Parameter Estimation in TV Image Restoration Using Variational Distribution Approximation

S. Derin Babacan, *Student Member, IEEE*, Rafael Molina, *Member, IEEE*, and Aggelos K. Katsaggelos, *Fellow, IEEE*

Abstract—In this paper, we propose novel algorithms for total variation (TV) based image restoration and parameter estimation utilizing variational distribution approximations. Within the hierarchical Bayesian formulation, the reconstructed image and the unknown hyperparameters for the image prior and the noise are simultaneously estimated. The proposed algorithms provide approximations to the posterior distributions of the latent variables using variational methods. We show that some of the current approaches to TV-based image restoration are special cases of our framework. Experimental results show that the proposed approaches provide competitive performance without any assumptions about unknown hyperparameters and clearly outperform existing methods when additional information is included.

Index Terms—Bayesian methods, image restoration, parameter estimation, total variation (TV), variational methods.

I. INTRODUCTION

IN MOST applications, the acquired images represent a degraded version of the original scene. These applications include astronomical imaging (e.g., using ground-based imaging systems or extraterrestrial observations of the earth and the planets), commercial photography, medical imaging (e.g., X-rays, digital angiograms, autoradiographs), and molecular and cellular bioimaging [2]–[4]. The degradation can be due to the atmospheric turbulence, the relative motion between the camera and the scene, and the finite resolution of the acquisition instrument.

A standard formulation of the image degradation model is given in matrix-vector form by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \qquad (1)$$

where the $N \times 1$ vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{n}$ represent, respectively, the original image, the available noisy and blurred image, and the noise with independent elements of variance $\sigma_{\mathbf{n}}^2 = \beta^{-1}$, and $\mathbf{H}$ represents the known blurring matrix. The images are assumed to be of size $m \times n = N$, and they are lexicographically ordered into $N \times 1$ vectors. The restoration problem calls for finding an

S. D. Babacan and A. K. Katsaggelos are with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, 60208-3118 USA (e-mail: sdb@northwestern.edu; aggk@eecs.northwestern.edu).

R. Molina is with the Departamento de Ciencias de la Computación e I.A. Universidad de Granada, 18071 Granada, Spain (e-mail: rms@decsai.ugr.es).

estimate of $\mathbf{x}$ given $\mathbf{y}$, $\mathbf{H}$, and knowledge about $\mathbf{n}$ and possibly $\mathbf{x}$ [2].

A number of approaches have been developed in providing solutions to the restoration problem (see, for example, [2], [3], [5], and references therein). A straightforward approach to the restoration problem is to use least squares estimation and select $\overline{\mathbf{x}}$, an estimate of the original image, as

$$\overline{\mathbf{x}} = \arg\max_{\mathbf{x}} \frac{1}{Z_{\text{noise}}(\beta)} \exp\left[-\frac{1}{2}\beta\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2\right] \qquad (2)$$

where $Z_{\text{noise}}(\beta) = (2\pi/\beta)^{N/2}$. However, as is well known, this approach does not lead to useful restorations in most cases. Use of prior knowledge about the original image can improve the restoration results. Within the Bayesian framework this knowledge is encapsulated as a prior distribution $\mathrm{p}(\mathbf{x})$.

A general model for the prior distribution $\mathrm{p}(\mathbf{x})$ is a Markov random field (MRF) which is characterized by its Gibbs distribution given by

$$\mathrm{p}(\mathbf{x}|\alpha) = \frac{1}{Z(\alpha)} \exp\{-\alpha F(\mathbf{x})\} \qquad (3)$$

where $Z(\alpha)$ is the partition function with a constant α and F is the energy function of the form $F(\mathbf{x}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{x})$, 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.

A critical issue is the choice of the energy function. In this paper we use the total variation (TV) image prior [6] whose energy function is the discrete version of the total variation integral defined as

$$F_{\text{TV}}(\mathbf{x}) = \int |\nabla(\mathbf{x})| \, d\mathbf{x}. \qquad (4)$$

We will explicitly write the form of the prior model in the next section.

If the hyperparameters α and β are known, following the Bayesian paradigm (see [7] for the unification of probabilistic and variational estimation), it is customary to select, as the restoration of $\mathbf{x}$, the image $\mathbf{x}_{(\alpha,\beta)}$ defined by

$$\mathbf{x}_{(\alpha,\beta)} = \arg\max_{\mathbf{x}} \mathrm{p}(\mathbf{x}|\alpha)\mathrm{p}(\mathbf{y}|\mathbf{x},\beta)$$
$$= \arg\min_{\mathbf{x}} \left[\alpha F_{\text{TV}}(\mathbf{x}) + \frac{\beta}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2\right]. \qquad (5)$$

Not much work has been reported in the literature on the joint parameter and image estimation when the parameters α and β are not known (see [5], [8] for recent developments in variational modeling and inference). Rudin *et al.* [6] consider the minimization of $F_{\text{TV}}(\mathbf{x})$ constrained by $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 = N\hat{\sigma}_{\mathbf{n}}^2$, where $\hat{\sigma}_{\mathbf{n}}^2$ represents an estimate of the noise variance,

and then proceed to estimate both the image and the associated Lagrange multiplier to this constrained optimization problem. Bertalmio *et al.* [9] make the Lagrange multiplier region dependent. Bioucas-Dias *et al.* [10], using their majorization-minimization approach [11], propose a Bayesian method to estimate the original image and α assuming that an estimate of the noise variance is available. To our knowledge no work has been reported on the simultaneous estimation of the parameters α and β and the image and also on the estimation of the uncertainty of those estimates (only point estimates of the parameters and image have been provided).

In this paper, we use the Bayesian paradigm to jointly estimate the image and unknown hyperparameters (α and β) in image restoration when the TV image prior is used. The estimation procedure will not provide only point estimates of the image and the hyperparameters but also the probability distributions that approximate the posterior distribution of the hyperparameters and the original image given the observation.

This paper is organized as follows. Section II presents a general description of the Bayesian modeling and inference of the TV restoration problem, which includes a brief discussion on estimation procedures (inference methods) that provide point or probability distribution estimates. The actual parameter hyperpriors, image prior, and observation models used in this paper are then presented in Section III. Section IV describes the variational approach to distribution approximation for TV image restoration and how inference is performed. We propose different approximations of the posterior distribution of the image and the unknown hyperparameters, and compare them to other approaches reported in the literature. Finally, in Section V, experimental results and comparisons with other methods are shown, and Section VI concludes the paper.

II. Bayesian Modeling and Inference

The Bayesian modeling of the TV restoration problem requires first the definition of a joint distribution $p(\alpha, \beta, \mathbf{x}, \mathbf{y})$ of the observation, $\mathbf{y}$, the unknown image, $\mathbf{x}$, and the hyperparameters α and β. To model the joint distribution, we utilize in this paper the hierarchical Bayesian paradigm (see, for example, [12]–[15]). In the hierarchical approach to image restoration, we have at least two stages. In the first stage, knowledge about the structural form of the observation noise and the structural behavior of the image is used in forming $p(\mathbf{y}|\mathbf{x}, \beta)$ and $p(\mathbf{x}|\alpha)$, respectively. These noise and image models depend on the unknown hyperparameters α and β. In the second stage, a hyperprior on the hyperparameters is defined, thus allowing for the incorporation of information about these hyperparameters into the process.

For $\alpha, \beta, \mathbf{x}, \mathbf{y}$, the following joint distribution is defined:

$$p(\alpha, \beta, \mathbf{x}, \mathbf{y}) = p(\alpha)p(\beta)p(\mathbf{x}|\alpha)p(\mathbf{y}|\mathbf{x}, \beta) \qquad (6)$$

and inference is based on $p(\alpha, \beta, \mathbf{x}|\mathbf{y})$.

Three crucial questions have to be addressed when modeling and performing inference for image restoration problems using the hierarchical Bayesian paradigm. The first one relates to the definition of $p(\alpha)$ and $p(\beta)$. We should be able to deal with the case of known hyperparameters which correspond to degenerate

distributions for $p(\alpha)$ and $p(\beta)$, but also with more realistic situations including the cases when some knowledge about these parameters is available or when only the observation $\mathbf{y}$ is available to estimate them.

The second crucial problem to be considered is to decide how inference will be carried out. A commonly used approach in image restoration (called the Evidence analysis [12]) consists of estimating the hyperparameters α, β by using

$$(\hat{\alpha}, \hat{\beta}) = \arg\max_{\alpha, \beta} p(\alpha, \beta|\mathbf{y})$$
$$= \arg\max_{\alpha, \beta} \int p(\alpha, \beta, \mathbf{x}|\mathbf{y})d\mathbf{x}$$
$$= \arg\max_{\alpha, \beta} \int p(\alpha, \beta, \mathbf{x}, \mathbf{y})d\mathbf{x} \qquad (7)$$

and then estimating the image by solving

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{x}|\hat{\alpha}, \hat{\beta}, \mathbf{y}). \qquad (8)$$

Another approach, also commonly used in image restoration, is the so called empirical analysis [16], which consists of calculating the restoration by solving

$$\overline{\mathbf{x}} = \arg\max_{\mathbf{x}} \int \int p(\alpha, \beta, \mathbf{x}, \mathbf{y})d\alpha d\beta. \qquad (9)$$

These inference procedures aim at optimizing a given function and not at obtaining posterior distributions that can be analyzed or simulated to obtain additional information about the quality of the estimates. Instead of having a distribution over all possible values of the parameters and the image, the above inference procedures choose a specific set of values. This means that we have neglected many other interpretations of the data. If the posterior is sharply peaked, other values of the hyperparameters and the image will have a much lower posterior probability but, if the posterior is broad, choosing a unique value will neglect many other choices of them with similar posterior probabilities.

The third crucial problem to be solved when using the Bayesian paradigm on TV image restoration is to decide how to calculate $p(\alpha, \beta, \mathbf{x}|\mathbf{y})$, which is in general a challenging task. An approach is provided by the variational distribution approximation. This approximation can be thought of as being between the Laplace approximation (see, for instance, [14] and [17]) and sampling methods [18]. The basic underlying idea, as will be explained later, is to approximate $p(\alpha, \beta, \mathbf{x}|\mathbf{y})$ with a simpler distribution. See the very interesting [19], [20] books [21], [22] and book chapter [23] for a comprehensive introduction to variational methods.

The last few years have seen a growing interest in the application of variational methods [19], [23] to inference problems. These methods attempt to approximate posterior distributions with the use of the Kullback–Leibler cross-entropy [24]. Application of variational methods to Bayesian inference problems include graphical models and neural networks [23], independent component analysis [19], mixtures of factor analyzers, linear dynamic systems, hidden Markov models [20], support vector machines [25] and blind deconvolution problems (see [15], [26], and [27]).

In this paper, we use a TV prior distribution for the image, and gamma distributions for the unknown parameter (hyperparameter) of the prior and the image formation noise. We apply variational methods to approximate the posterior probability of the unknown image and hyperparameters and propose two different approximations of the posterior distribution. We use the obtained posterior approximation to gain additional insight into the estimated hyperparameters and image.

III. HYPERPRIORS, PRIOR, AND OBSERVATION MODEL USED IN TV IMAGE DECONVOLUTION

We first describe the TV prior model as well as the observation model we use in the first stage of the hierarchical Bayesian paradigm. Then, since the prior and observation models depend on unknown hyperparameters, we proceed to explain the hyperprior distributions we utilize for these hyperparameters.

A. First Stage: Prior Models on Images

As image model we use the TV prior, given by

$$\mathrm{p}(\mathbf{x}|\alpha) \propto \frac{1}{Z_{\mathrm{TV}}(\alpha)} \exp\left[-\alpha \mathrm{TV}(\mathbf{x})\right] \tag{10}$$

where $Z_{\mathrm{TV}}(\alpha)$ is the partition function and

$$\mathrm{TV}(\mathbf{x}) = \sum_i \sqrt{\left(\Delta_i^h(\mathbf{x})\right)^2 + \left(\Delta_i^v(\mathbf{x})\right)^2} \tag{11}$$

where the operators $\Delta_i^h(\mathbf{x})$ and $\Delta_i^v(\mathbf{x})$ correspond to, respectively, the horizontal and vertical first order differences, at pixel i, that is, $\Delta_i^h(\mathbf{x}) = x_i - x_{l(i)}$ and $\Delta_i^v(\mathbf{x}) = x_i - x_{a(i)}$, with $l(i)$ and $a(i)$ denoting the nearest neighbors of i, to the left and above, respectively.

Unless we want to use very simple estimation procedures for the hyperparameter α, we need to calculate (approximate) the partition function $Z_{\mathrm{TV}}(\alpha)$. Using

$$\int\int \exp[-\alpha\sqrt{u^2+v^2}]dudv = 2\pi/\alpha^2 \tag{12}$$

we can utilize the following approximation of $\mathrm{p}(\mathbf{x}|\alpha)$ in (10) proposed in [11]:

$$\mathrm{p}(\mathbf{x}|\alpha) = c\alpha^{N/2} \exp\left[-\alpha \mathrm{TV}(\mathbf{x})\right] \tag{13}$$

where again N is the size of the original image $\mathbf{x}$, and c is a constant. Note that the idea of approximating partition functions in image priors to be able to estimate distribution parameters has also been used in [27].

The probability distribution corresponding to the observation model in (1) is given by

$$\mathrm{p}(\mathbf{y}|\mathbf{x},\beta) \propto \beta^{N/2} \exp\left[-\frac{\beta}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2\right]. \tag{14}$$

B. Second Stage: Hyperpriors on the Hyperparameters

A large part of the Bayesian literature is devoted to finding hyperprior distributions $\mathrm{p}(\alpha,\beta)$ for which $\mathrm{p}(\alpha,\beta,\mathbf{x}|\mathbf{y})$ can be either calculated in a straightforward way or be closely approximated. These are the so called conjugate priors [28] which have
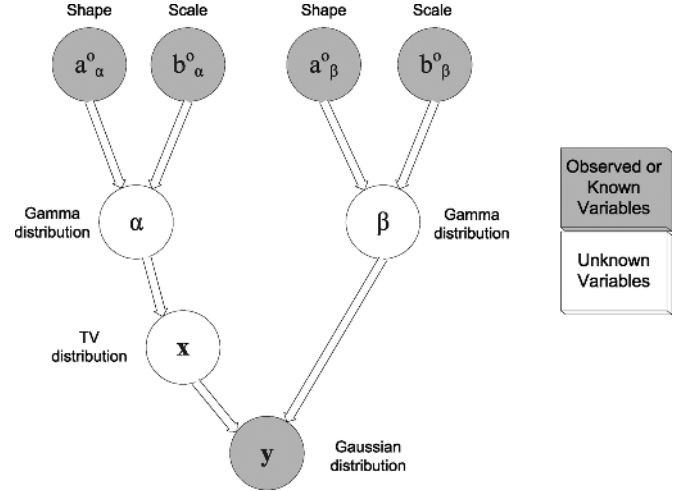

Fig. 1. Graphical model showing relationships between variables.

the intuitive feature of allowing one to begin with a certain functional form for the prior and end up with a posterior of the same functional form, but with the parameters updated by the sample information.

We will assume that each of the hyperparameters $\omega \in \{\alpha, \beta\}$ has as hyperprior the gamma distribution, $\Gamma(\omega|a_\omega^o, b_\omega^o)$, defined by

$$\mathrm{p}(\omega) = \Gamma\left(\omega|a_\omega^o, b_\omega^o\right) = \frac{(b_\omega^o)^{a_\omega^o}}{\Gamma\left(a_\omega^o\right)} \omega^{a_\omega^o - 1} \exp\left[-\omega b_\omega^o\right] \tag{15}$$

where $b_\omega^o > 0$ and $a_\omega^o > 0$ are, respectively, the scale and shape parameters, which are assumed to be known. We will discuss their calculation in Section V. The gamma distribution has the following mean, variance, and mode

$$E[\omega] = \frac{a_\omega^o}{b_\omega^o}, \quad Var[\omega] = \frac{a_\omega^o}{(b_\omega^o)^2}, \quad \mathrm{Mode}[\omega] = \frac{a_\omega^o - 1}{b_\omega^o}. \tag{16}$$

There are several important reasons for selecting Gamma distributions for the hyperpriors. First, the Gamma distribution is conjugate for the variance of the Gaussian, and, therefore, the posteriors will also have Gamma distributions in the Bayesian formulation. Second, as will be shown later, their update equations will exhibit interesting similarities to some previously derived results in the literature.

Finally, combining the first and second stages of the problem modeling we have the following global distribution:

$$\mathrm{p}(\alpha,\beta,\mathbf{x},\mathbf{y}) = \mathrm{p}(\alpha)\mathrm{p}(\beta)\mathrm{p}(\mathbf{x}|\alpha)\mathrm{p}(\mathbf{y}|\mathbf{x},\beta) \tag{17}$$

where $\mathrm{p}(\alpha)$, $\mathrm{p}(\beta)$, $\mathrm{p}(\mathbf{x}|\alpha)$, and $\mathrm{p}(\mathbf{y}|\mathbf{x},\beta)$ have been defined in (13)–(15). The joint probability model is shown in graphical form in Fig. 1 using a directed acyclic graph.

IV. BAYESIAN INFERENCE AND VARIATIONAL APPROXIMATION OF THE POSTERIOR DISTRIBUTION FOR TV IMAGE RESTORATION

The Bayesian paradigm dictates that inference on $(\alpha, \beta, \mathbf{x})$ should be based on

$$\mathrm{p}(\alpha,\beta,\mathbf{x}|\mathbf{y}) = \frac{\mathrm{p}(\alpha,\beta,\mathbf{x},\mathbf{y})}{\mathrm{p}(\mathbf{y})} \tag{18}$$

where $\mathrm{p}(\alpha,\beta,\mathbf{x},\mathbf{y})$ is given by (17).

Because $p(\alpha, \beta, \mathbf{x}|\mathbf{y})$ cannot be found in closed form, since

$$p(\mathbf{y}) = \int \int \int p(\alpha, \beta, \mathbf{x}, \mathbf{y}) d\mathbf{x} d\beta d\alpha \qquad (19)$$

cannot be calculated analytically, we apply variational methods to approximate this distribution by the distribution $q(\alpha, \beta, \mathbf{x})$. We utilize a mean field approximation for the posterior distributions of α, β, and $\mathbf{x}$ so that these posterior distributions are assumed to be independent given the observations. We will later show that particular selections of the distributions $q(\alpha, \beta)$ and $q(\mathbf{x})$ lead to the hyperparameters and image point estimates provided by the evidence and empirical analysis described in Section II. Notice, however, that unless the distributions $q(\alpha, \beta)$ and $q(\mathbf{x})$ are degenerate, the variational approximation provides us with additional information that goes beyond simple point estimates.

The variational criterion used to find $q(\alpha, \beta, \mathbf{x})$ is the minimization of the Kullback–Leibler divergence, given by

$$
\begin{aligned}
& C_{KL}\left(q(\alpha, \beta, \mathbf{x}) \| p(\alpha, \beta, \mathbf{x}|\mathbf{y})\right) \\
& = \int \int \int q(\alpha, \beta, \mathbf{x}) \log\left(\frac{q(\alpha, \beta, \mathbf{x})}{p(\alpha, \beta, \mathbf{x}|\mathbf{y})}\right) d\alpha d\beta d\mathbf{x} \\
& = \int \int \int q(\alpha, \beta, \mathbf{x}) \log\left(\frac{q(\alpha, \beta, \mathbf{x})}{p(\alpha, \beta, \mathbf{x}, \mathbf{y})}\right) d\alpha d\beta d\mathbf{x} + \text{const} \\
& = \mathcal{M}\left(q(\mathbf{x}, \alpha, \beta)\right) + \text{const}
\end{aligned} \qquad (20)
$$

which is always non-negative and equal to zero only when $q(\alpha, \beta, \mathbf{x}) = p(\alpha, \beta, \mathbf{x}|\mathbf{y})$.

Due to the form of the TV prior, the above integral is difficult to evaluate (note that also for the same reason the evidence and empirical estimates described in Section II are difficult to calculate). We can, however, majorize the TV prior by a function which renders the integral easier to calculate. Let us consider the following inequality, also used in [11], which states that, for any $w \geq 0$ and $z > 0$

$$\sqrt{wz} \leq \frac{w+z}{2} \Rightarrow \sqrt{w} \leq \frac{w+z}{2\sqrt{z}}. \qquad (21)$$

Let us also define for α, $\mathbf{x}$, and any N-dimensional vector $\mathbf{u} \in (R^+)^N$, with components u_i, $i = 1, \ldots, N$, the following functional:

$$
\begin{aligned}
M(\alpha, \mathbf{x}, \mathbf{u}) = {} & \alpha^{N/2} \\
& \times \exp\left[-\frac{\alpha}{2} \sum_i \frac{\left(\Delta_i^h(\mathbf{x})\right)^2 + \left(\Delta_i^v(\mathbf{x})\right)^2 + u_i}{\sqrt{u_i}} \right].
\end{aligned} \qquad (22)
$$

Now, using inequality (21) with $w = (\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2$ and $z = u_i$ and comparing (22) with (13), we obtain

$$p(\mathbf{x}|\alpha) \geq c \cdot M(\alpha, \mathbf{x}, \mathbf{u}). \qquad (23)$$

As will be shown later, vector $\mathbf{u}$ is a quantity that needs to be computed and has an intuitive interpretation related to the unknown image $\mathbf{x}$. Inequality (23) leads to the following lower bound for the joint probability distribution:

$$
\begin{aligned}
p(\alpha, \beta, \mathbf{x}, \mathbf{y}) \geq {} & c \cdot p(\alpha) p(\beta) M(\alpha, \mathbf{x}, \mathbf{u}) p(\mathbf{y}|\mathbf{x}, \beta) \\
& = F(\alpha, \beta, \mathbf{x}, \mathbf{u}, \mathbf{y}).
\end{aligned} \qquad (24)
$$

By defining

$$
\begin{aligned}
\tilde{\mathcal{M}}\left(q(\mathbf{x}, \alpha, \beta), \mathbf{u}\right) = {} & \int \int \int q(\alpha, \beta, \mathbf{x}) \\
& \times \log\left(\frac{q(\alpha, \beta, \mathbf{x})}{F(\alpha, \beta, \mathbf{x}, \mathbf{u}, \mathbf{y})}\right) d\alpha d\beta d\mathbf{x}
\end{aligned} \qquad (25)
$$

and utilizing inequality (24), we obtain

$$\mathcal{M}\left(q(\mathbf{x}, \alpha, \beta)\right) \leq \min_{\mathbf{u}} \tilde{\mathcal{M}}\left(q(\mathbf{x}, \alpha, \beta), \mathbf{u}\right). \qquad (26)$$

Therefore, by finding a sequence of distributions $\{q^k(\alpha, \beta, \mathbf{x})\}$ that monotonically decreases $\tilde{\mathcal{M}}(q(\mathbf{x}, \alpha, \beta), \mathbf{u})$ for a fixed $\mathbf{u}$ a sequence of an ever decreasing upper bound of $C_{KL}(q^k(\alpha, \beta, \mathbf{x}) \| p(\alpha, \beta, \mathbf{x}|\mathbf{y}))$ is also obtained due to (20). However, also minimizing $\tilde{\mathcal{M}}(q(\mathbf{x}, \alpha, \beta), \mathbf{u})$ with respect to $\mathbf{u}$ generates a sequence of vectors $\{\mathbf{u}^k\}$ that tightens the upper-bound for each distribution $q^k(\alpha, \beta, \mathbf{x})$. Therefore, the two sequences $\{q^k(\alpha, \beta, \mathbf{x})\}$ and $\{\mathbf{u}^k\}$ are coupled. We develop an iterative algorithm (Algorithm 1) to find such sequences.

Inequality (21) provides a local quadratic approximation to the TV prior. Had a fixed $\mathbf{u}^o$ with same elements been used a global conditional auto-regression model approximating the TV prior would have been obtained. Clearly, the procedure which updates $\mathbf{u}$ will provide a tighter upper bound for $\mathcal{M}(q(\mathbf{x}, \alpha, \beta))$, since we are using $\min_{\mathbf{u}} \tilde{\mathcal{M}}(q(\mathbf{x}, \alpha, \beta), \mathbf{u})$ instead of $\tilde{\mathcal{M}}(q(\mathbf{x}, \alpha, \beta), \mathbf{u}^o)$.

Finally, we note that the process to find the best posterior distribution approximation of the image in combination with $\mathbf{u}$ is a very natural extension of the majorization-minimization approach to function optimization (see [29]) and that local majorization has also been applied to variational logistic regression [30], as well as, to the inference of its parameters (see [31] and [32]).

The following algorithm can, therefore, be used for calculating the approximating posteriors $q(\alpha, \beta, \mathbf{x}) = q(\mathbf{x}) q(\alpha, \beta)$.

Algorithm 1

Posterior parameter and image distributions estimation in TV restoration using $q(\alpha, \beta, \mathbf{x}) = q(\alpha, \boldsymbol{\beta}) q(\mathbf{x})$.

Given $\mathbf{u}^1 \in (R^+)^N$ and $q^1(\alpha, \beta)$, an initial estimate of the distribution $q(\alpha, \beta)$, for $k = 1, 2, \ldots$ until a stopping criterion is met.

1) Find

$$
\begin{aligned}
q^k(\mathbf{x}) = {} & \arg\min_{q(\mathbf{x})} \int \int \int q^k(\alpha, \beta) q(\mathbf{x}) \\
& \times \log\left(\frac{q^k(\alpha, \beta) q(\mathbf{x})}{F(\alpha, \beta, \mathbf{x}, \mathbf{u}^k, \mathbf{y})}\right) d\alpha d\beta d\mathbf{x}.
\end{aligned} \qquad (27)
$$

2) Find

$$
\begin{aligned}
\mathbf{u}^{k+1} = {} & \arg\min_{\mathbf{u}} \int \int \int q^k(\alpha, \beta) q^k(\mathbf{x}) \\
& \times \log\left(\frac{q^k(\alpha, \beta) q^k(\mathbf{x})}{F(\alpha, \beta, \mathbf{x}, \mathbf{u}, \mathbf{y})}\right) d\alpha d\beta d\mathbf{x}.
\end{aligned} \qquad (28)
$$

3) Find

$$
q^{k+1}(\alpha,\beta) = \underset{q(\alpha,\beta)}{\arg\min} \int\int\int q(\alpha,\beta)q^k(\mathbf{x})
$$

$$
\times \log\left(\frac{q(\alpha,\beta)q^k(\mathbf{x})}{F(\alpha,\beta,\mathbf{x},\mathbf{u}^{k+1},\mathbf{y})}\right) d\alpha d\beta d\mathbf{x}. \quad (29)
$$

Set

$$
q(\alpha,\beta) = \lim_{k\to\infty} q^k(\alpha,\beta), \quad q(\mathbf{x}) = \lim_{k\to\infty} q^k(\mathbf{x}). \quad (30)
$$

Let us now further develop each of the steps of the above algorithm. To calculate $q^k(\mathbf{x})$, we observe that differentiating the integral on the right-hand side of (27) with respect to $q(\mathbf{x})$ and setting it equal to zero, we obtain

$$
q^k(\mathbf{x}) \propto \exp\left\{ E_{q^k(\alpha,\beta)}\left[\ln F(\alpha,\beta,\mathbf{x},\mathbf{u}^k)\right]\right\} \quad (31)
$$

which represents an N-dimensional Gaussian distribution with parameters

$$
\mathrm{cov}_{q^k(\mathbf{x})}[\mathbf{x}] = \left[E_{q^k(\beta)}[\beta]\mathbf{H}^t\mathbf{H} + E_{q^k(\alpha)}[\alpha](\Delta^h)^t W(\mathbf{u}^k)(\Delta^h) \right.
$$
$$
\left. + E_{q^k(\alpha)}[\alpha](\Delta^v)^t W(\mathbf{u}^k)(\Delta^v) \right]^{-1} \quad (32)
$$

and

$$
E_{q^k(\mathbf{x})}[\mathbf{x}] = \mathrm{cov}_{q^k(\mathbf{x})}[\mathbf{x}] E_{q^k(\beta)}[\beta]\mathbf{H}^t\mathbf{y} \quad (33)
$$

where $W(\mathbf{u}^k)$ is an $N \times N$ diagonal matrix of the form

$$
W(\mathbf{u}^k) = diag\left(\frac{1}{\sqrt{u_i^k}}\right), \quad i = 1,\ldots,N \quad (34)
$$

and Δ^h and Δ^v represent the $N \times N$ convolution matrices associated to the first order horizontal and vertical differences, respectively.

To calculate $\mathbf{u}^{k+1}$, we have from (28) that

$$
\mathbf{u}^{k+1} = \underset{\mathbf{u}}{\arg\min} \sum_i \frac{E_{q^k(\mathbf{x})}\left[\left(\Delta_i^h(\mathbf{x})\right)^2 + \left(\Delta_i^v(\mathbf{x})\right)^2\right] + u_i}{\sqrt{u_i}}
$$
$$
(35)
$$

and, consequently

$$
u_i^{k+1} = E_{q^k(\mathbf{x})}\left[\left(\Delta_i^h(\mathbf{x})\right)^2 + \left(\Delta_i^v(\mathbf{x})\right)^2\right], \quad i = 1,\ldots,N. \quad (36)
$$

Notice that $q^k(\alpha,\beta)$ is not required in calculating $\mathbf{u}^{k+1}$. It is clear from (36) that the vector $\mathbf{u}^{k+1}$ is a function of the spatial first order differences of the unknown image $\mathbf{x}$ under the distribution $q^k(\mathbf{x})$ and represents the local spatial activity of $\mathbf{x}$. Therefore, matrix $W(\mathbf{u}^k)$ in (34) can be interpreted as the *spatial adaptivity* matrix, since it controls the amount of smoothing at each pixel location depending on the strength of the intensity variation at that pixel, as expressed by the horizontal and vertical intensity gradients. That is, for the pixels with high spatial activity the corresponding entries of $W(\mathbf{u}^k)$ are very small or zero, which means that no smoothness is enforced, while for the pixels in a flat region the corresponding entries of $W(\mathbf{u}^k)$ are very large, which means that smoothness is enforced. This

matrix $W(\mathbf{u}^k)$ has also been referred to as the visibility matrix [33] since it describes the masking property of the human visual system, according to which noise is not visible in high spatial activity regions (its high frequencies are masked by the edges), while it is visible in the low spatial frequency (flat) regions. The visibility matrix and its complementary matrix $\mathbf{I} - W(\mathbf{u}^k)$ have been used in iterative image restoration in [34].

By differentiating the integral on the right hand side of (29) with respect to $q(\alpha,\beta)$ and setting it equal to zero, we obtain

$$
q^{k+1}(\alpha,\beta) \propto \exp\left\{ E_{q^k(\mathbf{x})}\left[\ln F(\alpha,\beta,\mathbf{x},\mathbf{u}^{k+1})\right]\right\} \quad (37)
$$

and thus

$$
q^{k+1}(\alpha,\beta) = q^{k+1}(\alpha)q^{k+1}(\beta) \quad (38)
$$

where $q^{k+1}(\alpha)$ and $q^{k+1}(\beta)$ are gamma distributions given respectively by

$$
q^{k+1}(\alpha) \propto \alpha^{N/2+a_\alpha^o-1}
$$
$$
\times \exp\left[-\alpha\left(\sum_i \sqrt{u_i^{k+1}} + b_\alpha^o\right)\right] \quad (39)
$$

$$
q^{k+1}(\beta) \propto \beta^{N/2+a_\beta^o-1}
$$
$$
\times \exp\left[-\beta\left(\frac{E_{q^k(\mathbf{x})}\|\mathbf{y}-\mathbf{Hx}\|^2}{2} + b_\beta^o\right)\right]. \quad (40)
$$

The means of these gamma distributions are given by

$$
E_{q^{k+1}(\alpha)}[\alpha] = \frac{N/2 + a_\alpha^o}{\sum_i \sqrt{u_i^{k+1}} + b_\alpha^o}
$$
$$
E_{q^{k+1}(\beta)}[\beta] = \frac{N/2 + a_\beta^o}{E_{q^k(\mathbf{x})}\|\mathbf{y}-\mathbf{Hx}\|^2/2 + b_\beta^o}. \quad (41)
$$

The calculation of $E_{q^k(\mathbf{x})}[\mathbf{x}]$, $\mathrm{cov}_{q^k(\mathbf{x})}[\mathbf{x}]$, $E_{q^k(\mathbf{x})}[(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2]$, and $E_{q^k(\mathbf{x})}[\|\mathbf{y}-\mathbf{Hx}\|^2]$ is carried out in Appendices I and II.

Note that we have

$$
\left(E_{q^{k+1}(\alpha)}[\alpha]\right)^{-1} = \frac{a_\alpha^o}{a_\alpha^o + \frac{N}{2}}\frac{b_\alpha^o}{a_\alpha^o}
$$
$$
+ \frac{N/2}{\left(a_\alpha^o + \frac{N}{2}\right)}\frac{2\sum_i \sqrt{u_i^{k+1}}}{N} \quad (42)
$$

$$
\left(E_{q^{k+1}(\beta)}[\beta]\right)^{-1} = \frac{a_\beta^o}{\left(a_\beta^o + \frac{N}{2}\right)}\frac{b_\beta^o}{a_\beta^o} + \frac{N/2}{\left(a_\beta^o + \frac{N}{2}\right)}
$$
$$
\times \frac{E_{q^k(\mathbf{x})}\left[\|\mathbf{y}-H\mathbf{x}\|^2\right]}{N} \quad (43)
$$

and thus

$$
\left(E_{q^{k+1}(\alpha)}[\alpha]\right)^{-1} = \gamma_\alpha\frac{1}{\bar{\alpha}^o} + (1-\gamma_\alpha)\frac{\sum_i \sqrt{u_i^{k+1}}}{N/2} \quad (44)
$$

$$
\left(E_{q^{k+1}(\beta)}[\beta]\right)^{-1} = \gamma_\beta\frac{1}{\bar{\beta}^o} + (1-\gamma_\beta)
$$
$$
\times \frac{E_{q^k(\mathbf{x})}\left[\|\mathbf{y}-H\mathbf{x}\|^2\right]}{N} \quad (45)
$$

where $\overline{\alpha}^o = a^o_\alpha / b^o_\alpha$, $\overline{\beta}^o = a^o_\beta / b^o_\beta$, and

$$\gamma_\alpha = \frac{a^o_\alpha}{a^o_\alpha + \frac{N}{2}}, \quad \gamma_\beta = \frac{a^o_\beta}{a^o_\beta + \frac{N}{2}}. \tag{46}$$

Equation (46) indicates that γ_α and γ_β, both taking values in the interval $[0,1]$, can be understood as normalized confidence parameters. As can be seen from (44) and (45), the inverses of the means of the hyperpriors are represented as convex combinations of their initial values and their maximum likelihood (ML) estimates. These ML estimates have been derived before either empirically or by using regularization formulations [34], [35]. According to (44) and (45), when they are equal to zero, no confidence is placed on the initial values of the hyperparameters and ML estimates are used, while when they are asymptotically equal to one, the prior knowledge of the mean is fully enforced, i.e., no estimation of the hyperparameters is performed.

Case of particular interest is when

$$a^o_\alpha = a^o_\beta = 1 \text{ and } b^o_\alpha = b^o_\beta = \infty \tag{47}$$

which corresponds to a flat hyperprior distribution. This type of hyperprior modeling makes the observation responsible for the whole estimation process.

In the proposed model, for estimating the posterior distribution of the image and the unknown hyperparameters no assumptions were made about $q(\mathbf{x})$ and $q(\alpha, \beta)$. We study now the case when $q(\mathbf{x})$ is a degenerate distribution, that is, a distribution which takes one value with probability one and the rest with probability zero. In the iterative procedure we describe next, we use $\mathbf{x}^k$ to denote the value $q^k(\mathbf{x})$ takes with probability one. We then have the following procedure.

Algorithm 2

Posterior parameter and image distributions estimation in TV restoration using $q(\alpha, \beta, \mathbf{x}) = q(\alpha, \beta)q(\mathbf{x})$ with $q(\mathbf{x})$ a degenerate distribution.

Given $q^1(\alpha, \beta)$, an initial estimate of the distribution $q(\alpha, \beta)$ and $\mathbf{u}^1 \in (R^+)^N$, for $k = 1, 2, \dots$ until a stopping criterion is met.

1) Calculate

$$\mathbf{x}^k = \left(E_{q^k(\beta)}[\beta] \mathbf{H}^t \mathbf{H} + E_{q^k(\alpha)}[\alpha](\Delta^h)^t W(\mathbf{u}^k)(\Delta^h) \right.$$
$$\left. + E_{q^k(\alpha)}[\alpha](\Delta^v)^t W(\mathbf{u}^k)(\Delta^v) \right)^{-1} E_{q^k(\beta)}[\beta] \mathbf{H}^t \mathbf{y}. \tag{48}$$

2) Calculate

$$\mathbf{u}^{k+1}_i = \left(\Delta^h_i(\mathbf{x}^k) \right)^2 + \left(\Delta^v_i(\mathbf{x}^k) \right)^2, \quad i = 1, \dots, N. \tag{49}$$

3) Calculate

$$q^{k+1}(\alpha, \beta) = q^{k+1}(\alpha) q^{k+1}(\beta) \tag{50}$$

where $q^{k+1}(\alpha)$ and $q^{k+1}(\beta)$ are gamma distributions given, respectively, by

$$q^{k+1}(\alpha) \propto \alpha^{N/2 + a^o_\alpha - 1} \exp\left[-\alpha \left(\sum_i \sqrt{u^{k+1}_i} + b^0_\alpha \right) \right] \tag{51}$$

$$q^{k+1}(\beta) \propto \beta^{N/2 + a^o_\beta} \exp\left[-\beta \left(\frac{\|\mathbf{y} - \mathbf{Hx}^k\|^2}{2} + b^o_\beta \right) \right]. \tag{52}$$

Set

$$q(\alpha, \beta) = \lim_{k \to \infty} q^k(\alpha, \beta), \quad \hat{\mathbf{x}} = \lim_{k \to \infty} \mathbf{x}^k. \tag{53}$$

Two additional factorizations of the distribution $q(\alpha, \beta, \mathbf{x})$ can be used. The first one corresponds to assuming that $q(\alpha, \beta)$ is a degenerate distribution. In this case, selecting as image estimate the mean value of the limiting $q(\mathbf{x})$ distribution in the corresponding algorithm is equivalent to performing the evidence analysis for the TV restoration problem. The second one corresponds to assuming that both $q(\alpha, \beta)$ and $q(\mathbf{x})$ are degenerate distributions. The corresponding algorithm is equivalent to maximizing alternatively in the hyperparameters and image the lower bound of $p(\alpha, \beta, \mathbf{x})$ given in (24). In other words, the estimation procedure is an iterated conditional mode (ICM) algorithm [36].

To end this section, we comment on two particular hyperparameter distributions $p(\alpha, \beta)$. The first one is obtained when both β and α are known quantities. Then Algorithm 2 with $\gamma_\alpha = \gamma_b = 1$, $\overline{\alpha}^o = \alpha$, and $\overline{\beta}^o = \beta$, provides the same solution with

$$\underline{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{ \frac{\beta}{2} \|\mathbf{y} - \mathbf{Hx}\|^2 + \alpha \text{TV}(\mathbf{x}) \right\}$$
$$= \arg\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{Hx}\|^2 + \frac{2\alpha}{\beta} \text{TV}(\mathbf{x}) \right\}. \tag{54}$$

If $\alpha = k/2$ with $k = 0.064$, the estimate of (54) is the one used in [11], and referred to as algorithm *BFO1* in Section V.

The second hyperparameter distribution $p(\alpha, \beta)$ is obtained when only β is known, that is, $\gamma_b = 1$, $\overline{\beta}^o = \beta$, and when $p(\alpha) \propto \alpha^{-1}$ and $\gamma_\alpha = 0$. Then Algorithm 2 at convergence provides [see (44)]

$$E_{q(\alpha)}[\alpha] = \frac{N/2}{\text{TV}(\hat{\mathbf{x}})} \tag{55}$$

and the solution for the image in (48) satisfies

$$-\beta \mathbf{H}^t(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}) + \frac{N/2}{\text{TV}(\hat{\mathbf{x}})} \times$$
$$\left[(\Delta^h)^t W(\hat{\mathbf{u}})(\Delta^h) + (\Delta^v)^t W(\hat{\mathbf{u}})(\Delta^v) \right] \hat{\mathbf{x}} = 0 \tag{56}$$

with

$$\hat{u}_i = \left(\Delta^h_i(\hat{\mathbf{x}}) \right)^2 + \left(\Delta^v_i(\hat{\mathbf{x}}) \right)^2, \quad i = 1, \dots, N. \tag{57}$$

Now, regularizing $W(\mathbf{u})$ by using $diag(1/\sqrt{u_i + \epsilon})$ where ϵ is a small positive constant to obtain a differentiable TV norm, we have

$$\left. \frac{\partial}{\partial \mathbf{x}} \text{TV}(\mathbf{x}) \right|_{\mathbf{x}=\hat{\mathbf{x}}} = \left[(\Delta^h)^t W(\hat{\mathbf{u}})(\Delta^h) + (\Delta^v)^t W(\hat{\mathbf{u}})(\Delta^v) \right] \hat{\mathbf{x}}. \tag{58}$$

Therefore, (56) can be rewritten as

$$-\beta \mathbf{H}^t(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}) + \left. \frac{N/2}{\text{TV}(\hat{\mathbf{x}})} \frac{\partial}{\partial \mathbf{x}} \text{TV}(\mathbf{x}) \right|_{\mathbf{x}=\hat{\mathbf{x}}} = 0. \tag{59}$$

That is, for this particular selection of $p(\alpha, \beta)$, Algorithm 2 provides the solution of

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{ \frac{\beta}{2} \|\mathbf{y} - \mathbf{Hx}\|^2 + \frac{N}{2} \log \mathrm{TV}(\mathbf{x}) \right\}. \quad (60)$$

Interestingly, this image estimate coincides with the image estimate proposed in [10], and referred to as algorithm *BFO2* in Section V, which is obtained as

$$\bar{\mathbf{x}} = \arg\max_{\mathbf{x}} \left\{ \log \left(\exp\left[-\frac{\beta}{2} \|\mathbf{y} - \mathbf{Hx}\|^2 \right] \right. \right.$$
$$\left. \left. \times \int \alpha^{N/2-1} \exp\left[-\alpha \mathrm{TV}(\mathbf{x}) \right] d\alpha \right) \right\}$$
$$= \arg\min_{\mathbf{x}} \left\{ \frac{\beta}{2} \|\mathbf{y} - \mathbf{Hx}\|^2 + \frac{N}{2} \log \mathrm{TV}(\mathbf{x}) \right\}. \quad (61)$$

Clearly, Algorithm 2 is a generalization of the algorithms presented in [11] and [10].

V. EXPERIMENTAL RESULTS

We performed a number of experiments to evaluate the performance of the proposed algorithms and also to compare them with other image restoration methods in the literature. We present results with Algorithm 1 (denoted by *ALG1*), Algorithm 2 (denoted by *ALG2*) and the TV-based approaches in [11] and [10], denoted (see the end of the previous section) by *BFO1* and *BFO2*, respectively. As already shown algorithms *BFO1* and *BFO2* are special cases of *ALG2*. We will elaborate on the differences and similarities of the methods in conjunction with the results. As in [11] and [10] we use a conjugate gradient algorithm (CG) to find the *BFO1* and *BFO2* image estimates.

We also included results obtained with the use of the algorithm in [16] which models the image distribution by a simultaneous autoregression (SAR) model [37] instead of a TV model and simultaneously estimates the prior and image hyperparameters. This algorithm will be denoted by *MOL* in the results. Comparing TV-based algorithms with this method provided useful insights about the proposed approaches.

In evaluating the upper bound of the performance of the proposed algorithms, we also provide results obtained by the algorithms denoted by *ALG1-TrueU*, *ALG2-TrueU*, *ALG1-True*, and *ALG2-True*. For the *ALG1-TrueU* and *ALG2-TrueU* algorithms, the noise variance $1/\beta$ is known (since we are dealing with synthetic experiments), and α and $\mathbf{u}$ are calculated using the original image [α from the equation $\alpha = (N/2)/\sum_i \sqrt{u_i}$ and $\mathbf{u}$ from (36) and (49)].

All three parameters are computed once, and, thus, they are not updated during the iterations. For the *ALG1-True* and *ALG2-True* algorithms α and β are treated as in *ALG1-TrueU* and *ALG2-TrueU*, but $\mathbf{u}$ is evaluated iteratively.

In our results, we provided the improvement in signal-to-noise ratio (ISNR) as an objective measure of the quality. The ISNR is defined as $10 \log_{10}(\|x - y\|^2 / \|x - \hat{x}\|^2)$, where x, y and $\hat{x}$ are the original, observed, and estimated images, respectively. In the tables we present in this section, we report the ISNR values, number of iterations, and estimated noise variances using a conjugate gradient (CG) approach [values in parentheses are obtained using a gradient descent (GD)

Fig. 2. (a) Lena image; degraded with a Gaussian shaped PSF with variance 9 and Gaussian noise of variance: (b) 0.16 ($\mathrm{BSNR} = 40$ dB), (c) 1.6 ($\mathrm{BSNR} = 30$ dB), (d) 1 ($\mathrm{BSNR} = 20$ dB).

approach to solve (33) and (48), as further discussed in Appendix I]. Note that since the parameter $1/\beta$ is not estimated by the algorithms *BFO1* and *BFO2*, but it is assumed known, the corresponding entries are denoted by "-". For all experiments, $\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 < 10^{-4}$ (or $\mathrm{E}_{q^k(\mathbf{x})}[\mathbf{x}]$ instead of $\mathbf{x}^k$) is used to terminate the algorithms, and a threshold of 10^{-6} is used to terminate the CG and GD iterations.

For the first set of experiments, we synthetically degraded the "Lena" and "Cameraman" images and the "Shepp–Logan" phantom with a Gaussian blur with variance 9 and additive Gaussian noise. We experimented with three noise levels, corresponding to blurred signal-to-noise ratios (BSNR) of 40, 30, and 20 dB. The original Lena image is shown in Fig. 2(a) and the degraded versions with the three noise levels in Fig. 2(b)–(d) (the corresponding noise variances are equal to 0.16, 1.6, and 16).

Flat hyperpriors on the hyperparameters are used as initial conditions, i.e., $a_\alpha^o = a_\beta^o = 0$ and $b_\alpha^o = b_\beta^o = 0$. The initially selected values for $\mathrm{E}_{q^1(\alpha)}[\alpha]$ and $\mathrm{E}_{q^1(\beta)}[\beta]$ for both *ALG1* and *ALG2* methods were equal to

$$\mathrm{E}_{q^1(\alpha)}[\alpha] = \frac{N/2}{\sum_i \sqrt{u_i}}, \quad \mathrm{E}_{q^1(\beta)}[\beta] = \frac{N}{\|\mathbf{y} - \mathbf{Hy}\|^2} \quad (62)$$

that is, we used the observations to initialize the hyperprior means. The observed image is used as the initial value of $\mathbf{x}$, and the initial value of $\mathbf{u}$ is calculated from this observed image. Note that the algorithms are initialized automatically without any manual input.

The ISNR values, the number of iterations, and the estimates of the noise variance $1/\beta$ are shown in Table I (it is noted that the true value of the noise variance is reported for the algorithms with the "True" suffix). In the second set of experiments, the

Fig. 3. Restorations of the Lena image blurred with a Gaussian PSF with variance 9 and 40-dB BSNR using the (a) *MOL* method (ISNR = 3.90 dB), (b) *BFO1* method (ISNR = 4.72 dB), (c) *BFO2* method (ISNR = 4.5 dB), (d) *ALG1* method (ISNR = 4.84 dB), and (e) *ALG2* method (ISNR = 4.64 dB).

Fig. 4. Restorations of the Lena image blurred with a Gaussian PSF with variance 9 and 20-dB BSNR using the (a) *MOL* method (ISNR = 2.45 dB), (b) *BFO1* method (ISNR = 3.02 dB), (c) *BFO2* method (ISNR = 2.47 dB), (d) *ALG1* method (ISNR = 3.06 dB), and (e) *ALG2* method (ISNR = 2.58 dB).

same images are degraded by a 9×9 uniform blur and additive Gaussian noise. The corresponding results are shown in Table II.

It is clear that knowledge of the noise and image parameters provides an advantage for *BFO1*; this method outperforms other methods in nearly all noise levels. However, both *ALG1* and *ALG2* result in comparable, in some cases even higher ISNR values, despite the fact that no prior information is assumed about the degradation process. We will later show that with the use of hyperpriors on the unknown hyperparameters higher ISNR values to the ones obtained by *BFO1* can be achieved by the *ALG1* and *ALG2* algorithms.

The important point to note here is that *ALG1* and *ALG2* generally perform better that *BFO2* and *MOL*. The proposed methods generally result in higher ISNR values than *BFO2*, although the noise variance is assumed to be known in *BFO2*. The *MOL* algorithm is outperformed by other methods in all experiments, although the noise variance is very accurately estimated.

This comparison clearly shows that the spatially adaptive deconvolution and noise removal achieved by TV-based restoration methods provides a significant improvement over methods like *MOL* which do not incorporate spatial adaptivity in the estimation procedure.

We also note that the proposed methods are robust to the initial values of the hyperparameters. For instance, when the algorithms are initialized using $\beta = 1$ and $\alpha = (0.064/2)\sigma^2$, as in [11], the resulting ISNR values are similar to the ones reported in Table I. For instance, for the 40-dB BSNR case with the Lena image, the ISNR values are 4.64 (4.75) dB and 4.34 (4.42) dB, and for the 20-dB BSNR case, the ISNR values are 2.88 (3.06) dB and 2.45 (2.51) dB for the *ALG1* and *ALG2* methods, respectively. These results show the robustness of the methods to parameter initialization.

Although the results in Table II are similar to the Gaussian blur case, we note some interesting differences. It is clear that

TABLE I
ISNR VALUES AND NUMBER OF ITERATIONS FOR THE LENA, CAMERAMAN, AND SHEPP–LOGAN IMAGES DEGRADED BY A GAUSSIAN BLUR WITH VARIANCE 9

		Lena			Cameraman			Shepp-Logan		
BSNR	Method	ISNR (dB)	iterations	$1/\beta$	ISNR (dB)	iterations	$1/\beta$	ISNR (dB)	iterations	$1/\beta$
40dB	MOL	3.90	26	0.16	2.73	35	0.30	3.67	40	0.45
	BFO1	4.72	20	-	3.51	19	-	7.07	19	-
	BFO2	4.50	19	-	3.27	16	-	5.88	17	-
	ALG1	4.78(4.84)	15(10)	0.16 (0.16)	3.39(3.38)	21(16)	0.30(0.30)	6.69(6.46)	36(35)	0.45(0.45)
	ALG2	4.49(4.64)	17(16)	0.16 (0.16)	3.26(3.28)	18(17)	0.30(0.30)	5.63(5.56)	20(21)	0.45(0.45)
	ALG1-TrueU	6.95	3	0.16	5.35	13	0.30	59.33	7	0.45
	ALG2-TrueU	6.95	4	0.16	5.33	18	0.30	58.91	7	0.45
	ALG1-True	4.84(4.89)	12(8)	0.16	3.49(3.44)	16(12)	0.30	6.60(6.42)	36(37)	0.45
	ALG2-True	4.65(4.86)	21(13)	0.16	3.49(3.49)	18(14)	0.30	5.95(6.09)	16(19)	0.45
30dB	MOL	3.13	27	1.62	2.14	35	3.05	2.91	49	4.53
	BFO1	3.87	24	-	2.89	18	-	5.15	18	-
	BFO2	3.56	21	-	2.47	16	-	3.94	16	-
	ALG1	3.87(4.03)	24(16)	1.60(1.60)	2.63(2.68)	26(23)	4.60(4.60)	4.31(4.30)	43(44)	4.60(4.60)
	ALG2	3.55(3.67)	20(21)	1.60(1.60)	2.41(2.49)	19(23)	4.78(4.70)	3.72(3.78)	16(39)	4.78(4.70)
	ALG1-TrueU	6.01	2	1.60	4.55	5	3.00	43.23	10	4.50
	ALG2-TrueU	6.01	2	1.60	4.55	6	3.00	36.38	9	4.50
	ALG1-True	3.94(4.12)	22(14)	1.60	2.86(2.91)	20(16)	3.00	4.31(4.31)	43(44)	4.50
	ALG2-True	3.81(3.95)	23(20)	1.60	2.87(2.95)	19(19)	3.00	4.13(4.21)	16(53)	4.50
20dB	MOL	2.45	4	16.05	1.64	3	30.23	2.24	3	45.11
	BFO1	3.02	20	-	2.13	16	-	3.56	17	-
	BFO2	2.47	23	-	2.23	29	-	2.20	13	-
	ALG1	2.87(3.06)	30(23)	16.22(16.20)	1.72(1.78)	38(36)	31.69(31.57)	1.85(1.86)	47(48)	48.66(48.61)
	ALG2	2.42(2.58)	23(34)	16.74(16.48)	1.42(1.58)	16(30)	33.52(32.62)	2.05(1.72)	12(53)	51.11(49.74)
	ALG1-TrueU	5.07	2	16.00	3.71	5	30.00	24.05	91	45.00
	ALG2-TrueU	5.07	2	16.00	3.70	6	30.00	21.56	72	45.00
	ALG1-True	3.05(3.25)	25(21)	16.00	2.13(2.20)	27(27)	30.00	2.20(2.21)	48(48)	45.00
	ALG2-True	2.92(3.04)	20(28)	16.00	2.10(2.20)	15(29)	30.00	2.49(2.28)	14(51)	45.00

ALG2 outperforms *ALG1* in high BSNRs, but it results in a lower ISNR in the low BSNR case. We can conclude that in the high BSNR case, where the noise level is low, exploiting additional information using the full variational formulation actually results in lower performance. However, using the full variational algorithm, i.e., *ALG1*, provides better image estimates in the low BSNR case. Another remark is that both algorithms fail to accurately estimate the noise variance when the noise level is very low at 40-dB BSNR, although the estimated noise variance is very close to the true value at high noise levels.

The results obtained with the use of GD and CG are comparable, although in most cases GD results in fewer iterations.

A fair comparison between *BFO1* and the proposed approaches can be made by looking at the performances of *ALG1-True* and *ALG2-True*. In most cases *ALG1-True* and *ALG2-True* outperform *BFO1*, while a smaller number of iterations is adequate for convergence. Additionally, *ALG1-TrueU* and *ALG2-TrueU* provide the upper bound in ISNR that can be achieved by TV-based restoration methods represented here.

Clearly, knowledge of the true value of the matrix **u** provides a significant advantage to the methods.

We next examine the effect of the introduction of additional information about the unknown hyperparameters through the use of the confidence parameters γ_α and γ_β on the performance of the algorithms. As we have already explained before, in the case of $\gamma_\alpha = \gamma_\beta = 0$, no information about the hyperparameters is available, and the observed image is responsible for the estimation of the hyperparameters and the image. However, one usually has some information about the original image and the degradation process. For example, off-line estimates of the image and noise variance can be computed, and provided to the algorithms. In our experiments, we provided the true image and noise variance to the algorithms and run the algorithms while varying the confidence parameters γ_α and γ_β from 0 to 1 in 0.1 intervals.

Table III shows the means of the posterior distributions of the hyperparameters, ISNR values, and the number of iterations obtained using *ALG1* for selected values of the confidence pa-

TABLE II
ISNR VALUES AND NUMBER OF ITERATIONS FOR THE LENA, CAMERAMAN, AND SHEPP–LOGAN IMAGES DEGRADED BY A 9×9 UNIFORM BLUR

		Lena			Cameraman			Shepp-Logan		
BSNR	Method	ISNR (dB)	iterations	$1/\beta$	ISNR (dB)	iterations	$1/\beta$	ISNR (dB)	iterations	$1/\beta$
40dB	*MOL*	6.79	21	0.17	6.16	25	0.31	5.82	27	0.48
	BFO1	8.34	8	-	8.55	8	-	14.22	12	-
	BFO2	8.35	9	-	8.25	12	-	12.01	12	-
	ALG1	8.42(8.24)	21(19)	0.12(0.12)	8.57(8.34)	28(24)	0.25(0.25)	13.69(13.32)	19(20)	0.40(0.40)
	ALG2	8.37(8.36)	16(15)	0.12(0.13)	8.46(8.29)	18(21)	0.27(0.27)	13.26(12.9)	15(17)	0.44(0.44)
	ALG1-TrueU	10.93	3	0.17	11.34	9	0.31	58.62	7	0.46
	ALG2-TrueU	10.93	3	0.17	11.33	13	0.31	58.41	7	0.46
	ALG1-True	8.48(8.43)	7(7)	0.17	8.58(8.19)	9(8)	0.31	13.53(13.52)	21(17)	0.46
	ALG2-True	8.46(8.40)	7(9)	0.17	8.53(8.51)	10(11)	0.31	12.14(12.68)	12(16)	0.46
30dB	*MOL*	4.62	30	1.71	3.98	39	3.22	4.10	41	4.69
	BFO1	6.08	11	-	5.68	10	-	8.88	12	-
	BFO2	5.64	12	-	4.65	14	-	6.91	13	-
	ALG1	5.89(5.89)	13(14)	1.49(1.49)	5.41(5.41)	20(19)	3.04(3.04)	7.77(7.75)	31(29)	4.60(4.60)
	ALG2	5.58(5.61)	14(14)	1.64(1.64)	4.38(4.39)	15(17)	3.60(3.60)	6.50(6.85)	15(29)	4.84(4.77)
	ALG1-TrueU	8.23	2	1.66	8.26	5	3.10	44.63	8	4.60
	ALG2-TrueU	8.23	2	1.66	8.26	6	3.10	38.01	7	4.60
	ALG1-True	5.96(5.97)	9(8)	1.66	5.70(5.72)	11(11)	3.10	7.72(7.70)	30(29)	4.60
	ALG2-True	6.02(6.04)	10(11)	1.66	5.61(5.64)	12(13)	3.10	7.14(7.44)	13(24)	4.60
20dB	*MOL*	2.94	18	17.07	2.26	4	32.42	2.66	5	45.68
	BFO1	4.09	14	-	3.31	14	-	5.57	16	-
	BFO2	4.14	16	-	2.12	20	-	2.95	14	-
	ALG1	3.72(3.83)	13(13)	17.00(16.91)	2.42(2.46)	23(22)	34.05(34.00)	3.01(3.03)	35(34)	49.97(49.88)
	ALG2	3.15(3.28)	12(22)	17.93(17.59)	1.94(2.12)	16(28)	36.52(35.56)	2.64(2.54)	12(40)	53.67(51.92)
	ALG1-TrueU	8.23	2	16.67	5.33	5	31.00	25.10	85	46.00
	ALG2-TrueU	8.23	2	16.67	5.33	6	31.00	22.97	68	46.00
	ALG1-True	5.95(5.96)	9(8)	16.67	3.33(3.38)	14(14)	31.00	3.51(3.51)	33(33)	46.00
	ALG2-True	6.02(6.04)	10(11)	16.67	3.25(3.32)	17(18)	31.00	3.34(3.37)	15(34)	46.00

TABLE III
POSTERIOR MEANS OF THE DISTRIBUTIONS OF THE HYPERPARAMETERS, ISNR, AND NUMBER OF ITERATIONS USING *ALG1* FOR THE LENA IMAGE WITH 40-dB AND 20-dB BSNR USING $\overline{\alpha}^o = 1/23.84$, AND $\overline{\beta}^o = 1/0.16$ AND $\overline{\beta}^o = 1/16$, RESPECTIVELY, FOR DIFFERENT VALUES OF γ_β AND γ_α

BSNR	γ_β	γ_α	$E[\beta]^{-1}$	$E[\alpha]^{-1}$	ISNR (dB)	iterations
40dB	1.0	1.0	0.16	23.84	4.75	13
	0.0	0.5	0.24	22.26	4.66	16
	0.5	0.0	0.20	20.30	4.68	16
	0.0	0.0	0.24	19.89	4.63	18
20dB	1.0	1.0	16.00	23.84	3.07	26
	0.0	0.5	16.17	20.83	3.01	28
	0.5	0.0	16.16	16.29	2.89	30
	0.0	0.0	16.34	16.29	2.88	30

rameters. The confidence values are selected to demonstrate the behavior of the algorithm in the following cases: 1) when full information about the image and noise variance is available, 2) when no information is provided, i.e., the observation is fully responsible for the restoration, 3) when some information about the image prior variance α is provided, and 4) when some information about the noise variance is provided. Moreover, the evolution of ISNR for the full set of confidence parameters is depicted in Fig. 5. A similar ISNR evolution is obtained for *ALG2* so its corresponding plot is not shown. It can be observed that the noise level changes the effect of the confidence parameters. In the low-noise case $(\mathrm{BSNR} = 40 \text{ dB})$, information about the noise variance affects the final ISNR more than the information about the image variance; there is almost no ISNR variance when $\gamma_\beta = 1$ and γ_α changes from 0 to 1. However, in the 20-dB BSNR case information about the image variance is more valuable than the noise variance. For a fixed γ_α, the ISNR value remains fixed for varying γ_β, whereas increasing the image variance confidence increases the obtained ISNR. It can be stated as a final remark that the algorithm is less successful at estimating the noise variance in low noise conditions, and less successful at
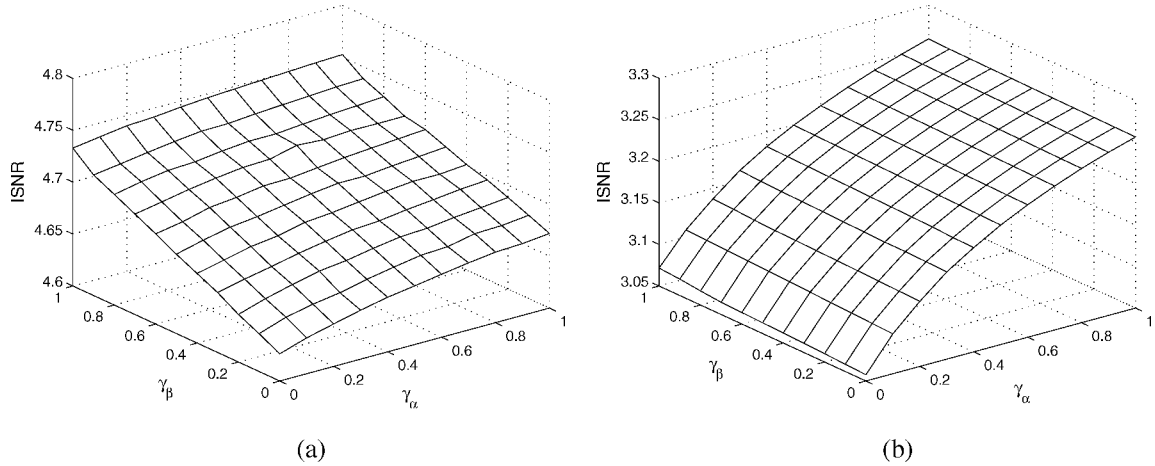
Fig. 5. Evolution of ISNR using *ALG1* for different values of γ_α and γ_β for the restoration of the Lena image blurred with a Gaussian with variance 9, and (a) $\mathrm{BSNR} = 40$ dB; (b)$\mathrm{BSNR} = 20$ dB.
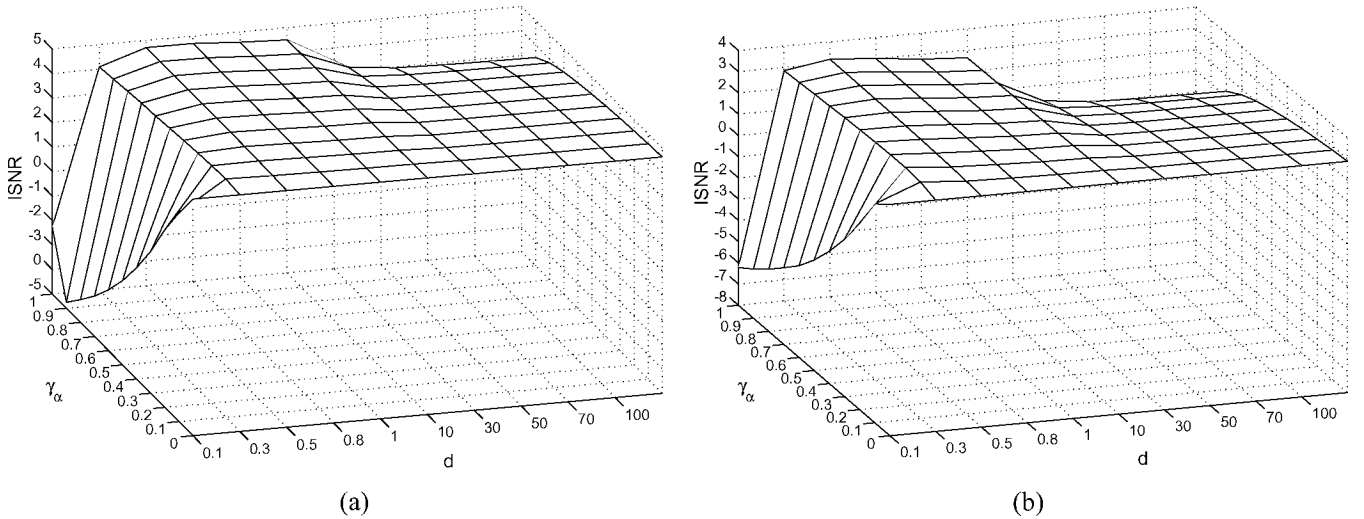

Fig. 6. Evolution of ISNR with varying γ_α and $\overline{\alpha}^o$ for Lena image degraded with Gaussian blur with variance 9 at (a) 40-dB BSNR and (b) 20-dB BSNR (note that $\overline{\alpha}^o = d \cdot \hat{\alpha}$).

estimating the image variance in high noise conditions. Therefore, information about the poorly estimated parameter helps to further increase the ISNR values. However, we should also state that the ISNR variation in these plots is small compared to the ISNR values (difference between the maximum and minimum ISNR values are 0.13 dB at 40-dB BSNR and 0.19 dB at 20-dB BSNR); therefore, we can see that the algorithm is robust to the estimated hyperparameter values in terms of the final restored image quality.

We will now examine the additional information provided by the variational approach and study how the distributions on the hyperparameters can be used to improve the results already obtained. We start our experiments by assuming flat hyperpriors, and applied algorithm *ALG2* to the Lena image degraded by a Gaussian blur of variance 9 and additive Gaussian noise at 40- and 20-dB BSNR, as we had before. This provides estimates of the noise and image variance, denoted by $\hat{\beta}$ and $\hat{\alpha}$, respectively. Next we run algorithm *ALG2* multiple times on the same degraded image with different initial hyperparameters: The final estimated noise variance of the algorithm is used without update, i.e., $\overline{\beta}^o = \hat{\beta}$ and $\gamma_\beta = 1$. By moving γ_α in

[0,1] and selecting the hyperprior mean as $\overline{\alpha}^o = d \cdot \hat{\alpha}$, where d is in the range $[10^{-1}, 10^2]$, we obtain the ISNR evolution graphs shown in Fig. 6(a) for the 40-dB BSNR case and Fig. 6(b) for the 20-dB BSNR case. It should be noted that the range of ISNR values obtained by this experiment includes the best ISNR achieved with known hyperparameter values, depicted in Table I, corresponding to *ALG2-True*. Thus, as expected, the results by *ALG1-True* and *ALG2-True* are included in the case when different selections of the gamma hyperpriors on the hyperparameters are used. A few remarks can be made by examining at Fig. 6. First, the algorithms are very robust with respect to the parameter α, since even in the case $\overline{\alpha}^o = 100 \cdot \hat{\alpha}$ the resulting ISNR value is very close to the highest achievable value. Second, one can conclude that the distribution of α is not sharply peaked at one value, and, therefore, multiple values of this parameter can be used in the restoration process without greatly affecting the performance of the algorithm.

Overall, the experimental results demonstrate that algorithms *ALG1* and *ALG2* provide comparable performance to the existing TV-based approaches even though no prior knowledge about the image and degradation process is assumed, and out-

perform them if prior knowledge is utilized. It is also clear that TV-based approaches result in higher quality restorations than nonspatially adaptive restoration methods. Another important point to be made is that with the developed framework, we can draw different estimates for the unknown hyperparameters from their estimated distributions and, thus, assign a degree of trust to the results and potentially achieve improved restoration results. The major distinction between the proposed algorithms *ALG1* and *ALG2* is that *ALG1* provides the approximation to the posterior distribution of the unknown image. For scientific applications for which a confidence value for a restoration is important (i.e., restoration of astronomical or medical images), *ALG1* can provide such information through the use of this posterior distribution. On the other hand, when images are restored for, for example, consumer applications *ALG2* can be the algorithm of choice.

The proposed algorithms are computationally more intensive than nonspatially adaptive restoration methods since (33) and (48) cannot be solved by direct inversion in the frequency domain and numerical approaches are utilized. Typically, the MATLAB implementations of our algorithms required on the average about 2–5 min on a 3.20-GHz Xeon PC for 256 × 256 images. Note that the running time of the algorithms can be improved by utilizing preconditioning methods (see, for example, [38]–[41]).

VI. CONCLUSION

We have presented two new methods for the simultaneous estimation of the image and the unknown hyperparameters in TV-based image restoration problems. We adopt a variational approach to provide approximations to the posterior distributions of the unknown variables. Utilizing this variational framework, different values from the posterior distributions can be drawn as estimates to the latent variables and prior information about the degradation process and the unknowns can be incorporated into the estimation process to increase the performance of the algorithms. We have analyzed the proposed methods and demonstrated that some of the current methods in TV-based image restoration are special cases of our formulation. Experimental results are provided to show the performance of the methods in the case where information about the degradation process and the unknown variables is not available, and when some information can be provided for improved performance.

APPENDIX I
CALCULATION OF THE IMAGE ESTIMATES IN ALGORITHMS 1 AND 2

To obtain the image estimates, the mean of the distribution $q^k(\mathbf{x})$ in (33) is used in Algorithm 1 and the point estimate in (48) is used in Algorithm 2. The estimation of the quantities can be carried out by the gradient descent (GD) or the conjugate gradient (CG) methods. Note that by using the GD or the CG methods we avoid the calculation of the inverse of the covariance matrix. Our descriptions will be specifically for Algorithm 1. However, the same results apply to Algorithm 2. We next describe the specific GD steps applied to the solution of

$$\mathbf{A}^k \mathbf{x} = \mathrm{E}_{q^k(\beta)}[\beta]\mathbf{H}^t \mathbf{y} \qquad (A1)$$

where

$$\mathbf{A}^k = \left(\mathrm{E}_{q^k(\beta)}[\beta]\mathbf{H}^t\mathbf{H} + \mathrm{E}_{q^k(\alpha)}[\alpha](\Delta^h)^t W(\mathbf{u}^k)(\Delta^h) \right.$$
$$\left. + \mathrm{E}_{q^k(\alpha)}[\alpha](\Delta^v)^t W(\mathbf{u}^k)(\Delta^v) \right). \quad (A2)$$

In the description that follows, we use the notation $i : +1$, $i : +2$, $i : +3$, $i : +4$ to denote the four pixels around pixel i (if $i = (u, v)$ they correspond to $(u + 1, v)$, $(u, v + 1)$, $(u - 1, v)$, and $(u, v - 1)$, respectively).

We now expand the matrix $\mathrm{E}_{q^k(\alpha)}[\alpha](\Delta^h)^t W(\mathbf{u}^k)(\Delta^h) + \mathrm{E}_{q^k(\alpha)}[\alpha](\Delta^v)^t W(\mathbf{u}^k)(\Delta^v)$ and calculate $[\mathbf{A}^k\mathbf{x}]$ at position i, $[\mathbf{A}^k\mathbf{x}]_i$. We have

$$[\mathbf{A}^k\mathbf{x}]_i = \left[\mathrm{E}_{q^k(\beta)}[\beta]\mathbf{H}^t\mathbf{H}\mathbf{x} \right]_i + 2\mathrm{E}_{q^k(\alpha)}[\alpha]\frac{1}{\sqrt{u_i^k}}\mathbf{x}_i$$
$$+ \mathrm{E}_{q^k(\alpha)}[\alpha]\frac{1}{\sqrt{u_{i:+3}^k}}\mathbf{x}_i + \mathrm{E}_{q^k(\alpha)}[\alpha]\frac{1}{\sqrt{u_{i:+4}^k}}\mathbf{x}_i$$
$$- \mathrm{E}_{q^k(\alpha)}[\alpha]\frac{1}{\sqrt{u_i^k}}\mathbf{x}_{i:+1} - \mathrm{E}_{q^k(\alpha)}[\alpha]\frac{1}{\sqrt{u_i^k}}\mathbf{x}_{i:+2}$$
$$- \mathrm{E}_{q^k(\alpha)}[\alpha]\frac{1}{\sqrt{u_{i:+3}^k}}\mathbf{x}_{i:+3}$$
$$- \mathrm{E}_{q^k(\alpha)}[\alpha]\frac{1}{\sqrt{u_{i:+4}^k}}\mathbf{x}_{i:+4}. \qquad (A3)$$

Let us now define

$$\mathbf{z}_j(\mathbf{u}^k) = \mathrm{E}_{q^k(\alpha)}[\alpha]\frac{1}{\sqrt{u_j^k}}. \qquad (A4)$$

Using this, we obtain

$$[\mathbf{A}^k\mathbf{x}]_i = \left[\mathrm{E}_{q^k(\beta)}[\beta]\mathbf{H}^t\mathbf{H}\mathbf{x} \right]_i + 2\mathbf{z}_i(\mathbf{u}^k)\mathbf{x}_i + \mathbf{z}_{i:+3}(\mathbf{u}^k)\mathbf{x}_i$$
$$+ \mathbf{z}_{i:+4}(\mathbf{u}^k)\mathbf{x}_i - \mathbf{z}_i(\mathbf{u}^k)(\mathbf{x}_{i:+1} + \mathbf{x}_{i:+2})$$
$$- \mathbf{z}_{i:+3}(\mathbf{u}^k)\mathbf{x}_{i:+3} - \mathbf{z}_{i:+4}(\mathbf{u}^k)\mathbf{x}_{i:+4}. \qquad (A5)$$

Combining with (A1), we obtain

$$2\mathbf{z}_i(\mathbf{u}^k)\mathbf{x}_i + \mathbf{z}_{i:+3}(\mathbf{u}^k)\mathbf{x}_i + \mathbf{z}_{i:+4}(\mathbf{u}^k)\mathbf{x}_i$$
$$= \mathbf{z}_i(\mathbf{u}^k)(\mathbf{x}_{i:+1} + \mathbf{x}_{i:+2}) + \mathbf{z}_{i:+3}(\mathbf{u}^k)\mathbf{x}_{i:+3}$$
$$+ \mathbf{z}_{i:+4}(\mathbf{u}^k)\mathbf{x}_{i:+4} + \mathrm{E}_{q^k(\beta)}[\beta]\left[\mathbf{H}^t(\mathbf{y} - \mathbf{H}\mathbf{x})\right]_i. \quad (A6)$$

Adding $\mathrm{E}_{q^k(\beta)}[\beta]\mathbf{x}_i$ to both sides of the above equation, we have

$$2\mathbf{z}_i(\mathbf{u}^k)\mathbf{x}_i + \mathbf{z}_{i:+3}(\mathbf{u}^k)\mathbf{x}_i + \mathbf{z}_{i:+4}(\mathbf{u}^k)\mathbf{x}_i + \mathrm{E}_{q^k(\beta)}[\beta]\mathbf{x}_i$$
$$= \mathbf{z}_i(\mathbf{u}^k)(\mathbf{x}_{i:+1} + \mathbf{x}_{i:+2}) + \mathbf{z}_{i:+3}(\mathbf{u}^k)\mathbf{x}_{i:+3}$$
$$+ \mathbf{z}_{i:+4}(\mathbf{u}^k)\mathbf{x}_{i:+4} + \mathrm{E}_{q^k(\beta)}[\beta]\left[\mathbf{H}^t(\mathbf{y} - \mathbf{H}\mathbf{x})\right]_i + \mathbf{x}_i].$$
$$(A7)$$

Let

$$\mathbf{r}_i(\mathbf{u}^k) = 2\mathbf{z}_i(\mathbf{u}^k) + \mathbf{z}_{i:+3}(\mathbf{u}^k) + \mathbf{z}_{i:+4}(\mathbf{u}^k) + \mathrm{E}_{q^k(\beta)}[\beta]. \quad (A8)$$

Finally, using this, we have to find the solution of

$$\mathbf{x}_i = \frac{\mathbf{z}_i(\mathbf{u}^k)}{\mathbf{r}_i(\mathbf{u}^k)}(\mathbf{x}_{i:+1} + \mathbf{x}_{i:+2}) + \frac{\mathbf{z}_{i:+3}(\mathbf{u}^k)}{\mathbf{r}_i(\mathbf{u}^k)}\mathbf{x}_{i:+3} + \frac{\mathbf{z}_{i:+4}(\mathbf{u}^k)}{\mathbf{r}_i(\mathbf{u}^k)}\mathbf{x}_{i:+4}$$
$$+ \frac{\mathrm{E}_{q^k(\beta)}[\beta]}{\mathbf{r}_i(\mathbf{u}^k)}\left[\left[\mathbf{H}^t(\mathbf{y} - \mathbf{H}\mathbf{x})\right]_i + \mathbf{x}_i\right] \quad (A9)$$

from which the GD iteration is obtained, that is

$$
\begin{aligned}
\mathbf{x}_i^{l+1} =& \frac{\mathbf{z}_i(\mathbf{u}^k)}{\mathbf{r}_i(\mathbf{u}^k)} \left(\mathbf{x}_{i:+1}^l + \mathbf{x}_{i:+2}^l \right) + \frac{\mathbf{z}_{i:+3}(\mathbf{u}^k)}{\mathbf{r}_i(\mathbf{u}^k)} \mathbf{x}_{i:+3}^l \\
&+ \frac{\mathbf{z}_{i:+4}(\mathbf{u}^k)}{\mathbf{r}_i(\mathbf{u}^k)} \mathbf{x}_{i:+4}^l + \frac{\mathrm{E}_{\mathrm{q}^k(\beta)}[\beta]}{\mathbf{r}_i(\mathbf{u}^k)} \\
&\times \left[\left[\mathbf{H}^t(\mathbf{y} - \mathbf{H}\mathbf{x}^l) \right]_i + \mathbf{x}_i^l \right], \quad \forall i.
\end{aligned} \tag{A10}
$$

Alternatively, a CG method can be applied. In our experiments we used the basic CG version shown in [42] to solve (A1). Note that several methods can be used (see, for instance, [38]–[40]) to calculate the TV image estimate without the use of the majorization of the TV prior.

APPENDIX II
CALCULATION OF REQUIRED EXPECTED
VALUES IN ALGORITHM 1

In this section we show how the calculations of $\mathbf{u}_i^{k+1}$ in (36) and $\mathrm{E}_{\mathrm{q}^{k+1}(\beta)}[\beta]$ in (41) are carried out. We first expand (36) to obtain

$$
\begin{aligned}
&\mathrm{E}_{\mathrm{q}^k(\mathbf{x})} \left[\left(\Delta_i^h(\mathbf{x}) \right)^2 + \left(\Delta_i^v(\mathbf{x}) \right)^2 \right] \\
&= \left(\Delta_i^h \left(\mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}] \right) \right)^2 \\
&\quad + \left(\left(\Delta_i^v \left(\mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}] \right) \right)^2 \right. \\
&\quad + \mathrm{E}_{\mathrm{q}^k(\mathbf{x})} \left[\left(\Delta_i^h \left(\mathbf{x} - \mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}] \right) \right)^2 \right] \\
&\quad + \mathrm{E}_{\mathrm{q}^k(\mathbf{x})} \left[\left(\Delta_i^v \left(\mathbf{x} - \mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}] \right) \right)^2 \right].
\end{aligned} \tag{A11}
$$

For (41), we have

$$
\begin{aligned}
\mathrm{E}_{\mathrm{q}^k(\mathbf{x})} \left[\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \right] =& \left\| \mathbf{y} - \mathbf{H}\mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}] \right\|^2 \\
&+ \mathrm{trace} \left(\mathrm{cov}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}] \mathbf{H}^t\mathbf{H} \right).
\end{aligned} \tag{A12}
$$

Therefore, $\mathrm{cov}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}]$ is explicitly needed to calculate these quantities. However, since the calculation of $\mathrm{cov}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}]$ is very intense, we propose the following approximation of the covariance matrix. We first approximate $W(\mathbf{u}^k)$ using

$$
W(\mathbf{u}^k) \approx z(\mathbf{u}^k)\mathbf{I} \tag{A13}
$$

where $z(\mathbf{u}^k)$ is calculated as the mean value of the diagonal values in $W(\mathbf{u}^k)$, that is

$$
z(\mathbf{u}^k) = \frac{1}{N} \sum_i \frac{1}{\sqrt{u_i^k}}. \tag{A14}
$$

We then approximate $\mathrm{cov}_{\mathrm{q}^k(\mathbf{x})}$ using

$$
\begin{aligned}
\mathrm{cov}_{\mathrm{q}^k(\mathbf{x})} \approx& \left(\mathrm{E}_{\mathrm{q}^k(\beta)}[\beta]\mathbf{H}^t\mathbf{H} + \mathrm{E}_{\mathrm{q}^k(\alpha)}[\alpha]z(\mathbf{u}^k)(\Delta^h)^t(\Delta^h) \right. \\
&\left. + \mathrm{E}_{\mathrm{q}^k(\alpha)}[\alpha]z(\mathbf{u}^k)(\Delta^v)^t(\Delta^v) \right)^{-1} \tag{A15} \\
=& \mathbf{B}^{-1}. \tag{A16}
\end{aligned}
$$

Note that the matrix $\mathbf{B}$ is a block circulant matrix with circulant blocks (BCCB); thus, computing its inverse can be performed in Fourier domain, which is very efficient [35].

Using this approximation, the last two terms in (A11) can be expressed as

$$
\begin{aligned}
&\mathrm{E}_{\mathrm{q}^k(\mathbf{x})} \left[\left(\Delta_i^h \left(\mathbf{x} - \mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}] \right) \right)^2 \right] \\
&+ \mathrm{E}_{\mathrm{q}^k(\mathbf{x})} \left[\left(\Delta_i^v \left(\mathbf{x} - \mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}] \right) \right)^2 \right] \\
&\approx \frac{1}{N} \mathrm{trace} \left[\mathbf{B}^{-1} \times \left((\Delta^h)^t(\Delta^h) + (\Delta^v)^t(\Delta^v) \right) \right].
\end{aligned} \tag{A17}
$$

Finally, we can approximate the last term in (A12) as follows:

$$
\mathrm{trace} \left[\mathrm{cov}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}]\mathbf{H}^t\mathbf{H} \right] \approx \mathrm{trace}[\mathbf{B}^{-1}\mathbf{H}^t\mathbf{H}]. \tag{A18}
$$

REFERENCES

[1] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Total variation image restoration and parameter estimation using variational posterior distribution approximation," presented at the Int. Conf. Image Processing, San Antonio, TX, Sep. 2007.
[2] A. K. Katsaggelos, Ed., *Digital Image Restoration*. New York: Springer-Verlag, 1991, vol. 23.
[3] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Process. Mag.*, vol. 14, no. 2, pp. 24–41, Mar. 1997.
[4] P. Sarder and A. Nehorai, "Deconvolution methods for 3D fluorescence microscopy images: An overview," *IEEE Signal Process. Mag.*, vol. 23, no. 3, pp. 32–45, May 2006.
[5] T. F. Chan and J. Shen, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. Philadelphia, PA: SIAM, 2005.
[6] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, pp. 259–268, 1992.
[7] A. B. Hamza, H. Krim, and G. B. Unal, "Unifying probabilistic and variational estimation," *IEEE Signal Process. Mag.*, vol. 19, no. 9, pp. 37–47, Sep. 2002.
[8] T. F. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," in *Handbook of Mathematical Models in Computer Vision*, N. Paragios, Y. Chen, and O. Faugeras, Eds. New York: Springer Verlag, 2005.
[9] M. Bertalmio, V. Caselles, B. Rougé, and A. Solé, "TV based image restoration with local constraints," *J. Sci. Comput.*, vol. 19, no. 1–3, pp. 95–122, Dec. 2003.
[10] J. Bioucas-Dias, M. Figueiredo, and J. Oliveira, "Adaptive total-variation image deconvolution: A majorization-minimization approach," presented at the EUSIPCO, Florence, Italy, Sep. 2006.
[11] J. Bioucas-Dias, M. Figueiredo, and J. Oliveira, "Total-variation image deconvolution: A majorization-minimization approach," presented at the ICASSP, Toulouse, France, May 2006.
[12] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 231–246, Feb. 1999.
[13] J. Mateos, A. Katsaggelos, and R. Molina, "A Bayesian approach to estimate and transmit regularization parameters for reducing blocking artifacts," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1200–1215, Jul. 2000.
[14] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, A. K. Katsaggelos, and J. Mateos, "Hyperparameter estimation in image restoration problems with partially-known blurs," *Opt. Eng.*, vol. 41, no. 8, pp. 1845–1854, Aug. 2002.
[15] R. Molina, J. Mateos, and A. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3715–3727, Dec. 2006.
[16] R. Molina, "On the hierarchical Bayesian approach to image restoration. Applications to astronomical images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 11, pp. 1122–1128, Nov. 1994.
[17] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, and A. K. Katsaggelos, "Hierarchical Bayesian image restoration for partially-known blur," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1784–1797, Aug. 2000.
[18] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan, "An introduction to MCMC for machine learning," *Mach. Learn.*, vol. 50, no. 1–2, pp. 5–43, Jan. 2003.

[19] J. Miskin, "Ensemble learning for independent component analysis," Ph.D. dissertation, Astrophysics Group, Univ. Cambridge, Cambridge, U.K., 2000.

[20] M. J. Beal, "Variational algorithms for approximate bayesian inference," Ph.D. dissertation, The Gatsby Computational Neuroscience Unit, Univ. College London, London, U.K., 2003.

[21] V. Smidl and A. Quinn, *The Variational Bayes Method in Signal Processing*. New York: Springer Verlag, 2005.

[22] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

[23] M. I. Jordan, Z. Ghahramani, T. S. Jaakola, and L. K. Saul, "An Introduction to variational methods for graphical models," in *Learn. Graph. Models*. Cambridge, MA: MIT Press, 1998, pp. 105–162.

[24] S. Kullback, *Information Theory and Statistics*. New York: Dover, 1959.

[25] C. M. Bishop and M. E. Tipping, "Variational relevance vector machine," in *Proc. 16th Conf. Uncertainty in Articial Intelligence*, 2000, pp. 46–53.

[26] J. W. Miskin and D. J. C. MacKay, "Ensemble learning for blind image separation and deconvolution," in *Advances in Independent Component Analysis*, M. Girolami, Ed. New York: Springer-Verlag, July 2000.

[27] A. C. Likas and N. P. Galatsanos, "A variational approach for Bayesian blind image deconvolution," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2222–2233, Aug. 2004.

[28] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York: Springer Verlag, 1985, ch. 3 and 4.

[29] K. Lange, "Optimization," in *Springer Texts in Statistics*. New York: Springer-Verlag, 2004.

[30] T. S. Jaakkola and M. I. Jordan, "Bayesian parameter estimation via variational methods," *Statist. Comput.*, vol. 10, no. 1, pp. 25–37, 2000.

[31] C. Bishop and M. Svensen, "Bayesian hierarchical mixtures of experts," in *Proc. 19th Annu. Conf. Uncertainty in Artificial Intelligence*, San Francisco, CA, 2003, pp. 57–64.

[32] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006.

[33] G. L. Anderson and A. N. Netravali, "Image restoration based on a subjective criterion," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, no. 6, pp. 845–853, Dec. 1976.

[34] A. K. Katsaggelos and M. G. Kang, "A spatially adaptive iterative algorithm for the restoration of astronomical images," *Proc. Int. J. Imag. Syst. Technol.*, vol. 6, no. 4, pp. 305–313, 1995.

[35] A. K. Katsaggelos, K. T. Lay, and N. P. Galatsanos, "A general framework for frequency domain multi-channel signal processing," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 417–420, Jul. 1993.

[36] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, vol. 48, no. 3, pp. 259–302, 1986.

[37] B. D. Ripley, *Spatial Statistics*. New York: Wiley, 1981, pp. 88–90.

[38] R. H. Chan, T. F. Chan, and C.-K. Wong, "Cosine transform based preconditioners for total variation deblurring," *IEEE Trans. Image Process.*, vol. 8, no. 10, pp. 1472–1478, Oct. 1999.

[39] C. R. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Trans. Image Process.*, vol. 7, no. 6, pp. 813–824, Jun. 1998.

[40] T. F. Chan, N. Ng, A. Yau, and A. Yip, "Superresolution image reconstruction using fast inpainting algorithms," *Appl. Comput. Harmon. Anal.*, vol. 23, no. 1, pp. 3–24, July 2007.

[41] M. K. Ng, H. Shen, E. Y. Lam, and L. Zhang, "A total variation regularization based super-resolution reconstruction algorithm for digital video," *EURASIP J. Adv. Signal Process.*, no. 74585, 2007.

[42] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer-Verlag, 1999.

S. Derin Babacan (S'02) was born in Istanbul, Turkey, in 1981. He received the B.Sc. degree from Bogazici University, Istanbul, in 2004, and the M.Sc. degree from Northwestern University, Evanston, IL, in 2006, where he is currently pursuing the Ph.D. degree in the Department of Electrical Engineering and Computer Science.

He is a Research Assistant with the Image and Video Processing Laboratory, Northwestern University. His primary research interests include image restoration, image and video compression, super resolution, and computer vision.

Rafael Molina (M'87) was born in 1957. He received the degree in mathematics (statistics) in 1979 and the Ph.D. degree in optimal design in linear models in 1983.

He became Professor of computer science and artificial intelligence at the University of Granada, Granada, Spain, in 2000. His areas of research interest are image restoration (applications to astronomy and medicine), parameter estimation in image restoration, super resolution of images and video, and blind deconvolution. He is currently the Head of the computer science and Artificial Intelligence Department, University of Granada.

Aggelos K. Katsaggelos (S'80–M'85–SM'92–F'98) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1979, and the M.S. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1981 and 1985, respectively.

In 1985, he joined the Department of Electrical and Computer Engineering at Northwestern University, Evanston, IL, where he is currently a Professor. He was the holder of the Ameritech Chair of Information Technology (1997–2003). He is also the Director of the Motorola Center for Communications and a member of the Academic Affiliate Staff, Department of Medicine, Evanston Hospital. He is the editor of *Digital Image Restoration* (Springer-Verlag, 1991), coauthor of *Rate-Distortion Based Video Compression* (Kluwer, 1997), co-editor of *Recovery Techniques for Image and Video Compression and Transmission* (Kluwer, 1998), co-author of *Super-Resolution for Images and Video* (Claypool, 2007) and *Joint Source-Channel Video Transmission* (Claypool, 2007), and the co-inventor of 12 patents

Dr. Katsaggelos has served the IEEE and other Professional Societies in many capacities. He is currently a member of the Publication Board of the IEEE PROCEEDINGS and has served as Editor-in-Chief of the *IEEE Signal Processing Magazine* (1997–2002) and a member of the Board of Governors of the IEEE Signal Processing Society (1999–2001). He is the recipient of the IEEE Third Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), an IEEE Signal Processing Society Best Paper Award (2001), an IEEE International Conference on Multimedia and Expo Paper Award (2006), and an IEEE International Conference on Image Processing Paper Award (2007). He is a Distinguished Lecturer of the IEEE Signal Processing Society (2007–2008).