Variational Bayesian Blind Deconvolution Using a Total Variation Prior

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

Abstract—In this paper, we present novel algorithms for total variation (TV) based blind deconvolution and parameter estimation utilizing a variational framework. Using a hierarchical Bayesian model, the unknown image, blur, and hyperparameters for the image, blur, and noise priors are estimated simultaneously. A variational inference approach is utilized so that approximations of the posterior distributions of the unknowns are obtained, thus providing a measure of the uncertainty of the estimates. Experimental results demonstrate that the proposed approaches provide higher restoration performance than non-TV-based methods without any assumptions about the unknown hyperparameters.

Index Terms—Bayesian methods, blind deconvolution, parameter estimation, total variation (TV), variational methods.

I. INTRODUCTION

IMAGE acquisition systems introduce blurring degradation to the acquired image. In many applications, it is desired to undo this process. Blind deconvolution refers to a class of problems when the original image is estimated from the degraded observations where the exact information about the degradation and noise is not available. The blind deconvolution problem is very challenging since it is hard to infer the original image and the unknown degradation only from the observed image. Moreover, the degradation is generally nonlinear (due to saturation, quantization, etc.) and spatially varying (lens imperfections, nonuniform motion, etc). However, most of the work in the literature approximates the degradation process by a linear spatially invariant (LSI) system, where the original image is convolved by the blur point spread function (PSF) and independent white Gaussian noise is added to the blurred image.

There are many applications where the PSF is unknown or partially known, where blind deconvolution is needed, such as astronomical imaging, remote sensing, microscopy, medical imaging, optics, photography, super-resolution applications, and motion tracking applications, among others.

A number of methods have been proposed to address the blind deconvolution problem. Reviews of the major approaches can be found in [2] and [3]. Blind deconvolution methods can be classified into two main categories based on the stage where the blur is identified. In the *a priori* blur identification methods, the PSF is estimated separately from the original image, and then used in an image restoration method [4]. The second category of methods, referred to as joint identification methods, provide an estimate of the original image and blur simultaneously [5]–[11]. Typically, these methods incorporate prior knowledge about the original image, degradation, and noise in the estimation process. This prior knowledge can be exploited with the use of convex sets and regularization techniques, or with the use of the Bayesian framework with prior models on the unknown parameters.

Methods based on the Bayesian formulation are of the most commonly used methods in blind deconvolution. Such methods introduce prior models on the image, blur, and their model parameters, which impose constraints on the estimates and act as regularizers. Simultaneous autoregressive (SAR), conditional autoregressive (CAR), and Gaussian models are some of the commonly used priors for the image and blur. With the use of these mathematical models one can try to model different types of blurs, like out-of-focus, motion, or Gaussian blurs, and different characteristics of the original image, such as smoothness and sharp edges.

Recently, there has been an interest in applying variational methods to the blind deconvolution problem. These methods aim at obtaining approximations to the posterior distributions of the unknowns with the use of the Kullback–Leibler (KL) divergence [12]. This variational methodology to the blind deconvolution problem in a Bayesian formulation has been utilized in [6], [7], [13], and [14].

In this paper, we propose to use variational methods for the blind deconvolution problem by incorporating a total variation (TV) function as the image prior, and a SAR model as the blur prior. Although the TV model has been used in a regularization formulation in blind deconvolution before (see, for example, [5]), to the best of our knowledge, no work has been reported on the simultaneous estimation of the model parameters, image, and blur. Previous works attempted to solve for the unknown image and the blur, but the model parameters are manually selected [5], [15]. Moreover, we cast the TV-based blind deconvolution into a Bayesian estimation problem, which provides advantages in blind deconvolution, such as means to estimate

S. D. Babacan and A. K. Katsaggelos are with the Department of Electrical Engineering and Computer Science, Northwestern University, 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, Spain (e-mail: rms@decsai.ugr.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

the uncertainties of the estimates. We develop two novel variational methods based on the hierarchical Bayesian formulation, and provide approximations to the posterior distributions of the image, blur, and model parameters rather than point estimates.

This paper is organized as follows. In Section II, we present the hierarchical Bayesian model and the prior models on the observation, the image and the blur. Section III describes the variational approximation method utilized in the Bayesian inference. We present experimental results in Section IV, and conclusions are drawn in Section V.

II. HIERARCHICAL BAYESIAN MODELING

The image degradation model is often presented as a discrete linear and spatially invariant system, which can be expressed in matrix-vector form as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1}$$

where the 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 unknown block-circulant blurring matrix formed by the degradation system with impulse response $\mathbf{h}$. The images are of size $N = n \times m$ so that the vectors $\mathbf{y}$ and $\mathbf{x}$ are of size $N \times 1$ and the matrix $\mathbf{H}$ is of size $N \times N$. Note that (1) can also be written as $\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{n}$ by forming the matrix $\mathbf{X}$ similarly to $\mathbf{H}$. The blind deconvolution problem calls for finding estimates of $\mathbf{x}$ and $\mathbf{h}$ given $\mathbf{y}$, and using knowledge about $\mathbf{n}$ and possibly $\mathbf{x}$ and $\mathbf{h}$.

In Bayesian models, all unknown parameters are treated as stochastic quantities and probability distributions are assigned to them. The unknown parameters $\mathbf{x}$ and $\mathbf{h}$ are assigned *prior* distributions $\mathrm{p}(\mathbf{x}|\alpha_{\mathrm{im}})$ and $\mathrm{p}(\mathbf{h}|\alpha_{\mathrm{bl}})$, which model the knowledge about the nature of the original image and the blur, respectively. The observation $\mathbf{y}$ is also a random process with the corresponding *conditional* distribution $\mathrm{p}(\mathbf{y}|\mathbf{x}, \mathbf{h}, \beta)$. Clearly, these distributions depend on the model parameters α_{im}, α_{bl}, and β, which are called *hyperparameters*. The meaning of the hyperparameters will become clear when the prior distributions are defined below. In this paper, we will denote the set of hyperparameters as $\Omega = (\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta)$.

The Bayesian modeling of this problem firstly requires the definition of the joint probability distribution of all unknown and observed quantities, which is factorized as

$$\mathrm{p}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, \mathbf{x}, \mathbf{h}, \mathbf{y}) = \mathrm{p}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta)\mathrm{p}(\mathbf{x}|\alpha_{\mathrm{im}})$$
$$\times \mathrm{p}(\mathbf{h}|\alpha_{\mathrm{bl}})\mathrm{p}(\mathbf{y}|\mathbf{x}, \mathbf{h}, \beta). \tag{2}$$

To alleviate the ill-posed nature of the blind deconvolution problem, prior knowledge about the unknown image and the blur is incorporated through the use of the prior distributions. If the hyperparameters are not assumed known, they have to be estimated simultaneously with the unknown parameters. To achieve this we utilize a hierarchical model which has two steps: In the first step, the *a priori* probability distributions $\mathrm{p}(\mathbf{h}|\alpha_{\mathrm{bl}})$ and $\mathrm{p}(\mathbf{x}|\alpha_{\mathrm{im}})$ and the conditional distribution $\mathrm{p}(\mathbf{y}|\mathbf{x}, \mathbf{h}, \beta)$ are formed that model the structure of the PSF, the original image, and the noise, respectively. In the second stage, *hyperpriors* on

the hyperparameters β, α_{im} and α_{bl} are defined to model the prior knowledge of their values.

In the next subsections, we first describe the prior models for the image and the PSF as well as the observation model we use in the first stage of the hierarchical Bayesian paradigm. We then proceed to explain the hyperprior distributions on the hyperparameters.

A. *First Stage: Prior Models on the Observation, PSF and Image*

We assume that the degradation noise is independent and Gaussian with zero mean and variance equal to β^{-1}, and consequently we have

$$\mathrm{p}(\mathbf{y}|\mathbf{x}, h, \beta) \propto \beta^{N/2} \exp\left[-\frac{\beta}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2\right]. \tag{3}$$

For the image prior we adopt the TV function, that is

$$\mathrm{p}(\mathbf{x}|\alpha_{\mathrm{im}}) \propto \frac{1}{Z_{\mathrm{TV}}(\alpha_{\mathrm{im}})} \exp\left[-\alpha_{\mathrm{im}}\mathrm{TV}(\mathbf{x})\right] \tag{4}$$

where $Z_{\mathrm{TV}}(\alpha_{\mathrm{im}})$ is the partition function. The TV function is defined as

$$\mathrm{TV}(\mathbf{x}) = \sum_i \sqrt{\left(\Delta_i^h(\mathbf{x})\right)^2 + \left(\Delta_i^v(\mathbf{x})\right)^2} \tag{5}$$

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. In other words, $\Delta_i^h(\mathbf{x}) = x_i - x_{l(i)}$ and $\Delta_i^v(\mathbf{x}) = x_i - x_{a(i)}$, with $l(i)$ and $a(i)$ denoting the nearest horizontal and vertical neighbors of pixel i, respectively. The TV prior has become very popular recently in the restoration literature because of its edge-preserving property by not over-penalizing discontinuities in the image while imposing smoothness [16]. Note that the TV prior is an improper prior (see, for example, [17]), but if integrated in an adequate affine subspace (a hyperplane), the density is normalizable.

The calculation of the partition function $Z_{\mathrm{TV}}(\alpha_{\mathrm{im}}) = \int \exp[-\alpha_{\mathrm{im}}\mathrm{TV}(\mathbf{x})]d\mathbf{x}$ in (4) presents a major difficulty. We can, however, approximate it by using [18]

$$\int \int \exp[-\alpha_{\mathrm{im}}\sqrt{s^2 + t^2}]dsdt = 2\pi/\alpha_{\mathrm{im}}^2. \tag{6}$$

Therefore, the TV prior can be approximated as

$$\mathrm{p}(\mathbf{x}|\alpha_{\mathrm{im}}) = c\alpha_{\mathrm{im}}^{N/2} \exp\left[-\alpha_{\mathrm{im}}\mathrm{TV}(\mathbf{x})\right] \tag{7}$$

with c a constant.

We utilize the SAR model for the blur prior, that is

$$\mathrm{p}(\mathbf{h}|\alpha_{\mathrm{bl}}) \propto \alpha_{\mathrm{bl}}^{M/2} \exp\left\{-\frac{1}{2}\alpha_{\mathrm{bl}}\|\mathbf{C}\mathbf{h}\|^2\right\} \tag{8}$$

where $\mathbf{C}$ denotes the discrete Laplacian operator, $\alpha_{\mathrm{bl}}^{-1}$ is the variance of the Gaussian distribution, and M is the support of the blur, which is assumed to be the same as the image support. Note that in (8), M should in theory be replaced by $M-1$, because $\mathbf{C}^T\mathbf{C}$ is singular. The SAR model is very efficient in estimating smooth PSFs, for instance, a Gaussian PSF modeling long-term atmospheric turbulence. Our selection of the

SAR prior is based on the fact that we aim at restoring images which have been blurred with smoothly varying PSFs. As we will show in the experiments, for such PSFs the proposed prior works better than TV-based blur priors (e.g., [5] and [15]). On the other hand, a TV blur prior models better piecewise smooth priors such as the rectangular shaped and out of focus blurs. This is in agreement with the fact that TV models are better image priors than autoregressive models.

B. Second Stage: Hyperpriors on the Hyperparameters

The hyperparameters are important in determining the performance of the algorithms to a great extent. In most previous work, the hyperparameters are assumed known. However, this requires a significant amount of supervision in the restoration process. To ameliorate this problem, in this work they are assumed unknown and are simultaneously estimated by introducing a second stage in the Bayesian model.

Finding the form of the hyperprior distributions that allows for easy calculation of the posterior distribution $p(\Omega, \mathbf{x}, \mathbf{h} | \mathbf{y})$ is a major problem in Bayesian literature. A desired property for the hyperprior is to be conjugate [19], that is, to have the same functional form with the product $p(\mathbf{x} | \alpha_{\mathrm{im}}) p(\mathbf{h} | \alpha_{\mathrm{bl}}) p(\mathbf{y} | \mathbf{x}, \mathbf{h}, \beta)$, so that the posterior distribution will have the same functional form as the prior distribution, only the parameters will be updated by the sample information.

Based on the above, we utilize the Gamma distribution for the hyperparameters α_{im}, α_{bl} and β, since it is the conjugate prior for the inverse variance (precision) of the Gaussian distribution. The Gamma distribution is defined by

$$p(\omega) = \Gamma\left(\omega | a_\omega^o, b_\omega^o\right) = \frac{(b_\omega^o)^{-a_\omega^o}}{\Gamma(a_\omega^o)} \omega^{a_\omega^o - 1} \exp\left[-\frac{\omega}{b_\omega^o}\right] \quad (9)$$

where $\omega > 0$ denotes a hyperparameter, $b_\omega^o > 0$ is the scale parameter, and $a_\omega^o > 0$ is the shape parameter, both of which are assumed to be known and introduce our prior knowledge on the hyperparameters. We discuss the selection of the shape and scale parameters in the experimental section. The gamma distribution has the following mean, variance, and mode:

$$E[\omega] = a_\omega^o b_\omega^o, \ \ \mathrm{Var}[\omega] = a_\omega^o (b_\omega^o)^2, \ \ \mathrm{Mode}[\omega] = (a_\omega^o - 1) b_\omega^o. \quad (10)$$

Note that in addition to the advantage of the already described conjugacy property, the Gamma distribution allows for the incorporation of more vague or precise knowledge about the precision parameters. By simply replacing a_ω^o by $a_\omega^o \cdot \lambda$ and b_ω^o by b_ω^o / λ, another Gamma distribution with the same mean but with variance $a_\omega^o b_\omega^o \cdot \lambda$ can be obtained. Therefore, by varying λ we maintain the same mean of the precision parameter ω but can vary the confidence on this mean.

Finally, by combining the first and second stage of the hierarchical Bayesian model, the joint distribution in (2) can be defined. The dependencies in this joint probability model are shown in graphical form in Fig. 1 using a directed acyclic graph.
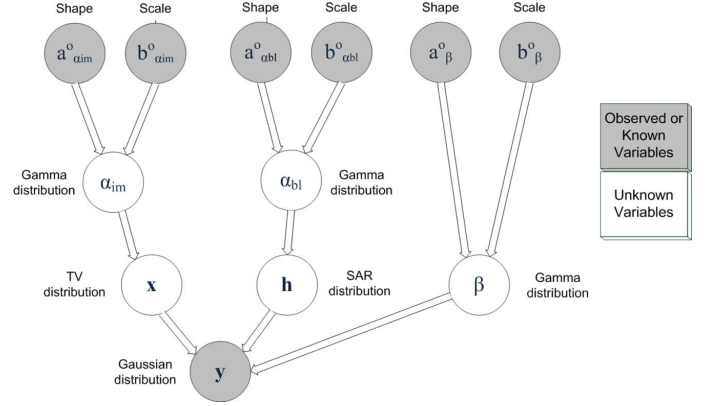

Fig. 1. Graphical model showing relationships between variables.

III. BAYESIAN INFERENCE AND VARIATIONAL APPROXIMATION OF THE POSTERIOR DISTRIBUTIONS

We will denote the set of all unknowns by $\Theta = (\Omega, \mathbf{x}, \mathbf{h}) = (\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, \mathbf{x}, \mathbf{h})$. As is widely known, Bayesian inference is based on the posterior distribution

$$p(\Theta | \mathbf{y}) = p(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, \mathbf{x}, \mathbf{h} | \mathbf{y}) = \frac{p(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, \mathbf{x}, \mathbf{h}, \mathbf{y})}{p(\mathbf{y})} \quad (11)$$

where $p(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, \mathbf{x}, \mathbf{h}, \mathbf{y})$ is given by (2). However, the posterior $p(\Theta | \mathbf{y})$ is intractable, since

$$p(\mathbf{y}) = \int \int \int \int \int p(\alpha, \beta, \mathbf{x}, \mathbf{h}, \mathbf{y}) d\mathbf{h} d\mathbf{x} d\beta d\alpha_{\mathrm{bl}} d\alpha_{\mathrm{im}} \quad (12)$$

cannot be calculated analytically. Therefore, we consider an approximation of $p(\Theta | \mathbf{y})$ by a simpler tractable distribution $q(\Theta)$ following the variational methodology [20]. The distribution $q(\Theta)$ will be found by minimizing the KL divergence, given by [12], [21]

$$\begin{aligned} C_{KL}\left(q(\Theta) \| p(\Theta | \mathbf{y})\right) &= \int q(\Theta) \log\left(\frac{q(\Theta)}{p(\Theta | \mathbf{y})}\right) d\Theta \\ &= \int q(\Theta) \log\left(\frac{q(\Theta)}{p(\Theta, \mathbf{y})}\right) d\Theta + \mathrm{const} \quad (13) \end{aligned}$$

which is always nonnegative and equal to zero only when $q(\Theta) = p(\Theta | \mathbf{y})$. In order to obtain a tractable approximation, the family of distributions $q(\Theta)$ are restricted utilizing the mean field approximation [22] so that $q(\Theta) = q(\Omega) q(\mathbf{x}) q(\mathbf{h})$, where $q(\Omega) = q(\alpha_{\mathrm{im}}) q(\alpha_{\mathrm{bl}}) q(\beta)$.

However, the use of the TV prior makes the integral in (13) difficult to evaluate even with this factorization. Therefore, a majorization of the TV prior is utilized to find an upper bound of the KL divergence. First, we define the following functional $M(\alpha_{\mathrm{im}}, \mathbf{x}, \mathbf{u})$, for α_{im}, $\mathbf{x}$, and any N-dimensional vector $\mathbf{u} \in (R^+)^N$

$$M(\alpha_{\mathrm{im}}, \mathbf{x}, \mathbf{u}) = \alpha_{\mathrm{im}}^{N/2}$$
$$\times \exp\left[-\frac{\alpha_{\mathrm{im}}}{2} \sum_i \frac{\left(\Delta_i^h(\mathbf{x})\right)^2 + \left(\Delta_i^v(\mathbf{x})\right)^2 + u_i}{\sqrt{u_i}}\right]. \quad (14)$$

Now, using the following inequality for $w \geq 0$ and $z > 0$

$$\sqrt{wz} \leq \frac{w+z}{2} \Rightarrow \sqrt{w} \leq \frac{w+z}{2\sqrt{z}} \qquad (15)$$

we obtain from (7)

$$\begin{aligned} &\exp\left[-\alpha_{\mathrm{im}} \mathrm{TV}(\mathbf{x})\right] \\ &= \exp\left[-\alpha_{\mathrm{im}} \sum_i \sqrt{\left(\Delta_i^h(\mathbf{x})\right)^2 + \left(\Delta_i^v(\mathbf{x})\right)^2}\right] \\ &\geq \exp\left[-\frac{\alpha_{\mathrm{im}}}{2} \sum_i \frac{\left(\Delta_i^h(\mathbf{x})\right)^2 + \left(\Delta_i^v(\mathbf{x})\right)^2 + u_i}{\sqrt{u_i}}\right] \end{aligned} \qquad (16)$$

which leads to the following lower bound for the image prior:

$$\mathrm{p}(\mathbf{x}|\alpha_{\mathrm{im}}) \geq \mathrm{c} \cdot \mathrm{M}(\alpha_{\mathrm{im}}, \mathbf{x}, \mathbf{u}) \qquad (17)$$

and the following lower bound for the joint probability distribution:

$$\begin{aligned} \mathrm{p}(\Theta, \mathbf{y}) &\geq \mathrm{c} \cdot \mathrm{p}(\Omega) \mathrm{M}(\alpha_{\mathrm{im}}, \mathbf{x}, \mathbf{u}) \mathrm{p}(\mathbf{h}|\alpha_{\mathrm{bl}}) \mathrm{p}(\mathbf{y}|\mathbf{x}, \mathbf{h}, \beta) \\ &= \mathrm{F}(\Theta, \mathbf{u}, \mathbf{y}). \end{aligned} \qquad (18)$$

For $\theta \in \{\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, \mathbf{x}, \mathbf{h}\}$ let us denote by Θ_θ, the subset of Θ with θ removed; for instance, if $\theta = \mathbf{x}$, $\Theta_{\mathbf{x}} = (\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, \mathbf{h})$. Then, utilizing the lower bound $\mathrm{F}(\Theta, \mathbf{u}, \mathbf{y})$ for the joint probability distribution in (13) we obtain an upper bound for the KL divergence as follows:

$$\begin{aligned} \mathcal{M}(\mathrm{q}(\Theta)) &= \int \mathrm{q}(\Theta) \log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta, \mathbf{y})}\right) d\Theta \\ &\leq \int \mathrm{q}(\theta) \left(\int \mathrm{q}(\Theta_\theta) \log\left(\frac{\mathrm{q}(\theta)\mathrm{q}(\Theta_\theta)}{\mathrm{F}(\Theta, \mathbf{u}, \mathbf{y})}\right) d\Theta_\theta\right) d\theta \\ &= \tilde{\mathcal{M}}(\mathrm{q}(\Theta), u). \end{aligned} \qquad (19)$$

Therefore, we minimize this upper bound instead of minimizing the KL divergence in (13). Note that the form of the inequality in (19) suggests an alternating (cyclic) optimization strategy where the algorithm cycles through the unknown distributions and replaces each with a revised estimate given by the minimum of (19) with the other distributions held constant. Thus, given $\mathrm{q}(\Theta_\theta)$, the posterior approximation $\mathrm{q}(\theta)$ can be computed by solving

$$\hat{\mathrm{q}}(\theta) = \arg\min_{\mathrm{q}(\theta)} C_{KL}\left(\mathrm{q}(\Theta_\theta)\mathrm{q}(\theta)\|\mathrm{F}(\Theta, \mathbf{u}, \mathbf{y})\right). \qquad (20)$$

In order to solve this equation, we note that differentiating the integral on the right hand side in (19) with respect to $\mathrm{q}(\theta)$ results in (see [23, Eq. (2.28)])

$$\hat{\mathrm{q}}(\theta) = \mathrm{const} \times \exp\left(\mathrm{E}_{\mathrm{q}(\Theta_\theta)}\left[\log \mathrm{F}(\Theta, \mathbf{u}, \mathbf{y})\right]\right) \qquad (21)$$

where

$$\mathrm{E}_{\mathrm{q}(\Theta_\theta)}\left[\log \mathrm{F}(\Theta, \mathbf{u}, \mathbf{y})\right] = \int \log \mathrm{F}(\Theta, \mathbf{u}, \mathbf{y})\mathrm{q}(\Theta_\theta)d\Theta_\theta.$$

We obtain the following iterative procedure to find $\mathrm{q}(\Theta)$ by applying this minimization to each unknown in an alternating way:

Algorithm 1

Given $\mathrm{q}^1(\mathbf{h})$, $\mathrm{q}^1(\alpha_{\mathrm{im}})$, $\mathrm{q}^1(\alpha_{\mathrm{bl}})$, and $\mathrm{q}^1(\beta)$, initial estimates of the distributions $\mathrm{q}(\mathbf{h})$, $\mathrm{q}(\alpha_{\mathrm{im}})$, $\mathrm{q}(\alpha_{\mathrm{bl}})$, and $\mathrm{q}(\beta)$,

for $k = 1, 2, \ldots$ until a stopping criterion is met:
  1) Find

$$\begin{aligned} \mathrm{q}^k(\mathbf{x}) = \arg\min_{\mathrm{q}(\mathbf{x})} \int\int \mathrm{q}^k(\Theta_{\mathbf{x}})\mathrm{q}(\mathbf{x}) \\ \times \log\left(\frac{\mathrm{q}^k(\Theta_{\mathbf{x}})\mathrm{q}(\mathbf{x})}{\mathrm{F}(\Theta_{\mathbf{x}}^k, \mathbf{x}, \mathbf{u}^k, \mathbf{y})}\right) d\Theta_{\mathbf{x}}d\mathbf{x}. \end{aligned} \qquad (22)$$

  2) Find

$$\begin{aligned} \mathrm{q}^{k+1}(\mathbf{h}) = \arg\min_{\mathrm{q}(\mathbf{h})} \int\int \mathrm{q}^k(\Theta_{\mathbf{h}})\mathrm{q}(\mathbf{h}) \\ \times \log\left(\frac{\mathrm{q}^k(\Theta_{\mathbf{h}})\mathrm{q}(\mathbf{h})}{\mathrm{F}(\Theta_{\mathbf{h}}^k, \mathbf{h}, \mathbf{u}^k, \mathbf{y})}\right) d\Theta_{\mathbf{h}}d\mathbf{h}. \end{aligned} \qquad (23)$$

  3) Find

$$\begin{aligned} \mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} \int \mathrm{q}^k(\Theta_{\mathbf{h}})\mathrm{q}^{k+1}(\mathbf{h}) \\ \times \log\left(\frac{\mathrm{q}^k(\Theta_{\mathbf{h}})\mathrm{q}^{k+1}(\mathbf{h})}{\mathrm{F}(\Theta_{\mathbf{h}}^k, \mathbf{h}^{k+1}, \mathbf{u}, \mathbf{y})}\right) d\Theta. \end{aligned} \qquad (24)$$

  4) Find

$$\begin{aligned} \mathrm{q}^{k+1}(\Omega) = \arg\min_{\mathrm{q}(\Omega)} \int\int \mathrm{q}^k(\Theta_\Omega)\mathrm{q}(\Omega) \\ \times \log\left(\frac{\mathrm{q}^k(\Theta_\Omega)\mathrm{q}(\Omega)}{\mathrm{F}(\Theta_\Omega^k, \Omega, \mathbf{u}^k, \mathbf{y})}\right) d\Theta_\Omega d\Omega. \end{aligned} \qquad (25)$$

Now we proceed to state the solutions at each step of the algorithm [(22)–(25)] explicitly. For simplicity, we will use the following notations $\mathrm{E}^k(\mathbf{x}) = \mathrm{E}_{\mathrm{q}^k(\mathbf{x})}(\mathbf{x})$, $\mathrm{cov}^k(\mathbf{x}) = \mathrm{cov}_{\mathrm{q}^k(\mathbf{x})}(\mathbf{x})$, $\mathrm{E}^k(\mathbf{h}) = \mathrm{E}_{\mathrm{q}^k(\mathbf{h})}[\mathbf{h}]$, $\mathrm{E}^k(\mathbf{H}) = \mathrm{E}_{\mathrm{q}^k(\mathbf{h})}(\mathbf{H})$, $\mathrm{cov}^k(\mathbf{h}) = \mathrm{cov}_{\mathrm{q}^k(\mathbf{h})}(\mathbf{h})$, $\mathrm{E}^k(\alpha_{\mathrm{im}}) = \mathrm{E}_{\mathrm{q}^k(\alpha_{\mathrm{im}})}(\alpha_{\mathrm{im}})$, $\mathrm{E}^k(\alpha_{\mathrm{bl}}) = \mathrm{E}_{\mathrm{q}^k(\alpha_{\mathrm{bl}})}(\alpha_{\mathrm{bl}})$ and $\mathrm{E}^k(\beta) = \mathrm{E}_{\mathrm{q}^k(\beta)}(\beta)$.

From (21), it can be shown that $\mathrm{q}^k(\mathbf{x})$ is an N-dimensional Gaussian distribution, rewritten as

$$\mathrm{q}^k(\mathbf{x}) = \mathcal{N}\left(\mathbf{x}|\mathrm{E}^k(\mathbf{x}), \mathrm{cov}^k(\mathbf{x})\right).$$

The covariance and mean of this normal distribution can be calculated from (22) as

$$\begin{aligned} \mathrm{cov}^k(\mathbf{x}) = \big(&\mathrm{E}^k(\beta)\mathrm{E}^k(\mathbf{H})^t\mathrm{E}^k(\mathbf{H}) + \mathrm{E}^k(\alpha_{\mathrm{im}})(\Delta^h)^t W(\mathbf{u}^k)(\Delta^h) \\ &+ \mathrm{E}^k(\alpha_{\mathrm{im}})(\Delta^v)^t W(\mathbf{u}^k)(\Delta^v) + N\mathrm{E}^k(\beta)\mathrm{cov}^k(\mathbf{h})\big)^{-1} \end{aligned} \qquad (26)$$

$$\mathrm{E}^k(\mathbf{x}) = \mathrm{cov}^k(\mathbf{x})\mathrm{E}^k(\beta)\mathrm{E}^k(\mathbf{H})^t\mathbf{y} \qquad (27)$$

where $(\cdot)^t$ is the transpose and $W(\mathbf{u})$ is the $N \times N$ diagonal matrix of the form

$$W(\mathbf{u}) = \text{diag}\left(\frac{1}{\sqrt{u_i^k}}\right), \quad i = 1, \ldots, N. \qquad (28)$$

Similarly to $\text{q}^k(\mathbf{x})$, $\text{q}^k(\mathbf{h})$ is an M-dimensional Gaussian distribution, given by

$$\text{q}^{k+1}(\mathbf{h}) = \mathcal{N}\left(\mathbf{h}|\text{E}^{k+1}(\mathbf{h}), \text{cov}^{k+1}(\mathbf{h})\right) \qquad (29)$$

with

$$\text{cov}^{k+1}(\mathbf{h}) = \left(\text{E}^k(\alpha_{\text{bl}})\mathbf{C}^t\mathbf{C} + \text{E}^k(\beta)\text{E}^k(\mathbf{X})^t\text{E}^k(\mathbf{X})\right.$$
$$\left. + N\text{E}^k(\beta)\text{cov}^k(\mathbf{x})\right)^{-1} \qquad (30)$$

and

$$\text{E}^{k+1}(\mathbf{h}) = \text{cov}^{k+1}(\mathbf{h})\text{E}^k(\beta)\text{E}^{k+1}(\mathbf{X})^t\mathbf{y}. \qquad (31)$$

It is worth emphasizing here that we did not assume *a priori* that $\text{q}^k(\mathbf{x})$ and $\text{q}^k(\mathbf{h})$ are Gaussian distributions. This result is derived due to the minimization of the KL divergence with respect to all possible distributions according to the factorization $\text{q}(\Theta) = \text{q}(\alpha_{\text{im}})\text{q}(\alpha_{\text{bl}})\text{q}(\beta)\text{q}(\mathbf{x})\text{q}(\mathbf{h})$ [24]. Note also that the image estimate in (27) is very similar to the image estimate proposed in [5] within a regularization framework; the uncertainty term $N\beta^k\text{cov}_{\text{q}^k(\mathbf{h})}[\mathbf{h}]$ is missing, however, in [5]. As we will see in the experimental results, this formulation will provide improved restoration results.

In step 4 of the algorithm, we find $\mathbf{u}^{k+1}$ from (24), given by

$$\mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} \sum_i \frac{\text{E}_{\text{q}^k(\mathbf{x})}\left[\left(\Delta_i^h(\mathbf{x})\right)^2 + \left(\Delta_i^v(\mathbf{x})\right)^2\right] + u_i}{\sqrt{u_i}}. \qquad (32)$$

Therefore, $\mathbf{u}^{k+1}$ can be obtained as

$$\mathbf{u}_i^{k+1} = \text{E}_{\text{q}^k(\mathbf{x})}\left[\left(\Delta_i^h(\mathbf{x})\right)^2 + \left(\Delta_i^v(\mathbf{x})\right)^2\right], \quad i = 1, \ldots, N \qquad (33)$$

where

$$\text{E}_{\text{q}^k(\mathbf{x})}\left[\left(\Delta_i^h(\mathbf{x})\right)^2 + \left(\Delta_i^v(\mathbf{x})\right)^2\right]$$
$$= \left(\Delta_i^h\left(\text{E}^k(\mathbf{x})\right)\right)^2 + \left(\left(\Delta_i^v\left(\text{E}^k(\mathbf{x})\right)\right)^2\right.$$
$$+ \text{E}_{\text{q}^k(\mathbf{x})}\left[\left(\Delta_i^h\left(\mathbf{x} - \text{E}^k(\mathbf{x})\right)\right)^2\right]$$
$$+ \text{E}_{\text{q}^k(\mathbf{x})}\left[\left(\Delta_i^v\left(\mathbf{x} - \text{E}^k(\mathbf{x})\right)\right)^2\right] \qquad (34)$$

and

$$\text{E}_{\text{q}^k(\mathbf{x})}\left[\left(\Delta_i^h\left(\mathbf{x} - \text{E}^k(\mathbf{x})\right)\right)^2\right] + \text{E}_{\text{q}^k(\mathbf{x})}\left[\left(\Delta_i^v\left(\mathbf{x} - \text{E}^k(\mathbf{x})\right)\right)^2\right]$$
$$= \frac{1}{N}\text{trace}\left[\text{cov}^k(\mathbf{x}) \times \left((\Delta^h)^t(\Delta^h) + (\Delta^v)^t(\Delta^v)\right)\right]. \qquad (35)$$

It is clear that the vector $\mathbf{u}^{k+1}$ in (33) represents the local spatial activity of $\mathbf{x}$ using its distribution approximation $\text{q}^k(\mathbf{x})$.

Consequently, the matrix $W(\mathbf{u}^{k+1})$ in (28) is the *spatial adaptivity matrix* at iteration $k + 1$ and it controls the smoothing applied to the unknown image in the restoration process. For instance, at strong edges where the vector $\mathbf{u}^{k+1}$ has large values, $W(\mathbf{u}^{k+1})$ will have small values so the amount of smoothing is decreased. On the other hand, in smooth regions the corresponding entry of $W(\mathbf{u}^{k+1})$ will be very high, so smoothness is enforced. This property of the restoration is also in accordance with the fact that noise is perceived as more visible in uniform regions than at edges (masking effect of the human visual system). Note also that the spatial adaptivity matrix is also referred to as the visibility matrix [25] and has been utilized in some image restoration approaches (see, for instance, [26] and [27]).

After finding estimates of the posterior distributions of the image and blur, we find the estimates for the hyperpriors at the last step of the algorithm. For $\omega \in \{\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta\}$, evaluating (25) using (21) results in

$$\text{q}^{k+1}(\omega) \propto \exp \text{E}_{\text{q}^k(\mathbf{x})\text{q}^{k+1}(\mathbf{h})\text{q}(\Omega_\omega)}$$
$$\left[\log \text{F}\left(\Omega_\omega^k, \omega, \mathbf{x}^k, \mathbf{h}^{k+1}, \mathbf{u}^{k+1}, \mathbf{y}\right)\right].$$

Evaluating this explicitly, we obtain

$$\text{E}_{\text{q}^k(\mathbf{x})\text{q}^{k+1}(\mathbf{h})}\left[\log \text{F}(\Theta)\right]$$
$$= \text{const} + \sum_{\omega \in \{\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta\}} \left((a_\omega^o - 1)\log \omega - \omega/b_\omega^o\right)$$
$$+ \frac{N}{2}\log \alpha_{\text{im}} + \frac{M}{2}\log \alpha_{\text{bl}} + \frac{N}{2}\log \beta$$
$$- \frac{1}{2}\alpha_{\text{im}}\text{E}_{\text{q}^k(\mathbf{x})}\left[\sum_i \frac{\left(\Delta_i^h(\mathbf{x})\right)^2 + \left(\Delta_i^v(\mathbf{x})\right)^2 + u_i^{k+1}}{\sqrt{u_i^{k+1}}}\right]$$
$$- \frac{1}{2}\alpha_{\text{bl}}\text{E}_{\text{q}^{k+1}(\mathbf{h})}\left[\|\mathbf{Ch}\|^2\right]$$
$$- \frac{1}{2}\beta\text{E}_{\text{q}^k(\mathbf{x})\text{q}^{k+1}(\mathbf{h})}\left[\|\mathbf{y} - \mathbf{Hx}\|^2\right] \qquad (36)$$

where

$$\text{E}_{\text{q}^k(\mathbf{x})}\left[\sum_i \frac{\left(\Delta_i^h(\mathbf{x})\right)^2 + \left(\Delta_i^v(\mathbf{x})\right)^2 + u_i^{k+1}}{\sqrt{u_i^{k+1}}}\right] = 2\sum_i \sqrt{u_i^{k+1}} \qquad (37)$$

$$\text{E}_{\text{q}^{k+1}(\mathbf{h})}\left[\|\mathbf{Ch}\|^2\right] = \left\|\mathbf{C}\text{E}^{k+1}(\mathbf{h})\right\|^2 + \text{trace}\left(\mathbf{C}^t\mathbf{C}\text{cov}^k(\mathbf{h})\right) \qquad (38)$$

and

$$\text{E}_{\text{q}^k(\mathbf{x})\text{q}^{k+1}(\mathbf{h})}\left[\|\mathbf{y} - \mathbf{Hx}\|^2\right]$$
$$= \left\|\mathbf{y} - \text{E}^{k+1}(\mathbf{h})\text{E}^k(\mathbf{x})\right\|^2 + \text{trace}\left(N\text{cov}^k(\mathbf{x})\text{cov}^{k+1}(\mathbf{h})\right)$$
$$+ \text{trace}\left(\text{E}^k(\mathbf{X})^t\text{E}^k(\mathbf{X})\text{cov}^{k+1}(\mathbf{h})\right)$$
$$+ \text{trace}\left(\text{E}^{k+1}(\mathbf{H})^t\text{E}^{k+1}(\mathbf{H})\text{cov}^k(\mathbf{x})\right). \qquad (39)$$

It can be seen from (36) that all hyperparameters have gamma distributions, given by

$$q^{k+1}(\alpha_{\text{im}}) \propto \alpha_{\text{im}}^{N/2+a_{\alpha_{\text{im}}}^o - 1}$$
$$\times \exp\left[-\alpha_{\text{im}}\left(1/b_{\alpha_{\text{im}}}^o + \sum_i \sqrt{u_i^{k+1}}\right)\right] \quad (40)$$

$$q^{k+1}(\alpha_{\text{bl}}) \propto \alpha_{\text{bl}}^{M/2+a_{\alpha_{\text{bl}}}^o - 1}$$
$$\times \exp\left[-\alpha_{\text{bl}}\left(1/b_{\alpha_{\text{bl}}}^o + \frac{\mathrm{E}_{q^{k+1}(\mathbf{h})}\left[\|\mathbf{Ch}\|^2\right]}{2}\right)\right] \quad (41)$$

$$q^{k+1}(\beta) \propto \beta^{N/2+a_\beta^o - 1}$$
$$\times \exp\left[-\beta\left(1/b_\beta^o + \frac{\mathrm{E}_{q^k(\mathbf{x})q^{k+1}(\mathbf{h})}\left[\|\mathbf{y}-\mathbf{Hx}\|^2\right]}{2}\right)\right]. \quad (42)$$

The means of these gamma distributions can be found using (10) and are represented as follows:

$$\left(\mathrm{E}^{k+1}(\alpha_{\text{im}})\right)^{-1} = \gamma_{\alpha_{\text{im}}}\frac{1}{\overline{\alpha}_{\text{im}}^o} + (1-\gamma_{\alpha_{\text{im}}})\frac{\sum_i\sqrt{u_i^{k+1}}}{N/2} \quad (43)$$

$$\left(\mathrm{E}^{k+1}(\alpha_{\text{bl}})\right)^{-1} = \gamma_{\alpha_{\text{bl}}}\frac{1}{\overline{\alpha}_{\text{bl}}^o} + (1-\gamma_{\alpha_{\text{bl}}})$$
$$\times \frac{\mathrm{E}_{q^{k+1}(\mathbf{h})}\left[\|\mathbf{Ch}\|^2\right]}{M} \quad (44)$$

$$\left(\mathrm{E}^{k+1}(\beta)\right)^{-1} = \gamma_\beta\frac{1}{\overline{\beta}^o} + (1-\gamma_\beta)$$
$$\times \frac{\mathrm{E}_{q^k(\mathbf{x})q^{k+1}(\mathbf{h})}\left[\|\mathbf{y}-\mathbf{Hx}\|^2\right]}{N} \quad (45)$$

where $\overline{\alpha}_{\text{im}}^o = a_{\alpha_{\text{im}}}^o/b_{\alpha_{\text{im}}}^o$, $\overline{\alpha}_{\text{bl}} = a_{\alpha_{\text{bl}}}^o/b_{\alpha_{\text{bl}}}^o$ and $\overline{\beta}^o = a_\beta^o/b_\beta^o$ and

$$\gamma_{\alpha_{\text{im}}} = \frac{a_{\alpha_{\text{im}}}^o}{a_{\alpha_{\text{im}}}^o + \frac{N}{2}}, \quad \gamma_{\alpha_{\text{bl}}} = \frac{a_{\alpha_{\text{bl}}}^o}{a_{\alpha_{\text{bl}}}^o + \frac{M}{2}}, \quad \gamma_\beta = \frac{a_\beta^o}{a_\beta^o + \frac{N}{2}}. \quad (46)$$

The parameters $\gamma_{\alpha_{\text{im}}}, \gamma_{\alpha_{\text{bl}}}$, and γ_β can be understood as normalized confidence parameters, as can be seen from (43)–(46) and they take values in the interval [0,1]. Therefore, the means of the posterior distributions of the hyperparameters are convex combinations of the prior hyperparameter values and their maximum likelihood (ML) estimates. When the confidence parameters are asymptotically equal to zero no confidence is placed on the initial values of the hyperparameters, and their ML estimates are used. On the other hand, a value asymptotically equal to one will result in no update on the hyperparameters, so that the algorithm will fully rely on the given initial parameters. In this case, no estimation of the hyperparameters is performed.

In Algorithm 1, no assumptions were imposed on the posterior approximations q(**x**) and q(**h**). We can, however, assume that these distributions are *degenerate*, i.e., distributions which take one value with probability one and the rest of the values with probability zero. We can obtain another algorithm under this assumption which is similar to algorithm 1. In this second algorithm, the value of the KL divergence is again decreased at each update step, but not by the maximum possible amount as was the case in algorithm 1.

Utilizing the fact that the distributions on **x** and **h** are degenerate, that is

$$q(\mathbf{x}) = \delta(\mathbf{x}-\underline{\mathbf{x}}) = \begin{cases} 1, & \text{if } \mathbf{x} = \underline{\mathbf{x}} \\ 0, & \text{otherwise} \end{cases} \quad (47)$$

$$q(\mathbf{h}) = \delta(\mathbf{h}-\underline{\mathbf{h}}) = \begin{cases} 1, & \text{if } \mathbf{h} = \underline{\mathbf{h}} \\ 0, & \text{otherwise} \end{cases} \quad (48)$$

with $\delta(\cdot)$ the delta function, we obtain the following algorithm 2, where we use $\mathbf{x}^k$ and $\mathbf{h}^k$ to denote the values $q^k(\mathbf{x})$ and $q^k(\mathbf{h})$ take with probability one, respectively, that is, $q^k(\mathbf{x}) = \delta(\mathbf{x}-\mathbf{x}^k)$ and $q^k(\mathbf{h}) = \delta(\mathbf{h}-\mathbf{h}^k)$.

Algorithm 2

Given $q^1(\mathbf{h})$, $q^1(\alpha_{\text{im}})$, $q^1(\alpha_{\text{bl}})$, and $q^1(\beta)$ the initial estimates of the distributions $q(\mathbf{h})$, $q(\alpha_{\text{im}})$, $q(\alpha_{\text{bl}})$ and $q(\beta)$, with $q^1(\mathbf{h})$ a degenerate distribution on $\mathbf{h}^1$,

for $k = 1, 2, \ldots$ until a stopping criterion is met:

1) Calculate

$$\mathbf{x}^k = \left(\mathrm{E}^k(\beta)(\mathbf{H}^k)^t\mathbf{H}^k + \mathrm{E}^k(\alpha_{\text{im}})(\Delta^h)^t W(\mathbf{u}^k)(\Delta^h)\right.$$
$$\left. + \mathrm{E}^k(\alpha_{\text{im}})(\Delta^v)^t W(\mathbf{u}^k)(\Delta^v)\right)^{-1} \mathrm{E}^k(\beta)(\mathbf{H}^k)^t\mathbf{y}. \quad (49)$$

2) Calculate

$$\mathbf{h}^k = \left(\mathrm{E}^k(\alpha_{\text{bl}})\mathbf{C}^t\mathbf{C} + \mathrm{E}^k(\beta)(\mathbf{X}^k)^t\mathbf{X}^k\right)^{-1} \mathrm{E}^k(\beta)(\mathbf{X}^k)^t\mathbf{x}. \quad (50)$$

3) Calculate

$$\mathbf{u}_i^{k+1} = \left(\Delta_i^h(\mathbf{x}^k)\right)^2 + \left(\Delta_i^v(\mathbf{x}^k)\right)^2, \quad i = 1, \ldots, N. \quad (51)$$

4) Calculate

$$q^{k+1}(\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta) = q^{k+1}(\alpha_{\text{im}})q^{k+1}(\alpha_{\text{bl}})q^{k+1}(\beta) \quad (52)$$

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

$$q^{k+1}(\alpha_{\text{im}}) \propto \alpha_{\text{im}}^{N/2+a_{\alpha_{\text{im}}}^o - 1}$$
$$\times \exp\left[-\alpha_{\text{im}}\left(1/b_{\alpha_{\text{im}}}^o + \sum_i\sqrt{u_i^{k+1}}\right)\right] \quad (53)$$

$$q^{k+1}(\alpha_{\text{bl}}) \propto \alpha_{\text{bl}}^{M/2+a_{\alpha_{\text{bl}}}^o - 1}$$
$$\times \exp\left[-\alpha_{\text{bl}}\left(1/b_{\alpha_{\text{bl}}}^o + \frac{\|\mathbf{Ch}^k\|^2}{2}\right)\right] \quad (54)$$

$$q^{k+1}(\beta) \propto \beta^{N/2+a_\beta^o - 1}$$
$$\times \exp\left[-\beta\left(1/b_\beta^o + \frac{\|\mathbf{y}-\mathbf{H}^k\mathbf{x}^k\|^2}{2}\right)\right]. \quad (55)$$

Set

$$q(\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta) = \lim_{k\to\infty} q^k(\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta),$$
$$\hat{\mathbf{x}} = \lim_{k\to\infty}\mathbf{x}^k, \hat{\mathbf{h}} = \lim_{k\to\infty}\mathbf{h}^k. \quad (56)$$

The update equations for the inverse of the means of the hyperparameters are then obtained from (53)–(55) as follows:

$$\left(\mathrm{E}^{k+1}(\alpha_{\mathrm{im}})\right)^{-1} = \gamma_{\alpha_{\mathrm{im}}}\frac{1}{\overline{\alpha}_{\mathrm{im}}^{o}} + (1 - \gamma_{\alpha_{\mathrm{im}}})\frac{\sum_i \sqrt{u_i^{k+1}}}{N/2} \quad (57)$$

$$\left(\mathrm{E}^{k+1}(\alpha_{\mathrm{bl}})\right)^{-1} = \gamma_{\alpha_{\mathrm{bl}}}\frac{1}{\overline{\alpha}_{\mathrm{bl}}^{o}} + (1 - \gamma_{\alpha_{\mathrm{bl}}})\frac{\|\mathbf{C}\mathbf{h}^k\|^2}{M} \quad (58)$$

$$\left(\mathrm{E}^{k+1}(\beta)\right)^{-1} = \gamma_{\beta}\frac{1}{\overline{\beta}^{o}} + (1 - \gamma_{\beta})\frac{\|\mathbf{y} - \mathbf{H}^k\mathbf{x}^k\|^2}{N}. \quad (59)$$

It is clear that using degenerate distributions for $\mathbf{x}$ and $\mathbf{h}$ in Algorithm 2 removes the uncertainty terms of the image and blur estimates. We will show in the experimental results section that these uncertainty terms (the covariances of $\mathbf{x}$ and $\mathbf{h}$) help to improve the restoration performance in high-noise cases, where the image and blur estimates can be poor. The poor estimation of one variable can influence the estimation of the other unknowns because of the alternating optimization procedure, and the overall performance of the algorithm will be affected. By estimating the full posterior distribution instead of the points corresponding to the maximum probability, the uncertainty of the estimates can be used to ameliorate these effects in the estimation of the unknowns. On the other hand, at low-noise cases where the estimates of the unknowns are more precise, Algorithm 2 results in better restorations.

Summarizing, Algorithm 1 iterates between (27), (31), (33), and (43)–(45), whereas Algorithm 2 iterates between (49), (50), (51) and (57)–(59) until convergence. Finally, a few remarks are needed for the calculation of the image and blur estimates. The blur estimates in (31) and (50) can be calculated by assuming block circulant with circulant sub-matrices (BCCB) matrices for $\mathbf{X}$ and $\mathbf{C}$, and finding the solutions in the Fourier domain, which is very efficient [28]. However, finding closed form solutions for the systems in (27) and (49) is practically very difficult because the BCCB assumption is not valid due to $\mathbf{W}$, and the high dimensionality of the matrices makes it hard to find the inverses. Therefore, we find numerical solutions by using a gradient descent (GD) algorithm which is very similar to the one proposed in [26] with small modifications. Other numerical techniques, such as conjugate gradient, can also be employed. Note that improved convergence and speed can be achieved by utilizing preconditioning methods (see, for example, [29] and [30]).

However, note that $\mathrm{cov}^k(\mathbf{x})$ is explicitly needed to calculate the quantities u_i^{k+1}, $\mathrm{cov}^k(\mathbf{h})$, and $\mathrm{E}_{\mathrm{q}^k(\mathbf{x})\mathrm{q}^{k+1}(\mathbf{h})}[\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2]$ in Algorithm 1. Since calculating this matrix is computationally very inefficient, we utilize an approximation to this inverse which is proposed for the image restoration problem in [31], where $W(\mathbf{u}^k)$ in (26) is replaced by $z(\mathbf{u}^k)\mathbf{I}$ with $z(\mathbf{u}^k)$ being the mean value of the diagonal values in $W(\mathbf{u}^k)$. Specifically

$$\begin{aligned}\mathrm{cov}^k(\mathbf{x}) &\approx \big(\mathrm{E}^k(\beta)\mathrm{E}^k(\mathbf{H})^t\mathrm{E}^k(\mathbf{H}) + \mathrm{E}^k(\alpha_{\mathrm{im}})z(\mathbf{u}^k)(\Delta^h)^t(\Delta^h)\\ &\quad + \mathrm{E}^k(\alpha_{\mathrm{im}})z(\mathbf{u}^k)(\Delta^v)^t(\Delta^v)\big)^{-1}\\ &= \mathbf{B}^{-1}.\end{aligned} \quad (60)$$

With this approximation matrix, $\mathbf{B}$ becomes a BCCB matrix, thus, computing its inverse can be performed in the Fourier domain. We, therefore, replace $\mathrm{cov}^k(\mathbf{x})$ by $\mathbf{B}^{-1}$ in (30), (35), and

(39). However, it should be noted that (60) is only used to calculate trace matrix values and in (30). Our experiments indicate that these values are very small when compared to other terms, so the effect of this approximation is negligible. Note also that this approximation is used only in Algorithm 1 and it is not needed in Algorithm 2.

We conclude this section by analyzing two very important issues: the convergence of the proposed algorithms and the closeness of the posterior approximations provided by algorithms 1 and 2 to the true posterior distribution of the unknown hyperparameters, image, and blur.

First we note that inequalities (15) and (16) provide the basis for approximating the TV prior image distribution by a Gaussian distribution in order to carry out analytical (not based on simulation) Bayesian analysis. From (19), we have

$$\mathcal{M}\left(\mathrm{q}(\Theta)\right) \leq \tilde{\mathcal{M}}\left(\mathrm{q}(\Theta), \mathbf{u}\right) \quad (61)$$

and so algorithms 1 and 2 provide a sequence of distributions $\{\mathrm{q}^k(\Theta)\}$ and a sequence of vectors $\{\mathbf{u}^k\}$ that satisfy

$$\tilde{\mathcal{M}}\left(\mathrm{q}^k(\Theta), \mathbf{u}^k\right) \geq \tilde{\mathcal{M}}\left(\mathrm{q}^k(\Theta), \mathbf{u}^{k+1}\right) \geq \tilde{\mathcal{M}}\left(\mathrm{q}^{k+1}(\Theta), \mathbf{u}^{k+1}\right). \quad (62)$$

Notice that when we decrease the value of the posterior approximation $\tilde{\mathcal{M}}(\mathrm{q}(\Theta), \mathbf{u}^{k+1})$, we obtain $\mathrm{q}^{k+1}(\Theta)$ which provides a decreased upper bound of $\mathcal{M}(\mathrm{q}^{k+1}(\Theta))$. Furthermore, minimizing $\tilde{\mathcal{M}}(\mathrm{q}^k(\Theta), \mathbf{u})$ with respect to $\mathbf{u}$ generates a new vector $\mathbf{u}^k$ that tightens the upper-bound of $\mathcal{M}(\mathrm{q}^k(\Theta))$. Consequently, algorithms 1 and 2 provide sequences of ever decreasing upper bounds. These sequences are bounded from below by $-\log \mathrm{p}(\mathbf{y})$ [see (13)], and, consequently, they converge.

Let us now examine the quality of the estimated posterior distributions. We only analyze here the type of the posterior distribution approximation obtained by algorithm 1; the discussion about algorithm 2 is very similar since in the iterative procedure we only use the mean and do not take into account its uncertainty. Inequality (15) provides a local quadratic approximation to the TV prior. Using always $\mathbf{u}^o$ with all its entries being equal is equivalent to utilizing a fixed global conditional auto-regression model to approximate the TV image prior. Clearly, the procedure which updates $\mathbf{u}$ (even if all its components are the same) will provide a tighter upper bound for $\mathcal{M}(\mathrm{q}(\Theta))$.

Let us also comment on the proximity of the estimated posterior distributions to the true posteriors. By using a different majorization of $\mathrm{TV}(\mathbf{x})$ from the one used in inequality (15), we obtain different approximations of the TV image prior. A major advantage of the one used in the paper is that it results in a quadratic approximation which is easy to analyze analytically. The closeness of the variational approximation to the true posterior in two or more dimensions is still an open question. Notice, however, that we have proved the optimality, in the divergence sense, of the obtained approximation among a given class of Gaussian distributions. Insightful comments on when the variational approximation may be tight can be found in [32] (see also [33] and [34]). A discussion on approximate Bayesian inference using variational methods and its comparison with other bounds can be found in [20].
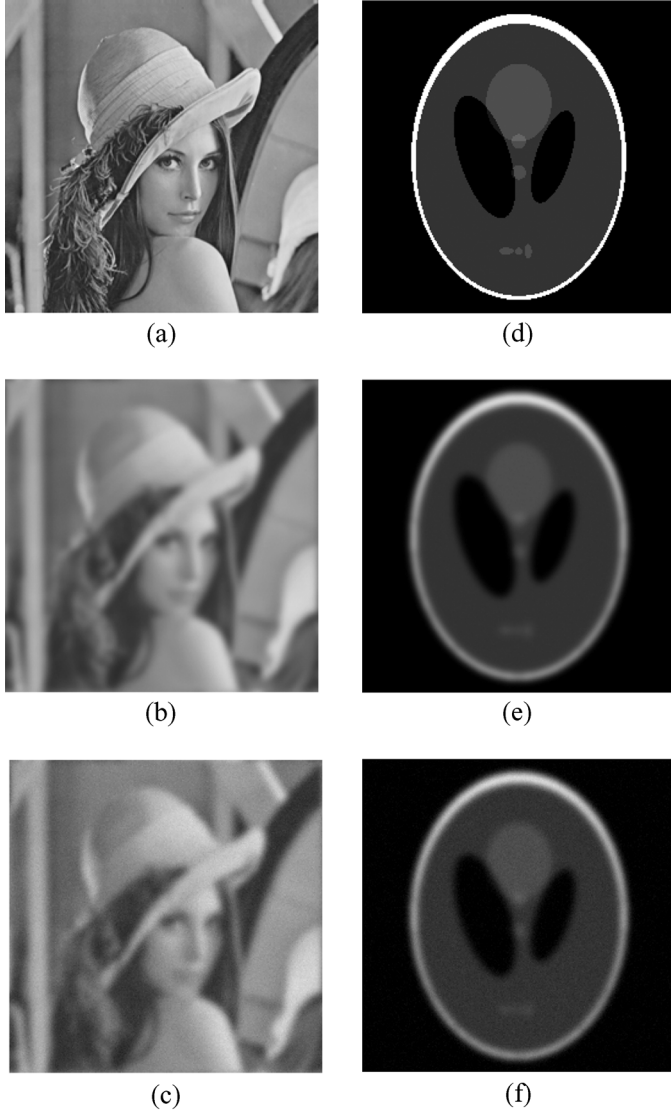
Fig. 2. (a) Lena image; degraded with a Gaussian shaped PSF with variance 9 and Gaussian noise of variance: (b) 0.16 ($\mathrm{BSNR} = 40$ dB), (c) 16 ($\mathrm{BSNR} = 20$ dB), (d) Shepp–Logan phantom; degraded with a Gaussian shaped PSF with variance 9 and Gaussian noise of variance: (e) 0.18 ($\mathrm{BSNR} = 40$ dB), (f) 18 ($\mathrm{BSNR} = 20$ dB).

IV. EXPERIMENTAL RESULTS

In this section, we present both synthetic and real blind deconvolution examples to demonstrate the performance of the algorithms. In the results reported below, we will denote Algorithm 1 by *TV1*, and Algorithm 2, where the distributions q($\mathbf{x}$) and q($\mathbf{h}$) are both degenerate, by *TV2*. We compare our algorithms with two other blind deconvolution algorithms based on variational approximations proposed in [7], which use SAR models for both the image and the blur. These algorithms are denoted by *SAR1* and *SAR2*, where *SAR1* provides an approximation to the full posterior distribution of the image and the blur, and *SAR2* assumes degenerate distributions. Comparing the proposed algorithms with *SAR1* and *SAR2* provides a measure of the effectiveness of the proposed TV image prior, and also the performance of the spatially adaptive deconvolution compared to non-spatially adaptive restoration methods. In the synthetic experi-

BSNR	Method	Lena		Cameraman		Shepp-Logan	
		ISNR (dB)	iterations	ISNR (dB)	iterations	ISNR (dB)	iterations
40dB	*TV1*	2.53	85	1.82	92	3.07	200
	TV2	2.95	200	1.73	200	3.36	200
	SAR1	1.35	63	1.03	66	1.20	121
	SAR2	1.43	78	1.01	89	1.35	180
	TV1-NB	4.33	9	2.96	11	4.16	28
	TV2-NB	4.31	9	2.95	11	4.15	28
20dB	*TV1*	2.62	81	1.70	5	2.47	8
	TV2	-32.50	500	-40.89	392	-23.88	476
	SAR1	1.62	80	1.16	98	1.53	146
	SAR2	-11.32	54	-8.83	80	-6.59	29
	TV1-NB	3.31	11	2.42	12	4.28	17
	TV2-NB	3.29	11	2.41	12	4.27	17

ments, we also include the results from the nonblind versions of our algorithms, where the blur is assumed to be known and only the image and the hyperparameters are estimated during iterations. These nonblind algorithms will be denoted as *TV1-NB* and *TV2-NB*.

For the first set of our experiments, "Lena," "Cameraman," and "Shepp–Logan" phantom images are blurred with a Gaussian-shaped function with variance 9, and white Gaussian noise is added to obtain degraded images with blurred-signal-to-noise ratios (BSNR) of 20 and 40 dB. The original and degraded "Lena" images and "Shepp–Logan" phantoms are shown in Fig. 2. The initial values for the *TV1* and *TV2* algorithms are chosen as follows: The observed image $\mathbf{y}$ is used as the initial estimate of $\mathbf{x}^1$, and a Gaussian function with variance 4 as the initial estimate $\mathbf{h}^1$ of the blur. The covariance matrices $\mathrm{cov}^1(\mathbf{h})$ and $\mathrm{cov}^1(\mathbf{x})$ are set equal to zero. The initial values β^1, α_{im}^1, and α_{bl}^1 are calculated according to (43)–(45), assuming degenerate distributions, and the initial value $\mathbf{u}^1$ is calculated using $\mathbf{x}^1$ in (51). It should be emphasized that except from the initial value of the blur, all parameters are automatically estimated from the observed image. For the *SAR1* and *SAR2* algorithms, the same initial blur is used, and other parameters are calculated also automatically from the observed image [7].

In this first set of experiments, we set all confidence parameters equal to zero, i.e., the observation is made fully responsible for the estimation process. The quantitative results are shown in Table I, where ISNR is defined as $10\log_{10}(\|\mathbf{x} - \mathbf{y}\|^2/\|\mathbf{x} - \hat{\mathbf{x}}\|^2)$, where $\mathbf{x}$, $\mathbf{y}$, and $\hat{\mathbf{x}}$ represent the original, observed, and estimated images, respectively. For all experiments, $\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2/\|\mathbf{x}^{k-1}\|^2 < 10^{-5}$ (or $\mathrm{E}^k(\mathbf{x})$ instead of $\mathbf{x}^k$) is used to terminate the algorithms, and a threshold of 10^{-5} is used to terminate the GD iterations. The corresponding restoration results for the "Lena" image are shown in Fig. 3 for the 40-dB BSNR case, and in Fig. 4 for the 20-dB BSNR case.

A few remarks can be made by examining the ISNR values in Table I and the restorations visually. First, note that the nonblind algorithms *TV1-NB* and *TV2-NB* result in higher ISNR values than the blind ones, as expected, although the resulting images

Fig. 3. Restorations of the Lena image blurred with a Gaussian PSF with variance 9 and 40-dB BSNR using the (a) *TV1* algorithm ($\mathrm{ISNR} = 2.53$ dB), (b) *TV2* algorithm ($\mathrm{ISNR} = 2.95$ dB), (c) *SAR1* algorithm ($\mathrm{ISNR} = 1.35$ dB), (d) *SAR2* algorithm ($\mathrm{ISNR} = 1.43$ dB), (e) *TV1-NB* algorithm ($\mathrm{ISNR} = 4.33$ dB), and (f) *TV2-NB* algorithm ($\mathrm{ISNR} = 4.31$ dB).

Fig. 4. Restorations of the Lena image blurred with a Gaussian PSF with variance 9 and 20-dB BSNR using the (a) *TV1* algorithm ($\mathrm{ISNR} = 2.62$ dB), (b) *TV2* algorithm ($\mathrm{ISNR} = -32.50$ dB), (c) *SAR1* algorithm ($\mathrm{ISNR} = 1.62$ dB), (d) *SAR2* algorithm ($\mathrm{ISNR} = -11.32$ dB), (e) *TV1-NB* algorithm ($\mathrm{ISNR} = 3.31$ dB), and (f) *TV2-NB* algorithm ($\mathrm{ISNR} = 3.29$ dB).

are visually comparable. Secondly, algorithms *TV1* and *TV2* result in higher ISNR values for all images and noise levels than the *SAR*-based algorithms. Visually, the TV-based algorithms result in sharper restorations, and in addition the ringing artifacts are reduced. Another important point is that algorithms *TV2* and *SAR2* fail to converge to successful restorations for the 20-dB BSNR case. On the other hand, algorithms *TV1* and *SAR1* result in acceptable restorations in this case. As can be seen in Fig. 4(a) and (c), *TV1* succeeds at removing the blur and reducing the ringing artifacts providing a better restored image than the *SAR1* algorithm.

The differences between the *TV*-based and *SAR*-based algorithms are clearer in the restoration of the Shepp–Logan phantom, which are shown in Fig. 5 for the 40-dB BSNR case, and in Fig. 6 for the 20-dB BSNR case. Algorithms *TV1* and *TV2* clearly outperform the *SAR* algorithms in terms of preserving and recovering the edges, whereas the ringing
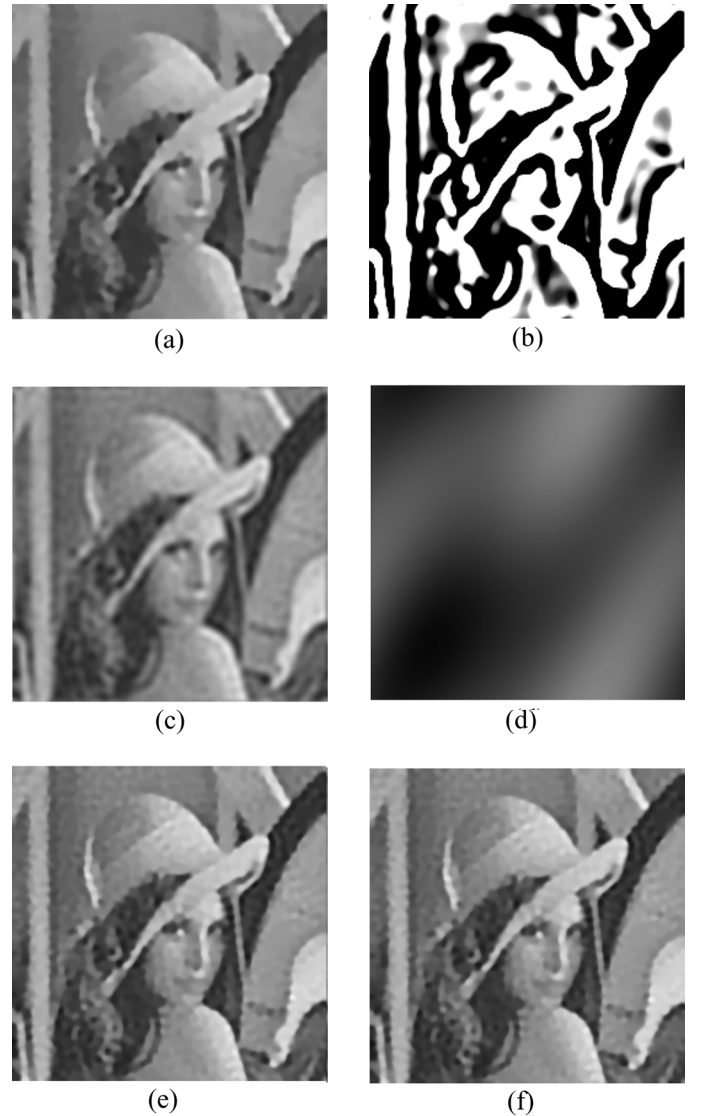
artifacts are more visible at 40-dB BSNR than at 20-dB BSNR. Again, algorithms *TV2* and *SAR2* fail to converge to meaningful restorations for the 20-dB BSNR case. Note that in other cases the restorations by *TV1* and *TV2* are very close to nonblind restoration results, except for some ringing artifacts resulting from estimation errors in the PSF.

A possible reason that algorithms *TV2* and *SAR2* fail to provide meaningful restorations for the 20-dB BSNR case is the lack of the uncertainty terms $\mathrm{cov}^k(\mathbf{h})$ and $\mathrm{cov}^k(\mathbf{x})$ in (49) and (50), respectively. In this case, the matrices that are inverted in (49) and (50) become worse conditioned than the matrices being inverted in algorithms TV1 [in (26) and (30)] and SAR1, thus degrading the quality of the restorations.

We note here that the proposed algorithms are quite robust to the initial selected value of the blur. When a Gaussian with variance 2 is chosen as $\mathbf{h}^1$, the ISNR values are 1.80 dB for *TV1*
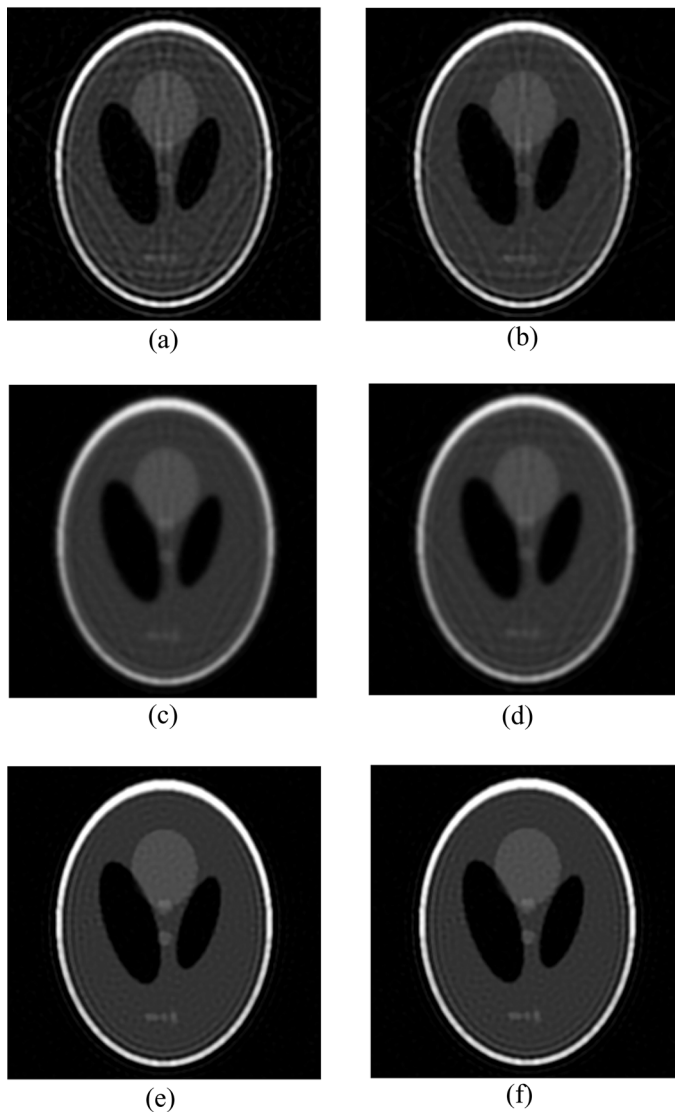
Fig. 5. Restorations of the Shepp–Logan phantom blurred with a Gaussian PSF with variance 9 and 40-dB BSNR using the (a) *TV1* algorithm ($\mathrm{ISNR} = 3.07$ dB), (b) *TV2* algorithm ($\mathrm{ISNR} = 3.36$ dB), (c) *SAR1* algorithm ($\mathrm{ISNR} = 1.20$ dB), (d) *SAR2* algorithm ($\mathrm{ISNR} = 1.35$ dB), (e) *TV1-NB* algorithm ($\mathrm{ISNR} = 4.16$ dB), and (f) *TV2-NB* algorithm ($\mathrm{ISNR} = 4.15$ dB).

Fig. 6. Restorations of the Shepp–Logan phantom blurred with a Gaussian PSF with variance 9 and 20-dB BSNR using the (a) *TV1* algorithm ($\mathrm{ISNR} = 2.47$ dB), (b) *TV2* algorithm ($\mathrm{ISNR} = -23.88$ dB), (c) *SAR1* algorithm ($\mathrm{ISNR} = 1.53$ dB), (d) *SAR2* algorithm ($\mathrm{ISNR} = -6.59$ dB), (e) *TV1-NB* algorithm ($\mathrm{ISNR} = 4.28$ dB), and (f) *TV2-NB* algorithm ($\mathrm{ISNR} = 4.27$ dB).

and 2.75 dB for *TV2* for 40-dB BSNR, and 1.71 for *TV1* and -33.36 dB for *TV2* for 20-dB BSNR, similarly to the results in Table I.

One-dimensional slices through the origin of the estimated blurs for all algorithms corresponding to the restoration of the "Lena" image are shown in Fig. 7. It is clear that all algorithms provide accurate estimates of the true PSF for both noise levels. As already mentioned, *TV2* and *SAR2* fail to converge to meaningful PSF and image estimates at 20-dB BSNR.

Before proceeding with the next set of experiments, we compare the SAR blur prior with the TV prior on the blur, as used in [5]. The algorithms proposed in [5] and [15] place TV priors both on the unknown image and blur and follow a regularization-based restoration procedure to estimate them. The hyperparameters are selected manually for optimal performance. Therefore, to facilitate a comparison between these algorithms and
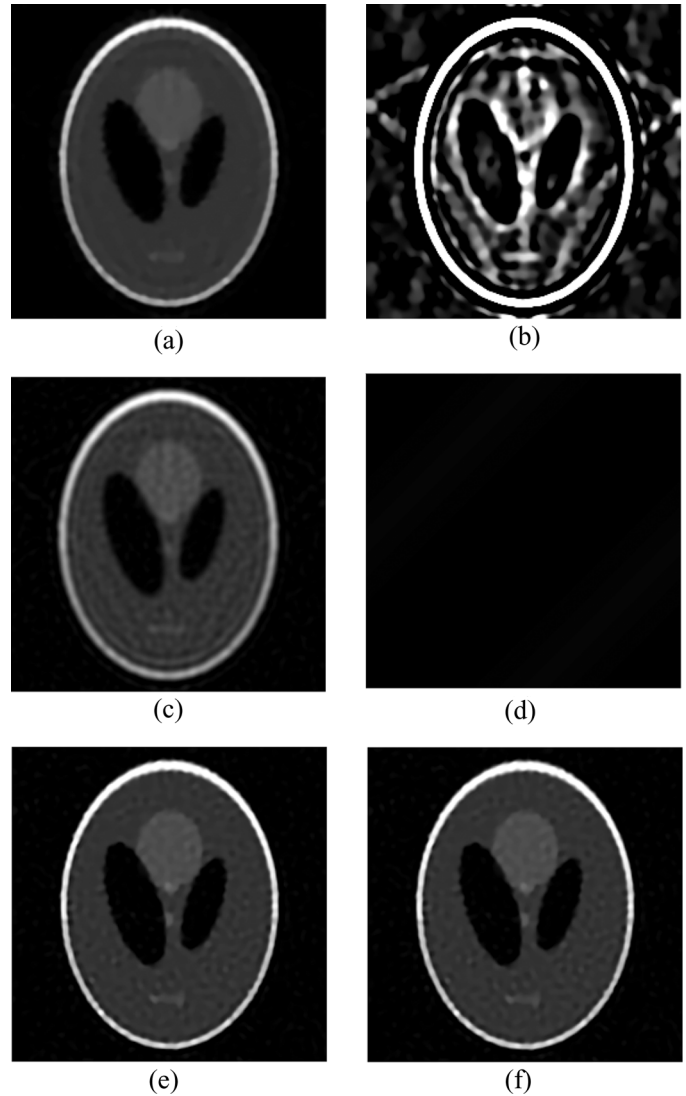
the proposed ones, which estimate the hyperparameters from the observed data, we calculate the optimal parameters from the original image and blur. Obviously, this is impossible in a practical setting, but it provides the best possible restoration result that can be obtained by [5]. The restored images are shown in Fig. 8(a) and (b) for the $\mathrm{BSNR} = 40$ dB and $\mathrm{BSNR} = 20$ dB cases, respectively. The corresponding ISNR values are 2.18 dB at 500 iterations and 2.26 dB at 79 iterations. One-dimensional slices through the origin of the estimated blurs for these cases are shown in Fig. 8(c). We note here that although the ISNR values are comparable with the ones resulting from the proposed algorithms reported in Table I, it is clear by examining Fig. 8 that at both noise levels the algorithm fails to identify and remove the blur accurately, and, therefore, the restored images are still blurry. The ISNR improvement can be attributed to the denoising performed by the TV prior on the image. The convergence is extremely slow in the $\mathrm{BSNR} = 40$ dB case. In the
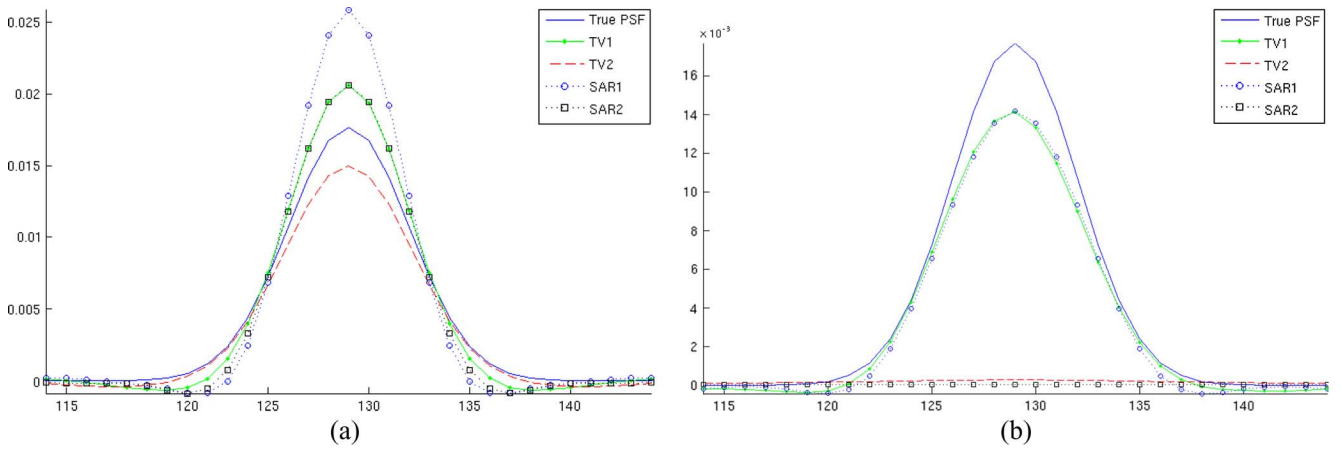
Fig. 7.   One-dimensional slices through the origin of the original and estimated PSFs in the restoration of the Lena image degraded by a Gaussian with variance 9 and (a) $\mathrm{BSNR} = 40$ dB, and (b) $\mathrm{BSNR} = 20$ dB, with algorithms *TV1*, *TV2*, *SAR1*, and *SAR2*.
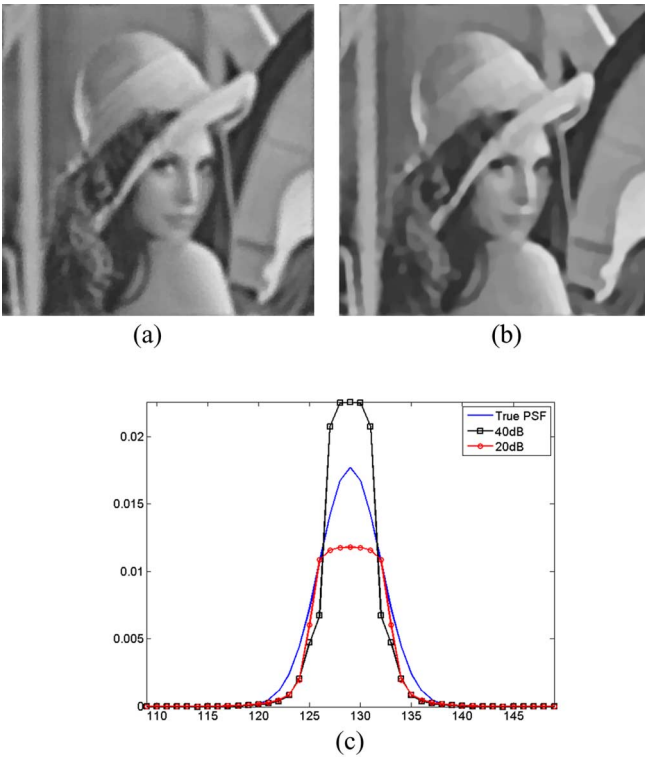

Fig. 8.   Restorations of the Lena image blurred with a Gaussian PSF with variance 9 using a TV blur prior and fixed optimal parameters as in [5]. (a) Restoration at 40-dB BSNR ($\mathrm{ISNR} = 2.18$ dB), (b) restoration at 20-dB BSNR ($\mathrm{ISNR} = 2.26$ dB), (c) estimated blur PSFs.

TABLE II
ISNR VALUES AND NUMBER OF ITERATIONS FOR THE LENA, CAMERAMAN AND
SHEPP–LOGAN IMAGES DEGRADED BY A GAUSSIAN BLUR WITH VARIANCE 5

		Lena		Cameraman		Shepp-Logan	
BSNR	Method	ISNR (dB)	iterations	ISNR (dB)	iterations	ISNR (dB)	iterations
40dB	*TV1*	3.19	200	1.66	73	2.05	137
	TV2	3.29	115	2.49	58	3.79	200
	SAR1	1.26	53	0.90	53	1.24	157
	SAR2	1.45	77	0.99	86	1.51	200
	TV1-NB	4.98	10	3.50	12	7.57	43
	TV2-NB	4.93	10	3.48	12	7.29	39
20dB	*TV1*	1.39	189	1.43	136	2.09	200
	TV2	-45.20	436	-42.54	297	-26.00	478
	SAR1	1.14	87	0.87	76	1.24	200
	SAR2	-13.15	55	-10.02	73	-7.87	27
	TV1-NB	2.92	10	2.40	12	4.68	16
	TV2-NB	2.83	11	2.37	12	4.65	16

shaped PSF with variance 5, and the initial estimate of the blur, $\mathbf{h}^1$, is a Gaussian PSF with variance 2. The corresponding ISNR values of the restorations are shown in Table II. As expected, all algorithms provide better restorations in this case, although the noise variances are higher compared to the first set of experiments to obtain the same BSNRs. Similarly to the first experiment, algorithms *TV1* and *TV2* result in better restoration performance both in terms of ISNR and visual quality.

Before proceeding with the next experiment, an important observation has to be made. We noticed in our experiments that the quality of the estimation of $\mathbf{u}$ is a very important factor in the performance of the algorithms. For example, in the case of "Lena" with $\mathrm{BSNR} = 40$ dB and Gaussian PSF with variance 5, if we run the algorithms *TV1* and *TV2* by calculating $\mathbf{u}$ from the original image, we obtain ISNR values of 3.52 and 3.60 dB, respectively. Other cases showed similar improvements. Thus, knowledge about this parameter greatly improves the ISNR performance (a similar conclusion is drawn in [26] and [31]). This also confirms that the decrease in the performance of the algorithms in the presence of high noise, e.g., $\mathrm{BSNR} = 20$ dB, is due to the fact that the spatial variations in the image; hence, the parameter $\mathbf{u}$, cannot be estimated well. This is also observed in
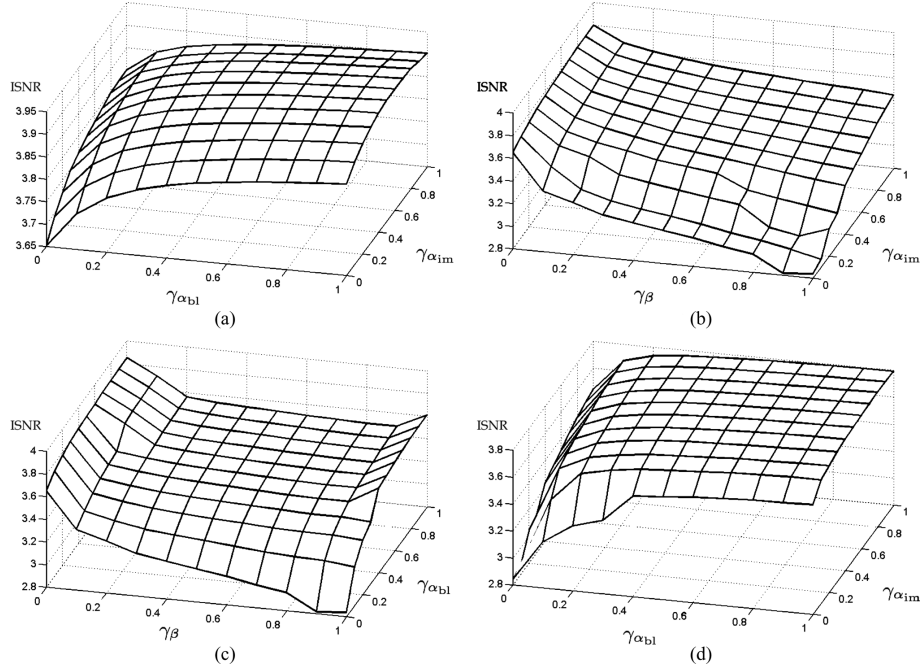
$\mathrm{BSNR} = 20$ dB case, the estimated PSF is very similar to an out-of-focus blur, indicating that the algorithm fails to identify the smooth nature of the PSF. These results are also in agreement with the ones reported in [5]. Based on the above, it is reasonable to conclude that for smooth PSFs such as a Gaussian, the proposed algorithms with the SAR blur prior outperform algorithms utilizing a TV blur prior, given also the fact that all required parameters are calculated from the observed image in an automated fashion.

In the second set of experiments, we tested the algorithms with a less severe blur. The images are blurred with a Gaussian

Fig. 9. ISNR evolution for different values of the confidence parameters for Algorithm 1 (*TV1*) applied to the "Lena" image degraded by a Gaussian with variance 9 and $\mathrm{BSNR} = 40$ dB. (a) For fixed $\gamma_\beta = 0$, (b) for fixed $\gamma_{\alpha_{\mathrm{bl}}} = 0$, (c) for fixed $\gamma_{\alpha_{\mathrm{im}}} = 0$, and (d) for fixed $\gamma_\beta = 1$.

[35] and several solutions are proposed for TV image restoration. In this work, we adapted a simple smoothing of the gradient of the image with a small Gaussian PSF (with variance 1) which largely improves both the performance and the convergence of the algorithms. Therefore, it is safe to claim that in high noise-cases, incorporating robust gradient estimation methods, such as [35] and [36], will further improve the performance of the proposed algorithms.

We now examine the effect of prior information on the performance of the proposed algorithms in the third set of experiments. Generally, some information about the values of the hyperparameters is available and can be utilized in the restoration to improve performance. For instance, the noise variance can be estimated quite accurately when a part of the image has uniform color. The image variance is more difficult to estimate from a single degraded observation. However, a set of images with similar characteristics can be used to acquire an estimate for this parameter. If an estimate of the image variance can be provided, the PSF variance can be approximated using this value (see [37] for details).

In addition to the prior knowledge on the hyperparameters, constraints on the blur estimates can also be imposed. Positivity and symmetry constraints are the most common ones, and it is known that they can significantly improve the convergence of the algorithms and the quality of the estimates [5]. Although such hard constraints have not been directly incorporated in our Bayesian framework, they can in practice improve the restoration results of the proposed algorithms *TV1* and *TV2* as shown experimentally next.

For simulation purposes, we calculated the values of the hyperparameters from the original image and PSF to be used as prior hyperparameter values. Then, using these prior values, we applied *TV1* to the "Lena" image degraded by a Gaussian PSF

TABLE III
POSTERIOR MEANS OF THE DISTRIBUTIONS OF THE HYPERPARAMETERS, ISNR, AND NUMBER OF ITERATIONS USING *TV1* FOR THE LENA IMAGE WITH 40-dB BSNR USING $\overline{\alpha_{\mathrm{im}}}^o = 0.042$, $\overline{\alpha_{\mathrm{bl}}}^o = 4.6 \times 10^8$, AND $\overline{\beta}^o = 6.25$, FOR DIFFERENT VALUES OF $\gamma_{\alpha_{\mathrm{im}}}$, $\gamma_{\alpha_{\mathrm{im}}}$ AND γ_β.

$\gamma_{\alpha_{\mathrm{im}}}$	$\gamma_{\alpha_{\mathrm{bl}}}$	γ_β	$E[\alpha_{\mathrm{im}}]$	$E[\alpha_{\mathrm{bl}}]$	$E[\beta]$	ISNR (dB)	iterations
0	0	0	0.088	3.3×10^8	5.63	3.65	32
0	1	0	0.086	4.6×10^8	5.62	3.85	38
1	1	0	0.041	4.6×10^8	5.75	3.90	51
1	0	0	0.041	3.7×10^8	5.76	3.80	51
0.6	1	0	0.051	4.6×10^8	5.72	3.92	45
0.8	1	1	0.046	4.6×10^8	6.25	3.80	82

and 40-dB BSNR with varying confidence parameters and obtained the ISNR evolution graphs shown in Fig. 9. To show the improved restoration performance and the best achievable ISNR, we applied positivity and symmetry constraints to the estimated PSF at each iteration as in [5]. Additionally, the support of the blur is estimated at each iteration using the first zero-crossing of the PSF from its center, and values outside this estimated support are set equal to zero. Selected ISNR values from these graphs with the estimated hyperparameters are shown in Table III. We included cases corresponding to the best ISNR values when (a) information about the noise variance is available, (b) information about only the PSF variance is available, (c) information about only the image variance is available, and (d) information about all hyperparameters is available. It is clear that if some information on the hyperparameters is available, biasing the algorithm towards these hyperparameters leads to improved ISNR values. However, it is interesting that incorporating the knowledge about the true value of the noise variance decreases the quality of the restorations; thus, it is better to put

(a)   (b)   (c)   (d)

Fig. 10.   Some restorations of the Lena image blurred with a Gaussian PSF with variance 9 and 40-dB BSNR using the *TV1* algorithm utilizing prior knowledge through confidence parameters and positivity and support constraints on the estimated blur. (a) $\gamma_{\alpha_{\mathrm{im}}} = \gamma_{\alpha_{\mathrm{bl}}} = \gamma_\beta = 0.0$ (ISNR $= 3.65$ dB), (b) $\gamma_{\alpha_{\mathrm{im}}} = 0$, $\gamma_{\alpha_{\mathrm{bl}}} = 1, \gamma_\beta = 0$ (ISNR $= 3.85$ dB), (c) $\gamma_{\alpha_{\mathrm{im}}} = 0.6, \gamma_{\alpha_{\mathrm{bl}}} = 1, \gamma_\beta = 0$ (ISNR $= 3.92$ dB), and (d) $\gamma_{\alpha_{\mathrm{im}}} = 0.8, \gamma_{\alpha_{\mathrm{bl}}} = 1, \gamma_\beta = 1$ (ISNR $= 3.80$ dB).
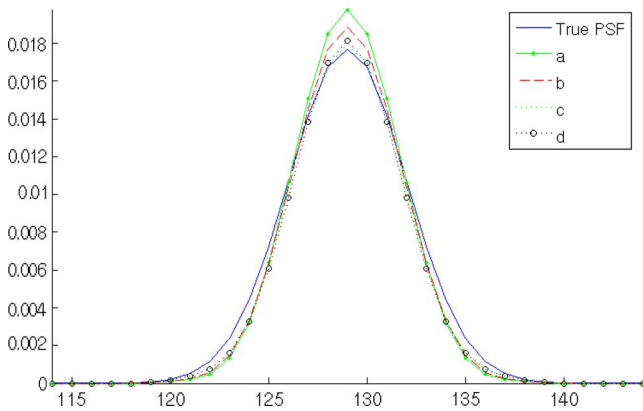

Fig. 11.   One-dimensional slices through the origin of the original and estimated PSFs in the restoration of the Lena image degraded by a Gaussian with variance 9 and $\mathrm{BSNR} = 40$ dB with algorithm *TV1*. (a) True PSF, Estimated PSF with (b) $\gamma_{\alpha_{\mathrm{im}}} = \gamma_{\alpha_{\mathrm{bl}}} = \gamma_\beta = 0.0$, (c) $\gamma_{\alpha_{\mathrm{im}}} = 0, \gamma_{\alpha_{\mathrm{bl}}} = 1, \gamma_\beta = 0$, (d) $\gamma_{\alpha_{\mathrm{im}}} = 0.6, \gamma_{\alpha_{\mathrm{bl}}} = 1, \gamma_\beta = 0$, and (e) $\gamma_{\alpha_{\mathrm{im}}} = 0.8, \gamma_{\alpha_{\mathrm{bl}}} = 1, \gamma_\beta = 1$.

no confidence on this parameter and let the algorithms adaptively select it at each iteration. On the other hand, we note that at convergence, the $E^k(\beta)^{-1}$ almost always converged to a value very close to the noise variance.

It should also be emphasized that the most critical hyperparameter is α_{bl}. It is clear from Fig. 9 and Table III that incorporating information about this parameter greatly increases the performance of the algorithm, and that the best ISNR is achieved when $\gamma_{\alpha_{\mathrm{bl}}} = 1$ is used. Restoration results with these confidence parameters are shown in Fig. 10. Note that the restoration quality is almost as high as the one achieved by the nonblind algorithms [see Fig. 3(e) for comparison]. One-dimensional slices of the estimated blurs corresponding to these cases are shown in Fig. 11, where it can be seen that the estimated PSFs are much closer to the true PSF than the ones in Fig. 7. Overall, it is clear from the results that, as expected, the performance of the algorithms can be largely increased when some information about these hyperparameters is provided and certain constraints on the estimated blur are imposed.

In our last set of experiments, the algorithms are applied to a real image of Saturn, which was taken at the Calar Alto Observatory in Spain, shown in Fig. 12(a). There is no exact expression
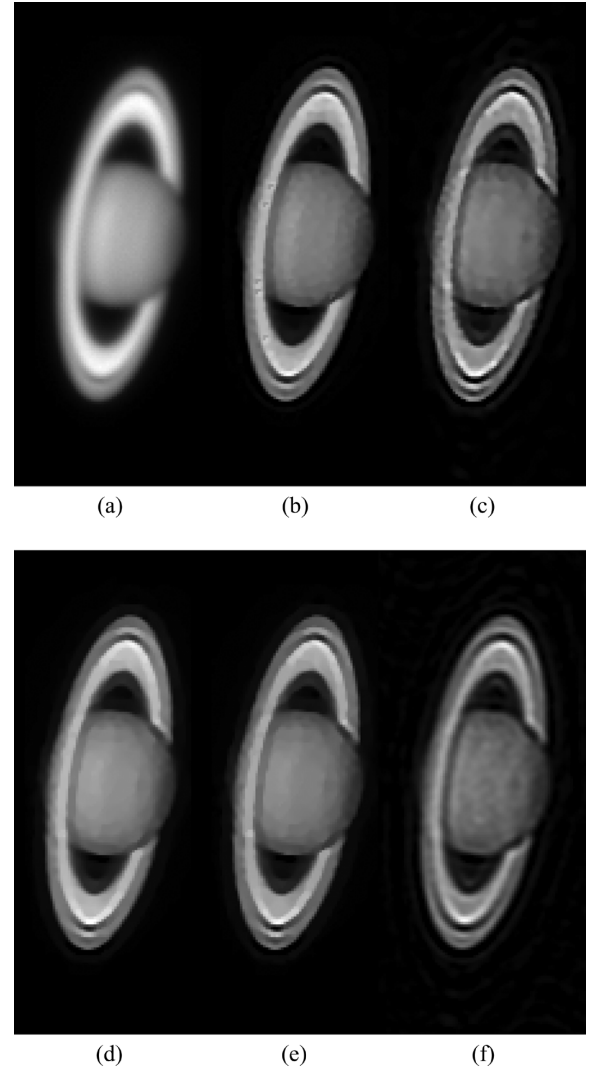

(a)   (b)   (c)

(d)   (e)   (f)

Fig. 12.   (a) Observed saturn image, (b) nonblind restoration with *TV2-NB*, (c) restoration with *TV2* with $\gamma_{\alpha_{\mathrm{im}}} = \gamma_{\alpha_{\mathrm{bl}}} = \gamma_\beta = 0.0$, (d) restoration with *TV1* with $\gamma_{\alpha_{\mathrm{im}}} = 0.8, \gamma_{\alpha_{\mathrm{bl}}} = 0.1$, and $\gamma_\beta = 0.8$; (e) restoration with *TV2* with $\gamma_{\alpha_{\mathrm{im}}} = 0.8, \gamma_{\alpha_{\mathrm{bl}}} = 0.1$, and $\gamma_\beta = 0.8$; (f) restoration with *SAR1* with $\gamma_{\alpha_{\mathrm{im}}} = 0.8, \gamma_{\alpha_{\mathrm{bl}}} = 0.1$, and $\gamma_\beta = 0.8$.

for the shape of the PSF for this image; however, the following approximation is suggested in [38] and [39]
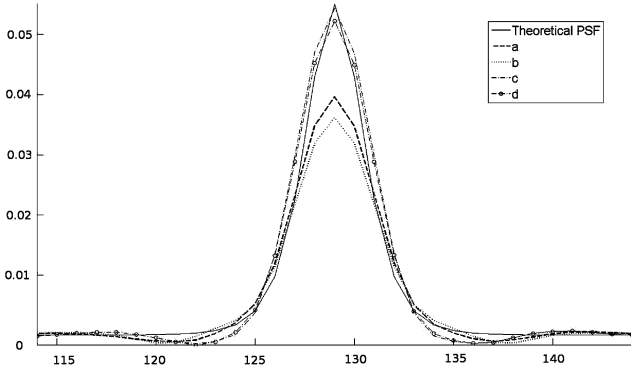
Fig. 13. One-dimensional slices through the origin of the theoretical and estimated PSFs in the restoration of the Saturn image. Estimated PSF (a) using *TV2* with $\gamma_{\alpha_{\mathrm{im}}} = \gamma_{\alpha_{\mathrm{bl}}} = \gamma_\beta = 0.0$, and with $\gamma_{\alpha_{\mathrm{im}}} = 0.8$, $\gamma_{\alpha_{\mathrm{bl}}} = 0.1$, and $\gamma_\beta = 0.8$ using (b) *SAR1*, (c) *TV1* and (d) *TV2*.

$$h(r) \propto \left(1 + \frac{r^2}{R^2}\right)^{-\delta} \tag{63}$$

with $\delta = 3$ and $R = 3.4$. The nonblind restoration result using *TV2-NB* with this theoretical PSF is shown in Fig. 12(b). This image is restored first by *TV1* and *TV2* with zero confidences placed on the prior values, i.e., $\gamma_{\alpha_{\mathrm{im}}} = \gamma_{\alpha_{\mathrm{bl}}} = \gamma_\beta = 0$. The initial blur is selected as a Gaussian shaped PSF with variance 1. Our experiments show that *TV2* gives a reasonably good restoration result, shown in Fig. 12(c), whereas *TV1* does not adequately remove the blur.

However, as in the previous experiment, the quality of the restorations can be improved by utilizing prior knowledge about the parameters. We used $\overline{\beta}^o = 8.16$, $\overline{\alpha}^o_{\mathrm{im}} = 0.24$, and $\overline{\alpha}^o_{\mathrm{bl}} = 1.6 \times 10^8$ as prior hyperparameter values, which are obtained by running *TV2-NB* with the PSF in (63). By selecting $\gamma_{\alpha_{\mathrm{im}}} = 0.8$, $\gamma_{\alpha_{\mathrm{bl}}} = 0.1$, and $\gamma_\beta = 0.8$, we obtain the restorations shown in Fig. 12(d) with *TV1* and Fig. 12(d) with *TV2*. As a comparison, the restoration result with *SAR1* with the same confidence parameters is shown in Fig. 12(e). Note that *TV*-based approaches are more successful at removing the blur while providing smooth restorations with less ringing. The estimated PSFs corresponding to these cases as well as the theoretical PSF is shown in Fig. 13. It is clear that the estimated PSFs by the proposed algorithms are much closer to the theoretical PSF than the *SAR1* result, even when no prior knowledge is incorporated.

We conclude this section by commenting on the computational complexity of the algorithms. The proposed algorithms are computationally more intensive than *SAR*-based restoration methods since (27) and (49) cannot be solved by direct inversion in the frequency domain and iterative numerical approaches are needed. Typically, the MATLAB implementations of our algorithms required on the average about 20 s per iteration 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, [29] and [30]), or splitting techniques [40].

V. CONCLUSION

In this paper, we presented a novel TV-based blind deconvolution methodology where the unknown image, blur and the hyperparameters are estimated simultaneously. The blind deconvolution problem is formulated using a hierarchical Bayesian model, and variational inference is utilized to approximate the posterior distributions of the unknown parameters rather than point estimates. Approximating the posterior distribution makes evaluating the uncertainty of the estimates possible. Two algorithms are provided resulting from this approach. It is shown that the unknown parameters of the Bayesian formulation can be calculated automatically using only the observation or using also prior knowledge with different confidence values to improve the performance of the algorithms. Experimental results demonstrated that the proposed approaches result in high-quality restorations in both synthetic and real image experiments.

REFERENCES

[1] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Total variation blind deconvolution using a variational approach to parameter, image, and blur estimation," presented at the EUSIPCO, Poznan, Poland, Sep. 2007.

[2] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Process. Mag.*, vol. 13, no. 3, pp. 43–64, Mar. 1996.

[3] T. E. Bishop, S. D. Babacan, B. Amizic, A. K. Katsaggelos, T. Chan, and R. Molina, "Blind image deconvolution: Problem formulation and existing approaches," in *Blind Image Deconvolution: Theory and Applications*, P. Campisi and K. Egiazarian, Eds. Boca Raton, FL: CRC, 2007, ch. 1.

[4] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. Freeman, "Removing camera shake from a single photograph," in *Proc. SIGGRAPH*, Boston, MA, 2006, vol. 25, pp. 787–794.

[5] T. F. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 370–375, Mar. 1998.

[6] 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, 2004.

[7] R. Molina, J. Mateos, and A. K. 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.

[8] J. H. Money and S. H. Kang, "Total variation minimizing blind deconvolution with shock filter reference," *Image Vis. Comput.*, vol. 26, no. 2, pp. 302–314, 2008.

[9] L. Bar, N. Sochen, and N. Kiryati, "Variational pairing of image segmentation and blind restoration," in *Proc. 8th Eur. Conf. Computer Vision*, Prague, Czech Republic, May 2004, vol. Part II, LNCS 3022, pp. 166–177.

[10] M. Bronstein, A. Bronstein, M. Zibulevsky, and Y. Zeevi, "Blind deconvolution of images using optimal sparse representations," *IEEE Trans. Image Process.*, vol. 14, no. 6, pp. 726–736, Jun. 2005.

[11] M. S. C. Almeida and L. B. Almeida, "Blind deblurring of natural images," presented at the ICASSP, Las Vegas, NV, Apr. 2008.

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

[13] 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, Jul. 2000.

[14] K. Z. Adami, "Variational methods in Bayesian deconvolution," *PHYSTAT*, vol. C030908, TUGT002.

[15] Y. L. You and M. Kaveh, "Blind image restoration by anisotropic regularization," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 396–407, Mar. 1999.

[16] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, pp. 259–268, 1992.

[17] J. Bernardo and A. Smith, *Bayesian Theory*. New York: Wiley, 1994.

[18] 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.

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

[20] M. Beal, "Variational Algorithms for Approximate Bayesian Inference," Ph.D. dissertation, The Gatsby Computational Neuroscience Unit, Univ. College London, London, U.K., 2003.

[21] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, pp. 79–86, 1951.

[22] G. Parisi, *Statistical Field Theory*. Redwood City, CA: Addison-Wesley, 1988.

[23] J. Miskin, "Ensemble Learning for Independent Component Analysis," Ph.D. dissertation, Astrophysics Group, Univ. Cambridge, Cambridge, U.K., 2000.

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

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

[26] S. D. Babacan, R. Molina, and A. Katsaggelos, "Parameter estimation in TV image restoration using variational distribution approximation," *IEEE Trans. Image Process.*, no. 3, pp. 326–339, Mar. 2008.

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

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] M. I. Jordan, Z. Ghahramani, T. S. Jaakola, and L. K. Saul, "An introduction to variational methods for graphical models," in *Learning in Graphical Models*. Cambrige, MA: MIT Press, 1998, pp. 105–162.

[33] A. Ilin and H. Valpola, "On the effect of the form of the posterior approximation in variational learning of ica models," *Neural Process. Lett.*, vol. 22, pp. 183–204, 2005.

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

[35] W. Zhao and A. Pope, "Image restoration under significant additive noise," *IEEE Signal Process. Lett.*, vol. 14, no. 6, pp. 401–404, Jun. 2007.

[36] J. C. Brailean and A. K. Katsaggelos, "Noise robust spatial gradient estimation for use in displacement estimation," in *Proc. Int. Conf. Image Processing*, Washington, DC, 1995, vol. 1, p. 211.

[37] Y. L. You and M. Kaveh, "A regularization approach to joint blur and image restoration," *IEEE Trans. Image Process.*, vol. 5, no. 3, pp. 416–428, Mar. 1996.

[38] A. F. J. Moffat, "A theoretical investigation of focal stellar images in the photographic emulsion and application to photographic photometry," *Astron. Astrophys.*, vol. 3, pp. 455–461, 1969.

[39] R. Molina and B. D. Ripley, "Using spatial models as priors in astronomical image analysis," *J. Appl. Statist.*, vol. 16, pp. 193–206, 1989.

[40] Y. Wang, W. Yin, and Y. Zhang, "A fast algorithm for image deblurring with total variation regularization," CAAM Tech. Rep. TR07-10 (2007), Rice Univ., Houston, TX, 2007.

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.

Mr. Babacan is the recipient of an IEEE International Conference on Image Processing Paper Award (2007).

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 a 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.

Aggelos K. Katsaggelos (S'87–M'87–SM'92–F'98) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of 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, 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 Seamless Communications and a member of the Academic Affiliate Staff, Department of Medicine, Evanston Hospital. He has published extensively in the areas of signal processing, multimedia transmission, and computer vision. He is the Editor of *Digital Image Restoration* (Springer-Verlag, 1991), coauthor of *Rate-Distortion Based Video Compression* (Kluwer, 1997), coeditor of *Recovery Techniques for Image and Video Compression and Transmission* (Kluwer. 1998), and coauthor of *Super-Resolution for Images and Video* (Claypool, 2007) and *Joint Source-Channel Video Transmission* (Claypool, 2007). He is also the coinventor of twelve international 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).