A Variational Approach for Bayesian Blind Image Deconvolution

Aristidis C. Likas, *Senior Member, IEEE,* and Nikolas P. Galatsanos, *Senior Member, IEEE*

***Abstract*—In this paper, the blind image deconvolution (BID) problem is addressed using the Bayesian framework. In order to solve for the proposed Bayesian model, we present a new methodology based on a variational approximation, which has been recently introduced for several machine learning problems, and can be viewed as a generalization of the expectation maximization (EM) algorithm. This methodology reaps all the benefits of a "full Bayesian model" while bypassing some of its difficulties. We present three algorithms that solve the proposed Bayesian problem in closed form and can be implemented in the discrete Fourier domain. This makes them very cost effective even for very large images. We demonstrate with numerical experiments that these algorithms yield promising improvements as compared to previous BID algorithms. Furthermore, the proposed methodology is quite general with potential application to other Bayesian models for this and other imaging problems.**

***Index Terms*—Bayesian parameter estimation, blind deconvolution, graphical models, image restoration, variational methods.**

I. Introduction

THE blind image deconvolution (BID) problem is a difficult and challenging problem because from the observed image it is hard to uniquely define the convolved signals. Nevertheless, there are many applications where the observed images have been blurred either by an unknown or a partially known point spread function (PSF). Such examples can be found in astronomy and remote sensing where the atmospheric turbulence cannot be exactly measured, in medical imaging where the PSF of different instruments has to be measured and thus is subject to errors, in photography where the PSF of the lens used to obtain the image is unknown or approximately known, etc.

A plethora of methods has been proposed to address this problem; see [1] for a seven-year-old survey of this problem. Since, in BID, the observed data are not sufficient to specify the convolved functions, most recent methods attempt to incorporate in the BID algorithm some prior knowledge about these functions. Since it is very hard to track the properties of the PSF and the image simultaneously, several BID methods attempt to impose constraints on the image and the PSF in an alternating fashion. In other words, such approaches cycle between two (the image and the PSF) estimation steps. In the image estimation step, the image is estimated assuming that the PSF is fixed to its last estimate from the PSF estimation step. In the PSF estimation step, the PSF is estimated assuming the image to be fixed to its last estimate from the image estimation step. This decouples the nonlinear observation model in BID into two linear observation models that are easy to solve. Algorithms of this nature that use a deterministic framework to introduce *a priori* knowledge in the form of convex sets, "classical" regularization, regularization with anisotropic diffusion functionals, and fuzzy soft constraints were proposed in [5]–[7] and [15], respectively.

A probabilistic framework using maximum likelihood (ML) estimation was applied to the BID problem in [2]–[4] using the expectation maximization (EM) algorithm [11]. However, the ML formulation does not allow the incorporation of prior knowledge, which is essential in order to reduce the degrees of freedom of the available observations in BID. As a result, in order to make these algorithms to work in practice, a number of deterministic constraints such as the PSF support and symmetry had to be used. These constraints, although they make intuitive sense, strictly speaking, cannot be justified theoretically by the ML framework.

In [8]–[10], the Bayesian formulation is used for a special case of the BID problem where the PSF was assumed partially known. In this case, the PSF was assumed to be given by the sum of a known deterministic component and an unknown stochastic component. In these works, two strategies were adopted in order to bypass the above-mentioned difficulties in writing down the probabilistic law relating the observations and the quantities to be estimated. First, in [8], the stochastic model that relates the observations with the quantities to be estimated was simplified. The direct dependence on the unknown image of the statistics of the additive noise component due to the PSF uncertainty was removed. This made possible to write down in closed form the probabilistic law that relates the observations with the quantities to be estimated and extend the EM algorithm in [3], [4], and [24] to this problem. Second, in [9] and [10], the use of the above-mentioned probabilistic law was bypassed by integrating out the dependence of the unknown image to the observations. More specifically, a Laplace approximation of the Bayesian integral that appears in this formulation was used. In spite of this, it was reported in [9] that the accuracy of the obtained estimates of the statistics of the errors in the PSF and the image could vary significantly, depending on the initialization. Thus, using the Bayesian approach in [9], it is impossible to obtain accurate restorations unless accurate prior knowledge about either the statistics of the error in the PSF or the image is available in the form of hyper-priors [10].

The Bayesian framework is a very powerful and flexible methodology for estimation and detection problems because

Manuscript received June 16, 2003; revised December 28, 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Alfred O. Hero.

The authors are with the Department of Computer Science, University of Ioannina, GR 45110, Ioannina, Greece (e-mail: arly@cs.uoi.gr; galatsanos@cs.uoi.gr).

Digital Object Identifier 10.1109/TSP.2004.831119

1053-587X/04$20.00 © 2004 IEEE

it provides a structured way to include prior knowledge concerning the quantities to be estimated. Furthermore, both the Bayesian methodology and its application to practical problems have recently experienced an explosive growth; see, for example, [12]–[14]. In spite of this, the application of this methodology to the BID problem remains elusive mainly due to the nonlinearity of the observation model. This makes intractable the computation of the joint probability density function (PDF) of the image and the PSF, given the observations. One way to bypass this problem is to employ in a Bayesian framework the technique of alternating between estimating the image and the PSF while keeping the other constant, as previously described. The main advantage of such a strategy is that it linearizes the observation model, and then, it is easy to apply the Bayesian framework. However, clearly, this is a suboptimal strategy. Another approach to bypass this problem could be to use Markov chain Monte Carlo (MCMC) techniques to generate samples from this elusive conditional PDF and then estimate the required parameters from the statistics of those samples. However, MCMC techniques are notoriously computational intensive, and furthermore, there is no universally accepted criterion or methodology to decide when to terminate [13].

In what follows, we propose to use a new methodology termed "variational" to adress the Bayesian BID problem in a computationally efficient way, resorting neither to the suboptimal linearization by alternating between the assumption that the image and the PSF are constant as previously explained nor to MCMC. The proposed approach is a generalization of both the ML framework in [2]–[4] and [24] and the partially known PSF model in [8]–[10]. The variational methodology that we use was first introduced in the machine learning community to solve Bayesian inference problems with complex probabilistic models; see, for example, [17], [19], [20], [22], and [23]. In the machine learning community, the term *graphical models* has been coined in such cases since a graph can be used to represent the dependencies among the random variables of the models, and the computations required for Bayesian inference can be greatly facilitated based on the structure of this graph. It has also been shown that the variational approach can be viewed as a generalization of the EM algorithm [16]. In [21], a similar methodology to the variational, which is termed ensemble learning, is used by Miskin and MacKay to address BID in a Bayesian framework. However, the approach in [21] uses a different model for both the image and the PSF. This model assumes that the image pixels are independent identically distributed and, thus, does not capture the between pixel correlations of natural images. Furthermore, our model allows simplified calculations in the frequency domain. This greatly facilitates the implementation of our approach for realistic high-resolution images. We believe that the approach in [21] cannot be applied to large images.

The rest of this paper is organized as follows: In Section II, we provide the background on variational methods; in Section III, we present the Bayesian model that we propose for the BID problem and the resulting variational functional; in Section IV, two iterative algorithms are presented that can be used to solve for this model, and we provide numerical experiments indicating the superiority of the proposed algorithms as compared with previous BID approaches; finally, in Section V, we provide our conclusions and suggestions for future work.

II. Background on Variational Methods

The variational framework constitutes a generalization of the well-known expectation maximization (EM) algorithm for likelihood maximization in Bayesian estimation problems with "hidden variables." The EM algorithm has been proved to be a valuable tool for many problems, since it provides an elegant approach to bypass difficult optimization and integrations required in Bayesian estimation problems. In order to efficiently apply the EM algorithm, two requirements should be fulfilled [11]: i) In the E-step, we should be able to compute the conditional PDF of the "hidden variables" given the observation data. ii) In the M-step, it is highly preferable to have analytical formulas for the update equations of the parameters. Nevertheless, in many problems, it is not possible to meet the above requirements and several variants of the basic EM algorithm have emerged. For example, a variant of the EM algorithm, called the "generalized EM" (GEM), proposes a partial M step in which the likelihood always improves. In many cases, partial implementation of the E step is also natural. An algorithm along such lines was investigated in [16].

The most difficult situation for applying the EM algorithm emerges when it is not possible to specify the conditional PDF of the hidden variables given the observed data that is required in the E-step. In such cases, the implementation of the EM algorithm is not possible. This significantly restricts the range of problems where EM can be applied. To overcome this serious shortcoming of the EM algorithm, the variational methodology was developed [17]. In addition, it can be shown that EM naturally arises as a special case of the variational methodology.

Assume an estimation problem where x and s are the observed and hidden variables, respectively, and θ are the model parameters to be estimated. All PDFs are parameterized by the parameters θ, i.e., $p(x;\theta)$, $p(s,x;\theta)$, and $p(s|x;\theta)$, and we omit θ for brevity in what follows.

For an *arbitrary PDF* $q(s)$ of the hidden variables s, it is easy to show that

$$\log p(x) + E_q(\log q(s)) = E_q(\log p(x,s)) + E_q(\log q(s)) - E_q(\log p(s|x))$$

where E_q denotes the expectation with respect to $q(s)$. The above equation can be written as

$$L(\theta) + E_q(\log q(s)) = E_q(\log p(x,s)) + KL(q(s)||p(s|x))$$

where $L(\theta) = \log p(x;\theta)$ is the likelihood of the unknown parameters, and $KL(q(s)||p(s|x))$ is the Kullback–Liebler distance between $q(s)$ and $p(s|x)$.

Rearranging the previous equation, we obtain

$$\mathbf{F}(q,\theta) = L(\theta) - KL(q(s)||p(s|x)) = E_q(\log p(x,s)) + H(q) \tag{1}$$

where $H(q)$ is the entropy of $q(s)$. From (1), it is clear that $\mathbf{F}(q,\theta)$ provides a *lower bound* for the likelihood of θ param-

eterized by the family of PDFs $q(s)$ since $KL(q(s)||p(s|x)) \geq 0$. When $q^*(s) = p(s|x;\theta)$, the lower bound becomes exact: $\mathbf{F}(q^*,\theta) = L(\theta)$. Using this framework, EM can then be viewed as a special case when $q^*(s) = p(s|x;\theta)$.

However, the previous framework allows us, based on (1), to find a local maximum of $L(\theta)$ *using an arbitrary* PDF $q(s)$. This is a very useful generalization because it bypasses one of the main restrictions of EM that of exactly knowing $p(s|x)$. The variational method works to maximize the lower bound $\mathbf{F}(q,\theta)$ with respect to both θ and q. This is justified by a theorem in [16], stating that if $\mathbf{F}(q,\theta)$ has a local maximum at $q^*(s)$ and θ^*, then $L(\theta)$ has a local maximum at θ^*. Furthermore, if $\mathbf{F}(q,\theta)$ has a global maximum at $q^*(s)$ and θ^*, then $L(\theta)$ has a global maximum at θ^*. Consequently, the variational EM approach can be described as follows:

$$\text{E-step: } q^{(t+1)} = \arg\max_q \mathbf{F}\left(q, \theta^{(t)}\right)$$
$$\text{M-step: } \theta^{(t+1)} = \arg\max_\theta \mathbf{F}\left(q^{(t+1)}, \theta\right).$$

This iterative approach increases at each iteration t, the value of the bound $\mathbf{F}(q,\theta)$, until a local maximum is attained.

III. Variational Blind Deconvolution

A. Variational Functional $\mathbf{F}(q,\theta)$

In what follows, we apply the variational approach to the Bayesian formulation of the *blind deconvolution* problem. The observations are given by

$$g = h * f + w = H \cdot f + w = F \cdot h + w \tag{2}$$

and we assume the $N \times 1$ vector g to be the observed variables, the $N \times 1$ vectors f, h are the hidden variables, w is Gaussian noise, and H and F are the $N \times N$ convolution matrices. We assume Gaussian PDFs for the priors of f and h. In other words, we assume $p(f) = N(\mu_f, \Sigma_f)$, $p(h) = N(\mu_h, \Sigma_h)$, and $p(w) = N(0, \Sigma_w)$. Thus, the parameters are $\theta = [\mu_f, \Sigma_f, \mu_h, \Sigma_h, \Sigma_w]^T$. The dependencies of the parameters and the random variables for the BID problem can be represented by the graph in Fig. 1.

The key difficulty with the above blind deconvolution problem is that the posterior PDF $p(f,h|g;\theta)$ of the hidden variables f and h given the observations g *is unknown*. This fact makes impossible the direct application of the EM algorithm. However, with the *variational approximation* described in the previous section, it is possible to bypass this difficulty. More specifically, we *select a factorized form for* $q(s)$ that employs Gaussian components

$$q(s) = q(h,f) = q(h)q(f) = N(m_{f^q}, C_{f^q})N(m_{h^q}, C_{h^q}) \tag{3}$$

where $\theta_q = [m_{f^q}, C_{f^q}, m_{h^q}, C_{h^q}]^T$ are the parameters of $q(s)$.

This choice for $q(s)$ can be justified because it leads to a *tractable* variational formulation that allows for the variational bound $\mathbf{F}(q,\theta)$ (1) to be specified analytically in the discrete Fourier domain (DFT) domain if circulant covariance matrices are used. From the right-hand side of (1), we have

$$\mathrm{F}(q,\theta) = E_q\left(\log\left(p(x,s)\right) + H(q)\right. \tag{4}$$

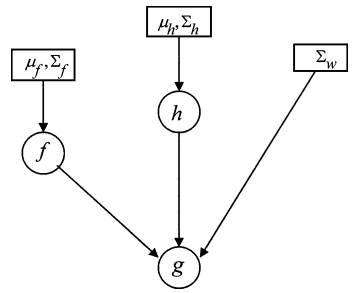

Fig. 1. Graphical model describing the data-generation process for the blind deconvolution problem considered in this paper.

where $p(x,s) = p(g,f,h) = p(g|f,h) \cdot p(f) \cdot p(h)$ with $p(g|h,f) = N(h * f, \Sigma_w)$.

The variational approach requires the computation of the expectation (Gaussian integral) in (4) with respect to $q(s)$. In order to facilitate computations for large images, we will assume circulant convolutions in (2) and that matrices Σ_f, Σ_h, Σ_w, C_{f^q}, and C_{h^q} are circulant. This allows an easy implementation in the DFT. Computing the expectation $E_q(\log(p(g,f,h))$ as well as the entropy of $q(s)$, we can write the result in the DFT domain as (5), shown at the bottom of the next page (the derivation is described in Appendix A), where $S_{f^q}(k)$, $S_{h^q}(k)$, $\Lambda_f(k)$, $\Lambda_h(k)$, and $\Lambda_w(k)$ are the eigenvalues of the $N \times N$ circulant covariance matrices C_{f^q}, C_{h^q}, Σ_f, Σ_h, and Σ_w, respectively. In addition, $G(k)$, $M_{f^q}(k)$, and $M_{h^q}(k)$ are the DFT coefficients of the vectors g, m_{f^q}, and m_{h^q}, respectively.

B. Maximization of the Variational Bound $\mathbf{F}(q,\theta)$

In analogy to the conventional EM framework, the maximization of the variational bound $\mathbf{F}(q,\theta)$ can be implemented in two steps, as described in the end of Section II. In the **E-step**, the parameters $\theta_q = [m_{f^q}, C_{f^q}, m_{h^q}, C_{h^q}]^t$ of $q(s)$ are updated. Three approaches have been considered for this update. The first approach (called VAR1) is based on the direct maximization of $\mathbf{F}(q,\theta)$ with respect to the parameters θ_q. It can be easily shown that such maximization can be performed analytically by setting the gradient of $\mathbf{F}(q,\theta)$ with respect to each parameter equal to zero, thus obtaining the update equations for $m_{f^q}^{(t+1)}$, $C_{f^q}^{(t+1)}$, $m_{h^q}^{(t+1)}$, and $C_{h^q}^{(t+1)}$. The detailed formulas of this approach are given in Appendix B.

In the second approach (called VAR2), we assume that $q(f) = p(f|g;h)$ and $q(h) = p(h|g;f)$. When h or f are assumed known, the observation model in (2) is linear. Thus, for Gaussians priors on h, f, and Gaussian noise w, the conditionals of h and f, given the observations, are Gaussians $p(f|h,g) = N(m_{f/g}, C_{f/g})$, $p(h|f,g) = N(m_{h/g}, C_{h/g})$ with known means and covariances, which are given by (see [3] and [4])

$$\begin{aligned} m_{f/g} &= \mu_f + \Sigma_f \cdot H^t \cdot (H \cdot \Sigma_f \cdot H^t + \Sigma_n)^{-1} \cdot (g - H \cdot \mu_f) \\ C_{f/g} &= \Sigma_f - \Sigma_f \cdot H^t \cdot (H \cdot \Sigma_f \cdot H^t + \Sigma_n)^{-1} \cdot H \cdot \Sigma_f \end{aligned} \tag{6}$$

$$\begin{aligned} m_{h/g} &= \mu_h + \Sigma_h \cdot F^t \cdot (F \cdot \Sigma_h \cdot F^t + \Sigma_n)^{-1} \cdot (g - F \cdot \mu_h) \\ C_{h/g} &= \Sigma_h - \Sigma_h \cdot F^t \cdot (F \cdot \Sigma_h \cdot F^t + \Sigma_n)^{-1} \cdot F \cdot \Sigma_h. \end{aligned} \tag{7}$$

Therefore, we set $m_{f^q}^{(t+1)} = m_{f/g}$, $C_{f^q}^{(t+1)} = C_{f/g}$, $m_{h^q}^{(t+1)} = m_{h/g}$, and $C_{h^q}^{(t+1)} = C_{h/g}$.

Since, in the above equations, we do not know the values of h and f, we use their current estimates $m_{h^q}^{(t)}$ and $m_{f^q}^{(t)}$. It must also be noted that all computations take place in the DFT domain. A disadvantage of this approach is that the update equations of the parameters θ_q do not theoretically guarantee the increase of the variational bound $\mathbf{F}(q,\theta)$. Nevertheless, the numerical experiments have shown that this is not a problem in practice, since in all experiments, the update equations resulted in an increase of $\mathbf{F}(q,\theta)$.

In the **M-step**, the parameters θ_q are considered fixed, and (5) is maximized with respect to the parameters θ, leading to the following update equations:

$$\begin{aligned}\mu_f^{(t+1)} &= \mu_{f^q}^{(t+1)}\\ \mu_h^{(t+1)} &= \mu_{h^q}^{(t+1)}\\ \Sigma_f^{(t+1)} &= C_{f^q}^{(t+1)} \quad \text{and}\\ \Sigma_h^{(t+1)} &= C_{h^q}^{(t+1)}\end{aligned} \tag{8}$$

for both approaches VAR1 and VAR2. The covariance of the noise is updated for the VAR1 and VAR2 approaches according to

$$\begin{aligned}\Lambda_w^{(t+1)}(k) = \frac{1}{N}\Big(|G(k)|^2 - 2\mathrm{Re}\Big\{M_{f^q}^{(t+1)}(k)M_{h^q}^{(t+1)}(k)G^*(k)\Big\}\Big)\\ + N\left(S_{f^q}^{(t+1)}(k) + \frac{1}{N}\left|M_{f^q}^{(t+1)}(k)\right|^2\right)\\ \times\left(S_{h^q}^{(t+1)}(k) + \frac{1}{N}\left|M_{h^q}^{(t+1)}(k)\right|^2\right)\end{aligned} \tag{9}$$

for $k = 0, 1 \ldots N-1$, where $\Lambda_w^{(t+1)}(k)$, $S_{f^q}^{(t+1)}(k)$, $S_{h^q}^{(t+1)}(k)$, $M_{f^q}^{(t+1)}(k)$, $M_{h^q}^{(t+1)}(k)$, and $G(k)$ are defined as previously. The detailed derivations of the formulas for the parameter updates of our models are given in Appendix B.

In the third approach (called VAR3), the optimization of the function $\mathbf{F}(q,\theta)$ at each iteration is done in *two stages*, assuming f and h to be constant in an alternating fashion. At the first stage of each iteration, f is assumed a random variable, and the parameters associated with f are updated, while h is kept constant. In the second stage, the reverse happens. More specifically, at the E-step of the first stage, since h is assumed deterministic, we have that $q(s) = q(f)$, and from (1), the new variational bound can be written

$$\mathbf{F}_f\left(q(f),\theta\right) = E_{q(f)}\left(\log p(g,f)\right) + H\left(q(f)\right) \tag{10}$$

where $\theta = [\mu_f, \Sigma_f, \Sigma_w]^T$. $\mathbf{F}_f(q,\theta)$ can be easily obtained from $\mathbf{F}(q,\theta)$ in (5) by replacing $M_{h^q}(k)$ with $H(k)$, setting $S_{h^q}(k) = 0$, and dropping the all the terms that contain $\Lambda_h(k)$. From (1), it is clear that in this case, setting $q(s) = q(f) = p(f|g;h)$ [given by (6)] leads to maximization of $\mathbf{F}_f(q,\theta)$ with respect to $q(f)$. In the M-step of the first stage, in order to maximize $\mathbf{F}_f(q,\theta)$ with respect to θ, it suffices to maximize $E_{p(f/g;h)}(\log p(g,f))$ since the entropy term is not a function of θ. Thus, the first stage reduces to the "classical" EM for the linear model $g = Hf + w$, which is also known as the "iterative Wiener filter"; see, for example, [3]. In the second stage of the VAR3 method, the role of f and h is interchanged, and the computations are similar. In other words, the variational bound $\mathbf{F}_h(q,\theta)$ (where $\theta = [\mu_h, \Sigma_h, \Sigma_w]^T$) is obtained from $\mathbf{F}(q,\theta)$ in (5) by replacing $M_{f^q}(k)$ with $F(k)$, $S_{f^q}(k) = 0$ and dropping all the terms that contain $\Lambda_f(k)$. The parameters of $p(h|g;f)$, in this case, are updated by (7).

For the VAR3 approach, the M-step updates specified in (8) still hold for both stages. However, the update of $\Lambda_w^{(t+1)}(k)$ in

$$\begin{aligned}\mathrm{F}(q,\theta) = C &- \frac{1}{2}\sum_{k=0}^{N-1}\left(\log\Lambda_w(k) + \log\Lambda_f(k) + \log\Lambda_h(k) - \log S_{f^q}(k) - \log S_{h^q}(k)\right)\\ &- \frac{1}{2}\sum_{k=0}^{N-1}\frac{\overbrace{\frac{1}{N}\left(|G(k)|^2 - 2\mathrm{Re}\left\{M_{f^q}(k)M_{h^q}(k)G^*(k)\right\}\right)}^{A_1(k)}}{\Lambda_w(k)}\\ &- \frac{1}{2}\sum_{k=0}^{N-1}\frac{\overbrace{N\left(S_{f^q}(k) + \frac{1}{N}|M_{f^q}(k)|^2\right)\left(S_{h^q}(k) + \frac{1}{N}|M_{h^q}(k)|^2\right)}^{A_2(k)}}{\Lambda_w(k)}\\ &- \frac{1}{2}\sum_{k=0}^{N-1}\frac{\overbrace{\left(S_{f^q}(k) + \frac{1}{N}|M_{f^q}(k)|^2\right) + \frac{1}{N}\left(|M_f(k)|^2 - 2\mathrm{Re}\left\{M_f^*(k)M_{f^q}(k)\right\}\right)}^{B(k)}}{\Lambda_f(k)}\\ &- \frac{1}{2}\sum_{k=0}^{N-1}\frac{\overbrace{\left(S_{h^q}(k) + \frac{1}{N}|M_{h^q}(k)|^2\right) + \frac{1}{N}\left(|M_h(k)|^2 - 2\mathrm{Re}\left\{M_h^*(k)M_{h^q}(k)\right\}\right)}^{C(k)}}{\Lambda_h(k)}\end{aligned} \tag{5}$$

the stage where h is considered deterministic and known is obtained from (9) by following the same rules as the ones used to obtain $\mathbf{F}_f(q,\theta)$ from $\mathbf{F}(q,\theta)$. This yields the update

$$\Lambda_w^{(t+1)}(k) = \frac{1}{N}\left(|G(k)|^2 - 2\mathrm{Re}\left\{M_{f^q}^{(t+1)}(k)H(k)G^*(k)\right\}\right) + |H(k)|^2\left(S_{f^q}^{(t+1)}(k) + \frac{1}{N}\left|M_{f^q}^{(t+1)}(k)\right|^2\right). \quad (11)$$

Similarly, the update $\Lambda_w(k)$ in the stage where f is considered deterministic and known is

$$\Lambda_w^{(t+1)}(k) = \frac{1}{N}\left(|G(k)|^2 - 2\mathrm{Re}\left\{F(k)M_{h^q}^{(t+1)}(k)G^*(k)\right\}\right) + |F(k)|^2\left(S_{h^q}^{(t+1)}(k) + \frac{1}{N}\left|M_{h^q}^{(t+1)}(k)\right|^2\right). \quad (12)$$

It is worth noting that the VAR3 approach, since it uses linear models, can be also derived without the variational principle by applying the "classical" EM (iterative Wiener filter) twice: once for f using as data-generation model $g = Hf + w$ with H known and once for h using as data-generation model $g = Fh + w$ with F known. From a Bayesian inference point of view, clearly, VAR3 is suboptimal since it alternates between the assumptions that f is random and h deterministic and vice-versa.

IV. Numerical Experiments

In our experiments, we used a simultaneously autoregressive (SAR) model [18] for the image; in other words, we assumed $p(f) \propto (\alpha)^{((N-1)/2)}\exp(-\frac{1}{2}\alpha\|Qf\|^2)$, where Q is the circulant matrix that represents the convolution with the Laplacian operator. For h, we assume $p(h) = N(m_h, \beta^2 I)$ and, for the noise, $p(n) = N(0, \sigma^2 I)$. Therefore, the parameters to be estimated are α, m_h, β, and σ^2.

The following five approaches have been implemented and compared:

i) variational method VAR1;
ii) variational method VAR2 (with $q(f) = p(f|h,g)$ and $q(h) = p(h|f,g)$);
iii) variational approach VAR3 in which h and f are estimated in an alternating fashion (Since the VAR3 approach, in contrast with the VAR1 and VAR2 methods, does not use a "full Bayesian" model, it serves as the comparison benchmark for the value of such model.);
iv) Bayesian approach for partially known blurs (PKN) as described in [9]
v) iterative Wiener filter (ITW) as described in [3] where only the parameters α and σ^2 are estimated.

The ITW, since it does not attempt to estimate the PSF, is expected to give always inferior results. However, it serves as a baseline that demonstrates the difficulty of each BID case we show in our experiments.

As a metric of performance for both the estimated image and the PSF the improvement in the signal-to-noise ratio (ISNR) was used. This metric is defined for the image as $\mathrm{ISNR}_f = \log 10(\|f-g\|^2/\|f-\hat{f}\|^2)$, where $\hat{f}$ is the restored image and, for the PSF, as $\mathrm{ISNR}_h = \log 10(\|h-h_{\mathrm{in}}\|^2/\|h-\hat{h}\|^2)$, where h_{in} and $\hat{h}$ are the initial guess and the estimate of the PSF, respectively. Two series of experiments were performed: first, with PSFs that were partially known, in other words, corrupted with random error and, second, with PSFs that were completely unknown.

A. Partially Known Case

Since, in many practical cases, the PSF is not completely unknown, in this series of experiments, we consider that the PSF is partially known [8]–[10], i.e., it is the sum of a deterministic component and a random component: $h = h_0 + \Delta h$. The Bayesian model that we use in this paper includes the partially known PSF case as a special case. Thus, in this experiment, we compared the proposed variational approaches with previous Bayesian formulations designed for this problem. The deterministic component h_0 was selected to have a Gaussian shape with support 31×31 pixels given by the formula $h_0(k,m) = \exp((k^2/\sigma_X^2) + (m^2/\sigma_Y^2))$ with $k, m = -15 \ldots 15$ that is also normalized to one such that $\sum_{k=-15}^{15}\sum_{m=-15}^{15} h_0(k,m) = 1$. The width and the shape of the Gaussian are defined by the variances, which were set at $\sigma_X^2 = \sigma_Y^2 = 20$. For the random component Δh, we used white Gaussian noise with $p(\Delta h) = N(0, \beta^2 I)$. In these experiments, since $m_h = h_0$ is known, the parameters to be estimated are α, β, and σ^2.

The following three cases were examined where, in each case, a degraded image was created by considering the following values for the noise and the PSF: i) $\sigma = 10^{-2}$, $\beta = 10^{-4}$, ii) $\sigma = 10^{-3}$, $\beta = 10^{-4}$, iii) $\sigma = 10^{-4}$, $\beta = 10^{-4}$. In all experiments and for all tested methods, the initial values of the parameters were $\hat{\alpha} = 500$, $\widehat{\beta^2} = 10^{-7}$, and $\widehat{\sigma^2} = 10^{-5}$. The obtained ISNR_f values of the restored images are summarized in Table I. Table I clearly indicates the superior restoration performance of the proposed variational methods (VAR1 and VAR2) as compared with both the partially known (PKN) method and the VAR3 approach. As expected, the improvement becomes more significant when the standard deviation of the PSF noise β becomes comparable with the standard deviation of the additive noise σ. In addition, as the noise in the PSF becomes larger, the benefits of compensating for the PSF increase as compared with using the ITW. It must be noted that, as also reported in [9], the PKN method is very sensitive to initialization of β and σ, and it did not converge in the third experiment. It is also interesting to mention that the first two variational schemes provide similar reconstruction results in all tested cases. In Fig. 2, we provide the images for the case $\sigma = 10^{-3}$, $\beta = 10^{-4}$.

B. Unknown Case

In this series of experiments, we assumed that the PSF is unknown; however, an initial estimate is available. In this experiment, an additional image was used to test the proposed algorithm. An initial estimate of the PSF was used for restoration with the iterative Wiener (ITW), and the same estimate was also used as the initial value of the PSF mean for the three variational (VAR1, VAR2, VAR3) methods. More specifically, the degraded image was generated by blurring with a Gaussian-shaped PSF h_{true}, as before, and additive Gaussian noise with variance $\sigma_g^2 = 10^{-6}$. The initial PSF estimate h_{init} was also assumed Gaussian shaped

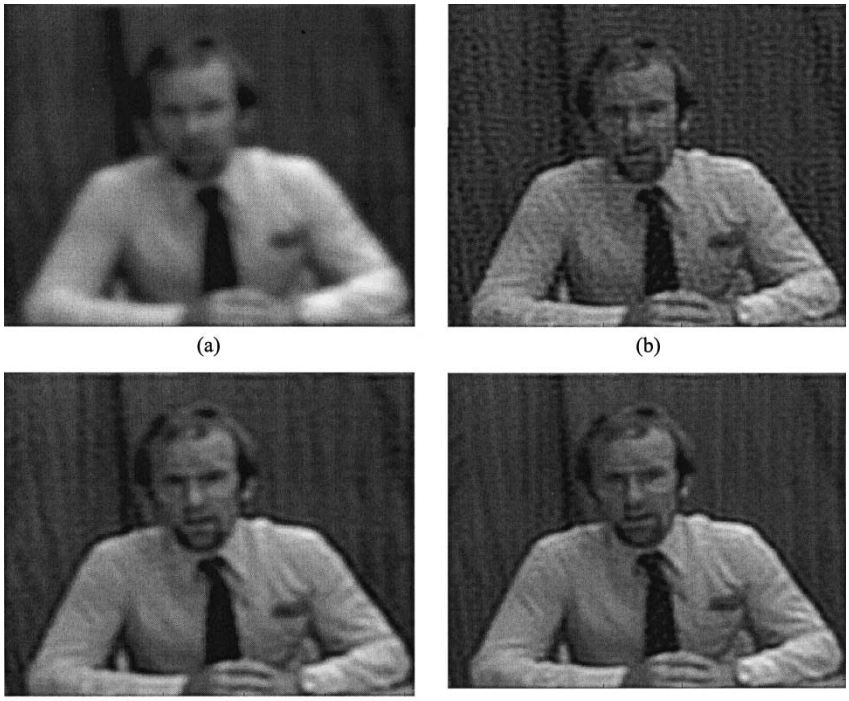

Fig. 2. Images from Table I case with $\sigma = 10^{-3}$, $\beta = 10^{-4}$. (a) Degraded image. (b) ITW, ISNR= 2.8 dB. (c) PKN, ISNR= 3.0 dB. (d) VAR2, ISNR= 3.9 dB.

TABLE I
ISNR$_f$ VALUES FOR THE PARTIALLY KNOWN EXPERIMENTS

	$\sigma = 10^{-2}, \beta = 10^{-4}$	$\sigma = 10^{-3}, \beta = 10^{-4}$	$\sigma = 10^{-4}, \beta = 10^{-4}$
VAR1	2.6dB	3.9dB	**4.8dB**
VAR2	2.6dB	3.9dB	**4.9dB**
VAR3	2.5dB	2.9dB	3.0dB
PKN	2.1dB	3.0dB	No convergence
ITW	2.5dB	2.8dB	1.64dB

but with different variances than those used to generate the images. Furthermore, the support of the true PSF is unknown. For this experiment, the unknown parameters to be estimated are α, m_h, β, and σ^2. The PKN method was not tested for this set of experiments since it is expected to yield suboptimal results because it is based on a different PSF model. Two cases were examined, and the results are presented in Table II along with the obtained ISNR values after 500 iterations of the algorithm. The PSF initializations h^1_{init} and h^2_{init} for these two experiments were chosen such that $||h_{\text{true}} - h^1_{\text{init}}|| = ||h_{\text{true}} - h^2_{\text{init}}||$, where h_{true} is the true PSF, which we are trying to infer.

In Figs. 3 and 4, we provide the images for cases 1 and 2 of Table II. In Fig. 5, we show the images that resulted from the experiments tabulated in Table III case 1, where the "Lena" image has been used.

From this set of numerical experiments, it is clear that the VAR1 approach is superior to both the VAR2 and VAR3 approaches in terms of both ISNR$_f$ and ISNR$_h$. This is expected since both VAR2 and VAR3 are suboptimal in a certain sense. VAR2, since it is in the E-step, does not explicitly optimize $\mathbf{F}(q, \theta)$ with respect to $q(s)$ and VAR3 since it does not use the "full Bayesian" model, as previously explained. Nevertheless, we observed in all our experiments, all methods increased monotonically the variational bound $\mathbf{F}(q, \theta)$. This is somewhat surprising since the VAR2 method does not optimize $\mathbf{F}(q, \theta)$ in the E-step, and the VAR3 method optimizes $\mathbf{F}_f(q, \theta)$ and $\mathbf{F}_h(q, \theta)$ in an alternating fashion.

V. CONCLUSIONS AND FUTURE WORK

In this paper, the blind image deconvolution (BID) problem was addressed using a Bayesian model with priors for both the image and the point spread function. Such a model was deemed necessary to reduce the degrees of freedom between the estimated signals and the observed data. However, for such a model, even with the simple Gaussians priors that used in this paper, it is impossible to write explicitly the probabilistic law that relates the convolving functions given the observations required for Bayesian inference. To bypass this difficulty, a variational approach was used, and we derived three algorithms that solved the proposed Bayesian model. We demonstrated with numerical experiments that the proposed variational BID algorithms provide superior performance in all tested scenarios compared with previous methods. The main shortcoming of the variational methodology is the fact that there is no analytical way to evaluate the tightness of the variational bound. Recently, methods based on Monte Carlo sampling and integration have been proposed to address this issue [23]. However, the main drawback of such methods is, on the one hand, computational complexity and, on the other hand, convergence assessment of the Markov

TABLE II
ISNRs OF ESTIMATED IMAGES AND PSF WITH THE "TREVOR" IMAGE

	case 1		case 2	
	Generating PSF $\sigma_X^2 = 20, \sigma_Y^2 = 20$	Initialization $\sigma_X^2 = 12, \sigma_Y^2 = 12$	Generating PSF $\sigma_X^2 = 20, \sigma_Y^2 = 20$	Initialization $\sigma_X^2 = 40, \sigma_Y^2 = 40$
	$ISNR_f$	$ISNR_h$	$ISNR_f$	$ISNR_h$
VAR1	3.18dB	7.45dB	1.63dB	2.92dB
VAR2	1.8dB	-6.54dB	1.59dB	2.36dB
VAR3	2.24dB	-0.59dB	1.53dB	2.52dB
ITW	2.25dB	NA	-15.7dB	NA

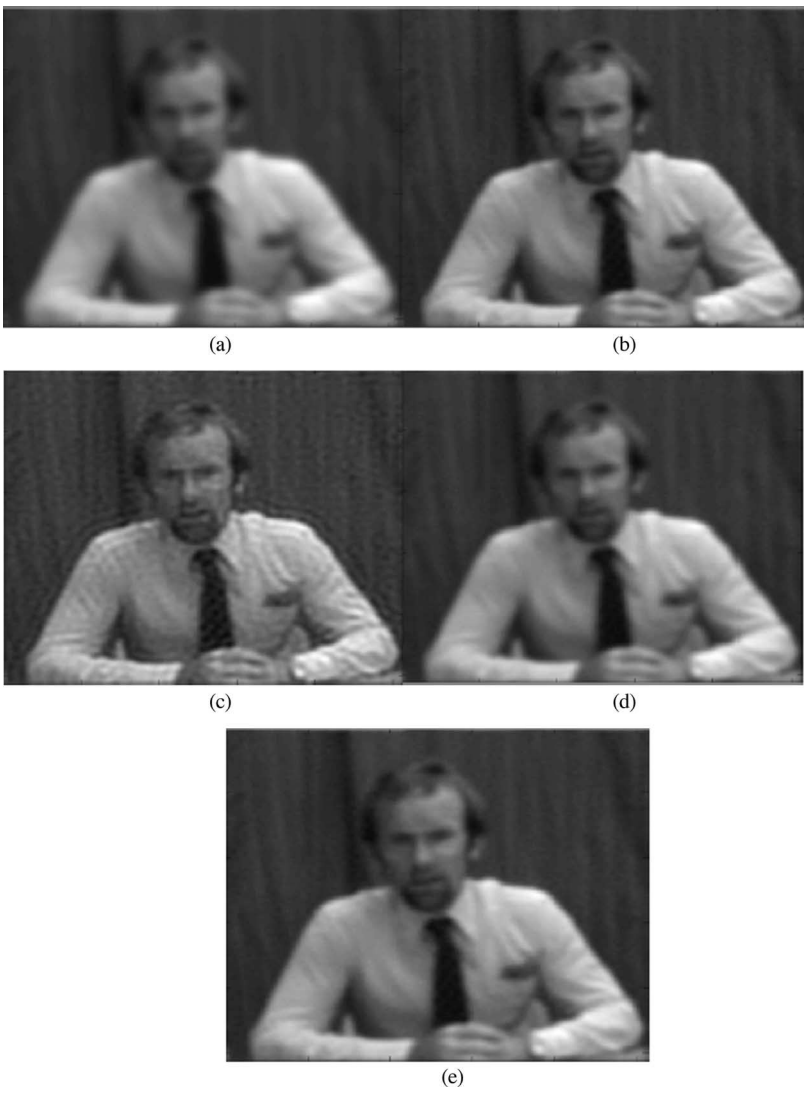

Fig. 3. Images from Table II case 1. (a) Degraded. (b) ITW, $\mathrm{ISNR}_f = 2.25$ dB. (c) VAR1, $\mathrm{ISNR}_f = 3.18$ dB (d) VAR2, $\mathrm{ISNR}_f = 1.8$ dB (e) VAR3, $\mathrm{ISNR}_f = 2.24$ dB.

chain. Thus, clearly, this is an area where more research is required in order to implement efficient strategies to evaluate the tightness of this bound. Furthermore, research on methods to optimize this bound is also necessary. In spite of this, the proposed methodology is quite general, and it can be used with other Bayesian models for this and other imaging problems. We plan in the very near future to apply the variational methodology to the BID problem with more sophisticated prior models that capture salient properties of the image and the PSF such as dc gain, nonstationarity, positivity, and spatial support.

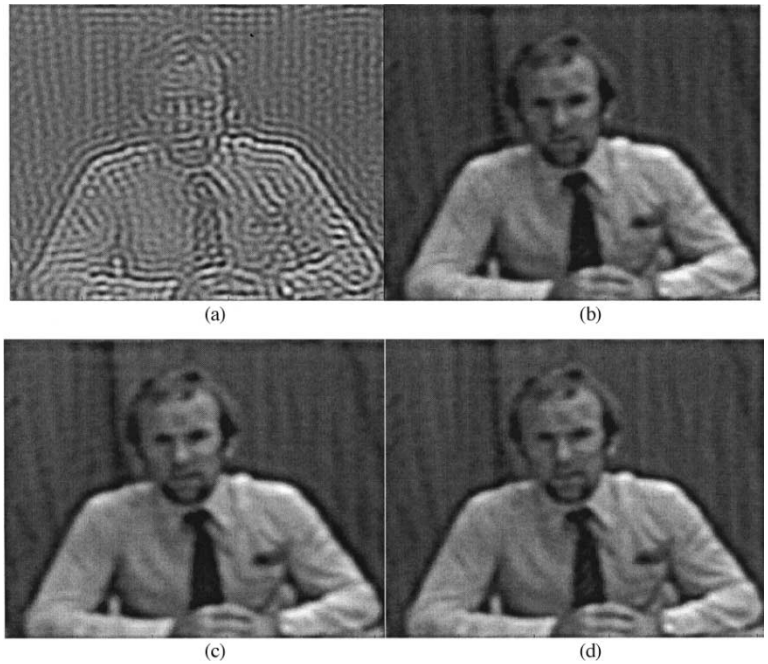

Fig. 4. Restored images from Table II case 2. (a) ITW, $\text{ISNR}_f = -15.7$ dB. (b) VAR1 $\text{ISNR}_f = 1.63$ dB. (c) VAR2, $\text{ISNR}_f = 1.59$ dB. (d) VAR3, $\text{ISNR}_f = 1.56$ dB.

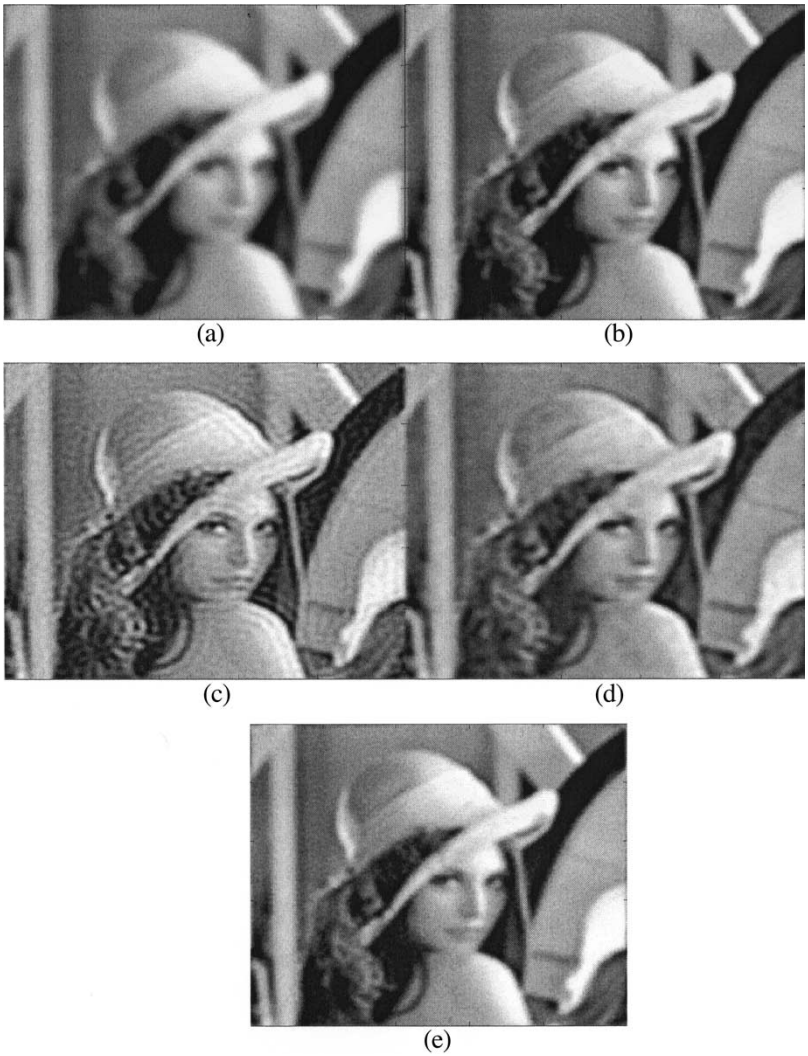

Fig. 5. Images from Table III case 1. (a) Degraded. (b) ITW, $\text{ISNR}_f = 2.73$ dB. (c) VAR1, $\text{ISNR}_f = 3.94$ dB. (d) VAR2, $\text{ISNR}_f = 2.37$ dB. (e) VAR3, $\text{ISNR}_f = 2.68$ dB.

TABLE III
FINAL ISNRS OF ESTIMATED IMAGES AND PSF FOR THE EXPERIMENTS WITH THE "LENA" IMAGE

	Case (1) Generating PSF $\sigma_X^2 = 20, \sigma_Y^2 = 20$ Initialization $\sigma_X^2 = 12, \sigma_Y^2 = 12$		Case (2) Generating PSF $\sigma_X^2 = 20, \sigma_Y^2 = 20$ Initialization $\sigma_X^2 = 40, \sigma_Y^2 = 40$	
	$ISNR_f$	$ISNR_h$	$ISNR_f$	$ISNR_h$
VAR1	3.94dB	7.23dB	3.37dB	4.81dB
VAR2	2.37dB	-4.87dB	3.14dB	2.87dB
VAR3	2.68dB	-1.28dB	2.43dB	2.69dB
ITW	2.73dB	NA	-18.01dB	NA

APPENDIX A
COMPUTATION OF THE VARIATIONAL BOUND $\mathbf{F}(q,\theta)$

From (1), we have that

$$\mathbf{F}(q,\theta) = E_q\left(\log p(x,s)\right) + H(q) \tag{A.1}$$

where $q(s) = q(f)q(h) = N(m_{f^q}, C_{f^q})N(m_{h^q}, C_{h^q})$, $p(x,s) = p(g,f,h) = p(g|f,h)\cdot p(f)\cdot p(h)$ (with $p(g|h,f) = N(h*f,\Sigma_w)$), and $H(q) = E_q(\log q(s))$ is the entropy of $q(s)$.

The implementation of the variational EM requires the computation of the Gaussian integrals appearing in (A.1). The integrand of the first part of (A.1) is given by

$$\begin{aligned}
&\log p(g,f,h)\\
&\quad = \log p(g|f,h) + \log p(f) + \log p(h)\\
&\quad = K_1 - \frac{1}{2}\\
&\qquad \times \Bigg\{\log|\Sigma_w| + \underbrace{(g-h*f)^t\Sigma_w^{-1}(g-h*f)}_{b1}\\
&\qquad\quad + \log|\Sigma_f| + \underbrace{(f-\mu_f)^t\Sigma_f^{-1}(f-\mu_f)}_{b2}\\
&\qquad\quad + \log|\Sigma_h| + \underbrace{(h-\mu_h)^t\Sigma_h^{-1}(h-\mu_h)}_{b3}\Bigg\}
\end{aligned} \tag{A.2}$$

where K_1 is a constant. The terms that are not constant in this integration with respect to the hidden variables are called $E_q(b_i)$ with $i = 1$, 2, and 3. These terms can be computed as

$$E_q(b_1) = E_q\left(\underbrace{g^t\Sigma_w^{-1}g}_{I_1} - \underbrace{(h*f)^t\Sigma_w^{-1}g - g^t\Sigma_w^{-1}(h*f)}_{I_2} + \underbrace{(h*f)^t\Sigma_w^{-1}(h*f)}_{I_3}\right). \tag{A.3}$$

These are the terms that must be integrated with respect to $q(h)$ and $q(f)$. The last one using the interchangeability of the convolution and its matrix vector representation is given by

$$\begin{aligned}
E_q(I_3) &= \int\int (h*f)^t\Sigma_w^{-1}(h*f)\cdot q(h)\cdot q(f)\cdot df\cdot dh\\
&= \int\left(\int f^t\cdot H^t\cdot\Sigma_w^{-1}\cdot H\cdot f\cdot q(f)\cdot df\right)\cdot q(h)\cdot dh\\
&= \int\left(\int \operatorname{trace}\left(H^t\cdot\Sigma_w^{-1}\cdot H\cdot f\cdot f^t\right)\cdot q(f)\cdot df\right)\\
&\qquad \cdot q(h)\cdot dh\\
&= \int \operatorname{trace}\left(H^t\cdot\Sigma_w^{-1}\cdot H\cdot\left(C_{f^q} + m_{f^q}m_{f^q}^t\right)\right)\\
&\qquad \cdot q(h)\cdot dh.
\end{aligned} \tag{A.4}$$

To compute this integral, we resort to the fact that these matrices are circulant and have common eigenvectors given by the discrete Fourier transform (DFT). Furthermore, for a circulant matrix C, it holds that $WCW^{-1} = \Lambda$, where Λ is the diagonal matrix containing the eigenvalues, and W is the DFT matrix. This decomposition can be also written as $(1/N)WCW^* = \Lambda$, where W^* denotes the conjugate since $W^{-1} = (1/N)W^*$; see, for example, [3]. Using these properties of circulant matrices we can write (A.5), shown at the bottom of the page.

In (A.5), $S_{f^q}(k)$, $S_{h^q}(k)$, and $\Lambda_w(k)$ are the eigenvalues of the covariance matrices C_{f^q}, C_{h^q}, and Σ_w. $M_{f^q}(k)$ and $M_{h^q}(k)$ are the DFTs of the vectors m_{f^q} and m_{h^q}, respectively. The remaining terms $E_q(I_1 + I_2)$ of (A.3) can be computed similarly; see (A.6), shown at the bottom of the next page.

As a result, for the term $E_q(b_1)$ we can write (A.7), shown at the bottom of the next page. The other terms $E_q(b_2)$ and $E_q(b_3)$ are similarly computed as (A.8), shown at the bottom of the next page, and (A.9), also shown at the bottom of the next page. The computation of $H(q)$ is easy because of the Gaussian choice for $q(f)$ and $q(h)$. In essence, we have to compute the sum of the

$$\begin{aligned}
E_q(I_3) &= \int\sum_{k=0}^{N-1}\left(\frac{|H(k)|^2}{\Lambda_w(k)}\left(S_{f^q}(k) + \frac{1}{N}|M_{f^q}(k)|^2\right)\right)\cdot q(h)\cdot dh\\
&= \sum_{k=0}^{N-1}\frac{\left(S_{f^q}(k) + \frac{1}{N}|M_{f^q}(k)|^2\right)\left(S_{h^q}(k) + \frac{1}{N}|M_{h^q}(k)|^2\right)}{\Lambda_w(k)}.
\end{aligned} \tag{A.5}$$

entropies $E_q(J_i)$ with $i = 1$ and 2 of two Gaussian pdfs, which is given by

$$\begin{aligned} H(q) &= -E_q\left\{\underbrace{\log q(f)}_{J_1}+\underbrace{\log q(h)}_{J_2}\right\} \\ &= E_q(J_1)+E_q(J_2) \\ &= -C+N+\frac{1}{2}\left(\log|C_{f^q}|+\log|C_{h^q}|\right). \end{aligned} \quad \text{(A.10)}$$

Replacing (A.7)–(A.10) into (A.2) results in (5) for $\mathbf{F}(q,\theta)$.

Appendix B
Maximization of $\mathbf{F}(q,\theta)$

We wish to maximize $\mathbf{F}(q,\theta)$ with respect to parameters θ_q and θ, where θ_q are the parameters that define $q(\cdot)$. Since we are not bounded by the EM framework that contains E and M steps, we can do this optimization any way we wish. However, in analogy to the EM framework, we have adopted the following two steps that we call the E and M steps:

E-step (update of θ_q):

$$\theta_q^{t+1} = \arg\max_{\theta^q}\left\{F(\theta_q,\theta^t)\right\}.$$

M-step (update of θ):

$$\theta^{t+1} = \arg\max_{\theta}\left\{F\left(\theta_q^{t+1},\theta\right)\right\}.$$

In the M-step, in order to find the parameters θ that maximize F, we need to find the derivatives $(\partial F(q,\theta))/(\partial\theta)$ and set them to zero. From (5), we have

$$\begin{aligned} \frac{\partial F(q,\theta)}{\partial\Lambda_w(k)} = 0 \Rightarrow & \frac{1}{\Lambda_w(k)} - \frac{A_1(k)+A_2(k)}{(\Lambda_w(k))^2} = 0 \\ \Rightarrow & \Lambda_w(k) = A_1(k)+A_2(k) \\ & \text{for } k = 0,1\ldots N-1. \end{aligned}$$

Similarly, we get $\Lambda_f(k) = B(k)$ and $\Lambda_h(k) = C(k)$ for $k = 0,1\ldots N-1$.

$$\frac{\partial F(q,\theta)}{\partial M_f(k)} = 0 \Rightarrow M_f(k) = M_{f^q}(k) \text{ and}$$

$$\frac{\partial F(q,\theta)}{\partial M_h(l)} = 0 \Rightarrow M_h(l) = M_{h^q}(l), \quad \text{for } k = 0,1\ldots N-1.$$

Thus, we can compute the unknown parameters $\theta^{(t+1)}$ as

$$M_f^{(t+1)}(k) = M_{f^q}^{(t+1)}(k) \quad \text{(B.1)}$$

$$M_h^{(t+1)}(k) = M_{h^q}^{(t+1)}(k) \quad \text{(B.2)}$$

$$\begin{aligned} \Lambda_f^{(t+1)}(k) = & \left(S_{f^q}^{(t+1)}(k)+\frac{1}{N}\left|M_{f^q}^{(t+1)}(k)\right|^2\right) \\ & + \frac{1}{N}\left(\left|M_f^{(t+1)}(k)\right|^2\right. \\ & \left. -2\mathrm{Re}\left\{\left(M_f^{(t+1)}(k)\right)^* M_{f^q}^{(t+1)}(k)\right\}\right) \\ = & S_{f^q}^{(t+1)}(k). \end{aligned} \quad \text{(B.3)}$$

For similar reasons

$$\Lambda_h^{(t+1)}(k) = S_{h^q}^{(t+1)}(k) \quad \text{(B.4)}$$

$$\begin{aligned} \Lambda_w^{(t+1)}(k) = & \frac{1}{N}\left(|G(k)|^2\right. \\ & \left. -2\mathrm{Re}\left\{M_{f^q}^{(t+1)}(k)M_{h^q}^{(t+1)}(k)G^*(k)\right\}\right) \\ & + \left(S_{f^q}^{(t+1)}(k)+\frac{1}{N}\left|M_{f^q}^{(t+1)}(k)\right|^2\right) \\ & \times\left(S_{h^q}^{(t+1)}(k)+\frac{1}{N}\left|M_{h^q}^{(t+1)}(k)\right|^2\right) \end{aligned} \quad \text{(B.5)}$$

for $k = 0,1\ldots N-1$.

In our experiments, we have used an SAR prior [12] for the image model; thus, $p(f) \propto (\alpha)^{((N-1)/2)}\exp(-(1/2)\|Qf\|^2)$, $p(h) = N(m_h,\beta I)$, and $p(n) = N(0,\sigma^2 I)$, where Q is the

$$E_q(I_2+I_1) = \frac{1}{N}\sum_{k=0}^{N-1}\left\{\frac{|G(k)|^2+M_{f^q}(k)M_{h^q}(k)G^*(k)+M_{f^q}^*(k)M_{h^q}^*(k)G(k)}{\Lambda_w(k)}\right\} \quad \text{(A.6)}$$

$$\begin{aligned} E_q(b_1) = \sum_{k=0}^{N-1}\Bigg\{ & \frac{\frac{1}{N}\left(|G(k)|^2+M_{f^q}(k)M_{h^q}(k)G^*(k)+M_{f^q}^*(k)M_{h^q}^*(k)G(k)\right)}{\Lambda_w(k)} \\ & + \frac{N\left(S_{f^q}(k)+\frac{1}{N}|M_{f^q}(k)|^2\right)\left(S_{h^q}(k)+\frac{1}{N}|M_{h^q}(k)|^2\right)}{\Lambda_w(k)}\Bigg\} \end{aligned} \quad \text{(A.7)}$$

$$E_q(b_2) = \sum_{k=0}^{N-1}\frac{\left(S_{f^q}(k)+\frac{1}{N}|M_{f^q}(k)|^2\right)+\frac{1}{N}\left(M_f^*(k)M_{f^q}(k)+M_f(k)M_{f^q}^*(k)+|M_f(k)|^2\right)}{\Lambda_f(k)} \quad \text{(A.8)}$$

$$E_q(b_3) = \sum_{k=0}^{N-1}\frac{\left(S_{h^q}(k)+\frac{1}{N}|M_{h^q}(k)|^2\right)+\frac{1}{N}\left(M_h^*(k)M_{h^q}(k)+M_h(k)M_{h^q}^*(k)+|M_h(k)|^2\right)}{\Lambda_h(k)} \quad \text{(A.9)}$$

$$\mathrm{Re}\left(M_{f^q}^{(t+1)}(k)\right) = \frac{\mathrm{Re}\left(M_{h^q}^{(t)}(k)\right)\mathrm{Re}\left(G(k)\right) + \mathrm{Im}\left(M_{h^q}^{(t)}(k)\right)\mathrm{Im}\left(G(k)\right)}{\alpha\sigma^2\left|Q(k)\right|^2 + NS_{h^q}^{(t)}(k) + \left|M_{h^q}^{(t)}(k)\right|^2} \tag{B.9}$$

$$\mathrm{Im}\left(M_{f^q}^{(t+1)}(k)\right) = \frac{-\mathrm{Im}\left(M_{h^q}^{(t)}(k)\right)\mathrm{Re}\left(G(k)\right) + \mathrm{Re}\left(M_{h^q}^{(t)}(k)\right)\mathrm{Im}\left(G(k)\right)}{\alpha\sigma^2\left|Q(k)\right|^2 + NS_{h^q}^{(t)}(k) + \left|M_{h^q}^{(t)}(k)\right|^2} \tag{B.10}$$

$$S_{f^q}^{(t+1)}(k) = \frac{\sigma^2}{\alpha\sigma^2\left|Q(k)\right|^2 + NS_{h^q}^{(t)}(k) + \left|M_{h^q}^{(t)}(k)\right|^2} \tag{B.11}$$

$$\mathrm{Re}\left(M_{h^q}^{(t+1)}(k)\right) = \frac{\beta\left[\mathrm{Re}\left(M_{f^q}^{(t+1)}(k)\right)\mathrm{Re}\left(G(k)\right) + \mathrm{Im}\left(M_{f^q}^{(t+1)}(k)\right)\mathrm{Im}\left(G(k)\right)\right] + \sigma^2\mathrm{Re}\left(M_h(k)\right)}{\sigma^2 + \beta NS_{f^q}^{(t+1)}(k) + \beta\left|M_{f^q}^{(t+1)}(k)\right|^2} \tag{B.12}$$

$$\mathrm{Im}\left(M_{h^q}^{(t+1)}(k)\right) = \frac{\beta\left[\mathrm{Re}\left(M_{f^q}^{(t+1)}(k)\right)\mathrm{Im}\left(G(k)\right) - \mathrm{Im}\left(M_{f^q}^{(t+1)}(k)\right)\mathrm{Re}\left(G(k)\right)\right] + \sigma^2\mathrm{Im}\left(M_h(k)\right)}{\sigma^2 + \beta NS_{f^q}^{(t+1)}(k) + \beta\left|M_{f^q}^{(t+1)}(k)\right|^2} \tag{B.13}$$

$$S_{h^q}^{(t+1)}(k) = \frac{\beta\sigma^2}{\sigma^2 + \beta NS_{f^q}^{(t+1)}(k) + \beta\left|M_{f^q}^{(t+1)}(k)\right|^2} \tag{B.14}$$

circulant matrix that represents the convolution with the Laplacian operator. Therefore, the unknown parameter vector θ to be estimated contains the parameters α, β, and σ^2. Because of the circulant properties, it holds that $\Lambda_f(k) = (1/(\alpha|Q(k)|^2))$, $\Lambda_h(k) = \beta$, and $\Lambda_w(k) = \sigma^2$. Based on these assumptions, the general equations (B.1)–(B.5) for the updates at the M-step take the specific following form:

M-step

$$\alpha^{(t+1)} = \left[\frac{1}{N-1}\sum_{k=0}^{N-1}\left(S_{f^q}^{(t+1)}(k) + \frac{1}{N}\left|M_{f^q}^{(t+1)}(k)\right|^2\right) \times\left|Q(k)\right|^2\right]^{-1} \tag{B.6}$$

$$\beta^{(t+1)} = \frac{1}{N}\sum_{k=0}^{N-1}\left(S_{f^q}^{(t+1)}(k) + \frac{1}{N}\left|M_{f^q}^{(t+1)}(k)\right|^2\right) \tag{B.7}$$

$$\begin{aligned}(\sigma^2)^{(t+1)} = \frac{1}{N}\Bigg[&\sum_{k=0}^{N-1}\Big(\left|G(k)\right|^2 \\ &-2\mathrm{Re}\Big\{M_{f^q}^{(t+1)}(k)M_{h^q}^{(t+1)}(k)G^*(k)\Big\}\Big) \\ &+N\sum_{k=0}^{N-1}\left(S_{f^q}^{(t+1)}(k) + \frac{1}{N}\left|M_{f^q}^{(t+1)}(k)\right|^2\right) \\ &\times\left(S_{h^q}^{(t+1)}(k) + \frac{1}{N}\left|M_{h^q}^{(t+1)}(k)\right|^2\right)\Bigg].\end{aligned} \tag{B.8}$$

For the VAR3 approach, the updates for α and β remain the same. However, to obtain the updates for the noise variance, we apply the same rules that were previously used to obtain the variational bounds $\mathbf{F}_f(q,\theta)$ and $\mathbf{F}_h(q,\theta)$ from the bound $\mathbf{F}(q,\theta)$ in (5).

For the VAR1 approach, the update equations for the parameters θ_q of $q(s)$ (which are complex in the DFT domain) are easily obtained by equating the corresponding gradient of $\mathbf{F}(q,\theta)$ to zero. This yields the following update equations for $k = 0,\ldots,N-1$:

E-step (VAR1 approach): We have (B.9)–(B.14), shown at the top of the page.

ACKNOWLEDGMENT

The authors acknowledge Dr. N. Vlassis, University of Amsterdam, Amsterdam, The Netherlands, for his insightful comments on the variational methodology.

REFERENCES

[1] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Mag.*, vol. 13, pp. 43–64, May 1996.

[2] R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Identification and restoration of noisy blurred images using the expectation-maximization algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1180–1191, July 1990.

[3] A. K. Katsaggelos and K. T. Lay, "Maximum likelihood identification and restoration of images using the expectation-maximization algorithm," in *Digital Image Restoration*, A. K. Katsaggelos, Ed. New York: Springer-Verlag, 1991, ch. 6.

[4] ——, "Maximum likelihood blur identification and image restoration using the EM algorithm," *IEEE Trans. Signal Processing*, vol. 39, pp. 729–733, Mar. 1991.

[5] Y. Yang, N. P. Galatsanos, and H. Stark, "Projection based blind deconvolution," *J. Opt. Soc. Amer.—A*, vol. 11, no. 9, pp. 2401–2409, Sept. 1994.

[6] Y. L. You and M. Kaveh, "A regularization approach to joint blur identification and image restoration," *IEEE Trans. Image Processing*, vol. 5, pp. 416–428, Mar. 1996.

[7] ——, "Blind image restoration by anisotropic regularization," *IEEE Trans. Image Processing*, vol. 8, pp. 396–407, Mar. 1999.

[8] V. N. Mesarovic, N. P. Galatsanos, and M. N. Wernick, "Iterative LMMSE restoration of partially-known blurs," *J. Opt. Soc. Amer.—A*, vol. 17, pp. 711–723, Apr. 2000.

[9] N. P. Galatsanos, V. N. Mesarovic, R. M. Molina, and A. K. Katsaggelos, "Hierarchical Bayesian image restoration from partially-known blurs," *IEEE Trans. Image Processing*, vol. 9, pp. 1784–1797, Oct. 2000.

[10] N. P. Galatsanos, V. N. Mesarovic, R. M. Molina, J. Mateos, and A. K. Katsaggelos, "Hyper-parameter estimation using gamma hyper-priors in image restoration from partially-known blurs," *Opt. Eng.*, vol. 41, no. 8, pp. 1845–1854, Aug. 2002.

[11] A. D. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the E-M algorithm," *J. R. Stat. Soc.*, vol. B39, pp. 1–37, 1977.

[12] C. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, Second ed. New York: Springer Verlag, June 2001.
[13] B. Carlin and T. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, Second ed. Boca Raton, FL: CRC, 2000.
[14] *Bayesian Inference for Inverse Problems*, vol. 3459, Proceedings of SPIE—The International Society for Optical Engineering, A. M. Djafari, Ed., July 1998.
[15] K. H. Yap, L. Guan, and W. Liu, "A recursive soft decision approach to blind image deconvolution," *IEEE Trans. Signal Processing*, vol. 51, pp. 515–526, Feb. 2003.
[16] R. M. Neal and G. E. Hinton, "A view of the E-M algorithm that justifies incremental, sparse and other variants," in *Learning in Graphical Models*, M. I. Jordan, Ed. Cambridge, MA: MIT Press, 1998, pp. 355–368.
[17] 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*, M. I. Jordan, Ed. Cambridge, MA: MIT Press, 1998, pp. 105–162.
[18] R. Molina and B. D. Ripley, "Using spatial models as priors in astronomical images analysis," *J. Appl. Stat.*, vol. 16, pp. 193–206, 1989.
[19] T. S. Jaakkola, "Variational methods for inference and learning in graphical models," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, MA, 1997.
[20] M. Cassidy and W. Penny, "Bayesian nonstationary autoregressive models for biomedical signal analysis," *IEEE Trans. Biomed. Eng.*, vol. 49, pp. 1142–1152, Oct. 2002.
[21] 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.
[22] Z. Ghahramani and M. J. Beal, "Variational inference for Bayesian mixtures of factor analyzers," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2000, vol. 12, pp. 449–455.
[23] M. J. Beal, "Variational algorithms for approximate Bayesian inference," PhD. dissertation, Gatsby Computational Neuroscience Unit, Univ. College London, London, U.K., 2003.
[24] K. T. Lay and A. K. Katsaggelos, "Image identification and restoration based on the expectation-maximization algorithm," *Opt. Eng.*, vol. 29, pp. 436–445, May 1990.

Aristidis C. Likas (S'91–M'96–SM'03) received the Diploma degree in electrical engineering and the Ph.D. degree in electrical and computer engineering, both from the National Technical University of Athens, Athens, Greece.

Since 1996, he has been with the Department of Computer Science, University of Ioannina, Ioannina, Greece, where he is currently an Assistant Professor. His research interests include neural networks, machine learning, statistical signal processing, and bioinformatics.

Nikolas P. Galatsanos (SM'95) received the Diploma of electrical engineering from the National Technical University of Athens, Athens, Greece, in 1982 and the M.S.E.E. and Ph.D. degrees from the Electrical and Computer Engineering Department, University of Wisconsin, Madison, in 1984 and 1989, respectively.

He was on the faculty of the Electrical and Computer Engineering Department, Illinois Institute of Technology, Chicago, from 1989 to 2002. Presently, he is a Professor with the Department of Computer Science, University of Ioannina, Ioninna, Greece. His research interests center on image processing and machine learning problems for medical imaging, bioinformatics, and visual communications applications. He has coedited a book, with A. K. Katsaggelos, entitled *Image Recovery Techniques for Image and Video Compression and Transmission* (Boston, MA: Kluwer, Oct. 1998).

Dr. Galatsanos has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and the IEEE SIGNAL PROCESSING MAGAZINE.