_{\mathcal{I} \subseteq \{1,2,\cdots,K\}} |h(\mathcal{A}_l \cap \mathcal{Z}_{\mathcal{I}} \cap \mathcal{C}_l(\mathcal{I}))| = 0.$$

Thus, A5 holds under the conditions of the Corollary. □

ACKNOWLEDGMENT

The authors thank C. Abbey for providing a preprint of [8], for the pointer to [10], and for insightful suggestions.

REFERENCES

[1] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Processing*, vol. 2, pp. 296–310, July 1993.

[2] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 6, pp. 298–311, Feb. 1997.

[3] A. H. Delaney and Y. Bresler, "Globally convergent edge-preserving regularized reconstruction: An application to limited-angle tomography," *IEEE Trans. Image Processing*, vol. 7, pp. 204–221, Feb. 1998.

[4] M. Nilkolova, J. Idier, and A. Mohammad-Djafari, "Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF," *IEEE Trans. Image Processing*, vol. 7, pp. 571–585, Apr. 1998.

[5] J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Trans. Image Processing*, vol. 5, pp. 493–506, Mar. 1996.

[6] M. Lavielle, "A stochastic algorithm for parametric and nonparametric estimation in the case of incomplete data," *Signal Process.*, vol. 42, no. 1, pp. 3–17, 1995.

[7] L. Bedini, E. Salerno, and A. Tonazzini, "Edge-preserving tomographic reconstruction from Gaussian data using a Gibbs prior and a generalized expectation-maximization algorithm," *Int. J. Imag., Syst., Technol.*, vol. 5, no. 3, pp. 231–238, 1994.

[8] C. K. Abbey, E. Clarkson, H. H. Barrett, S. P. Mueller, and F. J. Rybicki, "Approximate distributions for maximum likelihood and maximum a posteriori estimates under a Gaussian noise model," in *Information Processing in Medical Imaging*, J. Duncan and G. Gindl, Eds. Berlin, Germany: Springer-Verlag, 1997, pp. 167–175.

[9] C. K. Abbey, E. Clarkson, H. H. Barrett, S. P. Müller, and F. J. Rybicki, "A method for approximating the density of maximum likelihood and maximum a posteriori estimates under a Gaussian noise model," *Med. Imag. Anal.*, vol. 2, no. 4, pp. 395–403, 1998.

[10] O. E. Barndorff-Nielsen, "On a formula for the distribution of the maximum-likelihood estimator," *Biometrika*, vol. 70, no. 2, pp. 343–365, 1983.

[11] V. E. Johnson, "A model for segmentation and analysis of noisy images," *J. Amer. Statist. Assoc.*, vol. 89, pp. 230–241, Mar. 1994.

[12] P. J. Bickel and K. A. Doksum, *Mathematical Statistics*. Oakland, CA: Holden-Day, 1977.

[13] J. A. Fessler, "On transformations of random vectors," Commun. Signal Process. Lab., Dept. Elect. Eng. Comput. Sci., Univ. Michigan, Ann Arbor, MI, http://www.eecs.umich.edu/~fessler, Aug., 1998.

[14] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Trans. Med. Imag.*, vol. 9, pp. 439–446, Dec. 1990.

[15] K. J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*. New York: Wiley, 1990.

[16] K. Lange, "Corrections," *IEEE Trans. Med. Imag.*, vol. 10, p. 288, June 1991.

Jeffrey A. Fessler (S'83–M'90) received the B.S.E.E. degree from Purdue University, West Lafayette, IN, in 1985, and the M.S.E.E., M.S., and Ph.D. degrees from Stanford University, Stanford, CA, in 1986, 1989, and 1990, respectively.

From 1985 to 1988, he was a National Science Foundation Graduate Fellow at Stanford University. Since then, he has been with the University of Michigan, Ann Arbor. From 1991 to 1992, he was a Department of Energy Alexander Hollaender Post-Doctoral Fellow in the Division of Nuclear Medicine. From 1993 to 1995, he was an Assistant Professor in the Nuclear Medicine and the Bioengineering Program. Since 1995, he has been with the Department of Electrical Engineering and Computer Science, where he is now an Associate Professor. His research interests are in statistical aspects of imaging problems.

Hakan Erdogan (S'92–M'95) received the B.S. degree in electrical engineering and mathematics from Middle East Technical University, Ankara, Turkey, in 1993, and the M.S. and Ph.D. degrees in electrical engineering: systems from the University of Michigan, Ann Arbor, in 1995 and 1999, respectively.

He was a Research Assistant with the University of Michigan working in the area of statistical image reconstruction methods for transmission and emission tomography. He joined IBM T.J. Watson Research Labs, Yorktown Heights, NY, in 1999. His research interests are tomographic image reconstruction and speech recognition.

Wei Biao Wu received the M.S. degree from the University of Michigan, Ann Arbor, in 1999, and is currently pursuing the Ph.D. degree in the Department of Statistics.

His research interests are random dynamical systems, bootstrap, Markov chains and martingales, and limit theorems in probability theory.